Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Estimation and Application of the Tail Index

Author: **Markéta Pokorná**

Supervisor: **PhDr. Boril Šopov, MSc., LL.M.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature.

The author hereby declares that all the sources and literature used have been properly cited.

The author hereby declares that this thesis was not used to receive a degree from any other institution.

Prague, May 10, 2016

Signature

Acknowledgements

I wish to express my sincere thanks to my academic supervisor PhDr. Boril Šopov, MSc., LL.M. who provided highly appreciated guidance, helpful comments and suggestions.

I am especially grateful to all who supported me during my studies and during the work on this thesis.

Bibliography reference

POKORNÁ, Markéta. *Estimation and Application of Tail Index.* Praha, 2016. 53 p. Bachelor thesis. Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: PhDr. Boril Šopov, MSc., LL.M.

Extent of the thesis: 91,323 characters (spaces included)

Abstract

Examining the nature of extreme values plays an important role in financial risk management. This thesis investigates tail behaviour of distribution of returns using the framework of univariate Extreme Value Theory. The empirical research was conducted on the S&P 500 index and its seven constituents. The goal of this thesis was to use the Hill method to estimate the tail index of the series which characterizes the tail behaviour, especially the speed of the tail decay. To select the tail threshold several graphical methods were performed as they represent empirical measures of model stability. Classical Hill plots as well as alternative Hill plots and smoothing procedure were presented. The threshold choice based on stable regions in the graphs was found to be highly subjective. Hill method modified by Huisman was used instead and the results confirmed that the classical Hill method yields estimates which overestimate the tail thickness. All the examined series were found to have heavy tails with polynomial tail decay. This thesis stressed the need to model the left and the right tail separately as both extreme losses and profits are important depending on whether an investor takes a long or a short position on portfolio. Finally, the tail index was used to demonstrate the need to compute the expected losses for certain quantiles instead of simply the minimum losses as expressed by Value at Risk.

JEL Classification C14, C58, G32
Keywords Extreme Value Theory, Tail Index, tail behaviour, risk

Author's e-mail marketa.pokorna.jh@gmail.com
Supervisor's e-mail boril.sopov@gmail.com

Abstrakt

Zkoumání extrémních hodnot je nezbytnou součástí finančního risk managementu. Tato práce prověřuje vlastnosti chvostů pomocí jednorozměrného rámce Teorie extrémních hodnot. Empirický výzkum byl proveden na indexu S&P 500 a sedmi jeho konstituentech. Cílem této práce bylo odhadnout pomocí Hillovy metody chvostový index, který charakterizuje chování chvostů a určuje rychlost jejich svažování. K výběru chvostového prahu bylo využito několika grafických metod, jelikož jsou empirickými ukazateli stability modelu. Byl použit klasický Hillův graf a také alternativní graf s vyhlazovací technikou. Výběr prahu založen na stabilních plochách těchto grafů byl shledán vysoce subjektivním. Byla tedy použita Hillova metoda pozměněna Huismanem a výsledky potvrdily, že klasická Hillova metoda přináší odhady, které nadhodnocují tloušťku chvostů. Bylo zjištěno, že všechny zkoumané akcie mají těžké chvosty s polynomickým svažováním. V práci bylo zdůrazněno, že je potřeba modelovat zvlášť levý a pravý chvost, jelikož jsou důležité extrémní ztráty i zisky v závislosti na tom, zda investor drží krátkou či dlouhou pozici. Chvostový index byl také použit k ilustraci potřeby počítat očekávané ztráty místo pouhých minimálních možných ztrát, které vyjadřuje Value at Risk.

Klasifikace JEL	C14, C58, G32
Klíčová slova	Teorie extrémních hodnot, chvostový index, vlastnosti chvostů, risk

E-mail autora	marketa.pokorna.jh@gmail.com
E-mail vedoucího práce	boril.sopov@gmail.com

Contents

List of Tables

List of Figures

Acronyms

pdf probability distribution function

df distribution function

i.i.d. independent, identically distributed

sd standard deviation

CLT Central Limit Theorem

ML maximum likelihood

OLS ordinary least squares

WLS weighted least squares

FGLS feasible generalized least squares

GARCH Generalized Autoregressive Conditional Heteroscedasticity

ARCH Autoregressive Conditional Heteroscedasticity

ACF autocorrelation function

EVT Extreme Value Theory

GEV Generalized Extreme Value

GPD Generalized Pareto Distribution

VaR Value at Risk

ES Expected Shortfall

BMM Block Maxima Method

PoT Peaks over Threshold

MDA Maximum domain of attraction

CAPM Capital Asset Pricing Model

Chapter 1

Introduction

The highly volatile period in the markets that recurred in the last decade has had a major effect on the importance of risk management for all institutions, firms and individuals. Not only market crashes have lead to an adoption of more cautious approach towards asset management. An important part was also played by well-known losses suffered due to derivative mismanagement, such as the case of Procter & Gamble, or the losses experienced through risky investment strategy as in Orange County. However, more effort is being made to hedge against types of risks which cannot be influenced or prevented by diversification. They arise unexpectedly with low probability but with consequences that can be of an unprecedented severity. That is the reason why their predictability is generally on the front burner. These extreme risk events include natural disasters such as 100-year floods and major earthquakes, as well as substantial financial losses. Estimating the probability of extreme events poses unique difficulties not only because of the scarcity of the data due to the rare nature of these events. It is also necessary to shed the assumption of normality, which does not hold. These reasons gave rise to the Extreme Value Theory (EVT) framework.

Extreme Value Theory is a concept proposed to deal with modelling the effect of extreme events. It is based on sound statistical methods which are designed to estimate probabilities of extreme events. Such methods use limited range of data of such events that occurred in the past and are suitable for predicting events even more extreme than those previously observed. Therefore it gives the necessary guidance to evaluate the possibility of risk in improbable events. This methodology for statistical modelling provides researchers with a useful tool in the fields of risk management, insurance, finance and even biology and hydrology. Interestingly, the latter were the ones that initiated research in the field of Extreme Value Theory.

Within the framework of this theory, the purpose is to estimate distribution parameters so as to be able to make forecasts about the nature of risk in future periods. However, no assumption is made about all observations from the distribution. The focus is made directly on the tails without paying attention to the center of the distribution. This ensures that the modelled tail behaviour expresses as much information as possible about the disposition of the extreme values. Therefore the tail model is not directly influenced by observations which are of a regular occurrence. Compared to the classical risk measure, Value at Risk, the EVT does not assume that a certain distribution function holds globally and its interests lie in the whole tails and not just at the point where the tails begin. The idea is quite aptly expressed by DuMouchel (1983). He states that the natural way to model the tail behaviour of data is to let the tails "speak for themselves". Therefore, this thesis is focused on estimation of the Tail Index, a parameter from which the speed of tail decay can be directly inferred. This tail decay gives then a useful measure of the tail thickness, as the faster the decay, the lighter the tail.

While in the perfect state the knowledge of occurrence and incidence of extreme events would be desirable, applying the Extreme Value Theory represents many challenges. Among these, the most pertinent ones are the choice of the method of how to estimate parameters and the decision on the starting point of the tail. In the empirical part of this thesis real financial data will be examined. As hypothesized, the data would be far from the Normal distribution assumption. Financial stock market index S&P 500 will be subject to the analysis as well as some individual stocks as they might reflect rare events unequally. When choosing the tail cut-off point, several graphical methods will be employed and it shall be discussed, whether they can serve to decide where the tail begins and what role subjectivity plays therein. Another subject of interest are the differences between the left and the right tail. Faster decay of one tail tells us about the imbalance of the positive and negative extreme values. Therefore, both the minima (losses) and maxima (profits) come under scrutiny. The final hypothesis is that the tail behaviour of financial time series does change over time. For this reason the sample will be divided into shorter time spans to see if there was a period with fatter tails.

The thesis is organized as follows: Chapter 2 gives an overview of important literature contributing to both Extreme Value Theory and Tail Index estimation. In Chapter 3, attention is paid to the theoretical framework of the EVT. The tail index estimation method is described as well as graphical threshold selection techniques. Then the section about the tail index application gives guidelines about what can be inferred from the tail index values. The model is applied in practice in Chapter 4 which includes the initial data analysis and consequent tail index estimation. Finally, Chapter 5 concludes.

Chapter 2

Literature Review

The first foundations of the Extreme Value Theory were given by statisticians Ronald Fisher and his student Leonard H.C. Tippett (Fisher and Tippett, 1928). Their theorem described extreme order statistics' convergence in distribution. However, the rigorous proof was given a few years later by a Soviet mathematician Boris V. Gnedenko (Gnedenko, 1943) and therefore the first theorem of EVT is referred to as Fisher-Tipeett, Gnedenko. A further step was made by dividing the limit distribution into 3 families of generalised extreme value distribution. First type was described by Emil J. Gumbel (Gumbel, 1958), a German mathematician who devoted major part of his work to examine climate and hydrology. Further type of extreme value distributions was recognized by Fisher and Tippett based on work of Maurice Fréchet (Fréchet, 1927) and the last one was identified by Fréchet and described later in detail by a Swedish mathematician Waloddi Weibull (Weibull, 1951).

The first extreme value theory applications included environmental issues and mainly used values of annual maxima as the extremes to be modelled. The concern was for example to estimate the probability that the maximum river flow level next year will exceed a certain level or even all previous levels. Later, however, the focus has shifted from climate issues to insurance business and annual maxima were dominated by a new method called Peaks Over Threshold. Along with this new approach the EVT foundations had to be strengthened. The influential research was accomplished by Balkema and de Haan (1974) and Pickands III (1975). The former examined the limit distribution types of residual life time considering a bulb, the latter presented a method for inference about the upper tail of a probability distribution function. Thus it became common to consider extremes as values exceeding a certain threshold and Pickands-Balkema-de Haan is considered to be the second theorem of EVT.

With this extended theory some more applications in biology areas still appeared. A useful application has been made in the research of atmosphere pollution. Smith (1989) conducted a detailed analysis of ground-level of ozone using data form Houston, Texas. Another study by De Haan (1994) made use of EVT in studying the rises of sea level. His theory gave direction to determine a safe height of sea dykes in the Netherlands so that a flood would appear only once in 10,000 years. Because of the fact that the environmental disasters are closely connected with insurance data, the focus has moved on the insurance business as well. Rootzén and Tajvidi (1997) examined wind storm extremes using meteorological information and at the same time analysing large claims in insurance.

In the last few decades the interest in extreme values and the risk involved has shifted into financial domain. Risk management of financial portfolios has proved to be of high importance due to several events which revealed the unexpectedness and severity of extreme price movements in the markets. One of the first events triggering consciousness was the big market crash in October 1987 which brought financial crisis and contributed to a consideration of appropriate financial measures. Consequently a set of minimum capital requirements for banks was established by Basel I (and later by Basel II) set of rules. The following history of recurring crises indicated that excluding extreme events from models could bring much higher losses experienced during next crisis and even a collapse of institutions. Measures such as Value at Risk (VaR) were implemented at several financial institutions. Further shocks followed with Brazilian stock exchange crash in 1989, with 1997 Asian financial crisis, 2002 stock market downturn and the financial crisis of 2007-2008 even caused reconstruction of Basel I rules. VaR technique, which measures financial risk while considering amount of potential loss, probability of that amount of loss and time horizon, has been altered many times since. Jorion (1996) refers to the need to control financial risks better and shows how to improve the accuracy of VaR estimates.

Contrary to the VaR, EVT emphasizes maxima and minima and makes no assumptions about the original distribution thus modelling the tails of the distribution. During the 1990s the VaR approach has been extended by the extreme value theory in order not to examine only the central observations of the distribution. The basic idea was to turn away from normal distribution approximation. Among others, Longin (2000) presented an application of EVT to compute the VaR, where he utilises EVT to create an approach that covers normal market conditions as well as financial crises. The improved VaR approach was also applied by Ho et al. (2000) who modelled the tails of return distributions of Asian financial markets counting in the period of previous Asian financial turmoil. He found that the VaR estimates generated by the EVT differed substantially from those of traditional methods thus taking into account the fat tails. Therefore the mixed approach is widely used as the standard VaR method falls short on fitting the maxima and minima because of their scarcity.

Further contributions of the EVT include a better understanding of data by exploring the distribution of the tails in detail. An important question, however, rises when segregating the extreme values from the rest of the observations. A decision rule has to be employed to reasonably state where the tail begins. Another choice has to be made to decide on a method of estimating the parameters. An overview of approaches dealing with these options will be given considering the targeted area of this work.

This thesis is focused on the shape parameter of the limiting distribution, because the parameter demonstrates the behaviour of the limiting distribution uniquely and helps us understand the decay of the tails. The first general approach to inference about tails was proposed by a statistician Bruce M. Hill (Hill, 1975) and his Hill's estimator of the tail index (reciprocal of shape parameter). He stressed the importance to draw inference about the behaviour of the tails. At the same time he suggested that it might be done without assuming that a certain distribution function in its parametric form holds globally. His method is based on order statistics and a non parametric approach is conditioned upon the values of order statistics which exceed a certain threshold. Then the parameters' estimates are obtained by conditional maximum likelihood estimation. An application of Hill's estimate has become established due to its straightforward interpretation and computational simplicity. Already in the original paper (Hill, 1975) it was clearly stated that the proposed methods depend on a subjective choice of the threshold or in other words on the number of extreme value statistics. By setting the threshold high, fewer extreme observations are employed and therefore cause high variance of the tail index estimator. On the contrary, setting the threshold too low can lead to a biased estimator. Thus the threshold choice matters significantly and basically creates the trade-off between high variance and bias.

Other simple estimators were constructed in the way that they depend on a certain number of upper order statistics. Pickands III (1975) estimates the distribution parameters by a simple percentile method reaching again an estimator based on extreme observations. Dekkers et al. (1989) builds on the Hill estimator to propose a moment estimator (referred to as the DEdH estimator). Although these estimators were proved to be consistent and asymptotically normal, no clear rule on their superiority was derived until the first comparison of tail indices (De Haan, 1998). Here the Hill, Pickands and DEdH estimators were compared in terms of the asymptotic mean square error. The results suggest that there is no estimator which simply outperforms the others since the outcome depends on the distribution parameters and for different situations we get different superior estimators.

The subsequent research in the area focused on refinements of the original estimators but also originated methods pursuing estimators with better properties. Considerable amount of research is dealing with threshold selection. This topic attracts attention because segregating a certain number of extremes is a necessary step which requires a proper rationale behind. Sometimes the threshold is arbitrarily set at a certain quantile (e.g. 95%). However, the number of extreme order statistics is an essential aspect in estimating the tail index. A thorough consideration is therefore needed because selecting the threshold too high could cause high variance of the estimator. On the other hand, setting the threshold too low causes the estimator to be biased as observations which do not belong to extremes might also be considered. The original Hill's approach suggested looking for a stable region of estimates in Hill plot (a graph depicting estimates depending on the number of order statistics used). However, this basic plot is often not sufficiently informative (see Embrechts et al., 1997, p. 343) and therefore other techniques are required.

Resnick and Stărică (1997) propose a different method called the alternative Hill plot based on logarithmic horizontal scale. They find this plot more useful in terms of how much information it demonstrates. Here the parts of graph corresponding to high or low number of extreme order statistics are rescaled thus enabling more obvious interpretation of the graph. Moreover, the authors suggest additional smoothing procedure through averaging the Hill estimator so that the variance of the estimator is reduced and the volatility of the plot mitigated. Nevertheless, using both original and alternative Hill plot and their comparison is advised. Even more refined method is proposed by Drees et al. (2000) who built a measure based on the occupation time of the plot around the true value of the estimator. Their results demonstrate that the alternative Hill plot is superior to the original one unless the data comes from Pareto distribution.

Other decisions on threshold selection can be reached by non-graphical methods. One of the most commonly used is determining the cut-off level such that the asymptotic mean squared error of the estimator is minimized. However, this approach is based on asymptotic behaviour and provides little guidance about finite samples. More advanced methods are based on choosing the cut-off level endogenously (therefore it complies with the available data) and they do so by subsampling and bootstrap techniques. These were described by Hall (1990) and further developed by Danielsson et al. (2001) who solve the problem of choosing the threshold in subsamples by introducing a control variate. This forms an accurate approach which entails arbitrary choice of only the number of Monte Carlo replications.

One of the other issues connected with the tail index estimator is correcting for its bias. The main idea is to introduce the second order framework for the tail form and estimate both first and second order parameters. Several new estimators have been proposed where the bias is reduced and at the same time the variance not inflated or even decreased as well (Gomes et al., 2008). An alternative method was introduced by Huisman et al. (2001) who reduce the small sample bias by an estimator obtained as a weighted average of Hill estimators with weights derived from least squares estimation. As well as bias reduction, dealing with the estimator's rate of convergence is a concern of desired good performance. Slow rate of convergence represents a big shortcoming of the estimator because it would lack accuracy in smaller samples. The general use of most simple estimators is penalized by a slow rate of convergence. Hall (1982) describes an estimator with an algebraic rather than logarithmic convergence rate. This, however, comes at the expense of additional assumptions about the underlying distribution. As bias and rate of convergence are concerned, the best statistical properties of the estimator are desired. Despite the awareness of this fact, this thesis focuses on simple estimators with general use because of their simpler practical implementation.

If an application of extreme value theory shall be considered, one of the strongest assumptions to face is that the data must be independent and identically distributed. Nevertheless, if extremes of financial time series are to be modelled, this assumption might not be met. If the dependence structure is not taken into account, incorrect estimates might be obtained, which might lead to unexpected losses or on the contrary to overly conservative behaviour. Therefore an explicit procedure was devised to model the dependence by Generalized Autoregressive Conditional Heteroscedasticity (GARCH) family models and then apply the EVT for the innovations which are believed to be independent and identically distributed. This method is called Conditional Extreme Value Theory and as shown by McNeil and Frey (2000) the conditional approach dominates the unconditional one and works well also for market risk measures such as Value at Risk or Expected Shortfall. According to McNeil and Frey (2000) this method is robust enough to obtain good estimates even in case of misspecification of the GARCH model up to a certain degree. If the dependence is not modelled, EVT can still be applied to financial time series. However, simplifying assumption of independence would have to be used and even though the estimates would be consistent the standard errors should be expected to be over-optimistically small (McNeil et al., 2005).

Apart from the original estimators and their fine-tuning, also other alternative estimators have been proposed. Peaks Over Random Threshold (PORT) is a methodology suggesting estimators based on a sample of excesses over a random threshold which enables to estimate high quantiles. Also other methods should be mentioned such as the geometric, Kernel, or QQ estimators. These are, however, outside of the scope of this thesis.

Chapter 3

Theoretical Framework and Methodology

3.1 Measuring risk

Financial risk management is concerned with forecasting situations which might effect the invested wealth and thus endanger individuals or institutions. Even though more advanced methods have been under scrutiny in the last three decades, it has always been indisputable that risk has to be quantified.

The variance (and its square root, the standard deviation (SD)) is one of the first measures used to quantify risk of a portfolio as described by Markowitz (1952) who was awarded a Nobel prize for this contribution. The so called mean-variance optimization theory is a modern quantitative tool used for selecting a portfolio by maximizing the expected return with respect to a certain level of risk. The variance is a reasonable measure to quantify risk but only if the returns are considered to be normally distributed. This is due to the fact that it takes into account the dispersion of values from their mean. It might, however, not be a suitable tool to describe general distributions of returns as it does not examine the tails directly.

A more reflective standard to describe the risk was developed after the 1987 crash and is called Value at Risk (VaR). VaR is defined as the minimum amount of loss that occurs over a certain time period for a given portfolio and at a quantile p.

Denoted mathematically,

$$VaR_p(X) = inf\{x \in \mathbb{R} : P(X > x) \leq 1-p\} = inf\{x \in \mathbb{R} : F_X(x) \geq p\} \quad (3.1)$$

where the random variable X represents losses (expressed as positive values) and F_X is the cumulative distribution function of X.

Value at Risk played a key role in quantifying the capital requirements as a protection against market risk in the Basel II Accord. The VaR provides an estimate about risks yet again it does not describe the tail itself. Another common criticism of VaR is that it is not subadditive, which means that VaR of a portfolio of investments might be greater than the sum of different VaRs for the same but separate investments

$$VaR_p(X + Y) > VaR_p(X) + VaR_p(Y) \quad (3.2)$$

This contradicts the explanation of lowering the risk by diversification. Apart from subadditivity, VaR satisfies all conditions for a coherent risk measure as defined by Artzner et al. (1999). However, it has been shown that VaR for financial returns distribution is subadditive in the tail and thus is suitable for practical applications (Danielsson et al., 2005). Even though VaR still is a standard indicator of a distribution's tail cut-off, it should be much more significant to examine what happens further in the tail, i.e. when losses exceed VaR.

An alternative coherent risk measure to Value at Risk is called the Expected Shortfall (ES). Compared to VaR, it moreover considers the shape of the tail. At certain confidence level α the Expected Shortfall is defined as the expected value of VaRs which exceed VaR_p:

$$ES_p(X) = \frac{1}{1-p} \int_p^1 VaR_z(X) dz \quad (3.3)$$

The simple contribution of Expected Shortfall can be characterized by providing more insight into the tails as we shift from asking about the minimum loss arising in p per cent of cases to the expected loss arising in p per cent of cases.

Despite their shortcomings, variance, VaR and ES are statistical tolls which are, to a certain extent, suitable for decision making. They are straightforward in the way that they encompass the information about the risk in a single number. When considering a statistical perspective, the biggest obstacle in their application is to construct an appropriate estimate for the tails of the underlying distribution. Once we have that kind of estimate, the VaR and ES are rather simple to calculate. However, there are some reasons why these measures alone might not be sufficient for the tail inference.

To summarize the main issues arising even when considering the simplest case

$$X_1, X_2, ...X_n \overset{iid}{\sim} F$$

where we want to draw appropriate inference about the tail of F:

1. the estimates should desirably not be restricted in the range of the so far observed values
2. the extreme observations are rare thus forming only a small sample for the tail behaviour to be modelled
3. more accurate density estimation is needed in the tails

These facts cause the basis for extrapolation to be insufficient and therefore lead to tail models which employ distributions that are justified asymptotically.

3.2 Extreme Value Theory

Extreme Value Theory is a valuable tool to model distribution of extreme observations of returns. This thesis will be focused on modelling univariate extremes of distribution of returns, sometimes called the Profit & Loss distribution function (P&L). There are two basic perspectives of how to identify the extreme observations. The theoretical grounds establishing the Peaks over Threshold (PoT) methodology as well as the second one, Block Maxima Method (BMM), will be described in this section. Extreme distribution functions will be given to model extremes in both approaches.

Even though the theory has its foundations in full parametric modelling of extremes considering their asymptotic behaviour, there has recently been a shift towards computationally simpler semi-parametric models. These, namely the Hill's approach, will be described afterwards and employed in the empirical part of this thesis. Their estimation is achieved under a fairly general framework thus making the computation simple and straightforward.

Throughout this section it is assumed that we have a sample of n *independent, identically distributed* (i.i.d.) random variables which come from a cumulative distribution function F. Later the possible case of stationary, weakly dependent random variables will be considered. For the sample $(X_1, X_2, ..., X_n)$ the following notation will be used for the descending order statistics:

$$X_{1,n} \geq X_{2,n} ... \geq X_{n,n}.$$

In this thesis all the extremes will be considered positive values and referred to as maxima no matter if the right tail (profits) or the left tail (losses) of a probability distribution function is being described. This is made possible because

$$min(X_1, ..., X_n) = -max(-X_1, ..., -X_n)$$

The Extreme Value Theory is concerned with fitting a distribution for the tails without making assumptions about the center of the probability distribution. Therefore, the following topic is a cornerstone of this theory.

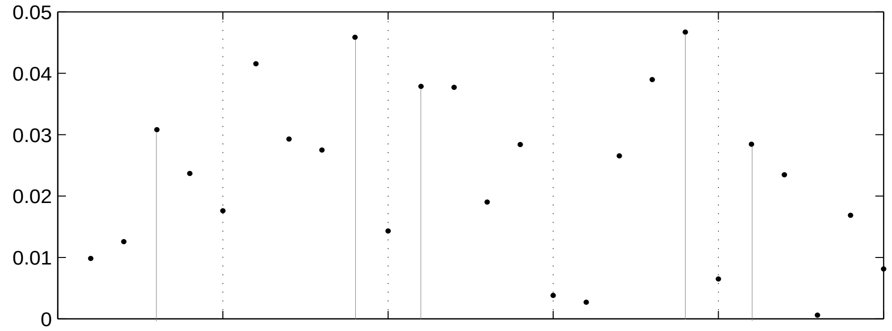
3.2.1 Identifying extremes: Block Maxima versus Peaks-over Threshold methods

The decision upon distinguishing the extremes from the centre of the distribution is one of the most significant ones in extreme value analysis. Inference about the extremes is connected with a certain point of view on the nature of the extremes. That could mean considering largest values either overall or of some fixed blocks. Even though some more advanced approaches have developed, these main 2 will be described and contrasted.

The Block Maxima Method (BMM) is the original approach used at the beginning of EVT applications and it identifies the maxima as the largest values from certain blocks of the data. The applications of this method cover mainly environmental sciences and considered 12-month periods' extreme values to account for seasonality. Therefore this approach is sometimes also called *annual maxima method*. Surely the size of the blocks can differ from 1 year and is to be chosen with maximum care and reasonable interpretation. The sample of size n is usually divided into k subsamples of m observations and the most extreme value is taken from each subsample k. This procedure is done in a way that there is a sufficient number k of extremes coming from large enough subsamples (size m). Then the distribution of the sequence (M_k) of block maxima is studied.

The subjective choice here is of the number of blocks or alternatively the block size because $n = m \times k$. The main critic of this approach is concerned

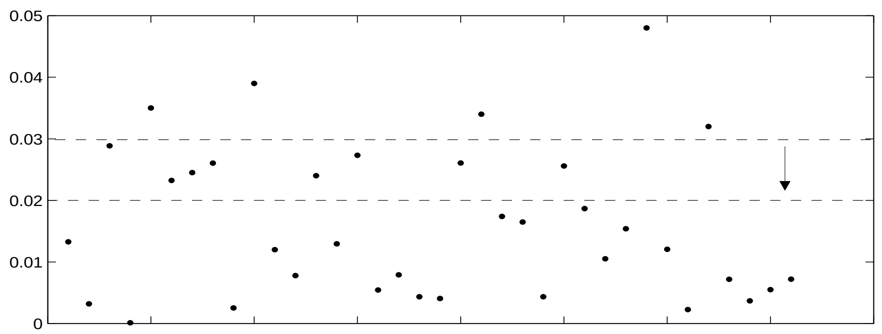
Figure 3.1: Block maxima method example



Source: Author's computations.

with how it deals with the data. As the far extremes are especially rare, the aim is not to disregard some of them. However, that might be the case here because only the largest value is taken from a block thus not counting in other possibly important extreme observations. An example can be seen in Figure 3.1.

Figure 3.2: Peaks over threshold example



Source: Author's computations.

The Peaks over Threshold (PoT) is a completely different approach in that it overcomes the data wasteful nature of BMM. It considers only observations that exceed a certain threshold and then these excesses are to be modelled. Because it uses the data more efficiently than BMM it prevails in most of the recent application studies. However, the threshold selection has a huge impact on the estimation of relevant parameters as it represents a delicate trade-off between variance and bias of the estimator. By setting the cut-off level too high, the sample of extremes contains too few observations causing high variance of estimates. By contrast when lowering the starting point of the tail, observations from the central part of the density might be included. Consequently, fitting a distribution to the tail might produce biased estimates because observations with centre-of-the-density behaviour might be encompassed. An example of PoT can be seen in Figure 3.2. It can be noticed that by slightly lowering the threshold the sample of exceedances increases significantly. The proposed techniques about how to deal with threshold selection issue will be described in a section together with Hill's method employed by this thesis.

3.2.2 Generalized Extreme Value distribution

The first extreme value distribution is called the GEV distribution and is closely connected with the Block Maxima Method. Now a probability framework for block maxima will be stated, following Coles et al. (2001). For a random sample $X_1, ..., X_n \overset{i.i.d.}{\sim} F$, the maxima are defined as

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

Then M_n is distributed as

$$\begin{aligned} Pr\{M_n \leq x\} &= Pr\{X_1 \leq x, ..., X_n \leq x\} && (3.4)\\ &= Pr\{X_1 \leq x\} \times ... \times Pr\{X_n \leq x\} && (3.5)\\ &= F(x)^n && (3.6) \end{aligned}$$

Here, F is unknown so the idea is to approximate F^n by a limit distribution as $n \to \infty$. To look for the distribution, a great similarity comes here when considering the Central Limit Theorem (CLT) for sums of random variables, where the normal distribution plays the key role. Defining the sums of random variables $X_1, X_2, ...X_n$ as $S_n = X_1 + ... + X_n$, the CLT states that as n goes to infinity, normalized sums (where sequence $a_n = nE(X_1)$ and $b_n = \sqrt{Var(X_1)}$) converge in distribution to standard normal distribution Φ:

$$\lim_{n\to\infty} P\left(\frac{S_n - a_n}{b_n} \leq x\right) = \Phi(x), \quad x \in \mathbb{R}.$$

In the same sense, the maxima need to be rescaled as well to obtain a non-degenerate distribution function (does not take on only a single value). The methodology here is followed from the models as described by McNeil et al. (2005).

The convergence of normalized maxima means that there exist suitable sequences $\{c_n\} > 0$ and $\{b_n\}$ such that

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \leqslant x\right) = \lim_{n\to\infty} F^n\left(c_n x + d_n\right) = H\left(x\right) \tag{3.7}$$

For the $H(x)$ to be a non-degenerate distribution function, it has to be found in the GEV family of distributions.

Definition 3.1 (The generalized extreme value (GEV) distribution (as definded by McNeil et al. (2005), p.265)). The cumulative distribution function of GEV distribution is given by

$$H_\xi\left(x\right) = \begin{cases} exp\left(-\left(1+\xi x\right)^{-1/\xi}\right), & \xi \neq 0, \\ exp\left(-e^{-x}\right), & \xi = 0, \end{cases}$$

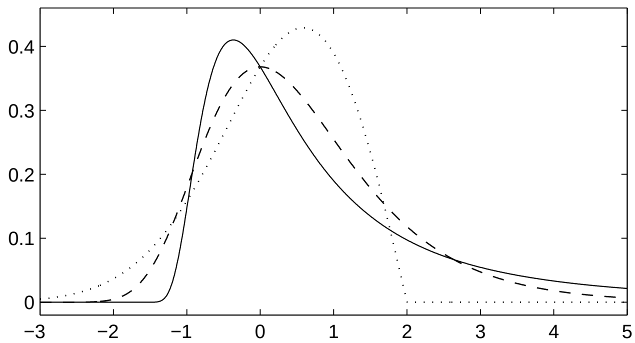
where $1+\xi x > 0$. A three parameter family is obtained by defining $H_{\xi,\mu,\sigma}\left(x\right) := H_\xi\left(\left(x-\mu\right)/\sigma\right)$ for a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$.

Here the location and scale parameters are used to represent the norming constants d_n and c_n which are unknown. This distribution is called generalized because its parametric form summarizes 3 types of distributions belonging to the GEV family. The specific family can be recognized by the parameter ξ, called the shape parameter, which is the primary parameter of interest here. Each distribution family is defined by the shape parameter up to location and scaling. For $\xi > 0$ the distribution family is called a Fréchet case and it describes heavy-tailed distributions which are bounded from below. It contains distributions such as the Pareto or the Student t-distributions, which have tails of negative polynomial decay with infinite right endpoint. The case $\xi = 0$ characterizes the Gumbel class of light-tailed distributions. These include distributions with an exponential tail behaviour such as the gamma, exponential or normal distribution. Finally, the situation when $\xi < 0$ is referred to as the Weibull case which describes short-tailed distributions with finite right endpoint (e.g. beta distribution). The use of these distributions in statistical modelling is enabled because they are continuous in ξ. Therefore, the GEV can be fitted to the block maxima by maximum likelihood method. Despite the disadvantages of block maxima method, the GEV can deal with the data clustering provided a long enough time horizon is considered Coles et al. (2001).

The densities of all three GEV distribution types can be seen in Figure 3.3. It depicts the finite right endpoint of Weibull case and infinite right endpoints

Figure 3.3: GEV probability distribution functions

The dashed line is the Gumbel case ($\xi = 0$), the solid line corresponds to the Fréchet case ($\xi = 0.6$) and the dotted line represents the Weibull case ($\xi = -0.6$). For all the cases $\mu = 0$ and $\sigma = 1$ was used.



Source: Author's computations.

of Fréchet and Gumbel, where the Fréchet tail's decay is much slower. The part which GEV plays in EVT is vindicated by the following definition and theorem.

Definition 3.2 (Maximum domain of attraction (MDA) as defined by McNeil et al. (2005), p.266). If 3.7 holds for some non-degenerate distribution function H, then F is said to be in the maximum domain of attraction of H, written $F \in MDA(H)$.

Theorem 3.1 (Fisher-Tippett, Gnedenko as stated by McNeil et al. (2005), p.266). *If $F \in MDA(H)$ for some non-degenerate distribution function H then H must be a distribution of type H_ξ, i.e. a GEV distribution.*

This result describes the asymptotic distribution of the extremes by stating that when maxima have a limit, it will be in the GEV family. Interestingly, it does not ensure the existence of the limit itself so attention needs to be paid to models where the limit does not exist. (e.g. the Poisson distribution case (Coles et al., 2001)). For the purpose of this thesis, only the case when $\xi \geq 0$ will be considered as the last case is ruled out because financial losses (considered as positive values) cannot be bounded from above.

The Fréchet case For $\xi > 0$ the extremes' distributions can be approximated by the Fréchet case limit distribution. It is possible to simply describe them in terms of slowly varying or regularly varying functions. A formal definition of slowly varying function is provided for the purpose of the next theorem.

Definition 3.3 (Slowly varying function as defined by McNeil et al. (2005), p.268). A positive, Lebesgue-measurable function L on $(0, \infty)$ is slowly varying at ∞ if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \qquad t > 0. \tag{3.8}$$

Theorem 3.2 (Fréchet MDA, Gnedenko as stated in McNeil et al. (2005), p.268). *For* $\xi > 0$*,*

$$F \in MDA(H_\xi) \Leftrightarrow \bar{F}_x = x^{-1/\xi} L(x) \tag{3.9}$$

for some function L *slowly varying at* ∞*.*

This theorem describes the behaviour of distribution F and the form of the tail $\bar{F}$ in the Fréchet case. The expression on the right characterizes regularly varying functions, that means they are power functions multiplied by slowly varying functions. The above stated power function has a negative index of variation $-1/\xi$ which says that the tail decays with rate

$$\alpha = 1/\xi$$

and this rate is the so called *tail index* which is the main parameter of interest in this thesis. Among the most mentioned distributions belonging to this class are the Pareto, Fréchet, F or Student t distribution.

3.2.3 Generalized Pareto Distribution

The block method (annual maxima) and fitting GEV can prove inefficient if there is larger sample of data at hand. The other widely used method to identify extremes is peaks over threshold (or alternatively, r-largest order statistics) and the exceedances over a threshold are modelled by the generalized Pareto distribution.

Definition 3.4 (GPD as defined by McNeil et al. (2005), p.275). The cumulative distribution function of the GPD is given by

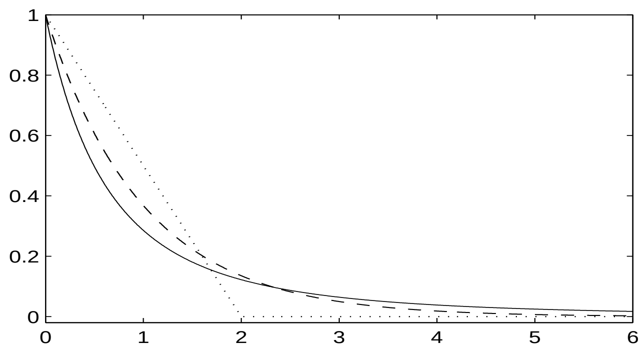
$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - exp(-x/\beta), & \xi = 0, \end{cases}$$

where $\beta > 0$, and $x \geqslant 0$ when $\xi \geqslant 0$ and $0 \leqslant x \leqslant -\beta/\xi$ when $\xi < 0$. The parameters ξ and β are referred to as, the *shape* and *scale* parameter, respectively.

Here again 3 different families of GPD can be characterized by the shape parameter ξ. It indicates the heaviness of the tail and from larger ξ we can infer a heavier tail.

Figure 3.4: GPD probability distribution functions

The dashed line suits an exponential distribution ($\xi = 0$), the solid line represents a Pareto type I distribution ($\xi = 0.6$) and the dotted line corresponds to a Pareto type II distribution ($\xi = -0.5$). For all the cases the scale parameter $\beta = 1$ was used.



Source: Author's computations.

The case when $\xi > 0$ corresponds to Pareto type I distributions, for which the tails decrease as a polynomial. For $\xi = 0$ an exponential type of distribution is obtained, i.e. distributions with exponential tail decay (such as Normal, exponential) and the case when $\xi < 0$ fits the short-tailed Pareto type II distribution. In addition, for fixed x they are continuous in ξ. GPD is a model for threshold excesses distribution and we get the estimates by fitting the distribution to the excess extremes by maximum likelihood. Considering the domain of attraction it can be written that $G_{\xi,\beta} \in MDA(H_\xi)$ (for $\xi > 0$ follows from Theorem 3.2), i.e. the GPD is in maximum domain of attraction of GEV. For the core theorem of fitting the excesses over a threshold, the following definitions are needed.

Definition 3.5 (Excess distribution over threshold u as defined in McNeil et al. (2005), p.276). Let X be a random variable with distribution function F. The excess distribution over the threshold u has distribution function

$$F_u(x) = P(X - u \leqslant x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{3.10}$$

for $0 \leqslant x < x_F - u$, where $x_F \leqslant \infty$ is the right endpoint of F.

Definition 3.6 (Mean excess function as defined in McNeil et al. (2005), p.276). The mean excess function of a random variable X with finite mean is given by

$$e(u) = E(X - u|X > u). \tag{3.11}$$

The distribution function F_u expresses how the excess values are distributed above the threshold u conditioning on the fact that random variable X is above u. The excess distribution function (df) is well known in survival analysis, where it describes the probability of failing in the period $(u, u + x]$ given that no fail occurred until u. The mean excess function expresses the mean of F_u which is defined provided $\xi < 1$ and is given by $E(x) = \beta/(1-\xi)$. An important characteristic of GPD is the linearity of the mean excess function in the threshold u. This property is quite useful for analysing whether the GPD suits the data and will be used in the initial data analysis. The following principal theorem (often called the second theorem of EVT) gives a mathematical result of the limiting excess distribution being the GPD.

Theorem 3.3 (Pickands-Balkema-de Haan as stated in McNeil et al. (2005), p.277). *We can find a (positive-measurable function) β_u such that*

$$\lim_{u \to x_F} \sup_{0 \leqslant x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0, \tag{3.12}$$

if and only if $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$.

This theorem summarizes the domain of attraction for all ξ (unlike the BMM case where the Fréchet case was represented by Theorem 3.2 (the Fréchet MDA, Gnedenko)). It states that distributions which are in maximum domain of attraction of GEV distribution form a large class of distributions whose excess distributions converges to the GPD. According to McNeil et al. (2005) all continuous distributions which are regularly used in statistics belong to the set of dfs in MDA(H_ξ) for some ξ. Therefore Pickands-Balkema-de Haan is an important result which says that the GPD is a recognized and adequately established distribution function for modelling extremes exceeding high thresholds. Moreover, the shape parameter in GPD is the same shape parameter ξ as in the GEV distribution. That means that returns whose block maxima can be fitted by a GEV with certain shape parameter ξ_0 can indeed, for a threshold u high enough, be fitted by a GPD with ξ_0 as well.

Modelling excesses by GPD For reasons stated above, the distribution of extremes will be assumed to belong to MDA(H_ξ). Even though the excesses over a high threshold u might not be distributed exactly as GPD, it will be assumed that for some high enough threshold $F_u(x) = G_{\xi,\beta}(x)$ for some $\xi \in \mathbb{R}$ and $\beta > 0$. The distribution of excesses can be written as $\bar{F}(x) = \bar{F}(u)\,\bar{F}_u(x-u)$. Then, the GPD is fitted to the excesses by maximum likelihood (ML) assuming the data to be independent and identically distributed, thus getting an empirical tail estimator for the distribution of excess:

$$\hat{\bar{F}}(x) = \frac{N_u}{N}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \tag{3.13}$$

where N_u represents the number of excesses over a threshold u in the sample of size N.

3.3 Tail index and semi-parametric methods

Until now the main two parametric approaches to modelling extreme values were introduced; fitting the GEV distribution to block maxima and fitting the GPD to excesses over a certain threshold. In this thesis the focus is moved from these fully parametric methods towards semi-parametric methods. Completely non-parametric models obviously have some weaknesses in the sense that they cannot deal with out-of-sample quantiles therefore making data extrapolation impossible. Although this might represent a serious drawback for forecasting, such methods still keep their usefulness in analysing the observed data. By making some stronger *a priori* assumptions about the data generating process semi-parametric single-index models represent a desired compromise. Within the semi-parametric approach, the extreme observations are chosen as either excesses over a fixed threshold or directly as k out of the n available observations. The parametric fitting as described above is simplified to the estimation of only one key parameter: shape parameter ξ or the *tail index* expressed as

$$\alpha = 1/\xi.$$

Therefore no precise parametric model is fitted to the data and no scale or location parameters need to be estimated. Only by assuming that the excess distribution function is in the maximum domain of attraction of GEV and using the appropriate EVT methodology, the shape parameter is to be estimated.

3.3.1 Hill's method

In the Hill's method, the underlying distribution of losses is assumed to be in the maximum domain of attraction of some distribution belonging to the Fréchet case of GEV. That means that fat-tails are suspected beforehand, expecting $\xi > 0$. From the Theorem 3.2 (the Fréchet MDA, Gnedenko), it follows that the distribution can be expressed in the means of some slowly varying function L and parameter α:

$$\bar{F}(x) = L(x)x^{-\alpha} \tag{3.14}$$

Here, the parameter of interest is directly the tail index α which is the reciprocal of the shape parameter ξ. The estimator is based on independent identically distributed data arranged in order of decreasing size: $X_{1,n} \geqslant ... \geqslant X_{n,n}$. As originally derived by Hill (1975), the Hill's tail index estimator is usually obtained by maximum likelihood estimation and expressed as

$$\hat{\alpha}^{(H)}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k+1,n} \right)^{-1}, \qquad 1 \leqslant k \leqslant n. \tag{3.15}$$

Alternatively, one can compute the estimate of the shape parameter as

$$\hat{\xi}^{(H)}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k+1,n}, \qquad 1 \leqslant k \leqslant n. \tag{3.16}$$

Weak consistency is ensured for *iid* data and even for weakly dependent data (Hsing, 1991) so that if $k \to \infty$ and $k/n \to 0$ for $n \to \infty$, then $\hat{\alpha}^{(H)} \xrightarrow{P} \alpha$. For X_n which is *iid* the estimator is also strongly consistent and asymptotically normal (McNeil et al., 2005).

3.3.2 Graphical threshold selection

Looking at some graphical representations might help to decide on a reasonable threshold value ($X_{k+1,n}$). First indicator is the mean excess function which should be considered in the initial data analyses. The fact that the mean excess function is linear while considering different thresholds u is an evidence that the GDP model for the particular levels is valid. Moreover, if the linear line has an upward trend, it indicates the GPD model with shape parameter $\xi > 0$. If the plot shows horizontal linear line, then the exponential distribution might be the right fit. Therefore it is suggested to look for linearity in the **sample**

mean excess plot:

$$(X_{i,n}, e_n(X_{i,n})) : \qquad 1 \leqslant i \leqslant n,$$

where $X_{i,n}$ is the ith upper order statistic and $e_n(X_{i,n})$ is the empirical estimator of the mean excess function from Definition 3.6 where $X_{i,n}$ represents a certain threshold u:

$$e_n(u) = \frac{\sum_{j=1}^{n} (X_j - u) I_{\{X_j > u\}}}{\sum_{j=1}^{n} I_{\{X_j > u\}}}. \tag{3.17}$$

Another indicator of a suitable threshold is the co called **Hill plot**, a graph of varying Hill shape estimates depending on the cut-off level, as already described by Hill (1975). The graph is given by

$$\left(k, \hat{\xi}_{k,n}^{(H)}\right) : \qquad k = 1, ..., n,$$

and here the stability of the estimate is desired thus a stable region in the graph will be looked for. Usually it is suggested to look for the stable part in the upper $1 - 5\%$ of the order statistics. Most helpful is to use the Hill plot when the data comes directly from Pareto or close to Pareto distributions. Then the graph provides a clear evidence on the value of the estimator (Drees et al., 2000). However, in other cases the plot might be quite volatile and not show any steady region therefore it is not clear which part of the graph gives the best accuracy. A great improvement to this method was made by Resnick and Stărică (1997) who propose an alternative method: the **altHill plot** where the x axis is rescaled. The graph is given by

$$\left(\theta, H_{\lceil n^{\theta} \rceil, n}\right), \qquad 0 \leqslant \theta < 1,$$

where $H = \hat{\xi}_{k,n}$ and $\lceil n^{\theta} \rceil$ stands for an integer equal to n^{θ} or the smallest integer greater than n^{θ}. The purpose of the changed scaling is to give more space in the graph to the smaller values of k which is the concern here because only a small number of k upper order statistics are believed to be the far tail observations. Drees et al. (2000) examined the superiority between the classical Hill and alternative Hill method. By proposing a measure for indicating the occupation time which a plot spends in the neighbourhood of the true estimate, they compared the two approaches. Their results show that the classical Hill plot is better when dealing with Pareto-like distribution. Otherwise there is a clear superiority of the *alt*Hill plot. However, in practice it is advised to use both plots (classical and *alt*Hill) and draw inference from their comparison. Another refinement technique was introduced by Resnick and Stărică (1997). They propose smoothing the shape parameter estimator ($\hat{\xi} = 1/\hat{\alpha}$) by a simple averaging:

$$av\left(\hat{\xi}_{k,n}\right) := \frac{1}{(u-1)\,k} \sum_{p=k+1}^{uk} \hat{\xi}_{p,n}, \tag{3.18}$$

where $u > 1$. By this adjustment, volatility of the plot is decreased and a stable region might appear. This estimate provides a good basis for other computations which would for example try to lower the bias by bootstrap techniques. However, if there is a substantial bias in the estimate, the averaging does not help to reveal more information. The choice of u has to be made in a way that the asymptotic variance is decreased (choose u high enough) but this is restricted by the sample size n therefore Resnick and Stărică (1997) suggest to take u between $n^{0.1}$ and $n^{0.2}$.

McNeil and Frey (2000) point out that the threshold selection is not such an issue when fitting GDP parametrically because the method is robust for $k > 50$. However, when applying the Hill's method it is necessary to choose an appropriate low number k of upper order statistics to curb the bias. To draw inference from the Hill plots, the following procedure is suggested for this thesis:

1. Compute the Hill shape estimates $\hat{\xi}$ and make a classical Hill plot
2. Zoom in the classical plots so that the plot can be seen for the upper 10% of the order statistics
3. Take the Hill estimates of shape parameter ($\hat{\xi} = 1/\hat{\alpha}$) and apply the smoothing procedure to get the $av\left(\hat{\xi}_{k,n}\right)$
4. Take the Hill shape estimates $\hat{\xi}$ and plot them in the alternative Hill plot and take the smoothed estimates $av\left(\hat{\xi}_{k,n}\right)$ and plot them in the same alternative Hill plot for comparison

3.3.3 Huisman's method

In the empirical research the Hill's method has been commonly used. This is mainly earned by the estimator's good asymptotic properties and simple implementation. However, it is biased in samples that are rather small in size. Even though no extremely short samples will be used in this thesis, a method that corrects for this problem is relevant here. The reason for it is that it enables splitting the long enough sample and apply the method for the separated subsamples to check whether the tail behaviour changes over time. An approach which corrects for the estimator's bias that arises due to small sample at hand was proposed by Huisman et al. (2001). The method yields estimates whose small bias is close to that of estimates for especially long time horizons. Moreover, it overcomes the shortcoming of graphical methods: their subjectivity. Here the estimate is not based on a certain number of order statistics but it utilizes information from many estimates that are based on different number of values above a threshold. Therefore the final estimate is a weighted average of a whole class of classical Hill estimators. In this thesis the methodology of Huisman will be followed and therefore the computations will be performed for the inverse of the tail index, i.e. the shape parameter ξ.

In this method the Hill shape estimates are computed for various thresholds k up to a reasonable last threshold taken into account, denoted κ. The basic idea is that for a sufficiently small k, it is possible to approximate the bias of the shape parameter estimate by a linear function depending on k. Then, the vector of $\{(\xi(k)) : k = 1, ..., \kappa\}$ can be used to estimate the following regression:

$$\xi(k) = \beta_0 + \beta_1 k + \varepsilon(k), \quad k = 1, ..., \kappa. \tag{3.19}$$

For k getting close to 0, an unbiased estimate would be reached with value equal to the intercept β_0.

There are 2 major obstacles to estimating the regression by ordinary least squares (OLS). The first one is that the error terms are heteroscedastic because the variance of Hill shape estimator is changing in k as derived by Hall (1990): $Var(\xi(k)) \approx \frac{1}{k\alpha^2}$. However, this problem can be treated by weighted least squares (WLS). The second issue is that variables $\xi(k)$ are correlated because they are computed from observations that overlap. As a result, the usual standard errors are not applicable. Not to neglect these two matters, the following procedure proposed by Huisman et al. (2001) will be used. It corrects for heteroscedasticity by WLS and also an appropriate computation of standard errors is proposed.

Weighted least squares The Equation 3.19 can be expressed by matrix notation as follows:

$$\boldsymbol{\xi}^* = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}, \tag{3.20}$$

where the vector $\boldsymbol{\xi}^*$ denotes $\xi(k), k = 1, ..., \kappa$, $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is a vector of regression coefficients and $\boldsymbol{Z}$ is a $(\kappa \times 2)$ matrix:

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & \kappa \end{pmatrix}. \tag{3.21}$$

To deal with the heteroscedasticity in error term, Huisman et al. (2001) suggest the following WLS matrix:

$$\boldsymbol{W} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sqrt{\kappa} \end{pmatrix}.$$

This weighting matrix reflects weights which are proportional to the reciprocal of the variance of dependent variable (the Hill shape estimate). Therefore the heteroscedasticity is settled by giving less weight to the observations with higher variance, that means Hill shape estimates computed based on fewer order statistics. Here it should also be noted that the optimal weighting matrix is unknown and should be estimated by feasible generalized least squares (FGLS). The greater efficiency of FGLS over OLS is ensured when the variance matrix is known (Greene, 2003, p.217). In case that it is unknown, the comparison is not that straightforward. Simulation results of Huisman et al. (2001) will be taken into account, where it is stated that in finite samples GLS is less accurate. Therefore it will not be performed in this thesis either and the matrix presented above will be used for WLS. Also the problem arising from correlated dependent variables has to be considered because the real matrix would not be diagonal. However, this requires the knowledge of the whole variance-covariance matrix and therefore this thesis will apply the simplified procedure with diagonal matrix keeping in mind that the estimates are not the most efficient.

Using the Huisman's methodology with weights given by the weighting matrix W gives the following estimates for the vector of coefficients from Equation 3.20

$$\hat{\beta}_{WLS} = (Z'W'WZ)^{-1}Z'W'W\gamma^*. \tag{3.22}$$

The Hill shape estimate is then the first element of the estimated vector $\hat{\beta_{WLS}}$ and the desired tail index is its reciprocal value.

Standard errors for Huisman's estimator As stated before, the dependent variables $\xi(k)$ of regression in Equation 3.19 are correlated. The reason for it is that any 2 estimates $\xi(i)$ and $\xi(j)$, where $i < j$, are based on i common observations. Therefore the following method was proposed as usual standard errors are not applicable (see Appendix in Huisman et al., 2001, p. 215). The general idea behind the correction is that the Hill estimator is in fact a linear combination of order statistics $(Y_1, ..., Y_n)$ where $Y_i = ln(X_i)$. Then for some $(\kappa) \times (\kappa + 1)$ transformation matrix $\boldsymbol{A}$, the vector $\boldsymbol{\xi^*}$ can be expressed as $\boldsymbol{\xi^*} = \boldsymbol{Ay}$ where $\boldsymbol{\xi^*} = (\xi(k) : k = 1, ..., \kappa)$. For the Hill's shape estimator proposed by Huisman in Equation 3.22, the covariance matrix is expressed as

$$cov(\beta_{WLS}) = (Z'W'WZ)^{-1}Z'W'W\Omega W'WZ(Z'W'WZ)^{-1}, \tag{3.23}$$

where $\boldsymbol{\Omega} = \boldsymbol{A\Sigma A'}$ is a covariance matrix for the shape estimates $\xi(k)$. It can be computed from the transformation matrix $\boldsymbol{A}$ and the matrix $\boldsymbol{\Sigma}$ which is the covariance of order statistics Y_i. These can be obtained as follows.

The transformation matrix A is derived from the conventional Hill shape estimator:

$$A = \begin{pmatrix} 0 & \dots & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & \dots & 0 & 0 & 0 & -1 & 1/2 & 1/2 \\ 0 & \dots & 0 & 0 & -1 & 1/3 & 1/3 & 1/3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1/\kappa & \dots & 1/\kappa & 1/\kappa & 1/\kappa & 1/\kappa & 1/\kappa \end{pmatrix} \tag{3.24}$$

The covariance matrix Σ is computed based on asymptotic normality of order statistics and covariances between two order statistics are equal to

$$cov\,(i,j) = \frac{p(i)(1-p(j))}{nf_z(\mu(i))f_z(\mu(j))} \tag{3.25}$$

for $i \leq j$. Here $p(i)$ can be approximated by i/n. Furthermore, for this calculation step it is assumed that the original data $X(i)$ comes from a Pareto distribution for $i = 1, ..., \kappa$. Then the density is given by $f(x) = \alpha b^{\alpha} x^{-\alpha-1}$ for $x \geq b$ and mean is given by $\mu(i) = ln((1-p(i))^{-1/\alpha}$ where α is approximated by the inverse of the estimated shape parameter ξ. As the matrix Σ is fully defined now, covariance matrix Ω can be obtained and the covariance in Equation 3.23 computed to get the appropriate standard errors.

3.4 Application of the tail index

To summarize the main assumptions for an application of the tail index, the basic framework will be restated. It is assumed that the the extreme observations are independent, identically distributed and their distribution is in the maximum domain of attraction of some GEV distribution. Then the tail can be expressed in the following form where ξ (or, alternatively, $\alpha = 1/\xi$) is the parameter of interest:

$$\bar{F}(x) = x^{-1/\xi}L(x), \quad x > 0, \tag{3.26}$$

where $L(x)$ is a slowly varying function and when multiplied by the power function $x^{-1/\xi}$, the tail is expressed by a regularly varying function.[1] Also by assuming only the Fréchet case, $1/\xi = \alpha > 0$, the underlying distribution is suspected to be heavy-tailed. From the expression in 3.26 the semi-parametric characteristic of the estimation can be clarified as the constituent with ξ represents the parametric part and the function L the non-parametric one. To estimate the parameter, k upper order statistics from the sample of size n are employed. Here, 2 assumptions are made. Firstly, one should use adequate amount of statistics, that means that $k(n) \to \infty$ as $n \to \infty$. Secondly, the focus should be only on the upper order statistics and therefore $k(n)/n \to 0$ as $n \to \infty$.

Even though we should not be tempted by the 'Single number tells it all' caveat, the interpretation of the tail index is very straightforward. As it represents the exponent of regular variation, higher tail index α indicates faster tail decay, i.e. lighter tail character. On the other hand, the larger the shape parameter ξ the slower the decline towards the tails which gives evidence on large number of extreme observations. Another interesting result which can be inferred directly from the tail index α is the number of finite moments. If the extreme observations follow GPD distribution, then the tail index tells us how many finite moments exist as $E(X^k) = \infty$ for $k \geqslant \alpha$. Therefore, from index α

[1] Terminology note: ξ is called the shape parameter and $\alpha = 1/\xi$ is the tail index which is sometimes also called the exponent of regular variation. Throughout this thesis terms of both parameters are used depending on the situation.

smaller than 1 no inference can be made about expected losses. For hypotheses testing, asymptotic normality of the Hill's estimator can be obtained, although only by imposing second order conditions for regular variation. However, if the goal is to test whether the underlying distribution is thin-tailed ($H_0 : \alpha = 0$), it is not possible because $\alpha > 0$ is assumed in the Hill's estimation beforehand. Thus the Hill's estimator is designed for the fat-tailed case which cannot be tested afterwards.

After the tail index estimator has been computed it can immediately be used to obtain an estimate for the tail:

$$\widehat{\bar{F}}(x) = \frac{k}{n}\left(\frac{x}{X_{k,n}}\right)^{-\hat{\alpha}_{k,n}} \tag{3.27}$$

and the corresponding p-quantile x_p:

$$\widehat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-1/\hat{\alpha}_{k,n}} X_{k,n}. \tag{3.28}$$

Also an estimator of excess distribution function $F_u(x-u), x \geqslant 0$, can be obtained through $F_u(x-u) = 1 - \bar{F}(x)/\bar{F}(u)$. Now, as the distribution function of extreme returns is specified, VaR analysis can be implemented with $VaR_q = F_q^{-1}$. Unlike a historical simulation method for computing VaR, the EVT provides a framework suitable also for out-of sample-forecasts and therefore the semi-parametric VaR can be expressed as

$$VaR_p = \left(\frac{n}{k}(1-p)\right)^{-\hat{\xi}_{k,n}} X_{k,n}. \tag{3.29}$$

Another useful application of the shape parameter is that it effectively determines the limit ratio of the Expected Shortfall and Value at Risk. It means that for very high quantiles, ξ can be used to compute the proportional relationship of ES and VaR as:

$$\lim_{p\to 1}\frac{ES_p}{VaR_p} = \begin{cases} (1-\xi)^{-1}, & \xi \neq 0, \\ 1, & \xi = 0. \end{cases} \tag{3.30}$$

Finally, an important stylized fact of financial returns has to be discussed as it contradicts the basic assumption of the EVT theoretical framework. It is well known that financial returns are dependent considering the second moments. Even though this dependence could be removed by a model where risk forecast is conditioned upon present market information, the unconditional model is still suitable for forecasts in longer horizons (Danielsson and De Vries, 2000). In the EVT conception, a 2-step procedure called the Conditional EVT could be employed. By the so called data prewhitening in the first step, the dependencies could be removed by one of the GARCH family models. This would yield standardized residuals that are approximately independent, identically distributed and EVT could be applied on them in the second step.

However, this procedure could be subject to a serious model misspecification in the first step. McNeil and Frey (2000) claim that the results are not sensitive to a minor misspecification of the GARCH model. In recent literature the 2-step approach occasionally appears, but the choice of the model is often not reasoned enough and simply followed by the straightforward GARCH(1,1). By this, the model is likely to be severely misspecified as the return series seldom follow a simple symmetric model and need to be examined by a more thorough analysis. Another problem arises for model specification as the autocorrelation function (ACF) cannot be used to determine the number of lags in the model. This is stressed by Mikosch (2003) who contends that the lack of higher moments in some financial series causes the confidence bands for autocorrelation to be underestimated.

On the other hand Danielsson and De Vries (2000) argue that if a longer time horizon is considered, the unconditional approach might be the preferable one. Moreover, a 2-step conditional EVT method could be opposed from another point of view. As it is not directly the returns that are to be modelled in the second step of the analysis, a substantial amount of information could be lost. In case the interest is to model both the left and right tail, using standardized residuals instead of returns would bring a completely different sample size for these separate calculations than is the corresponding number of losses and profits. In this thesis, therefore, the unconditional approach will be used with the knowledge that dependence brings consequences on the estimator's properties. During the estimation the time series is believed to be stationary with distribution F in the maximum domain of attraction of some extreme value distribution. Then the EVT can still be applied by making a simplifying assumption of independent extremes. This procedure will still bring consistent estimates, even though the standard errors might be too small.

Chapter 4

Empirical Research

The overall aim of the empirical part of this thesis is to examine tail characteristics of particular stock index and main stocks from the index. It includes the initial data analysis and examining the properties of data at hand. Then, the tail index will be computed for all the series chosen for this study.

4.1 Data Analysis

For the empirical analysis the following data was used as obtained from Thomson Reuters database[1] in the form of daily close prices. If the price series are denoted as $(P_0, P_1, ..., P_n)$ then the log-returns can be computed by logarithmic transformation as $X_t = log(P_t/P_{t-1})$ and the resulting return series $(X_1, ..., X_n)$ is then used in the consequent analysis. Log-returns were used to obtain series for the following stocks in the specified length.

Table 4.1: Data analysed

full name	symbol	ticker	length used
Standard & Poors 500 Index	S&P 500	SPX	21.3.1980 - 29.2.2016
Exxon Mobile Corporation	Exxon	XOM	20.3.1980 - 29.2.2016
JPMorgan Chase & Co.	JPMorgan	JPM	19.3.1980 - 29.2.2016
International Business Machines Corp.	IBM	IBM	19.3.1980 - 29.2.2016
Johnson & Johnson	Johnson	JNJ	18.3.1980 - 29.2.2016
General Electric Company	GE	GE	20.3.1980 - 29.2.2016
The Procter & Gamble Company	P&G	PG	20.3.1980 - 29.2.2016
The Walt Disney Company	Disney	DIS	19.3.1980 - 29.2.2016

[1]For the company's website, see www.thomsonreuters.com.

Unlike other studies which focus on a specific geographical layout of the stocks in their analysis, here the focus is made on different industries. Firstly, the base of the analysis comprises of examining the Standard & Poors 500 Index, which reliably represents the U.S. equities market. It is the leading standard in reflecting the U.S. market conditions and is designed to show the risk and return nature of the large market capitalization companies. Market capitalization is the feature that simply determines companies' weights in the index. Also the index is constituted in a way that all main industries have their representation. This is the reason why both the overall index and separate stocks exemplifying different industries will be analysed. In the following table 7 important stocks are linked to the corresponding industries and the industry weight in the index is noted.

Table 4.2: Industries and selected stocks

Industry	Sector weight in S&P 500	Stock
Information technology	20.4%	IBM
Financials	15.6%	JPMorgan
Health care	14.7%	Johnson
Consumer discretionary	12.9%	Disney
Consumer staples	10.7%	P&G
Industrials	10.1%	GE
Energy	6.6%	Exxon

The S&P 500 comprises of 504 leading stocks with large market capitalization and the stocks used in this thesis are all in the top 31 places according to their market capitalization.[2] The analysed stocks are chosen from seven significant industries and they are among the leading companies in the particular industry. In each industry the constituent with the largest weight was taken provided it had the desired length, otherwise the one with next largest weight was taken.

When choosing the length of the sample, several considerations have to be made. Firstly the goal is to include large amount of observations due to statistical properties of the estimators. Also it is desirable for the time span to include important market crashes because the properties of the tails are to be examined. Therefore, the time series begin in March 1980 and they constitute samples of approximately 36 years. Finally, this date was chosen so that all

[2] As of March 16, 2016.

the stocks were also constituents of the S&P 500 Index for the whole period.[3] As well as using the full sample, the analysis will be conducted also on shorter subsamples. The original sample is divided into 3 shorter time spans to examine whether the tail characteristics change over time. By splitting the sample, 12-year long periods of 3022 observations are obtained. They are believed to be adequate for analysis of extremes as each of them contains big market losses. The first period's results will be affected by the 1987 market crash. The next series will be influenced by the new millennium's volatility and the 2002 stock market downturn. The last one then shows the consequences of 2007-2008 financial crisis. The next table presents the periods in question.

Table 4.3: Sample and its division

	time span	observations
full sample	March 1980 - February 2016	9066
period 1	March 1980 - February 1992	3022
period 2	March 1992 - February 2004	3022
period 3	March 2004 - February 2016	3022

Table 4.4 shows the main descriptive statistics of the full data set. The sample mean of all the financial time series is positive and close to zero. The skewness is negative for all the series, which suggests a longer left tail. The last statistic, the excess kurtosis, expresses the kurtosis as a value above the kurtosis of the Normal distribution, which is equal to 3. Here the values are positive, therefore all the data is suspected to be far from normally distributed.

Table 4.4: Descriptive statistics

Stock	Max	Min	Mean	Sd	Skewness	Exc. Kurtosis
S&P 500	0.1096	-0.2290	0.0003	0.0113	-1.1527	26.4378
Exxon	0.1648	-0.2669	0.0003	0.0150	-0.4670	18.9910
JPMorgan	0.2239	-0.3246	0.0003	0.0233	-0.1168	14.5341
IBM	0.1237	-0.2609	0.0002	0.0170	-0.3841	12.6859
Johnson	0.1154	-0.2028	0.0005	0.0147	-0.3433	9.0296
GE	0.1798	-0.1922	0.0004	0.0173	-0.0933	9.0217
P&G	0.2005	-0.3601	0.0004	0.0148	-2.5303	70.1627
Disney	0.1748	-0.3438	0.0005	0.0195	-0.8166	19.4202

[3]Due to a later date of some stocks' incorporation to the index, sample could not be taken to cover also the period of 1970s, which would certainly be well-founded for the analysis of extreme events because of the oil shocks and substantial market losses.

When looking at maxima and minima of the samples, it can be directly seen that the minima are in absolute value much larger than the maxima, in case of S&P 500, IBM or Disney they are almost twice the amount. This could already be a sign of some distributional imbalance in the sample. The largest losses in these samples occurred on October 19, 1987 and were exceptionally severe compared to other losses. The only exception is P&G at which the largest loss was experienced on March 7, 2000, when P&G went through an organizational crisis itself. As far as the maxima are concerned (the largest profits), for all the stocks they were attained during either 1987 or $2008 - 2009$ crises. This surely indicates highly volatile periods as extreme losses appeared in these periods as well.

The highest sample variance from these stocks is detected in JPMorgan ($Var = 0.0005$). Annualized variances are computed using the value of 252 for the number of trading days in a year: $Var_{annual} = Var_{daily} \times 252$. The results are shown in Table 4.5 and it is clear that all the annualized variances of particular stocks considerably exceed the annualized variance of the full index. JPMorgan's sample annual variance is 13.64% while the S&P 500's is just 3.2%.

Table 4.5: Annualized sample variances

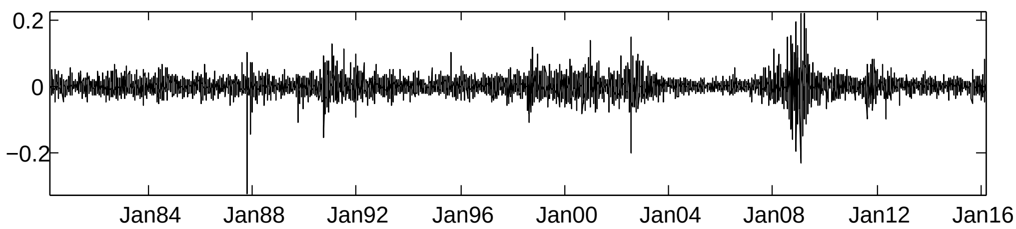
stock	sd	variance	annualized variance
S&P 500	0.0113	0.0001	3.20%
Exxon	0.0150	0.0002	5.69%
JPMorgan	0.0233	0.0005	13.64%
IBM	0.0170	0.0003	7.30%
Johnson	0.0147	0.0002	5.46%
GE	0.0173	0.0003	7.52%
P&G	0.0148	0.0002	5.55%
Disney	0.0195	0.0004	9.55%

Descriptive statistics of the individual subsamples can be found in Appendix B. Period 3 seems to be the most volatile period with annualized variance of JPMorgan being high at 16.38%. Even though the largest loss did not occur in this period, the largest profit of 22% was gained. On the other hand, for IBM and Disney the annualized volatility is higher in the 2nd period (11.83 and 11.61 per cent, respectively). Interestingly, there are shifts in skewness for different periods. Unlike in the full sample or the 1st period, some of the stocks turn to be positively skewed in periods 2 and 3 (e.g. JPMorgan and

Johnson). The excess kurtosis is overall the highest in period 1 when it is probably influenced by the severe losses that all stocks suffered during 1987. Only for P&G, the series is also substantially leptokurtic in period 2 when its largest losses from the year 2000 are involved.

The plotted return series, which can be seen in Appendix A, exhibit clusters of extreme values. Some of the biggest clusters of extremes can be seen in 1987, at the end of 1990s and the beginning of the new millennium. Then another period of high volatility comes in 2002 and the most noticeable cluster appears at the end of 2008. Especially large clusters can be noticed with P&G stock in 2000 or JPMorgan stock in 2008. Generally, larger clusters appear in individual stock return series rather than in the whole index S&P 500. Volatility clustering is a stylized fact (Andersen et al., 2009), which comes as no surprise in financial time series (as well as the fat-tailedness indicated by the sample kurtosis).

Figure 4.1: Log-returns of JPMorgan stock

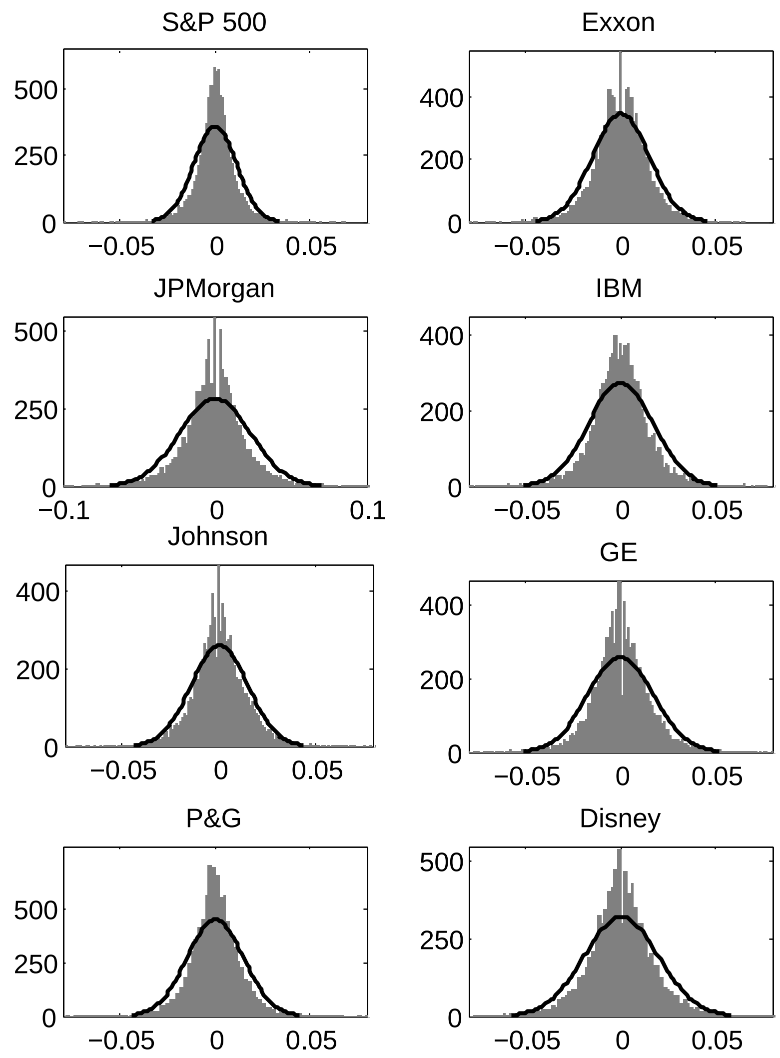


Source: Author's computations.

The next step is to examine the behaviour of the log-returns by formal tests. Firstly, the stationarity of the process is checked. The original stock price series was not stationary as the mean of such series clearly changes over time. However, by the logarithmic transformation of prices to log-returns, the stationarity is believed to be a valid assumption. This is also confirmed by the KPSS test, where the null hypothesis of stationarity cannot be rejected at significance level of 0.01. In the shorter subsamples (time span of 12 years) similar results were obtained, therefore stationarity cannot be rejected for these intervals either. Complementary tables showing tests results can be found in Appendix B.

Next the normality of the series is tested by Jarque-Bera test which compares the 3rd and 4th moment with the Normal distribution. Here the null hypothesis of coincident moments is strongly rejected with p-value smaller than 0.001. However, the higher moments might not be finite thus additional methods are considered to reject normality. Histograms are plotted in Figure 4.2 with the density of the Normal distribution fitted to the data. The normal density is only evaluated at the particular histogram bins and therefore does not represent the classical smooth probability distribution function (pdf) curve. It can be seen that the histograms are symmetric in the neighbourhood of zero, showing a high peak in the center and heavy tails on both sides. It can be clearly seen that the normal distribution is not a good fit for the data as the center of the density contains much more observations. On the other hand, data is lacking in between the center and the tails when compared to the normal

Figure 4.2: Histogram of log-returns with fitted Normal distribution

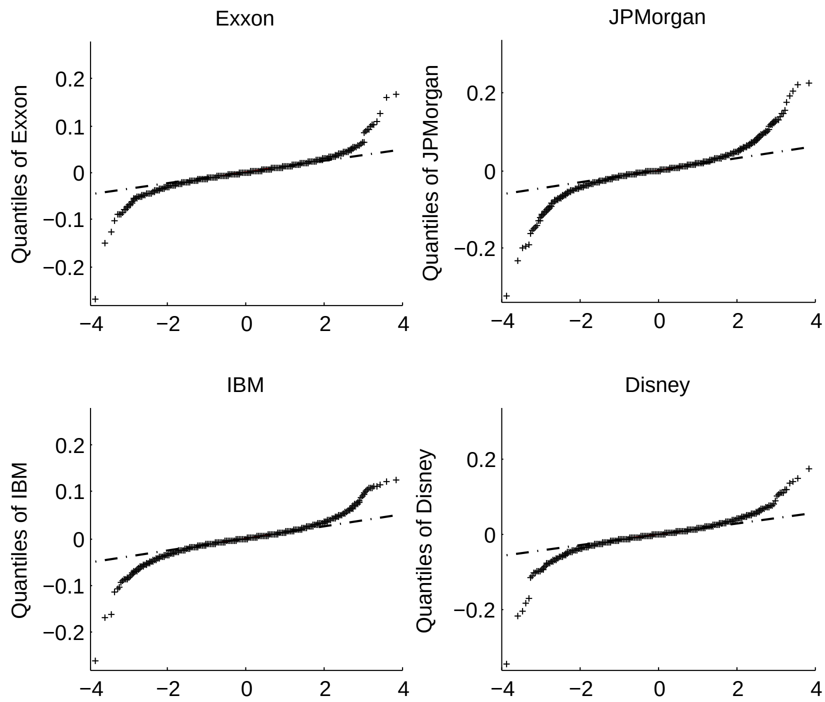


Source: Author's computations.

case. Finally, the tails of the sample distribution are longer and far extremes outside of the scope of normality can be spotted.

Also, the Q-Q plots are used (selected Q-Q plots are reported in Figure 4.3) and they show clearly that in the tails the quantiles of the empirical distribution get far away from the quantiles of the Normal distribution. Finally, properties

Figure 4.3: Q-Q plots against the Normal distribution



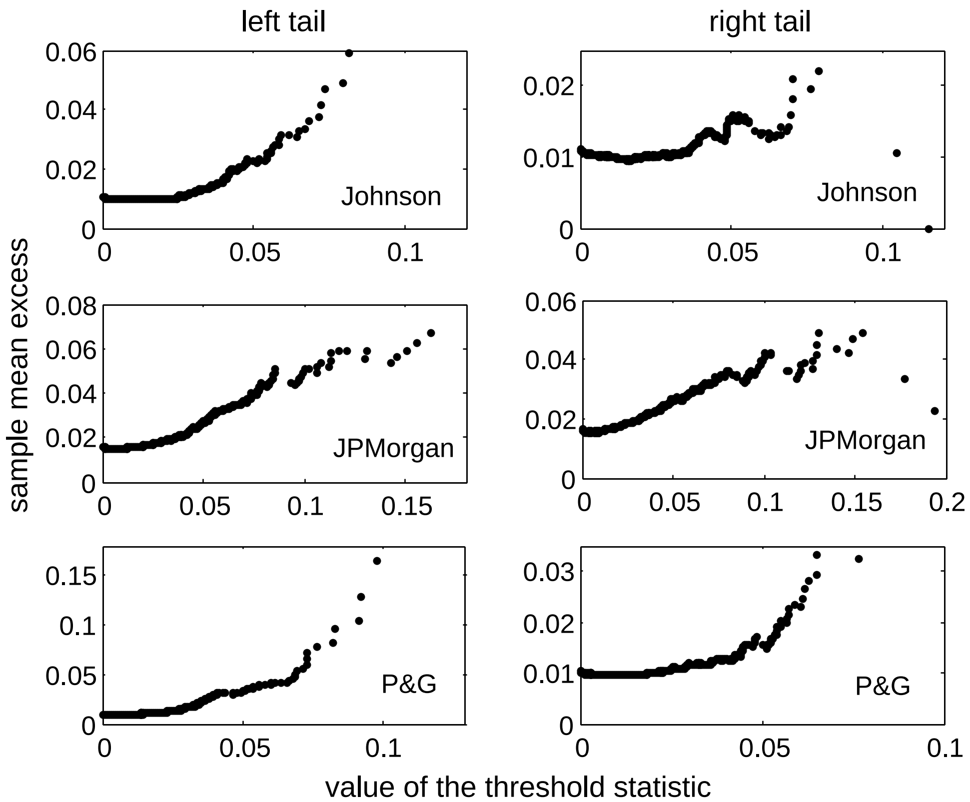
Source: Author's computations.

of the tails are examined in terms of residual heteroscedasticity. Because the extremes appear in clusters, some dependence in the squared residuals is suspected. Engle's Autoregressive Conditional Heteroscedasticity (ARCH) test is employed to see whether there is any dependence (ARCH effects) in the residuals. ARCH test is conducted at a 1% significance level using lags equal to 1, 3 and 7. The null hypothesis of no conditional heteroscedasticity is rejected for all the lag structures, which means that there is some level of dependence. However, as discussed in the theoretical part, this dependence will not be modelled. Now, the stationary series will be used for the next analysis with the simplifying assumption of independence. The Extreme Value Theory will be applied and the properties of the Hill estimator will be inferred with the notice that the series are weakly dependent.

4.2 Application of EVT

The initial part of the extreme value analysis includes looking for a tail threshold by using several graphical methods. At first the mean excess plots are employed. As described in the theoretical part, the sample mean excess function is plotted here for different threshold values. Interesting examples of these plots are presented in Figure 4.4, plots for the whole data samples are given in Appendix A. The first characteristic of the plots to be noticed is their up-

Figure 4.4: Sample mean excess plot for selected stocks

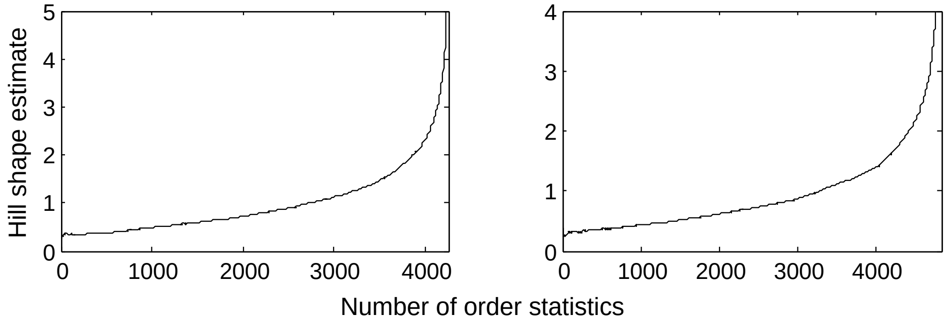


Source: Author's computations.

ward slope. This indicates a positive value of the estimated shape parameter and thus the fat-tailedness of the underlying distribution. The theory gives guidelines on the inference of the appropriate threshold from the point where approximate linearity of the graph begins. From that point, the observations are said to be suitable to be fitted with the GPD model. Here it should be noted that the plotted values for a few highest thresholds should not be considered in this suitability analysis as they express the mean excess of only a few most extreme values.

Generally, it can be seen that the mean excesses in the right tail (profits) are smaller in value than is the case of the left tail (losses). Only JPMorgan right tail's excesses are comparable with the left tail provided that the far extremes are neglected. If the starting point of the linearity is searched for, a seemingly clear situation comes in the Johnson's losses. 2.5% is the value of the loss from which the GPD distribution seems to be a reasonable fit. This cut-off point would separate 345 extremes which represent about 8% of the sample of losses. This, however, is a large proportion for the extreme value model as $3-5\%$ of observations usually fit the distribution. After a closer examination, a slight change of the slope around 4% might indicate the threshold there. This is therefore not objective here to decide by a simple look at the graph. Completely different situation arises in Johnson's profits. After an initial rather stable phase, there is no linearity to be spotted in the graph. Two considerable slumps dominate the graph and it is quite possible that this sample is not suitable to be fitted by GPD distribution. In case of JPMorgan approximate linearity seems to fit the graph but without any clear starting point. In P&G's profits, linear character is visible but it sharply changes the slope for the most extreme 1% of the sample. In the case of all other stocks, acceptance of the GPD model can be believed. However, in most cases no obvious decision can be made on the threshold value. This gives the conclusion that sample mean excess plot cannot be used as such to decide on the threshold value which would be directly used to compute the Hill's estimate. The next graphical method

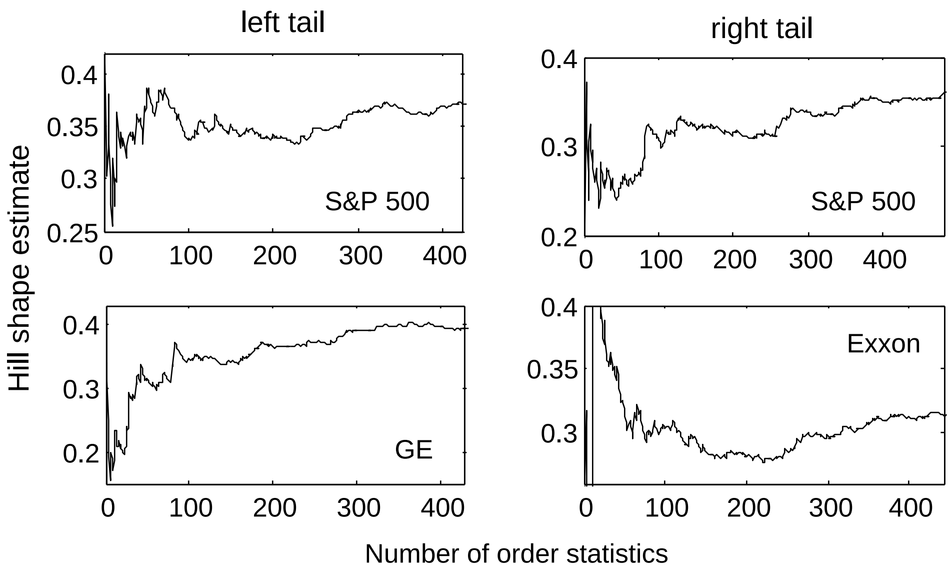
Figure 4.5: S&P 500 Hill plot



Source: Author's computations.

to be used in the extreme value analysis is examination of the Hill plots. In the Figure 4.5 the classical Hill plot of S&P 500 can be seen. In this plot, stable region of the Hill shape estimates is looked for. However, the plot is known to be the most efficient only if the data comes from Pareto or close to Pareto distribution. Here the graph does not show any extensive stable area of estimates and therefore the data is not distributed exactly as Pareto. The same holds also for the other stocks and so the graph will be zoomed in the upper 10% of the particular samples to look for stability there. The classical Hill plots

Figure 4.6: Hill plot for the upper part of the sample



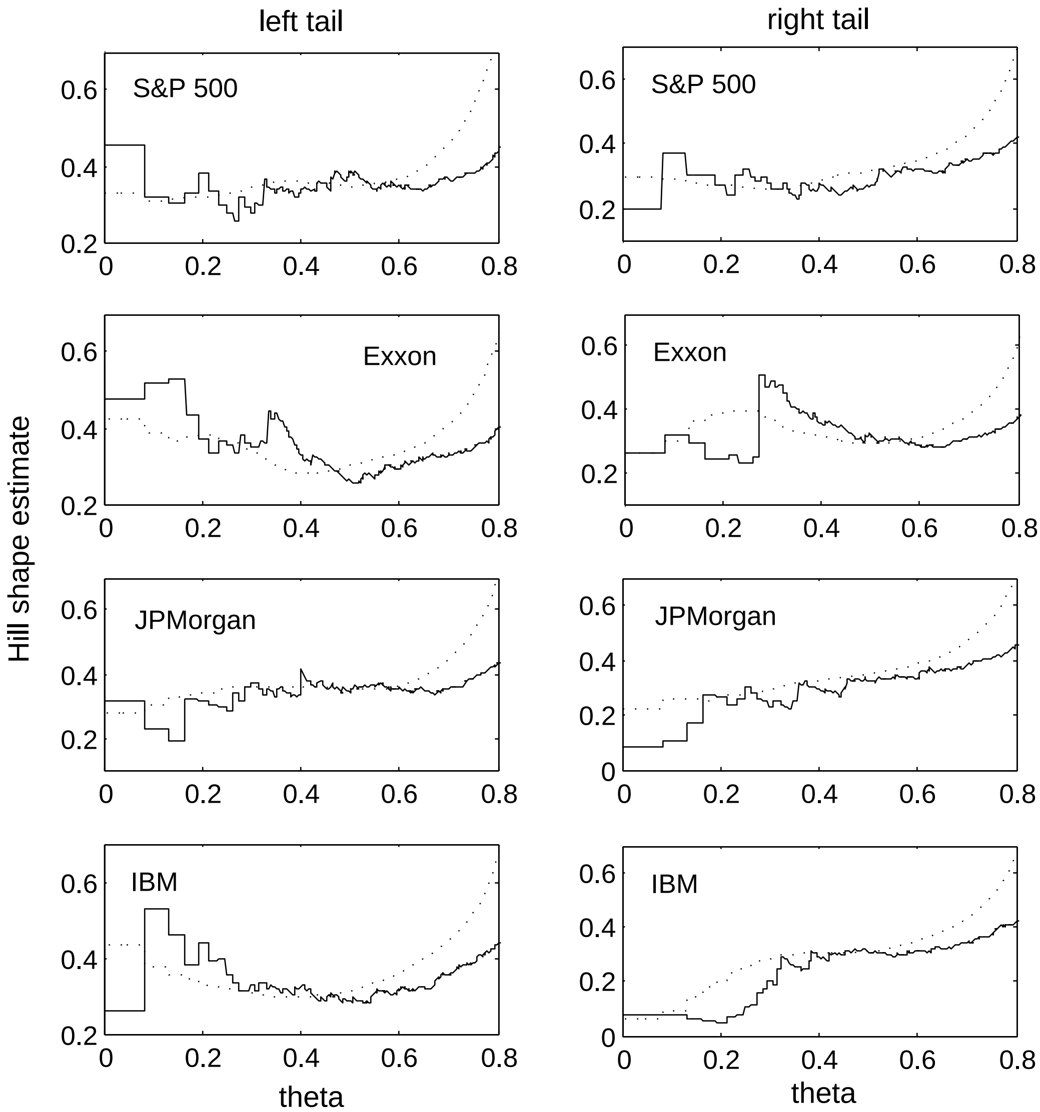
Source: Author's computations.

in Figure 4.6 are restricted only for the upper part of the whole sample because the stable area and so the threshold is assumed to lie there. In the example of S&P 500 this zoomed plot gives much better picture. At some places the estimates seem to gain stability. For the losses, an estimate of about 0.34 is suggested by the graph, using 4% of the sample statistics. For the right tail a stable area could indicate an estimate around 0.32 but still it is not that clear because the volatility of the plot is not adequately tamed.

Next, the alternative Hill plot will be presented. Unlike the zoomed graphs above, it does not neglect certain part of the sample. It does, however, give a large portion of the display space to the hill estimates that are based on quite a small number of upper order statistics. Moreover, these estimates are

Figure 4.7: Alternative and smoothed alternative Hill plots

The graph shows alternative Hill plot with θ on the x axis which uses N^θ number of order statistics for computing the estimate (the solid line). The dashed line represents the smoothed alternative Hill plot (see Equation 3.18.)



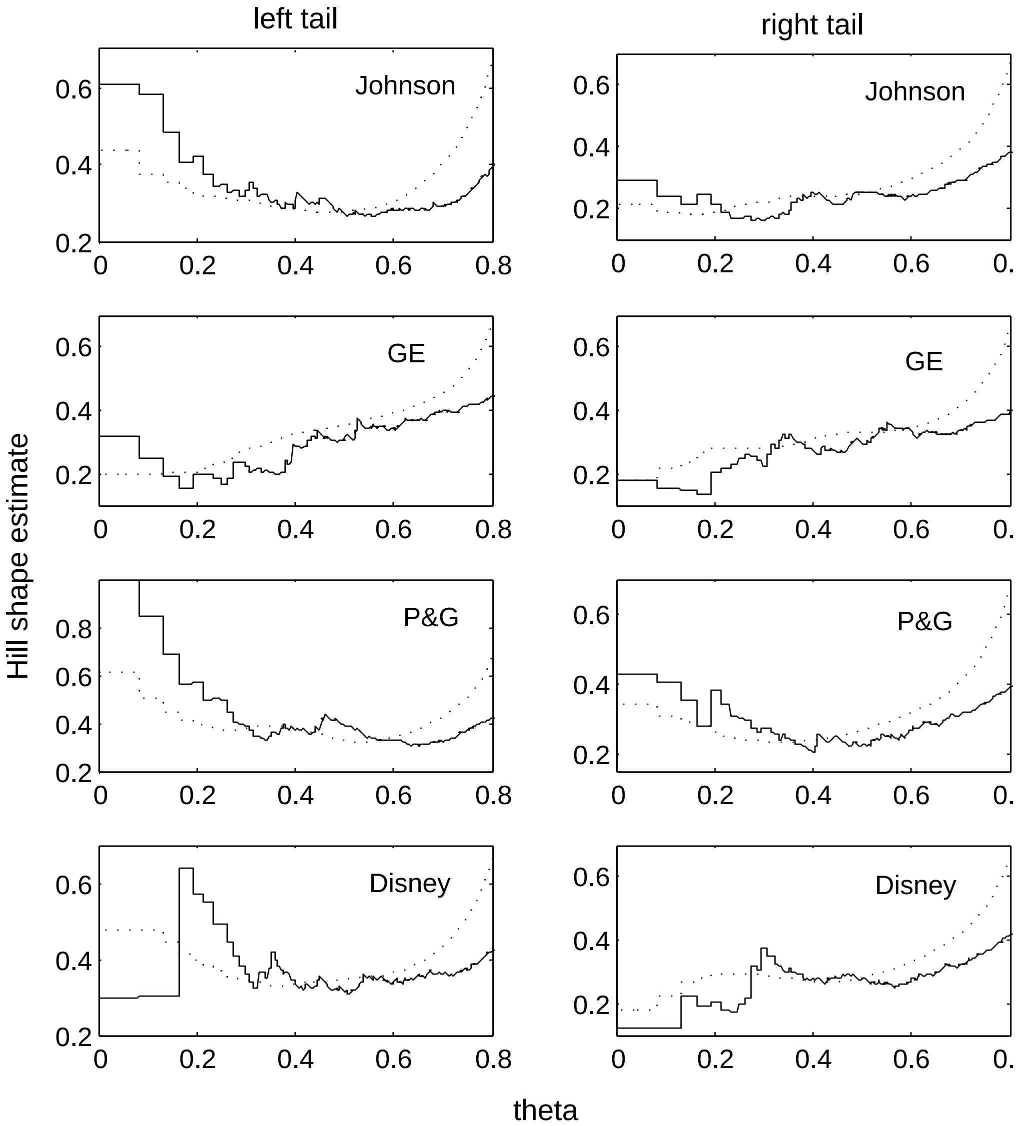
Source: Author's computations.

smoothed as described in the theoretical part (Equation 3.18). The plots can be seen in Figure 4.7 and Figure 4.8.

The alternative Hill plots appear rather volatile too, especially on the left-hand side. This part of graph is not relevant here because it represents estimates computed from less than about 10 upper order statistics. A reasonable area to look for stability is for θ between 0.3 and 0.6 as it approximately matches the top $3 - 5\%$ of the sample. The sharp rise of the plot in IBM and Disney

Figure 4.8: Alternative and smoothed alternative Hill plots cont.

The graph shows alternative Hill plot with θ on the x axis which uses N^{θ} number of order statistics for computing the estimate (the solid line). The dashed line represents the smoothed alternative Hill plot (see Equation 3.18.)



Source: Author's computations.

losses or Exxon profits can be explained by a minor discontinuity in extreme values (a sudden jump) as can be seen in the Q-Q plots in Figure 4.3. The alternative Hill plots also show the smoothed estimates.[4] The smoothing shows stable regions for most of the stocks. It performs well for S&P 500, JPMorgan

[4]For the smoothing $u = 4$ was used (see Equation 3.18) as it is just between $n^{0.1}$ and $n^{0.2}$ for all the samples, as suggested by Resnick and Stărică (1997).

losses, IBM, Johnson, P&G and Disney. It proves to be especially useful in case of Exxon losses, where it reveals a stable region. That would be hidden by the alternative Hill plot and also not clear form the classical plot. In Exxon profits stability arises after the initial jump is neglected. This also provides much more information compared to the classical plot in Figure 4.6.

On the other hand, the smoothing for JPMorgan profits as well as for GE overall does not seem to serve its purpose. It is gradually increasing with no steady area. Therefore, in these cases the attention needs to be drawn back to the classical or simple non-smoothed alternative Hill plots. Here again, the volatility of the graph causes troubles and the decision would have to be based on subjectivity. To sum up what has been learned from the graphical threshold selection techniques, the main findings could be pointed out. The smoothing procedure might bring desired stable areas. Still in some cases, it proved to be highly insufficient and the threshold would have to be chosen from the original volatile plots. Finally, even when the smoothing provides stable areas, they still often include a slight decline or increase and therefore cannot determine an estimate with satisfactory precision.

As the graphical methods did not prove to be sufficient for threshold selection, the Hill method modified by Huisman is used. Here the estimates do not require any choice on where the tail begins which makes the procedure automatic. The resulting estimates are presented in Table 4.6 and Table 4.7. The estimates are reported for the shape parameter with the appropriate standard errors computed also by the Huisman's method. Then the tail index estimate is reported as it is simply the reciprocal of the shape parameter. The results are given for both the left tail (losses) and the right tail (profits) as well as for the full sample length and three shorter periods.[5]

It can be directly seen that the estimated values of the tail index are positive which indicates heavy-tailed distributions. The estimate of the shape parameter, whose higher value means fatter tails, is for the full sample the highest for General Electric and IBM losses and JPMorgan profits with values 0.31, 0.302 and 0.299, respectively. For all the stocks the shape parameter estimate is the highest in the 3rd period for both the left and the right tail. This confirms that it was the most volatile period as was already signified by the descriptive statistics. The heaviest tails from all the analysed stocks were detected in the 3rd period in JPMorgan's returns. The estimated shape parameter is 0.386

[5]The full sample covers period 3/1980 - 2/2016, period 1 is 3/1980 - 2/1992, period 2 is 3/1992 - 2/2004 and period 3 is 3/2004 - 2/2016.

Table 4.6: Shape parameter and tail index estimates obtained by Huisman's method

	stock	left tail		right tail	
		$\hat{\xi}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\alpha}$
full sample	S&P 500	0.2677	3.7360	0.2524	3.9615
		(0.0856)		(0.0655)	
period 1		0.2527	3.9572	0.1966	5.0864
		(0.0505)		(0.0406)	
period 2		0.2044	4.8924	0.2481	4.0300
		(0.0861)		(0.0818)	
period 3		0.3067	3.2604	0.3118	3.2073
		(0.0813)		(0.0667)	
full sample	Exxon	0.2511	3.9820	0.2324	4.3031
		(0.0355)		(0.0942)	
period 1		0.2620	3.8171	0.2032	4.9206
		(0.06490		(0.0835)	
period 2		0.2027	4.9332	0.2243	4.4593
		(0.0574)		(0.0860)	
period 3		0.3071	3.2562	0.2712	3.6872
		(0.0597)		(0.0807)	
full sample	JPMorgan	0.2755	3.6301	0.2991	3.3438
		(0.0291)		(0.0299)	
period 1		0.2377	4.2066	0.2538	3.9399
		(0.0886)		(0.0328)	
period 2		0.2039	4.9040	0.2023	4.9432
		(0.0448)		(0.0245)	
period 3		0.3860	2.5906	0.4396	2.2746
		(0.0479)		(0.0349)	
full sample	IBM	0.3021	3.3099	0.2688	3.7197
		(0.0289)		(0.0285)	
period 1		0.2501	3.9990	0.1904	5.2519
		(0.0502)		(0.0409)	
period 2		0.2593	3.8558	0.2536	3.9435
		(0.0615)		(0.0495)	
period 3		0.3341	2.9928	0.2916	3.4294
		(0.0463)		(0.0389)	
full sample	Johnson	0.2193	4.5606	0.1911	5.2324
		(0.0742)		(0.0302)	
period 1		0.2282	4.3825	0.2211	4.5222
		(0.1004)		(0.0533)	
period 2		0.1917	5.2152	0.1651	6.0585
		(0.0729)		(0.0314)	
period 3		0.2592	3.8580	0.3006	3.3272
		(0.0463)		(0.0389)	

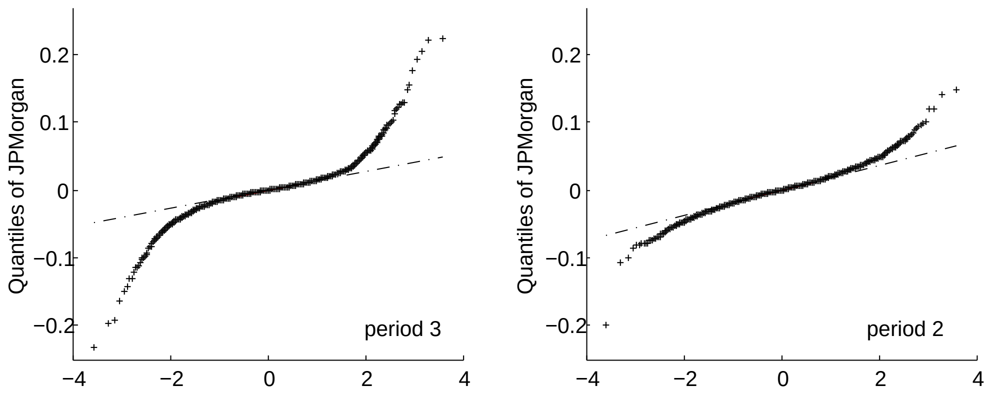
Table 4.7: Shape parameter and tail index estimates obtained by Huisman's method cont.

	stock	left tail		right tail	
		$\hat{\xi}$	$\hat{\alpha}$	shape $\hat{\xi}$	$\hat{\alpha}$
full sample	General Electric	0.3100	3.2255	0.2579	3.8772
		(0.0737)		(0.0285)	
period 1		0.2395	4.1755	0.2009	4.9772
		(0.0880)		(0.0857)	
period 2		0.2603	3.8422	0.2111	4.7377
		(0.0946)		(0.0577)	
period 3		0.4067	2.4591	0.3649	2.7405
		(0.0476)		(0.0598)	
full sample	P&G	0.2704	3.6985	0.2240	4.4643
		(0.0595)		(0.0377)	
period 1		0.2435	4.1067	0.1921	5.2064
		(0.0909)		(0.0623)	
period 2		0.2897	3.4523	0.2295	4.3580
		(0.0996)		(0.0428)	
period 3		0.3054	3.2744	0.2390	4.1838
		(0.0954)		(0.0842)	
full sample	Disney	0.2862	3.4938	0.2558	3.9089
		(0.0210)		(0.0118)	
period 1		0.3054	3.2745	0.2145	4.6615
		(0.0520)		(0.0163)	
period 2		0.2413	4.1447	0.2269	4.4080
		(0.0305)		(0.0139)	
period 3		0.3345	2.9895	0.2987	3.3482
		(0.0507)		(0.0308)	

for losses and even 0.44 for profits. This is also in line with the high sample standard deviation ($sd_{per3} = 0.0255$). The Figure 4.9 shows a noteworthy difference between the Q-Q plots of JPMorgan's returns in the 3rd versus the 2nd period. While in the first mentioned the observations spread out far away from the normal case, in period 2 the picture is not that striking and explains the lower shape parameter estimates. Apart from JPMorgan, considerably fat tail was also found in General Electric losses within the 3rd period. The thickness of the tails of the whole index S&P 500 seems to be smaller than for individual stocks. The only exceptions are Exxon Mobile and Johnson for which the tails seem to be lighter than is the case of the full index.

The next result is that the left tails are heavier than the right tails which holds for almost all the stocks. This is also supported by the Q-Q plots, which divert more on the side of losses and the sample mean excess functions where

Figure 4.9: Q-Q plots against the Normal distribution for JPMorgan returns in the 3rd and 2nd period



Source: Author's computations.

the mean excesses were generally smaller for the right tail, with the exception of JPMorgan. The shape estimates of JPMorgan are also an exception here because the right tail is estimated to be heavier than the left tail ($\hat{\xi}_{losses} = 0.28$ and $\hat{\xi}_{profits} = 0.3$). Interestingly, this holds also for the S&P 500 index in the last two periods.

Generally, the concern about the thickness of the left tail is quite intuitive as it implies the downside risk and therefore a direct decline in value of a portfolio. This is the case of long position on a portfolio when value is gained when the stock price goes up and on the contrary a price decrease represents the danger. On the other hand there is the case of derivatives. Their value depends on performance of a certain underlying asset. This relationship can of course be also negative so a speculator can expect a decline in value of the underlying asset and thus a benefit from the derivative. Such an example could be a forward contract where the seller anticipates a fall in price of the asset and thus assumes a short position. Distinguishing between the downside and upside risk is on of the major drawbacks of the Capital Asset Pricing Model (CAPM), which together with the modern portfolio theory bases the risk measure on standard deviation of asset returns. For the reasons stated above, it seems essential to analyse both the minima and maxima and assess the downside and upside risk separately. As the results of this study suggest, some stocks might be more suitable for long position while other stocks are more inclined to benefit from the short position and speculative behaviour.

Another important judgement about the underlying distribution can be made from the tail index estimate. From the estimated value $\hat{\alpha}$, one can directly infer the number of finite moments. All the estimates in this thesis correspond to finite variance models. Mostly, the series are characterized by three finite moments. In case of Johnson the underlying distribution seems to have four finite moments. This finding is again an evidence against the Normal distribution being the underlying distribution function. In that case, the number of moments would tend to infinity and the shape parameter of zero would mean exponential rather than polynomial tail decay.

As far as the properties of these estimators are concerned, their consistency is ensured even though the series were weakly dependent. If the results from Huisman's methods are compared with the estimates suggested by stable regions in the alternative smoothed Hill plots, there are substantial discrepancies. This supports the belief that the original Hill estimates are afflicted by bias. To give examples, the shape estimate in the plots appears to be around 0.35 versus the Huisman's 0.26 for the left tail of S&P 500, and similarly 0.38 versus 0.27 for JPMorgan. For the profits, the plots indicate a value about 0.26 for Johnson versus 0.19 by Huisman. In the same sense higher values can be found for all the stocks for which the plots give reasonable stable areas (i.e. not in GE and JPMorgan's profits). This fully corresponds with Huisman's results where the authors contend that the classical Hill estimates are upward biased and that their method appropriately curbs the bias.

The positive estimates obtained by the Hill's method modified by Huisman express that the underlying distribution has heavy tails. Yet this is a result that cannot be tested because fat tails ($\xi > 0$) are assumed beforehand in the Hill's method. This is, however, believed to be a good assumption since the descriptive statistics, histograms and QQ plots suggest heavy tails as well. As far as other applications of the tail index are concerned, the semi-parametric nature of the estimation makes it possible to construct the estimate for the tail or the corresponding p-quantile. Nevertheless, these estimates (as well as VaR) will not be obtained in this thesis as they require a direct choice of the threshold which is then used for the estimate computation.

As described in the theoretical part, the shape estimates can also serve to determine the limit ratio of Expected Shortfall and Value at Risk (see Equation 3.30). The computed values using the shape estimates are reported in Table 4.8. All the values here are greater than one, ranging between 1.2 and 1.78. It means that in the high quantiles close to one, the ES is $1.2-1.78$ higher than the VaR. This stresses the importance of not considering only VaR itself

Table 4.8: Limit ratio of ES to VaR: $\lim_{p \to 1} \frac{ES_p}{VaR_p}$

	stock	left tail	right tail	stock	left tail	right tail
full sample	S&P 500	1.3655	1.3377	Johnson	1.2809	1.2363
period 1		1.3382	1.2447		1.2956	1.2839
period 2		1.2569	1.3300		1.2372	1.1977
period 3		1.4424	1.4530		1.3499	1.4297
full sample	Exxon	1.3353	1.3027	GE	1.4493	1.3476
period 1		1.3550	1.2551		1.3149	1.2514
period 2		1.2542	1.2891		1.3518	1.2675
period 3		1.4432	1.3721		1.6854	1.5745
full sample	JPM	1.3802	1.4267	P&G	1.3706	1.2887
period 1		1.3119	1.3401		1.3219	1.2377
period 2		1.2561	1.2536		1.4078	1.2978
period 3		1.6287	1.7846		1.4397	1.3141
full sample	IBM	1.4329	1.3677	Disney	1.4010	1.3438
period 1		1.3334	1.2352		1.4397	1.2731
period 2		1.3502	1.3397		1.3180	1.2934
period 3		1.5018	1.4116		1.5026	1.4259

as it states only the minimum loss. The Expected Shortfall appears to be a more informative measure because it describes the expected loss given a certain quantile. Here the results show that it is relevant to take ES into account. For the whole index S&P 500 the expected losses over the full period are in fact at the far tail 36.55% higher than the minimum losses. In case of GE in the 3rd period, the expected losses are even 68.54% higher than the minimum losses indicated by VaR.

Regarding lower quantiles, the shape parameter cannot be used to express the relationship between ES and VaR. As mentioned above, the EVT will not be used to compute VaR nor ES here because of the threshold selection problem. However, a comparison of classical VaRs is presented in Table 4.9. Firstly, one-day Value at Risk is computed based on the so called historical simulation. That means the quantiles are derived from the empirical distribution function so the approach is completely non-parametric.

Next, VaR is given assuming the underlying distribution is Normal and the mean and variance are used according to the series' descriptive statistics. As the Normal distribution is symmetric, the values with positive sign apply to the right tail and values with a negative sign would correspond to the left tail. It can be said that the values of the Normal distribution roughly comply with the values of the empirical VaR until the 97.5% quantile. For the higher quantiles, it is clear that the normal case underestimates the extreme negative and positive returns.

Finally, the VaR is computed based on the assumption that the data is t-distributed where the computed tail index α is used for the number of degrees of freedom as it expresses the number of finite moments. Even though the Student's t distribution is symmetric around zero, the values are presented here separately because of a different estimated tail index for both cases. The values show that the VaR here is closer to the empirical values compared to the normal case but still it underpredicts the extremes in the right tail and in most cases also in the left tail. In case of P&G the values of VaR based on t-distribution are fairly close to the empirical ones and therefore the underlying distribution could behave as Student's t in the tails (although with different left and right tail thickness). For the other stocks the t-distribution does not seem to be a good fit, especially JPMorgan's high quantiles are greatly underestimated. On the other hand, S&P's high quantiles appear to be overestimated when assuming the t-distribution and so the tails of S&P are probably lighter than Student t distributed. Overall, the series seem to follow a distribution with thicker tail than Student t with the exception of P&G for which it seems approximately a good fit. On the other hand S&P 500 follows a heavy-tailed distribution, yet with lighter tail characteristics than those of Student t distribution. These judgements were, of course, made with the assumption that the empirical quantiles are a good representation of the underlying distribution. In fact, this might not be the case as the empirical quantiles could differ from the ones of the underlying distribution.

Table 4.9: Value at Risk

	S&P	Exxon	JPM	IBM	Johnson	GE	P&G	Disney
quantile	Empirical VaR for the left tail							
0.95	-0.0168	-0.0224	-0.0335	-0.0249	-0.0226	-0.0247	-0.0209	-0.0278
0.975	-0.0224	-0.0286	-0.0435	-0.0334	-0.0285	-0.0333	-0.0273	-0.0362
0.99	-0.0301	-0.0393	-0.0601	-0.0462	-0.0371	-0.0466	-0.0351	-0.0506
0.995	-0.0391	-0.0463	-0.0768	-0.0566	-0.0435	-0.0600	-0.0442	-0.0626
0.999	-0.0701	-0.0790	-0.1455	-0.0866	-0.0713	-0.1010	-0.0766	-0.1023
quantile	Empirical VaR for the right tail							
0.95	0.0166	0.0226	0.0344	0.0258	0.0240	0.0262	0.0226	0.0301
0.975	0.0219	0.0290	0.0465	0.0339	0.0308	0.0340	0.0290	0.0397
0.99	0.0291	0.0371	0.0661	0.0461	0.0390	0.0464	0.0388	0.0523
0.995	0.0379	0.0446	0.0852	0.0576	0.0478	0.0605	0.0451	0.0623
0.999	0.0557	0.0924	0.1403	0.1045	0.0692	0.0988	0.0651	0.1111
quantile	VaR for Normal distribution							
0.95	0.0182	0.0244	0.0380	0.0278	0.0237	0.0280	0.0240	0.0315
0.975	0.0218	0.0291	0.0453	0.0331	0.0284	0.0335	0.0287	0.0376
0.99	0.0259	0.0346	0.0538	0.0394	0.0338	0.0398	0.0341	0.0448
0.995	0.0287	0.0384	0.0597	0.0436	0.0374	0.0441	0.0378	0.0496
0.999	0.0345	0.0461	0.0716	0.0524	0.0450	0.0530	0.0455	0.0596
quantile	VaR for Student t distribution for the left tail							
0.95	-0.0218	-0.0213	-0.0220	-0.0227	-0.0206	-0.0229	-0.0218	-0.0222
0.975	-0.0286	-0.0278	-0.0289	-0.0302	-0.0265	-0.0306	-0.0287	-0.0294
0.99	-0.0390	-0.0376	-0.0397	-0.0422	-0.0350	-0.0430	-0.0392	-0.0407
0.995	-0.0483	-0.0462	-0.0494	-0.0533	-0.0424	-0.0545	-0.0487	-0.0509
0.999	-0.0771	-0.0721	-0.0797	-0.0892	-0.0635	-0.0923	-0.0780	-0.0833
quantile	VaR for Student t distribution for the right tail							
0.95	0.0213	0.0208	0.0225	0.0218	0.0199	0.0215	0.0207	0.0215
0.975	0.0278	0.0270	0.0301	0.0286	0.0254	0.0281	0.0267	0.0280
0.99	0.0376	0.0360	0.0419	0.0391	0.0330	0.0381	0.0354	0.0379
0.995	0.0463	0.0439	0.0529	0.0485	0.0394	0.0470	0.0429	0.0468
0.999	0.0724	0.0669	0.0881	0.0774	0.0570	0.0740	0.0647	0.0734

Chapter 5

Conclusion

The Extreme Value Theory offers a useful tool for modelling the distribution of extreme values. The theory provides methods to model the tails specifically without paying attention to the center of the distribution. This approach incorporates as much information about the tails as possible and makes the best use of the small number of extremes previously observed. This thesis was focused on financial returns of particular stocks and so it dealt with risk management in a univariate distribution framework.

To identify the extremes in an observed sample, the method of exceedances over a certain threshold was chosen over selecting maxima from time blocks. Therefore the statistical inference was related to the Generalized Pareto Distribution. To fit the distribution model, a semi-parametric approach was used. Here the main parameter of interest was the tail index and its reciprocal value (the shape parameter) because it accurately characterizes the tail behaviour. The higher the shape parameter, the heavier the tail. Hill's method was used for the parameter estimation as it is well tried in practice. It was also chosen because of its general use and computational simplicity.

In the empirical analysis, financial returns of S&P 500 index and its seven constituents were examined over a period of 36 years beginning in 1980. The stationarity of the data complied with the basic assumption of the theory and even though the data was weakly dependent, the method still yielded consistent estimates. The initial data analysis provided a sound base for the assumption of a fat-tailed distribution as it was indicated by the Q-Q plots, histograms and sample leptokurtosis.

In this theory, the segregation of extreme values from the central part of the distribution is a crucial step. As part of the Extreme Value Theory analysis, several graphical methods were applied to choose an appropriate points where the tails begin. These methods represent empirical measures of model stability. In the mean excess plot, an upward linearity was searched for because it indicates a reasonable model fit. However, these plots did not serve well in all cases to reveal for which part of data the Pareto model would be suitable. Then the Hill plots were utilized and a stable region was looked for as it should show the tail threshold. The classical Hill plots proved to be inefficient and so they were zoomed in because the upper $3-5\%$ of the sample usually fit the extreme value model. This step was also ineffective in mitigating the volatility of the plot. As well as that, the graphs were still rather unstable when an alternative plot with rescaled horizontal axis was used and so a smoothing procedure was carried out. These smoothed graphs mostly revealed stable regions from which an appropriate Hill's shape parameter estimate could be deduced. However, some of them revealed no stable region thus making the decision impossible. Overall, the graphical methods were obviously dependent on a subjective judgement regardless of whether they provided stable regions.

As the choice of the threshold form graphical methods was assessed to be troublesome, a Hill method modified by Huisman was utilized. No selection of the tail cut-off point was necessary there as it is a regression-based technique which is also believed to correct for the Hill's estimator small sample bias. The resulting shape parameter estimates are in line with Huisman's results as they prove that the classical Hill's estimates chosen from the Hill plots are upward-biased and thus overestimate the tail thickness. Most importantly, all the estimate values are positive and therefore indicate that the data come from a Pareto type distribution with a polynomial tail decay. The tail index also directly expresses the number of finite moments of the underlying distribution. It has been shown that all the analysed series are from finite variance models and even have three or four finite moments.

For all the stocks, both the left tail (minima) and the right tail (maxima) were modelled. This is due to the fact that not only extreme losses can cause trouble. Extreme positive returns should put investors on alert in case when they use derivatives for speculations and have a short position on portfolio. This gives reason why extreme profits also came under scrutiny in this thesis. The estimation is performed for the full sample as well as for three subperiods to see whether the tail behaviour changes over time. The left tail proved to be overall heavier in all the stocks except for JPMorgan Chase returns. The results state that the 3rd period analysed (March 2004 - February 2016) was the one with heaviest tails, which corresponds to a high volatility in this period.

The shape parameter also determines a limit ratio of the Expected Shortfall and Value at Risk in high quantiles close to one. Therefore the estimated parameters served to stress the importance of computing an expected loss at a certain quantile instead of only the minimum loss expressed by Value at Risk. Value at Risk comparison clearly shows that the Normal distribution underestimates the high quantiles and is not a suitable assumption for the underlying distribution.

In this thesis no tail estimate was formed as the subjectivity of the threshold selection was rather avoided. In the future research it is suggested to use a more refined technique to choose the threshold such as by bootstrapping. This will enable a proper tail and quantile estimation which will serve to forecast the risk in next periods. Also, an advance to multivariate framework would be advisable in order to examine the componentwise maxima by a multivariate extreme value model.

Bibliography

Andersen, T. G., Davis, R. A., Kreiss, J.-P., and Mikosch, T. V. (2009). *Handbook of financial time series*. Berlin: Springer. ISBN 978-3-540-71296-1.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3): pp. 203–228. doi: 10.1111/1467-9965.00068.

Balkema, A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, **2**(5): pp. 792–804. Available at http://www.jstor.org/stable/2959306.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. London: Springer. ISBN: 978-1-84996-874-4.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, **76**(2): pp. 226–248. doi:10.1006/jmva.2000.1903.

Danielsson, J. and De Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales of Economics and Statistics*, 60: pp. 239–270. Available at http://www.jstor.org/stable/20076262.

Danielsson, J., Jorgensen, B. N., Mandira, S., Samorodnitsky, G., and De Vries, C. G. (2005). Subadditivity re–examined: the case for value-at-risk. *Financial Markets Group, London School of Economics and Political Science*. Available at https://ideas.repec.org/a/eee/insuma/v44y2009i2p143-145.html.

De Haan, L. (1994). Extreme value statistics. *Extreme Value Theory and Applications*, 1: p. 93. doi: 10.1007/978-1-4613-3638-9$_6$.

De Haan, P. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, **52**(1): pp. 60–70. doi: 10.1111/1467-9574.00068.

Dekkers, A. L., Einmahl, J. H., and De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, **17**(4): pp. 1833–1855. Available at http://www.jstor.org/stable/2241667.

Drees, H., De Haan, L., and Resnick, S. (2000). How to make a hill plot. *Annals of Statistics*, **28**(1): pp. 254–274. Available at http://www.jstor.org/stable/2673989.

DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics*, **11**(4): pp. 1019–1031. Available at http://www.jstor.org/stable/2241294.

Embrechts, P., Klueppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Berlin [etc.]: Springer. doi: 10.1007/978-3-642-33483-2.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**(2): pp. 180–190. http://dx.doi.org/10.1017/S0305004100015681.

Fréchet, M. (1927). Sur les ensembles compacts de fonctions mesurables. *Fundamenta Mathematicae*, **9**(1): pp. 25–32. Available at http://eudml.org/doc/215135.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44**(3): pp. 423–453. doi: 10.2307/1968974.

Gomes, I. M., De Haan, L., and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**(1): pp. 31–52. doi: 10.1111/j.1467-9868.2007.00620.x.

Greene, W. H. (2003). *Econometric analysis*. Upper Saddle River, N.J.: Prentice Hall. ISBN 0-13-066189-9.

Gumbel, E. (1958). *Statistics of extremes*. New York: Columbia University Press. ISBN 0-486-43604-7.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44: pp. 37–42. Available at http://www.jstor.org/stable/2984706.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, **32**(2): pp. 177–203. doi:10.1016/0047-259X(90)90080-2.

Hill (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**(5): pp. 1163–1174. doi:10.1214/aos/1176343247.

Ho, L.-C., Burridge, P., Cadle, J., and Theobald, M. (2000). Value-at-risk: Applying the extreme value approach to asian markets in the recent financial turmoil. *Pacific-Basin Finance Journal*, **8**(2): pp. 249–275. doi:10.1016/S0927-538X(00)00008-1.

Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics*, **19**(3): pp. 1547–1569. Available at http://www.jstor.org/stable/2241962.

Huisman, R., Koedijk, K. G., Kool, C. J. M., and Palm, F. (2001). Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, **19**(2): pp. 208–216. doi: 10.1198/073500101316970421.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial Analysts Journal*, **52**(6): pp. 47–56. http://dx.doi.org/10.2469/faj.v52.n6.2039.

Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking and Finance*, **24**(7): pp. 1097–1130. doi: 10.1016/S0378-4266(99)00077-1.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7**(1): pp. 77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management.* Princeton, N.J.: Princeton University Press. ISBN 0-691-12255-5.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, **7**(3-4): pp. 271–300. doi:10.1016/S0927-5398(00)00012-8.

Mikosch, T. (2003). *Modeling dependence and tails of financial time series*, chapter Extreme Values in Finance, Telecommunications and the Environment, pages pp. 185–286. Chapman and Hall. ISBN: 978-0-203-48335-0.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, **3**(1): pp. 119–131. Available at http://www.jstor.org/stable/2958083.

Resnick, S. and Stărică, C. (1997). Smoothing the hill estimator. *Advances in Applied Probability*, **29**(1): pp. 271–293. doi: 10.2307/1427870.

Rootzén, H. and Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, 1997(1): pp. 70–94. doi: 10.1080/03461238.1997.10413979.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, **4**(4): pp. 367–377. Available at http://www.jstor.org/stable/2245845.

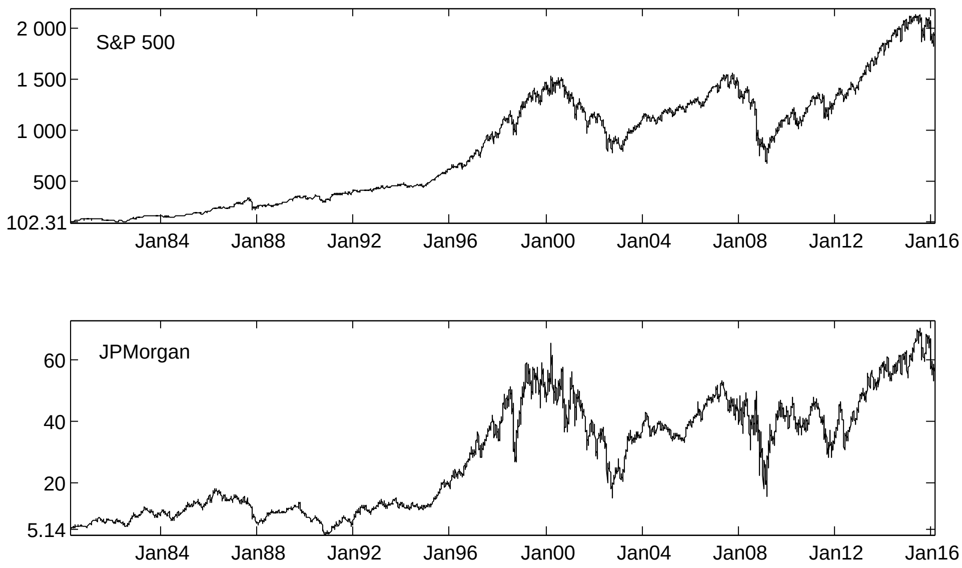
Weibull, W. (1951). A statistical ditribution function of wide applicability. *Journal of Applied Mechanics*, 18: pp. 293–297.

Appendix A

Figures

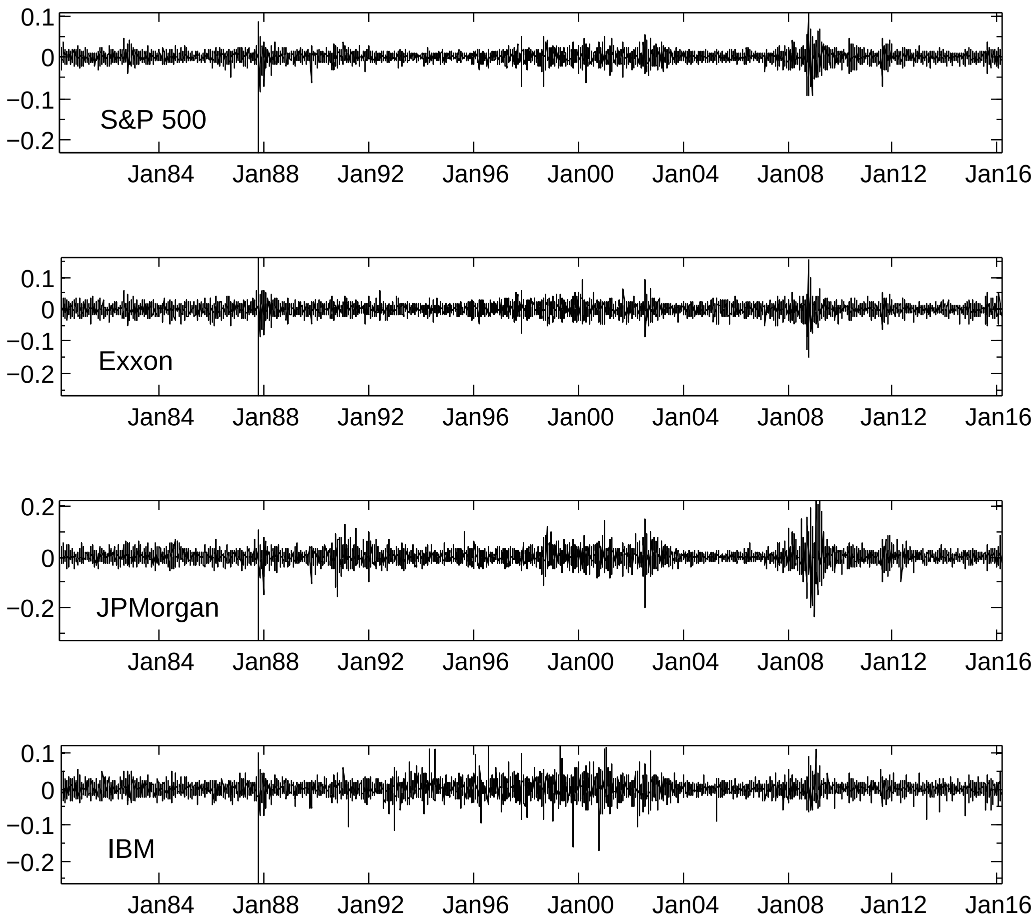
A.1 Data Analysis

Figure A.1: Prices (in USD) for the whole index and selected stock



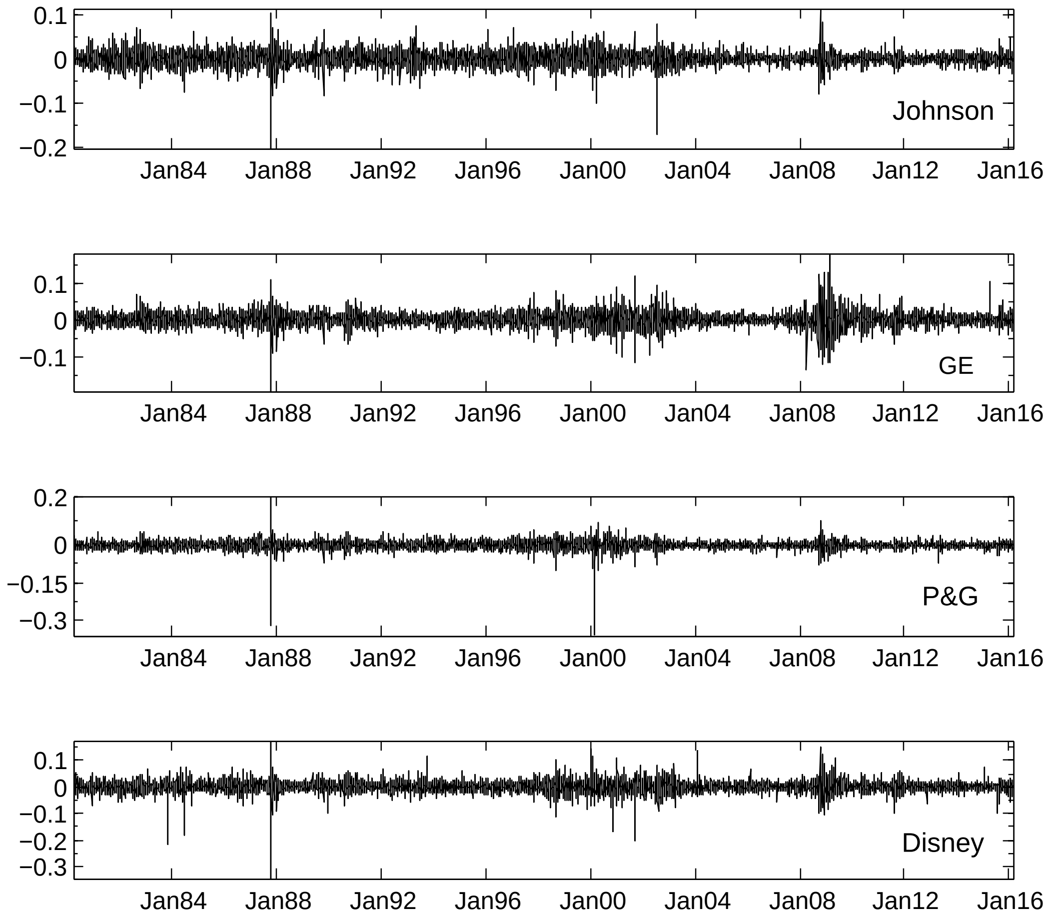
Source: Author's computations.

Figure A.2: Log-returns



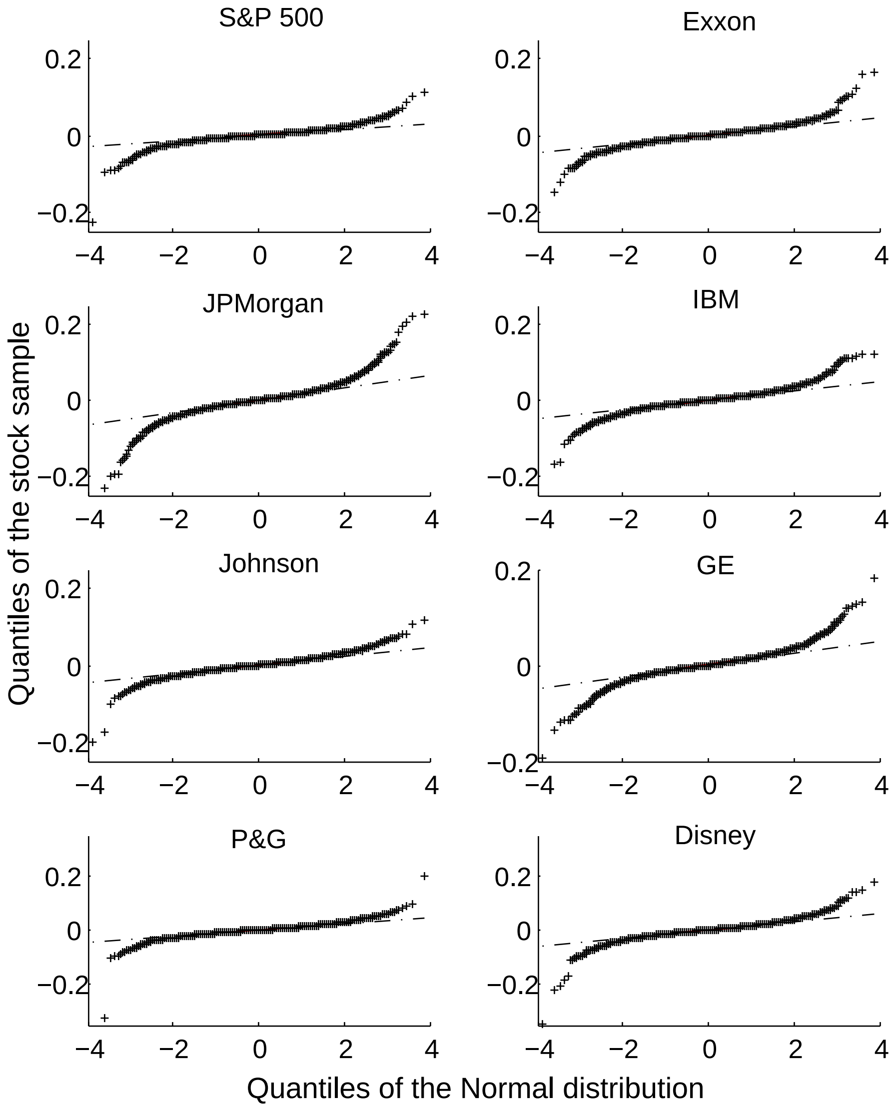
Source: Author's computations.

Figure A.3: Log-returns cont.



Source: Author's computations.

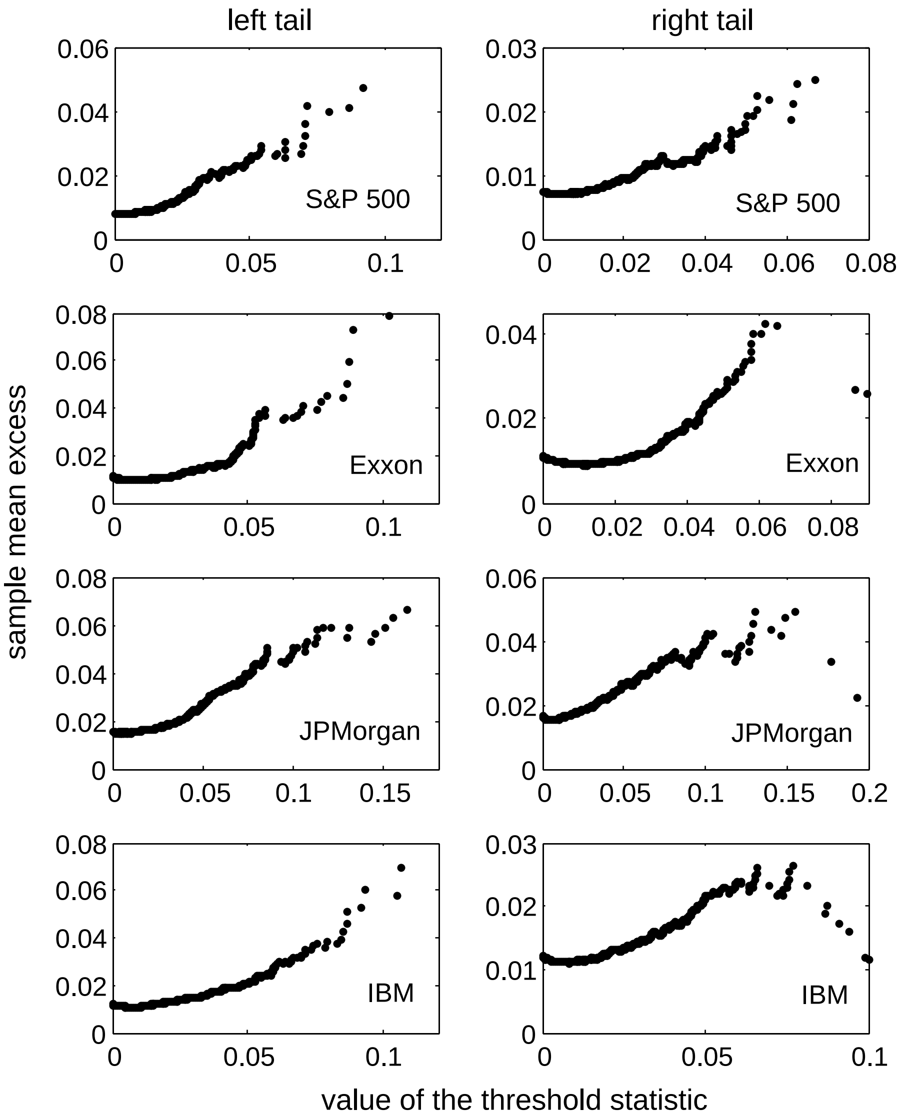
Figure A.4: Q-Q plot of log-returns against the Normal distribution



Source: Author's computations.

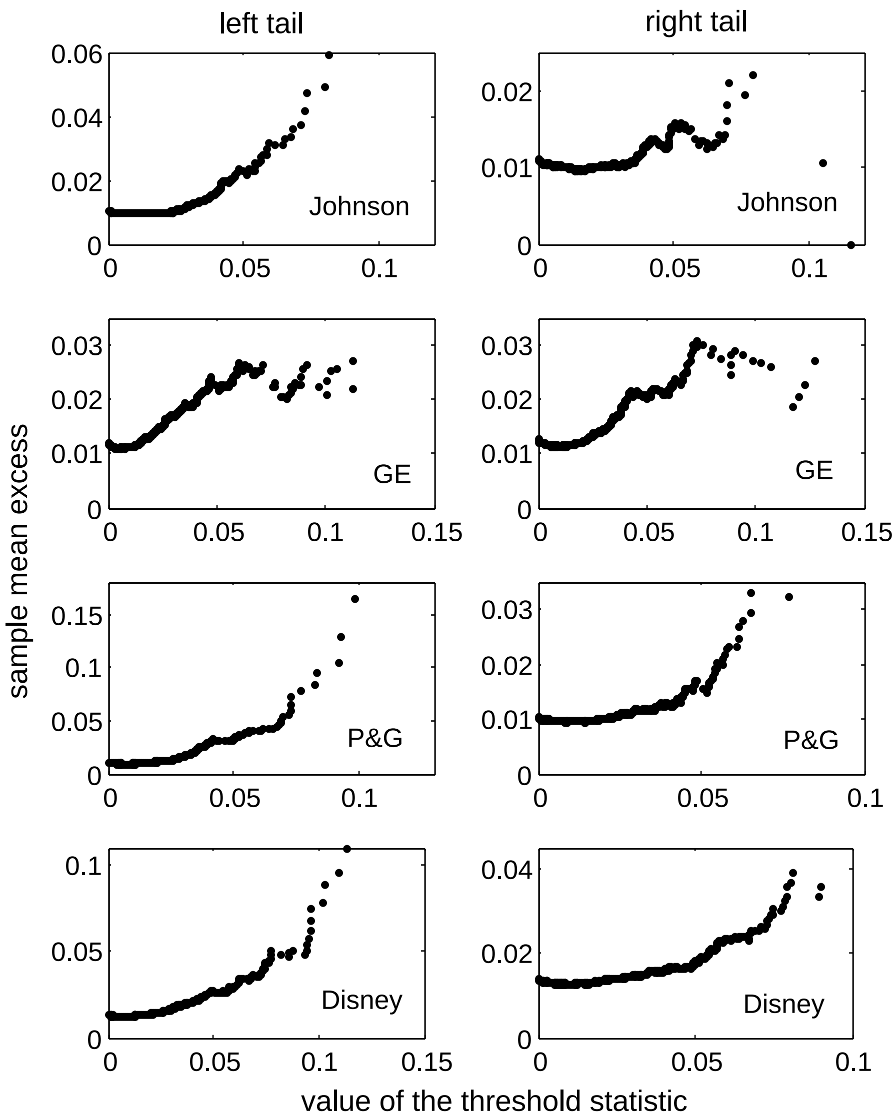
A.2 EVT

Figure A.5: Sample mean excess plot



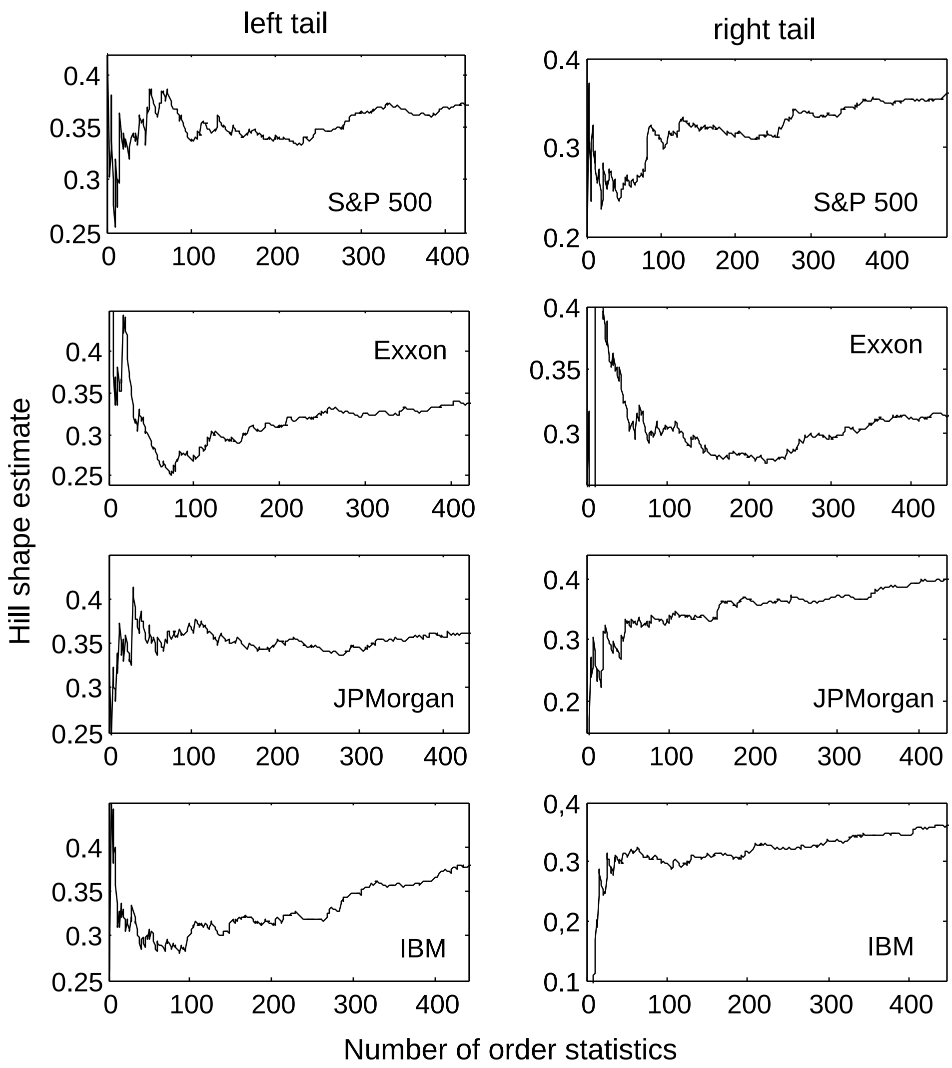
Source: Author's computations.

Figure A.6: Sample mean excess plot cont.



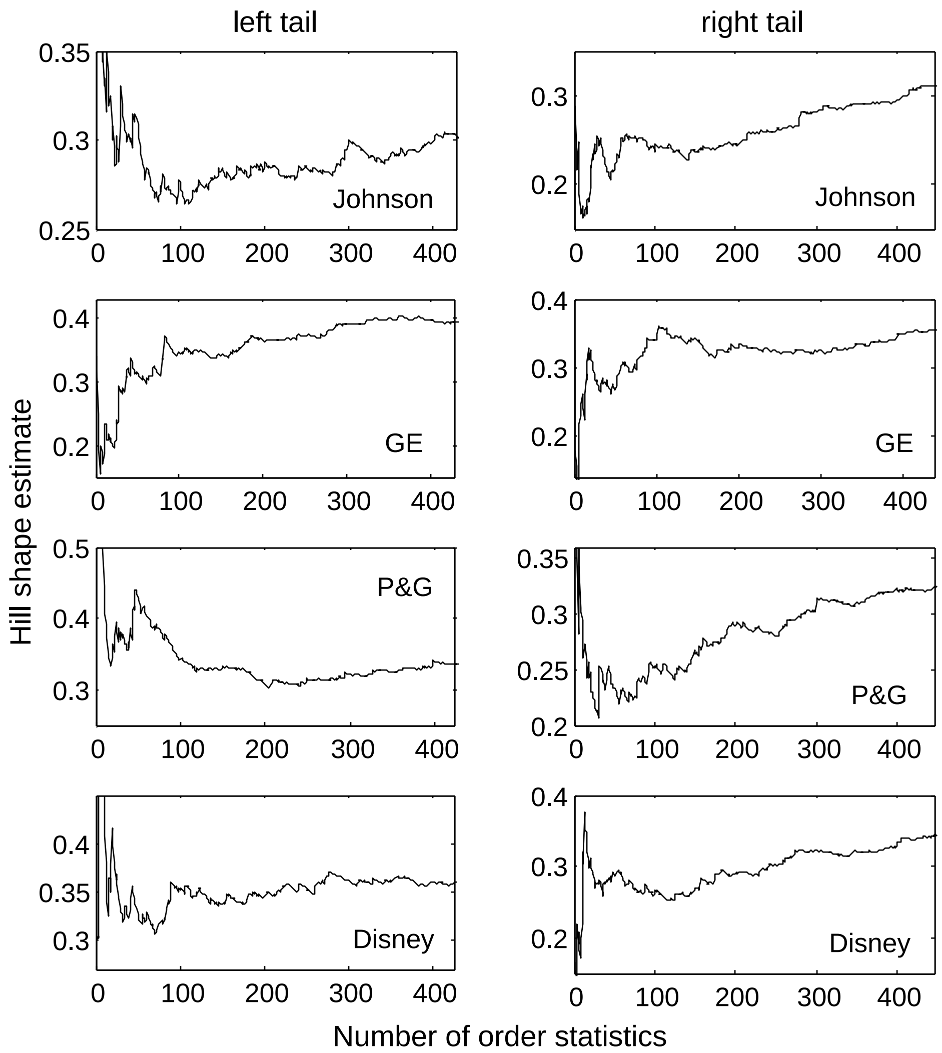
Source: Author's computations.

Figure A.7: Hill plot for the upper part of the sample



Source: Author's computations.

Figure A.8: Hill plot for the upper part of the sample cont.



Source: Author's computations.

Appendix B

Tables

Table B.1: Descriptive statistics of the subsamples

Stock	Max	Min	Mean	Sd	Skewness	Exc. Kurtosis
period 1: 3/1980 - 2/1992						
S&P 500	0.09	-0.23	0.0005	0.0107	-3.41	73.88
Exxon	0.16	-0.27	0.0004	0.0149	-1.51	40.86
JPMorgan	0.13	-0.32	0.0003	0.0210	-1.22	22.73
IBM	0.10	-0.26	0.0001	0.0148	-1.75	34.49
Johnson	0.10	-0.20	0.0007	0.0167	-0.54	9.07
GE	0.11	-0.19	0.0006	0.0155	-0.47	10.39
P&G	0.20	-0.33	0.0006	0.0153	-2.39	78.32
Disney	0.17	-0.34	0.0009	0.0198	-2.25	40.29
period 2: 3/1992 - 2/2004						
S&P 500	0.06	-0.07	0.0003	0.0107	-0.11	3.79
Exxon	0.09	-0.09	0.0004	0.0146	0.05	2.74
JPMorgan	0.15	-0.20	0.0004	0.0231	0.09	4.56
IBM	0.12	-0.17	0.0005	0.0217	0.03	5.49
Johnson	0.08	-0.17	0.0005	0.0164	-0.33	5.49
GE	0.12	-0.11	0.0005	0.0177	0.04	3.62
P&G	0.09	-0.36	0.0005	0.0175	-2.99	61.50
Disney	0.14	-0.20	0.0002	0.0215	-0.13	6.82
period 3: 3/2004 - 2/2016						
S&P 500	0.11	-0.09	0.0002	0.0123	-0.33	11.12
Exxon	0.16	-0.15	0.0002	0.0156	0.02	13.25
JPMorgan	0.22	-0.23	0.0001	0.0255	0.35	16.54
IBM	0.11	-0.09	0.0001	0.0135	-0.21	5.99
Johnson	0.12	-0.08	0.0002	0.0101	0.49	11.22
GE	0.18	-0.14	0.0000	0.0185	0.03	12.39
P&G	0.10	-0.08	0.0001	0.0110	-0.21	7.34
Disney	0.15	-0.10	0.0004	0.0168	0.17	8.21

Table B.2: KPSS test for stationarity

H_0: series is stationary

Stock	statistic	critical value (1% sig. level)	null rejection
S&P 500	0.0366	0.216	No
Exxon	0.0186	0.216	No
JPMorgan	0.0261	0.216	No
IBM	0.0738	0.216	No
Johnson	0.0268	0.216	No
GE	0.0538	0.216	No
P&G	0.0218	0.216	No
Disney	0.0569	0.216	No

Table B.3: Jarque-Bera test for normality

H_0: skewness is zero and excess kurtosis is zero

Stock	statistic	critical value (1% sig. level)	null rejection
S&P 500	266038	9.3238	Yes
Exxon	136568	9.3238	Yes
JPMorgan	79817	9.3238	Yes
IBM	61015	9.3238	Yes
Johnson	30978	9.3238	Yes
GE	30758	9.3238	Yes
P&G	1869266	9.3238	Yes
Disney	143474	9.3238	Yes

Table B.4: Engle's ARCH test

H_0: there are no ARCH effects

Stock	lag	statistic	critical value (1% sig. level)	null rejection
S&P 500	lag=1	169	6.63	Yes
Exxon		1180	6.63	Yes
JPMorgan		530	6.63	Yes
IBM		167	6.63	Yes
Johnson		435	6.63	Yes
GE		599	6.63	Yes
P&G		285	6.63	Yes
Disney		119	6.63	Yes
S&P 500	lag=3	533	11.34	Yes
Exxon		1299	11.34	Yes
JPMorgan		767	11.34	Yes
IBM		212	11.34	Yes
Johnson		538	11.34	Yes
GE		1043	11.34	Yes
P&G		298	11.34	Yes
Disney		447	11.34	Yes
S&P 500	lag=7	751	18.48	Yes
Exxon		1423	18.48	Yes
JPMorgan		996	18.48	Yes
IBM		303	18.48	Yes
Johnson		617	18.48	Yes
GE		1468	18.48	Yes
P&G		307	18.48	Yes
Disney		473	18.48	Yes

Bachelor Thesis Proposal

Author	Markéta Pokorná
Supervisor	PhDr. Boril Šopov, MSc., LL.M.
Proposed topic	Estimation and Application of the Tail Index

Topic characteristics Recently it has been shown that a good performance of models in periods with extreme events is of a great importance. Since financial returns do not seem to be normally distributed, the Extreme Value Theory (EVT) and its theorems serve to deal with this issue.

Hypotheses

1. Models with normally distributed innovations do not capture leptokurtic real data properly.
2. Hill's method is not the most preferable, because of its non-objectivity disadvantage.

Methodology The main interest of this study is application of Extreme Value Theory and estimation of the so called tail index. From the tail index inference can be made about the heaviness of tails. It is therefore a suitable tool to be used in financial risk management. The estimation of tail index by the original Hill's method will be described. Then its disadvantages will be stated and graphical methods of threshold selection will be discussed. In the empirical part several stocks will be chosen and the theory will be applied to the log-return series. The stylized facts of financial time series will be taken into account and analysed properly. Then the tail index will be computed for each series based on the Hill's method and other appropriately chosen procedures which will be considered based on better objectivity compared to Hill's method. The tail index will be computed also for subsamples of shorter time span to see whether the tail characteristics change over time.

Outline

1. Introduction
2. Literature Review
3. Theory and Methodology
   (a) EVT background
   (b) Hill's method
4. Empirical part
5. Conclusion
6. References

Core bibliography

1. Artzner, P., Delbaen, F., Eber, J.-M., and Health, D. (1999). Coherent measures of risk. *Mathematical finance*, **9(3)**: 203-228.
2. Drees, H., De Haan, L., and Resnick, S. (2000). How to make a hill plot. *The Annals of Statistics*, pages 1833-1855.
3. Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.
4. Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, **3(5)**, 1163-1174.
5. McNeil, A., Frey, R., and Embrechts, P. (2005). Quantitative risk management. *Princeton Series in Finance, Princeton,* 10.
6. Resnick, S., and Stărică, C. (1997). Smoothing the hill estimator. *Advances in Applied Probability*, pages 271-293.

Author Supervisor