Measuring Risk with Extreme Value Theory

A Master Thesis Presented

by

Vladimir Georgescu

(513294)

to

Prof. Dr. Wolfgang Härdle

CASE - Center of Applied Statistics and Economics

Humboldt University, Berlin

in partial fulfillment of the requirements

for the degree of

Master of Science in Economics and Management Science

Berlin, 27. May, 2009

Contents

1 Introduction

Risk measurement is of great importance in various fields, and Finance is no exception. Particularly relevant is the case of risks that occur very rarely, but produce extreme losses. The study of such risks is difficult because of many reasons: the scarcity of data, the need to predict events that may exceed the known history through the effects they produce, the inappropriateness of some statistical methods, to name a few. A modern branch of Statistics addresses this challenge, namely the Extreme Value Theory (EVT). In this framework, the tails of the distributions become much more relevant, and that is where the attention is focused. As an improvement to "classic" - central tendency analysis, EVT acknowledges the impact of heavy tails and earns usually better results in risk estimation.

This Master thesis presents the methods of EVT in a stepwise fashion and illustrates them with applications. It is organised as follows: the second chapter introduces risk measures, namely the widespread Value-at-Risk and the increasingly important expected shortfall. In chapter three, the dataset that is used is described and some preliminary analysis on the character of the data is performed. The next chapter is dedicated to the estimation methods, and two main groups are presented: the block maxima and the peaks-over-threshold (POT). In the POT case, the analysis is further divided into two approaches, the Generalised Pareto approach and the Hill Estimator approach. Chapter five concludes. The graphical support is intended to make the rather technical presentation of the topic more visual and also to provide a background for analysis using Matlab as statistical software.

2 Risk Measures

McNeil (1999) states that measuring a risk means summarising its distribution with a number known as a risk measure. The mean or standard deviation, for instance, measure some aspects of the risk, but do not provide information about the extreme risk. The Value-at-Risk (VaR) is the standard measure for downside risk, and the elements that define it are the maximum probable loss on a given portfolio over a period of time, given some specific confidence level. It provides an upper limit for a loss, which is exceeded only on a small proportion of occasions. An alternative risk measure is introduced in this section - the expected shortfall (ES). It is a more strict measure, since it is defined as the expected value of the exceedances over the Value-at-Risk. The comments and explanations to the definitions from this chapter rely on Franke, Härdle and Hafner (2008).

Definition 1 (Quantile Function)**.**
If F is a distribution function, we call the generalised inverse

$$F^{-1}(\gamma) = \inf\{x \in \mathbb{R};\ F(x) \geq \gamma\},\ \ 0 < \gamma < 1,$$

the quantile function*. It then holds that* $\mathrm{P}\{X_1 \leq F^{-1}(\gamma)\} = \gamma$*, i.e.,* $F^{-1}(\gamma)$ *is the* γ-quantile *of the distribution* F*.*

If F is strictly monotonic increasing and continuous, then F^{-1} is the generalised inverse of F.

Definition 2 (Value-at-Risk and Expected Shortfall)**.**
Let $0 < q < 1$*, and let* F *be the distribution of the loss* X *of a financial investment within a given time period, for example, one day or 10 trading days. Typical values for* q *are* $q = 0.95$ *and* $q = 0.99$*.*
a) The Value-at-Risk *(VaR) is the q-quantile*

$$VaR_q(X) = x_q = F^{-1}(q).$$

b) The expected shortfall *(S_q) is defined as*

$$S_q = \mathsf{E}\{X | X > x_q\}.$$

The most commonly used measurement today for quantifiyng market risk is the Value-at-Risk. However, in the future it is very likely that the expected shortfall will be at least as important.

Definition 3 (Coherent Risk Measure)**.**
A coherent risk measure *is a real-valued function $\rho : \mathbb{R} \rightarrow \mathbb{R}$ of real-valued random variables, which models the losses, with the following characteristics:*

(A1)	$X \geq Y$ *a.s.* $\Longrightarrow \rho(X) \geq \rho(Y)$	*(Monotonicity)*
(A2)	$\rho(X + Y) \leq \rho(X) + \rho(Y)$	*(Subadditivity)*
(A3)	$\rho(\lambda X) = \lambda \rho(X)$ *for* $\lambda \geq 0$	*(Positive homogeneity)*
(A4)	$\rho(X + a) = \rho(X) + a$	*(Translation equivariance)*

These characteristics correspond to intuitive obvious requirements of a market risk measurement:
(A1) When the loss from investment X is always larger or equal to the loss from investment Y, then the risk from investment X is also larger or equal to the risk from investment Y.
(A2) The risk of a portfolio consisting of investments in X and Y is at most as large as the sum of the individual risks (diversification of the risk).
(A3) When an investment is multiplied, then the risk is also multiplied accordingly.
(A4) By adding a risk free investment, i.e., a non-random investment with known losses a ($a < 0$, when the investment has fixed payments), to a portfolio, the risk changes by exactly a.
The VaR does not meet condition (A2) in certain situations. Let X and Y, for example, be i.i.d. and both can take on the value 0 or 100 with probabilities $\mathrm{P}(X = 0) = \mathrm{P}(Y = 0) = p$ and $\mathrm{P}(X = 100) = \mathrm{P}(Y = 100) = 1 - p$. Then $X + Y$ can be 0, 100 and 200 with probabilities $\mathrm{P}(X + Y = 0) = p^2$, $\mathrm{P}(X + Y = 100) = 2p(1 - p)$

and $\mathrm{P}(X + Y = 200) = (1 - p)^2$ respectively. For $p^2 < q < p$ and $q < 1 - (1 - p)^2$, e.g., for $p = 0.96, q = 0.95$, it holds that

$$VaR_q(X) = VaR_q(Y) = 0, \text{ but } VaR_q(X + Y) = 100.$$

The expected shortfall, on the other hand, is a coherent risk measure that always fulfills all four conditions. It also gives a more intuitive view of the actual risk of extreme losses than the Value-at-Risk. The VaR only depends on the probability of losses above the q-quantile x_q, but it does not say anything about whether these losses are always just a little above the threshold x_q or whether there are also losses that are much larger than x_q that need to be taken into account. In contrast the expected shortfall is the expected value of the potential losses above x_q and depends on the actual size of the losses.

The Value-at-Risk is simply a quantile and can be, for example, estimated as a sample quantile $\hat{F}_n^{-1}(q)$, where $\hat{F}_n(x)$ is the empirical distribution of a sample of negative values, i.e., losses, from the past. This particular estimator of $q \approx 1$, which is for the typical VaR-level of 0.95 and 0.99, is often too optimistic. Alternative VaR estimators, which have the possibility of reflecting extreme losses better, are the peaks-over-thershold (POT) and the Hill quantile estimators. Analogous estimators for the expected shortfall can be also derived.

3 Data Description

The methods of Extreme Value Theory that are described in the following chapters are illustrated with the help of programs written in Matlab software. In some cases, the data for these programs is generated randomly whereas in others there is a main dataset that is referred to. From this initial dataset, data is derived for the purpose of further applications.

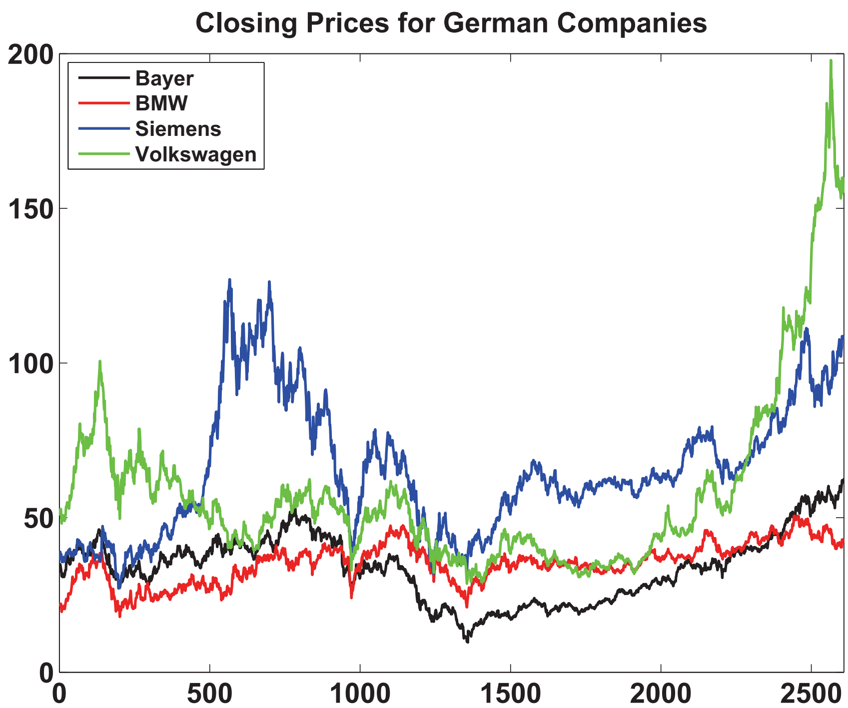

Figure 1: Closing prices of stocks: Bayer, BMW, Siemens, Volkswagen. Time period: 1. January 1998 - 31. December 2007

The data is represented by the closing stock prices of 4 DAX companies, namely Bayer, BMW, Siemens and Volkswagen, for the period 1. January 1998 - 31. December 2007, on daily basis. For each company, there are 2608 observations, collected from Monday to Friday. For non-trading days, the prices considered were the previous ones. A number of 2543 trading days were recorded on the Frankfurt Stock Exchange (FSE) for the period under review. The source for the data is Datastream. From the initial dataset, an equally-weighted portfolio is built, consisting of one of each of the 4 stocks.

Figure 1 plots the closing prices of Bayer, BMW, Siemens and Volkswagen.

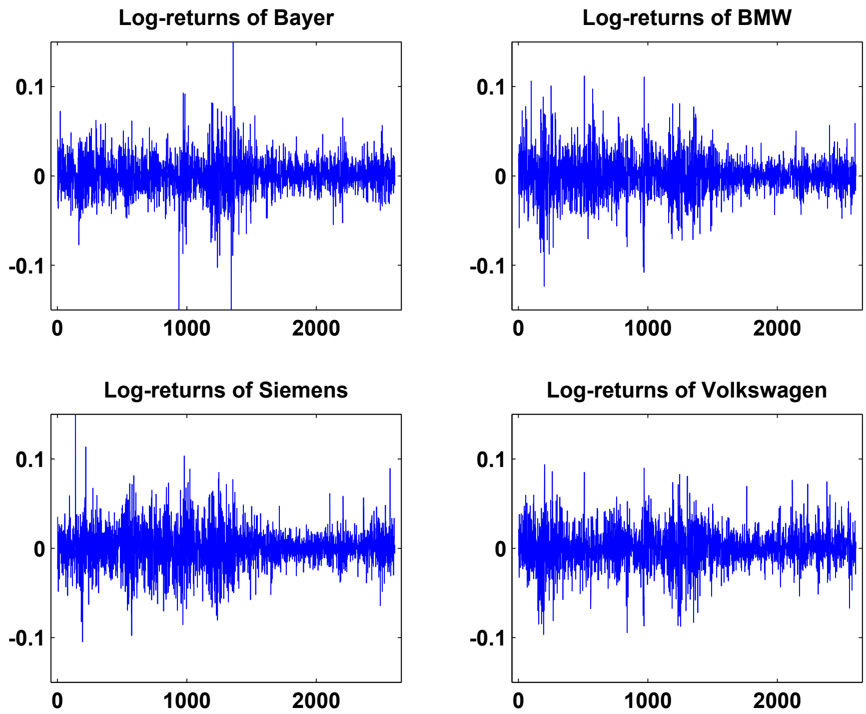

Figure 2: Daily log-returns for Bayer, BMW, Siemens and Volkswagen. Time period: 1. January 1998 to 31. December 2007

The daily log-returns for the companies are computed according to the formula:

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right),$$

where P_t is the level of the stock price on day t.

Figure 2 displays the log-returns graphically. The scale is the same for all companies in order to make the results comparable. The daily log-returns usually fall within the boundaries of a 15 percent decrease and a 15 percent increase, respectively, with the exception of some outilers displayed by the log-returns of the Bayer and Siemens stocks. The plot of the log-returns of the portfolio is displayed in Figure 3. The volatility, as expected, is visibly lower than in the individual cases, reflecting diversification.

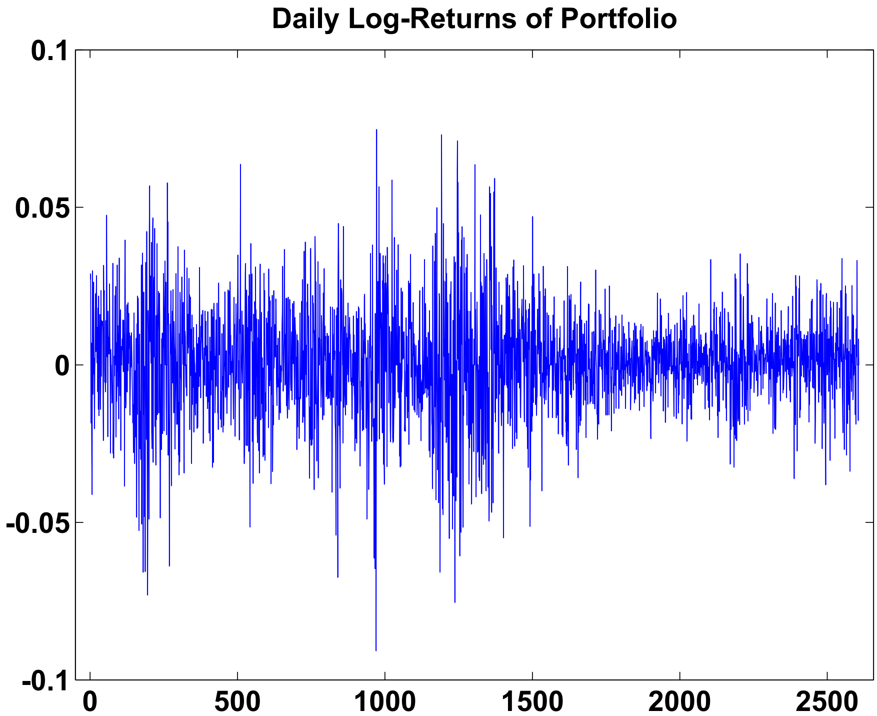

Figure 3: Daily log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen. Time period: 1. January 1998 to 31. December 2007

Name	Minimum	Maximum	Mean	Median	Standard Deviation
Bayer	-0.1844	0.3225	0.0003	0.0000	0.0209
BMW	-0.1238	0.1119	0.0003	0.0000	0.0214
Siemens	-0.1048	0.1566	0.0004	0.0000	0.0227
Volkswagen	-0.0965	0.0937	0.0004	0.0000	0.0211
Portfolio	-0.0908	0.0747	0.0004	0.0006	0.0171

Table 1: Summary statistics for the log-return daily data. Time period: 1. January 1998 to 31. December 2007

Table 1 provides descriptive statistics for the log-returns of the individual companies and the portfolio, respectively. Considering the standard deviation as a measure for volatility, the Siemens log-returns display the largest volatility, despite the fact that Bayer has larger outliers. The boundaries between minimum and maximum are narrowest and also the volatility is the lowest in the case of the equally-weighted portfolio. In terms of profitability, the average log-return is positive in all cases.

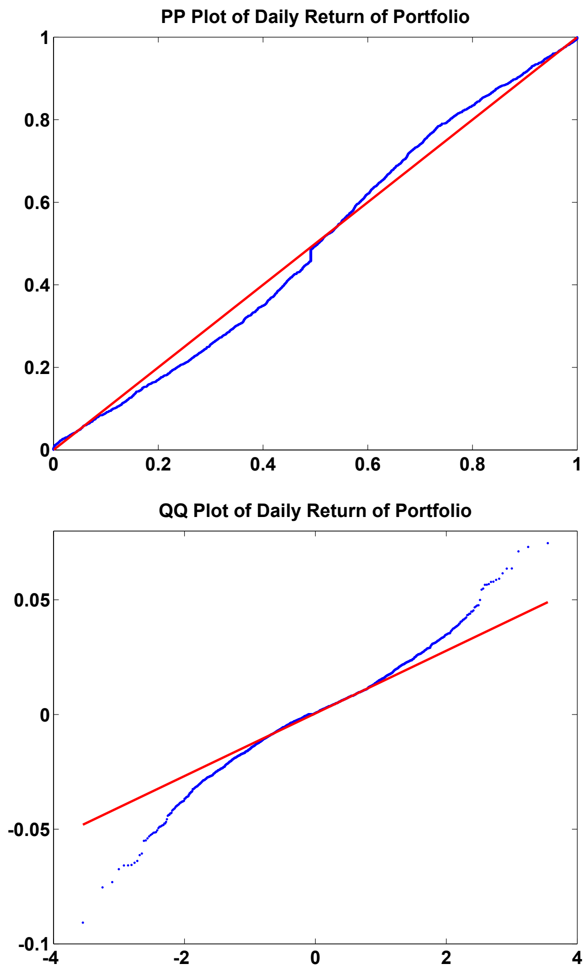

Figure 4: PP plot and QQ plot of the portfolio of Bayer, BMW, Siemens and Volkswagen. Time period: 1. January 1998 to 31. December 2007.

Figure 4 shows the plot of the probability distribution function and the quantile plot for the pdf $F(x) = \Phi(x)$ of the standard normal, for the portfolio of Bayer, BMW,

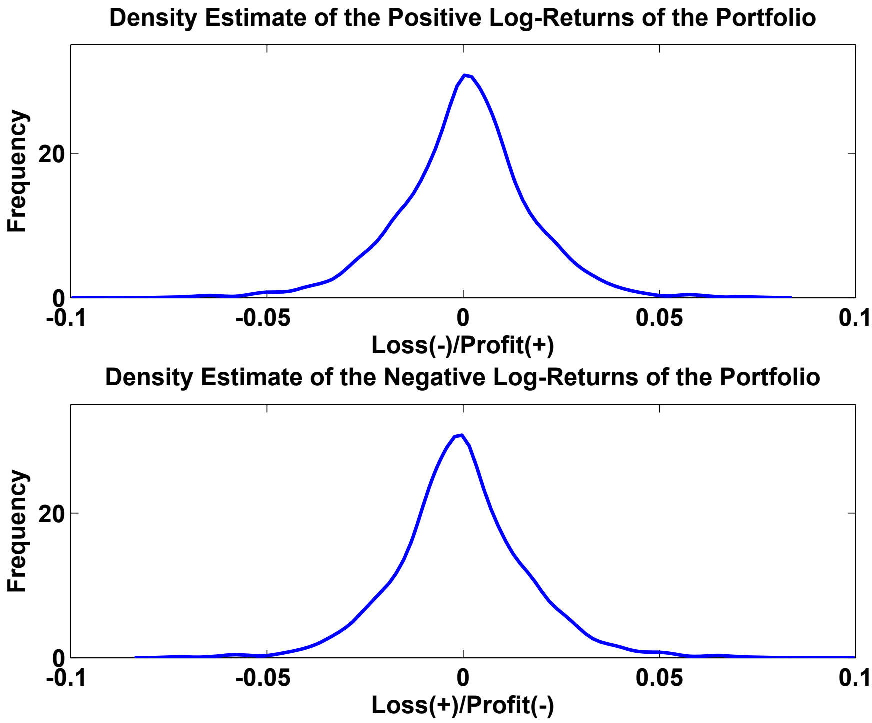

Figure 5: Kernel density estimate of the positive and negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen. Time period: 1. January 1998 to 31. December 2007

Siemens and Volkswagen. The deviations from the straight line show that the data is not normally distributed.

Figure 5 displays the kernel density estimate of the positive and negative log-returns of the portfolio, respectively. The reason for considering negative log-returns is that while one is interested in the loss side of the distribution, the risk measures are expressed as positive numbers. So basically, when trying to estimate the Value-at-Risk and the expected shortfall, the losses are depicted as positive quantities, and the profits as negative ones.

4 Estimation Methods

In this section, the main methods used for estimating the Value-at-Risk and the expected shortfall will be presented. According to McNeil (1999), there are two main types of models that deal with extreme values.

The first group and the oldest consists of the *block maxima models.* They represent models designed for the largest observations from samples of identically distributed observations.

The second group of models is considered to be more modern and it consists of the *peaks-over-treshold* (POT) models. They concern all the large observations which exceed a certain threshold. Within the POT class of models, one may choose between two different approaches. One is fully parametric and based on the generalised Pareto distribution (GPD), while the other is semi-parametric and is centered around the Hill estimator. The author regards the POT models as being generally the most useful in practice, since they are able to use the often limited data available on extreme values in a more efficient way.

The theorems, definitions and the lemma presented in this chapter are well established in the study of EVT. The proofs, comments and explanations in general rely to a great extent on Franke, Härdle and Hafner (2008).

4.1 The Block Maxima Method

Consider the stochastic behaviour of the maximum $M_n = \max(X_1, \ldots, X_n)$ of n identically distributed random variables $X_1, \ldots, X_n$ with cumulative distribution function (cdf) $F(x)$. From a risk management perspective $X_t = -Z_t$ is the negative return at day t. The cdf of M_n is

$$\mathrm{P}(M_n \leq x) = \mathrm{P}(X_1 \leq x, \ldots, X_n \leq x) = \prod_{t=1}^{n} \mathrm{P}(X_t \leq x) = F^n(x). \tag{1}$$

Only unbounded random variables X_t are considered, i.e. $F(x) < 1$ for all $x < \infty$. Obviously it holds that $F^n(x) \to 0$ for all x, when $n \to \infty$, and thus $M_n \xrightarrow{P} \infty$.

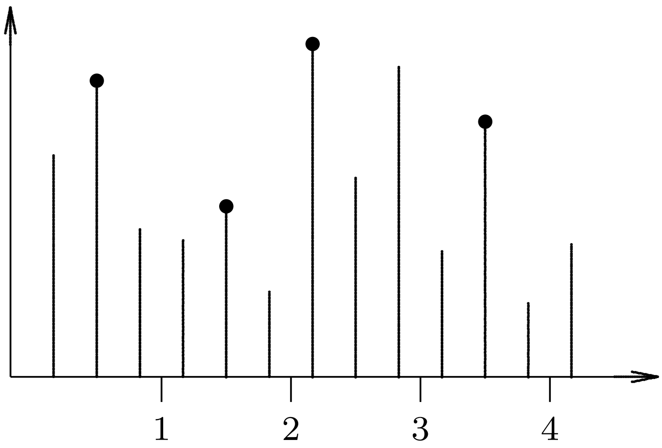

Figure 6: Block-maxima

Figure 6 illustrates the concept of block maxima, where the block size is n=3. The maximum of n unbounded random variables may become arbitrary large. In order to achieve a non-degenerate behaviour limit, M_n has to be standardised in a suitable fashion. For an analysis of asymptotics one needs the limit law for the block maxima M_n.

Definition 4 (Maximum Domain of Attraction).
The random variable X_t belongs to the maximum domain of attraction *(MDA) of a non-degenerate distribution G, if for suitable sequences $c_n > 0, d_n$ it holds that:*

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G \quad \textit{for} \;\; n \to \infty,$$

i.e. $F^n(c_n x + d_n) \to G(x)$ at all continuity points x of the cdf $G(x)$.

It turns out that only a few distributions G can be considered as the asymptotic limit distribution of the standardised maximum M_n. They are referred to as the *extreme value distributions*. These are the following three distribution functions:

Fréchet: $G_{1,\alpha}(x) = \exp\{-x^{-\alpha}\},\ x \geq 0$, for $\alpha > 0$,

Gumbel: $G_0(x) = \exp\{-e^{-x}\},\ x \in \mathbb{R}$,

Weibull: $G_{2,\alpha}(x) = \exp\{-|x|^{-\alpha}\},\ x \leq 0$, for $\alpha < 0$.

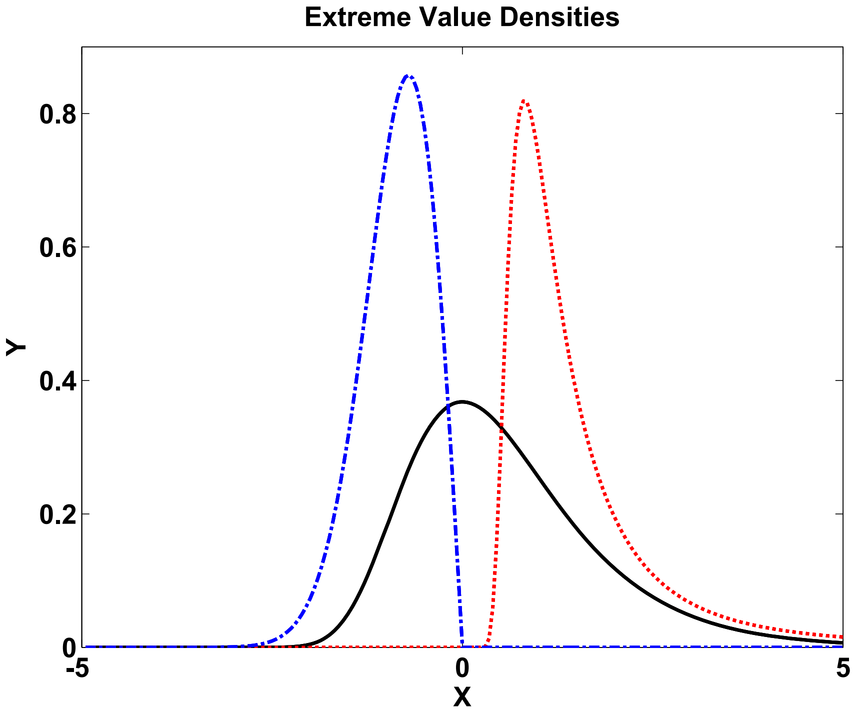

Figure 7: Gumbel distribution (solid line), Fréchet distribution with parameter $\alpha = 2$ (dotted line) and Weibull distribution with parameter $\alpha = -2$ (dash-dot line).

The Fréchet distributions are concentrated on the non-negative real numbers $[0, \infty)$, while the Weibull distribution, on the other hand, on $(-\infty, 0]$,. The Gumbel distributed random variables can attain any real number. Figure 7 displays the density function of the Gumbel distribution, the Fréchet distribution with parameter $\alpha = 2$ and the Weibull distribution with parameter $\alpha = -2$. Jenkinson and von Mises suggested a parametric representation for the three standard distributions:

Definition 5 (Generalised Extreme Value Distributions)**.**
The generalised extreme value distribution *(GEV* = generalised extreme value*) with the* form parameter $\gamma \in \mathbb{R}$ *has the distribution function:*

$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\},\ 1+\gamma x > 0$ *for* $\gamma \neq 0$

$G_0(x) = \exp\{-e^{-x}\},\ x \in \mathbb{R}$

G_0 is the Gumbel distribution, whereas $G_\gamma, \gamma \neq 0$ is linked to the Fréchet- and Weibull distributions by the following relationships:

$$G_\gamma\left(\frac{x-1}{\gamma}\right) = G_{1,1/\gamma}(x) \text{ for } \gamma > 0,$$

$$G_\gamma\left(-\frac{x+1}{\gamma}\right) = G_{2,-1/\gamma}(x) \text{ for } \gamma < 0.$$

This definition describes the standard form of the GEV distributions. In general the centre and the scale can be changed in order to obtain other GEV distributions: $G(x) = G_\gamma(\frac{x-\mu}{\sigma})$ with the form parameter γ, the location parameter $\mu \in \mathbb{R}$ and the scale parameter $\sigma > 0$. For asymptotic theory this does not matter since the standardised sequences c_n, d_n can be always chosen so that the asymptotic distribution G has the standard form ($\mu = 0, \sigma = 1$). McNeil (1999) considers the GEV to be the natural limit distribution for normalized maxima.

An important result of the asymptotic distribution of the maximum M_n is the *Fisher-Tippett theorem*:

Theorem 1 (Fisher and Tippett (1928))**.**
If sequences $c_n > 0, d_n$ *exist and a non-degenerate distribution* G*, so that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G \quad \text{for } n \to \infty,$$

then G *is a GEV distribution.*

Proof:

As a form of clarification the basic ideas used to prove this central result are outlined. Let $t > 0$, and $[z]$ represent the integer part of z. Since $F^{[nt]}$ is the distribution function of $M_{[nt]}$, due to our assumptions on the asymptotic distribution of M_n it holds that

$$F^{[nt]}(c_{[nt]}x + d_{[nt]}) \longrightarrow G(x) \text{ for } [nt] \to \infty, \quad \text{i.e.} \quad n \to \infty.$$

On the other hand it also holds that

$$F^{[nt]}(c_n x + d_n) = \{F^n(c_n x + d_n)\}^{\frac{[nt]}{n}} \longrightarrow G^t(x) \text{ for } n \to \infty.$$

In other words this means that

$$\frac{M_{[nt]} - d_{[nt]}}{c_{[nt]}} \xrightarrow{\mathcal{L}} G \ , \qquad \frac{M_{[nt]} - d_n}{c_n} \xrightarrow{\mathcal{L}} G^t$$

for $n \to \infty$. According to the Lemma, which is stated below, this is only possible when

$$\frac{c_n}{c_{[nt]}} \longrightarrow b(t) \geq 0, \qquad \frac{d_n - d_{[nt]}}{c_{[nt]}} \longrightarrow a(t)$$

and

$$G^t(x) = G\{b(t)x + a(t)\}, \ t > 0, \ x \in \mathbb{R}. \tag{2}$$

This relationship holds for arbitrary values t. We use it in particular for arbitrary t, s and $s \cdot t$ and obtain

$$b(st) = b(s)\ b(t), \ a(st) = b(t)a(s) + a(t). \tag{3}$$

The functional equations (2), (3) for $G(x), b(t), a(t)$ have only one solution, when G is one of the distributions $G_0, G_{1,\alpha}$ or $G_{2,\alpha}$, that is, G must be a GEV distribution.

□

McNeil (1999) compares the significance of the Fisher-Tipett theorem in the study of maxima with that of the central limit theorem in the case of sums and averages. The essential result of the theorem is the fact that the only limiting distribution that is possible for (normalized) block maxima is the GEV.

Lemma 1 (Convergence Type Theorem).
Let $U_1, U_2, \ldots, V, W$ be random variables, $b_n, \beta_n > 0$, $a_n, \alpha_n \in \mathbb{R}$. If

$$\frac{U_n - a_n}{b_n} \xrightarrow{\mathcal{L}} V$$

in distribution for $n \to \infty$, then it holds that:

$$\frac{U_n - \alpha_n}{\beta_n} \xrightarrow{\mathcal{L}} W \quad \textit{if and only if} \quad \frac{b_n}{\beta_n} \longrightarrow b \geq 0, \; \frac{a_n - \alpha_n}{\beta_n} \longrightarrow a \in \mathbb{R}.$$

In this case W has the same distribution as $bV + a$.

Notice that the GEV distributions are identical to the so called *max-stable* distributions, by which for all $n \geq 1$ the maximum M_n of n i.i.d. random variables $X_1, \ldots, X_n$ have the same distribution as $c_n X_1 + d_n$ for appropriately chosen $c_n > 0, d_n$.

Figure 8 shows the normal plot for the special case $F(x) = \Phi(x)$ with computer generated random variables that have a Gumbel distribution, Fréchet distribution with parameter $\alpha = 2$ and a Weibull distribution with parameter $\alpha = -2$ respectively. The differences with the normally distributed random variables, which would have approximately a straight line in a normal plot, can be clearly seen. Figure 9 compares the cdf of the pseudo random variables from the Gumbel distribution against the theoretical Gumbel distribution.

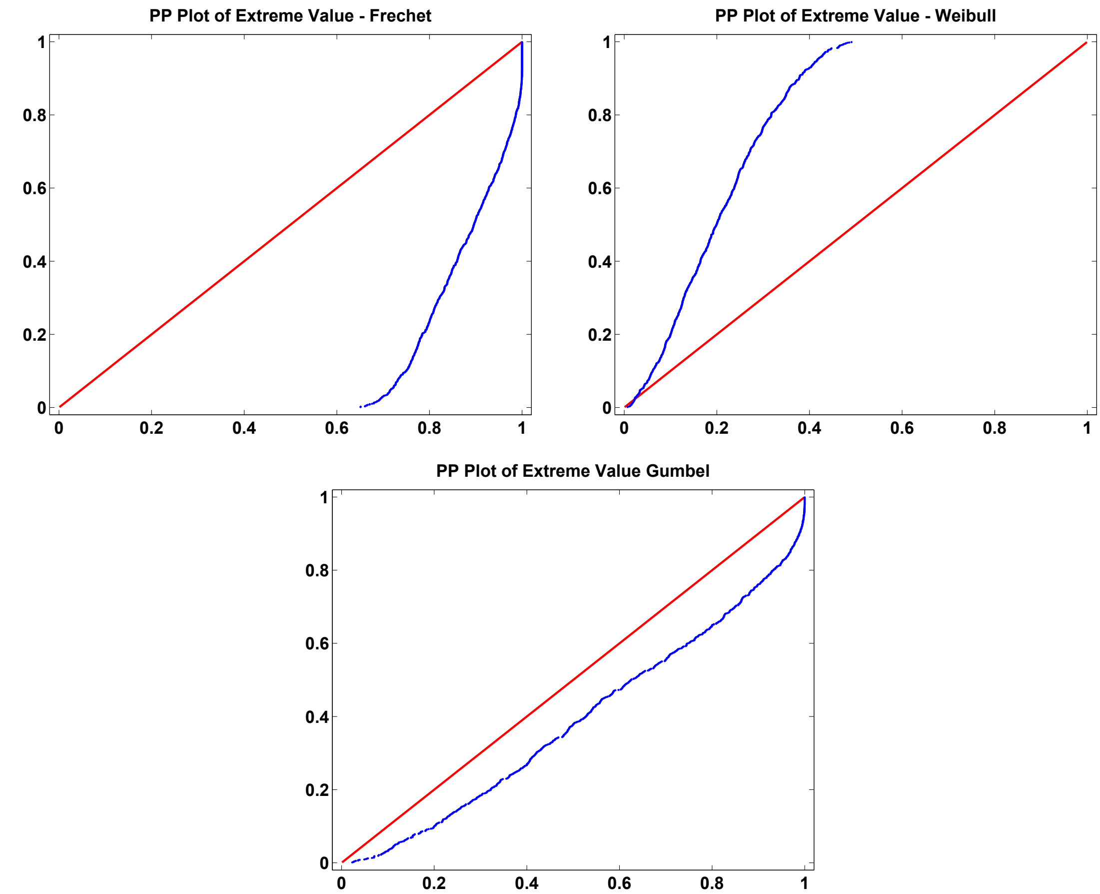

Figure 8: PP plot for the normal distribution and pseudo random variables with extreme value distributions. Fréchet (upper left), Weibull (upper right) and Gumbel (below).

Identifying the type of the limit (GEV) distribution

If the maximum of i.i.d. random variables converges in distribution after being appropriately standardised, then the question arises which of the three GEV distributions is the asymptotic distribution. The deciding factor is how fast the probability for extremely large observations decreases beyond a threshold x, when x increases. Since this exceedance probability plays an important role in extreme value theory,

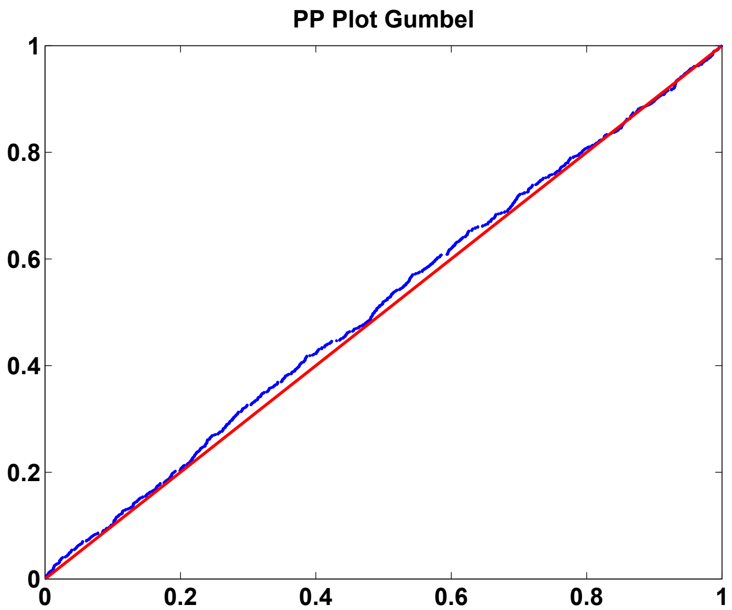

Figure 9: PP plot of the pseudo random variables with Gumbel distribution against theoretical Gumbel distribution.

some more notations are introduced:

$$\overline{F}(x) = \mathrm{P}(X_1 > x) = 1 - F(x).$$

The relationship between the exceedance probability $\overline{F}(x)$ and the distribution of the maxima M_n will become clear with the following theorem.

Theorem 2.

a) For $0 \leq \tau \leq \infty$ and every sequence of real numbers $u_n, n \geq 1$, it holds for $n \to \infty$ that

$$n\overline{F}(u_n) \to \tau \quad \textit{if and only if} \quad \mathrm{P}(M_n \leq u_n) \to e^{-\tau}.$$

b) F belongs to the maximum domain of attraction of the GEV distribution G with the standardised sequences c_n, d_n exactly when $n \to \infty$

$$n\overline{F}(c_n x + d_n) \to -\log G(x) \quad \textit{for all} \;\; x \in \mathbb{R}.$$

The exceedance probability of the Fréchet distribution $G_{1,\alpha}$ behaves like $1/x^\alpha$ for $x \to \infty$, because the exponential function around 0 is approximately linear, i.e.,

$$\overline{G}_{1,\alpha}(x) = \frac{1}{x^\alpha}\{1 + \mathcal{o}(1)\} \qquad \text{for} \quad x \to \infty.$$

Essentially all of the distributions that belong to the MDA of this Fréchet distribution show the same behaviour; $x^\alpha \overline{F}(x)$ is almost constant for $x \to \infty$, or more specifically, it is a slowly varying function.

Definition 6 (Slowly Varying Functions)**.**
A positive measurable function L in $(0, \infty)$ is called slowly varying*, if for all $t > 0$*

$$\frac{L(tx)}{L(x)} \to 1 \qquad \textit{for} \quad x \to \infty.$$

Typical slowly varying functions are, in addition to constants, logarithmic growth rates, for example $L(x) = \log(1 + x), x > 0$.

Theorem 3 (MDA of the Frechét Distribution)**.**
F belongs to the maximum domain of attraction of the Fréchet distribution $G_{1,\alpha}$ for some $\alpha > 0$, if and only if $x^\alpha \overline{F}(x) = L(x)$ is a slowly varying function. The random variables X_t with the distribution function F are unbounded, i.e., $F(x) < 1$ for all $x < \infty$, and it holds that

$$\frac{M_n}{c_n} \xrightarrow{\mathcal{L}} G_{1,\alpha}$$

with $c_n = F^{-1}(1 - \frac{1}{n})$.

For the description of the standardised sequence c_n, the following notation was used: c_n is an extreme quantile of the distribution F, and it holds that $\overline{F}(c_n) = \mathrm{P}(X_t > c_n) = 1/n$.

There is a corresponding criterion for the MDA of the Weibull distribution that can be shown using the relationship $G_{2,\alpha}(-x^{-1}) = G_{1,\alpha}(x), \ x > 0$. Random variables, whose maxima are asymptotically Weibull distributed, are by all means bounded, i.e., a constant $c < \infty$ exists, such that $X_t \leq c$ holds with probability 1 ($P(X_t \leq c) = 1$). Therefore, in financial applications they are only interesting in special situations where using a type of hedging strategy, the loss, which can result from an investment, is limited. Below mainly cases where the losses are unlimited will be discussed; cases in which losses are limited can be treated in a similar fashion.

Fréchet distributions appear as asymptotic distributions of the maxima of those random variables whose probability of values beyond x only slowly decreases with x, whereas only bounded random variables belong to the maximum domain of attraction of Weibull distributions. Many known distributions such as the exponential or the normal distribution do not belong to either one of the groups. It is likely that in such cases the distribution of the appropriate standardised maxima converges to a Gumbel distribution.

Theorem 4 (MDA of the Gumbel Distribution)**.**

The distribution function F of the unbounded random variables X_t belongs to the maximum domain of attraction of the Gumbel distribution if measurable scaling functions $c(x), g(x) > 0$ as well as an absolute continuous function $e(x) > 0$ exist with $c(x) \to c > 0,\ g(x) \to 1, e'(x) \to 0$ for $x \to \infty$ so that for $z < \infty$

$$\overline{F}(x) = c(x) \exp\left(-\int_z^x \frac{g(y)}{e(y)} dy\right), \quad z < x < \infty.$$

In this case it holds that

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G_0$$

with $d_n = F^{-1}\left(1 - \frac{1}{n}\right)$ and $c_n = e(d_n)$.

As a function $e(x)$, the *average excess function* can be used:

$$e(x) = \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(y)\, dy, \quad x < \infty,$$

which will be considered in more detail in the following.

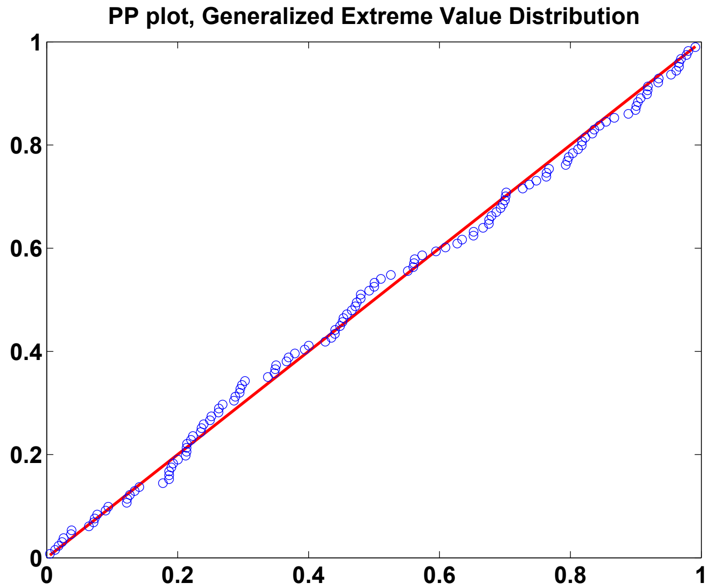

Figure 10: PP plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalized Extreme Value Distribution with parameter $\gamma = 0.15$ estimated with block maxima method. Time period: 1. January 1998 to 31. December 2007.

The exponential distribution with parameter λ has the distribution function $F(x) = 1 - e^{-\lambda x}, x \geq 0$, so that $\overline{F}(x) = e^{-\lambda x}$ fulfills the conditions stipulated in the theorem with $c(x) = 1,\ g(x) = 1,\ z = 0$ and $e(x) = 1/\lambda$. The maximum M_n of n independent exponentially distributed random variables with parameter λ thus converges in distribution to the Gumbel distribution:

$$\lambda\left(M_n - \frac{1}{\lambda}\ \log\ n\right) \xrightarrow{\mathcal{L}} G_0 \quad \text{for} \quad n \to \infty.$$

It can be shown, for example, that the normal distribution also belongs to the maximum domain of attraction of the Gumbel distribution. If, for example, M_n is the maximum of n independent standard normally distributed random variables,

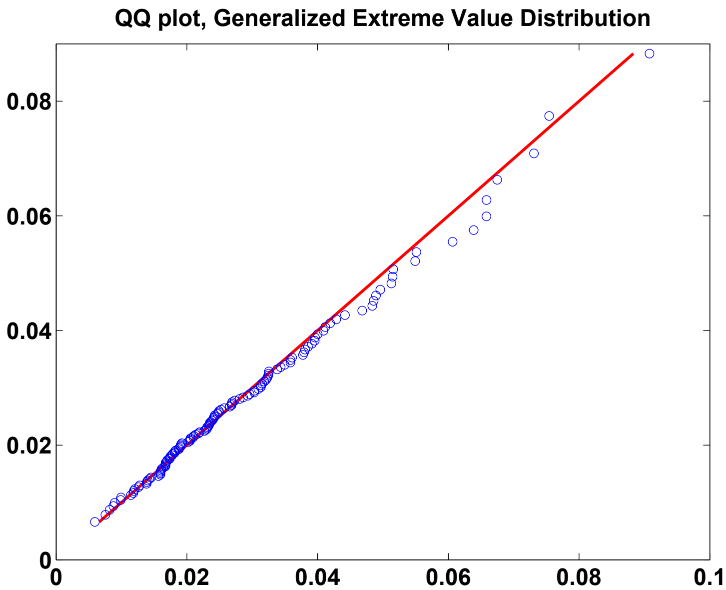

Figure 11: QQ plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalized Extreme Value Distribution with parameter $\gamma = 0.15$ estimated with block maxima method. Time period: 1. January 1998 to 31. December 2007.

then it holds that

$$\sqrt{2 \ \log \ n}(M_n - d_n) \xrightarrow{\mathcal{L}} G_0 \quad \text{for} \quad n \to \infty$$

$$\text{with} \qquad d_n = \sqrt{2 \ \log \ n} - \frac{\log \log \ n + \log(4\pi)}{2\sqrt{2 \ \log \ n}}.$$

Figure 10 represents the PP plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalised Extreme Value distribution. The risk measures are expressed as positive numbers, even though they depict losses. The goal is to find the maxima of the distribution of negative log-returns. The block maxima method is used, with a block size of n=20. The data is fitted to the GEV distribution, and a parameter $\gamma = 0.15$ is estimated. Figure 11 represents the QQ plot of the data against the GEV distribution and the same method is used. As it results from the two plots, the empirical data fits quite well to the GEV distribution with parameter $\gamma = 0.15$.

4.2 The Peaks-Over-Threshold (POT) Method

There is a tight relationship between the asymptotic behaviour of the maxima of random variables and the distribution of the corresponding excesses which builds the foundation for an important estimation method in extreme value statistics. In general it deals with observations crossing a specified threshold u, situation which is depicted in Figure 12.

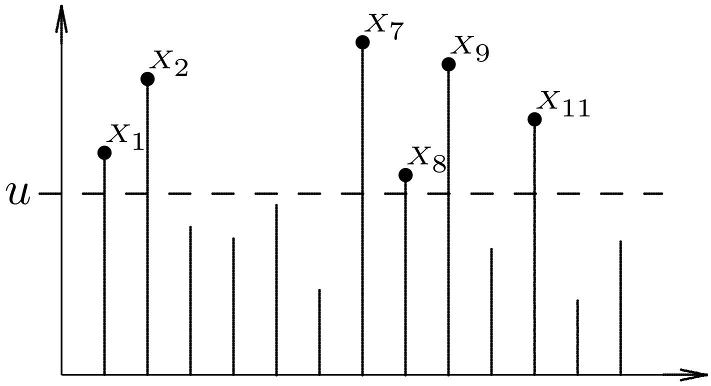

Figure 12: Excesses over a threshold u.

The distribution F_u of the observations crossing the threshold is defined as follows:

Definition 7 (Excess Distribution).
Let u be an arbitrary threshold and F a distribution function of an unbounded random variable X.

a) $F_u(x) = \mathrm{P}\{X - u \le x \mid X > u\} = \{F(u+x) - F(u)\}/\overline{F}(u),\ 0 \le x < \infty$ *is called the* excess distribution *beyond the threshold u.*

b) $e(u) = \mathsf{E}\{X - u \mid X > u\},\ 0 < u < \infty$*, is the* average excess function.

With partial integration it follows that this definition of the average excess function together with the following Theorem 5 agrees with:

$$e(u) = \int_u^\infty \frac{\overline{F}(y)}{\overline{F}(u)} dy.$$

If Δ_u is a random variable with the distribution function F_u, then its expectation is $\mathsf{E}\Delta_u = e(u)$.

Theorem 5 (Mean Excess Function)**.**
X is a positive, unbounded random variable with an absolutely continuous distribution function F.
a) The average excess function $e(u)$ identifies F exactly:

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\left(-\int_0^x \frac{1}{e(u)} du\right), \quad x > 0.$$

b) If F is contained in the MDA of the Fréchet distribution $G_{1,\alpha}$, then $e(u)$ is approximately linear for $u \to \infty$: $e(u) = \frac{1}{\alpha-1}\, u\{1 + o(1)\}$.

Throughout the rest of the chapter $X, X_1, \ldots, X_n$ are unbounded, i.i.d. random variables with distribution function F.
Notation: $X_{(1)} \leq \ldots \leq X_{(n)}$ and $X^{(1)} \geq \ldots \geq X^{(n)}$ represent the *order statistics*, that is, the data is sorted according to increasing or decreasing size. Obviously then $X_{(1)} = X^{(n)},\ X_{(n)} = X^{(1)}$ etc.

4.2.1 The Generalised Pareto Approach

McNeil (1999) underlines the advantages of this style of analysis. It is important that simple parametric formulae are obtained for the measures of risk, and it is relatively easy to estimate the statistical error by means of the maximum likelihood techniques. The Generalised Pareto distribution (GPD) is another probability distribution, but, in the risk management context, it is equally or more important than the Normal distribution. The Normal distribution is not suitable for addressing extreme loss, because its tails are too thin.

Another member of the distributions in the maximum domain of attraction of the Fréchet distribution $G_{1,\alpha}$ is the *Pareto distribution* with the distribution function

$$W_{1,\alpha}(x) = 1 - \frac{1}{x^{\alpha}}, x \geq 1, \alpha > 0,$$

as well as all other distributions with *Pareto tails*, i.e., with

$$\overline{F}(x) = \frac{\kappa}{x^{\alpha}}\{1 + \mathcal{O}(1)\} \quad \text{for} \quad x \to \infty.$$

Since $\overline{F}^{-1}(\gamma)$ for $\gamma \approx 1$ behaves here like $(\kappa/\gamma)^{1/\alpha}$, c_n for $n \to \infty$ is identical to $(\kappa n)^{1/\alpha}$, and

$$\frac{M_n}{(\kappa n)^{1/\alpha}} \xrightarrow{\mathcal{L}} G_{1,\alpha} \quad \text{for } n \to \infty.$$

Definition 8 (Generalised Pareto Distribution)**.**
The generalised Pareto distribution *(GP =* generalised Pareto*) with parameters $\beta > 0$, γ has the distribution function*

$$W_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} \quad \textit{for} \quad \begin{cases} x \geq 0 & \textit{if} \quad \gamma > 0 \\ 0 \leq x \leq \frac{-\beta}{\gamma} & \textit{if} \quad \gamma < 0, \end{cases}$$

and

$$W_{0,\beta}(x) = 1 - e^{-\frac{1}{\beta}x}, \ x \geq 0.$$

$W_{\gamma}(x) = W_{\gamma,1}(x)$ *is called the* generalised standard Pareto distribution *or* standardised *GP distribution.*

Figure 13 shows the generalised standard Pareto distribution with parameters $\gamma =$

$0.5, 0$ and -0.5 respectively.

For $\gamma = 0$ the standardised GP distribution is an exponential distribution with parameter 1. For $\gamma > 0$ it is a Pareto distribution $W_{1,\alpha}$ with the parameter $\alpha = 1/\gamma$. For $\gamma < 0$ the GP distribution is also referred to as a *Beta distribution* and has the distribution function $W_{2,\alpha} = 1 - (-x)^{-\alpha}, -1 \leq x \leq 0, \alpha < 0$.

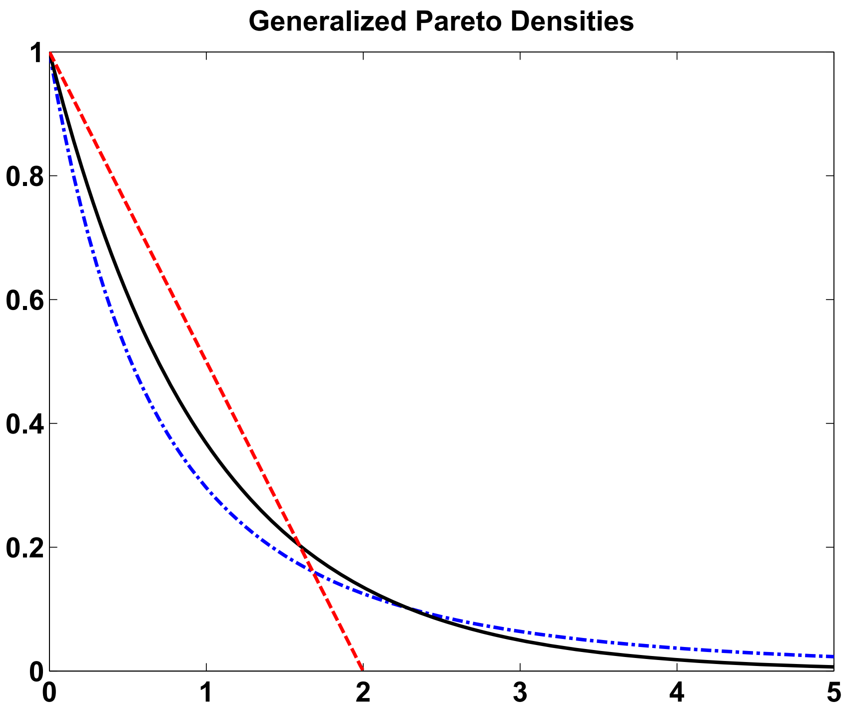

Figure 13: Standard Pareto distribution ($\beta = 1$) with parameter $\gamma = 0.5$ (dash-dot line), 0 (solid line) and -0.5 (broken line).

Theorem 6 (MDA of GEV distribution)**.**

The distribution F is contained in the MDA of the GEV distribution G_γ with the form parameter $\gamma \geq 0$, exactly when for a measurable function $\beta(u) > 0$ and the GP distribution $W_{\gamma,\beta}$ it holds that:

$$\sup_{x \geq 0} |F_u(x) - W_{\gamma,\beta(u)}(x)| \rightarrow 0 \, for \;\; u \rightarrow \infty.$$

A corresponding result also holds for the case when $\gamma < 0$, in which case the supremum of x must be taken for those $0 < W_{\gamma,\beta(u)}(x) < 1$.
For the generalised Pareto distribution $F = W_{\gamma,\beta}$ it holds for every finite threshold $u > 0$

$$F_u(x) = W_{\gamma,\beta+\gamma u}(x) \quad \text{for} \quad \begin{cases} x \geq 0 & \text{if} \quad \gamma \geq 0 \\ 0 \leq x < -\frac{\beta}{\gamma} - u & \text{if} \quad \gamma < 0, \end{cases}$$

In this case $\beta(u) = \beta + \gamma\, u$.

Definition 9 (Empirical Average Excess Function)**.**
Let $K_n(u) = \{j \leq n;\ X_j > u\}$ be the index of the observations outside of the threshold u, and let $N(u) = \#K_n(u)$ be their total number and

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(X_j \leq x)$$

the empirical distribution function, $\overline{\hat{F}_n} = 1 - \hat{F}_n$.

$$\begin{aligned} e_n(u) = \int_u^\infty \overline{\hat{F}_n}(y)dy / \overline{\hat{F}_n}(u) &= \frac{1}{N(u)} \sum_{j \in K_n(u)} (X_j - u) \\ &= \frac{1}{N(u)} \sum_{j=1}^{n} max\{(X_j - u), 0\} \end{aligned}$$

is called the empirical average excess function.

$e_n(u)$ estimates the average excess function $e(u)$ from Section 4.1.
In the rest of this section and in the following one estimators for extreme value characteristics will be discussed, such as the exceedance probabilities $\overline{F}(x) = 1 - F(x)$ for values x or the extreme quantile $F^{-1}(q)$ for $q \approx 1$.
First, only distributions F are considered that are contained in the MDA of a GEV distribution $G_\gamma,\ \gamma \geq 0$. The corresponding random variables are thus unbounded.

Definition 10 (Excess over Threshold)**.**
Let $K_n(u)$ and $N(u)$ be, as before, the index and total number of observations beyond

the threshold u respectively. The excess *over the threshold u is defined as the random variables $Y_l, l = 1, \ldots, N(u)$, with*

$$\{Y_1, \ldots, Y_{N(u)}\} = \{X_j - u;\ j \in K_n(u)\} = \{X^{(1)} - u, \ldots, X^{(N(u))} - u\}.$$

The excesses $Y_l,\ l \leq \mathrm{N}(u)$ describe by how much the observations, which are larger than u, go beyond the threshold u. The POT method assumes that these excesses are the basic information source for the initial data. From the definition it immediately follows that $Y_1, \ldots, Y_{N(u)}$ are i.i.d. random variables with distribution F_u given their random total number $\mathrm{N}(u)$, i.e., the excess distribution from Definition 7 is the actual distribution of the excesses. Due to Theorem 6 it also holds that $F_u(y) \approx W_{\gamma,\beta(u)}(y)$ for a GP distribution $W_{\gamma,\beta(u)}$ and all sufficiently large u.

The problem of estimating the exceedance probability $\overline{F}(x)$ for large x is first considered. A natural estimator is $\overline{\hat{F}}_n(x)$, the cdf at x is replaced with the empirical distribution function. For large x, however, the empirical distribution function varies a lot because it is determined by the few extreme observations which are located around x. The effective size of the sub-sample of extreme, large observations is too small to use a pure non-parametric estimator such as the empirical distribution function. We therefore use the following relationship among the extreme exceedance probability $\overline{F}(x)$, the exceedance probability $\overline{F}(u)$ for a large, but not extremely large threshold and the excess distribution. Due to Definition 7 the excess distribution is

$$\begin{aligned} \overline{F}_u(y) &= \mathrm{P}(X - u > y \mid X > u) = \overline{F}(y+u)/\overline{F}(u), \qquad \text{i.e.} \\ \overline{F}(x) &= \overline{F}(u) \cdot \overline{F_u}(x-u),\ \ u < x < \infty. \end{aligned} \tag{4}$$

For large u and using Theorem 6 we can approximate F_u with $W_{\gamma,\beta}$ for appropriately chosen γ, β. $F(u)$ is replaced with the empirical distribution function $\hat{F}_n(u)$ at the threshold u, for which due to the definition of $\mathrm{N}(u)$ it holds that

$$\hat{F}_n(u) = \frac{n - \mathrm{N}(u)}{n} = 1 - \frac{\mathrm{N}(u)}{n}.$$

For u itself this is a useful approximation, but not for the values x, which are clearly larger than the average sized threshold u. The estimator $1 - \hat{F}_n(x)$ of $\overline{F}(x)$ for extreme x only depends on a few observations and is therefore too unreliable. For this reason the POT method uses the identity (4) for $\overline{F}(x)$ and replaces both factors on the right hand side with their corresponding approximations, whereby the unknown parameter of the generalised Pareto distribution is replaced with a suitable estimator.

Definition 11 (POT Estimator)**.**
The POT estimator $\overline{F}^{\wedge}(x)$ *for the exceedance probability* $\overline{F}(x)$*, for large* x*, is*

$$\overline{F}^{\wedge}(x) = \frac{N(u)}{n}\,\overline{W}_{\hat{\gamma},\hat{\beta}}(x-u) = \frac{N(u)}{n}\left\{1 + \frac{\hat{\gamma}(x-u)}{\hat{\beta}}\right\}^{-1/\hat{\gamma}}, \ u < x < \infty,$$

whereby $\hat{\gamma}, \hat{\beta}$ *are suitable estimators for* γ *and* β *respectively.*

Maximum likelihood estimation of $\hat{\gamma}, \hat{\beta}$

$\hat{\gamma}, \hat{\beta}$ can be, for example, calculated as maximum likelihood estimators from the excesses $Y_1, \ldots, Y_{N(u)}$. First let's consider the case where $\mathrm{N}(u) = m$ is a constant and where $Y_1, \ldots, Y_m$ is a sample of i.i.d. random variables with the distribution $W_{\gamma,\beta}, \gamma > 0,$. Thus $W_{\gamma,\beta}$ is literally a Pareto distribution and has the probability density

$$p(y) = \frac{1}{\beta}\left(1 + \frac{\gamma y}{\beta}\right)^{-\frac{1}{\gamma}-1}, \ x \geq 0.$$

Therefore, the log likelihood function is

$$\ell(\gamma, \beta \mid Y_1, \ldots, Y_m) = -m \log \beta - \left(\frac{1}{\gamma} + 1\right) \sum_{j=1}^{m} \log\left(1 + \frac{\gamma}{\beta} Y_j\right).$$

By maximising this function with respect to γ, β we obtain the maximum likelihood (ML) estimator $\hat{\gamma}, \hat{\beta}$. Analogously we could also define the ML estimator for the parameter of the generalised Pareto distribution using $\gamma \leq 0$.

Theorem 7.
For all $\gamma > -\frac{1}{2}$ it holds for $m \to \infty$

$$\sqrt{m}\left(\hat{\gamma} - \gamma,\ \frac{\hat{\beta}}{\beta} - 1\right) \xrightarrow{\mathcal{L}} N_2(0, D^{-1}),$$

with $D = (1+\gamma)\begin{pmatrix} 1+\gamma & -1 \\ -1 & 2 \end{pmatrix}$, i.e. $(\hat{\gamma}, \hat{\beta})$ are asymptotically normally distributed. In addition they are asymptotically efficient estimators.

In our initial problem $m = \mathrm{N}(u)$ was random. Here the estimators we have just defined, $\hat{\gamma}$ and $\hat{\beta}$, are the conditional ML estimators given $\mathrm{N}(u)$. The asymptotic distribution theory is also known in this case; in order to avoid an asymptotic bias, $\overline{F}$ must fulfill an additional regularity condition. After we find an estimator for the exceedance probability and thus a cdf for large x, we immediately obtain an estimator for the extreme quantile.

Estimating the Value-at-Risk

Definition 12 (POT Quantile Estimator)**.**
The POT Quantile estimator $\hat{x}_q$ *for the q-quantile $x_q = F^{-1}(q)$ is the solution to $\overline{F}^{\wedge}(\hat{x}_q) = 1 - q$, i.e.*

$$\hat{x}_q = u + \frac{\hat{\beta}}{\hat{\gamma}}\left[\left\{\frac{n}{N(u)}(1-q)\right\}^{-\hat{\gamma}} - 1\right].$$

This quantile is actually the Value-at-Risk, so

$$V\hat{a}R_q = u + \frac{\hat{\beta}}{\hat{\gamma}}\left[\left\{\frac{n}{\mathrm{N}(u)}(1-q)\right\}^{-\hat{\gamma}} - 1\right].$$

Comparison to the empirical quantile

These estimators can be compared with the usual sample quantiles. To do this a threshold value u is selected so that exactly k excesses lie beyond u, that is $\mathrm{N}(u) = k > n(1-q)$ and thus $u = X^{(k+1)}$. The POT quantile estimator that is dependent on the choice of u respectively k is

$$\hat{x}_{q,k} = X^{(k+1)} + \frac{\hat{\beta}_k}{\hat{\gamma}_k}\left[\left\{\frac{n}{k}(1-q)\right\}^{-\hat{\gamma}_k} - 1\right],$$

where $\hat{\gamma}_k, \hat{\beta}_k$ is the ML estimator, dependent on the choice of k, for γ and β. The corresponding sample quantile is

$$\hat{x}_q^s = X^{([n(1-q)]+1)}.$$

This is in approximate agreement with $\hat{x}_{q,k}$ when the minimal value $k = [n(1-q)]+1$ is chosen for k. Simulation studies show that the value k_0 of k, which minimises the mean squared error $\mathrm{MSE}(\hat{x}_{q,k}) = \mathsf{E}(\hat{x}_{q,k} - x_q)^2$, is much larger than $[n(1-q)] + 1$, i.e., the POT estimator for x_q differs distinctly from the sample quantile $\hat{x}_q^s$ and is superior to it with respect to the mean squared error when the thresholds u or k are respectively chosen.

We are interested in the threshold u, for which the mean squared error of $\hat{x}_q$ is as small as possible. The error can be split into the variance and the squared bias of $\hat{x}_q$:

$$\mathrm{MSE}(\hat{x}_q) = \mathsf{E}(\hat{x}_q - x_q)^2 = \mathsf{Var}(\hat{x}_q) + \{\mathsf{E}(\hat{x}_q) - x_q\}^2.$$

The Mean Square Error dilemma

Unfortunately the two components of the mean squared error move in opposite directions when the threshold u used in calculating the POT quantile estimators is varied. The following bias variance dilemma needs to be considered:

- when u is too large, there are few excesses $Y_l,\ l \leq \mathrm{N}(u)$, and the estimator's variance is too large,

- when u is too small, the approximation of the excess distribution using a generalised Pareto distribution is not good enough, and the bias $\mathsf{E}(\hat{x}_q) - x_q$ is no longer reliable.

An essential aid in selecting an appropriate threshold u is the average excess plot, which is approximately linear beyond the appropriate threshold. This has already been discussed in Theorem 5, when one considers the relationship between the Fréchet distribution as the asymptotic distribution of the maxima and the Pareto distribution as the asymptotic distribution of the excesses. It is supported by the following result for the Pareto and exponential distributions $W_{\gamma,\beta}, \gamma \geq 0$,.

Theorem 8.

Let Z be a $W_{\gamma,\beta}$ distributed random variable with $0 \leq \gamma < 1$. The average excess function is linear:

$$e(u) = \mathsf{E}\{Z - u | Z > u\} = \frac{\beta + \gamma u}{1 + \gamma}, \quad u \geq 0, \quad \textit{for} \ 0 \leq \gamma < 1.$$

With the usual parametrization of the Pareto distribution $\gamma = \frac{1}{\alpha}$, i.e., the condition $\gamma < 1$ means that $\alpha > 1$ and thus $\mathsf{E}|Z| < \infty$.

This result motivates the following application in choosing the threshold: select the threshold u of the POT estimator so that the empirical average excess function $e_n(v)$ for values $v \geq u$ is approximately linear. An appropriate u is chosen by considering the average excess plots, where it is recommended that the largest points $(X_{(k)}, e_n(X_{(k)}))$, $k \approx n$, along the righthand edge of the plot be excluded, since their large variability for the most part distorts the optical impression.

Estimating the expected shortfall

The expected shortfall is closely related to the average excess function when $u = x_q$,

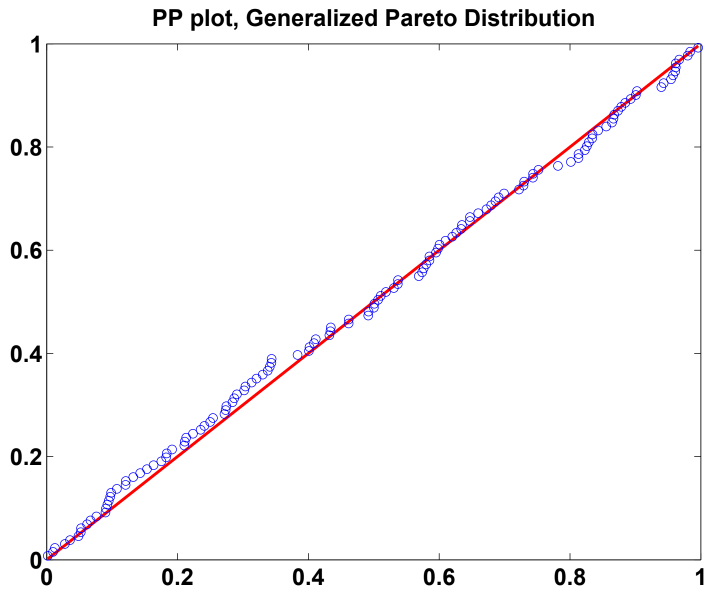

Figure 14: PP plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalized Pareto Distribution with parameter γ = -0.0242 estimated with the POT method, generalised Pareto approach. Time period: 1. January 1998 to 31. December 2007.

as immediately can be seen from the definition:

$$S_q = e(x_q) + x_q.$$

Only the POT estimator for the expected shortfall S_q is considered. Since $F_u(x) \approx W_{\gamma,\beta}(x)$ for a sufficiently large threshold u, it holds from Theorem 5, b) with $\alpha = 1/\gamma$

$$e(v) \approx \frac{\beta + (v-u)\gamma}{1-\gamma} \quad \text{for} \quad v > u.$$

Therefore, for $x_q > u$ we have

$$\frac{S_q}{x_q} = 1 + \frac{e(x_q)}{x_q} \approx \frac{1}{1-\gamma} + \frac{\beta - \gamma u}{x_q(1-\gamma)}.$$

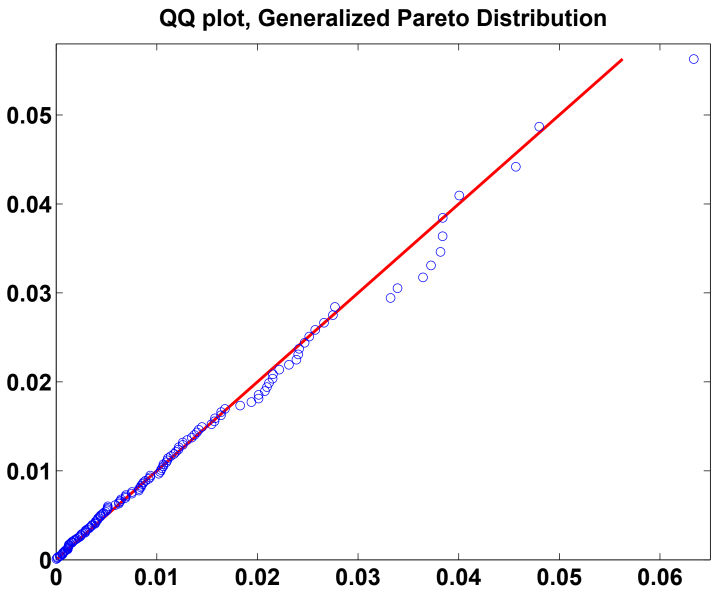

Figure 15: QQ plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalized Pareto Distribution with parameter γ = -0.0242 estimated with the POT method, generalised Pareto approach. Time period: 1. January 1998 to 31. December 2007.

The *POT estimator for the expected shortfall* S_q is thus

$$\hat{S}_{q,u} = \frac{\hat{x}_q}{1-\hat{\gamma}} + \frac{\hat{\beta}-\hat{\gamma}u}{1-\hat{\gamma}},$$

where $\hat{x}_q$ is the POT quantile estimator.

Given that $\hat{x}_q$ is actually the Value-at-Risk, one can rewrite:

$$\hat{S}_{q,u} = \frac{V\hat{a}R_q}{1-\hat{\gamma}} + \frac{\hat{\beta}-\hat{\gamma}u}{1-\hat{\gamma}}.$$

Figure 14 represents the PP plot of 130 tail values of negative log-returns of the portfolio of Bayer, BMW, Siemens and Volkswagen against the Generalised Pareto distribution. Negative log-returns are used, following the same logic from the previious method. The peaks-over-threshold method is used, with the Generalised Pareto approach. The data is fitted to the GPD, and a parameter γ = -0.0242 is estimated. Figure 15 represents the QQ plot of the negative log-returns of the portfolio against the GPD and the method used is the same. Also in the case of the POT method, the empirical data fits quite well to the GPD with parameter γ = -0.0242.

4.2.2 The Hill Estimator Approach

The POT method for estimating the exceedance probability and the extreme quantiles can be used on data with cdf that is in the MDA of a Gumbel or a Fréchet distribution, as long as the expected value is finite. Even for extreme financial data, this estimator seems reasonable based on empirical evidence. A classic alternative to the generalised Pareto approach POT estimator is the Hill estimator. It is only useful for distributions with slowly decaying tails, such as those in the MDA of the Fréchet distribution, and in simulations it often performs worse in comparison to the POT estimator.
In this section it is assumed that the data $X_1, \ldots, X_n$ are i.i.d. with a distribution function F in the MDA of $G_{1,\alpha}$ for some $\alpha > 0$. Due to Theorem 3 this is the case when $\overline{F}(x) = x^{-\alpha}L(x)$ with a slowly varying function L. The tapering behaviour of $\overline{F}(x) = \mathrm{P}(X_t > x)$ for increasing x is mainly determined by the so called *tail exponents* α. The starting point of the Hill method is the following estimator for α.

Definition 13 (Hill estimator)**.**
$X^{(1)} \geq X^{(2)} \geq \ldots \geq X^{(n)}$ *are the order statistics in decreasing order. The* Hill estimator $\hat{\alpha}_H$ *of the tail exponents* α *for a suitable* $k = k(n)$ *is*

$$\hat{\alpha}_H = \left(\frac{1}{k} \sum_{j=1}^{k} \log\ X^{(j)} - \log\ X^{(k)} \right)^{-1}.$$

The form of the estimator can be seen from the following simple special case. In general it holds that $\overline{F}(x) = L(x)/(x^{\alpha})$, but here it is assumed that with a fixed

$c > 0$ $L(x) = c^\alpha$ is constant. Set $V_j = \log(X_j/c)$, it holds that

$$\mathrm{P}(V_j > v) = \mathrm{P}(X_j > ce^v) = \overline{F}(ce^v) = \frac{c^\alpha}{(ce^v)^\alpha} = e^{-\alpha v},\ y \geq 0,$$

$V_1, \ldots, V_n$ are therefore independent exponentially distributed random variables with parameter α. As is well known it holds that $\alpha = (\mathsf{E}V_j)^{-1}$, and the ML estimator $\hat{\alpha}$ for α is $1/\overline{V}_n$, where $\overline{V}_n$ stands for the sample average of $V_1, \ldots, V_n$, thus,

$$\hat{\alpha} = \frac{1}{\overline{V}_n} = \left(\frac{1}{n}\sum_{j=1}^{n} \log(X_j/c)\right)^{-1} = \left(\frac{1}{n}\sum_{j=1}^{n} \log\ X^{(j)} - \log\ c\right)^{-1},$$

where for the last equation only the order of addition was changed. $\hat{\alpha}$ is already similar to the Hill estimator. In general it of course only holds that $\overline{F}(x) \approx \frac{c^\alpha}{x^\alpha}$ for sufficiently large x. The argument for the special case is similar for the largest observations $X^{(1)} \geq X^{(2)} \geq \ldots \geq X^{(k)} \geq u$ beyond the threshold u, so that only the k largest order statistics enter the definition of the Hill estimator.

The Hill estimator is consistent, that is it converges in probability to α when $n, k \to \infty$ such that $k/n \to 0$. Under an additional condition it can also be shown that $\sqrt{k}(\hat{\alpha}_H - \alpha) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \alpha^2)$, i.e., $\hat{\alpha}_H$ is asymptotically normally distributed.

Similar to the POT estimator when considering the Hill estimator the question regarding the choice of the threshold $u = X^{(k)}$ comes into play, since the observations located beyond it enter the estimation. Once again we have a bias variance dilemma:

- When k is too small, only a few observations influence $\hat{\alpha}_H$, and the variance of the estimator, which is α^2/k asymptotically, becomes too large,

- when k is too large, the assumption underlying the derivation of the estimator, i.e., that $L(x)$ is approximately constant for all $x \geq X^{(k)}$, is in general not well met and the bias $\mathsf{E}\hat{\alpha}_H - \alpha$ becomes too large.

Based on the fundamentals of the Hill estimator for the tail exponents α we obtain direct estimators for the exceedance probability $\overline{F}(x)$ and for the quantiles of F.

Since $\overline{F}(x) = x^{-\alpha}L(x)$ with a slowly varying function L, it holds for large $x \geq X^{(k)}$ that:

$$\frac{\overline{F}(x)}{\overline{F}(X^{(k)})} = \frac{L(x)}{L(X^{(k)})}\left(\frac{X^{(k)}}{x}\right)^{\alpha} \approx \left(\frac{X^{(k)}}{x}\right)^{\alpha}, \tag{5}$$

Because exactly one portion k/n of the data is larger or equal to the order statistic $X^{(k)}$, this is the $(1 - k/n)$ sample quantile. Therefore, the empirical distribution function takes on the value $1 - k/n$ at $X^{(k)}$, since it uniformly converges to the distribution function F, for sufficiently large n, a k that is not too large in comparison to n yields: $F(X^{(k)}) \approx 1 - k/n$, i.e., $\overline{F}(X^{(k)}) \approx k/n$. Substituting this into (5), a *Hill estimator is obtained for the exceedance probability* $\overline{F}(x)$:

$$\overline{F}^{\wedge}_{H}(x) = \frac{k}{n}\left(\frac{X^{(k)}}{x}\right)^{\hat{\alpha}_H}$$

By inverting this estimator, the *Hill quantile estimator* is obtained for the $q-$ quantile x_q with $q \approx 1$:

$$\begin{aligned}\hat{x}_{q,H} &= X^{(k)}\left\{\frac{n}{k}(1-q)\right\}^{-1/\hat{\alpha}_H} \\ &= X^{(k)} + X^{(k)}\left[\left\{\frac{n}{k}(1-q)\right\}^{-\hat{\gamma}_H} - 1\right]\end{aligned}$$

with $\hat{\gamma}_H = 1/\hat{\alpha}_H$, where the second representation clearly shows the similarities and differences to the POT quantile estimator.

In Value-at-Risk notation,

$$V\hat{a}R_{q,H} = X^{(k)} + X^{(k)}\left[\left\{\frac{n}{k}(1-q)\right\}^{-\hat{\gamma}_H} - 1\right]$$

The Hill quantile estimator is also illustrated with the help of a Matlab program.

5 Conclusion

Extreme Value Theory provides a scientific approach for a difficult practical problem, namely that of estimating extreme risks when only little data is available. Due to its assumptions of heavy tails and the extreme value distributions that are used, EVT outperforms other methods which tend to be very imprecise or underestimate the real risk. For these reasons, EVT is important in the theory and practice of a category of problems that is normally hard to address. However, there are limitations to its efectiveness, and the main problem of the lack of data still leaves the precision matter uncertain.

The paper represents an overview of the building blocks and methods of EVT. Two relevant risk measures have been intorduced, the Value-at-Risk and the expected shortfall. The data used for illustration has been described. A presentation of the main methods of EVT has been made, with the support of graphical tools. In the block maxima method, the key aspect is finding the limit distribution for the maxima, which facilitates thereafter the estimation of the risk through the suggested risk measures. In the case of the peaks-over-threshold method, two approaches are presented. The first one is the Generalised Pareto approach which earns expressions for Value-at-Risk and expected shortfall in parametric fashion. The approach based on the Hill Estimator is semi-parametric and the most important result presented here is the formula for Value-at-Risk estimation.

The main resource used, as pointed out throughout the paper, is the book by Franke, Härdle and Hafner (2008). A review of the relevant literature has also been performed. The main contribution of this paper is represented by the Matlab implementation of statistical methods. The programs and plots provide a background for the study of EVT using software packages.

References

Alfarano, S. and Lux, T. (2004), Extreme value theory as a theoretical background for power law behavior, *Claudio Cioffi-Revilla, editor, Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe.*

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999), Coherent Measures of Risk, *Mathematical Finance, 9(3), 203 – 228.*

Dowd, K. (2005), *Measuring Market Risk*, Wiley Finance.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Embrechts, P., Resnick, S. and Samorodnitsky, G. (1999), Extreme Value Theory as a Risk Management Tool, *North American Actuarial Journal, Vol. 3, No. 2, pp. 30-41.*

Embrechts, P., Resnick, S. and Samorodnitsky, G. (1998), Living on the Edge, *RISK Magazine, 11(1), 96 – 100.*

Fernandez, V. (2003), Extreme Value Theory and Value at Risk, *Revista de Analisis Economico, Vol. 18 No.1, pp.57-85.*

Franke, J., Härdle, W., and Hafner, C. (2008), *Statistics of Financial Markets: an Introduction. 2nd extended ed.*, Springer, Heidelberg.

Gilli, M. and Këllezi, E. (2006), An Application of Extreme Value Theory for Measuring Financial Risk, *Computational Economics 27(1), 2006, 1 – 23.*

Härdle, W., Simar, L.(2003), *Applied Multivariate Statistical Analysis*, Springer, Heidelberg.

Longin, F.M. (2000), From Value at Risk to Stress Testing: The Extreme Value Approach, *Journal of Banking & Finance 24, 1097 – 1130.*

McNeil, A.J., (1999), Extreme Value Theory for Risk Managers, *Internal Modeling and CAD II. London: Risk Books, pp. 93 - 113.*

McNeil, A.J., Frey, R. (2000), Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach, *Journal of Empirical Finance.*

Tsevas, G., Panaretos, J. (1998), Extreme Value Theory and its Applications to Financial Risk Management, *Proceeding of the 4th Hellenic European Conference on Computer Mathematics and its Applications, 509-516.*

List of Figures

List of Tables

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated resources. All passages, which are literally or in general matter taken out of publications or other resources, are marked as such.

Student's signature:

Name (in capitals):

Date of submission: