Technische Universiteit Delft
Faculteit Elektrotechniek, Wiskunde en Informatica
Delft Institute of Applied Mathematics

Risk measures of heavy-tailed data

Thesis on behalf of
Delft Institute of Applied Mathematics
as part of acquiring

the grade of

BACHELOR OF SCIENCE
in
Applied Mathematics

by

Marco Kirana

Delft, The Netherlands
July 2016

BSc thesis Applied Mathematics

"Risk measures of heavy-tailed data"

Marco Kirana

Technische Universiteit Delft

Supervisor

Dr. J.-J. Cai

Other committee members

Drs. E.M. van Elderen

...

Dr. D.C. Gijswijt

...

July, 2016

Delft

Abstract

The main aim for this project is to measure the risk of heavy-tailed data. The mean does not need to exist for this data. This study is part of the statistical branch called 'Extreme value theory'. In this theory we focus on the tail of the distribution, where the outcomes are extreme events. We start with explaining the basics of Extreme value theory and point out the importance of this study in reality. We then present and analyse the 'Generalized Pareto' distribution and the 'Peak over threshold' method which is necessary to analyse the tail. Then we introduce the reader to a couple of Risk measurement tools in order to quantify the risk of the data. At last we have a real application with Tsunami data of Japan where we make use of our findings in previous chapter to measure the risk of this data. This project shows that we can measure the risk of heavy-tailed data.

Contents

Chapter 1

Introduction

How disastrous would it be if a flood occurred in your country? Or a Tsunami, earthquake etc. The events above are called 'extreme' events. When such phenomena happen the results are usually very destructive, but the probability that such an extreme event occurs is of course very small and might not even happen in an entire lifetime. However if we would entirely neglect it and we are unlucky enough to experience one of those events the after-effect can be horrendous.
The goal of this project is to measure the risk of heavy-tailed data and even make conclusions on which type of event is more disastrous by just looking at the output of the quantification.

In chapter 2 we introduce and work with the Peak over Threshold method. This method shows that for a certain assumption of the random variable X we can state that the tail of X is approximately distributed as the generalized Pareto distribution (GPD). Furthermore we make use of the method of maximum likelihood and the method of moments for parameter estimation of GPD and the tail of X. We end chapter 2 with the estimation of a 'high' quantile.
In chapter 3 we are going to present 2 types of risk measurements called Value at Risk (VaR) and Expected shortfall (ES). We work with ES and a modification of it in order to measure the risk of our data.
In chapter 4 we have a real application of Tsunami data of Japan where we quantify the risk of the data containing the maximum water height of Tsunami's in the years 1403-2011. At the end of this project we have shown that we are able to measure risk of heavy-tailed data.

Chapter 2

Tail modeling

Our focus in this project is the tail of heavy-tailed data, where all the 'extreme' events are. We wish to model the tail and have knowledge on how it behaves. First we are going to show that the tail behaviour of a certain distribution can be captured. Then we are going to make use of the **Generalized Pareto Distribution (*GPD*)** to estimate a parameter that is needed in the tail study process. We begin with defining a distribution for the maximum of a set of i.i.d (independent identically distributed) random variables. Let $X_1, X_2, ..., X_n$ be i.i.d random variables, each with the distribution function $F(x) = P(X_i < x)$ and define $X_{\max} = \max\{X_1, X_2, ..., X_n\}$. Then we have for the distribution function of $X_{\max}$:

$$\begin{aligned} F_{X_{\max}} &= P(X_1 < x, X_2 < x, ..., X_n < x) \\ &= P(X_1 < x)P(X_2 < x) \cdot ... \cdot P(X_n < x) = F^n(x). \end{aligned}$$

Define $x^* = \sup\{x : F(x) < 1\}$ as the supremum (or right end point) and note that:

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 0 & \text{if } x < x^* \\ 1 & \text{if } x \geq x^* \end{cases}$$

The limit above has a degenerate behaviour (a limit with only 2 outcomes), we need to standardize $X_{\max}$ in some way...
We are going to make use of (1.1.1) from [L. de Haan(2006)], we will denote this as a theorem:

Theorem 2.1 *Suppose that there exists a sequence of constants $a_n > 0$ and b_n real $(n = 1, 2, ...)$, such that*

$$\lim_{n\to\infty} P\left(\frac{X_{\max} - b_n}{a_n}\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x),$$

*with $G(x)$ a non-degenerate distribution. $G(x)$ is called the **Extreme value distribution**. If all the above holds then we say that F is in **the domain attraction of G**, $F \in D(G)$. If F is the distribution function of a random variable X we can also say hat $X \in D(G)$.*

If F is in the domain of G we say that the behaviour of the tail of F is captured by G. This means that all the information of the tail can be retrieved from G. Now we are going to identify the Extreme value distribution G, making use of a theorem (1.1.3) of [L. de Haan(2006)] by Fisher and Tippet(1928) and Gnedenko(1943):

Theorem 2.2 *The class of extreme value distributions is $G_\gamma(ax + b)$ with $a > 0$, b real, where*

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}), \quad 1 + \gamma x > 0, \tag{2.1}$$

with γ real and where for $\gamma = 0$ the right hand side is interpreted as $exp(-e^{-x})$.

Definition 2.3 *γ in (2.1) is called the **Extreme value index**.*

Remark (1) *If we can estimate γ we can capture the behaviour of the tail of F.*

The Extreme value index (EVI) is a parameter γ that measures the weight of the right tail $\overline{F}(x) = 1 - F(x) = P(X > x)$ for large values x. There are 3 different categories for γ which we will denote below:

(1) $\boldsymbol{\gamma < 0}$:
This means that the right tail is light, F has a finite right endpoint x^*.
(2) $\boldsymbol{\gamma = 0}$:
This means that the right tail is of an exponential type, the right endpoint can be either finite or infinite.
(3) $\boldsymbol{\gamma > 0}$:
This means that the right tail is heavy, F has an infinite endpoint.

Reminder
We focus on data that is heavy-tailed, so we only look at cases where $\gamma > 0$ which corresponds to (3) above.

To continue with the tail modeling we are going to make use of Theorem 1.2.5 (Only the $\gamma > 0$ part) from [L. de Haan(2006)]:

Theorem 2.4 *The distribution function F is in the domain of attraction of the extreme value distribution $G \iff$ for some positive function σ,*

$$\lim_{t \to x^*} \frac{1 - F(t + x\sigma(t))}{1 - F(t)} = (1 + \gamma x)^{-\frac{1}{\gamma}}, \tag{2.2}$$

for all x with $1 + \gamma x > 0$. If (2.2) holds for some $\sigma > 0$, then it also holds with $\sigma(t) = \gamma t$ for $\gamma > 0$.

Rewriting (2.2) gives the following theorem:

Theorem 2.5 *$X \in D(G) \iff$ there exists a positive function $\sigma(t)$ such that*

$$\lim_{t \to x^*} P\left(\frac{X - t}{\sigma(t)} > x | X > t\right) = (1 + \gamma x)^{-\frac{1}{\gamma}}, \tag{2.3}$$

for $0 < x < \frac{1}{\max(0, -\gamma)}$.

Theorem 2.5 is the base of the **Peak over Threshold approach**, it states that the conditional distribution of $\frac{X-t}{f(t)}$ given that $X > t$ is GPD (limit) distributed. If we subtract the right side of the limit equation in Theorem 2.5 from the number 1 we have: $1 - (1 + \gamma x)^{-\frac{1}{\gamma}}$. This is the distribution function of the $GPD_{(0,1,\gamma)}$ distribution which we will introduce now.
The probability density function of the $GPD_{(\mu,\sigma,\gamma)}$ is:

$$f_{(\mu,\sigma,\gamma)}(x) = \begin{cases} \frac{1}{\sigma}(1 + \frac{\gamma(x-\mu)}{\sigma})^{-(\frac{1}{\gamma}+1)} & \text{if } \gamma \neq 0 \\ \frac{1}{\sigma}\exp(-\frac{x-\mu}{\sigma}) & \text{if } \gamma = 0 \end{cases} \qquad \begin{cases} x \in [\mu, \infty), & \text{when } \gamma \geq 0 \\ x \in [\mu, \mu - \frac{\sigma}{\gamma}], & \text{when } \gamma < 0 \end{cases}$$

For $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ and $\gamma \in (-\infty, \infty)$.
The distribution function of the $GPD_{(\mu,\sigma,\gamma)}$ is:

$$F_{(\mu,\sigma,\gamma)}(x) = \begin{cases} 1 - (1 + \frac{\gamma(x-\mu)}{\sigma})^{-\frac{1}{\gamma}} & \text{if } \gamma \neq 0 \\ 1 - \exp(-\frac{x-\mu}{\sigma}) & \text{if } \gamma = 0 \end{cases} \qquad \begin{cases} x \in [\mu, \infty), & \text{when } \gamma \geq 0 \\ x \in [\mu, \mu - \frac{\sigma}{\gamma}], & \text{when } \gamma < 0 \end{cases}$$

For $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ and $\gamma \in (-\infty, \infty)$.
As you can see the GPD has 3 parameters: μ, σ and γ which are called the location, scale and shape parameter respectively. The location parameter is used to shift the distribution, the scale parameter is used to make the distribution more spread out and the shape parameter changes the 'shape' of the distribution. The higher the value of γ the slower the tail of the distribution decays. Now let's discuss the different types of distributions that can occur for different values of γ. If $\gamma > 0$, then the GPD changes to a parametrized Pareto distribution, for $\gamma = 0$ we have a exponential distribution and for $\gamma < 0$ we have a Pareto type II distribution.

To prevent confusion in the future we are going to introduce notation when denoting random variables:

Notation
We are going to use Y when we refer to a random variable that is $GPD_{(0,\sigma,\gamma)}$ distributed.
For X we state only that $F \in D(G)$ for F the distribution function of X (thus $X \in D(G)$).
For both X and Y we use x as the realizations of the random variables.

The notation above will be stated every time we work with those random variables for extra clarification. Let's continue with rewriting Theorem 2.5 such that we can clearly see what the POT-method states. Let $y = x\sigma(t) + t$ in Theorem 2.5, then we roughly have for t large:

$$P(X > y | X > t) \approx \left(1 + \frac{\gamma(y-t)}{\sigma(t)}\right)^{-\frac{1}{\gamma}}. \tag{2.4}$$

Note that $1 - (1 + \frac{\gamma(y-t)}{\sigma(t)})^{-\frac{1}{\gamma}}$ is the distribution function of the $GPD_{(t,\sigma(t),\gamma)}$ distribution.

Remark (2) *The above shows that if we have a certain threshold t high we can model the tail of a distribution with the GPD.*

At last we are going to find an approximation of the conditional probability density function $f_X(X|X > t)$. Starting from (2.4), differentiating left and right and simple equation manipulations gives us:

$$f_{X|X>t}(y) \approx \frac{1}{\sigma(t)}(1 + \frac{\gamma}{\sigma(t)}(y-t))^{-(\frac{1}{\gamma}+1)}. \tag{2.5}$$

(2.5) is necessary in order to find the correct risk measure in chapter 3. Our next step is to estimate the parameters of Y and the parameter for the tail of X.

2.1 Parameter estimation of Y

We are going to estimate the 2 parameters of Y that is $GPD_{(0,\sigma,\gamma)}$ distributed. We are going to make use of 2 methods to obtain the estimators for the parameters:

- The method of Maximum Likelihood estimation (MLE)
- The method of Moments (MOM)

The first one is the most widely used for statistical estimation but can be troublesome to calculate. The second method is the easy to compute but the estimators can be biased. The MLE is asymptotically

efficient and does better than the MOM in terms of smaller error when both are unbiased. In the future of this Project we will make use of both the MLE and MOM.

2.1.1 Making use of the Maximum Likelihood Estimation method (MLE)

The maximum likelihood estimation method is based on selecting values of the model parameters such that the likelihood function will be maximized. It maximizes the probability of the sample data given the chosen probability distribution model.
For $Y \sim GPD_{(0,\sigma,\gamma)}$ we have the following probability density function:

$$f_{(0,\sigma,\gamma)}(x|\sigma,\gamma) = \frac{1}{\sigma}(1+\frac{\gamma x}{\sigma})^{-(\frac{1}{\gamma}+1)},$$

where x is a realisation of Y and x_i's are realisations of the Y_i's respectively.
So in our case with the n random variables $Y_1, Y_2, ..., Y_n$ our joint density function is the product of their marginal densities. We have the likelihood function $L(\sigma,\gamma)$:

$$L(\sigma,\gamma) = f_{(0,\sigma,\gamma)}(x_1,x_2,...,x_n|\sigma,\gamma) = \prod_{i=1}^{n}\frac{1}{\sigma}(1+\frac{\gamma x_i}{\sigma})^{-(\frac{1}{\gamma}+1)}$$

Taking the logarithm of $L(\sigma,\gamma)$ above gives the log-likelihood function $l(\sigma,\gamma)$:

$$l(\sigma,\gamma) = \sum_{i=1}^{n} log\left(\frac{1}{\sigma}(1+\frac{\gamma x_i}{\sigma})^{-(\frac{1}{\gamma}+1)}\right)$$

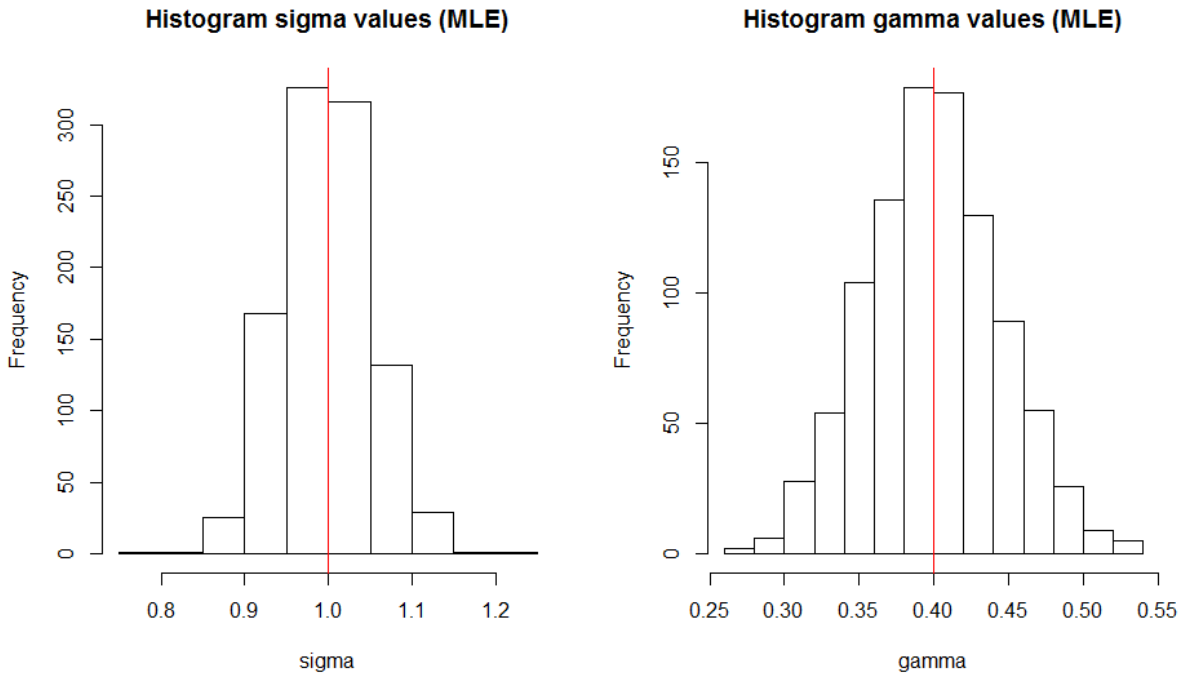
Taking partial derivatives with respect to σ and γ and setting these equations equal to zero results in the following equation system:

$$\begin{cases} \sum_{i=1}^{n}\frac{x_i}{\sigma+\gamma x_i} = \frac{1}{\gamma(\gamma+1)}\sum_{i=1}^{n} log(1+\frac{\gamma x_i}{\sigma}) \\ \sum_{i=1}^{n}\frac{x_i}{\sigma+\gamma x_i} = \frac{n}{1+\gamma} \end{cases}$$

Simplify a little bit with substitution:

$$\begin{cases} n = \frac{1}{\gamma}\sum_{i=1}^{n} log(1+\frac{\gamma x_i}{\sigma}) \\ \sum_{i=1}^{n}\frac{x_i}{\sigma+\gamma x_i} = \frac{n}{1+\gamma} \end{cases} \tag{2.6}$$

The equation system in (2.6) cannot be written in explicit form. The best thing that we can do now is use those equations as our estimators. We are going to make use of the statistical software R to estimate σ and γ. We start with a sample size of $n = 1000$, and let $\sigma = 1$, $\gamma = 0.4$. We start with simulating n values from the $GPD_{(0,1,0.4)}$ distribution which will be our sample data and solve the equations above in terms of σ and γ, which are estimates. We repeat this process a 1000 times which results in 1000 predicted values of our parameters that are plotted in histograms in Figure 2.1:



These are histograms of the values of the $\hat{\sigma}$'s and $\hat{\gamma}$'s given through simulation. The red lines indicate the true values of the parameters.

Figure 2.1: Histograms of the parameter estimations with MLE. The simulation-data is obtained from the $GPD_{(0,1,0.4)}$ distribution.

We can clearly see that the mean of the $\hat{\sigma}$'s and $\hat{\gamma}$'s are close to their true values. The method of maximum likelihood estimation is an accurate method for parameter estimation of Y.

2.1.2 Making use of the Method of Moments (MOM)

The method of moments estimation is based on the law of large numbers which states that the average of all the results obtained from a large number of trials should be close to the expected value, and will converge to the expected value when the number of trials reaches infinity.
The advantage of using the method of moments is that the estimators can be easily expressed in explicit form. We have seen that with the MLE method we needed a computer to calculate the estimators, which is not necessary with the MOM.

Suppose we have i.i.d random variables $Y_1, Y_2, ..., Y_n \sim GPD_{(0,\sigma,\gamma)}$. Now we are going to find the first and second moment which are defined as the expectation of Y, $\mathrm{E}[Y]$, and the expectation of Y^2, $\mathrm{E}[Y^2]$ respectively.

The mean and variance of the $GPD_{(0,\sigma,\gamma)}$ distribution:

$$E[Y] = \frac{\sigma}{1-\gamma}$$

$$Var[Y] = \frac{\sigma^2}{(1-\gamma)^2 \cdot (1-2\gamma)}$$

Note that the variance needs to be positive. The only part that can make the variance negative is if $(1-2\gamma)$ is negative, so we have to make the constraint $\boldsymbol{\gamma < 0.5}$.

E[Y] above is our first moment. Now we only need the second moment to start with the method of moments. To obtain the second moment we start from the definition of variance. Simple calculations give:

$$\mathrm{E}[Y^2] = \frac{2\sigma^2}{(1-\gamma)(1-2\gamma)}$$

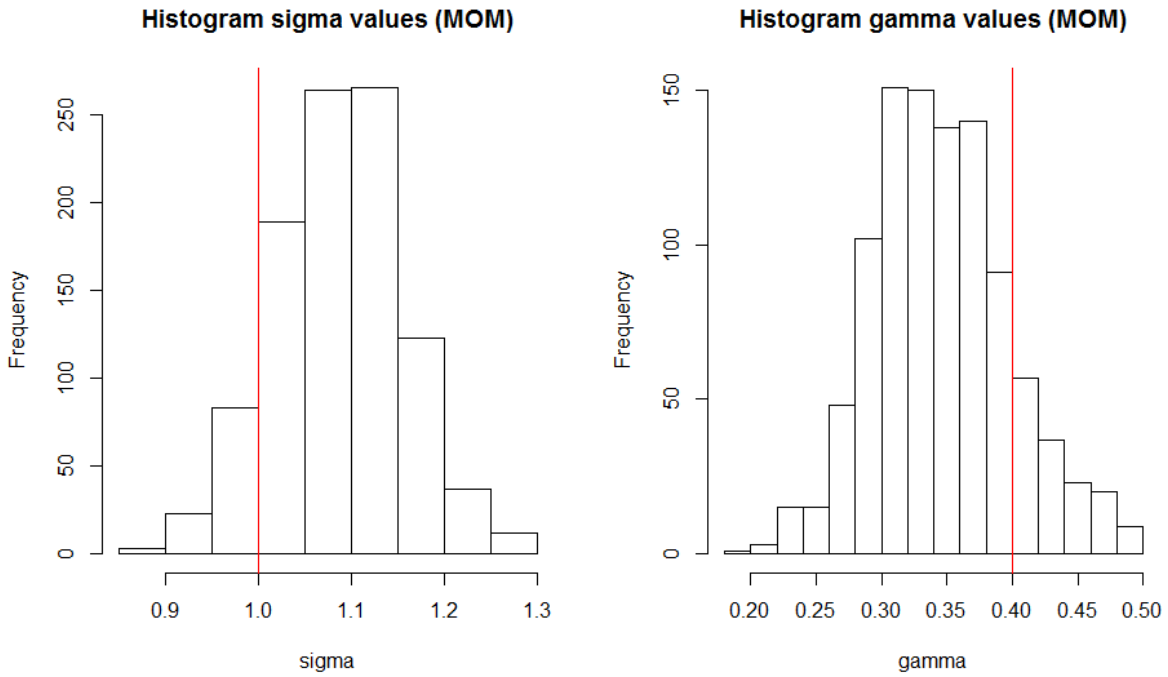
Let's name the first and the second moment μ_1 and μ_2 respectively, so:

$$\mu_1 = \frac{\sigma}{1-\gamma}$$
$$\mu_2 = \frac{2\sigma^2}{(1-\gamma)(1-2\gamma)}$$

Expressing σ and γ in terms of μ_1 and μ_2 gives us the following estimators:

$$\hat{\sigma} = \frac{\mu_1\mu_2}{2(\mu_1 - \mu_1^2)} \tag{2.7}$$
$$\hat{\gamma} = \frac{\mu_2 - 2\mu_1^2}{2(\mu_2 - \mu_1^2)} \tag{2.8}$$

Now we are going to obtain our estimations through simulation, again we make use of R. We choose the values of σ and γ to be 1 and 0.4 respectively. We start with a sample size of $n = 1000$ and let $\sigma = 1$, $\gamma = 0.4$. We start with simulating n values from the $GPD_{(0,1,0.4)}$ distribution which will be our sample data. From this sample data we calculate our parameter estimations with (2.7) and (2.8). This process will be repeated a 1000 times which results in 1000 predicted values of our parameters. We have the following results, shown in the histograms in Figure 2.2:



These are histograms of the values of the $\hat{\sigma}$'s and $\hat{\gamma}$'s given through simulation. The red lines indicate the true values of the parameters.

Figure 2.2: Histograms of the parameter estimations with MOM. The simulation-data is obtained from the $GPD_{(0,1,0.4)}$ distribution.

The mean of the $\hat{\sigma}$'s and $\hat{\gamma}$'s are close to their true values. This means that the method of moments is also an accurate way for parameter estimation of Y with $\gamma < 0.5$.

2.2 Parameter estimation for the tail of X

We are going to estimate the shape parameter γ for the tail of a random variable $X \in D(G)$. Estimating γ gives us the behaviour of the tail. In this section the sample data is from the Generalized Extreme Value (GEV) distribution, which has distribution function:

$$F_{(\mu,\sigma,\gamma)}(x) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\} \qquad \begin{cases} x \in [\mu - \frac{\sigma}{\gamma}, \infty), & \text{when } \gamma > 0 \\ x \in (-\infty, \infty), & \text{when } \gamma = 0 \\ x \in (-\infty, \mu - \frac{\sigma}{\gamma}], & \text{when } \gamma < 0 \end{cases}$$

$\mu \in \mathbb{R}, \sigma > 0$, and $\gamma \in \mathbb{R}$.
We need the $GEV_{(0,\sigma,\gamma)}$ to be heavy-tailed so we set $\gamma > 0$, this corresponds to the Fréchet distribution.

2.2.1 Modeling tails with the Maximum Likelihood Estimation method (MLE)

Remember the system of equations (2.6) which we had to solve to get our estimators. A problem that we can encounter is that the term $\frac{\gamma x_i}{\sigma}$ in the logarithm can get very large. This will lead to problems in finding an estimation for the parameters in R. When we focus on the tails we work with realisations x_i that are very large, this means that we need a method to deal with this problem. We are going to use a method in [Zhou(2009)]. Making use of a threshold t is necessary to focus on the tails of

the distribution, see Remark (2). First we define k:= k(n) as a sequence of integers with $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$ when $n \to \infty$. Define $X_{1,n}, X_{2,n}, ..., X_{n,n}$ as the order statistics of the i.i.d sample $X_1, X_2, ..., X_n$ ($X_{1,n} < ... < X_{n,n}$). We take $t = X_{n-k,n}$ ($k << n$) as the threshold, and we define the **empirical excesses** as $Z_i := X_{n-i+1,n} - X_{n-k,n}$ for $i = 1, ..., k$. Note that from (2.2) we can say that $X - t|X > t \sim GPD_{(0,\sigma(t),\gamma)}$, so in our case: $Z_i \sim GPD_{(0,\sigma(t),\gamma)}$ for $i = 1, ..., n$ and $t = X_{n-k,n}$.

We are going to estimate our parameters with Z_i as our random sample instead of the X_i sample. Note that the Z_i's are i.i.d because they are a linear combination of the X_i's. So now (2.6) becomes:

$$\begin{cases} \frac{1}{k}\sum_{i=1}^{k} log(1 + \frac{\gamma}{\sigma(t)} z_i) = \gamma \\ \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1+\frac{\gamma}{\sigma(t)} z_i} = \frac{1}{1+\gamma} \end{cases} \tag{2.9}$$

With $z_i, ..., z_k$ realisations of $Z_1, ..., Z_k$ respectively.
Substituting the first equation of (2.9) into the second one gives us the following equation:

$$\left(1 + \frac{1}{k}\sum_{i=1}^{k} log(1 + \frac{\gamma}{\sigma(t)} z_i)\right) \cdot \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + \frac{\gamma}{\sigma(t)} z_i} = 1$$

Now let $l = \frac{\gamma}{\sigma(t)}$ and define:

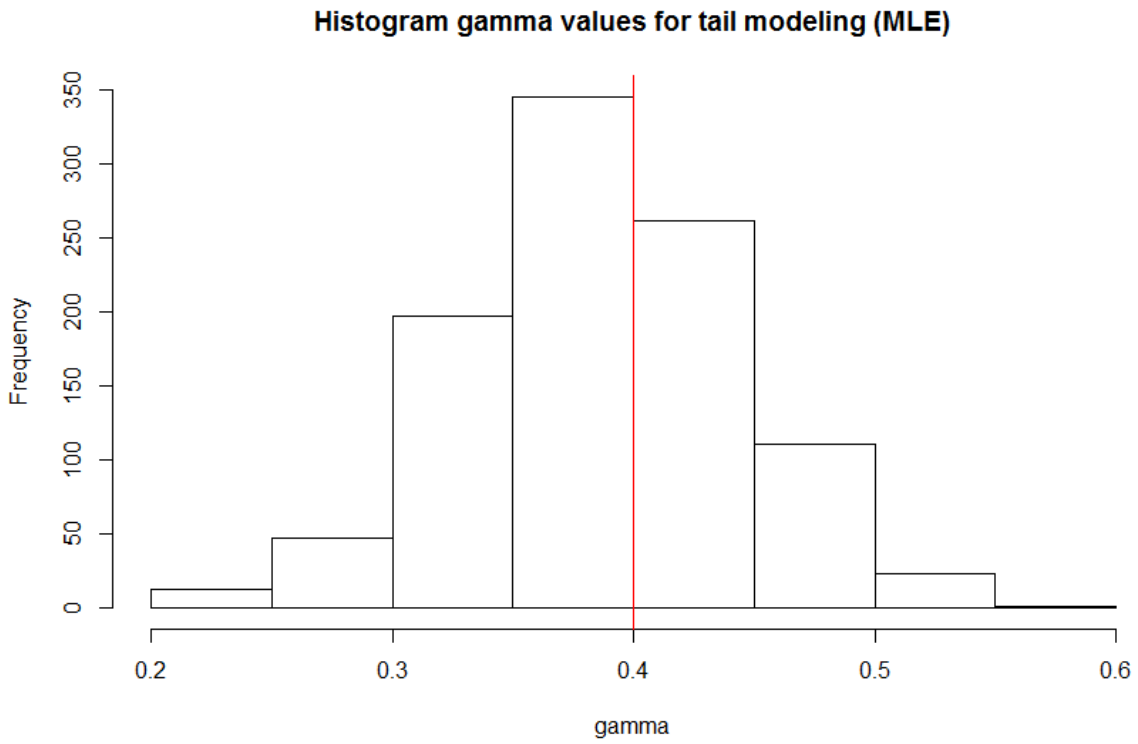
$$\begin{aligned} f_n(l) &= \left(1 + \frac{1}{k}\sum_{i=1}^{k} log(1 + l z_i)\right) \\ g_n(l) &= \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + l z_i} \\ h_n(l) &= f_n(l) g_n(l) - 1 \end{aligned}$$

This means that finding the root l^* of $h_n(l^*) = 0$ gives us:

$$\hat{\gamma} = f_n(l^*) - 1 \tag{2.10}$$

$$\hat{\sigma}(t) = \frac{\hat{\gamma}}{l^*} \tag{2.11}$$

For the simulation we set $n = 5000$ as the sample size and simulate the sample data from the $GEV_{(0,1,0.4)}$ distribution. To determine the best value of k we look at Figure 2.5 at the end of this section. For $k = 600$ we see that the mean of all the $\hat{\gamma}$'s is close to 0.4 thus we choose this value of k. For the simulation we compute (2.10) from the sample data and repeat this process a 1000 times which gives us 1000 predicted values of γ. The results are shown in the histogram in Figure 2.3:



This is the histogram of the $\hat{\gamma}$'s resulted through 1000 simulations. The red line indicates the true value of γ.

Figure 2.3: Histogram of the parameter estimations with MLE for tail modeling, empirical excesses Z_i as the sample data

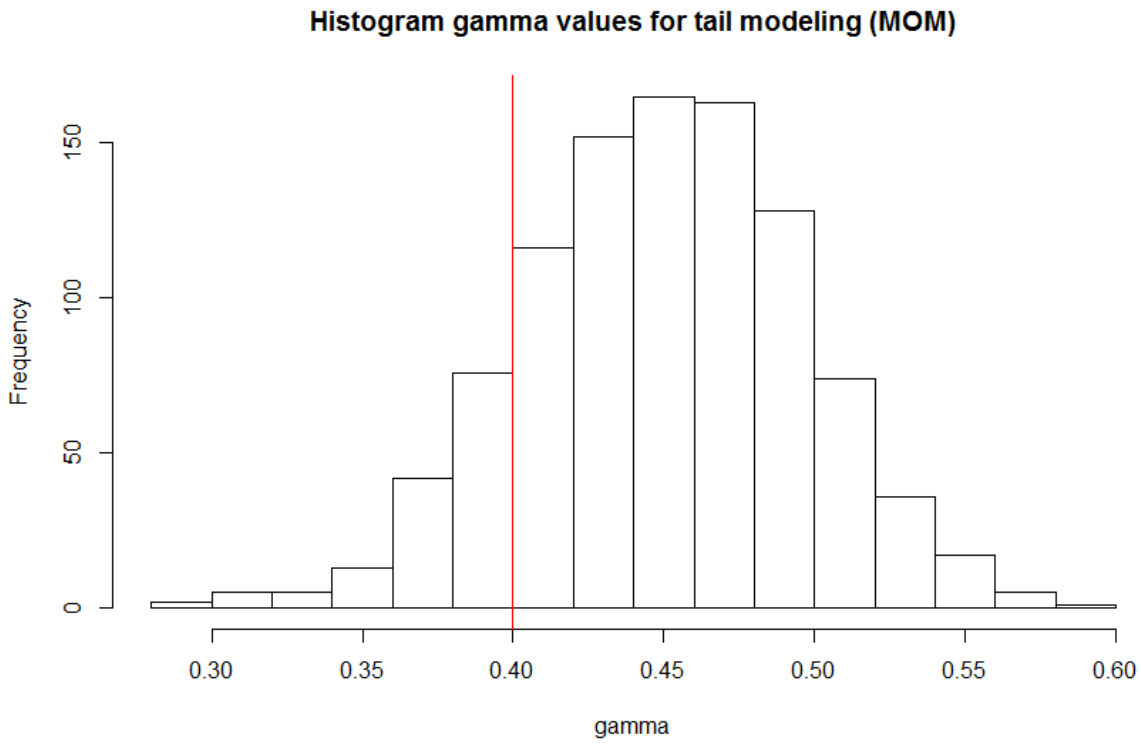
The mean of the $\hat{\gamma}$'s has an approximate value of 0.38 which is close to the true value of 0.4. So we conclude with the help of the MLE-method that the tail of a $GEV_{(0,1,0.4)}$-distribution can be modeled by the GPD.

2.2.2 Modeling tails with the Method of Moments (MOM)

When we are making use of the MOM to model tails we need a different approach compared to section 2.1.2. For the estimation of the shape parameter we are going to use (3.5.9) in [L. de Haan(2006)]:

$$\hat{\gamma}^{\text{MOM}} = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}, \tag{2.12}$$

With $M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1}(log(X_{n-i,n}) - log(X_{n-k,n})))^j \quad j = 1, 2.$
First we are going to sort the data of our total sample (which has size $n = 5000$ and is $GEV_{(0,1,0.4)}$ distributed). To determine the best value of k we look at Figure 2.5. We chose $k = 600$ so the sample size corresponds with the one we chose for the MLE. In the simulation we compute (2.12) from the sample data and repeat this a 1000 times. This results in the following histogram in Figure 2.4:



This is the histogram of the $\hat{\gamma}$'s resulted through 1000 simulations. The red line indicates the true value of γ.

Figure 2.4: Histogram of the parameter estimation with MOM for tail modeling. The 600 largest values as the sample data

The mean of the $\hat{\gamma}$'s has an approximate value of 0.45 and is close to the true value of 0.4. So we conclude with the help of the MOM that we can estimate the tail of a $GEV_{(0,1,0.4)}$-distribution with the GPD.

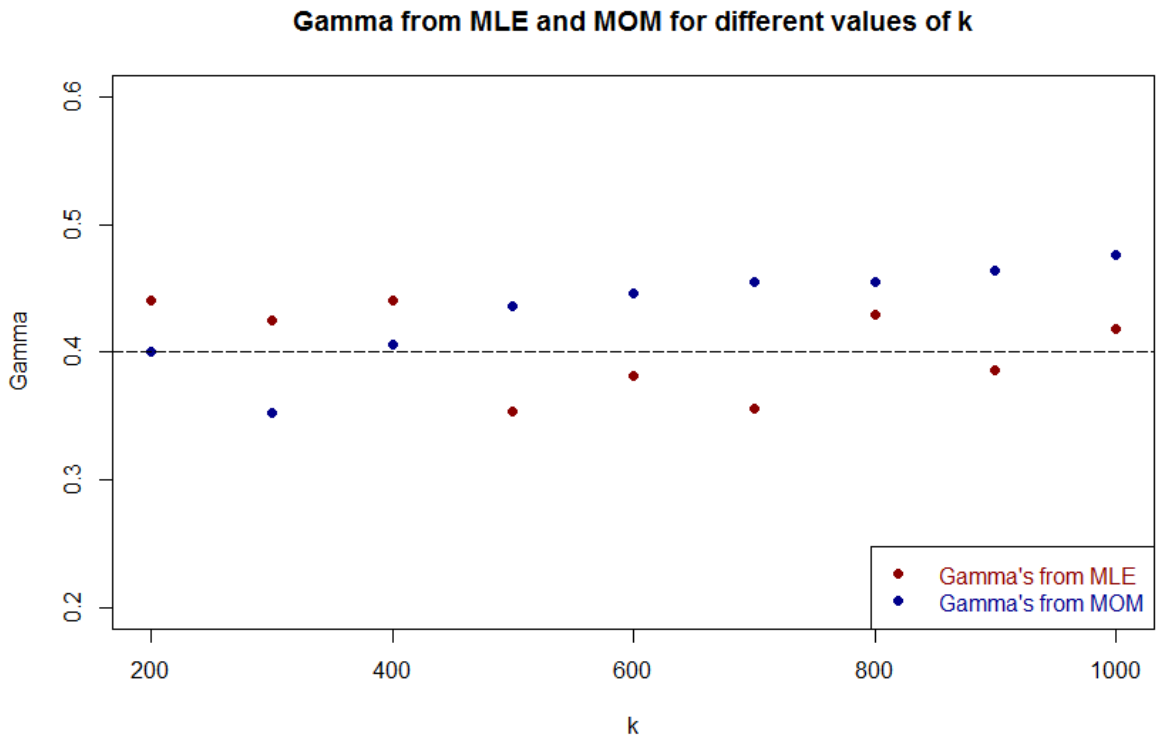


Figure 2.5: Graph of the mean of all the $\hat{\gamma}$'s for different values of k. Data is from the $GEV_{(0,\sigma,\gamma)}$ distribution.

We calculated the mean of the $\hat{\gamma}$'s above with only 5 $\hat{\gamma}$'s per k. The reason this number is so low is because the MLE is very sensitive and often fails to find a estimation for the parameters. This means there is some bias for the mean of the γ's in Figure 2.5.

2.3 High quantile estimation

Next we start with the estimation of a 'high' quantile. First we define a different way to describe a quantile and then we are going to make use of relation (1.1.20) in [L. de Haan(2006)]:

Definition 2.6 *Define the high quantile U as:*

$$U(t) = Q(1 - \frac{1}{t}) \tag{2.13}$$

With $Q(1 - \frac{1}{t})$ as the value x such that $F_X(x) = P(X \leq x) = 1 - \frac{1}{t}$.

Theorem 2.7 *For $\gamma \in \mathbb{R}, F \in D(G)$. There exists a positive function a such that for $x > 0$,*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \tag{2.14}$$

In this section we are going to estimate the quantile $U(\frac{1}{p})$ for p very small which has been introduced by definition (2.6). The reason why this is necessary is because we are going to use risk measures which

contain these quantiles. We are going to make use of the MLE-method and MOM again, first we are going to rewrite theorem (2.7) into:

$$U(tx) \approx U(t) + a(t)\frac{(\frac{x}{t})^{\gamma} - 1}{\gamma}, \quad x > t. \tag{2.15}$$

Let $t := \frac{n}{k}$ and $tx = \frac{1}{p}$ with $k := k(n)$ as a sequence of integers with $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$ when $n \to \infty$. Then we have for (2.15):

$$U(\frac{1}{p}) \approx U(\frac{n}{k}) + a(\frac{n}{k})\frac{(\frac{k}{np})^{\gamma} - 1}{\gamma} \tag{2.16}$$

2.3.1 Estimating the quantile with MLE

Remember that we used empirical excesses to get our estimators and that we can estimate γ with (2.10) and $a(\frac{n}{k})$ by (2.11). Filling the estimators in (2.16) gives us the following estimate:

$$\hat{U}^{\text{MLE}}\left(\frac{1}{p}\right) = X_{n-k,n} + \hat{\sigma}^{\text{MLE}}(X_{n-k,n})\frac{(\frac{k}{np})^{\hat{\gamma}^{\text{MLE}}} - 1}{\hat{\gamma}^{\text{MLE}}} \tag{2.17}$$

2.3.2 Estimating the quantile with MOM

For the MOM we still need an estimator for $a(\frac{n}{k})$. We are going to make use of (4.2.3) and theorem 4.2.1 from [L. de Haan(2006)] where it states that we can use the estimator:

$$\hat{\sigma}^{\text{MOM}} = X_{n-k,n} \cdot M_n^{(1)} \cdot (1 - \hat{\gamma}_-) \tag{2.18}$$

With $\hat{\gamma}_- = 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}$ and $M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1}(log(X_{n-i,n}) - log(X_{n-k,n})))^j$, $j = 1, 2$.
We estimate γ with (2.12) and $a(\frac{n}{k})$ with (2.18). Filling the estimators in (2.16) gives us the estimate:

$$\hat{U}^{\text{MOM}}\left(\frac{1}{p}\right) = X_{n-k,n} + \hat{\sigma}^{\text{MOM}}(X_{n-k,n})\frac{(\frac{k}{np})^{\hat{\gamma}^{\text{MOM}}} - 1}{\hat{\gamma}^{\text{MOM}}} \tag{2.19}$$

Chapter 3

Risk measurements

3.1 VaR and ES

In order to quantify the risk of heavy-tailed data we need some risk measurement tools. The two most popular ones are Value at Risk (VaR) and Expected shortfall (ES). The most commonly used in the financial world is VaR. But the VaR has been criticized that it fails to accurately measure risk. We will come to that later in this section after defining and explaining both.

In this project we won't work with loss data from investments but we look at Tsunami data instead. This shows that a risk measurement is versatile and can be used for a lot of situations. Let's start with introducing a certain distribution called the **loss distribution**, which we will denote by X. If we are in the right tail of X then we concede a big loss, and the more we go to the right the bigger the loss. And when we are in the left tail we have a win or a negative loss. The Tsunami data corresponds to a one-sided loss distribution where we denote X as the maximum water height above sea level, which is always greater or equal to zero (see chapter 4). With the loss distribution introduced we present the formal definition of VaR:

For a confidence level $p \in (0,1)$ ($p = 0.05$ is standard),

$$\mathrm{VaR}_p(\mathrm{X}) = \inf\{l \in \mathbb{R} : \mathrm{P}(\mathrm{X} > l) \leq p\} \tag{3.1}$$

The definition states that there is a probability of p that we lose more than VaR units of money. Now let's take a look at the expected shortfall, the formal definition is:

For a confidence level $p \in (0,1)$ ($p = 0.05$ is standard),

$$\begin{aligned} \mathrm{ES}_p(X) &= \mathrm{E}[X|X > \mathrm{VaR}_p] \\ &= \mathrm{E}[X|X > Q(1-p)] \end{aligned} \tag{3.2}$$

The expected shortfall (ES) takes the average value of all the values Exceeding the VaR and is also referred as the Conditional VaR.

If we make use of ES when we work with heavy-tailed data it's possible we result in a undefined solution. This is the case if the mean does not exist for the data. The reason why the solution can be undefined is because the tail decays so slowly that ES (which is the integral over the tail) is also undefined. This problem does not occur when using VaR. Note also that calculating VaR is way easier than ES. ES is a integral and can be problematic to solve. It seems like making use of VaR is more profitable at first glance, but the VaR has a couple of disadvantages compared to ES.

The first one is that the VaR is not coherent in a state where normality is not assumed, VaR is not sub-additive. ES is coherent under any assumption. Non sub-additivity means that in a portfolio a

combination of stocks the VaR not necessarily reduce risk compared to having separate portfolios with a stock each. This means that if we are in a non-normal world (which is more assumable than normality in financial markets) and we use the VaR to quantify risk from the combined portfolio we can have a bigger risk than from the separate portfolios, which is of course wrong. For the second disadvantage we look at Figure 3.1:

Different loss distributions

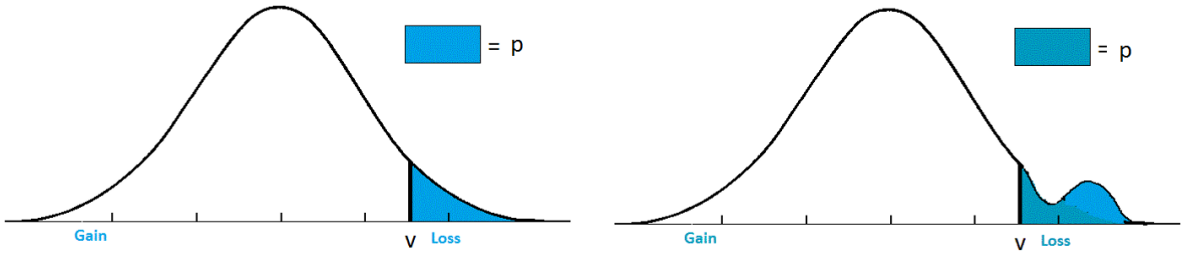


Figure 3.1: Loss distributions with different tail shapes

We can see that VaR fails to correctly quantify the risk in the loss distributions above. In the loss distribution on the right the tail is more 'extreme' but the VaR gives the same value for both the loss distributions. This is because the VaR only looks at the quantile which could lead to underestimation of risk. ES takes the shape of the loss distribution in consideration,

We are going to make use of ES or a modification like the expected log shortfall (introduced in (3.13)).

3.2 Finding a risk measure

In this section we are going to find risk measures for different assumptions of distribution and values that the shape parameter γ can take. We split our problem in two parts:

- Situation 1: $0 < \gamma < 1$
- Situation 2: $\gamma \geq 1$

Situation 1 is the case where we work with heavy-tailed data (γ smaller than 1). In situation 2 we work with super heavy-tailed data ($\gamma \geq 1$) with no upper-bound for γ.

3.2.1 Situation 1: $0 < \gamma < 1$

In this section we find a risk measure for heavy-tailed data with $0 < \gamma < 1$. First we are going to find a risk measure when we work with the random variable $Y \sim GPD_{(0,\sigma,\gamma)}$. Then we do the same but now with the random variable $X \in D(G)$.

3.2.1.1 $Y \sim GPD_{(0,\sigma,\gamma)}$

We need to express the quantile function $Q(1-p)$ in terms of our parameters. Let Y be $GPD_{(0,\sigma,\gamma)}$ distributed with realization x. The distribution function of Y is known and we get the following equation:

$$
\begin{aligned}
1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} &= 1 - p \\
&\vdots \\
x &= \frac{\sigma(p^{-\gamma} - 1)}{\gamma} \quad =: Q(1-p)
\end{aligned} \tag{3.3}
$$

Now we can compute the expected shortfall:

$$
\begin{aligned}
\text{ES} &= E[Y|Y > Q(1-p)] \\
&= \frac{E[Y \cdot \mathbb{1}_{\{Y>Q(1-p)\}}]}{p} \\
&= \frac{1}{p}\int_{Q(1-p)}^{\infty} \frac{x}{\sigma}(1 + \frac{\gamma x}{\sigma})^{-(\frac{1}{\gamma}+1)}dx
\end{aligned} \tag{3.4}
$$

We are going to make use of the substitution method with $u = 1 + \frac{\gamma x}{\sigma}$. This results in: $\begin{cases} \frac{\text{du}}{\text{dx}} = \frac{\gamma}{\sigma} \\ x = \frac{(u-1)\sigma}{\gamma} \end{cases}$

Substitution in (3.4) gives:

$$
\text{ES} = \frac{1}{p}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty} \frac{(u-1)\sigma}{\gamma^2} u^{-(\frac{1}{\gamma}+1)}du \tag{3.5}
$$

Note that (3.5) is undefined for $\gamma \geq 1$. In our case where $0 < \gamma < 1$ we can solve the integral:

$$
\text{ES} = \frac{1}{p}\left[\frac{\sigma}{\gamma(1-\gamma)}(1 + \frac{\gamma Q(1-p)}{\sigma})^{-\frac{1}{\gamma}+1} - \frac{\sigma}{\gamma}(1 + \frac{\gamma Q(1-p)}{\sigma})^{-\frac{1}{\gamma}}\right] \tag{3.6}
$$

Substituting (3.3) in (3.6):

$$
\text{ES} = \frac{\sigma}{\gamma}\left[\frac{1}{1-\gamma} \cdot p^{-\gamma} - 1\right] \tag{3.7}
$$

We can rewrite (3.7) back in terms of $Q(1-p)$, which results in:

$$
\text{ES} = \frac{Q(1-p)}{1-\gamma} + \frac{\sigma}{1-\gamma}, \tag{3.8}
$$

for $0 < \gamma < 1$ and $Y \sim GPD_{(0,\sigma,\gamma)}$.

3.2.1.2 $X \in D(G)$

For simplicity we replace $Q(1-p)$ with t in (3.2) (we can replace $Q(1-p)$ with t at all times):

$$
\text{ES} = \text{E}[X|X > t] \tag{3.9}
$$

with t a threshold. Note that we have an approximation of $f_X(X|X > t)$ from (2.5) which we will use for (3.9). Let's start with the definition of E[X|X>t]:

$$
\begin{aligned}
\text{E}[X|X > t] &= \int_t^{\infty} x f_{X|X>t}(x)dx \\
&\approx \int_t^{\infty} x \frac{1}{\sigma(t)}\left(1 + \frac{\gamma}{\sigma(t)}(x-t)\right)^{-(\frac{1}{\gamma}+1)} dx
\end{aligned} \tag{3.10}
$$

We make use of the substitution rule with $u = 1+\frac{\gamma}{\sigma(t)}(x-t)$. This results in: $\begin{cases} \frac{du}{dx} = \frac{\gamma}{\sigma(t)} \iff dx = du \cdot \frac{\sigma(t)}{\gamma} \\ x = \frac{(u-1)\sigma(t)}{\gamma} + t \end{cases}$

Substituting into (3.10) leads us to the following:

$$\int_1^\infty \frac{1}{\gamma} \cdot \left(\frac{(u-1)\sigma(t)}{\gamma} + t \right) u^{-(\frac{1}{\gamma}+1)} du \tag{3.11}$$

Note that the integral above is undefined if $\gamma \geq 1$. In our case where $0 < \gamma < 1$ we can solve the integral:

$$\begin{aligned} \mathrm{E}[X|X > t] &\approx \frac{\sigma(t)}{1-\gamma} + t \\ &= \frac{t}{1-\gamma} \end{aligned}$$

The last equality holds because of Theorem 2.4 where it's stated that if $\gamma > 0$ holds we can substitute $\sigma(t)$ with γt. Substituting $Q(1-p)$ back gives us:

$$\mathrm{ES} \approx \frac{Q(1-p)}{1-\gamma}, \tag{3.12}$$

for $0 < \gamma < 1$ and $X \in D(G)$.

3.2.2 Situation 2: $\gamma \geq 1$

In this section we are going to find a risk measure when the data is super heavy-tailed ($\gamma \geq 1$), making use of a transformation of the data. Again we start first with finding a risk measure when we work with the random variable $Y \sim GPD_{(0,\sigma,\gamma)}$. Then we do the same but now with the random variable $X \in D(G)$.

3.2.2.1 $Y \sim GPD_{(0,\sigma,\gamma)}$

We make use of the Expected log Shortfall (ELS) which is a log-transformation of the data. The definition of ELS is:

Definition 3.1 *The Expected log Shortfall is defined as:*

$$\mathrm{ELS} = \mathrm{E}[log(Y)|Y > Q(1-p)], \tag{3.13}$$

with Y a random variable and x a realisation of Y.

Let's parametrize ELS:

$$\mathrm{ELS} = \frac{1}{p} \int_{Q(1-p)}^\infty \frac{log(x)}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-(\frac{1}{\gamma}+1)} dx \tag{3.14}$$

We make use of the substitution rule with $u = 1 + \frac{\gamma x}{\sigma}$, this results in: $\begin{cases} \frac{du}{dx} = \frac{\gamma}{\sigma} \iff dx = du \cdot \frac{\sigma}{\gamma} \\ x = \frac{(u-1)\sigma}{\gamma} \end{cases}$

Substituting in (3.14) gives

$$\begin{aligned} \mathrm{ELS} &= \frac{1}{p\gamma} \int_{1+\frac{\gamma Q(1-p)}{\sigma}}^\infty log(\frac{(u-1)\sigma}{\gamma}) u^{-(\frac{1}{\gamma}+1)} du \\ &= \frac{1}{p\gamma} \left\{ \int_{1+\frac{\gamma Q(1-p)}{\sigma}}^\infty log((u-1)\sigma) u^{-(\frac{1}{\gamma}+1)} du - \int_{1+\frac{\gamma Q(1-p)}{\sigma}}^\infty log(\gamma) u^{-(\frac{1}{\gamma}+1)} du \right\} \end{aligned} \tag{3.15}$$

The second integral in the brackets in (3.15) above can be easily calculated so we focus on the first integral form now. let's use partial integration on the first integral of the brackets.

$$\begin{aligned}\frac{1}{p\gamma}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty} log((u-1)\sigma)u^{-(\frac{1}{\gamma}+1)}du &= \Big[log((u-1)\sigma)\cdot u^{-\frac{1}{\gamma}}\cdot -\gamma\Big]_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty} + \frac{\gamma}{\sigma}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty}\frac{u^{-\frac{1}{\gamma}}}{u-1}du \\ &= log(\gamma Q(1-p))\cdot\left(1+\frac{\gamma Q(1-p)}{\sigma}\right)^{-\frac{1}{\gamma}}\cdot\gamma + \frac{\gamma}{\sigma}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty}\frac{u^{-\frac{1}{\gamma}}}{u-1}du \end{aligned} \tag{3.16}$$

The integral in (3.16) is hard to solve, thus we are going to find a close upper bound. Note that in our project we have a high value for $Q(1-p)$. This means that in the integral we can safely say that $u > 2 \iff \frac{u}{2} > 1$ which results in the inequality $\frac{1}{u-1} \leq \frac{1}{u-\frac{u}{2}} = \frac{2}{u}$ for $u > 2$. This gives us the following inequality:

$$\begin{aligned}\frac{1}{p\gamma}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty} log((u-1)\sigma)u^{-(\frac{1}{\gamma}+1)}du &\leq log(\gamma Q(1-p))\cdot\left(1+\frac{\gamma Q(1-p)}{\sigma}\right)^{-\frac{1}{\gamma}}\cdot\gamma + \frac{\gamma}{\sigma}\int_{1+\frac{\gamma Q(1-p)}{\sigma}}^{\infty}\frac{2u^{-\frac{1}{\gamma}}}{u}du \\ &= log(\gamma Q(1-p))\cdot\left(1+\frac{\gamma Q(1-p)}{\sigma}\right)^{-\frac{1}{\gamma}}\cdot\gamma + \frac{2\gamma^2}{\sigma}\left(1+\frac{\gamma Q(1-p)}{\sigma}\right)^{-\frac{1}{\gamma}}\end{aligned}$$

So now we finally have an upper bound for the first integral in (3.15) which we will use to approximate ELS. Substituting (3.3) for $Q(1-p)$ and making simple equation manipulations gives us the following upper bound for ELS:

$$\begin{aligned}\text{ELS} &\leq log(\sigma(p^{-\gamma}-1)) + \frac{2\gamma}{\sigma} - log(\gamma) \\ &= log(Q(1-p)) + \frac{2\gamma}{\sigma}\end{aligned} \tag{3.17}$$

Note that the inequality comes from giving the upperbound $\frac{2\gamma}{\sigma}$ to the second integral in (3.16). This upperbound is very small compared to the first term in (3.17) so we can safely make the approximation:

$$\text{ELS} \approx log(Q(1-p)), \tag{3.18}$$

for $\gamma \geq 1$ and $Y \sim GPD_{(0,\sigma,\gamma)}$.

3.2.2.2 $X \in D(G)$

In this case we also make use of ELS to find a risk measure. We have (2.5) as the approximation of the probability density function. For simplicity we replace $Q(1-p)$ with t (we can replace $Q(1-p)$ with t at all times):

$$\begin{aligned}\text{ELS} &= \text{E}[log(X)|X > Q(1-p)] \\ &\approx \int_t^{\infty}\frac{log(X)}{\sigma(t)}\left(1+\frac{\gamma}{\sigma(t)}(x-t)\right)^{-(\frac{1}{\gamma}+1)}dx\end{aligned} \tag{3.19}$$

We make use of the substitution rule with $u = 1+\frac{\gamma}{\sigma(t)}(x-t)$. This results in: $\begin{cases} \frac{du}{dx} = \frac{\gamma}{\sigma(t)} \iff dx = du \cdot \frac{\sigma(t)}{\gamma}. \\ x = \frac{(u-1)\sigma(t)}{\gamma} + t \end{cases}$.

Substituting in (3.19) gives:

$$\begin{aligned} \text{ELS} &\approx \int_1^\infty log\left(\frac{(u-1)\sigma(t)}{\gamma} + t\right) \cdot \frac{1}{\gamma} u^{-(\frac{1}{\gamma}+1)} du \\ &\vdots \\ &= log(t) + \frac{1}{\gamma}\int_1^\infty log\left(\frac{(u-1)\sigma(t)}{\gamma t} + 1\right) \cdot u^{-(\frac{1}{\gamma}+1)} du \end{aligned}$$

We are going to show that the second term is approximately zero, let's start with partial integration:

$$\begin{aligned} \frac{1}{\gamma}\int_1^\infty log\left(\frac{(u-1)\sigma(t)}{\gamma t} + 1\right) \cdot u^{-(\frac{1}{\gamma}+1)} du = &\left[-log(1 + \frac{\sigma(t)}{\gamma t}(u-1)) \cdot u^{-\frac{1}{\gamma}}\right]_1^\infty \\ &+ \int_1^\infty \frac{\frac{\sigma(t)}{\gamma t}}{1 + \frac{\sigma(t)}{\gamma t}(u-1)} \cdot u^{-\frac{1}{\gamma}} du \end{aligned} \tag{3.20}$$

Note that the first term is equal to zero. This means that we only need to compute the integral on the right side of (3.20). We make use of Theorem 2.4 which states that when $\gamma > 0$ holds we can make the substitution $\sigma(t) = \gamma t$. By doing this we have $\frac{\frac{\sigma(t)}{\gamma t}}{1+\frac{\sigma(t)}{\gamma t}(u-1)} = \frac{1}{u}$. So (3.20) becomes:

$$\frac{1}{\gamma}\int_1^\infty log\left(\frac{(u-1)\sigma(t)}{\gamma t} + 1\right) \cdot u^{-(\frac{1}{\gamma}+1)} du = \int_1^\infty u^{-(\frac{1}{\gamma}+1)} du = \gamma$$

We then have for ELS:

$$\text{ELS} \approx log(Q(1-p)) + \gamma \approx log(Q(1-p)) \tag{3.21}$$

The last approximation can be made because $Q(1-p)$ is a large number, which results in $log(Q(1-p))$ also being a large number. let's state our findings again:

$$\text{ELS} \approx log(Q(1-p)), \tag{3.22}$$

for $\gamma \geq 1$ and $X \in D(G)$.

We finally have risk measures for both the situations and the different assumptions of the random variable we work with. In the next chapter we will work with real data that is of course heavy-tailed. We are going to find a risk measure and have a real-life application of our project. We end this chapter with a summary table of all the situations and different assumptions of distributions we work with.

Risk measures		
Situation	Distribution assumption	Risk measure
$0 < \gamma < 1$	$Y \sim GPD_{(0,\sigma,\gamma)}$	ES $= \frac{Q(1-p)}{1-\gamma} + \frac{\sigma}{1-\gamma}$
	$X \in D(G)$	ES $\approx \frac{Q(1-p)}{1-\gamma}$
$\gamma \geq 1$	$Y \sim GPD_{(0,\sigma,\gamma)}$	ES $\approx log(Q(1-p))$
	$X \in D(G)$	ELS $\approx log(Q(1-p))$

Table 3.1 Risk measures with different assumptions.

This table will help us find a risk measure based on type of distribution and the value of γ. Note that the assumption $Y \sim GPD_{(0,\sigma,\gamma)}$ is the special case of the assumption $X \in D(G)$.

Chapter 4

Application with Tsunami data of Japan

At last we are going to quantify the risk of data (maximum water height above sea-level) from Tsunami events in Japan in the time-period 1403-2011. The dataset we are going to use is acquired from the 'National Centers for environmental information' [NOAA(1403-2011)]. This dataset has a size of 207 containing the date, location and water height above sea level in meters. The size of the data is small compared with the sizes we chose when performing the simulations in this project but it's still sufficient for application. Let's take a look at the data first, see if we can spot any extreme events with inspection.

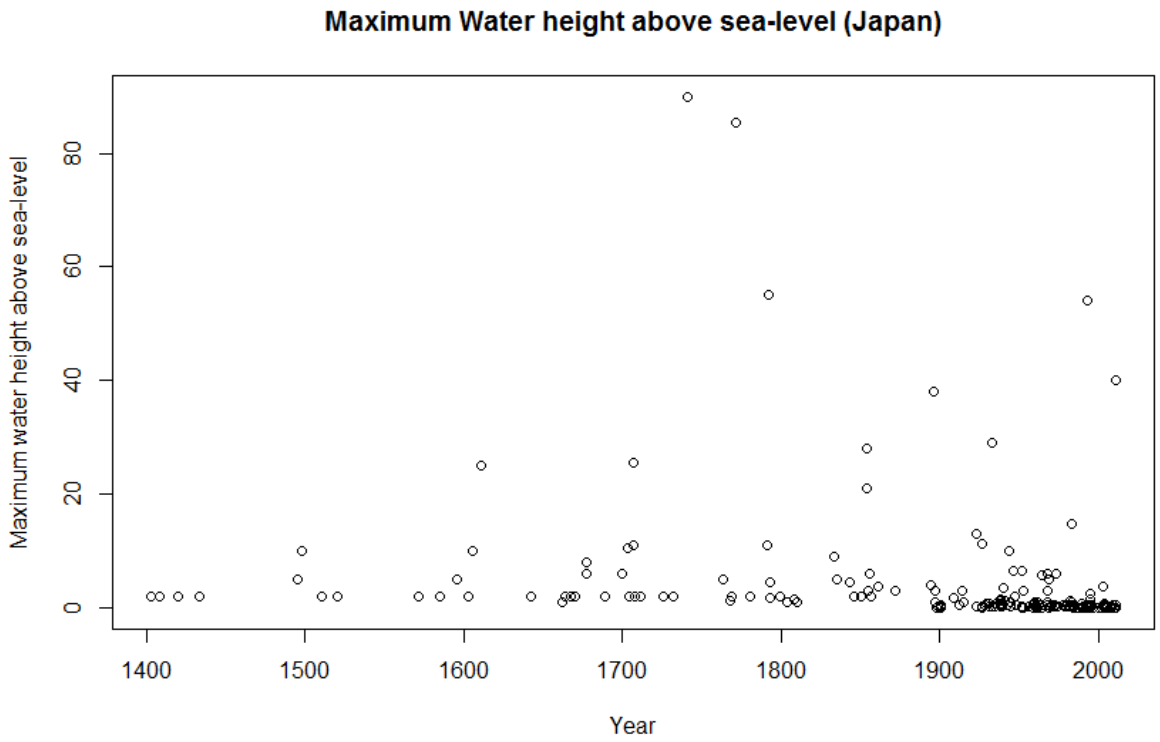


Figure 4.1: Graph of all the maximum water heights above sea-level from the time period 1403-2011

There are 2 values in the time period 1700-1800 that can be considered 'extreme' which are 90 meters in

1741 (W. Hokkaido Island,Japan) and 85.4 meters in 1771 (Ryukyu Islands,Japan). The amount of deaths are 2000 and 13486 respectively. One of the most recent Tsunami events took place at the Honshu Islands in 2011, and had a maximum water height of 38.90 meters above sea-level. The number of confirmed deaths is 15891 according to Japan's National Police Agency. Most people died by drowning and more than 2500 people are still reported missing. The average maximum water height is 4.1 meters, so the 2 events with the highest water height described above can definitely be labeled as 'extreme events'. Our first step is to estimate the parameters γ and $\sigma(x_{n-k,n})$ with $x_{n-k,n}$ the realisation of $X_{n-k,n}$. To choose the value of k we look at the graph in Figure 4.2:

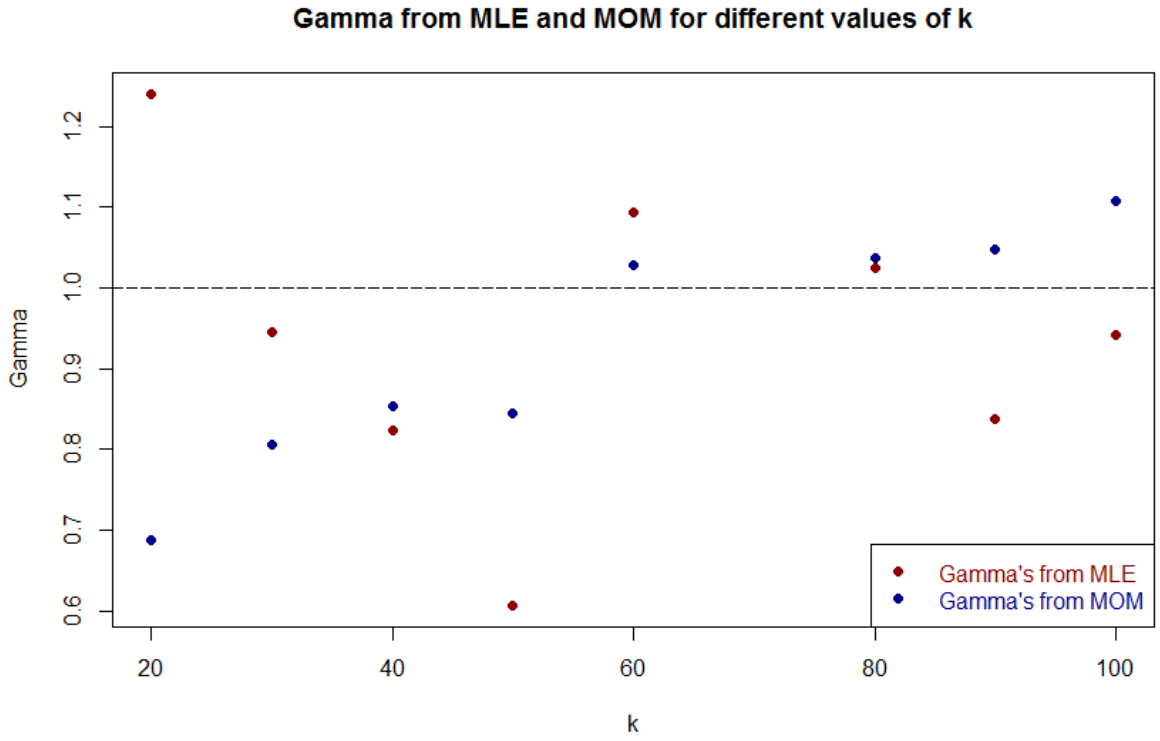


Figure 4.2: Graph of all the $\hat{\gamma}$'s for different values of k. Data is from [NOAA(1403-2011)].

For $k = 80$ both values of $\hat{\gamma}$ are close to each other and above 1, this is the best choice for k. Making use of both the MLE and MOM we estimate the parameters which gives the results shown below:

$$\hat{\gamma}^{\text{MLE}} = 1.03$$
$$\hat{\sigma}^{\text{MLE}}(x_{n-k,n}) = 1.54$$

$$\hat{\gamma}^{\text{MOM}} = 1.04$$
$$\hat{\sigma}^{\text{MOM}}(x_{n-k,n}) = 1.56$$

For the estimation of $\sigma(x_{n-k,n})$ we make use of Theorem 2.4 where it's stated that $\sigma(x_{n-k,n}) = \hat{\gamma} \cdot x_{n-k,n}$. With the chosen threshold k we result that both the estimates indicate that γ is bigger than 1, thus we make use of (3.22) as the risk measure. Making use of (2.17) and (2.19) which are the estimators of $Q(1-p)$ gives us the following (We set $p = \frac{1}{n} = \frac{1}{207}$):

$$\hat{Q}^{\text{MLE}}(0.995) \approx 136.93 \tag{4.1}$$
$$\hat{Q}^{\text{MOM}}(0.995) \approx 146.94$$

Substituting (4.1) in (3.22) gives us the following risk measurement:

$$\begin{aligned} \text{ELS}^{\text{MLE}} &\approx 4.92 \\ \text{ELS}^{\text{MOM}} &\approx 4.99 \end{aligned} \tag{4.2}$$

We bootstrapped the data (with a bootstrap sample size of 207) and estimated ELS^{MOM} 500 times which gave us the confidence interval of $[4.93, 5.03]$ for the ELS^{MOM}. Getting a confidence interval for ELS^{MLE} is very troublesome because it really depends on the data if we can find a solution for the parameters. We can bootstrap with very small samples but the confidence interval for ELS^{MLE} will have a huge bias. (4.2) is our quantification of risk of our data, focused on extreme tsunami events. The value indicates the estimate of the average log maximum water height given that the water height exceeds the estimated threshold $\hat{Q}(0.995)$. In the future we can also quantify the risk of extreme tsunami events of other countries based on their data (maximum water height above sea-level). This means that we can actually compare which type of event is more disastrous based on the computations we made above. We finally quantified risk and our project ends here with this result.

Chapter 5

Conclusion

Our research started with the Peak over threshold in chapter 2 where we obtained a way to study the tails. We have shown that if $X \in D(G)$ holds for a random variable X, the tail of X can be approximated with the GPD distribution. We stumbled upon problems with the parameter estimations obtained from the MLE, the MOM did not have problems but is less accurate in general. After introducing and specifying which risk measurements we can use for different kinds of assumptions we measured the data from Tsunami's (maximum water height above sea-level) occurred in Japan in the years 1403-2011. The project has shown that we can measure the risk of heavy-tailed data and we have completed our task.

Bibliography

[L. de Haan(2006)] A. Ferreira L. de Haan. *Extreme value theory An introduction.* Springer, New York, 2006.

[NOAA(1403-2011)] NOAA. Dataset water heights japan. *doi:10.7289/V5PN93H7*, 1403-2011.

[Zhou(2009)] Chen Zhou. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4):794–815, 2009.