A comparative study of VaR and ES using extreme value theory

Klara Andersson

Bachelor's thesis

Department of Economics
Lund University
Sweden

May 2020

Supervisor: Birger Nilsson

Abstract

Using data from OMXS30, we study which of the models block maxima and peaks-over-threshold, based on extreme value theory, are the most accurate when estimating the risk measures Value-at-Risk and Expected Shortfall. To perform this analysis, the risk measures are backtested. The extreme observations are fitted to the generalized extreme value distribution and the generalized Pareto distribution using maximum likelihood estimation. This study conclude that when estimating Value at Risk, block maxima is a more accurate model. When estimating Expected Shortfall, the difference between the models is not statistical significant.

Acknowledgement

I would like to thank my supervisor Birger Nilsson for excellent guidance throughout my thesis work.

Contents

Abbreviations

BM	block maxima
ES	Expected Shortfall
GEV	generalized extreme value
GPD	generalized Pareto distribution
MLE	maximum likelihood estimation
PBdH	Pickands-Balkema-de Haan
POT	peaks-over-threshold
VaR	Value-at-Risk

1 Introduction

The prices of stocks change over time and sometimes they change substantially over a short time period. The most recent example is the stock price drop in the spring 2020, caused by the coronavirus, e.g., the OMXS index dropped by 11.1% in one day (SVT, 2020). Other examples when the price dropped heavily, are during the "Black Monday" on October 19^{th} 1987, where Dow Jones dropped by over 20% (Sundkvist, 2014) and during the financial crisis 2008 (Ohlin, 2018). A natural question is then, how much can the price possibly fall? How bad can it be? Have we already seen the worst case scenario or can it be even worse in the future? Thus, we want to be able to predict how much the prices can drop in the future.

To predict how much the price can fall, we use different risk measures. Two risk measures are Value at Risk (VaR) and Expected Shortfall (ES). VaR is equal to the smallest loss such that the probability of obtaining a greater loss, is less than or equal to some predetermined probability α. Further, ES can be summarized as the average of the losses that are greater than VaR. Hence, when calculating VaR, a lower limit of "the worst losses" is obtained, while when calculating ES the average of these "worst losses" is produced (see Embrechts et al. 2005 and Hull 2018).

How do we calculate VaR and ES? A number of models exist for this purpose. Here, we will focus on two different models based on extreme value theory. Extreme value theory is used to analyze events that happen rarely, i.e., extreme events. In our setting, rare events consist of large drops in the stock prices. Such events do not occur frequently, but looking at historical data we can assume that they will happen sooner or later (Dowd, 2005).

The two models based on extreme value theory are called block maxima (BM) and peaks-over-threshold (POT). Both models have the same objective; fit a distribution to the sample of extreme observations. However, the models assume that the data follow different distributions. Also, which observations from the orignial sample that should be considered as extreme, differs in the two models (see Coles 2001 and Dowd 2005).

The purpose of this thesis is to compare these two models and examine which one is most accurate when estimating VaR and ES in the Swedish stock market. Below, some previous studies in this field are presented. Note that none of the studies below analyze the Swedish market. Hence, this study contributes to a more complete picture of the accuracy of VaR and ES estimates based on extreme value theory.

Thus, we are interested in possible differences of the VaR and ES estimates of the two models BM and POT. Cerovic and Karadzic (2015), who analyzed the Montenegrin stock market index for the period 2004-2014, concluded that POT passed the Kupiec test, while BM did not. Here, BM underestimated VaR too much. Similarly, Marinelli et al. (2007) establish that POT is significantly better than BM. Here the analysis was done on S&P500 for 1990-2004 and on 1998-2004 for some other series. However, in this study BM was over-conservative and produced fewer violations of VaR than expected. For all stocks in this study, BM was rejected at 95% level of confidence, while POT was not. On the contrary, neither the $\text{VaR}_{0.99}$ estimates of BM or POT is rejected in Bekiros and Georgoutsos (2005) for 95% level of confidence where the data consists of daily returns from 1997 to 2001 of the Dow Jones Industrial average and the Cyprus Stock Exchange indices. Note that they use Christoffersen's test and not Kupiec. In Embrechts et al. (2005) an argument why POT is better than BM is formulated:

> The block maxima method ... has the major defect that it is very wasteful of data; to perform our analyses we retrain only the maximum losses in large blocks (Embrechts et al., 2005, p. 275).

As mention earlier, BM and POT rely on the assumption that the extreme observations follow a certain distribution; denote the distribution in BM by H and the distribution in POT by G. H and G are two different distributions, but they have the same purpose: model the distribution of the extreme losses. In particular, we can note that a parameter, denoted ξ, is contained in both distributions. ξ should therefore take similar values (and same sign) in the two distributions (Coles, 2001). In Gilli and Këllezi (2006), a positive ξ is obtained in both models. Dowd (2005) and Embrechts et al. (2005) express that the case $\xi < 0$ is often not of great interest since most of financial data are more heavily tailed. This is strengthened by the results in Gilli and Këllezi (2006). In Cotter (2006), BM is modeled for the European, Asian and American markets. Similarly to Gilli and Këllezi (2006), ξ is positive in Cotter (2006), with one exception in the Asian market. Hence, Cotter (2006) conclude that the data follow the Fréchet distribution.

Summarizing the results in this study, we conclude that BM is significantly more accurate than POT for estimating VaR. When estimating ES, BM seems to perform better than POT. However, in this case the difference is non-significant, for the 95% level of confidence. Further, the parameter ξ becomes positive in BM and negative in POT.

Further, this thesis starts with a section of the theory of VaR and ES. Thereafter, the extreme value theory is introduced. We proceed with the underlying theory of the estimation of the distributions and end this part with the theory of backtesting VaR and ES. Before the results are given, the data and methodology used are presented. We close the thesis with a discussion of the results.

2 Theory

2.1 Value at Risk and Expected Shortfall

Value at Risk (VaR) and Expected Shortfall (ES) are two widely used risk measures. Since ES is defined in terms of VaR, we begin with the definition of VaR. VaR is defined as

$$\mathrm{VaR}_\alpha = \inf\{l \in \mathbb{R} : \Pr(\mathrm{L} > l) \leq 1 - \alpha\}, \tag{1}$$

where $\alpha \in (0, 1)$ is the chosen confidence level and L is a stochastic loss variable. In words, this can be interpreted as we want to find the smallest number l such that the probability of obtaining a loss greater than l is less than or equal to $1-\alpha$. The time horizon we will consider here is one day. If the distribution of L is continuous (which we will assume), an equivalent definition to (1) of VaR is

$$\Pr(L > \mathrm{VaR}_\alpha) = 1 - \alpha, \tag{2}$$

i.e., we want to find the value of VaR such that the probability of obtaining a loss greater than VaR is equal to $1-\alpha$ (Embrechts et al., 2005). Further, the definition of ES is

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_x \, dx, \tag{3}$$

which can be seen as taking the average VaR over all confidence levels $x \in (\alpha, 1)$. If the loss distribution is continuous, then ES can also be defined by

$$\mathrm{ES}_\alpha = \mathbb{E}(L \mid L \geq \mathrm{VaR}_\alpha). \tag{4}$$

Then the definitions (3) and (4) are equivalent (Embrechts et al., 2005).

As mention earlier, both risk measures are widely used. However, one crucial argument of choosing ES over VaR is that ES measures the size of potential losses greater than VaR. When using VaR, we only detect the lower limit of "the worst losses" (Embrechts et al., 2005).

2.2 Extreme value theory

Extreme value theory handles, as the name suggests, the theory of extreme events. In finance, an extreme event can be when the stock price decreases rapidly. Of course, we want to be able to analyze such events, or in other words, we want to find the distribution of these extreme events. However, there is a difficulty when working with extreme events. Extreme events are, by nature, uncommon and therefore we have relatively few samples to rely on when performing the estimation. This implies that long original samples are needed. Another problem is to define what is considered as an extreme event. There are two methods for defining extreme events. In the first method, the original data are dived into blocks, e.g. each month or year is a block. Then the largest observation in each block is extracted and inference is performed on this new sample. This method is called block maxima. In the second method, all observations above a certain threshold are regarded as extreme. This method is called peaks-over-threshold (Dowd, 2005). Below, these two methods are described in more detail.

2.2.1 Block maxima

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent distributed random variables, where X_i, $i = 1, \ldots, n$ has some unknown distribution function F. Define $M_n = \max\{X_1, X_2, \ldots, X_n\}$, i.e., M_n is the block maxima. The objective is now to find the distribution of M_n. Since X has distribution function F, we could argue that

$$\begin{aligned} \Pr(M_n \leq x) &= \Pr(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \\ &= \Pr(X_1 \leq x)\Pr(X_2 \leq x) \ldots \Pr(X_n \leq x) \\ &= (F(x))^n, \end{aligned} \tag{5}$$

but since F is unknown, this is to no help (Coles, 2001). Fortunately, the Fisher-Tippett theorem comes here into play. This theorem states that M_n converges in distribution to the generalized extreme value (GEV) family of distributions as $n \to \infty$. The GEV family of distributions are given by

$$H_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left(- \left(1 + \xi \dfrac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}} \right) & \xi \neq 0 \\ \exp\left(- \exp\left(\dfrac{x - \mu}{\sigma} \right)\right) & \xi = 0, \end{cases} \tag{6}$$

provided that $1+\xi(x-\mu)/\sigma > 0$. The three parameters μ, σ and ξ are the location parameter, the scale parameter and the shape parameter. As the names suggest, μ indicates where on the axis the distribution of M_n is located and σ illustrate the width of the distribution. Finally, ξ is a measure of the shape of the distribution. If $\xi > 0$, then M_n follows the Fréchet distribution, which is a distribution with heavy tails. Next, when $\xi = 0$, M_n follows the Gumbel distribution, which has exponential tails. For $\xi < 0$, the tails of the distribution is lighter than in the normal distribution. When modeling financial data the cases $\xi > 0$ and $\xi = 0$

are of most interest, since this sort of data rarely have tails lighter than normal (Dowd, 2005). Also, when $\xi < 0$, the distribution has an upper end-point; $\mu - \sigma/\xi$ (Coles, 2001).

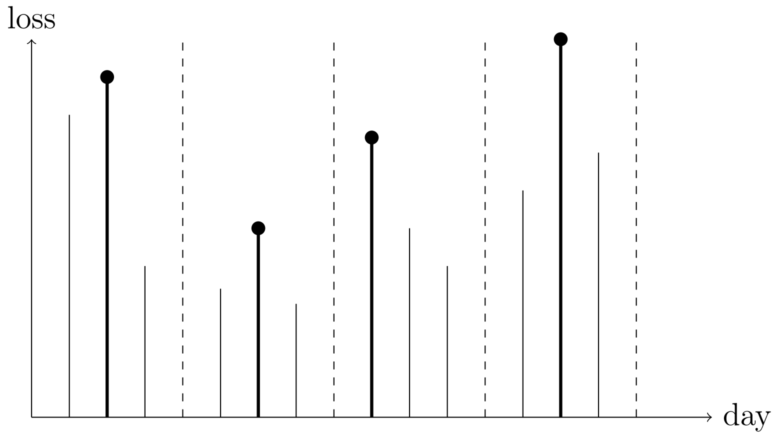

Figure 1: In the BM model, the largest observation in each block is considered as an extreme event, i.e., the block maxima. Inference is then done on the sample of block maxima.

Further, we can find the quantiles of the GEV distribution by setting $H(x) = p$, where p is some chosen probability, and solve for x. Thus, by inverting (6) we get

$$x_p = \begin{cases} \mu - \dfrac{\sigma}{\xi}(1 - (-\ln p)^{-\xi}) & \xi \neq 0 \\ \mu - \sigma \ln(-\ln p) & \xi = 0. \end{cases} \tag{7}$$

It should be emphasized that x_p is the quantile of the extreme value distribution $H(x)$ and not of the parent distribution $F(x)$ (Dowd, 2005).

Recall now the definition of VaR,

$$\Pr(L > \mathrm{VaR}_\alpha) = 1 - \alpha,$$

which is equivalent to

$$\Pr(L \leq \mathrm{VaR}_\alpha) = \alpha,$$

which also can be written as $F(\mathrm{VaR}_\alpha) = \alpha$. Hence, VaR_α can be seen as a quantile of F. Since F is unknown, it is not possible to find the quantile by simply inverting F. However, since $\Pr(M_n \leq x) = (F(x))^n$, and we now know that $\Pr(M_n \leq x)$ can be approximated by H for large n, we get

$$p = H(\mathrm{VaR}_\alpha) = \Pr(M_n \leq \mathrm{VaR}_\alpha) = (F(\mathrm{VaR}_\alpha))^n = \alpha^n,$$

where α is the confidence level for the corresponding VaR. Thus, by plugging in $p = \alpha^n$ in (7), we get

$$\mathrm{VaR}_\alpha = \begin{cases} \mu - \dfrac{\sigma}{\xi}(1 - (-n \ln \alpha)^{-\xi}) & \xi \neq 0 \\ \mu - \sigma \ln(-n \ln \alpha) & \xi = 0, \end{cases} \tag{8}$$

which is the desired quantile (Dowd, 2005).

The next step is to find a formula for ES. Recall the definition of ES in (3). However, instead of trying to solve this integral analytically, we approximate ES by numeric integration, i.e., we calculate the average VaR for a number of confidence levels from α to 1. So far, we have assumed that α could be any value in $(0, 1)$. However, we will use $\alpha = 0.99$ in this study. Because we are dealing with extreme value theory, we need to choose a high confidence level. Otherwise, the formulas for VaR and ES will be inaccurate (Dowd, 2005).

2.2.2 Peaks-over-threshold

Let $X_1, X_2, \ldots, X_N$ be a sequence of random variables that are independent and identically distributed with some unknown distribution function F. Then, the POT method can be summarized as that we want to find the distribution of all excess observations that are greater than some chosen threshold u. Denote this distribution by $F_u(x)$. Thus, we have

$$F_u(x) = \Pr(X - u \leq x) \mid X > u) = \frac{F(u + x) - F(u)}{1 - F(u)}. \tag{9}$$

However, F_u is unknown since F is unknown (Dowd, 2005). Fortunately, a theorem regarding a limit distribution solves the situation.

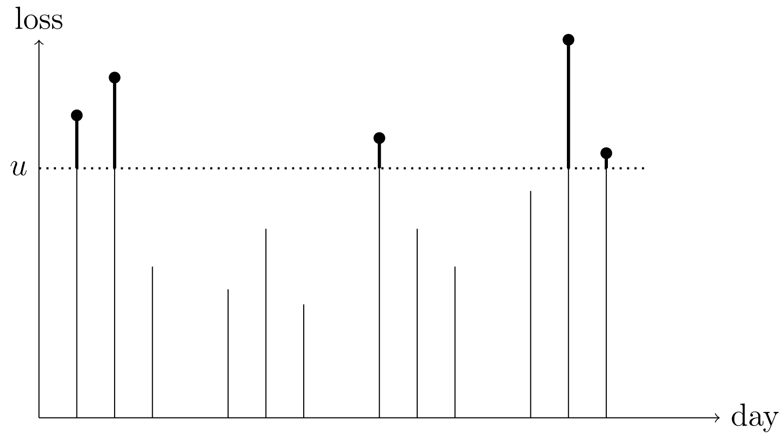

Figure 2: In the POT model, each observation above the threshold u is considered as an extreme event. The excess losses are then used for inference.

Theorem 2.1 (Pickands-Balkema-de Haan, (Embrechts et al., 2005))**.** *We can find a (positive measurable) function β_u, thats depend on the threshold u, such that*

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} \left| F_u(x) - G_{\xi,\beta_u}(x) \right| = 0,$$

if and only if F is in the domain of attraction of H_ξ, $\xi \in \mathbb{R}$. Here, $x_F \leq \infty$ is the right endpoint of F and G is the distribution function of GPD,

$$G_{\xi,\beta_u}(x) = \begin{cases} 1 - (1 + \xi x/\beta_u)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-x/\beta_u) & \xi = 0. \end{cases} \tag{10}$$

Remark. *A function F belongs to maximum domain of attraction of H if*

$$\lim_{n \to \infty} Pr((M_n - d_n)/c_n) = \lim_{n \to \infty} F^n(c_n x + d_n) = H(x),$$

for some non-degenerated function H and sequences of real constants (c_n) and (d_n), $c_n > 0$ for all n (Embrechts et al., 2005).

Thus, we can approximate $F_u(x)$ by G_{ξ,β_u} and (9) can now be written as

$$G_{\xi,\beta_u}(x) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

or equivalently

$$F(x+u) = G_{\xi,\beta_u}(x)(1-F(u)) + F(u). \tag{11}$$

In (11), the distribution function of $F(x+u)$ is stated. By a change of variable from $x+u$ to x, we obtain the distribution of the losses greater than u, namely

$$F(x) = G_{\xi,\beta_u}(x-u)(1-F(u)) + F(u).$$

Before it is possible to obtain an explicit formula for $F(x)$, we have to calculate $F(u) = \Pr(X \leq u)$. Denote the total number of observations by N and the number of observations that are greater than u by N_u. Then $F(u)$ can be approximated by

$$F(u) = \Pr(X \leq u) = \frac{N - N_u}{N} = 1 - \frac{N_u}{N}. \tag{12}$$

Now, by (10) and (12), we have

$$F(x) = \left(1 - \left(1 - \frac{N_u}{N}\right)\right)\left(1 - \left(1 + \xi\frac{x-u}{\beta_u}\right)^{-1/\xi}\right) + \left(1 - \frac{N_u}{N}\right),$$

which can be simplified to

$$F(x) = 1 - \frac{N_u}{N}\left(1 + \xi\frac{x-u}{\beta_u}\right)^{-1/\xi} \tag{13}$$

when $\xi \neq 0$. When $\xi = 0$, the same method is applied. However, we us

$$G_{\beta_u}(x) = 1 - \exp\left(-\frac{x}{\beta_u}\right)$$

instead of

$$G_{\xi,\beta_u}(x) = 1 - \left(1 + \xi\frac{x}{\beta_u}\right)^{-1/\xi}.$$

Thus, we get

$$F(x) = 1 - \frac{N_u}{N}\exp\left(-\frac{x-u}{\beta_u}\right), \tag{14}$$

when $\xi = 0$ (Dowd, 2005).

To obtain VaR, we want to find the quantile of (13), respectively for (14), for the confidence level α. By setting $F(\mathrm{VaR}_\alpha) = \alpha$ and solving for VaR_α in (13) and (14), we get

$$\mathrm{VaR}_\alpha = \begin{cases} u + \frac{\beta_u}{\xi}\left(\left(\frac{N}{N_u}(1-\alpha)\right)^{-\xi} - 1\right) & \xi \neq 0 \\ u + \beta_u \ln\left(\frac{N}{N_u}(1-\alpha)\right) & \xi = 0. \end{cases} \tag{15}$$

Finally, we derive the formula for ES. By (3) and (15) we have

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 u + \frac{\beta_u}{\xi}\left(\left(\frac{N}{N_u}(1-x)\right)^{-\xi} - 1\right) dx = \frac{\mathrm{VaR}_\alpha}{1-\xi} + \frac{\beta_u - \xi u}{1-\xi}, \tag{16}$$

for $\xi \neq 0$. For $\xi = 0$, we have

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 u + \beta_u \ln\left(\frac{N}{N_u}(1-x)\right) dx = \mathrm{VaR}_\alpha + \beta_u, \tag{17}$$

(Dowd, 2005).

2.2.3 Threshold selection

So far, we have taken the threshold u as given. However, the decision of u is not straightforward. According to the PBdH theorem, the choice of u should be as high as possible. On the other hand, a higher threshold will produce fewer extreme observations, resulting in high variances for the estimates of the parameters of the distribution. Hence, a threshold as low as possible is desirable, but that still fulfills the limit assumption of the theorem reasonably well. A simulation study in McNeil and Frey (2000) concludes that when 10% of the observations are regarded as extreme, is the optimal level. Another method for choosing the threshold is given in Coles (2001), which is explained below.

If $X \sim$ GPD, the expected value of X is given by

$$\mathbb{E}(X) = \frac{\beta_u}{1-\xi}.$$

Then the conditional expectation is given by

$$\mathbb{E}(X - u_0 \mid X > u_0) = \frac{\beta_{u_0}}{1-\xi},$$

for some threshold u_0. If u_0 is sufficiently large to satisfy the limit assumption of the PBdH theorem, then the same holds for any $u > u_0$. Further, it can then be shown that

$$\mathbb{E}(X - u \mid X > u) = \frac{\beta_u}{1-\xi} = \frac{\beta_{u_0} + \xi u}{1-\xi}. \tag{18}$$

for $u > u_0$. Thus, we see in (18) that $\mathbb{E}(X - u \mid X > u)$ is a linear function in u for $u > u_0$. The empirical counterpart of the conditional expectation is the sample mean of the threshold excesses, i.e.,

$$\frac{1}{N_u}\sum_{i=1}^{N_u}(x_i - u).$$

Then, it follows that the so called mean residual life plot

$$\left\{\left(u, \frac{1}{N_u}\sum_{i=1}^{N_u}(x_{(i)} - u)\right) : u < x_{\max}\right\},$$

where $x_{(i)}$, $i = 1, \ldots, N_u$ is the ordered sample, should be approximately linear for $u > u_0$. Hence, as threshold, we should choose the smallest u such that the mean residual life plot is linear (Coles, 2001).

2.3 Maximum likelihood estimation

To estimate the parameters (μ, σ, ξ) in GEV and (β_u, ξ) in GPD, the likelihood function is used (Coles, 2001). This function is defined by

$$L(\bar{\theta} \mid \bar{x}) = \prod_{i=1}^{n} f_{\bar{\theta}}(x_i), \tag{19}$$

where $\bar{\theta}$ is the vector of parameters to be estimated, $\bar{x} = (x_1, x_2, \ldots, x_n)$ are the data points and $f_{\bar{\theta}}$ is the probability density function. The likelihood function is then differentiated with respect to $\bar{\theta}$ in order to find the $\bar{\theta}$ that maximizes the function. However, usually the log-likelihood function

$$l(\bar{\theta} \mid \bar{x}) = \sum_{i=1}^{n} \ln f_{\bar{\theta}}(x_i) \tag{20}$$

is used instead since it is easier to differentiate this function. Because the logarithm is an increasing injective function, it follows that (19) and (20) are equivalent in this case (Evans and Rosenthal, 2010). Following this procedure for the GEV distribution, we obtain, for $\xi \neq 0$,

$$l(\mu, \sigma, \xi \mid \bar{z}) = -n \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \ln\left(1 + \xi \frac{z_i - \mu}{\sigma}\right) - \sum_{i=1}^{n} \left(1 + \xi \frac{z_i - \mu}{\sigma}\right)^{-1/\xi},$$

if $1 + \xi(z_i - \mu)/\sigma > 0$ for $i = 1, 2, \ldots, n$. $\bar{z}$ is the sample of block maxima. If this inequality is violated for some i, the log-likelihood function equals $-\infty$. This scenario corresponds to that one of the data points is located beyond the distribution's end-point. For $\xi = 0$, we get

$$l(\sigma, \xi \mid \bar{z}) = -n \ln \sigma - \sum_{i=1}^{n} \left(\frac{z_i - \mu}{\sigma}\right) - \sum_{i=1}^{n} \exp\left(- \frac{z_i - \mu}{\sigma}\right).$$

Further, the estimation of the parameters in the GPD distribution is done similarly. Denote the excess losses over u by $\bar{y}$. Then, the log-likelihood is given by

$$l(\beta_u, \xi \mid \bar{y}) = -N_u \ln \beta_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{N_u} \ln\left(1 + \xi \frac{y_i}{\beta_u}\right)$$

provided that $\xi \neq 0$ and $(1 + \xi y_i/\beta_u) > 0$ for $i = 1, 2, \ldots, N_u$. For the case $\xi = 0$, the log-likelihood function is equal to

$$l(\beta_u \mid \bar{y}) = -N_u \ln \beta_u - \frac{1}{\beta_u} \sum_{i=1}^{N_u} y_i,$$

(Coles, 2001).

2.4 Backtesting

When backtesting a risk measure, the purpose is to examine if the underlying model that estimates the risk measure is accurate (Hull, 2018).

2.4.1 Backtesting VaR

In this thesis, we consider $\mathrm{VaR}_{0.99}$. By (2), we have

$$\Pr(L > \mathrm{VaR}_{0.99}) = 0.01.$$

Hence, we expect that the losses will exceed VaR 1% of the times. In reality, we cannot expect that the actual number of violations of VaR is exactly equal to 1%. Thus, the question is then how much the actual frequency can deviate from the expected frequency and still be considered as acceptable. To answer this question, we use the Kupiec test (Hull, 2018).

Consider the probability mass function of the Binomial distribution,

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k},$$

which is interpreted as the probability of obtaining k successes out of n trials, with the probability of success equal to p. In our setting, a violation of VaR is counted as success, p is than equal to $1-\alpha$ and n is the total number of observations. The cumulative distribution function is then given by

$$\Pr(X \leq k) = \sum_{i=1}^{k} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}.$$

The test, called the Kupiec test, is performed as follows. If the actual number of violations is smaller than the expected number, we calculate $\Pr(X \leq k)$, where k is the actual number of violations. On the contrary, if the actual number of violations is greater than the expected number, $\Pr(X \geq k)$ is calculated, where again, k is the actual number of violations. Finally, these probabilities are compared to the chosen significant level. If the probability is less than the significant level, the model is rejected, otherwise it is not (Hull, 2018). To apply a two-sided test, we find a confidence interval of the number of expected violations of VaR. This can be done by using the inverse of binomial distribution function (Dowd, 2005).

2.4.2 Backtesting ES

A backtest is performed to see how accurate the estimates of ES are. Here, we use a backtest introduced by Acerbi and Szekely (2014). Recall the definition of ES in (4). Then we have

$$\mathrm{ES}_{\alpha,t} = \mathbb{E}(L_t \mid L_t > \mathrm{VaR}_{\alpha,t}) = \frac{\mathbb{E}(L_t I_t)}{\mathbb{E}(I_t)} = \frac{\mathbb{E}(L_t I_t)}{1-\alpha},$$

where I_t is the indicator function

$$I_t = \begin{cases} 1 \text{ if } L_t > \mathrm{VaR}_{\alpha,t}, \\ 0 \text{ if } L_t \leq \mathrm{VaR}_{\alpha,t}. \end{cases}$$

As mention earlier, we want to test if ES is correctly estimated for all days or if it is incorrectly estimated for some day, i.e., if it is under- or overestimated for some day. Hence, the null hypothesis is that ES is correctly estimated by the model for all days and the alternative hypothesis is that it is incorrectly estimated for at least one day (Acerbi and Szekely, 2014).

Further, define the test statistic Z by

$$Z = -\frac{1}{T(1-\alpha)} \sum_{t=1}^{T} \frac{L_t I_t}{ES_{\alpha,t}} + 1.$$

Under the null hypothesis, the ratio $L_t I_t / ES_{\alpha,t}$ equals one and since we expect $T(1-\alpha)$ exceedances of VaR, we have that $Z = 0$. Instead, if $ES_{\alpha,t}$ is incorrectly estimated for some day, $L_t I_t / ES_{\alpha,t}$ is not equal to one and we obtain $Z > 0$ or $Z < 0$ depending on if ES is over- or underestimated (Acerbi and Szekely, 2014).

Of course, we can not expect to obtain exactly $Z = 0$ even if ES is reasonable correctly estimated. Instead, we rely on the critical values to decide if the model is acceptable. Acerbi and Szekely (2014) present critical values for a range of distributions for $\alpha = 0.975$ and $T = 250$. However, we are interested in the critical values for $\alpha = 0.99$. One can then ask if the critical values for $\alpha = 0.975$ and $\alpha = 0.99$ coincide? Also, Acerbi and Szekely (2014) only report critical values for the left tail of the distribution. Since we are also interested in the right tail, we simulate the distribution of Z to obtain this value.

Acerbi and Szekely (2014) state that the critical values are stable for different distribution. We assume that the losses follow a normal distribution with mean 0 and standard deviation 1. Z is then simulated (using Monte Carlo simulation, see Dowd 2005) 500,000 times with $T = 625$ and $\alpha = 0.99$[1]. T is chosen to 625 because then the distribution of Z for $\alpha = 0.99$ seems to coincide with the distribution of Z for $\alpha = 0.975$ in Acerbi and Szekely (2014). If T is not increased, the distribution of Z seems to converge to the degenerated distribution located at the point 1 when $\alpha \to 1$. Table 1 illustrate the lower and upper critical values for a two-sided test for a range of significant levels.

sign. level	lower	upper
10%	-0.70	0.59
5%	-0.86	0.70
1%	-1.05	0.84

Table 1: Critical values for different significant levels. Lower is the critical value for underestimation and upper is the critical value for overestimation.

3 Data

The data consists of daily closing prices of the OMXS30 index (see Nasdaq 2020) from January 1^{th} 1988 to December 31^{st} 2016, in total 7281 observations. The data were downloaded from Swedish House of Finance – FinBas (2020) at March 25^{th} 2020. The closing prices were then transformed to daily changes in percent

[1]MATLAB was used to perform the simulation of Z (MATLAB, 2019).

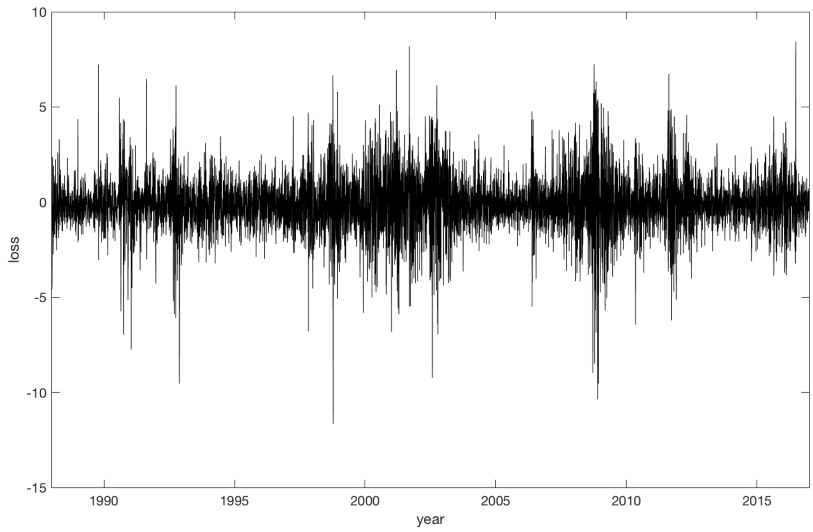

Figure 3: Daily percentage losses of OMXS30 from January 1^{th} 1988 to December 31^{st} 2016.

by

$$L_i = -100\left(\frac{P_i - P_{i-1}}{P_{i-1}}\right)$$

where P_i is the price of the index at day i and L_i is the loss at day i. A plot of the daily losses are shown in figure 3. Note that we interpret losses as positive and consequently gains as negative. Descriptive statistics of the data can be found in table 2.

Descriptive statistics	
number of obs.	7281
mean	-0.047
st. dev.	1.444
skewness	-0.197
kurtosis	4.434
min.	-11.652
max.	8.424

Table 2: Descriptive statistics of the OMXS30 index, for the period January 1^{st} 1988 – December 31^{st} 2016.

4 Method

The data are divided into in-sample and out-sample periods according to table 3, where the number of observations in each period is given in parentheses.

in-sample	out-sample
1988-2011 (6028)	2012 (250)
1989-2012 (6026)	2013 (250)
1990-2013 (6025)	2014 (249)
1991-2014 (6024)	2015 (251)
1992-2015 (6023)	2016 (253)

Table 3: In-sample and out-sample periods for the data.

For each in-sample period, the parameters of the GEV and GPD distributions are calculated, i.e., the parameters are updated once a year. To find the parameters (μ, σ, ξ) of the GEV distribution, extract the greatest loss each quarter and then apply the MLE to this new data set[2].

In the GPD, the threshold has to be decided. Recall the threshold selection process in section 2.2.3. The mean residual life plot is illustrated in figure 4. The graph seems to be approximately linear for $u = 5$ to 7, implying that we should chose $u = 5$. However, according to McNeil and Frey (2000), the threshold should be chosen such that 10% of the observations are categorized as extreme events. Choosing $u = 5$, only 0.3% of the observations are regarded as extreme. But the graph is also approximately linear for $u = 2$ to 4. If $u = 2$, 6.4% of the observations are regarded as extreme, which is closer to the 10% McNeil and Frey (2000) proposed. Thus, the threshold was chosen to 2%. The MLE is then applied to the losses greater than 2% to find (β_u, ξ)[3].

Then $\text{VaR}_{0.99}$ is calculated by (8) and (15) for each out-sample period using the parameters of the corresponding in-sample period. Count the number of violations of VaR, i.e., how many losses are greater than VaR. Finally, apply the Kupiec test to the number of violations for $\alpha = 0,99$, at the 95% level of confidence.

[2] The "in2extRemes" library in R was used for the MLE (Gilleland and Katz, 2016).
[3] The "in2extRemes" library in R was used for the MLE (Gilleland and Katz, 2016).

At last, calculate $ES_{0.99}$ numerically for the GEV distribution and analytically for the GPD. Before $ES_{0.99}$ is backtested, the distribution of the test statistic Z was simulated. Thereafter, $ES_{0.99}$ can backtested by the test of Acerbi and Szekely (2014).

5 Results

5.1 Block maxima

We have 112 quarterly block maxima M_n, with on average $n = 62.76$ observations in each block. The MLE of the parameters (μ, σ, ξ) for the GEV distribution is given in table 4. Based on these estimates, VaR is calculated by (8). Note that VaR of year 2012 is calculated by the estimates from 1988-2011, VaR of 2013 is based on the estimate of 1989-2012, etc. The estimates of VaR are presented in table 5. In total, we have 1253 observations in the out-sample periods 2012-2016. Therefore we expect $0.01 \cdot 1253 = 12.53$ violations of VaR under this period. Actual number of violations became 12 for $\xi \neq 0$ and 9 for $\xi = 0$. A 95% confidence interval for the number of violations is given by $[6, 20]$. Hence in both cases, the numbers of violations fall in this interval.

Further, ES is calculated. The estimates of ES are also found in table 5. ES is then backtested. For the case $\xi \neq 0$, the test statistic z is equal to 0.048 and for $\xi = 0$, we have $z = 0.222$. Reviewing table 1, we see that none of the cases can be rejected for the 95% level of confidence.

	$\xi \neq 0$			$\xi = 0$	
	μ	σ	ξ	μ	σ
1988-2011	2.543	1.044	0.174	2.645	1.135
1989-2012	2.568	1.066	0.161	2.664	1.150
1990-2013	2.523	1.022	0.167	2.619	1.107
1991-2014	2.469	0.991	0.185	2.572	1.084
1992-2015	2.481	0.968	0.177	2.577	1.053

Table 4: MLE of the parameters in the GEV and Gumbel distributions.

	VaR		ES	
	$\xi \neq 0$	$\xi = 0$	$\xi \neq 0$	$\xi = 0$
2012	3.044	3.168	4.198	4.168
2013	3.078	3.194	4.235	4.207
2014	3.013	3.130	4.133	4.106
2015	2.945	3.072	4.058	4.028
2016	2.946	3.063	4.019	3.991

Table 5: VaR and ES for the years 2012-2016 under the GEV distribution.

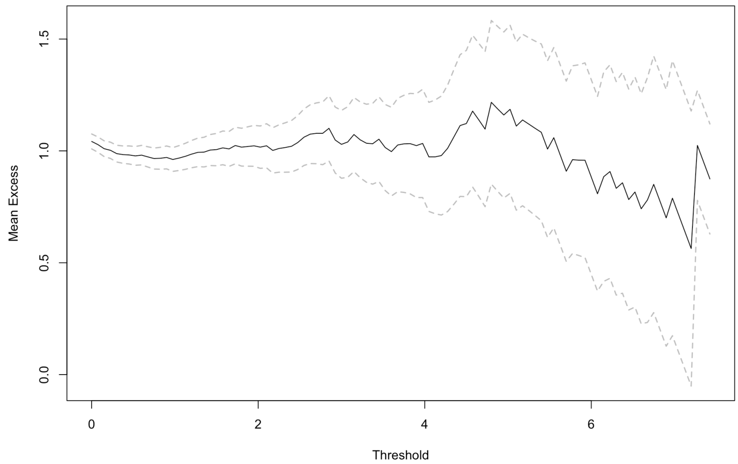

Figure 4: Mean residual life plot of the financial data with 95% confidence interval.

5.2 Peaks-over-threshold

We proceed with the results of the POT model. The frequency of extreme observations for each in- and out-sample period is given in table 6. We can note that the frequency is greater for the in-sample periods compared with the out-sample periods, with one exception. Further, the MLE of the parameters of the GPD is given in table 7. The estimates of VaR and ES are then calculated by the same in-sample and out-sample periods as in the previous model. The results are presented in table 8.

Finally VaR and ES are backtested again. Regarding VaR, we get 4 violations for both $\xi \neq 0$ and $\xi = 0$. Since $4 \notin [6, 20]$, we reject this model. For ES, we get $z = 0.661$ for $\xi \neq 0$ and $z = 0.654$ for $\xi = 0$. Again, consulting table 1, both

$z = 0.661$ and $z = 0.654$ fall in the non-rejection region for 95% level of confidence.

in-sample		out-sample	
year	freq	year	freq
1988-2011	6.9%	2012	5.2%
1989-2012	7.1%	2013	1.2%
1990-2013	7.0%	2014	1.6%
1991-2014	6.7%	2015	7.6%
1992-2015	6.9%	2016	4.7%

Table 6: The frequency of threshold exceedances for each in-sample and out-sample period.

	$\xi \neq 0$		$\xi = 0$
	β_u	ξ	β_u
1988-2011	1.077	-0.024	1.052
1989-2012	1.080	-0.027	1.052
1990-2013	1.079	-0.038	1.040
1991-2014	1.072	-0.033	1.038
1992-2015	1.045	-0.027	1.018

Table 7: MLE of the parameters in the GPD distribution.

6 Discussion

The purpose was to study which one of the models BM och POT that gives the most accurate estimates of VaR and ES. To evaluate the estimates of VaR and ES, backtesting was performed. When backtesting VaR, the POT model was rejected, while the BM model was not rejected. When VaR has been backtested in other studies, such as Cerovic and Karadzic (2015) and Marinelli et al. (2007), BM was outperformed by POT. Also, recall the argument of choosing POT over BM by Embrechts et al. (2005),

> The block maxima method . . . has the major defect that it is very

	VaR		ES	
	$\xi \neq 0$	$\xi = 0$	$\xi \neq 0$	$\xi = 0$
2012	4.038	4.037	5.043	5.089
2013	4.055	4.054	5.053	5.106
2014	4.030	4.029	4.996	5.069
2015	3.983	3.981	4.959	5.019
2016	3.961	3.959	4.927	4.977

Table 8: VaR and ES for the years 2012-2016 under the GPD distribution.

> wasteful of data; to perform our analyses we retrain only the maximum losses in large blocks. (Embrechts et al., 2005, p. 275)

Thus, many other studies have come to the conclusion that POT is better than BM, which is the opposite of the result in this study. However, we should observe that the analysis in this report is based on a greater number of observations than in Cerovic and Karadzic (2015) and Marinelli et al. (2007). We should also note that the advantage in POT compared to BM, as Embrechts et al. (2005) express, is that POT does not throw away useful (i.e., extreme) data. When the number of observations increases, it is possible that this advantage in POT is of less importance. Still, this argument does not fully explain why BM performs better than POT. The argument of a greater number of observations implies that the methods should perform equally well. At the same time, Bekiros and Georgoutsos (2005) use a relatively short sample period of five years and conclude that none of the methods can be rejected.

Further, the frequency of losses that exceed the threshold is greater in the in-sample periods compared to the corresponding out-sample periods, with one exception. This could indicate that the risk is overestimated in the out-sample periods. Because POT extract the extreme losses more effectively than BM (see Embrechts et al. 2005), this could be an argument why POT overestimates the risk even more than BM. To avoid this type of overestimation, a longer data set could be used. Another approach could be to combine the two models; divide the data into blocks and then apply POT but with a maximum number of observations allowed to be drawn from each block.

In addition, none of the models were rejected when ES was backtested at the 95% level of confidence. This can be seen as that ES is not equally sensitive as VaR to the over-estimation of risk, which is the case especially in the POT model. This would then imply that VaR is better choice for the evaluation of the two models.

Further, before estimating VaR and ES, the parameters of the distributions must be estimated: (μ, σ, ξ) in GEV and (β_u, ξ) in GPD. GEV and GPD are two different distributions, but in this context their purpose is the same; they show the distribution of extreme losses. Also, the tail-index ξ is the same parameter in the two distributions. Theoretically, we should get the same estimation of ξ in the methods (Coles, 2001). A weaker hypothesis is that the sign of ξ should be the same in the models. This is also the case in Gilli and Këllezi (2006). However, in this study we get $\xi > 0$ in BM and $\xi < 0$ in POT. When $\xi < 0$, the tail of the distribution is shorter. Since BM performs better in the backtesting of VaR, this indicate that the true distribution of the extreme losses has a thicker tail, i.e., $\xi > 0$. This would imply that BM is a more accurate model than POT.

Additionally, it is possible that the choice of α influence the result. As mention earlier, a high α needs to be chosen for the formulas of VaR and ES to be accurate (see Dowd 2005). Here, we let $\alpha = 0.99$, but an even higher α should be even better. Since POT extract the extreme events more efficient than BM, it is possibly that POT is more sensitive to the choice of α than BM. A solution is then to chose a higher α. On the other hand, this would imply that fewer observations will be considered as extreme, which also can lead to poor estimates.

7 Conclusion

How much can the stock price fall, was asked in the beginning. To answer this question, we use the two risk measures VaR and ES. However, how the risk measures should be estimated is not straightforward; there exist several methods for this purpose (see Hull 2018). Here, we focus on the methods BM and POT, which are based on extreme value theory. Hence, the purpose was to analyze which one of the the two methods BM and POT that produced the most accurate estimates of VaR and ES. To be able to do this analysis, we performed backtesting on VaR and ES. The results showed that BM gives more accurate estimates than POT. This result does not coincide with the general conclusion of which method that is the most accurate (see Dowd 2005 and Embrechts et al. 2005).

To improve the study a larger data set could be used. However, as mention earlier, the data set in this study have already more observations than in other studies, such as Bekiros and Georgoutsos (2005) and Marinelli et al. (2007). Also, none of the other studies mentioned earlier, perform the analysis on OMXS30 and it is possible that distribution changes over different stocks and indices. However, to be able to establish the result in this study, more research has to be done, e.g., use a different data set or use another method than MLE for the estimation of the distributions. Choosing another value of α could also bring some new light on the discussion.

References

Acerbi, C. and Szekely, B. (2014), 'Backtesting expected shortfall', *MSCI, https://www.msci.com/documents/10199/22aa9922-f874-4060-b77a-0f0e267a489b* .

Bekiros, S. D. and Georgoutsos, D. A. (2005), 'Estimation of value-at-risk by extreme value and conventional methods: a comparative evaluation of their predictive performance', *Journal of International Financial Markets, Institutions and Money* **15**(3), 209–228.

Cerovic, J. and Karadzic, V. (2015), 'Extrem value theory in emerging markets: evidence from the montenegrin stock exchange', *Economic Annals* **60**(206), 87–116.

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

Cotter, D. J. (2006), 'Extrem value estimation of boom and crash statistics', *European Journal of Finance* **12**(6-7), 553–566.

Dowd, K. (2005), *Measuring market risk*, second edn, Wiley.

Embrechts, P., Frey, R. and McNeil, A. J. (2005), *Quantitative Risk Management Concepts, techniques and tools*, Princeton University Press Princeton and Oxford.

Evans, M. J. and Rosenthal, J. S. (2010), *Probability and statistics the science of uncertainty*, second edn, Freeman.

finbas stockholm house of finance (2020), 'Omxs30, omx stockholm 30, 1988-01-01-2016-12-31, daily, last', *https://data.houseoffinance.se/finbas/indices* .

Gilleland, E. and Katz, R. W. (2016), in2extremes: Into the r package extremes - extreme value analysis for weather and climate applications, Technical report, National Center for Atmospheric Research, NCAR/TN-523+STR, 102 pp., D0I: 10.5065/D65T3HP2.

Gilli, M. and Këllezi, E. (2006), 'An application of extreme value theory for measuring financial risk', *Computational Economics* **27**(2–3), 207–228.

Hull, J. C. (2018), *Risk Management and Financial Institutions*, fifth edn, Wiley.

Marinelli, C., D'addona, S. and T.Rachev, S. (2007), 'A comparison of some univariate models for value-at-risk and expected shorfall', *International Journal of Theoretical and Applied Finance* **10**(6), 1043–1075.

MATLAB (2019), *version 9.7.0.1319299 (R2019b) Update 5*, The MathWorks Inc, Natick, Massachusetts.

McNeil, A. J. and Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**(3-4), 271–300.

Nasdaq (2020), 'Omx stockholm 30 index (omxs30)', https://indexes.nasdaqomx.com/Index/Overview/OMXS30, [2020-05-26].

Ohlin, J. (2018), 'Finanskrisen - vad var det som hände?', SVT, September 15, https://www.svt.se/nyheter/ekonomi/finanskrisen-vad-var-det-som-hande, [2020-05-14].

Sundkvist, F. (2014), 'Den svarta måndagen', Svenska Dagbladet, July 10, https://www.svd.se/den-svarta-mandagen.

SVT (2020), 'Värsta börsraset i modern tid', March 12, https://www.svt.se/nyheter/snabbkollen/raset-pa-borsen-nu-tio-procent, [2020-05-14].