On Value-at-Risk and the more extreme
A study on quantitative market risk measurements

Abstract
Inline with the third pillar of the Basel accords, quantitative market risk measurements are investigate and evaluated comparing JP Morgan’s RiskMetrics and Bollerslev’s GARCH with the Peek over Threshold and Block Maxima approaches from the Extreme Value Theory framework. Value-at-Risk and Expected Shortfall (Conditional Value-at-Risk), with 95% and 99% confidence, is predicted for 25 years of the OMXS30. The study finds Bollerslev’s suggested t distribution to be a more appropriate distributional assumption, but no evidence to prefer the GARCH to the RiskMetrics. The more demanding Extreme Value Theory procedures trail behind as they are found wasteful of data and more difficult to backtest and therefore evaluate.

Key words: Expected shortfall, EVT, RiskMetrics, GARCH, Value-at-Risk, Basel
1 Introduction
Value-at-Risk (VaR) was introduced by JP Morgan when the RiskMetrics approach was developed, in 1989, as a management tool for risk supervising trading books. The need for a single figure to represent the aggregated risk for one or several departments or trading books was needed and so invented. RiskMetrics used lagged returns, standard deviations and correlations in a fixed-parametric model to conduct one-day ahead, out of sample, forecasts to predict the maximum loss, given a high confidence. As the empirical field of application grow and got legislatively adopted in the Basel Accords, the models got more sophisticated and complex, many still based on the same basic concepts. Engle’s and Bollerslev’s ARCH and GARCH models soon found its’ successful use in the field of risk management, using maximum likelihood to estimate (the earlier fixed) parameters for a more agile model fit. VaR is a point estimate with the tendency to underestimate the actual loss in extreme events. The need to understand these more extreme situations is why the Expected Shortfall (ES), also referred to as CVaR (the expected loss conditioned upon that the VaR is violated), has therefore become more and more important to measure, on a stand along basis or complementing VaR. Since the tail-observations are nominally extreme they can be suspected to follow its’ own distribution, which emphasizes the need to studying the more rare and extreme movements isolated from the more common scenarios. Extreme Value Theory (EVT) provides the tools to estimate these possible distributions that the most extreme observations can be suggested belonging to and applications has been presented to calculate both VaR and ES on these distributions.

The aim of this study is to compare VaR and ES predictions using RiskMetrics, GARCH and EVT. In EVT two methods are used, namely Block Maxima (BM) and Peeks over Threshold (POT). The RiskMetrics and GARCH are used assuming standard normal Gaussian and the t distribution. The 95% and 99% confidence levels are studied. The study investigates the losses of long holdings on the OMXS30 between 10th of April 1990 to 7th May 2015. The preferred procedure are concluded from all four investigated procedures by examining model fit statistics and VaR and ES predictions and backtesting. Earlier research is presented in 1.1, the theoretical framework is outlined in section 2, followed by data presentation and methodology in

1 Shorted investments follows the same methodology.
section 3 and 4 succeeded by the result (5) and conclusions (Error! Reference source not found.).

1.1 Earlier Research

Value-at-Risk is a compulsory review reported to regulatory committees in every country in the European union, with many applicants outside the supervision of the Basel committee. The second Basel (2004) accord outlines the directive for market risk reporting and suggests a baseline, primitive method to account for the risk held by the institutions, who are encouraged to derive more sophisticated procedures to benefit from more accurate measures. This framework is extended within Basel III (2011) and will be more enhanced in the soon to be realised Basel IV. JP Morgan and Reuters (1996), establishing the RiskMetrics Group and were pioneers in the area of market risk measures, thus before the second accord, which Guldimann revised in 2000 to spread the framework freely to the entire financial sector. As the field grew, more dynamic and complex methods were suggested, many utilizing parametric models as GARCH to predict VaR and ES, as shown by Angelidis, et. al. (2004). Engle, R.F. (2001), Jánský, et. Al. (2011), Jiménez-Martín et. al. (2009) and Orhan et. al (2011), just to name a few. Accounting for the undetermined uncertainty in the excess tail, extreme value procedures has grown in field of market risk management and studies of different procedures and comparisons have been released for the same risk measurements such as Tsay (2005, 2009), Zhao, X. (2009), Qinlu, C. (2014). The field of risk management is ever so important and expanding incessantly, never has the prediction of losses be this beneficial.
2 Theoretical framework

Studying financial time series that represents the value of a portfolio or a single instrument, using logarithmic prices, at time $t$ is denoted $V_t$ and is treated as a random variable and the return, value change, is the profit-and-loss (P&L) or the return distribution. In EVT, in line with risk management practitioners’ commonly used methodology, the P&L distribution is used to denote the loss function after dropping the $P$ and considering the negative return function to study the upper (right hand side) distributional tail. Given a horizon $\ell$, measured discreetly in days, the loss function $L(\ell)$ of a holding during a given period $[t, t + \ell]$ is denoted as

$$L_{[t,t+\ell]} := -(V_{t+\ell} - V_t).$$

Here $\ell$ is 1, corresponding to the approximate daily percentage change known as the logarithmic return or continuously compounded return of the losses. The loss function $L(\ell)$ is assumed an iid random series where $t = 1, 2, ..., T$. The same iid assumption yields the non-transformed return series in the VaR estimation. (McNeil, Frey, Embrechts 2005, p. 25-28) The confidence level denoted $\alpha \in (0,1)$ that represents distributional quantiles in VaR and ES estimations, will be set to .95 and .99 in line with the common notation of risk management practitioners. Therefore $1 - \alpha$, .05 and .01, represents the probability of a VaR exceedance; a violation (McNeil, Frey, Embrechts 2005, p. 38).

2.1 Volatility predicting procedures

In this subsection VaR and ES calculations using volatility predictions form GARCH and RiskMetrics are presented together with the necessary error distribution assumptions and backtesting procedures. In this section original log-return series considered.

2.1.1 Error distributions

The joint density function of the random variable $r_t$ can be written as the product of conditional density functions conditioned upon the previous observations as

$$F_{\tau_1, \tau_2, ..., \tau_T}(r_1, r_2, ..., r_T) = \left(\prod_{j=2}^{T} f_{r_j|r_1, r_2, ..., r_{j-1}}(r_j|r_1, r_2, ..., r_{j-1})\right) f_{r_1}(r_1).$$

In order to estimate the parameters $\alpha_0$, $\alpha_1$ and $\beta_1$ we assume conditional normality $Z_t \sim N(0,1)$ and the initial values of $\sigma_t^2$ is set to the unconditional variance according
to equation 17. Given that the normal assumption of the error term is true, the joint
conditional distribution of $r_t, t = 2, 3, ..., T$, conditioning on $r_1, ..., r_{t-1}$, is

$$

f_{r_1|r_2, ..., r_{t-1}}(r_t|r_1, r_2, ..., r_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{r_t^2}{2\sigma_t^2}\right\},
$$

where the marginal density $r_t$ is dropped. The parameter estimates is found by
maximizing the log-likelihood function

$$

\mathcal{L}(r_t|\alpha_0, \alpha_1, \beta_1, \nu) = f_{r_2, ..., r_t|r_1, \sigma_t^2}(r_2, ..., r_t|r_1, \sigma_t^2) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left(\log(\sigma_t^2) + \frac{r_t^2}{\sigma_t^2}\right).
$$

Furthermore, the error term can be assumed to follow different distributions, below
Bollerslev’s (1987) suggested t distribution is studied. The probability density
function is given by

$$

\mathcal{L}(\alpha_0, \alpha_1, \beta_1) = T \log\left(\frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(v - 2)}}\right) - \frac{1}{2} \sum_{t=1}^{T} \log \sigma_t^2 - \left(\frac{\nu + 1}{2}\right) \sum_{t=1}^{T} \log\left(1 + \frac{r_t^2}{(v - 2)\sigma_t^2}\right),
$$

where $\nu > 2$, $\Gamma(\cdot)$ the gamma function. The lagged conditional volatility and degrees
of freedom standardization $\left(\sqrt{(v - 2)/\nu}\right)$ are introduced to the Student’s t
distribution. As $\nu \rightarrow \infty$ the t distribution converges asymptotically to the normal
distribution. (Bollerslev, 1987)

### 2.1.2 Value at Risk

Value-at-Risk (VaR) is a risk measure for apprehending and monitoring market risk.
VaR can easily be explained, from the financial institutions’ point of view, as the
potential maximum loss given a high probability or, in the eyes of the regulators’ as,
the minimum loss under extraordinary market circumstances. Both are statements
explaining the same, the one to use is dependence upon on the practitioner.

Consider a portfolio of, or a single, financial instrument, a fixed prediction horizon $h$
and $F_L(l) = \Pr(L \leq l)$ as the corresponding loss function. As $F_L$ is assumed
continuous and, in most models of interest, unbounded, to measure the maximum loss
is of no interest as it is simply infinity given by $\inf\{l \in \mathbb{R}: F_L(l) = 0\}$. VaR use this
property as an advantage when calculating the maximum loss given a high probability, or the other way around if you wish. Furthermore, denote a confidence level $\alpha$ (in this paper .95 and .99) the VaR of a financial instrument, given by the smallest return $l$ such that the probability that the loss $L$ exceeds $l$ is no larger than $(1 - \alpha)$, is defined as

$$VaR_\alpha = \Pr(L > l) \leq 1 - \alpha = F_L(l) \geq \alpha,$$

so, VaR is the quantile of the loss distribution. The commonly used confidence levels are 0.95 and 0.99 for regular market risk management use. By definition VaR does not give any information about the severity of a $1 - \alpha$ scenario, which is considered an obvious drawback and there is a need for complementary procedures. (McNeil, Frey, Embrechts 2005, p. 37-39)

Taking the expectation of the returns distribution into account is to distinguish between $VaR_\alpha$, with no mean remission, and $VaR_{\alpha, mean} := VaR_\alpha - \mu_L$. As the time horizon is short and the $\mu_L$ is close to zero the mean correction can be comfortably disregarded (McNeil, Frey, Embrechts 2005, p. 38).

Under the normal assumption VaR$_\alpha$ is given by the loss distribution $F_L$ with variance $\sigma^2$ and fixed $\alpha \in (0,1)$ is defined as,

$$VaR_\alpha = \sigma_t \phi^{-1}(\alpha),$$

where $\sigma_t$ is forecasted volatility, $\phi^{-1}(\alpha)$ denotes the standard Gaussian quintile at $\alpha$ and since $F_L$ is strictly increasing $F_L(VaR_\alpha) = \alpha$, therefore

$$\Pr(L \leq VaR_\alpha) = \Pr \left( \frac{L - \mu}{\sigma} \leq \phi^{-1}(\alpha) \right) = \phi(\phi^{-1}(\alpha)) = \alpha,$$

given that the returns are assumed linear.

If the loss distribution violates the normality assumption, which is not uncommon in empirical studies on financial time series, the data is often better modelled by a distribution with higher kurtosis, yielding higher probability for the more extreme values; thus thicker tails. Using the t distribution suggests the returns function, such as $(L - \mu)/\sigma$ in equation 7, is standard t with $v$ degrees of freedom. Denoting the loss model $L \sim t(v, \mu, \sigma^2)$, the first moment is $E(L) = \mu$ and the second $var(L) = \sqrt{\sigma^2 (v - 2)/v}$, when $v > 2$ and $\sigma$ is the standard deviation yielding

$$VaR_\alpha = \sigma_t t^{-1}_v(\alpha),$$

(9)
where $t_v$ is the distribution function of the standard t. (McNeil, Frey, Embrechts 2005, p. 39-40)

### 2.1.3 Expected Shortfall

The ES application on VaR estimates is both straightforward and arbitrary, given the VaR estimate and assumed distribution. For a loss $L$ with distribution function $F_L$ the ES at confidence level $\alpha$ is defined as

$$ES_\alpha = E(L|L > VaR_u) = \frac{1}{1 - \alpha} \int_{\alpha}^{1} VaR_u(F_L)du,$$

where $u \geq \alpha$ meaning all observations beyond the quantile corresponding to the specified confidence level $\alpha$ in the assumed distribution function. The integral gives the tail-expectation for the confidence level associated with the VaR interpretation minimum loss given a low confidence. Suppose that the loss distribution is normal with zero mean and variance $\sigma^2$ then

$$ES_\alpha(L) = \frac{1}{1 - \alpha} \int_{\alpha}^{1} \phi(l)dl$$

where the $\phi$ is the normal density. Assuming that the loss distribution in $L_t = (L - \mu)/\sigma$ is standard t, with $v$ degrees of freedom, where $v > 1$. Then ES of the standard t distribution is calculated by integration to be

$$ES_\alpha(L_t) = \sigma_t \sqrt{v/(v-2)} \frac{g_v(t_v^{-1}(\alpha^{th}))}{1 - \alpha^{th}} \left(\frac{(v-2) + (t_v^{-1}(\alpha^{th}))^2}{v-1}\right).$$

here $t_v$ is the $\alpha^{th}$ quantile and $g_v$ the density of Bollerslev’s suggested t distribution. (McNeil, Frey, Embrechts 2005, p. 44-46)

### 2.1.4 Volatility predictions using GARCH

Engle’s (1982) breakthrough, presenting the Auto Regressive Conditional Heteroscedasticity (ARCH) model steams from the empirical notion of volatility clustering observed in financial returns and takes this into account predicting conditional volatility. If the returns distribution $r_t$, for $t = 1, 2, ..., T$, is strictly stationary and with variance $\sigma^2_t$ the process can be presented as

$$r_t = \mu + \epsilon_t,$$

where the mean $\mu$ is assumed zero and the error term $\epsilon_t$ is dependent on time $t$ according to
\[ \varepsilon_t = \sigma_t z_t, \]  
(14)

here \( Z_t \) is strict white noise, WN(0,1), and \( \sigma_t \) is the conditional standard deviation in period \( t \) from the utilized ARCH(\( q \)) with innovations from the assumed error distribution. (Engle, 1982)

The original ARCH(\( q \)) is given with an estimated constant and the lagged squared returns \( r_t^2 \) since it is an unbiased estimator of \( \sigma^2_{t-1} \).

\[ \sigma^2_{t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2. \]  
(15)

This under the assumption that \( \alpha_0 \) and \( \alpha_i \) is strictly positive for \( i = 1, 2, ..., q \).

Bollerslev (1986) later Generalized the ARCH model adding the lagged conditional variance according to

\[ \sigma^2_{t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j}, \]  
(16)

where \( \alpha_0, \alpha_i \) and \( \beta_j \) is assumed to be strictly positive for all \( i = 1, 2, ..., q \) and \( j = 1, 2, ..., p \) to ensure positive variance. The GARCH(1,1) is obtained as \( q = p = 1 \), also \( \alpha_i + \beta_j < 1 \) must be satisfied to ensure positive variance. The initial exogenous conditional variance used to estimate the GARCH is obtained by assuming the unconditional variance, calculated as \( h \to \infty \) the conditional variance converges to the unconditional variance as

\[ r_{t+h}^2 \to \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}. \]  
(17)

For forecasting, the \( \sigma^2_{t+1} \) is based on the infinite history of the process. In practice an approximation is made because of the finite information in the observed random variable \( r \), the forecast for \( h = 1 \) is given as

\[ \sigma^2_{t+1} = \alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma^2_t. \]  
(18)

In practice, parameter estimates usually yields \( \alpha_1 < .25 \) and \( \beta_1 > .7 \). A \( \beta_1 \) close to 1 implies a clustering of persistent, slow changing, volatility, whereas a high value on \( \alpha_1 \) suggests that volatility reacts rapidly to market movements. (Bollerslev 1986, Taylor 1986, McNeil, Frey, Embrechts 2005)
2.1.5 Volatility predictions using RiskMetrics

The collapsed exponentially weighted moving average (EWMA) used by RiskMetrics for the random variable $r$ gives

$$
\sigma_{t|t-1}^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2,
$$

(19)

where the variable $r$ is decomposed as familiarly from the previous section

$$
r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \rightarrow IID \mathcal{N}(0,1).
$$

(20)

In equation 19, $\sigma_t^2$ is the volatility at time $t$, lambda is the weight factor showing the decay in the expanded $MA$ series. The fixed lambda controls the relative impact between previous volatility and return. The use of lambda ensures positive variance and the non-stationarity is avoided as no parameters are estimated, the reader is encourage to view documentation on the IGARCH for more information on this. The findings of the RiskMetrics group is that the $EWMA$ provides the most accurate volatility forecasts of the realized variance as $\lambda = 0.94$ for daily data and $\lambda = 0.97$ for monthly. In line with these findings, $\lambda = 0.94$ is used. As the model base the result on that the random variable $r$ contains infinite history and this in not empirically true, the forecast estimator is approximated according to

$$
\sigma_{t+1}^2 = 0.94\sigma_t^2 + 0.06r_t^2.
$$

(21)

(RiskMetrics, 1996)

2.1.6 Examining model fit

At time $t$, the $VaR_{\alpha}^{t,h}$ quantile is estimated for the period $t + h$ that yield a prediction. Assuming a continuous loss distribution $Pr(L_{t+h} > VaR_{\alpha}^{t,h}) = 1 - \alpha$ and by construction to be exceeded (violated) precisely $(1 - \alpha)$% of observed times, referred to as the nominal size (Orhan & Köksal, 2011). The violation is recorded using an indicator function to denote 1 for the observed loss $L_{t+h}$ being larger then the prediction $VaR_{\alpha}^{t,h}$ as

$$
I_{t+h}^{(\alpha)} := I_{\{L_{t+h} > VaR_{\alpha}^{t,h}\}}.
$$

(22)

If the predicted $VaR_{\alpha}^{t,h}$ is a one-step-ahead prediction from a volatility prediction model the indicators, recording violations, should behave similar to an iid Bernoulli random variable with violations close to $(1 - \alpha)$. The recorded violations are referred to as the empirical size and should be binomial with expectation $N(1 - \alpha)$ (McNeil, Frey, Embrechts 2005, p. 55). The unconditional coverage is tested using
Kupiec likelihood ratio test to examine if the empirical size deviates from the nominal size according to

\[ LR_{uc} = 2 \ln \left( 1 - \frac{V}{N} \right)^{N-V} \left( \frac{V}{N} \right)^{V} - 2 \ln((1 - \alpha)^{N-V} \alpha^V), \]  

where \( LR_{uc} \sim \chi^2(1) \) and as the empirical \( \frac{V}{N} \) deviates from \( \alpha \) the test statistic grows larger. Rejecting the null hypothesis suggested a model misspecification that yields an incorrect empirical size.

To evaluate ES is similar as for VaR. The volatility residual \( S_{t+h} \) is constructed by observing \( ES_{t+h} \) and tested for zero mean. Moreover, \( S_{t+h} \) is a martingale difference series defined as \( S_{t+h} = (L_{t+h} - ES_{t+h})I_{t+h} \) where \( I_{t+h} \) is the indicator function in Equation 23 and \( S_{t+h} \) satisfies \( E(S_{t+1}|F_t) = 0 \), lined out as

\[ S_{t+h} = \sigma_{t+h}(L_{t+h} - ES_{t+h})I_{\{L_{t+h}>VaR_{t+h}\}}. \]  

In practice the success of the expected shortfall estimation is examined by testing for

\[ E(S_{t+1}) = E \left( (L_{t+1} - ES_{t+1})I_{t+1} \right) = 0, \]  

where the successful model does not reject the zero mean. The violation residuals, defined as

\[ R_{t+1} = \frac{S_{t+1}}{\sigma_{t+1}}, \]  

are observed to provide more insight on the coverage. (McNeil, Frey, Embrechts 2005, p. 55, 163)

### 2.2 Extreme Value Theory

EVT is used for ES and VaR calculations as well as the variance predicting procedures. Two procedures are investigated to extract the extreme observations that underlay, the estimations of extreme distributions and \( VaR_{EVT} \) and \( ES_{EVT} \) calculations namely Block Maxima (BM) and Peek over Threshold (POT). Roughly explained, the risk measures are calculated upon the estimated limiting distribution, providing VaR and ES estimate for a given high confidence level \( \alpha \). The notation in the section follows Tsay (2009) here a single return of order statistics \( r_j \) and the return series \( X_n \), for \( n = 1, 2, \ldots N \), that makes up the df \( F(X) \).
The left figure illustrates the BM approach where \( x_2, x_5, x_7, x_{11} \) represents the maximum for block 1, 2, 3 and 4. The figure on the right hand side illustrates the POT procedure where the limiting distribution is extracted using a high threshold \( u \) to arrive at \( x_1, x_2, x_5, x_9, x_{11} \). It is obvious how the methods utilized different extraction methods and the importance of specifying the sorting criteria.

Consider \( n \) returns, serially independent with a common cumulative distribution function \( F(x) \), where the minimum return, by order statistics, is \( r_1 = \min_{1 \leq j \leq n} \{ r_j \} \) and the maximum \( r_n = \max_{1 \leq j \leq n} \{ r_j \} \). As long positions are considered the loss function as \( r_1 = -\max_{1 \leq j \leq n} \{ -r_j \} = \{ r^c \} \), where \( c \) denotes the sign change. Negative log returns are therefore used to determine the maxima and distribution of the losses. The CDF, denoted \( F_{n,n}(x) \), is given by

\[
F_{n,n}(x) = \Pr[r_{(x)} \leq x] = \prod_{j=1}^{n} F(x) = [F(x)]^n
\]  

(27)

The true CDF \( F(x) \) of \( r_1 \) is of course never known hence \( F_{n,n}(x) \) of \( r_{(n)} \) is unknown. On the contrary, as \( n \) increases \( F_{n,n}(x) \) becomes degenerated as \( F_{n,n}(x) \to 0 \) for \( x < u \) and \( F_{n,n}(x) \to 1 \) for \( x \geq u \) as \( n \) goes to infinity. EVT aims to determine the sequence \( \{ \mu_n \} \) and \( \{ \sigma_n \} \), where \( \sigma_n > 0 \), so that the distribution of \( r_{(n*)} \equiv (r_{(n)} - \mu_n)/\sigma_n \) converges to one of three possible nondegenerate distribution as \( n \to \infty \), where * identifies the maximum. The sequence \( \{ \mu_n \} \) is a series determining the location and \( \{ \sigma_n \} \) a series of scaling factors. For the limiting distributions, the sharp sequence \( \{ \xi_n \} \) will govern the tail behaviour. The different approaches to extract the extreme observations estimates different distribution. For the Block Maxima approach, the variables are assumed to be Generalized Extreme Value (GEV) distributed and for the Peek over Threshold approach Generalized Pareto distributed (GPD), in association with the GEV. The extraction methods and tail-distribution estimation processes are accounted now outlined.
Figure 2 Probability distribution functions of the generalized extreme value distributions for right tailed losses. The solid line is Gumbel distribution ($\xi = 0$), the dotted line Weibull distribution $\xi = -0.5$ and the dashed line is for the Fréchet distribution with $\xi = 0.9$. Observe the infinite tails of the Gumbel and Fréchet with exponential and a power function decays, in contrast to the finite tailed Weibull.

2.2.1 Block Maxima

In general terms, consider a sequence of stationary iid random returns $X_n$, for $n = 1, 2, \ldots N$, that consists of continues compounded losses. Lets have $S_n = X_1 + \cdots + X_N$ as the sum of iid random variables $X_1, X_2, \ldots$, the appropriate normalization uses sequences of normalizing constants to normalize the sum as $(S_n - \mu_n)/\sigma_n$, defined by $\mu_n = nE(x_n)$ and $\sigma_n = \sqrt{var(x_n)}$, we have

$$\lim_{n \to \infty} Pr \left( \frac{S_n - \mu_n}{\sigma_n} \leq x \right) = \phi(x).$$

(28)

Limit distributions for normalizations within EVT consists of a number of $M_{n,j} = \max(X_{n,1}, \ldots, X_{n,M})$ maxima that are iid random variables referred to as block maxima. Then the distribution function of the standard GEV is given by

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

(29)

where $1 + \xi x > 0$. GEV consists of distributions defined by $H_{\xi,\mu,\sigma}(x) := H_\xi((x - \mu)/\sigma)$, where $\mu$ is the parameter controlling the location and where $\sigma(>0)$ is scaling the distribution. The $\xi$ is the shape parameter in $H_\xi$ that defines the distribution as one of three types of GEV distributions as
\[ H_\xi(x) = \begin{cases} \text{Fréchet: } & \xi > 0 \\ \text{Gumbel: } & \xi = 0 \\ \text{Weibull: } & \xi < 0. \end{cases} \] (30)

The Weibull is a so-called short tailed distribution, with finite right endpoint, in contrast to the Fréchet and the Gumbel for whom the end points are infinite. The latter two only differ in the decay where the Fréchet is the slower, allowing for larger kurtosis. (McNeil, Frey, Embrechts 2005, p. 265-276)

Assuming that the block maxima \( M_n \) of iid rvs converge in distribution by normalization meaning that there exist sequences of real constants \( \mu_n \) and \( \sigma_n \) where \( \sigma_n > 0 \) that gives

\[
\lim_{n \to \infty} \Pr(M_n - \mu_n) / \sigma_n \leq x = \lim_{n \to \infty} F^n(\sigma_n x + \mu_n) = H(x). \quad (31)
\]

For \( H(x) \) being a non-degenerate distribution function, \( F \) is said to be in the maximum domain of attraction of \( H, F \in MDA(H) \). Therefore, by the Fisher-Tippett, Gnedenko theorem, \( H(x) \) must be a distribution of type \( H_\xi(x) \). For further insight of this theorem, also known as the “convergence of types theorem”, the reader is referred to Emberchts, Klüppelberg and Mikosch (1997) for in-depth theoretical consultation.

In practice, for some return variable \( r_j \), where \( j = 1, 2, ..., T \), that consists of data from an unknown distribution \( F \) within the MDA of \( H_\xi \), for some \( \xi \). The theory implies that the true distribution of the \( m \) number of block maximum \( M_n \) can be approximated by fitting the GEV distribution \( H_{\xi,\mu,\sigma} \), given a large enough \( m \) as in Figure 1. To achieve this, data is continually divided as

\[
\left\{ r_1, \ldots, r_n, r_{n+1}, \ldots, r_{2n}, r_{2n+1}, \ldots, r_{3n}, \ldots, r_{(m-1)n+1}, \ldots, r_{nm} \right\}, \quad (32)
\]

where the observed data is \( r_{im+j} \), where \( 1 \leq j \leq n, \ i = 0, 1, ..., m - 1 \) and for simplicity assuming \( N = nm \), separating the series into \( m \) non-overlapping subsamples with \( n \) observations in each block, where the block maximum of the \( j \)th block is denoted as \( M_{n,j} \). The established subsample maxima \( M_{n,j} \) are preferably referred to as a sample of from an EVT distribution (Tsay, 2005, p. 6). The intention is to have \( n \) sufficiently large to asymptotically apply the extreme value theory to each subsample \( M_{n,m} = (r_{im+j} - \mu_n) / \sigma_n \) and therefore the block maximum observations \( M_{n,j} \) to asymptotically follow an extreme value distribution (Longin, 1999, p. 1105).
The estimation of the GEV, denoted $h_{\xi, \mu, \sigma}$, is done by maximizing the following expression

$$l(\xi, \mu, \sigma; M_{n,1}, ..., M_{n,m}) = \sum_{i=1}^{m} \ln h_{\xi, \mu, \sigma}(M_{n,i})$$

$$= -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \ln \left(1 + \xi \frac{M_{n,i} - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \xi \frac{M_{n,i} - \mu}{\sigma}\right)^{-1/\xi},$$

where $\sigma > 0$ and $1 + \xi (M_{n,i} - \mu)/\sigma > 0$ for all $i$. Estimation under maximum likelihood assumes independent observations and as $n$ is large, the $M_{n,i}$ observations are also independent. An obvious trade-off appears estimating the GEV using block maxima when determining the number and size of the blocks ($m$ and $n$, respectively), whereas a larger $n$ gives more accurate approximation of the GEV distribution with low parameter bias, whereas larger $m$ leads to more blocks and lower variance in the estimated GEV parameters. (McNeil, Frey, Embrechts 2005, p. 267-275)

The relation between $m$ and $n$ is dictated by the subsample size $n$, usually $n = 21, 63, 125, 252$ which correspond to one observation every (approximate trading-) month, quarter, semester and year. Christoffersen et al. (1998) suggest a block size of 10 to 15 trading days to appropriately model the S&P500, while Longin (2000) suggests 21 trading days, the later more appropriate as of the theoretical connection. The preferred BM model is determined by model fit evaluation based on procedures such as parameter evaluation, GEV fitted residuals and quantile-to-quantile plotting, covered in a latter section. (Tsay, 2005, p. 9-11)

### 2.2.2 Risk measurement application on the GEV distribution

Given the parameter estimates from the BM procedure and the block size $n$, the VaR is obtained from the estimated asymptotic distribution of the losses according to

$$1 - p^* = \begin{cases} 
\exp \left[- \left(1 + \frac{\xi_n (M^*_n - \mu_n)}{\sigma_n}\right)^{-1/\xi_n}\right] & \text{if } \xi_n \neq 0, \\
\exp \left[- \exp \left(- \frac{M^*_n - \mu_n}{\sigma_n}\right)\right] & \text{if } \xi_n = 0,
\end{cases}$$

where $p^*$ is a small upper tail probability of the given GEV distribution that indicates the potential loss and $M^*_n$ the $(1 - p^*)^{th}$ quantile from the subsample. Because return series are usually serially uncorrelated or have weak serial correlations the following probabilistic relationship is used to obtain VaR as
1 - p = Pr\left(M_{n,j} \leq M_n^*\right) = \left[Pr\left(M_t \leq M_t^*\right)\right]^n.  \hspace{1cm} (35)

This leading to

\[VaR_\alpha = \begin{cases} 
\mu - \frac{\sigma}{\xi} \left\{1 - \left[-n \ln(1 - p)\right]^\xi\right\} & \text{if } \xi \neq 0 \\
\mu - \sigma \ln\left[-n \ln(1 - p)\right] & \text{if } \xi = 0,
\end{cases} \hspace{1cm} (36)\]

note that n represents the number of observations in each block in this representation.

### 2.2.3 Peek over Threshold

On of the drawbbacks of the BM procedure is the wastefulness of data, only retaining one observation in a large block of observations. Threshold exceedance models on the other hand, filter out the desired extreme observations beyond a specified threshold. The threshold and the chance to exceed this level controlled by probabilistic law, meaning that the Peek over Threshold (POT) procedure considers the conditional distribution of \(x = r_t - u\), where \(x\) is the exceeding observations at time \(t\) over some threshold \(u\). The conditional (excess) distribution for \(r \leq x + u\) given \(r > u\) is defined as

\[
Pr(r \leq x + u|r > u) = \frac{Pr(\mu x + \sigma x + u)}{Pr(r > u)} = \frac{Pr(\mu x + \sigma x + u) - Pr(\mu x + u)}{1 - Pr(\mu x + u)}. \hspace{1cm} (37)
\]

In contrast to the GEV, methods for threshold exceedance utilize the General Pareto Distribution (GPD) defined as

\[
\hat{G}_{\xi,\beta}(x) = \begin{cases} 
1 - \left(1 + \frac{\xi x}{\psi(u)}\right)^{-1/\xi} & \xi \neq 0 \\
1 - \exp\left(-x/\psi(u)\right) & \xi = 0,
\end{cases} \hspace{1cm} (38)
\]

where \(\psi(u) = \sigma + \xi(u - \mu)\). Here \(\psi(u) > 0\) and \(x \geq 0\) will be \(\xi \geq 0\) and \(0 \leq x \leq -\psi(u)/\xi\) when \(\xi < 0\). The same parameter representation is used as for the GEV where \(\xi\), \(\sigma\) and \(\mu\) refers to shape, scale and location parameter respectively and as familiar there are three sub-distribution depending on the estimated shape parameter according to

\[
\hat{G}_{\xi,\beta}(x) = \begin{cases} 
Pareto: \xi > 0 
Exponential: \xi = 0 
Pareto type II: \xi < 0,
\end{cases} \hspace{1cm} (39)
\]

when \(\xi < 0\) the distributions have finite right endpoints, such as uniform distributions. For \(\xi = 0\) the distributions are medium-tailed, as the normal and log-normal distributions. Lastly, \(\xi > 0\) holds heavy-tailed distributions such as the t-distribution. As the distributions for normalized maxima converge to a GEV distribution by utilizing the convergence of types theorem. Pickands (1975), Balkema
nad de Haan (1974) showed that for non-degenerate distribution function $G(x)$ and a large threshold $u$, $G_{\xi,\beta}$ will be in the maximum domain of attraction $G_{\xi,\beta} \in \text{MDA}(H_{\xi})$ and therefore also converge in distribution to one of three GPD functions described. Denoting the exceedance observations $X_1, \ldots, X_{N_u}$, that is, $X_t = R_t - u$ for $t = 1, 2, \ldots, N_u$, the log-likelihood for the GPD is calculated

$$\ln L(\xi, \psi(u); X_1, \ldots, X_{N_u}) = \sum_{j=1}^{N_u} \ln g_{\xi,\psi(u)}(X_j)$$

subject to the constraints $\psi(u) > 0$ and $1 + \xi Y_j / \psi(u) > 0$ for all $j$. (McNeil, Frey, Embrechts 2005, p. 275-278, Tsay, 2005, p. 20)

### 2.2.4 Risk measurement applications on the GPD

For $x \geq u$, we have that

$$F(x) = \Pr(X > u) \Pr(X > x|X > u)$$

$$= F(u) \Pr(X - u > x - u|X > u)$$

$$= F(u) F_u(x - u)$$

$$= F(u) \left(1 + \frac{\xi}{\psi(u)}\right)^{-1/\xi}.$$  

Which, if $F(u)$ is known, gives the tail probabilities. In practise $F(u)$ is estimated by $N_u/n$ that is the proportion of sample values above the threshold $u$. For a reliable estimate of $F(u)$ implies assuming a sufficiently large proportion of observations above the threshold. To obtain a high quantile, $\alpha = (1 - p)$, of the underlying distribution this expression is inverted to interpreted as a VaR. VaR is then equal to

$$\text{VaR}_\alpha = q_x(F) = u + \frac{\psi(u)}{\xi} \left(\frac{1 - (1 - p)}{F(u)}\right)^{-\xi} - 1.$$  

The associated ES is calculated, using the $\text{VaR}_\alpha$ estimated in Equation 43 and assuming $\xi < 1$, as

$$\text{ES}_\alpha = \frac{1}{1 - \alpha} \int_0^1 q_x(F) \, dx = \frac{\text{VaR}_\alpha}{1 - \xi} + \frac{\psi(u) - \xi u}{1 - \xi}.$$  

(McNeil, Frey, Embrechts 2005, p. 282-284)
2.2.5 Threshold block size selection and EVT model fit

For the POT method, choosing the threshold $u$, over which the extreme value distribution should be estimated, is crucial. The threshold needs to be high enough for the asymptotic theory of GDP to be valid while still extracting a sample so large to minimize the parameter bias. The selection relays on guidelines, in combination with model fit statistics to decide and confirm the chosen level. The literature suggests a sample mean excess plot defined by the points

$$ (u, e_n(u)), x_1^n < u < x^n_n, $$

where $e_n(u)$ is the sample mean excess function defined as

$$ e_n(u) = \frac{\sum_{i=k}^{n}(x_i^n - u)}{n - k + 1}, k = \min\{|x_i^n > u|, $$

and $n - k + 1$ is the number of observations exceeding the threshold $u$. In addition to the mean excess plot, algorithmic summary procedures investigating the model fit that run the GDP estimation procedure for an increasing threshold level (or a decreasing number of exceedance) and plots the estimated shape parameter and the empirical high quantile, these investigative tools will also be consulted. Desired in the graphical threshold selection tools are linearity in the high quantile, shape parameter or excess mean over some interval as the threshold $u$ is increased. For the BM procedure, no corresponding selection tools exist to decide the block size why the model fit statistics will rule out maladjusted models. (Coles, 2001, p. 79-86, Gilli & Këllezi, 2006, p. 16)

To investigate model fit there are different model diagnostic tools to evaluate the two EVT procedures. Examining the GDP model fit for the POT procedure the quantile-to-quantile (QQ-) plot, to investigate the residuals of GDP quantiles and ordered data, and the tail plot to compare the estimated distribution with tail observations. To evaluate the GEV model fit for the BM procedure a residual scatter and quantile-to-quantile (QQ-) plot will be consulted to determine how well the model fit the empirical distribution function to evaluate the exponential (for the scatter) and linear (for the QQ-plot) coherence. (Tsay, 2009, p. 13, 21, 26)
3 Data

The data is downloaded from the web page of Nasdaq OMX and spans 10\textsuperscript{th} of April 1990 to 7\textsuperscript{th} May 2015. The descriptive statistics is presented in Figure 3. The mean and median is about zero as expected, there is some negative skewness indicating that the losses (left tail) are somewhat larger than the profits which is also visible in the histogram. The Jarque-Bera test for normality rejects nonnormality because of a large kurtosis and some skewness.

In Figure 4 the return series of this period is presented, showing clear signs of heteroscedasticity and bell shaped volatility clustering.

![Figure 3](image1.png)

\textit{Figure 3} Descriptive statistics and histogram for the investigated OMXS30 data series indicate non-normality as of the Jarque-Bera test and estimated high kurtosis. The approximate zero mean indicated no trend exists, but some skewness as of the measure.

![Figure 4](image2.png)

\textit{Figure 4} The plotted returns over time show heteroskedasticity and clustered volatility. The economic crisis in early and late nineties, early 2000s and late to 2000s and its aftemaths are easily distinguished. There is indication of some lag structure in the material as low volatility periods are followed by the same and high volatility are followed by high, but declining, volatility as typical for financial return series.
4 Methodology

Kevin Sheppard’s MFE toolbox is used in Matlab R2015a to calculate the VaR and ES estimates for RiskMetrics and GARCH procedures. Regarding the EVT procedures, the evir package is used for RStudio, version 0.98.1102, is used to estimate the extreme value VaR and ES.

For GARCH and RiskMetrics, one-step ahead volatility forecasts are estimated utilizing proposed techniques by a rolling window. The VaR and ES predictions are calculated using this forecasted volatility. Both measures are compared to the observed return in the predicted time period $t + 1$ and tested for unconditional coverage, and ES zero-mean given that is VaR exceeded. The confidence level $\alpha$ and the assumed error distribution will vary, testing all combinations of 95th and 99th confidence in combination with the Gaussian and Bollerslev’s suggested t distribution. Both families of procedures are affected by the size of the rolling window and as $n$ is smaller, estimations are more responsive but suffers greater parameter errors. A window of 500 observations, which approximately corresponds to two trading years, will be used to tackle this trade-off. Regarding the EVT procedures, the block size / threshold level are determined by parameter and model fit procedures to counter the apparent trade-off between having enough observations to apply the EVT framework and estimating stable parameters for the estimated extreme distributions. In contrast to the volatility predicting procedures, the entire dataset will be used to estimate the EVT models, yielding an obvious difference, due to the framework and purpose of the EVT.

Numeric predictions from all covered procedures, and assumptions, will be presented and discussed to yield a true empirical result in combination with the model evaluations.
5 Results

This section presents the concluding results for the volatility predicting processes and extreme value theory, in descending order. As a reminder, losses are considered and are made positive by negative transformation, why VaR and ES estimates are presented as positives. All Value-at-Risk and Expected shortfall measurements are based on a long OMXS30 holding of 1 000 000 SEK.

5.1 Volatility predicting processes

The MFE estimation engine is used to estimate the conditional volatility, used for volatility predictions. Figure 5 describes the conditional volatility over the investigated timeframe. Different periods experience different volatility and lagged time-dependence suspected to present in the material, which would reject homoscedasticity. This motivates the choice of models that accounts for changing volatility. The periods of 1994 to 1997 and 2004 to 2006 are both in the return series and conditional volatility seen as slow-moving periods followed by a chock, this slow moving market conditions are rare.

Alter the distributional assumption between the normal and Bollerslev’s suggested t distribution is commonly done in financial econometrics. Plotting the degrees of freedom suggests that the majority of the investigated data is non-normal. As mentioned in the theoretical part, the t distribution converged asymptotically to the normal as $v \to \infty$, this might be experienced in the mid to end nineties where the degrees of freedom hits a restraining ceiling, probability enforce by the package.

Figure 5 In correspondence with Figure 4, the conditional volatility in the investigated OMXS30 is varying with the periods of economic turbulence and lower in-between. Here the first 499 observations are used to calculate the conditional volatility for day 500 that also is the start of the graph.
developer, at about $v = 10000$. The degrees of freedom where $v > 50$ are marked as zeros in Figure 6 to indicate the censored observations. Moreover, the degrees of freedom are all $2 < v$ by construction. Please consider the Appendix for the original plot, this is not presented here as the extreme estimates dwarfs the remainder.

Figure 6 The $v$ degrees of freedom for the estimated $t$ distribution, using the rolling window of 500 observations. As mentioned in 2.1.1 the $t$ distribution converges asymptotically to the normal distribution as $v \to \infty$. This is the likely to be the case for some periods in the mid-nineties as the degrees of freedom literary goes through the roof, original graph is Figure 19 in the appendix. As the degrees of freedom vary, from its floor of 3, this upholds the $t$ as a suggested distribution.

In the outlined results from both volatility-predicting processes the coverage is of primary interest and a non-rejection is indicating that the proper model fits can’t be rejected. The conducted tests are nominal size = empirical size for VaR and $E[ES_{residual}] = 0$ for the ES predictions. Descriptive examples and graphs will be provided in the following sections. The coverage is tested for the entire observed time period in contrast to the parameter estimates and risk measure predictions that represent the last observed rolling window. The reader is advised to visit the Appendix for the full graphical outline.

5.1.1 RiskMetrics results

To illustrate the results Figure 7 describes a cutout of the return series, in black, together with the VaR and ES estimates, in blue and red. The risk measures should by construction produce one violation every hundred observation, as the confidence is
The coverage is tested as it is impossible observe this by optical inspection.

For the same process the ES residuals are presented in Figure 8. The residual series should have a zero-mean around the ES estimates to not reject a good model fit. And by inspection, the residuals are suspected not to deviate to that extent of rejecting the zero-mean null hypothesis, this tested in Table 1.

To summarize the results, the RiskMetrics process is quite successful considering the tested coverage and yields VaR and ES predictions that increase as the confidence is increased and the normal is substituted for the t distribution, in line with expectations. The coverage is only rejected in two cases, both considering the normal distribution, for 95% ES and 99% VaR.
<table>
<thead>
<tr>
<th>RiskMetrics</th>
<th>Prediction</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VaR</td>
<td>ES</td>
</tr>
<tr>
<td>95% Normal t</td>
<td>8 573</td>
<td>10 751</td>
</tr>
<tr>
<td>99% Normal t</td>
<td>12 125</td>
<td>13 892</td>
</tr>
</tbody>
</table>

*Table 1 The RiskMetrics models’ predicted VaR and ES for the last model estimated by the rolling window, including tested coverage for the entire dataset. For the VaR and ES coverage the violation ratio and calculated mean is presented, *, ** and *** indicates significance on the 10%, 5% and 1% level respectively.*

Under the normal distribution, considering 99% coverage, the RiskMetrics process underestimates the risk and therefore produces a violation-proportion that is significantly larger then .01. When the coverage is altered to 95% under normality the VaR coverage is accepted, while the ES residual mean is not zero, which is hard to tell by inspecting the ES residual plot, Figure 9, but rejected by the test.

![Figure 9 Standardized ES residuals with significantly tested non-zero mean, indicating a model misspecification.](image)

The model assuming normality and 99% confidence does not reject the ES coverage while rejecting the VaR, ES residuals presented in Figure 10. By inspecting the ES residual plot, the residual distribution seems to be skewed and by descriptive statistics the skewness is .7078. The residual mean is .0423 (95% confidence interval: -.0357, .1204) and the cause for non-rejection is suspected to be the small sample size of 77 observed violations. In contrast to the RiskMetrics model assuming normality for 95% coverage, where the ES coverage is rejected.
To calculate the investigated risk measures using RiskMetrics and altering between 95% and 99% confidence, the models assuming the t distribution are superior, for this period of the OMXS30, testing the VaR unconditional coverage and ES zero-mean.

5.1.2 GARCH results

In contrast to the RiskMetrics process, GARCH models estimate model parameters for every prediction to adjust every estimated model to the changing aspects in the investigated material. A $\beta_1$ close to 1 and a low $\alpha_1$ suggests persistent volatility clustering, as seen in Figure 11 that describes the parameter estimates for the GARCH process under the t distributional assumption and 99% confidence. Varying parameter estimates indicates changing characteristics over the observed time span. Period where the estimated $\beta_1$ decreases is often countered by an increase in $\alpha_1$ or in $\alpha_0$. The constant $\alpha_0$ is normally small as the variation in conditional volatility is picked up in $\alpha_1$ and $\beta_1$, because of this the period in late 2005 is of interest as an obvious drop in $\alpha_1$ and $\beta_1$ is observed together with an increase in $\alpha_0$, here the process might correct the estimates as the rolling window transitions from a low-volatility period to a more turbulent. Again, the earlier described rare market conditions in the periods of 1994-1997 and 2004-2006 gives remarkable parameter estimates. Looking at the original index, these periods are characterised as a slow but steady inclining market with very little volatility over such long time that the GARCH have time to incorporate these characteristics’. As the market experience’s a chock, both in 1997 and 2005, the conditional volatility peaks, $\beta_1$ decreases and $\alpha_0$ increases drastically to account for the variation, as $\alpha_1$ remains small. This particular behaviour is especially true the latter case.
The parameter estimates and predictions are found in Table 2 together with the tested coverage for the entire time series. As the same forecast is used predicting both 95 and 99% VaR and ES, the parameter estimates are the same for both distributional assumptions. The estimates are all the same for an unchanged distributional assumption. The t distributional assumption puts slightly more weight on $\beta_1$, which is favourable but not of great concern as the difference is at the second decimal.

<table>
<thead>
<tr>
<th>Parameter estimate</th>
<th>Constant</th>
<th>ARCH</th>
<th>GARCH</th>
<th>Prediction</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>Normal</td>
<td>.0143</td>
<td>.0923</td>
<td>.8151</td>
<td>8 228</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>.0129</td>
<td>.0943</td>
<td>.8248</td>
<td>8 913</td>
</tr>
<tr>
<td>99%</td>
<td>Normal</td>
<td>.0143</td>
<td>.0923</td>
<td>.8151</td>
<td>11 637</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>.0129</td>
<td>.0943</td>
<td>.8248</td>
<td>12 584</td>
</tr>
</tbody>
</table>

As for the RiskMetrics procedure the only rejected coverage is found when normality is assumed, in the case of 99% coverage for VaR and 95% coverage for ES. Considering the predictions, the VaR and ES is increasing as the normal is substituted for the t distribution and the confidence is increased, in line with the expectation. For 99% GARCH-n, the conclusion established is the same for as for the RiskMetrics procedure. The non-rejected ES coverage is suspected to be caused by the low power yielded from the small sample, here the skewness is estimated to .6872, the residual mean is .0334 (95% CI: -.0470, .1139) for the same number of observations. Also as for the RiskMetrics models, the preferred GARCH models used to predict the investigated risk measures, for the investigated period of the OMXS30, considering
the VaR unconditional coverage and the ES zero-mean are the models assuming Bollerslev’s suggested t distribution, regardless of the confidence level $\alpha$.

5.2 Extreme value processes

The EVT framework is by construction wasteful regarding data in several ways and to different extents depending on the procedure. To accumulate enough extreme observations for the EVT theory to apply the entire data series will be used. First, to estimate an EVT distribution for the losses, the profits (positive returns) are initially separated from the material as according to Figure 12.

![Figure 12 Losses, separated from the return series, for EVT estimation. The characteristics from the original data series are unmistakably the same, still experiencing heteroskedasticity and volatility clustering.](image)

The risk measures are calculated using the parameters of the estimated extreme value distributions.

5.2.1 Block Maxima results

The GEV distribution is fitted to the loss data altering the block size corresponding to an approximate trading month, quarter, semester and year with 21, 63, 125 and 252 observations respectively. The parameter estimates, standard errors with in brackets and risk measurements are presented in Table 3. Investigating the scatter- and QQ-plots does show that the estimated models fit the data well, the scatter gives the impression of exponentially distributed residuals and the ordered against exponential quantiles do not deviate from the constructed linearity.
Table 3 Extreme value distribution estimated parameters and risk measure predictions presented for all investigated block sizes n and corresponding number m. The Shape parameter is not significant for n = 21, also for n = 125 and n = 63 the small sample jeopardises the EVT application.

<table>
<thead>
<tr>
<th>Parameter estimates Prediction</th>
<th>Location (µ)</th>
<th>Scale (σ)</th>
<th>Shape (ξ)</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 21</td>
<td>1.8991***</td>
<td>.8653***</td>
<td>.0211</td>
<td>18 349</td>
<td>32 675</td>
</tr>
<tr>
<td>m = 299</td>
<td>(.2385)</td>
<td>(.1878)</td>
<td>(.2808)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 63</td>
<td>1.2307***</td>
<td>.4834***</td>
<td>.3598***</td>
<td>7 682</td>
<td>14 708</td>
</tr>
<tr>
<td>m = 99</td>
<td>(.0704)</td>
<td>(.0620)</td>
<td>(.1275)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 125</td>
<td>1.0513***</td>
<td>.4105***</td>
<td>.3662***</td>
<td>4 980</td>
<td>9 614</td>
</tr>
<tr>
<td>m = 50</td>
<td>(.0424)</td>
<td>(.0374)</td>
<td>(.0890)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 252</td>
<td>.8748***</td>
<td>.3667***</td>
<td>.3182***</td>
<td>2 328</td>
<td>5 798</td>
</tr>
<tr>
<td>m = 24</td>
<td>(.0268)</td>
<td>(.0229)</td>
<td>(.0624)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The fitted GEV distributions yield all significant parameter estimates except for the shape parameter when to block size is set to n = 21, this insignificance yields this model’s exclusion. As there is no coverage to be tested, because the GEV procedure only provides one VaR estimates for each used confidence level, significantly more weight is put on the model fit and the assumption that the EVT framework is applicable. Furthermore, as the model fit is appropriate for the remaining models the analysis will proceed only considering the GEV model with n = 63 as the other two models include few enough blocks to doubt the applicability of the EVT framework.

5.2.2 Peek over Threshold results

To be able to calculate VaR and ES using the POT procedure the threshold-exceeding portion needs to be identified by concluding the empirical threshold. Figure 13 and Figure 14 investigates a threshold high enough for the subsample to only include the extreme observations while still producing stable estimates for the shape parameter and a high quantile respectively, for increasing thresholds. In Figure 15 the excess mean is investigated for stability while increasing the threshold. In all three plots, linearity beyond the set threshold is desired and several thresholds levels are tested because of inconsistency in the stability plots and their arbitrary and vague indications and interpretations.
Figure 13: The algorithmic summary procedure running the GDP estimation procedure iterating an increasing threshold level (or a decreasing number of exceedance), plotting the estimated shape parameter. Linearity after a suggested $u$ is desired for a high enough threshold to only consider the extreme observations. This is suggested at $u = 1.45$ to have a high enough threshold, followed by linearity before non-linear departure.

In Figure 13 the shape parameter can be considered stable until a threshold level of $u = 1.45$ where about $n_u = 120$ observed losses remains to estimate the GPD.

Figure 14: The algorithmic summary procedure running the GDP estimation procedure for an increasing threshold level (or a decreasing number of exceedance), plotting the empirical high quantile. Linearity following a suggested $u$ is desired for a high enough threshold to only consider the extreme observations.

In Figure 14, investigating the stability of a high quantile, the estimates are stable before $n_u = 48$ observed losses, which suggests a threshold level at $u = 1.7$ to have linearity in the estimates before the line departures and $n_u = 80$ observations available for GPD estimation at this point.
The mean excess plot in Figure 15 is without any obvious departure from 1 to 2 and 2 to 3. At $u = 3$ the extreme value theory is not feasible as only $n_u = 8$ excess observations remains. This enables a threshold set at $u = 1$ and $u = 2$ where $n_u = 304$ and $n_u = 43$ excess observations are available for model estimation. As there is an obvious trade-off between a high threshold, for the EVT to apply, and having enough observations for the parameter estimates to be unbiased, the aggregated suggested thresholds will be used to estimate the GDP and examined using model fit procedures to rule-out not suitable threshold before calculation the VaR and ES. The suggested threshold levels are presented in Table 4.

<table>
<thead>
<tr>
<th>Threshold, $u$=</th>
<th>1</th>
<th>1.35</th>
<th>1.45</th>
<th>1.7</th>
<th>2</th>
</tr>
</thead>
</table>

*Table 4 Model fit investigated threshold levels concluded from the high-quantile-, threshold- and mean excess plots.*

Consulting the QQ-plot and tail plot, to investigate the GPD model fit for the different thresholds, no model is an obvious deviating fit by inspection why all four thresholds are used to calculate the risk measures presented in Table 5.
Parameter estimates and risk measure predictions presented for all investigated thresholds $u$ and corresponding $n_u$ extracted extreme observations. The parameter significance and the shape parameter coefficient are of special interest to conclude the preferred model from which to consider the predicted risk measures. For the OMXS30 the threshold $u = 1$ yields significant scale and shape parameter and is therefore the preferred model, using $n_u = 304$ extreme observations to estimate a heavy tailed limiting distribution. Parameter estimates, standard errors within brackets and risk measures from the POT procedure are presented in Table 5. Further analysis will proceed only considering the model with threshold $u = 1$ because of the parameter significance and good model fit according to the graphical assessment tools, all presented in the appendix. As the location parameter is non-significant further analysis will conducted with causation. As the Shape parameter is positive, and significantly different from zero, the distribution of extreme observations are suggested to have a heavy tailed distribution such as the t distribution or similar. Figure 16 describes the estimated extreme distribution clearly illustrating the slow decay, here VaR 95% and 99% are marked with solid lines and ES 95% and 99% as dashed lines, in the order of appearance.

Table 5 GPD estimated parameters and risk measure predictions presented for all investigated thresholds $u$ and corresponding $n_u$ extracted extreme observations. The parameter significance and the shape parameter coefficient are of special interest to conclude the preferred model from which to consider the predicted risk measures. For the OMXS30 the threshold $u = 1$ yields significant scale and shape parameter and is therefore the preferred model, using $n_u = 304$ extreme observations to estimate a heavy tailed limiting distribution.

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>Location ($\mu$)</th>
<th>Scale ($\sigma$)</th>
<th>Shape ($\xi$)</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
<th>ES 95%</th>
<th>ES 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u = 1$</td>
<td>-0.1589</td>
<td>0.3164***</td>
<td>0.1218*</td>
<td>9 898</td>
<td>15 503</td>
<td>13 490</td>
<td>19 870</td>
</tr>
<tr>
<td></td>
<td>$n_u = 304$</td>
<td>(.1948)</td>
<td>(.0853)</td>
<td>(.0667)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = 1.35$</td>
<td>-0.0668</td>
<td>0.2919**</td>
<td>0.1348</td>
<td>11 448</td>
<td>16 197</td>
<td>14 502</td>
<td>19 991</td>
</tr>
<tr>
<td></td>
<td>$n_u = 150$</td>
<td>(.3940)</td>
<td>(.1384)</td>
<td>(.1003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = 1.45$</td>
<td>-0.6799</td>
<td>0.4885*</td>
<td>0.0436</td>
<td>9 740</td>
<td>17 539</td>
<td>14 631</td>
<td>22 786</td>
</tr>
<tr>
<td></td>
<td>$n_u = 116$</td>
<td>(.6496)</td>
<td>(.2454)</td>
<td>(.0995)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = 1.7$</td>
<td>.0443</td>
<td>.1952</td>
<td>.1989</td>
<td>14 663</td>
<td>17 484</td>
<td>16 520</td>
<td>20 041</td>
</tr>
<tr>
<td></td>
<td>$n_u = 80$</td>
<td>(.6194)</td>
<td>(.1647)</td>
<td>(.1588)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = 2$</td>
<td>-1.0818</td>
<td>.5853</td>
<td>.0219</td>
<td>8 614</td>
<td>17 793</td>
<td>14 343</td>
<td>23 728</td>
</tr>
<tr>
<td></td>
<td>$n_u = 43$</td>
<td>(2.1700)</td>
<td>(.7043)</td>
<td>(0.1998)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 16 An illustrative example putting the VaR and ES predications in relation to one another. The VaR 95% and 99% are represented with solid lines, from left to right, as is the ES 95% and 99% represented with dashed lines. The solid slope represents the theoretical extreme value distribution using the parameter estimates from the preferred GPD when $u = 1$. 

31
To summarize the results, looking at the concluded volatility predicting processed described in Table 6 the parameter estimates are quite close. The GARCH model puts relatively less weight on the lagged volatility relative to the RiskMetrics making the GARCH model more flexible to react to rapid market change. Looking at the predicted VaR it is interesting to note that the distance between the 95% and 99% predictions are larger for the RiskMetrics process in comparison to the narrower GARCH predictions, also the GARCH estimates are included in the span of the RiskMetrics predictions. If this is true in repeated sampling, with support from forecasting evaluation tool such as the mean square error (MSE), such a result should indicate that the GARCH predictions are more precise, simply a recommendation for further studies. In contrast to the nested parameter estimates and VaR predictions, the ES predictions are solely larger for the RiskMetrics model. As the deviation between the predictions are considered small for the different processes (.0428 and .0630 for the 95% and 99% confidence respectively, which makes a difference about 428 and 630 SEK given a portfolio of 1 000 000 SEK). Given this difference the RiskMetrics model is considered the more risk avers prediction the Expected Shortfall in this last model. From the institutional perspective, a too risk averse model would require the investor to hold more liquidity that may cause an opportunity loss.

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>Constant</th>
<th>Return^2</th>
<th>Volatility Predictions</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
<th>ES 95%</th>
<th>ES 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskMetrics - t</td>
<td>0</td>
<td>.06</td>
<td>.94</td>
<td>8 609</td>
<td>13 088</td>
<td>11 110</td>
<td>16 344</td>
</tr>
<tr>
<td>GARCH - t</td>
<td>.0129</td>
<td>.0943</td>
<td>.8248</td>
<td>8 913</td>
<td>12 584</td>
<td>10 682</td>
<td>15 714</td>
</tr>
</tbody>
</table>

Table 6 Parameter estimates and predicted risk measures from the concluded volatility predicting processes. The parameter results shows that the GARCH – t put relatively less weight on the lagged volatility making it a more flexible model to corporate rapid change in the return series. The GARCH VaR predictions are found with in the span of the RiskMetrics’, which also yields larger ES predictions.

Regarding the EVT procedures, the approaches should yield different models as of the different extraction methods, referring to Figure 1. Table 7 presents the estimated EVT distributions with parameter and risk measure estimates.

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>Location (µ)</th>
<th>Scale (σ)</th>
<th>Shape (ξ)</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
<th>ES 95%</th>
<th>ES 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM</td>
<td>n = 63</td>
<td>m = 99</td>
<td>1.2307</td>
<td>.4834</td>
<td>.3598</td>
<td>7 682</td>
<td>14 708</td>
</tr>
<tr>
<td>POT</td>
<td>µ = 1</td>
<td>n_s = 304</td>
<td>-.1589</td>
<td>.3164</td>
<td>.1218</td>
<td>9 898</td>
<td>15 503</td>
</tr>
</tbody>
</table>

Table 7 The different extraction methods and procedures yield different number of extreme observations and
parameter estimates. The similar size and sign of the shape parameter indicate a slow decay and coherence between the processes. The POT procedure is more risk averse for the compared VaR 95% and 99%.

As the BM and POT utilizes different extraction methods and converge in different distributional families, different distributions are expected, especially as the number of observations and the actual observations differ. The similarities in sign and size of the shape parameters indicate that both processes suggest a slow power-decay, similar to a t distribution, among the extreme observations. Figure 17 illustrates the estimated extreme value distributions from the BM and POT procedures, estimating a generalized extreme value distribution and a general Pareto distribution, respectively. The solid line producing a tailed distribution represents the estimates GEV distribution using the parameters from Table 7, together with solid vertical lines representing the VaR 95% and 99% for this procedure. The dashed slope is the GPD outlined used the parameter estimates from Table 7, with corresponding VaR 95% and 99% as dashed vertical lines and dotted lines representing 95% and 99% ES.

![Figure 17](image.png)

*Figure 17 Presenting the estimated GEV and GPD, from the BM and POT procedures, respectively. The Solid distribution is the fitted GEV distribution with 95% and 99% VaR as solid vertical lines from left to right as the confidence increases. The dashed sloped line represents the fitted GPD with corresponding VaR 95% and 99% as dashed lines and 95% and 99% ES dotted, from left to right as the confidence increases. All input are found in Table 7. The shape of the distributions differs heavily but similar risk measure estimates, notable but expected. The GPD estimate is the more risk avers comparing with the GEV estimates.*

The VaR estimates are similar but deviate somewhat, the GPD are more risk avert for the investigated material. The risk measure difference is suspected to decrease as the confidence is increased, when comparing the difference for both confidence levels and converge as \( \alpha \to \infty \).

### 5.3 Concluding results

Stating the results from the concluded models gives several VaR and ES estimates outlined in Table 8 for VaR and Table 9 for ES, all considering a long position of
1 000 000 \textit{SEK} invested in the OMXS30. Regarding VaR predictions, the RiskMetrics and GARCH predictions are close for both levels of confidence amplifying their common properties. Furthermore, relative to the GARCH predictions, the RiskMetrics VaR is lower on the 95% level and higher on the 99% level, this can be an effect of the RiskMetrics’ heavier weight on the volatility term. The EVT procedures gives quite deviating predictions on the 95% level, shrinking the gap as the confidence is increased, as illustrated in Figure 17. Overall, for the 95% confidence level the EVT models’ predictions are outliers around the volatility predicting processes predictions. No clear pattern can be found for on the 99% level among the VaR estimates.

<table>
<thead>
<tr>
<th></th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskMetrics - t</td>
<td>8 609</td>
<td>13 088</td>
</tr>
<tr>
<td>GARCH - t</td>
<td>8 913</td>
<td>12 584</td>
</tr>
<tr>
<td>\text{BM}_{n=63, m=99}</td>
<td>7 682</td>
<td>14 708</td>
</tr>
<tr>
<td>\text{POT}_{u=1, n=304}</td>
<td>9 898</td>
<td>15 503</td>
</tr>
</tbody>
</table>

\textit{Table 8 Preferred models, from all investigated procedure, predicting VaR for a portfolio of one million SEK.}

The characteristics’ from the VaR predictions does reflect the ES estimates when comparing the RiskMetrics and GARCH. The RiskMetrics estimates are larger then from the GARCH and relatively slightly increasing. The EVT estimates for ES are clearly separated and larger then the other processes, all seen in Table 9.

<table>
<thead>
<tr>
<th></th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskMetrics - t</td>
<td>11 110</td>
<td>16 344</td>
</tr>
<tr>
<td>GARCH - t</td>
<td>10 682</td>
<td>15 714</td>
</tr>
<tr>
<td>\text{POT}_{u=1, n=304}</td>
<td>14 502</td>
<td>19 871</td>
</tr>
</tbody>
</table>

\textit{Table 9 Preferred models, from all investigated procedure, predicting ES for a portfolio of one million SEK.}
To further illustrate the differences in the VaR and ES estimates, the predicted losses are plotted as a scatter in Figure 18. Regarding the VaR estimates, the predictions are spread out indicating different distributional tails with different rates of decay. There are more similarities for the ES estimates.

Figure 18 Predicted VaR and ES from all investigated procedures. On the axes are the losses estimated by each procedure given the confidence level. The circles and stars represent VaR and ES estimates respectively, together with a solid 45-degree line. For the VaR estimates the predictions are spread out and indicates slower of faster rates of decay as the prediction is over or under the 45-degree line. For the ES estimated the EVT is far of indicating a heavy tail in comparison to the remainder.

6 Conclusion

Regarding the volatility predictions, the t distribution yields the preferred models due to the concluded non-normality in the return series. The parametric model, GARCH, is not concluded a more appropriate model for the investigated data measuring VaR and ES with 95% and 99% confidence compared to the fixed-parametric RiskMetrics model. Regarding the EVT models, The Peek over Threshold procedure is preferred over the Block Maxima procedure, due to its less wasteful extraction method, that increased the applicability of the extreme value theory. The volatility prediction processes are preferred before the EVT procedures as these models are more easily applied, modified and backtested.
6.1 Recommendations for further studies

The volatility predicting processes can be modified in a variety of ways, especially the GARCH framework. A wider analysis should be conducted to conclude a, perhaps, more appropriate model fit. To conclude the preferred EVT procedure using backtesting analysis is not possible in the outline. A Bootstrap or Monte Carlo application is suggested to provide repeated sampling to carry out backtesting procedures for a limited data series.
7 References


ISBN web 92-9197-859-0.


8 Appendix

Volatility prediction processes

Degrees of freedom from the t distribution

Figure 19 Degrees of freedom for the t distribution, showing the otherwise excluded observations in Figure 6. The estimation is likely to halt at \( \nu = 10000 \) where the distribution is close to normal.

RiskMetrics appendix

VaR and ES predictions
ES residuals
GARCH appendix

Parameter estimates

GARCH parameter estimates for n with 95 percent confidence

GARCH parameter estimates for t with 99 percent confidence
VaR and ES predictions
ES residuals
Extreme value theory Appendix

BM model fit plots

Block size $n = 21$

Block size $n = 63$
Block size $n = 125$

Block size $n = 252$
POT model fit plots

Threshold \( u = 1 \)
Threshold $u = 1.35$
Threshold $u = 1.45$
Threshold $u = 1.7$
Threshold $u = 2$