Niklas Ruusunen

Extreme Value Theory in Tail Risk Forecasting

Laskentatoimen ja Rahoituksen Akateeminen Yksikkö
Kauppatieteiden Kandidaatti

Vaasa 2025

VAASAN YLIOPISTO
Laskentatoimen ja Rahoituksen Akateeminen Yksikkö
Tekijä: Niklas Ruusunen
Tutkielman nimi: Extreme Value Theory in Tail Risk Forecasting
Tutkinto: Kauppatieteiden Kandidaatti
Oppiaine: Laskentatoimi ja Rahoitus
Työn ohjaaja: Maria Zhukova
Valmistumisvuosi: 2025 **Sivumäärä:** 38

TIIVISTELMÄ:

Kandidaatin tutkielma pyrkii selittämään kattavasti rahoitusmarkkinoilla käytetyt riski- ja ennustusmallit painottaen niiden teoreettista taustaa Portfolio teorian, ääriarvo teorian ja normaali jakautuneiden osaketuottojen teorian kautta, ja syventyen yleisesti käytettyihin Value-at-Risk ja Expected Shortfall (Conditional Value-at-Risk) malleihin, joista keskustellaan riskin ennustamisen näkökulmasta. Tutkielman ensisijaisena tarkastelun aiheena on Ääriarvoteorian (EVT) implikaatiot ja sovelletut menetelmät edellä mainittuihin riski malleihin. Hienovaraisempina malleina, ääriarvoteoriaan perustuvat Value-at-Risk ja Expected Shortfall mallit oletetaan toimivan muita vaihtoehtoisia malleja, kuten normaalijakautuneita osaketuottoja olettavaa mallia paremmin. Tutkielman ytimessä on reflektoida aiempaa mallien suorituskykyä vertailevaa tutkimus aineistoa löytääkseen vastauksen ääriarvoteoriaan perustuvien mallien toimivuuteen verrattuna vaihtoehtoihin ja tämän implikaatiota riskien hallinnalle.

AVAINSANAT: Extreme Value Theory, Value-at-Risk, Expected Shortfall, Tail Risk, Forecasting

Table of Contents

Pictures

No table of figures entries found.

Charts

Tables

No table of figures entries found.

1 Introduction

AI disclaimer: This thesis has utilized Grammarly and Chat-GPT to check the language of the text and to suggest ways of fixing the language of the text.

Tail risk is an essential concept in financial risk management that deals with misestimating the probabilities of highly impactful and negative tail events. This misestimation of risk is caused by the tails of a distribution F_X being fatter than estimated; the probability density of the tails events is greater than estimated. The risk of misestimating tail risk makes the modelling and forecasting of tail risk essential for risk management, as under (over) estimating, for example, the expected loss of a portfolio, can lead to severe issues and inefficiencies, which affect investors and institutions alike.

The main challenges in tail risk modelling are that tail events are rare and, therefore, difficult to observe and quantify. Additionally, stock returns and other financial data are nonnormally distributed (Richardson, Smith. 1993), and thus model and distribution selection become highly relevant and difficult. Traditional risk management approaches, such as Value-at-Risk (VaR), do not correctly account for such tail risks due to estimating a single point of the distribution, not the rest of the tail (Hull. 2018. p. 273). As a result, they may underestimate the likelihood and impact of rare tail events. More intricate methods of tail risk modelling, on the other hand, provide a more accurate and comprehensive view of the risk by modelling the distribution and or the tail of the distribution more accurately.

An extremely relevant theory for tail risk modelling is found in extreme value theory (EVT), which was first founded by Gumbel (1935). EVT is part of probability theory and a statistical approach that focuses on the extreme values of a distribution, and, subsequently, provides a framework for modelling tail events. EVT assumes that the extreme values of a given distribution follow a different distribution than the rest of the data, and therefore, require a separate modelling approach to the rest of the distribution.

In finance, and other regions where EVT is utilized, like engineering, the tails of a distribution are often more important than the central region for modelling rare events and their inherent risks.

This thesis examines the existing empirical literature on tail risk models and the utilization of EVT in tail risk modelling. The angle of approach in this thesis is to examine the advantages and possible disadvantages of EVT utilization.

1.1 Objective

The objective and hypothesis are to show that EVT-based tail risk models, especially ones fitted with the Peak Over Threshold (POT) method are superior to the common normally distributed methods of VaR and Expected Shortfall (ES) and even compared to other nonnormally distributed VaR and ES methods, which differ in complexity. Thus, the hypotheses of this thesis are:

$\boldsymbol{H_1}$: EVT-based models give a significant advantage in tail risk modelling compared to normally distributed models.

$\boldsymbol{H_2}$: EVT-based models are on par or better than other more complex tail risk models.

The core purpose of this study is to show the advantages of EVT-based tail risk models compared to the most basic, and thus, most used tail risk models, and why they should be adopted more widely. A secondary purpose is also to establish EVT-based models as competitive choices when considering more intricate models for tail risk estimation. With a wider adaption of more intricate tail risk models, like the ones utilizing EVT, financial institutions and investors alike can protect themselves more effectively from severe losses and other tail risk-related adverse events, that can cause massive losses of capital.

1.2 Structure

This thesis starts by first introducing tail risk management in finance and discussing its theoretical basis. The second part of the thesis then discusses the common approaches of tail risk modelling, these being VaR and ES. For both models, a few common approaches are shown, and their advantages and disadvantages are discussed. The models are introduced with their static versions, but subsequently, forecasting these models is also discussed.

The third part of the thesis discusses the application of EVT-based solutions to tail risk modelling to the methods discussed in the prior part. Additionally, the theoretical basis of EVT is also discussed. The last part of the thesis is a literature review of the research done on EVT-based models and their performance when compared to other measures used for modelling tail risk.

2 Tail Risk in Finance

Tail risk refers to low probability extreme events, which, as referred by the name, lie in the tails of a given distribution of a random variable, such as stock returns. Tail risk arises from various sources, including financial market volatility, geopolitical events, macroeconomic shocks, and natural disasters. These factors and others can significantly impact financial markets, leading to extreme market movements and price fluctuations. For example, the global financial crisis (GFC) of 2008 was a result of a combination of factors, including subprime mortgage defaults, regulatory failures, and liquidity problems, which led to the actualization of tail risk. A more recent example of tail risk actualizing through an unpredicted event was the COVID-19 crash in the spring of 2020. Markets reacted rapidly to the unpredicted implications of a global pandemic, causing stocks to plunge in a matter of weeks and days. Due to the unpredictable nature of the stock market, these tail risk events will continue to happen, and thus, finding measures for mitigating their risks for portfolios is of high importance.

An important factor in tail risk is also that it is a risk that cannot be estimated well with a normal distribution. While Bachelier (1900) was the first to assume that changes in stock prices followed the normal distribution through Brownian motion over 100 years ago, research has evolved since and shown signs of his hypothesis being false. The Q-Q plot in **Figure 1** shows a deviation between theoretical and actual values at the tail ends, hinting at non-normality. Similarly, comparing the histogram of actual returns in **Figure 2** to the theoretical values of **Figure 3** shows similar results, as the actual data is considerably more leptokurtic than the theoretical normally distributed data (excess kurtosis of 12.9647), and it is also slightly negatively skewed (skewness of -0.7925) entailing that the mass of the probability is positive leaning, but the negative tail is greater. From this observation follow the notion that the probability density of extreme negative events would be greater than predicted by a normal distribution.

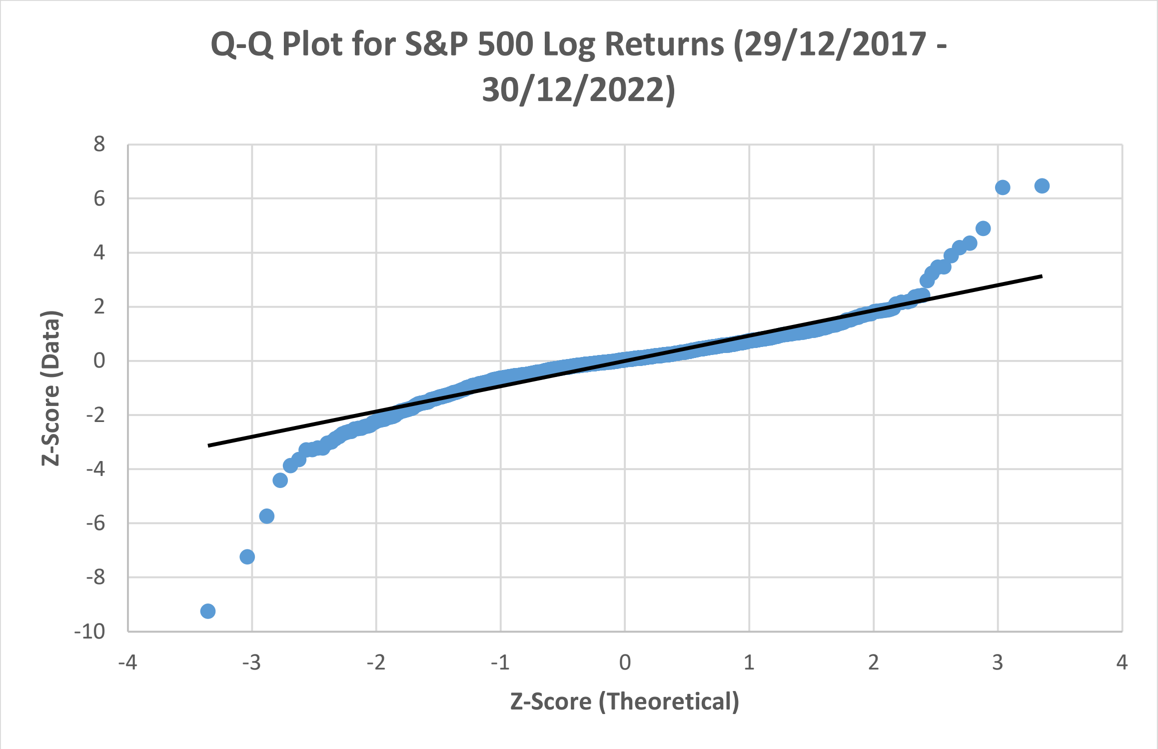

Figure 1 Q-Q Plot for S&P 500 Log Returns

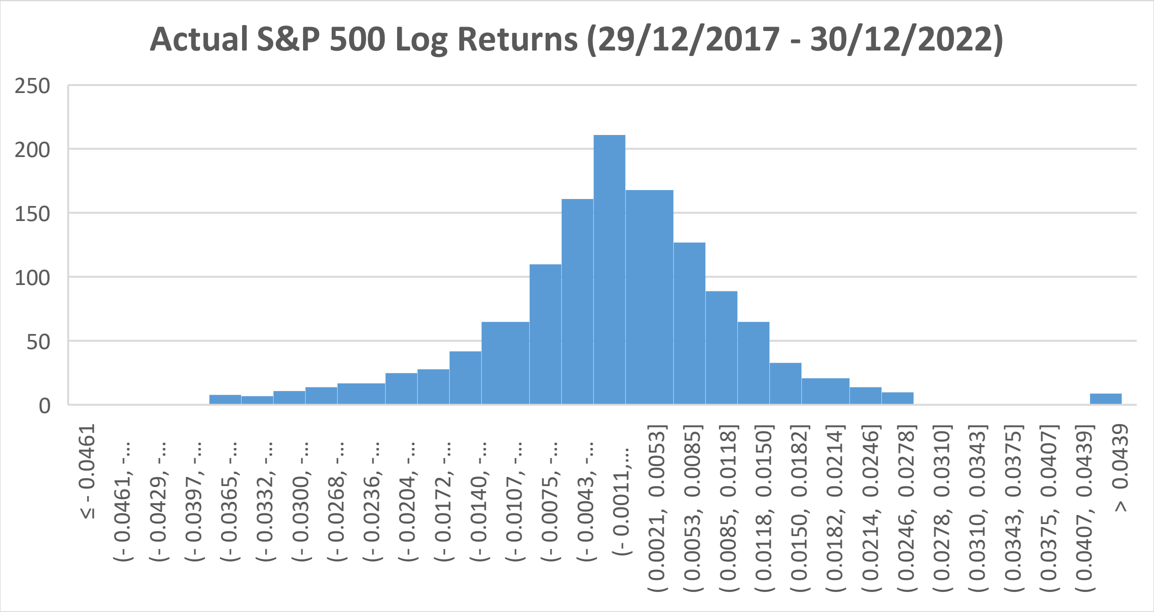

Figure 2 Histogram of Observed S&P 500 Log Returns

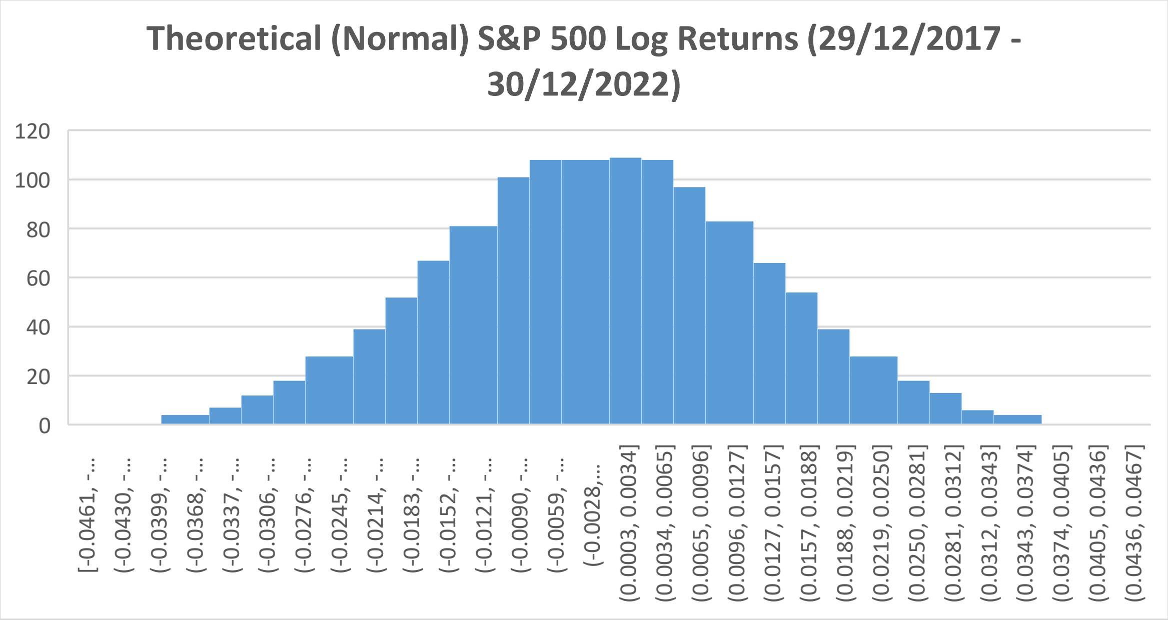

Figure 3 Histogram of Theoretical (Normal) S&P 500 Log Returns

As tail risk models concentrate on the tails of a distribution F_X, a distribution with a fatter tail, like a Student-t distribution with low degrees of freedom, can give better results. In addition, with Richardson and Smith (1993) finding that stock returns are nonnormally distributed, utilizing other distributions than normal is likely feasible. Other nonnormal distributions in addition to the Student-t distribution, like Laplace and Exponential distributions, can also be utilized.

A method of undermining the limitations of normal distributions and others that underestimate the fatness of the tails of the distribution is by considering the tail(s) as separate distributions from the body of the distribution. This method is employed in Extreme Value Theory-based approaches, which are the topic of this thesis and will be discussed later.

2.1 Theoretical Basis of Essential Concepts

Tail risk modelling and management are closely related to the assumption of normality in stock returns and through that, models and theories that follow the assumption of normality. In addition, tail risk modelling obviously is closely related to theories and models of reducing portfolio risk, mainly through the risk component of volatility or standard deviation. The most prominent theory, which assumes normality and is based on reducing risk is the modern portfolio theory, which was developed by Markowitz. (1952). Portfolio theory's main tenant is based on the idea that investors can reduce risk by diversifying their portfolios across different uncorrelated or weakly correlated assets. Markowitz found that the risk for a two or n-asset portfolio is lesser than their weighted average if their correlation is less than one. In the case of 2 assets, this can be formulated as:

$$\sigma_p^2 = \sigma_a^2 w^2 + \sigma_b^2 (1-w)^2 + 2w(1-w)\sigma_a \sigma_b \rho_{ab},$$

Where w is the weight of asset a, σ_x^2 is the variance of each asset and ρ_{ab} is the correlation between the assets. Markowitz showed that by combining assets with different returns and risks, investors could achieve a higher return for a given level of risk or a lower risk for a given level of return.

For a portfolio of greater size than 2 assets, the portfolio variance formula of n assets takes the form of:

$$\sigma_p^2 = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,n}^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix},$$

Where $\begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}$ is a matrix of asset weights in a portfolio with a sum of 1 and $\begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,n}^2 \end{bmatrix}$ is the covariance matrix of the assets in said portfolio.

In the context of tail risk, portfolio theory suggests that investors can reduce their exposure to tail risk by diversifying across different assets and asset classes. Tail risk modelling can help investors to identify and manage the risks associated with extreme market movements. By modelling the tails of the distribution of asset returns, investors can estimate the probability and magnitude of extreme events and design strategies to protect against them, mainly by hedging the tail risk.

3 Common Tail Risk Models

3.1 Value-at-Risk

Value-at-risk (VaR) is a common risk management tool used in finance that measures the maximum loss that an investment portfolio is expected to suffer over a given period of time with a given probability. Banks and other financial institutions typically use it to measure and manage the risk of their portfolios. As noted by Hull (2018. P. 269-270), JPMorgan was credited for creating and popularizing VaR as a risk measure with better understandability than prior methods.

VaR is calculated by evaluating the potential losses of an investment portfolio over a specified period of time under different market scenarios. The results are then used to determine the maximum amount of loss that a portfolio could suffer over a given period of time. VaR is also used to identify areas of potential risk in an investment portfolio and to develop risk management strategies to protect against losses.

According to the Basel Committee on Banking Supervision (2009) and Hull (2018. P. 347-398), in the Basel regulations, VaR is an essential component of the capital adequacy framework for banks. Basel II and III regulations require banks to hold a certain amount of capital based on their VaR estimates for various risk categories, such as credit, market, and operational risks.

Basel II introduced the internal VaR model, allowing banks to use their internal models to estimate VaR, subject to specific minimum standards and supervisory review. This approach recognizes that banks have different risk profiles and that a one-size-fits-all approach to capital adequacy may not be appropriate.

Basel III builds on Basel II by introducing several enhancements to the VaR framework. One fundamental change is introducing a stressed VaR measure, which requires banks to estimate VaR under extreme market conditions, such as a severe recession or financial crisis. This helps to ensure that banks have sufficient capital to withstand adverse market conditions.

In addition, Basel III requires banks to backtest their VaR models to ensure that they are accurate and reliable. Banks must also disclose information about their VaR models and risk management practices to regulators and the public.

3.1.1 Historic Value-at-Risk

The historic approach to Value at Risk is the simplest method of measuring Value at Risk. As stated by McNeil, Frey, Embrechs (2015. P. 64) and Hull (2018. P. 269-272), historic VaR for a confidence level $\alpha \in (0,1)$ measures the quantile $q_\alpha(F_X)$ of a distribution F_X of historic returns series $X = (X_t, X_{t-1}, \dots, X_{t-n})$ where $X_t = \frac{P_t}{P_{t-1}} - 1$ or $X_t = \log\left(\frac{P_t}{P_{t-1}}\right)$. As a static measure for VaR measured from historic data, its uses for forecasting future VaR are poor if we assume that historic returns won’t repeat in the future. However, because the historical method assumes no distribution F_X, it may offer better information about the underlying distribution. In practice, historical versions of VaR are the most popular ones (Hull. 2018. P. 293-298).

3.1.2 Monte Carlo Simulation Value-at-Risk

An alternative approach to measuring Value-at-Risk is by finding a value which matches a given quantile $q_\alpha(F_R)$ is done by a Monte Carlo simulation. In a Monte Carlo simulation, we simulate the Brownian motion of stock returns or prices for a period T from a given mean μ and volatility σ. A simulation of possible outcomes is yielded, and from this, VaR can be estimated as the quantile $q_\alpha(F_X)$ of the simulated returns distribution. This method too, has the advantage of not assuming a distribution. It is best used for any underlying asset which might be hard to model otherwise, like exotic derivatives.

3.1.3 Parametric Value-at-Risk

Parametric (Variance-Covariance) Value-at-Risk assumes that the returns follow some distribution F_X, most commonly the normal distribution. Thus, if normality is assumed, as Bachelier (1900) assumed, a returns distribution F_X is suspected of following a normal distribution $X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2. For short time horizons, μ is assumed to be zero, which leads to VaR being proportional to σ (Hull. 2018). According to McNeil, Frey and Embrechs (2015. P. 65), the $\alpha \in (0,1)$ confidence level parametric VaR can be formulated as:

$$VaR_\alpha = -\mu + \sigma\Phi^{-1}(\alpha),$$

where $\Phi^{-1}(\alpha)$ is the α-quantile of the normal distribution Φ. We can note that the VaR model returns positive values, as in risk management, we estimate losses, which tend to be shown as positive values. In its basic form, the parametric VaR measures the Value-at-Risk for a time period $\Delta_t = 1$. Because of this, the previous VaR model is referred to as one day ahead VaR. Introducing a time period term Δ_t to VaR leads to VaR taking the form:

$$VaR_\alpha = -\mu\Delta_t + \sigma\sqrt{\Delta_t}\Phi^{-1}(\alpha).$$

In addition to the normal distribution, other distributions can also be used for VaR. A commonly used parametric VaR approach is the Student-t distributed VaR. If a returns distribution F_X is assumed to follow a Student-t distribution $X \sim t(df, \mu, \sigma^2)$ with df degrees of freedom, μ mean and σ^2 variance given by $\sigma_X^2 = \frac{df\quad^{2}}{df-2}$ when $df > 2$, the $\alpha \in (0,1)$ confidence level parametric VaR, as shown by McNeil, Frey, Embrechs (2015. P. 66), can be formulated as:

$$VaR_\alpha = -\mu + \sigma t_{df}^{-1}(\alpha),$$

where $t_{df}^{-1}(\alpha)$ is the α-quantile of the Student-t distribution t. The degrees of freedom df can be estimated by $df = n - k$ or more fittingly, by utilizing statistic tests and maximizing the goodness of fit of a t distribution to the data given a certain df.

3.2 Expected Shortfall

Expected shortfall (ES), also known as Conditional Value at Risk (CVaR), is a modification of the parametric VaR model, which considers the expected loss given that we exceed the quantile α of the VaR model. According to McNeil, Frey and Embrechs (2015. P. 69-71), ES is expressed as

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(X)du.$$

ES can be considered the average loss we expect to have given we exceed the threshold given by VaR, weighted by the probability of each loss.

In the case that we assume a normal distribution of stock returns F_X, parametric ES can be formulated as

$$ES_\alpha = -\mu + \sigma\frac{\phi\big(\Phi^{-1}(\alpha)\big)}{1-\alpha},$$

Where $\Phi^{-1}(\alpha)$ is the α-quantile of the normal distribution Φ and $\phi(q_\alpha)$ Is the normal probability density function

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

As stated by McNeil, Frey and Embrechs (2015. P. 72) Student-t distributed version can be formulated as

$$ES_{\alpha} = -\mu + \sigma * \left(\frac{g_{df}\left(t_{df}^{-1}(\alpha)\right)}{1-\alpha} \frac{df + \left(t_{df}^{-1}(\alpha)\right)^{2}}{df - 1} \right),$$

Where df is the degrees of freedom, $t_{df}^{-1}(\alpha)$ is the α-quantile of the Student-t distribution t and $g_{df}(x)$ is the t-distributed probability density function

$$g_{df}(x) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\,\Gamma\left(\frac{df}{2}\right)} \left(1 + \frac{x^{2}}{df}\right)^{-\frac{df+1}{2}},$$

Where $\Gamma(z)$ is a gamma function

$$\Gamma(x) = \int_{0}^{\infty} t^{x-1} e^{-t} dt.$$

3.3 Forecasting Value-at-Risk and Expected Shortfall

In practical use, tail risk models, such as VaR or ES, are used for forecasting tail risk by utilizing different forecasting methods. The static methods mentioned before are combined with volatility forecasting models such as Generalized Autoregressive Conditional Heteroscedastic (GARCH) models first introduced by Bollerslev, T. (1986). The basic GARCH (1, 1) model can be formulated as

$$\sigma_{t}^{2} = \omega + \alpha \epsilon_{t-1}^{2} + \beta \sigma_{t-1}^{2},$$

Where ω is the intercept term, which represents the variance at $t = 0$, α is the coefficient of the autoregressive term, which represents the impact of past realized variance on current variance, $\epsilon_{t-1}^{2} = (R_{t-1} - \mu_{t-1})^{2}$ the squared residual of the returns or realized variance, β is the coefficient of the moving average term, which represents the impact of past GARCH values and σ_{t-1}^{2} is the variance predicted at $t-1$. As β

multiplies the previous σ_{t-1}^2, the first term σ_1^2 is different, this being the long-run variance $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$. The GARCH model can be additionally fitted with additional exogeneous variables when research the impact exogenous factors on volatility. As shown by Hull, J. (2018. P. 229-233), the parameters ω, α and β are estimated by maximizing the log-likelihood function:

$$L(\epsilon_t^2, \sigma_t^2) = \ln\left(\frac{1}{\sqrt{2\pi\sigma_t^2}} e^{\frac{-\sigma_t^2}{2\epsilon_t^2}}\right),$$

Which is subject to $\omega \geq 0, \alpha \geq 0, \beta \geq 0$ and $\alpha + \beta < 1$.

GARCH also has other variations like EGARCH, TGARCH and GJR GARCH. However, they will not be discussed further as only the basic knowledge of GARCH is needed for understanding tail risk forecasting. Additionally, in forecasting VaR or ES with volatility being estimated by GARCH, the mean μ can also be estimated with a simple moving average or exponential moving average.

The forecasted VaR or ES yielded by the calculations provides a more useful measure of risk than a constant model. The model's accuracy may depend on how far the measure is forecasted, with 1 day ahead GARCH providing the best approximations for volatility. The difference between the actual loss (negative returns) of a stock index and a fitted Value-at-Risk model can be observed in **Figure 4**.

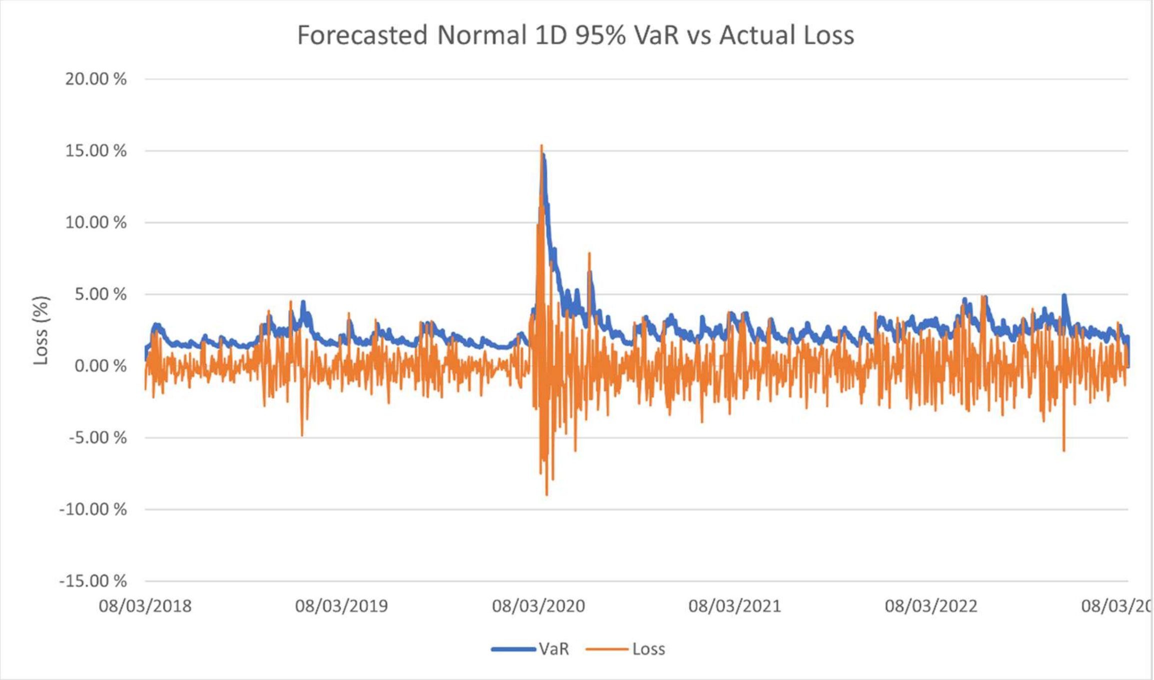

Figure 4 Forecasted Russel 2000 95% VaR

3.4 Measuring the performance of Value-at-Risk and Expected Shortfall

The performance of VaR and ES models is performed by back-testing. Back-testing, simply put is calculating how a model or a measure currently in use would have worked in the past (Hull, J. 2018. p. 285). Back-testing is vital for VaR and ES models because we want to know if the parameters and expectations of the model are over or under-tuned, this being if the model is violated too often or too seldom. In benchmarking and measuring a VaR model, for example, we expect that a 95% VaR model is violated (exceeded) statically around 5% of the time and for $\alpha\%$ VaR, we expect model violation around a $(1-\alpha)\%$ of the time.

A popular model for backtesting was proposed by Kupiec (1995). The backtesting model assumes that if the probability of an exception in a VaR model is p, then it should be assumed that m exceptions are observed in a sample size of n, which leads to

$$-2\ln((1-p)^{n-m}p^m)+2\ln\left(\left(1-\frac{m}{n}\right)^{n-m}\left(\frac{m}{n}\right)^m\right)\sim\chi^2(1).$$

The discussion on differing probability distribution models comes back when testing the prediction ability and the backtesting results of the different models. When fitted on OMXH25 data, it can be observed by comparing the normal 99% VaR of **Figure 5** to the Student-t 99% VaR of **Figure 6** that the normally distributed model is more sensitive to violations of the model.

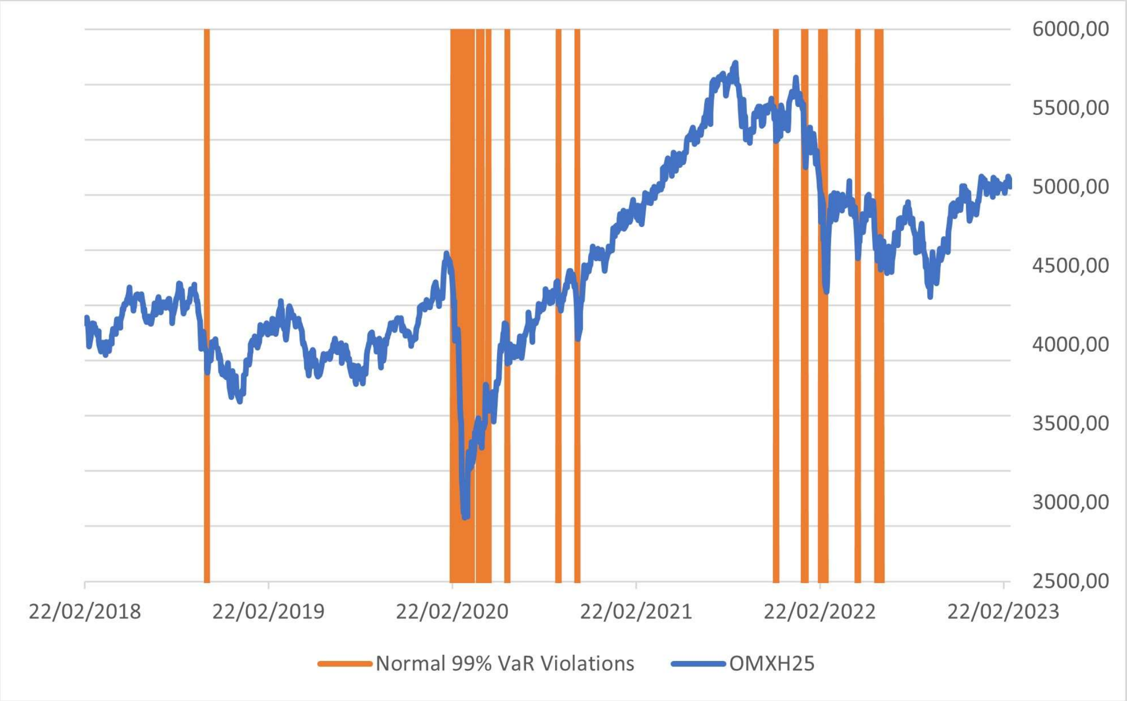

Figure 5 Normal 99% VaR Violations

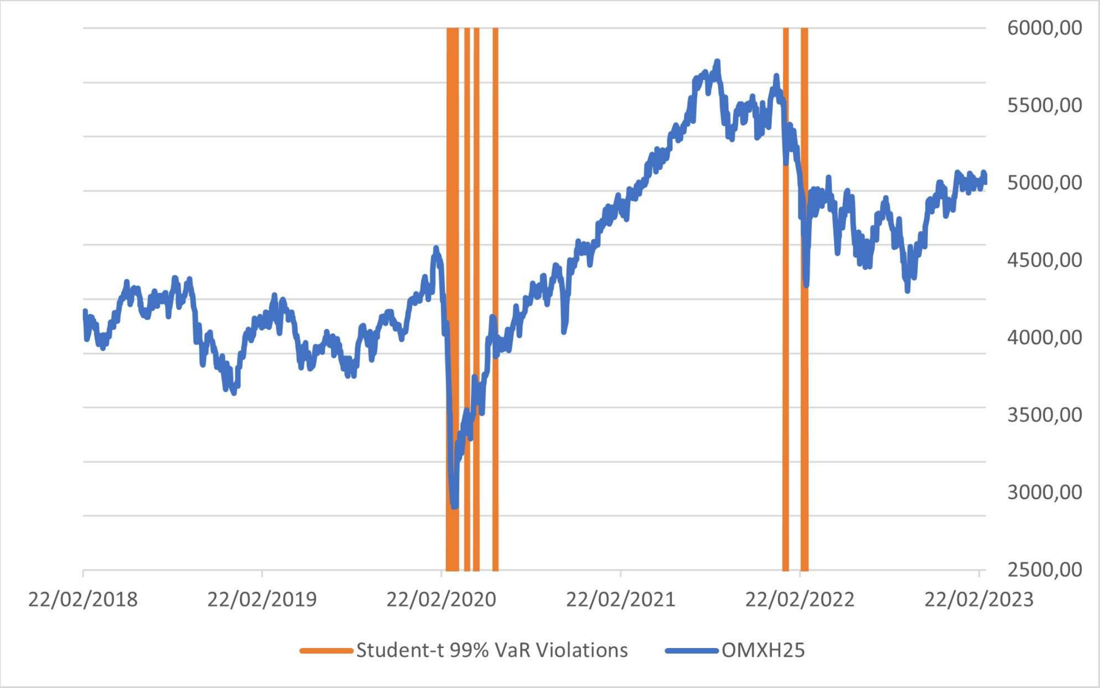

Figure 6 Student-t 99% VaR Violations

3.5 Comparing Value-at-Risk and Expected Shortfall

The primary advantage of Value-at-Risk in tail risk management is its understandability; VaR can provide a simple and easily understood measure of risk. VaR can be used to assess tail risk by measuring the maximum potential loss that could occur in a specific time period. This can be used to help managers make decisions about risk management strategies. VaR can also be used to compare the risk of multiple investments and to evaluate the effectiveness of risk management strategies. The simplicity of VaR eases these measures. Another advantage, especially relevant for banking regulatory purposes (Basel III), is the ease of backtesting for VaR.

However, as noted by Hull (2018. P. 273-274) and Allen, Singh and Powell (2013), VaR has some drawbacks in tail risk management. Firstly, VaR does not consider the potential for extreme events that occur with a very low probability of occurrence. This means that VaR may not capture the full extent of the potential risk from tail events. Secondly, VaR does not consider the potential for correlations between different investments and markets, which can lead to an underestimation of risk.

Expected Shortfall (ES) is a risk management metric that provides a more realistic view of portfolio risk than other measures, such as Value at Risk (VaR), as found by Hull (2018. P. 274) and by Allen, D., Singh, A., Powell, R. (2013). ES calculates the expected loss beyond the VaR level and estimates the potential for significant losses in extreme market conditions. By capturing the tail risk of a portfolio, ES offers a more accurate picture of the probability of losses in adverse market conditions.

While VaR focuses on the likelihood of a specific loss level over a given time horizon, ES considers the severity of losses that may occur beyond the VaR level. This makes ES a more robust measure of risk as it estimates the expected loss in the worst-case scenarios. Therefore, ES is essential for portfolio managers and investors who need to estimate potential losses and prepare for worst-case scenarios. The difference between the models can be observed from **Figure 7**.

However, calculating ES can be complex and time-consuming compared to VaR. The calculation of ES requires estimating the conditional expectation of portfolio losses beyond the VaR level, which often involves complex mathematical models and simulations. Additionally, the accuracy of ES estimates depends on the quality and availability of data used in the calculation, which can also add to the complexity of the calculation process.

Furthermore, backtesting ES is challenging compared to VaR. Backtesting involves evaluating the accuracy of the calculated risk metric by comparing the expected losses with the actual losses observed in the market. However, backtesting ES requires simulating and testing extreme market conditions, which can be challenging and often results in a limited amount of data for testing.

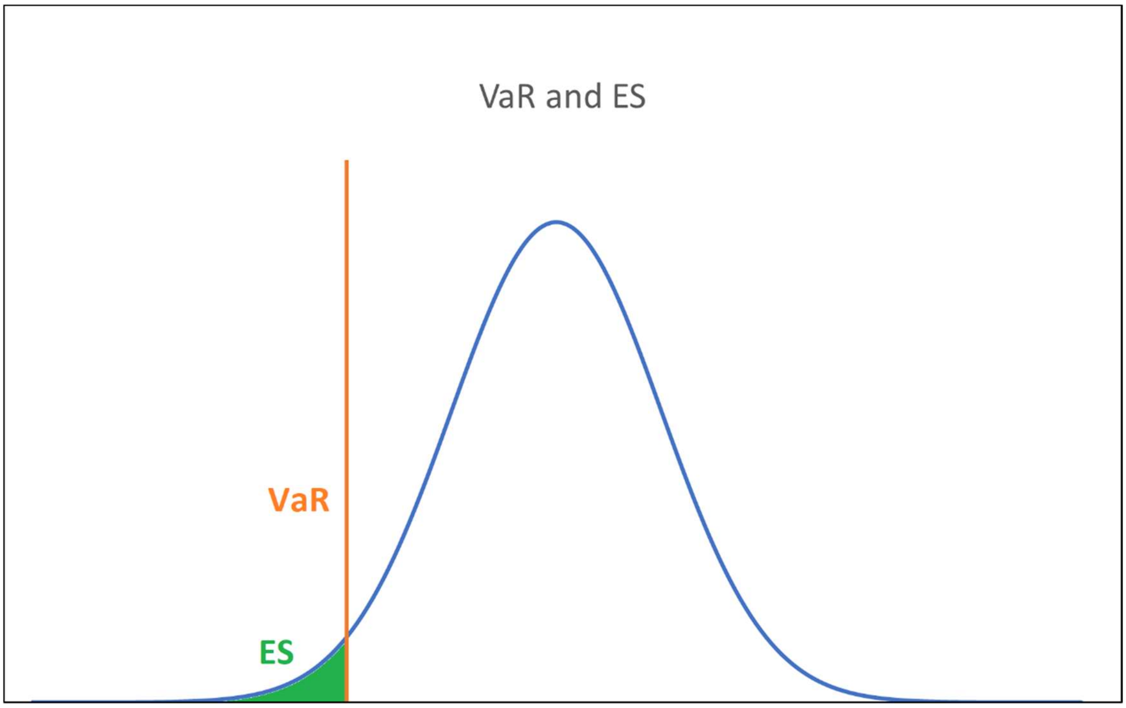

Figure 7 VaR and ES comparison

4 Extreme Value Theory in Finance

Extreme value theory (EVT) is a statistical methodology that examines the behaviour of rare events or extreme values. The theory was pioneered by Gumbel (1935) in addition to Tippett (1925), and Fisher and Tippett (1928), whose ideas led the way in extreme value statistics and distributions. Although EVT was first used for hydrological data and yarn breakage, EVT was later introduced to finance and financial tail risk management. EVT aims to identify and model the distribution of extreme events, such as stock market crashes or natural disasters, that may significantly impact financial markets. EVT is based on the theory of extreme values, a branch of probability theory that focuses on the distribution of a random variable's maximum or minimum values.

In finance, EVT has been utilized to estimate the tail risk of investments, which is the likelihood of extreme events occurring that may lead to significant losses. EVT provides a framework for estimating the probability of extreme events and developing strategies to manage the associated risk. EVT can be used to model the distribution of market returns, identify potential outliers, and estimate the probability of extreme events occurring in the future.

One of the critical applications of EVT in finance is the estimation of value at risk (VaR). EVT is used to estimate VaR by modelling the tail distribution of the investment returns, which is typically assumed to follow a fat-tailed distribution. The fat-tailed distribution reflects a higher probability of extreme events than a normal distribution, which assumes that extreme events are unlikely to occur.

4.1 Applying EVT to Financial Tail Risk Models

A common method for fitting a Generalized Pareto Distribution (GPD) to stock returns data is the Peak Over Threshold (POT) method. POT is a standard approach for utilizing EVT with financial data, which is also why it is also noted by McNeil, Frey and Embrechs, (2015. P .135-153). This paper will concentrate on the POT approach, which is widely utilized in VaR and ES methods. In POT, a threshold value u is decided, and values that violate this threshold or "peaks" are studied while applying the method, hence the name peak over the threshold. In the POT approach, extreme events are identified as those that exceed the threshold $u,$ which is typically set based on historical data or expert judgment. The POT approach is beneficial for modelling rare events that do not fit well into traditional statistical models. By focusing on the tail of the distribution, the POT approach can provide a more accurate estimate of the probability of extreme events, as argued by Allen, Singh and Powell (2013), which can be used to estimate risk measures such as VaR and ES.

The GPD is of the form:

$$G_{\xi,\beta}(Y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta} Y\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - e^{-\frac{Y}{\beta}} & if\ \xi = 0 \end{cases}$$

Where $\beta > 0, Y \geq 0$ when $\xi \geq 0$ and $-\frac{\beta}{\xi} \geq Y \geq 0$ if $\xi < 0$. A representation of a GPD can be seen in **Figure 8**. For a distribution function F_X of losses (returns) X, the excess distribution over the threshold u is given by:

$$F_u(Y) = P(X - u \leq Y | X > u) = \frac{F(Y + u) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)},$$

For $0 < Y < x_F - u$ where $x_F \leq \infty$ is the rightmost tail of the distribution and $Y = X - u$ and $F(u)$ is the conditional excess distribution function. The mean excess function of Y with finite mean is given by

$$e(u) = E(X - u | X > u).$$

For an extreme distribution with threshold u, we can assume that $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leq X < x_F - u$ and $\xi \in \mathbb{R}$ and $\beta > 0$. For $x \geq u$

$$\bar{F}(x) = P(X > u)P(X > x|X > u) = \bar{F}(u)\bar{F}_u(x-u) = \bar{F}(u)\left(1 + \xi\frac{x-u}{\beta}\right)^{-\frac{1}{\xi}},$$

Where given $F(u)$, is the formula for tail probabilities. The inverse of the prior is the POT Value-at-Risk, which is formulated as

$$VaR_\alpha = q_\alpha(F_X) = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right),$$

Where u is the threshold, β is the magnitude parameter for which $\beta > 0$, ξ is the shape parameter which is usually $0.4 \geq \xi \geq 0.1$ for financial data and $\bar{F}(u) = \frac{n_u}{n}$ where n_u is the number of samples where $R_t > u$. As stated by Allen, Singh and Powell (2015. P. 154-155), for the EVT VaR to work properly, $\alpha \geq \bar{F}(u)$ must hold true. The parameters β and ξ are estimated by maximizing log-likelihood. For the log-likelihood maximization, only the n_u samples where $X_t > u$ are considered. The log-likelihood function is as follows:

$$L(\beta, \xi, Y_i, \ldots, Y_{n_u}) = -n_u \ln(\beta) - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{n_u} \ln\left(1 + \xi\frac{Y_i}{\beta}\right),$$

Where $Y_i = X_t - u$. The log likelihood function is subject to constraints $\beta > 0$ and $1 + \frac{\xi Y_i}{\beta} > 0$.

According to McNeil, Frey and Embrechts (2015. P. 154-155), EVT can also be used for estimating ES with the assumption of $\xi < 1$. ES can be formulated as

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_x(F_X)\,dx = \frac{VaR_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi},$$

Where VaR_α is the VaR estimated by the previous EVT VaR formula.

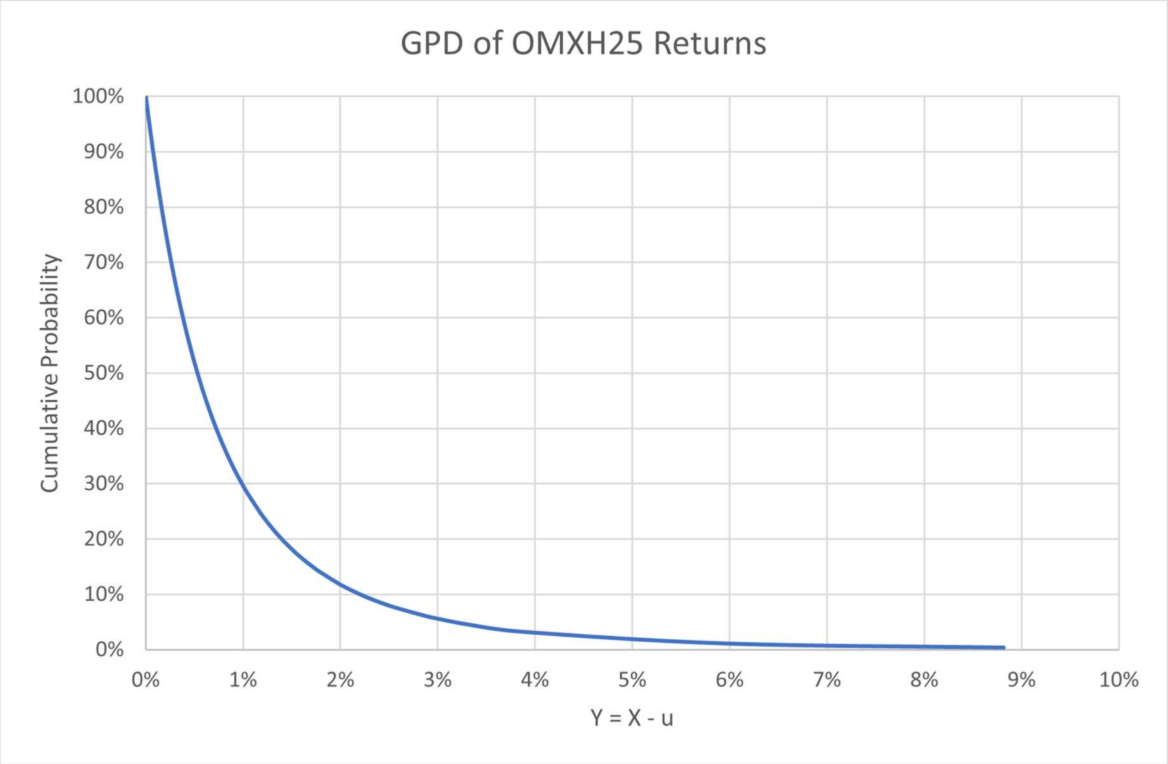

Figure 8 Generalized Pareto Distribution of OMXH25 Returns

An EVT Value-at-Risk can also be estimated non-parametrically without assuming no tail model with the Hill estimator (Hill. 1975; Nieto and Ruiz. 2016) given $\xi > 0$, yielding a VaR model:

$$VaR_\alpha = q_\alpha(F_X) = u\left(\frac{\alpha}{\bar{F}(u)}\right)^{-\xi},$$

Where $\xi = \frac{1}{n_u}\sum_{j=1}^{k}\left(\log(R_j) - \log(R_{j+1})\right)$. This model however is subject to small sample bias, which can be corrected with a linear regression estimation of ξ on Hill estimates of differing tail observations. Additionally, McNeil and Frey (2000) found that the Hill estimator tended to be more unstable than the ones yielded by the GPD approaches, thus making the GPD approach a more stable solution for modelling tails of a given distribution.

5 Literature Review of Extreme Value Theory in Tail Risk Forecasting

GPDs are a more flexible distribution that can capture the heavy-tailed nature of financial returns. The Peaks Over Threshold (POT) method is a popular EVT technique focusing on observations exceeding a particular threshold value. The GPD can be fit to these exceedances, and the resulting model can be used to estimate the tail risk measures such as VaR and ES.

In contrast, normal and t distributions assume that the returns are normally distributed or have a symmetric t-distribution. This assumption often results in underestimating the extreme tail risk. This can lead to significant financial losses, as the tails of the distribution can contain a large portion of the risk.

Using a violation-based backtesting method to evaluate the dynamic EVT VaR (VaR forecast), Allen, Singh and Powell (2013) tested both static and dynamic methods for predicting tail risk in market and volatility indices with the focus on the dynamic. In their paper, Allen, Singh and Powell (2013) apply POT in combination with GARCH (1,1) to forecast tail risk and apply the static methods shown before. This method applies POT to the distribution of residuals gained by first applying the GARCH model and a constant extreme value threshold u. This yields an estimator $VaR(Z)_\alpha$ $(ES(Z)_\alpha)$ that can be combined with the volatility and mean forecasting methods to gain

$$VaR_\alpha = -\mu_{t+1} + \sigma_{t+1} VaR(Z)_\alpha.$$

They found that the dynamic EVT-based model performs overall better than other compared models of a normally distributed GARCH (1,1) VaR and RiskMetrics, especially with 1% VaR for the FTSE 100 and 5% for the S&P 500. The other methods are rejected by their test. However, they interestingly found that none of the methods worked for the volatility indices. They find this is due to extreme variation in values of both volatility indices, leading to overestimation in the VaR model. The performance of the Expected Shortfall models was not tested as as mentioned before, backtesting them is difficult. All in all, they found that EVT models tend to be better than the alternatives.

In their paper, Zikovic and Aktan (2011), researched the performance of different VaR and ES models when used on market index and commodity data with a focus on the performance of BRW simulation models. The BRW models with an optimal decay factor λ_{Opt} for which $\left|C_t\left(\lambda_{Opt}\right)\right| = \min\left|C_t(\lambda)\right|$ are as follows:

$$C_t(\lambda) = \begin{cases} 1 + \left(L_t - VaR_{t,\alpha}\right)^2 & if\ L_t > VaR_{t,\alpha} \\ 0 & if\ L_t \leq VaR_{t,\alpha} \end{cases},$$

$$VaR_{t,\alpha} = \sum_{j=t-N}^{t} r_j I\left(\sum_{i=1}^{N} f_i(\lambda; N) I\left(r_{t-i} \leq r_j\right) = \alpha\right).$$

Their backtesting results showed that widely used VaR methods like historical simulation and RiskMetrics, but found that conditional and unconditional EVT GARCH approach tended to outperform. An additional generalized Pareto distribution test also showed performance but failed the backtest for the NIKKEI, FTSE and CAC indexes. They however also found that the BRW models dependent on the decay factor optimization, tended to outperform the other models even the EVT GARCH model in some cases. Thus, the research by Zikovic and Aktan (2011) again showed outperformance of the EVT-based models, especially when compared with GARCH models, but also found that BRW models could prove an alternative for EVT.

A paper by Gencay, Selcuk and Ulugulyagci (2003) also researched the performance of EVT-based models compared to alternatives on a dataset of Turkish stock returns and S&P 500 stock returns. The research tested both an adaptive (rolling window) and a nonadaptive (static) GPD model, finding that in the case of the Turkish data, EVT-based models generally outperformed the others on the left tail of the distribution at 97.5th quantile or greater. When tested on the S&P 500 data, they found that EVT-based methods tended to perform the best with the adaptive GPD being one of the best for almost all quantiles from the 95th to the 99.9th and the static GPD starting to outperform at 97.5th and greater quantiles. The research by Gencay, Selcuk and Ulugulyagci (2003) once again showed EVT generally outperforming the other with the Variance-Covariance or the traditional parametric VaR method generally performing the worst.

Research on EVT-based models was also performed by Danielsson and Moritomo (2000) on data from TOPIX, S&P 500, JPY/USD, WTI and TSE2. They compared a normally distributed model, a normal GARCH model Studen-t GARCH model and an EVT model, finding EVT to produce stable and accurate forecasts while finding that GARCH-based models have lesser accuracy and greater instability. They however note that models with longer memory would be suited for GARCH, while EVT is preferred in other cases.

An analysis of the effectiveness of differing Value-at-Risk measures was conducted by Bhattacharya and Ritolia (2006). With data on Indian stocks and comparing a static historical model with static and dynamic (GARCH) normal and EVT models, they found that only the EVT models succeeded in the backtest with the dynamic EVT performing the best, giving more support to the hypothesis of this thesis.

Similar results were found by Bilandi and Kudla (2016), who researched the performance of EVT-based models compared to historical simulation and parametric methods like EWMA and differing GARCH models. Their results show that EVT models outperformed while historical simulations performed better than the non-EVT models. They however note that while the EVT models outperformed, their harder implementation may make them less accessible for use.

In their research, Tong, Diao and Li (2024) compared GARCH-EVT hybrid VaR models with normal and nonnormal filters to parametric alternatives on a data set of Chinese stock returns. Their research found that hybrid EVT models with heavy-tailed Student-t distribution-based GARCH filters generally outperformed others. Additionally, they found that normally and skewed normally distributed models underperformed, while Student-t and skewed Student-t models had greater performance than the normally distributed parametric models but underperformed the EVT models. Their findings provide additional support for both the hypothesis and a method of improving the aforementioned EVT models with heavy-tailed GARCH filters.

Research by Exhaust and Just (2020) on GARCH-EVT VaR models with differing thresholds. Their research, while not comparing EVT to different models, found that the GARCH-EVT model performed well in all thresholds and that the choice of the tail threshold did not improve the accuracy of the models. This implies that while implementing EVT-based methods may be difficult, the choice in threshold parameters is easier and there may be no need to optimize beyond a standard 95th quantile threshold.

Aridi, Hooi and Cheong (2024) researched the comparative performance between low- and high-frequency conditional EVT models. Their research tested EVT models with GARCH, realized variance and HAR approaches on a date from the COVID-19 period. They found that HAR-EVT models outperformed the other EVT VaR models with HAR-EGARCH-EVT models showing superior performance. They additionally tested the performance of the EVT ES models using a traffic light backtest, which provided results in line with the VaR backtests for the pre-COVID-19-crisis period, but lower performance during the crisis period compared to other EVT-GARCH models. Their research combines EVT with the heterogeneous market hypothesis (HMH) which states that investors follow the nonlinear behaviour of heterogeneous market participants in places where one price does not reflect every investor's information even with all investors having access to similar information, evolving the research and methodology of EVT models.

The previous research discussed has shown that EVT-based models for VaR and ES have outperformed the competing models. The performance of these models over the normally distributed ones gives positive support for H_1 , and their general outperformance against other complex models shows support for H_2, but with a caveat of their being other models like the optimized BRW that can stand on par with the EVT models. A significant factor is their tendency to outperform the most widely used methods of VaR and ES. Despite the fluctuating nature of the best tail risk models in academia, a general move towards EVT-based tail risk methods should be advantageous for future risk management.

6 Conclusion

This thesis reviewed the standard methods of modelling tail risk with VaR and ES and how these can be enhanced with EVT-based approaches (mainly POT). The history, theoretical background, how the models are applied, and their advantages and disadvantages are discussed, finding that different methods fit different situations.

From previous research, we gain insight into the performance of EVT-based models compared to the standard approaches and to some more sophisticated approaches. The previous research shows that EVT-based models tend to perform better than the simpler standard models and better than or on par with more advanced methods, thus showing that risk management of several institutions and other actors who still rely on the tail risk methods based on normal distributions and other simpler approaches like historical simulation could be enhanced with the utilization of EVT -based methods. The advantage of EVT comes from its approach of treating the tails of a given distribution as separate from the rest of the distribution, which when applied, yields better estimates for tail risk and a better grounding for hedging this risk.

References

Allen, D., Singh, A., Powell, R. (2013). EVT and tail-risk modelling: Evidence from market indices and volatility series. The North American Journal of Economics and Finance. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1062940813000259

Aridi, N., Hooi, T. and Cheong, C. (2024). A comparative VaR analysis between low-frequency and high-frequency conditional EVT models during COVID-19 crisis. Cogent Economics & Finance. Retrieved 4.3.2025 from: https://www.tandfonline.com/doi/full/10.1080/23322039.2024.2377495#abstract

Bachelier, L. (1900). Théorie de la speculation. Annales scientifiques de l'École normale supérieure. Retrieved 4.3.2025 from: http://www.numdam.org/item/10.24033/asens.476.pdf

Basel Committee on Banking Supervision. (2009). Enhancements to the Basel II framework. BIS. Retrieved 4.3.2025 from: https://www.bis.org/publ/bcbs157.pdf

Bhattacharyaa, M. and Ritolia, G. (2006). Conditional VaR using EVT – Towards a planned margin scheme. International Review of Financial Analysis. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1057521906000688#sec7

Bilandi, M. and Kudla, J. (2016). Comparing the Precision of Different Methods of Estimating VaR with a Focus on EVT. Oxford Journal of Finance and Risk Perspectives. Retrieved 4.3.2025 from: https://www.researchgate.net/profile/Majid-Bilandi/publication/311696017_COMPARING_THE_PRECISION_OF_DIFFERENT_

METHODS_OF_ESTIMATING_VAR_WITH_A_FOCUS_ON_EVT_COMPARING_THE_PRECISION_OF_DIFFERENT_METHODS_OF_ESTIMATING_VAR_WITH_A_FOCUS_ON_EVT_112/links/58552b5a08ae8f695555ee72/COMPARING-THE-PRECISION-OF-DIFFERENT-METHODS-OF-ESTIMATING-VAR-WITH-A-FOCUS-ON-EVT-COMPARING-THE-PRECISION-OF-DIFFERENT-METHODS-OF-ESTIMATING-VAR-WITH-A-FOCUS-ON-EVT-112.pdf

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/abs/pii/0304407686900631

Danielsson, J. and Moritomo, J. (2000). Forecasting extreme financial risk: a critical analysis of practical methods for the Japanese market. Monetary Economic Studies. Retrieved 4.3.2025 from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=531e7029c4f3d74ef8a3b8df4a04108c823ac695

Echaust, K. and Just, M. (2020). Value at Risk Estimation Using the GARCH-EVT Approach with Optimal Tail Selection. Mathematics. Retrieved 4.3.2025 from: https://www.mdpi.com/2227-7390/8/1/114

Fisher, R. A., Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society. Retrieved 4.3.2025 from: https://scispace.com/pdf/limiting-forms-of-the-frequency-distribution-of-the-largest-5ecr8qpsvp.pdf

Gencay, R., Selcuk, F. and Ulugulyagci, A. (2003). High volatility, thick tails and extreme value theory in value-at-risk estimation. Insurance: Mathematics and Economics.

Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0167668703001677#SEC3

Gumbel, E. (1935). Les valeurs extrêmes des distributions statistiques. Annales de l'I. H. P. Retrieved 4.3.2025 from: http://www.numdam.org/item/AIHP_1935__5_2_115_0.pdf.

Hill, B. (1975). A Simple General Approach to Inference About the Tail of a Distribution. The Annals of Statistics. Retrieved 4.3.2025 from: https://www-jstor-org.proxy.uwasa.fi/stable/2958370?seq=1

Hull, J. (2018). Risk Management and Financial Institutions. Wiley Finance Series.

Kupiec, H. (1995). Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives. Retrieved 4.3.2025 from: https://web.archive.org/web/20170809061153id_/http://www.systemic-risk-hub.org/papers/bibliography/Kupiec_1995_Preview.pdf

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance. Retrieved 31.3.2023 from: https://www-jstor-org.proxy.uwasa.fi/stable/2975974?origin=crossref

McNeil, A., Frey, R., Embrechs, P. (2015). Quantitative Risk Management Concepts, Techniques and Tools. Princeton Series in Finance.

McNeil, A., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0927539800000128

Nieto, M., Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. International Journal of Forecasting. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S016920701500120X

Richardson, M., Smith, T. (1993). A Test for Multivariate Normality in Stock Returns. The Journal of Business. Retrieved 4.3.2025 from: https://web-s-ebscohost-com.proxy.uwasa.fi/ehost/pdfviewer/pdfviewer?vid=0&sid=1cacd7c2-9088-468c-96d4-ef20786e1284%40redis

Tippett, L. H. C. (1925). On the Extreme Individuals and the Range of Samples Taken from a Normal Population. Biometrika. Retrieved 4.3.2025 from: https://www-jstor-org.proxy.uwasa.fi/stable/2332087?seq=1

Tong, B., Diao, X., Li, X. (2024). Forecasting VaRs via hybrid EVT with normal and non-normal filters: A comparative analysis from the Chinese stock market. Pacific-Basin Finance Journal. Retrieved 4.3.2025 from: https://www.sciencedirect.com/science/article/abs/pii/S0927538X24000222

Zikovic and Aktan (2011). Decay factor optimisation in time weighted simulation — Evaluating VaR performance. International Journal of Forecasting. Retrieved 4.3.2025 from: https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0169207011000045