UNIVERSITY OF TURKU
Turku School of Economics

Enhancing Value at Risk Models Using Extreme Value Theory in Extreme Market Conditions

Empirical Evidence from Finland during COVID-19 crisis

Bachelor's thesis
in Accounting and Finance

Author:
Miro Ikäheimonen

Supervisor:
D.Sc. Oana Apostol

13/05/2025
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Bachelor's thesis

Subject: Accounting and Finance
Author: Miro Ikäheimonen
Title: Enhancing Value at Risk Models Using Extreme Value Theory in Extreme Market Conditions
Supervisor: D.Sc. Oana Apostol
Number of pages: 45
Date: 13/05/2025

Periods of extreme market volatility, such as the COVID-19 crisis, challenge the reliability of traditional risk management tools in the financial sector. Sudden and severe downturns reveal the limitations of widely used models, raising questions about how well financial institutions can anticipate and prepare for such events.

One commonly used tool in risk management has been Value at Risk (VaR), which estimates the potential loss in value of a portfolio over a given period. However, standard VaR models, particularly those based on the assumption of normally distributed returns, often underestimate the likelihood and impact of rare, extreme events. Although recent regulations, such as Basel III, have shifted focus toward more conservative measures like Expected Shortfall, VaR remains a widely recognized and applied metric. Improving its reliability during turbulent times is therefore both a practical and academic concern.

This thesis explores how Extreme Value Theory (EVT) can enhance the performance of VaR during times of financial stress, using the COVID-19 market crash in Finland as a case study. EVT is specifically designed to model rare and extreme outcomes, offering tools that may better capture the risks traditional models overlook.

The aim of the thesis is to examine whether EVT can improve the accuracy of VaR in extreme market conditions. The analysis compares different EVT approaches, such as the Peaks-Over-Threshold (POT) and Block Maxima methods, and evaluates how their results differ from conventional VaR calculations. The backtesting results indicate that EVT-based VaR, particularly the POT method, outperforms traditional VaR approaches during the COVID-19 crisis in Finland. Among the traditional models, the Historical Simulation VaR demonstrated the best performance. The findings contribute to the broader discussion on risk modeling by providing insight into the strengths and weaknesses of both traditional and EVT-based methods in periods of market distress.

Key words: Risk Management, Value at Risk, Extreme Value Theory

Kanditutkielma

Oppiaine: Laskentatoimi ja rahoitus
Tekijä: Miro Ikäheimonen
Otsikko: Ääriarvoteorian hyödyntäminen Value-at-Risk -mallien määrittämisessä
Ohjaaja: KTT Oana Apostol
Sivumäärä: 45
Date: 13/05/2025

Äärimmäiset osakemarkkinoiden heilahtelujaksot, kuten COVID-19-pandemiaan liittynyt kriisi, haastavat perinteisten riskienhallintamallien luotettavuuden rahoitussektorilla. Äkilliset ja voimakkaat markkinaliikkeet paljastavat laajasti käytettyjen riskienhallinnassa käytettyjen mallien rajoituksia ja herättävät kysymyksiä niiden kyvystä ennakoida ja hallita harvinaisia, mutta suuria riskejä.

Yksi yleisimmin käytetyistä riskienhallinnan malleita on Value at Risk (VaR), joka arvioi sijoitussalkun mahdollisen arvonmenetyksen tietyllä todennäköisyystasolla ja aikavälillä. Perinteiset VaR-mallit, erityisesti normaalijakaumaan perustuvat menetelmät, aliarvioivat usein äärimmäisten tapahtumien todennäköisyyksiä ja vaikutuksia. Vaikka uudempi sääntely, kuten Basel III -kehikko, on siirtänyt painopistettä kohti varovaisempia menetelmiä, kuten odotettavissa olevaa tappiota (Expected Shortfall), VaR säilyttää asemansa keskeisenä ja laajasti sovellettuna riskimittarina. Sen luotettavuuden parantaminen kriisiaikoina onkin sekä käytännöllisesti että akateemisesti merkittävä tavoite.

Tämä tutkielma tarkastelee, kuinka ääriarvoteorian (Extreme Value Theoryn, EVT) soveltaminen voi parantaa VaR-mallien tarkkuutta poikkeuksellisissa markkinaolosuhteissa. Ääriarvoteoria tarjoaa työkaluja harvinaisten ja äärimmäisten arvojen mallintamiseen, mikä voi auttaa tunnistamaan riskejä, jotka jäävät perinteisten mallien ulottumattomiin.

Tutkielman tavoitteena on selvittää, voidaanko ääriarvoteorian avulla parantaa VaR-mallien suorituskykyä äärimmäisissä markkinatilanteissa. Analyysissa vertaillaan ääriarvoteorian eri lähestymistapoja, kuten Peaks-Over-Threshold (POT) ja Block Maxima -menetelmiä, sekä arvioidaan niiden antamia tuloksia suhteessa perinteisiin VaR-laskelmiin. Testitulokset osoittavat, että ääriarvoteoria-pohjaiset VaR-mallit, erityisesti POT-menetelmä, suoriutuivat paremmin kuin perinteiset VaR-mallit COVID-19-kriisin aikana Suomessa. Perinteisistä malleista historialliseen dataan perustuva VaR tuotti parhaan tuloksen backtestauksessa. Tutkimus tarjoaa näkökulmaa riskimallinnuksen kehittämiseen suomalaisilla rahoitusmarkkinoilla COVID-19-kriisin aikana.

Avainsanat: Riskienhallinta, Value at Risk, Ääriarvoteoria

CONTENTS

FIGURES

TABLES

1 Introduction

Risk management plays a central role in financial decision-making by helping institutions anticipate and prepare for potential losses. In particular, its importance becomes especially apparent during periods of heightened uncertainty and market turbulence. (McNeil, Frey and Embrechts, 2015, 7-16.) To manage risk, financial institutions rely on quantitative models to estimate potential losses. Among these tools, Value-at-Risk (VaR) has become one of the most commonly used metrics (McNeil et al., 2015, 64; Hull, 2018, 271). However, many researchers (e.g. Jorion (2006) and Ben Ayed and Ben Hassen (2024)) have pointed out that traditional VaR models tend to underestimate the likelihood of rare extreme losses, precisely the events that pose the greatest threat during financial crises.

The COVID-19 pandemic, a global health crisis that began in early 2020, triggered an unprecedented shock to global financial markets. Stock prices plummeted, volatility surged, and risk management frameworks were put to the test (Ashraf, 2020). One common reason for this underestimation lies in a core assumption behind many traditional VaR models: that asset returns follow a normal distribution (Alexander, 2009, 138-139). The normal distribution is a symmetric, bell-shaped curve where most values cluster around the mean, and the probability of extreme outcomes decreases rapidly (Williams, n.d.). It is widely used in finance but often underestimates rare, extreme events. While this assumption simplifies modeling and is appropriate in calm markets, it fails to represent the reality of financial crises, where extreme losses occur more frequently than predicted by the normal model. This raises an important question: can alternative risk estimation techniques better account for extreme market conditions?

One approach that has been suggested to improve risk estimates during extreme market conditions is Extreme Value Theory (EVT). EVT is a statistical method that focuses specifically on rare and unusually large losses, those events that traditional models often fail to capture. Instead of assuming that market movements follow typical patterns, EVT looks at the most extreme changes, which tend to happen during financial crises (Longin, 2000). By focusing on the tails of the distribution, EVT can complement VaR by offering more accurate estimates of potential losses under extreme stress.

EVT has shown promising results in previous studies, such as Longin (2000) and McNeil and Frey (2000), during major international crises. However, its application to the Finnish stock market, particularly during COVID-19-induced volatility, has not yet been a subject of research. This thesis contributes to the existing literature by evaluating how EVT-based VaR models perform compared to traditional methods, using data from an exceptionally turbulent period in Finnish financial markets.

This thesis investigates whether EVT-based VaR models provide a more effective risk management tool compared to traditional VaR methods. Unlike conventional approaches, EVT is designed to model the behavior of extreme losses by focusing specifically on the tails of the distribution. By applying EVT's two primary methods, the Block Maxima model and the Peaks-Over-Threshold (POT) model, this study aims to determine whether these techniques yield more accurate risk estimates under extreme financial conditions.

1.1 Research Question

This study aims to answer the following research question:

- Can Extreme Value Theory (EVT) improve the accuracy of Value-at-Risk (VaR) models in extreme market conditions?

By addressing this question, the study seeks to contribute to the ongoing discussion about improving risk estimation in financial markets. The study adopts an empirical approach that compares the performance of traditional and EVT-based VaR models under conditions of market stress. The analysis is grounded in real market data and focuses on model evaluation through statistical backtesting techniques.

The empirical context of the study is the Finnish stock market, with a focus on the period of the COVID-19 crisis. The analysis uses OMXHPI index data spanning from 2005 to 2020. The data, sourced from Refinitiv Workspace, is divided into two segments: the estimation period (from 1st of August 2005 to 31st of January 2020), which is used to fit the models and estimate parameters, and the backtesting period (February 3rd 2020 – July 31st 2020), which evaluates the models' performance during the COVID-19 financial crisis. The performance of EVT-based VaR models will be assessed by comparing them with traditional VaR methodologies, applying two different backtesting techniques to evaluate their predictive accuracy.

1.2 Structure

This thesis is organized as follows. Section 2 introduces the theoretical foundations of market risk, Value-at-Risk (VaR), and Extreme Value Theory (EVT). It outlines three VaR estimation methods: normal linear, historical, and Monte Carlo simulation. It also presents the EVT framework through its two main approaches: the Block Maxima and Peaks-Over-Threshold models. Section 3 describes the data processing steps and the methodology for implementing both traditional and EVT-based VaR models. It also explains the backtesting techniques used to assess model accuracy. Section 4 introduces the data and presents the empirical results. It describes the Finnish stock market dataset used in the study and compares the performance of traditional and EVT-based VaR models during the COVID-19 crisis. The final section summarizes the main findings, compares performance of traditional VaR methods to EVT-based VaR, and suggests directions for future research. This structure supports the overall goal of evaluating whether EVT improves risk estimation under extreme market conditions.

2 Theory

This section provides the theoretical framework for the thesis by introducing key concepts related to market risk, Value at Risk (VaR), and Extreme Value Theory (EVT). The section begins by defining market risk and outlining its significance in financial decision-making. It then discusses the fundamentals of VaR, including its historical development and various calculation methods: normal linear, historical simulation, and Monte Carlo simulation. After that, the section introduces EVT as an alternative approach to improve risk estimation in extreme market conditions and explains its two main modeling techniques: the Block Maxima and Peaks-Over-Threshold methods. Finally, the section introduces previous studies on the performance of EVT-based VaR compared to other VaR calculation methods.

2.1 Market Risk and Its Measurement

Financial risk is often divided into five main types of risk, which are market risk, credit risk, liquidity risk, legal risk and operational risk (Lai and Xing, 2008, 306). In this thesis, I primarily focus on market risk, as it is most closely associated with the subject of this thesis. Market risk occurs when prices of different financial assets, such as stocks, bonds, or derivatives change as new information becomes available to the market or supply and demand changes. For most financial instruments, price changes are usually relatively stable under normal conditions, with sudden large movements happening only occasionally. Because of this, market risk models often assume a continuous distribution, such as normal distribution or exponential distribution. (Miller, 2019, 6.) A continuous distribution describes the likelihood of different values for a variable that can take on any value within a certain range. We find the chance of the variable being between two numbers by looking at the area under a curve. The total area under the curve is always 1. (Devore, 2012, 138-139). Since many financial instruments have extensive historical data, this data can be used to assess and measure market risk (Miller, 2019, 6).

In classical finance theory such as Markowitz (1952), risk is defined with standard deviation of returns or, in other words, volatility. Today, various methods have been developed to measure market risk, such as VaR and Expected Shortfall, as well as statistical measures like skewness and kurtosis, which describe the shape of return distributions. Skewness describes the asymmetry of a probability distribution: a negative value suggests a longer tail on the left side, a positive value points to a longer right tail, and a value near zero implies a roughly symmetric shape (Freeman, Anderson, Sweeney, Williams and Shoesmith, 2017, 60-61). Kurtosis is a statistical measure that describes the shape of a probability distribution by quantifying the relative heaviness of its tails and sharpness of its peak in comparison to a normal distribution (DeCarlo, 1997).

Still, many risk managers also use standard deviation along side with these newer methods. Standard deviation also plays a crucial role in calculating these methods, such as VaR, which is discussed more broadly later in this thesis (Miller, 2019). Despite the wide array of available risk measures, VaR remains especially interesting due to its intuitive interpretation and widespread adoption in both regulatory frameworks and practical risk management (McNeil et al., 2015, 64; Sarykalin, Serraino and Uryasev, 2008). Unlike purely statistical metrics, VaR provides a single, quantifiable estimate of potential loss at a given confidence level, making it highly actionable for decision-making and capital allocation (McNeil et al., 2015, 23).

2.2 Value at Risk

This section provides an overview of the Value at Risk (VaR) measure and explores its most common calculation methods. It begins with a general introduction to the concept and purpose of VaR, followed by a closer look at three widely used approaches: the Normal Linear VaR, Historical VaR, and Monte Carlo Simulated VaR.

2.2.1 General Overview of the Value at Risk Measure

Value at Risk is one of the most commonly used risk metric in financial institutions today (McNeil et al., 2015, 64). It originates from J.P. Morgan's RiskMetrics system from 1980's where the bank's chairman requested a daily one-page risk report covering the bank's entire trading portfolio from his staff. In 1994, J.P. Morgan made RiskMetrics public, providing free data online, which spurred improvements and widespread adoption of VaR by financial institutions. (Lai and Xing, 2008, 307.) In 1996, the Basel Committee introduced an amendment to Basel I regulation to set capital requirements for market risks. In the amendment, Value at Risk was regulated as a quantitative standard for calculating market risk on a daily basis (Basel Committee on Banking Supervision, 1996, 44).

VaR provided a single, interpretable number that quantified risk across different asset classes, making it useful for risk managers, regulators, and executives. However, after the 2008 financial crisis, VaR as the primary market risk management method began to be criticized, and the Basel 2.5 framework introduced in 2009 supplemented VaR with additional measures: Stressed VaR, which uses stressed historical data [1]; the Incremental Risk Charge, addressing default and credit migration risks; and capital requirements for

[1] Stressed historical data refers to market data from periods of financial stress, such as the 2008 crisis, used to capture extreme but plausible market conditions.

Securitizations, aimed at better reflecting the risks associated with complex structured products. (McNeil et al., 2015, 23-24.)

Value at Risk introduces the following statement: "We are X percent confident that our losses will not exceed V dollars over time T, with a $(1 - X)\%$ probability that losses will exceed V." Here, the value V represents the VaR of the portfolio or instrument. V is determined as a function based on other two parameters in the statement: the time horizon (T) and the confidence level (X). In other words, VaR is the loss level during a time period T that we are X% certain will not exceed. (Hull, 2018, 271.) McNeil et al. (2015) defines VaR using the following formula:

$$\mathrm{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.$$

Here $\alpha \in (0, 1)$ is the confidence level, L represents portfolio loss, l is the loss threshold, and $F_L(l)$ is the cumulative distribution of L, meaning the probability that the loss is at most l. Also here inf refers to the greatest lower bound of a set, which is, the largest value that is less than or equal to all elements satisfying the given condition. VaR is the smallest loss l where the chance of a larger loss is at most $1 - \alpha$. Typical values for α are $\alpha = 0.95$ or $\alpha = 0.99$, and in market risk management, the time horizon is usually one or ten days. (McNeil et al., 2015, 64; Alexander, 2009, 14-15.) Figure 1 visualizes the concept of calculating VaR from the probability distribution of portfolio returns. Since losses are just negative gains, the VaR level (-V) represents the worst expected loss at a given confidence level (X%). (Hull, 2018, 271.)

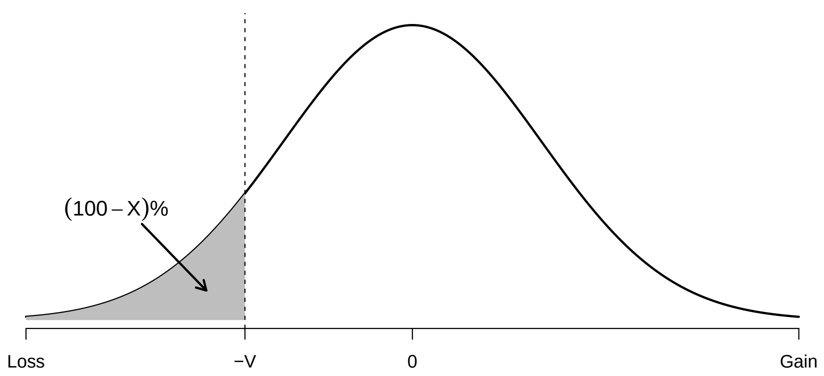

Figure 1: VaR from the probability distribution of portfolio returns

There are several methods to calculate VaR, each with its own approach and assumptions. Choosing the method depends on the structure and needs of the portfolio or instrument. The potential for losses results from exposure to various risk factors and the distribution of these risk factors. This distinction is reflected in risk management systems, which can be divided into models for exposure and models for the distributions of risk factors. (Jorion, 2006, 247-249.) In this context, a risk factor refers to any variable that influences the value of a financial asset or portfolio and contributes to overall risk, such as stock market index returns, interest rates, or company-specific stock returns (Alexander, 2009, 41-42). Risk factor distribution is usually divided into parametric and non-parametric approaches. In some special cases combination of these two can also be used. Parametric methods involve estimating the distribution of returns and using this estimated distribution to make risk forecasts. Most commonly, the normal distribution or the Student t-distribution is used, but other distributions, such as the Pareto-positive stable distribution, can also be applied. (Danielsson, 2011, 93.) Another example of a parametric approach is Monte Carlo simulation. However, Monte Carlo simulation can also be nonparametric, by generating a large number of hypothetical return scenarios based on historical data or other assumptions about asset price movements. This depends on whether it assumes a predefined return distribution or purely relies on historical data to generate scenarios (Alexander, 2009). There are also several other parametric methods, and for example Nadarajah, Chan and Longin (2016) present 35 different parametric methods or distributions to calculate VaR.

In contrast, non-parametric risk forecasting does not use assumptions about the underlying return distribution but instead relies on historical data to estimate risk. The most common nonparametric approach is historical simulation, which uses the empirical distribution of past returns to assess future risk. (Danielsson, 2011, 93.)

Although risk factor distributions are crucial in VaR estimation, as mentioned earlier, another key component of risk assessment is how exposure is measured. This can be done using different valuation methods that determine how changes in asset prices affect the overall portfolio. Exposure can be divided into Local-Valuation Methods and Full-Valuation Methods. Local-Valuation Method assumes that risk factors, such as stock returns or interest rates, follow a normal distribution and that portfolio risk can be approximated using linear (delta) exposure to these factors. Full-Valuation Method re-prices the portfolio across multiple scenarios, directly simulating how asset prices respond under different market conditions. Unlike the Delta-Normal method, it does not rely on linear approximations and can accommodate non-linear relationships. (Jorion, 2006, 247-254.)

Different VaR methods are often classified based on how they model risk exposure and risk factor distributions (Jorion, 2006, 247-254). In this thesis, I focus on three approaches to risk factor modeling: the normal distribution (parametric), historical simulation (non-parametric), and Monte Carlo simulation (which can be parametric or non-parametric, depending on implementation).

2.2.2 Normal Linear VaR

The Normal Linear Value at Risk (VaR) is based on the assumption that the portfolio returns are normally distributed and that changes in the portfolio value can be calculated linearly based on its risk factors. In practice, this means that the risk factor returns of the portfolio and their dependencies can be represented using a covariance matrix. This provides a mathematical approach to describing how the returns of different assets move with each other. The model is also referred to as covariance VaR, delta-normal VaR, or parametric VaR in the literature. (Alexander, 2009, 41-42.) The detailed calculation of Normal Linear VaR, along with the relevant formulas, is discussed in Section 3.2

One of the main strengths of the Normal Linear VaR model is its simplicity and computational efficiency. Because it relies on straightforward mathematical formula, it is well-suited for large portfolios and daily risk reporting. The model also offers intuitive insights into how volatility and correlations affect portfolio risk. However, the key assumptions of normality and linearity limit its accuracy in real-world scenarios. In particular, the model tends to underestimate extreme losses, as it does not account for fat tails or the nonlinear behavior of instruments such as options. (Jorion, 2006, 127–137.)

2.2.3 Historical VaR

Historical VaR uses historical changes in market and instrument prices to formulate a distribution of potential future portfolio profits and losses. From this, the VaR is calculated as the threshold loss that is exceeded only in $1 - \alpha$ proportion of cases, where α is the confidence level. (Linsmeier and Pearson, 2000, 50.) Historical VaR is a nonparametric method that does not rely on predefined assumptions about the distribution of risk factors, which is one of its main advantages. Unlike the Normal Linear VaR model, it does not impose parametric constraints, allowing a more flexible and empirically driven approach to risk estimation. (Alexander, 2009, 141-142.)

The distribution of these profits and losses is formulated by taking the current portfolio and applying historical changes in market factors over the last N days. In other words, N sets of hypothetical market factors are generated, with their values based on current market prices and past changes. (Linsmeier and Pearson, 2000, 50.) Historical VaR is therefore based on the assumption that the distribution continues to hold within the prediction period (Tsay, 2010, 338). In other words, historical VaR assumes that market behavior remains the same in the future as in the historical period used in the calculations. This can be problematic in periods of extreme volatility and market crashes, such as the financial crisis or the COVID-19 crisis.

2.2.4 Monte Carlo simulated VaR

Monte Carlo simulation is highly flexible technique used to solve many different financial problems (Alexander, 2009, 201). Monte Carlo simulation is a parametric method used to estimate risk in VaR calculations. It works by simulating random movements in risk factors based on assumed probability distributions, generating future values using stochastic processes instead of relying directly on historical data, as done in historical simulation. This method models changes in risk factors from estimated parametric distributions, and positions are re-priced using full valuation across multiple simulated scenarios to estimate potential portfolio outcomes. (Jorion, 2006; Alexander, 2009). A pseudo-random number generator is employed to generate thousands or even hundreds of thousands of hypothetical changes in the risk factors. These hypothetical changes are then applied to the current portfolio to estimate potential profits and losses across numerous simulated scenarios. The resulting distribution of possible portfolio outcomes is used to calculate the VaR. (Linsmeier and Pearson, 2000, 56.) This method is particularly useful for non-linear portfolios and derivative instruments, where traditional VaR models do not provide accurate estimates (Alexander, 2009, 44-45).

A key advantage of Monte Carlo VaR is its ability to capture volatility variations, heavy tails, and extreme market conditions, which the Normal Linear VaR models often do not account for (Jorion, 2006, 266-267). This makes it particularly useful for non-linear portfolios and derivative instruments. However, the main challenges of the method is its high computational cost and model risk, as the accuracy of the simulations depends on the quality of the stochastic process and a sufficiently large number of iterations (Alexander, 2009; Jorion, 2006).

2.3 Extreme Value Theory

This section provides an overview of Extreme Value Theory (EVT) and its application in financial risk modeling. EVT is particularly useful for assessing the risk of rare but severe market events that fall outside the scope of standard distribution assumptions. The section begins with a general overview of EVT, followed by a more detailed examination of its two primary approaches: the Block-Maxima model and the Peaks-Over-Threshold (POT) model.

2.3.1 General Overview of the Extreme Value Theory

In financial markets, extreme price changes in the market can lead to significant losses for investors and financial institutions. As discussed in the previous section, traditional risk measures, such as VaR, are often based on the normal distribution, which assumes that the majority of returns are concentrated around average fluctuations and that extreme losses are very rare. Extraordinary events, like the 2008 financial crisis or the COVID-19 pandemic in 2020, play a central role in finance, particularly in risk management and financial regulation (Longin, 2000). Extreme Value Theory (EVT) establishes limit laws for extreme values in large datasets. It provides key results that describe the behavior of sample maxima and minima, upper-order statistics (for example the kth largest value in a sample), and values that exceed high thresholds. (McNeil et al., 2015, 135.) EVT is a general statistical framework developed independently of financial applications and is widely used in areas such as hydrology, insurance, and environmental science (Coles, 2001, 1-2). In finance, its main relevance lies in modeling the tail of the distribution of financial risk-factor changes, where traditional models often fail (Longin, 2000). In this thesis, the focus is on how EVT can be applied to market risk through its integration into VaR estimation.

Unlike conventional VaR methods that rely on the entire return distribution, EVT-based approaches concentrate specifically on extreme outcomes and provide a more accurate assessment of tail risk (Longin, 2000, 1998.) In other words, EVT is used to model and estimate the behavior of the distribution's tails (Hull, 2018, 307). Empirical research by Cont (2001) shows that the return distributions in financial markets often exhibit heavy tails, meaning that large losses occur more frequently than the normal distribution would predict. As mentioned, in these situations, EVT offers an effective statistical approach for modeling heavy tails and extreme values.

As mentioned earlier, EVT is usually divided into two models. Block Maxima model focuses on the most extreme values observed within large datasets consisting of identically distributed samples. This model divides historical data into periods of equal lengths and determines the worst return in each period. Other model is known as the Peaks-Over-Threshold (POT). It uses only returns from the whole dataset that exceeds a certain high level or threshold. POT is often more popular out of the two, since it is generally considered to be the most useful for practical applications (McNeil et al., 2015; Miller, 2019). Next, Section 2.3.2 will provide a more detailed examination for the Block Maxima model and Section 2.3.3 for the Peaks-Over-Threshold model.

2.3.2 Block-Maxima model

Block-Maxima method is a fundamental approach in EVT, where observed dataset is divided into multiple blocks, and the maximum value from each block is selected for further analysis. The method examines the asymptotic behavior of the distribution of extreme values, specifically the maximum or minimum observations, as the sample size approaches infinity (Marimoutou, Raggad and Trabelsi, 2009). The following sections are based on (Embrechts, Klüppelberg and Mikosch, 1997, 114-115), a widely referenced work that has been used in numerous studies and books on the subject, such as in McNeil (1998), Gençay and Selçuk (2004), Kuester, Mittnik and Paolella (2005), and McNeil et al. (2015).

Suppose we have n independent and identically distributed (i.i.d.) random variables X_1, X_2, $\ldots$, X_n. They share a common cumulative distribution function $F(x)$. We define the maximum value of this sample:

$$M_n = \max(X_1, \ldots, X_n), \quad n \geq 2.$$

Since M_n is the largest observed value in the sample, its cumulative distribution function is obtained as follows:

$$P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = F^n(x),$$

where $x \in \mathbb{R}$ and $n \in \mathbb{N}$. This means that the maximum is at most x if all the sample values are at most x. So what happens when the sample size becomes large? Extreme values tend to appear near the upper limit of the distribution. This means that the asymptotic behavior of M_n is influenced by the right tail of the cumulative distribution function $F(x)$ near its highest value. We denote this by

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}.$$

Here x_F denotes the right end point of F. In other words x_F is the smallest upper bound such that the distribution function $F(x)$ does not reach the value 1. From here we can see that if $x < x_F$ then the probability that the sample maximum remains below x approaches zero as $n \to \infty$:

$$P(M_n \leq x) = F^n(x) \to 0.$$

Furthermore, if $x \geq x_F$ and $x_F < \infty$, we get:

$$P(M_n \leq x) = F^n(x) = 1.$$

In practical terms, if the distribution has a clear upper bound x_F, then the largest sample values will increasingly concentrate near this bound as the sample size grows:

$$M_n \to x_F, \quad \text{as } n \to \infty,$$

where $x_F \leq \infty$.

As $n \to \infty$, the shape of the distribution's tail determines which distribution the sample maximum M_n follows. The theorem of Fisher and Tippett (1928), later supplemented by Gnedenko (1943), introduces limit laws for this sample maximum. Let (X_n) be a sequence of i.i.d. random variables with a cumulative distribution function $F(x)$. If there exist norming constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that:

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H, \quad \text{as } n \to \infty,$$

where $M_n = \max(X_1, ..., X_n)$ is the sample maximum, then the limiting distribution $H(x)$ belongs to one of the following three types: Gumbel distribution, Fréchet distribution, and Weibull distribution.

In McNeil et al. (2015, 139-141) the use of these distributions are described as follows: The Gumbel distribution has an exponentially decreasing tail. This class includes the normal and exponential distributions, as well as some with moderately heavy tails, such as the lognormal distribution, however, these are not considered heavy-tailed. Its cumulative distribution function is given by:

$$\Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$$

The Fréchet distribution is used to model heavy-tailed distributions, like Pareto, Student's t, and log-gamma, which are common in finance and catastrophe risk analysis. It's relevant when extreme values can get really large since its tails follow a power function $x - \alpha$, meaning exceptionally high values stay relatively probable. These distributions are defined by regularly varying and slowly varying functions. The Fréchet distribution is given by:

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \quad \alpha > 0 \end{cases}$$

In the context of EVT, the Weibull distribution corresponds to distributions with a finite right endpoint $x_F \to \infty$. This case is generally less relevant in financial modeling, compared to Gumbel distribution and Fréchet distribution, particularly for market risk, since loss models with infinite support are typically preferred. Its cumulative distribution function is given by:

$$\Psi_\alpha(x) = \begin{cases} \exp\left[-(-x)^{-\alpha}\right], & x \leq 0, \quad \alpha < 0 \\ 1, & x > 0. \end{cases}$$

Since manually selecting the correct distribution can be hard and subjective, generalized extreme value (GEV) distribution combines the three forms into a single model (Coles, 2001, 47-49). Embrechts et al. (1997, 316) presents GEV in following form

$$H_{\xi;\mu,\psi}(x) = \exp\left\{ -\left(1+\xi\frac{x-\mu}{\psi}\right)^{-1/\xi} \right\}, \quad 1+\xi\frac{x-\mu}{\psi} > 0, \quad x \in \mathbb{R}. \tag{1}$$

Here μ is location parameter, which determines where the center of the distribution is. ψ is scale parameter and it tells how wide the distribution is and ξ expresses the shape of the distribution (Coles, 2001, 47-49). The value of ξ classifies the GEV distribution into three types of distributions we talked before: $\xi > 0$ corresponds to the Fréchet distribution, $\xi = 0$ corresponds to the Gumbel distribution and $\xi < 0$ corresponds to the Weibull distribution (McNeil et al., 2015, 136). The standardized version of the GEV distribution is also presented for example in Embrechts et al. (1997, 152) where it is assumed that $\mu = 1$ and $\psi = 0$. This simplifies the distribution, making it appear in the following form:

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

where $1 + \xi x > 0$.

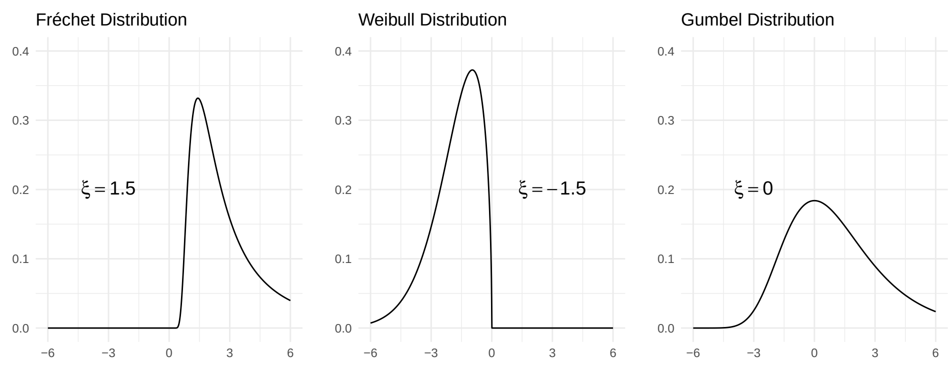

Figure 2: Plots of cumulative distribution functions for different distributions.
The scale parameter is $\psi = 2$ in each case, and the shape parameter ξ is as shown in the image.

However, the Block Maxima method has a significant drawback: it discards a large portion of the available data, as only the maximum value from each large block is retained for analysis, even though one block could contain more extreme events. (Coles, 2001, 74; McNeil et al., 2015, 146). In practical applications, observations are usually not entirely independent; for example, extreme values of sea level heights are often observed close together due to natural phenomena. Especially in such cases, a lot of data from exceptional years with multiple extreme observations remains unused (Vuorinen, 2020, 14).

2.3.3 Peaks-Over-Threshold model

Due to the inefficiency of the Block Maxima method, threshold exceedance methods have largely replaced it in practice, as they utilize all data points that exceed a specified high threshold. Usually, the Generalized Pareto distribution (GPD) is used to model the distribution of these exceedances (McNeil et al., 2015, 146-147).

In EVT, the GPD is a fundamental distribution used to model threshold exceedances. Its formation is closely related to the GEV distribution, which was talked about in Section 2.3.2. Suppose again that we have n i.i.d. random variables $X_1, X_2, \ldots, X_n$, which have common distribution function F. The block maximum $M_n = \max(X_1, X_2, \ldots, X_n)$ asymptotically follows the distribution $G(z)$, which has the form

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}.$$

It can be observed that this equation is identical to Equation (1), except that the parameters μ, σ and ξ determine the location, scale, and tail behavior of the extreme values in the distribution, respectively. Let X represent any of the variables in $X_1, X_2, \ldots, X_n$. When we analyze values that exceed a sufficiently large threshold u, the distribution of $(X - u)$, given that $X > u$, can be approximated by a function:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{2}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$, $y > 0$, and ξ is again the shape parameter. There is a clear connection between the GEV and GPD distributions: the distribution of Block Maxima follows the GEV distribution, and its parameters determine the GPD distribution of threshold exceedances. Notably, the shape parameter ξ remains the same in both the GEV and GPD models. In contrast, the scale parameter $\tilde{\sigma}$ of the GPD distribution depends on the threshold, with parameter u in $\tilde{\sigma} = \sigma + \xi(u - \mu)$. The shape parameter is important because the behavior of the GPD distribution depends particularly on the shape parameter ξ: If $\xi < 0$ the excess variable y is upper-bounded, meaning that extreme values have a clear maximum limit and the distribution is short-tailed. If $\xi = 0$, the GPD simplifies to an exponential distribution:

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y > 0.$$

This corresponds to a situation where the distribution of extreme values decreases exponentially. If $\xi > 0$, the distribution is unbounded, meaning that extreme values can grow arbitrarily large. (Coles, 2001, 74-76.)

The selection of an appropriate threshold level is a critical step in the POT method to ensure the generation of accurate results. If the threshold is set too high, the number of exceedances becomes insufficient, leading to greater variability and reduced reliability in the estimation of extreme value parameters. Conversely, if the threshold is set too low, the GPD may fail to accurately characterize the behavior of exceedances, thereby producing biased parameter estimates. Typically, the threshold is determined based on the model's goodness-to-fit to the data, which we will discuss later in Section 3.3.1 (Barbel and Holger, 2004).

Mean excess function, also referred to as mean residual life function, quantifies the expected amount by which a random variable X exceeds a specified threshold u, given that X is already greater than u. The mean excess function of a random variable X with a finite mean is defined as:

$$e(u) = E(X - u \mid X > u),$$

where $e(u)$ represents the expected excess over the threshold u. (McNeil et al., 2015, 148). As previously discussed in this section, the POT method assumes that the exceedances of the largest observations above the threshold u follow the GPD. The mean excess function of the GPD as presented in Coles (2001, 79) and McNeil et al. (2015, 148), is given by:

$$e(u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma + \xi u}{1-\xi}, \quad \xi < 1,$$

where again ξ represents the shape parameter and σ denotes the scale parameter. If the GPD provides a good fit, the mean excess function $e(u)$, should increase approximately linearly as a function of u, once the threshold is sufficiently high. To determine the threshold u, we analyze the sample's behavior in mean excess function. When u is set too low, the function often displays curvature, or in some cases, remains constant, which may suggest an exponential tail structure rather than a heavy-tailed distribution. Conversely, when u is set too high, the number of exceedances is significantly reduced, leading to increased variance in parameter estimates and, consequently, lower reliability. For a proper range of u the function displays linear growth, suggesting that the GPD is a suitable model for excess values over this threshold. The optimal threshold u is chosen at the point where the function first demonstrates linearity while also ensuring a sufficient number of exceedances for reliable parameter estimation. (Coles, 2001, 79-80.)

Scarrott and MacDonald (2012) defines the Hill plot as a method specifically designed for heavy-tailed distributions ($\xi > 0$), based on the Hill estimator introduced in Hill (1975). The Hill estimator is designed to estimate the shape parameter ξ in cases where the tail of the distribution exhibits Pareto-type behavior. Let $X_1, X_2, ..., X_n$ denote a dataset sorted in ascending order, from which the largest $k + 1$ values are selected. The tail index is defined as $\xi = \frac{1}{\alpha}$ (Barbel and Holger, 2004). Larger α corresponds to a smaller ξ indicating a lighter-tailed distribution, whereas a smaller α leads to larger ξ, signifying a heavier-tailed distribution. The Hill estimator, H_k, computes the average logarithmic difference between these top values and is given by:

$$H_k = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{(n-i+1)} - \log X_{(n-k)}\right).$$

Hill plot visualizes the value of the Hill estimator for different choices of k, where k represents the number of the largest observations included. An optimal threshold u is at the point where the Hill plot stabilizes. (Barbel and Holger, 2004; Scarrott and MacDonald, 2012).

Other ways to estimate the right threshold include probabilistic methods, such as using the exponential distribution or plug-in estimators; computational approaches, like bootstrap-based methods and algorithmic techniques; and mixture models, which incorporate bulk models with varying levels of parametric assumptions (Scarrott and MacDonald, 2012).

2.4 Review of Empirical Applications of EVT in VaR Estimation

EVT has been widely employed in the literature to improve the estimation of Value at Risk (VaR), particularly during periods of financial stress. This section reviews key empirical studies that have applied EVT to financial return series, covering both the Block Maxima and the Peaks-Over-Threshold (POT) approaches.

One of the earliest and most influential applications of EVT to VaR estimation is by Danielsson and Vries (2000), who used a semi-parametric POT method to estimate the tail index and quantify extreme losses. This approach combines historical simulation (non-parametric) to model the central part of the distribution with the POT method to capture the tails. Their research shows that while traditional models may perform reasonably at moderate risk levels, they tend to significantly underestimate losses in the distribution's tails, particularly at the 99% VaR level. The model is empirically validated through backtesting on simulated portfolios and demonstrates more accurate prediction of rare losses.

Another foundational study is McNeil and Frey (2000), who introduced a two-step approach for VaR estimation by combining GARCH[2] models for volatility clustering with the POT method for tail risk modeling. Their analysis uses financial return data from the U.S. and German markets, the USD–GBP exchange rate, and gold prices. Backtesting results show that the GARCH–EVT model provides more reliable VaR estimates than standard normal-based methods, especially during periods of high market turbulence.

In contrast to POT-based methods, Longin (2000) applied the Block Maxima approach by fitting a Generalized Extreme Value (GEV) distribution to extreme daily returns of the S&P 500 and SBF 240 indices. The study compares Historical VaR, Normal Linear VaR, and Block Maxima based VaR. It concludes that the EVT-based approach effectively captures rare and extreme market movements by modeling the distribution tails. Compared to traditional methods, EVT-based VaR is preferred as it enables reliable out-of-sample forecasting, reduces model risk by avoiding rigid distributional assumptions, and explicitly accounts for extreme event risk that standard models like GARCH or historical simulation often overlook.

More relevant to this thesis, Ben Ayed and Ben Hassen (2024) and Aridi, Hooi and Cheong (2023) investigate the performance of EVT-based VaR during the COVID-19 crisis. Ben Ayed and Ben Hassen (2024) focus on Islamic stock markets and aim to identify the most accurate VaR method for calculating the Minimum Capital Requirement, which is the amount of capital financial institutions are required to hold under Basel II and Basel II.5 regulations to cover market risk. Their study applies the POT method for EVT-based VaR estimation and concludes that EVT performs better across most indices during the crisis period. Aridi et al. (2023), in turn, evaluate the Malaysian Shariah-compliant stock market using a conditional EVT method, which combines the POT approach with GARCH- and EGARCH-filtered return series. Their results also show that conditional EVT-based VaR outperforms VaR estimates based solely on GARCH or EGARCH models.

The existing literature strongly supports the use of EVT to enhance the accuracy of VaR estimates, especially under extreme market conditions. POT-based models are more commonly employed than Block Maxima based ones, as they typically offer greater flexibility and a better fit for observed tail behavior. Most studies include empirical backtesting, consistently demonstrating that EVT-based models outperform traditional VaR methods in capturing rare losses. Nevertheless, challenges remain, particularly regarding threshold selection in POT models, the limitations imposed by small sample sizes in GEV estimation, and the practical implementation of these models in real-time risk systems. This thesis contributes to the literature by applying both EVT methods to a new dataset,

[2] GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models how volatility changes over time in financial time series. Instead of assuming constant variance, GARCH assumes that today's volatility depends on past squared returns (shocks) and past volatility estimates (Engle, 1982).

along with a comprehensive backtesting procedure during the COVID-19 crisis, thereby offering updated insights into their comparative performance.

3 Methodology

This section presents the research methodology used to analyze the effectiveness of Extreme Value Theory (EVT) in enhancing Value-at-Risk (VaR) models during extreme market conditions, particularly in the Finnish stock market during the COVID-19 crisis. The methodology outlines the data selection process, the different VaR calculation techniques, the EVT framework, and the approach for backtesting and comparing model performance.

3.1 Research Design

To answer the research question, this study follows a structured approach. First, financial market data from the Finnish stock exchange (OMXHPI) during the COVID-19 crisis is collected and analyzed to understand return distributions. This data is discussed in the Section 4. Second, traditional VaR models (normal linear, historical, and Monte Carlo simulated VaR) are implemented and compared. Third, EVT-based approaches, including the Block Maxima and Peaks-Over-Threshold methods, are used to model extreme losses. Then, a comparative analysis between traditional and EVT-based VaR models is conducted, using backtesting techniques to assess model performance. Finally, robustness checks are performed to validate the findings. This section is structured as follows: Section 3.2 outlines the traditional VaR calculation methods used as benchmarks and Section 3.3 presents the EVT-based models and explains the estimation process. Finally, Section 3.4 introduces the backtesting methods used in the study.

3.2 Traditional Value at Risk calculation methods

As defined earlier the traditional VaR calculation methods are Normal Linear VaR, Historical VaR and Monte Carlo simulated VaR. Important assumption with all VaR models is that the portfolio remains the same over risk horizon (Alexander, 2009, 142).

Jorion (2006, 107) defined steps for calculating Normal Linear VaR following way: Parameters needed are portfolio's current mark to market, standard deviation of the portfolio, time horizon or holding period (usually 1 or 10 trading days) and confidence level (usually 95% or 99%). In the calculation critical value from the chosen distribution at chosen confidence level is used (e.g. 99% confidence level yields a 2,33 factor, assuming normal distribution). For example, if mark to market is 100$ million, standard deviation is 15%, time horizon 10 days, and confidence level 99%, then VaR is calculated:

$$VaR = \$1000000 \times 0,15 \times \sqrt{\frac{10}{252}} \times 2,33 = \$69600.$$

In other words, the worst potential loss with these parameters is \$69600.

The historical VaR method estimates potential losses by using the empirical distribution of past returns. Let us now take a practical example of how this could be computed: Suppose that we want to calculate the historical VaR at a 5% confidence level for the OMXH index. As a reference period (denoted by N in Section 2.2.3), we select the logarithmic returns of the last 100 trading days for the index. These returns are then sorted from best to worst. VaR at 5% means that 5% of the time, the loss is greater than the VaR value. In other words, in this case we select the fifth worst daily return. The historical VaR is then calculated as follows:

$$VaR = R_{5\%} \times \text{portfolio value}$$

Now, let's assume the portfolio value is \$1000000, and the 5th worst daily return is -2.5%. The corresponding VaR would be:

$$VaR_{5\%} = -2.5\% \times \$1000000 = \$25000$$

This means that with a 5% probability, the portfolio could lose at least \$25,000 in one day, based on historical data. (Linsmeier and Pearson, 2000; Jorion, 2006, 108-110).

The calculation method of Monte Carlo VaR is similar to historical VaR, but it differs in that it does not rely directly on past returns. Instead, it simulates possible future market scenarios using stochastic processes. Again, we choose a 5% confidence level for calculating the Monte Carlo VaR for the OMXH index. Similar to before, we take the last 100 days as the reference period and determine the logarithmic returns of OMXH for these days. Calculate the mean (μ) and standard deviation (σ) of the daily logarithmic returns. Here μ represents the average daily return of OMXH and σ represents the dispersion of returns, i.e., volatility. We then generate 10 000 simulated daily returns from a normal distribution:

$$R_{\text{sim}} = \mu + \sigma \times Z,$$

where Z is a random number drawn from a standard normal distribution $N(0, 1)$. This can be done, for example, using the R programming language. The simulated portfolio values are obtained by multiplying the initial portfolio value (again, for example, \$1000000) by the simulated returns:

$$V_{\text{sim}} = V_0 \times e^{R_{\text{sim}}}$$

The 10,000 simulated portfolio values are then sorted from largest to smallest. Since we are using a 5% confidence level, we select the 500th smallest portfolio value from the simulation results. For example, if the 500th worst simulated return is −3.4%, then the Monte Carlo simulated VaR is:

$$VaR_{\text{Monte Carlo}} = \$1000000 \times (-3.4\%) = \$34000$$

This means that with 95% confidence, the portfolio loss should not exceed \$34000 in a single day based on the Monte Carlo simulation. (Linsmeier and Pearson, 2000).

3.3 Extreme Value Theory for VaR enhancement

3.3.1 Block Maxima Approach

The theoretical foundation of the Block Maxima approach is discussed in Section 2.3.2. This method employs the Generalized Extreme Value (GEV) distribution to model extreme values in financial markets. In this section, the Block Maxima approach is implemented for VaR estimation.

The effectiveness of the Block Maxima method largely depends on the chosen time interval. Selecting the number and size of blocks (denoted as m and n, respectively) involves a trade-off. Generally, a larger n enhances the approximation of the Block Maxima distribution by the GEV distribution, reducing bias in parameter estimates. Conversely, increasing m provides more Block Maxima observations for maximum likelihood estimation, thereby reducing the variance of the parameter estimates (McNeil et al., 2015, 143). In this thesis, I employ monthly blocks, yielding 174 blocks from the dataset, from which the maximum loss within each block is extracted for analysis.

Once the largest losses from each block are identified, the next step is to fit them to a GEV distribution, as defined in Equation (1). The parameters of the GEV distribution are estimated using maximum likelihood estimation, which is the most commonly used method for GEV parameter estimation. The GEV parameters are obtained by maximizing the log-likelihood function with respect to the given parameters (Roslan, Chin and Gabda, 2020). Ailliot, Thompson and Thomson (2011) presents the log-likelihood function for random sample $X_1, X_2, \ldots, X_i$ as follows:

$$\ln L(\theta) = -n\ln(\sigma) - \sum_{i=1}^{n}\left(1 - \xi\frac{X_i - \mu}{\sigma}\right)^{\frac{1}{\xi}} + \left(\frac{1}{\xi} - 1\right)\sum_{i=1}^{n}\ln\left(1 - \xi\frac{X_i - \mu}{\sigma}\right),$$

where n is the number of observations, X_i represents the observed maxima, and the remaining parameters are as defined in Equation (1). The maximum likelihood estimator corresponds to the value θ that maximizes $\ln L(\theta)$. This estimation process involves numerically optimizing this function to obtain the best-fitting parameter values. In this thesis I utilize the *fgev* function from Stephenson (2002) R package *evd* to estimate GEV parameters. The *fgev* function also calculates standard errors by a numerical approximation, which are then used to create confidence intervals, helping to assess the uncertainty and reliability of the estimates (Stephenson, 2002).

After obtaining the parameter estimates, it is essential to validate the fitted GEV model to ensure that it adequately describes the observed block maxima. The goodness-of-fit is evaluated using graphical methods, namely Probability-Probability (P-P) and Quantile-Quantile (Q-Q) plots. A P-P plot is a visual tool for assessing how well a given probability distribution fits a set of observed data (Mushkudiani and Einmahl, 2007). In this case, P-P compares the empirical cumulative distribution function of the observed maxima with the fitted GEV cumulative distribution function. A Q–Q plot is also a graphical method for comparing observed data distribution with a theoretical reference distribution. It plots the empirical quantiles of the dataset against the theoretical quantiles expected under the assumed distribution. If the points in the Q–Q plot follow a straight line, it suggests that the data aligns well with the reference distribution. Deviations from linearity indicate potential discrepancies, suggesting that the assumed distribution may not be a good fit (McNeil et al., 2015, 85). In this case, Q-Q plots assess whether the quantiles of the observed data align with the theoretical GEV quantiles. Both P-P and Q-Q plots can also be enhanced from the *evd* package.

Finally, once the GEV model is deemed appropriate, the estimated parameters are used to compute VaR at confidence levels of $\alpha = 95\%$ and $\alpha = 99\%$. GEV-based VaR is calculated using the following equation presented in Szubzda and Chlebus (2019):

$$VaR(\alpha) = \begin{cases} \mu_n - \frac{\sigma_n}{\xi_n}\left(1 - (-n\ln(\alpha))^{-\xi_n}\right), & \text{if } \xi > 0 \text{ (Fréchet)} \\ \mu_n - \sigma_n \ln(-\ln(\alpha)), & \text{if } \xi = 0 \text{ (Gumbel),} \end{cases}$$

where again the parameters are the same as in equation (1).

3.3.2 Peaks-Over-Threshold (POT) Approach

As discussed in Section (2.3.3) on the Peaks-Over-Threshold (POT) approach, selecting an appropriate threshold level u is critical for obtaining reliable and accurate results. Proper threshold selection ensures that the model effectively captures extreme values while maintaining statistical robustness. Two primary methods for determining the optimal threshold u were introduced: the Mean Excess Function and the Hill Plot. In this study, both of them are employed for more accurate threshold estimation. These two methods are chosen due to their widespread use in extreme value theory research and its ability to provide a visually intuitive assessment of whether the GPD assumption holds.

Once the threshold is determined, the exceedances are modeled using the GPD, as defined in Equation (2). The subsequent estimation procedure follows the same steps outlined in Section 3.3.1, where Maximum Likelihood Estimation is applied using the *fgev* function in R to estimate the GPD parameters. To validate the model fit, P-P plots and Q-Q plots, introduced in Section 3.3.1, are utilized to assess whether the fitted GPD model accurately describes the exceedances. If the model aligns well with empirical exceedances in these diagnostics, it can be deemed suitable for VaR estimation.

Once the GPD parameters are estimated, they are used to compute VaR at confidence levels $\alpha = 0.95$ and $\alpha = 0.99$. Szubzda and Chlebus (2019) presents the VaR formula under the GPD framework as follows:

$$VaR(\alpha) = u + \frac{\sigma}{\xi}\left[\left(\frac{n}{N_u}(1-\alpha)\right)^{-\xi} - 1\right], \tag{3}$$

where n is the total number of return observations in the dataset and N_u represents the number of return rates exceeding the threshold u.

3.4 Backtesting Methods

Backtesting is a statistical framework which verifies that actual losses are in line with projected ones (Jorion, 2006, 139). It is an evaluation of how effectively the current method for calculating the measure would have performed based on past data. Backtesting involves examining how frequently the actual loss exceeds the estimated VaR over a given time period when VaR is calculated using the chosen methodology. Times where the actual loss surpasses the VaR estimate are referred to as exceptions. Say we choose to calculate one-day VaR with 1% confidence. If exceptions occur around 1% of the time, the VaR methodology is likely reliable. However, if they happen 7% of the time, VaR is probably underestimated, leading to insufficient capital. Conversely, if exceptions occur only 0.3% of the time, VaR may be overestimated, resulting in excessive capital requirements (Hull, 2018, 285-286). Alexander (2009, 334) represents this mathematically by the Bernoulli variable, where the assumption is that VaR exceedances follow an i.i.d. Bernoulli process. In the Bernoulli process a variable takes values 1 or 0. Here, value 1 means exception. We define an indicator function $I_{a,t}$ that marks whether daily returns or profits and losses

(P&L) exceed the VaR threshold set at the 100α% confidence level by

$$I_{a,t+1} = \begin{cases} 1, & \text{if } Y_{t+1} < -\text{VaR}_{1,a,t}, \\ 0, & \text{otherwise.} \end{cases}$$

Here Y_{t+1} is the realized daily return or P&L on the portfolio from time t, when the VaR estimate is made, to time $t+1$. Christoffersen (1998) introduced that for the VaR model to be accurate, the sequence I_{t+1} should satisfy unconditional coverage (matching the expected exceedance rate α with the probability of exceptions) and independence (past exceedances do not predict future ones). If both conditions are met at the same time, the sequence genuinely follows an i.i.d. Bernoulli process and the VaR forecast is considered to have valid conditional coverage (Alexander, 2009, 333-337).

VaR backtesting is often divided into unconditional coverage tests and conditional coverage tests (Jorion, 2006). The most common unconditional coverage test introduced by (Kupiec et al., 1995) is the Propotion of Failure test (POF-test). It test if the reported VaR is violated more or less than 100α% of the time. The formula for the POF test is following:

$$\text{POF} = 2\ln\left[\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{n-I(\alpha)}\left(\frac{\hat{\alpha}}{\alpha}\right)^{I(\alpha)}\right], \tag{4}$$

where

$$\hat{\alpha} = \frac{1}{n}I(\alpha).$$

Here also n is the number of observations, I_α is number of exceedances, and α is the given confidence level (Zhang and Nadarajah, 2018, 3619). The POF test statistic follows a likelihood ratio test, which is asymptotically chi-squared (χ^2) distributed with one degree of freedom. Under the null hypothesis ($H_0 : \hat{\alpha} = \alpha$), the POF test statistic is compared to the critical value of the distribution with a given confidence level. If the POF value is greater than the critical value, the model is rejected, indicating a mismatch between the expected and observed exceedances. This statistical approach ensures that the VaR model accurately reflects risk by verifying whether exceedance frequencies align with theoretical probabilities (Jorion, 2006, 146-147).

The biggest flaw of the unconditional coverage test is that the test only examines the total number of exceedances, but it does not take into account the temporal dependence of exceedances. Christoffersen (1998) introduced conditional coverage test aims to assess whether VaR exceedances are temporally independent or occur in clusters, which may indicate shortcomings in the VaR model. The test combines the independence statistic with Kupiec's POF - test. The test formula is:

$$LR = LR_{ind} + POF,$$

where

$$LR_{ind} = -2\ln\left[(1-\pi)^{N_{00}+N_{01}}\pi^{N_{01}+N_{11}}\right] + 2\ln\left[(1-\pi_0)^{N_{00}}\pi_0^{N_{01}}(1-\pi_1)^{N_{10}}\pi_1^{N_{11}}\right].$$

and POF is defined as in the Equation (4). Each test component follows a chi-squared χ^2 distribution with one degree of freedom independently. When combined, the total test statistic follows a chi-squared distribution with two degrees of freedom. The test begins by defining an indicator function I_t, which is 0 if VaR is not exception and 1 if it is. A contingency table is constructed to track the number of days (N_{ij}) in which a specific transition occurs between consecutive days: N_{00} (no exception followed by no exception), N_{01} (no exception followed by a exception), N_{10} (exception followed by no exception), and N_{11} (exception followed by another exception). To test independence, transition probabilities are computed: $\pi_0 = \frac{N_{01}}{N_{00}+N_{01}}$, the probability of a violation given that the previous day had no violation, and $\pi_1 = \frac{N_{11}}{N_{10}+N_{11}}$, the probability of a violation given that the previous day had a violation. If exceedances are truly independent, then $\pi_0 = \pi_1$. The result from the LR_{ind} and POF are independently compared to critical value from the $\chi^2(1)$ and the combination to critical value from $\chi^2(2)$ to find out if the VaR model correctly predicts the number of exceedances and whether those exceedances occur independently over time. (Jorion, 2006; Zhang and Nadarajah, 2018)

Other VaR backtesting methods are for example Kupiec's TUFF test (Time until first failure), Time between failures likelihood ratio test, density forecast tests, Lopez's magnitude loss function and the Basel Committee's traffic light test (Zhang and Nadarajah, 2018). However, in this thesis, I use only Kupiec's POF test and Christoffersen's joint test due to their simplicity and widespread recognition.

4 Data and Results

Assuming that the Value at Risk (VaR) estimate is based on historical data, it is necessary to define an estimation period, which determines the sample used to evaluate the VaR parameters (Alexander, 2009, 332). In this thesis, the estimation period comprises OMXHPI closing price data spanning from August 1, 2005, to July 31, 2020. This dataset has been sourced from Eikon Refinitiv. The returns were transformed into continuously compounded returns using the logarithmic return formula as presented by (Tsay, 2010, 5):

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

The log-returns are segmented into two distinct parts: the first 3785 observations (from August 1 2005 to January 31 2020) are employed for estimating the model and its parameters, while the final 130 observations (from February 3 2020 to July 31 2020) are utilized for backtesting. The whole dataset is presented in Figure 3 where the left figure displays the prices of OMXHPI, and the right figure the log-returns. The dashed line separates the sample data from the backtesting data. We can see that during the almost 15-year sample period there have been one major market crash during the financial crises from 2008 to 2009. Figure 4 presents prices and log-returns of OMXHPI during the COVID-19, which is used as backtesting period. If we compare log-returns from Figure 3 to log-returns in Figure 4 we can see a substantial increase in volatility during the early months of 2020. This reflects the extreme uncertainty and market stress triggered by the pandemic and provides a suitable environment to evaluate the performance of different Value-at-Risk models under extreme market conditions.

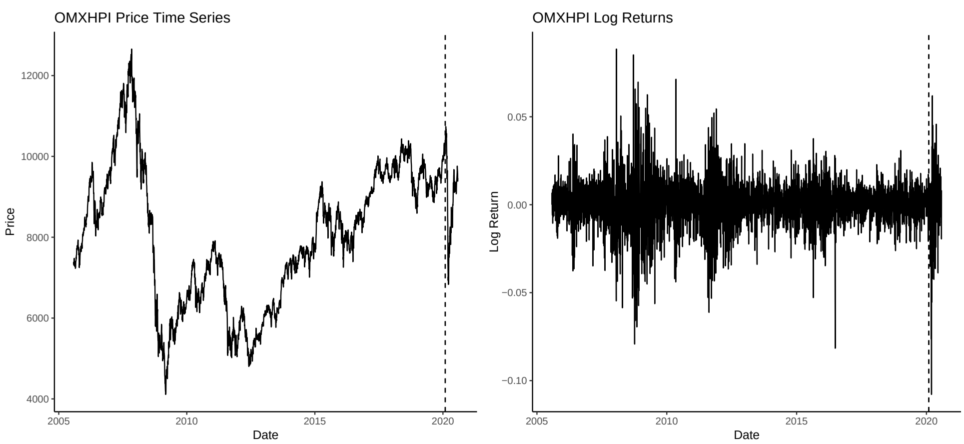

Figure 3: Price and log-return Time Series whole dataset

We have to address the relatively low number of VaR exceedances, particularly at the 99% confidence level, which limits the statistical power of the backtesting results presen-

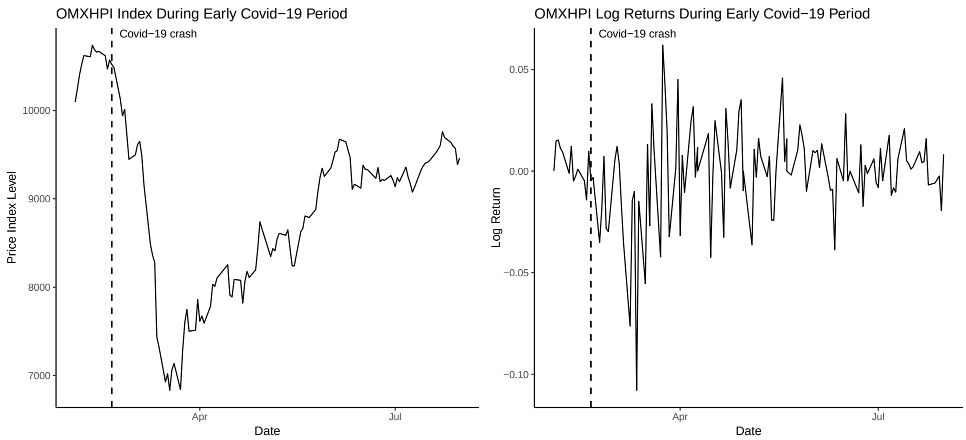

Figure 4: Price and log-return Time Series COVID-19 dataset

ted in Table 2 and in Table 4. This is a known issue when working with high-confidence VaR models, where exceptions are inherently rare. Additionally, focusing the analysis specifically on the COVID-19 crisis reduces the overall sample size to only 130 observations. However, this narrower time frame is a deliberate choice as the COVID-19 period represents an extraordinary stress scenario, offering a unique opportunity to evaluate the performance of different VaR models under extreme market conditions. Our focus on an extreme period accentuates the differences, but also means the results are tailored to that unusual context. For example, Basel III states to use 12 months or 250 observations to backtesting models (Basel Committee on Banking Supervision, 2019). Had the sample been, for example, the 250 days including calmer periods around the crisis, the test outcomes could differ. While this limits the generalizability of the results in this thesis, it provides valuable insight into model behavior during a financial crisis, which is highly relevant from a risk management perspective.

Table 1 reports descriptive statistics for the OMXHPI log-return series across three time periods: the 15-year pre-COVID period, the COVID crisis window, and the entire dataset. The standard deviation increased markedly during the COVID period relative to the longer pre-COVID window, indicating a substantial rise in market volatility. Maximum and minimum returns also highlight this elevated risk: the COVID period exhibits the most negative return and a comparatively lower positive maximum. All three return distributions exhibit negative skewness, suggesting that downside-tail events are more extreme than their upside counterparts. This asymmetry can be particularly pronounced during COVID (skewness $= -1.1411$), indicating a higher frequency of extreme negative returns. Additionally, kurtosis values significantly exceeding the normal benchmark of 3 confirm the presence of heavy tails, implying an elevated likelihood of extreme outcomes. Finally, the Jarque–Bera test, which tests if the dataset follows normal distribution, strongly rejects the null hypothesis of normality in all cases ($p < 0.001$) (McNeil et al., 2015, 85). This justifies the use of EVT-based approaches, which are specifically designed to capture the non-normal behavior observed in the tails of return distributions.

Table 1: Descriptive statistics for the OMXHPI return series across different time periods.

Statistic	15-Year Data	COVID Period	Whole Dataset
Mean	0.019	−0.050	0.017
Std. Dev.	0.014	0.022	0.014
Min	−0.089	−0.108	−0.108
Max	0.093	0.062	0.093
Skewness	−0.099	−1.114	−0.227
Kurtosis	7.582	7.541	8.124
Jarque–Bera	3316.6	138.6	4133.8
p-value	< 0.001	< 0.001	< 0.001

4.1 Backtesting and Performance Evaluation of Traditional VaR methods

Table 2 presents backtesting results for VaR models applied to OMXHPI log-returns during the COVID period, using Normal Linear, Historical, and Monte Carlo methods. Model performance is assessed based on expected versus actual exceedances and two standard backtesting tests: the unconditional coverage test (LR_{uc}) and the conditional coverage test (LR_{cc}).

The Normal Linear VaR model performs poorly, significantly underestimating risk and failing both tests. It has 17 exceedances at 95%, which is nearly three times over the expected 6 exceedances. This shows it underestimated risk and also suggests potential clustering of exceedances, as also the Christoffersen's test is rejected. Figures 3 and 4, which present the log-return time series of the full sample period and the COVID-19 backtesting period respectively, further support this interpretation. Both figures exhibit signs of volatility clustering, with particularly elevated clustering observed during the COVID-19 period. The Historical VaR model shows solid performance overall, failing no tests at the 95% level and only marginally failing at the 99% level, indicating that it generally provides more reliable coverage than the normal model. The Monte Carlo VaR model performs moderately: it underestimates risk at the 95% level but shows acceptable results at the 99% level. Overall, the historical method offers the most robust performance under the conditions of the COVID period, followed by Monte Carlo method, while the Normal Linear model proves inadequate.

Table 2: VaR backtest results OMXHPI for COVID (31.7.2005-31.7.2020 used as sample data)

The table presents backtesting results for VaR at 95% and 99% confidence levels using Normal Linear, Historical, and Monte Carlo VaR methods. Expected and actual exceedance refer to the theoretical and observed number of VaR breaches. LR_{uc} and LR_{cc} denote the test statistics for unconditional and conditional coverage, respectively. The decision indicates whether the null hypothesis H_0 of correct VaR model specification is rejected.

	Normal Linear VaR		Historical VaR		Monte Carlo VaR	
	$VaR_{95\%}$	$VaR_{99\%}$	$VaR_{95\%}$	$VaR_{99\%}$	$VaR_{95\%}$	$VaR_{99\%}$
Expected exceedance	6	1	6	6	6	1
Actual VaR exceedance	17	12	7	2	15	3
LR_{uc}	12.786	33.03072	0.04809234	4.375934	8.686364	1.640043
Decision	Reject H_0	Reject H_0	Fail to Reject H_0	Reject H_0	Reject H_0	Fail to Reject H_0
LR_{cc}	13.08864	33.04817	0.8584621	4.439429	8.73276	1.782914
Decision	Reject H_0	Reject H_0	Fail to Reject H_0	Fail to Reject H_0	Reject H_0	Fail to Reject H_0

4.2 Backtesting Results of EVT-Based VaR

4.2.1 Backtesting Block Maxima method

The parameters estimated from the sample data for the Block Maxima method, are presented in Table 3. These parameters are calculated for the GEV distribution, defined in Equation (1).

Table 3: GEV parameter estimates.

The location parameter μ determines the central tendency, the scale parameter ψ reflects the spread of the distribution, and the shape parameter ξ characterizes the tail behavior.

Parameters	**Location** μ	**Scale** ψ	**Shape** ξ
Estimate	-0.06076018	0.03236066	-0.57590907

It is noteworthy that the GEV model yields a strongly negative shape parameter ξ, indicating that the distribution belongs to the Weibull family. A negative ξ implies that the tail of the distribution is bounded, meaning that extreme losses are assumed to have a finite upper limit and occur less frequently than they might in reality. Consequently, the model underestimates the probability of extreme negative returns, leading to downward-biased VaR estimates. As a result, the VaR estimates are not sufficiently negative, which further underestimates the risk. This causes a lack of observed exceedances in the out-of-sample period, despite the potential for significant losses to occur. This, in turn, undermines the reliability of backtesting procedures, as statistical tests such as Kupiec's or Christoffersen's cannot be performed in the absence of violations. For example, the GEV-based $VaR_{99\%}$ was calculated to be -0.1399, while the most negative return in the COVID sample data was only -0.1078. As no observations fall below the VaR threshold, backtesting becomes infeasible. The goodness-to-fit tests back this observation. The model's goodness-to-fit was evaluated using PP and QQ plots based on the GEV distribution fitted to monthly minima. These diagnostic plots, presented in Figure 5, reveal notable deviations from the theoretical diagonal, especially in the lower tail. This indicates a poor model fit precisely in the region most critical for assessing extreme losses and estimating VaR.

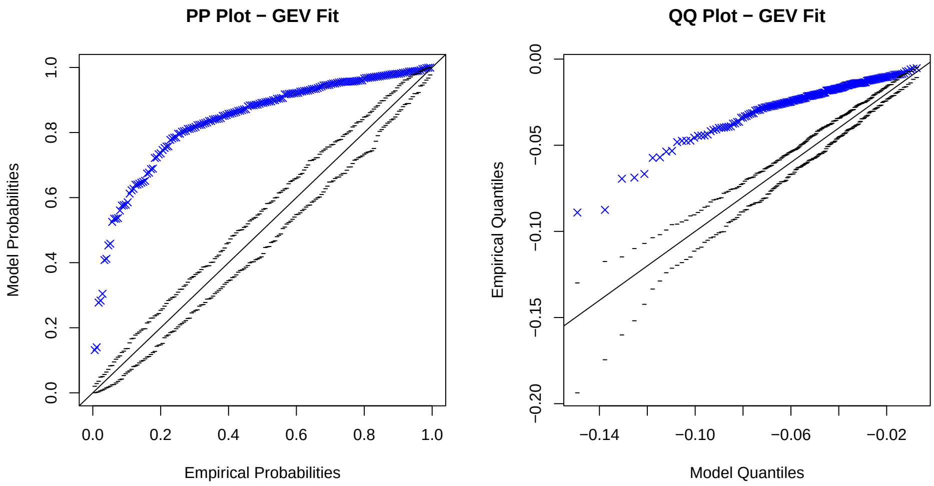

Figure 5: Goodness-to-fit GEV distribution

4.2.2 Backtesting Peaks-Over-Threshold method

Nonetheless, the VaR estimates derived from the POT method yielded meaningful results. The optimal threshold is determined through an analysis of both the Mean Excess Plot and the Hill Plot, as discussed in detail in Section 2.3.3. The Mean Excess Plot presented in the Figure 6 demonstrates a distinct change in behavior across the threshold range. For lower values of u, the plot appears erratic and lacks a clear trend, suggesting that the data in this region do not conform well to the assumptions of the POT method. As the threshold increases, particularly around $u = -0.03$, the mean excess function begins to exhibit a more linear and gently increasing pattern. This behavior aligns with the theoretical expectation for data that follow a Generalized Pareto Distribution above the threshold. Beyond this point, the function eventually turns downward, indicating a breakdown of the GPD fit for higher thresholds. The Hill Plot presented in the Figure 7 complements this by showing that the estimates remain relatively stable between $u = -0.04$ and $u = -0.015$, after which the plot becomes highly volatile. Since $u = -0.03$ lies within the region where both the Mean Excess Plot is approximately linear and the Hill Plot is stable, it represents an appropriate and theoretically justified threshold for the application of the POT method.

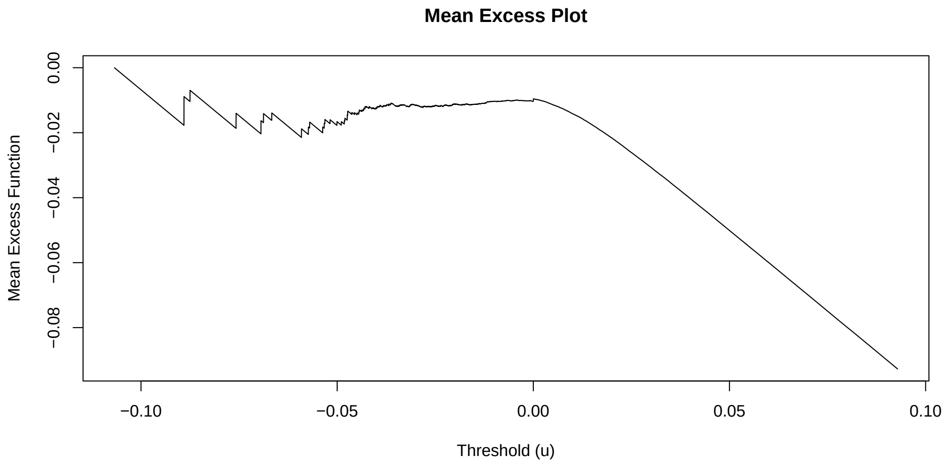

Figure 6: Mean Excess Plot for the estimation of optimal threshold

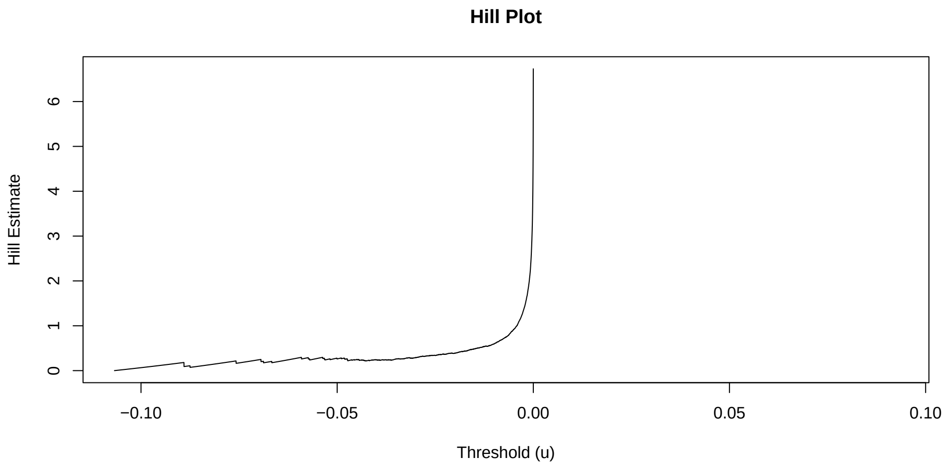

Figure 7: Hill Plot for the estimation of optimal threshold

The goodness-to-fit for the GPD is also assessed using the PP and QQ plots shown in Figure 8. In the PP plot (left panel), the empirical probabilities are plotted against the model probabilities. The points closely follow the 45-degree reference line, indicating that the model-predicted probabilities are in strong agreement with the observed data. This alignment suggests a good overall fit, particularly in the central region of the distribution. The QQ plot (right panel), which compares the empirical and theoretical quantiles, also shows a generally good fit, especially in the lower to mid quantile range. However, some deviation is observed in the upper tail, where the empirical quantiles exceed the model quantiles. These discrepancies, particularly the red crosses diverging from the diagonal line, suggest that the GPD may slightly underestimate extreme values in the tail. Overall, both plots indicate that the GPD provides a reasonable fit to the data, though with minor limitations in capturing the most extreme observations.

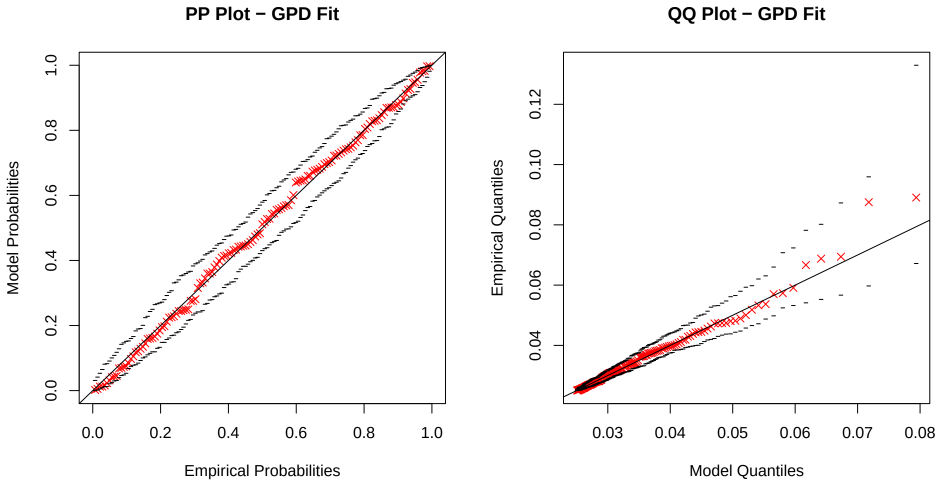

Figure 8: Goodness-to-fit GPD distribution

Table 4 presents the backtesting results of the POT based VaR model for the OMXHPI log-return series over the period 31.7.2005 to 31.7.2020. The model was tested at two confidence levels: 95% and 99%, using backtesting techniques presented in the Section 3.4 to evaluate its accuracy in forecasting extreme losses. At the 95% confidence level, the model predicted 6 exceedances over the test period, while only 2 were actually observed. The Kupiec test statistic ($LR_{uc} = 4.447$) leads to a rejection of the null hypothesis H_0, suggesting that the actual number of exceedances is statistically inconsistent with the expected rate. However, the Christoffersen test ($LR_{cc} = 4.510$) fails to reject H_0, indicating that while the unconditional coverage is not ideal, the exceedances that do occur are not clustered and thus exhibit the required independence. This implies that the model is likely too conservative at the 95% level, slightly overestimating risk.

In contrast, at the 99% confidence level, the POT model performs very well. The expected number of exceedances was 1, which matches the actual observed value exactly. Both Kupiec ($LR_{uc} = 0.076$) and Christoffersen ($LR_{cc} = 0.092$) tests fail to reject the null hypothesis, indicating that the model provides statistically accurate and independent exceedance forecasts at this more extreme threshold. This outcome suggests a well-calibrated tail risk estimation and supports the use of EVT-based methods in capturing rare but impactful losses.

Overall, the results highlight the strength of the POT-based VaR model in modeling tail risk, especially at higher confidence levels where capturing the most extreme losses is critical. The results also indicate that the method may lean slightly toward conservative estimates at lower confidence levels, which is often preferred in risk management to avoid underestimation of potential losses.

Table 4: EVT Based VaR backtest results using Peaks Over Threshold method for OMXHPI (31.7.2005 – 31.7.2020)

The table reports backtesting results for VaR at 95% and 99% confidence levels using the POT method. Expected and actual exceedance refer to the predicted and observed number of VaR breaches. LR_{uc} and LR_{cc} are test statistics for unconditional and conditional coverage, respectively. The decision indicates whether the null hypothesis H_0 of correct VaR model specification is rejected.

Peaks-Over-Threshold based VaR		
	$\text{VaR}_{95\%}$	$\text{VaR}_{99\%}$
Expected exceedance	6	1
Actual VaR exceedance	2	1
LR_{uc}	4.447392	0.07597023
Decision	Reject H_0	Fail to Reject H_0
LR_{cc}	4.510386	0.09159539
Decision	Fail to Reject H_0	Fail to Reject H_0

5 Conclusions

This thesis evaluated whether Extreme Value Theory (EVT) improves the performance of Value At Risk (VaR) models during extreme market conditions, using empirical data from the Finnish stock market during the COVID-19 crisis. Traditional VaR methods, Normal Linear, historical simulation, and Monte Carlo simulation, were compared with EVT-based approaches, which are the Block Maxima and Peaks-Over-Threshold (POT) methods. The backtesting results demonstrate that EVT-based models, particularly the POT method, provided more accurate risk estimates under extreme conditions. Traditional VaR models, especially the Normal Linear approach, consistently underestimated tail risk, leading to frequent VaR violations. This is primarily due to their assumption of normally distributed returns, which fails to capture the heavy-tailed nature of financial return distributions observed during crisis periods. Among the traditional methods, Historical VaR performed the best, having nearly correct coverage at 95% confidence level. As Historical VaR inherently accounts for extreme past events in its empirical distribution, it benefits from including prior crises, like the 2008 financial crisis, in its sample, thus anticipating extreme moves.

In contrast, the POT method focuses specifically on modeling the extreme tail of the distribution by using all observations that exceed a high threshold. This allowed it to retain more relevant data and model extreme losses more accurately. As a result, it produced fewer backtesting violations and aligned more closely with realized losses during the COVID-19 market shock. These advantages are supported by EVT theory, which recognizes POT as more efficient and reliable for modeling rare events compared to the Block Maxima method.

These findings are consistent with many prior researches, including studies by Longin (2000), McNeil and Frey (2000), and Danielsson and Vries (2000), which have shown that EVT-based VaR methods generally outperform traditional models during periods of market stress. The superior performance of the POT method observed in this thesis echoes earlier results in both developed and emerging market contexts. These studies also emphasized that traditional VaR models tend to underestimate tail risk, particularly under crisis conditions. The empirical results further support the prevailing view in the literature (e.g., Coles (2001); McNeil et al. (2015)) that the POT method is more suitable for practical risk management than the Block Maxima approach, particularly when the available data set is moderate in size. Applying EVT to Finnish market data during the COVID-19 crisis, this study contributes new empirical evidence to an unexplored region and confirms that EVT's theoretical advantages also hold in smaller, less liquid markets.

However, the literature also highlights persistent challenges, including the difficulty of threshold selection in POT models, the sensitivity of Block Maxima to small sample sizes, and the limited evidence on real-world implementation through empirical backtesting. This thesis addresses these issues directly. For threshold selection in the POT method, two established graphical techniques were used: the Mean Excess function and the Hill plot to determine an appropriate level. This improves the reliability of the estimation and avoids using arbitrary or fixed threshold values that could distort results. The weakness of the Block Maxima method in small samples was also confirmed empirically, as it failed to produce viable VaR estimates under the COVID-19 market conditions. This supports theoretical concerns raised in the literature (e.g., Coles (2001)) and offers new evidence from a Nordic market context. Furthermore, a comprehensive backtesting procedure was applied using Kupiec's and Christoffersen's tests, ensuring that all models were evaluated under consistent conditions. These steps enhance the practical relevance of the findings and demonstrate that POT-based EVT models can be implemented effectively in real-world risk assessment.

Despite these contributions, several limitations remain. The analysis was limited to a single index and a relatively short time frame, which may affect the generalizability of the results. In addition, the assumption of independent and identically distributed returns may not fully hold during crisis periods, and the models did not incorporate time-varying volatility, which could further improve their accuracy.

However, the strong performance of the POT method, both in theory and empirical testing, supports the conclusion that EVT provides better risk estimation than traditional VaR methods under extreme market conditions. Therefore, we can answer the research question following way: Value at Risk (VaR) based on Extreme Value Theory (EVT) provides a better risk estimation than traditional VaR methods in extreme market conditions.

From a practical perspective, financial institutions could benefit from incorporating EVT-based VaR models, particularly the POT method, into their risk management practices to improve tail risk estimation under stressed conditions. This may strengthen capital adequacy frameworks and enhance preparedness for market shocks. This study focused exclusively on VaR as a risk management measure. Future research could extend the analysis by applying EVT to other risk metrics, such as Expected Shortfall. Additionally, examining different asset classes, such as options or multi-asset portfolios, would broaden the applicability of the findings. Another promising direction would be to develop hybrid models that integrate EVT with copula functions or advanced time series volatility models like EGARCH, to enhance the accuracy of tail risk prediction.

References

Ailliot, P., Thompson, C. – Thomson, P. (2011) Mixed methods for fitting the gev distribution. *Water Resources Research*, vol. 47 (5).

Alexander, C. (2009) *Market Risk Analysis Volume IV: Value-at-Risk Models*. Wiley.

Aridi, N. A., Hooi, T. S. – Cheong, C. W. (2023) The var evaluation of shariah stock market in malaysia during covid-19 pandemic by using conditional evt method. *International Journal of Business and Society*, vol. 24 (3), 1079 – 1098.

Ashraf, B. N. (2020) Stock markets' reaction to covid-19: Cases or fatalities? *Research in International Business and Finance*, vol. 54.

Barbel, F. – Holger, R. (2004) Extreme values in finance, telecommunications, and the environment. Monographs on statistics and applied probability ; 99, Seminaire europeen de statistique (5th : 2001 : Gothenburg, Sweden)., Chapman Hall/CRC, Boca Raton, FL.

Basel Committee on Banking Supervision (1996) Amendment to the capital accord to incorporate market risks. Tech. rep., Bank for International Settlements (BIS), URL: `https://www.bis.org/bcbs/publ/d457.pdf`.

Basel Committee on Banking Supervision (2019) Minimum Capital Requirements for Market Risk. Tech. rep., Bank for International Settlements (BIS).

Ben Ayed, W. – Ben Hassen, R. (2024) The basel 2.5 capital regulatory framework and the covid-19 crisis: evidence from the ethical investment market. *PSU Research Review*, vol. 8 (3), 727 – 748.

Christoffersen, P. F. (1998) Evaluating interval forecasts. *International economic review (Philadelphia)*, vol. 39 (4), 841–862.

Coles, S. (2001) *An introduction to statistical modeling of extreme values*. Springer series in statistics, Springer, Bristol.

Cont, R. (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, vol. 1 (2), 223–236.

Danielsson, J. (2011) *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*, vol. 588 of *Wiley finance series*. Wiley, Newark, 1 edn.

Danielsson, J. – Vries, C. G. D. (2000) Value-at-risk and extreme returns. *Annales d'Économie et de Statistique*, (60), 239–270.

DeCarlo, L. (1997) On the meaning and use of kurtosis. *Psychological Methods*, vol. 2, 292–307.

Devore, J. (2012) *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole, Cengage Learning.

Embrechts, P., Klüppelberg, C. – Mikosch, T. (1997) *Modelling Extremal Events : for Insurance and Finance*. Stochastic Modelling and Applied Probability, 33, Springer Berlin Heidelberg, Berlin, Heidelberg, corr. 4. print. edn.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, vol. 50 (4), 987–1007.

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*, vol. 24 (2), 180–190.

Freeman, J., Anderson, D., Sweeney, D., Williams, T. – Shoesmith, E. (2017) *Statistics For Business and Economics*. Cengage, United States, 4 edn.

Gençay, R. – Selçuk, F. (2004) Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting*, vol. 20 (2), 287–303.

Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, vol. 44 (3), 423–453.

Haan, L. d. L. – Ferreira, A. (2006) *Extreme value theory : an introduction*. Springer series in operations research and financial engineering, Springer, New York.

Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of statistics*, vol. 3 (5), 1163–1174.

Hosking, J. R. M., Wallis, J. R. – Wood, E. F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, vol. 27 (3), 251–261.

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

Jorion, P. (2006) *Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial*

Risk. McGraw Hill LLC.

Kuester, K., Mittnik, S. – Paolella, M. S. (2005) Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, vol. 4 (1), 53–89.

Kupiec, P. H. et al. (1995) *Techniques for verifying the accuracy of risk measurement models*, vol. 95. Division of Research and Statistics, Division of Monetary Affairs, Federal . . . .

Lai, T. L. – Xing, H. (2008) *Statistical models and methods for financial markets*. Springer texts in statistics, Springer, New York.

Linsmeier, T. J. – Pearson, N. D. (2000) Value at risk. *Financial Analysts Journal*, vol. 56, 47 – 67.

Longin, F. M. (2000) From value at risk to stress testing: The extreme value approach. *Journal of Banking Finance*, vol. 24 (7), 1097–1130.

Marimoutou, V., Raggad, B. – Trabelsi, A. (2009) Extreme value theory and value at risk: Application to oil market. *Energy Economics*, vol. 31 (4), 519–530.

Markowitz, H. (1952) Portfolio selection. *The Journal of Finance*, vol. 7 (1), 77–91.

McNeil, A. J. (1998) Calculating quantile risk measures for financial return series using extreme value theory. Report, ETH Zürich, Departement Mathematik.

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

McNeil, A. J., Frey, R. – Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, USA.

Miller, M. B. M. B. (2019) *Quantitative financial risk management*. Wiley finance series, Wiley, Hoboken, New Jersey.

Mushkudiani, N. A. – Einmahl, J. H. (2007) Generalized probability-probability plots. *Journal of Statistical Planning and Inference*, vol. 137 (3), 738 – 752.

Nadarajah, S., Chan, S. – Longin, F. (2016) Estimation methods for value at risk. In *Extreme Events in Finance*, 283–356, John Wiley Sons, Incorporated, United States.

Roslan, R., Chin, S. N. – Gabda, D. (2020) Parameter estimations of the generalized extreme value distributions for small sample size. *Mathematics and Statistics*, vol. 8, 47–51.

Sarykalin, S., Serraino, G. – Uryasev, S. (2008) Value- at-risk vs conditional value-at-risk in risk management and optimization. *Tutorials in Operations Research.*

Scarrott, C. – MacDonald, A. (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, vol. 10 (1), 33–60.

Solari, S., Egüen, M., Polo, M. J. – Losada, M. A. (2017) Peaks over threshold (pot): A methodology for automatic threshold estimation using goodness of fit p-value. *Water Resources Research*, vol. 53 (4), 2833–2849.

Stephenson, A. G. (2002) evd: Extreme value distributions. *R News*, vol. 2 (2), 31–32.

Szubzda, F. – Chlebus, M. (2019) Comparison of block maxima and peaks over threshold value-at-risk models for market risk in various economic conditions. *Central European economic journal*, vol. 6 (53), 70–85.

Tsay, R. S. (2010) *Analysis of financial time series.* Wiley series in probability and statistics, Wiley, Hoboken, NJ, 3rd edition edn.

Vuorinen, J. (2020) Ääriarvoteoria ja value-at-risk ja niiden soveltaminen riskienhallintaan. *Matematiikan ja tilastotieteen laitos Turun yliopisto.*

Williams, R. (n.d.) Normal distribution. `https://www3.nd.edu/~rwilliam/stats1/`, pDF document.

Zhang, Y. – Nadarajah, S. (2018) A review of backtesting for value at risk. *Communications in Statistics - Theory and Methods*, vol. 47 (15), 3616–3639.

Disclosing the use of AI:

Artificial intelligence tools, including ChatGPT, were utilized to improve the clarity, grammar, structure, and wording of the text. Additionally, the abstract was translated into Finnish using AI-assisted translation. Some of the R code used in the empirical analysis was developed with the assistance of AI tools.