Robust Portfolio Optimization with Expected Shortfall

Daniel Isaksson

May 30, 2016

Abstract

This thesis project studies robust portfolio optimization with Expected Shortfall applied to a reference portfolio consisting of Swedish linear assets with stocks and a bond index. Specifically, the classical robust optimization definition, focusing on uncertainties in parameters, is extended to also include uncertainties in log-return distribution. My contribution to the robust optimization community is to study portfolio optimization with Expected Shortfall with log-returns modeled by either elliptical distributions or by a normal copula with asymmetric marginal distributions. The robust optimization problem is solved with worst-case parameters from box and ellipsoidal uncertainty sets constructed from historical data and may be used when an investor has a more conservative view on the market than history suggests.

With elliptically distributed log-returns, the optimization problem is equivalent to Markowitz mean-variance optimization, connected through the risk aversion coefficient. The results show that the optimal holding vector is almost independent of elliptical distribution used to model log-returns, while Expected Shortfall is strongly dependent on elliptical distribution with higher Expected Shortfall as a result of fatter distribution tails.

To model the tails of the log-returns asymmetrically, generalized Pareto distributions are used together with a normal copula to capture multivariate dependence. In this case, the optimization problem is not equivalent to Markowitz mean-variance optimization and the advantages of using Expected Shortfall as risk measure are utilized. With the asymmetric log-return model there is a noticeable difference in optimal holding vector compared to the elliptical distributed model. Furthermore the Expected Shortfall increases, which follows from better modeled distribution tails.

The general conclusions in this thesis project is that portfolio optimization with Expected Shortfall is an important problem being advantageous over Markowitz mean-variance optimization problem when log-returns are modeled with asymmetric distributions. The major drawback of portfolio optimization with Expected Shortfall is that it is a simulation based optimization problem introducing statistical uncertainty, and if the log-returns are drawn from a copula the simulation process involves more steps which potentially can make the program slower than drawing from an elliptical distribution. Thus, portfolio optimization with Expected Shortfall is appropriate to employ when trades are made on daily basis.

Keywords: Robust Portfolio Optimization, Risk Management, Expected Shortfall, Elliptical Distributions, GARCH model, Normal Copula, Hybrid Generalized Pareto-Empirical-Generalized Pareto Marginals, Markowitz Mean-Variance Optimization, Contribution Expected Shortfall

Acknowledgements

I would like to thank my colleagues at SAS Institute for all supporting and encouraging conversations. In particular I would like to thank my supervisor Jimmy Skoglund for introducing me to the topic of robust portfolio optimization, for always being a helping hand and a great source of inspiration. I would also like to thank my supervisor Professor Henrik Hult at the Royal Institute of Technology for valuable support throughout this thesis project.

Stockholm, May 30, 2016

Daniel Isaksson

Contents

1 Introduction **1**
- 1.1 Introduction to Portfolio Optimization 2
  - 1.1.1 Markowitz Mean-Variance Optimization Problem . . . 3
  - 1.1.2 Reference Portfolio and Benchmark Solution 4
  - 1.1.3 Stylized Assumptions in Markowitz Optimization Problem 7

2 Portfolio Optimization with Expected Shortfall **9**
- 2.1 Risk Measures 9
  - 2.1.1 Value-at-Risk 11
  - 2.1.2 Expected Shortfall 12
  - 2.1.3 Contribution Expected Shortfall 12
  - 2.1.4 Properties of the quantile function 14
- 2.2 Portfolio Optimization Problem Formulation 14
  - 2.2.1 Remark: Elliptically distributed log-returns 16
  - 2.2.2 General Problem Assumptions 17
- 2.3 Rockafellar and Uryasev simplification 18
- 2.4 Application on Reference Portfolio 21

3 Portfolio Optimization Under Uncertainty **25**
- 3.1 Model Uncertainty 26
  - 3.1.1 Introduction to Robust Optimization 27
  - 3.1.2 Robust Optimization with Elliptical Distributions . . . 28
    - 3.1.2.1 Solution with Normal Distributed Log-returns 32
    - 3.1.2.2 Solution with Student's t Distributed Log-returns 33
  - 3.1.3 Robust Optimization with Asymmetric Distributions . 35
    - 3.1.3.1 The univariate generalized Pareto distribution 35
    - 3.1.3.2 Copula dependence 37
    - 3.1.3.3 Sample Preparation with GARCH(1,1) Model 38
    - 3.1.3.4 Solution with Normal Copula and Hybrid GPD-Empirical-GPD Log-return Marginals 40
- 3.2 Statistical Uncertainty 42

3.2.1 Statistical Uncertainty in the Portfolio Optimization Problem . . . 43

4 Analysis and Conclusions **45**
4.1 Analysis of Results . . . 45
4.1.1 Elliptically distributed log-returns . . . 46
4.1.2 Asymmetric Distributed Log-returns . . . 50
4.1.3 Analysis of Solutions with Euler Risk Decomposition . 53
4.2 Comments on Worst-Case Scenario Based Robust Optimization 57
4.3 More Comments and Further Investigation . . . 57

Appendix A Optimization Parameters **61**

Appendix B Convergence as Function of Sample Size D **63**

Appendix C Reference Solutions **65**
C.1 Normal Distributed Log-returns . . . 65
C.2 Student's t Distributed Log-returns . . . 66

Appendix D GARCH and GPD Parameters **69**

Chapter 1

Introduction

The goal of this thesis project is to introduce and solve an alternative robust portfolio optimization problem to Markowitz classical mean-variance optimization problem. Classical robust optimization focus on parameter uncertainty for a given distribution and the contribution of this thesis project is to extend the robust optimization to include uncertainties in log-return distribution as well. The alternative portfolio optimization problem is solved with both elliptical distributions and asymmetric log-return distributions, applied to a reference portfolio consisting of stocks and a bond index from the Swedish market.

The thesis project is organized as follows. In Chapter 1, I introduce general concepts of portfolio optimization and give a brief historical background on the development of modern portfolio optimization. A reference portfolio is then constructed and the traditional portfolio optimization problem by Markowitz [12] is solved to obtain a benchmark solution from historical data. The chapter concludes with arguments on why Markowitz mean-variance optimization problem is not particularly good in financial mathematics due to its narrow area of practical application.

In Chapter 2 I introduce risk measures and based on a risk measure commonly used in financial risk management called Expected Shortfall I formulate a problem better fit for modern portfolio optimization. With theorems provided by Rockafellar and Uryasev [15], the portfolio optimization problem is approximated as a convex linear program that can be solved with standard optimization algorithms. I then show that Markowitz mean-variance optimization problem is a special case of the portfolio optimization problem with Expected Shortfall, connected with a risk aversion coefficient.

Chapter 3 deals with optimization under uncertainties and the concept of robust optimization is presented, being central when solving optimization problems with uncertainty in parameters. My contribution to the literature of robust portfolio optimization is that I first perform a case study under different elliptical distributions and then study robust portfolio optimization with an asymmetric hybrid generalized Pareto-Empirical-Generalized Pareto distribution. To analyze the statistical uncertainty in the results, the bootstrap procedure is applied to calculate standard errors for the holdings and the Expected Shortfall.

In Chapter 4 I analyze the results obtained throughout the thesis project and compare them to the benchmark Markowitz solution. I analyze the properties of the optimization problem and conclude in which areas the optimization problem is particularly applicable, being advantageous over Markowitz mean-variance optimization problem. The thesis ends with remarks on areas where further investigation can be done.

1.1 Introduction to Portfolio Optimization

Suppose an investor has initial capital V_0 at time $t = 0$ that should be invested in a market consisting of risky assets. The aim of the investments is often to, at some future time $T = 1$, have gained at least as much money as the investor would have gained if the capital was invested in a risk-free asset with certain interest rate R_0 instead. Since the assets on the market have unknown future values, the investor cannot be certain of the future portfolio value V_1 and hence the investor is exposed to risk. To limit this risk, the investor might pose limitations on how much the future portfolio value is allowed to vary. The more risk averse the investor is, the smaller the expected value of the future portfolio typically is and hence there is a trade-off between expected future portfolio value and risk exposure.

Traditionally the log-returns

$$R_t = \log\left(\frac{S_t}{S_{t-1}}\right), \tag{1.1}$$

of an asset, S_t being the asset price at time t, are assumed to have the Markov property from day to day and are often assumed to be weakly dependent and close to independent and identically distributed. Therefore, by letting V_1 be a function of assets' log-returns, $V_1 = f(\mathbf{R}_1)$ for some multivariate function f, one may construct approximately independent copies of V_1 as $\{f(\mathbf{R}_{-n+1}), \ldots, f(\mathbf{R}_0)\}$. Hence, investors try to predict future portfolio value by analyzing historical log-returns. An alternative approach which is equally good is to work directly with the log-returns and maximize the expected future portfolio log-return since this is equivalent of maximizing the expected future portfolio value. This approach is used in this thesis project. Furthermore, since the amount of initial capital V_0 is only a scaling factor, it is common practice to set $V_0 = 1$ so that the solution is expressed as proportions of the total capital invested rather than expressed as monetary units.

1.1.1 Markowitz Mean-Variance Optimization Problem

Modern portfolio optimization was first introduced by Markowitz [12] in 1952 with what is often referred to as Markowitz mean-variance optimization problem. The idea is to maximize the expected return, subject to the constraint that the variance of the portfolio must be smaller than some predetermined tolerance level T. Assume an investor has the possibility to invest in n assets and let $\mathbf{R} = (R_1, \ldots, R_n)$ be a random vector of log-returns at time $t = 1$ corresponding to each of the assets as defined by (1.1). The covariance matrix of returns for the n assets is further given by Σ which is assumed to be symmetric and positive semi-definite and thus also invertible. Let $\mathbf{w} = (w_1, \ldots, w_n)$ be a vector of holdings, or weights, of the initial capital invested in each asset. If the future portfolio log-return is denoted by $X = \mathbf{w}^T\mathbf{R}$, then Markowitz mean-variance optimization problem can be stated

$$\begin{aligned} \max_{\mathbf{w}} \quad & \mathbb{E}[X] \\ \text{Subject to } & \mathbf{w}^T\Sigma\mathbf{w} \leq T \\ & \sum_{i=1}^{n} w_i = V_0. \end{aligned}$$

Depending on the value of T the optimal solution will differ and plotting the expected future portfolio log-return versus its standard deviation for different values of T gives what is known as the efficient frontier.

Markowitz mean-variance optimization problem can be stated in various ways and an alternative formulation is the mean-variance trade-off formulation

$$\begin{aligned} \max_{\mathbf{w}} \quad & \mathbf{w}^T\mu - \frac{c}{2V_0}\mathbf{w}^T\Sigma\mathbf{w} \\ \text{Subject to } & \sum_{i=1}^{n} w_i = V_0 \end{aligned} \tag{1.2}$$

where μ is the mean log-return vector and the trade-off parameter $c > 0$ is a dimensionless constant that should be interpreted as a risk aversion coefficient. The trade-off problem is convex and has analytical solution

$$\mathbf{w} = \frac{V_0}{c}\Sigma^{-1}\left(\mu - \frac{\max\{\mathbf{1}^T\Sigma^{-1}\mu - c, 0\}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\mathbf{1}\right).$$

Additional constraints can easily be added to Markowitz mean-variance optimization problem where lower and upper bounds on each weight w_i are often used. A particularly common constraint is to set the lower bounds of each weight to zero, implying that only long positions in the n assets are allowed. The optimization problem then becomes

$$
\begin{aligned}
\max_{\mathbf{w}} \quad & \mathbf{w}^T\mu - \frac{c}{2V_0}\mathbf{w}^T\Sigma\mathbf{w} \\
\text{Subject to} \quad & \sum_{i=1}^{n} w_i = V_0 \\
& w_i \geq 0, \quad i = 1, \dots, n.
\end{aligned}
\tag{1.3}
$$

Since the introduction of Markowitz mean-variance optimization problem, portfolio optimization has been developed with more complex problem formulations to improve the optimization further. This thesis project studies portfolio optimization of one of these alternative optimization problems, which do not rely on some assumptions that must hold in Markowitz mean-variance optimization problem. These assumptions are presented in Section 1.1.3.

Before presenting the alternative optimization problem a solution to (1.3) is calculated by applying the problem to a reference portfolio that will be used throughout the thesis project. The solution will work as a benchmark solution to compare future solution with to analyze the sensitivity of portfolio optimization.

1.1.2 Reference Portfolio and Benchmark Solution

To be able to solve Markowitz mean-variance optimization problem (1.3), historical log-return data from assets in a reference portfolio is needed so that estimates of μ and Σ can be calculated. This introduces two questions:

- Which assets are of interest in the reference portfolio?
- From which historical time period should the data be collected?

The answer depends on the investor. In this thesis project the reference portfolio consists of linear assets being stocks and a bond index from the Swedish market. The assets included in the reference portfolio are listed in Table 1.1 and are chosen to reflect some of the largest companies on the Swedish market, diversified over a large range of business areas. The idea is to include high and low volatile assets in different business sectors that are more and less dependent on the current state of economy. An investor should then have good control of deciding what risk level he is willing to be exposed to, depending on how the portfolio weights are chosen. The OMRX Total Bond Index, henceforth abbreviated the Total Bond Index, is included in the reference portfolio to have a position which can be considered close to riskless with small expected log-return.

The historical time period used for collecting asset price data is chosen to be January 2, 2007 until January 22, 2016. The time period includes the global 2007 financial crisis, having its peak between approximately 2007 – 2009 according to the time line in [8], where risky assets typically are more correlated, and some years afterwards where the market is starting to rise again and the assets are less correlated.

Table 1.1: Swedish assets included in the reference portfolio and their corresponding business areas.

Asset name	Business Area
AstraZeneca	Health Care
Ericsson A	Technology
Hennes & Mauritz B	Retail
ICA Gruppen	Retail
Nordea Bank	Banks
SAS	Travel & Leisure
SSAB A	Basic Resources
Swedish Match	Personal & Household Goods
TeliaSonera	Telecommunications
Volvo	Industrial Goods & Services
Total Bond Index	Government bonds

The historical data is obtained from the Nasdaq OMX web page and consists of daily prices for each of the assets in the reference portfolio[1]. Figure 1.1 depicts the historical price developments for the stocks to the left and the Total Bond Index to the right.

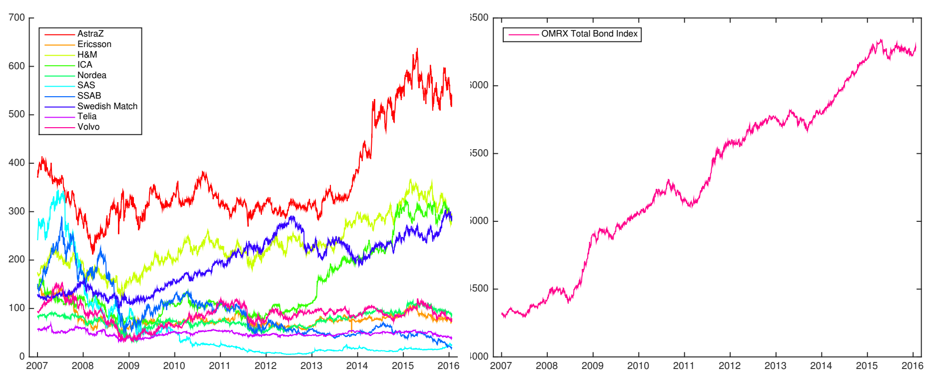

Figure 1.1: Price developent for the stocks (left) and the bond index (right) in the reference portfolio during January 2, 2007-January 22, 2016.

From Figure 1.1 one can make several observations. For instance that

[1]7 data points for OMRX Total Bond Index was missing in the historical data set and were linearly interpolated. The data is automatically adjusted for asset splits.

some companies managed the financial crisis better than others, for instance Nordea Bank, and that some assets have historically had a positive trend, such as Swedish Match, while other assets have had a negative trend, such as SAS. There are furthermore some assets that have changed very little in value during the time period of interest, for instance TeliaSonera, while other assets have been more volatile, for instance ICA Gruppen. All these behaviors were desired in the construction of the reference portfolio. Ultimately, the assets included in a portfolio is up to the investor to decide and the reference portfolio in this thesis project could have been selected differently. The general conclusions in the thesis project do however not directly depend on the reference portfolio itself other than the fact that the weight vector and Expected Shortfall would change if other assets were used.

The optimization problem (1.3) is a quadratic programming problem and can be solved with standard solving algorithms. The solution is presented in Table 1.2 when applied to the reference portfolio with risk aversion coefficient $c = 5.33$ (turns out later why this particular value was chosen), $V_0 = 1$ and empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$. The solution was calculated using the function `quadprog` in Optimization Toolbox in Matlab version 9.0 on a Core i5 CPU 2.60 GHz Laptop with 8 GB of RAM using the interior point method. See Appendix A for numerical values of the empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$.

Table 1.2: Benchmark solution to Markowitz mean-variance optimization problem (1.3) applied to the reference portfolio in Table 1.1 with empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$.

Asset name	Weight
AstraZeneca	0.0001
Ericsson A	0
Hennes & Mauritz B	0.0080
ICA Gruppen	0.0666
Nordea Bank	0
SAS	0
SSAB A	0
Swedish Match	0.1436
TeliaSonera	0
Volvo	0
Total Bond Index	0.7816

As can be seen from Table 1.2, most assets in the portfolio will not be invested in and about 78% of the initial capital is invested in the Total Bond Index, about 14% is invested in Swedish Match, approximately 7% is invested in ICA Gruppen and the remaining 1% is invested in Hennes & Mauritz B. With the benchmark solution $\mathbf{w}_{BM}$, the expected daily future portfolio log-return is

$$\mathbb{E}[\mathbf{w}_{BM}^T \mathbf{R}] = \mathbf{w}_{BM}^T \widehat{\mu} = 2.0242 \cdot 10^{-4} \tag{1.4}$$

corresponding to a yearly[2] expected portfolio log-return of $251 \cdot \mathbf{w}_{BM}^T \widehat{\mu} \approx 5.08\%$ and the daily portfolio variance is

$$\sigma_p^2 = \mathbf{w}_{BM}^T \widehat{\Sigma} \mathbf{w}_{BM}^T = 7.8354 \cdot 10^{-6}, \tag{1.5}$$

meaning that the yearly portfolio variance is approximately 0.01%.

1.1.3 Stylized Assumptions in Markowitz Optimization Problem

Markowitz mean-variance optimization problem (1.3) is theoretically important, but when applied to real financial portfolios it suffers from some stylized assumptions that must hold for the problem formulation to be applicable. The first stylized assumption is that the log-returns have elliptical distribution, which is not accurate enough when studying historical data, especially during times of financial crisis where asset prices become more volatile and more correlated. Furthermore, elliptical distributions are symmetric which often is not the case for historical log-returns where either the loss or the gain tail of the distribution is more prominent than the other tail. Therefore, parts of the empirical log-return distribution information is lost when modeling with elliptical distributions. The second stylized assumption is that the portfolio is linear, which is a major limitation. Often highly non-linear assets such as options or other financial derivatives are included in financial portfolios, that cannot be taken into consideration by Markowitz mean-variance optimization problem. With these drawbacks taken into account, alternative portfolio optimization problems have recently become popular to study. This thesis project focus on the first drawback discussed but the problem can easily be extended to include non-linear assets in the reference portfolio as well.

[2]There are about 251 trading days in a year on the Swedish market.

Chapter 2

Portfolio Optimization with Expected Shortfall

With the introduction to portfolio optimization in Chapter 1 I have so far concluded that Markowitz mean-variance optimization problem (1.3) is not sufficient to employ in portfolio optimization unless the log-returns are elliptically distributed, but due to its historical importance it may be used as a benchmark problem. In this chapter, I introduce the concept of risk measures and construct an alternative optimization problem based on a risk measure better fit for modern portfolio optimization. The aim is to construct an appropriate portfolio optimization problem that does not rely on the stylized assumptions that must hold for Markowitz mean-variance optimization problem to be considered good.

2.1 Risk Measures

Financial risk can be measured in several different ways, but it is often desirable to measure risk in monetary units, meaning that the risk is expressed as buffer capital that needs to be added to a portfolio to protect it from undesired outcomes. In Markowitz mean-variance optimization problem, the risk is measured as the variance of the future portfolio return. However, variance is not considered to be a very good risk measure in finance since it is defined as the expected squared deviation from the mean value and thus does not make difference between positive deviations, portfolio gain, and negative deviations, portfolio loss. Furthermore, standard deviation can only be considered accurate enough to be translated into monetary risk if the future value of the portfolio value is approximately normal distributed. This is often a too strict assumption that simplifies the real portfolio log-return distribution too much. Instead, it is often desirable to utilize a risk measure that make difference between good and bad deviations from the expected future portfolio value. In this section I first present basic risk measure theory and later specify two risk measures widely used in financial risk management.

Generally, let $\rho(X)$ be a function measuring the risk of a stochastic variable X. Different risk measures have different properties and below is a list of such mathematical properties that are considered to be useful or desirable, together with brief explanations on how they can be interpreted.

1. **Translation invariance**. $\rho(cR_0 + X) = -c + \rho(X)$ for $c \in \mathcal{R}$. This means that adding the amount c with risk-free interest rate R_0 to a portfolio reduces the risk equally much.

2. **Monotonicity**. If $X_2 < X_1$, then $\rho(X_1) \leq \rho(X_2)$. This means that if we know for certain that a portfolio X_1 is greater than a portfolio X_2 at a future time, then the first portfolio is considered to be less risky.

3. **Convexity**. $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$, for any real $\lambda \in [0, 1]$. The risk measure rewards diversification, which means that it takes into account that it often is wise to spread out the investment in several risky positions, rather than investing everything in one.

4. **Normalization**. $\rho(0) = 0$. This means that it is acceptable to not invest in risky assets at all, so that an empty portfolio is riskless.

5. **Positive homogeneity**. $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda \geq 0$. This means that for instance investing twice as much in one position is twice as risky.

6. **Subadditivity**. $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$. This property should also be interpreted as that the risk measure rewards diversification. A company consisting of two business units is interpreted as less risky compared to the two units considered as separate companies.

The reader is refered to the book [5, Ch. 6] by Hult, Lindskog, Hammarlid and Rehn for a more thorough presentation on general risk measure theory and more comments on the above properties, as well as more on why variance is considered a bad risk measure in finance.

A risk measure with the properties translation invariance and monotonicity is said to be a monetary measure of risk, and a risk measure considered to replace variance in Markowitz mean-variance optimization problem should satisfy at least these two properties. A risk measure that in addition to translation invariance and monotonicity also satisfies convexity is said to be a convex risk measure. The convex risk measure family is thus a subset of the monetary risk measure family. Finally, a third risk measure family is the coherent risk measure, where $\rho(X)$ satisfies the properties translation invariance, monotonicity, positive homogeneity and subadditivity. It is easy to see that a risk measure satisfying positive homogeneity also satisfies normalization by setting $\lambda = 0$. It can further be shown that positive homogeneity and convexity together implies subadditivity but not the reverse. Hence, a coherent risk measure is also a convex risk measure but the opposite does not generally hold, so the coherent risk measure family is a subset of the convex risk measure family and thus also a subset of the monetary risk measure family. When choosing an appropriate risk measure for a portfolio optimization problem replacing Markowitz mean-variance optimization problem one may therefore consider both convex and coherent risk measures to be at least as good as monetary risk measures. Below, two risk measures that are commonly used in risk management are presented that are considered good in financial mathematics.

2.1.1 Value-at-Risk

The first risk measure presented is the Value-at-Risk, abbreviated VaR. Value-at-Risk satisfies translation invariance, monotonicity and positive homogeneity and is hence a monetary risk measure. The Value-at-Risk at level $p \in (0,1)$ of a stochastic variable X is defined as

$$VaR_p(X) = \min\{m \in \mathbb{R} : P(mR_0 + X < 0) \leq p\}.$$

If X is assumed to have right continuous and increasing cumulative distribution function $F(x)$, which will be the case in this thesis project, it follows that

$$\begin{aligned} VaR_p(X) &= \min\{m \in \mathbb{R} : P(-X > mR_o) \leq p\} \\ &= \min\{m \in \mathbb{R} : P(L \leq m) \geq 1-p\} = F_L^{-1}(1-p) \end{aligned}$$

where $L = -X/R_0$ should be interpreted as the discounted loss. Hence, the Value-at-Risk of the stochastic variable X is the $(1-p)$-level quantile of the associated discounted loss L.

A direct advantage with Value-at-Risk over traditional variance is that it can be used when variance is not a relevant risk measure, for instance when the expected value of a distribution does not represent a fair picture of the distributional appearance. Furthermore, since Value-at-Risk computes the $(1-p)$-level quantile of the discounted loss it accounts only for large losses but not large gains and hence Value-at-Risk makes difference between negative and positive deviations from the expected future portfolio value. However, since Value-at-Risk only look at a particular level $(1-p)$ of the loss quantile, traders can take advantage of this to "hide" risky investments by making the losses more extreme so that the potential loss ends up further out in the loss distribution tail so that it is not discovered by Value-at-Risk. With this strategy, very risky portfolios could seem acceptable that otherwise would not have been if the risk was visible for risk managers. This might result in that companies are exposed to extremely large loss scenarios, although with relatively small probability, but with the possible outcome that the company suffers a huge loss and possibly bankruptcy.

2.1.2 Expected Shortfall

A better risk measure than Value-at-Risk in the sense that it takes into account all losses located in the whole $(1-p)$-level quantile tail of the loss distribution is Expected Shortfall. For a stochastic variable X, the Expected Shortfall, abbreviated ES, at level $p \in (0,1)$ is defined as

$$ES_p(X) = \frac{1}{p}\int_0^p VaR_u(X)du = \frac{1}{1-q}\int_q^1 VaR_u(L)du, \qquad (2.1)$$

where $q = 1-p$. Expected Shortfall appears under several different names in literature and with minor changes it is goes under the names Conditional Value-at-Risk (CVaR), Average Value-at-Risk (AVaR), Tail Value-at-Risk (TVaR), Expected Tail Loss (ETL) and Tail Conditional Expectation (TCE). Similarly to Value-at-Risk, if X has right continuous and increasing distribution function $F(x)$ then the Expected Shortfall has representation

$$ES_p(X) = \frac{1}{p}\int_0^p F_L^{-1}(1-u)du.$$

Since Expected Shortfall is defined through Value-at-Risk, it inherits the properties of Value-at-Risk, being translation invariance, monotonicity and positive homogeneity. It can further be shown that Expected Shortfall also satisfies subadditivity and is hence a coherent risk measure.

2.1.3 Contribution Expected Shortfall

In addition to the total portfolio risk, the investor might be interested in how much each asset in the reference portfolio contributes to the total risk when allocating the capital. The contributions go under the names Marginal Risk, Contribution Risk and Euler Allocations/Decomposition and originate from Euler's decomposition theorem. The theorem states that a function f of n variables $x_1, \ldots, x_n$ with continuous partial derivatives is homogenous of degree k if and only if the equation

$$\sum_{i=1}^n x_i \frac{\partial f}{\partial x_i} = kf$$

holds for all $x_1, \ldots, x_n \in \mathcal{D}$ being an open domain. There exists several different approaches for decomposing the total portfolio risk but a natural decomposition given a homogenous risk measure would based on Euler's decomposition theorem be

$$\rho(X) = \sum_{i=1}^n w_i \frac{\partial \rho(X)}{\partial w_i}. \qquad (2.2)$$

This is natural since exactly the total risk is decomposed, meaning that the sum of all contributions is exactly the total risk. Expected Shortfall is (positive) homogenous of degree 1 and a decomposition of the total portfolio Expected Shortfall is

$$ES_p(X) = \sum_{i=1}^{n} U_i \frac{\partial ES_p(X)}{\partial U_i}$$

where $\mathbf{U} = (U_1, \dots, U_n)^T$ is a vector of sensitivities defined by

$$U_i = \sum_{j=1}^{n} w_j \frac{\partial F_j(\mathbf{S}(t))}{\partial \mathbf{S}_i(t)},$$

w_j being the asset weight and $F_j(\mathbf{S}(t))$ is a financial instrument. The derivative $\partial ES_p(X)/\partial U_i$ should then be interpreted as the Euler Allocation for asset i.

Skoglund and Chen show in [18] that if the log-returns are assumed to have multivariate normal distribution, the Euler Allocations can be calculated analytically as

$$\frac{\partial ES_p(X)}{\partial \mathbf{U}} = \lambda(\Phi^{-1}(q)) \frac{\partial \sigma_p}{\partial \mathbf{U}}. \tag{2.3}$$

Here $\lambda(x)$ is the hazard function for the normal distribution, defined as

$$\lambda(x) = \frac{\phi(x)}{1 - \Phi(x)}. \tag{2.4}$$

and

$$\frac{\partial \sigma_p}{\partial \mathbf{U}} = (\mathbf{U}^T \Sigma \mathbf{U})^{-1/2} \Sigma \mathbf{U}.$$

If the log-returns are modeled by their empirical distribution, the derivative $\partial ES_p(X)/\partial \mathbf{U}$ does not exist since the portfolio distribution is discrete. In this case, we may use the results of Tasche [21] and approximate the derivative as

$$D_{ES_p(X)}(w_i) = \frac{\sum_{d=1}^{D} \mathbb{E}[\mathbf{L}_{i,\cdot} | L_d > L_s]}{\sum_{d=1}^{D} \mathbf{1}_{(L_d > L_s)}} \tag{2.5}$$

where $\{L_{i,d}\}_{i=1,d=1}^{n,D}$ are the loss components, n being the number of assets and D the amount of available loss scenarios, $\mathbf{L}_{i,\cdot} = (L_{i,1}, \dots, L_{i,D})'$, L_s is the loss sample point corresponding to the p-level Value-at-Risk and

$$\mathbf{1}_{(L_d > L_s)} = \begin{cases} 1, & L_d > L_s \\ 0, & \text{otherwise.} \end{cases}$$

In this thesis project the investment horizon is one day and the risk-free interest rate can be approximated to zero, which implies that $L_i = -R_i$, i.e. the log-returns with switched sign. With linear assets, the portfolio loss scenarios are then obtained by multiplying the weight vector with the loss vector, $L = \mathbf{w}^T\mathbf{L} = -\mathbf{w}^T\mathbf{R}$. Furthermore, by ordering the portfolio losses such that $L_1 > L_2 > \cdots > L_D$ then the portfolio Value-at-Risk at level p, defined as the p-level loss quantile, is

$$VaR_p(X) = L_s = L_{[Dp]+1} \tag{2.6}$$

where $[Dp]$ is the integer part of Dp.

2.1.4 Properties of the quantile function

Since the quantile function turns out to play a central role in both Value-at-Risk and Expected Shortfall, I here state two properties of the quantile function that will be used later.

Proposition 2.1 *If $g : \mathbb{R} \to \mathbb{R}$ is a nondecreasing function and left continuous, then for any random variable Z it holds that $F^{-1}_{g(Z)}(p) = g(F^{-1}_Z(p))$ for all $p \in (0, 1)$.*

Proposition 2.2 *For any random variable X with continuous and strictly increasing density function F_X, $F^{-1}_{-X}(p) = -F^{-1}_X(1 - p)$ for all $p \in (0, 1)$.*

I refer the reader to Hult et al.[5, pp. 170-172] for complete proofs.

2.2 Portfolio Optimization Problem Formulation

In optimization, we are generally interested in finding the global extreme point to a function under some constraining requirements. Convex optimization problems is a particularly nice family of optimization problems where there exists a unique optimal solution being the global extreme point. Since Expected Shortfall is a coherent risk measure, it is also convex and hence an appealing risk measure to employ in financial optimization problems, opposite to Value-at-Risk which does not satisfy the convexity property and may therefore produce several local extreme points. With this in mind, it becomes natural to use Expected Shortfall rather than Value-at-Risk as risk measure in a portfolio optimization problem. Additionally, since Expected Shortfall does not demand elliptically distributed log-returns or linear portfolios, it does not rely on the stylized assumptions that must be assumed in Markowitz mean-variance optimization problem. In this section I formulate the portfolio optimization problem based on minimizing Expected Shortfall that will be the subject of study in the remaining part of this thesis project.

Simply minimizing the Expected Shortfall in a portfolio optimization problem is trivial without constraining equations and in a market consisting of only risky assets the solution would be to not invest at all. Therefore, a reasonable initial constraint would be that the entire initial capital V_0 must be invested in the market. Additionally, it is common to constrain the asset allocations to long positions, i.e. $w_i \geq 0$, $i = 1, \ldots, n$. The portfolio optimization problem can then be stated mathematically as

$$\begin{aligned} \min_{\mathbf{w}} \quad & ES_p(\mathbf{w}^T \mathbf{R}) \\ \text{Subject to } & \sum_{i=1}^{n} w_i = V_0 \\ & w_i \geq 0, \quad i = 1, \ldots, n, \end{aligned}$$

where n is the number of assets available in the portfolio.

Often, investors are not interested in only minimizing the risk, but they want some reward for exposing their capital to risk. Therefore, investors add the constraint that the future expected portfolio (log)-return should be larger some threshold (log-)return θ. The optimization problem then becomes

$$\begin{aligned} \min_{\mathbf{w}} \quad & ES_p(\mathbf{w}^T \mathbf{R}) \\ \text{Subject to } & \mathbf{w}^T \mu \geq \theta \\ & \sum_{i=1}^{n} w_i = V_0 \\ & w_i \geq 0, \quad i = 1, \ldots, n. \end{aligned} \tag{2.7}$$

The above problem formulation will be the foundation of this thesis project. Next I remark on a special case when (2.7) have the same solution as the benchmark solution to Markowitz mean-variance optimization problem and then I state a few general model assumptions that is used when later solving (2.7) numerically. In Section 2.3 I show that the portfolio optimization problem can be approximated and solved by ordinary linear programming.

In section 2.1.3 it was said that a natural decomposition of the total portfolio risk is obtained by calculating the Euler Allocations since then the risk contributions sum up to exactly the total risk. But this property can be achieved for any risk decomposition method by simply normalizing. However, Tasche proves in [20] that Euler decomposition is the only risk decomposition method that is consistent with (local) portfolio optimization which motivates the benefits of using Euler decomposition in this thesis project.

In portfolio optimization the Euler Allocations have a second field of application in addition to measuring the risk contribution from a specific asset to the total portfolio risk. The famous Sharpe ratio, introduced by Sharpe [17], relates the expected portfolio (log)-return to the risk as

$$S = \frac{\mathbf{w}^T \mu}{\rho} \tag{2.8}$$

and the larger Sharpe ratio the better is the investment. Similarly to the decomposition of the portfolio risk a natural decomposition of the Sharpe ratio is to define

$$S_i^* = \frac{w_i \mu_i}{w_i \frac{\partial \rho}{\partial w_i}} \tag{2.9}$$

as the marginal Sharpe ratio of asset i, $i = 1, \ldots, n$. An asset should then be considered a good investment if the marginal Sharpe ratio is large. With the marginal Sharpe ratio as argument for investing in an asset or not, it can be seen that the only information an investor requires for (local) portfolio optimization is the expected log-return and Euler Allocation. This observation will be used later when attempting to analyze the optimal solutions to the portfolio optimization problem with Expected Shortfall.

2.2.1 Remark: Elliptically distributed log-returns

The following remark points out a special case where portfolio optimization problem (2.7) can be simplified to a modified version of Markowitz mean-variance problem (1.3). This is important since the solution to (2.7) should in this case be the same as the benchmark solution in Table 1.2.

Consider a scenario where the vector of log-returns $\mathbf{R}$ is assumed to be elliptically distributed with mean μ and covariance matrix Σ. Then $\mathbf{R}$ can be represented as

$$\mathbf{R} \stackrel{d}{=} \mu + WA\mathbf{Z}$$

where A is a matrix such that $AA^T = \Sigma$ and $\mathbf{Z} \in N(\mathbf{0}, I_d)$, I_d being the d-dimensional identity matrix, is independent of $W \geq 0$. The portfolio log-return is then

$$\mathbf{w}^T\mathbf{R} = \mathbf{w}^T\mu + \sqrt{\mathbf{w}^T\Sigma\mathbf{w}}WZ_1 \tag{2.10}$$

where I have used a standard property for spherically distributed random variables and Z_1 is univariate standard normal distributed. Minimizing the Expected Shortfall then yields

$$\begin{aligned}
\min_{\mathbf{w}} ES_p(\mathbf{w}^T\mathbf{R}) &= \min_{\mathbf{w}} -\frac{1}{p}\int_0^p F^{-1}_{\mathbf{w}^T\mathbf{R}}(u)du \\
&= \min_{\mathbf{w}} -\mathbf{w}^T\mu - \sqrt{\mathbf{w}^T\Sigma\mathbf{w}}\frac{1}{p}\int_0^p F^{-1}_{WZ_1}(u)du
\end{aligned}$$

where I have used both Proposition 2.1 and Proposition 2.2. Now, since minimizing $-f(\mathbf{w})$ corresponds to maximizing $f(\mathbf{w})$, portfolio optimization problem (2.7) with elliptically distributed log-returns can be written

$$
\begin{aligned}
\max_{\mathbf{w}} \quad & \mathbf{w}^T\mu - \frac{c}{2V_0}\sqrt{\mathbf{w}^T\Sigma\mathbf{w}} \\
\text{Subject to } & \mathbf{w}^T\mu \geq \theta \\
& \sum_{i=1}^{n} w_i = V_0 \\
& w_i \geq 0, \quad i = 1, \ldots, n,
\end{aligned}
$$

where

$$c = -\frac{2V_0}{p}\int_0^p F^{-1}_{WZ_1}(u)du. \tag{2.11}$$

Note the similarity between the above special case and Markowitz mean-variance optimization problem (1.3). Optimization problem (2.7) can in the special case considered hence be rewritten as an optimization problem with solution equivalent to that of the standard quadratic optimization problem.

2.2.2 General Problem Assumptions

This section presents some assumptions that need to be specified for the portfolio optimization problem (2.7) to be numerically solvable. The assumptions stated are used when solving the portfolio optimization problem unless otherwise explicitly stated.

Assumption 1 The investor's initial capital $V_0 = 1$. This was already assumed when calculating the benchmark solution and is used to simplify the analysis work of the solutions since the asset weights w_i can be seen as proportions of the whole initial capital rather than monetary weights. The solution is then easily interpreted for arbitrary initial capital V_0.

Assumption 2 The portfolio lives for one day. This means that the goal is to invest optimally for one day ahead, which is convenient since the historical data consists of daily log-returns calculated using (1.1).

Assumption 3 $p = 0.01$ ($q = 0.99$). Banks often use the p-value $p = 0.005$ but other common levels are $p = 0.01$ and $p = 0.05$. The smaller value p, the more unlikely it is to be exposed to a loss larger than the calculated Expected Shortfall. $p = 0.01$ means that with a portfolio life of one day, one would expect the loss to be larger than the Expected Shortfall in 1 out of every 100 days. Plugging in $p = 0.01$ in (2.11) yields with Z_1 being standard normal distributed and $W = 1$

$$c = -\frac{2}{0.01}\int_0^{0.01} \Phi^{-1}(u)du = \frac{2\phi(\Phi^{-1}(0.01))}{0.01} = 5.33, \tag{2.12}$$

where $\phi(x)$ denotes the probability density function of the standard normal distribution. This explains why the risk aversion coefficient $c = 5.33$ was used earlier in the benchmark solution.

Assumption 4 The threshold for acceptable expected daily portfolio log-return is $\theta = 2.0242 \cdot 10^{-4}$ corresponding to 5.08% yearly log-return. This is the same expected portfolio log-return as for the benchmark solution calculated in (1.4). The value of θ is ultimately up to the investor to decide, being a trade-off between return and risk, where larger θ could expose the investor to larger risk. The impact on the optimal solution of varying θ is analyzed in Section 4.1.

When analyzing the Expected Shortfall of a portfolio it is common to look at the relative Expected Shortfall defined as the ratio

$$ES_{\text{rel}} = \frac{ES}{\text{Portfolio market value}}.$$

With Assumption 1 the portfolio market value is 1 implying that in this thesis project the relative Expected Shortfall equals the Expected Shortfall calculated when solving the portfolio optimization problem (2.7) which simplifies the analysis.

2.3 Rockafellar and Uryasev simplification

Since the definition of Expected Shortfall involves an integral of the Value-at-Risk, being the quantile function of the loss distribution, it is difficult to work directly with and optimize (2.1). Rockafellar and Uryasev provide in [15] two theorems that simplify problem formulation (2.7) much. This section presents this simplification.

The key is to characterize $\text{ES}_p(X)$ and $\text{VaR}_p(X)$ with the auxiliary function $H_p(\mathbf{w}, \alpha)$ on $\mathcal{W}, \mathbb{R}^m$ defined by

$$H_p(\mathbf{w}, \alpha) = \alpha + \frac{1}{1-q} \int\limits_{\mathbf{R} \in \mathbb{R}^m} [L(\mathbf{w}, \mathbf{R}) - \alpha]^+ f(\mathbf{R}) d\mathbf{R} \tag{2.13}$$

where

$$[x]^+ = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

and $f(\mathbf{R})$ is the probability density function of the log-return vector $\mathbf{R}$. The theorems regarding the characterization are central in this thesis project and are stated below.

Theorem 2.3 *As a function of α, $H_p(\mathbf{w},\alpha)$ is convex and continuously differentiable. The ES_p of the loss associated with any $\mathbf{w} \in \mathcal{W}$ can be determined from the formula*

$$ES_p(X) = \min_{\alpha\in\mathbb{R}} H_p(\mathbf{w},\alpha). \tag{2.14}$$

In this formula, the set consisting of the values α for which the minimum is attained, namely

$$A_p(\mathbf{w}) = \operatorname*{argmin}_{\alpha\in\mathbb{R}} H_p(\mathbf{w},\alpha),$$

is a nonempty and closed bounded interval (perhaps reducing to a single point), and the VaR_p of the loss is given by

$$VaR_p(X) = \textit{left endpoint of } A_p(\mathbf{w}). \tag{2.15}$$

In particular, one always has

$$VaR_p(X) = \operatorname*{argmin}_{\alpha\in\mathbb{R}} H_p(\mathbf{w},\alpha), \quad ES_p(X) = H_p(\mathbf{w}, VaR_p(\mathbf{w})).$$

The theorem relies on the assumption that the cumulative loss distribution function F_L is continuous.

Proof. $H_p(\mathbf{w},\alpha)$ is convex from its definition (2.13) and has derivative

$$\frac{\partial}{\partial\alpha} H_p(\mathbf{w},\alpha) = 1 + \frac{1}{1-q}\left(F_L(\mathbf{w},\alpha) - 1\right) = \frac{1}{1-q}\left(F_L(\mathbf{w},\alpha) - q\right).$$

Therefore, the values of α that minimize $H_p(\mathbf{w},\alpha)$, i.e. the set $A_p(\mathbf{w})$, are those for which $F_L(\mathbf{w},\alpha) = q$. These values form a non-empty and closed interval since $F_L(\mathbf{w},\alpha)$ is continuous and nondecreasing with limits 0 as $\alpha \to -\infty$ and 1 as $\alpha \to \infty$. This proves (2.15). Furthermore,

$$\min_{\alpha\in\mathbb{R}} H_p(\mathbf{w},\alpha) = H_p(\mathbf{w},\alpha_q(\mathbf{w})) = \alpha_q(\mathbf{w}) + \frac{1}{1-q} \int\limits_{\mathbf{R}\in\mathbb{R}^m} [L(\mathbf{w},\mathbf{R}) - \alpha_q(\mathbf{w})]^+ f(\mathbf{R}) d\mathbf{R},$$

and the integral is simplified as

$$\begin{aligned}
\int\limits_{\mathbf{R}\in\mathbb{R}^m} [L(\mathbf{w},\mathbf{R}) - \alpha_q(\mathbf{w})]^+ f(\mathbf{R}) d\mathbf{R} &= \int\limits_{L(\mathbf{w},\mathbf{R})\geq\alpha_q(\mathbf{w})} [L(\mathbf{w},\mathbf{R}) - \alpha_q(\mathbf{w})] f(\mathbf{R}) d\mathbf{R} \\
&= \int\limits_{L(\mathbf{w},\alpha)\geq\alpha_q(\mathbf{w})} L(\mathbf{w},\alpha) f(\mathbf{R}) d\mathbf{R} - \alpha_q(\mathbf{w}) \int\limits_{L(\mathbf{w},\alpha)\geq\alpha_q(\mathbf{w})} f(\mathbf{R}) d\mathbf{R} \\
&= (1-q) ES_p(X) - \alpha_q(\mathbf{w})\left(1 - F_L(\mathbf{w},\alpha_q)\right),
\end{aligned}$$

where in the last equality I have used the definition of ES_p. Moreover, $F_L(\mathbf{w},\alpha_q) = q$ thus yielding

$$\min_{\alpha\in\mathbb{R}} H_p(\mathbf{w},\alpha) = ES_p(X),$$

which proves equation (2.14) and completes the proof. □

Theorem 2.4 *Minimizing the ES_p of the loss associated with all $\mathbf{w} \in \mathcal{W}$ is equivalent to minimizing $H_p(\mathbf{w}, \alpha)$ over all $(\mathbf{w}, \alpha) \in \mathcal{W} \times \mathbb{R}$, in the sense that*

$$\min_{\mathbf{w} \in \mathcal{W}} ES_p(X) = \min_{(\mathbf{w},\alpha) \in \mathcal{W} \times \mathbb{R}} H_p(\mathbf{w}, \alpha),$$

where, moreover, a pair $(\mathbf{w}^*, \alpha^*)$ achieves the second minimum if and only if $\mathbf{w}^*$ achieves the first minimum and $\alpha^* \in A_p(\mathbf{w}^*)$. In particular, therefore, in circumstances where the interval $A_p(\mathbf{w}^*)$ reduces to a single point (as is typical), the minimization of $H(\mathbf{w}, \alpha)$ over $(\mathbf{w}, \alpha) \in \mathcal{W} \times \mathbb{R}$ produces a pair $(\mathbf{w}^*, \alpha^*)$, not necessarily unique, such that $\mathbf{w}^*$ minimizes ES_p and α^* gives the corresponding VaR_p.

Furthermore, $H_p(\mathbf{w}, \alpha)$ is convex with respect to $(\mathbf{w}, \alpha)$, and ES_p is convex with respect to $\mathbf{w}$, when $L(\mathbf{w}, \mathbf{R})$ is convex with respect to $\mathbf{w}$, in which case, if the constraints are such that $\mathcal{W}$ is convex, the joint minimization is an instance of convex programming.

Proof. The initial claims in the first section follows directly from Theorem 1 and by realizing that minimization of $H_p(\mathbf{w}, \alpha)$ with respect to $(\mathbf{w}, \alpha) \in \mathcal{W} \times \mathbb{R}$ can be carried out by first minimizing over $\alpha \in \mathbb{R}$ for fixed $\mathbf{w}$ and then minimizing over $\mathbf{w} \in \mathcal{W}$.

Proving the claim in the second section starts by the observation that $H_p(\mathbf{w}, \alpha)$ is convex with respect to $(\mathbf{w}, \alpha)$ when $[L(\mathbf{w}, \mathbf{R}) - \alpha]^+$ is convex. Since a decomposition of two convex functions is convex, this is true when $L(\mathbf{w}, \alpha)$ is convex with respect to $\mathbf{w}$. The convexity of ES_p follows from the fact that minimizing an extended real-valued convex function of two variables with respect to one of these variables results in a convex function of the remaining variable or by recalling that Expected Shortfall is a coherent risk measure and thus satisfies the convexity property. □

From Theorem 2.3 and 2.4 it follows that instead of minimizing Expected Shortfall directly through its definition (2.1) one can equivalently minimize $H_p(\mathbf{w}, \alpha)$ defined by (2.13). With convex loss function $L(\mathbf{w}, \mathbf{R})$ this is particularly nice since then the optimization problem becomes a convex program.

By sampling D samples from the probability density function $f(\mathbf{R})$ of the return vector, Rockafellar and Uryasev then argues that the integral in (2.13) can be approximated with the sum

$$H_p(\mathbf{w}, \alpha) \approx \tilde{H}_p(\mathbf{w}, \alpha) = \alpha + \frac{1}{(1-q)D} \sum_{d=1}^{D} [L(\mathbf{w}, \mathbf{R}_d) - \alpha]^+$$

which is convex and piecewise linear in α.

Finally, by introducing auxiliary variables z_d, $d = 1, \ldots, D$ the original problem formulation (2.7) can be approximated as

$$
\begin{aligned}
&\min_{\mathbf{w},\alpha} \quad \alpha + \frac{1}{(1-q)D}\sum_{d=1}^{D} z_d \\
&\text{Subject to } \mathbf{w}^T\mu \geq \theta \\
&\qquad\qquad z_d \geq 0, \quad d = 1,\dots,D \\
&\qquad\qquad -L(\mathbf{w},\mathbf{R}_d) + \alpha + z_d \geq 0, \quad d = 1,\dots,D \\
&\qquad\qquad \sum_{i=1}^{n} w_i = V_0 \\
&\qquad\qquad w_i \geq 0, \quad i = 1,\dots,n.
\end{aligned}
\tag{2.16}
$$

where n is as always the number of assets available in the reference portfolio. The above problem formulation is a standard result in portfolio optimization with Expected Shortfall and is for instance presented by Skoglund and Chen in [19, pp. 156-157]. The problem is a convex linear program with standardized numerical solving algorithms such as the Simplex or interior point method [3, Ch. 5,10], when $L(\mathbf{w},\mathbf{R})$ is convex. This is for instance true for the reference portfolio used in this thesis project, where $L(\mathbf{w},\mathbf{R}) = -\mathbf{w}^T\mathbf{R}$ is linear.

2.4 Application on Reference Portfolio

In this section portfolio optimization problem (2.16) is solved applied to the reference portfolio in the special case where the solution theoretically should equal the benchmark solution. To arrive at this scenario, the parameters in (2.16) must be cleverly chosen to coincide with the parameters in the benchmark problem formulation (1.3).

Since the benchmark solution relies on the assumption of multivariate normal distributed log-returns, this must be assumed in this case as well. To compute loss scenarios $L(\mathbf{w},\mathbf{R}_d)$, I should hence simulate log-return vectors $\mathbf{R}$ from a multivariate normal distribution with parameters μ and Σ estimated from the historical data presented graphically in Figure 1.1. Also, $q = 0.99$ since $p = 0.01$. For the sum to be approximately equal to the integral in (2.13), I should choose D big. In this application I use $D = 15,000$[1]. The solution presented in Table 2.1 was calculated using the function `linprog` in Optimization Toolbox in Matlab version 9.0 on a Core i5 CPU 2.60 GHz Laptop with 8 GB of RAM with the empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$. The last row corresponds to the total portfolio Expected Shortfall the investor is exposed to if investing according to the optimal solution.

[1]In Appendix B I evaluate how large D must be for the simulated solutions to converge and have a narrow confidence interval.

Table 2.1: Solution to (2.16) with log-returns simulated from multivariate normal distribution with empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$.

Asset name	Weight
AstraZeneca	0
Ericsson A	0
Hennes & Mauritz B	0.0066
ICA Gruppen	0.0693
Nordea Bank	0
SAS	0
SSAB A	0
Swedish Match	0.1422
TeliaSonera	0
Volvo	0
Total Bond Index	0.7819
Expected Shortfall	0.0072

By comparing the optimal solution in Table 2.1 to the benchmark solution in Table 1.1 it is seen that the solutions are almost the same. The small differences in the weights exist due to numerical simulations of log-returns and from the approximation of the integral to a sum.

Skoglund and Chen present under the assumption of normal distributed log-returns two equations in [19, pp. 28-29] to transform the portfolio variance to Value-at-Risk according to

$$VaR_p(X) = \Phi^{-1}(q)\sigma_p \tag{2.17}$$

where $\Phi^{-1}(x)$ is the normal quantile function and portfolio variance to Expected Shortfall as

$$ES_p(X) = \sigma_p \lambda \left(\frac{VaR_p(X)}{\sigma_p} \right). \tag{2.18}$$

where $\lambda(x)$ is again the hazard function for the normal distribution defined in (2.4). Hence, by recalling the portfolio variance for the benchmark solution calculated in (1.5), the benchmark Value-at-Risk becomes

$$VaR_{0.01}(\mathbf{w}_{BM}^T \mathbf{R}) = 0.0065$$

and the benchmark Expected Shortfall is

$$ES_{0.01}(\mathbf{w}_{BM}^T \mathbf{R}) = 0.0075.$$

The Expected Shortfall by investing according to Table 2.1 is close to the benchmark Expected Shortfall. Therefore I can verify the claim that in the special case of multivariate normal distributed log-returns, portfolio optimization problem (2.7) is indeed identical to Markowitz mean-variance optimization problem (1.3). Furthermore, since the two optimization problems are connected only by the risk aversion coefficient c, calculated as (2.11) for all elliptical distributions, it is clear that Markowitz mean-variance optimization problem is a special case of portfolio optimization with Expected Shortfall for all elliptical distributions. Since portfolio optimization with Expected Shortfall does not rely on the stylized assumptions that Markowitz mean-variance optimization problem requires, (2.7) is a more general problem formulation that can be applied to more general investment situations with possibility to include non-linear assets or model the log-returns with asymmetric distributions for instance. In this thesis project I focus on the modeling of asset log-returns with non-elliptical distributions.

Chapter 3

Portfolio Optimization Under Uncertainty

Recall that the future portfolio log-return is a linear function of portfolio weights w_i and random log-returns R_i, $i = 1, \ldots, n$. Since the set of historical log-returns is finite, there will be uncertainty in estimated parameters such as the expected log-return vector and covariance matrix. Parameter uncertainty might in turn cause large errors in the final portfolio decision deviating much from the true optimal asset allocations. Parameter uncertainty can be thought of as a sub class to the larger family of model uncertainty, with may also include uncertainty in distribution model. Distribution uncertainty originates from the fact that the true distribution of empirical log-returns is unknown and must be approximated. The approximation can often not be made perfect and there exists uncertainty in the distribution approximation.

As mentioned in Section 2.2.1, when log-returns are assumed to have elliptical distribution, portfolio optimization problem (2.7) has solution that is equal to that of Markowitz mean-variance optimization problem. In this special case, there is no need to simulate log-return outcomes since the solution only depends on the expected log-return vector and covariance matrix. When the log-returns are assumed to be other than elliptically distributeds problem (2.7) does not have analytical solution and simulations of log-returns is necessary. The simulations introduce another source of uncertainty called statistical uncertainty. By the nature of simulations, one cannot be certain that the outcomes reflect a fair picture of the distribution they were drawn from, which is why statistical uncertainty is a problem. The remaining part of this thesis project will study portfolio optimization with Expected Shortfall under both model and statistical uncertainty and present approaches on how to handle optimization under those uncertainties.

3.1 Model Uncertainty

Markowitz mean-variance optimization problem assumes the expected log-return vector μ and covariance matrix Σ to be known and the same holds for the portfolio optimization problem (2.7). However in real applications the parameters are often estimated from historical market data, as in this thesis project. Since there is only a limited amount of historical data available on the market, the expected log-return vector has to be estimated by the empirical mean vector and similarly the covariance matrix must also be approximated. One cannot be certain that the approximated parameters equals the true market parameters and hence there exists model uncertainty in the parameters when solving the portfolio optimization problem yielding uncertainty in the optimal solution.

A long established problem in portfolio optimization is referred to as the problem of error maximization, discussed by Scherer in [16, pp. 185-186]. The problem origins from that the optimization algorithm tends to select assets with the best properties, in this case high log-return and low variance and correlation, and not select assets with the worst properties. These are the assets where estimation errors in μ and Σ are likely to be largest, with strong dependence on outliers in the data. Hence the optimal solution will have strong dependence on parameter uncertainty, where positive estimation error leads to over-weighted assets and negative estimation error leads to under-weighted assets.

In addition to parameter uncertainty, the multivariate distribution of the empirical log-returns can generally only be approximated, leading to uncertainty in distribution in the model. Different distributions are more or less successful in modeling historical data and where some are better at modeling the tails of the empirical distribution, others are better at capturing the behavior in the center of the empirical distribution. A perfect distribution fit on the entire empirical distribution is generally not possible to find and hence uncertainty in distribution must be regarded as a relevant factor in the model.

In the 1990s, two approaches were developed to tackle model uncertainty. One approach is called robust statistics, which involves removing or down-weighting what is thought of as being outliers in the empirical data set. The second approach is the concept of robust optimization, which will be considered in this thesis project. Robust optimization can intuitively be thought of as attempting to optimize the worst-case scenario given a confidence region. Traditionally, parameter uncertainty in Markowitz mean-variance optimization problem has received a lot of attention from the robust optimization community but less has been said about distribution uncertainty. This thesis project contributes to the robust optimization community by studying worst-case scenario based robust optimization of portfolio optimization problem (2.16) under different distribution models. First I will perform a case study on robust optimization under elliptical distributions and then move on to study robust optimization under asymmetric log-return distributions.

A third possibility for model uncertainty is that the model itself is wrong. For instance, there might exist liquidity risk in assets, invalidating the translation invariance property of the risk measure, included in the portfolio which is not covered by the model, or it could be that the covariance matrix is dependent on time and state of economy. These types of model uncertainties can be harder to evaluate directly without changing the entire problem formulation and will not be considered in this thesis project.

3.1.1 Introduction to Robust Optimization

This section introduces robust optimization and specifies what is meant by worst-case scenario based robust optimization. Generally, let $f(\mathbf{w}; X, \mathbf{a})$ and $g_k(\mathbf{w}; X, \mathbf{b})$, $k = 1, \ldots, m$ be functions of the weight vector $\mathbf{w}$, where X is a stochastic variable with cumulative distribution function $F_X(x)$ and $\mathbf{a}, \mathbf{b}$ are parameters and consider the optimization problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}; X, \mathbf{a}) \\ \text{Subject to } & g_k(\mathbf{w}; X, \mathbf{b}) \leq g_{k,0}, \quad k = 1, \ldots, m, \end{aligned}$$

where $g_{k,0}$ $k = 1, \ldots, m$ are constants. If $f(\mathbf{w}; X, \mathbf{a})$ is a convex function and the constraining equations span a convex region, there exists a unique global optimal solution that can be found using standard optimization algorithms.

Now suppose that there is some uncertainty in the parameters $\mathbf{a}, \mathbf{b}$ and in the cumulative distribution function $F_X(x)$ and it is only known that $\mathbf{a} \in \mathcal{A}$, $\mathbf{b} \in \mathcal{B}$ and $F_X(x) \in \mathcal{F}$ for some sets $\mathcal{A}, \mathcal{B}$ and family of distributions $\mathcal{F}$. The problem formulation then becomes

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}; X, \mathbf{a}) \\ \text{Subject to } & g_k(\mathbf{w}; X, \mathbf{b}) \leq g_{k,0}, \quad k = 1, \ldots, m \\ & \mathbf{a} \in \mathcal{A}, \mathbf{b} \in \mathcal{B}, F_X(x) \in \mathcal{F}. \end{aligned}$$

Since ordinary optimization problems require parameters to be known, this problem cannot be solved with traditional optimization methods. Instead, the concept of robust optimization has been developed. The general objective of robust optimization is to compute solutions that ensure feasibility independent of the parameters and distribution included in the uncertainty sets. Note that there is a key difference between robust optimization and sensitivity analysis, which is typically applied after the optimization to study the change in objective function under small variations in the underlying data. With robust optimization, the goal is instead to compute solutions when the uncertainty is included in the problem formulation, prior to optimization. Similarly to ordinary optimization, if $f(\mathbf{w}; X, \mathbf{a})$ and $g_k(\mathbf{w}; X, \mathbf{b})$

are convex functions and $\mathcal{A}, \mathcal{B}, \mathcal{F}$ are convex sets, the robust optimization problem is particularly nice since when an extreme point is found, we know it is the global optimal solution.

There are different approaches on how to solve a robust optimization problem, and one of the most commonly used approaches in portfolio optimization is the worst-case scenario based robust optimization approach introduced by Tütüncü and König in [22]. They argue it is a good idea to solve the robust optimization problem by first finding the worst-case scenarios for the parameters given the uncertainty sets and then consider the resulting optimization problem with these worst-case scenario parameters and solve it with ordinary optimization algorithms. For instance, if the worst-case scenario is attained for the smallest $\mathbf{a}$ in the uncertainty set $\mathcal{A}$ and for the largest $\mathbf{b}$ in the uncertainty set $\mathcal{B}$ then the worst-case scenario based robust optimization problem would be

$$\begin{aligned} &\min_{\mathbf{w}, \mathbf{a} \in \mathcal{A}} \quad f(\mathbf{w}; X, \mathbf{a}) \\ &\text{Subject to } \max_{\mathbf{b} \in \mathcal{B}} g_k(\mathbf{w}; X, \mathbf{b}) \leq g_{k,0}, \quad k = 1, \ldots, m \end{aligned}$$

under the additional constraint that X is drawn from the worst-case distribution, measured in some way, that is included in the distribution uncertainty set $\mathcal{F}$.

Since the worst-case scenario based robust optimization approach is widely used in robust portfolio optimization problems, it is used in this thesis project as well. Section 4.2 discuss in greater detail the effect of interpreting robust optimization as finding the worst-case scenario and then optimize the resulting worst-case scenario optimization problem.

3.1.2 Robust Optimization with Elliptical Distributions

Before presenting the worst-case scenario based robust version of (2.16) I mention some work that historically has been done on robust optimization of Markowitz mean-variance optimization problem. Lobo and Boyd discuss in [11] a robust variant of Markowitz mean-variance problem with uncertain covariance matrix Σ. The problem is presented as

$$\begin{aligned} \min_{\mathbf{w}} \max_{\Sigma \in \mathcal{S}} \quad & \mathbf{w}^T \Sigma \mathbf{w} \\ \text{Subject to } & \mathbf{w}^T \mu \geq R_{min} \\ & \sum_{i=1}^{n} w_i = 1 \\ & w_i \geq w_{min} \end{aligned}$$

and aims at finding the portfolio weights that minimize the total portfolio variance with the worst-case covariance matrix. The authors assume that the investor is ambiguous about the covariance matrix and have several possible candidates. The robust problem is discussed with box and ellipsoidal constraints on the covariance matrix, mathematically as

$$\begin{aligned}\mathcal{S}_{box} &= \left\{\Sigma \in \mathbb{S}_{+}^{n} | \Sigma_{ij,low} \leq \Sigma_{ij} \leq \Sigma_{ij,high}, i, j = 1, \ldots, n\right\} \\ \mathcal{S}_{ellipsoid} &= \left\{\Sigma \in \mathbb{S}_{+}^{n} | (\mathbf{s} - \widehat{\mathbf{s}})^{T} Q (\mathbf{s} - \widehat{\mathbf{s}}) \leq 1\right\},\end{aligned}$$

where $\mathbf{s}$ is a vector representation of the upper triangular elements of Σ, $\widehat{\mathbf{s}}$ is the corresponding empirically estimated vector and Q is the second moment of the Wishart distribution for the empirically estimated covariance matrix. A similar robust optimization problem is also presented where instead the box and ellipsoidal uncertainty sets constrain the mean log-return vector μ, mathematically written as

$$\begin{aligned}\mathcal{M}_{box} &= \{\mu \in \mathbb{R}^{n} | \mu_{i,low} \leq \mu_{i} \leq \mu_{i,high}, i = 1, \ldots, n\} \\ \mathcal{M}_{ellipsoid} &= \left\{\mu \in \mathbb{R}^{n} | (\mu - \widehat{\mu})^{T} S^{-1} (\mu - \widehat{\mu}) \leq 1\right\},\end{aligned}$$

where S is a scaled version of Σ and $\widehat{\mu}$ is the empirically estimated expected log-return vector. Both the box and ellipsoidal uncertainty set are convex sets so they are attractive from an optimization perspective. Ye, Parpas and Rustem [23] takes the problem formulation by Lobo and Boyd further and propose a problem where both μ and Σ are uncertain so that the optimization problem becomes

$$\begin{aligned}\min_{\mathbf{w}} \max_{\Sigma \in \mathcal{S}} \quad & \mathbf{w}^{T} \Sigma \mathbf{w} \\ \text{Subject to} \quad & \min_{\mu \in \mathcal{M}} \mathbf{w}^{T} \mu \geq R_{min} \\ & \sum_{i=1}^{n} w_{i} = 1 \\ & w_{i} \geq w_{min}\end{aligned}$$

and the objective is to find the portfolio weights that minimize the risk given worst-case scenarios in both expected log-return and covariance.

The worst-case scenario based robust version of problem (2.16) is now easy to formulate in a similar manner. Let the covariance matrix $\widehat{\Sigma}$ and the mean log-return vector $\widehat{\mu}$ be estimates from historical data of the uncertain parameters Σ and μ and assume we know that the true parameters are somewhere in the uncertainty sets $\mathcal{M}, \mathcal{S}$. The robust version of portfolio optimization problem (2.16) with linear loss function is then given by

$$
\begin{aligned}
\min_{\mathbf{w},\alpha} \quad & \alpha + \frac{1}{(1-q)D}\sum_{d=1}^{D} z_d \\
\text{Subject to} \quad & \min_{\mu\in\mathcal{M}} \mathbf{w}^T\mu \geq \theta \\
& z_d \geq 0, \quad d = 1,\dots,D \\
& \min_{\mu\in\mathcal{M}} \max_{\Sigma\in\mathcal{S}} \mathbf{w}^T\mathbf{R}_d + \alpha + z_d \geq 0, \quad d = 1,\dots,D \\
& \sum_{i=1}^{n} \mathbf{w}_i = V_0 \\
& w_i \geq 0, \quad i = 1,\dots,n
\end{aligned}
\tag{3.1}
$$

and the log-return vectors $\mathbf{R}_d$, $d = 1,\dots,D$ are simulated from some distribution that is interpreted as the worst one of all distributions included in the distribution uncertainty set $\mathcal{F}$.

To solve (3.1) numerically, the uncertainty sets $\mathcal{M},\mathcal{S}$ must be specified so that the worst-case scenarios can be found, and the distribution uncertainty set $\mathcal{F}$ must be defined. I begin by constructing the uncertainty sets $\mathcal{M},\mathcal{S}$.

To ensure that the robust optimization problem remains convex, I let $\mathcal{M},\mathcal{S}$ be boxes or ellipsoids as discussed by Lobo and Boyd. One approach to construct a box uncertainty set for μ is to add or subtract some term $\pm\epsilon|\widehat{\mu}_i|$ to each empirically estimated element $\widehat{\mu}_i$. This yields the box uncertainty set

$$\mathcal{M}_{box} = \{\mu \in \mathbb{R}^n : \mu = \widehat{\mu} \pm \epsilon|\widehat{\mu}|\}$$

and the worst-case scenario expected log-return vector would then be

$$\mu_{\text{min}}^{(\text{box})} = \widehat{\mu} - \epsilon|\widehat{\mu}|. \tag{3.2}$$

In this thesis project I have set $\epsilon = 0.20$. A reasonable sanity check is to investigate if $\mu_{\text{min}}^{(\text{box})}$ lies within the approximate 95% confidence interval calculated from the historical data to make sure that it is not unrealistically far from the empirically estimated parameter. With the data used in this thesis project and $\epsilon = 0.20$, $\mu_{\text{min}}^{(\text{box})}$ does indeed lie within the 95% confidence interval.

To calculate the worst-case scenario covariance matrix requires a bit more thought process. It is not possible to employ the same strategy as when calculating a box uncertainty set for μ since this could result in a non-invertible matrix which is not desired. There are several methods to define the worst-case covariance matrix and the general idea is that one wants to increase the eigenvalues and rotate the eigenvectors by some angle so that the resulting matrix is scaled and oriented worse than the original covariance matrix. One approach that does this implicitly is to study the historical price development for the assets in the reference portfolio and locate a time period where the assets seem the most correlated. Typically this occurs during times of financial crisis. Looking at the asset price development in Figure 1.1, it seems as the assets in the reference portfolio are the most correlated during 2007 – 2009 where most asset prices have a negative trend, which fits well with the time line for the financial crisis presented in [8]. Therefore, the worst-case scenario covariance matrix originating from a box uncertainty set is defined with basis on the financial crisis and taken as the covariance matrix estimated from historical data from the first trading day of 2007 until the last trading day of 2009, i.e.

$$\Sigma_{\max}^{(\text{box})} = \widehat{\Sigma} \quad \text{between January 2, 2007 – December 30, 2009.} \tag{3.3}$$

Since an empirically estimated covariance matrix is always positive semi-definite, this approach ensures an invertible worst-case scenario covariance matrix. Throughout the remaining part of the thesis project I refer to the worst-case scenario parameters $\mu_{\min}^{(\text{box})}$ and $\Sigma_{\max}^{(\text{box})}$ given by (3.2) and (3.3) respectively as "box uncertainty parameters". See Appendix A for numerical values.

With ellipsoidal uncertainty sets the elements depend on each other which makes it harder to find the worst-case scenarios $\mu_{\min}$ and $\Sigma_{\max}$. The value of one element restricts the range of values the other elements can take so it is not possible to find the worst-case scenario for each element separately without taking into account the other element values. One approach to find the worst-case scenario expected log-return vector originating from an ellipsoidal uncertainty set is however to solve the optimization problem

$$\mu_{\min}^{(\text{ellipsoidal})} = \begin{cases} \min\limits_{\mu} & \mu - \widehat{\mu} \\ \text{Subject to} & (\mu - \widehat{\mu})^T S^{-1} (\mu - \widehat{\mu}) \leq 1. \end{cases} \tag{3.4}$$

In this thesis project, the above optimization problem is solved with $S^{-1} = 100\Sigma^{-1}$ to limit the resulting worst-case scenario parameter $\mu_{\min}$ to not deviate too much from $\widehat{\mu}$. In a similar way, the worst-case scenario covariance matrix originating from an ellipsoidal uncertainty set can be obtained by solving the optimization problem

$$\Sigma_{\max}^{(\text{ellipsoidal})} = \begin{cases} \max\limits_{\mathbf{s}} & \mathbf{s} - \widehat{\mathbf{s}} \\ \text{Subject to} & (\mathbf{s} - \widehat{\mathbf{s}})^T Q (\mathbf{s} - \widehat{\mathbf{s}}) \leq 1 \\ & \mathbf{s} \leq \frac{(n-1)\widehat{\mathbf{s}}}{\chi^2_{0.975}} \\ & \operatorname{diag}(\Sigma_{\max}^{(\text{ellipsoidal})}) \geq 0. \end{cases} \tag{3.5}$$

The two last constraints are posed to ensure that $\Sigma_{\max}$ is not too far from $\widehat{\Sigma}$ and so that all variances are positive. In this thesis project the second moment of the Wishart distribution, Q, is estimated by simulating $10,000$ covariance matrices from the Wishart distribution and then calculate the covariance matrix for all pairwise elements in the simulated covariance matrices. The solution to (3.5) might not be positive semi-definite which is a requirement for the portfolio optimization problem to be solvable. In that case Rebonato and Jäckel present in [14] a general methodology to find the closest symmetric and positive semi-definite matrix given a non positive semi-definite matrix that can be used to obtain a feasible covariance matrix. The method involves spectral decomposition of the matrix and setting negative eigenvalues to zero. Throughout the remaining part of the thesis project I refer to the worst-case parameters $\mu_{\text{min}}^{(\text{ellipsoidal})}$ and $\Sigma_{\text{max}}^{(\text{ellipsoidal})}$ given by (3.4) and (3.5) respectively as "ellipsoidal uncertainty parameters". See Appendix A for numerical values.

Regarding the distribution uncertainty set $\mathcal{F}$, the thesis project first considers elliptical distributions with the historically popular normal distribution and Student's t distribution with different degrees of freedom, i.e.

$$\mathcal{F}_{elliptical} = \{N(\mu, \Sigma), \quad t(\mu, \Sigma, \nu)\}, \tag{3.6}$$

and later on considers asymmetric log-return distributions with the generalized Pareto distribution.

For the elliptical distributions in $\mathcal{F}_{elliptical}$, it is easy to determine the worst-case scenario distribution. Since the Student's t distribution has heavier tails than the normal distribution, the normal distribution is ruled out directly as the worst-case scenario distribution, having lower probability of encountering extremely large losses. For the Student's t distribution, the variance is undefined for $\nu < 2$, infinite for $\nu = 2$ and for increasing $\nu > 2$ the variance is decreasing. Hence, the smaller the degrees of freedom, the heavier the tails of the distribution and the greater is the variance and Expected Shortfall. The worst-case scenario distribution that ensures that variance exists is hence Student's t distribution with $\lim\limits_{\nu \searrow 2} \nu \approx 2.1$ degrees of freedom.

Since not much work has been done by the robust optimization community on uncertainty in distribution, it would be interesting to analyze the behavior of the solution to (3.1) under different distributions. Therefore, instead of simply solving (3.1) with the worst-case distribution in $\mathcal{F}_{elliptical}$, this thesis project contributes to the robust portfolio optimization community with a case study by solving the problem for each of the distributions in $\mathcal{F}_{elliptical}$ with both box and ellipsoidal uncertainty parameters. This is done in the two following sub sections.

3.1.2.1 Solution with Normal Distributed Log-returns

This section presents solutions to the robust portfolio optimization problem (3.1) applied to the reference portfolio with log-return vectors $\mathbf{R}_d$, $d =$

$1, \ldots, D$ drawn from the multivariate $N(\mu_{\min}, \Sigma_{\max})$ distribution. The left column in Table 3.1 presents the solution with $D = 15,000$ simulated log-return vector samples and box uncertainty parameters. The right column of Table 3.1 presents the solution with ellipsoidal uncertainty parameters. The last row in each column corresponds to the Expected Shortfall when investing according to the corresponding column. Since the robust portfolio optimization problem is still convex and linear, the problem can be solved by the same algorithm as in Section 2.4 for the non-robust problem formulation.

Table 3.1: Solutions to (3.1) applied to the reference portfolio with $N(\mu_{\min}, \Sigma_{\max})$ distributed log-returns. Left column with box uncertainty parameters. Right column with ellipsoidal uncertainty parameters.

	Box uncertainty	**Ellipsoidal uncertainty**
Asset name	Weight	Weight
AstraZeneca	0	0
Ericsson A	0	0
Hennes & Mauritz B	0	0.0024
ICA Gruppen	0.1509	0.0051
Nordea Bank	0	0
SAS	0	0
SSAB A	0	0
Swedish Match	0.3485	0.0032
TeliaSonera	0	0
Volvo	0	0
Total Bond Index	0.5006	0.9892
Expected Shortfall	0.0208	0.0051

Note that since the log-return vector is assumed to have multivariate normal distribution, the problem is equivalent to a Markowitz mean-variance optimization problem according to the remark in Section 2.2.1. Hence, instead of simulating log-returns and introducing statistical uncertainty to the solution, it would be possible to solve (1.3) with the same worst-case parameters and risk aversion coefficient $c = 5.33$ calculated in (2.12) and avoid statistical uncertainty. Those solutions are presented in Appendix C.1 as reference solutions to validate the accuracy of the simulated solutions.

3.1.2.2 Solution with Student's t Distributed Log-returns

This section presents the robust solutions to (3.1) applied to the reference portfolio with $15,000$ log-return vectors simulated from multivariate Student's t distributions with different degrees of freedom ν. Maximum-Likelihood estimation of the degrees of freedom to the historical log-return data in the reference portfolio gives $\widehat{\nu} = 3.58$. Hence low degrees of freedom when simulating log-returns from a Student's t distribution is reasonable. It is also interesting to study larger degrees of freedom as well to observe if the robust solution converges to the robust solution with normal distributed log-returns. Therefore I decide to solve (3.1) with $\nu = \{2.1, 3.58, 10, 20\}$ in this thesis project. With box uncertainty parameters, the solutions for varying degrees of freedom are presented in Table 3.2. With ellipsoidal uncertainty parameters the corresponding solutions are presented in Table 3.3.

Table 3.2: Robust solutions to (3.1) with multivariate Student's t distributed log-returns, box uncertainty parameters and varying degrees of freedom.

Asset name	$\nu = 2.1$ Weight	$\nu = 3.58$ Weight	$\nu = 10$ Weight	$\nu = 20$ Weight
AstraZeneca	0	0	0	0
Ericsson A	0	0	0	0
Hennes & Mauritz B	0	0	0	0
ICA Gruppen	0.1494	0.1414	0.1454	0.1512
Nordea Bank	0	0	0	0
SAS	0	0	0	0
SSAB A	0	0	0	0
Swedish Match	0.3495	0.3551	0.3523	0.3482
TeliaSonera	0	0	0	0
Volvo	0	0	0	0
Total Bond Index	0.5011	0.5035	0.5023	0.5005
Expected Shortfall	0.0902	0.0427	0.0260	0.0231

Table 3.3: Robust solutions to (3.1) with multivariate Student's t distributed log-returns, ellipsoidal uncertainty parameters and varying degrees of freedom.

Asset name	$\nu = 2.1$ Weight	$\nu = 3.58$ Weight	$\nu = 10$ Weight	$\nu = 20$ Weight
AstraZeneca	0	0.0004	0	0
Ericsson A	0	0	0	0
Hennes & Mauritz B	0.0012	0.0041	0	0.0003
ICA Gruppen	0.0010	0	0.0006	0.0006
Nordea Bank	0.0007	0	0.0011	0.0024
SAS	0	0	0	0
SSAB A	0	0	0	0
Swedish Match	0.0091	0.0075	0.0024	0.0056
TeliaSonera	0	0	0	0
Volvo	0	0	0	0
Total Bond Index	0.9880	0.9884	0.9919	0.9911
Expected Shortfall	0.0237	0.0113	0.0063	0.0056

Similarly to the case with normal distributed log-returns, to avoid introducing statistical uncertainty in the solutions, it is possible to solve Markowitz mean-variance optimization problem (1.3) with the same worst-case parameters and risk aversion coefficient c calculated using (2.11) with the Student's t quantile function. Those solutions are presented in Appendix C.2 as reference solutions to validate the accuracy of the simulated solutions.

3.1.3 Robust Optimization with Asymmetric Distributions

A well known problem when fitting distributions to empirical data with Maximum-Likelihood estimation is that most focus is on fitting the center of the distribution and less focus is on the tails. This is because by definition more observations are located in the center than in the tails of the empirical distribution. This behavior is unfortunate when calculating risk measures such as Expected Shortfall which depends heavily on the loss tail of the modeled distribution. One attempt to solve this problem is to model the distributional tails and center separately.

3.1.3.1 The univariate generalized Pareto distribution

Modeling the tails and center of a univariate empirical distribution separately results in an asymmetric log-return distribution and there is no longer a simple analytical relationship between Markowitz mean-variance optimization problem and the portfolio optimization problem with Expected Shortfall. This is of course of interest because then problem formulation (2.7) is a unique portfolio optimization problem not identical to a Markowitz mean-variance optimization problem.

A distribution that is popular to use for modeling the tails of an empirical log-return distribution is the generalized Pareto distribution (GPD). More correctly, given independent and identically distributed residuals $r_1, \ldots, r_n$ with unknown distribution function F and a high threshold u^h, the excesses $r_k - u^h$ often turns out to be well modeled by the generalized Pareto distribution. For a shape parameter $\gamma > 0$ and a scale parameter $\beta > 0$, the generalized Pareto distribution is given by

$$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}}, \quad x \geq 0$$

and for $\gamma = 0$ it is given by

$$G_{\gamma,\beta}(x) = 1 - e^{-\frac{x}{\beta}}, \quad x \geq 0.$$

The choice of a suitably high threshold u^h is crucial, but often hard, for the generalized Pareto distribution to be a good fit to the excesses. If u^h is too large then few observations can be used for parameter estimation yielding poor estimates with large variation. If instead u^h is too small then more observations can be used to estimate the parameters but modeling the excesses with the generalized Pareto distribution becomes questionable. One approach to estimate the parameters is to pick some u_1^h far out in the empirical distribution, say the 90% quantile of the empirical distribution, and estimate the parameters with Maximum Likelihood estimation. Then a new threshold u_2^h is chosen farther out in the empirical distribution, say the 91% quantile, and the parameters are estimated once again with Maximum Likelihood estimation. The procedure is repeated and the threshold u^h is chosen as the first candidate for which the parameters to the generalized Pareto distribution are stable from that point on. This method relies on the assumption that the parameters will converge before we are not too far out in the distribution tail. With data originating from the historical log-returns in this thesis project the method could not be used because the parameters did not converge until very far out in the distribution tail. Instead, the threshold were chosen as the 90% empirical quantile, i.e. $u^h = F_n^{-1}(0.90)$, for each asset in the reference portfolio[1]. In Section 4.3 I mention another strategy that could be used to choose better values of u^h.

Note that the generalized Pareto distribution models excesses above a high threshold u^h which means that we are looking at the upper tail of the log-return distribution, but Expected Shortfall only depends on the lower tail of the log-return distribution. Luckily, if the lower tail of the log-return distribution consists solely of negative values below some negative threshold u^l one may simply look at the absolute values and the "excess" of an observation in the lower tail is then the positive distance from the observation to u^l and one can model the lower tail of the log-return distribution with a generalized Pareto distribution as well. Levine makes use of this strategy in [9] when modeling the lower tail of the total monthly return distribution for the A-rated 7- to 10-year corporate bond component of Citi's U.S. Broad Investment Grade Bond Index from January 1980 to August 2008. Similarly to the upper threshold u^h, the lower thresholds are chosen as the 10% empirical quantile, i.e. $u^l = F_n^{-1}(0.10)$, for each asset in the reference portfolio.

Since generalized Pareto distributions can only be used to model the tails of a distribution a third distribution must be used as a bridge to connect the two tails with the center of the distribution. In this thesis project, this is done with the empirical distribution.

A mathematical expression for the entire underlying distribution will be derived by considering two scenarios and then combining them. Consider first the scenario where the excesses above some threshold u^h of a distribution consisting of independent and identically distributed random variables is modeled with a generalized Pareto distribution $G^h_{\gamma,\beta}(x - u^h)$. For $x \geq u^{\mathrm{h}}$

[1] Skoglund and Nyström show in [13] that Maximum Likelihood estimation of generalized Pareto distribution parameters is not sensitive of the choice of threshold value.

the cumulative distribution function of the underlying distribution is given by

$$\begin{aligned} P(X \leq x) &= P(X \leq u^h) + P(u^h \leq X \leq x) \\ &= P(X \leq u^h) + (1 - P(X \leq u^h))G^h_{\gamma,\beta}(x - u^h). \end{aligned}$$

The probability $P(X \leq u^h)$ is estimated by the empirical distribution function as $P(X \leq u^h) \approx F_n(u^h)$. For $x < u^h$ the remaining underlying distribution is modeled with the empirical distribution $F_n(x)$. The cumulative distribution function for the entire underlying distribution is then modeled as

$$F(x) = \begin{cases} F_n(x), & x < u^h \\ F_n(u^h) + (1 - F_n(u^h))G^h_{\gamma,\beta}(x - u^h), & x \geq u^h. \end{cases}$$

Now consider the reversed scenario where the "excesses" of the independent and identically distributed random variables below some threshold u^l is modeled by a generalized Pareto distribution $G^l_{\gamma,\beta}(u^l - x)$ and for $x > u^l$ the remaining underlying distribution is modeled with the empirical distribution $F_n(x)$. The cumulative distribution function for the entire underlying distribution is in this case modeled as

$$F(x) = \begin{cases} F_n(u^l)\left(1 - G^l_{\gamma,\beta}(u^l - x)\right), & x \leq u^l \\ F_n(x), & x > u^l. \end{cases}$$

By combining the two scenarios, the cumulative distribution function for the entire underlying distribution is given by

$$F(x) = \begin{cases} F_n(u^l)\left(1 - G^l_{\gamma,\beta}(u^l - x)\right), & x \leq u^l \\ F_n(x), & u^l < x < u^h \\ F_n(u^h) + (1 - F_n(u^h))G^h_{\gamma,\beta}(x - u^h), & x \geq u^h. \end{cases} \tag{3.7}$$

where the shape and scale parameters γ and β may differ between the two generalized Pareto distribution functions $G^l_{\gamma,\beta}(x)$ and $G^h_{\gamma,\beta}(x)$. The cumulative distribution function (3.7) is asymmetric and will be referred to in this thesis project as the hybrid GPD-Empirical-GPD distribution.

3.1.3.2 Copula dependence

The generalized Pareto distribution is univariate so for each asset in the reference portfolio two unique generalized Pareto distributions and the empirical distribution must be modeled to obtain unique univariate cumulative distribution functions. To capture the dependence between the different assets a copula may be used. A copula can be explained as a function that couples a joint distribution function to univariate marginal distribution functions and it is from the joint distribution function where the multivariate dependence between the univariate random variables is inherited. A copula is constructed by combining probability and quantile transforms. If X is a random variable with continuous distribution function F then the probability transform states that $F(X)$ has uniform distribution on $(0,1)$, i.e. $F(X)$ is $U(0,1)$. If U is a random variable with uniform distribution on $(0,1)$ and G is some distribution function then the quantile transform states that $G^{-1}(U)$ has distribution function G, i.e. $P(G^{-1}(U) \leq x) = G(x)$. Using the probability transform it follows that for a random vector $\mathbf{U} = (U_1, \ldots, U_n)$ whose components have uniform distribution on $(0,1)$ a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ whose components have marginal distributions $F_1, \ldots, F_n$ can be constructed as

$$\mathbf{X} = (F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n)). \tag{3.8}$$

Now, to describe the dependence between the components of $\mathbf{X}$ one may instead describe the dependence between the components of $\mathbf{U}$ and use (3.8). The distribution function $C(u_1, \ldots, u_n)$ is called a copula and using the quantile transform it follows that

$$\begin{aligned} C(F_1(x_1), \ldots, F_n(x_n)) &= P(U_1 \leq F_1(x_1), \ldots, U_n \leq F_n(x_n) \\ &= P(F_1^{-1}(U_1) \leq x_1, \ldots, F_n^{-1}(U_n) \leq x_n) \\ &= F(x_1, \ldots, x_n) \end{aligned}$$

where F is the joint multivariate distribution of $\mathbf{X}$.

Since the components of $\mathbf{U}$ are transformed to components in $\mathbf{X}$ using the marginal distributions of $\mathbf{X}$ it is important to realize that the measure of dependence between the random variables must be invariant to this transformation. Ordinary linear correlation is not invariant to this transformation which makes it necessary to find such a measure. One measure that has the desired property is the rank correlation Kendall's tau. Kendall's tau is defined on the random vector (X_1, X_2) as

$$\tau(X_1, X_2) = P\left((X_1 - X_1')(X_2 - X_2') > 0\right) - P\left((X_1 - X_1')(X_2 - X_2') < 0\right)$$

where (X_1', X_2') is an independent copy of (X_1, X_2). Lindskog, McNeil and Schmock show in [10] that if (X_1, X_2) have an elliptical distribution then the relationship

$$\tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho) \tag{3.9}$$

holds under some technical conditions. The relationship makes it easy to convert the linear correlation matrix C into a transformation invariant rank correlation matrix C_τ that can be used when simulating from an elliptical copula.

3.1.3.3 Sample Preparation with GARCH(1,1) Model

The validity of fitting a generalized Pareto distribution to excesses in a distribution tail relies heavily on the assumption of independent and identically distributed random variables. Hence, to model the tails of the empirical distributions to each asset in the reference portfolio with generalized Pareto distribution it is of great importance that the data has the right properties. When initially defining log-returns in (1.1) it was mentioned that log-returns are often assumed to be weakly dependent and close to independent and identically distributed. In financial time series analysis of risk factors (here the log-returns), stylized facts are effects that are commonly observed in the data. Josefsson summarizes the stylized facts in [7] as

1. Return series are not independent and identically distributed although they show little serial correlation.
2. Series of absolute or squared returns show significant serial correlation.
3. Conditional expected returns are close to zero.
4. Volatility appears to vary over time.
5. Return series are leptokurtic[2] or heavy tailed.
6. Extreme returns appear in clusters.

The stylized facts are problematic since they imply that the empirical log-returns cannot be considered independent and identically distributed and hence the excesses in the distribution tails cannot be modeled directly with generalized Pareto distributions. Some sample preparation has to be done, typically by filtering the data through a time series model. A popular model used for filtering financial log-return data is the Generalized Autoregressive Conditional Heteroscedasticity, GARCH, model.

The general GARCH(p,q) process is defined in the following way. Let $\{Z_t\}$ be independent and identically distributed $N(0,1)$. Then $\{X_t\}$ is called a GARCH(p,q) process if

$$X_t = \mu + \sigma_t Z_t, \quad t \in \mathbb{Z}$$

where $\{\sigma_t\}$ is the nonnegative process

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

and $\alpha_0 > 0$, α_j, $\beta_j \geq 0$, $j = 1, 2, \ldots$. For more on GARCH(p,q) processes and other time series models I refer to [1] by Brockwell and Davis. Note from the GARCH model that if X_t is interpreted as the log-return at time t then the model is a linear combination of squared historical log-returns

[2]A leptokurtic distribution has fatter tails and higher peak than the curvature found in a normal distribution.

which fits stylized fact 2. Also, the volatility is dependent on time and on previous volatilities, which fits stylized facts 4 and 6. This is an intuitive motivation on why GARCH models often fits historical log-return time series well. Furthermore, by applying Hölder's inequality when calculating the kurtosis of $\sigma_t Z_t$,

$$k(\sigma_t Z_t) = \frac{\mathbb{E}[(\sigma_t Z_t)^4]}{\mathbb{E}[(\sigma_t Z_t)^2]^2} = k(Z_t)\frac{\mathbb{E}[\sigma_t^4]}{\mathbb{E}[\sigma_t^2]^2} \geq k(Z_t)$$

it is seen that it is greater than the kurtosis of Z_t. Hence, even though the residuals are assumed to be normal distributed, the GARCH(p,q) model takes into account that log-returns often have fatter tails than the normal distribution. This fits well with stylized fact 5.

There are methods for fitting general GARCH(p,q) models to time series data which require us to find p, q, $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$. In this thesis project I settle with fitting GARCH(1,1) models to the historical log-returns, which makes the process easier. GARCH(1,1) models are used since they are often good enough to filter financial data and the parameters can for instance be approximated by Maximum Likelihood estimation.

When the parameters have been estimated the standardized residuals $\{Z_t\}$ are obtained as

$$Z_t = \frac{R_t - \widehat{\mu}}{\widehat{\sigma}_t}$$

and if the GARCH filtration is successful then the stylized facts should be accounted for, meaning that the standardized residuals should be independent and identically distributed.

In Appendix D I analyze the standardized residuals to the filtered empirical log-returns and conclude that the distribution tails are asymmetric and heavier than for the standard normal distribution assumed in the definition of the GARCH process. In practice this is not a problem since the standardized residuals can be modeled by any suitable distribution. This motivates the idea of modeling the tails with generalized Pareto distributions.

3.1.3.4 Solution with Normal Copula and Hybrid GPD-Empirical-GPD Log-return Marginals

In this section I solve the robust portfolio optimization problem (3.1) with worst-case scenario parameters originating from both box and ellipsoidal uncertainty sets and log-returns simulated from a normal copula with hybrid GPD-Empirical-GPD marginal distributions. Skoglund and Wei note in [19, p. 132] that the choice of copula has a second-order effect when modeling market risk factors such as stock log-returns, the most important factor being to model stochastic volatility and volatility clustering which is accounted for in the GARCH(p,q) model. Hence choosing the normal copula for dependency structure is not the most important decision in the modeling procedure and if a Student's t copula were to be chosen instead, it would outperform the normal copula model only by a small margin. In Appendix D, Table D.1, I present the parameters for the GARCH(1,1) models used for filtering the historical log-returns to standardized residuals. With help of Figure D.1 I argue on basis of the stylized facts for financial time series why the GARCH(1,1) models have filtered the empirical log-return data well into standardized independent and identically distributed residuals. Secondly I use Figure D.2 to describe why the standardized residuals are not well modeled with the standard normal distribution which motivates the use of hybrid GPD-Empirical-GPD distributions instead. In Table D.2 I present values of estimated thresholds and parameters for the generalized Pareto distributions when fitted to the standardized residuals and in Figure D.3 the estimated generalized Pareto distribution functions are plotted to show that the models approximate the true data well.

In Section 3.1.2.1 and Section 3.1.2.2 it was easy to simulate log-returns by generating random variables from the multivariate normal and Student's t distribution. With hybrid GPD-Empirical-GPD marginal distributions and dependence between the standardized residuals from a normal copula the simulation of log-returns is not as straight forward. Below I provide an algorithm that can be used to simulate log-returns in this case.

Algorithm 1 Simulation of log-returns from normal copula with hybrid GPD-Empirical-GPD marginal distributions

1: Simulate dependent random variables $x_1, \ldots, x_n$ from the multivariate normal distribution $N(0, C_\tau)$, where C_τ is the Kendall's tau rank correlation matrix with each element calculated with (3.9).

2: Transform $x_1, \ldots, x_n$ to $U(0,1)$ distributed random variables as $u_i = \Phi(x_i)$ where $\Phi(x)$ is the standard normal cumulative distribution function.

3: The standardized residuals $Z_i = F_i^{-1}(u_i)$, $i = 1, \ldots, n$, where the quantile function F_i^{-1} is the inverse function of (3.7) for asset i with parameters found in Table D.2.

4: The simulated log-returns $R_i = \sigma_{t+1}^i Z_i + \widehat{\mu}_i$, $i = 1, \ldots, n$, where σ_{t+1}^i is the forecasted volatility obtained from the GARCH(1,1) model with parameters in Table D.1 and $\widehat{\mu}_i$ is the empirically estimated mean log-return value for asset i.

Simulation from a copula can often be written in a similar algorithm as the one above, specified for the particular task it should be used for. Cherubini, Luciano and Vecchiato write in [2, Ch. 6] several algorithms to generate samples from other types of copulas, for instance the Student's t copula but also archimedean copulas such as the Clayton, Gumbel and Frank copula.

To generate the log-returns needed for the optimization problem the above algorithm is repeated 15,000 times to obtain the appropriate sample size D. Note that log-returns with worst-case scenario parameters are as easy to generate by simply replacing C_τ with a corresponding worst-case scenario rank correlation matrix in step 1 and replace the forcasted standard deviation σ^i_{t+1} and the empirically estimated μ_i with corresponding worst-case scenario forcasted standard deviation and worst-case scenario expected log-return in step 4. To solve the robust portfolio optimization problem (3.1) the simulated log-returns can then be directly inserted in the same optimization solving algorithm as when solving with elliptically distributed log-returns since the optimization problem itself is unchanged. Table 3.4 presents the optimal weight vector and portfolio Expected Shortfall when solving (3.1) with both box and ellipsoidal uncertainty parameters.

Table 3.4: Solutions to (3.1) applied to the reference portfolio with log-returns modeled by a normal copula with hybrid GPD-Empirical-GPD marginals. Left column with box uncertainty parameters and right column with ellipsoidal uncertainty parameters.

	Box uncertainty	**Ellipsoidal uncertainty**
Asset name	Weight	Weight
AstraZeneca	0	0
Ericsson A	0	0
Hennes & Mauritz B	0.0123	0.0037
ICA Gruppen	0.1855	0
Nordea Bank	0	0
SAS	0	0
SSAB A	0	0
Swedish Match	0.3201	0.0080
TeliaSonera	0	0
Volvo	0	0
Total Bond Index	0.4822	0.9884
Expected Shortfall	0.0233	0.0056

3.2 Statistical Uncertainty

Portfolio optimization problem (2.7), or equivalently problem (2.16), has in the case of elliptically distributed log-returns with known parameters μ and Σ analytical solution that corresponds to the solution to Markowitz mean-variance optimization problem. Under other model assumptions the problem might however not have analytical solution. This applies for instance to log-returns simulated from the normal copula with hybrid GPD-Empirical-GPD marginals, and an optimal solution must be numerically calculated from simulations. Simulating D log-return vectors $\mathbf{R}$ several times would then produce different solutions since statistical uncertainty is present. A popular strategy to evaluate the sensitivity of a solution in the presence of statistical uncertainty is the bootstrap strategy, discussed in [6] by James, Witten, Hastie and Tibshirani.

The idea behind bootstrapping is to estimate the accuracy of some estimated quantity κ when it is not possible or practical to create new samples from the original population. This could for instance be the case when the original distribution is unknown, because it is too expensive or too time consuming or because it is in other ways impossible to generate new data samples. With the reference portfolio in this thesis project it is impossible to obtain more historical daily log-return data for the time period of interest and simulating log-returns from the normal copula with hybrid GPD-Empirical-GPD marginal distributions is quite time consuming. Therefore the bootstrap procedure is appropriate to use to estimate standard errors for the portfolio weights. In the (non-parametric) bootstrap procedure one obtains new distinct artificial data sets by repeatedly sampling with replacement from the original data set. That is, from the original data set $\mathbf{X}$ we repeatedly draw and replace elements to construct new data sets $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(B)}$, for B being big, where in each artificial data set $\tilde{\mathbf{X}}^{(b)}$ a specific element in $\mathbf{X}$ may occur multiple or no times. From each artificial data set the desired quantity is calculated so that we have B samples $\kappa^{(1)}, \dots, \kappa^{(B)}$ originating from the bootstrapped data sets. An estimate of the standard error can then be calculated as

$$SE_B(\kappa) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\kappa^{(b)} - \frac{1}{B}\sum_{b'=1}^{B}\kappa^{(b')}\right)^2}. \tag{3.10}$$

Small standard error implies that the original data set $\mathbf{X}$ is large enough to produce a precise solution.

3.2.1 Statistical Uncertainty in the Portfolio Optimization Problem

In this section I apply the bootstrap strategy when solving (3.1) with log-returns simulated from the normal copula with hybrid GPD-Empirical-GPD marginal distributions so that standard errors for each of the weights in Table 3.4 can be estimated. In this case what is referred to as the original data set $\mathbf{X}$ would be a set $\underline{\mathbf{R}}$ consisting of D log-return vectors $\mathbf{R}$ simulated from the normal copula. Drawing with replacement one element from $\mathbf{X}$ would further correspond to drawing with replacement one log-return vector $\mathbf{R}$ from $\underline{\mathbf{R}}$ and the desired quantity κ that should be calculated for each new artificial data set $\tilde{\underline{\mathbf{R}}}^{(b)}$ is the optimal weight vector $\mathbf{w}^{(b)}$ and portfolio Expexted Shortfall $ES_p^{(b)}(X)$. This results in B samples of optimal weight vectors $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(B)}$ and portfolio Expected Shortfall $ES_p^{(1)}(X), \dots, ES_p^{(B)}(X)$ and the standard errors can then be calculated by using (3.10).

In this application, I let $B = 1000$. With $D = 15,000$ simulated log-returns in the original data set $\underset{\sim}{\mathbf{R}}$ this means that I should draw with replacement $15,000$ samples of log-return vectors, calculate the optimal solution to (3.1) and repeat the procedure 1000 times. The standard errors for each asset weight w_i can then be calculated with (3.10). Table 3.5 summarizes each of the standard errors when conducting the described procedure for both box and ellipsoidal uncertainty parameters so that standard errors are estimated for every weight and the Expected Shortfall in Table 3.4.

Table 3.5: Standard errors calculated with the bootstrap procedure when solving (3.1) with original log-returns simulated from the normal copula with hybrid GPD-Empirical-GPD marginal distributions. Left column with box uncertainty parameters. Right column with ellipsoidal uncertainty parameters.

	Box uncertainty	**Ellipsoidal uncertainty**
Asset name	Standard error	Standard error
AstraZeneca	0	0.0019
Ericsson A	0	0
Hennes & Mauritz B	0.0148	0.0016
ICA Gruppen	0.0111	0.0024
Nordea Bank	0	0.0004
SAS	0	0
SSAB A	0	0
Swedish Match	0.0078	0.0037
TeliaSonera	0	0.0001
Volvo	0	0.0001
Total Bond Index	0.0092	0.0018
Expected Shortfall	$6.6754 \cdot 10^{-4}$	$1.7177 \cdot 10^{-4}$

A short analysis of the results is presented in Section 4.1.2 together with a comparison of the asymmetric solutions in Table 3.4 with the solutions in Table 3.1.

Chapter 4

Analysis and Conclusions

In this chapter the results found in the previous chapter are analyzed. I also give comments on the interpretation of robust optimization as worst-case scenario based optimization. The thesis project ends with comments on alternative approaches that could have been made and comments on areas of further investigation is mentioned.

4.1 Analysis of Results

In this section the results obtained throughout the thesis project are analyzed and compared to gain knowledge of how the different model assumptions investigated affect the optimal solution to the robust portfolio optimization problem with Expected Shortfall.

An initial observation valid for all solutions is that even though 11 assets are included in the reference portfolio, only a few of them are actually invested in. This could initially be seen as counter-intuitive since the resulting portfolio would be more diversified if the initial capital were spread across more assets and the investor could become exposed to less risk. The key to understand this behavior is however to recall that with highly correlated assets the idea of a well diversified portfolio lose its meaning. As an example, consider a portfolio consisting of two assets with correlation 1 but different expected log-returns. In that case the risk would not decrease if half the capital were allocated in each of the assets since they are perfectly correlated. It would hence be better to invest everything in the asset with the higher expected log-return. The same principle holds for the reference portfolio in this thesis project and diversifying the initial capital more than suggested by the optimal solution would decrease the expected portfolio log-return more than the decrease in Expected Shortfall and make the solution sub-optimal. If a more distributed portfolio allocation is preferred by the investor then additional lower and upper bounds on the weights could be added to the portfolio optimization problem formulation. As an example, pension funds often have restrictions on the portfolio weights in different asset classes and industries.

4.1.1 Elliptically distributed log-returns

Here the different solutions with elliptically distributed log-returns will be compared and analyzed. More specifically, the benchmark solution in Table 1.2 and the solutions in Table 2.1 and Table 3.1 through Table (3.3) are investigated. Since portfolio optimization with Expected Shortfall is directly related to Markowitz mean-variance optimization with elliptically distributed log-returns I may compare the robust solutions to (2.16) to the benchmark solution to analyze the effect of model uncertainty in portfolio optimization.

The first observation that I make is that with box uncertainty parameters the robust solutions resembles the benchmark solution more than with ellipsoidal uncertainty parameters. This is seen by observing that the solutions in the left columns of Table 3.1 and Table 3.2 have similar structure as the benchmark solution in Table 1.2 while the solutions in the right columns of Table 3.1 and Table 3.3 differs more from the benchmark solution. With box uncertainty parameters the initial capital is invested in the same assets as in the benchmark solution and the only difference is that the robust solutions are a bit more diversified. With ellipsoidal uncertainty sets the wast majority of the initial capital is invested in the Total Bond Index and approximately one percent of the capital is diversified across a few other assets. Note that this result dependents on how much $\mu_{\min}$ and $\Sigma_{\max}$ deviate from the empirically estimated parameters where larger deviations lead to larger solution deviation from the benchmark solution. I can therefore conclude from the optimal solutions that the ellipsoidal uncertainty parameters deviate more from the empirically estimated parameters than the box uncertainty parameters, which is confirmed by comparing the numerical values in Appendix A. The results from robust optimization with box uncertainty parameters can be seen as a stress test of the optimization problem where the investor has a more conservative view on the market than the historical data suggests, while robust optimization with ellipsoidal uncertainty parameters could be seen as exposing the portfolio optimization problem to a large financial crisis. A conclusion independent of the particular parameters used in this thesis project is that portfolio optimization with Expected Shortfall is sensitive to parameter uncertainty. This can be concluded since the robust optimal solutions deviate notably from the benchmark solution regardless of which worst-case scenario parameters that are used. If the optimization problem would had been insensitive to parameter uncertainty then the robust solutions would have been close to each other and close to the benchmark solution. It is therefore of great importance to take into consideration uncertainties in parameters when solving a portfolio optimization problem. Quite small variations in parameters can have huge impact on the optimal solution and the risk exposure.

With $\mu_{\min}$ and $\Sigma_{\max}$ held fixed, the optimal robust solutions seem to be almost independent of the underlying log-return distribution. This can for instance be seen by comparing the left (right) column of Table 3.1 with Table 3.2 (3.3) or equivalently by comparing the reference solutions in Appendix C. The reference solutions suggest that the optimal solution is almost independent of the multivariate elliptical distribution used for modeling log-returns. The multivariate distribution model only have significant impact on the Expected Shortfall where Student's t distributed log-returns yield larger risk than normal distributed log-returns. These two observations should come as no surprise. The constraint $\mathbf{w}^T\mu \geq \theta$ is independent of the log-return distribution and the optimal weight vector is insensitive of the log-return distribution. On the contrary, the risk is directly dependent on the distribution model since fatter tails implies a greater probability of encountering extremely negative (and positive) log-returns which increases the Expected Shortfall. The Student's t distribution with low degrees of freedom have fatter tails than the normal distribution and therefore Student's t distributed log-returns imply larger Expected Shortfall. Additionally, as the degrees of freedom increases the Expected Shortfall should converge to the Expected Shortfall for normal distributed log-returns. This behavior is verified by solving (3.1) with Student's t distributed log-returns and plot Expected Shortfall as function of the degrees of freedom. Figure 4.1 depicts this study with a reference level of Expected Shortfall for normal distributed log-returns included to observe the convergence of Expected Shortfall.

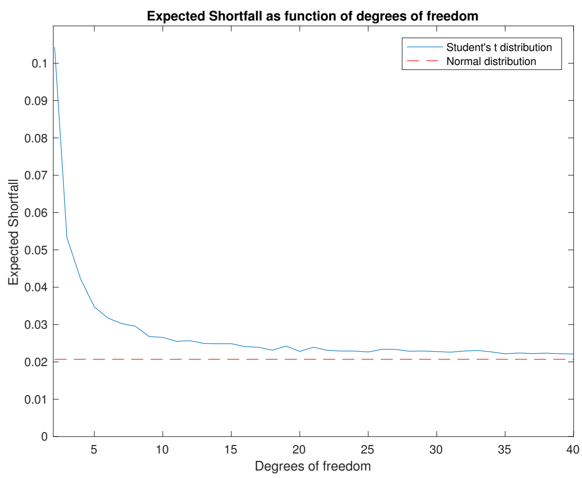

Figure 4.1: Expected Shortfall with Student's t distributed log-returns is converging for increasing degrees of freedom to Expected Shortfall with normal distributed log-returns.

Until now the threshold for acceptable expected daily log-return has been held constant to $\theta = 2.0242 \cdot 10^{-4}$ (5.08% expected yearly log-return). Figure 4.2 presents robust solutions to (3.1) with box uncertainty parameters and both multivariate normal and Student's t distributed log-returns for varying θ. As can be seen, the solutions are almost identical with small differences originating from statistical uncertainty and small differences in the risk aversion coefficient. Hence the choice of distribution has little impact on the optimal weights regardless of the level of θ. As long as θ is feasible[1] the optimal solution is affected very little by different elliptical distributions and portfolio optimization with Expected Shortfall is insensitive to which elliptical distribution model that is used.

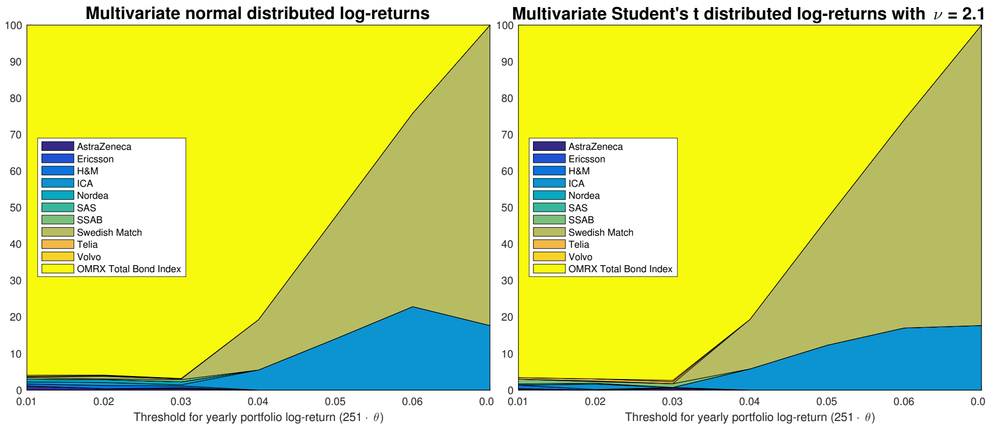

Figure 4.2: *Left*: Optimal robust solutions to (3.1) with box uncertainty parameters and multivariate normal distributed log-returns as function of θ. *Right*: Corresponding solutions with multivariate Student's t distributed log-returns and 2.1 degrees of freedom.

Previously I concluded that Expected Shortfall depends on the log-return distribution and the heavier tails of the distribution the larger Expected Shortfall. It is further known that Expected Shortfall increases when θ increases as a result of a less diversified portfolio. It would be interesting to analyze how Expected Shortfall depends on both these factors simultaneously and if some structure can be found in the dependence. Figure 4.3 visualize Expected Shortfall when solving (3.1) for different elliptical distributions with box uncertainty parameters.

[1]If θ is too large then the constraint $\mathbf{w}^T\mu \geq \theta$ cannot be satisfied and the portfolio optimization problem has no solution.

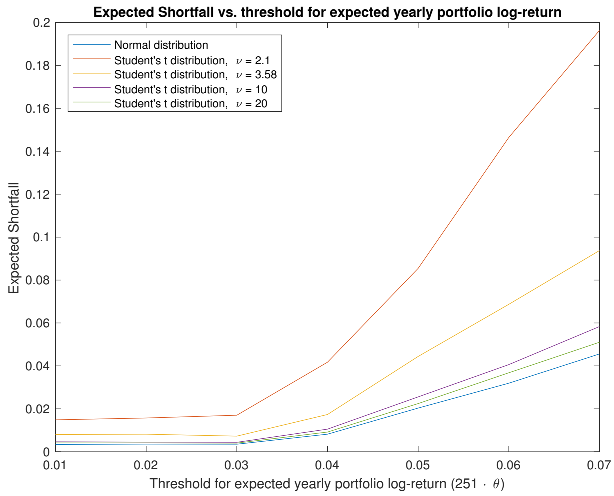

Figure 4.3: Expected Shortfall for problem (3.1) for different elliptically distributed log-returns as function of the threshold for expected portfolio log-return θ.

Two interesting observations can be made. The first is that given a log-return distribution, Expected Shortfall increases marginally when the threshold for acceptable expected yearly portfolio log-return is less than 3% and afterwards it increases more pronounced. This behavior should be specific for the reference portfolio used in this thesis project and for another reference portfolio the behavior could differ. Looking at Figure 4.2 one notice that an expected yearly portfolio log-return of 3% seems to be a breakpoint in the optimal weight vector as well. It is stable and almost unchanged for expected yearly portfolio log-returns less than 3% but then changes continuously for larger values. This explains the behavior of Expected Shortfall. I conclude that for the specific reference portfolio used in this thesis project the investor can increase the acceptable expected yearly portfolio log-return up to 3% in the robust portfolio optimization problem with box uncertainty parameters without increasing the Expected Shortfall since the optimal asset allocations are unchanged. With acceptable expected yearly portfolio log-return greater than 3% the investor must accept that the Expected Shortfall will increase more rapidly. The second interesting observation from Figure (4.3) is that there seems to be some structure in the ratio $ES^{\text{Student's t}}/ES^{\text{Normal}}$ and the figure looks like a folding fan. Since the optimal weight vector is almost unchanged between different elliptical distributions, the dependence between $ES^{\text{Student's t}}$ and ES^{Normal} for a given threshold θ must be possible to explain from how Expected Shortfall is calculated analytically for elliptical distributions. Recalling that the portfolio log-return under elliptical distributions can be written as (2.10), the portfolio Expected Shortfall is

$$ES_p(\mathbf{w}^T\mathbf{R}) = -\mathbf{w}^T\mu + \sqrt{\mathbf{w}^T\Sigma\mathbf{w}}ES_p(WZ_1), \quad (4.1)$$

where WZ_1 determines which standard elliptical distribution is used. Now, since $\mathbf{w}$ is almost unchanged between different elliptical distributions and the parameters μ and Σ are not dependent on the distribution, it is differences in Expected Shortfall for standard elliptical distributions, $ES_p(WZ_1)$, that determines the ratio $ES^{\text{Student's t}}/ES^{\text{Normal}}$. The ratio should therefore be approximately proportional to $ES_p^{\text{Standard Student's t}}/ES_p^{\text{Standard Normal}}$ where the numerator depends on the degrees of freedom.

4.1.2 Asymmetric Distributed Log-returns

I continue the analysis of the results by comparing the robust solutions between asymmetric and elliptically distributed log-returns. This allows me to conclude what impact the implementation of portfolio optimization problem (2.7) in favor of Markowitz mean-variance optimization problem (1.3) has, since it can handle asymmetric distributions. If an asymmetric distribution model have visible impact on the optimal solution then I can conclude that problem (2.7) is important since it changes the investment strategy. Something that was not achieved by altering between elliptical distributions.

First of all I give a short comment on the GARCH(1,1) filtering and generalized Pareto distribution fit to the log-return distribution tails. When modeling historical log-returns, the stylized facts in Section 3.1.3.3 should be accounted for in a good model. Modeling the log-returns directly with an elliptical distribution may take into account the heavy tailed behavior of the log-returns if the multivariate Student's t distribution is used, but for instance volatility clustering and significant serial autocorrelation is not accounted for explicitly. By filtering the log-returns with a GARCH(1,1) model, the heavy tailed behavior is captured, as well as volatility clustering and serial correlation. Hence, more stylized facts are taken into consideration which should result in a better model. A GARCH(1,1) model together with generalized Pareto distribution models in the tails to account for asymmetry is thus a much more accurate model than an elliptical distribution model. The optimization results when simulating log-returns from a normal copula with hybrid GPD-Empirical-GPD marginals should therefore be more reliable than the results when simulating from elliptical distributions.

Since the log-returns are simulated from a normal copula, the robust solutions in Table 3.4 will be compared to the solutions in Table 3.1 where the log-returns have multivariate normal distribution. Comparing the tables shows that with both box and ellipsoidal uncertainty parameters the Expected Shortfall is greater when simulating log-returns from the normal copula. The result is expected. As has been discussed before, the multivariate normal distribution has not fat enough tails to be good at modeling historical log-return tails and therefore the Expected Shortfall calculated with multivariate normal distribution will be smaller than the true Expected Shortfall. Modeling the tails with generalized Pareto distributions solves this problem and the tails are better modeled with fatter tails which results in larger Expected Shortfall. In this sense, the Expected Shortfall in Table 3.4 should be thought of as closer to reality than the Expected Shortfall in Table 3.1.

It is not only the Expected Shortfall that changes when simulating log-returns from a normal copula with asymmetric marginals instead of simulating from a multivariate normal distribution. The optimal asset weights change notably as well, seen by comparing the solutions in Table 3.1 and Table 3.4. The general structure of the investments is very similar between the tables, but quite large differences can be seen in the amount of capital invested when comparing individual assets. This behavior was not seen when changing between elliptical distributions. It can for instance be seen that with box uncertainty parameters a switch to simulating with the normal copula results in an investment increase of approximately 4% in ICA Gruppen and 1% in Hennes & Mauritz B and a decrease of approximately 3% in Swedish Match and 2% in the Total Bond Index. These can be significant changes. For instance, if \$1 million is invested then there is a shift of additional \$40,000 invested in ICA Gruppen. The optimal solutions with box uncertainty parameters differ more than with ellipsoidal uncertainty parameters but this is likely due to that with ellipsoidal uncertainty parameters the optimal solution suggests that almost all initial capital should be invested in one asset and this decision cannot be changed much without violating the constraint $\mathbf{w}^T\mu \geq \theta$. In Section 4.1.3 I make an attempt to explain the differences in optimal solution by considering changes in each asset's risk contribution to the entire portfolio's Expected Shortfall.

With elliptically distributed log-returns the optimal solutions to (3.1) could be compared to the analytical solutions in Appendix C where (1.3) is solved with risk aversion coefficient calculated by (2.11) to find an appropriate sample size. In that case it was concluded that with $D = 15,000$ log-return samples the simulated solutions are close to the analytical solutions. With hybrid GPD-Empirical-GPD distributed log-returns the solutions cannot be compared with analytical solutions. Instead, standard errors to the optimal solutions in Table 3.4 were calculated using the bootstrap method with results presented in Table 3.5. I begin by analyzing the solutions when box uncertainty parameters were used. By comparing the magnitudes of the standard errors with the magnitudes of the asset weights I see that most uncertainty lies in the Hennes & Mauritz B investment. The standard error is larger than the investment and the true optimal weight might differ quite much from the one suggested and it is not possible to reject at a 95% confidence level that the true optimal investment is zero. In this sense, larger sample size D of log-returns is desired to lower the standard error of the optimal weight on that particular asset. On the other hand, the portfolio Expected Shortfall has very small standard error so the uncertainty in the Hennes & Mauritz B investment seems to have little influence on the uncertainty in Expected Shortfall. In this sense larger sample size seems unnecessary. I conclude that solving (3.1) with $15,000$ log-return samples simulated from a normal copula with hybrid GPD-Empirical-GPD marginal distributions yields acceptable accuracy, but increasing the sample size a bit further could increase the accuracy in the Hennes & Mauritz B investment if this is important to the investor.

When it comes to the results in the right columns of Table 3.4 and 3.5 where ellipsoidal uncertainty parameters have been used, the standard errors points out that it is not possible to reject that the true optimal weights for Hennes & Mauritz B and Swedish Match are zero. This means that the optimal solution with ellipsoidal uncertainty parameters could be to invest everything in the Total Bond Index. The result is easily interpreted. With large volatility and correlations and small expected log-returns on the risky assets in a portfolio, a conservative investor would, if he is forced to invest, invest all his capital in the least risky asset, being the Total Bond Index. The standard error for the portfolio Expected Shortfall is furthermore small and I conclude that with ellipsoidal uncertainty parameters, $15,000$ log-return samples simulated from a normal copula with hybrid GPD-Empirical-GPD marginal distributions is large enough to obtain accurate solutions regardless of which parameters are used.

In Section 4.1.1 I concluded that portfolio optimization with Expected Shortfall is sensitive to uncertainties in the parameters μ and Σ. Now, after analyzing the optimal robust solutions with different types of log-return distribution models, I will extend the general conclusion and say something about sensitivity to distribution. It has already been seen that the optimal weight vector is quite insensitive to different elliptical log-return distributions. However, changing from an elliptical to an asymmetric distribution changes the optimal weight vector significantly. I therefore conclude that portfolio optimization with Expected Shortfall is sensitive to distribution uncertainty - if both elliptical and asymmetric distributions are included in the distribution uncertainty set $\mathcal{F}$. This is an interesting conclusions since the robust optimization community has not put much focus on distribution uncertainty. I can thus conclude that problem (2.7) indeed is important since not only parameter uncertainty but also distribution uncertainty affects the optimal solution.

The major drawback with simulating log-returns from a normal copula with hybrid GPD-Empirical-GPD marginal distributions is that the investor is required to use an algorithm such as Algorithm 1. The algorithm is potentially more time consuming than simulating log-returns from a simple multivariate elliptical distribution since the data must be treated in four steps rather than just one. With many risk factors the additional steps 2-

4 in Algorithm 1 can potentially take relatively long time. Therefore, this approach requires careful implementation for high frequency trading optimization but works perfectly fine in this application when trades are made once a day and can be used in favor of the simpler multivariate distribution approach. The reasoning can be taken further and it can be concluded that if it is of great importance to model the log-returns well, then portfolio optimization with Expected Shortfall is appropriate to use since the log-returns can be modeled by copulas with different marginal distributions. The cost is that the optimization problem becomes computationally heavier, is more time consuming and introduces statistical uncertainty since simulations has to be made. If a fast optimization procedure is important then Markowitz mean-variance optimization problem might be better to consider. An exception is if the reference portfolio includes non-linear assets where portfolio optimization with Expected Shortfall must be used or the assets must be approximated by linear functions.

4.1.3 Analysis of Solutions with Euler Risk Decomposition

Up to this point I have motivated the appearance of the optimal weight vectors on basis of the expected log-return and volatility of each asset. Positive expected log-return and small volatility should make an asset attractive, but with ellipsoidal uncertainty parameters, the optimal weight vector suggests to invest a small part of the capital in assets with negative expected log-return. In an attempt to explain this behavior, I will analyze the Euler Allocations defined in section 2.1.3 together with the expected log-return vector, having the marginal Sharpe ratio (2.9) in mind.

In this thesis project, all financial instruments are linear assets, i.e. $F_j(\mathbf{S}(t)) = S_j(t)$ meaning that the sensitivities $\mathbf{U} = \mathbf{w}$ and the Euler Allocation defined in (2.3) for normal distributed log-returns is simplified to

$$\frac{\partial ES_p(X)}{\partial \mathbf{U}} = \frac{\partial ES_p(X)}{\partial \mathbf{w}} = \lambda(\Phi^{-1}(q))(\mathbf{w}^T\Sigma\mathbf{w})^{-1/2}\Sigma\mathbf{w}$$

Table 4.1 presents robust Euler Allocations in the case where $\mathbf{w}$ is solved with box uncertainty parameters. The table also includes $\mu_{\text{min}}^{(\text{box})}$ on yearly basis in percentage to make the analysis more convenient. For ellipsoidal uncertainty parameters, the corresponding Euler Allocations are presented in Table 4.2. Recall that the optimal weight vectors are given in Table 3.1.

Table 4.1: Euler Allocations calculated under the assumption of normal distributed log-returns with box uncertainty parameters and the expected yearly log-return in percent.

Multivariate normal distribution - Box uncertainty

Asset name	$\partial ES_p(X)/\partial w_i$	$\mu_{\min}^{(\text{box})}$ (% per year)
AstraZeneca	0.0139	3.3057
Ericsson A	0.0243	−7.7406
Hennes & Mauritz B	0.0201	4.6686
ICA Gruppen	0.0355	6.0308
Nordea Bank	0.0293	0.4732
SAS	0.0283	−30.1588
SSAB A	0.0314	−27.2710
Swedish Match	0.0461	7.2086
TeliaSonera	0.0196	−4.8182
Volvo	0.0306	−2.3136
Total Bond Index	−0.0004	3.3135

Table 4.2: Euler Allocations calculated under the assumption of normal distributed log-returns with ellipsoidal uncertainty parameters and the expected yearly log-return in percent.

Multivariate normal distribution - Ellipsoidal uncertainty

Asset name	$\partial ES_p(X)/\partial w_i$	$\mu_{\min}^{(\text{ellipsoidal})}$ (% per year)
AstraZeneca	−0.0035	−11.4241
Ericsson A	−0.0160	−42.7518
Hennes & Mauritz B	−0.0131	−20.5889
ICA Gruppen	−0.0072	−12.5113
Nordea Bank	−0.0251	−39.4785
SAS	−0.0193	−80.5988
SSAB A	−0.0293	−73.9089
Swedish Match	−0.0021	−6.9987
TeliaSonera	−0.0121	−29.2233
Volvo	−0.0260	−46.4973
Total Bond Index	0.0053	5.2738

Now recall that the optimal solution with box uncertainty parameters according to Table 3.1 is to invest approximately 50% of the initial capital in the Total Bond Index, approximately 35% in Swedish Match and the remaining 15% in ICA Gruppen. Looking at the yearly expected log-return one might wonder why Hennes & Mauritz B is not invested in, having larger expected log-return than the Total Bond Index. The answer lies in the Euler Allocations, where it is seen that the Total Bond Index works as a hedge against all other assets in the reference portfolio, hence decreasing the total portfolio Expected Shortfall.

With ellipsoidal uncertainty parameters, the optimal solution is according to Table 3.1 to invest approximately 99% in the Total Bond Index and the remaining 1% split between ICA Gruppen, Swedish Match and Hennes & Mauritz B, all three having negative expected log-return. Investing in assets having negative expected log-return could seem counter intuitive but is explained by fact that the assets act as hedges against investments in the Total Bond Index, as seen in Table 4.2. Hence, by allowing the expected total portfolio log-return to decrease a little, still being larger than θ, by investing 1% in assets with negative expected log-return the investor is able to decrease the portfolio's Expected Shortfall.

Equation (2.5) and (2.6) can be used to approximate the Euler Allocations when the log-returns are simulated from the normal copula with hybrid GPD-Empirical-GPD marginal distributions. With $p = 0.01$ and sample size $D = 15,000$ it follows that $VaR_p(X) = L_{151}$, the 151th largest simulated loss. The Euler Allocation for asset i is then the mean of the 150 largest losses for that asset. With box uncertainty parameters the Euler Allocations are presented in Table 4.3 and with ellipsoidal uncertainty parameters the Euler Allocations are presented in Table 4.4. Recall that the optimal weight vectors are presented in Table 3.4.

Table 4.3: Euler Allocations calculated for log-returns simulated from a normal copula with hybrid GPD-Empirical-GPD marginal distributions with box uncertainty parameters and the expected yearly log-return in percent.

Normal copula - Box uncertainty

Asset name	$D_{ES_p(X)}(w_i)$	$\mu_{\text{min}}^{(\text{box})}$ (% per year)
AstraZeneca	0.0097	3.3057
Ericsson A	0.0164	-7.7406
Hennes & Mauritz B	0.0177	4.6686
ICA Gruppen	0.0360	6.0308
Nordea Bank	0.0246	0.4732
SAS	0.0185	-30.1588
SSAB A	0.0281	-27.2710
Swedish Match	0.0518	7.2086
TeliaSonera	0.0122	-4.8182
Volvo	0.0213	-2.3136
Total Bond Index	-0.0003	3.3135

Table 4.4: Euler Allocations calculated for log-returns simulated from a normal copula with hybrid GPD-Empirical-GPD marginal distributions with ellipsoidal uncertainty parameters and the expected yearly log-return in percent.

Normal copula - Ellipsoidal uncertainty

Asset name	$D_{ES_p(X)}(w_i)$	$\mu_{\min}^{(\text{ellipsoidal})}$ (% per year)
AstraZeneca	-0.0008	-11.4241
Ericsson A	-0.0102	-42.7518
Hennes & Mauritz B	-0.0097	-20.5889
ICA Gruppen	-0.0040	-12.5113
Nordea Bank	-0.0203	-39.4785
SAS	-0.0147	-80.5988
SSAB A	-0.0211	-73.9089
Swedish Match	-0.0013	-6.9987
TeliaSonera	-0.0089	-29.2233
Volvo	-0.0219	-46.4973
Total Bond Index	0.0057	5.2738

Similar conclusions that were drawn from analyzing the Euler Allocations with multivariate normal distributed log-returns can be drawn from Table 4.3 and Table 4.4 as well. More interesting is instead to see if the changes in Euler Allocations can motivate the differences in optimal solutions between Table 3.1 and Table (3.4). With ellipsoidal uncertainty parameters it is hard to tell whether the differences in solutions depend on the change of underlying distribution or being an artifact of statistical uncertainty so I focus on the solutions with box uncertainty parameters. As was previously noted, the investment in Swedish Match decreases by 3% when going from the multivariate normal model to the normal copula model. Furthermore, the Euler Allocation increases compared to the multivariate normal model. This seems reasonable that if there is an increase in Euler Allocation, i.e. an increase in the risk contribution, the consequence is that less capital is invested in that asset for a risk averse investor. The same principle holds for the investment changes for Hennes & Mauritz B and the Total Bond Index as well. Surprisingly, the investment in ICA Gruppen increases by 4% when switching to the normal copula even though the risk contribution increases for that asset. This behavior goes against the intuitive conclusion that an increase in risk contribution makes an asset less attractive.

I conclude that combining Euler Allocations and expected log-return, having in mind the marginal Sharpe ratio, can be used as a great tool when analyzing the structure behind optimal portfolio solutions but should be used together with other methods to understand the entire structure.

4.2 Comments on Worst-Case Scenario Based Robust Optimization

This thesis project has, as commonly done in optimization, interpreted robust optimization as worst-case scenario based optimization, presented in Section 3.1.1. This section discuss the effect of this interpretation on the optimization problem.

Firstly, worst-case scenario based robust optimization should be seen as a conservative optimization approach. When the worst-case scenarios from the uncertainty sets $\mathcal{M}, \mathcal{S}$ are used in the optimization problem, it means that the investor has a more negative view on the market than what the empirically estimated parameters suggests. The investor expects the log-returns of the assets in his portfolio to be worse than it has been historically and believes that the assets are more correlated than seen in the covariance matrix. In short, the investor has a conservative view on the market and optimizes his asset allocations according to his view on the market.

One negative aspect with worst-case scenario based robust optimization is that the approach is very sensitive to outliers in the historical data. As an example, with a 95% confidence interval used as uncertainty set, the worst-case scenario is more extreme when outliers are present in the historical data than what the worst-case scenario would be with no outliers. This results in an even more conservative view on the market for the investor. One approach to solve this problem is to search the data for outliers and remove them prior to constructing the 95% confidence interval.

A positive aspect with the worst-case scenario based robust optimization approach is that the solution is quite stable if re-solving the problem periodically as more historical data is available. This is positive in the sense that the investor does not need to re-balance his asset allocations very often and trading fees can be held low for the investor.

An alternative approach to worst-case scenario based robust optimization could be to grid the uncertainty sets and solve the optimization problem for each combination of realizations in each uncertainty set and then analyze how the solution changes as function of parameter location within the uncertainty sets. This approach would however be very time consuming and perhaps computationally impossible to solve with many assets in the portfolio and fine grids of the uncertainty sets.

4.3 More Comments and Further Investigation

The thesis project ends by a few remarks on some alternative approaches that could have been used and discuss areas where further investigation could be of interest.

In this thesis project the worst-case box uncertainty log-return $\mu_{\min}^{(\text{box})}$

is constructed as in (3.2) by subtracting 20% of each element in the vector. The result is then compared to the lower limit of the 95% confidence interval to see whether the worst-case scenario is reasonable or not. An alternative approach is to define $\mu_{\min}^{(\text{box})}$ directly as the lower limit of the 95% confidence interval. With the particular historical data in this thesis project this approach resulted in that all elements in $\mu_{\min}^{(\text{box})}$ were negative. This yields a robust optimization problem without solutions since the constraint $\mathbf{w}^T\mu_{\min} \geq \theta$ cannot be satisfied with long positions. In light of this, this alternative approach was not used in the thesis project but could be more intuitive to use in other applications or with other reference portfolios.

All solutions in this thesis project are based on the same reference portfolio with assets listed in Table 1.1. It would however be interesting to analyze how different reference portfolios consisting of other types of financial assets affect the optimal solution. For instance, how is the optimal solution affected if the amount of financial assets included in the reference portfolio increases? It seems reasonable that larger reference portfolios would decrease the Expected Shortfall since the optimization algorithm has more assets to choose from. However, as has been seen in all solutions throughout this thesis project, the optimal solution often turns out to be to investment in a small sub group of the possible assets. Therefore, including more assets in the reference portfolio does not necessarily decrease the risk and could at worst result in a less time efficient program that can be crucial for the investor. What can be said is that including more assets in the reference portfolio does at least not increase the financial risk. The perhaps most interesting area of further investigation would be to extend the reference portfolio to include other types of financial assets such as foreign currencies or non-linear financial derivatives such as options. This would introduce possibilities of more complex hedging opportunities than possible with linear assets. It is furthermore interesting because non-linear portfolios makes full use of the risk measure Expected Shortfall since Markowitz mean-variance optimization problem cannot handle non-linear financial assets. In that case the non-linear assets have to be approximated by linear functions. Thirdly it would be interesting from a practical point of view to include non-linear assets since it is common that investors have non-linear assets in their portfolios.

In this thesis project, the p-value has been held constant to 1% but can of course be altered as well. Decreasing the p-value should however only increase the Expected Shortfall or the investor must decrease the threshold for acceptable expected log-return to keep the risk constant. Hence the investor has to find a p-level where he feels confident with both the expected portfolio log-return and the risk he is exposed to.

Another interesting area of further investigation is to study the effect of different copulas on the optimal solution. In this thesis project, the normal copula was used together with hybrid GPD-Empirical-GPD marginal distributions but the dependency structure between different log-returns could just as well had been modeled by another copula. However, as was mentioned earlier, according to Skoglund and Nyström the choice of copula has a second-order effect on the model accuracy when modeling market risk factors. If a Student's t copula were used instead, the results had probably only been marginally improved.

The last area that I will mention where further investigation could be made is in the decision of the upper and lower thresholds for the generalized Pareto distributions. In this thesis project the thresholds were chosen as the 90% and 10% empirical quantiles respectively but no investigation was made to establish whether these thresholds were good representatives of where the tails begin. A better and more consistent approach to find the thresholds is to use a parametric estimation method such as the Hill's tail-index estimator [4]. This could improve the results a bit further but also make the program more time consuming.

Different approaches taken in a portfolio optimization problem, whether it is about which problem formulation to use, which models to use or which parameters to use, is ultimately up to the investor to decide and a trade-off between model accuracy and time efficiency/computational possibility always has to be considered.

Appendix A

Optimization Parameters

When solving the portfolio optimization problem (2.16) the parameters μ and Σ are needed for instance in the process of simulating log-returns. Since the portfolio lives for one day, μ is the expected daily log-return and Σ is the covariance matrix for daily log-returns. In the thesis project I begin by solving the portfolio optimization problem with empirically estimated parameters from historical data and then solve the robust portfolio optimization problem (3.1) after manipulating the empirical parameters to obtain two different kinds of worst-case scenario parameters. The worst-case scenario box uncertainty parameters are defined by (3.2) and (3.3) and the worst-case scenario ellipsoidal uncertainty parameters are defined by (3.4) and (3.5). In this Appendix I clarify the parameters by writing out their explicit numerical values.

The numerical values of the empirical parameters estimated for the time period January 2, 2007 until January 22, 2016 from daily log-return data to the assets in the reference portfolio are

$$\widehat{\mu} = 10^{-4}\begin{bmatrix}1.6463 & -2.5699 & 2.3250 & 3.0034 & 0.2357 & -10.0129 & -9.0541 & 3.5899 & -1.5997 & -0.7681 & 1.6501\end{bmatrix}^T$$

$$\widehat{\Sigma} = 10^{-4}\begin{bmatrix}
2.2076 & 0.8564 & 0.7462 & 0.5434 & 0.8873 & 0.9597 & 0.9827 & 0.6127 & 0.7489 & 0.8089 & -0.0201 \\
0.8564 & 6.7548 & 1.4816 & 0.8375 & 2.3592 & 1.8613 & 2.5425 & 0.8313 & 1.3073 & 2.5546 & -0.0688 \\
0.7462 & 1.4816 & 2.6549 & 0.8594 & 1.9553 & 1.7679 & 2.1897 & 0.7987 & 1.2000 & 2.0837 & -0.0606 \\
0.5434 & 0.8375 & 0.8594 & 3.2819 & 1.2290 & 1.2987 & 1.4293 & 0.4836 & 0.8364 & 1.3421 & -0.0424 \\
0.8873 & 2.3592 & 1.9553 & 1.2290 & 4.9191 & 2.7372 & 3.5812 & 1.0444 & 1.8956 & 3.3907 & -0.1052 \\
0.9597 & 1.8613 & 1.7679 & 1.2987 & 2.7372 & 13.5689 & 3.0979 & 0.7071 & 1.5030 & 2.9935 & -0.0821 \\
0.9827 & 2.5425 & 2.1897 & 1.4293 & 3.5812 & 3.0979 & 8.3021 & 1.0916 & 1.9876 & 4.4259 & -0.1217 \\
0.6127 & 0.8313 & 0.7987 & 0.4836 & 1.0444 & 0.7071 & 1.0916 & 2.4675 & 0.7539 & 1.0733 & -0.0271 \\
0.7489 & 1.3073 & 1.2000 & 0.8364 & 1.8956 & 1.5030 & 1.9876 & 0.7539 & 2.6576 & 1.9369 & -0.0538 \\
0.8089 & 2.5546 & 2.0837 & 1.3421 & 3.3907 & 2.9935 & 4.4259 & 1.0733 & 1.9369 & 5.8128 & -0.1094 \\
-0.0201 & -0.0688 & -0.0606 & -0.0424 & -0.1052 & -0.0821 & -0.1217 & -0.0271 & -0.0538 & -0.1094 & 0.0195
\end{bmatrix}.$$

The numerical values of the box uncertainty parameters defined by (3.2) and (3.3) are

$$\mu_{\min}^{(\mathrm{box})} = 10^{-4}\begin{bmatrix}1.3170 & -3.0839 & 1.8600 & 2.4027 & 0.1885 & -12.0155 & -10.8649 & 2.8719 & -1.9196 & -0.9217 & 1.3201\end{bmatrix}^T$$

$$\Sigma_{\max}^{(\mathrm{box})} = 10^{-4}\begin{bmatrix}
3.3115 & 1.2960 & 0.9321 & 0.8164 & 1.0944 & 1.7779 & 1.2788 & 0.8737 & 0.9871 & 0.9052 & -0.0278 \\
1.2960 & 8.4445 & 2.3581 & 1.4760 & 4.2318 & 3.2969 & 4.3982 & 1.5645 & 2.1447 & 4.3579 & -0.0883 \\
0.9321 & 2.3581 & 3.9946 & 1.3956 & 3.1897 & 3.1002 & 3.6072 & 1.2078 & 1.7751 & 3.4598 & -0.0772 \\
0.8164 & 1.4760 & 1.3956 & 5.1444 & 1.9984 & 2.4270 & 2.4435 & 0.7163 & 1.3759 & 2.4078 & -0.0562 \\
1.0944 & 4.2318 & 3.1897 & 1.9984 & 8.9431 & 5.1257 & 6.1330 & 1.8146 & 3.0073 & 6.0763 & -0.1361 \\
1.7779 & 3.2969 & 3.1002 & 2.4270 & 5.1257 & 18.4093 & 5.4142 & 1.4777 & 2.9255 & 5.3976 & -0.1051 \\
1.2788 & 4.3982 & 3.6072 & 2.4435 & 6.1330 & 5.4142 & 13.4753 & 1.8367 & 2.9929 & 7.6285 & -0.1678 \\
0.8737 & 1.5645 & 1.2078 & 0.7163 & 1.8146 & 1.4777 & 1.8367 & 3.7902 & 1.1495 & 1.7437 & -0.0398 \\
0.9871 & 2.1447 & 1.7751 & 1.3759 & 3.0073 & 2.9255 & 2.9929 & 1.1495 & 4.6483 & 3.0786 & -0.0602 \\
0.9052 & 4.3579 & 3.4598 & 2.4078 & 6.0763 & 5.3976 & 7.6285 & 1.7437 & 3.0786 & 9.6670 & -0.1439 \\
-0.0278 & -0.0883 & -0.0772 & -0.0562 & -0.1361 & -0.1051 & -0.1678 & -0.0398 & -0.0602 & -0.1439 & 0.0234
\end{bmatrix}.$$

The numerical values of the ellipsoidal uncertainty parameters defined by (3.4) and (3.5) are

$$\mu_{\min}^{(\mathrm{ellipsoidal})} = 10^{-4}\begin{bmatrix}-4.5514 & -17.0326 & -8.2027 & -4.9846 & -15.7285 & -32.1111 & -29.4458 & -2.7883 & -11.6428 & -18.5248 & 2.1011\end{bmatrix}^T$$

$$\Sigma_{\max}^{(\mathrm{ellipsoidal})} = 10^{-4}\begin{bmatrix}
4.5494 & 1.7649 & 1.5377 & 1.1198 & 1.8286 & 1.9776 & 2.0250 & 1.2627 & 1.5434 & 1.6669 & -0.0413 \\
1.7649 & 13.9200 & 3.0532 & 1.7258 & 4.8617 & 3.8357 & 5.2395 & 1.7130 & 2.6940 & 5.2644 & -0.1417 \\
1.5377 & 3.0532 & 5.4712 & 1.7710 & 4.0294 & 3.6433 & 4.5124 & 1.6459 & 2.4729 & 4.2941 & -0.1248 \\
1.1198 & 1.7258 & 1.7710 & 6.7631 & 2.5327 & 2.6763 & 2.9455 & 0.9966 & 1.7237 & 2.7658 & -0.0873 \\
1.8286 & 4.8617 & 4.0294 & 2.5327 & 10.1372 & 5.6408 & 7.3801 & 2.1522 & 3.9064 & 6.9874 & -0.2168 \\
1.9776 & 3.8357 & 3.6433 & 2.6763 & 5.6408 & 27.9625 & 6.3842 & 1.4572 & 3.0972 & 6.1689 & -0.1693 \\
2.0250 & 5.2395 & 4.5124 & 2.9455 & 7.3801 & 6.3842 & 17.1087 & 2.2496 & 4.0960 & 9.1208 & -0.2508 \\
1.2627 & 1.7130 & 1.6459 & 0.9966 & 2.1522 & 1.4572 & 2.2496 & 5.0849 & 1.5537 & 2.2118 & -0.0558 \\
1.5434 & 2.6940 & 2.4729 & 1.7237 & 3.9064 & 3.0972 & 4.0960 & 1.5537 & 5.4767 & 3.9915 & -0.1109 \\
1.6669 & 5.2644 & 4.2941 & 2.7658 & 6.9874 & 6.1689 & 9.1208 & 2.2118 & 3.9915 & 11.9787 & -0.2254 \\
-0.0413 & -0.1417 & -0.1248 & -0.0873 & -0.2168 & -0.1693 & -0.2508 & -0.0558 & -0.1109 & -0.2254 & 0.0402
\end{bmatrix}.$$

By comparing the different parameters it is easy to see that the worst-case scenario parameters indeed are worse than the empirically estimated parameters. The worst-case scenario expected log-return vectors are more negative than the empirically estimated expected log-return vector and the worst-case scenario covariance matrices have larger pairwise covariances than the empirically estimated covariance matrix. It is also easy to observe that the ellipsoidal uncertainty parameters are more extreme than the box uncertainty parameters. Hence, if the box uncertainty parameters are seen as stress tests of the empirically estimated parameters, then the ellipsoidal uncertainty parameters could be seen as parameters from a large global financial crisis.

Appendix B

Convergence as Function of Sample Size D

This chapter studies how large the sample size D of simulated log-returns must be for the solutions to the portfolio optimization problem (2.16) to be precise and have small statistical uncertainty. While holding all other parameters constant, let us first increase the sample size until the solution weight vector no longer have visible oscillations and then find the sample size D^* where further increments no longer improve the solution accuracy much. Figure B.1 presents the solution weight vector to the problem (2.16) as function of sample size D. The simulated log-returns come from a multivariate normal distribution with empirically estimated parameters $\widehat{\mu}$ and $\widehat{\Sigma}$.

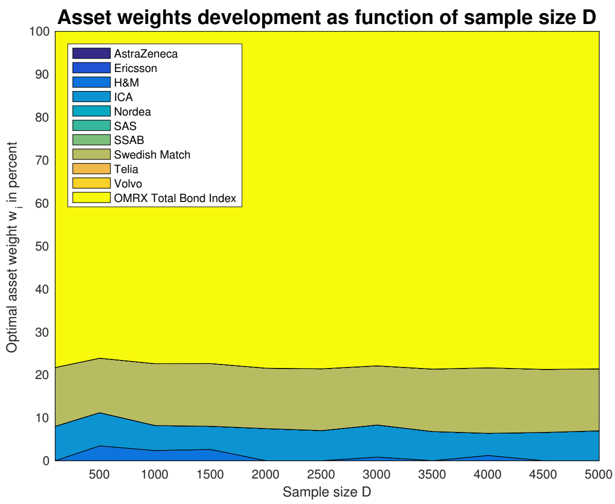

Figure B.1: The optimal weight vector as function of the sample size D of simulated log-returns.

From the figure it seems as the solution converges quite fast, perhaps as early as for $D^* = 1000$ samples. However, if convergence of Expexted Shortfall as function of sample size is investigated the behavior is different. This behavior is presented in Figure B.2 and depicts approximate 95% confidence intervals for Expected Shortfall when solving (2.16) 100 times each for increasing sample size. With multivariate normal distributed log-returns the Expected Shortfall seems to have converged with narrow confidence interval for approximately $D^* = 10,000$ and for Student's t distributed log-returns with 2.1 degrees of freedom the threshold is approximately $D^* = 15,000$ samples. Both observations are far above the initial one of $D^* = 1000$. For an investor, accuracy in both optimal weight vector and Expected Shortfall is important and the sample size must be chosen sufficiently large for both factors to be accurate estimates of the true values. Therefore, when solving the portfolio optimization problem I should use a step size of at least 10,000 when simulating log-returns from the multivariate normal distribution and 15,000 when simulating from the Student's t distribution. For simplicity I use $D = 15,000$ in both cases throughout this thesis project.

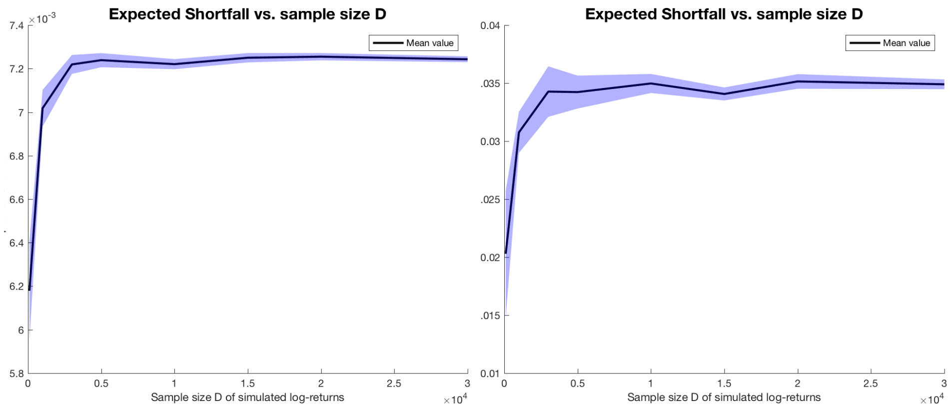

Figure B.2: 95% confidence intervals for Expected Shortfall as function of the sample size D for (left) multivariate normal and (right) multivariate Student's t distributed log-returns with 2.1 degrees of freedom.

Appendix C

Reference Solutions

This Appendix presents reference solutions to the robust portfolio optimization problem (3.1) when instead Markowitz mean-variance optimization problem (1.3) is solved. By doing this, statistical uncertainty is avoided. The two optimization problems are connected through the risk aversion coefficient c which for elliptically distributed log-returns is calculated as (2.11). The reference portfolio Expected Shortfall in the last row of each table is calculated using (2.17) and (2.18).

C.1 Normal Distributed Log-returns

Under the assumption of multivariate normal distributed log-returns, the risk aversion coefficient c in Markowitz mean-variance optimization problem is already calculated in (2.12) to be $c = 5.33$. With the same worst-case scenario parameters μ_{min} and Σ_{max} used to obtain the two solutions in Table 3.1, the corresponding reference solutions to Markowitz mean-variance problem (1.3) are presented in Table C.1.

Table C.1: Solutions to (1.3) with multivariate normal distributed log-returns. Left column with box uncertainty parameters and right column with ellipsoidal uncertainty parameters.

	Box uncertainty	**Ellipsoidal uncertainty**
Asset name	Weight	Weight
AstraZeneca	0	0.0015
Ericsson A	0	0.0004
Hennes & Mauritz B	0	0.0006
ICA Gruppen	0.1479	0.0015
Nordea Bank	0	0.0004
SAS	0	0
SSAB A	0	0.0002
Swedish Match	0.3506	0.0029
TeliaSonera	0	0.0004
Volvo	0	0.0004
Total Bond Index	0.5015	0.9917
Expected Shortfall	0.0213	0.0052

C.2 Student's t Distributed Log-returns

Under the assumption of multivariate Student's t distributed log-returns, the risk aversion coefficient c is defined by (2.11), and the Student's t quantile function depends on the degrees of freedom. Table C.2 presents the different values of the risk aversion coefficient for $\nu = 2.1, 3.58, 10, 20$.

Table C.2: The risk aversion coefficient calculated numerically using (2.11) with the Student's t quantile function for different degrees of freedom.

Degrees of freedom ν	2.1	3.58	10	20
Risk aversion coefficient c	25.2278	11.5312	6.7265	5.9538

Note that as the degrees of freedom increases, the risk aversion coefficient decreases. This behavior is expected and c should converge to the risk aversion coefficient for normal distributed log-returns, i.e. converge to $c = 5.33$.

With the same box uncertainty parameters that were used to obtain the solutions in Table 3.2, the corresponding reference solutions to Markowitz mean-variance optimization problem are presented in Table C.3. Table C.4 presents in turn the reference solutions to the simulated solutions with ellipsoidal uncertainty parameters in Table 3.3.

Table C.3: Solutions to (1.3) with multivariate Student's t distributed log-returns with different degrees of freedom and box uncertainty parameters.

Asset name	$\nu = 2.1$ Weight	$\nu = 3.58$ Weight	$\nu = 10$ Weight	$\nu = 20$ Weight
AstraZeneca	0	0	0	0
Ericsson A	0	0	0	0
Hennes & Mauritz B	0	0	0	0
ICA Gruppen	0.1479	0.1479	0.1479	0.1479
Nordea Bank	0	0	0	0
SAS	0	0	0	0
SSAB A	0	0	0	0
Swedish Match	0.3506	0.3506	0.3506	0.3506
TeliaSonera	0	0	0	0
Volvo	0	0	0	0
Total Bond Index	0.5015	0.5015	0.5015	0.5015
Expected Shortfall	0.0914	0.0447	0.0265	0.0235

Table C.4: Solutions to (1.3) with multivariate Student's t distributed log-returns with different degrees of freedom and ellipsoidal uncertainty parameters.

Asset name	$\nu = 2.1$ Weight	$\nu = 3.58$ Weight	$\nu = 10$ Weight	$\nu = 20$ Weight
AstraZeneca	0.0001	0.0012	0.0014	0.0015
Ericsson A	0	0.0003	0.0004	0.0004
Hennes & Mauritz B	0.0027	0.0007	0.0006	0.0006
ICA Gruppen	0.0035	0.0020	0.0016	0.0016
Nordea Bank	0.0001	0.0004	0.0004	0.0004
SAS	0	0	0	0
SSAB A	0	0.0001	0.0002	0.0002
Swedish Match	0.0044	0.0036	0.0030	0.0029
TeliaSonera	0	0.0003	0.0003	0.0003
Volvo	0	0.0004	0.0004	0.0004
Total Bond Index	0.9892	0.9910	0.9916	0.9916
Expected Shortfall	0.0238	0.0112	0.0065	0.0058

Appendix D

GARCH and GPD Parameters

The generalized Pareto distribution $G_{\gamma,\beta}(x)$ is given by

$$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}}, \quad x \geq 0$$

for some shape parameter $\gamma > 0$ and scale parameter $\beta > 0$ and

$$G_{\gamma,\beta}(x) = 1 - e^{-\frac{x}{\beta}}, \quad x \geq 0$$

if $\gamma = 0$.

Modeling excesses in the tail of a distribution over a suitably high threshold u with a generalized Pareto distribution is possible if the underlying data is independent and identically distributed. The log-returns in the reference portfolio can only be assumed to be weakly dependent and close to independent and identically distributed and typically suffers from the stylized facts presented in Section 3.1.3.3. To obtain independent and identically distributed data the historical log-returns to each asset in the reference portfolio are filtered with fitted GARCH(1,1) models, also defined in Section 3.1.3.3. The GARCH(1,1) parameters were numerically estimated with the Maximum Likelihood method using the function `estimate` in Matlab and are presented in Table D.1 together with other statistical quantities of interest.

Table D.1: Maximum Likelihood estimated parameters to the GARCH(1,1) models used for filtering historical log-returns to standardized residuals.

Asset name	Parameter estimate	Standard error	t statistic
AstraZeneca	$\alpha_0 = 7.28056 \cdot 10^{-6}$	$5.95667 \cdot 10^{-7}$	12.2225
	$\alpha_1 = 0.0931352$	0.00721758	12.9039
	$\beta_1 = 0.875473$	0.00857937	102.044
Ericsson A	$\alpha_0 = 0.000272561$	$2.47196 \cdot 10^{-5}$	11.0261
	$\alpha_1 = 0.165719$	0.0223635	7.41023
	$\beta_1 = 0.411872$	0.0546948	7.53037
Hennes & Mauritz B	$\alpha_0 = 1.24182 \cdot 10^{-5}$	$1.52487 \cdot 10^{-6}$	8.14376
	$\alpha_1 = 0.0763801$	0.00970769	7.868
	$\beta_1 = 0.874378$	0.0140866	62.0714
ICA Gruppen	$\alpha_0 = 1.56923 \cdot 10^{-5}$	$2.08814 \cdot 10^{-6}$	7.51496
	$\alpha_1 = 0.168988$	0.0114856	14.713
	$\beta_1 = 0.794041$	0.0133304	59.566
Nordea Bank	$\alpha_0 = 4.66648 \cdot 10^{-6}$	$1.23123 \cdot 10^{-6}$	3.7901
	$\alpha_1 = 0.0646265$	0.00657526	9.82874
	$\beta_1 = 0.923464$	0.00792685	116.498
SAS	$\alpha_0 = 4.38621 \cdot 10^{-5}$	$5.27758 \cdot 10^{-6}$	8.31103
	$\alpha_1 = 0.118331$	0.0067498	17.5311
	$\beta_1 = 0.858769$	0.00767209	111.934
SSAB A	$\alpha_0 = 6.72565 \cdot 10^{-6}$	$8.38453 \cdot 10^{-7}$	8.0215
	$\alpha_1 = 0.041997$	0.00414664	10.1279
	$\beta_1 = 0.948355$	0.00437622	216.707
Swedish Match	$\alpha_0 = 1.22517 \cdot 10^{-5}$	$1.64289 \cdot 10^{-6}$	7.45739
	$\alpha_1 = 0.103236$	0.0101418	10.1793
	$\beta_1 = 0.847721$	0.0149104	56.8544
TeliaSonera	$\alpha_0 = 4.23416 \cdot 10^{-6}$	$6.64369 \cdot 10^{-7}$	6.37321
	$\alpha_1 = 0.079082$	0.0052433	15.0825
	$\beta_1 = 0.909637$	0.00517679	175.715
Volvo	$\alpha_0 = 7.43774 \cdot -6$	$1.28612 \cdot -6$	5.78307
	$\alpha_1 = 0.0647348$	0.00877252	7.37927
	$\beta_1 = 0.921605$	0.00983831	93.6752
Total Bond Index	$\alpha_0 = 2 \cdot 10^{-7}$	$9.32652 \cdot 10^{-8}$	2.14442
	$\alpha_1 = 0.0865343$	0.0102208	8.46645
	$\beta_1 = 0.813981$	0.00897679	90.6761

Figure D.1 depicts realizations and histograms of the standardized residuals in the two upper plots and sample autocorrelation functions of squared log-returns and squared standardized residuals in the two lower plots for each asset in the reference portfolio. According to stylized fact 2 for financial risk factor time series the squared log-returns should show sign of autocorrelation which can be observed for every asset in the lower left plots. After filtering the log-returns the goal is to have squared residuals without autocorrelation. This seems to hold for the standardized residuals by observing the lower right plot for every asset and hence it is concluded that the GARCH(1,1) filtration has been successful. Looking now at the histograms of the standardized residuals, the distributions are leptokurtic, or heavy tailed, and hence the standard normal distribution is not appropriate to use as model distribution.

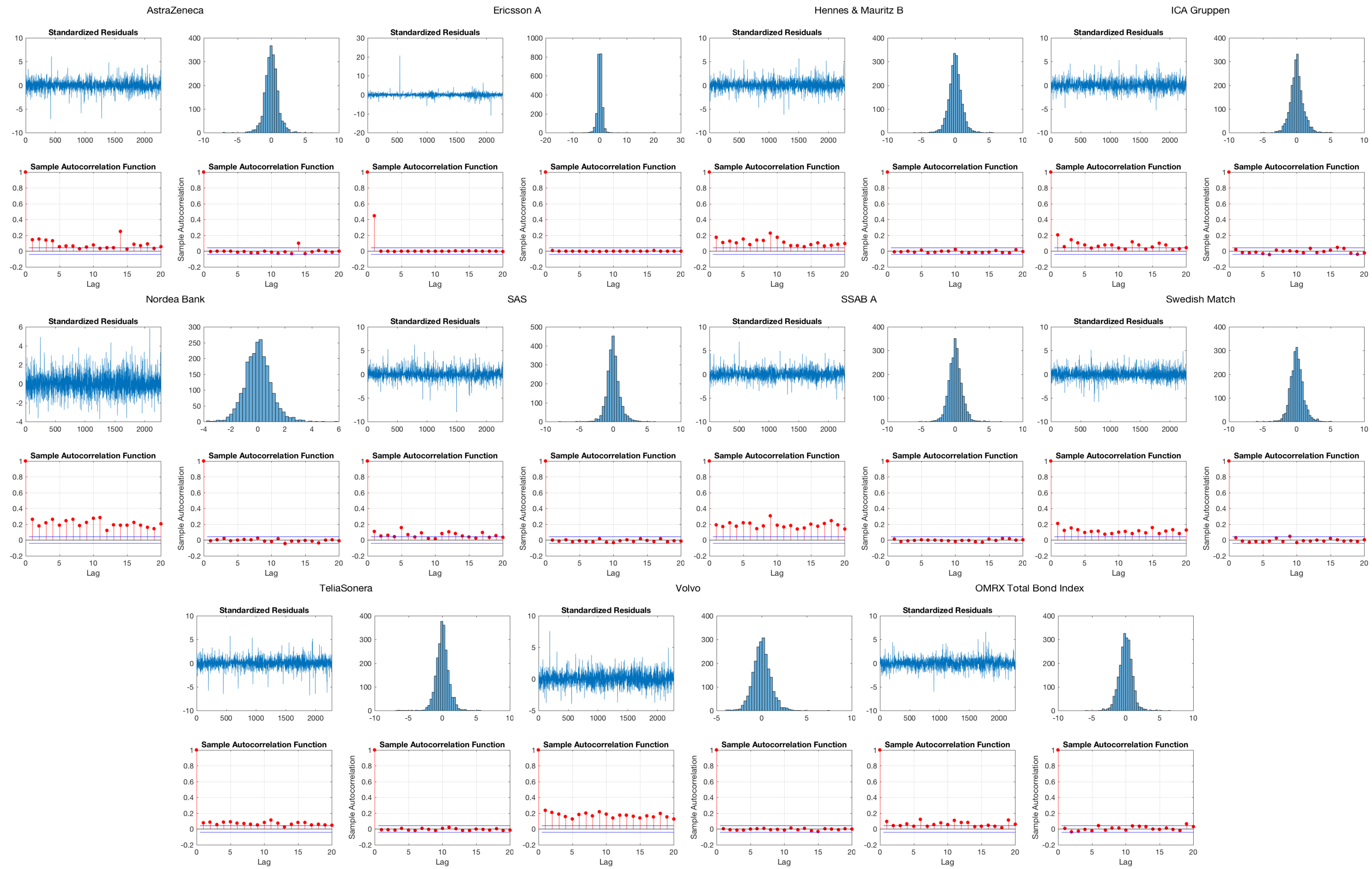

Figure D.1: Each sub figure consists of realizations and histograms of the standardized residuals and sample autocorrelation functions of squared log-returns and squared standardized residuals.

The same conclusion can be drawn by observing Figure D.2 which shows Quantile-Quantile plots for the empirical residual quantile function plotted against the standard normal quantile function. A good fit should produce a straight line and a reverted S-shaped curve indicates that the tails are heavier than for the standard normal distribution. As can be seen, reverted S-shapes appear in every plot indicating that the standardized residuals have distributions with fatter tails than the standard normal distribution. Notice also that the S-shapes are asymmetric, meaning that the two tails of the residual distribution are of different size and length. Therefore, an elliptical distribution with fat tails, for instance the Student's t distribution, would not capture the entire empirical residual distribution since elliptical distributions are symmetric. Thus generalized Pareto distributions are appropriate to model the tails of the residual distributions since this can give an asymmetric distribution.

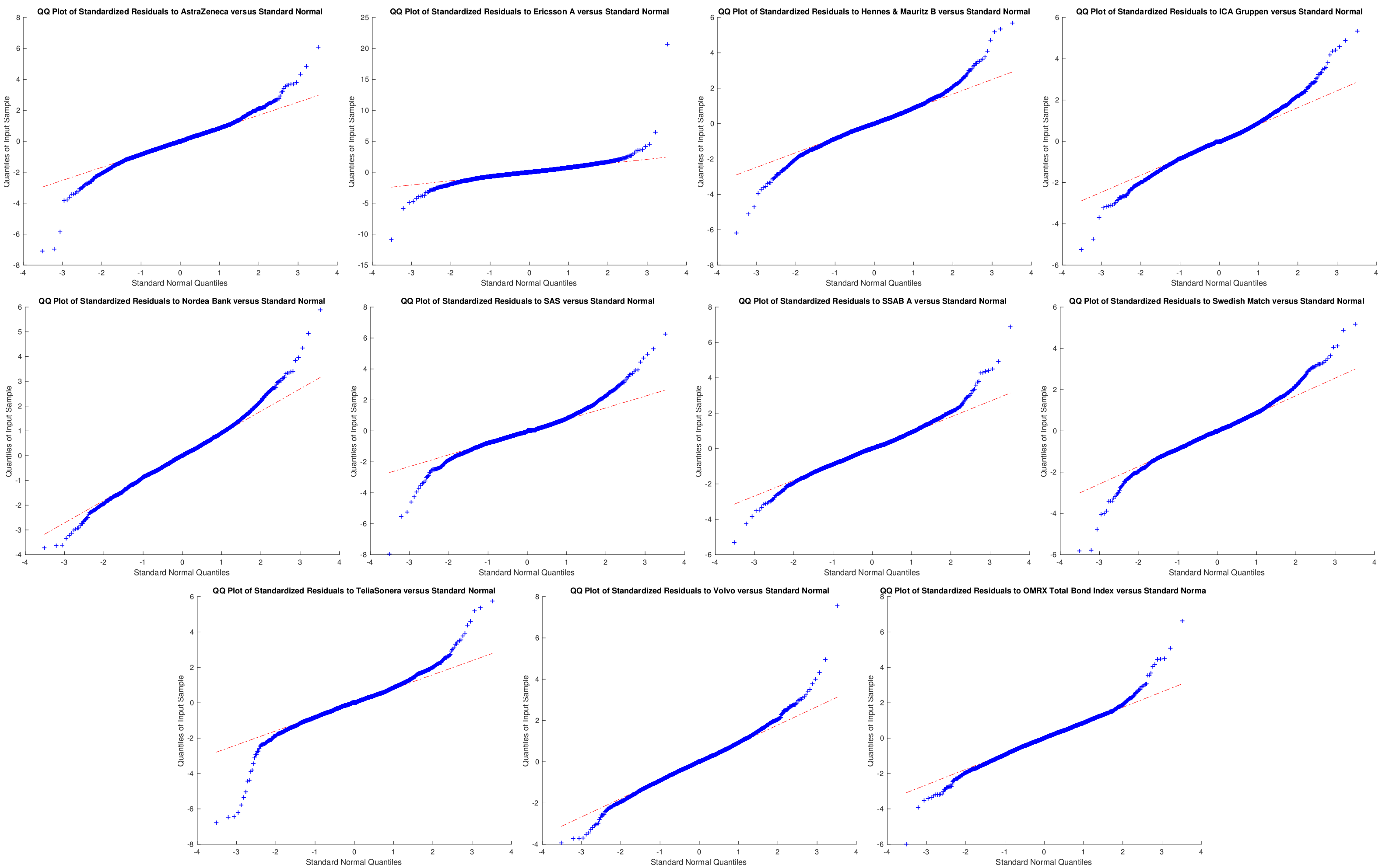

Figure D.2: Each sub figure consists of Quantile-Quantile plots of the empirical quantile function for the standardized residuals versus the standard normal quantile function.

Table D.2 presents Maximum Likelihood estimated parameters to the modeled generalized Pareto distributions with tail thresholds chosen as $u^h = F_n^{-1}(0.90)$ and $u^l = F_n^{-1}(0.10)$ respectively for each asset in the reference portfolio.

Table D.2: Maximum Likelihood estimated generalized Pareto distribution parameters to standardized residuals calculated for each asset in the reference portfolio. The thresholds are chosen as $u^h = F_n^{-1}(0.90)$ and $u^l = F_n^{-1}(0.10)$.

Asset name	Threshold	$\widehat{\gamma}$	$\widehat{\beta}$
AstraZeneca	$u^l = -1.0957$	0.1348	0.5897
	$u^h = 1.0984$	−0.0127	0.7010
Ericsson A	$u^l = -0.9369$	0.1829	0.5844
	$u^h = 0.9762$	0.2569	0.4286
Hennes & Mauritz B	$u^l = -1.1497$	0.1112	0.5477
	$u^h = 1.1377$	0.0701	0.6366
ICA Gruppen	$u^l = -1.1299$	−0.0201	0.6120
	$u^h = 1.2474$	0.0527	0.6210
Nordea Bank	$u^l = -1.1740$	−0.0745	0.5647
	$u^h = 1.2112$	0.0104	0.6627
SAS	$u^l = -1.0576$	0.1897	0.5085
	$u^h = 1.1622$	0.0712	0.6935
SSAB A	$u^l = -1.1388$	0.0557	0.5270
	$u^h = 1.2375$	0.1244	0.5508
Swedish Match	$u^l = -1.1020$	0.1509	0.5094
	$u^h = 1.1974$	−0.0139	0.6881
TeliaSonera	$u^l = -1.0573$	0.2619	0.4897
	$u^h = 1.1462$	0.0709	0.6137
Volvo	$u^l = -1.1655$	−0.0123	0.5449
	$u^h = 1.2166$	0.0637	0.5970
Total Bond Index	$u^l = -1.2251$	0.0695	0.5012
	$u^h = 1.1286$	0.2358	0.4493

The thresholds u^l and u^h and parameters $\widehat{\gamma}$ and $\widehat{\beta}$ are used in Section 3.1.3.4 when solving the robust optimization problem (3.1). They are used in Algorithm 1 step 3 when calculating F_i^{-1} being the inverse of the hybrid GPD-Empirical-GPD distribution function (3.7).

Figure D.3 displays the generalized Pareto distributions with parameters in Table D.2 together with the empirical cumulative distribution for the standardized residuals. The estimated tail distributions are found to fit the empirical residual data well in the tails and the hybrid GPD-Empirical-GPD distributions will reflect the entire residual distribution much more accurate than the standard normal or Student's t distribution.

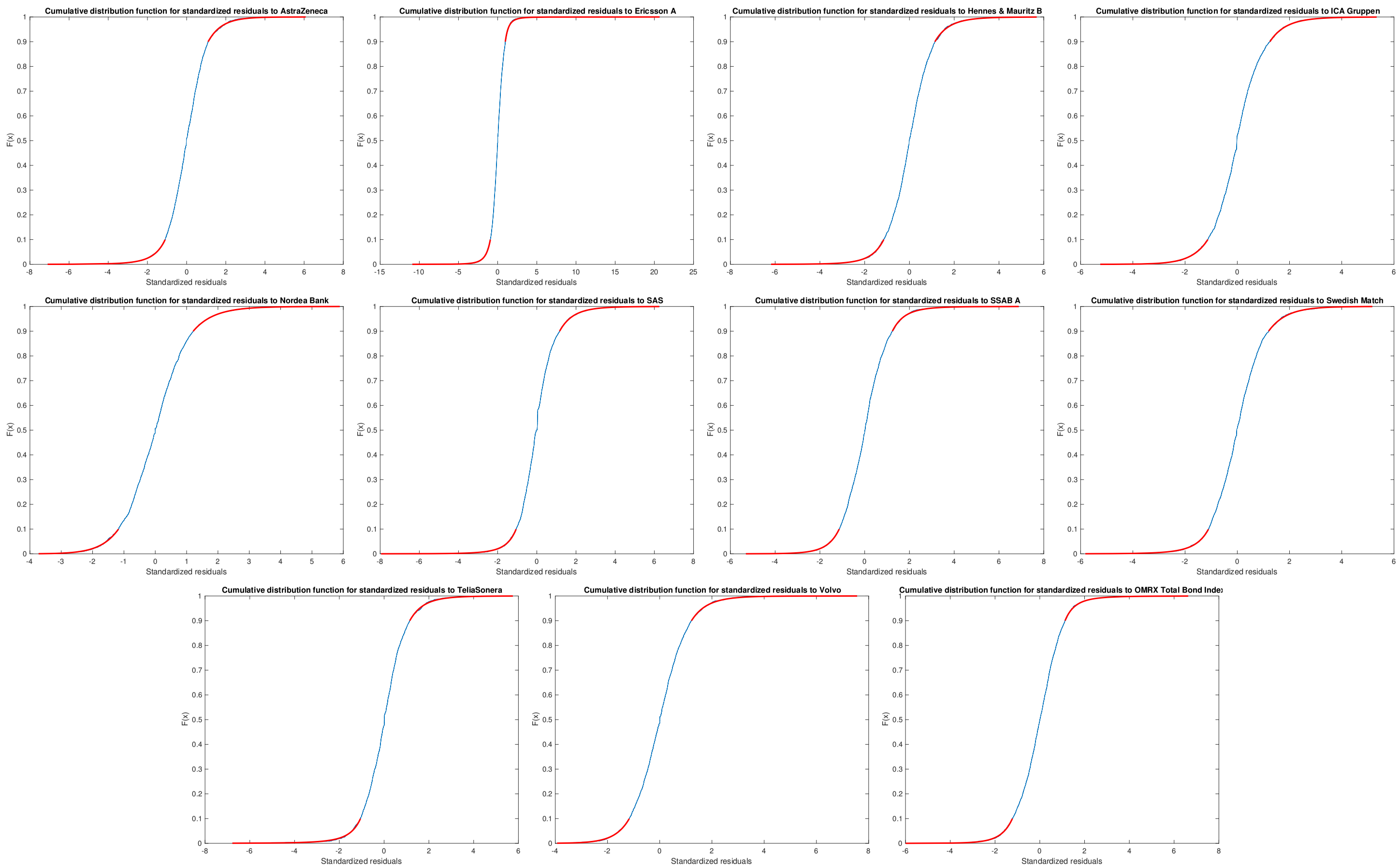

Figure D.3: Empirical cumulative distribution functions for the standardized residuals for each asset in the reference portfolio together with estimated generalized Pareto distributions in the distribution tails.

Bibliography

[1] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. Springer Science & Business Media, 2006.

[2] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004.

[3] Igor Griva, Stephen G Nash, and Ariela Sofer. *Linear and nonlinear optimization*. Siam, 2009.

[4] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174. Institute of Mathematical Statistics, 1975.

[5] Henrik Hult, Filip Lindskog, Ola Hammarlid, and Carl Johan Rehn. *Risk and portfolio analysis: Principles and methods*. Springer Science & Business Media, 2012.

[6] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[7] Marcus Josefsson. A copula evt-based approach for measuring tail related risk: Applied on the Swedish market. Master's thesis, Royal Institute of Technology, 2004.

[8] Patrick Kingsley. Financial crisis: timeline. *The Guardian*, 7 August 2012.

[9] Damon Levine. Modeling tail behavior with extreme value theory. *Risk Management*, (17):14–18, September 2009.

[10] Filip Lindskog, Alexander Mcneil, and Uwe Schmock. *Kendall's tau for elliptical distributions*. Springer, 2003.

[11] Miguel Sousa Lobo and Stephen Boyd. The worst-case risk of a portfolio. *Unpublished manuscript. Available from http://faculty.fuqua.duke.edu/%7Emlobo/bio/researchfiles/rsk-bnd.pdf*, 2000.

[12] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91. American Finance Association, Wiley, 1952.

[13] Kaj Nyström and Jimmy Skoglund. Efficient filtering of financial time series and extreme value theory. *Journal of Risk*, 7(2):63–84, 2005.

[14] Riccardo Rebonato and Peter Jäckel. The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Available at SSRN 1969689*, 1999.

[15] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[16] Bernd Scherer. *Portfolio Construction and Risk Budgeting, 5 ed.* Risk Books, London, Great Britain, 2015.

[17] William F Sharpe. The sharpe ratio. *The journal of portfolio management*, 21(1):49–58. Institutional Investor Journals, 1994.

[18] Jimmy Skoglund and Wei Chen. Risk contributions, information and reverse stress testing. *The Journal of Risk Model Validation*, 3(2):61–77. Incisive Media Plc, 2009.

[19] Jimmy Skoglund and Wei Chen. *Financial Risk Management - Applications in Market, Credit, Asset and Liability Management and Firmwide Risk*. Wiley Finance Series, 2015.

[20] Dirk Tasche. Risk contributions and performance measurement. *Report of the Lehrstuhl für mathematische Statistik, TU München.* Citeseer, 1999.

[21] Dirk Tasche. Conditional expectation as quantile derivative. *arXiv preprint math/0104190*, 2001.

[22] Reha H Tütüncü and M Koenig. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187. Springer, 2004.

[23] Kai Ye, Panos Parpas, and Berç Rustem. Robust portfolio optimization: a conic programming approach. *Computational Optimization and Applications*, 52(2):463–481. Springer, 2012.