An Empirical Study of Modern Portfolio Optimization

ERIK LAGERSTRÖM
MICHAEL MAGNE SCHRAB
An Empirical Study of Modern Portfolio Optimization

ERIK LAGERSTRÖM
MICHAEL MAGNE SCHRAB
Abstract

*Mean variance optimization* has shortcomings making the strategy far from optimal from an investor’s perspective. The purpose of the study is to conduct an empirical investigation as to how modern methods of portfolio optimization address the shortcomings associated with mean variance optimization. *Equal risk contribution*, the *Most diversified portfolio* and a modification of the *Minimum variance portfolio* are considered as alternatives to the mean variance model. Portfolio optimization models introduced are explained in detail and solved using the optimization algorithms *Cyclical coordinate descent* and *Alternating direction method of multipliers*. Through implementation and backtesting using a diverse set of indices representing various asset classes, the study shows that the mean variance model suffers from high turnover and sensitivity to input parameters in comparison to the modern alternatives. The sophisticated asset allocation models equal risk contribution and the most diversified portfolio do not rely on expected return as an input parameter, which is seen as an advantage, and are not affected to the same extent by the shortcomings associated with mean variance optimization. The paper concludes by discussing the findings critically and suggesting ideas for further research.

**Keywords:** mean variance optimization, portfolio theory, asset allocation strategies, equal risk contribution, most diversified portfolio, empirical study, back test
Sammanfattning

Acknowledgments

We would like to extend our thanks to our external partner Erik Penser Bank, more specifically Jonas Thulin, Jonatan Werner and Mats Carlsson, for insightful input and providing data as well as theoretical aid.

Additionally, we would like to thank our advisor professor Henrik Hult at the KTH Department of Mathematics for providing ideas and thoughts on how to approach as well as solve some of the mathematical problems encountered. We would also like to thank Anja Janssen at the KTH Department of Mathematics for providing useful and inspirational papers on portfolio optimization and for helping us come up with the idea for this thesis.

Lastly, the authors would like to thank one another. Writing this thesis has been challenging and insightful but also a lot of fun. We both agree that the quality would not have been the same had the thesis been written by a single author. Furthermore, our family and friends have been there for us, supporting and providing critical thinking to our judgements and writing, something the authors are truly grateful for. It would not have been the same without your support!
Contents

1 Introduction ........................................... 7
  1.1 Background ...................................... 7
  1.2 Mean variance optimization .......................... 8
    1.2.1 Closed form solution ........................... 8
    1.2.2 Quadratic programming .......................... 8
    1.2.3 Shortcomings ................................ 9
  1.3 Traditional alternative investment strategies ......... 9
    1.3.1 Equally weighted portfolio .................... 9
    1.3.2 Market cap weighted portfolio .................. 9
  1.4 Purpose ........................................ 10
  1.5 Previous research ................................ 10
  1.6 Limitations ..................................... 10

2 Objective functions and constraints .................. 11
  2.1 Global minimum variance .......................... 11
    2.1.1 Mathematical definition ........................ 11
    2.1.2 Example: Three asset universe ................. 12
    2.1.3 Additional constraints .......................... 12
    2.1.4 Asset concentration ............................ 12
    2.1.5 Diversification constraint ...................... 13
  2.2 Most diversified portfolio ....................... 13
    2.2.1 Mathematical definition ........................ 13
    2.2.2 Details of the diversification ratio ............ 13
    2.2.3 Example: Two asset universe .................... 14
    2.2.4 Example: Three asset universe .................. 14
    2.2.5 Other properties ................................ 15
    2.2.6 Asset concentration ............................ 15
  2.3 Equal risk contribution ........................... 15
    2.3.1 Mathematical definition ........................ 15
    2.3.2 Moving away from equal risk .................... 16
    2.3.3 Adding weight constraints ........................ 16
    2.3.4 Risk measure .................................. 16

3 Optimization algorithms ............................ 18
  3.1 Coordinate descent .............................. 18
    3.1.1 Cyclical coordinate descent .................... 19
  3.2 Alternating direction method of multipliers ....... 20
    3.2.1 Dual decomposition ............................ 20
    3.2.2 Method of multipliers .......................... 21
    3.2.3 Alternating direction method of multipliers .... 22
    3.2.4 Scaled form .................................... 22
    3.2.5 Notes on convergence .......................... 22
    3.2.6 Constrained convex optimization .................. 23
    3.2.7 Example: Mean variance quadratic program ....... 24
    3.2.8 Proximal operator ................................ 24
    3.2.9 Dykstra’s algorithm ............................ 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Methodology and Data</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Methodology</td>
<td>27</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Model evaluation</td>
<td>27</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Performance evaluation</td>
<td>28</td>
</tr>
<tr>
<td>4.2</td>
<td>Data</td>
<td>29</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Expected return</td>
<td>29</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Covariance matrix</td>
<td>29</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Analysis of data</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>Results</td>
<td>33</td>
</tr>
<tr>
<td>5.1</td>
<td>Global minimum variance</td>
<td>33</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Optimization program</td>
<td>33</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Solution: Quadratic programming</td>
<td>33</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Application to data</td>
<td>34</td>
</tr>
<tr>
<td>5.2</td>
<td>Most diversified</td>
<td>34</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Optimization program</td>
<td>34</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Solution: ADMM</td>
<td>35</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Application to data</td>
<td>36</td>
</tr>
<tr>
<td>5.3</td>
<td>Equal risk contribution</td>
<td>37</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Optimization program</td>
<td>37</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Solution: ADMM</td>
<td>37</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Moving away from equal risk</td>
<td>39</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Application to data</td>
<td>39</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Solution: CCD</td>
<td>39</td>
</tr>
<tr>
<td>5.4</td>
<td>Evaluation of traditional strategies</td>
<td>40</td>
</tr>
<tr>
<td>5.5</td>
<td>Evaluation of new strategies</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>Discussion</td>
<td>48</td>
</tr>
<tr>
<td>6.1</td>
<td>Portfolio turnover rate</td>
<td>48</td>
</tr>
<tr>
<td>6.2</td>
<td>Asset concentration</td>
<td>48</td>
</tr>
<tr>
<td>6.3</td>
<td>Adding noise to the data</td>
<td>49</td>
</tr>
<tr>
<td>6.4</td>
<td>Model parameters</td>
<td>49</td>
</tr>
<tr>
<td>6.5</td>
<td>Alternative risk measures</td>
<td>49</td>
</tr>
<tr>
<td>6.6</td>
<td>Further research</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>51</td>
</tr>
</tbody>
</table>
1 Introduction

The portfolio optimization problem is widely studied topic in financial mathematics and well known to both academics and finance industry professionals. Traditional methods using quadratic programming to maximize expected return while minimizing variance (Mean variance optimization, MVO) have become highly popular among asset managers. The industry traction gained by the mean variance model can be explained by several factors; it is simplistic, tractable, and computationally feasible. Potential replacement models proved to be too complex and difficult to implement, leading these models to be rejected by the finance industry. While the mean variance model has become well established in the finance industry in recent years, it has become evident that the model has shortcomings [1].

Ever since the introduction of MVO, academics and industry professionals have presented candidate models for its replacement [2]. However, these potential alternatives have not achieved comparable levels of success within the finance industry. In some cases, alternative optimization programs made use of parameters difficult to estimate. In other cases, the optimization program itself was not solvable with currently available analytical and numerical methods. Newly introduced portfolio optimization models often had good mathematical properties and financial relevance but lacked in practical aspects.

As a result of recent developments in the fields of machine learning and mathematical optimization, numerical methods capable of solving large scale complex portfolio optimization problems are available. Perrin and Roncalli demonstrate the capability of the optimization algorithms Alternating direction method of multipliers (ADMM) and Cyclical coordinate descent (CCD) in solving modern portfolio optimization programs [2]. The work of Perrin and Roncalli is the basis of this thesis which intends to study the the properties of alternatives to mean variance optimization.

1.1 Background

The distribution of capital among an investors chosen assets is the goal of portfolio optimization. In financial terms, the problem is known as asset allocation. How much should an asset manager buy or sell of a specific asset? How big of a weight, as a fraction, of the available capital, should each asset be given? An intelligent investor would prefer an asset allocation providing the greatest return on capital with minimal associated risk. However, what most investors are, or should be, aware of is that high returns often come with a higher risk. Investors may also be interested in a portfolio with minimal risk while still being fully invested.

An argument that speaks in favor of moving away from a strategy based on intuition towards a systematic way of allocating assets is the removal of human judgement. Using a robust and reproducible way of allocating capital has the advantage of not being clouded by feelings, personal judgement or skepticism. In other words, the use of a systematic allocation strategy eliminates the risks associated with psychological missteps and confirmation bias posed by humans in asset allocation. However, the greatest returns are most likely obtained when human judgement is combined with a robust portfolio optimization strategy.
1.2 Mean variance optimization

Asset allocation originated with the work of Harry Markowitz and his 1952 paper on mean variance optimization. An investor with a desire to construct an efficient portfolio will want to maximize expected returns for a given risk level \([3]\). In an efficient portfolio, increasing expected returns always comes at a cost of increasing volatility. A mean variance approach to portfolio optimization involves solving the quadratic programming (QP) problem

\[
x^*(\gamma) = \arg\min_x \frac{1}{2} x^\top \Sigma x - \gamma x^\top \mu \\
\text{s.t. } 1_n^\top x = 1
\]

with a vector of assets \(x \in \mathbb{R}^n\), the associated covariance matrix \(\Sigma \in \mathbb{R}^{n \times n}\), expected returns \(\mu \in \mathbb{R}^n\), and the risk aversion parameter \(\gamma\). Decreasing \(\gamma\) leads to increased risk aversion with the investor being as risk averse as possible as \(\gamma\) tends towards zero. The constraint \(1_n^\top x = 1\) requires all available funds to be invested in the assets \(x\), no cash left over. Other common constraints are

- No short selling \(x_i \geq 0\)
- Weight bounds \(x^-_i \leq x_i \leq x^+_i\)

1.2.1 Closed form solution

Problem (1) is a convex optimization problem with equality constraints. The solution is found by way of Lagrange multipliers. To begin with, introduce the Lagrangian \(\mathcal{L}(x, \lambda)\)

\[
\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)
\]

At the optimal value \(x^*(\gamma)\), the gradients of \(f(x)\) and \(g(x)\) are parallel

\[
\nabla_{x, \lambda} \mathcal{L}(x, \lambda) = 0 \iff \begin{cases} 
\nabla_x f(x) = \lambda \nabla_x g(x) \\

\lambda = 0
\end{cases}
\]

Applying (5) to the mean variance program (1) leads to the following solution for the optimal weights \(x^*(\gamma)\)

\[
x = \frac{1}{2} \Sigma^{-1} (\gamma \mu - \lambda 1_n) \\
\lambda = \frac{\gamma 1_n^\top \Sigma^{-1} \mu - 1}{1_n^\top \Sigma^{-1} 1_n}
\]

1.2.2 Quadratic programming

A quadratic programming (QP) problem consists of finding the minimum of a quadratic function under linear constraints. The general form of a QP problem with variables \(x \in \mathbb{R}^n\) and constraints \(A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m\) is

\[
x^* = \arg\min_x \frac{1}{2} x^\top Q x - c^\top x \\
\text{s.t. } Ax \leq b
\]

The QP problem (1) is a special case with equality constraints and a positive definite covariance matrix \(\Sigma\). Problems of this type have closed form solutions that may be found analytically. There are many solvers available capable of computing solutions to general QP problems. Hence, optimization problems that can be written as a QP problem are easy to solve.
1.2.3 Shortcomings

Why would an asset manager want to stray from the traditional mean variance framework and consider more complex objective functions to construct portfolios? Imagine deriving a set of portfolio weights by solving the mean variance model. When adjusting the portfolio at a later date, new model input parameters will need to be computed. If the covariance matrix, or more importantly [4], the expected return has not changed significantly at that time, the portfolio weights should not change considerably. However, in the mean variance framework, there may still be considerable variation in the portfolio. This example describes one of the main drawbacks of the mean variance model; its sensitivity to input parameters [5].

A consequence of the mean variance model’s sensitivity to input parameters is the possibility of high portfolio turnover, which in turn infers the problem of high transaction costs. A good way of achieving high returns as an asset manager is to keep the transaction costs as low as possible, avoiding paying unnecessary spreads, fees and commissions for trading. However, with the volatile solutions produced by the mean variance framework, the investor may be forced to readjust the portfolio significantly even in the event of small changes in the market.

All portfolio allocation strategies aim to determine the optimal weights of specific assets to hold for a certain period of time. However, the input parameters to these calculations are based solely on historical data. Essentially, asset managers extrapolate from historical data. The use of parameter estimates based on historical data is not exclusive to asset managers utilizing the mean variance strategy. Given that historical returns and covariances are not guaranteed to accurately predict future values, the fewer model input parameters to estimate, the better.

1.3 Traditional alternative investment strategies

The following subsection makes note of investment strategies available as alternatives to portfolio optimization techniques. In addition to the methods in their standard form, investors can employ these strategies with modifications by setting their own readjustment windows and limits on maximum deviation from asset weights.

1.3.1 Equally weighted portfolio

A simple yet effective portfolio strategy is the equally weighted portfolio. The capital to be used for investment is divided equally among the chosen assets. Each time the portfolio is to be re-balanced, assets that have gained value are sold and assets that have lost value are bought. The re-balancing procedure is intended to bring the portfolio back to equal weight once new asset prices have been set by the market, thus starting every period with equal weights in all assets.

1.3.2 Market cap weighted portfolio

Another strategy used to form portfolios and indices is the method of weighted market capitalization. Each asset is given a weight equal to the quotient of its market capitalization and the total market capitalization of the portfolio. At periods of redistribution of capital, new quotients are calculated and the capital is allocated accordingly. Examples of indices using
a market-cap weighted philosophy are the S&P, Nasdaq 500, DAX and TOPIX. Needless to say, the strategy is widely used to track weighted movement of a collection of assets.

1.4 Purpose

The aim of this thesis is to conduct a thorough empirical investigation as to what extent sophisticated allocation strategies address the shortcomings of mean variance optimization. The Risk budgeting portfolio, Most diversified portfolio, and a modification of the standard Minimum variance portfolio will represent the modern alternatives to traditional mean variance optimization. The foundation of the empirical investigation is a comparison of portfolio optimization model properties through backtesting. The specific drawback of mean variance optimization of interest to the study is the model’s sensitivity to input parameters. An essential measurement highly related to a portfolio optimization model’s sensitivity to input parameters is the turnover rate. Hence, a primary goal of the thesis is to investigate whether the modern portfolio optimization strategies have a lower average rate of turnover than the traditional mean variance strategy. In addition, a good portfolio optimization strategy should be robust in the sense that small changes in the market should not allow for considerable variation in the portfolio. Adding noise to the data set changing the day-to-day returns while leaving the large scale behavior of the market unaffected should not lead to considerable variation in the portfolio. The study also aims to determine if the modern portfolio optimization strategies are more robust than the mean variance model to small disturbances in the day-to-day returns by investigating the effect this has on portfolio turnover rate.

1.5 Previous research

The work of Perrin and Richard is the foundation of the study. Their paper Machine learning optimization algorithms and portfolio allocation provides theoretical insight on optimization algorithms and demonstrates solution methods to portfolio optimization problems of interest to this study [2]. Michaud shines light on the many issues associated with mean variance optimization of both mathematical and political nature [5]. Richard and Roncalli show many useful results related to risk budgeting portfolios [6]. Choueifaty and Coignard are the progenitors of the most diversified portfolio [7].

1.6 Limitations

The sample mean of the returns data is assumed to accurately estimate expected returns as needed for the mean variance model. Estimating expected returns in this fashion ignores the multivariate structure of the returns data. With this in mind, the study assumes the sample mean to be an acceptable estimator of the expected return. A volatility based risk measure is used both for the purposes of estimating asset risk and computing risk budgeting portfolios. The volatility based risk measure is assumed to accurately measure risk in the variety of assets comprising the data set.
2 Objective functions and constraints

In order to explain and set out what a specific strategy seeks to accomplish, allocation strategies are formulated using an objective function. The goal of an optimization program is to find the minimum or maximum of such a function. Together with the objective function, additional constraints are often used in order to restrict the solution to a specific domain of interest.

2.1 Global minimum variance

The global minimum variance (GMV), or minimum variance (MV) portfolio is a variation of the traditional mean variance portfolio. As the name suggests, the minimum variance portfolio seeks the asset weights $x$ minimizing portfolio variance, $x^\top \Sigma x$. Generally viewed as a simpler version of the mean variance portfolio, the minimum variance portfolio does not take expected returns into account. Hence, the model requires only one input parameter, the asset covariance matrix $\Sigma$.

Similarly to other portfolio optimization models, MV has both its strengths and weaknesses. One advantage of the model, perhaps the most important property to industry professionals, is that it is easy to understand the portfolio objective and that it is financially relevant. An asset with low variance will have a high portfolio weight while, conversely, an asset with high variance will have a low portfolio weight. Furthermore, the model is simple in the sense that it only uses one input parameter, the covariance matrix of the asset returns data. Turning our attention to some of the model’s drawbacks, in most cases the portfolio is concentrated on a small number of assets. In addition, it is a long/short portfolio which to many financial institutions is unacceptable [2]. A long/short portfolio is a portfolio where the investor is allowed to both buy and short sell assets.

2.1.1 Mathematical definition

The standard MV portfolio is the solution to the following QP problem with equality constraints

$$x^* = \arg \min_x \frac{1}{2} x^\top \Sigma x$$

$$\text{s.t. } 1_n^\top x = 1$$

(9)

The standard form of minimum variance is not widely used in practice for reasons mentioned above. Avoiding an estimate of the expected return turns out to be a considerable advantage of the MV portfolio in comparison to the mean variance portfolio (1). Estimators of the expected return are inaccurate, even in ideal cases when the data is independent and identically distributed [8].

The solution to the GMV program (9) is easy to derive using the method covered in section (1.2.1). The optimal weights are

$$x^* = \frac{\Sigma^{-1} 1_n}{1_n^\top \Sigma^{-1} 1_n}$$

(10)
2.1.2 Example: Three asset universe

In order to illustrate the fact that the GMV portfolio in its form (9) is a long/short portfolio, consider the following example. There are three assets in the universe with volatilities $\sigma_A = 0.2$, $\sigma_B = 0.3$, $\sigma_C = 0.1$, and correlation matrix $S$.

$$S = \begin{bmatrix}
1 & 0.6 & 0.5 \\
0.6 & 1 & 0.4 \\
0.5 & 0.4 & 1
\end{bmatrix}$$  \(11\)

The solution is found by plugging the numbers into (10). The optimal asset weights are

$$x_A = 0.028 \quad x_B = -0.035 \quad x_C = 1.007$$

Two significant drawbacks of the model are demonstrated by this example. The model heavily favours the assets with the lowest volatility, asset $C$, allocating nearly all the available capital to this asset. As expected, a short position was generated by the model, in this case asset $B$. Many financial institutions are not interested in short positions and would not accept an asset allocation such as the one in this example.

2.1.3 Additional constraints

A first attempt to remedy some of the undesirable properties of the minimum variance portfolio might be to add simple weight constraints. To that end, consider the following modification to the QP problem (9)

$$x^* = \arg \min_x \frac{1}{2} x^\top \Sigma x$$

s.t. \( \begin{cases} 
1_n^\top x = 1 \\
0_n \leq x \leq x^+ 
\end{cases} \quad 12\)

The investment problem no longer has a closed form solution and must be solved using numerical methods. One quickly comes to the conclusion that adding weight constraints is not the answer to the drawbacks of the minimum variance portfolio. Using the same setup as in the previous example, setting the maximum weight bound to 1 and 0.9 generates the following solutions, respectively

$$x_A = 0.00 \quad x_B = 0.00 \quad x_C = 1.00$$

$$x_A = 0.10 \quad x_B = 0.00 \quad x_C = 0.90$$

By adding weight constraints the model is prevented from producing long/short portfolios. However, the weights still tend to be concentrated on a few number of assets.

2.1.4 Asset concentration

An asset allocation concentrated on a few number of assets is not desirable from a risk management/diversification perspective. Financial regulations may also prohibit portfolios with too high of a weight in single assets. In order to address this, many professionals impose a relative weight bound based on a benchmark $b$

$$\delta^- b_i \leq x_i \leq \delta^+ b_i$$  \(13\)

However, this does not properly address the problem of diversification in the portfolio since the resulting weights depend on the benchmarks weights [2].
2.1.5 Diversification constraint

Perrin and Roncalli propose adding a diversification constraint to address the problem of diversification in the minimum variance portfolio [2]. The diversification constraint makes use of the Herfindahl index \( H(x) = \sum_{i=1}^{n} x_i^2 \), which assumes the value 1 in a single asset portfolio and the value \( \frac{1}{n} \) in an equally weighted portfolio. To introduce diversification in the model, a quantity known as number of effective bets, which is defined to be the inverse of the Herfindahl index, \( N(x) = H(x)^{-1} \), is constrained to be larger than some minimal value. The modified optimization program becomes

\[
x^* = \arg \min_x \frac{1}{2} x^\top \Sigma x \\
\text{s.t.} \left\{ \begin{array}{l}
1_n^\top x = 1 \\
0_n \leq x \leq x^+ \\
N(x) \geq N^-
\end{array} \right.
\]

This minimum variance program is of interest since it combines the property of minimizing the variance while at the same time incorporating a constraint on asset diversification.

2.2 Most diversified portfolio

In 2008, Choueifaty and Coignard proposed the most diversified portfolio (MDP), a portfolio based on a measure of diversification known as a diversification ratio [7]. In their paper, Towards maximum diversification, they present the portfolio as an alternative to equally weighted, market cap weighted and minimum variance portfolios.

2.2.1 Mathematical definition

Define the most diversified portfolio as the weights \( \sum_{i=1}^{n} x_i = 1 \) which maximize the diversification ratio \( D(x) \)

\[
D(x) = \frac{x^\top \sigma}{\sqrt{x^\top \Sigma x}}
\]

\( \sigma \) is a vector of asset volatilities and \( \Sigma \) is the asset covariance matrix. Choueifaty et al. show that the MDP is also the solution to the following minimization problem [9]

\[
x^* = \arg \min_x \frac{1}{2} \ln (x^\top \Sigma x) - \ln (x^\top \sigma) \\
\text{s.t.} \left\{ \begin{array}{l}
1_n^\top x = 1 \\
x \in \Omega
\end{array} \right.
\]

Under a set of constraints \( \Omega \). In most cases, the positive weight constraint is applied, meaning only long positions are allowed.

2.2.2 Details of the diversification ratio

Choueifaty introduced the diversification ratio (15) as a measure of portfolio diversification. A single-asset portfolio corresponds to a diversification ratio of 1. In general, the diversification ratio is greater than 1. To give an intuitive sense of how the diversification ratio works, consider the following examples given in Choueifaty and Coignard [7].
2.2.3 Example: Two asset universe

Consider a two-assets universe consisting of assets A and B. The volatilities $\sigma_A$ and $\sigma_B$ are 15 and 30 percent respectively. Assets A and B have a correlation strictly less than 1. In this case, diversification means both assets contribute equally to portfolio volatility. Therefore, the most diversified portfolio is the solution to the following equations

\begin{align*}
0.15x_A &= 0.3x_B \\
x_A + x_B &= 1
\end{align*}

The solution is $x_A = \frac{2}{3}$, $x_B = \frac{1}{3}$.

2.2.4 Example: Three asset universe

Consider a three asset universe consisting of assets A, B and C. Their respective weights in most diversified portfolio are $x_A$, $x_B$ and $x_C$. Assets A and B are both bank stocks with high correlation equal to 0.9. Asset C is a pharmaceutical stock with low correlation to assets A and B equal to 0.1. All assets are assumed to have the same volatility $\sigma$. Compute the weights in the most diversified portfolio.

From the data on the correlation and volatility, the covariance matrix is found to be

\[\Sigma = \sigma^2 \begin{bmatrix} 1 & 0.9 & 0.1 \\ 0.9 & 1 & 0.9 \\ 0.1 & 0.9 & 1 \end{bmatrix} = \sigma^2 V\]

By factoring out the volatility $\sigma$, the diversification ratio $D(x)$ is simplified in the following way

\[D(x) = \frac{1}{\sqrt{x^\top Vx}}\]

The numerator simplifies to $x^\top 1 = 1$, which follows from the constraints of the MDP. Maximizing the diversification ratio (20) is the same as minimizing

\[D(x) = x^\top Vx\]

The last result shows how maximizing the diversification ratio in a universe where all assets have the same volatility results in minimizing the variance.

Since all assets contribute equally to portfolio volatility, the constraint $x_A = x_B$ must be applied. The weights in the most diversified portfolio are found as the solution to the following QP problem

\[x^* = \arg\min_{x} x^\top Vx \quad \text{s.t.} \quad Ax = b\]

\[A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\]

Again, using the method covered in section (1.2.1) yields the following solution
The solution is $x_A = 0.263$, $x_B = 0.263$, $x_C = 0.474$.

2.2.5 Other properties

It is clear from the second example that if all assets have the same volatility, the most diversified portfolio is equal to the minimum variance portfolio. In addition, it is important to consider whether the covariance matrix is invertible since this property will affect the uniqueness of the solution [7].

Accurate estimation of the covariance matrix is vital when solving for the MDP with actual data since it is the only piece of information the model uses. The data set should ideally consist of a full market cycle of daily returns [7]. Naturally, this fact applies to the minimum variance portfolio and equal risk contribution as well.

2.2.6 Asset concentration

Similarly to the minimum variance portfolio, the MDP with linear constraints may yield weights concentrated on a small number of assets. Perrin and Roncalli suggest adding a weight diversification constraint $D(x) > D^-$ [2]. To achieve this, they consider the number of effective bets $N(x)$ and constrain it to be larger than some minimal value $N^-$.

$$
x^* = \arg \min_x \frac{1}{2} \ln (x^\top \Sigma x) - \ln (x^\top \sigma)
\begin{align*}
\text{s.t.} \quad & 1^\top x = 1 \\
& x \in \Omega \\
& N(x) \geq N^-
\end{align*}
$$

2.3 Equal risk contribution

Asset allocation in the traditional sense has focused mainly on finding the optimal way of allocating investor capital. Historically, when investigating alternatives to the mean variance framework, the two main strategies considered were minimum variance and the equally weighted portfolio. In 2008, Maillard, Roncalli and Teiletche proposed a strategy that can be viewed as a middle ground between minimum variance and the equally weighted portfolio [4]. In their paper, they suggest allocating capital such that the risk contribution of each asset in the portfolio contributes equally to the total risk of the portfolio. They called the strategy equal risk contribution (ERC). Strategies dealing with the allocation of risk rather than capital have gained traction among investors and go under the label of risk budgeting. In such allocation strategies, asset managers solve for how to allocate risk and then translate this to capital.

2.3.1 Mathematical definition

Consider a portfolio $x$ consisting of $n$ assets, $x = (x_1, x_2, \ldots, x_n)$, with $\sigma_i^2$ being the variance of asset $i$ and $\sigma_{ij}$ be the covariance of asset $i$ and $j$. Let $\Sigma$ be the covariance matrix. Maillard et al. define the total risk of the portfolio as $\sigma(x) = \sqrt{x^\top \Sigma x}$ [4]. The choice of risk measure is of course subjective and can be changed. Furthermore, they define the the marginal
risk contribution of asset $i$ as $\partial x_i \sigma(x) = \frac{\partial \sigma(x)}{\partial x_i}$. With this in mind, the problem of assets contributing equally to the total risk can be formalized in the following way

$$RC_i(x) = x_i \frac{\partial \sigma(x)}{\partial x_i} = x_j \frac{\partial \sigma(x)}{\partial x_j} = RC_j(x) \quad (26)$$

When moving away from the special cases with equal volatility and correlation between all assets, the problem no longer has a closed form solution and a numerical algorithm is needed to find the solution. Roncalli and Richard formulate the general risk budgeting approach to portfolio optimization as [6]

$$x_{RB} = \arg \min_{x} R(x)$$

s.t.\ 
$$\sum_{i=1}^{n} b_i \ln x_i \geq \kappa^*$$

$$1_n^T x = 1$$

$$x \geq 0_n$$ \quad (27)

Using the Lagrange formulation of (27), Roncalli and Richard (2019) show the former optimization program to be equivalent to (28) when a volatility based risk measure is considered.

$$x^* = \arg \min_{x} \frac{1}{2} x^T \Sigma x - \lambda \sum_{i=1}^{n} b_i \ln x_i$$ \quad (28)

$\lambda$ is any positive scalar and $b_i$ is the risk weight of asset $i$. Optimization program (28) has preferable properties when compared to program (27) with regards to solving the optimization problem, as the former is a logarithmic barrier problem with appealing characteristics [6]. Perrin and Roncalli state that the ERC portfolio is the scaled solution $x^*/(1_n^T x^*)$ of optimization program (28) [2].

### 2.3.2 Moving away from equal risk

When the risk budgets are all equal $b_i = \frac{1}{n}$, a special case of the risk budgeting portfolio known as equal risk contribution is obtained. Varying the risk budgets to allocate a certain risk to each asset produces a general risk budgeting portfolio.

### 2.3.3 Adding weight constraints

To comply with certain financial regulations limiting the maximum weight placed on a single asset it may be necessary to apply an additional constraint on the ERC portfolio (27)

$$x_{RB} = \arg \min_{x} R(x)$$

s.t.\ 
$$\sum_{i=1}^{n} b_i \ln x_i \geq \kappa^*$$

$$1_n^T x = 1$$

$$0_n \leq x \leq x^+$$ \quad (29)

### 2.3.4 Risk measure

There are some alternatives to choose from when considering which risk measure $R(x)$ to use for the purposes of computing the ERC portfolio. Richard and Roncalli mention the
following risk measures, one is a standard deviation based risk measure and the other is based on the portfolio variance [6].

\[ R(x) = -x^\top (\mu - r) + c \cdot \sqrt{x^\top \Sigma x} \]

\[ R(x) = \sqrt{x^\top \Sigma x} \] (30)

The standard deviation based risk measure measures the trade off between expected return and volatility. Since avoiding an estimate of the expected return is important for reasons mentioned previously, this risk measure will not be considered in this text.
3 Optimization algorithms

When solving complex problems in finance, it is not uncommon that one is trying to maximize or minimize a certain function. A simple example might be a producer attempting to maximize income as a function of units to produce, involving both the production and selling costs as well as the revenue generated from selling.

More complex problems might involve quadratic functions with constraints. For example, consider mean variance optimization, where an asset manager seeks to maximize expected return while keeping volatility minimal. In doing this, the asset manager is looking for the optimal solution. In some cases, these problems have closed-form solutions that can be found analytically.

However, as the complexity of the problem increases, a closed-form solution to the optimization may not exist. In this case, one can implement an optimization algorithm to find the solution numerically. Solving an optimization problem successfully depends on choosing a suitable algorithm. In most cases, the characteristics of the problem will determine what algorithm is appropriate. Is the objective function linear or non-linear? Is the program constrained? Is the objective function convex? These are some of the questions that need to be answered before choosing an optimization algorithm. The optimization algorithm will continue iterating until a convergence criteria is met, which can be a previously specified number of iterations or absolute error between two iterations. Termination rule and stopping criteria are two terms used interchangeably with convergence criteria.

3.1 Coordinate descent

Coordinate descent is an optimization algorithm that builds upon the concept of gradient descent. In order to understand how coordinate descent works, it is necessary to understand gradient descent. Consider the following generic unconstrained optimization problem consisting of finding the variable $x \in \mathbb{R}^n$ minimizing $f(x)$

$$x^* = \arg \min_x f(x)$$ (31)

Where $f(x) : \mathbb{R}^n \to \mathbb{R}$ is continuous and convex function. Gradient descent starts off with an initial guess $x^{(0)}$ and then works in a similar way to the descent method (32)

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)} = x^{(k)} - \eta D^{(k)}$$ (32)

$x^{(k)}$ is the approximate solution at the $k^{th}$ iteration, $\eta$ is the step size and $D^{(k)}$ is the direction in which you step [2]. In the case of gradient descent, the direction in which to step in is the gradient vector of $f(x)$ at the current point $x^{(k)}$.

The coordinate descent algorithm is a slight modification of gradient descent method. It is somewhat more granular in the sense that it works by minimizing the function along one coordinate at each iteration, instead of the whole plane.

$$x^{(k+1)}_i = x^{(k)}_i + \Delta x^{(k)}_i = x^{(k)}_i - \eta D^{(k)}_i$$ (33)

Algorithm 1 is a high level illustration of the coordinate descent procedure.
Algorithm 1: Coordinate descent, gradient formulation

Result: Solution to $x^* = \arg\min_x f(x)$

set the initial guess $x^0$ and the step size $\eta$

$k = 0$

$\text{convg} = \text{False}$

while not convg do

Pick a coordinate $i \in \{1, n\}$

$x_i^{(k+1)} \leftarrow x_i^{(k)} - \eta \nabla_i f(x^{(k)})$

$x_j^{(k+1)} \leftarrow x_j^{(k)}$ if $j \neq i$

$k \leftarrow k + 1$

$\text{convg} = \max_i \left| x_i^{(k+1)} - x_i^{(k)} \right| \leq \varepsilon$

end

Moving towards the coordinate descent algorithm rather than gradient descent has the advantage of transforming the problem to a scalar-valued problem from a vector-valued problem. This is preferable both from a computational and intuitive perspective.

The step length $\eta$ need not be fixed, but can be varying and set with a line search. Up until 2001, the coordinate descent algorithm could not deal with functions that were non-differentiable and/or non-convex. However, with the contributions of Tseng, this can now be achieved [10].

3.1.1 Cyclical coordinate descent

A problem that arises is how best to pick the coordinate $i$ for which to update the solution. There are ways to find the optimal coordinate $i$. However, this would require the gradient along each coordinate to be calculated. This causes the algorithm to no longer be efficient, and the computational complexity exceeds that of gradient descent [2].

In 2001, Tseng suggested a somewhat trivial, yet efficient approach to the problem of choosing the update coordinate. Namely, iterating over the coordinates cyclically. The method has come to be known as cyclical coordinate descent (CCD). Given the nature of the method, it ensures that all coordinates have been updated during one iteration cycle $[k - n + 1, \ldots, k]$. This approach is popular and widely used which can be explained by its simplicity and tractability. It is also computationally negligible to pick the coordinate. Algorithm 2 demonstrates the cyclical coordinate descent algorithm at a high level.
Algorithm 2: Cyclical coordinate descent

Result: Solution to $x^* = \arg \min_x f(x)$

Set the initial guess $x^0$

$k = 0$

$\text{convg} = \text{False}$

While not $\text{convg}$ do

For $i = 1 : n$ do

$x_i^{(k+1)} = \arg \min_x f\left(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right)$

End

$k ← k + 1$

$\text{convg} = \max_i \left| x_i^{(k+1)} - x_i^{(k)} \right| ≤ \varepsilon$

End

3.2 Alternating direction method of multipliers

Alternating direction method of multipliers (ADMM) belongs to the group of optimization algorithms known as augmented Lagrangian methods. The following section presents context and theory necessary to understand ADMM as outlined in Boyd et al. [11]. It is not intended to be a complete mathematical description. To begin with, two methods of optimization which are closely related to ADMM are presented, dual decomposition and method of multipliers.

3.2.1 Dual decomposition

Consider the following convex equality constrained optimization problem

$$x^* = \arg \min_x f(x)$$

s.t. $Ax = b$ \hspace{1cm} (34)

With variable $x \in \mathbb{R}^n$, constraints $A \in \mathbb{R}^{m \times n}$ and convex $f : \mathbb{R}^n \to \mathbb{R}$. Introduce the variable $y \in \mathbb{R}^m$ and the dual problem

Lagrangian $L(x, y) = f(x) + y^T (Ax - b)$

Dual function $g(y) = \inf_x L(x, y)$ \hspace{1cm} (35)

Dual problem $y^* = \arg \max_x g(x)$

The dual function $g(y)$ is the greatest lower bound of the Lagrangian (with respect to $x$). This quantity may also be expressed using the conjugate of $f$, $f^*$

$$g(y) = \inf_x L(x, y) = -f^* (-A^T y) - b^T y$$

$$f^*(y) = \arg \max_x y^T x - f(x)$$ \hspace{1cm} (36)

Assuming that strong duality holds and that there is a unique minimizer of $L(x, y^*)$, a primal optimal point $x^*$ may be obtained from a dual optimal point $y^*$ in the following way

$$x^* = \arg \min_x L(x, y^*)$$ \hspace{1cm} (37)
Apply gradient ascent to the dual problem

\[ y^{k+1} = y^k + \lambda^k \nabla g(y^k) \]  

for some small \( \lambda \), \( \nabla g(y^k) = Ax - b \) and \( \hat{x} = \arg \max_x L(x, y^k) \). The following reasoning leads to the expression for the gradient of the dual function

\[
\nabla g(y) = A \nabla f^*(-A^T y) - b \\
\n\nabla f^*(-A^T y) = \arg\min_z f(z) + y^T Az = \hat{x} 
\]

The dual ascent method to solve for problem (34) is

\[
x^{k+1} := \arg\min_x L(x, y^k) \\
y^{k+1} := y^k + \lambda^k (Ax^{k+1} - b) 
\]

The first step minimizes \( x \) and the second step updates the dual variable. The step sizes \( \lambda^k \), if chosen carefully, ensure that the dual function \( g(y) \) increases with every step, hence the name dual ascent.

If \( f(x) \) is assumed to be separable, meaning \( f(x) = f_1(x_1) + f_2(x_2) + \ldots f_n(x_n) \), \( x = (x_1, x_2, \ldots, x_n) \), then the Lagrangian is separable in \( x \)

\[
L(x, y) = L_1(x_1, y) + L_2(x_2, y) + \ldots + L_n(x_n, y) 
\]

Minimization in \( x \) in the dual ascent method now becomes minimization in \( n \) separate variables \( x = (x_1, x_2, \ldots, x_n) \).

With a separable objective function, the dual ascent method is referred to as dual decomposition. The \( x \)-update and dual update become the following

\[
x_i^{k+1} := \arg\min_{x_i} L_i(x_i, y^k), \quad i = 1, \ldots, N \\
y^{k+1} := y^k + \lambda^k \left( \sum_{i=1}^N A_i x_i^{k+1} - b \right) 
\]

Convergence of the dual decomposition method depends on the finiteness and convexity of the objective function \( f(x) \).

### 3.2.2 Method of multipliers

In an effort to build upon dual decomposition, consider the augmented Lagrangian for (34)

\[
L_\rho(x, y) = f(x) + y^T (Ax - b) + (\rho/2) \| Ax - b \|^2_2 
\]

\( \rho > 0 \) is known as the penalty parameter. By modifying the Lagrangian it is possible to avoid assumptions necessary on the objective function \( f(x) \) to yield convergence in the dual decomposition method, namely convexity and finiteness. With the augmented Lagrangian, problem (34) may be viewed as

\[
x^* = \arg\min_x f(x) + (\rho/2) \| Ax - b \|^2_2 \\
\text{s.t. } Ax = b 
\]
Applying the dual ascent method (40) to problem (44) yields the algorithm known as *method of multipliers*.

\[
x^{k+1} := \arg\min_x L_\rho (x, y^k) \\
y^{k+1} := y^k + \rho (Ax^{k+1} - b)
\]  \hspace{1cm} (45)

Note that the penalty parameter becomes the step size. Method of multipliers has been shown to converge under more general conditions than dual decomposition.

### 3.2.3 Alternating direction method of multipliers

ADMM is an attempt to blend the properties of decomposability of dual ascent with convergence of method of multipliers. The algorithm may be used to solve for optimization problems in the following form

\[
\{ x^*, y^* \} = \arg\min_{(x,y)} f_x(x) + f_y(y) \\
\text{s.t. } Ax + By = c
\]  \hspace{1cm} (46)

with \( x \in \mathbb{R}^n, z \in \mathbb{R}^m, A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{p \times m}, c \in \mathbb{R}^p \) and convex functions \( f_x(x) \) and \( f_y(y) \).

The augmented Lagrangian in this case is

\[
L_\rho(x, y, z) = f_x(x) + f_y(y) + z^T (Ax + By - c) + \frac{\rho}{2} \|Ax + By - c\|_2^2
\]  \hspace{1cm} (47)

The iterations are

\[
x^{k+1} := \arg\min_x L_\rho (x, y^k, z^k) \\
y^{k+1} := \arg\min_y L_\rho (x^{k+1}, y, z^k) \\
z^{k+1} := z^k + \rho (Ax^{k+1} + By^{k+1} - c)
\]  \hspace{1cm} (48)

### 3.2.4 Scaled form

ADMM has a scaled form which is more concise. Without going into much detail, set the scaled dual variable to \( u = (1/\rho)z \). The scaled form has iterations

\[
x^{k+1} := \arg\min_x \left( f(x) + \frac{\rho}{2} \|Ax + By^k - c + u^k\|_2^2 \right) \\
y^{k+1} := \arg\min_y \left( g(y) + \frac{\rho}{2} \|Ax^{k+1} + By - c + u^k\|_2^2 \right) \\
u^{k+1} := u^k + Ax^{k+1} + By^{k+1} - c
\]  \hspace{1cm} (49)

### 3.2.5 Notes on convergence

According to Boyd et al. ADMM is slow to converge to high accuracy but fast to converge to moderate accuracy [11]. Most practical applications of ADMM are ones where high accuracy is not a necessity. For the purposes of solving portfolio optimization problems, the accuracy of ADMM is sufficient since investors are often limited to buying whole numbers of assets. The primal residual \( r \) and dual residual \( s \) at iteration \( k + 1 \) are defined as
\[ r^{k+1} = Ax^{k+1} + By^{k+1} - c \]
\[ s^{k+1} = \rho A^T B (y^{k+1} - y^k) \]

Boyd et al. show that reasonable stopping criteria for ADMM are [11]
\[ \|r^k\|_2 \leq \epsilon^{\text{pri}} \quad \text{and} \quad \|s^k\|_2 \leq \epsilon^{\text{dual}} \]

The following termination rule will be applied to the optimization problems treated in this text.
\[ \max \left( \|x^{(k+1)} - y^{(k+1)}\|^2, \|y^{(k+1)} - y^{(k)}\|^2 \right) \leq \varepsilon \]

To speed up convergence of the algorithm and avoid the task of choosing an appropriate value for the penalty parameter (and step-size) \(\rho\), a variable penalty parameter \(\rho^k\) will be used. Boyd et al. make note of the following update scheme [11]
\[ \rho^{k+1} := \begin{cases} \tau^{\text{incr}} \rho^k & \text{if } \|r^k\|_2 > \mu \|s^k\|_2 \\ \rho^k / \tau^{\text{decr}} & \text{if } \|s^k\|_2 > \mu \|r^k\|_2 \\ \rho^k & \text{otherwise} \end{cases} \]

The typical values \(\tau^{\text{incr}} = \tau^{\text{decr}} = 2\) and \(\mu = 10\) will be applied.

### 3.2.6 Constrained convex optimization

It is possible to apply the ADMM method to a generic constrained optimization problem. The following example is given in Boyd et al. [11]
\[ x^* = \arg\min_x f(x) \quad \text{s.t. } x \in \Omega \]

with \(x \in \mathbb{R}^n\), convex \(f(x)\) and \(\Omega\). Problem (54) can be written in ADMM form (46) in the following way
\[ \{x^*, y^*\} = \arg\min_{(x,y)} f(x) + g(y) \quad \text{s.t. } x - y = 0 \]

\[ g(y) = 1_{\Omega}(y) = \begin{cases} 0, & \text{if } y \in \Omega \\ \infty, & \text{if } y \notin \Omega \end{cases} \]

\(g(y)\) is the convex indicator function of \(\Omega\). With the scaled dual variable, the augmented Lagrangian is
\[ L_\rho(x, y, u) = f(x) + g(y) + (\rho/2)\|x - y + u\|_2^2 \]

The iterations are
\[ x^{k+1} := \arg\min_x \left( f(x) + (\rho/2)\|x - y^k + u^k\|_2^2 \right) \]
\[ y^{k+1} := \Pi_{\Omega} (x^{k+1} + u^k) \]
\[ u^{k+1} := u^k + x^{k+1} - y^{k+1} \]

The y-update is the Euclidean projection onto \(\Omega\). This way of rewriting an optimization problem to ADMM form is very useful and will be applied in later sections.
3.2.7 Example: Mean variance quadratic program

The following example demonstrates how to apply ADMM to a mean variance optimization program which consists of a quadratic objective function together with equality and inequality constraints.

\[
x^*(\gamma) = \arg \min_x \frac{1}{2} x^\top \Sigma x - \gamma x^\top \mu
\]
\[
\text{s.t.} \quad \begin{cases} Ax = b \\ x \geq 0_n \end{cases}
\]

Identify the functions \( f(x) \) and \( g(y) \) in the ADMM form (55).

\[
f(x) = \frac{1}{2} x^\top \Sigma x - \gamma x^\top \mu, \quad \text{dom} f = \{x | Ax = b\}
\]

\[
g(y) = \begin{cases} 0, & \text{if } y \geq 0 \\ \infty, & \text{if } y < 0 \end{cases}
\]

The iterations to solve for this particular program are

\[
x^{k+1} := \arg\min_x \left( f(x) + \frac{\rho}{2} \| x - y^k + u^k \|^2_2 \right)
\]
\[
y^{k+1} := \left( x^{k+1} + u^k \right)_+
\]
\[
u^{k+1} := u^k + x^{k+1} - y^{k+1}
\]

It is possible to simplify the x-update by inserting the objective function and using the fact that the square of the Euclidean norm is equal to the transpose of the vector multiplied by the vector itself.

\[
x^{k+1} := \arg\min_x \left( \frac{1}{2} x^\top (\Sigma + \rho I) x - \gamma x^\top \mu + \frac{\rho}{2} (x - y^k + u^k)^\top (x - y^k + u^k) \right)
\]

To continue, factor the expression in terms of \( x \). The remaining constants are left out of the expression since they do not affect the optimal value \( x^{k+1} \).

\[
x^{k+1} := \arg\min_x \left( \frac{1}{2} x^\top (\Sigma + \rho I) x - \gamma x^\top \mu + \rho (y^k - u^k)^\top (y^k - u^k) \right)
\]

When the domain of the objective function \( f(x) \) is considered, the last minimization problem reduces to a QP problem with equality constraints. Using Lagrange multipliers, the following system of equations is found leading to the solution \( x^{k+1} \).

\[
\begin{bmatrix} \Sigma + \rho I & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} \\ \lambda \end{bmatrix} = \begin{bmatrix} \rho (y^k - u^k) - \gamma \mu \\ b \end{bmatrix}
\]

The y-update is the Euclidean projection onto the the domain \( x \geq 0 \).

3.2.8 Proximal operator

Certain variable updates in applications of ADMM will require evaluating the proximal operator, \( \text{prox}_f(v) \). Richard and Roncalli cover the proximal operator to the extent that it applies to evaluating risk budgeting portfolios [6]. The following section briefly reviews the theory of the proximal operator and provides some identities given by Richard and Roncalli.
that will be useful in later sections.

Consider a proper closed convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. The proximal operator is a function $\text{prox}_f(v) : \mathbb{R}^n \to \mathbb{R}^n$ which is defined as

$$\text{prox}_f(v) = x^* = \arg \min_x \left\{ f(x) + \frac{1}{2} \|x - v\|^2 \right\}$$  \hspace{1cm} (66)

The objective function $f_v(x) = f(x) + \frac{1}{2} \|x - v\|^2$ is strongly convex and therefore the proximal operator has a unique minimum for every $v$ [12].

Evaluating the proximal operator of the logarithmic barrier function will be necessary when solving for risk budgeting portfolios. To that end, consider the following minimization

$$\text{prox}_{-\ln(x)}(v) = x^* = \arg \min_x \left\{ -\ln(x) + \frac{1}{2} \|x - v\|^2 \right\}$$  \hspace{1cm} (67)

Optimization program (67) is unconstrained. Hence, simple techniques from calculus are all that is required to find the optimal value $x^*$. Since $x$ is scalar in this case, (67) simplifies to

$$\text{prox}_{-\ln(x)}(v) = x^* = \arg \min_x \left\{ -\ln(x) + \frac{1}{2} (x - v)^2 \right\}$$  \hspace{1cm} (68)

The first-order condition is

$$- \frac{1}{x} + x - v = 0$$  \hspace{1cm} (69)

Solving for $x$ and noting that $\ln x$ is defined for $x > 0$ yields

$$x^* = \frac{v + \sqrt{v^2 + 4}}{2}$$  \hspace{1cm} (70)

The convex indicator function is a useful tool to incorporate constraints into the objective function of an optimization program, which for a convex set $\Omega$ is defined as

$$f(x) = 1_{\Omega}(x) = \begin{cases} 0, & \text{if } x \in \Omega \\ \infty, & \text{if } x \notin \Omega \end{cases}$$  \hspace{1cm} (71)

In this case, the proximal operator is the Euclidean projection onto $\Omega$, $P_\Omega(v)$.

$$\text{prox}_f(v) = \arg \min_x \left\{ 1_\Omega(x) + \frac{1}{2} \|x - v\|^2 \right\} = P_\Omega(v)$$  \hspace{1cm} (72)

### 3.2.9 Dykstra’s algorithm

Dykstra’s algorithm is a way of numerically computing the proximal operator of convex sums of functions to which the proximal operator is known. Perrin and Roncalli formalize the problem in the following way [2]

$$x^* = \arg \min_x \left\{ \sum_{j=1}^{m} f_j(x) + \frac{1}{2} \|x - v\|^2 \right\}$$  \hspace{1cm} (73)
Assume that the proximal operator of the functions $f_j(x)$ in the summand are known. The extent to which Dykstra’s algorithm will be necessary in this text is limited to the case when the proximal operator of the convex function $f(x) = f_1(x) + f_2(x)$ needs to be computed. This particular optimization problem may be written as

$$ x^* = \arg \min_x f_1(x) + f_2(x) + \frac{1}{2} \|x - v\|^2 $$

$$ = \text{prox}_f(v) \quad (74) $$

A solution to the minimization problem (74) is provided by Dykstra’s algorithm [13]. The iterations are

$$ \begin{align*}
  x^{(k+1)} &= \text{prox}_{f_1}(y^{(k)} + p^{(k)}) \\
  p^{(k+1)} &= y^{(k)} + p^{(k)} - x^{(k+1)} \\
  y^{(k+1)} &= \text{prox}_{f_2}(x^{(k+1)} + q^{(k)}) \\
  q^{(k+1)} &= x^{(k+1)} + q^{(k)} - y^{(k+1)}
\end{align*} $$

(75)

with initial values $x^{(0)} = y^{(0)} = v$ and $p^{(0)} = q^{(0)} = 0_n$. 

26
4 Methodology and Data

For the purposes of evaluating new portfolio optimization and allocation strategies, the traditional strategies covered at the end of section (1) will be used as comparison when evaluating the three more modern investment problems covered in section (2). The portfolio optimization models referred to as the traditional strategies are the mean variance model, minimum variance model and the equally weighted portfolio. The term new strategies applies to the ones presented throughout section (2); equal risk contribution, most diversified portfolio and minimum variance. The most diversified portfolio and minimum variance portfolios will be solved with an added constraint on diversification. New portfolio optimization models will be checked for sensitivity to input parameters, to see whether the shortcomings of the mean variance approach have been addressed. Also, a historical comparison of all the investment problems will be done to evaluate their performance. In order to allow for extensive evaluation of the implemented allocation strategies, several metrics providing descriptive statistics will be calculated. All portfolios have been re-adjusted at the same dates to allow for a fair comparison.

The mean variance model will be solved both unconstrained and with a cap on the maximum turnover. Minimum variance will be solved unconstrained and with a constraint on the maximum weight in single assets. As mentioned above, these portfolio optimization models will serve as the traditional strategies. It is worth mentioning to avoid confusion that the minimum variance portfolio is included in both the new category as well as the traditional category. The two minimum variance models are distinguished by the new minimum variance model having an additional constraint on asset diversification.

4.1 Methodology

To clarify which properties of portfolio optimization models are of interest to this study, the following section outlines which evaluation methods will be applied.

4.1.1 Model evaluation

The primary method of strategy evaluation will be a comparison of portfolio turnover rate. A good portfolio optimization model should be robust in the sense that small changes in the market should not make for large changes in the composition of the portfolio. Furthermore, if an asset looks promising during a holding period, without major changes in market conditions, the asset should be promising for the next holding period as well. Keeping portfolio turnover to a minimum is also a good way of generating high returns by avoiding unnecessary transaction costs. Perrin and Richard define portfolio turnover as [2]

\[
\tau(x|\bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x}_i)^+ + \sum_{i=1}^{n} (\bar{x}_i - x_i)^+ = \sum_{i=1}^{n} |x_i - \bar{x}_i| \tag{76}
\]

In other words, the turnover rate can be thought of as the change between the new portfolio after re-balancing \(x\) and the old portfolio \(\bar{x}\).

A secondary method of portfolio evaluation will consist of adding noise to the data set and investigating the impact this has on the portfolio weights. Adding a zero mean Gaussian random variable to each asset \(a\) effectively creates a new data set with the same behaviour...
on a large scale as the original data set but with disturbances in the day-to-day returns. The Gaussian random variables have a variance equivalent to the variance of the asset to which it is added.

\[ a_i^{\text{noise}} = a_i + \epsilon \]  

(77)

\( \epsilon \in \mathcal{N}(0, \sigma_i^2) \) with \( \sigma_i \) being the standard deviation of the \( i^{th} \) asset.

By comparing the average weights over the investment period of each strategy with and without noise each strategy’s sensitivity to input parameters is evaluated.

\[ \Delta x = \sum_{i=1}^{n} x_i^{\text{noise}} - x_i \]  

(78)

With \( n \) being the total amount of portfolio readjustments.

### 4.1.2 Performance evaluation

As a tool for further model comparison, some general financial metrics such as performance and risk statistics will be computed for each strategy. _Compound annual growth rate_ (CAGR) is a rate of return which is defined as

\[ \text{CAGR} = \left( \frac{V_{\text{end}}}{V_{\text{begin}}} \right)^{1/N} - 1 \]  

(79)

\( V_{\text{end}} \) and \( V_{\text{begin}} \) are portfolio values and \( N \) is the number of investment periods.

Portfolio volatility is a very useful statistic for portfolio performance evaluation. For a set of portfolio weights \( x \), portfolio volatility is defined as

\[ \sigma = \sqrt{x^\top \Sigma x} \]  

(80)

Maximum drawdown (MDD) is another useful statistic measuring the volatility of a portfolio. MDD can be thought of as the greatest drop in portfolio value over the investment period

\[ \text{MDD}(\text{portfolio}) = \frac{P - L}{P} \]  

(81)

Where \( P \) is the peak portfolio value before the largest drop and \( L \) is the lowest portfolio value before a new high is established.

The _Sharpe ratio_ is a statistic which gives a broad overview of the risk adjusted return for an investment strategy. It is defined as

\[ \text{Sharpe} = \frac{R_p - R_f}{\sigma_p} \]  

(82)

Finally, _equity curves_ showing model performance over the investment period will be included for both the traditional strategies and the new strategies. It is worth noting that both CAGR and the Sharpe ratio are per allocation period and not per annum. A wide set of measurements provide a good overall picture of the strength and weaknesses associated with each allocation philosophy.
4.2 Data

The data set has been provided by Erik Penser Bank and consists of historical asset closing prices. A decision to work with a variety of assets was made in agreement with Erik Penser Bank to demonstrate the objective of each allocation strategy as clearly as possible. In order to avoid the effects of selection and survivorship bias, the data set consists of indices covering a broad spectrum of asset classes rather than individuals stocks, bonds or currency pairs. The indices cover the asset classes Swedish equities, global equities, alternative investments, short rates, long rates and commodities. These indices include higher as well as lower risk investments together with assets with higher and lower correlation which allows for interesting portfolios. The data has been extracted from Bloomberg and the respective tickers are SBX Index, NDUEACWF Index, HFRXGL Index, RXVX Index, RXBO Index and CRYTR Index.

4.2.1 Expected return

The standard estimator of the expected value of a random variable is the sample mean. A simple way of estimating expected returns for a mean variance asset allocation is the sample mean.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$  \hspace{1cm} (83)

In many cases, it may be difficult to justify using sample means as an estimator for expected returns of a set of assets. For instance, a set of stocks trading on the same market will inevitably be correlated in some way, which makes the problem of estimating returns a multivariate one. Michaud states that the sample mean is a poor estimator of the expected return due to the fact that it ignores the multivariate structure of the problem [5]. Setting expected returns according to the subjective view of the portfolio manager is also a possible solution. In this study, sample means are utilized and re-estimated at each re-balancing point.

An estimate of the expected return computed during portfolio re-balancing is assumed to accurately predict the expected return during the next holding period. There are arguments for and against extrapolating from historical data in this way. An asset increasing in value during a holding period does not imply the asset will keep increasing in value during the next holding period. As extrapolating from historical data is the standard method of parameter estimation for portfolio optimization, the same methods will be applied in this study. Given that historical extrapolation is prone to produce misleading estimators and the fact that mean variance optimization tends to maximize the effect of errors in the parameter estimation [5], looking towards alternative portfolio optimization strategies is well justified. If one insists on implementing portfolio optimization models utilizing the expected return as a parameter, more powerful estimation techniques should be considered.

4.2.2 Covariance matrix

The primary component in all optimization programs treated in this text is the covariance matrix $\Sigma$. The sample covariance matrix will be utilized as the covariance matrix’s estimator, which is defined as
It is known that the sample covariance matrix $Q$ is an unbiased estimator of the covariance matrix $\Sigma$. There are many possible estimation methods for correlation between a set of assets. For the purposes of this study, simple windows with a size of 100 days will be used for estimation of asset covariances. The asset holding period, the time between re-balancing dates, is also 100 days. It is important to keep the window size short to ensure the underlying time series is stationary but not too short to avoid losing predictive power.

4.2.3 Analysis of data

Financial time series have a tendency to go up and down over time. In times of financial turmoil, indices experience periods of increased volatility. Figures 1 and 2 show both the absolute development and relative development of the time series (indices) that make up the data set.
The two previous figures show the indices varying levels of volatility and return over time. As expected, the highest volatility is found in the stock indices and the lowest volatility is found in the rates.

Analyzing the data from a time series perspective, one quickly finds that the series are, as with most financial time series, non-stationary. This means that over time, the mean and variance of the different time series, for a fixed window size, are not constant. A commonly used way to make time series stationary is to transform the series to log differences of consecutive days $R_t$.

$$R_t = \ln \left( \frac{S_t}{S_{t-1}} \right)$$  \hspace{1cm} (85)

$S_t$ is the asset price at time $t$. Figure 3 shows the transformed time series.
By studying figure 3, one can deduce that the transformed series have approximately zero mean independently of size and location of the window. However, as with most financial time series, the transformed series do not have constant volatility over time. All indices in the data set experience volatility clustering, which is most apparent in the two top charts at 2009 (the financial crisis). A time series may be non-stationary as a whole while still being stationary during certain periods. Dividing the indices into windows of size 100 days results in many stationary sub-intervals. A window size of 100 effectively means that the portfolio is reallocated every 100 days and the calculations are based on the past 100 days of data. The past 100 days of covariances are assumed to accurately predict correlation for the next holding period. Justifying extrapolating from historical data in this fashion is only possible when the underlying sub-interval of the time series is stationary. The only sub-intervals violating the stationarity criterion of constant variance are the periods surrounding the financial crisis. These parts of the data set are assumed to not significantly negatively affect the results.

Daily log returns of each asset are used to estimate covariances and expected returns. Since the returns data consists of log returns, expected returns have been estimated by summing the past 100 days of log returns. Covariances and correlations have been estimated using built in functions of the Python library Numpy, which are just implementations of the sample covariance matrix (84).
5 Results

The following section is devoted to solving the portfolio optimization models introduced throughout section (2). Following each solution, the model in question will be applied to the data set discussed in section (4.2) in order to demonstrate its properties independently. The section ends with a strategy comparison and presentation of portfolio evaluation statistics. To refresh the memory of the reader, all sections begin with a brief introduction to the theory of each model. Solutions using ADMM make use of the strategy covered in section (3.2.6) showing how the algorithm may be applied to a general constrained convex optimization program.

5.1 Global minimum variance

Section (2.1) introduced a simplified version of mean variance optimization which does not include the expected return as a parameter, the minimum variance portfolio. As described in section (4.2.1), avoiding reliance on an estimator of the expected return is crucial to obtain more robust optimization programs. The following section presents a possible solution to the problem of asset diversification in the minimum variance portfolio.

5.1.1 Optimization program

To introduce diversification in GMV, Perrin and Roncalli propose adding a constraint on the minimum number of effective bets to the optimization program \[2\]. Such a constraint can be thought of as the minimum number of assets with a significant weight in the portfolio.

\[
x^* = \arg \min_x \frac{1}{2} x^T \Sigma x
\]

s.t. \[
\begin{align*}
1_n^T x &= 1 \\
0_n &\leq x \leq x^+ \\
\mathcal{N}(x) &\geq N^- 
\end{align*}
\]

One possible choice of diversification measure is the Herfindhal index.

\[
\mathcal{H}(x) = \sum_{i=1}^n x_i^2
\]

Which takes the value one in a single asset portfolio and \(1/n\) in an equally weighted portfolio. The number of effective bets is defined as the inverse of the Herfindhal index.

\[
\mathcal{N}(x) = \mathcal{H}(x)^{-1}
\]

5.1.2 Solution: Quadratic programming

Program (86) may be solved using quadratic programming. The most obvious issue initially seems to be the constraint on the minimum number of effective bets \(\mathcal{N}(x) \geq N^-\). However, notice how the function \(\mathcal{N}(x)\) may be written alternatively as

\[
\mathcal{N}(x) = \mathcal{H}(x)^{-1} = (x^T x)^{-1} = \frac{1}{x^T x}
\]

Instead of having the diversification constraint appear as a constraint, a penalty is added to the objective function
\[ x^*(\lambda) = \arg \min_x \frac{1}{2} x^\top \Sigma x + \lambda x^\top x \]
\[
\text{s.t. } \begin{cases} 
1_n^\top x = 1 \\
0_n \leq x \leq x^+ 
\end{cases}
\] (90)

Program (90) is a QP problem with equality and inequality constraints. The optimal solution depends on the parameter \( \lambda \geq 0 \), which in turn determines the number of effective bets in the solution. Since the effect of the diversification constraint is of interest, the maximum asset weight limit \( x^+ \) will remain equal to one in all solutions. Notice how setting \( \lambda = 0 \) results in the standard GMV portfolio and setting \( \lambda = \infty \) results in an equally weighted portfolio. Hence, the effective bet function \( N(x) \) is an increasing function of \( \lambda \). Perrin and Roncalli suggest using the bisection method to determine \( \lambda \) [2].

5.1.3 Application to data

Consider the following application of program (90) to the data set described in section (4.2).
To find a portfolio with a certain number of minimum effective bets, the non-linear equation \( N(\lambda) = N^* \) is solved using the bisection method. Table 1 summarizes the results. As the number of effective bets is increased, the solution becomes more diversified. Selecting a number of effective bets equal to the number of assets yields the equally weighted portfolio, which is to be expected.

<table>
<thead>
<tr>
<th>Number of effective bets</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish stocks</td>
<td>0</td>
<td>0.002</td>
<td>0.014</td>
<td>0.045</td>
<td>0.092</td>
<td>0.166</td>
</tr>
<tr>
<td>Global stocks</td>
<td>0</td>
<td>0.001</td>
<td>0.019</td>
<td>0.063</td>
<td>0.110</td>
<td>0.166</td>
</tr>
<tr>
<td>Alternative investments</td>
<td>0.05</td>
<td>0.148</td>
<td>0.255</td>
<td>0.242</td>
<td>0.211</td>
<td>0.167</td>
</tr>
<tr>
<td>Short rates</td>
<td>0.845</td>
<td>0.565</td>
<td>0.354</td>
<td>0.294</td>
<td>0.237</td>
<td>0.167</td>
</tr>
<tr>
<td>Long rates</td>
<td>0.105</td>
<td>0.281</td>
<td>0.329</td>
<td>0.290</td>
<td>0.238</td>
<td>0.167</td>
</tr>
<tr>
<td>Commodities</td>
<td>0</td>
<td>0.003</td>
<td>0.028</td>
<td>0.066</td>
<td>0.112</td>
<td>0.167</td>
</tr>
</tbody>
</table>

Table 1: Average portfolio weights depending on number of effective bets, minimum variance portfolio

5.2 Most diversified

The most diversified portfolio was introduced in section 2.2. In short, the goal of this particular model is to generate a set of weights maximizing a diversification measure. The following section presents a solution to the most diversified portfolio with an added diversification constraint.

5.2.1 Optimization program

To solve for the most diversified portfolio, consider the following optimization program

\[ x^* = \arg \min_x \frac{1}{2} \ln (x^\top \Sigma x) - \ln (x^\top \sigma) \]
\[
\text{s.t. } \begin{cases} 
1_n^\top x = 1 \\
x \in \Omega 
\end{cases}
\] (91)
If the constraints Ω are limited to weight constraints, such as the box constraint \(0_n \leq x \leq x^+\), the preceding program is easy to solve using non-linear constrained optimization methods. The package scientific python contains optimization methods using sequential quadratic programming that are able to solve problems of this type.

The most diversified portfolio suffers from poor diversification, allocating much of the weight to only a few assets. Diversification may be addressed in a way similar to how it was handled in the minimum variance portfolio, by introducing a diversification constraint and limiting it to be greater than some minimal value \(\mathcal{D}(x) \geq \mathcal{D}^-\). After limiting the number of effective bets to be greater than some minimal value, the new optimization program becomes

\[
x^* = \arg\min_x \frac{1}{2} \ln (x^\top \Sigma x) - \ln (x^\top \sigma)
\]

\[
\text{s.t. } \begin{align*}
1_n^\top x &= 1 \\
0_n \leq x &\leq x^+ \\
\mathcal{N}(x) &\geq \mathcal{N}^-
\end{align*}
\] (92)

### 5.2.2 Solution: ADMM

Incorporate the constraints into the objective function using the convex indicator function

\[
\{x^*, y^*\} = \arg\min_{(x,y)} f_{\text{MDP}}(x) + 1_{\Omega_1}(x) + 1_{\Omega_2}(y)
\]

\[
\text{s.t. } x = y
\]

\[
1_{\Omega}(x) = \begin{cases} 
0, & \text{if } x \in \Omega \\
\infty, & \text{if } x \notin \Omega 
\end{cases}
\] (94)

\(f_{\text{MDP}}(x)\) is the objective function of the most diversified portfolio, \(\Omega_1 = \{x \in \mathbb{R}^n : 1_n^\top x = 1\}\), and \(\Omega_2 = \{x \in \mathbb{R}^n : 0_n \leq x \leq x^+ \cap \mathcal{B}_2(0_n, \sqrt{\frac{1}{\mathcal{N}^-}})\}\). The domain \(\Omega_2\) describes the intersection between the weight constraints and the constraint on the minimum number of effective bets. Perrin and Richard show that the constraint on the minimum number of effective bets can be represented by the ball \(\mathcal{B}_2(0_n, \sqrt{\frac{1}{\mathcal{N}^-}})\) [2]. The ADMM iterations for optimization program (93) are

\[
x^{k+1} = \arg\min_x \left(f(x) + (\rho/2) \|x + y^k + u^k\|^2_2\right)
\]

\[
y^{k+1} = \arg\min_y \left(g(y) + (\rho/2) \|x^{k+1} + y + u^k\|^2_2\right)
\]

\[
u^{k+1} := u^k + x^{k+1} - y^{k+1}
\] (95)

With \(f(x)\) and \(g(y)\) set to

\[
f(x) = f_{\text{MDP}}(x) + 1_{\Omega_1}(x) \quad (96) \quad g(y) = 1_{\Omega_2}(y) \quad (97)
\]

The x-update reduces to a constrained non-linear optimization program

\[
x^{(k+1)} = \arg\min_x \frac{1}{2} \ln (x^\top \Sigma x) - \ln (x^\top \sigma) + \frac{\rho}{2} \|x + y^k + u^k\|^2_2
\]

\[
\text{s.t. } 1_n^\top x = 1
\] (98)
The $y$-update corresponds to the Euclidean projection onto the domain $\Omega_2$

$$y^{(k+1)} = P_{\Omega_2} \left( x^{(k+1)} + u^{(k)} \right) \quad (99)$$

Another way of viewing the $y$-update is as the proximal operator of $f_1(y) = 1_{\Omega_3}(y)$ and $f_2(y) = 1_{\Omega_4}(y)$

$$y^{k+1} = \arg\min_y f_1(y) + f_2(y) + \frac{1}{2} \| y - v \|^2_2$$

$$= \text{prox}_f(v) \quad (100)$$

where $\Omega_3 = B_2(0_n, \sqrt{\frac{1}{N-1}})$, $\Omega_4 = \{ x \in \mathbb{R}^n : 0_n \leq x \leq x^+ \}$ and $v = x^{(k+1)} + u^{(k)}$.

Computing the proximal operator of a convex sum of functions $f(y) = f_1(y) + f_2(y)$ is made possible by Dyktra’s algorithm, section (3.2.9). Perrin and Richard show the iterations to be [2]

$$\begin{cases}
  y^{(k+1)} = \frac{1}{\max \left( 1, \sqrt{\frac{1}{N-1}} \| w^{(k)} + z_1^{(k)} \|_2 \right)} (w^{(k)} + z_1^{(k)}) \\
  z_1^{(k+1)} = y^{(k)} + z_1^{(k)} - x^{(k+1)} \\
  w^{(k+1)} = \max \left( x^-, \min \left( y^{(k+1)} + z_2^{(k)}, x^+ \right) \right) \\
  z_2^{(k+1)} = y^{(k+1)} + z_2^{(k)} - w^{(k+1)}
\end{cases} \quad (101)$$

Initial values are set to $y^{(0)} = w^{(0)} = v$ and $z_1^{(0)} = z_2^{(0)} = 0_n$. The termination rule is $\| y^{(k+1)} - y^{(k)} \| < \varepsilon$.

### 5.2.3 Application to data

Consider the following application of program (92) to the data set described in section (4.2).

<table>
<thead>
<tr>
<th>Number of effective bets</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish stocks</td>
<td>0.006</td>
<td>0.020</td>
<td>0.036</td>
<td>0.063</td>
<td>0.124</td>
<td>0.147</td>
</tr>
<tr>
<td>Global stocks</td>
<td>0</td>
<td>0.001</td>
<td>0.012</td>
<td>0.029</td>
<td>0.096</td>
<td>0.137</td>
</tr>
<tr>
<td>Alternative investments</td>
<td>0.019</td>
<td>0.064</td>
<td>0.145</td>
<td>0.190</td>
<td>0.187</td>
<td>0.175</td>
</tr>
<tr>
<td>Short rates</td>
<td>0.901</td>
<td>0.644</td>
<td>0.377</td>
<td>0.286</td>
<td>0.190</td>
<td>0.177</td>
</tr>
<tr>
<td>Long rates</td>
<td>0.068</td>
<td>0.256</td>
<td>0.391</td>
<td>0.363</td>
<td>0.258</td>
<td>0.203</td>
</tr>
<tr>
<td>Commodities</td>
<td>0.005</td>
<td>0.015</td>
<td>0.039</td>
<td>0.069</td>
<td>0.145</td>
<td>0.161</td>
</tr>
</tbody>
</table>

Table 2: Average portfolio weights depending on number of effective bets, most diversified portfolio

The most diversified portfolio favours the index comprised of rates with short maturity since this asset has very low volatility. In this case, the portfolio is diversified in terms of each asset contributing approximately equally to portfolio volatility. However, the portfolio is not diversified with respect to asset concentration. Applying a constraint on the minimum number of effective bets improves the most diversified portfolio from a risk management and regulatory perspective, since it is no longer concentrated on just a single asset.
5.3 Equal risk contribution

Section (2.3) introduced the concept of an equal risk contribution portfolio, which is a special case of a more general risk budgeting portfolio. In the following section, two solutions to equal risk contribution are presented. One solution is found via cyclical coordinate descent and the other is found by way of alternating direction method of multipliers.

5.3.1 Optimization program

Richard and Roncalli formulate the general risk budgeting program as [6]

\[
x^\star(S, \Omega) = \arg \min_{x \in S \cap \Omega} R(x)
\]

s.t. \[
\sum_{i=1}^{n} b_i \ln x_i \geq \kappa \star
\]

They also make note of the following two results which allow for simplification of the last program.

- The logarithmic barrier constraints guarantees \( x \geq 0 \);
- There is only one constant \( \kappa^\star \) such that the constraint \( \sum_{i=1}^{n} b_i \ln x_i = 1 \) is satisfied.

\[
x^\star(\Omega, \kappa) = \arg \min \, R(x)
\]

s.t. \[
\sum_{i=1}^{n} b_i \ln x_i \geq \kappa
\]

To incorporate the constraint \( \sum_{i=1}^{n} b_i \ln x_i \geq \kappa \), the program is rewritten using the corresponding Lagrange formulation

\[
x^\star(\Omega, \lambda) = \arg \min \, R(x) - \lambda \sum_{i=1}^{n} b_i \ln x_i + 1_\Omega(x)
\]

The characteristics of the solution depends on what constraints \( \Omega \) are applied and the parameter \( \lambda \). A volatility based risk measure is applied \( R(x) = \sqrt{x^\top \Sigma x} \) and the risk budgets are set to \( b_i = \frac{1}{n} \).

5.3.2 Solution: ADMM

Incorporate the constraint \( \Omega \) into the objective function of program (104) using the convex indicator function

\[
x^\star(\Omega, \lambda) = \arg \min \, R(x) - \lambda \sum_{i=1}^{n} b_i \ln x_i + 1_\Omega(x)
\]

\[
1_\Omega(x) = \begin{cases} 0, & \text{if } x \in \Omega \\ \infty, & \text{if } x \notin \Omega \end{cases}
\]

Richard and Roncalli emphasize the importance of the parameter \( \lambda \) as the constrained risk budgeting portfolio is found to be the optimal value \( \lambda^\star \) such that the constraint \( \frac{1}{n} x = 1 \) holds [6]. They suggest applying the bisection algorithm to find \( \lambda^\star \) and ADMM to find
An alternative strategy consists of choosing a value for \( \lambda \) and then normalizing the resulting weights, which is the method that will be applied here.

\[
x_{\text{ERC}} = \frac{x^*(\lambda)}{1^T x^*(\lambda)}
\]  

(107)

The ADMM iterations for optimization program (105) are

\[
x^{k+1} := \arg\min_x \left( f(x) + \left( \frac{\rho}{2} \| x + y^k + u^k \|_2^2 \right) \right)
\]

\[
y^{k+1} := \arg\min_y \left( g(y) + \left( \frac{\rho}{2} \| x^{k+1} + y + u^k \|_2^2 \right) \right)
\]

\[
u^{k+1} := u^k + x^{k+1} - y^{k+1}
\]

(108)

With \( f(x) \) and \( g(y) \) set to

\[
f(x) = R(x) + \lambda \sum_{i=1}^n b_i \ln y_i
\]

(109)

\[
g(y) = -\lambda \sum_{i=1}^n b_i \ln y_i
\]

(110)

An iterative formula for the x-update is

\[
x^{(k+1)} = \arg\min \left\{ f(x) + \frac{\varphi}{2} \| x - y^{(k)} + u^{(k)} \|_2^2 \right\}
\]

(111)

Before substituting the general risk measure \( R(x) \) for an explicit expression, note that minimizing \( \sqrt{x^T \Sigma x} \) is the same as minimizing \( \frac{1}{2} x^T \Sigma x \). Consider the following simplification to the objective function \( f^{(k)}(x) \) of the x-update

\[
f^{(k)}(x) = \frac{1}{2} x^T \Sigma x + \frac{\varphi}{2} \left( x - y^{(k)} + u^{(k)} \right)^T \left( x - y^{(k)} + u^{(k)} \right)
\]

(112)

Which uses the fact that the Euclidean norm can be expressed as \( \| v \|_2^2 = v^T v \). Rewriting this expression to isolate the functions of \( x \) yields

\[
f^{(k)}(x) = \frac{1}{2} x^T \Sigma x + \frac{\varphi}{2} \left( x^T x - 2 x^T y^{(k)} + \left( v^{(k)} \right)^T y^{(k)} \right)
\]

(113)

where \( v^{(k)} = y^{(k)} - u^{(k)} \). Since \( \left( v^{(k)} \right)^T v^{(k)} \) does not depend on \( x \), it is left out of the objective function

\[
x^{(k+1)} = \arg\min \frac{1}{2} x^T \left( \Sigma + \varphi I_n \right) x - \varphi x^T \left( y^{(k)} - u^{(k)} \right)
\]

s.t. \( x \in \Omega \)

(114)

If the constraints \( \Omega \) are linear, the x-update becomes a standard QP problem. Since the constraints \( x \geq 0 \) and \( 1_n^T x = 1 \) are embedded in the optimization, the x-update may be solved unconstrained as well. In this case, an iterative formula for the x-update is

\[
x^{(k+1)} = \left( \Sigma + \varphi I_n \right)^{-1} \varphi \left( y^{(k)} - u^{(k)} \right)
\]

(115)
Which follows from setting the derivative with respect to \( x \) to zero and solving for the optimal value \( x^{(k+1)} \).

An iterative formula for the \( y \)-update is

\[
y^{(k+1)} = \arg \min \left\{ g(y) + \frac{\varphi}{2} \left\| x^{(k+1)} - y + u^{(k)} \right\|^2 \right\}
\]  ____1____ \right\} \tag{116}
\]

Which is the proximal operator of the logarithmic barrier function. Using identity (70) the solution is deduced to be

\[
y^{(k+1)}_i = \frac{x^{(k+1)}_i + u^{(k)}_i + \sqrt{(x^{(k+1)}_i + u^{(k)}_i)^2 + 4\lambda b_i \varphi^{-1}}}{2} \tag{117}
\]

An iterative formula for the dual variable update is

\[
u^{(k+1)} = u^{(k)} + \left( x^{(k)} - y^{(k)} \right) \tag{118}
\]

5.3.3 Moving away from equal risk

Changing the risk budget to allocate a certain risk to each asset is easy in the ADMM solution of the ERC portfolio.

5.3.4 Application to data

Consider the following application of program (104) to the data set described in section (4.2). For equal risk contribution, the risk budgets are set to

\[
b = \left[ \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right]
\]

For the risk budgeting portfolio, the stock indices are given a higher budget while the rates, commodities and alternative investments are given less weight.

\[
b = \left[ \frac{1}{3}, \frac{1}{3}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right]
\]

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Swedish stocks</th>
<th>Global stocks</th>
<th>Alt. investments</th>
<th>S. rates</th>
<th>L. rates</th>
<th>Commodities</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERC</td>
<td>0.010</td>
<td>0.011</td>
<td>0.050</td>
<td>0.781</td>
<td>0.137</td>
<td>0.011</td>
</tr>
<tr>
<td>RB</td>
<td>0.026</td>
<td>0.031</td>
<td>0.043</td>
<td>0.689</td>
<td>0.199</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 3: Effect of varying the risk budgets, risk budgeting portfolio

5.3.5 Solution: CCD

The equal risk contribution portfolio is the scaled solution to \( x^*/\left(1_n^T x^*\right) \) to

\[
x^* = \arg \min_x \frac{1}{2} x^T \Sigma x - \lambda \sum_{i=1}^n b_i \ln x_i
\]  ____2____ \frac{\lambda}{n} x_i^{-1} = 0 \text{ or } x_i (\Sigma x)_i - \frac{\lambda}{n} = 0. \]  ____3____ last expression, consider the vector \( x = (x_1, x_2, x_3)^T \) and covariance matrix
The matrix-vector product $\Sigma x$ is
\[
\Sigma x = \begin{bmatrix}
\sigma_i^2 x_1 + \rho_{1,2} \sigma_1 \sigma_2 x_2 + \rho_{1,3} \sigma_1 \sigma_3 x_3 \\
\rho_{1,2} \sigma_1 \sigma_2 x_1 + \sigma_2^2 x_2 + \rho_{2,3} \sigma_2 \sigma_3 x_3 \\
\rho_{1,3} \sigma_1 \sigma_3 x_1 + \rho_{2,3} \sigma_2 \sigma_3 x_2 + \sigma_3^2 x_3 
\end{bmatrix}
\] (121)

By selecting the $i^{th}$ row of the matrix-vector product $(\Sigma x)_i$ and multiplying by $x_i$ the following is obtained
\[
x_i^2 \sigma_i^2 + x_i \sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \frac{\lambda}{n} = 0 
\] (122)

The second order equation (122) is solved to construct a system of equations for all assets $i$. The result is the following update for the $i^{th}$ asset.
\[
x_i^{(k+1)} = \frac{-v_i^{(k+1)} + \sqrt{(v_i^{(k+1)})^2 + 4\lambda \sigma_i^2}}{2\sigma_i^2}
\] (123)

Where $v_i^{(k+1)}$ is defined as
\[
v_i^{(k+1)} = \sigma_i \sum_{j < i} x_j^{(k+1)} \rho_{i,j} \sigma_j + \sigma_i \sum_{j > i} x_j^{(k)} \rho_{i,j} \sigma_j 
\] (124)

Separating the information gained by already updated $x$ coordinates and the ones not yet updated.

### 5.4 Evaluation of traditional strategies

Figure 4 and Table 5.4 present statistics summarizing the performance of the traditional strategies. Equity curves and statistics depend on penalizing parameters, weight caps and constraints. With this in mind, several versions of the same strategy are not included in the results to avoid producing misleading statistics.
There is considerable variation in the performance of the optimization strategies. On the one hand, the yield of the models range from approximately 1.5 times the initial investment in the unconstrained minimum variance portfolio to nearly 4 times the initial investment in the turnover constrained mean variance portfolio. On the other hand, the increased yield results in greater volatility with 1.15% and 8.74% in the minimum variance and mean variance portfolios, respectively.

Figure 5 shows the average portfolio weights over the investment period. The mean variance strategy tends to allocate more weight to the risky assets, in this case the equity indices. As the rates have much lower volatility in comparison to the equity indices, it is not surprising that the minimum variance models allocate greater weight to these asset classes.
Figure 6 show the traditional strategies average turnover rate over the investment period. To give an intuitive sense of how turnover is measured, a net sell of 25% and net buy of 25% results in a turnover of 50%. Not surprisingly, the mean variance model has a very high average turnover, indicating that the strategy is expensive on a relative basis when compared to the other strategies.

Figure 7 shows the effect on average turnover when adding noise to the data set for each of the traditional strategies. The mean variance models are both affected to a greater extent than the minimum variance models. The equally weighted portfolio is omitted from both of these statistics since its turnover rate is determined by changes in the market only.
Figure 6: Average strategy turnover

Figure 7: Noise impact on weights, traditional strategies
5.5 Evaluation of new strategies

The following section proceeds with a presentation of statistics for the new portfolio optimization strategies. Figure 8 shows the equity curves of the new strategies. Statistics are found in Table 5.

![Equity curves](image)

**Figure 8: Equity curves, backtest of new strategies**

<table>
<thead>
<tr>
<th>Metric</th>
<th>ERC</th>
<th>MDP</th>
<th>MV Diversification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGR</td>
<td>0.2</td>
<td>0.3</td>
<td>0.47</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.5</td>
<td>2.05</td>
<td>1.9</td>
</tr>
<tr>
<td>Maximum drawdown</td>
<td>-2.4</td>
<td>-6.3</td>
<td>-6.9</td>
</tr>
<tr>
<td>Sharpe</td>
<td>-0.52</td>
<td>-0.05</td>
<td>-0.012</td>
</tr>
</tbody>
</table>

**Table 5: Statistics, backtest of new strategies**

Visual inspection of the equity curves show that the most diversified portfolio and minimum variance portfolios experience greater returns than equal risk contribution over the investment period. However, the most diversified portfolio and minimum variance portfolios suffer from increased volatility in comparison to equal risk contribution. The number of effective bets is set to four in both the most diversified and minimum variance portfolios.
The statistics show the same results quantitatively. Volatility is 4 times higher in the most diversified portfolio and minimum variance portfolio as in equal risk contribution. Maximum drawdown and Sharpe ratio further reinforce these results.

Average weights over the market period for each strategy are displayed in Figure 9.

![Average portfolio weights](image)

**Figure 9: Average strategy weight, backtest of proposed strategies**

The difference in volatility and return of each strategy is a result of portfolio composition. Figure 9 shows that equal risk contribution had a significant average weight of the available capital invested in short rates, which is often classified as low risk, on a relative basis. The minimum variance and most diversified portfolios were constructed using four minimum effective bets which resulted in more diversified portfolios in terms of asset concentration in comparison to equal risk contribution.

Figure 10 shows the average turnover rate over the investment period for the traditional strategies.
The new strategies are all conservative with a low average turnover rate. Equal risk contribution allocated a significant weight to the short rates with minor changes to the portfolio over the investment period, resulting in a very low average turnover rate. The most diversified suffers from the highest turnover rate among the new strategies at nearly 45%.

As for sensitivity towards disturbances in the input data, all three new strategies are relatively robust with an average change in weights on the order of magnitude $10^{-2}$. Values on the change in average weights are found in Figure 11.

Figure 10: Average turnover, backtest of proposed strategies
Figure 11: Noise impact on weights, backtest of proposed strategies
6 Discussion

While the traditional mean variance model yields the greatest returns over the period examined, as seen when comparing figures (4) and (8), it is important not to judge the portfolio optimization models treated in this paper simply on which provides the greatest returns. As the mean variance model is based on an unreliable estimate of the expected return, it is not a defensible strategy from a financial perspective. For the same reason, the asset allocation generated by the mean variance model is difficult to understand intuitively. Emphasis should be placed on finding a strategy which produces portfolios that are robust and rely on parameter estimates that are justifiable.

6.1 Portfolio turnover rate

One of the most important attributes in a good portfolio optimization model is a low rate of turnover. In other words, it is important to keep readjustment of the portfolio by purchasing and selling assets to a minimum. Section (5.4) clearly demonstrates one of the drawbacks of the traditional mean variance portfolio, its high turnover rate. Continuing on this track, seeing as the turnover of traditional mean variance optimization is greater than 150%, this strategy effectively, on average, reallocates 150%/2 = 75% of the portfolio at each readjustment. In addition, the mean variance strategy places a high portion of its capital in stocks, paying the resulting commission related to these re-allocations will amount to a substantial cost and thus negatively affect the returns it is able to generate.

Section (5.5) present figures on the turnover rate in the new strategies, specifically figure (10). Equal risk contribution is the clear winner with the lowest turnover rate. However, the data set for this study is quite mixed with regards to the risk levels. A volatility based risk measure was utilized and the short rates have the lowest volatility by far, which can be seen in Table 6. The ERC portfolio places most of its weight on the short rates as a result. Hence, the low turnover rate of ERC may be a consequence of the high weight it places on short rates. Figure (9) shows the average weights of all the new strategies over the investment period. The minimum variance and most diversified portfolios are both more diversified in terms of asset concentration, placing more weight on the stocks and other assets, although still a low portion on a relative basis when compared to mean variance optimization. Out of these two, the minimum variance portfolio performs better with a diversified asset allocation and low turnover rate.

<table>
<thead>
<tr>
<th>Asset</th>
<th>Swedish stocks</th>
<th>Global stocks</th>
<th>Alt. assets</th>
<th>S. rates</th>
<th>L. rates</th>
<th>Commodities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vol (%)</td>
<td>1.25</td>
<td>0.9</td>
<td>0.22</td>
<td>0.009</td>
<td>0.125</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Table 6: Daily asset return volatility

6.2 Asset concentration

The issue of asset concentration has been brought up at many points throughout this report. Both the minimum variance and most diversified optimization models were solved with a constraint on asset diversification, which is just another term for asset concentration. As mentioned many times previously, an asset allocation with weights concentrated on just a few number of assets is not desirable for many reasons. The reader may be asking themselves why asset concentration is not considered a problem in the risk budgeting portfolio?
Table 3 shows two asset allocations of a risk budgeting type of which one utilizes equal risk budgets, creating an ERC portfolio. Both of these specific risk budgeting portfolios allocate a small, but not insignificant, weight to the stock indices (on the order of $10^{-2}$). Asset allocations without the diversification constraint for the minimum variance and most diversified portfolios are available in Tables 1 and 2, corresponding to the column with 1 effective bet. In these cases, the stock indices are allocated an almost negligible weight, demonstrating the fact that asset concentration is not that much of an issue in the risk budgeting portfolio. Applying the same portfolio optimization models to a more homogeneous data set consisting of, for instance, equities only, would undoubtedly show that asset concentration is not an issue in a risk budgeting portfolio, which cannot be said for the minimum variance and most diversified models.

### 6.3 Adding noise to the data

Figures (7) and (11) show the difference in average weights between the data set with and without noise for the traditional strategies and new strategies, respectively. The impact of noise is clearly greater on the traditional strategies as the difference in average portfolio weights is on the order of magnitude $10^{-1}$. The same measurements for the new strategies resulted in a difference in average portfolio weights on the order of magnitude $10^{-2}$. The conclusion is that the new strategies are not as sensitive to slight changes in the market conditions, which is desirable for a portfolio optimization strategy.

### 6.4 Model parameters

A benefit of the new strategies is the reliance on solely one input parameter, the covariance matrix. The mean variance portfolio is an error maximizer in the sense that it allocates the highest weight to the assets with high expected return, negative correlation and low variance [5]. The assets with high expected return are also inevitably the ones with the highest error of estimation since high returns show up in the tail of the underlying return distribution. As mentioned earlier, estimates of the mean are inherently unreliable, even in ideal cases when the data is independent and identically distributed [8]. Due to this fact, the authors argue that the exclusion of expected return as a input parameter is beneficial in the sense that it results in a more robust portfolio optimization strategy.

### 6.5 Alternative risk measures

Section (2.3.1) mentions two risk measures available for the purposes of portfolio optimization. A volatility based risk measure has been utilized when constructing the risk budgeting portfolios, which is expressed as

$$ R(x) = \sqrt{x^\top \Sigma x} $$

There are alternative risk measures available, besides the volatility based one used for this study. Value at Risk (VaR) and Expected Shortfall (ES) are two common risk measures in quantitative finance. However, due to the nature of VAR and ES, they are inefficient from a computational efficiency perspective as the distributions for all assets would have to be estimated at every portfolio readjustment. Alternatively, empirical VAR and ES are estimated from data, avoiding the risky task of properly fitting a distribution to the returns.
data. As a further alternative, Richard and Roncalli also mention the following risk measure, incorporating the expected return parameter [6]

\[ \mathcal{R}(x) = -x^\top (\mu - r) + c \cdot \sqrt{x^\top \Sigma x} \]  

(126)

\( r \) is the risk free interest rate for the period and \( c \) is a constant indicating how much the volatility should be weighted. The many alternative risk measures available are too numerous to be listed here. In the end, the investor should choose a risk measure deemed to appropriately measure risk for the specific application and not attempt to reverse engineer and find a measure of risk producing favourable results.

### 6.6 Further research

The risk budgeting portfolio used for the purposes of model comparison was limited to the special case of equal risk budgets. Varying the risk budgets to allocate a specific risk budget to each asset lets the investor create a general risk budgeting portfolio. Changing the risk budget is an easy task in the ADMM solution of the risk budgeting portfolio. Further research could aim to determine to what extent the risk budgets affect the turnover rate and sensitivity to small day-to-day changes in the market. Another possible area of further research is investigating the effects of alternative risk measures, as the study was limited to a volatility based risk measure.

Catastrophic risks may be hidden in the tails of the distribution of the daily returns data. Further research should consider utilizing a risk measure better suited to extreme risks. A volatility based risk measure may have a tendency to underestimate risk in the assets. Expected shortfall is a suitable measurement of extreme risk that may be present in the tails of the returns distribution which unfortunately carries a cost of high computational complexity. It may be worth sacrificing some computational efficiency for more robust risk modeling.

The mean variance portfolios solved in this study relied on an expected return parameter found through estimation by sample means. In many cases, sample means are not a justifiable way of estimating expected returns due to the multivariate nature of the distribution of the returns data. Further research could aim to determine the effects on the mean variance portfolio when utilizing more powerful estimation techniques for expected returns, to create a more reliable parameter estimate.

Throughout the report, the covariance matrix has been utilized when measuring the correlation structure of the data. The reader should bear in mind that this is a measure of linear dependence between assets, not designed for capturing non-linear behavior. Given that dependencies between assets tend to change over time and tail dependencies behave differently than the bulk of the distribution, it would be of interest to investigate how correlation measures for non-linear dependence, such as Kendall’s tau or copulas would perform.
References


