Copenhagen Business School

Master Thesis

Finance and Investments

Hierarchical Risk Parity – A Hierarchical Clustering-Based Portfolio Optimization

---

Authors:

Christian Erik Kaae (120398)

Joonas Aleksanteri Karppinen (141992)

Supervisor:

Johan Stax Jakobsen

15/05 - 2022

Number of pages: 87

Number of characters: 169,734

Abstract

In 2016 López de Prado developed the Hierarchical Risk Parity (HRP), a risk-based portfolio optimization algorithm that uses graph theory and machine learning to construct diversified portfolios. The HRP is a function of the return covariance matrix, but unlike most traditional portfolio optimization strategies, such as Markowitz's mean-variance framework, the HRP does not require it to be positive-definite. This thesis aims to contribute to the existing literature on the performance of the HRP and improve the HRP by alleviating three of its counterintuitive features. This is achieved by developing a modified version of HRP and then comparing the HRP to the modified version of the HRP, the minimum-variance portfolio, and other risk-based allocation strategies in three separate analyses. First, the robustness is evaluated through a comprehensive set of simulated simple block-diagonal correlation matrices and randomized block-diagonal correlation matrices. Second, the risk is assessed using simulated return series and covariance matrices. Lastly, the performance is analyzed using a walk-forward analysis applied to the historical data of the 30 Industry Portfolios.

The combined results provide further evidence for the HRP's competitiveness. The HRP achieves lower out-of-sample variance in Monte Carlo simulations than the minimum-variance portfolio, equal risk contribution portfolio, inverse-variance portfolio, and equally weighted portfolio. However, the HRP shows less robustness towards misspecifications of the sample covariance matrix than other non-hierarchical allocations. Lastly, HRP outperforms the inverse-variance, equally weighted, and equal risk contribution portfolios on risk-based performance measures in the walk-forward analysis but underperforms relative to the minimum-variance portfolio.

Furthermore, the counterintuitive features of the HRP are solved by developing a modified version of the HRP, referred to as the HRP Topdown. The empirical analysis shows that the HRP Topdown significantly increases the returns and performs better from a risk-based perspective than the original HRP. However, the outperformance comes with a cost of higher transaction costs, limiting the practical implementation of the HRP Topdown method. Nonetheless, the benefit of the HRP Topdown method is that it replaces some of the counterintuitive and arbitrary choices made in the original HRP with rational alternatives.

Keywords: Asset Allocation, Hierarchical Clustering, Hierarchical Risk Parity, Graph Theory, Machine Learning, Portfolio Optimization

Contents

List of Figures

List of Tables

1 Introduction

One of the most recurrent financial problems is the issue of portfolio construction (López de Prado, 2016b), which refers to the process of efficiently allocating wealth among asset classes and securities (Maillard et al., 2010). The field of portfolio construction has a longstanding history in the academic literature. In 1952, Markowitz introduced what was to become one of the most influential frameworks of modern portfolio theory, namely the mean-variance framework. His framework builds on the assumption that a rational investor seeks to maximize the expected return for a given level of risk. An important implication of Markowitz's framework is that it is rarely optimal to allocate all capital to the assets with the highest expected returns. Instead, one should consider how the assets behave as a group, taking correlation into account when building the optimal portfolio (López de Prado, 2016b).

Over seventy years later, Markowitz's work is still at the heart of modern portfolio theory. However, despite its theoretical advantages, the solution often suffers from serious drawbacks in its practical implementation (Maillard et al., 2010). There are two fundamental problems when using Markowitz's framework in practice:

1. Estimating expected returns
2. Estimating a positive-definite covariance matrix

These issues relate to the fact that in Markowitz's framework, the portfolio optimization is solved as a deterministic problem – the estimation errors in the inputs are completely ignored (Kolm et al., 2014). However, this assumption of precise expected returns and covariance matrix estimates is not prudent.

Recently, several risk-based strategies have entered the field of portfolio theory in response to the difficulty of estimating the expected returns precisely. These new strategies focus on portfolio risk and disregard the expected return, eliminating some estimation errors. However, most risk-based methods still require the estimation of a covariance matrix. Although studies have shown that the estimation of expected returns drives the majority of the instability in a portfolio optimization compared to the estimation of the covariance matrix (Kallberg and Ziemba, 1984; Chopra, 1993; Merton, 1980; Jagannathan and Ma, 2003), this does not mean that one can disregard the issues that arise from relying on the covariance matrix when conducting a risk-based allocation analysis. In fact, the need for diversification increases with asset correlation. Moreover, higher correlation leads to more ill-conditioned covariance matrices, where the estimation errors may be large enough to offset any benefits of diversification – a phenomenon commonly referred to as Markowitz's curse (López de Prado, 2016b).

In response to Markowitz's curse, a myriad of research focusing on techniques to improve the robustness and reduce the numerical instability of the covariance matrix estimator has emerged. For example, one of the issues with using the sample covariance matrix is that it assigns equal weights to the most recent and the most distant observations and, therefore, may not reflect the current state of the market. Instead, one can use exponentially weighted variances, which gives greater weight to the most recent observation, thus expressing a more dynamic view of the market, where current market conditions are taken into consideration more accurately (Horasanlı and Fidan Kececi, 2007). Alternatively, other methods aim to transform the sample covariance matrix to make

it more stable. The most famous solution to a more robust covariance estimator is the shrinkage transformation proposed by Ledoit and Wolf (2003, 2004a,b), where the most extreme coefficients are pulled towards more central values, thereby systematically reducing estimation error where it matters most. Other methods have focused on Random Matrix Theory (RMT) to model and filter the randomness of correlations between assets.

Raffinot (2017), along with other researchers, contends that a correlation matrix is too complex to be fully analyzed and disregards the hierarchical structure of correlations. Thus, López de Prado (2016b) suggests a different approach to reduce estimation errors by grouping the data into natural groups using clustering methods and circumventing the need for the inversion of the covariance matrix. Clustering is a machine learning technique that groups data without the need for class labels and thus does not use category labels that tag objects with prior identifiers, i.e., no prior labels to the data (Jain, 2010). So while there are other ways of grouping data in finance, clustering benefits from not requiring any financial, economic, or psychological explanation for grouping (Papenbrock, 2011). One popular branch of clustering methods is hierarchical clustering which groups data using an inherent hierarchical structure. Hierarchical clustering has a filtering effect in correlation-based clustering, which finds the links in the correlation network that are important and puts them at the center of attention, resulting in more reliable outcomes Raffinot (2017); Tumminello et al. (2010).

To address the flaws of traditional portfolio optimization strategies, Lopéz de Prado (2016) introduced a new allocation strategy called Hierarchical Risk Parity (HRP). His strategy combines the advantages of risk-based models with hierarchical clustering and enables managers to compute a portfolio on a singular covariance matrix - an impossible feat for methods that rely on quadratic optimization such as Markowitz's mean-variance framework (López de Prado, 2016b). The HRP approach can be broken down into the following three stages:

1. **Tree clustering**: Grouping of similar assets into clusters, based on a proper distance metric and single linkage clustering.
2. **Quasi-diagonalization**: Reorganization of the rows and columns of the covariance matrix, so that the largest values lie along the diagonal.
3. **Recursive bisection**: Splitting the allocations through recursive bisection of the reordered covariance matrix.

In effect, HRP circumvents the issue of forecasting returns and the inversion of the covariance matrix, which should reduce the instability and concentration of the constructed portfolio while improving its performance.

1.1 Problem statement

In his seminal paper on the Hierarchical Risk Parity, López de Prado (2016b) was able to show that the HRP delivers lower out-of-sample variance than the minimum-variance portfolio and traditional risk parity's inverse-variance portfolio. While the results of López de Prado's out-of-sample simulations are promising, the study only compares HRP against two allocation strategies: the minimum-variance and inverse-variance portfolios. In addition, to compare different strategies, López de Prado uses only two out-of-sample performance

measures, variance and Sharpe ratio. Thus, this thesis will strive to add to the empirical evidence of the HRP performance by comparing it to more traditional risk-based portfolio allocation techniques and including a more comprehensive set of performance measures. The main research question of this thesis is therefore formulated as follows:

How does the HRP allocation technique perform compared to the minimum-variance portfolio and other risk-based allocation strategies?

Several sub-questions are posed to answer the above-stated research question:

- *How much is the HRP portfolio impacted by instability in the covariance matrix compared to non-hierarchical allocation strategies?*
- *How well do the portfolios perform out-of-sample, and how does the in-sample performance of different portfolios translate to out-of-sample performance?*
- *How well does the HRP perform on empirical data compared to non-hierarchical allocation strategies?*

Lastly, while López de Prado's (2016) HRP model is innovative and yields promising results, it also has its disadvantages, and several studies have pointed out some of its counterintuitive features (Raffinot, 2017, 2018; Lohre et al., 2020; Huang, 2020). Most noteworthy is that the hierarchical structure produced in the first stage is not considered when the bisection is performed in stage three, which induces some arbitrariness in the method. Hence, this thesis also contributes to the existing literature by connecting the first and third stages and alleviating other minor weaknesses of the original HRP approach by López de Prado (2016b).

1.2 Scope and Limitation

Since the HRP approach does not take the expected returns into account, the portfolios constructed using hierarchical clustering are compared to risk-based asset allocation strategies. The risk based allocation strategies used as a comparison throughout this thesis are limited to equally-weighted, minimum-variance, equal risk contribution, and inverse-variance portfolios. In addition, the hierarchical clustering methods used in this thesis are long-only by construction. Therefore, for the sake of comparability, the risk-based strategies are also limited to their long-only variants.

Arguably, other risk-based approaches, e.g., the most diversified portfolio or inverse-volatility, could have been included in the analysis. However, the four approaches mentioned above were deemed to be representative for the risk-based strategies. Also, limiting the number of strategies helps us retain the focus on the hierarchical clustering approaches and keeps the study concise. In addition, mean-variance optimizing portfolio have been excluded from the analysis. Mean-variance optimization has been shown to be especially difficult to implement due to challenges in estimating the expected returns. We have chosen to take the previous research regarding mean-variance optimization at face value, and therefore, we are focusing purely on risk-based asset allocation strategies that ignore the estimation of expected returns.

The investment universe used in the empirical analysis of this thesis consists of the 30 Industry Portfolios for the US stock market downloaded from French's Data Library, which includes Food, Beer, Smoke, Games, etc.[1] The dataset used ranges from January 2000 to February 2022. Therefore, it will include several stock market crashes, including the Dot-com bubble, the Financial Crisis, and the 2020 stock market crash caused by the Covid-19 pandemic. Although the analysis could be applied to more common indices such as the S&P500, Dow Jones Industrial Average, Russell 2000, MSCI World, etc., the 30 industry portfolios dataset is chosen to ease the data collection. Additionally, there is an illustrative advantage of using fewer assets, as it is easier to interpret, comprehend, and visualize.

Explicit transaction costs are disregarded since they are deemed unfeasible for the dataset used. However, the relative level of transaction costs required to implement different strategies is considered in the empirical analysis by measuring the average turnover. In addition, the risk-free rate is assumed to be zero throughout the study, and tax implications are not considered. Lastly, a common assumption in finance is that returns are assumed to be independent and identically distributed (iid) (Miller, 2018). This assumption allows for the famous square-root rule; the standard deviation increases with the square root of time (Miller, 2018). Thus, we impose the same assumption.

1.3 Structure

The remainder of the thesis is organized into eight sections, summarized below.

Section 1 – Literature Review

The first section presents the historical evolution of hierarchical clustering in asset allocation and outlines the literature that has built on López de Prado's (2016b) work.

Section 2 – Theory

The theoretical framework is presented in this section. First, Markowitz's seminal work on Modern Portfolio Theory is introduced, followed by an overview of different risk-based allocation strategies. Next, the issues caused by estimating the covariance matrix and its various underlying sources are presented. After that, a brief introduction to clustering theory is provided to give the necessary background to understand the HRP algorithm. Lastly, the HRP model by López de Prado (2016b) is introduced and explained.

Section 3 – Improving the HRP

Three of the most prominent weaknesses of the HRP strategy are presented. These include chaining, lack of a link between the first and third step of the HRP, and using an optimal number of clusters. Next, an alternative model that implements solutions to the three issues is presented. Lastly, the Elbow method is explained.

Section 4 – Monte Carlo Simulations

Two simulation studies are presented in this section. The first study analyzes and measures the impact of

[1] Website: mba.tuck.dartmouth.edu

covariance matrix instability on different portfolio optimization methods. The second study seeks to replicate López de Prado's (2016) findings on the out-of-sample performance of the various strategies while also providing insights into the in-sample performance.

Section 5 – Empirical Analysis

This section conducts an empirical comparison of the different portfolios using a walk-forward analysis. The analysis evaluates the performance of the different portfolios using a range of measures, including portfolio returns, weight concentrations, turnover, and several risk-based performance measures.

Section 6 – Discussion

The simulation studies and the empirical walk-forward analysis are discussed in this section. The section consists of four separate discussions, one for each analysis and a final discussion of the performance differences between the original HRP and the modified version introduced in section three.

Section 7 – Conclusion

This section rounds up the thesis and answers the thesis research question along with the sub-questions.

Section 8 – Further Research

The final section of this thesis covers some suggestions for further research on HRP and topics within hierarchical clustering.

2 Literature Review

This thesis builds on López de Prado's (2016b) work surrounding the hierarchical risk parity, which builds on previous work on the usage of hierarchal clustering in asset allocation. Thus, the following section presents the historical evolution of hierarchal clustering in asset allocation and outlines the literature that has built on López de Prado's (2016b) work.

2.1 The Origins of Hierarchical Clustering in Finance

The idea of grouping stocks in hierarchical ways can be dated back to the 1960s, when Benjamin F. King (1966) added a final footnote as advice for future research:

> It is believed that a convincing argument has been made for acceptance of the hypothesis that a small number of factors, market and industry, are sufficient to explain the essential comovement of a large group of stock prices; it is possible, however, that more satisfactory results could be obtained by methods that are distribution free. Here we are thinking of a factor-analytic analogue to median regression and non-parametric analysis of variance, where the measure of distance is something other than expected squared deviation. In future research we would probably seriously consider investing some time in the exploration of distribution free methods.

However, it was not until the end of the 1990s that hierarchical clustering started gaining traction in the field of financial research. In 1999, Mantegna published a seminal work on the hierarchical structure of financial markets that explored the topological arrangement of stocks in complex financial markets. The main finding of this study was that correlation-based clustering is able to give meaningful economic structure to a portfolio of stocks. Since this seminal study, many works have followed the research of clustering, correlations, hierarchies, and networks in financial markets.

Existing literature on clustering in portfolio management has considered clustering as a means of filtering and improving parameter estimation. For instance, Tola et al. (2008) argue that correlation-based clustering can be seen as a filtering procedure, where the correlation matrix is transformed into a new matrix containing a smaller number of distinct elements. Their study compared different filtering procedures, and the authors found that filtering based on clustering methods outperforms filtering methods relying on Random Matrix Theory (RMT). Similarly, Tumminello et al. (2010) contend that hierarchical clustering procedures are filtering procedures that act as an alternative to filtering procedures of the RMT and the shrinkage of a correlation matrix. The authors used 100 high-capitalization stocks of the New York Stock Exchange (NYSE) and their corresponding sector classifications and found that the hierarchical clustering based on the correlation matrix is able to detect clusters of stocks belonging to the same sectors or sub-sectors without the need for any supervision of the clustering procedure.

2.2 Literature on Hierarchical Risk Parity

In 2016 López de Prado explored a new approach to asset allocation called Hierarchical Risk Parity (HRP) which involved using machine learning and graph theory to infer a hierarchical relationship between assets. His approach uses the inferred hierarchical structure and inverse-variance allocation to allocate weight between clusters of correlated asset returns. Additionally, the HRP approach has the advantage that it avoids the inversion of the covariance matrix, which most traditional risk-based allocation techniques require and which often adds significant estimation errors. Using Monte Carlo experiments, López de Prado (2016b) was able to show that the HRP delivers less risky portfolios out-of-sample than traditional risk parity methods.

López de Prado (2016b) relied on simulated data to demonstrate the performance of the HRP. Since the seminal study, other authors have explored the performance of the HRP using empirical datasets. For example, Lau et al. (2017) applied the HRP to two different cross-asset universes: alternative risk premia and traditional market indices. Their results showed that the HRP is superior to most traditional risk-based portfolio construction methods in terms of risk-adjusted returns. Although the minimum-variance portfolio had a slightly higher risk-adjusted returns in one of the risk premia indices, the HRP portfolio was more diversified. Furthermore, their out-of-sample simulations demonstrated that the HRP produces lower out-of-sample volatilities and drawdowns than the minimum-variance portfolio. Moreover, Burggraf (2020) analyses the HRP in a large portfolio of cryptocurrencies. The out-of-sample simulations of their study reveal that the HRP outperforms traditional risk-minimization methods in terms of tail risk-adjusted return. Furthermore, other authors have used a variety of different data sets, e.g., individual stocks (Raffinot, 2017, 2018), different equity indices (Pfitzinger and Katzke, 2019; Jothimani and Bener, 2019; Huang, 2020), commodities ((Alipour et al., 2016; Molyboga, 2020), and multi-asset multi-factor investment universes (Lohre et al. (2020)Lohre et al., 2020; Kolrep et al., 2020).

Many authors have pointed out that the original HRP approach by López de Prado (2016b) does not fully exploit the information produced by the clustering process. Essentially, in the HRP approach, clustering is only used to reorder the assets, and the shape of the dendrogram is not considered during the bisection. As such, the bisection stage might separate highly correlated assets. Thus, instead of bisecting the tree, Raffinot (2017) introduced an approach that splits the allocations based on the tree structure, or the shape of the dendrogram, that represents the correlation clusters. This tree-based sectioning is also applied by Lohre et al. (2020) and Huang (2020). In a similar vein, Pfitzinger and Katzke (2019) extended the bisectioning of the HRP with an algorithm that takes the asset similarity into account. This algorithm considers a trade-off between symmetry and minimum dissimilarity, which is represented by a threshold τ. On the one end of the spectrum ($\tau = 1$), the allocation is performed precisely along the dendrogram, while on the other end ($\tau = 0$), the allocation is carried out using the naive bisection solely as in the original HRP. Overall, the conclusions of these studies indicate that the sectioning-based allocation seems to be superior to the naïve bisectioning of the HRP and produces more intuitive and stable allocations.

Some authors have extended the HRP approach by using cleaned/filtered covariance or correlation matrices. Molyboga (2020) applies a refined approach to HRP, named the modified HRP (MHRP). Firstly, his modified

approach substitutes the sample covariance matrix with a Ledoit-Wolf shrank exponentially weighted covariance matrix. Secondly, Molyboga (2020) incorporates an inverse-volatility allocation instead of an inverse-variance approach to enhance diversification. Lastly, he applies volatility targeting to improve diversification across time. Molyboga then conducts a large-scale Monte Carlo simulation on Barclays commodity trading advisor (CTA) portfolios over the 2002-2016 period. His findings show that the MHRP improves the out-of-sample Sharpe ratios by 50% while also reducing downside risk. Similarly, Jothimani and Bener (2019) replaced the historical correlation of the HRP with the Gerber statistic, a robust co-movement statistic for covariance matrix estimation. This hierarchical risk parity model based on Gerber statistic (HRP-GS) outperforms the original HRP based on the risk-return profile and modified Sharpe ratio.

Other modifications of the HRP have included using alternative distance and codependence metrics. For example, Barziy and Chlebus (2020) assessed the performance of the HRP subject to several codependence and distance metrics. In addition to the Pearson correlation coefficient, the alternative codependence metrics included; distance correlation, mutual information, and variation of information. They transform the codependence metrics to distance matrices using angular, absolute angular, and squared angular distance metrics. However, they find that the Pearson correlation codependence metric of the original HRP algorithm outperforms all the other codependence metrics in every scenario, raising doubts about the effectiveness of using alternative metrics. On the other hand, Lohre et al. (2020) compared different distance measures for the clustering process and found contradicting results. In addition to the standard correlation-based clustering of the original HRP, the authors evaluated a distance measure based on the lower tail dependence to include the notion of tail risk management. Assessing the performance of these two distance metrics, they found that strategies based on both distance metrics can compete with traditional risk parity strategies. Moreover, they discovered that the hierarchical risk parity strategy based on the lower tail dependence seems to be especially beneficial for tail risk management.

To summarize, studies have shown that the HRP delivers high risk-adjusted returns and comparatively diversified portfolios. In addition, a plethora of different modifications have been applied to the original HRP. Numerous alterations can be incorporated into the HRP concerning codependence and distance metrics, estimating the covariance or correlation matrix, different portfolio construction methods, and others. However, Schwendner et al. (2021) note a trade-off between precision and over-analyzing. Introducing more sophisticated modeling elicits different sets of assumptions and could potentially lead to an over-rigorous balancing of precision, bias, and time lag. Therefore, one needs to consider this trade-off carefully before including more intricacies in the model.

3 Theory

This chapter presents the theory underlying the thesis. It starts with an introduction to covariance, correlation, and eigenvalues. Then, Harry Markowitz's seminal work on Modern Portfolio Theory is presented, followed by an overview of different risk-based allocation strategies. Afterward, the issues caused by the estimation of the covariance matrix will be presented and discussed. Thereafter, clustering theory is presented to provide the necessary background to understand the HRP algorithm. Lastly, the hierarchical risk parity model from López de Prado's (2016) paper "Building Diversified Portfolios that Outperform Out-of-Sample" is introduced and explained.

3.1 Covariance, Correlation, and Eigenvalues

The terms covariance matrix, correlation matrix, and eigenvalue are used extensively throughout the thesis. Thus, brief explanations and definitions are provided in the following section.

Covariance

The covariance, σ_{XY}, of two random variables X and Y is defined as:

$$\sigma_{XY} = E[(X - E[X])(Y - E[Y])]$$

Assuming that the means are unknown, the sample covariance is estimated by:

$$\hat{\sigma}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

where $\bar{X}$ and $\bar{Y}$ represents the sample mean of X and Y. In the case where $X = Y$ the sample covariance equals the sample variance of the X (or Y), that is:

$$\hat{\sigma}_{XY} = \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The variance of a portfolio, σ_{PF}^2, with the $n \times 1$ vector of weights, $\mathbf{w}$, is defined as:

$$\sigma_{PF}^2 = \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$$

where $\mathbf{\Sigma}$ is the population covariance matrix, a square matrix that gives the covariance between each pair of elements of a given random vector. If we consider a N dimensional random vector, $\mathbf{X}$, then the covariance matrix is defined as:

$$\mathbf{\Sigma} = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\top\right]$$

However, the covariance matrix is generally unobserved and the sample covariance matrix is arguably the most commonly used estimator of the covariance matrix, defined as:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T$$

Correlation

Correlation is closely related to the concept of covariance. The correlation between two two variables X and Y, $\rho_{X,Y}$, is defined as:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Correlation can be thought of as a normalized covariance as it transforms covariance to a $[-1, 1]$ range with unified units that do not depend on the scaling of X and Y. Equivalently, a correlation matrix can be viewed as the covariance matrix of the standardized random variables (Miller, 2018). The correlation matrix is generally also unobserved and requires an estimator. The sample correlation matrix $\hat{\mathbf{C}}$ can be constructed by centering and normalizing the sample covariance matrix, that is:

$$\hat{\mathbf{C}} = diag(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}}\ \hat{\mathbf{\Sigma}}\ diag(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}}$$

where

$$diag(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\hat{\sigma}_{1,1}}} & & & \\ & \frac{1}{\sqrt{\hat{\sigma}_{2,2}}} & & \\ & & \ddots & \\ & & & \frac{1}{\sqrt{\hat{\sigma}_{n,n}}} \end{bmatrix}$$

Eigenvalues

Eigenvalues are necessary for discussing the numerical condition of the sample correlation and covariance matrix. Thus, a definition of eigenvalues and some of their properties will be presented, but their mathematical derivation is left out as it is not essential in this thesis. If we consider an $n \times n$ matrix $\mathbf{A}$ and a scalar λ, then λ is an eigenvalue of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v}$ in R^n such that:

$$\mathbf{Av} = \lambda\mathbf{v}$$

Additionally, the following properties of eigenvalues will be used later in the thesis. Let $\mathbf{A}$ be a $n \times n$ matrix:

- The matrix $\mathbf{A}$ has n eigenvalues counting multiplicity.
- The sum of the n eigenvalues of $\mathbf{A}$ is the same as the trace of of $\mathbf{A}$, that is, the sum of the diagonal elements of $\mathbf{A}$:

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

- The product of the n eigenvalues of A is the same as the determinant of A:

$$det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_n$$

3.2 Modern Portfolio Theory

Harry Markowitz conducted his first stream of work on Portfolio Theory in the 1950s. His work continues to shape how investors and managers approach diversification and portfolio construction and is commonly referred to as Modern Portfolio Theory (MPT). Markowitz (1952) abandoned the general idea that investment portfolios should solely aim at maximizing the expected return. Focusing only on return neglects the risk associated with different assets, which may have the same returns but dissimilar risks. Investors should instead construct portfolios by accounting for both risks and returns, a relationship that Markowitz (1952) labels the "expected return – variance of return rule (E-V rule)". Markowitz (1959) also pointed out that assets should not be viewed in isolation but rather on how they behave in groups (portfolios). When assets are combined correctly, they can diversify away part of the risks and give higher expected returns without increasing the risk.

Markowitz (1952) proposed that an efficient portfolio maximizes the expected return while minimizing the variance. In his approach, the investor creates an *efficient frontier* consisting of multiple portfolios, where the expected return is maximized at different risk levels. Figure 1 is a graphical representation of Markowitz's efficient frontier.

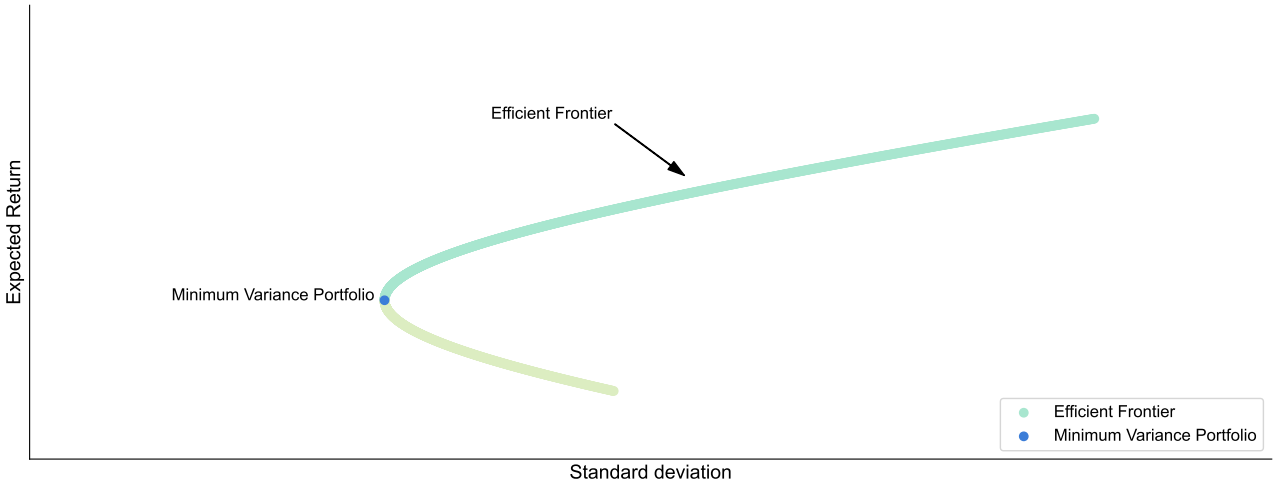

Figure 1: Illustration of the efficient frontier

After the efficient frontier is constructed, the investor has to decide which efficient portfolios she prefers. According to Markowitz (1959), the investor's degree of risk aversion will determine the optimal portfolio. Figure 2 illustrates the efficient frontier with three investors' risk-return indifference curves (C1, C2, C3).

A risk-averse investor, C3, has a steeper indifference curve than the less risk-averse investors. Thus, C3 requires a significant increase in expected return to compensate for a slight increase in standard deviation. Conversely, if

the investor is less risk-averse, C2, she requires less compensation for the increase in standard deviation. Thus, the risk-averse investor chooses a portfolio closer to the left, whereas a risk-taker chooses a portfolio higher up the frontier.

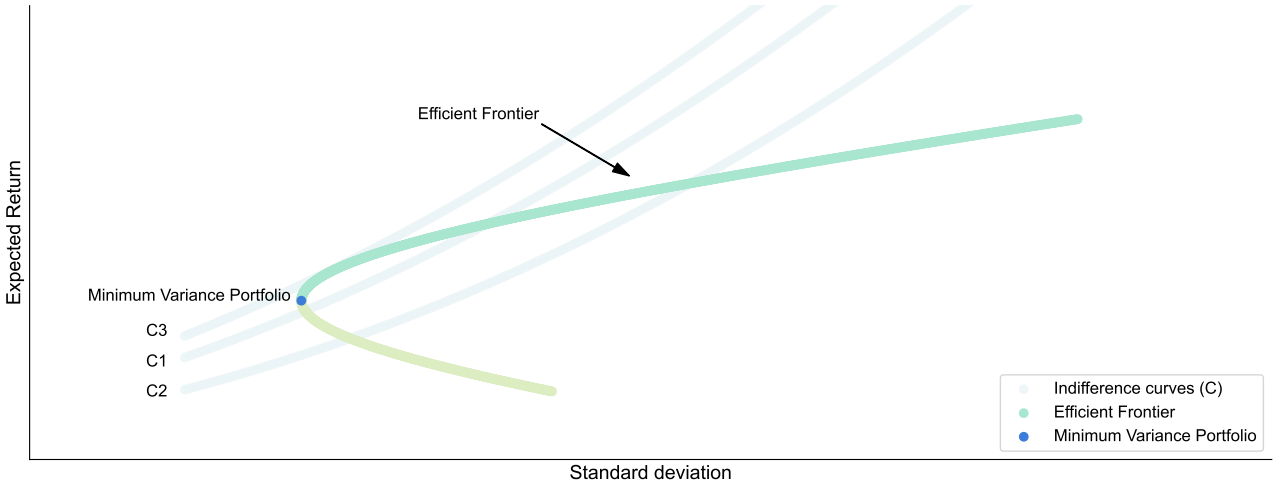

Figure 2: Illustration of the efficient frontier with investor risk preferences

Markowitz's (1952) approach is essentially a convex optimization problem based on the assumption that the covariance matrix is positive definite so that the variance can be calculated. The optimization problem is typically subject to inequality conditions (an upper and lower bound for the asset weights in the portfolio) and an equality condition (that the weights add up to one) (Bailey and de Prado, 2013). For example, as mentioned in the limitation, this thesis does not allow short sales and consequently only consider long-only solutions. Mathematically, if $\mathbf{w}$ represents the vector of portfolio weights, $\boldsymbol{\mu}$ the vector of expected return, and $\boldsymbol{\Sigma}$ is the covariance matrix, then Markowitz (1952) seeks to maximize the risk-return tradeoff, $\mathbf{w}^\top \boldsymbol{\mu} - \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$, under the constraints that the weights must sum to one, $\mathbf{w}^\top \mathbf{1} = 1$, and no shorting constraint, $w_i \geq 0$, as shown in Equation (1).

$$\begin{aligned} &\max_{\mathbf{w}} \ \mathbf{w}^\top \boldsymbol{\mu} - \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ &\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1 \\ &w_i \geq 0 \end{aligned} \tag{1}$$

where $\mathbf{w}^\top \boldsymbol{\mu}$ represents the portfolio's expected return and $\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$ its variance. To account for risk-aversion, c, the objective function can be expressed as:

$$\max_{\mathbf{w}} \ \mathbf{w}^\top \boldsymbol{\mu} - \frac{c}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$$

These problems have no analytic solution, and an optimization algorithm must be used. Markowitz (1987) developed an optimization algorithm for computing such a solution, which he named the "critical line algorithm" or CLA, which is explained in detail in Appendix A. However, this thesis will not focus on Markowitz's solution

since the Scipy library offers an optimization module called `optimize`, which is computationally efficient while still reaching an optimum.

3.2.1 Limitations of Markowitz's Model

In theory, Markowitz's model provides investors with a straightforward and intuitive framework to face the portfolio construction problem. However, it only applies if the risk-return estimates are incredibly accurate. In practice, the true input values for the optimization are not observable ex-ante (Braga, 2016). Consequently, Markowitz's method relies on estimates, which are at best an informed guess of the true values in the investment period, which is problematic as Markotwitz's approach is extremely sensitive to estimation errors (Michaud and Michaud, 2008). Different methods have been developed to improve portfolio optimization in response to the estimation risks of Markowitz's approach. There are two significant ways to deal with estimation risk: decrease the need for estimations or improve the estimates.

Markowitz's model takes two input parameters; the expected return and the covariance matrix. A vast literature exists on the robustifying estimation of both parameters. Arguably, the most famous model for estimating the expected returns is the Capital Asset Pricing Model (CAPM), developed by William Sharpe (1964). The CAPM suggests that there will be a consistent relationship between assets' expected returns and their systematic risk. The expected return of an asset i, $E[r_i]$, can be expressed through the CAPM as:

$$E[r_i] = r_f + \beta_i \cdot E[E(r_m) - r_f]$$

where r_f is the risk-free rate , β_i, is the beta of security i and $E[E(r_m) - r_f]$ is the expected market risk premium. The CAPM is a single factor model and has been extended to multifactor models, where other firm characteristics are used to explain the drivers behind the expected return. Similarly, several models for estimating or improving the covariance matrix estimator have been developed. As mentioned in the introduction, the most famous solution to a more robust covariance estimator is the shrinkage transformation proposed by Ledoit and Wolf (2003, 2004a,b). The shrinkage transformation pulls the most extreme coefficients towards more central values, thereby systematically reducing estimation error where it matters most.

While both the expected return and the covariance matrix are used in Markowitz's model, a part of the literature argues that error in the expected return vector is the biggest problem (Kallberg and Ziemba, 1984; Chopra, 1993; Merton, 1980; Jagannathan and Ma, 2003). For example, Kallberg and Ziemba (1984) found that estimation errors in the expected returns are about ten times as important as estimation errors in the covariance matrix. Thus, instead of trying to improve the expected return estimates, alternative models that do not rely on estimating expected returns could be favorable.

3.3 Risk-Based Asset Allocation Strategies

The global financial crisis, which started in 2008, pawed the way for new strategies in the field of asset management. The crisis resulted in a drop of around -50% in equities (Amadeo, 2012), and while many institutional investors had well-diversified portfolios, this diversification was not enough to protect them from the crisis. Correlations between asset classes increased considerably, and the promised diversification of portfolio optimization did not materialize. Since diversification was traditionally associated with Markowitz's model, this prompted a wave of criticism towards the Modern Portfolio Theory (Roncalli, 2013). In light of this criticism of Markowitz's model, several strategies that reduced the reliance on expected returns and focused on risk diversification emerged.

Multiple risk-based allocation strategies have been developed through extensive research. Arguably the most famous risk-based model is the minimum variance portfolio, which is an extension of Markowitz's framework. Alternatively, the equally weighted portfolio thus does not rely on any metrics and does not require any optimization. Lastly, there exists a branch of allocation strategies that solely relies on the estimated covariance matrix. The following sections will briefly introduce and explain the idea behind the following risk-based allocation strategies: Equal Weighting, Minimum-Variance, Equal Risk Contribution, and Inverse-Variance.

3.3.1 Equally-Weighted Portfolio

The equally weighted portfolio, also known as the 1/N portfolio or the naive portfolio, is the simplest and most straightforward allocation strategy. In an equally-weighted portfolio, all assets are given the same weight, and therefore, weights, w_i^{EW}, are calculated as

$$w_i^{EW} = \frac{1}{N} \qquad \text{for all } i = 1, ..., N \tag{2}$$

where N is the number of assets. The advantage of this strategy is that it can be determined without any estimations of returns, risk, or correlations (Braga, 2016). However, the assumption that this heuristic and straightforward method guarantees diversification can be misleading. If the individual risks of the portfolio constituents are significantly different, then equal weighting can lead to limited diversification of risks manifested in unintended risk concentrations to a limited number of high-volatility assets (Maillard et al., 2010).

Despite the simplicity, equally-weighted portfolios have been shown to outperform more sophisticated models out-of-sample. For example, DeMiguel et al. (2009) evaluated the out-of-sample performance of the equally weighted portfolio relative to 14 sophisticated extensions of Markowitz's mean-variance portfolio. Their findings show that none of the more sophisticated models consistently outperformed the equally-weighted strategy out-of-sample. With similar conclusions, Duchin and Levy (2009) found that due to estimation errors in Markowitz's mean-variance framework, the equally-weighted portfolio outperformed mean-variance portfolios out-of-sample for relatively small portfolios.

3.3.2 Minimum-Variance Portfolio

The minimum-variance (MV) portfolio is the portfolio that has the lowest variance of all portfolios. This portfolio is often called the *global minimum-variance portfolio* to emphasize that it is the portfolio with the lowest variance (ex-ante) among the universe of risky assets, disregarding any constraints on expected return (Munk, 2018). This portfolio plays a unique role in the efficient frontier. The portfolio can be determined solely based on a covariance matrix, and thus, it is the only portfolio on the efficient frontier that does not require expected returns as input (Lee, 2011).

Mathematically, the minimization problem can be formulated as follows

$$\begin{aligned} &\min_{\mathbf{w}} \ \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \\ &\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1 \end{aligned} \tag{3}$$

where $\mathbf{w}$ represents an $n \times 1$ weight vector, $\mathbf{\Sigma}$ represents the $n \times n$ covariance matrix, and $\mathbf{1}$ represents an all-ones vector of n elements. This constrained minimization problem can be solved using the Lagrangian method[2] and leads to the following analytical solution for the global minimum-variance portfolio:

$$\mathbf{w}_{GMVP} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}} \tag{4}$$

However, the closed-form solution defined above is only available when the global minimum-variance portfolio is subject to only one constraint: fully invested. Furthermore, in this thesis, a constraint on short selling, $w_i \geq 0$, is necessary to compare the different portfolios fairly. In this case, the optimization problem takes the following form (Braga, 2016):

$$\begin{aligned} \mathbf{w}_{MV} &= \underset{w}{\operatorname{argmin}} \ \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^\top \mathbf{1} = 1 \\ & w_i \geq 0 \end{aligned} \tag{5}$$

The inequality constraint prevents us from obtaining an analytical solution, and hence, the problem becomes a quadratic programming (QP) problem that needs to be solved numerically (Braga, 2016).

3.3.3 Risk Parity

Risk parity takes a risk-based approach to constructing portfolios, where the weights of the assets are determined by their risk contribution. Risk parity caught the attention of many investors after the recent financial crisis but was first introduced to the investment world by Bridgewater in 1996 and coined by Edward Qian in 2005 (Roncalli, 2013). Qian (2005) argued that risk is the diversification factor, meaning that if a portfolio can take a substantial loss due to a few assets, then it is not diversified.

[2]For a detailed derivation see, for example, Munk (2018).

Roncalli (2013) points out that the term *risk parity* is used differently across the academic literature and the financial industry. While some professionals use it to describe a weighting method where asset weights are inversely proportional to their volatility, others consider the risk parity portfolio equivalent to the equal risk contribution (ERC) portfolio. Conversely, risk parity is also regarded as equivalent to a risk budgeting (or RB) portfolio, where the risk budgets do not need to be equal (Roncalli, 2013). Following Roncalli (2013), this thesis considers risk parity a synonym for risk budgeting. If the risk budgets are equal, this corresponds to the ERC portfolio, which is more explicit and less overused by the investment industry. The following three sections will cover risk budgeting, ERC, and the inverse-variance portfolio.

3.3.4 Risk Budgeting

The fundamental idea of risk budgeting portfolio is to set the risk contribution from each component in the portfolio equal to the risk budget defined by the portfolio manager (Bruder and Roncalli, 2012). By defining w_i as the weight of the i^{th} asset, then the risk measure for the portfolio, $R(\mathbf{w})$, can be written as (Roncalli, 2013):

$$R(\mathbf{w}) = \sum_{i=1}^{n} w_i \cdot \frac{\partial R(\mathbf{w})}{\partial w_i} = \sum_{i=1}^{n} RC_i \tag{6}$$

where RC_i is the risk contribution of asset i.

3.3.5 Equal Risk Contribution Portfolio

The goal of ERC is to construct portfolios where every asset has the same risk contribution towards the overall portfolio risk (Roncalli, 2013). Conceptually, risk parity is similar to the equally weighted portfolio, but instead of having a portfolio perfectly diversified from a capital perspective, risk parity seeks to be perfectly diversified from a risk perspective.

In practice, this can be achieved by solving a system of equations where the risk contribution for all assets is set equal. Following Maillard et al. (2010), if there are short selling restrictions ($\mathbf{w} \in [0, 1]$), budget constraints (b), and a restriction on weights summing to one ($\mathbf{w}^\top \mathbf{1} = 1$), then the problem can be formulated as:

$$\mathbf{w} = \{\mathbf{w} \in [0, 1] : \mathbf{w}^\top \mathbf{1} = 1, w_i \cdot (\mathbf{\Sigma w})_i = b_i \cdot (\mathbf{w}^\top \mathbf{\Sigma w})\}. \tag{7}$$

Note that the restriction of weights summing to one is optional, but it works as a normalizing restriction, making ERC more comparable to other strategies (Maillard et al., 2010). With the following constraints, finding a closed-form solution for the ERC portfolio weights is not possible, as w_i is a function of the risk contributions which, by definition, depends on w_i. Therefore, numerical optimization is necessary. Maillard et al. (2010) writes the optimization problem as:

$$\begin{aligned} &\mathbf{w}_{ERC} = \text{argmin } f(\mathbf{w}) \\ &\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1 \\ &w_i \geq 0 \end{aligned} \tag{8}$$

where the objective function $f(\mathbf{w})$ is to minimize the variance of the (rescaled) risk contributions, given by the formula:

$$f(\mathbf{w}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_i(\mathbf{\Sigma w})_i - w_j(\mathbf{\Sigma w})_j \right)^2 \tag{9}$$

Only in the hypothetical case where all assets have the same correlations is it possible to carry out explicit solutions (Maillard et al., 2010). Under this assumption and using volatility as the portfolio risk measure, the ERC portfolio is equivalent to the inverse-volatility portfolio. If we assume that, in addition to uniform correlations, the assets have all the same Sharpe ratio; the ERC portfolio corresponds to the Maximum Sharpe Ratio or the tangency portfolio (Maillard et al., 2010).

3.3.6 Inverse-Variance Portfolio

Inverse-variance states that the portfolio weight of each asset is proportional to the inverse of its variance, namely

$$w_i^{IVA} = \frac{1/\sigma_i^2}{\sum_j (1/\sigma_j^2)} \text{ with } \sigma_i^2 = \Sigma_{ii} \tag{10}$$

Inverse-variance corresponds to minimum-variance when the covariance matrix is diagonal, that is, when all cross-correlations are assumed to be zero. With an additional assumption of equal expected returns, IVA becomes mean-variance optimal.

3.4 Covariance Instability

Most risk-based asset allocation models take the sample covariance matrix as the input parameter. However, the sample covariance matrix estimator often induces estimation error. Thus, the following section highlights the issues caused by the estimation of the covariance matrix. Highlighting which sources of instability the covariance matrix estimator brings to portfolio optimization.

The general issue with all covariance matrix estimators is that the true covariance matrix is unobserved. Consequently, the most common statistical method uses historical returns and computes the sample covariance matrix. However, the performance of the sample covariance matrix is questionable, especially when there are few samples compared to the dimension. In fact, in Ledoit and Wolf's (2004) paper *Honey, I Shrunk the Sample Covariance Matrix*, the authors explicitly state that the central message of their paper is that nobody should be using the sample covariance matrix for the purpose of portfolio optimization. The core problem is that the sampled covariance matrix tends to have extreme values due to an extreme amount of error, which the mean-variance optimization will latch onto, thus placing large bets on the most unreliable coefficients (Ledoit and Wolf, 2004b). This phenomenon is known as "error-maximization" (Michaud, 1989).

A three-asset simulation experiment illustrates the errors that arise when computing and using the sample covariance matrix. By defining the expected returns, volatilities, and correlation matrix, one can compute the true covariance matrix. The defined parameters are then used as inputs for the simulations that draw random samples from a multivariate normal distribution. Then by computing the sample covariance matrix from 100 observations and performing minimum-variance optimization 1000 times, the sensitivity of portfolio weights towards an ill-conditioned covariance matrix can be illustrated in a boxplot, as shown in Figure 3. The figure shows large undesired intervals for the estimated portfolio weights, highlighting the minimum-variance method's inaccuracy and sensitivity toward the sample covariance matrix estimator.

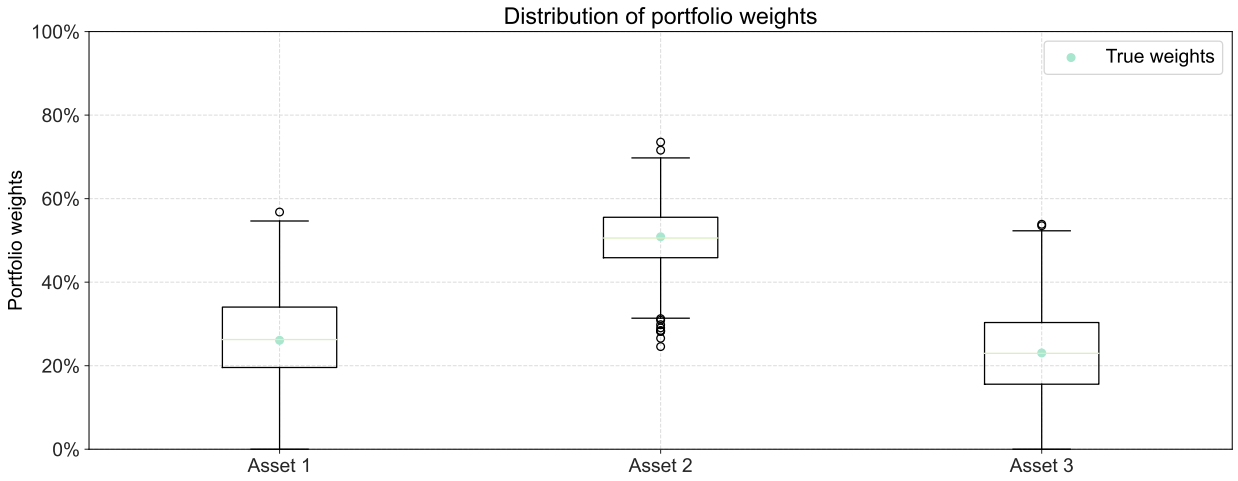

Figure 3: Boxplot illustration of the distribution of portfolio weights. The distribution is found by computing the sample covariance matrix of 100 observations drawn from a multivariate normal distribution and calculating the minimum-variance optimal portfolio weights 1000 times.

Many of the traditional portfolio optimization models require the inversion of a positive-definite covariance matrix. According to López de Prado (2016b), at least $\frac{1}{2}N(N-1)$ iid observations are required to estimate an invertible sample covariance matrix. For example, this would mean 1,275 observations or around five years of daily iid data for a covariance matrix of size 50, and it is unrealistic to assume that the correlation structures would remain invariant over five years Zakamulin (2015). Thus, estimating an invertible covariance matrix requires too many observations for a large portfolio, which leads investors to either rely on small portfolios, transform the portfolio, or use allocation strategies that do not require the covariance's inversion matrix. However, even if one could estimate a non-singular sample covariance matrix, the inversion of a sample covariance matrix is prone to significant errors when the matrix is numerically unstable (López de Prado, 2016b).

The numerical instability of the covariance matrix can be measured by the discrepancy between the maximum and minimum eigenvalues. The discrepancy, known as the *condition number*, is precisely defined as the absolute value of the ratio between its maximal and minimal (by moduli) eigenvalues (López de Prado, 2016b). It is desirable to have a small condition number since a large condition number would indicate that the covariance matrix is nearly singular. However, it is common for financial sample covariance matrices to exhibit large condition numbers, and the inverse of those covariance matrices leads to unstable solutions in asset allocation strategies López de Prado (2020). The instability associated with portfolio optimization techniques is driven primarily by:

- *Noise* in the input variables
- *Signal* structure that magnifies the estimation errors in the input variables

However, most literature focus on discussing noise-induced instability, whereas signal-induced instability is often ignored or misunderstood (López de Prado, 2020).

3.4.1 Instability Caused by Noise

The law of large numbers implies that in the classical statistical limit, i.e., when the number of observations $T \to \infty$, with a fixed number of variables, N, the covariance estimator $\hat{\mathbf{\Sigma}}$ converges to the true population covariance $\mathbf{\Sigma}$. Analogous to this, Anderson (1963) showed that with sample size approaching infinity and a fixed number of variables, the sample eigenvalues of a correlation matrix converge to the population eigenvalues. However, if the number of variables is not fixed, and the ratio $q = N/T$ (the ratio between number of variables and number of observations)[3] tends toward 1, the sample eigenvalues become noisy estimators of the population eigenvalues, no matter how large T is (Bun et al., 2017). More concretely, when $T \gg N$, each individual coefficient of the covariance matrix $\mathbf{C}$ can be estimated with negligible error. But if N is also large and the ratio q is close to 1, as it is often the case in financial applications, the sample estimator $\hat{\mathbf{C}}$ becomes very noisy. This phenomenon is also known as the *curse of dimensionality* (Bun et al., 2017).

To capture the effect of the curse of dimensionality, two Ukrainian mathematics Vladimir Marčenko and Leonid Pastur (1967) came up with a distribution to capture the increasing dimensionality, known as the *Marčenko-*

[3]Defining q as the ratio between N and T, the findings of Anderson (1963) described a situation where $q = 0$.

Pastur distribution. The Marčenko-Pastur distribution defines the eigenvalue spectrum in the large dimension limit (LDL), that is, in the case when both N and T become very large but their ratio $q = N/T$ is not infitesimal Bun et al. (2017). Let us consider a $T \times N$ matrix X of independent and identically distributed random observations, where the underlying process generating the observations has zero mean and variance σ^2. The eigenvalues λ of the real covariance matrix $\mathbf{C} = T^{-1}\mathbf{X}^\top\mathbf{X}$ asymptotically converge (as $N \to +\infty$ and $T \to +\infty$ with $1 < T/N < +\infty$) to the Marčenko-Pastur probability density function (PDF) (López de Prado, 2020):

$$f_\lambda(\lambda) = \begin{cases} \frac{T}{N}\frac{\sqrt{(\lambda_{max}-\lambda)(\lambda-\lambda_{min})}}{2\pi\lambda\sigma^2} & \text{if } \lambda \in [\lambda_{min}, \lambda_{max}] \\ 0 & \text{if } \lambda \notin [\lambda_{min}, \lambda_{max}] \end{cases}$$

where

$$\lambda_{max} = \sigma^2(1 + \sqrt{N/T})^2$$
$$\lambda_{min} = \sigma^2(1 - \sqrt{N/T})^2$$

This theoretical distribution allows us to infer the relationship between the empirical correlation matrix $\hat{\mathbf{C}}$ and the real correlation matrix $\mathbf{C}$ in terms of their eigenvalues and elucidates why empirical correlation matrices contain substantial amounts of noise. The deviations from the purely random matrix $\lambda \notin [\lambda_{min}, \lambda_{max}]$ are the ones containing the real information or *signal*, while $\lambda \in [\lambda_{min}, \lambda_{max}]$ are consistent with random behavior or *noise* (Laloux et al., 2000). Although returns from real financial data are not independent and the distribution of eigenvalues does not closely fit the Marčenko-Pastur, the distribution helps us understand what random noise in eigenvalues looks like (Pedersen et al., 2021).

Figure 4 illustrates the distribution of the Marčenko-Pastur PDF of three different ratios; $q = 0.5$, $q = 0.25$, and $q = 0.1$. From the figure we see that although the "true" correlation matrix $\mathbf{C}$ is an identity matrix, its empirical determination $\hat{\mathbf{C}}$ from finite number of observations generates non-trivial eigenvalues (Laloux et al., 2000). Furthermore, the dispersion in the eigenvalue spectrum of $\hat{\mathbf{C}}$ is an increasing function of the ratio q. As q increases, the lower bound of the Marčenko-Pastur distribution, λ_{min}, gets smaller, while the upper bound, λ_{max}, increases. Fundamentally, this means that instability associated with *noise* is regulated by the ratio $q = N/T$ (López de Prado, 2020).

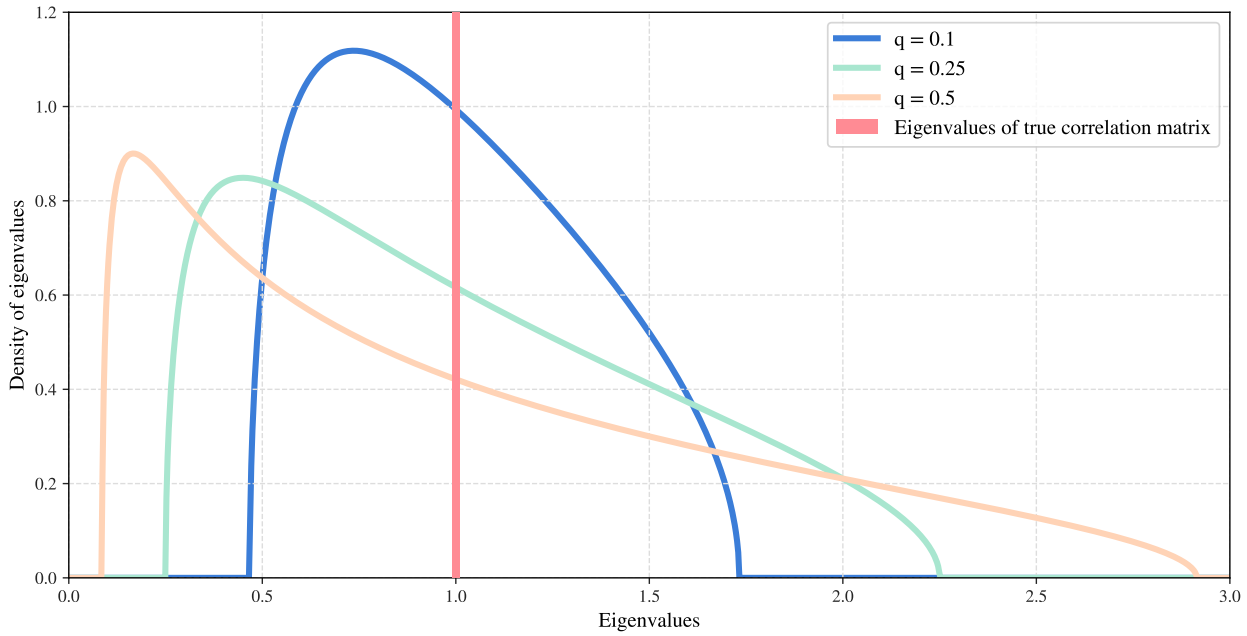

Figure 4: Marčenko-Pastur Distributions with $N = 500$ for three different ratios: $q = 0.1$, $q = 0.25$, and $q = 0.5$. The red line corresponds to a perfect estimation of the population eigenvalues.

A brief look into matrix inversion helps us understand the effect of noise on the stability of the covariance matrix. The inverse of an invertible matrix A is defined as

$$\mathbf{A}^{-1} = \frac{1}{det(\mathbf{A})} adj(\mathbf{A}) \tag{11}$$

where $det(\cdot)$ is the determinant and $adj(\cdot)$ is the adjugate matrix. As the ratio q approaches 1, the lower bound of the eigenvalues approaches 0. Since the sample covariance matrix is an $N \times N$ matrix, the determinant of the matrix is the product of the N eigenvalues. Hence, as the bulk of eigenvalues approach zero, the determinant of the sample covariance matrix also approaches zero. This has the effect that the inverse of the estimated covariance matrix, $\tilde{\mathbf{\Sigma}}^{-1}$, cannot be estimated robustly, and the weights in the optimization problems requiring an inversion of the covariance matrix become highly unstable (López de Prado, 2020).

3.4.2 Instability Caused by Signal

Signal is the structure of the data, and signal-induced instability is often overlooked because it goes against our intuition – signal is precisely what we are trying to extract (López de Prado, 2020). However, signal can cause issues, and since it is structural, the signal-induced instability cannot be reduced by sampling more observations (López de Prado, 2020). Signal-induced instability occurs due to the clustering behavior of highly-correlated variables within the sample covariance matrix. If a subset of assets exhibits greater correlation among themselves than to the rest of the assets within the correlation matrix, the subset forms a cluster within the correlation matrix. These clusters can end up affecting the entire sample covariance matrix. The higher the intra-cluster

correlation, the higher the condition number (López de Prado, 2020). As a result, the signal-induced instability of a covariance matrix stems from a few dominant clusters within the correlation matrix. Thus, the signal-induced instability should be avoided by utilizing clustering methods since these methods do not rely on the inversion of the covariance matrix.

As already alluded to in the noise section, empirical eigenvalues do not closely fit the Marčenko-Pastur distribution, and a visual illustration of the discrepancies helps us understand the signal. Thus, Figure 5 depicts the empirical eigenvalue density of 406 stocks of the S&P 500 between 1991 and 1996. The inset of the plot shows that the largest eigenvalue, corresponding to the market factor, is 25 times higher than the highest eigenvalue of the Marčenko-Pastur distribution. The abnormally large eigenvalue results from the fact that stocks tend to move up or down together (Almog and Shmueli, 2019). In effect, this large eigenvalue will push the other eigenvalues towards zero. As explained in Section 3.1, the trace of an $n \times n$ matrix, that is, the sum of its n diagonal elements, corresponds to the sum of its n eigenvalues (counting multiplicity). Because the trace of the correlation matrix is precisely n, the eigenvalue of the market factor can only increase at the expense of the other eigenvalues (López de Prado, 2016a). Hence, analogous to the noise-induced instability, a high signal will push the bulk of the eigenvalues toward zero and lead to unstable solutions in optimization problems requiring an inversion of the covariance matrix.

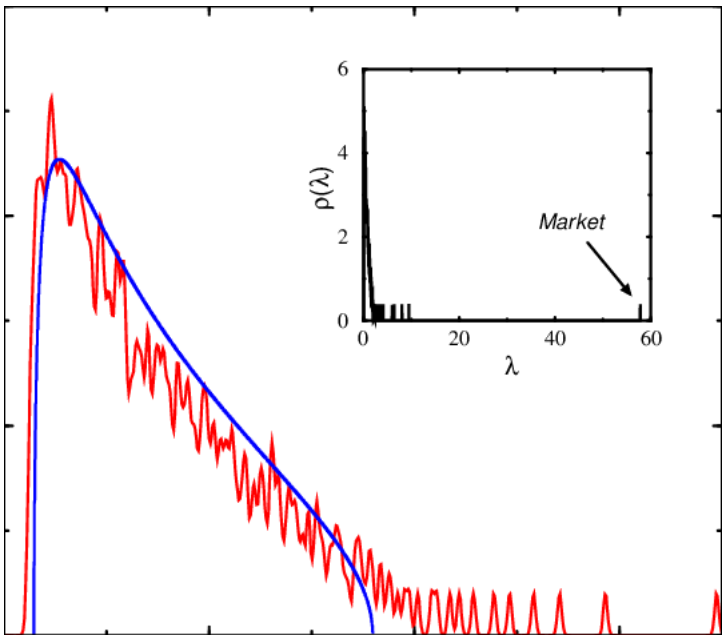

Figure 5: Empirical eigenvalue density of the S&P 500 during the years 1991-1996 with $N = 406$ and $T = 1300$ (red line). The empirical density is compared to the theoretical Marčenko-Pastur Distribution with the same ratio $q = N/T$ (blue line). The inset shows the presence of one very large eigenvalue, corresponding to the market mode (factor). Source: Potters et al. (2005)

3.5 From Geometric to Hierarchical Relationships

The problems regarding the instability have received a lot of attention in the past 20 years López de Prado (2016b); Kolm et al. (2014); Guigues (2011). López de Prado (2016b) postulates that a correlation matrix is too complex to be thoroughly analyzed and understood. As already alluded in the previous section, there are $\frac{1}{2}N(N-1)$ pairwise correlations for a portfolio of N assets. For example, there are 1,275 correlation coefficients between the constituents of a matrix of size 50, and the number proliferates as more constituents are added. Moreover, López de Prado (2016b) contends that the instability of the quadratic optimizers is caused by the absence of hierarchical structure. He claims that the correlation matrices lack the notation of hierarchy, which means that all investments are potential substitutes for each other. He also provides a concrete example to illustrate the lack of hierarchy in correlation matrices (López de Prado, 2016b, p. 4):

> For example, stocks could be grouped in terms of liquidity, size, industry and region, where stocks within a given group compete for allocations. In deciding the allocation to a large publicly-traded U.S. financial stock like J.P. Morgan, we will consider adding or reducing the allocation to another large publicly-traded U.S. bank like Goldman Sachs, rather than a small community bank in Switzerland, or a real estate holding in the Caribbean. And yet, to a correlation matrix, all investments are potential substitutes to each other.

Figure 6 visualizes the relationships implied by a correlation matrix of size 16 as a *complete graph*, that is, as a graph where every node is a potential substitute for another.

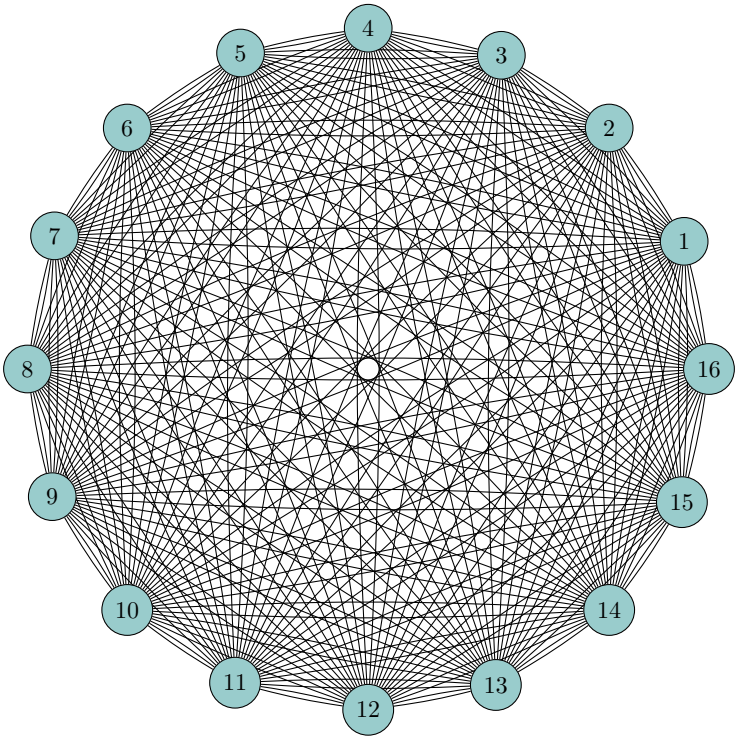

Figure 6: Visualization of the relationships implied by a correlation matrix of size 16. The correlation matrix is represented as a complete graph, as each asset is seen as substitutable with another.

In a similar vein, Tumminello et al. (2010) claim that a complete graph has too much information, and therefore, a "filtering" procedure can ease the analysis of a correlation matrix. Hence, instead of analyzing the full covariance matrix, López de Prado (2016b) proposes applying a hierarchical structure, known as a *tree*, to simplify the analysis. The tree structure, illustrated in Figure 7, has two desirable features (Raffinot, 2017; López de Prado, 2016b):

1. It filters the correlation-network and finds the links in the network that are important and removes the rest. As a result, the tree consists of only $N-1$ edges, whereas the original matrix has $\frac{1}{2}N(N-1)$ edges.
2. It distributes the weights using a top-down approach. This is intuitive and congruous with how many asset managers construct their strategic portfolios, e.g., from macro factors or asset classes to sectors to individual securities.

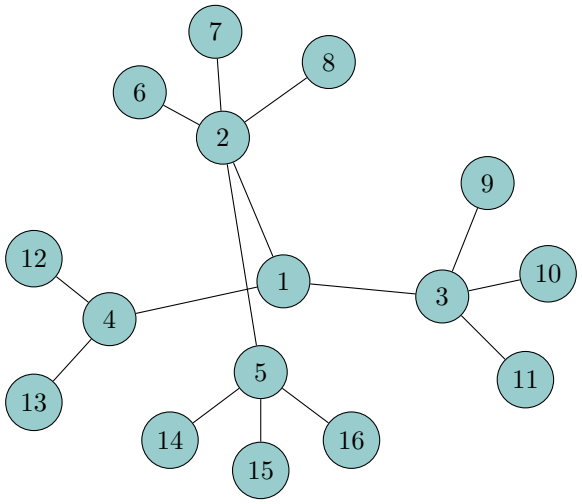

Figure 7: Visualization of the relationships implied by a correlation matrix of size 16. The correlation matrix is represented with a tree structure since it incorporates hierarchical relationships; thus, the assets are not seen as substitutable with another, as appose to the complete graph.

Following these arguments, López de Prado (2016b) proposed a portfolio allocation method called the Hierarchical Risk Parity (HRP). The HRP addresses the issues with the lack of hierarchy in correlation matrices by utilizing a hierarchical clustering of assets. Furthermore, it incorporates a heuristic risk-based asset allocation method and thus, avoids the problems with covariance matrix instability. The HRP does not require the inversion of a covariance matrix, and it can even handle singular covariance matrices, an impossible task for quadratic optimizers. Before exploring the proposed method in more detail, we will introduce clustering and other relevant terminology in the following sections.

3.6 Hierarchical Clustering

Raw data is challenging to interpret and understand, but arranging it into different groups is one of the most fundamental ways of understanding information (Jain, 2010). Cluster analysis is the study of methods and algorithms for grouping or clustering data by measured or perceived intrinsic characteristics or similarities. Clustering avoids class labels and thus does not use category labels that tag objects with prior identifiers. There are two broad categories of clustering algorithms: hierarchical and partitional (Jain, 2010). This study will only focus on hierarchical clustering.

Hierarchical clustering can be boiled down to three steps: (A) selecting the type of hierarchical clustering algorithm, (B) shaping the data to represent the objects' distance to each other, and (C) performing the clustering and initiating hierarchy. Thus, the second step involves selecting a suitable distance measure and assessing the data based on the chosen measurement. In the third step, the clusters are formed based on the distances determined in the second step and a linkage criterion, which defines the distance between clusters (Raffinot, 2017). There are two main types of hierarchical clustering algorithms (Jain, 2010):

- Agglomerative clustering (bottom-up): each data point is placed in its own cluster, and at each step of the algorithm, the two most similar clusters are merged into a new larger cluster. The algorithm is repeated until all clusters are members of one single large cluster, often displayed using a dendrogram.
- Divisive hierarchical clustering (top-down): all data points are put into one large cluster, and in each step of the algorithm, the current cluster is divided into two new clusters. The algorithm is repeated until all observations are in their own cluster.

Figure 8 illustrates the two clustering algorithms.

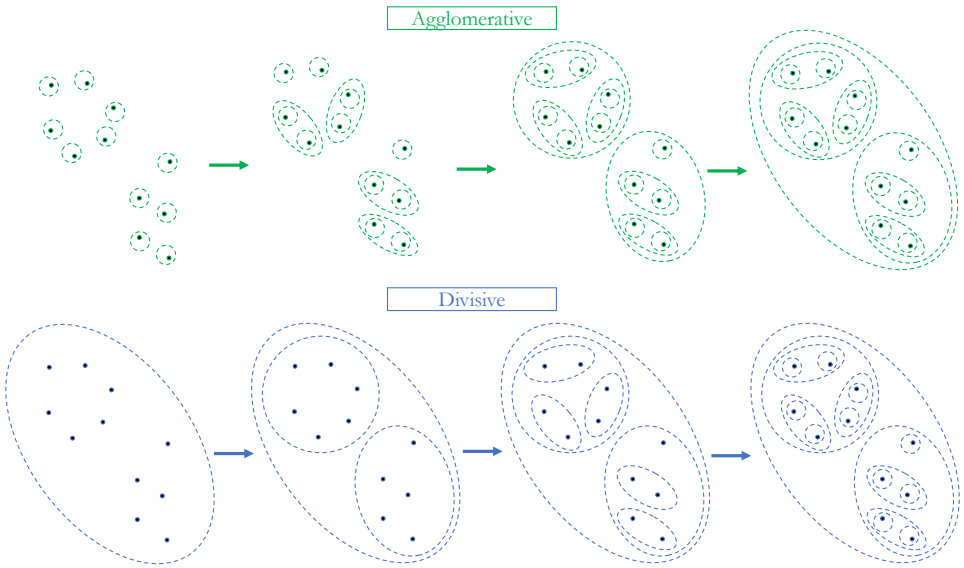

Figure 8: Step-by-step illustrations of Agglomerative and Divisive clustering. In the Agglomerative clustering algorithm, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Conversely, in the Divisive clustering algorithm, observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

While both methods have their advantages and disadvantages, this thesis will focus on agglomerative hierarchical clustering as it is the method that hierarchical risk parity utilizes.

3.6.1 Distance Measure for Clustering

Hierarchical clustering requires a measure of similarity (or dissimilarity) between clusters, regardless of the chosen clustering method. However, given the variety of features and scales, the distance measure needs to be selected with care (Jain et al., 1999). Thus, a brief explanation of distance metrics and López de Prado's (2016b) choice for transforming correlations to a distance metric is provided below.

Similarity-based clustering methods rely on a distance matrix that contains the pairwise distance between N objects based on a similarity (or dissimilarity) measure. If we consider a $T \times N$ matrix $\mathbf{X}$ of financial data, then similarity-based clustering methods require replacing $\mathbf{X}$ with a $N \times N$ dissimilarity matrix or distance matrix $\mathbf{D}$[4]. $\mathbf{D}$ is obtained by applying a dissimilarity measure to the data matrix $\mathbf{X}$. However, the correlation coefficient between assets cannot be used as a distance measure since it does not fulfill the properties that define a distance metric (Mantegna, 1999). A distance metric between two elements must satisfy the following properties:

- Symmetry: $d[x, y] = d[y, x]$
- Non-negativity: $d[x, y] \geq 0$
- Identity of indiscernibles: $d[x, y] = 0 \iff x = y$
- Triangle inequality: $d[x, z] \leq d[x, y] + d[y, z]$

Thus, López de Prado (2016b) suggests applying the following transformation to the correlation coefficients, $\rho_{i,j}$, to make it into a distance metric:

$$d_{i,j} = d[X_i, X_j] = \sqrt{\frac{1}{2}(1 - \rho_{i,j})} \tag{12}$$

This formulation allows for the computation of a $N \times N$ distance matrix $\mathbf{D}$, defined as:

$$\mathbf{D} = \{d_{i,j}\}_{i,j=1,\dots,N} = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & \cdots & \cdots & 0 \end{bmatrix} \tag{13}$$

For perfectly correlated assets, $\rho_{i,j} = 1$, the distance is zero. Conversely, perfectly negatively correlated assets, $\rho_{i,j} = -1$, have a distance of one.

[4]Note that dissimilarity and distance are used interchangeably in this thesis. The only difference between these two is that a dissimilarity measure does not have to satisfy the triangle inequality (Murphy, 2012). Thus, all distances are also dissimilarities.

3.6.2 Linkage Criteria

The linkage criterion determines the distance between clusters. There are four primary linkage criteria: single linkage, complete linkage, average linkage, and Ward's method. There are pros and cons to all the different linkage criteria. However, Papenbrock (2011) found that single linkage and Ward's method yielded superior performance with respect to the risk-adjusted-performance. Additionally, López de Prado's (2016b) HRP approach uses single linkage as its linkage criterion. As a result, this thesis will focus on these two methods, while Appendix B provides a description of complete and average linkage.

Single Linkage

In single linkage, also called the nearest-neighbor clustering, the distance between two clusters is the minimum distance between members of the two clusters (Raffinot, 2018). Let C_i and C_j represent two clusters. The distance d_{C_i,C_j} between C_i and C_j is calculated as the minimum of the set of pairwise observation distances $D(x, y)$ where one member of the pair, x, is in cluster C_i and the other, y, is in C_j. Then, for clusters C_i, C_j:

$$d_{C_i,C_j} = min_{x,y}\{D(x, y) \mid x \in C_i, y \in C_j\} \tag{14}$$

The method only requires that a single distance $D(x, y) \mid x \in C_i, y \in C_j$, to be small for two groups C_i and C_j for the clusters to be merged, regardless of the other observation dissimilarities between the groups. Overall, single-linkage is relatively efficient and straightforward to implement and allows for non-elliptical shapes. However, single-linkage is sensitive to outliers and susceptible to a phenomenon called *chaining* (Hastie et al., 2009), described in detail in Section 4.

Ward's Method

Ward's method, also called the minimum variance method, is an agglomerative clustering method proposed by Joe H. Ward Jr. in 1963. Ward's method assumes that a cluster is represented by its centroid (centre point of the cluster), and the proximity between two clusters is measured in terms of the increase in the squared error that results when the two clusters are merged (Tan et al., 2016). For clusters C_i, C_j, with m_i, m_j numbers of observations, respectively:

$$d_{Ci,Cj} = \frac{m_i, m_j}{m_i + m_j}\|c_i - c_j\|^2 \tag{15}$$

where c_i, c_j are the centroids of the clusters, defined as:

$$c_i = \frac{1}{m_i}\sum_{x \in C_i} x \tag{16}$$

Comparison

Figure 9 illustrates the two methods. However, the figure does provide any performance differences between the two methods. In fact, linkage criteria are difficult to compare as they cannot be theoretically compared since

it is practically impossible to describe the different approaches mathematically in a way that can be compared (Jain and Dubes, 1988). Instead, Papenbrock (2011) makes a comparison based on the intended use of the algorithms, leaving the choice to the user.

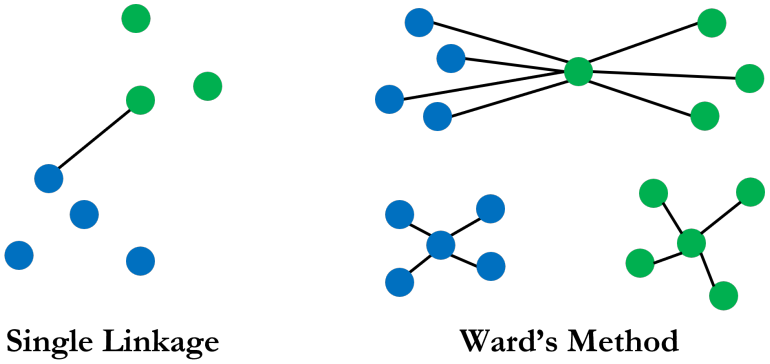

Figure 9: Illustration of single linkage and Ward's method.

Single linkage is sensitive towards outliers and is prone to form chained clusters, which leaves large clusters. However, it provides a robust estimation of the distance matrix while keeping as much as possible of the original structure. Additionally, single-linkage helps gain insights into the correlation structure between assets or separate assets that are different (Papenbrock, 2011). In contrast, Ward's method is less sensitive to outliers as it focuses on the center of clusters rather than individual objects. Ward's method is computationally heavier and is not guaranteed to produce better results. In fact, no linkage method is optimal in all cases but should be used depending on the data set and clustering objective. Figure 10 illustrates single linkage and Ward's method on four generated data sets with different structures[5]. The figure shows that Ward's method is computationally heavier as it takes longer to detect the clusters. Additionally, it clearly illustrates that the two methods create different clusters and that the choice forlinkage cr iteria should depend on the underlying data set.

[5]The first two datasets are generated using the scikit-learn library's *make_circles* and *make_blobs*, respectively. That generates 2D binary classification datasets that are challenging to certain algorithms. The third dataset is blobs with varied variances. Lastly, the last dataset is white noise, or data with no structure.

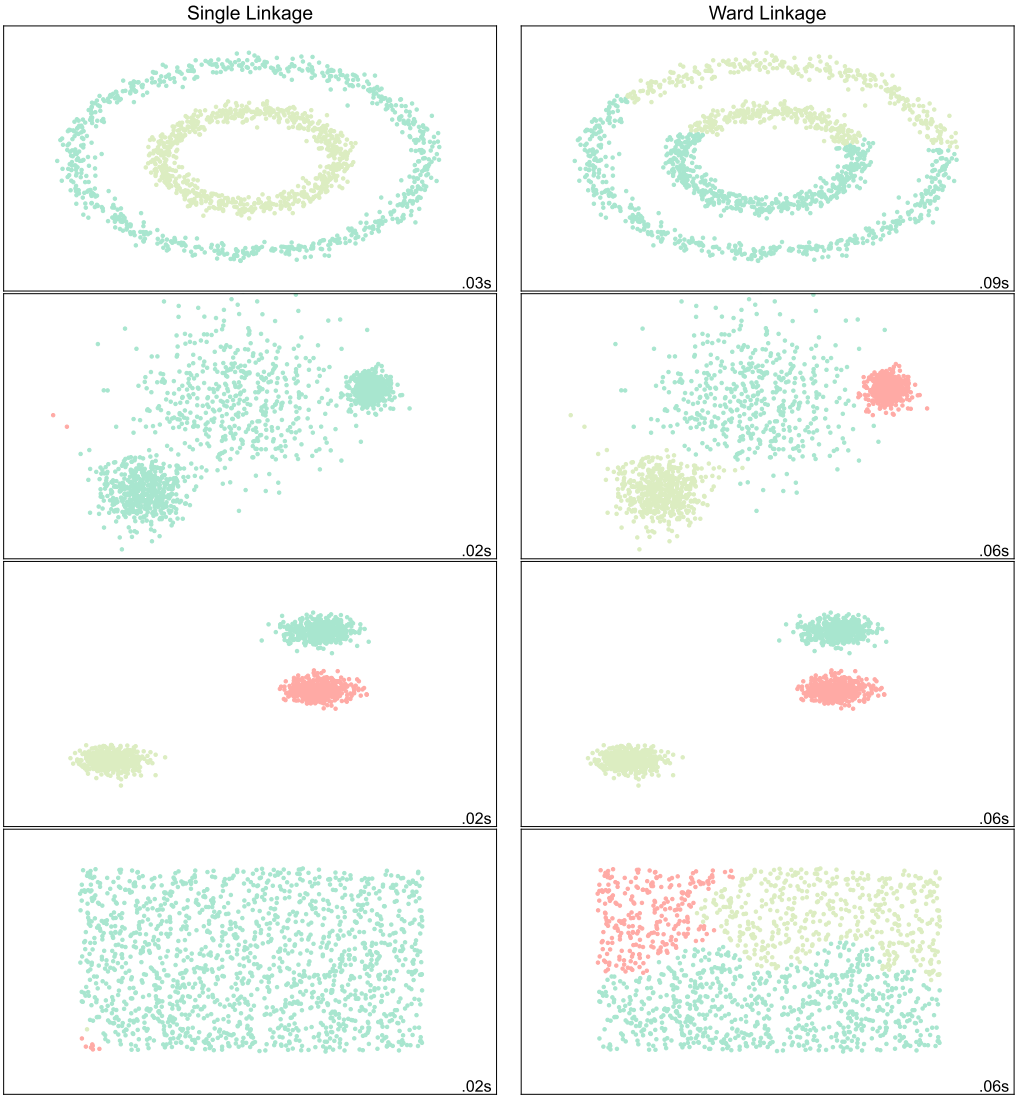

Figure 10: Illustration of the single linkage and Ward's method on four different datasets. The first three datasets have different embedded structures, whereas the last dataset is without any structure.

3.6.3 Dendrograms

The merging process of hierarchical clustering is usually represented graphically using a binary tree called a *dendrogram*. A dendrogram displays the relationships between clusters, their subclusters, and the order in which the clusters were merged. The height, depicted on the y-axis, reflects the dissimilarity between the joined clusters. This is also referred to as the *cophenetic distance* (Tan et al., 2016).[6]

The terminal nodes of a dendrogram represent individual observations, or *singleton clusters*, and are all plotted at zero height (Hastie et al., 2009). Clusters that are the most similar are merged first, represented in the dendrogram as the branch with the lowest height. As the agglomerative clustering methods possess a monotonicity property, the dissimilarity between merged clusters increases with the level of the merger until it reaches the

[6]The cophenetic distance only tells us about the dissimilarity between the two joined clusters. However, a dendrogram does not allow for inferences about the dissimilarity of two "random" objects or clusters (Lohre et al., 2020).

top of the dendrogram. The final merger at the top of the tree, called the root of the tree, represents a cluster containing all the data (Murphy, 2012).

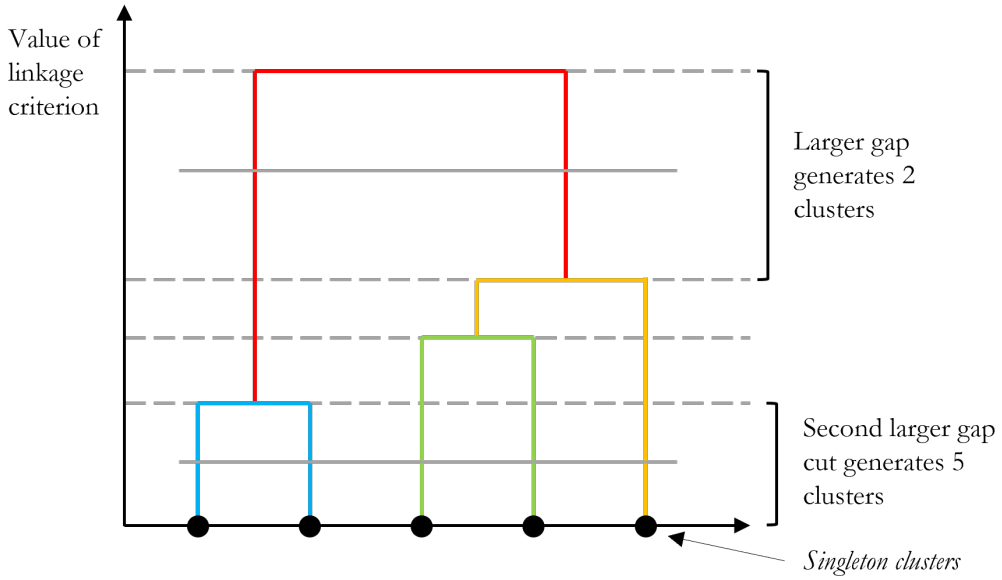

Figure 11: Illustration of a dendrogram with five observations. The colors of the vertical lines illustrate the clustering procedure, and the black dots illustrate the individual observations (*singleton clusters*).

3.7 Hierarchical Risk Parity (HRP)

The following section illustrates López de Prado's (2016b) HRP approach, starting with a quick overview of his arguments for the introduction and development of the HRP approach and an outline of the three stages of the approach. After that, three subsequent sections explain each of the three stages in detail.

3.7.1 Overview

In 2016 López de Prado introduced the HRP approach to address three major concerns of quadratic optimizers: instability, concentration, and underperformance. The instability problem of traditional allocation strategies is mitigated in the HRP approach, as it does not rely on the estimation of the expected returns or require the inversion of the covariance matrix. The concentration problem concerns mainly the minimum-variance allocation strategy, which drives extreme concentration because of its goal of minimizing the portfolio's variance. Lastly, although not explicitly stated, the underperformance issue may refer to many portfolio optimization methods failing to outperform the simple 1/N portfolio consistently (DeMiguel et al., 2009).

Unlike the traditional asset allocation strategies, HRP incorporates the notion of hierarchy, which is intuitively sensible. Instead of viewing all assets in a portfolio as substitutes, an investor would undoubtedly consider certain assets more as substitutes for one another and others more as complements to one another. For example, in a pure equity portfolio, an investor may group her stocks by their respective industries. This example portrays the intuition behind the notion of hierarchy in the HRP approach. However, it should be noted that the algorithm does not use qualitative aspects to cluster the assets, only the covariance matrix.

The HRP approach can be broken down into the following three stages:

1. **Tree clustering**: Group similar investments into clusters, based on a proper distance metric and using single linkage.
2. **Quasi-diagonalization**: Reorganize the rows and columns of the covariance matrix, so that the largest values lie along the diagonal.
3. **Recursive bisection**: Split allocations through recursive bisection of the reordered covariance matrix.

The three stages are described in detail in the following sections.

3.7.2 Tree Clustering

In the first stage of the HRP approach, a hierarchical clustering algorithm is used to group "similar" assets. Naturally, one has to decide what metric to use for similarity. López de Prado (2016b) uses correlation as the root of similarity between the assets. He starts with the correlation matrix and converts it into a distance matrix $\mathbf{D} = \{d_{i,j}\}_{i,j=1,\dots,N}$ where $d_{i,j} = \sqrt{\frac{1}{2}(1-\rho_{i,j})}$, as explained in detail in Section 3.6.1. However, this distance only considers assets as pairs, yet it would be advantageous to also consider the role that each asset plays in the context of the asset universe. Imagine two assets with low correlation, but their relationship with the rest of the assets may be similar. For example, A and B have low correlation, but the correlation between

A and C is similar to the correlation between B and C. Here, it would be beneficial to consider these assets more similar than initially thought. Therefore, López de Prado (2016b) takes the Euclidean distance between the columns of the original distance matrix $\mathbf{D}$ and forms the Euclidean distance matrix $\tilde{\mathbf{D}}$:

$$\tilde{\mathbf{D}} = \{\tilde{d}_{i,j}\}_{i,j=1,\dots,N} \text{ where } \tilde{d}_{i,j} = \sqrt{\sum_{n-1}^{N}(d_{n,i} - dn, j)^2}$$

The next step is to perform the clustering based on a linkage criterion. As mentioned, López de Prado (2016b) uses a single linkage based clustering algorithm for constructing the clusters. The clustering algorithm is illustrated in Figure 12 below.

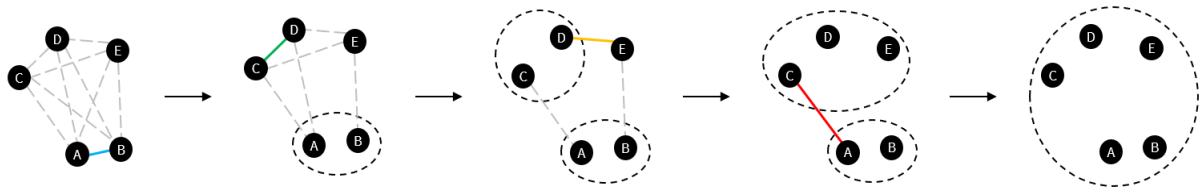

Figure 12: Illustration of clustering using single linkage as linkage criterion.

Additionally, a dendrogram can be constructed to illustrate the notion of hierarchical clustering. Thus, Figure 13 illustrates a dendrogram built on the hierarchical illustration in Figure 12.

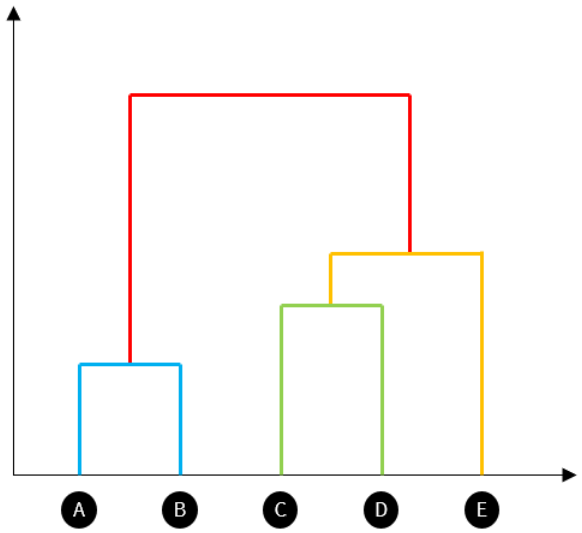

Figure 13: Illustration of a dendrogram based on the single linkage illustration in Figure 12.

The clustering process can be stored in a linkage matrix $\mathbf{Y}$. The matrix will have four columns and $(N-1)$

rows, where each row m represents one cluster and each column holds the following information:

$$\mathbf{Y} = (y_{m_1}, y_{m_2}, y_{m_3}, y_{m_4})$$

y_{m_1} = reference to the left constituent of cluster m

y_{m_2} = reference to the right constituent of cluster m

y_{m_3} = the distance between the two constituents of cluster m

y_{m_4} = the number of assets in cluster m

Thus, the linkage matrix of the tree in Figure 13 would have the following structure:

m	y_{m_1}	y_{m_2}	y_{m_3}	y_{m_4}
1	A	B	d_1	2
2	C	D	d_2	2
3	CD	E	d_3	3
4	AB	CDE	d_4	5

Note that CD, AB, and CDE are clusters and not assets and that column m is only used for illustration and would not be in the linkage matrix.

The first stage in the HRP approached is summarised below in an algorithm layout.

Algorithm 1 Hierarchical Clustering using HRP

1. Estimate $N \times N$ correlation matrix $\boldsymbol{\rho} = \{\rho_{i,j}\}_{i,j=1,\dots,N}$
2. Convert the correlation matrix into a distance matrix $\mathbf{D} = \{d\}_{i,j=1,\dots,N}$ where $d_{i,j} = \sqrt{\frac{1}{2}(1-\rho_{i,j})}$
3. Compute a Euclidean distance matrix $\tilde{\mathbf{D}} = \{\tilde{d}_{i,j}\}_{i,j=1,\dots,N}$ where $\tilde{d}_{i,j} = \tilde{d}(X_i, X_j) = \sqrt{\sum_{n=1}^{N}(d_{n,i} - d_{n,j})^2}$
4. Let us denote the set of clusters as U:
   (a) Construct the first cluster (i^*, j^*) as $U[1] = \underset{i,j}{\operatorname{argmin}}\, \tilde{D}(i,j)$
   (b) Using a single linkage, update the distance matrix by calculating the pairwise distances of the newly formed cluster and other items
   (c) Continue combining assets recursively until we are left with only one single cluster
5. Summarize the information in linkage matrix $\mathbf{Y} \in \mathbb{R}^{(N-1)\times 4}$

Note that the algorithm could be altered to use a different distance metric or linkage criterion.

3.7.3 Quasi-Diagonalization

In the second stage of the HRP algorithm, the rows and columns of the covariance matrix are reorganized so that the largest values lie along the diagonal. This quasi-diagonalization of the covariance matrix is based on the information from the clustering algorithm in the previous stage of the HRP approach. This way, the assets in the covariance matrix which are similar will be placed together, and dissimilar assets will be placed far apart.

As the HRP utilizes inverse-variance allocation, then by quasi-diagonalizing the covariance matrix, López de Prado (2016b) justifies the use of inverse-variance allocation as it minimizes the variance of uncorrelated time series.[7] To illustrate the effect of quasi-diagonalization, the 30 Industry Portfolios' sample correlation matrix is quasi-diagonalized and plotted along with the original sample correlation matrix in Figure 14. The figure shows the effect of undergoing this quasi-diagonalization, where the colors represent higher and lower values (correlation). Notice that the highest values lie near the diagonal.

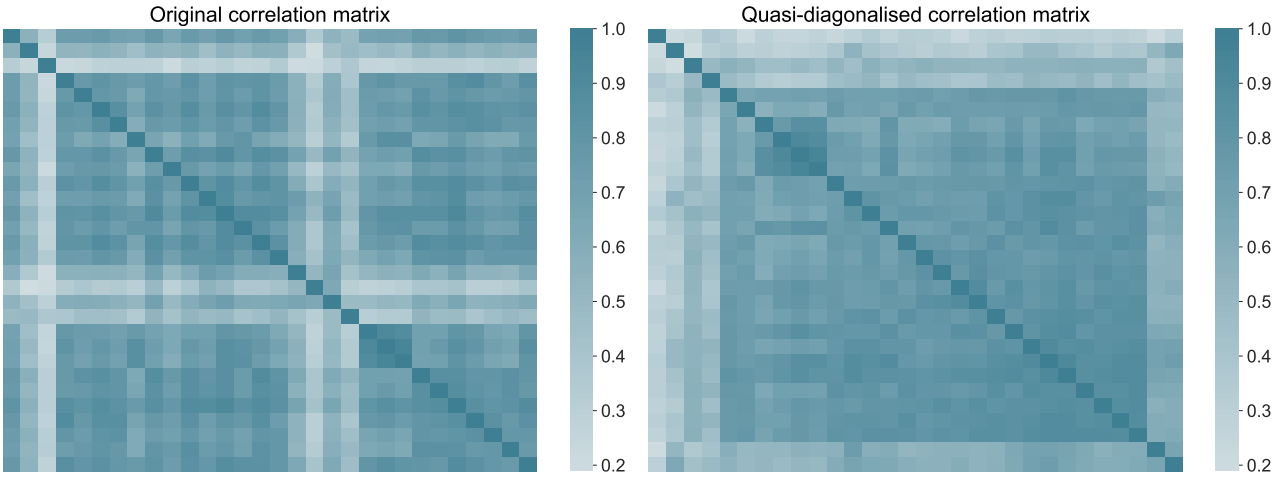

Figure 14: Illustration of the 30 Industry Portfolios' sample correlation matrix before and after quasi-diagonalization.

3.7.4 Naïve Recursive Bisection

In the final stage of the HRP algorithm, the recursive bisection algorithm allocates the portfolio weights. First, the algorithm is initialized by making a list, L, of items, L_0, and assigning a unit weight of one, $W = 1$, to each security in the list:

$$L = \{L_0\} \text{ with } L_0 = \{n\}_{n=\{1,...,N\}} \text{ and } W_n = 1 \text{ for all } n = \{1, ..., N\}$$

Then in a tree-like manner, the algorithm goes from top to bottom and allocates weight each time the clusters splits. The algorithm assumes the tree to be binary, where clusters are recursively split into two equally-sized sub-clusters. Note that this tree is not necessarily the tree produced in stage 1. Only the order of the assets produced in stage 2 is used.

The algorithm uses a top-down inverse-variance allocation to determine the final weights of each asset. Thus, the variance of the clusters needs to be determined and is found by:

$$Var\big(L_i^{(j)}\big) = \mathbf{w}_i^{(j)^\top} \mathbf{\Sigma}_i^S \mathbf{w}_i^{(j)} \tag{17}$$

[7] See López de Prado (2016b) for a proof.

where $L_i^{(j)}$ denotes the subset j of the item i, $\mathbf{\Sigma}_i^S$ is the covariance matrix between the constituents of the subset, and $\mathbf{w}_i^{(j)}$ denotes the weights of the securities in the sub-cluster. The weights of the securities in a given sub-cluster is determined using a bottom-up inverse-variance allocation, that is

$$w_i^{(j)} = \frac{diag[\mathbf{\Sigma}_i^S]^{-1}}{tr(diag[\mathbf{\Sigma}_i^S]^{-1})} \tag{18}$$

where diag[.] and tr[.] are the diagonal and trace operators defined in Section 3.1.

Finally, the relative weights of each sub-cluster are updated using a split factor, α_i, which is calculated using the inverse-variance allocation between the two clusters, which for two assets simplifies to their relative variances:

$$\alpha_1 = 1 - \frac{Var\big(L_i^{(1)}\big)}{Var\big(L_i^{(1)}\big) + Var\big(L_i^{(2)}\big)}$$
$$\alpha_2 = 1 - \alpha_1$$

Essentially, this means that if subset 1 has a higher variance than subset 2, the weight of subset 1 is scaled down. The weights in the sub-clusters, $\mathbf{W}_i^S$, are then updated (re-scaled) using the split factor

$$\mathbf{W}_i^S = \alpha_i \cdot \mathbf{W}_i^S$$

The algorithm is recursive, and the above-mentioned procedure repeats until the bottom of the tree is reached. For example, for a portfolio of five assets, the list of weights would be updated in the following manner:

$$\begin{aligned}[1,1,1,1,1] &\rightarrow [0.6, 0.6, 0.4, 0.4, 0.4] \\ &\rightarrow [0.36, 0.24, 0.4, 0.4, 0.4] \\ &\rightarrow [0.36, 0.24, 0.24, 0.16, 0.16] \\ &\rightarrow [0.36, 0.24, 0.24, 0.096, 0.064]\end{aligned}$$

Note that this example uses a constant $\alpha = 0.6$ which is unrealistic but simplifies the illustration of the procedure.

The final stage in the HRP approached is summarised below in an algorithm layout.

Algorithm 2 Naive Recursive Bisection

1. The algorithm is initialized by:
   (a) Setting the list of items: $L = \{L_0\}$, with $L_0 = \{n\}_{n=1,...,N}$
   (b) Assigning a unit weight to all items: $W_n = 1, \forall n = 1, ..., N$
2. If $|L_i| = 1, \forall L_i \in L$, then stop.
3. For each $L_i \in L$ such that $|L_i| > 1$:
   (a) Bisect L_i into two subsets, $L_i^{(1)} \cup L_i^{(2)} = L_i$, where $|L_i^{(1)}| = \text{int}[\frac{1}{2}|L_i|]$, and the order is preserved
   (b) Define the variance of $L_i^{(j)}, j = 1, 2$, as the quadratic form $Var\left(L_i^{(j)}\right) = {w_i^{(j)}}^\top \mathbf{\Sigma}_j^S w_i^{(j)}$, where $\mathbf{\Sigma}_j^S$ is the covariance matrix between the constituents of the subset $L_i^{(j)}$, and $w_i^{(j)} = \frac{diag[\mathbf{\Sigma}_j^S]^{-1}}{tr(diag[\mathbf{\Sigma}_j^S]^{-1})}$, where diag[.] and tr[.] are the diagonal and trace operators
   (c) Compute the split factor: $\alpha_i = 1 - \frac{Var\left(L_i^{(1)}\right)}{Var\left(L_i^{(1)}\right)+Var\left(L_i^{(2)}\right)}$, so that $0 \leq \alpha_i \leq 1$
   (d) Re-scale allocations W_n by a factor of $\alpha_i, \forall n \in L_i^{(1)}$
   (e) Re-scale allocations W_n by a factor of $(1 - \alpha_i), \forall n \in L_i^{(2)}$
4. Loop to step 2

Note that in $\text{int}[\frac{1}{2}|L_i|]$ the "int" function in is a floor function that rounds down to the nearest integer.

4 Improving the HRP

The following section highlights and explains three of the HRP approach's most notable flaws and weaknesses. After that, an alternative method developed to alleviate these issues is presented. Lastly, the Elbow method is introduced and explained.

4.1 Undesirable Properties of the HRP

Since López de Prado developed the HRP in 2016, several studies have identified a few flaws, weaknesses, and undesirable properties (Raffinot, 2018). Three of these issues are:

- Chaining
- Lack of link between first and third step
- Optimal number of clusters

The following sections will describe and illustrate these issues in detail.

Chaining

The HRP algorithm incorporates the single linkage method, which suffers from an effect known as *chaining*. The single linkage only requires that a single distance is small for two clusters to be joined, disregarding all the other members of the clusters. Therefore, it tends to combine observations linked by a series of close intermediate observations Hastie et al. (2009).

The chaining effect is illustrated in Figure 15. We clearly see three clusters encircled with red dashed lines in the figure. However, using the single linkage, the two clusters at the left of the figure are merged through two intermediate observations. This chaining effect makes the clusters long and too spread out (Raffinot, 2018), causing the clusters to have dissimilar assets.

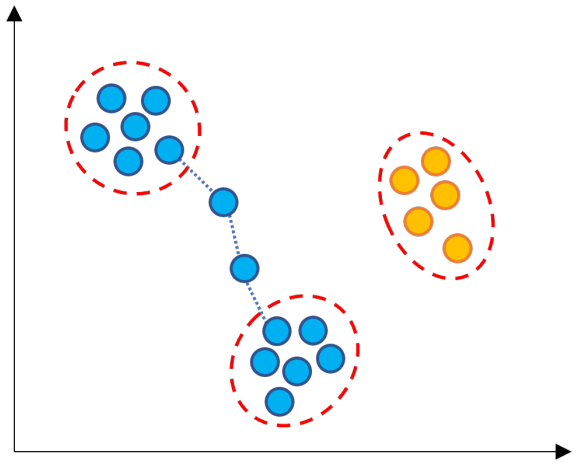

Figure 15: Illustration of the chaining effect. The two clusters with blue circles at the left are merged through two intermediate observations resulting in an unwanted large cluster instead of the three clusters encircled with red dashed lines.

Chaining can be problematic for the HRP portfolio construction since it could result in clusters containing considerably dissimilar assets that should not be in the same cluster. Effectively, including additional assets to an existing cluster changes the cluster's variance. Subsequently, as clusters compete for allocations based on their relative variances, the chaining of clusters will affect the allocated weights and result in suboptimal allocations.

Lack of Link Between the First and Third Stage

The hierarchical structure produced in the first stage is not considered when the bisection is performed in stage three; instead, only the order created in stage two matters. This counterintuitive feature has been pointed out by Raffinot (2017, 2018), Lohre et al. (2020), and Huang (2020). The lack of a link between the stages causes multiple issues illustrated and described below.

Figure 16 illustrates the issue using a simple case of four assets divided into two clusters. Logically, the dendrogram suggests that asset A is one cluster and B, C, and D are the other cluster (indicated by the black line). However, using recursive bisection, the HRP ignores the clusters' structure and only focuses on the number of assets and the order obtained at the bottom level. The effect is that assets A and B are merged into one cluster and C and D into the other cluster.

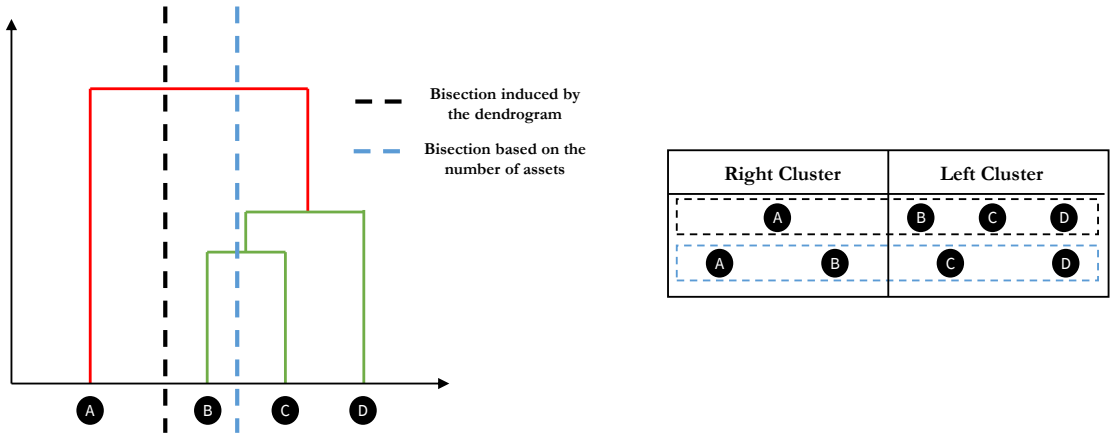

Figure 16: Illustration of bisection based on the dendrogram and the number of assets. The box on the right indicates the clusters' assets, where the bisection based on the dendrogram splits the assets 1-3, and the bisection based on the number of assets splits the assets 2-2.

Furthermore, the lack of a link between stages 1 and 3 yields the undesirable trait that the HRP is sensitive to the order of the assets, i.e., by swapping the placement of the assets in the covariance matrix, the allocated weight might change. For example, if we change the places of assets B and C in Figure 16 and update the covariance matrix accordingly, the underlying data does not change, and the allocated weights should remain the same. To illustrate this issue, Figure 17 shows the distribution of portfolio weights in a ten asset example. The distribution was constructed by resampling the order of the dataset, calculating a sample covariance matrix,

and calculating the optimal weights using HRP. This procedure was repeated 1000 times. The figure illustrates that by simply changing the order of the assets, the HRP allocates the weights differently, albeit the differences are relatively small.

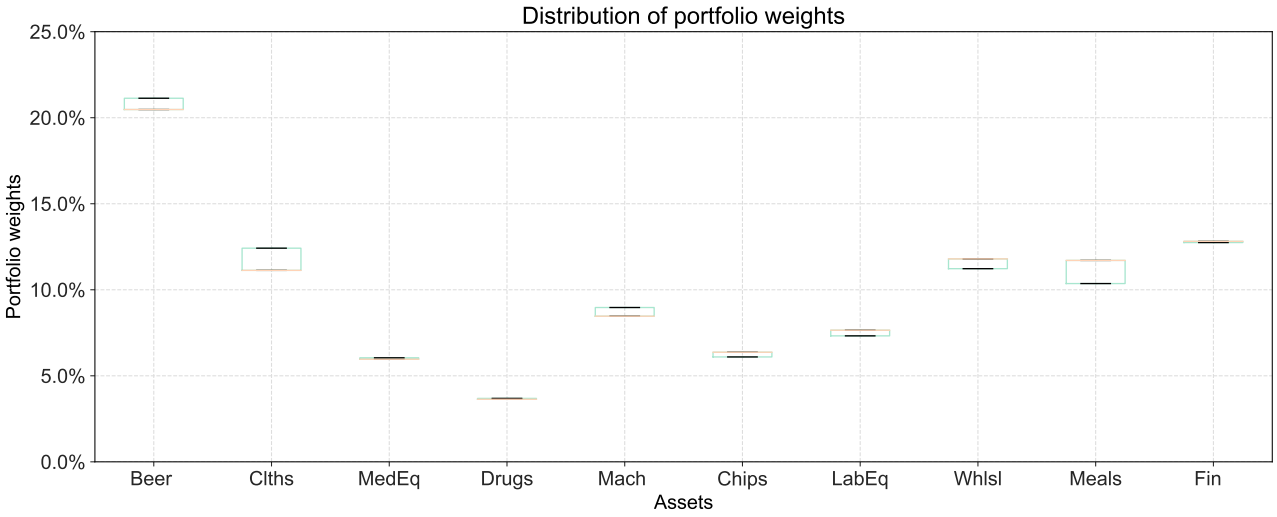

Figure 17: Boxplot illustration of the distribution of portfolio weights. The distribution was constructed by resampling the order of the dataset, calculating a sample covariance matrix, and calculating the optimal weights using HRP. This procedure was repeated 1000 times. The illustration highlights the HRP's sensitivity towards asset order.

Lastly, the recursive bisection splits the list of assets into two equally sized groups. However, if the number of assets is uneven, the algorithm naively splits the list so that the "right" subset contains the extra asset. Thus, the algorithm ignores any similarity, distance, or dissimilarity between the assets and only splits based on an arbitrary rule for uneven numbers.

Number of Clusters

In the HRP algorithm, each asset is treated as its own cluster, and the relevant number of clusters is not considered. According to Raffinot (2018), this can be seen as a form of overfitting, leading to potentially harmful results. Essentially, as the HRP grows the tree to a maximum depth, it creates singleton clusters at the lowest hierarchical level. As such, the clustering process might try to provide more information about the underlying structure than can be provided with reliability, and it will almost certainly lead to a sub-optimal number of clusters and, consequently, sub-optimal allocations (Huang, 2020).

4.2 HRP Topdown

Section 4.1 highlighted three main issues with López de Prado's (2016) HRP. We have developed an alternative method to alleviate these issues, which will be described in the following section. The new approach will be referred to as HRP Topdown, while López de Prado's algorithm will be referred to as the original HRP approach, or simply the HRP.

The HRP Topdown approach only consists of 2 stages; (1) tree clustering and (2) weight allocation. The tree

clustering is altered to overcome the chaining issue with the original HRP. Instead of using single-linkage, the HRP Topdown uses Ward's method, described in detail in Section 3.6.2. However, everything else in the tree clustering stage is equivalent to the original HRP (see Section 3.7.2) and yields the linkage matrix as output. To overcome the issue of the lack of a link between stages 1 and 3 in the original HRP method, the HRP Topdown uses the tree generated in the first stage to allocate weight instead of using a naive recursive bisection. As a result, the second step (quasi-diagonalization) in the original HRP is excluded. The final step of the HRP Topdown is allocating weights to the assets. The HRP Topdown allocates weights based on the tree from stage one. Thus, it goes from the top of the tree, and for each cluster, it allocates weights within the clusters based on inverse-variance allocation, calculates the total variance of the cluster, and finally allocates weight to each cluster using inverse-variance allocation. Furthermore, the HRP Topdown aims to overcome the issue with the number of clusters. Hence, instead of going through the entire tree, the HRP Topdown cuts off the tree after a specific number of clusters and splits the remaining weight equally amongst the final assets in each cluster. The optimal number of clusters is calculated using the Elbow method described below.

4.3 Elbow Method

The elbow method is a popular technique for determining the optimal number of clusters (López de Prado, 2020). Its key metric is a distortion score, W_k, given by the sum of square distances from each point to its assigned center. If the number of clusters, k, increases, the average distortion will decrease, each cluster will contain fewer constituents, and the constituents will be closer to their respective centroids. However, the marginal improvement in average distortion will eventually decline as k increases. The k at which the marginal improvement in distortion declines is called *the elbow*, at which we should stop dividing the data into more clusters (Dangeti, 2017). Typically, the elbow method is plotted as a cost function produced by the different values of k, as illustrated in Figure 18.

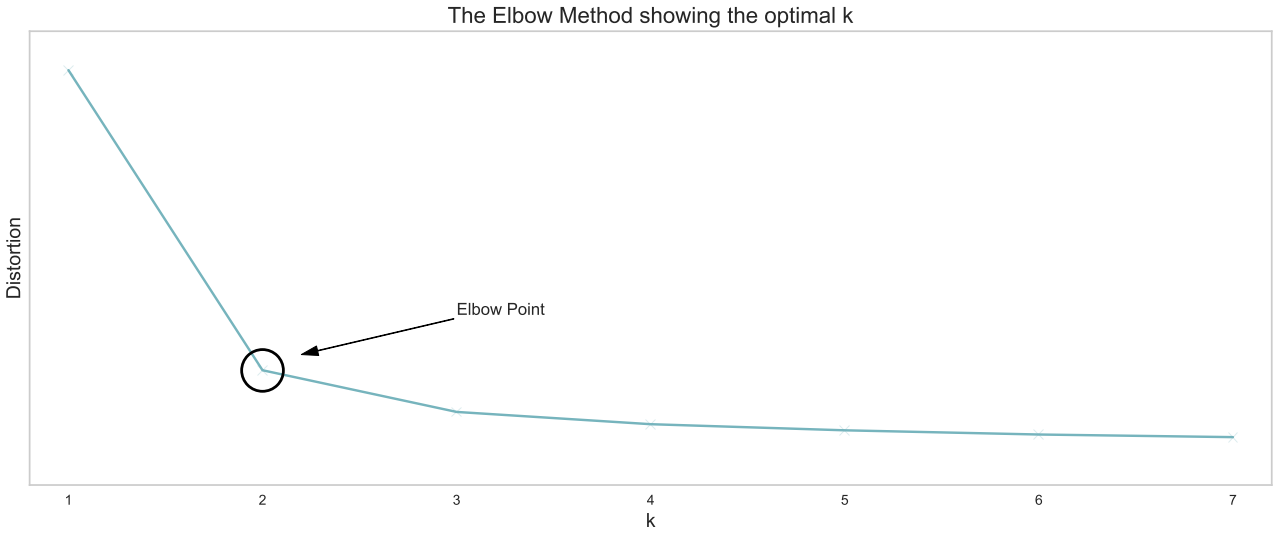

Figure 18: Illustration of the Elbow method. The method stops adding clusters when the marginal improvement in distortion starts to decline. In this case, the optimal number of clusters is two, as illustrated by the encircled point.

5 Monte Carlo Simulations

The following section consists of two analyses. The first analysis covers the impact of estimation error in the covariance matrix on different allocation strategies. It covers two different covariance matrices, one with a simple correlation structure and one with a more complex and realistic correlation structure. The analysis does not seek to provide economic intuition but highlights the models' robustness. The second analysis uses more economic-related performance measures to evaluate the in-sample and out-of-sample performance of the six different allocation strategies. The primary purpose of the analysis is to replicate López de Prado's (2016b) results with minor improvements to investigate the performance of the HRP approach further.

General assumptions and limitations

The HRP is long-only by construction. This implied constraint makes for an unfair comparison between strategies that allow shorting, as constrained portfolios, everything else equal, are less optimal. Hence, we aim to use a fairer comparison in this evaluation by including only long-only versions of different allocation methods. Additionally, many portfolio managers are subject to shorting constraints (Maillard et al., 2010). The benchmark methods included in the analysis are traditional risk-based and heuristic allocation strategies, namely equally-weighted, long-only minimum variance, ERC, and inverse-variance. Finally, the HRP Topdown approach requires an optimal number of clusters as input, calculated using the Elbow method. However, this is computationally heavy and significantly increases the running time. Thus, the optimal number of clusters is calculated on only one simulation and held constant for every iteration.

5.1 Estimation Error in Weight Allocation

Instability in the covariance matrix can affect the accuracy of different portfolio optimization methods, as mentioned in Section 3.4.This is problematic as it may corrupt the research results on portfolio optimization, as the strategies do not perform as optimally as intended. Thus, the following analysis seeks to measure the impact of covariance matrix instability on the weights of different portfolio optimization methods. The overall idea of the following analysis is to compare the difference between optimal portfolio weights computed using a true covariance matrix and a sample covariance matrix. The measurement will be done on a purely statistical basis, and the financial or economic implications are not considered in this analysis.

5.1.1 Methodology

The analysis follows the methodology outlined in López de Prado (2016a). To begin with, we need to define a true correlation matrix with some hierarchical structure. The *block-diagonal correlation matrix*, defined in López de Prado (2016a), is a highly stylized representation of an empirical correlation matrix. It has a distinct structure, where correlations are formed as blocks to the diagonal. As such, it represents an empirical correlation matrix that has a hierarchical structure that resembles clusters. The analysis will be performed based on two different correlation matrices: a *simple block-diagonal correlation matrix* and a *randomized block-diagonal correlation matrix*.

The simple block-diagonal correlation matrix is constructed as follows. First, we generate a correlation matrix of size 50 that has blocks representing the correlation structure in its diagonal. The size of each block is fixed at 10, and all intra-cluster correlations are assumed to be 0.5. Second, the inter-cluster correlations are set to zero to make the structure as simple as possible. Lastly, the correlation matrix is transformed into a true covariance matrix using random uniform volatilities. The simple block-diagonal correlation matrix is illustrated on the left of Figure 19.

The randomized block-diagonal correlation matrix is illustrated on the right of Figure 19 and represents a more realistic correlation structure. Similar to the simple block-diagonal matrix, the number of clusters is five. However, the size of each block is generated randomly to depict non-constant block sizes. The randomized block-diagonal correlation matrix also incorporates a stochastic component by including noise. In addition, the intra-cluster correlations are slightly increased. As discussed earlier in Section 3.4, adding more correlated assets and increasing the intra-cluster correlations exacerbates the signal-induced instability of the matrix. Therefore, the changes make the instability of the corresponding covariance matrix more pronounced and allow us to infer the effect of signal-induced instability on different allocation methods.

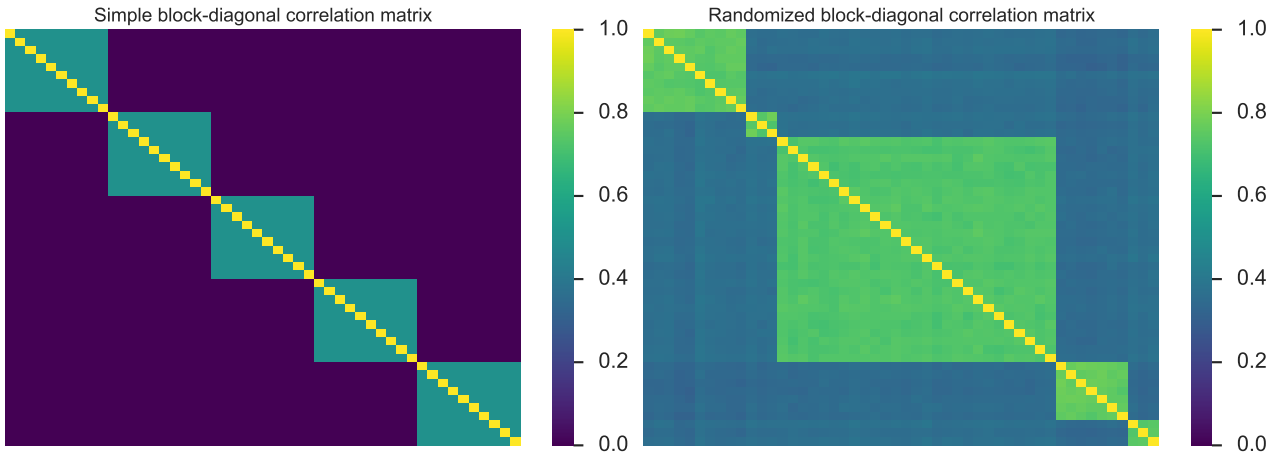

Figure 19: Heatmap illustration of a simple and randomized block-diagonal correlation matrix of 30 assets.

We can simulate returns based on the true covariance matrices by following these formulations of two stylized correlation structures. The true correlation matrices are transformed into true covariance matrices with standard deviations drawn from a random uniform distribution between 5% and 10%. The mean return is assumed to be zero since none of the objective functions require a return component. Using both true covariance matrices and setting the mean equal to zero, we simulate 10,000 return series of 260 observations. The number of observations represents one year of returns consistent with López de Prado's (2016b) out-of-sample Monte Carlo simulations.

Finally, the optimal weights generated by different sample covariance matrices are compared to the true weights generated by the true covariance matrix using the *root-mean-squared error (RMSE)*. The RMSE will reveal the effect of estimation error in the sample covariance matrix on the optimal portfolio weights, and it is defined as

follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(\mathbf{w}_{\hat{\mathbf{\Sigma}},i} - \mathbf{w}_{\mathbf{\Sigma}}\right)^2}{N}} \tag{19}$$

where N is the number of iterations, $\hat{\mathbf{\Sigma}}$ the sample covariance matrix, and $\mathbf{\Sigma}$ the true covariance matrices. The RMSEs are then multiplied by 100 for convenience and readability. The RMSE is a statistical measure and provides little economic intuition. However, economic intuition is not the purpose of the analysis. Thus, the economic-related performance measures are left for the subsequent analysis. Additionally, note that the equally-weighted portfolio is naturally excluded from this analysis since it does not rely on risk estimates.

5.1.2 Results

The results of the two different correlation matrices are presented in the following section. First, the distribution of eigenvalues is examined for the *simple block-diagonal correlation matrix*, followed by a presentation of the estimation error results. Second, the same analysis is performed using the *randomized block-diagonal correlation matrix*.

Simple Block-Diagonal Correlation Structure

Figure 20 shows the corresponding eigenvalues of sample correlation matrices. Since $T = 260$ observations are used to estimate the covariance for $N = 50$ variables, the ratio is $q = N/T = 0.2$. The ratio's deviation from zero indicates that the sample covariance matrix has some noise-induced instability which is also evident in Figure 20. From the figure, we also distinguish a clear signal representing the hierarchical structure of the correlation matrix. These signals push some of the eigenvalues toward zero, making the inversion of the sample covariance matrix slightly more unstable. The average condition number of the 10,000 simple correlation structure matrix simulations is 37.60.

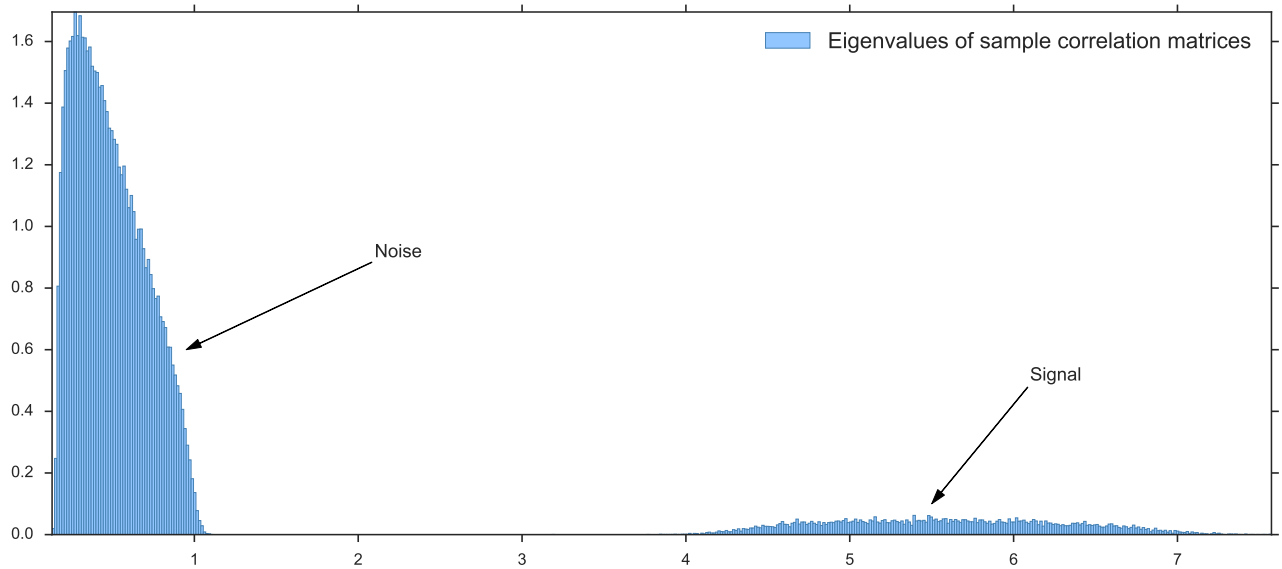

Figure 20: Illustration of the estimated eigenvalues of all the 10.000 simulations of the Simple Block-Diagonal Correlation Structure.

Table 1 presents the estimation errors of simulations based on the simple correlation structure. The MV portfolio has the highest estimation errors, resulting in an RMSE of 1.9300. The results show that the original HRP is sensitive to covariance misspecifications with the second-highest estimation error in the group. On the other hand, the HRP Topdown is more robust and has a considerably lower RMSE of 0.1806. It reduces the estimation errors by 91% compared to the MV and 82% compared to the original HRP. While the IV portfolio produces similar errors to the HRP Topdown portfolio, the equal risk contribution only has an RMSE of 0.0862. The low RMSE indicates that ERC is more robust towards estimation errors in the sample covariance matrix than the other strategies.

Table 1: RMSEs using the simple block-diagonal correlation structure

	IV	MV	ERC	HRP	HRP Topdown
RMSE	0.1788	1.9301	0.0862	1.0896	0.1806

Randomized Block-Diagonal Correlation Structure

Figure 21 displays the corresponding eigenvalues of sample correlation matrices generated from the randomized correlation structure. Although the ratio $q = N/T$ is still 0.2, the more complex hierarchical structure significantly distorts the eigenvalues. The reason is that there is considerably more signal-induced instability that can be traced back to two different sources. First, the large block in the middle of the randomized correlation structure enlarges the highest eigenvalues. This large block, representing a cluster, means that the corresponding assets are more heavily exposed to a common eigenvector. Thus, the eigenvalue associated with these assets explains a greater amount of the variance and decreases the other eigenvalues. Second, the intra-cluster correlations are slightly higher than in the simple structure, exacerbating the instability caused by the signal. As a result, the average condition number grows to 296.14, almost an eightfold increase from the simple structure.

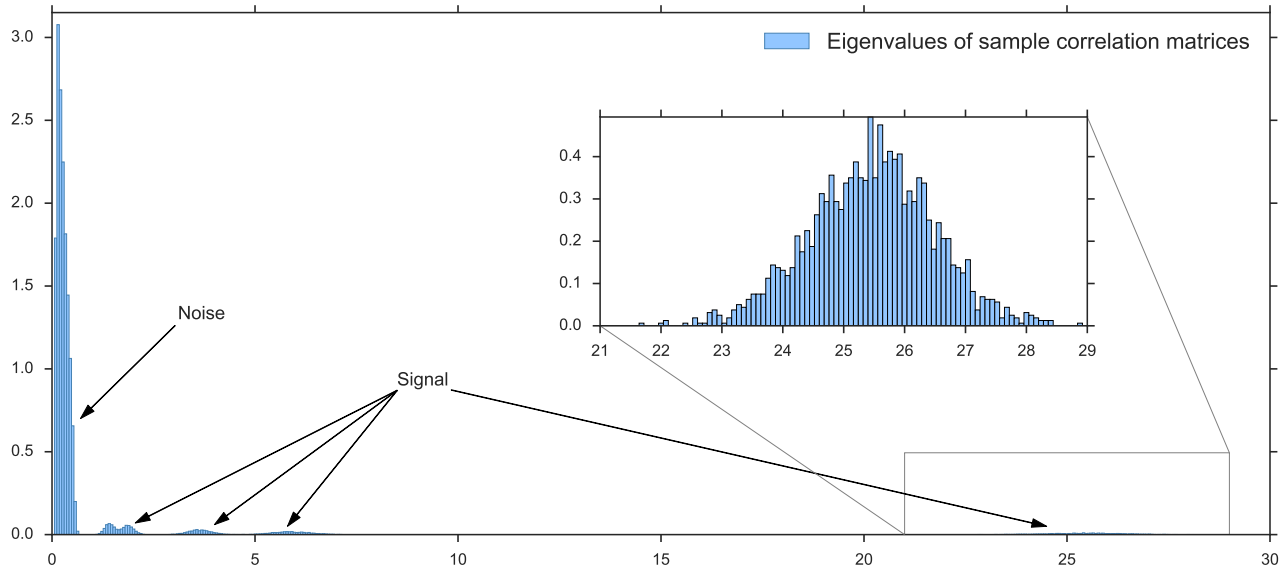

Figure 21: Illustration of the estimated eigenvalues of all the 10.000 simulations of the Randomized Block-Diagonal Correlation Structure

Table 2 shows the RMSEs from the more realistic randomized hierarchical structure simulations. The MV portfolio has the highest estimation errors in line with Table 1. Furthermore, its RMSE increases by 37.5% compared to the simpler correlation structure due to increased instability of the sample covariance matrix. The HRP Topdown portfolio also experience significant increase in the estimation errors. However, the IV, ERC, and HRP portfolios improves from the more complex structure. The IV portfolio does not rely on the entire covariance matrix to allocate weights but only on the diagonal, that is, the variances of different assets. Thus, the IV portfolio should not suffer from the more complex structure but only from misspecifications in the diagonal. Similar to the IV portfolio, the ERC portfolio shows strong robustness towards covariance misspecification consistent with previous research by Ardia et al. (2015), Neffelli (2018), and Sjöstrand et al. (2020). Lastly, the results indicate that the HRP portfolio benefits from the stronger structure in the sample covariance matrix.

Table 2: RMSEs using the randomized block-diagonal correlation structure

	IV	MV	ERC	HRP	HRP Topdown
RMSE	0.1612	2.6543	0.0762	0.9942	0.2371
Increase	-9.9%	37.5%	-11.6%	-8.8%	31.1%

5.1.3 Summary

The simulation studies show that the MV portfolio is the most sensitive strategy to errors in the sample covariance matrix. The optimal weight estimation errors deteriorate with increased sample covariance matrix instability. It is also worth noting that the MV portfolio used in the analysis constrains the weights to be non-negative, which reduces the deviation of weights in estimated optimal portfolios (Jagannathan and Ma, 2003).

The original HRP approach by López de Prado (2016b) has the second-highest RMSEs. This indicates that the HRP method might not be as stable as López de Prado argues. Although the HRP is not as sensitive to the instability of the sample covariance matrix as the MV portfolio, it is still affected by the changes in variances of different assets. Furthermore, the errors are significantly higher than the inverse-variance allocation, although both allocate based on an inverse proportion of the variances. The difference most likely stems from the HRP having a higher concentration in portfolio weights by construction, causing changes in variances to affect the allocations more.

The HRP Topdown method reduces the estimation errors significantly compared to the original HRP approach. The explanation for the reduction is twofold. First, the original approach treats all assets as singleton clusters, and as such, is highly dependent on individual estimated variances. This form of overfitting accentuates the effect of changes in variances between different simulations and results in high variation in individual weights. As the HRP Topdown method only considers the relevant clusters and then weights intra-clusters equally, it is only affected by changes in aggregated cluster variances. Therefore, it is more robust to changes in single variances and produces more stable allocations. Second, the single linkage criterion is more sensitive to outliers

and only requires that a single distance between two clusters is small for the clusters to be merged. In effect, means that minor changes in just one estimated correlation could result in different clusters. In contrast, Ward's method measures the proximity of clusters in terms of the distance between the centroids of the clusters. Hence, Ward's method is less sensitive to outliers and results in more balanced trees and allocation.

The ERC portfolio is the most stable across the simulations, followed by the inverse-variance portfolio. In addition to having the lowest estimation errors, both methods benefit from an increased instability in the sample covariance matrix. This is somewhat counterintuitive and a possible explanation is their low weight concentration as Ardia et al. (2015) argue that lower concentration of portfolio weights leads to less vulnerability to covariance misspecifications. However, a proper investigation into these results is necessary for a comprehensive explanation. This investigation is left for further research.

5.2 Out-of-Sample Simulations

One of López de Prado's (2016b) main findings was that the HRP delivers lower out-of-sample variance than CLA's minimum-variance.[8] He showed this by conducting Monte Carlo experiments. Thus, the goal of the following analysis is to replicate his findings on the out-of-sample performance while also providing insights into the in-sample performance. The analysis follows the methodology outlined in López de Prado (2016b), with a few additions and modifications.

5.2.1 Methodology

Simulations

The simulations are generated precisely as in López de Prado (2016b) to most accurately replicate his results. The sample covariance matrices are estimated using simulated return series, which are simulated 2000 times in the following steps. First, five series of random Gaussian returns with a mean of 0 and an arbitrary standard deviation of 1%[9] are generated. These five series represent different assets. Each set of returns consists of $260 \cdot 2 = 520$ observations, equivalent to 2 years of daily observations. As real returns are not cross-sectionally independent (López de Prado, 2016b), a random correlation structure is added to the generated data by taking the five series and adding random draws from a normal (Gaussian) distribution with 0 mean and 0.25% standard deviation. These five new series of correlated data are appended to the five original series, leaving us with ten assets (return series). Third, random shocks (jumps) are added to the generated returns. Two types of random shocks are included: common shocks to various investments and specific shocks to a single investment. The jump sizes are -50% and 200%. The common shock affects two assets simultaneously, and the specific shock affects one asset at a different time. Thus, a total of six shocks are added to the generated returns. Note that the shocks are always added in the second year of observations, observations 261-520. The simulations are then repeated 2000 times. Figure 22 visualizes a simulation of the ten series of returns (assets) with three subplots. The first plot contains all the assets, the second includes only the asset with shocks, and the third includes assets without shocks.

[8] Unlike (López de Prado, 2016b), we use the SLSQP optimization procedure to derive the minimum-variance portfolio as it is more common in practice. This does not change the interpretation of results since the optimization objective in both cases is to minimize the variance.

[9] Note that López de Prado (2016b) writes 10% but uses 1% in his code. As a daily standard deviation of 1% seems more reasonable than 10%, we assume that he meant to write 1% instead of 10%.

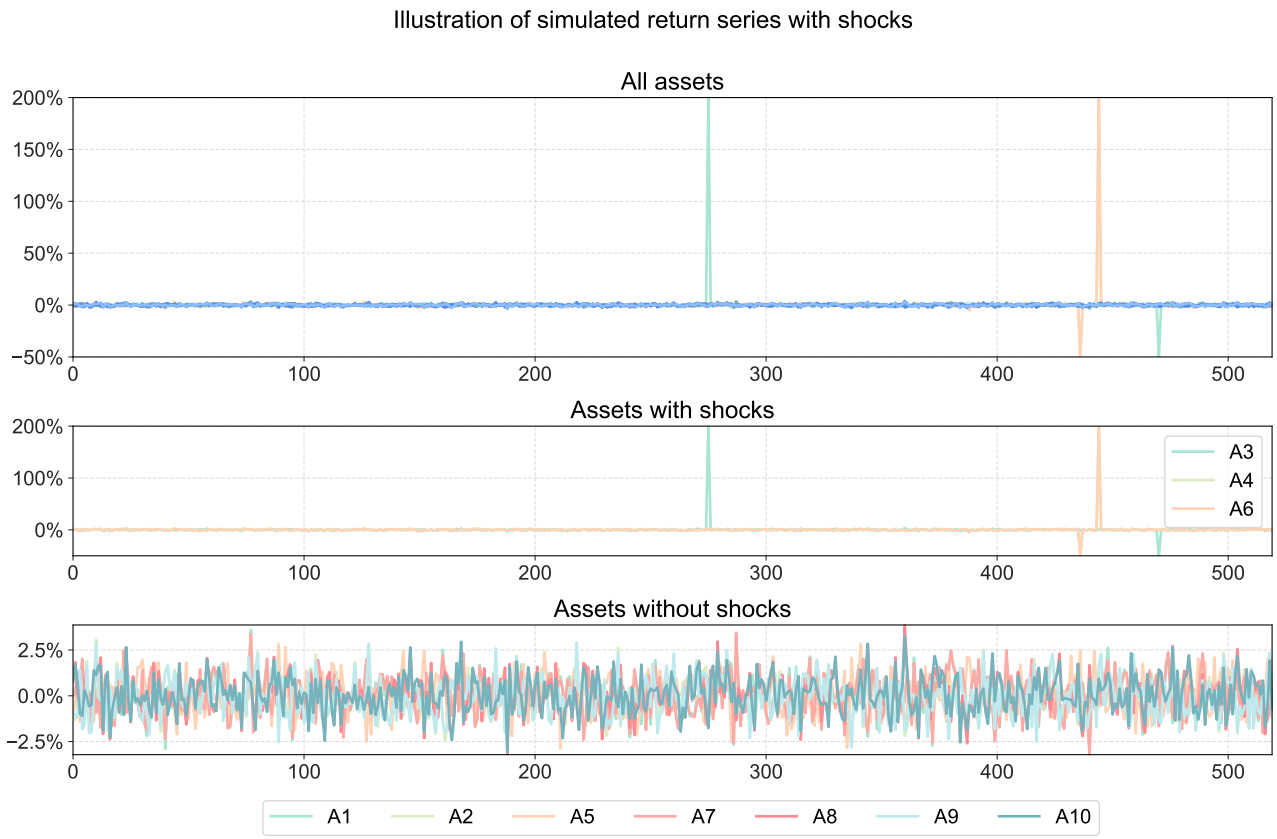

Figure 22: Illustration of 10 simulated return series containing shocks. The return series represents asset returns. The uppermost plot shows all ten return series in the same plot, whereas the middle plot only includes return series with shocks. Lastly, the bottom plot displays the return series that does not contain shocks.

In-sample and Out-of-sample estimations

The sample covariance matrices are estimated using 260 observations, representing the *in-sample* returns. For each allocation strategy, the optimal weights are calculated, and the in-sample variance, $\hat{\sigma}^2_{in}$, is estimated by:

$$\hat{\sigma}^2_{in} = \mathbf{w}^\top \hat{\mathbf{\Sigma}} \mathbf{w}$$

and stored in a dataframe. Next, the out-of-sample returns, r_{oos}, are calculated by:

$$r_{oos} = \mathbf{w}\mathbf{R}$$

where $\mathbf{R}$ represents the following 22 observations in the simulated return series. Then, these 22 returns are stored in a data frame, and the succeeding covariance matrix is estimated by shifting the simulated returns series by 22 observations. Thus, for each iteration in the algorithm, 12 in-sample variances and 260 out-of-sample daily returns are estimated and stored. The in-sample variances are annualized by multiplying by 260.

5.2.2 Results

In-sample Variance

For each iteration, 12 in-sample variances are computed. However, since the shocks are added after the first 260 observations, the sample variances of the portfolios are steadily increasing as the covariance matrix estimates start to incorporate the jumps. Figure 23 illustrates the increase in variance over time by taking an average of each of the 12 variance estimates across the 2000 simulations. The equally weighted portfolio has the highest slope, followed by the HRP Topdown portfolio, indicating their sensitivity towards variance outliers in the assets (assets with shocks). The rest of the portfolios are less impacted by the jumps and thus only increase slightly.

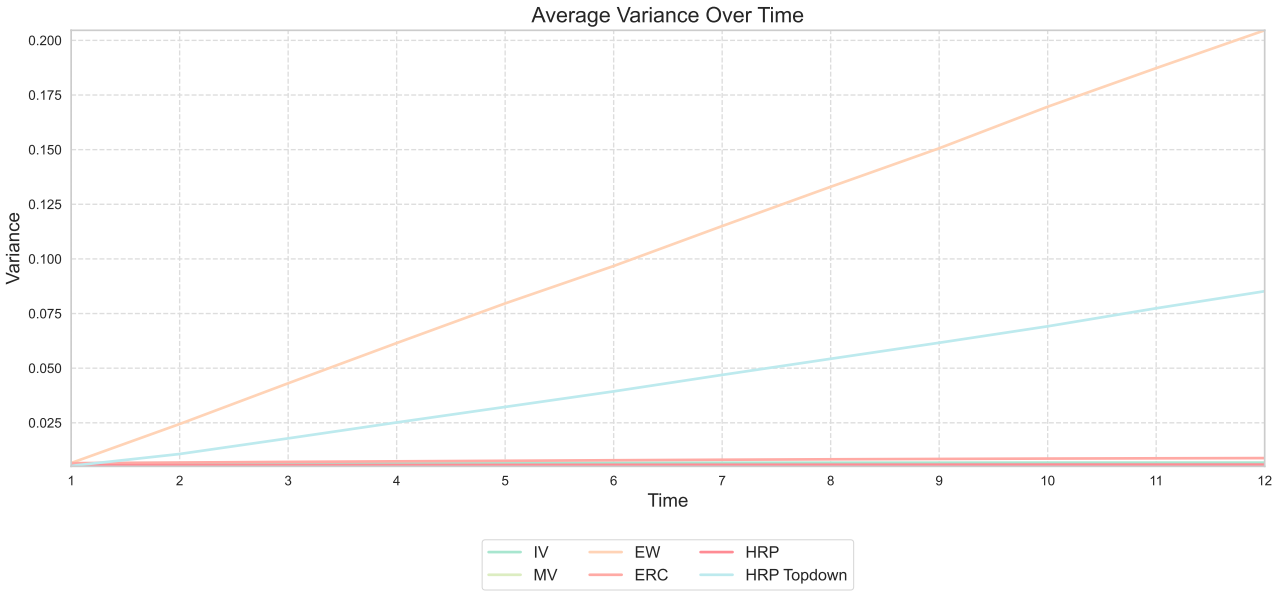

Figure 23: Illustration of the average in-sample variances of different portfolios across time and simulations. The variances are calculated as the averages of each of the 12 variance estimates across 2000 simulations.

The broad variance range of Figure 23 makes it difficult to differentiate between the variances of the portfolios with a flatter slope. Thus, Table 3 highlights the average first observed in-sample variance, last observed in-sample variance, and average in-sample variance across the 12 periods. The MV portfolio has the lowest in-sample variance across all estimates, closely followed by the original HRP. Conversely, the EW portfolio has the highest in-sample variance in all three cases. Intuitively, this makes sense since the EW portfolio does not aim to optimize any function that includes the variance. Hence, it will always allocate weights to the assets with shocks in the in-sample estimation.

Table 3: In-sample Variance

	IV	MV	EW	ERC	HRP	HRP Topdown
First	0.0065	0.0051	0.0066	0.0065	0.0055	0.0053
Last	0.0068	0.0055	0.2046	0.0089	0.0059	0.0852
Mean	0.0067	0.0054	0.1060	0.0079	0.0058	0.0438

Out-of-sample

López de Prado (2016b) estimates the out-of-sample variance based on yearly out-of-sample returns, r_{yearly}, which is found by aggregating the daily returns:

$$r_{yearly} = \prod_{i=1}^{260}(1 + r_i) - 1 \tag{20}$$

Thus, the out-of-sample variance, $\hat{\sigma}^2_{oos}$, is calculated based on a total of 2000 (number of iterations) yearly returns, that is:

$$\hat{\sigma}^2_{oos} = Var(r_{yearly}) \tag{21}$$

Table 5 displays the out-of-sample variance for each of the six portfolios. López de Prado (2016b) found that the out-of-sample variance of the HRP was 0.0671. Thus, our results align with his, and the slight deviation stems from the difference in iteration (our 2000 to his 10000). The table shows that the EW portfolio has the lowest out-of-sample variance, whereas the MV portfolio has the highest. In line with López de Prado (2016b), the HRP has a lower out-of-sample variance than the MV and IV portfolios. However, the HRP Topdown has an even lower variance out-of-sample than the original HRP portfolio.

Table 4: Out-of-sample Variance

	IV	MV	EW	ERC	HRP	HRP Topdown
$\hat{\sigma}^2_{oos}$	0.0930	0.0948	0.0136	0.0635	0.0619	0.0587

To elaborate and further understand the out-of-sample variance of the different allocation strategies, the returns are converted into prices and depicted in a fan chart shown in Figure 24. The y-range is set to $[0.7, 2.2]$. The narrow percentiles of the EW portfolio indicate that it has the lowest out-of-sample variance, in line with the results in Table 5. The MV portfolio has the largest spread, followed by the IV portfolio, while the HRP and HRP Topdown portfolios look relatively similar.

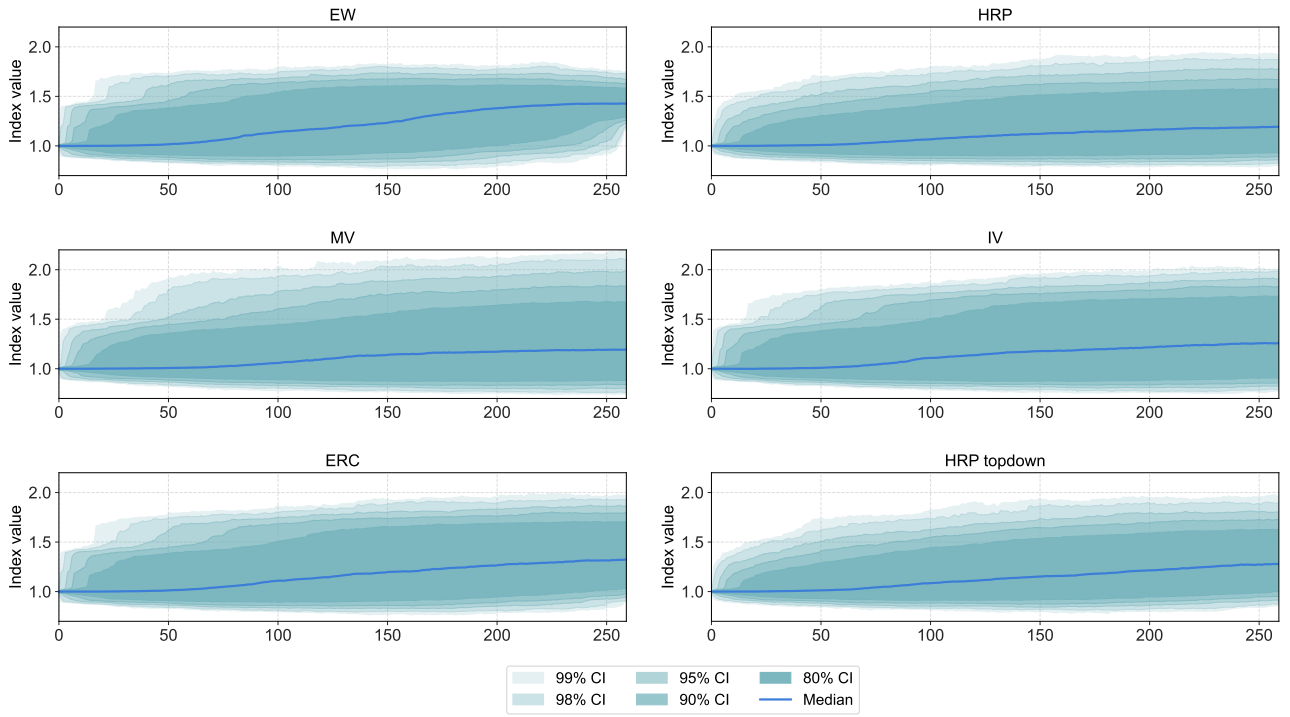

Figure 24: Illustration of the price development of the six different portfolios. The percentiles and medians are based on 2000 simulations, each yielding one year (260 observations) of returns as indicated by the x-axis. The returns are then converted to prices with an initial price of one.

The rather odd path behavior, especially the EW portfolio, stems from the shocks added to the simulated returns. The EW portfolio captures all shocks equivalently (same weight) in every iteration; thus, the overall return of the EW portfolio will be similar for each path, as indicated by the narrowing at the end. The broader interval in the middle of the price path stems from the fact that the shocks appear in different places for each simulation. The identically sized shocks make for an unfair evaluation of the EW portfolio out-of-sample variance as each path (iteration) becomes close to identical for the EW portfolio. Therefore, we also show the standard deviation of daily returns instead of yearly returns to alleviate this issue.

Figure 5 displays the out-of-sample variance of the daily returns multiplied by 260 to annualize and make them more interpretable. The table paints a different picture than Figure 24 and Table 5. In this case, the EW portfolio has the highest out-of-sample variance, while the HRP portfolio has the lowest. Additionally, the MV portfolio went from having the highest out-of-sample variance to the second-lowest.

Table 5: Out-of-sample Variance

	IV	MV	EW	ERC	HRP	HRP Topdown
$\hat{\sigma}^2_{oos}$	0.1271	0.1105	0.2180	0.1359	0.0854	0.1221

5.2.3 Summary

This analysis aimed to replicate the results of López de Prado's (2016b) out-of-sample Monte Carlo simulations. Albeit minor differences, the results are in line with his, showing that the HRP approach yields lower out-of-sample variance than the MV portfolio. Additionally, the HRP Topdown portfolio had an even lower out-of-sample variance than the original HRP, but the EW portfolio had the lowest. However, as mentioned in the analysis, López de Prado's (2016b) estimation approach had a few caveats, which resulted in an unfair measurement of the EW portfolio's out-of-sample variance. To alleviate the issue, we showed the out-of-sample variance based on daily observations instead of yearly cumulative returns. The new estimate for out-of-sample variance showed that the HRP portfolio provided the lowest out-of-sample variance, followed by the MV portfolio, which had the highest value when using López de Prado's (2016b) estimation approach. However, the two different variance methods might not be perfectly comparable. Alternatively, the shocks could have been altered to vary in each simulation, which would remove the bias of the EW portfolio, as it would no longer have the approximately same return series in each iteration. However, this is left for further research, as the main point of the analysis was to replicate López de Prado's (2016b) results.

The in-sample variance analysis showed that the MV portfolio had the lowest in-sample variance, closely followed by the HRP approach. It also highlighted how the EW and HRP Topdown portfolios were the only portfolios that failed to deliver consistently low variance after the occurrence of shocks in the period. The EW does not seek to minimize variance, and thus, the results come as no surprise. On the other hand, the HRP Topdown delivered lower in-sample variance than the HRP approach when the assets did not contain jumps but were significantly higher after. This is a result of the HRP approach's construction, as the weights of the final assets in each of the optimal clusters are distributed equally. Thus, it allocates higher weights to the assets containing the jumps. Finally, the in-sample variance highlighted that the MV portfolio has the lowest in-sample variance but not the lowest out-of-sample variance.

6 Empirical Analysis

In the following section, an empirical comparison of the different portfolios is conducted using a walk-forward analysis. While the results of the out-of-sample analysis seem promising, there is some arbitrariness in the construction of the return series. Hence, it is deemed interesting to evaluate the HRP using empirical data to assess a more concrete and practical case. The performance of different allocation methods is evaluated in terms of portfolio returns, weight concentrations, turnover, and several risk-based performance measures.

6.1 Methodology

A walk-forward analysis is a two-step process: optimation and testing. In the first step (optimization), the portfolio weights are calculated based on a subset of the data referred to as the training data. In this case, the training data is used to calculate the sample covariance matrix. In the second step (testing), the performance of the different strategies is evaluated using another subset of the data referred to as the evaluation data. (Pardo, 2011). The walk-forward analysis is usually performed on a sequence of periods, as shown in Figure 25.

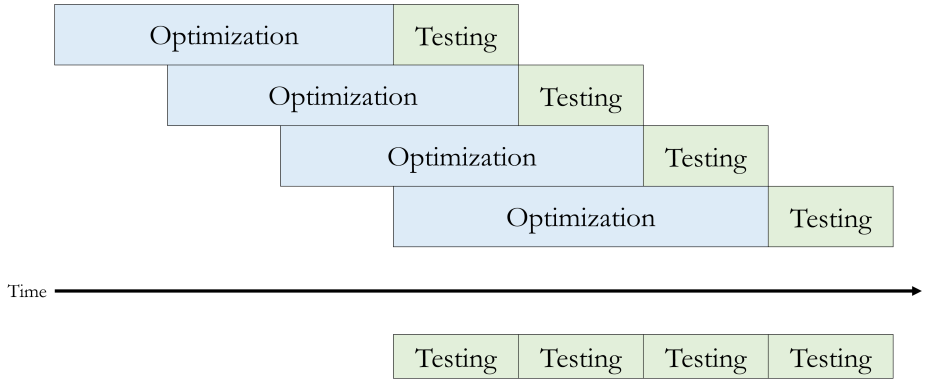

Figure 25: Illustrative diagram of the walk-forward analysis.

In practice, the dataset consists of daily return data of the 30 Industry Portfolios for the US stock market downloaded from French's Data Library, which includes Food, Beer, Smoke, Games, and others.[10] In order to capture both good and bad periods, the dataset will range from January 2000 to February 2022. The training period is set to one year, and the evaluation period is set to three months, as Pardo (2011) recommends setting the evaluation period to around 25-35% of the in-sample window. The analysis will use IV, MV, EW, ERC, HRP, and HRP Top-down allocation strategies to create six portfolios with matching names. The algorithm is set to capture the optimal weights, different out-of-sample performance measures, and the out-of-sample return series. Thus, the algorithm loops through the entire dataset and returns $\frac{266-12}{3} = 85$ consecutive out-of-sample results.

The following sections will analyze and explain the results of the walk-forward analysis, starting with a look into the structure of the sample correlation matrix, dendrogram, and the optimal number of clusters, followed by a

[10] Website: mba.tuck.dartmouth.edu

segment on the average weight allocations of the different strategies. Then the performance of the portfolios is analyzed by examining the returns and several risk measures. Lastly, the analysis is summarized with the most important findings.

6.2 Characteristics of the 30 Industry Portfolios

Most of the allocation strategies utilize the covariance matrix to determine the weight allocated to each asset. Thus, a brief look into the 30 Industry Portfolios' sample covariance and correlation matrix will provide insight into the basis of the allocations. Note that a different sample covariance matrix is calculated for each optimization period, but for practical reasons, only the sample covariance matrix of the entire period is illustrated. However, it should be sufficiently representative of the estimated sample covariance matrices. It is advantageous to plot the sample correlation matrix instead of the sample covariance matrix since it is easier to interpret and provides much of the same information as correlation is just a normalized version of covariance. Figure 26 displays the sample correlation matrix of the 30 Industry Portfolios. Overall the portfolios have a relatively high correlation with each other. However, Smoke, Coal, and Beer stick out as the portfolios with the lowest correlations. The correlation ranges between 0.23 and 0.89, with a mean value of 0.63.

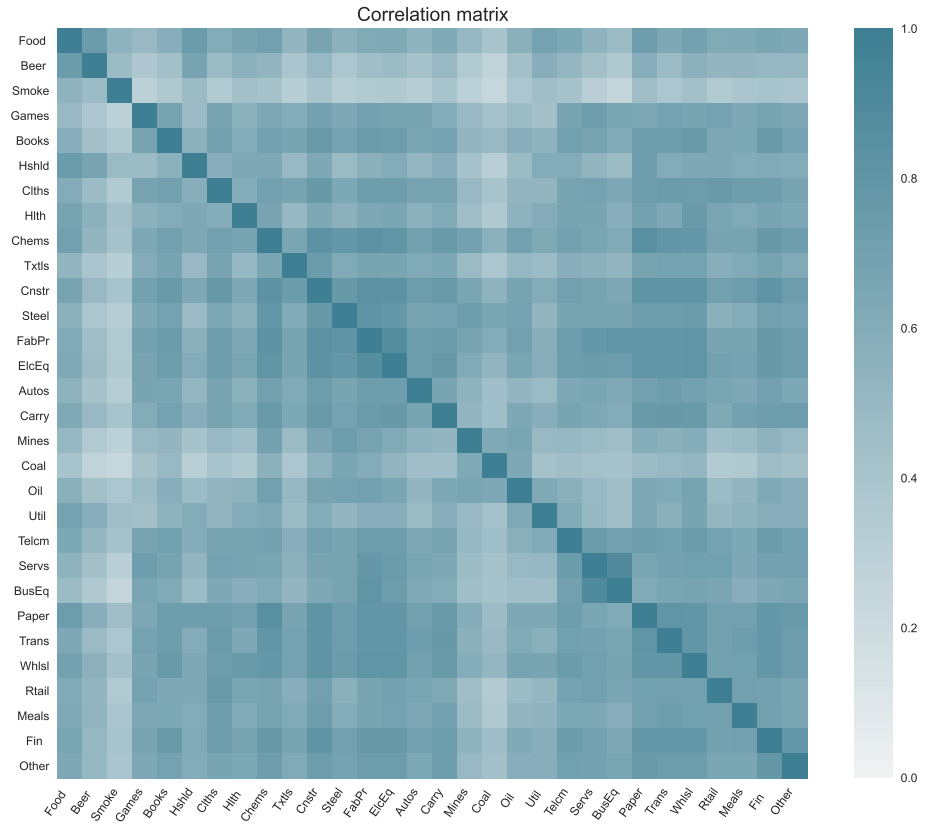

Figure 26: Illustration of the sample correlation matrix of the 30 Industry Portfolios based on daily observations from January 2000 to February 2022.

The numerical condition of the sample covariance matrix is represented by its condition number. Thus, a brief

look into the condition number of the calculated sample correlation matrices will provide insight into their numerical condition. Figure 27 illustrates the rolling and mean condition number of the 30 Industry Portfolios' correlation matrix. The sample correlation matrix's condition number is calculated for each optimization period in the walk-forward analysis as the highest eigenvalue divided by the lowest eigenvalue. The figure shows Markowitz's curse in action; as the correlation increases in crashes, the condition number also increases, which leads to unstable inverse correlation matrices. However, in these periods, the need for diversification is greatest.

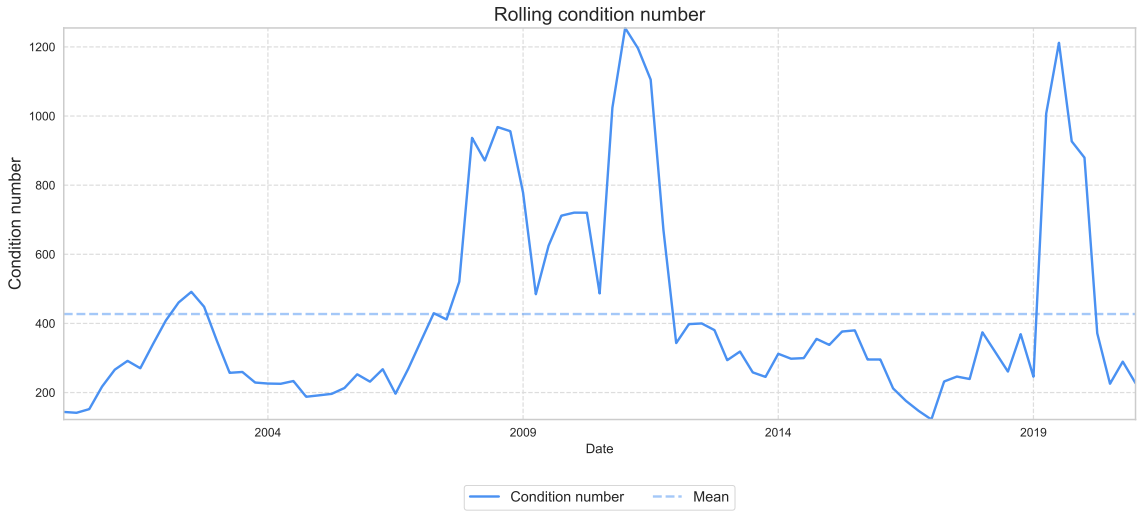

Figure 27: Illustration of the running condition number of the 30 Industry Portfolios' sample correlation matrix. For each optimization period in the walk-forward analysis, the condition number is calculated as the ratio between its maximal and minimal eigenvalue.

Clustering

The original HRP strategy relies on single linkage to cluster the assets. However, as mentioned in Section 4.1, single linkage is prone to chaining. This issue becomes apparent when applying linkage to the sample correlation-based distances of the entire period and displaying its dendrogram in Figure 28. Notice that the blue nodes in the middle of the dendrogram are sequentially joined to existing clusters, causing the chaining effect. For example, if we look at Figure 26, we see that Utilities (Util) and Business Equipment (BusEq) have a fairly low correlation. Yet, they are sequentially joined to the same large cluster by using a single linkage. Essentially, this means that Utilities and Business Equipment will compete against each other for allocations, although they are very dissimilar assets. Furthermore, this is against López de Prado's (2016b) argument that the solution of the HRP is intuitive.

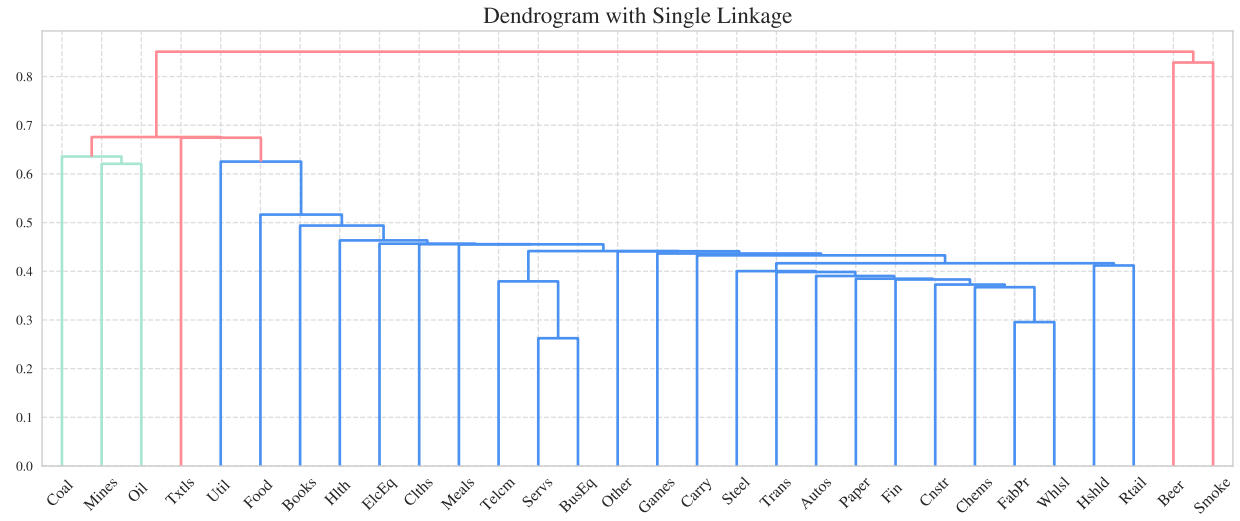

Figure 28: Dendrogram of the 30 Industry Portfolios's sample correlation-based distances using single linkage.

The original HRP method grows the tree to a maximum depth and creates singleton clusters at the lowest hierarchical level. As such, the clustering process might try to provide more information about the underlying structure than can be provided with reliability. Therefore, as mentioned in Section 4, a modified approach was developed and named HRP Topdown. It calculates the optimal number of clusters using the Elbow method discussed in Section 4.3 and displayed in Figure 29. The optimal number of clusters is calculated individually for each training period in the walk-forward analysis. However, for illustration purposes, the Elbow method is applied once for the sample correlation matrix of the entire dataset. Here, the optimal number of clusters is estimated to be nine, as indicated in the figure by the dashed black line. The actual values for the optimal number of clusters in the empirical analysis range between six and 11.

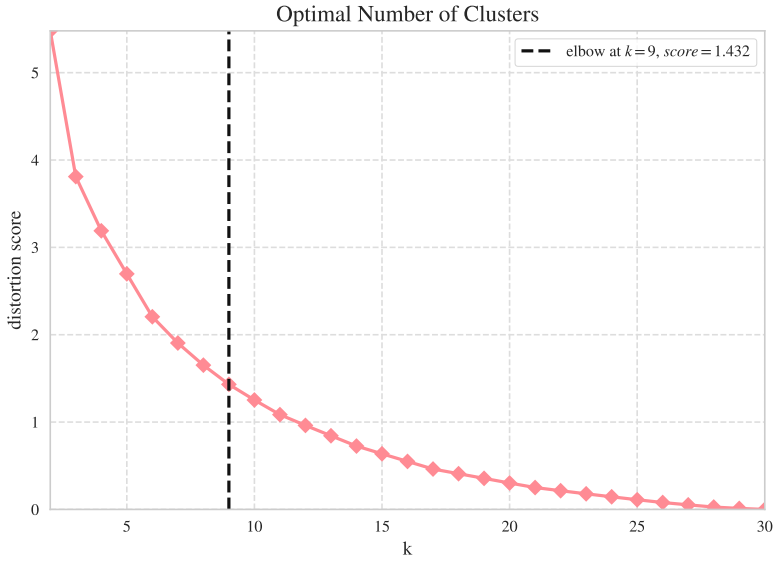

Figure 29: Illustration of the Elbow method on the 30 Industry Portfolios' sample correlation matrix.

In addition to calculating the optimal number of clusters, the HRP Topdown portfolio uses Ward's method as linkage criteria. Ward's method is applied to avoid chaining in the clustering process and create a more balanced tree. Hierarchical clustering of the 30 Industry Portfolios' sample correlation-based distances using Ward's method is shown in Figure 30. The figure illustrates that Ward's method produces a more balanced tree and avoids the chaining of clusters. The dashed black line represents the cut-off line that shows the cophenetic distance for the optimal number of clusters. It is worth noting that the Elbow method does not tell the cophenetic distance where to draw the cut-off line. However, the line can be estimated at a cophenetic distance of 0.64 since the number of clusters with this cut-off distance is nine.

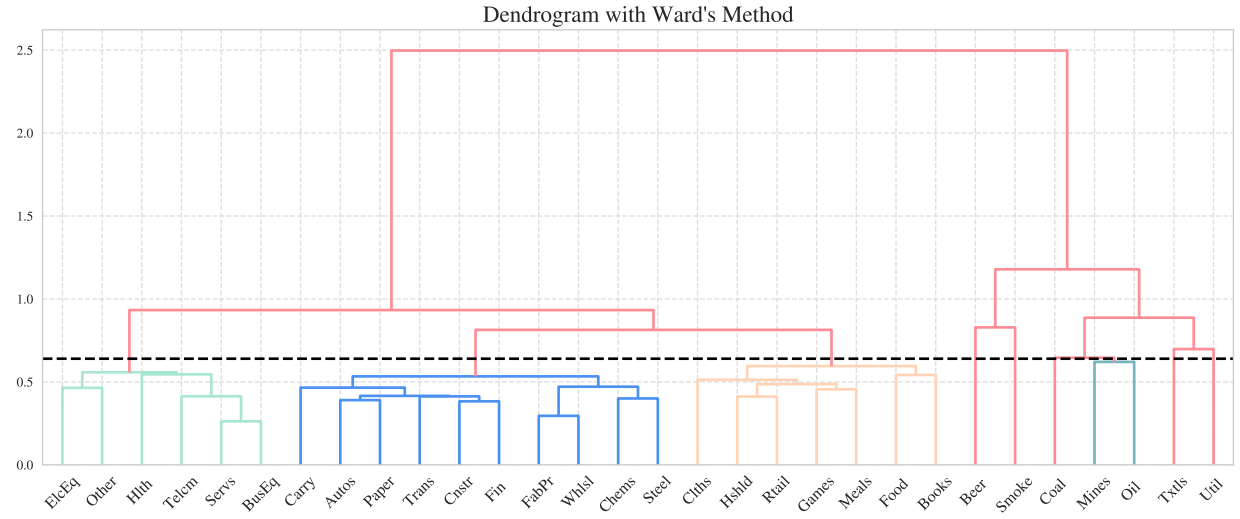

Figure 30: Illustration of the 30 Industry Portfolios' dendrogram using Wards method and with an optimal number of clusters based on the Elbow method. The red lines illustrate the nine clusters, and the dashed black line indicates the cut-off level.

6.3 Comparison Measures

In the following section, the performance and risk measures that are used in the subsequent analysis are introduced. The focus will be on risk-based performance measures. The measures included are Average Turnover per Rebalancing, Sharpe Ratio, Adjusted Sharpe Ratio, Maximum Drawdown, Sum of Squared Portfolio Weights, Skewness, Kurtosis, VaR, and CVaR.

The Average Turnover per Rebalancing

Turnover is a measure of the fluctuations in portfolio weights. While there are many different options available for calculating the turnover, in this thesis, we use the average turnover per rebalancing (ATR) applied by Raffinot (2017, 2018):

$$ATR = \frac{1}{F}\sum_{t=2}^{F}|w_{i,t} - w_{i,t-1}| \tag{22}$$

where F is the total number of rebalances during the testing period and $w_{i,t}$ is the weight of asset i at time

t. The average turnover per rebalancing acts as a proxy for the transaction costs. Incorporating transaction costs into a performance analysis can turn an ostensibly profitable strategy into unprofitable. However, it is not feasible to make an exact comparison of transaction costs as there are multiple ways to implement the strategies, and transaction costs often depend on the scale of the investments. Nonetheless, the ATR will elucidate the relative level of transaction costs required to implement different strategies, as a high turnover can be associated with high transaction costs (DeMiguel et al., 2009).

Sharpe Ratio

The *Sharpe ratio* is a common performance measure, that focuses on risk-adjusted returns. Mathematically, the Sharpe ratio, SR, of a portfolio is defined as:

$$SR = \frac{\mu - r_f}{\sigma} \tag{23}$$

where μ represents the portfolios average return, σ represents the standard deviation of the portfolios' excess return, and r_f represents the risk-free rate. However, in practice, the population's μ and σ are unknown and estimated using the sample average of returns divided by the sample standard deviation of returns. The risk-free rate is assumed to be zero, as changing the risk-free rate will not impact the performance comparison since the increase in the risk-free rate will decrease all the Sharpe ratios with equal proportions. The annualized Sharpe ratio is found by multiplying Equation 23 with $\sqrt{252}$, as the daily returns are assumed to be iid and the number of trading days in a year is 252 (consistent with the dataset).

Adjusted Sharpe Ratio

Transaction costs can highly affect the return of the portfolios. Thus, the Sharpe ratio is adjusted to account for transaction costs, estimated as a function of ATR in this thesis. In line with Lassance (2021), the transaction costs are assumed to be proportional to the ATR with a cost of 20 basis points. The ATR calculates the average turnover rate of the portfolios, and since the portfolios are rebalanced four times a year, the yearly transaction cost can be estimated as:

$$TC = 4 \cdot ATR \cdot 0.2\%$$

Thus, the Adjusted Sharpe Ratio, ASR, is given by:

$$ASR = \frac{\bar{r}^{pf}_{daily} - TC/252}{\hat{\sigma}^{pf}_{daily}} \cdot \sqrt{252} \tag{24}$$

Maximum Drawdown

Drawdown can be defined as the proportional loss from the last local maximum to the next local minimum (Leal and Mendes, 2005). To put it simply, it measures the decline of a portfolio's value from peak to trough and is typically given in absolute terms, that is:

$$Drawdown = \left| \frac{\text{Trough Value} - \text{Peak Value}}{\text{Peak Value}} \right| \tag{25}$$

Maximum drawdown, MDD, is then defined as the largest drawdown in the value of a portfolio during a specified time horizon. Maximum drawdown is an intuitive measure of tail-risk, widely used in practice, and easily measurable. However, despite its intuitiveness and simplicity, maximum drawdown has its limitations. Maximum drawdown only measures the size of the largest loss without considering the frequency or the time frame of large losses (Gray and Vogel, 2013). For example, if strategy A is assumed to have an MDD of -30% and strategy B an MDD of -40%, the maximum drawdown only informs us that strategy A is less risky. However, strategy A could also have multiple drawdowns of -20%, whereas strategy B does not have any other drawdowns. Despite this, strategy A would still outperform B in terms of MDD. On the other hand, if we assume that two strategies, C and D, have the same MDD, a maximum drawdown would indicate that these two strategies are equally risky. Yet, the drawdown of C could have taken five months, whereas the drawdown of D occurred in a week, making it objectively riskier. For these reasons, one needs to be careful when interpreting the maximum drawdown measure.

Sum of Squared Portfolio Weights

The sum of squared portfolio weights (SSPW) is used to measure the average diversification level of portfolio weights across out-of-sample forecasts. Raffinot (2017, 2018) defines the SSPW as:

$$SSPW = \frac{1}{F} \sum_{t=2}^{F} \sum_{i=1}^{N} w_{i,t}^2 \tag{26}$$

where F denotes the overall number of portfolio rebalances and $w_{i,t}$ is the weight of asset i at time t. The values of SSPW range from 0 to 1, where lower values represent less concentrated portfolios. This measure has also been used by Blume and Friend (1975) and Goetzmann and Kumar (2008) for approximating the level of portfolio diversification as a deviation from the market portfolio weights.

Skewness and Kurtosis

Skewness and Kurtosis help the investor describe and understand the distribution of the investment returns. Thus, these two measures are introduced. Skewness is a measure of the symmetry of a distribution. A normal distribution is symmetric and therefore has zero skewness. If the tail on the right side of the distribution is longer or fatter, it is a sign of positive skewness so that the mean and median will be greater than the mode. If the opposite is true, the skewness is negative.

The central moments for samples of size n are given by

$$m_r = \frac{1}{n} \sum (x_i - \bar{x})^r \tag{27}$$

Using the sample moments, the moment coefficient of skewness can be formulated as

$$g_1 = \frac{m_3}{m_2^{3/2}} \tag{28}$$

However, the sample moments are biased estimators of the population moments. Adjusting for non-asymptotic sample sizes, the sample skewness can be defined as (Joanes & Gill, 1998):[11]

$$Skewness = \frac{\sqrt{n(n-1)}}{n-2} g_1 \tag{29}$$

Most financial assets have skewed return distributions. More precisely, a negative skewness is apparent in daily stock returns since large price falls are commonly ensued by further price falls as investors try to close their positions to limit losses (Alexander, 2008). Kurtosis measures the tailedness of the distribution. Positive kurtosis indicates that a distribution is peaked, narrow, and with thick tails. The opposite is true for negative kurtosis. The kurtosis of a normal distribution is three. As kurtosis is commonly compared to the kurtosis of a normal distribution, three is often subtracted from the kurtosis. This is called the *excess kurtosis*. The moment coefficient of (excess) kurtosis is defined as:

$$g_2 = \frac{m_4}{m_2^2} - 3 \tag{30}$$

Again, the moment coefficients are biased. Removing the biases results in the following formulation for the sample excess kurtosis (Joanes and Gill, 1998):

$$Kurtosis = \frac{n-1}{(n-2)(n-3)}[(n+1)g_2 + 6] \tag{31}$$

The return distributions of most financial assets are *leptokurtic*. A leptokurtic distribution has positive excess kurtosis. Essentially, it indicates the presence of wide outliers than in the normal distribution. Alternatively, a distribution with a negative excess kurtosis is called *platykurtic* (Munk, 2018).

Value-at-Risk

Value-at-Risk is a measure of tail risk. Many investors are especially concerned about the left tail of return distributions, as it explains the potential extreme losses which may occur. If returns are assumed to be normally distributed, then by denoting the mean return of the portfolio as μ and the standard deviation of the portfolio as σ, VaR is defined as:

$$VaR = \mu + \sigma N^{-1}(\alpha) \tag{32}$$

[11] These are also the equations applied by pandas.DataFrame.skew and pandas.DataFrame.kurtosis functions used in the analysis.

where α is the confidence level and $N^{-1}(.)$ is the inverse of a cumulative normal distribution. However, the returns of the portfolios are not normally distributed, so the normal formulation of VaR is not an efficient measure of risk. Therefore, the most popular approach for calculating VaR is known as the *historical simulation* (Hull, 2012). The approach estimates the distribution of portfolio changes from a finite number of historical observations. Essentially, the VaR is found by calculating the portfolio's historical returns, shorting them in ascending order, and finally selecting the $N(= \alpha \cdot \textit{number of observations})$'th return as the VaR estimate. Thus, the approach builds on the assumption that the empirical probability distribution estimated for market variables over the immediately preceding period is a good proxy for the behavior of the market variables for the subsequent period. However, market variables' behavior is nonstationary, making the estimates of percentiles of the distribution subject to error (Hull, 2012). VaR is an attractive measure due to its simplicity. However, VaR has its caveats. It can be criticized for not including the magnitude of losses when the VaR is exceeded. In addition, VaR is not a coherent risk measure. Coherent risk measures satisfy four conditions, namely, monotonicity, translation invariance, homogeneity, and subadditivity conditions. VaR satisfies the first three conditions. However, it does not satisfy the subadditivity condition. Subadditivity means that the risk measure for two merged assets should not be greater than the sum of their individual risk measures before merging (Hull, 2012). Essentially, diversification reduces risk, but VaR does not consider diversification benefits.

Conditional Value-at-Risk

To overcome the shortcomings of Value-at-Risk, a closely related tail risk measure, the Conditional Value-at-Risk (CVaR)[12], is often used. CVaR measures the expected loss during time T conditional on the loss being greater than the VaR (Hull, 2012). When losses are assumed to be normally distributed with mean μ and standard deviation σ, CVaR with a confidence level of α is given by

$$CVaR = \mu + \sigma \frac{e^{-Y^2/2}}{\sqrt{2\pi} \cdot (1 - \alpha)} \tag{33}$$

where Y is the α'th percentile point of the standard normal distribution (Hull, 2012). However, similar to VaR, the most popular approach for estimating CVaR is the *historical simulation* approach. The approach follows the same methodology as with the historical simulation of the VaR. However, where the VaR was estimated as the N'th worst loss, the CVaR can be estimated by averaging the losses worse than VaR, the $N - 1$ worst losses. Unlike VaR, CVaR is a coherent risk measure and recognizes the benefits of diversification. However, it has the disadvantage that it is more difficult to understand than the VaR (Hull, 2012).

6.4 Results

In this section, the results of the empirical walk-forward analysis are presented. First, the weight allocations of the six different portfolios are examined. Second, the distributions of returns and the accumulative returns are evaluated. Third, various risk-based measures are used to quantify the risk level of the different portfolios. Lastly, the empirical analysis is summarized in the final section.

[12] Also referred to as the Expected Shortfall (ES).

6.4.1 Weights

The portfolios' weight allocation in the assets is the only difference between them. Thus, before a thorough performance analysis, the portfolio weights are examined. Figure 31 displays the average weight allocated to the 30 industry portfolios for each of the six portfolios' at the moment of rebalancing. The size of a bubble indicates how much weight the portfolio has allocated to the asset – the bigger the bubble, the higher the allocation. The EW, ERC, IV, and HRP portfolios have relatively equal weight allocations, whereas the HRP Topdown and MV portfolios have more concentrated weights.

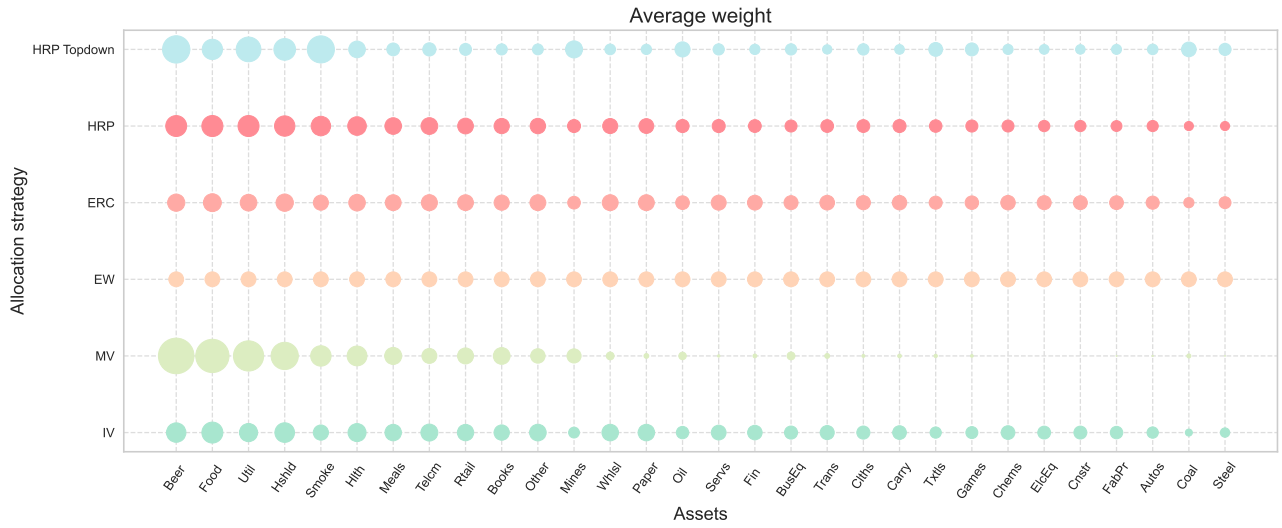

Figure 31: Illustration of the average weights allocated to the 30 industry portfolios for each of the six portfolios at the moment of rebalancing. The bubble size indicates how much weight the portfolio has allocated to the asset on average, i.e., the bigger the bubble, the higher the allocation.

Furthermore, instead of displaying the average allocation, the maximum and minimum weights at the rebalancing are shown in Table 6, along with the diversification measure SSPW. The table highlights the most extreme weight allocations and the overall diversification of each strategy. The SSPW indicates that the MV portfolio is the least diversified portfolio. Additionally, the MV portfolio has the highest single asset concentration of all the portfolios, with 58% allocated weight to a single asset in a portfolio of 30 assets. The HRP Topdown portfolio has the second-largest observed weight but a low SSPW value, indicating a diversified portfolio with a proneness for high allocations, in line with Figure 31. The original HRP has the third-largest observed weight allocation to a single asset and third-largest SSPW, indicating that it is less diversified than the IV, EW, and ERC portfolio, albeit more diversified than HRP Topdown and MV.

Table 6: Maximum, Minimum, and SSPW of Weight

	IV	MV	EW	ERC	HRP	HRP Topdown
Max	0.1028	**0.5788**	0.0333	0.0608	0.1462	0.3720
Min	0.0018	**0.0000**	0.0333	0.0080	0.0026	0.0023
SSPW	0.0417	**0.2395**	0.0333	0.0356	0.0468	0.0756

Lastly, Table 7 displays the ATR of the different allocation strategies. It highlights the potential transaction costs from portfolio turnover between rebalancing periods. The EW exhibits the lowest average turnover, followed by the ERC and the IV portfolios. The turnovers of these three portfolios are between 5.50% and 9.37%, indicating stable weight allocations. The MV portfolio has the highest turnover with an average of 43.99%, which is in line with the fact that it has the most concentrated portfolio weights. The average turnover of the HRP Topdown portfolio is 38.73%, the second-highest of the group. It is only a decrease of 4.48 percentage points from the turnover of the MV portfolio, indicating that the costs of maintaining the HRP Topdown portfolio can also be quite substantial. The original HRP, on the other hand, has an average turnover of 19.33%, less than half of the turnovers of the MV and the HRP Topdown portfolios.

Table 7: The Average Turnover per Rebalancing

	IV	MV	EW	ERC	HRP	HRP Topdown
ATR	0.0937	**0.4399**	0.0550	0.0656	0.1933	0.3873

6.4.2 Performance

Figure 32 shows how the six different portfolios would have performed if they were implemented at the start of 2000. The cumulative returns are indexed so that all portfolio values start with a value of one. The figure shows that all the strategies seem to follow each other closely. Only the MV and HRP Topdown portfolios substantially deviate from the other portfolios. The HRP Topdown portfolio seems to perform best overall but is outperformed by the MV portfolio from around 2015 to shortly after the Corona Crash[13] in march 2020. However, the HRP Topdown portfolio had a better bounce back and ended with the highest cumulative return over the period.

[13]Synonym for the stock market crash that occurred in 2020 due to the COVID-19 pandemic.

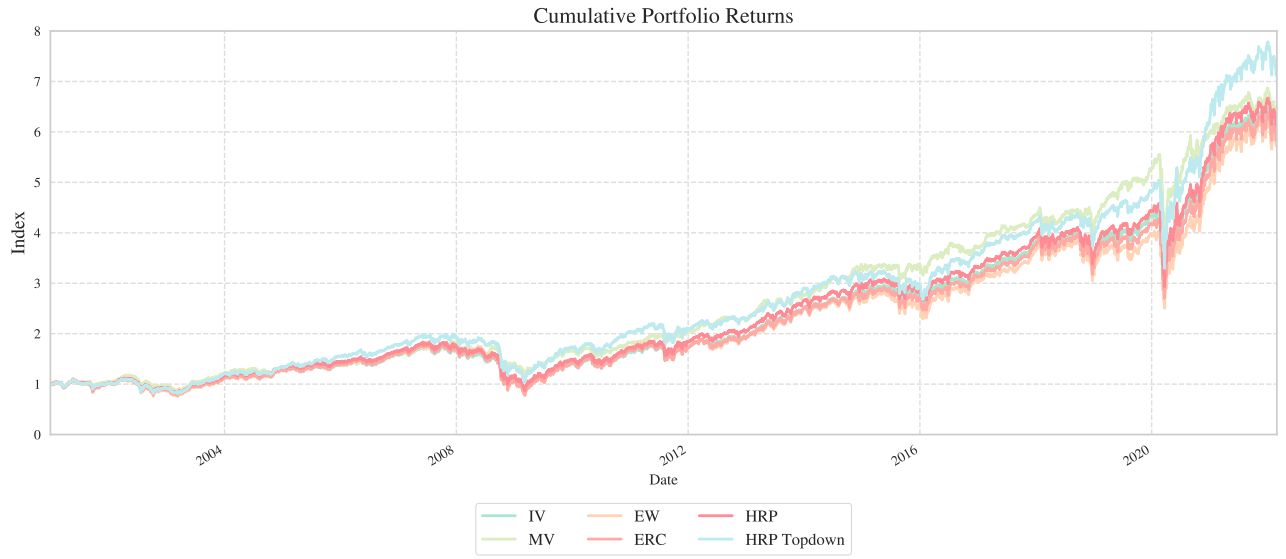

Figure 32: Cumulative returns of the six portfolios constructed in the walk-forward analysis from 2000 to 2022.

Figure 32 provides an insight into how the cumulative returns compare across the six portfolios. However, it lacks information on how the portfolios' daily returns differ from each other. To illustrate this, Figure 33 plots the distribution of the daily returns using a Gaussian kernel density estimate (KDE), along with kurtosis and skewness measures. All the portfolios have similar skewness, with MV and HRP Topdown as the largest outliers. Their negative skewness suggests that all the portfolios frequently make small gains and have a few large losses. The positive kurtosis suggests that abnormally extreme returns (outside three standard deviations) on either side of the mean occasionally occur. The combination of positive kurtosis and negative skewness implies higher risk because of the higher odds of negative outliers.

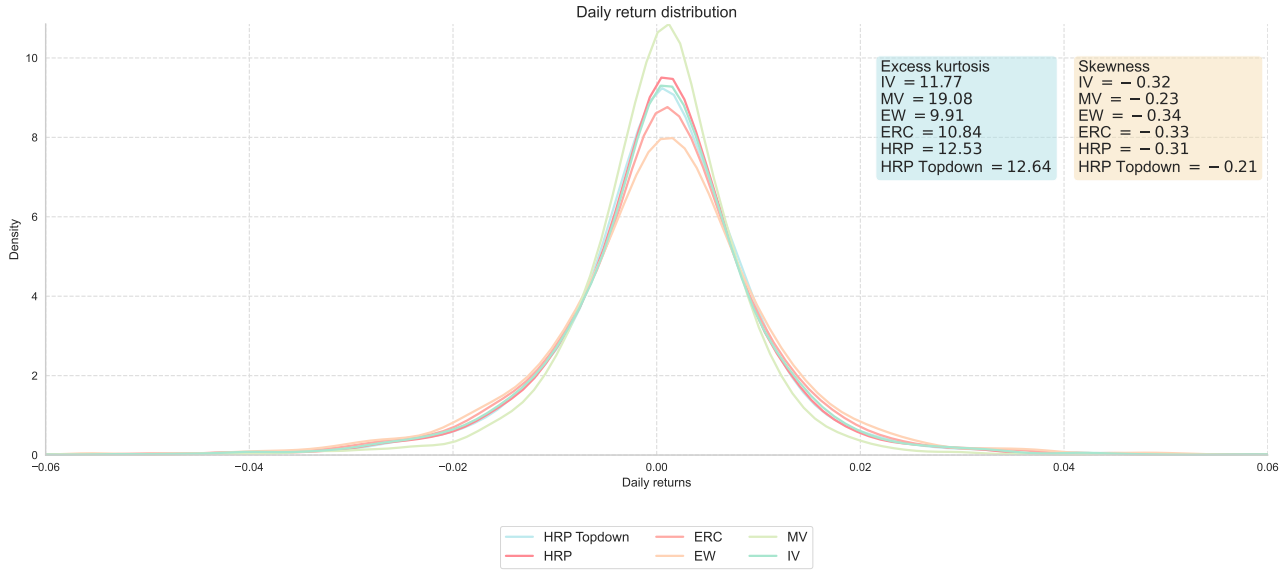

Figure 33: Illustration of six portfolios' distributions of the daily returns using a Gaussian kernel density estimate (KDE), along with kurtosis and skewness measures.

Lastly, the out-of-sample period (evaluation period) was three months in the walk-forward analysis. For each of these out-of-sample periods, the cumulative return was calculated. Table 8 provides the average, median, maximum, and minimum values of the out-of-sample cumulative return observations. The HRP Topdown portfolio had the highest average cumulative return, while the MV portfolio had the lowest. The equally weighted portfolio had the highest and lowest return observation, consistent with the fact that it is the only portfolio that does not take the variances of different assets into account and strive to allocate the portfolio weights based on a risk measure.

Table 8: Cumulative out-of-sample period return characteristics

	IV	MV	EW	ERC	HRP	HRP Topdown
Mean	0.0267	0.0259	0.0279	0.0272	0.0271	**0.0292**
Median	0.0333	0.0321	**0.0357**	0.0344	0.0301	0.0290
Max	0.2058	0.1803	**0.2228**	0.2143	0.2021	0.2124
Min	-0.2315	-0.2099	**-0.2633**	-0.2454	-0.2296	-0.1988

The last comment on the equally weighted portfolio leads us to the next part of the analysis. Returns alone are not a sufficient measure of portfolio performance, especially when it is not part of the portfolio optimization objective. Thus, the subsequent section will investigate the risk-based characteristics of the portfolios.

6.4.3 Risk Measures

There are many different risk measures, all intending to quantify the financial risk associated with the investment. As the simulation analysis mentioned and highlighted, López de Prado (2016b) uses variance as his measure for risk, illustrating that the HRP portfolio has lower out-of-sample variance than the IV portfolio. However, standard deviation can be expressed in the same units as the original values. Thus, the first risk measure is a rolling 252-day annualized standard deviation.

Figure 34 shows the rolling 252-day annualized standard deviation of the six portfolios. The EW portfolio has the highest volatility throughout the entire period, while the MV portfolio has the lowest. The other four portfolios seem to follow each other quite closely. Thus, to get a deeper understanding of the difference in the portfolios' volatility, their distributions are plotted in Figure 35.

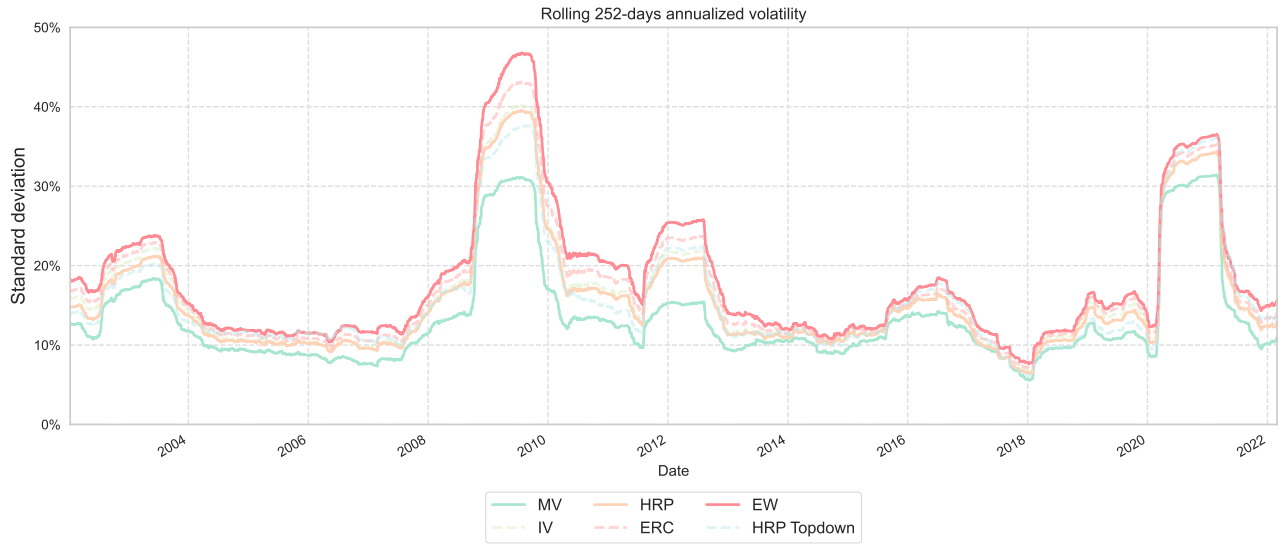

Figure 34: Illustration of the annualized rolling volatility using a 252-day window for each portfolio in the walk-forward analysis.

Figure 35 displays the annualized volatility distributions of the six portfolios. Unlike Figure 33, the distributions have long and varying tails, as all the portfolios experience periods with high volatility. The EW portfolio's volatility distribution has the longest right tail, while the MV has the shortest but a rather large hump around 30% volatility. The HRP and IV portfolios seem to have very similar volatility distributions, whereas the HRP Topdown has a slightly shorter right tail.

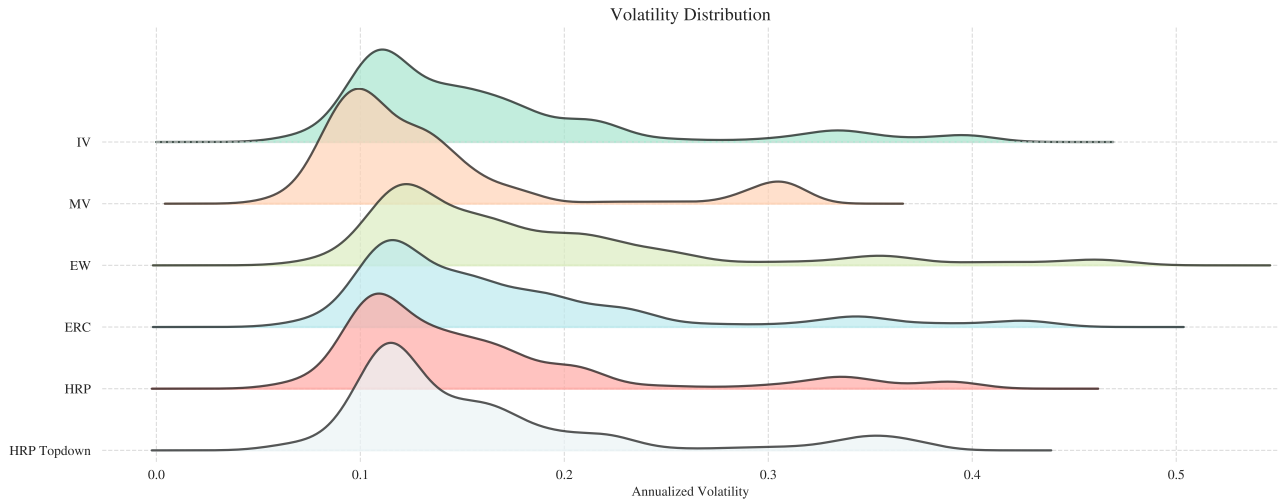

Figure 35: Illustration of the distribution of the annualized rolling volatility using a 252-day window for each portfolio in the walk-forward analysis.

Volatility is only one risk measure, and investors do not usually consider upside volatility as an issue. Thus, three different tail-risk measures are included in the analysis to evaluate the investment risk further. Table 9 displays the Sharpe ratio, adjusted Sharpe ratio, and three common risk measures. The MV portfolio has the highest

Sharpe ratio and adjusted Sharpe ratio, and the lowest value in all three risk measures. Thus, it outperforms the other portfolios on every measure in Table 9. The original HRP and HRP Topdown portfolios outperform the IV, EW, and ERC portfolios in all the measures. Moreover, the HRP Topdown slightly outperforms the original HRP in all measures except VaR. Lastly, the EW portfolio has the lowest Sharpe ratio and the highest value in all three risk measures.

Table 9: Risk measures

	Sharpe Ratio	Adjusted SR	VaR	CVaR	Max Drawdown
IV	0.5792	0.5750	0.2608	0.4290	0.5366
MV	**0.6868**	**0.6627**	**0.2050**	**0.3454**	**0.3777**
EW	0.5276	0.5254	0.2970	0.4847	0.5843
ERC	0.5543	0.5515	0.2760	0.4540	0.5624
HRP	0.5970	0.5881	0.2547	0.4193	0.5272
HRP Topdown	0.6451	0.6273	0.2566	0.4161	0.4778

6.5 Summary of the Empirical Analysis

The empirical analysis indicates that the MV and HRP Topdown portfolios seem to be the most competitive out of the six portfolios. However, the EW portfolio outperforms based on ATR and SSPW. On the other hand, the original HRP, IV, and ERC portfolios fail to outperform in any of the used performance measures. Thus, the following summary will focus on MV, HRP Topdown, and EW portfolios.

First, out of the three portfolios, the MV portfolio has the highest weight concentration and EW the lowest, while the HRP Topdown portfolio weight concentration is closer to the MV than the EW portfolio. The EW portfolio allocates without any optimization or objective, unlike the MV and HRP Topdown. Thus, the difference is not surprising. The weight allocation is the only difference between the portfolios and drives the performance differences. There are several ways to measure and compare performance. Based on Table 9, the MV portfolio performs best. It has the highest Sharpe ratio and adjusted Sharpe ratio and the lowest value in all the risk measures, closely followed by the HRP Topdown portfolio. However, as Figure 32 illustrates, it does not give the highest cumulative return since the HRP Topdown portfolio outperformed in the period after the Corona crash. Thus, an investor would have benefited from the HRP Topdown portfolio's higher period mean return, illustrated in Table 8. However, Figure 32 also indicates that the MV portfolio had a higher return until 2020, and since the investment horizon is somewhat arbitrary, the MV portfolio might appear as the best choice. Lastly, for ease of reference, all the portfolio methods included in the analysis are ranked in descending order in Table 10 for the different performance measures[14].

[14]Note that the HRP Todown is shortened to HRP TD.

Table 10: Ranking of all portfolio measures

Mean	**Sharpe Ratio**	**VaR**	**CVaR**	**Max Drawdown**	**SSPW**	**ATR**
MV	MV	MV	MV	MV	EW	EW
HRP TD	HRP TD	HRP TD	HRP TD	HRP TD	ERC	ERC
HRP	HRP	HRP	HRP	HRP	IV	IV
IV	IV	IV	IV	IV	HRP	HRP
ERC	ERC	ERC	ERC	ERC	HRP TD	HRP TD
EW	EW	EW	EW	EW	MV	MV

7 Discussion

The following section will delve into our results' meaning, importance, and relevance. The focus will be on explaining and evaluating our findings while showing how it relates to our research questions. The section is split into four subsections that follow the structure of the thesis' analysis with the addition of a discussion on the performance difference between the original HRP and HRP Topdown approach.

7.1 Estimation Error in Weight Allocation

Misspecifications of the covariance matrix should affect five of the six portfolio optimization methods since they use the covariance matrix as an input. Therefore, the estimation error analysis measured the impact of covariance matrix instability on the weights of different portfolio optimization methods. From the first part of the analysis, concerning the simple block-diagonal correlation structure, the results show that the ERC portfolio is more robust towards estimation errors in the sample covariance matrix than the other strategies. Conversely, the HRP portfolio had the second-highest RMSE, only outperforming the MV portfolio. The MV and HRP portfolios had significantly higher RMSEs than the IV, ERC, and HRP Topdown. The second part of the analysis concerning the randomized block-diagonal correlation matrix shows that only the MV and HRP Topdown portfolios were negatively affected by the more complex structure. In contrast, the IV, ERC, and HRP portfolios all improved.

The results are somewhat counterintuitive, as estimating a covariance matrix based on a more complex correlation structure should yield worse estimates than a simpler one. However, the results of the IV and the ERC portfolios are consistent with the findings of Ardia et al. (2015), who found that the ERC and IV are less sensitive toward correlation misspecifications. In addition, Sjöstrand et al. (2020) also provided empirical grounds for the ERC portfolio's resilience to estimation error when facing more complex structures. However, prior research fails to explain why this is the case thoroughly. It could be interesting to investigate this in detail, but we leave it for further research to explore.

While the findings show that the HRP approach has higher RMSE than the IV and ERC portfolios, it is still less than the MV portfolio. This finding is in line with the argument of López de Prado (2016b), who argues that the MV portfolio suffers from instability. Furthermore, the HRP is able to alleviate the instability issue since the estimation errors actually decrease with a more unstable covariance matrix. In addition, the RMSE is purely a statistical measure of robustness and, thus, does not provide any economic interpretation of the effects. As a result, the difference between portfolios RMSEs may not have any significant effect on the performance of the HRP approach in regards to improving returns or minimizing risk. Thus, the analysis provides little information on the practical implications of the HRP approach's performance.

7.2 Out-of-Sample Study

Risk is arguably one of the most fundamental performance measures within finance. Therefore, the out-of-sample analysis focuses on the risk-based performance measure variance. The analysis provides further evidence that

the HRP approach has a strong performance compared to the MV portfolio and the other risk-based allocation strategies. While the MV portfolio should produce the lowest variance in-sample, it does not necessarily yield the least risky portfolio when evaluated out-of-sample. The Monte Carlo simulations clearly show that the HRP approach delivers lower out-of-sample variance than the MV portfolio, in line with López de Prado's (2016b) findings.

However, as noted in the analysis, López de Prado's choice for estimating variance had some undesirable traits, which left the EW portfolio with the lowest variance. The original approach of López de Prado (2016b) computes the out-of-sample variance as the variance in yearly returns of each simulation, essentially measuring the variances between simulations instead of variances of returns. Therefore, the original method produces somewhat unrealistic results that ensue from the shocks added to the estimated returns. The shocks have identical sizes in all simulations, which means that the EW portfolio will capture and weigh the shocks equally in all the simulations. The equal weighting of shocks leads to lower changes in the yearly returns than the other allocation methods. Hence, in addition to López de Prado's choice of variance, the daily variance of the returns was also included in the analysis.

The findings build on existing evidence of López de Prado (2016b) by including additional risk-based allocation methods in the comparison. The out-of-sample results are consistent with the previous research when the original approach of López de Prado (2016b) is considered. Additionally, we show that the HRP portfolio leads to the lowest variance out-of-sample even when additional risk-based strategies are considered. However, when looking at the daily out-of-sample variances instead of the variances of yearly returns, the MV portfolio outperforms the IV portfolio, contrary to López de Prado's findings. Nonetheless, the HRP portfolio still outperforms all other allocation methods, conforming to the results of López de Prado. Thus, the daily variance estimates provide further evidence that the HRP approach performs well from a risk-based perspective. The results also show that filtering the covariance matrix with clustering seems beneficial since the HRP portfolio has lower out-of-sample variance than the IV portfolio, although they both use inverse-variance allocation.

The thesis also adds to the previous research by showing that the HRP portfolio has comparatively low variances in-sample, although it does not aim to optimize any objective function. The inverse-variance allocation of the HRP method is a heuristic method that corresponds to the MV portfolio when all cross-correlations are assumed to be zero. Therefore, the results seem to indicate that the HRP method is able to construct a portfolio with low variance, regardless of the strong assumption of no correlations between assets. While the in-sample variance has minor practical implications, the out-of-sample findings provide evidence that practitioners who seek to have lower variance should consider the HRP portfolio.

While the findings indicate that the HRP provides competitive results, the results' generalizability is limited by both parameters included and not included in the analysis. Practitioners should keep in mind that portfolio turnover is not considered in the study. Thus, if higher turnover comes as a corollary of the lower variance, the HRP might not be favourable to implement due to high transact ion costs. Additionally, the shocks added to the return series are somewhat unrealistic as a daily return of 200%, or -50%, is a relatively rare occurrence

in real stocks. Lastly, the simulated portfolio only consists of 10 assets. The number of assets chosen speaks in favor of the performance of the HRP portfolio. The instability of the covariance matrix will increase with the number of assets if the number of observations is held constant. The portfolios relying on the inversion of a covariance matrix should suffer from higher instability. Thus, the competitiveness of the HRP portfolio should increase with the number of assets, ceteris paribus, as it does not rely on the inversion of the covariance matrix.

7.3 Empirical Analysis

Simulations are an excellent tool for examining the performance of the HRP approach, but they lack the actuality of historical data. Therefore, the HRP approach's performance was also evaluated using empirical data through a walk-forward analysis. The walk-forward analysis measured the performance of the HRP approach using a range of different metrics. It provided evidence that the HRP portfolio is more diversified than the MV portfolio but less diversified than the IV, EW, and ERC portfolios, albeit the difference being significantly less than between the HRP and MV portfolios. Likewise, the HRP portfolio had less turnover than the MV portfolio but considerably higher than the IV, EW, and ERC portfolios. In terms of returns, the analysis did not provide evidence that the HRP portfolio had competitive performance, as it did not achieve the highest return in any measure. However, the walk-forward analysis provided further evidence that, from a risk-based perspective, the HRP portfolio performs well, exhibiting a comparatively high Sharpe ratio and low tail-risk.

While López de Prado (2016b) focused on simulations to measure the performance of the HRP approach, these results demonstrate that the HRP approach yields similar risk-based performance when applied to actual empirical data. Moreover, an analysis performed on actual data has arguably more practical implications than a simulation, as it accounts for factors that simulations might fail to replicate, such as behavioral biases. Thus, the analysis provides further evidence that practitioners who seek to reduce risk in their portfolio should consider the HRP portfolio. However, as the HRP falls short on a return-based performance, investors who seek to maximize return might consider a different approach to portfolio construction, such as the altered HRP Topdown approach. Lastly, the MV portfolio performed very well in the walk-forward analysis compared to the other two analyses and López de Prado's (2016b) findings. However, the results are consistent with Raffinot (2017), who also finds that the MV portfolio achieves better risk-adjusted performance than the HRP in all data sets used.

In practice, the applicability of the HRP method could be limited depending on the level of transaction costs. While the adjusted Sharpe ratios indicate that the penalty of higher turnover is negligible at 20bps per ATR unit, they are not representative of higher transaction costs or different transaction cost estimates. Furthermore, the high turnover of the HRP portfolio disputes López de Prado's (2016b) argument that the HRP method is stable.

To end with, the generalizability of the results is limited, as the walk-forward test only provides a single backtest. Although the walk-forward is done using a rolling-window, it still produces a single historical path. As such, it might not be representative of future performance. Furthermore, the choice of assets within the portfolio

has significant implications for the results. Firstly, evaluating a data set with more assets would exacerbate the instability of the sample covariance matrix if the number of observations used to estimate the covariance matrix is held constant. Secondly, a stronger correlation structure and more volatile assets could also affect the results as López de Prado (2016b) argues that the HRP's out-of-sample outperformance would benefit from these conditions. Lastly, the ratio between the length of the evaluation and optimization window was on the lower side of Pardo's (2011) recommendation of $25\% - 35\%$.

7.4 HRP versus HRP Topdown

Section 4 outlined the three eminent issues with López de Prado's (2016b) HRP approach. In brief, the HRP approach suffers from chaining, does not use the hierarchical structure constructed in stage 1 to allocate weights in stage 3, and does not consider the number of clusters. Therefore, an alternative approach was developed and named HRP Topdown to alleviate these issues. While no direct analysis of the performance differences was conducted, both approaches were used in our three analyses. Thus, their competitiveness can be inferred from their performance in those three analyses.

The first analysis shows that the HRP Topdown portfolio has lower sensitivity toward covariance misspecification, which suggest it is more robust than the original HRP. However, its estimation error significantly increases from the more complex correlation structure, whereas the original HRP decreases. The increase might stem from its equal distribution of the remaining weights after the optimal cluster cut-off. Regardless, the analysis is purely a statistical measure of robustness and has little economic interpretation.

The second analysis shows that the HRP Topdown portfolio delivered lower in-sample variance when the assets did not contain jumps. However, the original HRP portfolio delivered lower variance once jumps were introduced, as the HRP Topdown would allocate more weights to the assets containing jumps. In the out-of-sample performance, the HRP Topdown portfolio yielded a lower out-of-sample variance than the original HRP when the out-of-sample variance was measured on a yearly cumulative return basis. However, when the out-of-sample variance was measured by daily observations, the original HRP yielded a significantly lower out-of-sample variance than the HRP Topdown portfolio. These findings indicate that the HRP Topdown performs well when the overall variance is relatively stable. However, as it allocates higher weights to assets with higher variance than the original HRP due to its equal split of weights after the optimal number of clusters, it suffers when a few assets experience extreme jumps.

The overall impression of the empirical analysis is that the HRP Topdown portfolio performs better than the original HRP. The HRP Topdown delivered significantly higher returns, with lower variance and lower maximum drawdown but roughly twice as high ATR as the original HRP. In addition, the portfolios had similar VaR, CVaR, and excess kurtosis, but the HRP Topdown was less skewed. Finally, the HRP Topdown portfolio had a higher traditional and adjusted Sharpe Ratio than the HRP.

In conclusion, the HRP Topdown approach does not seem to deliver lower variance than the original HRP. However, it increases the return substantially and yields better risk-adjusted performance than the original

HRP. Nonetheless, higher returns and risk-adjusted performance come with a cost of higher turnover, which could limit the use of the HRP Topdown method in practice.

8 Conclusion

This thesis aims to add to the empirical evidence of the HRP performance by comparing it to more traditional risk-based portfolio allocation techniques and including a more comprehensive set of performance measures. Additionally, the thesis seeks to improve the HRP by alleviating three of its counterintuitive features.

This study finds that the HRP allocation technique performs competitively compared to the MV portfolio and other risk-based allocation strategies. The first Monte Carlo experiment shows that the HRP portfolio produces higher estimation errors than most other non-hierarchical allocation strategies regardless of the structure of the estimated covariance matrix. Thus, providing no evidence of stronger robustness than other on-hierarchical allocation strategies. However, the robustness is purely a statistical measure and has little economic interpretation, thus deemed less relevant in performance comparison. For example, The EW portfolio has no estimation errors, yet it might not have the highest return or the lowest risk. Instead, the second Monte Carlo experiment shows that the HRP portfolio delivers low in-sample and out-of-sample variance. While the MV portfolio yields the lowest in-sample variance, the HRP has lower out-of-sample variance. Lastly, the empirical analysis shows that the HRP outperforms the IV, EW, and ERC portfolios on risk-based performance measures. However, it fails to outperform the MV portfolio in this particular dataset and period.

Finally, the original HRP approach has some counterintuitive features that this study aimed to solve with a modified version called the HRP Topdown. The empirical analysis shows that the HRP Topdown significantly increases the returns and performs better from a risk-based perspective than the original HRP approach. However, the outperformance comes with a cost of higher transaction costs that could limit the practical application of the HRP Topdown method. Nonetheless, the benefit of the HRP Topdown method is that it replaces some of the counterintuitive and arbitrary choices made in the original HRP approach with rational alternatives.

9 Further Research

This study has accentuated the possibilities of using the HRP approach to improve portfolio performance and diversify using hierarchical clustering. Hopefully, this study will encourage future research on the HRP approach or, more generally, the usage of hierarchal clustering in asset allocation. There are many interesting topics within this area to research further, and the following section will highlight some of these topics.

Many alterations could be applied within the scope of hierarchical clustering. Firstly, different distance metrics could be used to evaluate different sources of financial data to find robust clusters. Secondly, the performance implications of different linkage criteria could be tested. Lastly, the hierarchical clustering methods described in this thesis are flexible and allow for multiple variations of the same ideas. For example, one could incorporate different portfolio construction methods within and between clusters and explore different combinations.

The error estimation analysis performed in Section 5.1 yielded interesting and slightly counterintuitive results, which could be explored further. In particular, the ERC improved from a more complex covariance structure. Here, further research could explore why this is the case and under which conditions the ERC portfolio suffers the most. Additionally, the effect of different covariance matrix estimators on estimation noise and performance of hierarchical clustering strategies could be explored.

This thesis's out-of-sample and empirical analysis focused on equities and a relatively small investment universe. An interesting alteration would be analyzing the HRP approach in a different investment universe that includes different asset classes or multi-factor strategies. Likewise, could the effect of increasing the size of the portfolios be investigated, as many investors prefer large diversified portfolios.

10 Appendix

A CLA

To simplify Markowitz's CLA algorithm figure 36 illustrates in four steps how his algorithm works for a three asset example. Let **a**, **b**, and **c** be three different assets with different levels of expected return and variance, where **b** has the highest return. The blue triangle is then the set of all possible allocations to **a**, **b**, and **c**, while the dashed line represents the set of all points (portfolios) with a given expected return and the yellow circle is the set of all points (portfolios) with a given variance. CLA starts at the security on the highest expected return line, which is the right corner where **b** lies. Then CLA begins adding in the next most attractive security based on the covariance matrix and the mean returns – security **a** in this case. It follows the *critical line* (green line) along **ab** until it finds a more efficient path and then it moves onto a new critical line, in this case, this occurs at point **l** along line **ab**. Then CLA follows the new *critical cine* along **lx** adding in security **c** until it finds a more efficient path or reaches the minimum possible risk portfolio – in this case this occurs at point **x** along line **lx**. The algorithm is now done, and the output is the asset weights in each corner portfolio, whereas the entire efficient frontier is composed of linear combinations of the corner portfolios.

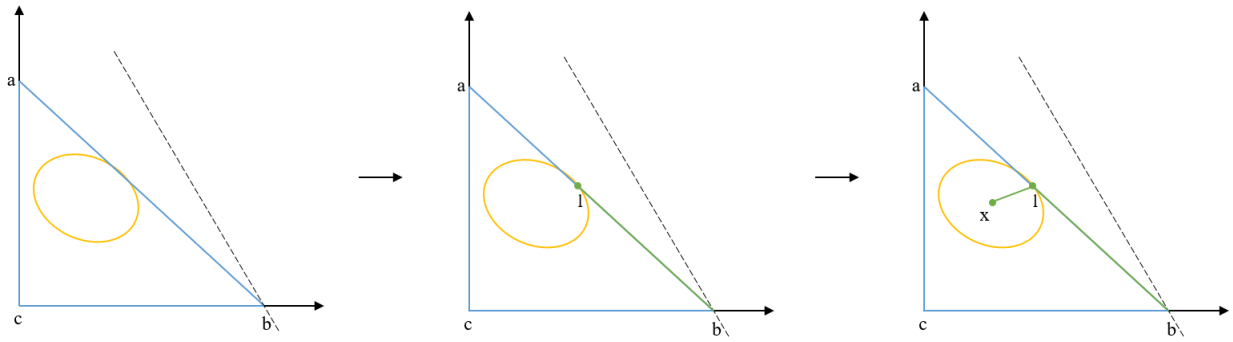

Figure 36: Illustration of CLA algorithm.

While the graph above might not look like the traditional efficient frontier, it represents the same information, and by rotating the graph 90 degrees, it becomes more evident, as shown in figure 37.

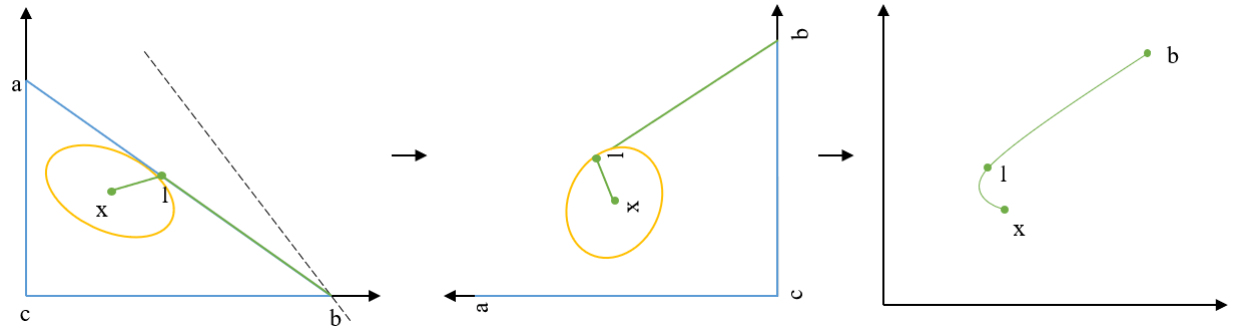

Figure 37: Illustration of how CLA converts to the efficient frontier when rotated.

In Markowitz's approach, the investor can thus use the linear combinations of the corner portfolios to create an efficient frontier, where expected return is maximized at different risk levels, and subsequently, choose the desired portfolio based on her risk preferences (Markowitz, 1959). The last graph in figure 37 illustrates Markowitz's efficient frontier, and highlights how a risk-averse investor can choose a portfolio closer to the left, whereas a risk-taker can choose something higher up the frontier, ultimately allocate her entire portfolio in asset **b**.

B Complete and Average Linkage

Complete Linkage

Complete linkage represents the opposite extreme of joining two clusters together. In complete linkage agglomerative clustering, the distance between two clusters is defined as the distance between the two most distant points in the clusters (Raffinot, 2018). For clusters C_i, C_j:

$$d_{C_i,C_j} = max_{x,y}\{D(x,y) \mid x \in C_i, y \in C_j\} \tag{34}$$

This is also called the furthest-neighbor clustering.

Whereas single linkage only requires that a single pair of objects to be close for the two clusters to merge, regardless of the similarity of the other members of the group, complete linkage requires that all the observations in the union of two clusters are relatively similar. Thus, complete linkage is prone to choosing cluster pairs whose merge has a small diameter, i.e., compact clusters. However, the method is fairly sensitive to outliers, and the produced clusters could violate the" closeness" property. Specifically, observations assigned to a cluster can be considerably closer to members of other clusters than its own cluster's members Hastie et al. (2009); Murphy (2012).

Average Linkage

Average linkage clustering uses the average of the distances between any two points in the clusters. For clusters C_i, C_j:

$$d_{C_i,C_j} = mean_{x,y}\{D(x,y) \mid x \in C_i, y \in C_j\} \tag{35}$$

Average linkage is a compromise between single and complete linkage and is considered more robust than them. Its advantage is that the method produces relatively compact clusters that are relatively far apart. However, unlike single or complete linkage, average linkage is not invariant to monotone transformations. Since average linkage averages the distances, the resulting clusters depend on the measurement scale. Conversely, single linkage and complete linkage do not alter the relative ordering of the distances. This invariance to monotone transformations is commonly used as an argument in favor of single or complete linkage over average linkage Hastie et al. (2009); Murphy (2012).

References

Alexander, C. (2008). *Market Risk Analysis, Quantitative Methods in Finance*. John Wiley & Sons, Chichester.

Alipour, E., Adolphs, C., Zaribafiyan, A., and Rounds, M. (2016). *Quantum-Inspired Hierarchical Risk Parity [White paper]*. 1QBit.

Almog, A. and Shmueli, E. (2019). Structural entropy: monitoring correlation-based networks over time with application to financial markets. *Scientific reports*, 9(1):1–13.

Amadeo, K. (2012). Stock Market Crash of 2008. *About.com*.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148.

Ardia, D., Bolliger, G., Boudt, K., and Gagnon Fleury, J.-P. (2015). The Impact of Covariance Misspecification in Risk-Based Portfolios. *Annals of Operation Research*, 254(1).

Bailey, D. H. and de Prado, M. (2013). An open-source implementation of the critical-line algorithm for portfolio optimization. *Algorithms*, 6(1):169–196.

Barziy, I. and Chlebus, M. (2020). HRP performance comparison in portfolio optimization under various codependence and distance metrics.

Blume, M. E. and Friend, I. (1975). The Asset Structure of Individual Portfolios and Some Implications for Utility Functions. *The Journal of Finance*, 30(2):585.

Braga, M. D. (2016). *Risk-Based Approaches to Asset Allocation: concepts and practical applications*. Springer.

Bruder, B. and Roncalli, T. (2012). Managing Risk Exposures Using the Risk Budgeting Approach. *SSRN Electronic Journal*.

Bun, J., Bouchaud, J. P., and Potters, M. (2017). Cleaning large correlation matrices: Tools from Random Matrix Theory. *Physics Reports*, 666:1–109.

Chopra, V. K. (1993). Improving optimization. *The Journal of Investing*, 2(3):51–59.

Dangeti, P. (2017). *Statistics for machine learning*. Packt Publishing Ltd.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Duchin, R. and Levy, H. (2009). Markowitz Versus the Talmudic Portfolio Diversification Strategies. *The Journal of Portfolio Management*, 35(2):71–74.

Goetzmann, W. N. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3):433–463.

Gray, W. R. and Vogel, J. (2013). Using Maximum Drawdowns to Capture Tail Risk. *SSRN Electronic Journal*.

Guigues, V. (2011). Sensitivity analysis and calibration of the covariance matrix for stable portfolio selection. *Computational Optimization and Applications*, 48(3):553–579.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York, New York, NY, second edition edition.

Horasanlı, M. and Fidan Kececi, N. (2007). Portfolio Selection by Using Time Varying Covariance Matrices. *Journal of Economic and Social Research*, 9:1–22.

Huang, W. (2020). Performance of Hierarchical Equal Risk Contribution Algorithm in China Market. *Available at SSRN 3695598*.

Hull, J. (2012). *Risk management and financial institutions,+ Web Site*, volume 733. John Wiley & Sons.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.

Joanes, D. N. and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):183–189.

Jothimani, D. and Bener, A. (2019). Risk Parity Models for Portfolio Optimization: A Study of the Toronto Stock Exchange. *Proceedings - 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications, Deep-ML 2019*, pages 27–32.

Kallberg, J. G. and Ziemba, W. T. (1984). Mis-specifications in portfolio selection problems. In *Risk and capital*, pages 74–87. Springer.

King, B. F. (1966). Market and Industry Factors in Stock Price Behavior on JSTOR. *The Journal of Business*, 39(1):139–190.

Kolm, P. N., Tütüncü, R., and Fabozzi, F. J. (2014). 60 Years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234(2):356–371.

Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J.-P. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 03(03):391–397.

Lassance, N. (2021). Maximizing the Out-of-Sample Sharpe Ratio. *Available at SSRN 3959708*.

Lau, A., Kolanovic, M., Lee, T., and Krishnamachari, R. (2017). Cross Asset Portfolios of Tradable Risk Premia Indices. *Global Quantitative and Derivatives Strategy, JP Morgan*.

Leal, R. P. C. and Mendes, B. V. d. M. (2005). Maximum Drawdown: Models and Applications. *The Journal of Alternative Investments*, 7(4):83–91.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2004b). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Lee, W. (2011). Risk-Based Asset Allocation: A New Answer to an Old Question? Old Question? *The Journal of Portfolio Management*, 37(4):11–28.

Lohre, H., Rother, C., and Schäfer, K. A. (2020). Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations. *Machine Learning for Asset Management: New Developments and Financial Applications*, pages 329–368.

López de Prado, M. (2016a). A robust estimator of the efficient frontier. *Available at SSRN 3469961*.

López de Prado, M. (2016b). Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4).

López de Prado, M. (2020). *Machine Learning for Asset Managers*. Cambridge University Press.

Maillard, S., Roncalli, T., and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

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

Markowitz, H. M. (1959). *Portfolio selection: Efficient Diversification of Investments*. Wiley, New York.

Markowitz, H. M. (1987). Mean-Variance Analysis in Portfolio Choice and Capital Markets Blackwell. *New York*.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*, 45:31–42.

Michaud, R. O. and Michaud, R. O. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxfrod University Press.

Miller, M. B. (2018). *Quantitative financial risk management*. John Wiley & Sons.

Molyboga, M. (2020). A Modified Hierarchical Risk Parity Framework for Portfolio Management. *The Journal of Financial Data Science*, 2(3):128–139.

Munk, C. (2018). Financial markets and investments. *Copenhagen, Denmark: Lecture notes*.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Neffelli, M. (2018). Target matrix estimators in risk-based portfolios. *Risks*, 6(4):125.

Papenbrock, J. (2011). *Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization*. PhD thesis, Karlsruher Institut für Technologie, Karlsruhe.

Pardo, R. (2011). *The evaluation and optimization of trading strategies*, volume 314. John Wiley & Sons.

Pedersen, L. H., Babu, A., and Levine, A. (2021). Enhanced Portfolio Optimization. *Financial Analysts Journal*, 77(2):124–151.

Pfitzinger, J. and Katzke, N. (2019). A constrained hierarchical risk parity algorithm with cluster-based capital allocation.

Potters, M., Bouchaud, J. P., and Laloux, L. (2005). Financial Applications of Random Matrix Theory: Old Laces and New Pieces. *Acta Physica Polonica B*, 36(9):2767–2784.

Qian, E. Y. (2005). Risk Parity Portfolios: Efficient Portfolios Through True Diversification. *Panagora Asset Management*.

Raffinot, T. (2017). Hierarchical Clustering-Based Asset Allocation. *The Journal of Portfolio Management*, 44(2):89–99.

Raffinot, T. (2018). The Hierarchical Equal Risk Contribution Portfolio. *SSRN Electronic Journal*.

Roncalli, T. (2013). Introduction to Risk Parity and Budgeting. *SSRN Electronic Journal*.

Schwendner, P., Papenbrock, J., Jaeger, M., and Krügel, S. (2021). Adaptive Seriational Risk Parity and Other Extensions for Heuristic Portfolio Construction Using Machine Learning and Graph Theory. *The Journal of Financial Data Science*, 3(4):65–83.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Sjöstrand, D., Behnejad, N., and Richter, M. (2020). *Exploration of Hierarchical Clustering in Long-Only Risk-Based Portfolio Optimization*. PhD thesis, CBS, Copenhagen.

Tan, P.-N., Steinbach, M., and Kumar, V. (2016). *Introduction to Data Mining*. Pearson Education India.

Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1):235–258.

Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244.

Zakamulin, V. (2015). A Test of Covariance-Matrix Forecasting Methods. *The Journal of Portfolio Management*, 41(3):97–108.