



DEGREE PROJECT IN MATHEMATICS,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2021

Hierarchical Clustering in Risk-Based Portfolio Construction

NATASHA NANAKORN

ELIN PALMGREN

Abstract

Following the global financial crisis, both risk-based and heuristic portfolio construction methods have received much attention from both academics and practitioners since these methods do not rely on the estimation of expected returns and as such are assumed to be more stable than Markowitz's traditional mean-variance portfolio. In 2016, López de Prado presented the Hierarchical Risk Parity (HRP), a new approach to portfolio construction which combines hierarchical clustering of assets with a heuristic risk-based allocation strategy in order to increase stability and improve out-of-sample performance. Using Monte Carlo simulations, López de Prado was able to demonstrate promising results.

This thesis attempts to evaluate HRP using walk-forward analysis and historical data from equity index and bond futures, against more realistic benchmark methods and using additional performance measures relevant to practitioners. The main conclusion is that applying hierarchical clustering to risk-based portfolio construction does indeed improve the out-of-sample return and Sharpe ratio. However, the resulting portfolio is also associated with a remarkably high turnover, which may indicate numerical instability and sensitivity to estimation errors. It is also identified that López de Prado's original HRP approach has an undesirable property and alternative approaches to HRP have consequently been developed. Compared to López de Prado's original HRP approach, these alternative approaches increase the Sharpe ratio with 10% and reduce the turnover with 60-65%. However, it should be noted that compared to more mainstream portfolios the turnover is still rather high, indicating that these alternative approaches to HRP are still somewhat unstable and sensitive to estimation errors.

Keywords

Portfolio construction, asset allocation, risk-based asset allocation, hierarchical clustering, agglomerative clustering, hierarchical risk parity, risk, volatility

Hierarkisk klustering för riskbaserad portföljallokering

Sammanfattning

Efter den globala finanskrisen har intresset för riskbaserade och heuristiska metoder för portföljallokering ökat inom såväl akademien som finansindustrin. Det ökade intresset grundar sig i att dessa metoder inte kräver estimering av förväntad avkastning och därför kan antas vara mer stabila än portföljer med grund i Markowitz moderna portföljteori. Lopéz de Prado presenterade 2016 en ny metod för portföljallokering, Hierarchical Risk Parity (HRP), som kombinerar hierarkisk klustering med en heuristisk riskbaserad portföljkonstruktion och vars syfte är att öka stabiliteten och förbättra avkastningen. Baserat på Monte Carlo-simuleringar har Lopéz de Prado lyckats påvisa lovande resultat.

Syftet med detta examensarbete är att utvärdera HRP med hjälp av walk-forward-analys och empirisk data från aktieindex- och obligationsterminer. I denna utvärdering jämförs HRP med andra vanliga portföljmetoder med avseende på prestandamått relevanta för portföljförvaltare. Den huvudsakliga slutsatsen är att tillämpning av hierarkisk klustering inom ramen för riskbaserad portföljallokering förbättrar såväl den absoluta avkastningen som Sharpekvoten. Däremot är det tydligt att vikterna i en HRP-portfölj har hög omsättning över tid, vilket kan tyda på numerisk instabilitet och hög känslighet för skattningsfel. Vidare har en oönskad egenskap i Lopéz de Prados ursprungliga HRP-metod identifierats, varför två alternativa HRP-metoder har utvecklats inom ramen för examensarbetet. Jämfört med Lopéz de Prados ursprungliga metod förbättrar de två alternativa metoderna Sharpekvoten med 10% och minskar omsättningen av portföljvikterna med 60-65%. Det bör dock understrykas att även de nya metoderna har en förhållandevis hög omsättning, vilket tyder på att numerisk instabilitet och hög känslighet för skattningsfel till viss del fortfarande kvarstår.

Nyckelord

Portföljallokering, portföljhantering, portföljmetoder, riskbaserad portföljallokering, hierarkisk klustering, agglomerativ klustering, risk, volatilitet

Acknowledgements

We would like to express our warmest gratitude to our supervisors at Lynx Asset Management, Florian Koch and Tobias Rydén, for their input, guidance and encouragement. We would also like to extend our thanks to Lynx Asset Management for giving us the opportunity to write our thesis and for providing us with data.

Lastly, we would also like to thank our supervisor at KTH Royal Institute of Technology, Boualem Djehiche, for his valuable feedback and advice.

Natasha Nanakorn & Elin Palmgren,
Stockholm, May 2021.

Contents

1	Introduction	10
1.1	Background	10
1.2	Research Aim	11
1.3	Scope & Limitations	12
1.4	Related Work	12
2	Financial Background	15
2.1	Financial Instruments	15
2.1.1	Forwards	15
2.1.2	Futures	16
2.2	Performance Measures	16
2.2.1	Sharpe Ratio	16
2.2.2	Turnover	17
2.2.3	Drawdown	17
3	Mathematical Background	18
3.1	Portfolio Construction	18
3.1.1	Mean-Variance Portfolio	18
3.1.2	Minimum Variance Portfolio	19
3.1.3	Risk Parity Portfolio	20
3.1.3.1	Risk Budgeting (RB) Portfolio	21
3.1.3.2	Equal Risk Contribution (ERC) Portfolio	22
3.1.4	Inverse-Variance Portfolio	23
3.1.5	Inverse-Volatility Portfolio	23
3.1.6	Equally Weighted (EW) Portfolio	24
3.2	Clustering	25
3.2.1	Hierarchical Clustering	26
3.2.1.1	Features	26
3.2.1.2	Distance Measures	27
3.2.1.3	Linkage Criteria	27
3.2.1.4	Dendrograms	28
3.2.1.5	Restriction on Growth	29
4	Methodology	31
4.1	Overview	31
4.2	Data	31
4.3	Modeling	32
4.4	Walk-Forward Analysis	33
4.5	Randomization of HRP	34
5	Analysis	35
5.1	Original HRP Algorithm	35
5.1.1	Overview	35
5.1.2	Algorithm	36
5.1.2.1	Tree Clustering	36
5.1.2.2	Quasi-Diagonalization	37

5.1.2.3	Recursive Bisection	39
5.1.3	Suggested Improvement	41
5.2	New Approaches to HRP	41
5.2.1	Bottom-Up Approach	42
5.2.2	Top-Down Approach	43
6	Results & Discussion	45
6.1	Evaluation of Original HRP Approach	45
6.2	Evaluation of New Approaches	49
6.3	Evaluation of New Approaches With Restriction on Growth	52
6.4	Summary	55
7	Conclusion	57
7.1	Research Aim	57
7.2	Future Work	58
A	Data	59
B	Portfolio Weights	61
C	Numeric Examples for New Approaches to HRP	63
C.1	Three Assets	63
C.1.1	Top-Down	63
C.1.2	Bottom-Up	64
C.1.3	Comparison	65
C.2	Four Assets	66
C.2.1	Top-Down	66
C.2.2	Bottom-Up	68
C.2.3	Comparison	70

List of Tables

1	Performance measures	48
2	Ranking of portfolio construction methods	49
3	Performance measures	52
4	Ranking of portfolio construction methods	52
5	Cluster descriptions	53
6	Performance measures	54
7	Performance measures	55
8	Ranking of portfolio construction methods	56
9	Descriptive metadata for each instrument in data set	59

List of Figures

1	Example dendrograms	28
2	Illustration to clarify clustering terminology	36
3	Dendrogram to explain clustering and linkage matrix	37
4	Correlation matrix before and after quasi-diagonalization	38
5	Comparison of tree structures in stage 1 and 3 of original HRP algorithm .	40
6	Sharpe ratio as a function of shrinkage coefficient	45
7	Sharpe ratio as a function of shrinkage coefficient	46
8	Share of bonds in portfolios evaluated	46
9	Equity curve	47
10	Rolling volatility	47
11	Drawdown	48
12	Equity curve	50
13	Comparison of trees constructed using single linkage and Ward's method . .	50
14	Average path length	51
15	Average silhouette score and within-cluster dissimilarity	52
16	Dendrogram with restriction on growth	53
17	Equity curve	54
18	Price series	60
19	Number of assets used to construct portfolios	60
20	Inverse-variance portfolio weights	61
21	HRP portfolio weights	61
22	Bottom-up (single) portfolio weights	62
23	Bottom-up (restricted) portfolio weights	62

Notation

N	Number of assets
n_i	Number of observations for asset i
$\mathbf{R} \in \mathbb{R}^N$	Vector of returns
$\boldsymbol{\mu} = E[\mathbf{R}] \in \mathbb{R}^N$	Expected value of returns
$\boldsymbol{\Sigma} = \text{Cov}[\mathbf{R}] \in \mathbb{R}^{N \times N}$	Covariance matrix of returns
$\mathbf{w} \in \mathbb{R}^N$	Vector of relative portfolio weights in portfolio of risky assets
$\sigma^2(\mathbf{w}) = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$	Variance of portfolio
$\sigma(\mathbf{w}) = \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$	Standard deviation or volatility of portfolio
$\sigma_i^2 = \text{Var}[R_i]$	Variance of asset i
$\sigma_i = \sqrt{\text{Var}[R_i]}$	Standard deviation or volatility of asset i
$\rho_{ij} = \frac{\text{Cov}[R_i, R_j]}{\sqrt{\text{Var}[R_i] \text{Var}[R_j]}}$	Correlation coefficient for assets i and j
r	Risk-free rate

1 Introduction

1.1 Background

Optimal portfolio construction refers to the process of efficiently allocating capital among a predefined set of assets or securities. (Maillard et al., 2010) The field of portfolio construction has been studied extensively by both academics and practitioners since the 1950s, when Markowitz first introduced his pioneering mean-variance approach to portfolio construction. (Markowitz, 1952, 1959) More than sixty years later, Markowitz’s work is still the foundation of modern portfolio theory. (Hult et al., 2010)

Of course, modern portfolio theory still revolves around Markowitz’s mean-variance approach to portfolio construction since it offers many advantages. One of the advantages commonly considered the most obvious is that the method is very intuitive and easy to understand. (Roncalli, 2013) Moreover, it is considered both elegant and powerful, given that it may be formulated as a convex optimization problem that can be solved analytically. (López de Prado, 2016)

Despite Markowitz’s work from the 1950’s still being considered the foundation of modern portfolio theory, the field of portfolio construction is still subject to extensive research. This is explained by the fact that although Markowitz’s mean-variance portfolio is optimal in theory, it may be somewhat flawed and unreliable in practice. (Kolm et al., 2010) In essence, these flaws are the result of two fundamental problems with Markowitz’s mean-variance approach to portfolio construction:

1. It requires the estimation of expected returns
2. It requires the estimation and, more importantly, the inversion of a positive-definite covariance matrix

The first problem has given rise to an upsurge in the use of risk-based optimization strategies such as risk budgeting and risk parity. (Maillard et al., 2010; Bruder and Roncalli, 2012) In particular, Markowitz’s mean-variance portfolio was subject to much criticism following the global financial crisis, as most institutional portfolios were mean-variance optimal but still demonstrated poor performance, resulting in severe losses for investors all over the world. (Roncalli, 2013) Moreover, studies have been able to demonstrate that the estimation of expected returns has a larger impact on the instability of the portfolio than does the estimation of the covariance matrix. (Merton, 1980; Jagannathan and Ma, 2003) A considerable fraction of investors has since preferred risk-based approaches to portfolio construction over methods relying on the estimation of expected returns. As a result, much research has been devoted to exploring the performance of these alternative portfolio construction methods that do not take returns into consideration. (Maillard et al., 2010)

Although these risk-based allocation strategies do not rely on the estimation of expected returns, they still require the inversion of a potentially ill-conditioned covariance matrix. In general, the covariance matrix is ill-conditioned when the assets are highly correlated, implying a high condition number. In turn, the inversion of an ill-conditioned matrix often results in an amplification of inherent estimation errors and ultimately results in a numerically unstable portfolio. In fact, studies have been able to demonstrate that the estimation errors may be large enough to offset any benefits of diversification. This phenomenon is nor-

ally referred to as Markowitz’s curse and naturally the resulting portfolio is characterized by underperformance under these conditions. (López de Prado, 2016)

As a result of the unreliability and underperformance of both Markowitz’s mean-variance portfolio and other risk-based optimization strategies, much research has been focused on strategies to improve the robustness and reduce the numerical instability of the covariance matrix. (Kolm et al., 2010) There is a large variety of different such strategies and extensions and all will not be covered in this thesis. However, some of the most studied extensions of the mean-variance approach to portfolio construction include taking into consideration transaction costs and tax effects (Stein, 2001; Litterman, 2003; Stein and Garland, 2008; Brandes et al., 2012), incorporating additional constraints into the quadratic optimization problem (Clarke et al., 2002), utilizing Bayesian priors (Black and Litterman, 1992) and increasing the stability of the inverse matrix using shrinkage (Ledoit and Wolf, 2003).

Other research has focused on exploring the performance of more heuristic approaches to portfolio construction, such as equal weighting, inverse-variance and inverse-volatility. As the names suggest, inverse-variance and inverse-volatility are both heuristic and risk-based. Despite the lack of mathematical rigor in these portfolio construction methods, there are studies that have been able to demonstrate superior performance when compared to traditional optimization strategies (see e.g. DeMiguel et al. (2009)). As the term ‘heuristic’ is used differently by different researchers, it should be noted that in this thesis all portfolio construction methods that are not based on optimization will be described as heuristic.

In 2016, López de Prado introduced a new asset allocation strategy attempting to circumvent the inherent flaws of the traditional optimization strategies. The presented Hierarchical Risk Parity (HRP) approach combines hierarchical clustering of assets with a heuristic risk-based allocation strategy, in the purpose of rendering the allocation more optimal. In brief, the approach operates in three stages:

1. Assets are clustered based on pairwise correlations to form a tree
2. Assets are rearranged according to this tree structure as to form a quasi-diagonal covariance matrix
3. Allocations are split between subsets of assets using inverse-variance allocation

In effect, HRP does not require either the forecasting of returns or the inversion of the covariance matrix. As a result, the approach is intended to address not only the instability of the constructed portfolio but also improve performance and reduce concentration. Using Monte Carlo simulations, López de Prado is able to demonstrate that HRP constructs portfolios with lower out-of-sample variance and higher out-of-sample return than both Markowitz’s minimum-variance portfolio and the original inverse-variance portfolio. (López de Prado, 2016)

1.2 Research Aim

López de Prado’s results are indeed promising but have only been produced using simulated data, compared against only two other portfolio construction methods and evaluated using only two out-of-sample performance measures. As a result, it is deemed interesting to evaluate the HRP approach using historical data, against more realistic portfolio construction methods and using additional performance measures. This thesis will also attempt to contribute with smaller improvements to López de Prado’s original HRP approach.

In summary, this thesis attempts to achieve the following:

1. Evaluate the out-of-sample performance of López de Prado’s HRP approach as measured by performance criteria relevant to practitioners and using empirical data
2. Analyze the properties of the HRP portfolio as compared to more mainstream portfolios
3. Identify and implement potential improvements to the HRP approach

1.3 Scope & Limitations

This thesis is written in cooperation with Lynx Asset Management (Lynx), a Swedish hedge fund located in Stockholm. More precisely, Lynx is a commodity trading advisor (CTA), which is commonly also referred to as a managed futures fund. As a result, Lynx mainly trades in the futures markets and underlying asset classes include fixed income, equities, bonds, and commodities. The data used for modeling and evaluating selected portfolio construction methods in this thesis will exclusively be price data from futures contracts, but from multiple markets, and with equity indices and bonds as underlying assets.

Moreover, since the HRP approach does not take into consideration expected returns, the resulting portfolios will be compared only against other risk-based portfolio construction methods. Restricted by the availability of price data, the analysis performed within the scope of this thesis will be performed on the last 30 years. All evaluation of the performance of different portfolio construction methods will be performed using a walk-forward analysis based on out-of-sample performance measures and closing prices subject to geometric rolling. In general, transaction costs are disregarded and the risk-free rate is set equal to zero.

1.4 Related Work

To begin with, it should be noted that despite HRP’s many appealing features, the existing literature on the matter is rather scarce, both in terms of academic literature that explore the mathematical properties and underlying assumptions of the method, and in terms of empirical literature that evaluate the out-of-sample performance. Burggraf and Vyas (2020) have highlighted that there is especially a lack of empirical literature on the out-of-sample performance of HRP applied to other asset classes than stocks or multi-asset. Of course, this implies that the scope of the related work is rather limited, but also that it is indeed highly relevant to further add to the literature addressing and building on the HRP approach to portfolio construction, especially applied to other asset classes, such as futures.

Many of the papers focusing on HRP evaluate how the method performs on a specific data set. For instance, Burggraf and Vyas (2020) applied HRP to a portfolio of crypto-currencies and found that HRP managed volatility and tail risk better than other risk-based allocation strategies such as inverse-volatility, minimum variance and maximum diversification. In order to further improve the robustness of the method, Burggraf and Vyas (2020) also applied shrinkage to the covariance matrix. Moreover, Lau et al. (2017) applied HRP to different cross-asset universes consisting of many tradable risk premia indices and confirmed that HRP delivered superior risk-adjusted returns compared to equal risk contribution (ERC), inverse-variance, maximum diversification and an equally weighted (EW) portfolio. Whilst

HRP and the minimum variance portfolio produced comparable risk-adjusted returns, HRP was more diversified than the minimum variance portfolio. Moreover, the same test was repeated on simulated data drawn from the historical distributions of the cross asset risk premia indices initially used and similar results were found.

Moreover, as López de Prado (2016) has pointed out, HRP is flexible in the sense that the method can be adapted to include features that are aligned with the user’s preferences. This can for instance be achieved by altering the distance measure or linkage criterion used, or by changing the portfolio construction method used both within and across clusters. Modifications of this kind have been the focus in several different papers. For example, Lohre et al. (2020) have considered two different distance measures; the distance measure used by López de Prado (2016) and a distance measure based on the lower tail dependence coefficient. The portfolio method used for allocating weights was also adjusted; instead of using inverse-variance, ERC and inverse-volatility were used. In this study, the overall objective was to test whether altering the distance measure might improve the tail risk management. This hypothesis was confirmed, but the results also suggested that the improved tail risk management comes at the cost of a higher turnover. The fact that the HRP approach implies a high out-of-sample turnover has been highlighted in other studies as well. Lohre et al. (2020) found that the turnover of the HRP portfolio was three times higher than that of traditional risk-parity strategies used as benchmarks. Burggraf and Vyas (2020) concluded that HRP and inverse-volatility behave similarly in terms of weight allocation but that HRP tend to make adjustments more frequently, resulting in a higher turnover.

HRP can also be generalized to employ other clustering methods than hierarchical clustering. For example, both Leon et al. (2017) and Duarte and De Castro (2020) have considered partitional clustering methods, i.e. clustering which creates a one-level partitioning of the objects, such as K-means, K-medoids and spectral clustering. Duarte and De Castro (2020) found that when compared to the minimum variance portfolio, the Ibovespa Index and the original HRP, partitional clustering based portfolio methods demonstrated the best average performance in terms of return and Sharpe ratio, but with slightly higher volatility, turnover and drawdown. (Duarte and De Castro, 2020) In contrast, Leon et al. (2017) found that a particular hierarchical clustering method referred to as Ward’s method outperformed all other clustering methods when evaluated using the performance measure Omega ratio.

Another subset of the literature building on López de Prado’s HRP approach consists of studies that make use of other portfolio construction methods, which clearly draw on the same basic ideas as HRP but with fundamentally different properties. One of the first such attempts to build on HRP was published in 2018 by Raffinot, who suggested a strategy referred to as Hierarchical Clustering Asset Allocation (HCAA). The alterations made to the HRP approach were twofold. On the one hand, Raffinot used different linkage criteria (simple, complete and average linkage, as well as Ward’s method) to construct the clusters and included a step where the optimal number of clusters is determined using the Gap index, instead of growing the tree to maximum depth as in the original algorithm. On the other hand, the capital allocation within and across clusters was computed using equal weighting, instead of inverse-variance. (Raffinot, 2018a)

Moreover, Raffinot (2018a) evaluated the resulting portfolios on three different data sets, which differed in terms of number of assets and composition of the universe (S&P sectors, multi-assets, and individual stocks). The study demonstrated that hierarchical clustering-based portfolios achieved statistically superior risk-adjusted returns compared to other more

traditional risk-based allocation strategies. Unfortunately, the evaluation of different linkage criteria was inconclusive. Raffinot stated as main conclusion that the estimation of the correlation matrix is the most crucial step and that the use of shrinkage should be explored further. (Raffinot, 2018a)

In a later paper, Raffinot (2018b) developed a new strategy, referred to as Hierarchical Equal Risk Contribution Portfolio (HERC), with the intention to retain the best from HRP and HCAA, respectively. Similarly to HCAA, HERC includes a selection of the optimal number of clusters based on the Gap index. However, similarly to HRP, HERC employs a top-down recursive division approach to asset allocation, but with a scaling factor based on risk contribution, allowing for other risk measures than variance only. The performances of HERC and HCAA were subsequently compared. Although HCAA was hard to beat, HERC performed better for some risk measures than others, and especially well for conditional drawdown at risk.

However, Huang (2020) found that when applying the HERC strategy, considering different risk measures, to the China stock market it was beaten by both the inverse-variance and the equally weighted portfolio. In the same paper, Huang (2020) also tried different linkage criteria and found that the single linkage criterion was especially useful when the objective was to put high weights on outliers, i.e. when it is important to separate assets which differ significantly from the rest.

Lastly, it should be noted that several papers have highlighted the fact that the tree retrieved from the hierarchical clustering in HRP is essentially not used to allocate weights to different assets; the hierarchical clustering is only performed to retrieve a re-ordered list of assets, corresponding to the leaves of the tree. The retrieved list is subsequently used in order to compute a new tree by recursively bisecting the list and allocating weights along this new tree. This recursive bisection has been raised as a disadvantage of HRP by e.g. Raffinot (2018a), Raffinot (2018b), Huang (2020), and Lohre et al. (2020). Lohre et al. (2020) even pressed on the fact that this operation is not by any means common practice within machine learning nor intuitive. Given this criticism, the model used in Lohre et al. (2020) as well as HCAA in Raffinot (2018a) and HERC in Raffinot (2018b) rely solely on the tree obtained from the hierarchical clustering in the first stage of the algorithm.

In summary, it should be noted that the literature evaluating and building on López de Prado’s HRP approach is relatively scarce. Although some studies have attempted to evaluate the HRP approach on a new and different data set, e.g. Burggraf and Vyas (2020), few have evaluated it on asset classes other than stocks or multi-asset. It is also apparent that the possibilities to make adjustments to the original HRP algorithm are numerous. The distance measure, linkage criterion and portfolio construction method used within and between clusters can be altered in different combinations. Moreover, the clustering itself can be altered by determining the optimal number of clusters beforehand or by making use of alternative clustering methods, such as partitional clustering. Another possibility is to make changes to how the covariance matrix is estimated and potentially shrunk. Overall, studies have confirmed that the HRP approach results in high risk-adjusted returns, but also observed high out-of-sample average turnover. Lastly, several studies have also highlighted the seemingly nonintuitive recursive bisection, where the tree constructed when clustering assets is in fact not used to allocate weights.

2 Financial Background

2.1 Financial Instruments

As already mentioned, the data used for modeling and evaluation of portfolio construction methods is price data for different kinds of futures contracts. Since futures are commonly described with forwards as background, both instrument types are described in this section.

2.1.1 Forwards

A forward contract is a financial derivative which specifies an agreement between two parties for delivery of an underlying asset at a specified delivery date for an agreed-upon price. (Bodie et al., 2014) The party who is long a forward contract is obliged to purchase the underlying asset on the delivery date for the agreed-upon price, whilst the party who is short a forward contract is obliged to deliver the underlying asset at the delivery date for the agreed upon-price. Moreover, forward contracts are entered into without cost and the delivery date is often referred to as the settlement day.

The proceeds between the two parties entering a forward contract can be described as follows. Assume that two parties enter the forward contract at time t with settlement day at time T and an agreed-upon price of $F_{t,m}$, where $m = T - t$. By letting the spot price of the underlying asset at time T be expressed as S_T , the profit (loss) at time T of the party who is long at the settlement day is expressed as:

$$S_T - F_{t,m}. \quad (1)$$

The profit (loss) of the party who is short at the settlement day is expressed as:

$$F_{t,m} - S_T. \quad (2)$$

The forwards price, i.e. $F_{t,m}$, varies slightly depending on whether the underlying asset has for example a dividend, dividend yield or holding costs such as storage costs associated with it. If the underlying asset does not have any dividends or holding costs, for instance a zero coupon bond, then the forwards price equals the expected spot price at maturity of the underlying asset:

$$F_{t,m} = S_t e^{r^*m}. \quad (3)$$

Worth noting is that although a forward contract technically entails the delivery of an underlying asset, the delivery rarely takes place as traders commonly close out their positions before the delivery date.

Moreover, there are two major risks associated with forward contracts, namely counter-party risk and liquidity risk. Counter-party risk relates to the possibility that one of the parties won't be able to meet their ends of the agreement, whilst liquidity risk refers to the possibility that traders will be unable to close out their positions before the delivery date. Both of these risks are largely managed by the futures markets when instead entering into a futures contract, which will be described in more detail in the next subsection. (Hull, 2015)

2.1.2 Futures

Futures contracts largely operate in the same way as forward contracts, but with standardized terms. To begin with, futures contracts operate on a marked-to-market basis and are traded on an exchange. The exchange standardizes the sort of contracts that may be traded in terms of underlying asset, contract size, settlement day and cash settlement versus physical delivery. This mitigates the liquidity risk associated with forward contracts as market participants are constrained to a limited number of contracts.

Futures contracts are also not held between two different parties who represent the different ends of the agreement, but rather between a trader and a clearinghouse. The clearinghouse becomes the seller of the contract for the trader who is long and the buyer of the contract for the trader who is short. As its role as trading partner, the clearinghouse also guarantees that the contract will not fall through, ultimately mitigating the counter-party risk.

The clearinghouse consequently carries a lot of risk and as a collateral, the clearing house requires that traders who are either long or short a futures contract deposit money into a security account commonly referred to as a margin account. Profits and losses reflecting changes in futures prices are subsequently settled daily in the margin account; a rise in futures prices benefits the traders who are long and vice versa. The process of daily settlements of the margin account is referred to as marking-to-market. This differs from forward contracts, where no money is exchanged until the settlement day. (Bodie et al., 2014; Hull, 2015)

2.2 Performance Measures

In order to evaluate and compare different portfolio construction methods, the performance of the constructed portfolios must be quantified using some specified measures. Beyond annualized return and volatility, the selected performance measures for this thesis are 2.2.1 Sharpe Ratio, 2.2.2 Turnover and 2.2.3 Drawdown.

2.2.1 Sharpe Ratio

Sharpe ratio is a common measure used to evaluate the performance of a portfolio. It is a measure of risk-adjusted return, where risk is defined as standard deviation. (Bodie et al., 2014) More precisely, the Sharpe ratio of a portfolio is defined as:

$$\text{Sharpe ratio} = \frac{\mathbf{w}^T \boldsymbol{\mu} - r}{\sigma(\mathbf{w})}. \quad (4)$$

Of course, the higher the Sharpe ratio, the better the risk-adjusted return. For portfolios holding futures exclusively, the risk-free return can be disregarded as no capital is invested when entering into the contract and thus the capital can be invested at the risk-free rate anyway. In practice, the Sharpe ratio is calculated as the sample average of returns divided by the sample standard deviation of returns, both estimated over the same time period. The annualized Sharpe ratio is determined by multiplying Equation (4) with $\sqrt{253}$, where 253 is assumed to be the number of trading days in a year.

2.2.2 Turnover

Turnover is a measure of the amount of trading necessary to implement a trading strategy. It is defined as the sum of the absolute value of trades across all available assets during the time horizon T divided by the number of assets and the number of observations:

$$\text{Turnover} = \frac{1}{n-1} \frac{1}{N} \sum_{t=2}^n \sum_{i=1}^N |w_{i,t} - w_{i,t-1}|. \quad (5)$$

In the above expression, $w_{i,t}$ is the portfolio weight in asset i at time t . A high turnover implies that the strategy will be expensive in the presence of transaction costs. (DeMiguel et al., 2009)

2.2.3 Drawdown

Drawdown is a measure that is defined as the cumulative loss from peak to trough of a portfolio's value, where both peak and trough are observed within some specified time horizon. More specifically, for a stochastic process X and a time horizon $T \in (0, \infty)$, the drawdown $D := \{D_t\}_{t \in (0, T)}$ is defined by:

$$D_t = \sup_{u \in [0, t]} X_u - X_t. \quad (6)$$

In turn, maximum drawdown is defined as the largest such decline in portfolio value during the specified time horizon T :

$$\text{Maximum drawdown} = \sup_{t \in [0, T]} \{D_t\}. \quad (7)$$

As such, maximum drawdown is a widely used indicator of downside risk and considered especially important by hedge funds and other commodity trading advisors. (Van Hemert et al., 2020) In fact, maximum drawdown grew increasingly important as an indicator of risk following the global financial crisis, where many investors found themselves caught in a so called liquidity trap and were forced to close positions although market conditions were far from favorable at the time. (Goldberg and Mahmoud, 2017) It became clear that under these circumstances, traditional measures of risk, such as volatility, are less relevant and consequently maximum drawdown provides a different perspective on understanding the risk of an investment. (Leal and Vaz de Melo Mendes, 2005)

In practice, drawdown is determined for discrete time points $t = 1, \dots, T$ and maximum drawdown is determined as the maximum such value. Commonly, the average drawdown is considered as well, which may be defined as:

$$\text{Average drawdown} = \frac{1}{T} \sum_{t=1}^T D_t. \quad (8)$$

3 Mathematical Background

3.1 Portfolio Construction

As already explained in the introduction, portfolio construction entails allocating capital among a predefined set of assets or securities. (Maillard et al., 2010) Markowitz’s pioneering work on the mean-variance approach to portfolio construction is still, more than sixty years later, the foundation of modern portfolio theory. (Markowitz, 1952; Hult et al., 2010) However, since Markowitz’s modern portfolio theory is somewhat flawed in practice, the field of portfolio construction is still subject to extensive research. (Kolm et al., 2010) In particular, the mean-variance portfolio was subject to much criticism following the global financial crisis (Roncalli, 2013) and since, a considerable fraction of investors has explored more heuristic and risk-based approaches to portfolio construction. As a result, much research has been devoted to exploring the performance of these alternative portfolio construction methods. (Maillard et al., 2010)

This section provides theory and background on Markowitz’s original mean-variance approach to portfolio construction as well as the most common risk-based and/or heuristic methods. The methods covered in this thesis are mean-variance, minimum variance, equal risk contribution (ERC), inverse-variance, inverse-volatility, and equal weighting (EW). Each method is described in both mathematical and more general terms. A special focus is placed on background, advantages, disadvantages, and how the different methods relate to one another. To avoid terminological confusion, the concepts of risk parity and risk budgeting are described as well.

3.1.1 Mean-Variance Portfolio

As described above, Markowitz’s mean-variance portfolio is commonly viewed as the foundation of modern portfolio theory. The approach is based on the idea that rational investors want to maximize the expected return for a given level of volatility or risk. Effectively, this implies selecting the portfolio weights that generate the portfolio where the expected return is maximized and the risk is minimized. (Maillard et al., 2010)

There are many advantages of Markowitz’s mean-variance approach to portfolio construction. One of the most obvious may be that the approach is very intuitive and easy to understand. (Roncalli, 2013) Moreover, it is considered both elegant and powerful, given that it may be formulated as a convex optimization problem that can be solved analytically. (López de Prado, 2016, 2019)

As already alluded to in the introduction, there are also several disadvantages associated with the mean-variance portfolio. More concretely, there are two main problems when applying Markowitz’s mean-variance approach to portfolio construction in practice. The first problem relates to noise in the input variables and the second problem relates to the method requiring the inversion of the covariance matrix. (López de Prado, 2019)

The latter is a problem when the covariance matrix is ill-conditioned, which is the case when the assets are highly correlated. (Bailey and López de Prado, 2012) The inversion of an ill-conditioned matrix magnifies the estimation errors in the input variables, i.e. small changes in the input variables will result in much larger changes in the portfolio weights. In other words, the resulting portfolio may be numerically unstable, sometimes to the extent that the benefits of diversification are offset by the estimation errors. (López de Prado, 2016) It

is also the case that mean-variance portfolios have a tendency to be highly concentrated in a small number of assets, implying limited diversification, which of course results in poor performance of the portfolio in the presence of significant estimation errors. (Maillard et al., 2010)

The formulation of the convex optimization problem is based on the assumption that the covariance matrix is positive definite. When this assumption holds, the convex optimization problem can be defined as follows for $c > 0$:

$$\text{maximize } \mathbf{w}^T \boldsymbol{\mu} - \frac{c}{2} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}. \quad (9)$$

Note that in the above expression c is the coefficient of risk aversion and reflects the investor's preferences with regard to the trade-off between maximizing return and minimizing risk. The larger the value of the constant c , the more risk avert is the investor. For a certain $c > 0$ there exists a certain combination of portfolio risk σ_p and return μ_p that is optimal. These pairs (μ_p, σ_p) make up the so called efficient frontier of mean-variance optimal portfolios. For $c > \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$, not all capital invested in the optimal portfolio of risky assets and the efficient frontier is a straight line. For $c \leq \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$ all capital is invested and the efficient frontier is a concave curve. (Hult et al., 2010)

However, above optimization problem is commonly subject to some constraint, the most frequent being that all capital has to be invested in this risky portfolio. If short-selling is allowed, the constraint $\sum_i |w_i| = 1$ is used. If short-selling is not allowed, the constraints $w_i > 0$ and $\sum_i w_i = 1$ are used. Without any constraints, the convex optimization problem in Equation (9) has the following analytical solution:

$$\mathbf{w} = \frac{1}{c} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \quad (10)$$

This solution is only defined up to some scaling factor α , which in Equation (10) is $\alpha = 1/c$. For this reason, it is possible to insert the binding budget constraint of $\mathbf{w}^T \mathbf{1} = \sum_i w_i = 1$ by simply normalizing the weights in Equation (10). The resulting normalized expression is given by Equation (12), in which the coefficient of risk aversion is cancelled out:

$$\mathbf{w}^T \mathbf{1} = \frac{1}{c} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \quad (11)$$

$$\mathbf{w}_{\text{norm}} = \frac{\mathbf{w}}{\mathbf{w}^T \mathbf{1}} = \frac{\frac{1}{c} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\frac{1}{c} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}. \quad (12)$$

3.1.2 Minimum Variance Portfolio

As mentioned briefly in the introduction, an increasing share of investors prefer portfolio construction methods whose solution is both robust and simple to compute. Since research has shown that the estimation error is much larger in the estimates of expected returns than in the estimate of the covariance matrix, (Merton, 1980; Jagannathan and Ma, 2003) the interest in risk-based allocation strategies has increased during the last decades. (Maillard et al., 2010) One such method is the minimum variance portfolio, which is based on the idea that an investor is only concerned with minimizing risk and not necessarily with maximizing returns. (Hult et al., 2010)

Of course, it can be argued that one of the main advantages with the minimum variance portfolio is indeed that it does not require the estimation of expected returns, since this should reasonably result in a more robust and numerically stable portfolio. (Maillard et al., 2010) Moreover, the minimum variance portfolio is in fact the only mean-variance efficient portfolio that does not require the estimation of expected returns, and is the portfolio on the efficient frontier with minimum variance. (Hult et al., 2010) Of course, this implies that the solution to the minimum variance portfolio is unique, up to a normalizing constant. (Maillard et al., 2010)

On the other hand, others argue that it is actually a disadvantage that the minimum variance portfolio does not take into consideration expected returns, since returns are indeed important to investors and asset managers in general. Another disadvantage is that the minimum variance portfolio is often highly concentrated in a small number of assets, since assets with low volatility are clearly favored. (Maillard et al., 2010) A study from 2017 has also shown that the minimum variance portfolio is very sensitive to errors in the estimate of the covariance matrix. (Ardia et al., 2017)

Mathematically, the minimum variance portfolio is the result of solving below convex optimization problem:

$$\text{minimize} \quad \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} \quad (13a)$$

$$\text{subject to} \quad \mathbf{w}^T \mathbf{1} = 1. \quad (13b)$$

The solution to this problem is unique and the resulting portfolio weights are given by the following expression:

$$\mathbf{w} = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}. \quad (14)$$

It should be noted that under the assumption that expected returns are equal for all assets considered for inclusion in a portfolio, the minimum variance portfolio and the mean-variance portfolio coincide. This is obvious through a comparison of above solution (Equation (14)) with that of the mean-variance problem (Equation (12)).

3.1.3 Risk Parity Portfolio

Risk parity is an example of a risk-based approach to portfolio construction and saw an upswing in popularity following the global financial crisis. The general idea is to construct a portfolio of assets where the weights are determined solely based on the risk of the assets. Since risk management became more important to investors in the aftermath of the global financial crisis, such a strategy is often considered both intuitive and interesting. (Bruder and Roncalli, 2012; Roncalli, 2013)

Although risk-based allocation strategies did not receive much attention until after the global financial crisis, the term appeared in financial literature already in 2005 and was most certainly used by practitioners before that, albeit perhaps to a more limited extent. However, it was not until 2010 when Maillard et al. presented a theoretical framework that the analytical properties of the approach were properly explored. (Maillard et al., 2010; Roncalli, 2013) Since then, several of the largest institutional investors have turned

to minimum variance or risk parity strategies and have been able to demonstrate strong performance of their portfolios. (Bruder and Roncalli, 2012)

However, it should be noted that there is some confusion with regard to the term 'risk parity'. Some use risk parity as an umbrella term encompassing several different portfolios constructed using risk budgeting techniques, including inverse-volatility and inverse-variance (e.g. López de Prado (2016)), while others use the term only to describe the portfolio construction method equal risk contribution (ERC), which is a special case of risk budgeting (e.g. Maillard et al. (2010)). In this thesis, the former definition will be used, in line with Roncalli (2013). As a result, both risk budgeting (RB) and equal risk contribution (ERC) will be presented in subsections 3.1.3.1 and 3.1.3.2. However, the inverse-volatility and inverse-variance portfolios will be treated in separate sections as they set themselves apart from ERC in being more heuristic. (Bruder and Roncalli, 2012)

3.1.3.1 Risk Budgeting (RB) Portfolio

The underlying idea behind risk budgeting is that the risk contribution from each asset in the portfolio is set equal to some risk budget set by the portfolio manager. (Bruder and Roncalli, 2012) As the name implies, the risk contribution of a particular asset i is defined as the share of the portfolio's total risk that can be attributed to that asset: (Maillard et al., 2010)

$$RC_i = w_i * \delta_{w_i} \sigma(\mathbf{w}). \quad (15)$$

It should be noted that in the above expression, volatility is used as risk measure. Volatility is indeed the most common measure of risk, but others such as Value-at-Risk may be used as well. In the above expression, $\delta_{w_i} \sigma(\mathbf{w})$ is the marginal risk contribution:

$$\delta_{w_i} \sigma(\mathbf{w}) = \frac{\delta \sigma(\mathbf{w})}{\delta w_i} = \frac{(\Sigma \mathbf{w})_i}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}} = \frac{w_i \sigma_i^2 + \sum_{j \neq i} w_j \sigma_i \sigma_j \rho_{ij}}{\sigma(\mathbf{w})}. \quad (16)$$

It should also be noted that the sum of the risk contributions for all assets in the portfolio is equal to the total volatility of the portfolio, i.e.:

$$\sum_{i=1}^N RC_i = \sum_{i=1}^N w_i \frac{(\Sigma \mathbf{w})_i}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}} = \frac{\mathbf{w}^T \Sigma \mathbf{w}}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}} = \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} = \sigma(\mathbf{w}). \quad (17)$$

Using above notation and letting b_i denote the risk budget for asset i , the RB portfolio can be expressed as below:

$$\mathbf{w} = \{\mathbf{w} \in [0, 1]^N : \mathbf{w}^T \mathbf{1} = 1, w_i * (\Sigma \mathbf{w})_i = b_i * (\mathbf{w}^T \Sigma \mathbf{w})\}. \quad (18)$$

Often, the above expression is accompanied by a set of constraints on the risk budget, namely that $b_i > 0$ and $\sum_{i=1}^N b_i = 1$. In the general case there is no analytical solution to the risk budgeting problem, but the weights can be expressed endogenously using β . Letting i denote an asset and p the portfolio of risky assets, β for asset i can be defined as:

$$\beta_i = \frac{\text{Cov}[R_i, R_p]}{\text{Var}[R_p]} = \frac{\text{Cov}[R_i, \sum_{j=1}^N w_j R_j]}{\sigma^2(\mathbf{w})} = \frac{\sum_{j=1}^N w_j \sigma_i \sigma_j \rho_{ij}}{\sigma^2(\mathbf{w})}. \quad (19)$$

The portfolio weights can subsequently be expressed as:

$$w_i = \frac{b_i/\beta_i}{\sum_{j=1}^N b_j/\beta_j}. \quad (20)$$

From above expression, it is clear that the weight allocated to each asset is proportional to its risk budget and inversely proportional to its beta. (Bruder and Roncalli, 2012)

As already alluded to in the introduction to risk parity, it is interesting to note that many other risk-based portfolios can be interpreted in terms of risk budgeting. For example, the minimum variance portfolio is an RB portfolio with risk budget for each asset set equal to its portfolio weight, i.e. $b_i = w_i$. (Raffinot, 2018a)

3.1.3.2 Equal Risk Contribution (ERC) Portfolio

When Maillard et al. presented a theoretical framework for risk parity in 2010, they actually explored the analytical properties of the equal risk contribution (ERC) portfolio. The underlying idea behind the ERC portfolio is that all assets in the portfolio should contribute equally to the overall risk of the portfolio. (Roncalli, 2013) In practice, this is achieved by either solving the risk budgeting problem in (18) with $b_i = \frac{1}{N} \forall i$, or by solving the system of equations where the risk contribution for all assets are set equal. With a binding budget constraint and not allowing for short-selling, the latter problem can be expressed as below: (Maillard et al., 2010)

$$\mathbf{w} = \{\mathbf{w} \in [0, 1]^N : \mathbf{w}^T \mathbf{1} = 1, w_i * (\Sigma \mathbf{w})_i = w_j * (\Sigma \mathbf{w})_j \forall i, j = 1, \dots, N\}. \quad (21)$$

It is clear that when the number of assets is large, so is the number of parameters in the above expression and that it is not possible to determine an explicit solution to the problem in the general case. Similarly to the general RB portfolio, the ERC portfolio weights can be expressed using β . Since $RC_i = RC_j = \frac{\sigma(\mathbf{w})}{N} \forall i, j = 1, \dots, N$, it follows that:

$$w_i = \frac{1/\beta_i}{\sum_{j=1}^N 1/\beta_j} = \frac{1/\beta_i}{N}. \quad (22)$$

Of course, this expression is still endogenous, since each asset's β depends on the portfolio as a whole. From this expression, it is clear that assets with high volatility or high correlation with other assets in the portfolio will receive less weight in the portfolio, which may be considered intuitive. (Maillard et al., 2010)

Moreover, there are two special cases that further add to the intuition and general understanding of the ERC portfolio. In the case with $N = 2$ assets and in the case with $N \geq 2$ assets but $\rho_{ij} = \rho \forall i, j = 1, \dots, N$, the ERC portfolio does not depend on the correlation of the assets and coincides with the inverse-volatility portfolio:

$$w_i = \frac{1/\sigma_i}{\sum_{j=1}^N 1/\sigma_j}. \quad (23)$$

Since the ERC portfolio is only defined up to a scaling constant, Equation (23) implies that the weight of each asset i is inversely proportional to its volatility. (Maillard et al., 2010)

As for all portfolio construction methods that do not incorporate estimates of future returns, the resulting portfolio is less sensitive to input parameters and estimation errors. (Merton, 1980; Jagannathan and Ma, 2003) Beyond this advantage and the valuable intuition behind the method, the most prominent advantage of the ERC portfolio is that it is supposed to maximize diversification of the portfolio's overall risk. This is interesting since several studies have pointed to the fact that diversification of risk can result in improved returns. (Booth and Fama, 1992; Fernholtz et al., 1998) In fact, the ERC portfolio may be thought of as a minimum variance portfolio subject to a constraint on a certain level of diversification. In 2010, Maillard et al. were able to present results implying that the volatility of the ERC portfolio is located somewhere between that of the minimum variance portfolio and that of the EW portfolio. Moreover, a study from 2017 has also shown that the ERC portfolio is rather robust to estimation errors in the estimate of the covariance matrix, at least compared to the minimum variance portfolio. (Ardia et al., 2017) Lastly, it has also been shown that the risk contribution can be a good predictor of how much each position will contribute to the portfolio's overall loss, especially in cases where the loss is large. (Qian, 2005)

3.1.4 Inverse-Variance Portfolio

Although the name of Lopéz de Prado's algorithm implies that it makes use of risk parity, it should be noted that the weights between and within clusters are scaled using inverse-variance. (Lopéz de Prado, 2016) The idea of the heuristic inverse-variance allocation is to limit the risk of the portfolio by allocating small weights to assets with large variance and large weights to assets with small variance. Not only does the inverse-variance allocation not require the estimation of future returns, but it also does not require the estimation of pairwise correlations between assets. As such, making use of heuristic approaches to portfolio construction is in general a simple way of circumventing the issue of high concentration and general instability in optimization-based methods such as mean-variance and minimum variance.

The portfolio weights of the inverse-variance portfolio are given by the following expression:

$$w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^N 1/\sigma_j^2}. \quad (24)$$

From above expression it is also clear that the inverse-variance portfolio is mean-variance optimal only under the assumption that all assets are uncorrelated and have equal expected returns. Similarly, the inverse-variance portfolio and the minimum variance portfolio are identical under the assumption that all assets are uncorrelated. (Hult et al., 2010) Moreover, it is interesting to note that the inverse-variance portfolio may also be considered as an RB portfolio with the following risk budgets: (Raffinot, 2018a)

$$b_i = w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^N 1/\sigma_j^2}. \quad (25)$$

3.1.5 Inverse-Volatility Portfolio

As the name implies, the inverse-volatility portfolio is very similar to the inverse-variance portfolio. The only difference is that emphasis is placed on volatility rather than variance.

As a result, the portfolio weights are given by the following expression: (Shimizu and Shiohama, 2020)

$$w_i = \frac{1/\sigma_i}{\sum_{j=1}^N 1/\sigma_j}. \quad (26)$$

The inverse-volatility portfolio is mean-variance optimal under the assumption that all assets are uncorrelated and have equal risk-adjusted returns, rather than equal returns as assumed by the inverse-variance portfolio. As already mentioned, the inverse-volatility portfolio also coincides with the ERC portfolio under the assumption that all assets have equal correlations. As for many portfolio construction methods, there exists some confusion with regard to terminology. For example, Leote de Carvalho et al. (2012) refer to the inverse-volatility portfolio as the equal-risk budget portfolio. It should also be noted that it could be argued for that inverse-volatility is less heuristic than inverse-variance since the assumption that all assets have equal risk-adjusted returns is more established than the assumption that all assets have equal expected returns. (Sharpe, 1964)

3.1.6 Equally Weighted (EW) Portfolio

The EW portfolio is probably the most basic example of a heuristic portfolio construction method and increased in popularity after the collapse of the internet bubble. (Roncagli, 2013) The idea is to allocate the same weights to all assets considered for inclusion in a portfolio and is therefore commonly also referred to as the naive portfolio or the 1/N portfolio. (Maillard et al., 2010)

$$w_i = \frac{1}{N} \quad \forall i = 1, \dots, N. \quad (27)$$

The advantages of this approach mainly include the advantages common to all heuristic methods. However, it may be considered a disadvantage that the EW portfolio does not incorporate any active bets on particular assets. In cases where the assets are associated with very different risk profiles, the strategy will also imply limited diversification since a large share of the portfolio's overall risk will be attributable to a limited number of high-volatility assets. (Maillard et al., 2010)

Despite these supposed disadvantages, the EW portfolio is actually rather widely used in practice (Bernartzi and Thaler, 2001; Windcliff and Boyle, 2004) and DeMiguel et al. (2009) have even been able to demonstrate that no mainstream portfolio construction methods are able to consistently outperform the EW portfolio.

Lastly, it should be noted that the EW portfolio is mean-variance optimal under the assumption that all assets have identical means, variances, and correlations. More specifically, the EW portfolio is the unique portfolio on the efficient frontier under these conditions, i.e. coincides with the minimum variance portfolio. (Hult et al., 2010) The EW portfolio may also be viewed as a risk budgeting portfolio with risk budgets $b_i = \frac{\beta_i}{N}$.

3.2 Clustering

Cluster analysis is a form of unsupervised learning, which in turn is a class of machine learning approaches in which no examples of previously solved tasks, commonly known as training data, are provided in order to assist the algorithm in solving the task. Unsupervised learning can be contrasted against supervised learning, in which training data is provided and used in order to infer statistical properties. (Hastie and Tibshirani, 2001) One practical implication of the inherent differences between supervised and unsupervised learning is that with supervised learning there exists a clear measure of success. More specifically, validation data can be used to determine how well the model's output compares to the true values. The measure of success can be used to judge the accuracy of a model and act as a basis for comparison between different models. For unsupervised learning, on the other hand, it is difficult to verify the output of the model since the true values are ambiguous. Of course, this makes it even more difficult to determine the quality of a model and, subsequently, to compare different models. Heuristic methods are consequently used for determining the quality of the results. (Hastie and Tibshirani, 2001)

When it comes to classification, the distinction is drawn between unsupervised and supervised classification, where unsupervised classification is more commonly referred to as clustering. In supervised classification, the algorithm is provided with training data consisting of labelled objects and the task is to label new objects, drawing upon what the algorithm has learnt from the training data. In the case of clustering, the task is to directly label a given set of objects, where the cluster assertions are solely data driven and thereby not based on any prior information. As a result, clustering commonly entails both identifying useful clusters as well as labelling the objects with respect to the clusters that have been identified. Clustering can thus be especially useful whenever there is a need of grouping objects but little prior information about the data exists. (Jain et al., 1999)

More specifically, clustering is the process of grouping a collection of objects into clusters based on a select set of features associated with those objects. Clustering can be used for different purposes, but the general aim is to divide objects into clusters in such a way that similar objects are placed within the same cluster whilst dissimilar objects are placed in different clusters. The degree of similarity between two objects is determined with regard to the objects' features and a formal definition of similarity or dissimilarity. The dissimilarity between two objects is often a distance measure such as the squared Euclidean distance between the two objects' feature vectors. For the remainder of this thesis, it is assumed that a formal definition of dissimilarity, rather than similarity, is used and that this is indeed a distance measure. The extent to which a distance measure is appropriate is largely dependent on the data and the goal. For example, it may make less sense to use the squared Euclidean distance as a measure of dissimilarity for non-quantitative features. (Hastie and Tibshirani, 2001)

Moreover, clustering can be divided into two main classes - partitional clustering and hierarchical clustering. Partitional clustering creates a one-level partitioning of the objects and the most common methods are k-means and k-medoids. (Jain et al., 1999) Hierarchical clustering, on the other hand, creates a hierarchy of multiple binary partitions. Since hierarchical clustering is one of the main elements of this thesis, the next section is dedicated to a more elaborate description of the topic.

3.2.1 Hierarchical Clustering

Hierarchical clustering classifies data into nested binary clusters, resulting in a hierarchical structure or tree. One of the advantages of hierarchical clustering is that it takes into account the fact that complex systems, such as financial markets, can exhibit clear hierarchical structures which strongly affect the dynamics of complex systems. (Simon, 1962) Hierarchical clustering provides a quantitative description of these hierarchies, which can contribute to a more adequate modeling of a system. (Raffinot, 2018a)

However, hierarchical clustering requires that the objects to be classified have numerical measurements on a set of features. As a result, each object has its own feature vector, which jointly constitute the feature matrix, where each row corresponds to an object and each column corresponds to a feature. The analysis is subsequently performed on the rows of this matrix. These features are used to compute the distances between the objects, which in turn serve as the basis for the clustering - objects with similar features are clustered together and objects with dissimilar features are placed in different clusters. (Murtagh and Contreras, 2017) In the setting of portfolio construction, the correlation matrix can for instance serve as the basis of the analysis, where each asset represents an object and features can be defined as a function of the pair-wise correlations between the assets.

Moreover, hierarchical clustering can be described in terms of two main steps: (1) structuring the data to represent the objects' distance to each other and (2) conducting the clustering and introducing hierarchy. The first step entails deciding on a suitable distance measure and evaluating the data based on that measure. In practice, this involves computing the distances between feature vectors. The second step entails forming the clusters based on the distances determined in the first step and some linkage criterion, which defines the distance between clusters. (Murtagh and Contreras, 2017)

Lastly, it is worth mentioning that there exists two main types of hierarchical clustering - agglomerative and divisive clustering. Agglomerative clustering is commonly described as a bottom-up approach and divisive clustering as a top-down approach. (Jain et al., 1999) This thesis will only treat agglomerative clustering, which starts with every object representing a singleton cluster. At each iteration, two clusters are combined, resulting in the number of clusters being reduced by one for each iteration until only one cluster remains. In the final iteration, the root of the tree is formed. Which clusters to combine at each iteration is determined by the distances between the clusters present in that iteration. In general, the two clusters with minimum distance between them are combined into one new cluster. However, different linkage criteria define the distances between clusters differently, and consequently the use of different linkage criteria can generate very different trees.

In the following subsections, 3.2.1.1, 3.2.1.2 and 3.2.1.3, the features, distance measures and linkage criteria reviewed in this thesis are described. Subsection 3.2.1.4 covers the dendrogram, which is a type of tree diagram commonly used to graphically represent hierarchical clustering, and subsection 3.2.1.5 treats the topic of restricting the growth of the tree.

3.2.1.1 Features

Hierarchical clustering is based on the features of the objects to be clustered. The feature used in Lopéz de Prado's HRP is based on the correlation matrix and is commonly referred to as the correlation distance index. It is defined as a matrix $D \in \mathbb{R}^{N \times N}$, where each element is defined as below for $i, j = 1, \dots, N$:

$$d(x_i, x_j) = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}. \quad (28)$$

3.2.1.2 Distance Measures

In order to perform the clustering, a way to measure the dissimilarity between clusters is needed. As mentioned previously, this is often measured by a distance measure and in this thesis assumed to be a distance measure. In general, a distance measure $\tilde{d} : B \rightarrow [0, \infty)$ between any pair of objects x_i and x_j satisfies the following properties: (Murtagh and Contreras, 2017)

- Symmetry: $\tilde{d}(x_i, x_j) = \tilde{d}(x_j, x_i)$
- Non-negativity: $\tilde{d}(x_i, x_j) \geq 0$
- Triangular inequality: $\tilde{d}(x_i, x_j) \leq \tilde{d}(x_i, x_k) + \tilde{d}(x_k, x_j)$
- $\tilde{d}(x_i, x_j) = 0$ if and only if $i = j$

It should be noted that it is only the triangular inequality that separates a distance measure from a dissimilarity measure. Moreover, using this notation, $B = \{(x_i, x_j) \mid i, j = 1, \dots, N\}$ and N is the number of objects considered. In López de Prado's original HRP approach, the Euclidean distance is used as distance measure. It is calculated between any two columns of the matrix D of features to obtain $\tilde{D} \in \mathbb{R}^{N \times N}$, where each element is defined as below for $i, j = 1, \dots, N$:

$$\tilde{d}(x_i, x_j) = \sqrt{\sum_{n=1}^N (d_{n,i} - d_{n,j})^2}. \quad (29)$$

3.2.1.3 Linkage Criteria

The linkage criterion determines the distance between clusters. Two different linkage criteria are reviewed in this thesis, namely single linkage and Ward's method. For below definitions, let $x \in C_i$ and $y \in C_j$ be objects contained in clusters C_i and C_j , respectively, and $\tilde{d}(x, y)$ be the distance between the two objects x and y , which in this thesis is taken to be the distance defined in Equation (29).

Single Linkage

The single linkage criterion defines the distance between two clusters as the minimum distance between any two objects in the clusters. The distance between clusters C_i and C_j is defined as:

$$d_{C_i, C_j} = \min\{\tilde{d}(x, y) \mid x \in C_i, y \in C_j\}. \quad (30)$$

This linkage criterion is sensitive to outliers and can lead to long clusters and ultimately an unbalanced tree, which is a well-known problem commonly referred to as chaining. (Lohre et al., 2020) This is intuitive as clustering occurs whenever two objects belonging to different clusters are close to each other, regardless of the distances between other objects in the two

clusters. Moreover, the single linkage criterion is intrinsically related to the correlation-network method Minimum Spanning Tree; clusters formed using single linkage are namely subgraphs of the minimum spanning tree of a data set. (Jain et al., 1999)

Ward's Method

Unlike single linkage, Ward's method is based on the centers of the clusters rather than a graph representation. (Murtagh and Contreras, 2017) The distance between two clusters is defined as the increase in the sum of the squared error when two clusters are merged. If m_i and m_j are the number of objects in clusters C_i and C_j , respectively, and \mathbf{c}_i and \mathbf{c}_j are the centers of the feature vectors associated with the objects in each cluster, then the distance between clusters C_i and C_j is defined as:

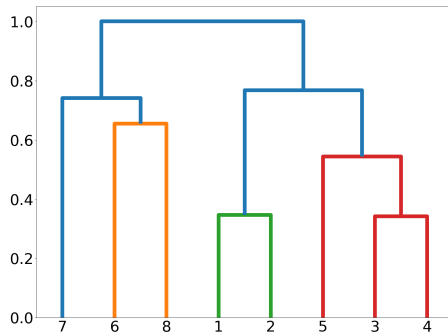
$$d_{C_i, C_j} = \frac{m_i m_j}{m_i + m_j} |\mathbf{c}_i - \mathbf{c}_j|^2. \quad (31)$$

The center of a set of vectors is defined as the average of these vectors. Note that Ward's method is independent of the choice of distance measure \tilde{d} . Focusing on the center of clusters rather than individual objects, Ward's method is less sensitive to outliers. Moreover, Ward's method aims to find compact clusters with similar number of objects. (Lohre et al., 2020)

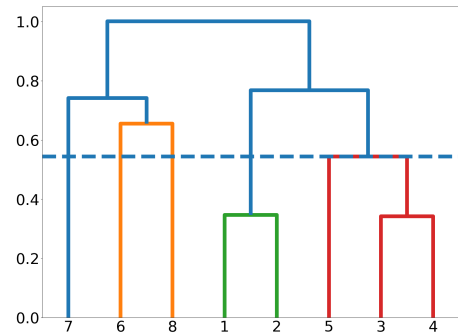
3.2.1.4 Dendrograms

As already alluded to, the hierarchical structure resulting from hierarchical clustering takes the shape of a tree. In turn, the resulting tree can be visualized using a dendrogram, which is the illustration of a binary tree where each partition is associated with a level of hierarchy. An example of such a dendrogram is presented in Figure 1 (a).

In a dendrogram, the y-axis represents the distance between clusters. This distance is defined by the linkage criterion and referred to as the cophenetic distance. (Raffinot, 2018a) Each level of the dendrogram represents a certain clustering, i.e. a merge of two clusters. As such, the hierarchical structure visualized in a dendrogram represents an ordering of the merges of clusters. The first merge of clusters is performed for the two clusters that are the most similar, and thus the closest to one another, and is consequently situated the furthest down in the dendrogram. The final merge of clusters corresponds to the root of the tree and is situated furthest up in the dendrogram. (Hastie and Tibshirani, 2001)



(a) Example of dendrogram



(b) Example of dendrogram with a cut-off

Figure 1: Example dendrograms with 8 objects, with and without cut-off line.

3.2.1.5 Restriction on Growth

Although agglomerative clustering has many desirable properties, it should be noted that it imposes a binary hierarchical structure regardless of whether such a structure actually exists in the data. The implication may be that the clustering provides more information about the underlying structure than what can be provided with reliability and can in a sense resemble a form of ‘overfitting’, which in turn can cause instability.

In essence, this issue is the result of agglomerative clustering not taking into consideration the optimal number of clusters. As such, the issue can be amended by imposing a restriction on the growth of the tree, which implies ‘cutting’ the dendrogram at a certain cophenetic distance, neglecting all clusters formed below this distance and viewing all assets in the neglected clusters as part of the same cluster. As a result, selecting the cophenetic distance for which the tree should be cut is equivalent to selecting the optimal number of clusters to have at the lowest hierarchical level of the tree. It should be noted that in this setting, ‘cluster’ refers to a group of objects and not a merge of two clusters. Figure 1 (b) shows a dendrogram with a cut-off line at a cophenetic distance of approximately 0.55. In Figure 1 (b), objects 7, 6 and 8 are singleton clusters, objects 1 and 2 belong to the same cluster, and objects 5, 3 and 4 belong to the same cluster. The number of clusters with this cut-off distance is consequently five.

Of course, the key question is where to draw the cut-off line. By considering the clusters formed at each cophenetic distance corresponding to a merger and evaluating those with respect to certain scoring functions which aim to judge the quality of the clustering, an optimal cophenetic distance to cut the tree can be determined and, ultimately, an optimal number of clusters. In practice, however, an integral part of the process of deciding the optimal number of clusters is qualitative judgement by subject-matter experts. As such, the selection of the optimal number of clusters is a holistic evaluation. Below follows definitions of two of the most common scoring functions and a measure of how balanced a tree is.

Silhouette scores

The silhouette score is a measure of how similar an object is to other objects within its own cluster compared to objects in the closest neighboring cluster. As such, each object receives one silhouette score. For object i , let a_i be the average distance between object i and other objects in its cluster and let b_i be the average distance between object i and other objects in the closest neighboring cluster to object i . (Tibshirani et al., 2001) The silhouette score for object i can then be defined as:

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \in [-1, 1]. \quad (32)$$

An average silhouette score of value 0 for a cluster can be interpreted as an indication that the cluster under consideration overlaps with its closest neighboring cluster. Since each object has a silhouette score, the score can also be used to reallocate specific objects to better clusters. (Kaufman and Rousseeuw, 1990)

Lastly, it should be noted that silhouette scores are not defined in the two cases where there exists only one cluster containing all objects (because then b_i is not well-defined) or where there exists singleton clusters (because then a_i is not well-defined). (Kaufman and Rousseeuw, 1990) As a result, Kaufman and Rousseeuw (1990), who developed the silhouette score, suggest to set $s_i = 0$ for singleton clusters whilst the silhouette score for the case where there exists only one cluster remains undefined.

Elbow method

The elbow method is a heuristic used to determine the appropriate number of clusters. The method consists of plotting the within-cluster dissimilarity W_K as a function of the number of clusters K . The within-cluster dissimilarity is defined as:

$$W_K = \sum_{k=1}^K \sum_{\mathbf{x} \in C_k} (\mathbf{x} - \mathbf{c}_k)^T (\mathbf{x} - \mathbf{c}_k). \quad (33)$$

In the above expression, C_k is the k :th cluster, \mathbf{c}_k is the center of the k :th cluster and \mathbf{x} is a feature vector associated with an object in the k :th cluster. In general, W_K will naturally decrease as the number of clusters increases, since the distance between any arbitrary feature vector and its corresponding center will decrease. The idea is to choose K such that a kink in the curve can be identified. The rationale for the method is that a low W_K is desirable, but that adding another cluster will at some point only lead to a very limited improvement in W_K , since the improvements are diminishing. Identifying the kink is therefore a heuristic to determine K such that the subsequent improvements in W_K are so small that they most likely correspond to splitting a cluster that ‘naturally’ exists in the data, rather than splitting the data into its natural clusters. (Hastie and Tibshirani, 2001)

Average path length

The path between a leaf node n_k and the root node n_1 is defined as the sequence of nodes n_1, \dots, n_k such that n_i is the parent of n_{i+1} for $1 \leq i < k$. The length of the path between the leaf and the root is subsequently the number of edges in the path. Moreover, the average path length is the average of all the leaves’ path lengths. (Weiss, 2019) Of course, an unbalanced tree will, in general, have a larger average path length than a more balanced tree.

4 Methodology

This chapter presents the methodology used in this thesis. In general, the methodology employed has been iterative and, as a result, the thesis can be described as consisting of three separate parts:

1. To begin with, López de Prado's original HRP algorithm has been evaluated on a certain data set using certain performance measures and against certain other portfolio construction methods.
2. Secondly, the same algorithm has been analyzed and reviewed in order to understand in detail how it operates and what theoretical properties it has. During this analysis, an undesirable property was identified. As a result, alternative approaches to HRP have been developed and evaluated.
3. Thirdly, a potential improvement to the approaches developed in part 2 has been explored and implemented. The resulting approaches developed have been evaluated.

These three separate evaluations of the performance of different portfolio construction methods are reflected in section 6 Results & Discussion. The rest of this chapter is devoted to describing the procedure used to evaluate the different portfolio construction methods in greater detail, since all three evaluations are performed identically. Section 4.1 provides an overview and brief summary of the evaluation. Section 4.2 describes the price data used for the evaluation. Section 4.3 describes how the price data is used to compute returns and produce a good estimate of the covariance matrix. Section 4.4 describes the method and algorithm used to evaluate the performance of the different portfolio construction methods using walk-forward analysis. Lastly, section 4.5 describes how a randomization element is introduced into the original HRP approach, in order to mitigate the effects of the undesirable property of the original approach.

4.1 Overview

Since López de Prado only evaluated the out-of-sample performance of the original HRP algorithm using Monte Carlo simulations, it is deemed highly interesting to evaluate it using empirical data. Furthermore, the only performance measures taken into consideration were volatility, mean return and Sharpe ratio, and the only other portfolio construction methods used as benchmark were mean-variance and inverse-variance. In this evaluation, both additional performance measures and additional benchmark methods are taken into consideration. The portfolio construction methods included are minimum variance, ERC, inverse-variance, inverse-volatility, and EW. They are selected on the basis that they are the most common risk-based and/or heuristic portfolio construction methods. The performance measures used to evaluate the portfolios are annualized mean return, annualized volatility, annualized Sharpe ratio, maximum drawdown, average drawdown and turnover.

4.2 Data

The data used are closing prices for a maximum of $N = 38$ financial instruments during the time period starting 1977-11-28 and ending 2021-03-05. It should be noted that not all assets included in the data set have price series starting as early as 1977-11-28. Moreover, the convention of 253 trading days per year is used.

The financial instruments used are futures contracts and it should be noted that futures cover several different asset classes that may have very different average volatility and return. More specifically, the data set will include 13 bond futures and 25 equity index futures trading on 15 different markets. Naturally, equity index futures have both higher return and higher volatility compared to bond futures, which should be taken into consideration when evaluating the results. In general, futures are also less correlated than stocks, which is the most common investment universe used for this type of evaluation of portfolios. Lastly, the fact that bond futures have demonstrated strong performance throughout the time period covered should also be taken into consideration when evaluating the results, since this benefits portfolio construction methods that favor low volatility. See Appendix A for descriptive metadata for the full data set.

4.3 Modeling

As described in section 2 Financial Background, a futures contract always has an expiration date. At any point in time, more than one futures contract on the same underlying asset but with different expiration dates are trading in the market simultaneously. Typically, traders wish to avoid settlement, be it cash or physical, and instead maintain the same position even after expiry of the contract. As a result, traders will often roll over a futures contract about to expire, i.e. the near-month contract, to the corresponding futures contract in a further-out month. In practice, this implies closing out the former position and entering into the latter one. (Carchano and Pardo, 2009; Chan, 2013)

For modeling purposes, rolling of contracts is used to attain one single time series with historical prices of the different futures contracts included in the data set. Of course, there are different rollover strategies and the most common include non-adjusted, arithmetically adjusted and geometrically adjusted. The latter two adjust the price series as to eliminate the jump in price at the date of rollover. Such futures contracts with adjusted price series are commonly referred to as continuous futures contracts. In this thesis, geometrically adjusted price series are used. In practice, the whole price series is backward adjusted, one contract at a time, beginning with the contract that was most recently rolled over. At the time of the most recent rollover date, the ratio between the price of the current near-month contract and that of the preceding contract is calculated and multiplied to the whole price series of the earlier contract. This procedure is then repeated for all previous rollover dates until the whole price series has been adjusted and consequently all price jumps have been eliminated. It should be noted that the adjusted price series no longer corresponds to the actual historical prices. (Chan, 2013)

From the adjusted price series, log-returns $\{R_t\}_{t=1}^n$ are calculated:

$$R_t = \log\left(\frac{P_t}{P_{t-1}}\right). \quad (34)$$

Once log-returns are determined, the covariance matrix of the returns is to be estimated. Due the heteroskedasticity of returns it is desirable to give more weight to recent observations, whilst still incorporating information from the whole time-series into the estimate. Therefore the covariance matrix is estimated using an exponentially weighted moving average, implying that the covariance at time t between two assets x and y is estimated as:

$$\text{Cov}_t(x, y) = \frac{\sum_{i=0}^n (1 - \alpha)^i (x_{t-i} - \bar{x})(y_{t-i} - \bar{y})}{\sum_{i=0}^n (1 - \alpha)^i}. \quad (35)$$

In the above expression, α represents the decay factor and is set to $\alpha = 2/254$, where 253 is the average number of trading days in a year. Moreover, x_t and y_t are defined as the smoothed log-returns of the assets x and y at time t , and \bar{x} and \bar{y} are defined as the average smoothed log returns of the assets x and y . The first covariance estimate considered is the one such that at least 246 pair-wise observations exist prior to and including the time at which the correlation is estimated and n represents the number of pair-wise observations that exist prior to and including the time t . The number 246 is determined to be as close to one year as possible. Moreover, smoothed log-returns are defined as the simple rolling average of three of these daily log-returns:

$$Z_t = \frac{1}{3} \sum_{i=t-2}^t R_i \text{ for } t = 3, \dots, n. \quad (36)$$

Smoothed returns in (36) are preferable to the non-smoothed returns in (34) since the smoothed returns provide a more just reflection of the correlation between instruments trading on different exchanges. Of course, exchanges in different parts of the world have different and to some extent non-overlapping opening hours. For example, events in the afternoon, U.S. west coast time, will not be reflected in Asian markets on the same trading day but rather on the following trading day.

As explained in both section 1.1 Background and 3.1 Portfolio Construction, many portfolio construction methods require the inversion of the covariance matrix. In the presence of estimation errors, this operation amplifies these errors, possibly resulting in significant instability in the resulting portfolio weights. In order to mitigate this effect, the covariance matrix can be transformed using linear shrinkage, which serves as to pull the most extreme values towards the most central ones. (Ledoit and Wolf, 2004) For the sake of this evaluation, the covariance matrix is shrunk towards the diagonal of the covariance matrix:

$$\Sigma_\alpha = (1 - \alpha) * \Sigma + \alpha * \text{diag}(\Sigma). \quad (37)$$

The challenge is to determine the optimal shrinkage coefficient $\alpha \in [0, 1]$. (Ledoit and Wolf, 2004) For this reason, each portfolio construction method is evaluated for different values of α , and the value of α that results in the best performance measures for each method will be used going forward. As measure of performance, Sharpe ratio is considered most interesting.

4.4 Walk-Forward Analysis

Portfolio construction methods based on optimization produce portfolios that are optimized for the particular data set used to determine the portfolio weights. For example, it is given that the mean-variance portfolio should have the highest Sharpe ratio and that the minimum variance portfolio should have the lowest variance when evaluated on the same data set used to determine the portfolio weights, i.e. evaluated in-sample. However, when these portfolios are evaluated using a different data set, i.e. evaluated out-of-sample, the same does not necessarily hold true. As such, out-of-sample evaluation is far more reliable than in-sample

evaluation. Similarly, heuristic portfolio construction methods are calculated based on a particular data set, and may not perform as intended out-of-sample. (Pardo, 2008)

However, one out-of-sample evaluation of the performance of a portfolio does not guarantee that a certain portfolio construction method will continue to outperform over time and under changing market conditions. In order to say that a portfolio construction method is indeed robust and not the result of overfitting, walk-forward analysis is used. If one walk-forward entails calculating portfolio weights using one subset of the data (training data) and subsequently evaluating the performance of the constructed portfolio using another subset of the data (evaluation data), then a walk-forward analysis is a set of sequential walk-forwards performed on different subsets of the data such that the evaluation periods do not overlap. (Pardo, 2008)

In this thesis, the training data used to determine portfolio weights is not defined explicitly, given that the covariance matrix is estimated using an exponential moving average. Instead, exponential weights are placed on the data with a decay factor set to $\alpha = 2/254$, implying that correlations will be largely based on price data from the past year. The resulting portfolio is subsequently evaluated on one day's price data. In practice, the algorithm entails looping through the data set such that one portfolio is constructed each day t , where portfolio weights are calculated mainly using price data from time points $t-253, \dots, t$ and subsequently evaluated using return data as defined in (34) from time point $t+2$. This can be viewed as calculating portfolio weights based on days up until and including day t , trading on day $t+1$ and then collecting returns on day $t+2$. As such, this procedure can also be interpreted as applying daily rebalancing of the portfolios.

In total, portfolios are constructed for 8086 time points in the interval 1989-12-21 - 2021-03-05, using each of the portfolio construction methods included in the comparison. Beyond the different HRP portfolios, the minimum variance, ERC, inverse-variance, inverse-volatility and EW portfolios are evaluated. It should be noted that amongst these portfolios, the minimum variance portfolio is the only one where short-selling is allowed in its unconstrained form. Since it is well known that portfolios subject to constraints are less optimal than unconstrained portfolios, (Clarke et al., 2002) the minimum variance portfolio subject to the constraint that short-selling is not allowed is included in the comparison. This constrained portfolio will going forward be referred to as the long-only minimum variance portfolio. The performance measures used to evaluate the portfolios are annualized mean log-return, annualized volatility, annualized Sharpe ratio, maximum drawdown, average drawdown and turnover. For the definitions of these measures, see 2 Financial Background.

4.5 Randomization of HRP

In order to mitigate the effects of the undesirable property of the original HRP algorithm, a randomization element is introduced. The original HRP algorithm is run 10 times, resulting in 10 different portfolio weights for each asset in the data set. When evaluating the performance of the original HRP approach, the average of these 10 portfolio weights is used for each asset. Since this implies a rather large number of runs, smaller changes to the numerical implementation of López de Prado's HRP algorithm have also been implemented in order to reduce the run time. These changes entailed using Numba, an open-source just-in-time (JIT) compiler that compiles Python code into C. (Lam et al., 2015)

5 Analysis

This section is devoted to algorithmic descriptions of both the original HRP approach and the new approaches to HRP developed in this thesis. The overarching aim is to understand the properties of the original HRP approach. In turn, this understanding is used as motivation for developing the new approaches to HRP. Section 5.1 describes the original HRP approach and section 5.2 describes the new approaches to HRP, with a particular focus on how they differ from both each other and from the original approach.

5.1 Original HRP Algorithm

The aim of this section is to understand and explain how the HRP algorithm operates, in order to draw conclusions about its mathematical rigor and correctness. This is motivated by the fact that although the python code for the HRP algorithm is available online, López de Prado does not provide much intuition behind the code. The analysis will consequently entail both description and problematization of the algorithm. Ultimately, suggestions on how the algorithm can be improved will be presented.

Section 5.1.1 gives an introduction to the HRP approach and the intuition behind it, section 5.1.2 describes the algorithm in more detail, one stage at a time, and section 5.1.3 summarizes the suggested improvements.

5.1.1 Overview

The HRP approach was first introduced and published in 2016 with the aim of addressing two fundamental problems with modern portfolio theory: Instability and underperformance. These problems have been treated in more detail in section 3.1 Portfolio Construction. In conformity with other heuristic allocation strategies, the HRP approach mitigates these problems by not incorporating returns or requiring the inversion of the covariance matrix. López de Prado performs Monte Carlo simulations and is able to demonstrate that the HRP portfolio has lower volatility and higher Sharpe ratio than both Markowitz's traditional mean-variance portfolio and the inverse-variance portfolio when evaluated out-of-sample.

As demonstrated by López de Prado's simulation, the HRP approach does not only provide stability and superior performance, but may also be considered intuitive. In traditional portfolio construction, the covariance matrix does not incorporate the notion of hierarchy. Instead, all instruments considered for a portfolio are viewed as substitutable for one another. However, in practice, an investor constructing his or her portfolio will most certainly view certain instruments more as substitutes for one another and others more as complements to one another. For example, for a portfolio comprised of only stocks, an investor may choose to group stocks in terms of geography or industry. Although the HRP algorithm does not take into consideration qualitative aspects when clustering instruments, this example effectively illustrates the intuition behind the algorithm.

The algorithm operates in three stages:

1. Tree clustering
2. Quasi-diagonalization
3. Recursive bisection

In the first stage, assets are clustered to form a tree. In the second stage, assets are re-ordered as to reflect this hierarchical structure. When the columns and the rows of the covariance matrix are re-ordered in the same way, the result is a quasi-diagonal matrix. Lastly, in the third stage, allocations are split between subsets of assets using inverse-variance allocation.

5.1.2 Algorithm

5.1.2.1 Tree Clustering

In the first stage of the algorithm, assets are combined into a hierarchical structure of clusters using agglomerative clustering. A cluster is defined as the formation of two constituents, where a constituent may be either a singleton cluster or another cluster containing several assets. The two constituents of a cluster may also be referred to as left and right child, respectively. Of course, with N assets, $N-1$ clusters are formed and can be illustrated using a tree structure. The term branch will be used to describe all assets contained within a certain cluster, rather than simply the two constituents.

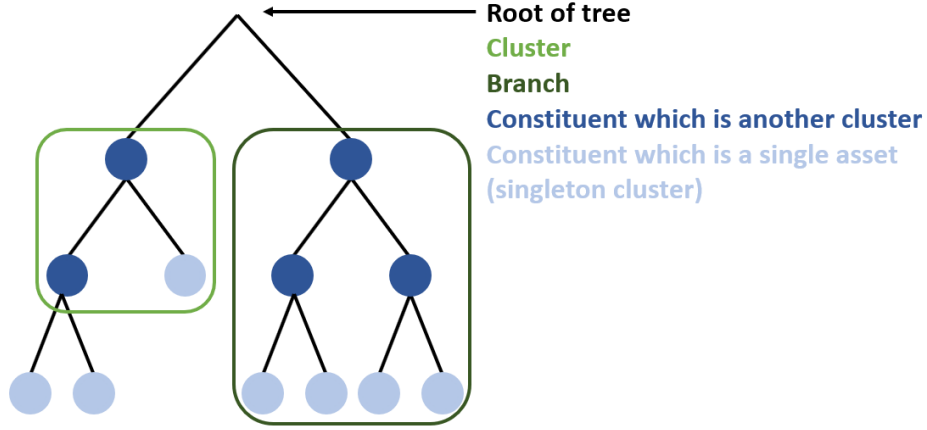


Figure 2: Illustration to clarify clustering terminology

The algorithm is described in pseudocode in Algorithm 1. In essence, assets that have similar correlation to all other assets will be considered 'close' to each other and consequently be clustered.

Algorithm 1: Hierarchical clustering

Input : Price data ($\mathbf{X} \in \mathbb{R}^{N \times n}$)

Output: Linkage matrix (\mathbf{L})

- 1 Estimate correlation matrix $\boldsymbol{\rho} = \{\rho_{i,j}\}_{i,j=1,\dots,N}$ based on \mathbf{X}
 - 2 Compute feature matrix $\mathbf{D} = \{d_{i,j}\}_{i,j=1,\dots,N} \in \mathbb{R}^{N \times N}$, where $d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$
 - 3 Compute distance matrix $\tilde{\mathbf{D}} = \{\tilde{d}_{i,j}\}_{i,j=1,\dots,N} \in \mathbb{R}^{N \times N}$, where $\tilde{d}_{i,j} = \sqrt{\sum_{n=1}^N (d_{n,i} - d_{n,j})^2}$
 - 4 Cluster all assets and summarize cluster information in linkage matrix $\mathbf{L} \in \mathbb{R}^{(N-1) \times 4}$
-

Although the algorithm is rather straightforward, the linkage matrix may deserve further explanation. Each row m in the linkage matrix $L \in \mathbb{R}^{(N-1) \times 4}$ represents one cluster:

$$L = (y_{m,1} \ y_{m,2} \ y_{m,3} \ y_{m,4}) \quad \text{for } m = 1, \dots, N-1. \quad (38)$$

Element $y_{m,1}$ holds a reference to the left constituent of cluster m and element $y_{m,2}$ holds a reference to the right constituent of the same cluster. Element $y_{m,3}$ is the distance between the two constituents of cluster m . Lastly, element $y_{m,4}$ is the number of assets in cluster m . If there are N assets and these are referenced as $0, 1, \dots, N-1$, then the i :th cluster will be referenced as $N-1+i$. The two assets i and j with minimum distance \tilde{d} will be clustered first. Once clustered, the cluster is referred to as cluster 1 but also treated as an asset with index N . This means that another asset k will be clustered with cluster 1 next if the distance between the two items is the smallest. Since single linkage is used, the distance recorded between two clusters is the minimum distance between any two assets in the clusters. The last cluster to be created, i.e. cluster $(N-1)$, is naturally the root of the tree. This is clarified in Figure 3.

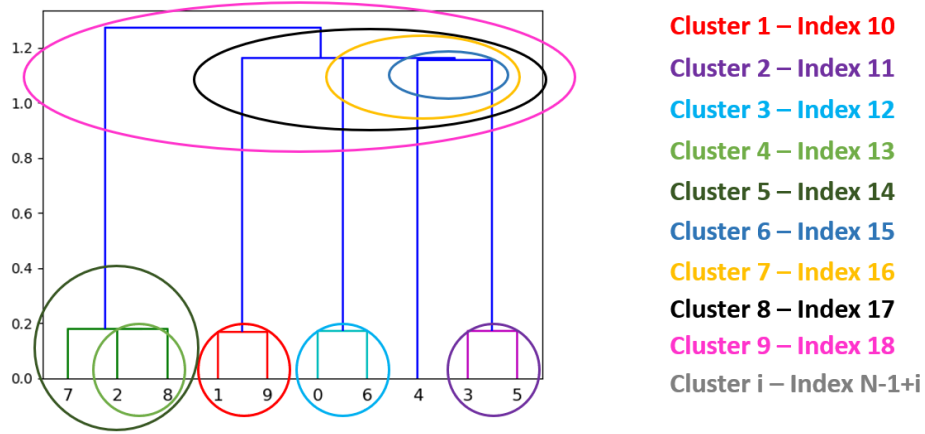


Figure 3: Illustration to clarify how the clustering works in practice and relates to the linkage matrix. The first row of the linkage matrix will contain information about cluster 1, which consists of assets 1 and 9. Cluster 5 is the first cluster to consist of not only single assets but another cluster. It consists of asset 7 and cluster 4, which in turn consists of assets 2 and 8. Consequently, $y_{5,1} = 7$ and $y_{5,2} = 13$ in the linkage matrix. The cluster or asset with the smallest index in a cluster will always be placed as the left child.

Possible changes to the algorithm include altering how the covariance matrix is estimated as well as the features, the distance metric \tilde{d} and the linkage criterion.

5.1.2.2 Quasi-Diagonalization

In the second stage of the algorithm, the rows and the columns of the covariance matrix are reorganized in order to reflect the hierarchical structure produced in the previous stage. The purpose with this reorganization is to obtain a quasi-diagonal matrix. As previously mentioned, the HRP algorithm makes use of inverse-variance allocation, which disregards correlations entirely and consequently is optimal only when assets can be assumed to be uncorrelated. By ensuring that the covariance matrix is quasi-diagonal, López de Prado means that it is more justifiable to disregard correlations. Continuing the example with

ten assets presented in Figure 3, the corresponding correlation matrix before and after the quasi-diagonalization is presented in Figure 4.

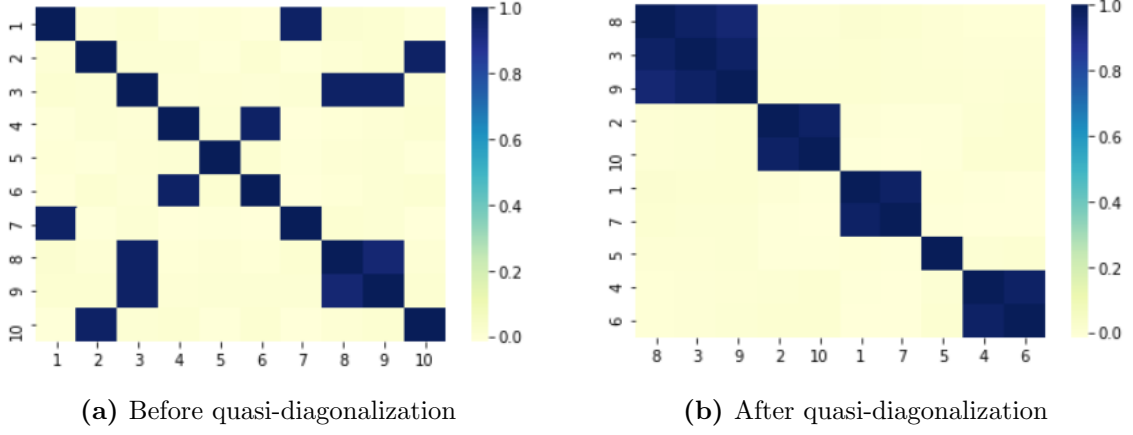


Figure 4: Correlation matrix before and after quasi-diagonalization for data simulated by López de Prado.

The algorithm is described in pseudocode in Algorithm 2. Although it may be considered a straightforward matter to obtain the order of the assets corresponding to the hierarchical structure constructed in the first stage, the algorithm itself is not. For this reason, some additional comments will be provided, beyond the algorithm itself.

Algorithm 2: Quasi-diagonalization

Input : Linkage matrix (**L**)
Output: Sorted list of assets (**sortIx**)

- 1 extract the first two values in the bottom row of the linkage matrix (root of tree)
- 2 place these in labelled column vector **sortIx**
- 3 **while** *there are still clusters left in the structure* **do**
- 4 make space for constituents in **sortIx**
- 5 find clusters in **sortIx**
- 6 replace constituents with their children
- 7 sort **sortIx** by index
- 8 re-index **sortIx**
- 9 **end**
- 10 return **sortIx** as a list

On line 3, ensuring that there are still clusters in the structure implies ensuring that at least one value in **sortIx** is larger than or equal to the number of assets N , since the first cluster receives index N , the second $N+1$, etc. On line 4, space is made in **sortIx** by adding one empty indexed line below each line item in **sortIx**. This is necessary for keeping track of the correct order of the assets. On line 5, clusters are found in the same way the condition in the while loop is checked, i.e. using indices. Line 6 reads that constituents are replaced with their children. In practice, each constituent is replaced with its left child and the corresponding right child is placed on the line below its sibling, created on line 4. On line 7, **sortIx** is sorted so that the order of the lines in **sortIx** corresponds to the order of the

indices. On line 8, **sortIx** is reordered so that each index corresponds to each line. This is necessary in cases where the space made on line 4 is not used, i.e. if a constituent was in fact an asset and not a cluster. In this case indices may be e.g. 0 2 3 4 6 instead of 0 1 2 3 4. Lastly, on line 10, **sortIx** is returned as a list.

5.1.2.3 Recursive Bisection

In the third and final stage of the algorithm, portfolio weights are allocated using inverse-variance through a tree structure. The algorithm is described in pseudocode below.

Algorithm 3: Recursive bisection

Input : Covariance matrix (Σ); Sorted list of assets (**sortIx**)
Output: Portfolio weights (**w**)

```

1 create a column vector w with indices equal to sortIx and all values equal to 1
2 create a list cItems containing one element, which corresponds to the list sortIx
3 while not all elements in cItems have size 1 do
4   | bisect elements in cItems in two subsets of equal size
5   | for two consecutive elements in cItems do
6   |   | calculate variance for each element using inverse-variance allocation
7   |   | calculate scaling factor
8   |   | rescale allocations for all items in each element using scaling factor
9   | end
10 end
11 return w
```

To begin with, all weights are set equal to one. On line 2, the list **cItems** is created. For the first iteration, this list holds only one element, which is the ordered list of assets generated in stage 2. In the example with ten assets, **cItems** will consequently be equal to $[[7, 2, 8, 1, 9, 0, 6, 4, 3, 5]]$. The while loop initiated on line 3 will continue as long as not all elements in **cItems** have length one, i.e. until **cItems** = $[[7],[2],[8],[1],[9],[0],[6],[4],[3],[5]]$. When this is the case, the algorithm has traversed top-down through the tree and reached the leaves of the tree, i.e. weights for all assets have been determined. For each iteration, the algorithm moves down one level in the tree by bisecting the elements in **cItems** in two subsets of equal size, which is performed on line 4.

However, it should be noted that the inverse-variance allocation of weights is not only applied top-down but also bottom-up. To be more precise, on line 6, inverse-variance is used bottom-up to determine the weights of the assets within a branch, and these weights are used exclusively to calculate the total variance of a particular branch of the tree. For example, branch K consisting of m assets would receive the following weights and total variance:

$$\tilde{\mathbf{w}}_{\mathbf{K}} = \{\tilde{w}_i\}_{i=1}^m \text{ where } \tilde{w}_i = \frac{1/\sigma_i^2}{\sum_{j=1}^m 1/\sigma_j^2} \quad (39)$$

$$\sigma_{\mathbf{K}}^2 = \tilde{\mathbf{w}}_{\mathbf{K}}^T \Sigma_{\mathbf{K}} \tilde{\mathbf{w}}_{\mathbf{K}}. \quad (40)$$

In turn, on lines 7-8, the total variance of a branch is used to allocate weights top-down through the tree. For two consecutive elements in **cItems**, i.e. for two branches located

next to each other in the tree on a particular hierarchical level, weights of the individual assets are scaled using a scaling factor α . For example, for the two consecutive branches J and K, the weights for all assets in each branch are scaled as follows:

$$\alpha_{J,K} = \frac{1/\sigma_J^2}{1/\sigma_J^2 + 1/\sigma_K^2} \quad (41)$$

$$\mathbf{w}_J = \mathbf{w}_J * \alpha_{J,K} \quad (42)$$

$$\mathbf{w}_K = \mathbf{w}_K * (1 - \alpha_{J,K}). \quad (43)$$

Moreover, it should be noted that the hierarchical structure or tree produced in stage 1 is not necessarily the tree used in this final stage. Instead, only the order of the assets produced in stage 2 is used and since the list of assets is bisected in equal subsets on line 4, the clusters created in stage 1 are not necessarily preserved. For the example with 10 assets, this is illustrated in Figure 5.

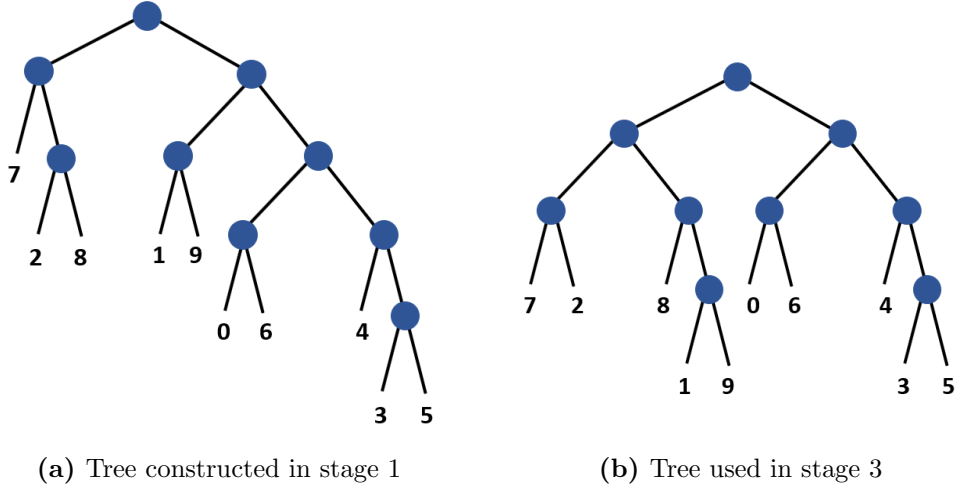


Figure 5: Comparison of tree structures in stage 1 and 3 of original HRP algorithm

The fact that the hierarchical structure produced in stage 1 is not used in stage 3 has been noted by Raffinot (2018a), Raffinot (2018b), Huang (2020) and Lau et al. (2017). This characteristic is not only odd but also implies that the original order of the assets in the data set matters. For example, swapping places between asset 1 and 2 in the original data set, and updating the rows and the columns of the covariance matrix accordingly, does not alter the data set itself and asset 1 and 2 should receive the same weights regardless of this swap. The recursive bisection in stage 3 of López de Prado’s HRP algorithm ensures that this is not necessarily the case. Of course, this is not a trait desirable for a portfolio construction method. A reasonable portfolio construction method should allocate the same weight to an asset regardless of where its information is located in the covariance matrix. As a result, altering this part of the algorithm so that it makes use of the tree produced in stage 1 is identified as a possible alteration to López de Prado’s original algorithm. Other possible changes include altering the portfolio construction method used in the algorithm, both within and between clusters.

5.1.3 Suggested Improvement

In summary, there is primarily one property of the original HRP approach identified as undesirable. This is the property that the order of the assets matters, i.e. that when swapping the places of two assets in the data set, the assets do not necessarily receive the same weights as before the swap. This property is the result of the recursive bisection in the third stage, which implies that the original hierarchical structure generated in the first stage is disregarded. As a result, it is deemed interesting to amend this and develop alternative algorithms where the hierarchical structure is preserved and actually used to allocate weights.

As already mentioned, it is also possible to alter several of the different parameters or methods used in the algorithm, such as linkage criterion, features, distance metric, clustering methods, and portfolio construction method used within and/or between clusters. Within the scope of this thesis, it is deemed most interesting to alter the way the covariance matrix is estimated, by using an exponentially weighted moving average, and to investigate whether the performance of the algorithm is affected by shrinkage.

5.2 New Approaches to HRP

As already explained, it is deemed odd that the tree generated in the first stage of the original HRP algorithm is in fact not used when allocating weights in the third and final stage of the algorithm. In practice, the fact that the hierarchical structure is not used for allocation has the undesirable effect that the order of the assets matters. For this reason, it is deemed interesting to investigate alternative algorithms where the hierarchical structure is indeed used when allocating weights. Two such alternative methods have been developed and will be described in more detail in sections 5.2.1 and 5.2.2. Going forward, these new approaches will be referred to as the bottom-up approach and the top-down approach, whilst López de Prado's algorithm first published in 2016 will be referred to as the original HRP approach. It should be noted that the names are related only to how the weights are calculated, not how the clustering is performed, since both are based on agglomerative clustering.

The two new methods both consist of two stages; (1) tree clustering and (2) weight allocation. The first stage is equivalent to the first stage of the original HRP approach described in detail in 5.1.2.1 and generates the linkage matrix as output. However, unlike the original HRP approach, both new approaches allocate weight using the tree retrieved from the hierarchical clustering in the first stage. As a result, the second stage of the original HRP approach ('quasi-diagonalization') is excluded and the third and final stage ('recursive bisection') is altered. The final stage of the two new approaches ('weight allocation') of course still takes the linkage matrix and the covariance matrix as input and generates the portfolio weights as output. However, in this second and final stage, the two new approaches differ in how they allocate weights. The first approach employs a bottom-up approach, whilst the second approach employs a top-down approach much similar to the one in the original HRP, but without the recursive bisection. Since the first stage is identical to that of the original HRP approach for both new approaches, only the second stage is described in more detail in the following subsections.

5.2.1 Bottom-Up Approach

The algorithm starts at the very bottom of the tree. In each iteration, two assets are combined to form a cluster. In subsequent iterations, the newly formed cluster is interpreted as a 'new' asset, that has replaced its constituents. In order to reflect the new set of assets, where two assets have been lumped into one, the covariance matrix must be updated. The new asset is defined as a weighted sum of its constituents, where the weights of the constituents are determined using inverse-variance weighting. The variance and the covariances are calculated accordingly. This new asset takes the place of its constituents and the number of rows and columns of the covariance matrix is consequently reduced by one. This process is continued until only one cluster remains and the last cluster naturally corresponds to the root of the tree. An algorithmic description is presented below.

Algorithm 4: Bottom-up approach

Input : Linkage matrix (\mathbf{L}); covariance matrix ($\mathbf{\Sigma}$)
Output: Portfolio weights (\mathbf{w})

- 1 add 2 columns to the linkage matrix, each containing a list of all assets in the left and right branch, respectively, of each cluster
- 2 create a column vector \mathbf{w} with all values equal to 1
- 3 **for** *each cluster, starting from the bottom* **do**
- 4 set left_child= $\mathbf{L}[\text{cluster},0]$ and right_child= $\mathbf{L}[\text{cluster},1]$
- 5 calculate the scaling factor α based on left_child and right_child
- 6 set left_branch equal to the constituents of left_child
- 7 set right_branch equal to the constituents of right_child
- 8 scale $\mathbf{w}[\text{left_branch}] = \mathbf{w}[\text{left_branch}] * \alpha$
- 9 scale $\mathbf{w}[\text{right_branch}] = \mathbf{w}[\text{right_branch}] * (1 - \alpha)$
- 10 update covariance matrix by viewing the newly formed cluster as another asset
- 11 **end**
- 12 **return** \mathbf{w}

In the very beginning, on line 1, two columns are added to the linkage matrix. These two columns are added in order to keep track of all assets included in a certain branch of the tree, which is required in order to scale weights on lines 8-9. On line 2, the portfolio weights are initiated with all values equal to one. On line 3, the traversal of the tree is initiated. This is represented by a for-loop which goes through all clusters, starting from the bottom of the tree. In practice, this corresponds to going through the linkage matrix, one row at a time, from top to bottom. For each cluster, the scaling factor is calculated on line 5. The scaling factor is determined as:

$$\alpha = \frac{1/\sigma_L^2}{1/\sigma_L^2 + 1/\sigma_R^2}. \quad (44)$$

In the above expression, α is the scaling factor, σ_L^2 is the variance of the left_child and σ_R^2 is the variance of the right_child. To be more specific, the left branch and the right branch are the set of asset(s) which together constitute the left_child and right_child respectively and the left_child and right_child are either an asset or clusters themselves. Since the algorithm traverses the tree bottom-up, the variance for each child will have been calculated in the previous iteration (see Equation (45) below). After computing the scaling factors, the weights are subsequently rescaled (see lines 8-9).

Once a new cluster is formed, it is to be viewed as a new asset which replaces its constituents. This 'new' asset, i.e. the cluster, can be interpreted as a weighted sum of its constituents, where the weights are determined by α and $(1-\alpha)$. This implies that the covariance matrix must be updated accordingly (see line 10). The covariance matrix is updated using the following expression:

$$\Sigma_{\text{new}} = A \Sigma A^T. \quad (45)$$

Σ is the covariance matrix from the previous iteration and Σ_{new} is the updated covariance matrix. The lumping of the `left_child` and `right_child` is consequently accomplished using the matrix A , which can be interpreted as the linear transformation of assets occurring when two assets are clustered together. In general, the matrix A will look something like the below:

$$A = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \alpha & \dots & 1-\alpha & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 \end{bmatrix} \quad (46)$$

Where the α and $(1-\alpha)$ elements have the indices $[\text{left_child}, \text{left_child}]$ and $[\text{left_child}, \text{right_child}]$, respectively. Moreover, when considering the iteration corresponding to cluster k , Σ will have the dimensions $(N-k+1, N-k+1)$, A will have the dimensions $(N-k, N-k+1)$ and Σ_{new} will have the dimensions $(N-k, N-k)$.

5.2.2 Top-Down Approach

The top-down approach starts from the root and traverses down the tree. In each iteration, the weights of all assets belonging to the left and the right branch of the current cluster, respectively, are scaled with a scaling factor determined using inverse-variance. Of course, computing this scaling factor requires determining the total variance of the left and right branch, respectively. What is fundamentally different between the bottom-up and the top-down approach, besides the obvious intuition behind the two, is how the variance of a branch of the tree is calculated. In the bottom-up approach, it is calculated using inverse-variance weighting on the left and the right child contained in that branch. Each of the children may be either an asset, for which the variance is given, or another cluster, for which the variance has been determined in a previous iteration. In the top-down, approach, on the other hand, the variance of a branch is calculated using inverse-variance weighting of all assets contained in the branch. Intuitively, this implies considering all assets contained in a branch as if part of the same cluster on the same hierarchical level, i.e. disregarding the hierarchical structure further down the tree.

Below follows an algorithmic description of the top-down approach.

Algorithm 5: Top-down approach

Input : Linkage matrix (\mathbf{L}); covariance matrix ($\mathbf{\Sigma}$)

Output: Portfolio weights (\mathbf{w})

```
1 add 2 columns to the linkage matrix, each containing a list of all assets in the left
  and right branch, respectively, of each cluster
2 create a column vector  $\mathbf{w}$  with all values equal to 1
3 for each cluster, starting from the root of the tree do
4   set left_child=L[cluster,0] and right_child=L[cluster,1]
5   calculate the scaling factor  $\alpha$  based on left_child and right_child
6   set left_branch equal to the constituents of left_child
7   set right_branch equal to the constituents of right_child
8   scale  $\mathbf{w}[\text{left\_branch}] = \mathbf{w}[\text{left\_branch}] * \alpha$ 
9   scale  $\mathbf{w}[\text{right\_branch}] = \mathbf{w}[\text{right\_branch}] * (1 - \alpha)$ 
10 end
11 return  $\mathbf{w}$ 
```

The algorithm is much similar to the algorithm for the bottom-up approach except for the fact that the algorithm starts at the root of the tree (line 3) and that the covariance matrix is not updated to reflect the formation of new assets. It is primarily on line 5 that the difference between the bottom-up approach and the top-down approach takes place. Similarly as to the bottom-up approach, the scaling factor for the top-down approach is calculated as:

$$\alpha = \frac{1/\sigma_L^2}{1/\sigma_L^2 + 1/\sigma_R^2}. \quad (47)$$

The difference, however, lies in how σ_R^2 and σ_L^2 are calculated. In the top-down approach, the variance of the left and the right branch, respectively, is calculated based on a covariance matrix which has been reduced from the original covariance matrix to only contain rows and columns corresponding to the assets contained in the relevant branch. By letting Σ_{red} be the reduced covariance matrix corresponding to the m assets contained in the left branch, the weights $\tilde{\mathbf{w}}$ used exclusively for calculating the variance of the left branch can be expressed as:

$$\tilde{\mathbf{w}} = \{\tilde{w}_i\}_{i=1}^m \quad \text{where } \tilde{w}_i = \frac{1/\sigma_i^2}{\sum_{j=1}^m 1/\sigma_j^2}. \quad (48)$$

Subsequently, the variance of the left branch, σ_L^2 , can be expressed as:

$$\sigma_L^2 = \tilde{\mathbf{w}}^T \Sigma_{\text{red}} \tilde{\mathbf{w}}. \quad (49)$$

Of course, this operation is performed also for the right branch of each cluster. It is clear that the variance of a branch containing more than two assets will be different when calculated using the bottom-up approach versus the top-down approach. In cases where the variance is different, so will the scaling factors be and in turn the portfolio weights as well. This is demonstrated for three and four assets in Appendix C.

6 Results & Discussion

This section presents and discusses the results of the three separate evaluations of the portfolio construction methods performed. Section 6.1 presents the result from the evaluation of the original HRP algorithm compared against select benchmark methods. Section 6.2 presents the results from the evaluation of the new approaches to HRP, developed in order to manage the undesirable property of the original HRP approach identified in section 5 Analysis. Section 6.3 presents the result from the evaluation of the new approaches to HRP, where restriction on growth has been implemented. All evaluations have been performed using the same procedure, described in detail in section 4 Methodology.

6.1 Evaluation of Original HRP Approach

To begin with, it is investigated which portfolio construction methods that would benefit from shrinkage of the covariance matrix. This is performed by plotting the performance measures mean log-return, volatility, Sharpe ratio, turnover, maximum drawdown and average drawdown as functions of the shrinkage coefficient α . Sharpe ratio as a function of α is deemed most interesting and is included below as Figure 6. From this figure, it is clear that minimum variance, with and without short-selling, ERC and HRP are affected by shrinkage. It is expected that both minimum variance and ERC should require shrinkage, but based on the work of Lopéz de Prado, the original hypothesis was that HRP would not be overly affected by shrinkage. Effectively, this implies that HRP is indeed somewhat sensitive to estimation errors in the covariance matrix, at least for this particular data set.

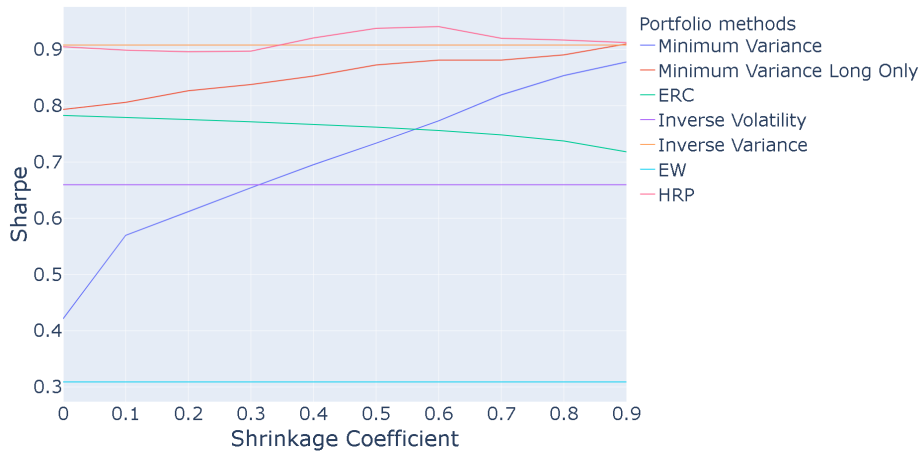


Figure 6: Sharpe ratio for each portfolio construction method as a function of α . The increment used is 0.1 on the interval $[0, 0.9]$.

Since a shrinkage coefficient $\alpha > 0.5$ implies forsaking too much information in the covariance matrix, the same plot as in Figure 6 but with increment 0.02 on the interval $[0, 0.5]$ is produced and included as Figure 7. From this figure, it is concluded that $\alpha = 0$ will be used for the ERC portfolio and $\alpha = 0.12$ will be used for HRP and minimum variance. Shrinkage of $\alpha = 0.14$ will be used for the long-only minimum variance as this value correspond to the first iteration where a warning regarding the convexity of the problem is not raised by the used solver, implying that the solver is unstable for lower values of α .

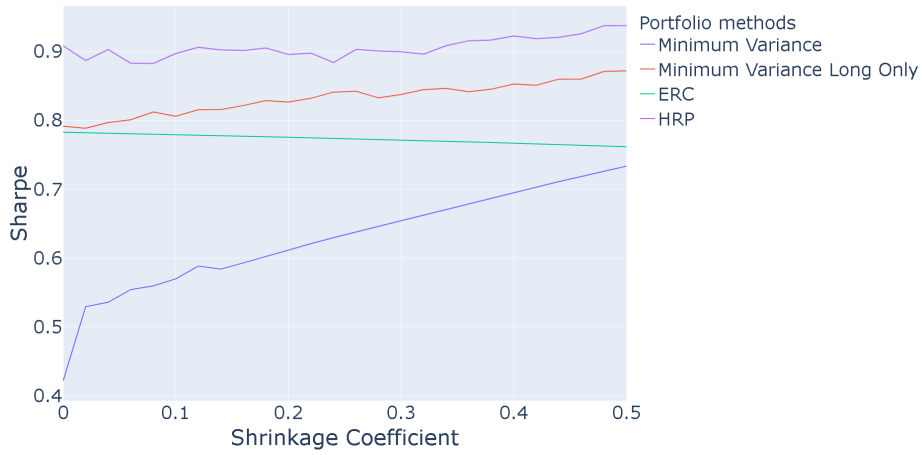


Figure 7: Sharpe ratio for the four portfolio construction methods in need of shrinkage as a function of α . The increment used is 0.02 on the interval $[0, 0.5]$.

In Figure 8, the share allocated to bond versus equity index futures is plotted for each portfolio construction method. The reason for excluding the minimum variance portfolio is that the portfolio allows for short-selling and thus a fair comparison is not possible. Moreover, it is clear that all methods except EW seem to favor bond futures over equity index futures, which makes sense since bond futures are associated with significantly less risk than equity index futures and all other portfolio construction methods favor assets with low risk by construction. Unlike the other portfolio construction methods, EW does not consider risk but weights all assets equally, resulting in a much higher share of equity futures. When analyzing the performance of portfolio construction methods that clearly favor low-risk assets such as bond futures, it should also be taken into consideration that bond futures have generally performed well during the time period considered.

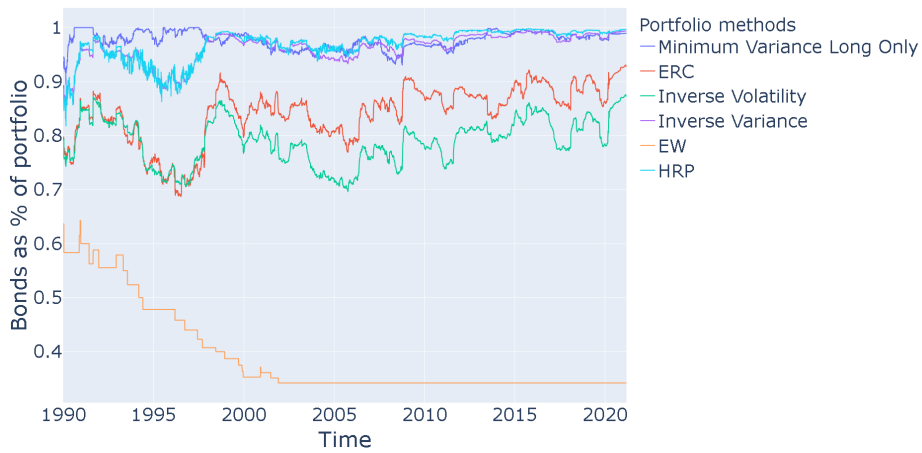


Figure 8: Share of portfolio consisting of bond futures for all portfolios constructed in walk-forward analysis except the minimum variance portfolio.

In Figure 9, the equity curve is presented. In this thesis, an equity curve is the graphical representation of the cumulative log-returns from the time point when the first portfolio

is constructed, i.e. 1989-12-21. The figure gives a general overview of how the different portfolios have performed with regard to return, volatility and drawdown and the results are not especially surprising. It is for instance expected that the minimum variance portfolios have the lowest return and that their curves are rather smooth, indicating low volatility. It is also apparent that inverse-variance and HRP are closely related, which is intuitive given that HRP makes use of inverse-variance allocation within and between clusters. Two other portfolios that are closely related to one another are ERC and inverse-volatility. This is also expected, as the risk measure used in this ERC portfolio is volatility. It is also expected that portfolios where the weights are inversely proportional to variance rather than volatility would generate lower returns, since they place even less weight on risky assets.

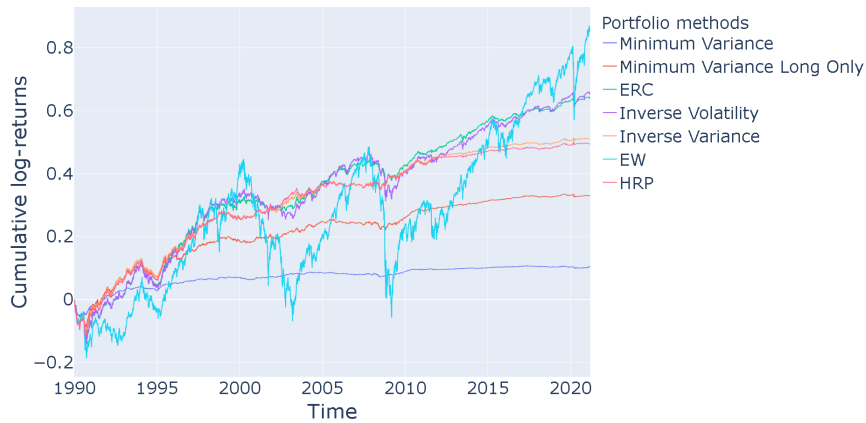


Figure 9: Equity curve, i.e. cumulative log-returns of portfolios constructed in walk-forward analysis using each portfolio construction method included in analysis.

In Figure 10, annualized rolling volatilities are presented. All of the portfolio methods except for EW seem to follow the same pattern and are of somewhat similar magnitude. The observation that the EW portfolio clearly sets itself apart from the rest is not considered surprising and can be traced back to the above-mentioned discussion regarding the overweight of bond futures in portfolio methods which favor low-risk assets.

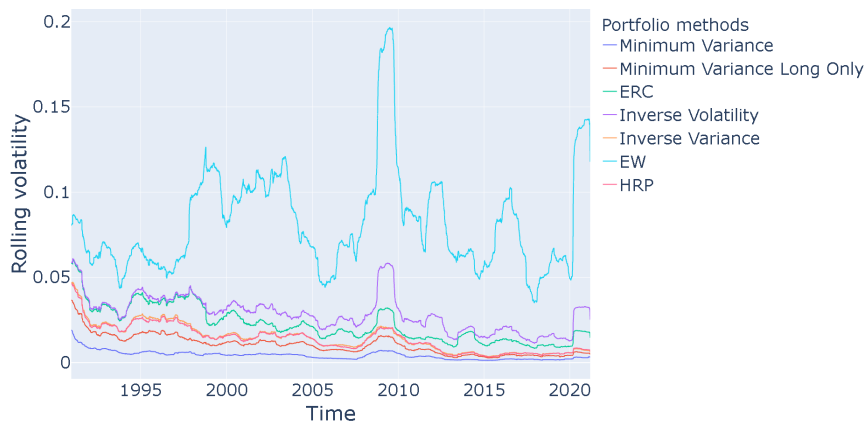


Figure 10: Annualized rolling volatility using a one year window for each portfolio construction method included in analysis.

In Figure 11, drawdowns are presented. It should be noted that in this figure drawdown is not defined as a percentage of the peak, which is often the case, but as an absolute value. The reason for not expressing drawdown as a percentage is that log-returns are used. The figure confirms the preliminary conclusions drawn from Figure 9, i.e. that EW suffers from large drawdowns, whilst the other portfolios are more stable over time. Unsurprisingly, the large drawdowns coincide with two rather particular events, namely the burst of the internet bubble in 2000 and the global financial crisis in 2008. The fact that all methods except for EW seem to be rather resilient to these two events does indeed support the claim that risk-based portfolio construction methods are robust and stable over time.

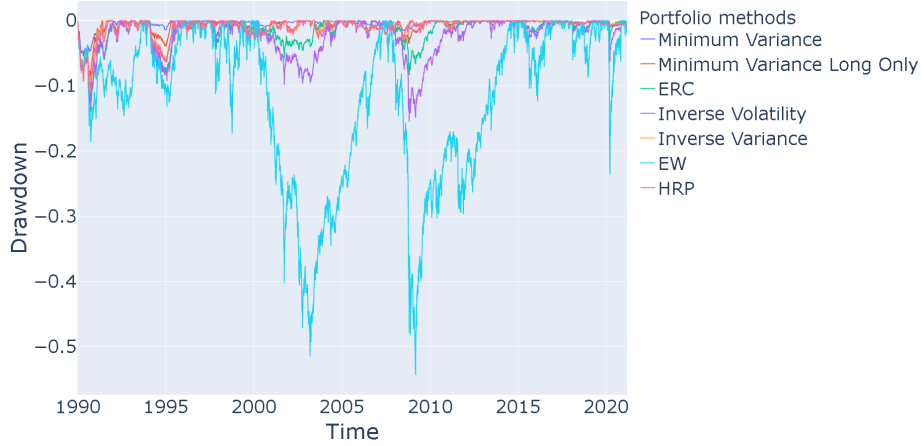


Figure 11: Drawdown at each point in time for each portfolio construction method included in analysis.

Beyond analyzing the performance measures of the portfolios over time, a single performance measure for each portfolio has been produced. These are presented in Table 1.

	Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Average drawdown	Maximum drawdown	Turnover
MinVar	3.30e-03	5.61e-03	0.59	-4.07e-03	-55.05e-03	307e-06
MinVar Long	10.31e-03	12.64e-03	0.82	-7.80e-03	-87.49e-03	1018e-06
ERC	20.00e-03	25.55e-03	0.78	-15.42e-03	-140.11e-03	246e-06
InvVol	20.49e-03	31.06e-03	0.66	-23.72e-03	-154.83e-03	106e-06
InvVar	15.91e-03	17.53e-03	0.91	-8.87e-03	-103.80e-03	176e-06
EW	26.94e-03	87.17e-03	0.31	-120.42e-03	-543.41e-03	4e-06
HRP	15.45e-03	17.03e-03	0.91	-9.59e-03	-110.85e-03	1443e-06

Table 1: Performance measures for all portfolio methods included in analysis.

In Table 2, the same performance measures are considered, but instead of presenting the values, the portfolio construction methods have been ranked from best to worst for each performance measure. In general, these results confirm the conclusions that have already been drawn based on the figures with performance over time for the portfolios.

Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Average drawdown	Maximum drawdown	Turnover
EW	MinVar	InvVar	MinVar	MinVar	EW
InvVol	MinVar Long	HRP	MinVar Long	MinVar Long	InvVol
ERC	HRP	MinVar Long	InvVar	InvVar	InvVar
InvVar	InvVar	ERC	HRP	HRP	ERC
HRP	ERC	InvVol	ERC	ERC	MinVar
MinVar Long	InvVol	MinVar	InvVol	InvVol	MinVar Long
MinVar	EW	EW	EW	EW	HRP

Table 2: Ranking of all portfolio methods included in analysis for the different performance measures.

However, from these two tables it is easy to draw the conclusion that there is no one method which consistently outperforms the other. Instead, the ranking depends on the performance measure considered. For example, the EW portfolio has the highest return, but also highest volatility and drawdown. Instead, the minimum variance portfolios outperform the others in terms of volatility and drawdown. This result is intuitive but still contradicts the results from Lopéz de Prado’s own simulations. Based on Lopéz de Prado’s work, one of the initial hypotheses was also that HRP would outperform the other portfolios in terms of Sharpe ratio, which is confirmed by this analysis as the difference in Sharpe ratio between inverse-variance and HRP is deemed negligible. However, it should also be noted that both HRP and the long-only minimum variance portfolio have remarkably high turnovers compared to the other portfolios. Even though some studies have highlighted that HRP has demonstrated a high turnover, it is somewhat unexpected that the turnover of the HRP portfolio is approximately 1.4x larger than that of the long-only minimum variance portfolio and more than 4x larger than that of the portfolio with the third highest turnover. One of Lopéz de Prado’s main points was namely that traditional optimization portfolios are numerically unstable and that HRP, in contrast to traditional optimization portfolios, delivers intuitive and stable results. The high turnover, on the other hand, indicates that HRP is not immune to the problem of numerical instability.

6.2 Evaluation of New Approaches

This section presents the result from the evaluation of the new approaches to HRP, developed in order to manage the undesirable property of the original HRP approach identified in section 5 Analysis. These approaches are referred to as the bottom-up approach and the top-down approach, respectively, and are applied using both single linkage and Ward’s method as linkage criterion and compared only against the original HRP approach.

It should be noted that the same shrinkage coefficient $\alpha = 0.12$ is used for all portfolio methods evaluated, as all of them are different versions of HRP. In Figure 12, the equity curve is presented, which gives a general idea of how the different portfolios have performed in terms of return, volatility and drawdown. It is interesting to note that no portfolio stands out in terms of volatility or drawdown but that the bottom-up and top-down approaches applied with single linkage demonstrate higher returns than the other portfolios.

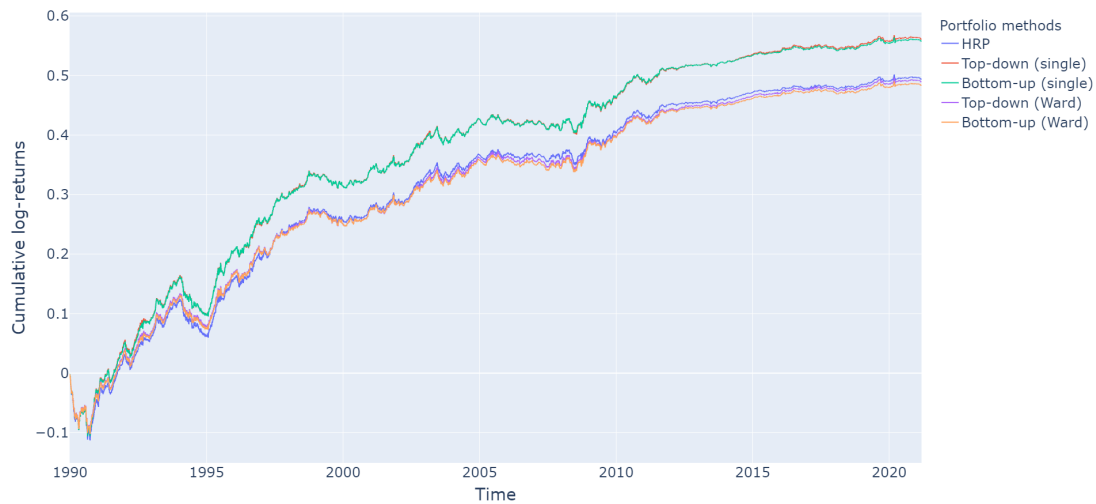
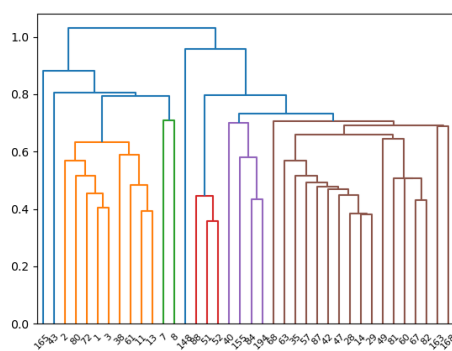
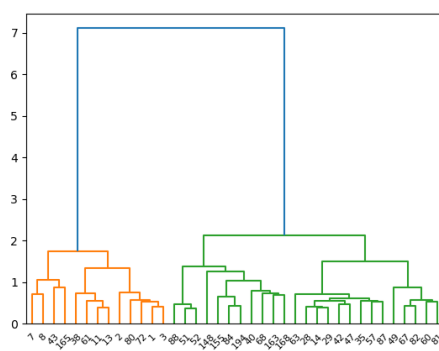


Figure 12: Equity curve, i.e. cumulative log-returns of portfolios constructed in walk-forward analysis using each portfolio construction method included in analysis

Although the original HRP algorithm applies single linkage, the return of the portfolio is more in line with the return of the portfolios constructed using Ward's method as linkage criterion. This behavior may be considered nonintuitive but upon reflection it is actually logical. Ward's method generally results in a more balanced tree than single linkage, which to the contrary often suffers from chaining. A similar tree as that produced using Ward's method is constructed in the third and final stage of the original HRP algorithm, where recursive bisection ensures that the original hierarchical structure is disregarded and weights are allocated using a more balanced and symmetrical tree. The dendrograms of the trees generated using single linkage and Ward's method, respectively, in the last iteration of the walk-forward analysis are presented below in Figure 13. It is clear that the tree constructed using Ward's method is slightly more balanced, but that the tree constructed using single linkage does not suffer severely from chaining.



(a) Tree constructed using single linkage



(b) Tree constructed using Ward's method

Figure 13: Comparison of trees constructed using single linkage and Ward's method as linkage criterion.

Figure 14 presents the average path length of the two different trees for all time points in the walk-forward analysis for which portfolios are constructed. To begin with, the graph confirms the initial conclusions drawn from Figure 13, i.e. that the tree constructed using Ward’s method is more balanced since the value of the average path length is lower. However, it is interesting to note that the average path length changes from one day to another for both linkage criteria, which implies that the tree is changed from one day to another. In practice, this means that the hierarchical structure may be changed by adding one single data point to the data used to estimate the covariance matrix. As a result, it is clear that the HRP approach is very sensitive to changes in the input data and consequently may be considered relatively unstable. Of course, this contradicts López de Prado’s claim that HRP delivers stable results. Moreover, it is expected that the tree constructed using single linkage is even more unstable than that constructed using Ward’s method, as reflected by the high variation in the average path length for the single linkage.

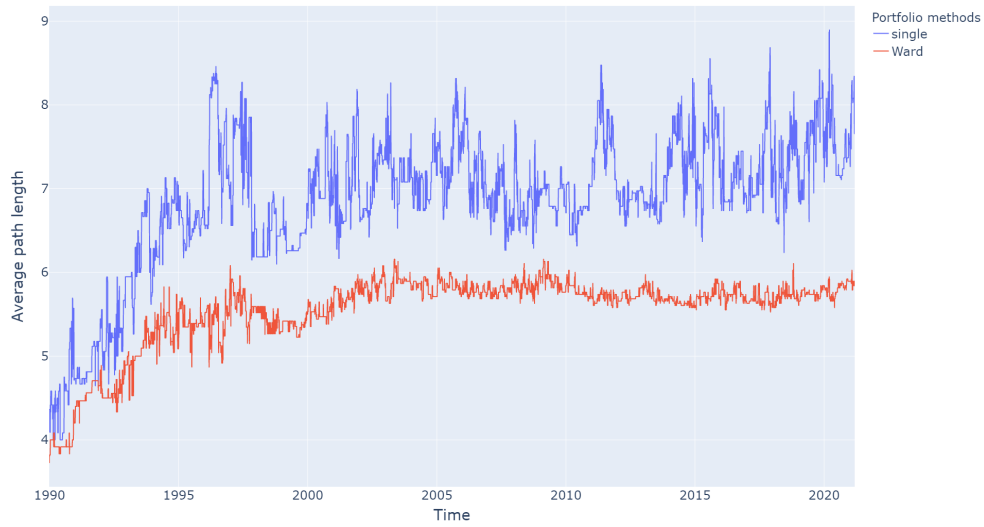


Figure 14: The average path length of the trees constructed using single linkage and Ward’s method, respectively.

The performance of each of the approaches is summarized in Table 3 and Table 4 below. In general, the results presented in Table 3 confirm the initial observations based on the equity curve. The portfolios constructed using the top-down and bottom-up approaches with single linkage perform better than the other three portfolios in terms of both annualized mean log-return and Sharpe ratio. Correspondingly, they also demonstrate slightly higher annualized volatility than the others. In terms of average and maximum drawdown, all five portfolios are considered to behave very similarly. However, it is clear that the original HRP approach results in a particularly high turnover compared to the others - almost 2.5x higher than the second highest. This behavior is also obvious from comparing the figures in Appendix B, which plot the portfolio weights over time. This result also refutes the hypothesis that the high turnover was caused by the tree suffering from chaining. However, it should be noted that all portfolios demonstrate rather higher turnovers, compared with the portfolios evaluated in the previous section. This property may be explained by the fact that the hierarchical structure appears to be unstable over time and sensitive to changes in the input data, as demonstrated by the average path length in Figure 14.

	Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Average drawdown	Maximum drawdown	Turnover
HRP	15.45e-03	17.03e-03	0.91	-9.59e-03	-110.85e-03	1443e-06
Top-down (single)	17.54e-03	17.56e-03	1.00	-10.13e-03	-104.52e-03	533e-06
Bottom-up (single)	17.43e-03	17.32e-03	1.01	-9.92e-03	-104.85e-03	591e-06
Top-down (Ward)	15.31e-03	16.86e-03	0.91	-9.70e-03	-98.92e-03	523e-06
Bottom-up (Ward)	15.12e-03	16.66e-03	0.91	-9.60e-03	-100.65e-03	552e-06

Table 3: Performance measures for all portfolio methods included in analysis.

Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Average drawdown	Maximum drawdown	Turnover
Top-down (single)	Bottom-up (Ward)	Bottom-up (single)	HRP	Top-down (Ward)	Top-down (Ward)
Bottom-up (single)	Top-down (Ward)	Top-down (single)	Bottom-up (Ward)	Bottom-up (Ward)	Top-down (single)
HRP	HRP	Bottom-up (Ward)	Top-down (Ward)	Top-down (single)	Bottom-up (Ward)
Top-down (Ward)	Bottom-up (single)	Top-down (Ward)	Bottom-up (single)	Bottom-up (single)	Bottom-up (single)
Bottom-up (Ward)	Top-down (single)	HRP	Top-down (single)	HRP	HRP

Table 4: Ranking of all portfolio methods included in analysis for the different performance measures.

6.3 Evaluation of New Approaches With Restriction on Growth

This section presents the result from the evaluation of the new approaches to HRP, where restriction on growth has been implemented in order to conclude whether this may result in reduced turnover. Given the results from section 6.2, the bottom-up and top-down approaches are implemented only using single linkage. As already explained in section 3.2.1 Hierarchical Clustering, restricting the growth of the tree implies cutting the tree at a certain cophenetic distance, neglecting all clusters formed below this distance and viewing all assets in the neglected clusters as part of the same cluster. As a result, selecting the cophenetic distance for which the tree should be cut is equivalent to selecting the number of clusters to have at the lowest hierarchical level of the tree. The selection of the optimal number of clusters is based on two measures, the average silhouette score and the within-cluster dissimilarity, calculated based on the full data set. Both measures indicate that 10 clusters are optimal, as presented in Figure 15.

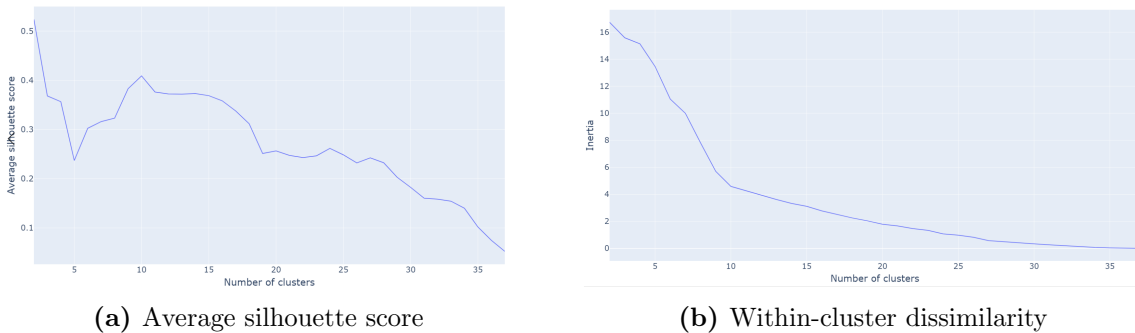


Figure 15: Scores used to evaluate how to restrict the growth of the tree, i.e. select the optimal number of clusters or the cophenetic distance for which the tree should be cut.

However, it is acknowledged that the optimal number of clusters is very much dependent on the research field and is often determined qualitatively by subject-matter experts. As a result, the number of clusters suggested by the average silhouette score and the within-cluster dissimilarity has been analyzed qualitatively by inspecting the resulting dendrogram, presented in Figure 16, and observing which assets are placed in the same clusters. In general, the clustering is very intuitive. The assets are first separated based on asset class, and then on geography. More details are presented in Table 5, where each cluster is described.

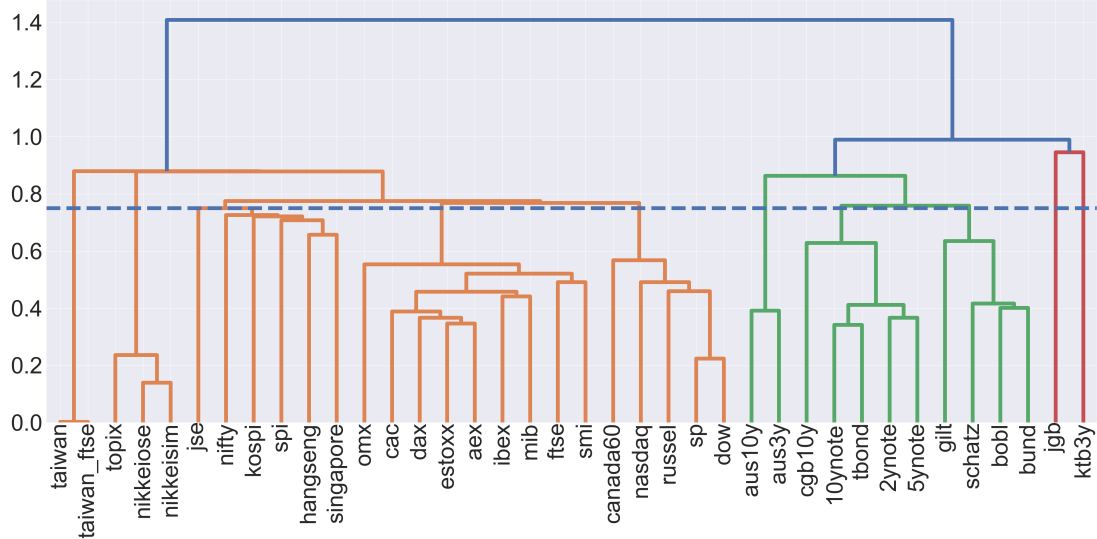


Figure 16: Dendrogram produced using single linkage. The dashed line corresponds to the cophenetic distance used to cut the tree, i.e. all clusters formed below this line are neglected and all assets are placed within one cluster. The cophenetic distance is approximately 0.75 and corresponds to 10 clusters.

Cluster	Assets	Description
1	taiwan, taiwan_ftse	The only two Taiwanese instruments
2	topix, nikkeiose, nikkeisim	The only Japanese equity index futures
3	jse, nifty, kospi, spi, hangseng, singapore	jse is the only African instrument spi is the only Oceanic equity index future The remaining are Asian equity index futures
4	omx, cac, dax, estoxx, aex, ibex, mib, ftse, smi	The only European equity index futures
5	canada60, nasdaq, russel, sp, dow	The only North American equity index futures
6	aus10y, aus3y	The only Oceanian bond futures
7	cgb10y, 10ynote, tbond, 2ynote, 5ynote	The only North American bond futures
8	gilt, schatz, bobl, bund	The only European bond futures
9	jgb	One single Asian bond future (Japan)
10	ktb3y	One single Asian bond future (Korea)

Table 5: Description of clusters resulting from having implemented 10 clusters and performed the clustering on the full data set. Cluster 1 is the first cluster when observing the dendrogram in Figure 16 from left to right. In this case, the term 'cluster' refers to a group of assets, and not the merging of two constituents as described in section 5 Analysis.

Since using 10 clusters is not only suggested when analyzing quantitative measures but also makes sense when analyzing the clusters qualitatively, the performance of the bottom-up and top-down approaches is evaluated with growth restricted to a maximum of 10 clusters. Since not all assets are present in the data set for the early dates on which portfolios are constructed, implying that for these dates one or several of these 10 clusters may be empty, the number of clusters is updated whenever a new asset is added to the data set.

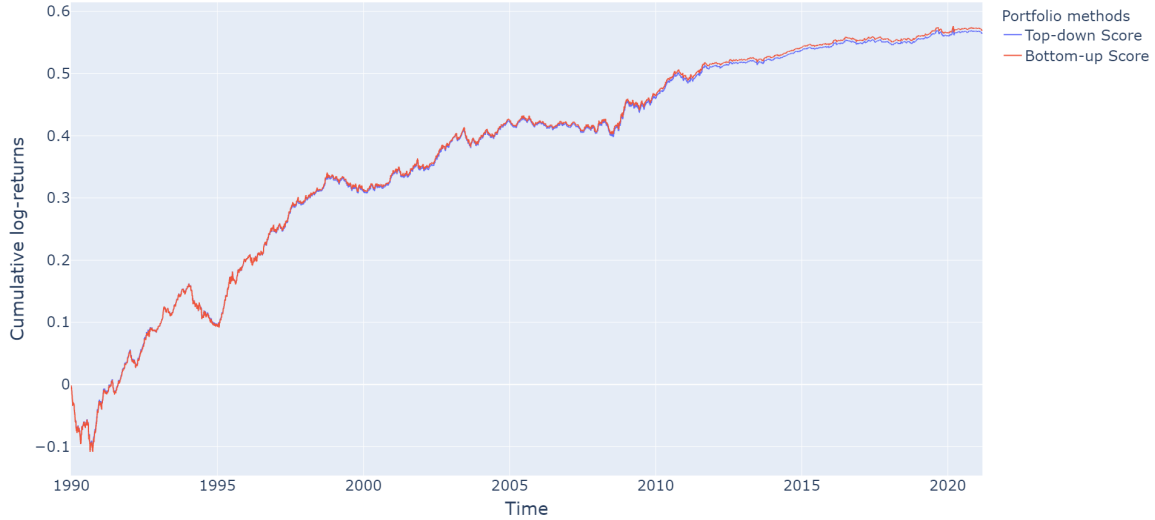


Figure 17: Equity curve, i.e. cumulative log-returns of portfolios constructed in walk-forward analysis using each portfolio construction method included in analysis.

It is clear from the equity curve in Figure 17 that both approaches behave very similarly, which is confirmed by the results in Table 6. It can be noted that the bottom-up approach has slightly higher annualized mean log-return, volatility and Sharpe ratio compared to the top-down approach. The bottom-up approach also has slightly higher turnover than the top-down approach. However, these differences are too small to be certain that this is truly the case and not the result of noise. It is also interesting to note that the restriction on growth has improved the turnover for both methods, albeit this improvement is very small.

	Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Average drawdown	Maximum drawdown	Turnover
Bottom-up	17.81e-03	17.63e-03	1.01	-10.13e-03	-107.09e-03	562e-06
Top-down	17.66e-03	17.59e-03	1.00	-10.13e-03	-105.40e-03	529e-06

Table 6: Performance measures for the bottom-up and the top-down approach with restriction on growth such that a maximum of 10 clusters are formed.

6.4 Summary

In summary, three separate evaluations have been performed. First, the original HRP approach was evaluated and compared to other more mainstream portfolio construction methods. Secondly, changes to the original HRP approach were made in order to fully make use of the hierarchical structure when allocating weights. As a result, two new approaches were developed - the bottom-up approach and the top-down approach. These two different approaches were implemented with both single linkage and Ward's method as linkage criterion. Thirdly, restriction on growth to a maximum of 10 clusters was implemented for the bottom-up and top-down approaches using single linkage. As a result, six new approaches to HRP have been evaluated in total:

- Bottom-up (single): Bottom-up approach with single linkage and no restriction on growth
- Top-down (single): Top-down approach with single linkage and no restriction on growth
- Bottom-up (Ward): Bottom-up approach with Ward's method as linkage and no restriction on growth
- Top-down (Ward): Top-down approach with Ward's method as linkage and no restriction on growth
- Bottom-up (restricted): Bottom-up approach with single linkage and restriction on growth
- Top-down (restricted): Top-down approach with single linkage and restriction on growth

In order to draw meaningful conclusions, it is interesting to compare the performances of all methods evaluated in this thesis. Table 7 presents the values for the performance measures deemed most relevant and Table 8 presents a ranking of the portfolio construction methods, from best to worst for each performance measure.

	Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Turnover
MinVar	3.30e-03	5.61e-03	0.59	307e-06
MinVar Long	10.33e-03	12.64e-03	0.82	1018e-06
ERC	20.00e-03	25.55e-03	0.78	246e-06
InvVol	20.49e-03	31.06e-03	0.66	106e-06
InvVar	15.91e-03	17.53e-03	0.91	176e-06
EW	26.94e-03	87.17e-03	0.31	4e-06
Original HRP	15.45e-03	17.03e-03	0.91	1443e-06
Top-down (single)	17.54e-03	17.56e-03	1.00	533e-06
Bottom-up (single)	17.43e-03	17.32e-03	1.01	591e-06
Top-down (Ward)	15.31e-03	16.86e-03	0.91	523e-06
Bottom-up (Ward)	15.12e-03	16.66e-03	0.91	552e-06
Top-down (restricted)	17.66e-03	17.59e-03	1.00	529e-06
Bottom-up (restricted)	17.81e-03	17.63e-03	1.01	562e-06

Table 7: Performance measures deemed most relevant for all portfolio methods evaluated in this thesis.

To begin with, it is interesting to note that all new methods implemented using single linkage outperform the original HRP approach, as well as all other benchmark portfolios, in terms of Sharpe ratio. The bottom-up approach with single linkage and restriction on growth has the highest Sharpe ratio of all methods evaluated, although it should be noted that all new methods implemented using single linkage are deemed comparable and the small differences insignificant.

In general, it is interesting to note that the performances of inverse-variance, the original HRP and the new approaches to HRP implemented using Ward's method are comparable for all measures except for turnover. The original HRP approach has a remarkably high turnover compared to all other methods except for the long-only minimum variance portfolio. However, one can argue that the difference in turnover between the long-only minimum variance portfolio and the original HRP portfolio is indeed also significant. Although the new approaches to HRP still have a rather high turnover, the values are 2-3x better than the original HRP approach. The restriction on growth improved the turnover slightly, but not significantly. Of course, a high turnover is especially problematic in the presence of high transaction costs, but nonetheless it is an indication of numerical instability of the portfolio.

Ann. mean log-return	Ann. volatility	Ann. Sharpe ratio	Turnover
EW	MinVar	Bottom-up (restricted)	EW
InvVol	MinVar Long	Bottom-up (single)	InvVol
ERC	Bottom-up (Ward)	Top-down (restricted)	InvVar
Bottom-up (restricted)	Top-down (Ward)	Top-down (single)	ERC
Top-down (restricted)	Original HRP	Top-down (Ward)	MinVar
Top-down (single)	Bottom-up (single)	InvVar	Top-down (Ward)
Bottom-up (single)	InvVar	Bottom-up (Ward)	Top-Down (restricted)
InvVar	Top-down (single)	Original HRP	Top-down (single)
Original HRP	Top-down (restricted)	MinVar Long	Bottom-up (Ward)
Top-down (Ward)	Bottom-up (restricted)	ERC	Bottom-up (restricted)
Bottom-up (Ward)	ERC	InvVol	Bottom-up (single)
MinVar Long	InvVol	MinVar	MinVar Long
MinVar	EW	EW	Original HRP

Table 8: Ranking of all portfolio methods evaluated in this thesis for the performance measures deemed most relevant.

7 Conclusion

To begin with, it is concluded that the original HRP outperforms all traditional benchmark portfolios, except for inverse-variance, in terms of Sharpe ratio for this particular data set and time period. However, HRP performs very much in line with inverse-variance in terms of annualized mean log-return, volatility and Sharpe ratio, which contradicts the results of Lopéz de Prado's simulations.

However, another key finding is that the original HRP approach has the undesirable property that the hierarchical structure produced using agglomerative clustering is not fully used when allocating weights between assets. This property results in a more balanced tree but also a remarkably high turnover. This result is not consistent with Lopéz de Prado's initial claim that the HRP portfolio is stable and indicates that it is indeed sensitive to errors in the estimation of the covariance matrix.

Alternative approaches to HRP, in which the hierarchical structure is fully used and the clustering is performed using single linkage, demonstrate superior out-of-sample performance compared to the original HRP approach in terms of annualized mean log-return, Sharpe ratio and turnover. However, although the turnover of these alternative approaches is only 35-40% of that of the original approach, it is still rather high compared to the other benchmark methods. This indicates that all approaches to HRP evaluated within the scope of this thesis are somewhat unstable and sensitive to estimation errors. Restricting the growth of the tree and effectively reducing the number of clusters did not improve the turnover significantly, at least not for the selected cophenetic distance.

In summary, it is concluded that applying hierarchical clustering to risk-based portfolio construction can improve the out-of-sample return and Sharpe ratio of the resulting portfolio. However, the resulting portfolio is also associated with high turnover, which may indicate numerical instability and sensitivity to errors in the estimate of the covariance matrix.

7.1 Research Aim

This section attempts to explain how the research aim stated in section 1.2 Research Aim has been achieved, using the methodology described in section 4 Methodology and the results and following discussion from section 6 Results & Discussion.

Evaluation of the Out-of-Sample Performance of HRP

Initial findings include that HRP is in fact affected by shrinkage, at least for this particular data set, and that HRP is closely related to inverse-variance in terms of annualized mean log-return, Sharpe ratio, volatility and drawdown. However, HRP and inverse-variance are both outperformed by EW, inverse-volatility and ERC in terms of annualized mean log-return and by minimum variance in terms of annualized volatility and drawdown. This may be considered intuitive but the latter contradicts the results from Lopéz de Prado's own simulations. They do, however, outperform all other methods in terms of Sharpe ratio, with a Sharpe ratio being approximately 14% higher than that of the portfolio with the third highest Sharpe ratio.

However, HRP separates itself from inverse-variance and all other benchmark portfolios with a remarkably high turnover. The turnover of the original HRP is approximately 1.4x higher than the portfolio with the second highest turnover, namely the long-only minimum variance portfolio, and more than 4x higher than that of the portfolio with the third highest turnover.

This is an especially surprising result given that the background used to motivate the intuition behind the strong out-of-sample performance of HRP is that traditional portfolios based on optimization are numerically unstable and therefore perform well in-sample but poorly out-of-sample. The high turnover of the HRP portfolio is not consistent with this reasoning and indicates that it suffers from numerical instability, at least for the data set used in this thesis.

Analysis of the Properties of HRP

Section 5 Analysis is devoted to the analysis of the algorithm and the properties of the HRP approach to portfolio construction. It is concluded that the recursive bisection used in the original HRP algorithm results in a more balanced tree, but also that the hierarchy identified using agglomerative clustering is not fully used when allocating weights to different assets. Analytically, this property implies that the order of the assets in the data set matter. For example, swapping places between two assets in the data set, and updating the rows and the columns of the covariance matrix accordingly, does not alter the data set itself and the two assets should receive the same weights regardless of this swap. The recursive bisection in Lopéz de Prado’s original algorithm ensures that this is not necessarily the case. Compared to more mainstream portfolio construction methods, this property is unconventional and consequently deemed undesirable.

Identification and Implementation of Potential Improvements to HRP

The recursive bisection in the original HRP algorithm also results in an inferior performance in terms of annualized mean log-return and Sharpe ratio compared to alternative approaches to HRP, where the clustering is performed using single linkage and the resulting hierarchy is fully used when allocating weights. It also appears to be a driver behind the remarkably high turnover of HRP, since the alternative approaches to HRP have a turnover that is 2-3x lower than that of the original HRP. As such, the bottom-up and top-down approaches to HRP implemented using single linkage and developed by the authors of this thesis are considered an improvement of the original HRP approach. Restricting the growth of the tree to a maximum of 10 clusters improved the Sharpe ratio and turnover slightly, but this improvement is not deemed statistically significant.

7.2 Future Work

For researchers and practitioners interested in studying this area, there exists several topics worth exploring further, some of which are stated below. To begin with, potential future research includes performing the same evaluation of HRP but using a data set consisting of other asset classes or covering other time periods. The notion of performance may also be broadened by incorporating other performance measures than the ones included in this thesis or to better reflect the practitioner’s objectives. Another option is to make use of other clustering methods than hierarchical clustering. Within the scope of hierarchical clustering, it would be interesting to evaluate different distance metrics and linkage criteria. Moreover, since weight allocation within and between clusters can be based on different portfolio construction methods, there are many different combinations that can be explored. It would also be interesting to investigate whether the optimal cophenetic distance associated with other scoring functions would have an effect on the turnover of the HRP portfolio. Lastly, another plausible area of future research entails investigating whether alterations to the walk-forward procedure may have an effect on the performance.

A Data

Index	Instrument	Asset class	Asset subclass	Region	Country	Exchange	Currency	Data available from
1	10ynote	Fixed Income	Bonds	North America	United States	CBOT	USD	1982-05-03
2	2ynote	Fixed Income	Bonds	North America	United States	CBOT	USD	1990-06-25
3	5ynote	Fixed Income	Bonds	North America	United States	CBOT	USD	1988-05-20
7	aus10y	Fixed Income	Bonds	Oceania	Australia	ASX	AUD	1989-01-01
8	aus3y	Fixed Income	Bonds	Oceania	Australia	ASX	AUD	1989-12-13
11	bobl	Fixed Income	Bonds	Europe	Germany	EUREX	EUR	1991-10-04
13	bund	Fixed Income	Bonds	Europe	Germany	EUREX	EUR	1988-09-29
14	cac	Equities	Indices	Europe	France	EURONEXT	EUR	1988-12-07
28	dax	Equities	Indices	Europe	Germany	EUREX	EUR	1990-11-23
29	estoxx	Equities	Indices	Europe	Germany	EUREX	EUR	1998-06-19
35	ftse	Equities	Indices	Europe	United Kingdom	LIFFE	GBP	1988-02-26
38	gilt	Fixed Income	Bonds	Europe	United Kingdom	LIFFE	GBP	1983-02-25
40	hangseng	Equities	Indices	Asia	Hong Kong	HKFE	HKD	1992-04-01
42	ibex	Equities	Indices	Europe	Spain	MEFF	EUR	1992-06-01
43	jgb	Fixed Income	Bonds	Asia	Japan	OSE	JPY	1985-10-21
47	mib	Equities	Indices	Europe	Italy	IDEM	EUR	1995-03-15
49	nasdaq	Equities	Indices	North America	United States	CME	USD	1996-04-09
51	nikkeiose	Equities	Indices	Asia	Japan	OSE	JPY	1988-09-05
52	nikkeisim	Equities	Indices	Asia	Japan	SGX	JPY	1989-09-27
57	omx	Equities	Indices	Europe	Sweden	OM	SEK	1993-05-21
60	russel	Equities	Indices	North America	United States	CME	USD	1993-02-03
61	schatz	Fixed Income	Bonds	Europe	Germany	EUREX	EUR	1997-03-07
63	smi	Equities	Indices	Europe	Switzerland	EUREX	CHF	1998-10-01
67	sp	Equities	Indices	North America	United States	CME	USD	1982-04-21
68	spi	Equities	Indices	Oceania	Australia	ASX	AUD	2000-05-02
72	tbond	Fixed Income	Bonds	North America	United States	CBOT	USD	1977-11-26
80	cgb10y	Fixed Income	Bonds	North America	Canada	ME	CAD	1989-09-15
81	canada60	Equities	Indices	North America	Canada	ME	CAD	1999-10-01
82	dow	Equities	Indices	North America	United States	CBOT	USD	1997-10-06
84	taiwan	Equities	Indices	Asia	Singapore	SGX	USD	1997-05-01
87	aex	Equities	Indices	Europe	Netherlands	LIFFE	EUR	1989-01-01
88	topix	Equities	Indices	Asia	Japan	OSE	JPY	1990-05-16
148	nifty	Equities	Indices	Asia	Singapore	SGX	USD	2000-11-01
155	kospi	Equities	Indices	Asia	Korea	KRX	KRW	1996-09-12
163	jse	Equities	Indices	Africa	South Africa	JSE	ZAR	1995-09-21
165	ktb3y	Fixed Income	Bonds	Asia	Korea	KRX	KRW	1999-12-03
168	singapore	Equities	Indices	Asia	Singapore	SGX	SGD	1998-10-29
194	taiwan_ftse	Equities	Indices	Asia	Singapore	SGX	USD	1997-05-01

Table 9: Descriptive metadata for each instrument in data set used to evaluate the HRP method. The 38 instruments included in the data set are equity index and bond futures with first day of trading during the year of 2000 at the latest.

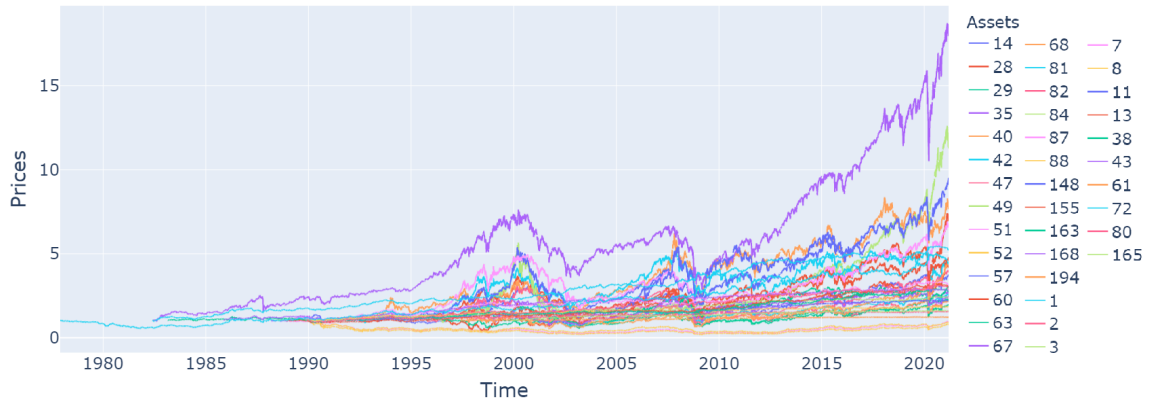


Figure 18: Price series for all instruments in the data set for the full time period of available data.

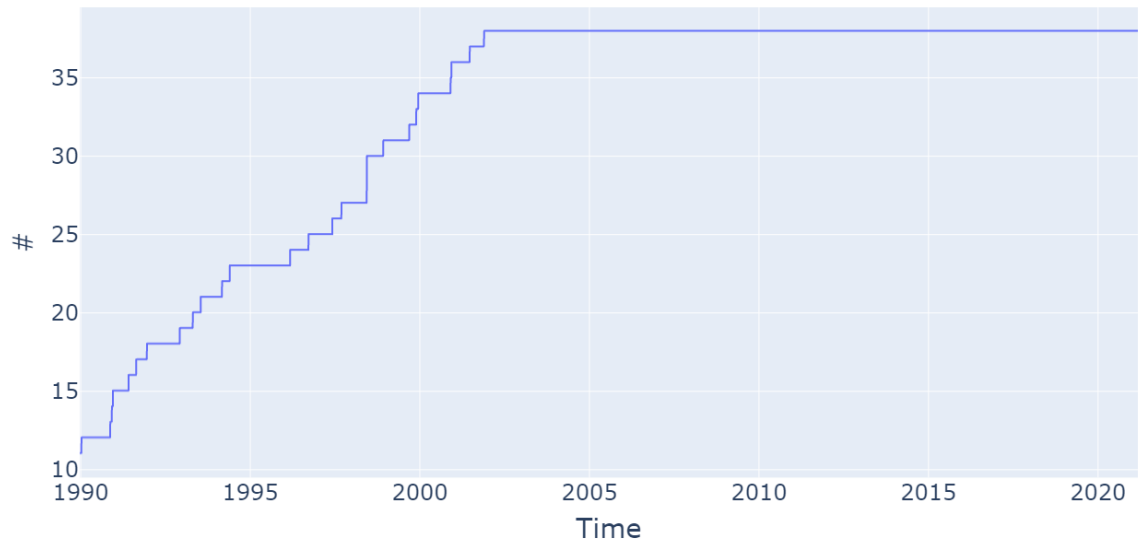


Figure 19: Number of assets used to construct portfolios in walk-forward analysis. Roughly one year of historical price data required for an instrument to be included in a portfolio constructed at a particular point in time. A portfolio is constructed at a particular point in time if and only if there exists at least ten instruments that can be included in the portfolio. The total number of instruments is 38.

B Portfolio Weights

Figures for the weights for all portfolio construction methods evaluated in this thesis have not been included in order to limit the length of the thesis. The weight figures have been included for the inverse-variance portfolio and the original HRP portfolio, since these two are the most similar methods out of the methods included in the first evaluation. Moreover, the weight plots for the bottom-up approach with single linkage, with and without restriction on growth, are included since the bottom-up approach has slightly higher turnover than the top-down and since single linkage results in higher turnover than Ward's method.

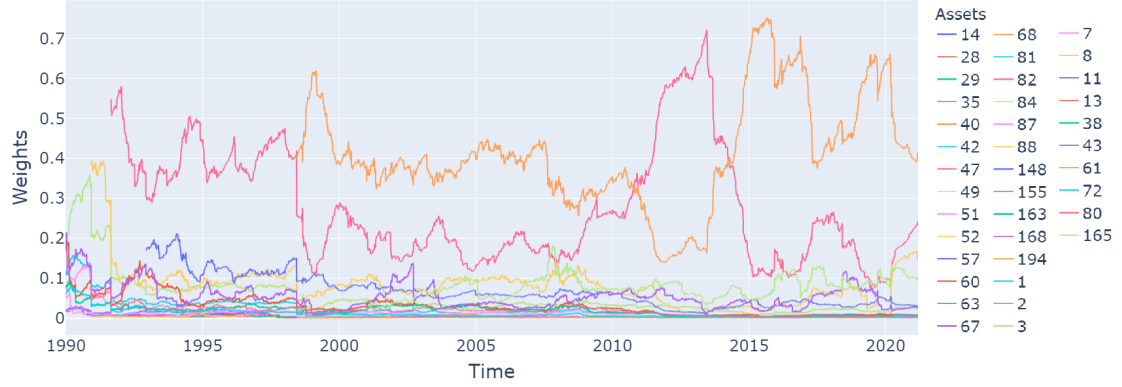


Figure 20: Portfolio weights for all portfolios constructed in walk-forward analysis using inverse-variance.

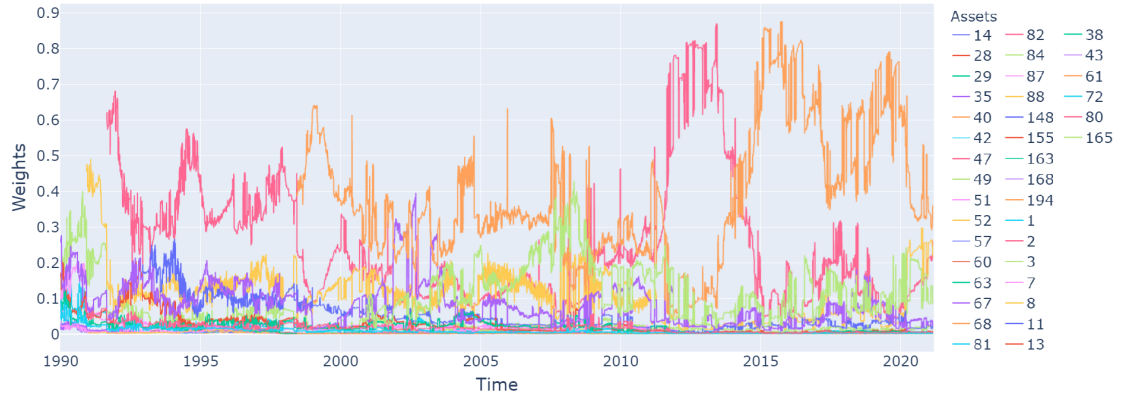


Figure 21: Portfolio weights for all portfolios constructed in walk-forward analysis using the original approach to HRP.

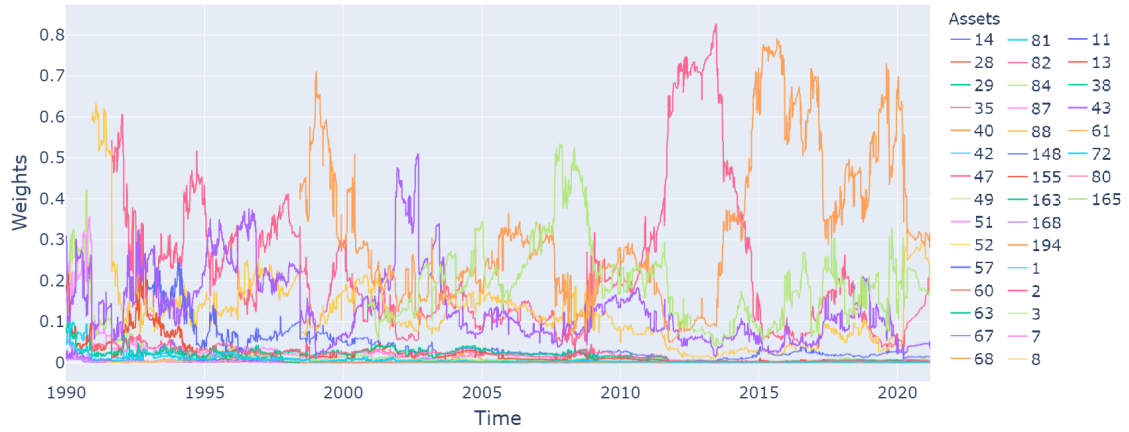


Figure 22: Portfolio weights for all portfolios constructed in walk-forward analysis using the bottom-up approach to HRP with single linkage as linkage criterion and no restriction on growth.

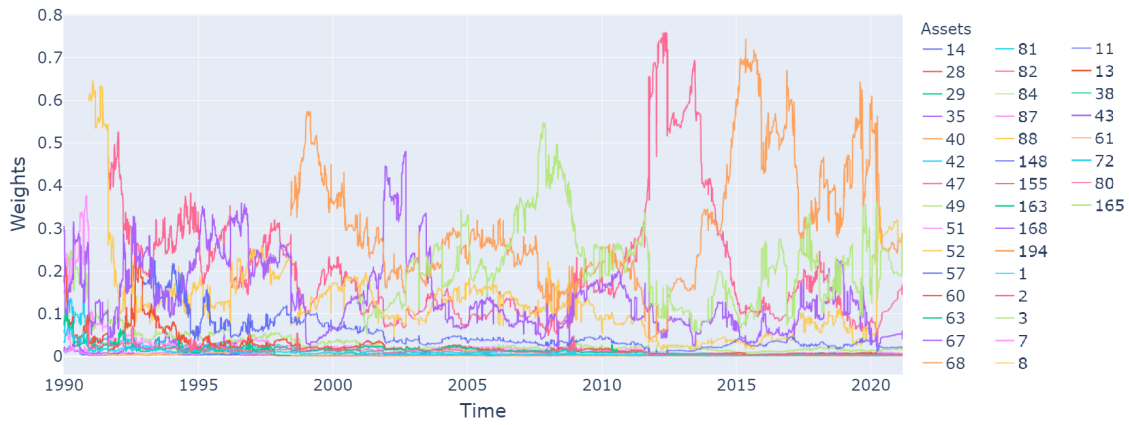


Figure 23: Portfolio weights for all portfolios constructed in walk-forward analysis using the bottom-up approach to HRP with single linkage as linkage criterion and restriction on growth.

C Numeric Examples for New Approaches to HRP

Notation

w_i = Weight of asset with index i

σ_i^2 = Variance of asset or cluster with index i

ρ_{ij} = Correlation coefficient between asset i and asset j

Σ_k = Covariance matrix after iteration k

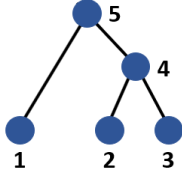
C.1 Three Assets

Initiation

$$w_1 = w_2 = w_3 = 1$$

$$\Sigma_0 = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \sigma_2\sigma_3\rho_{23} \\ \sigma_1\sigma_3\rho_{13} & \sigma_2\sigma_3\rho_{23} & \sigma_3^2 \end{pmatrix}$$

Assume that the correlations between the assets are such that the following tree is constructed using hierarchical clustering:



C.1.1 Top-Down

Iteration 1: Cluster 5

1. Calculate weights for assets in left branch according to inverse-variance

Since the left branch consists of only one asset, we know that $\tilde{w}_1 = 1$.

2. Determine the variance of the left branch

Since the left branch consists of only one asset, the variance of the left branch is equal to the variance of asset 1, σ_1^2 .

3. Calculate weights for assets in right branch according to inverse-variance

$$\tilde{w}_2 = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} \quad \# \text{ left child}$$

$$\tilde{w}_3 = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2} \quad \# \text{ right child}$$

$$\tilde{w}_4 = (\tilde{w}_2, \tilde{w}_3) \quad \# \text{ set } \tilde{w}_4 \text{ as column vector of all weights in cluster}$$

4. Calculate variance of the right branch

The weights calculated in the previous step, \tilde{w}_2 and \tilde{w}_3 , are used to determine the total variance of cluster 4, σ_4^2 .

$$\sigma_4^2 = \tilde{w}_4^T \Sigma_{2,3} \tilde{w}_4 = \tilde{w}_2^2 \sigma_2^2 + 2\tilde{w}_2\tilde{w}_3\sigma_2\sigma_3\rho_{23} + \tilde{w}_3^2 \sigma_3^2$$

In the above expression, $\Sigma_{2,3}$ is a matrix containing the rows and the columns of the covariance matrix corresponding to assets 2 and 3.

5. Determine scaling factor used to allocate weights between the left and the right branch

$$\alpha_1 = \frac{\sigma_4^2}{\sigma_1^2 + \sigma_4^2}$$

6. Scale weights

$$\begin{aligned} w_1 &= \alpha_1 * w_1 && \# \text{ left branch} \\ w_4 &= (w_2, w_3) && \# \text{ column vector of all assets in branch} \\ w_4 &= (1 - \alpha_1) * w_4 && \# \text{ right branch} \end{aligned}$$

Iteration 2: Cluster 4

Since cluster 4 consists only of assets and no other clusters, we skip steps 1-4 and jump directly to step 5.

5. Determine scaling factor used to allocate weights between the left and the right branch

$$\alpha_2 = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2}$$

6. Scale weights

$$\begin{aligned} w_2 &= \alpha_2 * w_2 && \# \text{ left branch} \\ w_3 &= (1 - \alpha_2) * w_3 && \# \text{ right branch} \end{aligned}$$

Final weights

$$\begin{aligned} w_1 &= \alpha_1 = \frac{\sigma_4^2}{\sigma_1^2 + \sigma_4^2} \\ w_2 &= (1 - \alpha_1) * \alpha_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_4^2} * \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} \\ w_3 &= (1 - \alpha_1) * (1 - \alpha_2) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_4^2} * \frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2} \end{aligned}$$

C.1.2 Bottom-Up

Iteration 1: Asset 2 and 3 are joint to form cluster 4

1. Determine weights for assets in cluster

$$\begin{aligned} \alpha_1 &= \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} && \# \text{ calculate scaling factor acc. to inverse-variance} \\ w_2 &= \alpha_1 * w_2 && \# \text{ scale weight of left child in cluster} \\ w_3 &= (1 - \alpha_1) * w_3 && \# \text{ scale weight of right child in cluster} \end{aligned}$$

2. Update the covariance matrix so that cluster 4 replaces assets 2 and 3

After each iteration, the formed cluster replaces its left and right branches. In the subsequent iteration, the newly formed cluster is treated as any other asset. The covariance matrix is updated using the following scaling factor:

$$A_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha_1 & 1 - \alpha_1 \end{pmatrix}$$

Finally, the new covariance matrix is determined:

$$\Sigma_1 = A_0 \Sigma_0 A_0^T = \begin{pmatrix} \sigma_1^2 & \alpha_1 \sigma_1 \sigma_2 \rho_{12} + (1 - \alpha_1) \sigma_1 \sigma_3 \rho_{13} \\ \alpha_1 \sigma_1 \sigma_2 \rho_{12} + (1 - \alpha_1) \sigma_1 \sigma_3 \rho_{13} & \sigma_4^2 \end{pmatrix}$$

where $\sigma_4^2 = \alpha_1^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \sigma_3^2$

Iteration 2: Asset 1 and cluster 4 are joint to form cluster 5 (root of tree)

1. Determine weights for assets in cluster

$$\begin{aligned}\alpha_2 &= \frac{\sigma_4^2}{\sigma_1^2 + \sigma_4^2} && \# \text{ calculate scaling factor acc. to inverse-variance} \\ w_1 &= \alpha_2 * w_1 && \# \text{ scale weight of left child of cluster} \\ w_4 &= (w_2, w_3) && \# \text{ set } w_4 \text{ as a column vector of all weights in cluster} \\ w_4 &= (1 - \alpha_2) * w_4 && \# \text{ scale weights of all assets in right branch of cluster}\end{aligned}$$

2. Update the covariance matrix so that cluster 5 replaces asset 1 and cluster 5

Since cluster 5 is the root of the tree, this procedure is equal to calculating the variance of the portfolio. The scaling factor used is:

$$A_1 = (\alpha_2 \ 1 - \alpha_2)$$

Finally, the variance is determine:

$$\Sigma_2 = A_1 \Sigma_1 A_1^T = \sigma_5^2 = \alpha_2^2 \sigma_1^2 + 2\alpha_2(1 - \alpha_2)(\alpha_1 \sigma_1 \sigma_2 \rho_{12} + (1 - \alpha_1) \sigma_1 \sigma_3 \rho_{13}) + (1 - \alpha_2)^2 \sigma_4^2$$

Final weights

$$\begin{aligned}w_1 &= \alpha_2 = \frac{\sigma_4^2}{\sigma_1^2 + \sigma_4^2} \\ w_2 &= (1 - \alpha_2) * \alpha_1 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_4^2} * \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} \\ w_3 &= (1 - \alpha_2) * (1 - \alpha_1) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_4^2} * \frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2}\end{aligned}$$

C.1.3 Comparison

It is clear that for $N = 3$ assets, the top-down approach and the bottom-up approach render identical weights. Furthermore, it is interesting to explore the special case of zero correlation between assets.

Special case: Zero correlation

In the special case of zero correlation between the assets, i.e. $\rho_{12} = \rho_{13} = \rho_{23} = 0$, above expressions for the weights can be simplified. To begin with, determine the variance of cluster 4, using notation used in the bottom-up approach:

$$\begin{aligned}\sigma_4^2 &= \alpha_1^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \sigma_3^2 = \{\rho_{23} = 0\} = \alpha_1^2 \sigma_2^2 + (1 - \alpha_1)^2 \sigma_3^2 \\ &= \{\alpha_1 = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2}\} = \left(\frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2}\right)^2 \sigma_2^2 + \left(\frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2}\right)^2 \sigma_3^2 = \frac{\sigma_2^2 \sigma_3^4 + \sigma_2^4 \sigma_3^2}{(\sigma_2^2 + \sigma_3^2)^2} = \sigma_2^2 \sigma_3^2 * \frac{\sigma_2^2 + \sigma_3^2}{(\sigma_2^2 + \sigma_3^2)^2} \\ &= \frac{\sigma_2^2 \sigma_3^2}{\sigma_2^2 + \sigma_3^2}\end{aligned}$$

The expression σ_4^2 is then inserted into the expression for the weight of asset 1:

$$w_1 = \alpha_2 = \frac{\sigma_4^2}{\sigma_1^2 + \sigma_4^2} = \frac{\frac{\sigma_2^2 \sigma_3^2}{\sigma_2^2 + \sigma_3^2}}{\sigma_1^2 + \frac{\sigma_2^2 \sigma_3^2}{\sigma_2^2 + \sigma_3^2}} = \frac{\sigma_2^2 \sigma_3^2}{\sigma_1^2 (\sigma_2^2 + \sigma_3^2) + \sigma_2^2 \sigma_3^2} = \frac{1}{\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_1^2}{\sigma_3^2} + 1} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}}$$

The same procedure can be repeated for assets 2 and 3 to show that the below holds:

$$w_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}}$$

$$w_3 = \frac{\frac{1}{\sigma_3^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}}$$

From these expressions, it is possible to draw the conclusion that, under the assumption of zero correlations, both the bottom-up and the top-down approach to HRP is equal to using inverse-variance weighting.

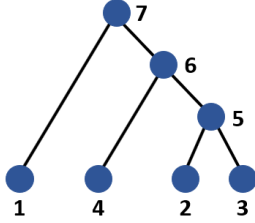
C.2 Four Assets

Initiation

$$w_1 = w_2 = w_3 = w_4 = 1$$

$$\Sigma_0 = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} & \sigma_1\sigma_4\rho_{14} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \sigma_2\sigma_3\rho_{23} & \sigma_2\sigma_4\rho_{24} \\ \sigma_1\sigma_3\rho_{13} & \sigma_2\sigma_3\rho_{23} & \sigma_3^2 & \sigma_3\sigma_4\rho_{34} \\ \sigma_1\sigma_4\rho_{14} & \sigma_2\sigma_4\rho_{24} & \sigma_3\sigma_4\rho_{34} & \sigma_4^2 \end{pmatrix}$$

Assume that the correlations between the assets are such that the following tree is constructed using hierarchical clustering:



C.2.1 Top-Down

Iteration 1: Cluster 7

1. Calculate weights for assets in left branch according to inverse-variance

Since the left branch consists of only one asset, we know that $\tilde{w}_1 = 1$.

2. Determine the variance of the left branch

Since the left branch consists of only one asset, the variance of the left branch is equal to the variance of asset 1, σ_1^2 .

3. Calculate weights for assets in right branch according to inverse-variance

$$\tilde{w}_i = \frac{\frac{1}{\sigma_i^2}}{\frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2} + \frac{1}{\sigma_4^2}} \quad \# \text{ calculate weights for all assets in right branch, i.e. for } i = 2, 3, 4$$

$$\tilde{w}_6 = (\tilde{w}_2, \tilde{w}_3, \tilde{w}_4) \quad \# \text{ set } w_6 \text{ as column vector of all weights in branch}$$

4. Calculate variance of the right branch

The weights calculated in the previous step, \tilde{w}_2 , \tilde{w}_3 and \tilde{w}_4 , are used to determine the total variance of cluster 6, σ_6^2 :

$$\sigma_6^2 = \tilde{w}_6^T \Sigma_{2,3,4} \tilde{w}_6$$

In the above expression, $\Sigma_{2,3,4}$ is a matrix containing the rows and the columns of the covariance matrix corresponding to assets in the branch.

5. Determine scaling factor used to allocate weights between the left and the right branch

$$\alpha_1 = \frac{\sigma_6^2}{\sigma_1^2 + \sigma_6^2}$$

6. Scale weights

$$\begin{aligned} w_1 &= \alpha_1 * w_1 && \# \text{ left branch} \\ w_6 &= (w_2, w_3, w_4) && \# \text{ column vector of all assets in right branch} \\ w_6 &= (1 - \alpha_1) * w_6 && \# \text{ right branch} \end{aligned}$$

Iteration 2: Cluster 6

1. Calculate weights for assets in left branch according to inverse-variance

Since the left branch consists of only one asset, we know that $\tilde{w}_4 = 1$.

2. Determine the variance of the left branch

Since the left branch consists of only one asset, the variance of the left branch is equal to the variance of asset 4, σ_4^2 .

3. Calculate weights for assets in right branch according to inverse-variance

$$\begin{aligned} \tilde{w}_i &= \frac{\frac{1}{\sigma_i^2}}{\frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}} && \# \text{ calculate weights for all assets in right branch, i.e. for } i = 2, 3 \\ \tilde{w}_5 &= (\tilde{w}_2, \tilde{w}_3) && \# \text{ set } w_5 \text{ as column vector of all weights in branch} \end{aligned}$$

4. Calculate variance of the right branch

The weights calculated in the previous step, \tilde{w}_2 and \tilde{w}_3 , are used to determine the total variance of cluster 5, σ_5^2 :

$$\sigma_5^2 = \tilde{w}_5^T \Sigma_{2,3} \tilde{w}_5$$

In the above expression, $\Sigma_{2,3}$ is a matrix containing the rows and the columns of the covariance matrix corresponding to assets in the branch.

5. Determine scaling factor used to allocate weights between the left and the right branch

$$\alpha_2 = \frac{\sigma_5^2}{\sigma_4^2 + \sigma_5^2}$$

6. Scale weights

$$\begin{aligned} w_4 &= \alpha_2 * w_4 && \# \text{ left branch} \\ w_5 &= (w_2, w_3) && \# \text{ column vector of all assets in right branch} \\ w_5 &= (1 - \alpha_2) * w_5 && \# \text{ right branch} \end{aligned}$$

Iteration 3: Cluster 5

Since cluster 5 consists only of assets and no other clusters, we skip steps 1-4 and jump directly to step 5.

5. Determine scaling factor used to allocate weights between the left and the right branch

$$\alpha_3 = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2}$$

6. Scale weights

$$\begin{aligned} w_2 &= \alpha_3 * w_2 && \# \text{ left branch} \\ w_3 &= (1 - \alpha_3) * w_3 && \# \text{ right branch} \end{aligned}$$

Final weights

$$\begin{aligned} w_1 &= \alpha_1 = \frac{\sigma_6^2}{\sigma_1^2 + \sigma_6^2} \\ w_2 &= (1 - \alpha_1) * (1 - \alpha_2) * \alpha_3 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} * \frac{\sigma_4^2}{\sigma_4^2 + \sigma_5^2} * \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} \\ w_3 &= (1 - \alpha_1) * (1 - \alpha_2) * (1 - \alpha_3) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} * \frac{\sigma_4^2}{\sigma_4^2 + \sigma_5^2} * \frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2} \\ w_4 &= (1 - \alpha_1) * \alpha_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} * \frac{\sigma_5^2}{\sigma_4^2 + \sigma_5^2} \end{aligned}$$

C.2.2 Bottom-Up

Iteration 1: Asset 2 and 3 are joint to form cluster 5

1. Determine weights for assets in cluster

$$\begin{aligned} \alpha_1 &= \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} && \# \text{ calculate scaling factor acc. to inverse-variance} \\ w_2 &= \alpha_1 * w_2 && \# \text{ scale weight of left child in cluster} \\ w_3 &= (1 - \alpha_1) * w_3 && \# \text{ scale weight of right child in cluster} \end{aligned}$$

2. Update the covariance matrix so that cluster 5 replaces assets 2 and 3

After each iteration, the formed cluster replaces its left and right branches. In the subsequent iteration, the newly formed cluster is treated as any other asset. The covariance matrix is updated using the following scaling factor:

$$A_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha_1 & 1 - \alpha_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finally, the new covariance matrix is determined:

$$\Sigma_1 = A_0 \Sigma_0 A_0^T = \begin{pmatrix} \sigma_1^2 & r_{15} & r_{14} \\ r_{15} & \sigma_5^2 & r_{45} \\ r_{14} & r_{45} & \sigma_4^2 \end{pmatrix}$$

where $\sigma_5^2 = \alpha_1^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \sigma_3^2$

Iteration 2: Asset 4 and cluster 5 are joint to form cluster 6

1. Determine weights for assets in cluster

$$\begin{aligned}
\alpha_2 &= \frac{\sigma_5^2}{\sigma_4^2 + \sigma_5^2} & \# \text{ calculate scaling factor acc. to inverse-variance} \\
w_4 &= \alpha_2 * w_4 & \# \text{ scale weight of left child of cluster} \\
w_5 &= (w_2, w_3) & \# \text{ set } w_5 \text{ as a column vector of all weights in cluster} \\
w_5 &= (1 - \alpha_2) * w_5 & \# \text{ scale weights of all assets in right branch of cluster}
\end{aligned}$$

2. Update the covariance matrix so that cluster 6 replaces asset 4 and cluster 5

After each iteration, the formed cluster replaces its left and right branches. In the subsequent iteration, the newly formed cluster is treated as any other asset. The covariance matrix is updated using the following scaling factor:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha_2 & 1 - \alpha_2 \end{pmatrix}$$

Finally, the new covariance matrix is determined:

$$\Sigma_2 = A_1 \Sigma_1 A_1^T = \begin{pmatrix} \sigma_1^2 & r_{16} \\ r_{16} & \sigma_6^2 \end{pmatrix}$$

where

$$\begin{aligned}
\sigma_6^2 &= \alpha_2^2 \sigma_5^2 + 2\alpha_2(1 - \alpha_2)r_{45} + (1 - \alpha_2)^2 \sigma_4^2 = \{\sigma_5^2 = \alpha_1^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \sigma_3^2\} \\
&= \alpha_1^2 \alpha_2^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\alpha_2^2 \sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \alpha_2^2 \sigma_3^2 + 2\alpha_2(1 - \alpha_2)r_{45} + (1 - \alpha_2)^2 \sigma_4^2 \\
&= \{r_{45} = \alpha_1\sigma_2\sigma_4\rho_{24} + (1 - \alpha_1)\sigma_3\sigma_4\rho_{34}\} = \alpha_1^2 \alpha_2^2 \sigma_2^2 + 2\alpha_1(1 - \alpha_1)\alpha_2^2 \sigma_2\sigma_3\rho_{23} + (1 - \alpha_1)^2 \alpha_2^2 \sigma_3^2 \\
&\quad + 2\alpha_1\alpha_2(1 - \alpha_2)\sigma_2\sigma_4\rho_{24} + 2(1 - \alpha_1)\alpha_2(1 - \alpha_2)\sigma_3\sigma_4\rho_{34} + (1 - \alpha_2)^2 \sigma_4^2
\end{aligned}$$

Iteration 3: Asset 1 and cluster 6 are joint to form cluster 7 (root of tree)

1. Determine weights for assets in cluster

$$\begin{aligned}
\alpha_3 &= \frac{\sigma_6^2}{\sigma_1^2 + \sigma_6^2} & \# \text{ calculate scaling factor acc. to inverse-variance} \\
w_1 &= \alpha_3 * w_1 & \# \text{ scale weight of left child of cluster} \\
w_6 &= (w_2, w_3, w_4) & \# \text{ set } w_6 \text{ as a column vector of all weights in cluster} \\
w_6 &= (1 - \alpha_3) * w_6 & \# \text{ scale weights of all assets in right branch of cluster}
\end{aligned}$$

2. Update the covariance matrix so that cluster 7 replaces asset 1 and cluster 6

Since cluster 7 is the root of the tree, this procedure is equal to calculating the variance of the portfolio. The scaling factor used is:

$$A_2 = (\alpha_3 \quad 1 - \alpha_3)$$

Finally, the variance is determined:

$$\Sigma_3 = A_2 \Sigma_2 A_2^T = \sigma_7^2 = \alpha_3^2 \sigma_1^2 + 2\alpha_3(1 - \alpha_3)r_{16} + (1 - \alpha_3)^2 \sigma_6^2$$

Final weights

$$\begin{aligned}
w_1 &= \alpha_3 = \frac{\sigma_6^2}{\sigma_1^2 + \sigma_6^2} \\
w_2 &= \alpha_1(1 - \alpha_2)(1 - \alpha_3) = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2} * \frac{\sigma_4^2}{\sigma_4^2 + \sigma_5^2} * \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} \\
w_3 &= (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3) = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_3^2} * \frac{\sigma_4^2}{\sigma_4^2 + \sigma_5^2} * \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2} \\
w_4 &= \alpha_2(1 - \alpha_3) = \frac{\sigma_5^2}{\sigma_4^2 + \sigma_5^2} * \frac{\sigma_1^2}{\sigma_1^2 + \sigma_6^2}
\end{aligned}$$

C.2.3 Comparison

When comparing the final weights produced using the top-down and the bottom-up approaches, respectively, it is clear that the expressions in terms of σ_1^2 , σ_2^2 , σ_3^2 , σ_4^2 , σ_5^2 and σ_6^2 are identical. However, that does not imply that the weights are in fact identical when calculated using the two different approaches. This is explained by the fact that the variance of a cluster consisting of at least one other cluster, σ_6^2 in this case, is calculated differently. In the bottom-up approach, the variance of cluster 6 is calculated using inverse-variance allocation between asset 4 and cluster 5, where the variance of cluster 5 is determined through inverse-variance allocation between assets 2 and 3. In the top-down approach, on the other hand, the variance of cluster 6 is calculated using inverse-variance allocation between all assets in the cluster, i.e. as if assets 2, 3 and 4 were all situated on the same hierarchical level of the tree. This is intuitive but all the more difficult to prove algebraically, given the complexity of the expressions. Instead, this is confirmed using Python for a number of different covariance matrices.

References

- D. Ardia, D. Bolliger, K. Boudt, and J.-P. Gagnon-Fleure. The impact of covariance misspecification in risk-based portfolios. *Annals of Operations Research*, 253:1–16, 2017.
- D. Bailey and M. López de Prado. Balanced baskets: A new approach to trading and hedging risks. *Journal of Investment Strategies*, 1(4):21–62, 2012.
- S. Bernartzi and R.H. Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98, 2001.
- F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48:28–43, 1992.
- Z. Bodie, A. Kane, and A. J. Marcus. *Investments*. McGraw-Hill Education, New York, 2014.
- D. Booth and E. Fama. Diversification and asset contributions. *Financial Analyst Journal*, 48(3):2632, 1992.
- Y. Brandes, I. Domowitz, and V. Serbin. *Transaction Costs and Equity Portfolio Capacity Analysis*. In B. Scherer, I& K. Winston, *The Oxford Handbook of Quantitative Asset Management*. OUP Oxford, 2012.
- B. Bruder and T. Roncalli. Managing risk exposures using the risk budgeting approach. *Risk Management eJournal*, 2012.
- T. Burggraf and A. Vyas. Beyond risk parity – a machine learning-based hierarchical risk parity approach on cryptocurrencies. *Finance Research Letters*, 38, 2020.
- O. Carchano and A. Pardo. Rolling over stock index futures contracts. *The Journal of Futures Markets*, 29(7):684–694, 2009.
- E. P. Chan. *Algorithmic Trading: Winning Strategies and Their Rationale*. John Wiley I& Sons, Inc., New Jersey, 2013.
- R. Clarke, H. De Silva, and S. Thorley. Portfolio constraints and the fundamental law of active management. *Financial Analysts Journal*, 58:48–66, 2002.
- V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22:1915–1953, 2009.
- F. Duarte and L. De Castro. A framework to perform asset allocation based on partitional clustering. *IEEE Access*, PP, 2020. doi: 10.1109/ACCESS.2020.3001944-2.
- R. Fernholtz, R. Garvy, and J. Hannon. Diversity-weighted indexing. *Journal of Portfolio Management*, 4(2):74–82, 1998.
- L. Goldberg and O. Mahmoud. Drawdown: From practice to theory and back again. *Mathematics and Financial Economics*, 11:275–297, 2017.
- T. Hastie and R. Tibshirani. *The elements of statistical learning: Data mining, inference, and prediction*. Springer., New Jersey, 2001.

- W. Huang. Performance of hierarchical equal risk contribution algorithm in china market. 2020.
- J.C. Hull. *Options, Futures and Other Derivatives*. Pearson Education Ltd, Harlow, 2015.
- H. Hult, F. Lindskog, O. Hammarlid, and C.J. Rehn. *Risk and Portfolio Analysis: Principles and Methods*. Springer, 2010.
- R. Jagannathan and R. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance, American Finance Association*, 58(4):1651–1684, 2003.
- A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. 31, 1999.
- L. Kaufman and P. J. Rousseeuw. *Finding groups in data: An introduction to cluster analysis*. Wiley, New York, 1990.
- P. N. Kolm, R. Tütüncü, and F.J. Fabozzi. 60 years of portfolio optimization. *European Journal of Operational Research*, 234(2):356–371, 2010.
- Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15*, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450340052. doi: 10.1145/2833157.2833162. URL <https://doi.org/10.1145/2833157.2833162>.
- A. Lau, M Kolanovic, R. Krishnamachari, and T. Lee. Cross asset portfolios of tradable risk premia indices. 2017.
- R. P. C. Leal and B. Vaz de Melo Mendes. Maximum drawdown: Models i& applications. *The Journal of Alternative Investments*, 7(4):83–91, 2005.
- O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.
- O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.
- D. Leon, A. Aragon, J. Sandoval, G. Hernandez, A. Arevalo, and J. Nino. Clustering algorithms for risk-adjusted portfolio clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, 108:1334–1343, 2017.
- R. Leote de Carvalho, X. Lu, and P. Moulin. Demystifying equity risk-based strategies: A simple alpha plus beta description. *The Journal of Portfolio Management*, 38:56–79, 2012.
- R. Litterman. *Modern Investment Management*. John wiley I& Sons Inc., New Jersey, 2003.
- H. Lohre, C. Rother, and Schäfer. *Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations*, chapter 9, pages 329–368. John Wiley I& Sons, Ltd, 2020. ISBN 9781119751182. doi: <https://doi.org/10.1002/9781119751182.ch9>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119751182.ch9>.

- M. Lopéz de Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):56–69, 2016.
- M. Lopéz de Prado. A robust estimator of the efficient frontier. *Cornell University*, 2019.
- S. Maillard, T. Roncalli, and J. Teiletche. The properties of equally weighted risk contributions portfolios. *The Journal of Portfolio Management*, 36(4):60–70, 2010.
- H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, New Haven, 1959.
- R.C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8:323–361, 1980.
- F. Murtagh and P. Contreras. Algorithms for hierarchical clustering: an overview, ii. *WIREs Data Mining Knowl Discov*, 7, 2017.
- R. Pardo. *The Evaluation and Optimization of Trading Strategies*, 2nd edition. John Wiley I& Sons, Inc., 2008.
- E. Qian. Risk parity portfolios: Efficient portfolios through true diversification. *Panagora Asset Management*, September, 2005.
- T. Raffinot. Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, 44(2):89–99, 2018a.
- T. Raffinot. The hierarchical equal risk contribution portfolio. *SSRN Electronic Journal*, 2018b. doi: 10.2139/ssrn.3237540.
- T. Roncalli. *Introduction to Risk Parity and Budgeting*. Chapman I& Hall, 2013.
- W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(23):425–442, 1964.
- H. Shimizu and T. Shiohama. Constructing inverse factor volatility portfolios: A risk-based asset allocation for factor investing. *International Review of Financial Analysis*, 68, 2020.
- Herbert A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962. ISSN 0003049X. URL <http://www.jstor.org/stable/985254>.
- D. M. Stein. Equity portfolio structure and design in the presence of taxes. *Journal of Wealth Management*, 4:37–42, 2001.
- D. M. Stein and J. P. Garland. *Investment Management for Taxable Investors*. In *F.J. Fabozzi. Handbook of finance (Vol. II)*. John wiley I& Sons Inc., 2008.
- R. Tibshirani, W. Guenther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B*, 63:411–423, 2001.
- O. Van Hemert, M. Ganz, C. R. Harvey, S. Rattray, E. Sanchez Martin, and D. Yawitch. Drawdowns. *The Journal of Portfolio Management*, 46(9):34–50, 2020.

S. Weiss. Lecture notes in csci335 software design and analysis ii, chapter 4: Trees, 2019.

H. Windcliff and P. Boyle. The $1/n$ pension plan puzzle. *North American Actuarial Journal*, 8(1):32–45, 2004.

TRITA-SCI-GRU 2021:200