Hierarchical Clustering To Improve Portfolio Tail Risk Characteristics

Adam Eidenvall

January 2021

Abstract

Many agree that estimating portfolio risks has better estimation possibilities, than estimations on returns. Therefore investors attempts to construct better, more efficient risk-managed portfolios by diversifying portfolios through factors rather than traditional asset classes. This entails very often in estimations of correlation matrices so complex it cannot be fully analyzed. Hierarchical clustering reduces the complexity, by only focusing on the correlations that matters.

Hierarchical clustering uses graph theory, linked to unsupervised machine learning techniques. Hierarchical clustering is obtained by the suggested data and is a formation of a recursive clustering. Several hierarchical clustering methods are presented and evaluated against traditional risk-based portfolios with focus on left hand tail risk. A regime shift, based on momentum is applied to minimize drawdowns. The portfolios are tested on simulated data derived from Bootstrapping simulations and on historical data in a Walk forward optimization process.

The results indicate that hierarchical clustering based portfolios are truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

Keywords: Hierarchical Clustering, Asset Allocation, Portfolio Construction, Graph Theory, Machine Learning, Risk Parity, Regime Shift, Bootstrapping, Walk Forward

Acknowledgements

This thesis was written during the fall of 2020 at the Centre of Mathematical Sciences at Lund University, in cooperation with OQAM at Malmö.

Would like to express my gratitude towards Thorbjörn Wallentin, my supervisor at OQAM. For bringing me the opportunity to write this thesis and his support and knowledge throughout the project. Would also like to thank my academic supervisor Erik Lindström for his valuable feedback and suggestions during this project.

Finally, I would like to address a tremendous thanks to my family and friends for their support throughout my entire studies at Lund University, LTH.

Lund, January 2021
Adam Eidenvall

Abbreviations

A summary of abbreviations used throughout the thesis:

AGNES	Agglomerative Hierarchical Clustering Algorithms
CVaR	Conditional Value at Risk
ERC	Equal Risk Contribution
EW	Equal Weighting Portfolio
HCAA	Hierarchical Clustering Asset Allocation
HERC	Hierarchical Equal Risk Contribution
HESRC	Hierarchical Expected Shortfall Risk Contribution
HRP	Hierarchical Risk Parity
HSDRC	Hierarchical Standard Deviation Risk Contribution
HVRC	Hierarchical Variance Risk Contribution
IV	Inverse Volatility Portfolio
MPT	Modern Portfolio Theory
MST	Minimum Spanning Tree
MVP	Minimum Variance Portfolio
PCA	Principal Component Analysis
RSHEM	Regime Shift Hierarchical Equal and Modern Portfolio Theory
RSHEV	Regime Shift Hierarchical Equal and Variance Risk Contribution
VaR	Value at Risk
VIX	Market Volatility Index
WF	Waterfall Weighting Portfolio

Contents

Chapter 1

Introduction

1.1 Background

For a long time, it's always been interesting for an investor to balance the highest possible return towards a minimum risk. One of the more commonly used mathematical frameworks in finance is the mean-variance analysis, also known as the modern portfolio theory (MPT). First introduced by Markowitz [16] in 1952 and is a so-called standard portfolio theory, who strives for an optimal risk and return trade-off by resorting simulated portfolios on Markowitz [16] seminal mean-variance paradigm. Over time, optimization towards such portfolios has increased in sensitivity regarding inputs of expected return. Resulting in a change of focus towards estimate risk instead of forecasting return. Portfolio strategies are constantly relevant and in attempts to construct better and more efficient risk-managed portfolios, investors diversifying not only using different asset classes but also by investing in different factor styles such as value and momentum. Many agree that trying to estimate returns generates unnecessary noise in relation to the estimates, but estimating portfolio risks has better estimation possibilities. Very often, however, this means estimating correlations / correlation matrices fraught with high complexity.

Meucci [17] introduced an innovative way of managing diversification through a principal component analysis (PCA). In a certain set of assets, PCA identifies the most important risk drivers to represent the different main components. The main components represent different main portfolios with uncorrelated sources of risk. In order to achieve a well-diversified portfolio, the total risk is distributed equally over the uncorrelated sources of risk. Uncorrelated risk sources that are more economically meaningful were later derived from Deguest, Meucci and Santangelo [4]. Suggesting resorting towards a minimum-torsion transformation instead, based on the statistical nature of PCA. Several different risk parity strategies have later been developed, see Lohre, Opfer and Orszàg [12], Bernardi, Leippold and Lohre [3] and Kolrep, Lohre, Radatz and Rother [9].

In the recent literature, risk parity allocation paradigms have been presented based on hierarchical cluster techniques. Strategies based on hierarchical clustering can be divided into two parts. First, based on a hierarchical clustering algorithm, assets in the selected investment universe are divided into different hierarchical clusters, represented in a tree-like map. Then based on the hierarchical structure is an allocation strategy applied to obtain the portfolio weights, which generates a meaningful degree of diversification. Something that is used in Lopez de Prado [13] technology, hierarchical risk parity (HRP). The algorithm behind HRP uses a distance measure to cluster a given set of asset class and style factors, then distribute the risk budget evenly along these clusters. These clusters are expected to provide more meaningful elements to facilitate the portfolio diversification, as they automatically address the structures dependency and considered as more related to natural building blocks than certain aggregate factors.

The correlation matrix is used to determine the dissimilarity measure in the clustering algorithm. Clusters based on correlation are meaningful networks in order to construct diversified portfolios with more reliable results, see Lillo, Mantegna and Tumminello [11] since estimations of covariance and correlation matrices are measures of differences. Lopez de Prado [13] argues in this regard, that correlation matrices are too complex to be fully analyzed. However, by constructing a minimum spanning tree (MST) first introduced by Prim [21], complexity can be reduced by focusing on the most relevant correlations. MST is linked to hierarchical clustering algorithm single linkage and thus reflects the hierarchical structure of the investigated assets.

HRP is not the first risk parity strategy based on hierarchical clustering technique, previously a hierarchical clustering-based waterfall approach by Papenbrock [19] had been presented. The assets are clustered in a hierarchical tree, then at each bisection starting from the top, the weights are split equally until there are no more bisection. However, since Lopez de Prado [13] presented hierarchical risk parity, interest in hierarchical risk parity strategies has increased. The HRP approach showed better performance on out of sample data, indicating that hierarchical clustering really helps to achieve an optimal weight distribution. Later, Raffinot [22] presented hierarchical clustering-based asset allocation (HCAA). Similar to Papenbrock [19] waterfall approach and Lopez de Prado [13] HRP technology, assets are clustered in a hierarchical tree. Further based on the obtained optimal numbers of clusters, the invested capital is allocated across cluster and within clusters at multiple hierarchical levels. It was followed up by another paper Raffinot [23], presenting the Hierarchical Equal

Risk Contribution (HERC) algorithm. HERC allocates capital equally within and across clusters. Raffinot [23] explains that the HERC portfolios are diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

1.2 Purpose

The main purpose in this thesis is to validate if a better diversified portfolio can be achieved with the help of Hierarchical Clustering Based Asset Allocation compared to more traditional asset allocation strategies. Risk factors will also be considered instead of just traditional assets. Focus will be to achieve a more efficient capital allocation regarding left tail risk. The thesis will briefly go over, if we see any difference in the conclusions since the outbrake of the Covid-19 pandemic in March 2020.

1.3 Outline

The outline for this thesis is:

Chapter 2: The selected data is presented to obtain the multi asset multi factor universe and the theory behind hierarchical clustering is explained, where various methods are declared in order to obtain the dendrogram from hierarchical clustering.

Chapter 3: The different allocation methods is presented, both traditional and based on hierarchical clustering techniques. Also, an explanation on the regime shift is given.

Chapter 4: This chapter introduce the different allocation methods that will be tested on simulated and historical data. Will also explain how to test the quantification of the dendrogram.

Chapter 5: The obtained results from the previous chapters are presented in this chapter.

Chapter 6: In this chapter, a discussion of the results and a comparison on how well the different models performed is given. The chapter also review some strengths and weaknesses and predictions to be improved in the future and ends with a short summary.

Chapter 2

Hierarchical Clustering Theory

This Chapter serves to give a brief explanation of the theory behind hierarchical clustering and how to obtain the tree formed dendrogram. Starting by presenting the chosen data and continue with the theory behind the diversification of the dendrogram. Various techniques to derive the dendrogram are explained and compared before an optimal number of clusters to the dendrogram is suggested. An example of a dandrogram can be seen in Figure 2.2

2.1 Multi Asset Multi Factor Universe

To establish the multi asset multi factor universe for our study we consider traditional asset classes - equities, commodities, bonds and credits, as well as factor investments with different style factors. The daily time series are collected with Bloomberg Terminal, a period from 25 May 2000 to 31 November 2020. To represent our global equities, both developed markets and emerging markets are considered. Our developed markets are represented by equity index for OMXS30, S&P 500, Nasdaq, Russell2000, Stoxx Euro 600, Nikkei 225, and FTSE100. MSCI EM, China, Brazil and India representing the emerging markets. For commodities we will consider gold, silver, copper and oil. Bonds will be represented by 10Y US government bonds and 10Y German government bonds. Investment grade and high yield representing the credit market

For our factor investment will we consider four different investment styles momentum, value, size and low volatility. Momentum is based on the idea that recently well performed assets are more likely to outperform recently underperformed assets. According to a given valuation metric, the idea value investing is based on cheap asset tend to outperform expensive assets. Size investors assume a significant upside growth potential for smaller companies that is unmatched by larger companies. Last, low volatility investing assumes that asset with low volatility will outperform the broad market in the long time. The different style factors are combined with the traditional asset classes. We combine momentum with equity, commodities, rates and foreign exchange. While style factor value combines with equity and foreign exchange. Last small cap and low volatility are both combined with asset class equity. In appendix A, a more detailed explanation is given about the chosen assets to the multi asset multi factor universe.

2.2 Minimum Spanning Tree

Markowitz [16] mean-variance approach is like many other portfolio optimization methods very sensitive to changes in input variables, where a small estimation error in the expected returns, can lead to a big allocation difference in the portfolio optimization. Correlations are less sensitive to estimation error, but a complete correlation matrix comes with big complexity of N nodes and $\frac{1}{2}N(N-1)$ edges. The full correlation matrix is visualized to the left in Figure 2.1. From Lopez de Prado [13] can we read that correlation and covariance matrices are too complex to be fully analyzed but lacks a hierarchical structure of interactions among the assets. The complexity can be reduced by dropping unnecessary edges and only focus on the edges between relevant correlations. Achieved from a well-known approach from graph theory, the so-called minimum spanning tree (MST). An algorithm to obtain the MST was first introduced by Prim [21]. In order to apply this algorithm, a distance measure must be defined. Where the correlation coefficient can be used, since it is a measure of dissimilarity between the assets. In the next sub-section, a more precise definition will be provided of the dissimilarity distance. Applying the dissimilarity measure on the correlation matrix forms the so-called dissimilarity matrix, from which the MST is derived from. The MST connects all vertices without any cycles and minimum numbers of edges, creating a subset of $N-1$ edges of the original graph. The complexity from the complete graph with $\frac{1}{2}N(N-1)$ edges is successfully reduced to a connected graph with $N-1$ edges, focused on essential information contained from the correlation matrix and introduces a hierarchical structure. The MST is visualized to the right in Figure 2.1.

Graph theory is linked to unsupervised machine learning and particularly is the hierarchical cluster algorithm called single linkage naturally related to the MST. There are many different linkage methods to choose from and the more common ones are described in the next sub-section. The hierarchical organization of the investigated assets are presented in a direct way thru the MST, as can be seen in Mantegna [15].

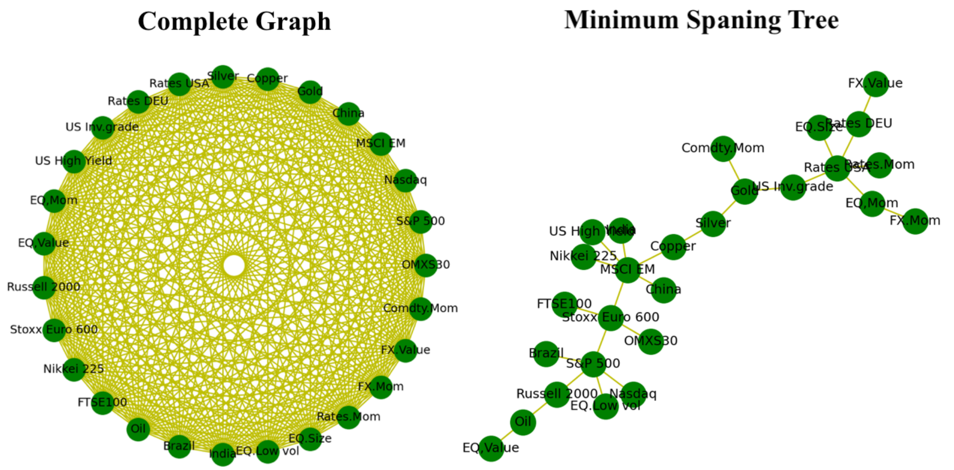

Figure 2.1: Edges between assets, representing the correlation between assets. Complete graph to the left and minimum spanning tree to the right

This results in a so-called dendrogram and is visualized as a tree structure and can be seen in Figure 2.2. Starting at the bottom of the tree, all objects being single-element cluster. Moving up the tree, objects that are similar to each other are combined into one cluster. The higher the height of the fusion, less similar are the objects. We can visualize the embedded hierarchical structure from the MST by comparing the original correlation matrix to the quasi-diagonalized correlation matrix, obtain by reordering rows and columns in the original correlation matrix according to the derived dendrogram. Both correlation matrices can be seen in Figure 2.3. The reordered correlation matrix results in a block-diagonal organization, where similar assets and style factors are placed together and dissimilar assets further apart. Indicating in a well-diversified matrix, which is important and certainly helpful in discerning meaningful building blocks to create diversify portfolios.

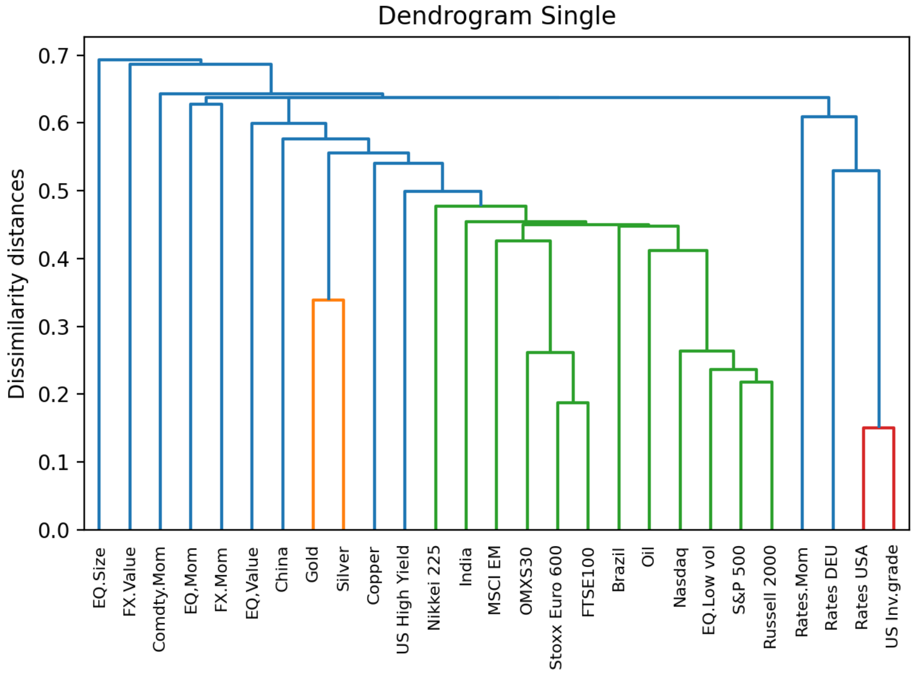

Figure 2.2: Dendrogram obtained with single linkage, associated with MST to the right in Figure 2.1

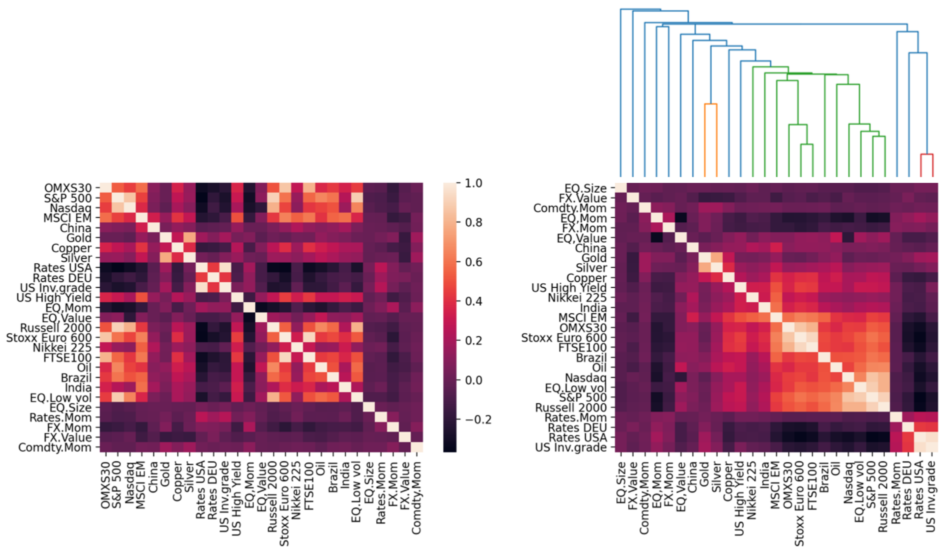

Figure 2.3: Original correlation matrix to the left and quasi-diagonalized correlation matrix to the right, based on the order obtained from Figure 2.2

2.3 Hierarchical Clustering Algorithm

Hierarchical clustering is a form of grouping algorithm, where a set of data points is divided into subsets or clusters. The main goal of the algorithms is to create clusters that are coherent with similarities internally, but clearly differ from each other externally. Assets within one cluster should be as similar as possible and different from assets within all other clusters. In particular, two main hierarchical cluster methods exist, Divisive and Agglomerative. In divisive clustering, all assets starts as one cluster and then divides the clusters until all assets are single-element clusters. Agglomerative clustering who is the more common one, known as AGNES starts with all assets as single-element clusters and then merge clusters until all assets are in the same cluster. The hierarchical clustering algorithm AGNES finds recursively nested clusters, resulting in a tree-based representation called dendrogram. All assets at the bottom are single-element cluster and the higher the height of the fusion, the less similar are the assets. The two methods obtain the same dendrogram, but AGNES is a faster algorithm. Time complexity is $\mathcal{O}(n^3)$ for AGNES and $\mathcal{O}(2^n)$ for Divisive, where n being the number of assets. AGNES will be used in this thesis. Divisive could be a better option, if all data points not being clustered.

The key in cluster analysis is the dissimilarity between the considered objects, obtain by the dissimilarity matrix. Assuming X is a $T \times N$ data matrix, T is observed returns for each of the N assets. where X_i denotes the return times series of asset i. AGNES and other dissimilarity-based clustering algorithms require a dissimilarity matrix. This means replacing the $T \times N$ data matrix X by an $N \times N$ dissimilarity matrix D, which is obtained by applying a dissimilarity measure to the data matrix X.

Definition 2.3.1 (Dissimilarity measure) *Let $X_i \quad i = 1, ..., N$ be the columns of a data matrix X, each containing T observations. A function $d : B \to [0, \infty]$ is called dissimilarity measure if*

$$\begin{aligned} d(X_i, X_j) \geq 0, \quad d(X_i, X_j) = d(X_j, X_i) \\ and \quad d(X_i, X_i) = 0 \quad \forall i, j = 1, ...N \end{aligned} \tag{2.1}$$

where $B = \{(X_i, X_j) \; i, j = 1, ..., N\}$.

B is the Cartesian product of items in $\{1, ..., i, ..., N\}$. If function d additionally satisfies the triangle inequality, dissimilarity matrix D defines a metric space. Euclidean and Manhattan distances are just two of the ways to define the dissimilarity measure. We will consider $d : B \to [0, 1]$ as defined by Lopez de Prado [13]:

$$d_{i,j} = d(X_i, X_j) = \sqrt{0.5(1 - p_{i,j})} \tag{2.2}$$

where $p_{i,j} = p_{i,j}(X_i, X_j)$, $p_{i,j}$ is the Pearson correlation coefficient. Lopez de Prado [13] verify that equation 2.2 is a dissimilarity measure, where perfectly correlated assets $p_{i,j} = 1$, we have $d = 0$ and for perfectly negatively correlated assets $p_{i,j} = -1$, we have $d = 1$. The dissimilarity among all single assets and factors are now contained in D.

Once the dissimilarity measure is defined, we have to define how to use this information to measure the distance between assets. We will use so-called linkage criterion. There is various criterion to choose from, but in Raffinot [22] we can see the more commonly ones are single linkage, complete linkage, average linkage and Ward's method.

Single linkage, also known as the nearest point algorithm, is the distance between two clusters defined as the minimal distance of any two elements in the clusters:

$$d_{C_i,C_j} = \min_{x \in C_i, y \in C_j} d(x, y) \tag{2.3}$$

where C_i and C_j denote two clusters and x and y represent elements within the clusters. In contrast to single linkage, Complete linkage also known as farthest point algorithm is defined as the maximum distance between any two elements of the clusters:

$$d_{C_i,C_j} = \max_{x \in C_i, y \in C_j} d(x, y) \tag{2.4}$$

average linkage computes the average of the distance of any elements in the two examined clusters:

$$d_{C_i,C_j} = \frac{1}{|C_i|} \frac{1}{|C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y) \tag{2.5}$$

where C_i and C_j are the carnality of cluster i and j. Lastly Ward's method minimizes the total within-cluster variance

$$d_{C_i,C_j} = \frac{|C_i||C_j|}{|C_i| + |C_j|} |\mu_i - \mu_j|^2 \tag{2.6}$$

where μ_i and μ_j is the center of clusters i and j. The choice of linkage criterion will affect our dendrogram and result in different tree structures. The different tree structures based on linkage criterion can be observed in Figure 2.4. We can see that single

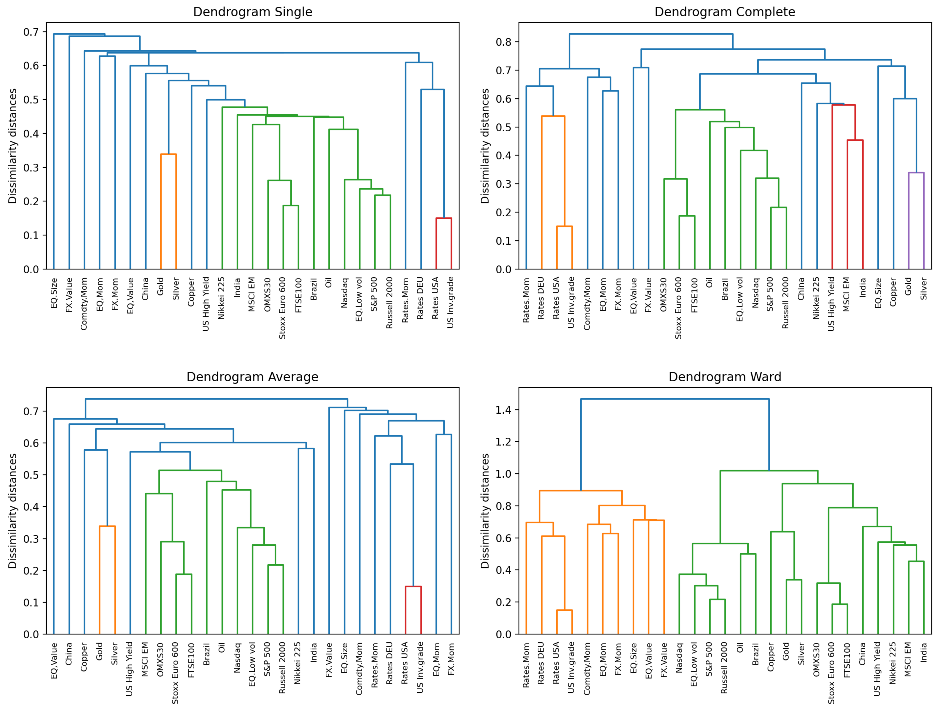

Figure 2.4: Comparison of linkage methods: single linkage (top left), complete linkage (top right), average linkage (bottom left) and Ward's method (bottom right)

linkage suffers from a well-known problem called chaining, which leads to long straggly clusters. Complete linkage and average linkage tend to result in compact clusters, but both methods together with single linkage are sensitive to outliers. Ward's method is an exception considering the sensitivity to outliers and also results in compact clusters of similar size, therefore further away from chaining. Making Ward's method to a popular choice among researchers. Papenbrock [19] describes why chaining becomes a problem, especially when it comes to the assets weight concentration. Affecting the allocation, when creating diversified portfolios. The different weight concentration defined by Papenbrock [19], based on the different linkage methods are visualized in Figure 2.5.

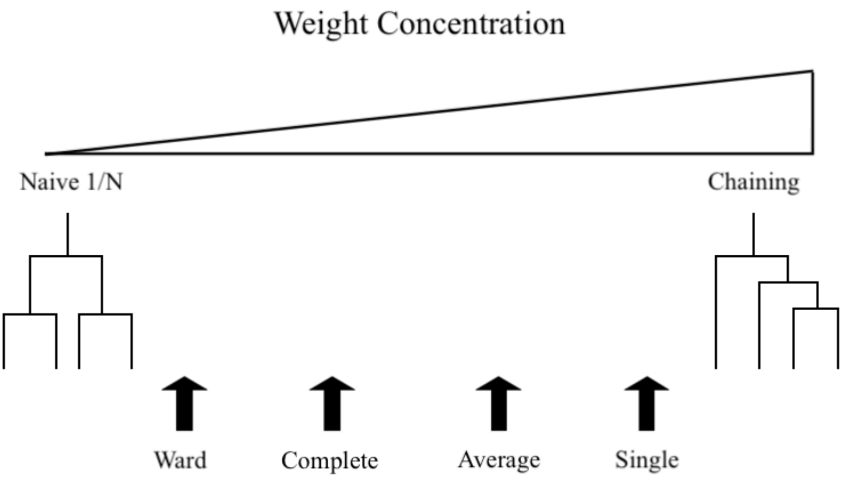

Figure 2.5: The weight concentration by using trees with different linkage methods in the cluster-based waterfall approach

2.4 Optimal Number of Clusters

Once the hierarchical clustering tree is fully grown and a complete dendrogram is accomplished, decisions can be made however to devote the structure of the dendrogram as it is. The recursive bisection in Lopez de Prado [13] HRP approach is only based on the assets order, obtained in the bottom level. Completely fully grown trees could come with some problems. Analyzing a large dataset makes the algorithm computationally slow when computing large trees and a risk of overfitting, where small inaccuracies in the data could lead to large estimation errors in the portfolio weights. Instead, we might want to consider the dendrogram only up to a certain level and not taking the whole hierarchical structure into account. This will make the algorithm faster but comes with a little loss of information. Optimal number of clusters is a trade-off between reducing noise, but at the same time maintaining the structure of the dendrogram. Cutting the dendrogram at a certain level will divide the assets and style factors into clusters. There are many ways to decide an optimal number of clusters, one could simply look at the dendrogram and decide from that. A statistical criterion could also be applied to determined the optimal numbers of clusters. Gap statistics was developed by Hastie, Tibshirani and Walther [8] and is a popular method, used by Raffinot [23] in the HERC algorithm. The total within intra-cluster variation is compared for different values of k with their expected values under null reference distribution of the data, where k being the number of clusters.

Suppose we have data with k clusters $C_1, C_2, ..., C_k$. C_r denoting the indices of observations in cluster r and the sum of pairwise distances D_r for cluster C_r, given by,

$$D_r = \sum_{i,i' \in C_r} d_{i,i'} \tag{2.7}$$

where $d_{i,j}$ is the Euclidean distance between data points i and j. Based on D_r, the cluster inertia W_k is then calculated as the within-cluster sum of squares around the cluster means

$$W_k = \sum_{r=1}^{k} \frac{D_r}{2n_r} \tag{2.8}$$

where n_r being the number of data points in cluster r. Once the cluster inertia is calculated, the Gap statistic is given by the following equation,

$$Gap_n(k) = E_n^*[\log(W_k)] - \log(W_k) \tag{2.9}$$

where E_n^* denotes expectation from the reference distribution based on n numbers of samples. The optimal numbers of clusters k will be the value that maximizes $Gap_n(k)$. Often will $Gap_n(k)$ results in an increasing function, resulting in a very large numbers of clusters, which we wanted to avoid. We should therefore select the number of clusters as the smallest value of k that maximizes our Gap statistic before the rate of change begins to slow down and more advantageous than one cluster. This is achieved by selecting the first k that satisfied

$$Gap_n(k) > Gap_n(k+1),\ Gap_n(k) > Gap_n(1),\ k = 2, ..., N \tag{2.10}$$

A visualization of the gap statistics method derived from equation 2.9 and applied on dendrogram in Figure 2.4 based on linkage criteria ward, can be seen to the left in Figure 2.6. Another way to decide the optimal numbers of clusters is to use the dendrograms dissimilarity distance. The dissimilarity distance for each node, starting from the top will represent the clusters inertia. The same idea as Hastie, Tibshirani and Walther [8] gap statistics, will the clusters inertia be compared to an expectation inertia obtained from a reference distribution, based on a dendrogram created from n numbers of random correlation matrix. In the same way as equation 2.9 is the optimal numbers of clusters, the cluster k that maximizes

$$Gap_n(k) = E_n^*[\log(Z_k)] - \log(Z_k) \tag{2.11}$$

where Z_k is cluster k inertia from the dissimilarity distance. The method is visualized to the right in Figure 2.6, derived from the dendrogram in Figure 2.4 based on linkage criteria ward. Equation 2.9 and 2.11 are basically the same equation, but will separate them to differentiate the two methods.

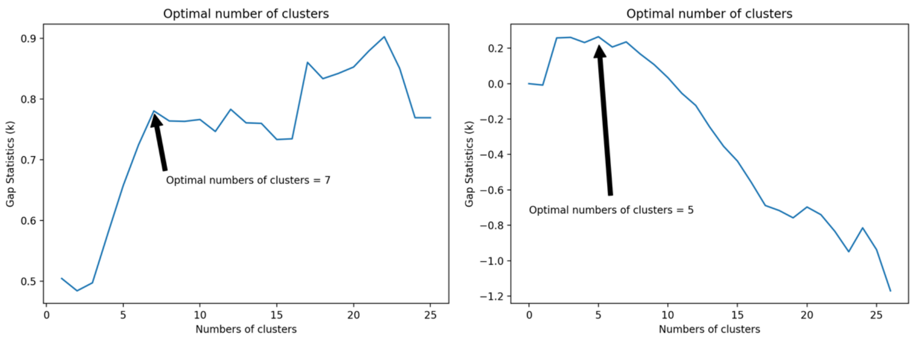

Figure 2.6: The suggested optimal number of clusters derived from equation 2.9 to the left and from equation 2.11 to the right

Chapter 3

Allocation Methods

After obtaining the dendrogram and an optimal number of clusters are established, one must decide how to allocate the invested capital in the assets. In this chapter will we go over the different allocation methods that will be used and tested in this thesis. Starting with some classical risk-based allocation techniques and then move on to allocation methods on the dendrogram. Finally, two regime shift allocations are applied.

3.1 Classic Risk-Based Allocation Techniques

We will consider Equal Weighting (EW), Minimum-Variance Portfolio (MVP) and Inverse Volatility (IV) as the classic risk-based allocation techniques, that will be compared in this thesis against allocations techniques based on hierarchical clustering.

3.1.1 Equal Weighting

A common expression is that you should divide your risk across assets. The simplest way is the allocation strategy Equal Weighting $(1/N)$. Where the same amount of the invested capital is given with the same importance to each asset. The weights are obtained by

$$\omega_i = \frac{1}{N} \tag{3.1}$$

where ω_i is the weights for asset i and N is the numbers of assets. EW gives a completely balanced allocation of the weights. But if the assets individual risk differs significant, EW might suffer from a concentrated risk allocation.

3.1.2 Minimum-Variance Portfolio

One of the more commonly used risk-based allocation techniques is the Minimum-Variance Portfolio and is derived from Markowitz [16] mean-variance approach. It selects the point furthest to the left on the efficient frontier, which indicates portfolio weighting with least risk. Figure 3.1 illustrated the MPT, where red dots indicate the minimum-variance portfolio and the maximum sharpe ratio portfolio. The MPT in Figure 3.1 is estimated on all 27 assets and are obtained using 27000 random portfolios. We can see that 27000 random portfolios are too few to give an optimal estimation, but good enough too avoid making the program to slow. The weights for the minimum-variance portfolio are determined by solving the quadratic optimization problem

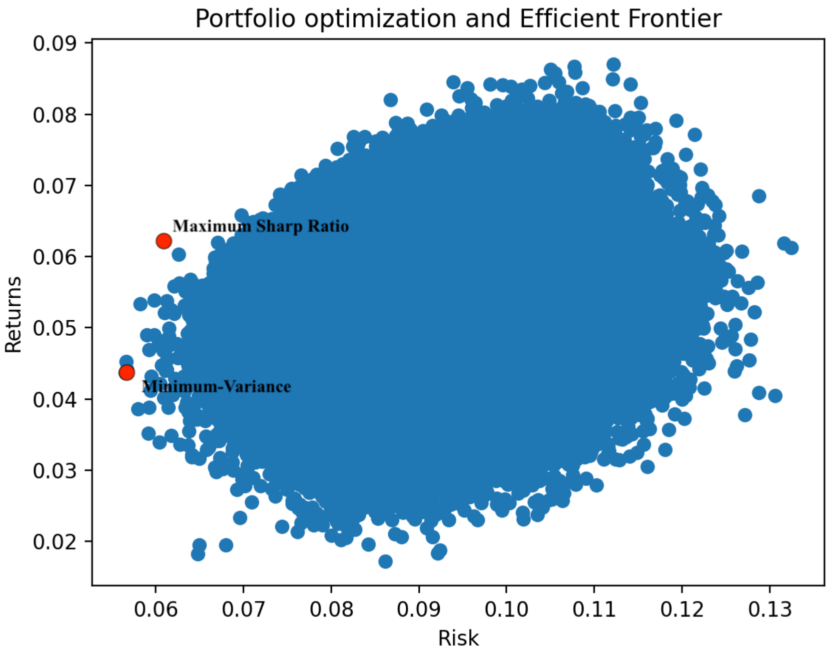

Figure 3.1: A simulated example over the MPT obtained from 27000 random portfolios of the 27 assets. The two red dots indicating the two portfolios minimum variance and maximum sharpe ratio

$$\arg\min_{\omega} = \frac{1}{2}\omega^T \Sigma \omega \quad \text{s.t.} \quad \mathbb{1}^\top \omega = 1 \tag{3.2}$$

where Σ is the covariance matrix. Maximum sharpe ratio portfolio is the portfolio on the efficient frontier that maximizes the ratio between return and risk. The weights in MVP are relatively easy to compute but very often is the portfolio concentrate with low-volatility assets.

3.1.3 Inverse Volatility

We will also constructing a portfolio based on the assets Inverse Volatility (IV) and is a so-called risk parity strategy which diversifies the risk, such that all assets contribute equally to the overall portfolio risk. Each of the assets are first weighted inversely proportional to its volatility, also known as the standard deviation σ_i given from

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{3.3}$$

x_i is the i^{th} point in the data set, $\bar{x}$ is the mean value of the data set and n is the number of data points in the data set. The assets are then divided against the portfolios total inversely proportional from all assets volatility. The weights are given by

$$\omega_i = \frac{1/\sigma_i}{\sum_{j=1}^{N} 1/\sigma_j} \quad \text{for} \quad i = 1, \ldots, N \tag{3.4}$$

This IV strategy often mention as the standard deviation is a simple risk parity strategy, but do not consider the correlations between the portfolio's assets.

3.2 Hierarchical Risk-Based Allocation Techniques

Hierarchical risk-based allocation techniques are very flexible and can be utilized in various ways. There are many functions that can be applied as risk contribution and methods to apply on the tree structure, obtained from hierarchical clustering. We will go over some allocation techniques derived on hierarchical clustering.

3.2.1 Hierarchical Risk Parity HRP

Lopes de Prado [13] HRP approach only consider the ensuing order of the assets, obtained from the hierarchical clustering dendrogram. Single linkage is used for HRP and is the best method, then HRP doesn't consider the different clusters in the dendrogram. An allocation technique is then suggested based on recursive bisection and inverse variance allocation. The method starts with all of the ordered assets as one set, with all unit weights equal to one. The sets are then bisected into two subsets and the subsets are then bisected until all subsets includes only one asset. At each step of the bisection, assets weights in the two subsets are re-scaled with a factor α and $(1-\alpha)$ respectively. We obtain factor α by calculating the inverse variance allocation of the two subsets. The subsets variance is also determined by using an inverse variance allocation within the subset. Algorithm 1 summarizes the procedure.

Algorithm 1: Recursive Bisection

---

1. Initialization
   (a) (a) Set the list of items: $L = \{L_0\}$, with $L_0 = \{n\}_{n=1,\dots,N}$ according to the order obtained by the dendrogram
   (b) Assign a unit weight to all assets: $\omega_n = 1, \quad \forall n = 1, \dots, N$
2. If $|L_i| = 1, \quad \forall L_i \in L$, then stop
3. For each $L_i \in L$ such that $|L_i| > 1$:
   (a) Bisect L_i into two subsets, $L_i^{(1)} \cup L_i^{(2)} = L_i$, where $|L_i^{(1)}| = \text{round}(\frac{1}{2}|L_i|)$, and the order is preserved
   (b) Define the variance of $L_i^{(j)}, j = 1, 2$ as $\tilde{V}_i^{(j)} = (\tilde{\omega}_i^{(j)})^T V_i^{(j)} \tilde{\omega}_i^{(j)}$, where $V_i^{(j)}$ is the covariance matrix between the constituents of the $L_i^{(j)}$ bisection and $\tilde{\omega}_i^{(j)}$ are the inverse variance weights
   (c) Compute the split factor: $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}}$
   (d) Re-scale allocations ω_n by a factor of $\alpha_i, \forall n \in L_i^{(1)}$
   (e) Re-scale allocations ω_n by a factor of $(1 - \alpha_i), \forall n \in L_i^{(2)}$
4. Loop to 2.

The introduced methodology is very flexible, we can for instance use different functions for obtaining ω and α in steps $3.(b)$ and $3.(c)$ of Algorithm 1. But the recursive

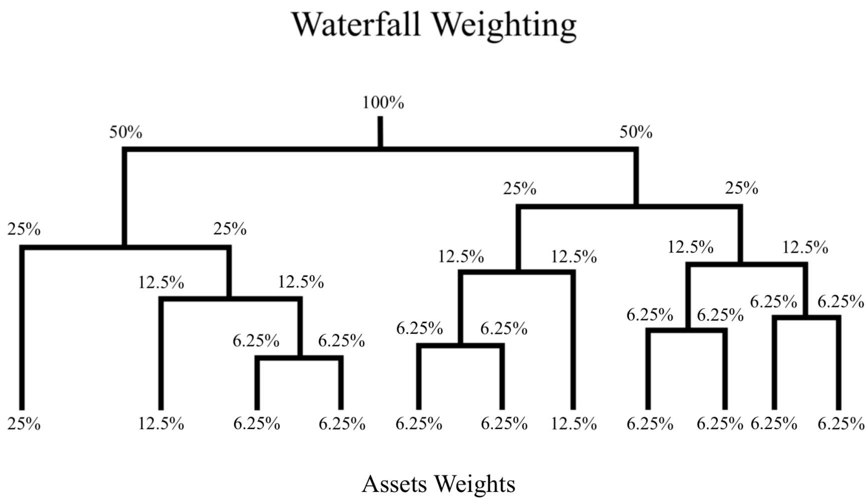

Figure 3.2: A visualization of an example over the cluster-based waterfall weighting scheme

bisection in HRP rather focuses on the obtained order at the bottom level of the dendrogram and ignores the obtained clusters at different levels of the dendrogram. The allocation algorithm gives a certain arbitrariness but can be advanced by splitting the allocations into clusters obtained from the dendrogram.

3.2.2 Hierarchical Clustering-Based Asset Allocation

Instead of using recursive bisection as an algorithm, one might want to consider the tree structure in the dendrogram. We will therefore divide our assets based on the dendrograms tree nodes, rather than cutting our assets in half based on the bottom level. Wards linkage criteria is considered in both Papenbrock [19] and Raffinot [23] approach. Papenbrock [19] waterfall approach starts on top of the tree, in the so-called root node with 100% of our invested capital. Then going from node to node in a top to down manner, the capital is divided in half to each new cluster until we are at the bottom and there are no more nodes. The weights are then given to each asset respectively from their percentage at the bottom. A simple visualization of the allocation technique is given in Figure 3.2.

Raffinot [22] suggest a similar weighting approach as the waterfall approach derived by Papenbrock [19], that allocates capital equally across cluster and within clusters obtained from the hierarchy in the dendrogram. The portfolio weights and risk contribution are obtained by solving a general case, the equal risk contribution portfolio

(ERC) introduced by Maillard, Roncalli, and Teiletche [14]. The approach solves the following optimization problem.

$$\omega^* = \arg\min f(\omega) \quad \text{s.t.} \quad \mathbb{1}^\top \omega = 1 \quad \text{and} \quad 0 \leq \omega \leq 1 \tag{3.5}$$

where

$$f(\omega) = \sum_{i=1}^{N} \sum_{j=1}^{N} (\omega_i(\Sigma\omega_i) - \omega_j(\Sigma\omega_j))^2 \tag{3.6}$$

The program basically minimizes the variance of the re-scaled risk contribution. The ERC portfolio is only ensured when the condition $f(\omega^*) = 0$, which is verified when $\omega_i(\Sigma\omega_i) = \omega_j(\Sigma\omega_j)$ for all i and j.

The approach was followed up by another paper, Raffinot [23] and introduced the Hierarchical Equal Risk Contribution (HERC) an algorithm that advance Lopez de Prado [13] algorithm described in Algorithm 1. The HERC algorithm is described in Algorithm 2. The algorithm goes from node to node, starting at the top of the dendrogram in the root node. Weights are assigned on each node based on the chosen methodology for the within and across-cluster allocation. In step 4 are the weights re-scaled within each cluster, obtained from the suggested optimal numbers of clusters. Step 4 should only be executed if an optimal number of clusters is used.

Algorithm 2: Clustering-based weight allocation

1. Perform hierarchical clustering and generate dendrogram
2. Assign all assets a unit weight $\omega_i = 1 \; \forall i = 1, ..., N$
3. For each dendrogram node (beginning from the top):
   (a) Determine the members of clusters C_1 and C_2 belonging to the two sub-branches of the according dendrogram node
   (b) Calculate the within cluster allocations $\tilde{\omega_1}$ and $\tilde{\omega_2}$ for C_1 and C_2 according to the chosen methodology
   (c) Based on the within cluster allocations $\tilde{\omega_1}$ and $\tilde{\omega_2}$ calculate the across cluster allocation α (splitting factor) for C_1 and $(1-\alpha)$ for C_2 according to the chosen methodology
   (d) For each asset in C_1 re-scale allocation ω by factor α
   (e) For each asset in C_2 re-scale allocation ω by factor $(1-\alpha)$
4. For each cluster containing more than one element:
   (a) Determine the members of the cluster
   (b) Calculate the within cluster allocation
   (c) For each asset in the cluster re-scale ω by the within cluster allocation
5. End

One can easily see that if step 4 is ignored, The HERC approach becomes the same as Papenbrock [19] waterfall approach, when equal weighting is the chosen methodology as risk contribution. There are various alternatives available as risk parity strategies, when choosing the methodology for risk contribution. We can simply use inverse volatility in equation 3.4 as allocation technique within and across clusters or use the inverse variance given by

$$\omega_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{N} 1/\sigma_j^2} \quad for \quad i = 1, ..., N \tag{3.7}$$

The left-hand tail risk can also be considered for the chosen methodology as risk contribution. Value at risk (VaR) is a measure for the minimum loss over the whole range of outcomes for a given presentence alpha in the tail. The VaR is given from

$$\mathrm{VaR}_{h,\alpha} = -x_{h,\alpha} \tag{3.8}$$

where

$$\mathbb{P}(X < -x_{h,\alpha}) = \alpha \tag{3.9}$$

X are the daily returns over h numbers of days. With VaR in mind, we can define the conditional value at risk (CVaR) also known as the expected shortfall (ES). Gives the average loss over the whole range of outcomes in the $\alpha\%$ of the tail. CVaR is defined by

$$\mathrm{CVaR}_{h,\alpha} = -\mathrm{E}(X|X < x_{h,\alpha}) = \frac{1}{\alpha}\int_{-\infty}^{x_{h,\alpha}} xf(x)dx \tag{3.10}$$

Applying a normal distribution to equation 3.9 gives us

$$\mathrm{CVaR}_{h,\alpha}(X) = \alpha^{-1}\varphi(\Phi^{-1}(\alpha))\sigma_h - \mu_h \tag{3.11}$$

φ is the standard normal p.d.f., $\Phi(x)$ is the standard normal c.d.f., so $\Phi^{-1}(\alpha)$ is the standard normal quantile. σ_h and μ is the normalized volatility and mean over the horizon h. CvaR can then be used as risk contribution within and across, where the weights are given from

$$\omega_i = \frac{\mathrm{CVaR}_{h,\alpha}(X_i)}{\sum_{j=1}^{N} \mathrm{CVaR}_{h,\alpha}(X_j)} \tag{3.12}$$

3.3 Regime Shift Allocation Based On Momentum

Momentum has offered investors some of the highest sharpe ratio on the market, Barroso and Santa-Clara [1]. However, the strategy is unappealing to investors who dislike negative skewness and kurtosis, since momentum strategies also has had some of the worst crashes. From Barroso and Santa-Clara [2] can we read that the risk of momentum is highly variable over time and can be predicted. By predicting the risk of momentum, crashes in the market can be minimized. We can do so by applying a regime shift, see Nystrup, Hansen, Larsen, Madsen, Lindström [18]. A so-called risk-on/risk-off switch. Based on the predicted risk of momentum can we tell when to switch the risk between on and off. Risk-on tell us that it is good times in the market, and we should therefore take a higher risk when distributing our investment over the assets. Risk-off indicates that there is a higher risk for the market to crash and we should therefore invest with more caution.

In order to predict the risk of momentum, will we use realized volatility of momentum obtained from daily returns. We will use the Market Volatility Index or often mention as VIX index, obtained from Bloomberg terminal. The VIX index is calculated in real time and uses the live prices of S&P 500 options and includes standard CBOE SPX options and weekly CBOE SPX options, expire on the third Friday of every month and every Friday respectively. An option must have an expiry date between 23 and 37 days to be considered for the VIX index. The VIX index is given from

$$\text{VIX} = \sigma \cdot 100 \tag{3.13}$$

with

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2 \tag{3.14}$$

and where T is the time to expiration, F is the forward index level derived from index option price, K_0 is the first strike below the forward index level F, K_i is the strike price of the i^{th} out-of-the-money option, R is the risk-free interest rate to expiration, $Q(K_i)$ is the mid-point of the bid-ask spread for option with strike K_i and ΔK_i is the interval between strike prices given from

$$\Delta Ki = \frac{K_{i+1} - K_{i-1}}{2} \tag{3.15}$$

The VIX value is visualized in Figure 3.3. Even if the VIX value is based on the American stock market it can be applied to the other markets and our multi asset multi factor universe. Since the American stock market is the largest and most impactful market, other markets tend to follow its development closely. A VIX value below 20 indicates that the market is calm, and the risk appetite should be high and a VIX value above 30 indicates that one should be more caution. The average VIX value is around 20 and we will use 20 as the value to determine the regime shift, where a VIX value below 20 indicates risk-on and a VIX value of 20 or higher indicates risk-off.

Based on earlier papers from Papenbrock [19], Markowitz [16] and Raffinot [22] will we now advance the previous mention and tested allocation methods, by construct two different allocation techniques based on regime shift. The first one will use Papenbrock [19] waterfall approach as allocation technique across cluster. Then for the

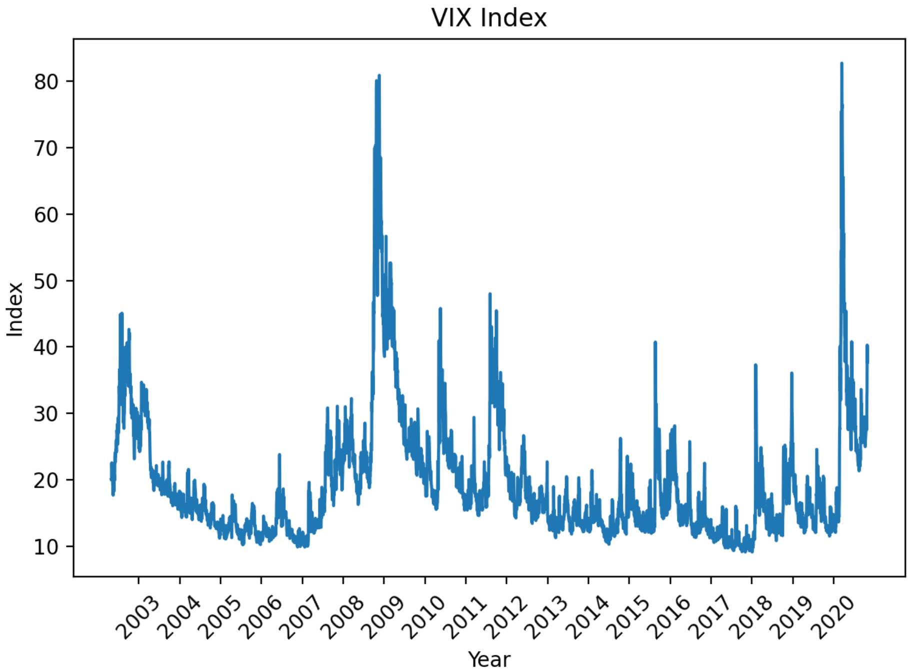

Figure 3.3: The VIX Index obtained from Bloomberg terminal, from 2000-05-25 to 2020-10-30

within cluster allocation Markowitz [16] MPT will be used, where the maximum-sharp ratio portfolio will be used when it is risk-on and the minimum-variance portfolio when the VIX value indicates risk-off. The allocation technique is summarized in Algorithm 3.

Algorithm 3: Regime shift allocation based on waterfall and MPT

1. Perform hierarchical clustering and generate dendrogram
2. Perform Waterfall weighting on each dendrogram node, beginning from the top to assign each asset their weight ω_i:
3. For each cluster containing more than one element:
   (a) Determine the members of the cluster
   (b) Calculate the total weight ω_{tot} from the within cluster weights
   (c) If ($VIX < 20$:): For each asset in the cluster re-scale ω_i by the maximum-sharp ratio portfolio based on the within assets divided by ω_{tot}
   (d) Else: For each asset in the cluster re-scale ω_i by the minimum-variance portfolio based on the within assets divided by ω_{tot}
4. End

For the second regime shift allocation method, will we switch between two different hierarchical clustering-based asset allocation. During risk-on will we tend towards more risky allocation and therefore allocate the invested capital given with equal risk contribution across and within clusters, the HERC allocation method. When indicating risk-off, will we be more careful and will therefore use variance as risk contribution across and within the clusters, the HVRC allocation method. In Appendices B are an example of the weights given for all of the different allocation methods, based on the whole data set.

Chapter 4

Model Design

Once the Portfolio weights are given, different allocation methods can be tested. A model with simulated data will be used to test the methods and their uncertainty before the methods will be appointed on a model with historical data, to study the uncertainty in the methods and the market. For a better understanding of the dendrograms stability over time, will we start by taking a look at the quantity of the dendrogram.

4.1 Quantifying of the Dendrogram

The robustness of the dendrogram is very complex. Where a small change can have a big impact on the quantify of the dendrogram, even if the dendrogram overall stays pretty much the same. The bottom order of the assets in the dendrogram could be the same even if assets change between clusters. The suggested optimal number of clusters can easily alternate between similar numbers from a period to another period. Even if this doesn't affect the outcome on the allocation weighting very much, does it affect the quantify of the dendrogram.

We will therefore not try to quantify the dendrogram, instead analyze the dendrogram over different time periods and be compared to each other. A time period around two years will be the interval that the dendrogram is derived from. The two different models to calculate the optimal numbers of clusters will also be used for analyzing the quantity of the dendrogram. The two models will generate their suggested optimal number of clusters based on the dendrogram over a two year interval. A rolling window will then transfer the interval one week forward and generate a new dendrogram and continue so over the complete time series. The two models will derive their optimal numbers of clusters based on comparison towards 100 random correlation matrices.

4.2 Bootstrap Simulation

In order to test the different allocation methods uncertainty will we use bootstrap simulations. The different allocation weighting methods will be derived from the obtained dendrogram, based on the complete time interval. With the use of 1000 times the number of assets to derive the two-portfolio weighting, obtained from MPT. The weights for the different allocation methods will further on be static for the whole simulation. Bootstrap simulation will then generate the new simulated prices. The sampled prices are generated by re-sampling observed data many times. Where random samples are drawn based on the original dataset to create simulated datasets. Bootstrap simulations core advantage is the fact that the method allows us to generate data without making assumptions about the distribution of the underlying data, Pažický [20].

A two year interval of known data will be considered when generating 1000 new simulate prices one month forward every day for each asset. The interval will then move one month forward and generate new simulated prices for the next month based on the new two year interval. The next simulated stock price is calculated from

$$P_t = P_{t-1} e^{\mu_t^*} \tag{4.1}$$

P_t is the stock price today, P_{t-1} is yesterday's stock price and μ_t^* is the randomly selected expected logarithmic daily return, drawn from the known data interval of logarithmic daily returns μ_t . The logarithmic daily return μ_t is calculated for each day in the interval by

$$\mu_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{4.2}$$

The allocation methods are then applied on the new simulated stock prices where the average value represent the statistically new value and all simulated daily returns represent the model's variance. All assets stock prices and the allocation methods portfolio value are initially set to 100\$.

4.3 Walk Forward

To estimate the uncertainty for the allocation methods and the markets uncertainty are the methods tested on real historical data, using a walk forward optimization process. From Liew [10] can we read that using walk forward optimization reduce the chance of overfitting in our backtest. In trading, overfitting is the process that designs a trading system that adapts so closely to the noise of historical data, that it becomes ineffective in the future. Noise are distractions that don't offer useful qualities, while signals are useful fundamental information. A walk forward optimization forces us to adjust our strategy parameters towards signals from the past by constantly testing our updated parameters in the out-of-sample data, Liew [10].

The process finds the optimal parameters for each of the different portfolios in a certain time period called in-sample. The obtained parameters are then applied over on the following time period called out-of-sample, which is not included in the in-sample time period. The time interval is then shifted forward so the new in-sample period ends where the previous out-of-sampled time period ended and continue so until the complete time period is covered. The in-sample time period consists of the latest two years back before the out-of-sample time period. The portfolio weights in the different allocation methods are estimated and updated before tested in the out-of-sample time period consisted of one month forward from the in-sample time period. All portfolios value is initially set to 100$ and the portfolios estimated from MPT are derived from 1000 times the number of assets portfolios.

The derived time series are then analyzed and compared to the other allocation methods. The different time series will be analyzed by annualized return, standard deviation, Sharpe ratio, Sortino ratio, maximum drawdown, Calmar ratio and conditional value at risk. Standard deviation is given from equation 3.3 and CVaR is given from equation 3.11. Annualized return is the geometric average amount of earned money each year over a given time period and is given from

$$\text{Annualized Return} = \left(\left(\frac{\text{Ending Value}}{\text{Beginning Value}} \right)^{\frac{1}{Y}} \right) - 1 \tag{4.3}$$

Y is the holding period in years. Sharpe ratio indicates the return of an investment compared to it's risk, where the ratio is the average return without the risk-free rate per unit of volatility and is given from

$$\text{Sharpe Ratio} = \frac{R_p - r_f}{\sigma_p} \tag{4.4}$$

R_p is the portfolios return, r_f is the risk-free rate and σ_p is the standard deviation of the portfolio's excess return. Sortino ratio is similar to the Sharpe ratio, taking the portfolio's return and subtracts the risk-free rate, and then divides that amount only by the portfolio's downside deviation. Sortino ratio is given from

$$\text{Sortino Ratio} = \frac{R_p - r_f}{\sigma_d} \tag{4.5}$$

σ_d is the standard deviation of the portfolio's excess return only from the downside. Maximum drawdown (MDD) indicates the downside risk over a specified time period. MDD is the maximum observed loss from a peak to the portfolios lowest value after the peak and before a new peak is attained. MDD is derived from

$$\text{MDD} = \frac{\text{Trough Value} - \text{Peak Value}}{\text{Peak Value}} \tag{4.6}$$

and Calmar ratio is a measure of the portfolios performance comparing the average annual rate of return against its maximum drawdown. The higher ratio, the better the portfolio perform on a risk-adjusted basis and is estimated from

$$\text{Calmar Ratio} = \frac{\text{Annualized Return}}{\text{MDD}} \tag{4.7}$$

In this thesis will we consider a good portfolio based on the portfolio's left hand tail risk. How well the portfolio performance in it's worst times. MDD and CVaR are considered as measures of great importance for balancing the portfolio left hand tail risk. A good portfolio must also generate a profit, where Annualized Return and Calmar ratio will be of high value to be consider as a good portfolio.

Chapter 5

Results

The results based on the earlier chapters are presented in this chapter. Starting with a brief look at the dendrogram over different time periods and then move on to the obtained results from the two methods, bootstrap simulation and walk forward optimization.

5.1 Dendrogram Over Time

The obtained dendrogram over the complete time series from 2000-05-25 to 2020-10-30 is visualized in Figure 5.1 and followed by obtained dendrogram over a two year period, from the first trading day in June to the last trading day in May two years forward. The obtained dendrograms are compiled in Figure 5.2 - 5.6. All Dendrogram are derived with ward as linkage criterion. Corresponding reordered correlation matrices for the dendrograms are visualized in Appendices C, Figure C.1-C.5.

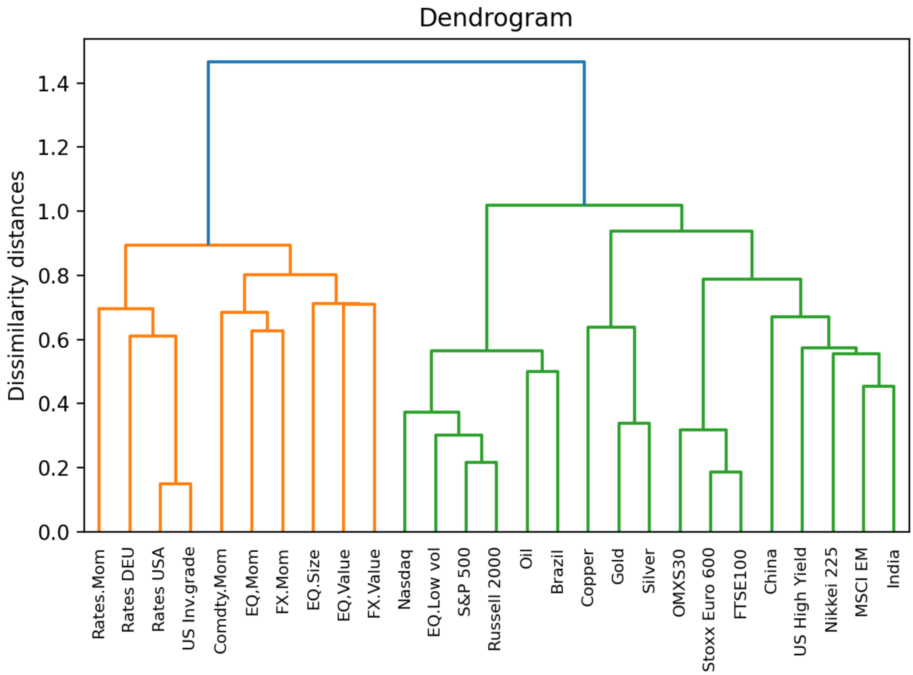

Figure 5.1: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2000-05-25 to 2020-10-30

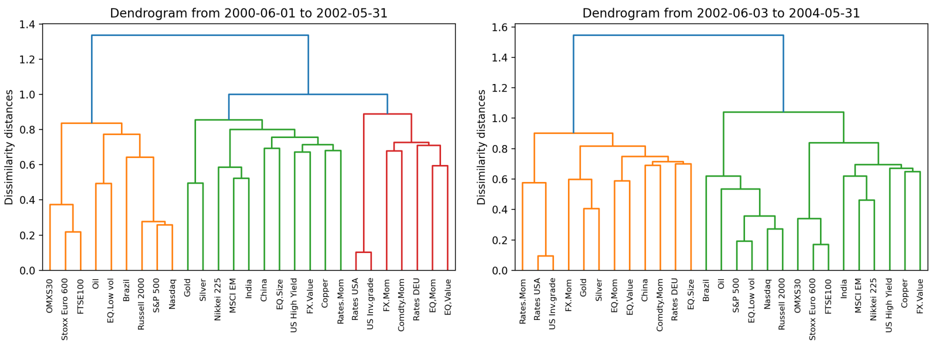

Figure 5.2: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2000-05-25 to 2002-05-31 to the left and from 2002-06-03 to 2004-05-31 to the right

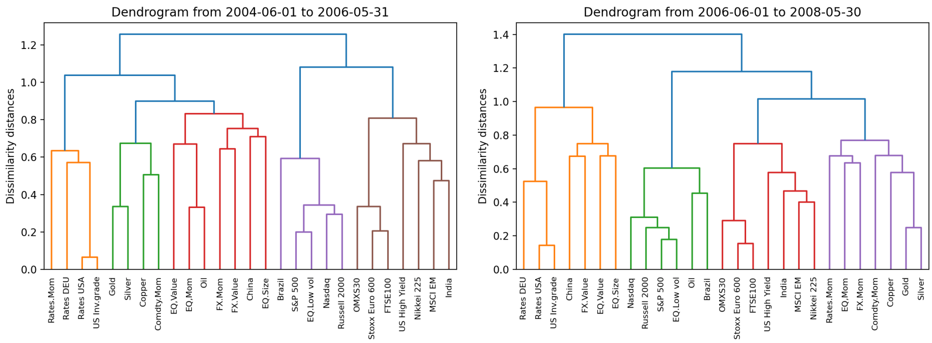

Figure 5.3: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2004-06-01 to 2006-05-31 to the left and from 2006-06-01 to 2008-05-30 to the right

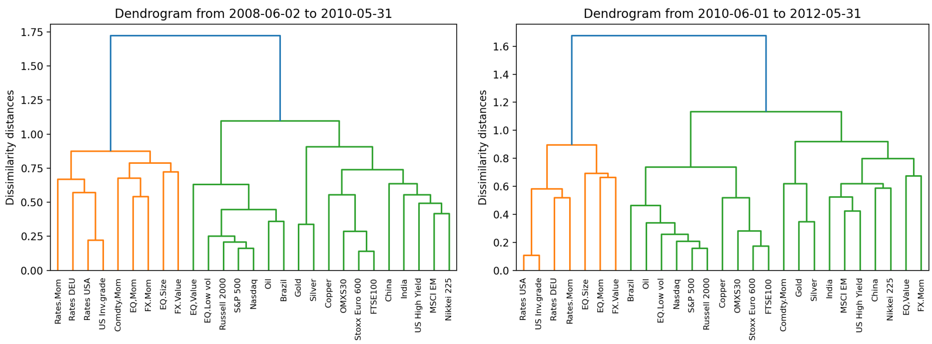

Figure 5.4: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2008-06-02 to 2010-05-31 to the left and from 2010-06-01 to 2012-05-31 to the right

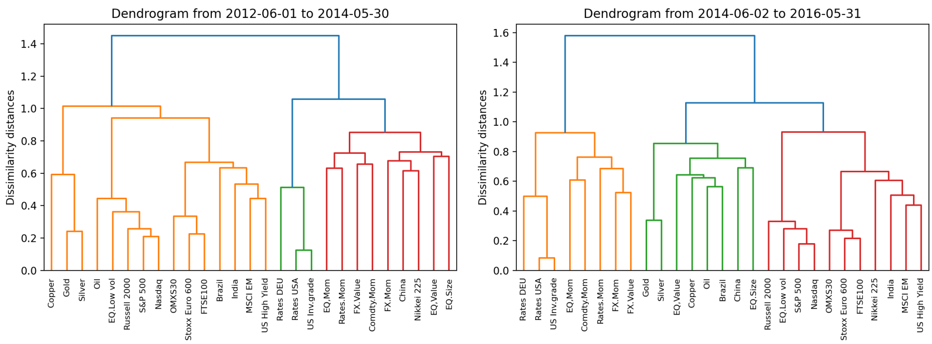

Figure 5.5: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2012-06-01 to 2014-05-30 to the left and from 2014-06-02 to 2016-05-30 to the right

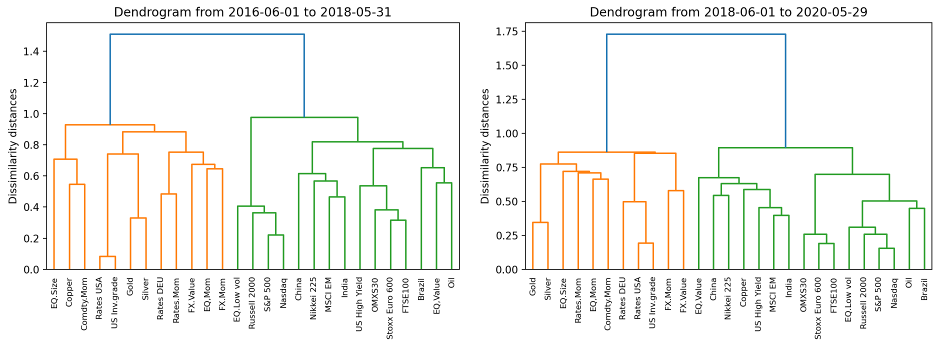

Figure 5.6: Obtained dendrogram using Ward's methods as linkage criterion over historical data from 2016-06-01 to 2018-05-31 to the left and from 2018-06-01 to 2020-05-29 to the right

Similar clusters recur several times during the different time periods in Figure 5.2 - 5.6. but not always grouped together with same cluster higher up in the dendrogram.

In Figure 5.7 can we see the two different models suggested numbers of clusters for the obtained dendrogram over two years, starting from 2000-05-25 to 2002-05-24 and then shifted forward one week at a time.

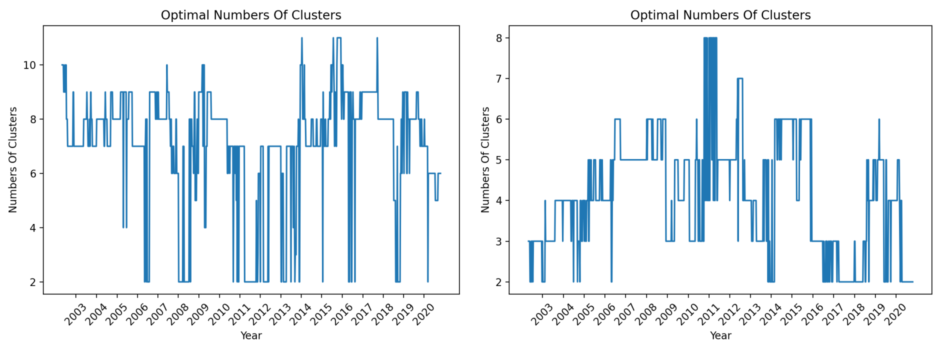

Figure 5.7: The suggested optimal number of clusters derived from equation 2.9 to the left and from equation 2.11 to the right based on data over a two year period

The two methods in Figure 5.7 differs from each other. Equation 2.9 indicates towards a higher numbers of clusters in comparison to equation 2.11. A high proportion of two suggested clusters, from both methods. Which could indicate in an incorrect suggestion.

5.2 Simulation

The different portfolios simulated index derived from the different allocation methods are compiled in Figure 5.8. Bootstapping simulations are used for the evolution of the assets next prices, with 1000 simulation on every day and each asset. Portfolio weights are calculated over the entire data time, from 2000-05-25 to 2020-10-30 and unchanged under the whole simulation. The next prices are simulated based on the historic data two years back.

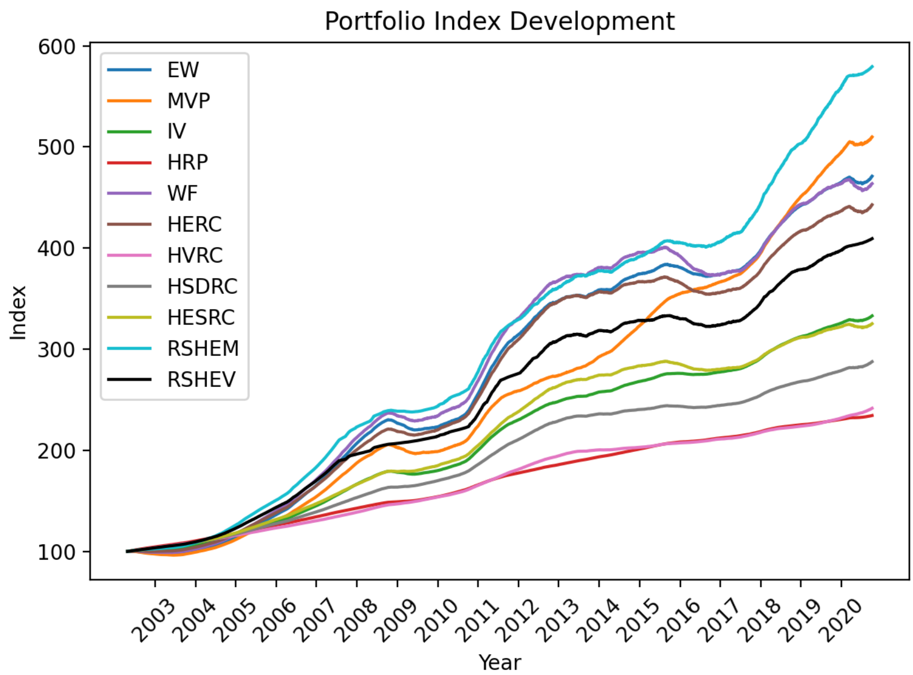

Figure 5.8: The different allocation methods simulated portfolio index obtained from bootstrap simulation, with Equal Weighting (EW), Minimum-Variance Portfolio (MVP), Inverse Volatility (IV), Hierarchical Risk Parity (HRP), Hierarchical Equal Risk Contribution (HERC), Hierarchical Variance Risk Contribution (HVRC), Hierarchical Standard Deviation Risk Contribution (HSDRC), Hierarchical Expected Shortfall Risk Contribution (HESRC), Regime Shift Hierarchical Equal with MPT (RSHEM) and Regime Shift Hierarchical Equal and Variance (RSHEV)

The simulated logarithmic returns obtained from all 1000 simulated portfolios, with bootstrapping are visualized in Figure 5.9-5.19. For the different allocation methods.

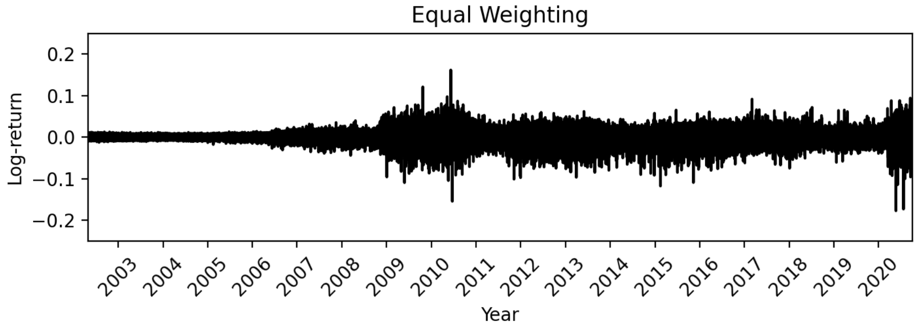

Figure 5.9: The logarithmic returns, obtained from bootstrap simulation for the equal weighting portfolio

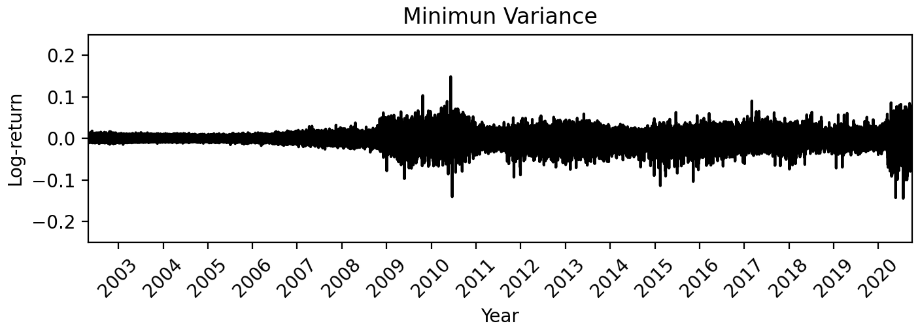

Figure 5.10: The logarithmic returns, obtained from bootstrap simulation for the minimum variance portfolio

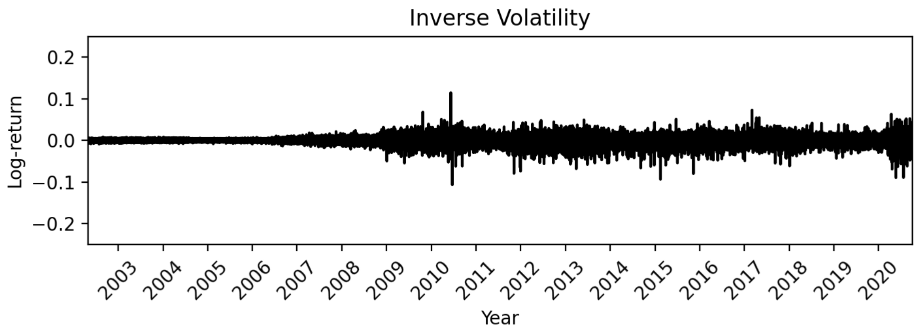

Figure 5.11: The logarithmic returns, obtained from bootstrap simulation for the inverse volatility portfolio

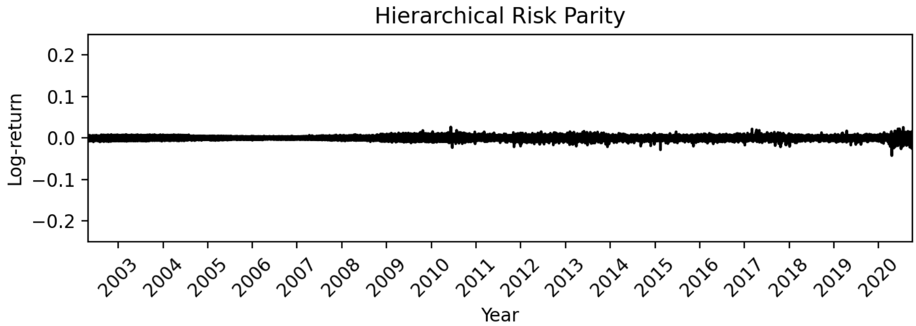

Figure 5.12: The logarithmic returns, obtained from bootstrap simulation for the hierarchical risk parity portfolio

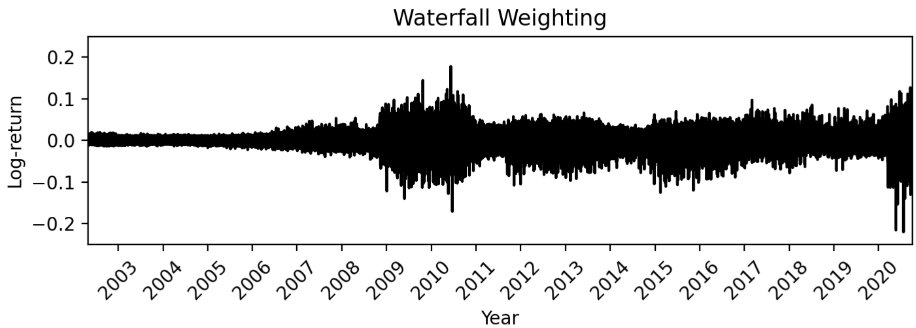

Figure 5.13: The logarithmic returns, obtained from bootstrap simulation for the waterfall weighting portfolio

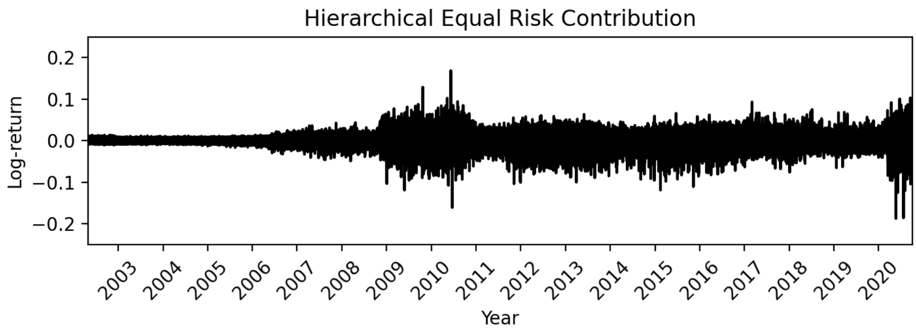

Figure 5.14: The logarithmic returns, obtained from bootstrap simulation for the hierarchical equal risk contribution portfolio

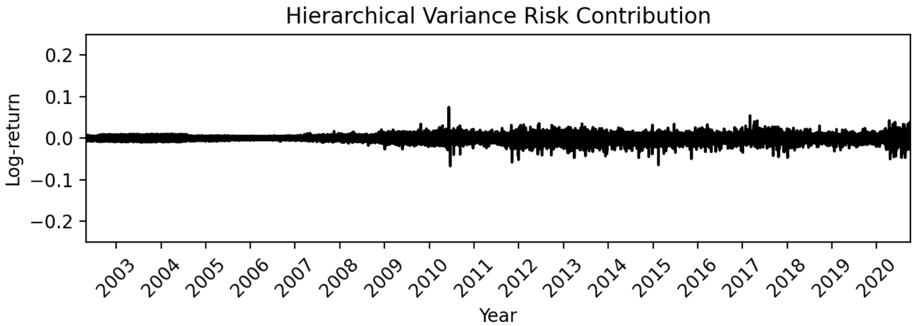

Figure 5.15: The logarithmic returns, obtained from bootstrap simulation for the hierarchical variance risk contribution portfolio

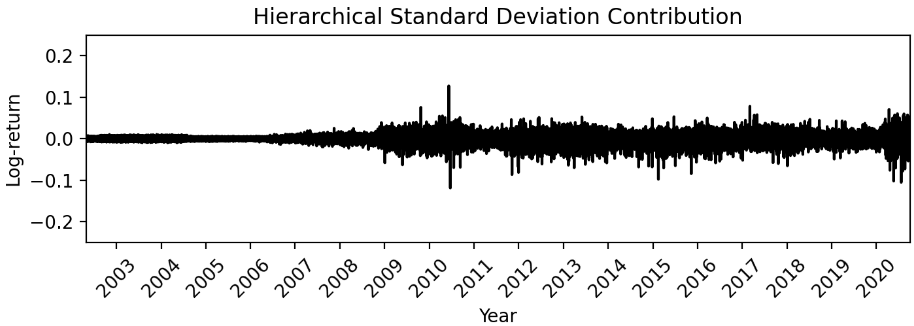

Figure 5.16: The logarithmic returns, obtained from bootstrap simulation for the hierarchical standard deviation risk contribution portfolio

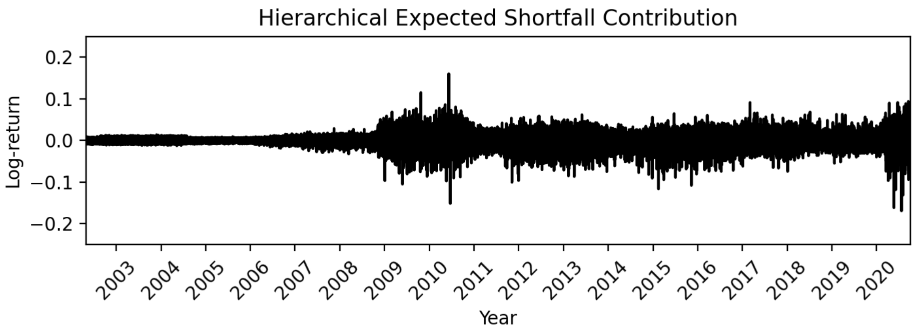

Figure 5.17: The logarithmic returns, obtained from bootstrap simulation for the hierarchical expected shortfall risk contribution portfolio

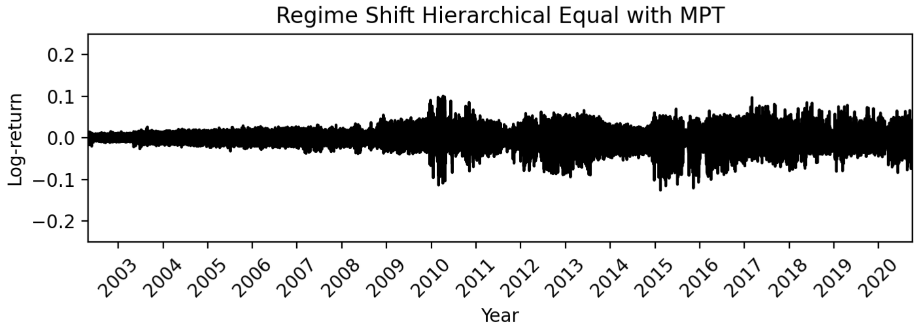

Figure 5.18: The logarithmic returns, obtained from bootstrap simulation for the regime shift portfolio based on hierarchical equal across cluster and MPT within

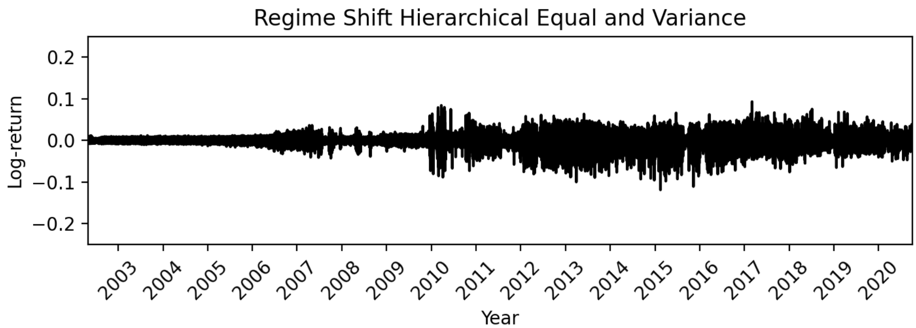

Figure 5.19: The logarithmic returns, obtained from bootstrap simulation for the regime shift portfolio based on hierarchical equal and variance risk contribution

5.3 Real-Value Test

The index over the different allocation methods in the real-value test are visualized in Figure 5.20, where we have used the Walk Forward process. The portfolios weights are derived over a two years in-sampled period and then applied on the upcoming out-of-sample period over a month. The periods are then shifted forward one month. The portfolios obtained from MPT are estimated with 1000 times number of assets.

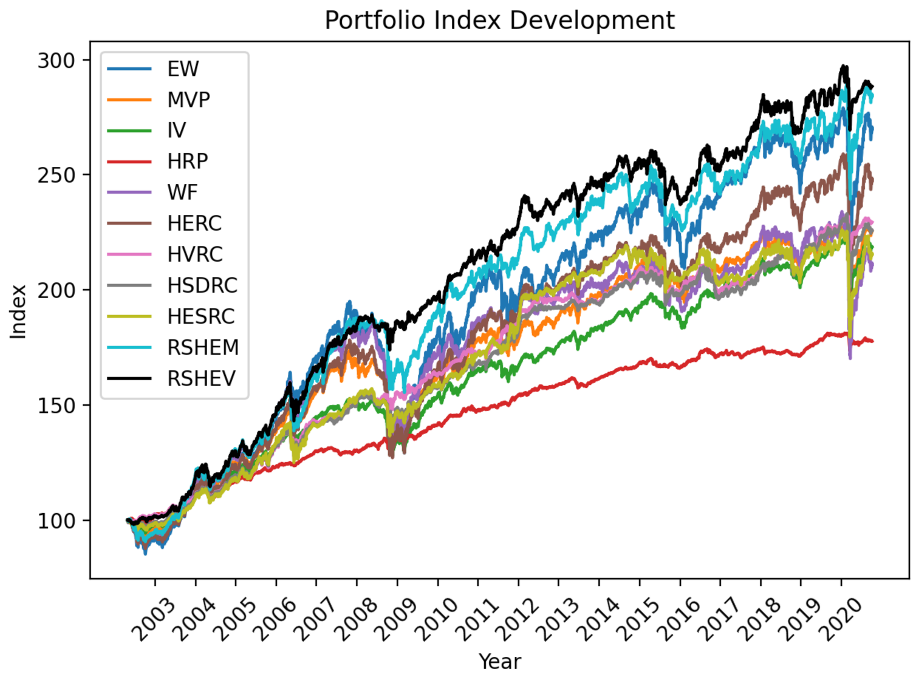

Figure 5.20: The different allocation methods portfolio index, obtained from walk forward process on historical data, with Equal Weighting (EW), Minimum-Variance Portfolio (MVP), Inverse Volatility (IV), Hierarchical Risk Parity (HRP), Hierarchical Equal Risk Contribution (HERC), Hierarchical Variance Risk Contribution (HVRC), Hierarchical Standard Deviation Risk Contribution (HSDRC), Hierarchical Expected Shortfall Risk Contribution (HESRC), Regime Shift Hierarchical Equal with MPT (RSHEM) and Regime Shift Hierarchical Equal and Variance (RSHEV)

The obtained logarithmic returns from the different allocation methods portfolios are visualized in Figure 5.21-5.24. Indicating the distribution for the different portfolios logarithmic returns over the entire test period.

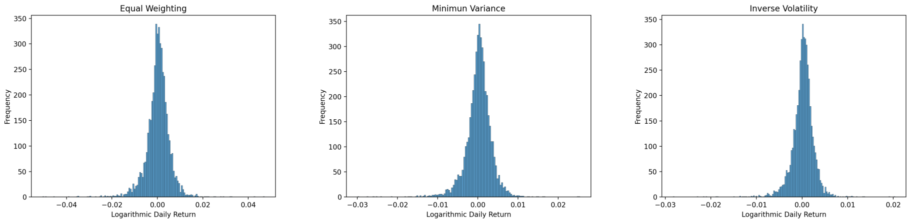

Figure 5.21: Distribution over the logarithmic returns obtained from the walk forward process regarding the equal weighting portfolio (left), minimum variance portfolio (center) and inverse volatility portfolio (right)

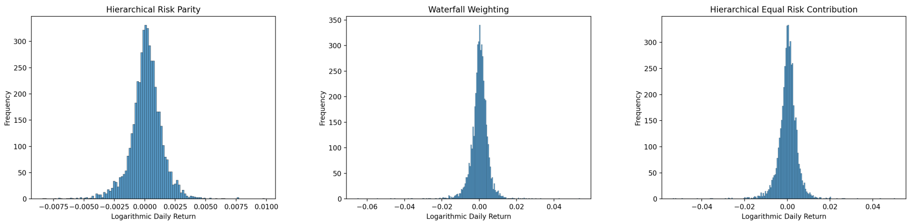

Figure 5.22: Distribution over the logarithmic returns obtained from the walk forward process regarding the hierarchical risk parity portfolio (left), waterfall weighting portfolio (center) and hierarchical equal risk contribution portfolio (right)

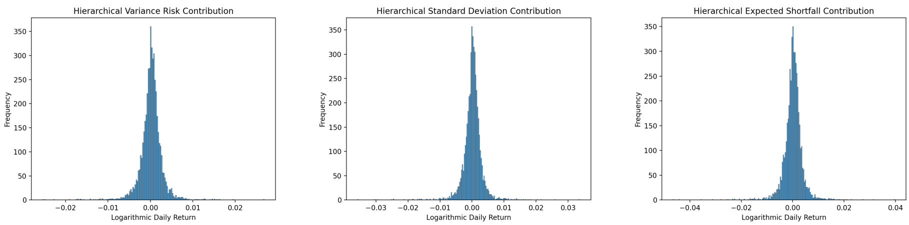

Figure 5.23: Distribution over the logarithmic returns obtained from the walk forward process regarding the hierarchical variance risk contribution portfolio (left), hierarchical standard deviation risk contribution portfolio (center) and hierarchical expected shortfall risk contribution portfolio (right)

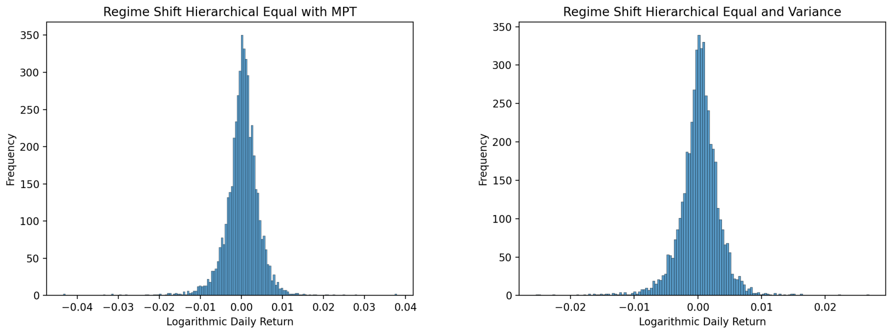

Figure 5.24: Distribution over the logarithmic returns obtained from the walk forward process regarding the regime shift portfolio based on hierarchical equal across cluster and MPT within (left) and the regime shift portfolio based on hierarchical equal and variance risk contribution (right)

Estimated values for annualized return, standard deviation, sharpe ratio, maximum drawdown, Calmar ratio, Sortino ratio and CVaR are compiled in Table 5.1

Table 5.1: Performance and risk statistics for the different allocation methods on the walk forward process from 2000-05-25 to 2020-10-30

	Classic Risk Based Allocation			Regime Shift Allocation		
	EW	MVP	IV	RSHEM	RSHEV	
Annualized Return	0.055	0.045	0.042	0.058	0.059	
Standard Deviation	1.00	0.56	0.40	0.85	0.67	
Sharpe Ratio	0.65	0.86	1.13	0.83	1.08	
Sortino Ratio	0.74	1.00	1.25	0.94	1.29	
Maximum Drawdown	0.34	0.24	0.16	0.18	0.10	
Calmar Ratio	0.16	0.19	0.27	0.33	0.59	
99% CVaR	2.64	1.50	1.05	2.21	1.73	
95% CVaR	2.03	1.15	0.81	1.70	1.32	
	Hierarchical Risk Based Allocation					
	HRP	WF	HERC	HVRC	HSDRC	HESRC
Annualized Return	0.032	0.041	0.051	0.046	0.045	0.043
Standard Deviation	0.19	0.83	0.84	0.43	0.50	0.65
Sharpe Ratio	1.59	0.51	0.67	1.14	0.96	0.68
Sortino Ratio	2.13	0.57	0.76	1.30	1.03	0.76
Maximum Drawdown	0.06	0.30	0.28	0.08	0.14	0.21
Calmar Ratio	0.58	0.14	0.18	0.57	0.33	0.21
99% CVaR	0.48	2.21	2.21	1.09	1.31	1.70
95% CVaR	0.37	1.70	1.70	0.84	1.01	1.31

5.4 One Year Of Covid-19

In Figure 5.25 are the allocation methods portfolio index illustrated over one year since the outbreak of Covid-19, from 2019-11-01 to 2020-10-30. Estimated values for maximum drawdown and CVaR are compiled in Table 5.2 estimated from the outbreak of Covid-19.

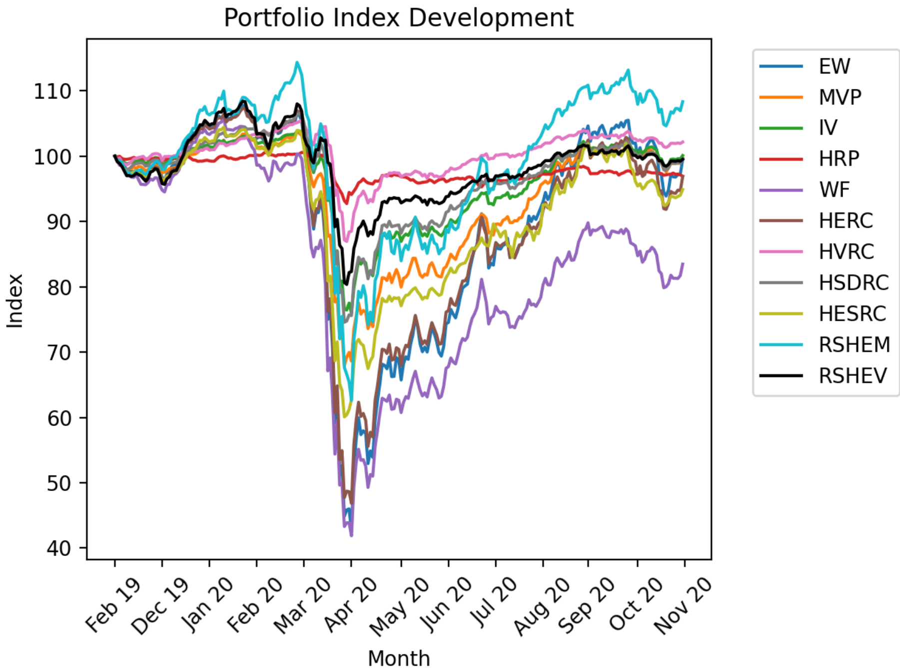

Figure 5.25: The different allocation methods portfolio index, obtained from walk forward process of the last year from 2019-11-01 to 2020-10-30

Table 5.2: Performance and risk statistics for the different allocation methods on the walk forward process from 2019-11-01 to 2020-10-30

	Classic Risk Based Allocation			Regime Shift Allocation		
	EW	MVP	IV	RSHEM	RSHEV	
Maximum Drawdown	0.61	0.34	0.27	0.45	0.26	
99% CVaR	5.86	3.32	2.42	4.88	2.50	
95% CVaR	4.54	2.57	1.87	3.77	1.93	
	Hierarchical Risk Based Allocation					
	HRP	WF	HERC	HVRC	HSDRC	HESRC
Maximum Drawdown	0.08	0.60	0.56	0.18	0.30	0.42
99% CVaR	0.76	5.15	5.44	1.71	2.84	3.81
95% CVaR	0.59	4.00	4.21	1.32	2.19	2.95

Chapter 6

Discussion

In this chapter will we discuss the models, parameters and results from previous chapters. Furthermore, will we reflect over the choices made in this study and give some suggestions for future research. Ending with a short summery of the thesis.

6.1 Choice of Model and Parameters

Before the result was achieved some choices had to be done, regarding the models and parameters. The choices made and its credibility will be discussed, before we go any further in on the result. Starting by taking a look at the selected data.

6.1.1 Data Selection

All data for the assets were retrieved from Bloomberg Terminal, a trustworthy source that many investors use. The exception is data representing factor investing in Equity Momentum and Equity Value who were extracted from constructed portfolios from French [7] momentum factor and Fama/French 3 Factors, based on Fama and French [6]. Also a reliable source.

The data in the created multi asset multi factor universe covers equities, commodities, bonds, credits and different factor investments, with investment styles momentum, value, size and low volatility. Which is a good distribution to represent the majority of investment styles. The majority of assets comes from equities. An even distribution over the different style factors and market factors could been used for a better outcome. Higher degree of combinations between asset and investment styles would affect the results to the better, but were difficult to come over for a long consecutive period of time.

6.1.2 Construction of Dendrogram

Logarithmic return were used to determine the dissimilarity distance, based on the assets correlations in equation 2.2. Correlations is the key to obtain diversity thru hierarchical clustering. Persons correlations, that were used on returns, gives an explanation on the short-term changes between assets and is a good measure for diversity between assets. Correlation based on last prices could improve valuations of long-term evolution and other formulas on returns, instead of Person coefficient could tell more if assets moves up and down together. It's important to know what one would like to measure and how to divide the assets, when constructing the dissimilarity distance. Different dissimilarity distances would result in different clusters with different information. Like correlations tells if assets and clusters move up and down together, tells cointegration if assets and clusters keeps the same distance to each other.

The thesis has considered two choices of linkage criterion, Ward's method and Single linkage, see equation 2.6 and 2.3. Single linkage were only used for HRP and is a good choice for HRP, then HRP only consider the bottom order of the assets. Ward's method results in compact clusters of similar size. Making it a good choice, since the goal in this thesis is to divide the assets into different clusters with hierarchical clustering and is confirmed by Papenbrock [19] and Raffinot [22].

6.1.3 Bootstrap Simulations

There are mainly two different methods to generate new price samples from historical data, Monte Carlo simulations and Bootstrap simulations. New prices with Monte Carlo simulation are produced by drawing from a hypothesized analytical distribution. Giving this method a replicated sample of prices that follows the same distributional properties as the original data. In this case and many other are the underlying distribution unknown. The advantage of new prices with bootstrap simulations, comes from that the method can derive estimates of standard errors and confidence intervals for complex estimators of the distribution. For most problems are the true confidence interval unknown. Bootstrapping is more accurate than standard intervals, obtained from sampling variance and assumptions of normality, DiCiccio and Efron [5].

The obtained logarithmic returns in Figures 5.9-5.19 all has their highest variance around 2010, due to the financial crisis around 2008. Since the Bootstrap simulation generates new prices based on the historical returns two years back, understandable that the variance from crises are shifted two years forward. Making the regime shift adjust two years earlier than the simulated market and can be seen from the HERC portfolio in Figure 5.8. A shorter time period could generate more relevant prices. A small effect from the outbreak of Covid-19, can be seen in the end of Figures 5.9-5.19 together with 2018.

6.1.4 Walk Forward

Using Walk Forward optimization allows us to always have the optimal portfolio weights for the different allocation methods and how one would invest in the future based on the past. The length of the out-of-sample period is set to one month, adjusting the new portfolios every month without taking transaction cost into account. An out-of-sample period of three months would reduce the rate of time purchasing new portfolio weights, but would cause unnecessary risk exposure during turmoil. The Walk Forward optimization are for the moment only optimizing the dendrogram towards the in-sample time. Optimizing more variables during the test period would improve the results, but would make the program slower.

6.2 Dendrogram

The dendrograms over time together with the corresponding reordered correlation matrices, according to the order derived from the dendrogram with Ward's methods, indicates that the assets are well diversified from hierarchical clustering. The corresponding reordered correlation matrices can be seen in Appendices C. During the different time periods similar groups of assets can be found together multiple times and in the dendrogram over the whole timeline, interpret that the hierarchical clustering method works for both longer and shorter time periods. Similar groups of clusters would be obtained if enough number of clusters were used. Such as OMXS30, Stoxx Euro 600 and FTSE100 or Russell 2000, S&P 500 and Nasdaq, are never far apart. A commodity cluster can also be seen with gold, silver and copper, while oil being one of the more unspecified assets together with Brazil.

It can be seen in Figure 2.6 that two methods to select an optimal number of clusters for the dendrogram varies. Method based on equation 2.9 varies from two to ten clusters with the majority between six to nine clusters. Method based on equation 2.11 varies from two to eight clusters with the majority between three to six clusters. Before defining the numbers of cluster to the dendrogram, it might be a good idea to double check the graphs in Figure 2.6 for each method. If the suggested number of clusters seems reasonable or it could be a wrongfully suggestion. Especially in the extreme cases.

6.3 Comparison Between the Methods

We will now move on to a closer look at the portfolios results, obtained from the different models and allocation methods.

6.3.1 Bootstrap vs Walk Forward

The obtained logarithmic returns from Walk Forward in Figure 5.21-5.24 have all similar variance on the logarithmic returns as their corresponding simulated portfolio in Figure 5.9-5.19 and within acceptable uncertainty. The simulated portfolios return are all much higher than the portfolio index obtained from historical values, mostly from a more stable increasing curve for the simulated portfolios and not as much drawdowns as in the portfolios based on historical data. Similar order based on return can be found between the different allocation methods in the obtained index value from simulated values and historical values. Simulated index value for RSHEM and MVP are much higher compared to the other portfolios respective to their values from Walk Forward. RSHEM and MVP are both constructed from MPT. Portfolio based on MPT has a statistically advantage when applied on simulated data based on the same data as MPT is constructed from. The increased return for HVRC comes from lower numbers of suggested numbers of clusters. Not generating the same amount of weighting towards assets with low variance.

6.3.2 Different Allocation Methods

HRP differ mostly from the other allocation methods, with a very low standard deviation, MDD and CVaR. Also have a high Sharpe ratio, Sortino ratio and Calmar ratio, but lack in the annualized return. WF is a basically a worse version of HERC on almost all parameters. HERC differ from the other hierarchical clustering methods, with a higher annualized return but inferior on the other parameters. HVRC, HSDRC and HESRC have similar values and understandable from its similar weighting, which can be seen in Appendices B Figure B.2 and B.3. Where a small edge in performance is given to HVRC.

Resembling annualized return between EW and HERC, with a small edge to EW. Comparing EW and HERC, a statistically better risk-adjusted performances can be found in HERC. In the same way can a statistically better risk-adjusted performances be found in HVRC in comparison to MVP and IV. Even if IV has a lower standard deviation and CVaR, since also a lower annualized return. The statistically better risk-adjusted performances in HERC and HVRC follows in the last year's performance, seen in Figure 5.25 and Table 5.2. The two portfolios based on regime shift RSHEM and RSHEV outperforms the other portfolios without transaction costs and resists big drawdowns. Where an advantage is given to the RSHEV on all parameters, mostly in CVaR and maximum drawdown. RSHEM could suffers from too few portfolio simulation for MPT, where an increased number would generate more reliable results.

6.3.3 Portfolio Suggestion

A good portfolio with focus on left hand tail risk is considered by low values of MDD and CVaR combined with high values of annualized return and Calmar ratio. As mention before does HRP outperform all the other portfolios except when it comes to annualized return, where it's by far the worst. What is considered a good portfolio is always a balance between how much risk one is willing to take in comparison to return, where Calmar ratio is a good measure for that. HRP has a high Calmar ratio, so is considered as a good portfolio with low risk.

My own recommendation between all of the portfolios would be RSHEV, which has some of the best values and doesn't lack in any of the other parameter. An interesting thought would be to combine a new portfolio. Based on regime shift, one would construct the portfolio just like RSHEM. But instead of weighting the portfolio like HVRC at risk-off, one would weight as HRP.

6.4 Improvements and Further Research

As mention earlier, improvements could be made in the selection of data and assets. With a more evenly spread over the different markets and assets with equally value in the beginning, would use the full potential from hierarchical clustering. As seen in Figure 3.1 the use of 27000 simulated portfolios to MPT was not optimal, an increased number of simulations would improve the results. A lot of improvements could be made in the Walk Forward optimization process, where a lot more parameters could be optimized during the in-sample test and also minimizing the transaction costs.

An interesting topic for further research, would be to investigate different methods to derive the correlation matrices. Where one could investigate if there is a better way to determine the similarity between assets and if this improves the diversity with hierarchical clustering. There are many ways to allocate the investment capital along the dendrogram and there are a lot more methods that could be tested. The regime shift successfully minimize big drawdowns for the portfolios, but miss out most of the increasing market that follows. It would be interesting to test if there is a more optimal way to switch from risk-off to risk-on after a big drawdown.

6.5 Summary

The purpose of this thesis was to validate if a better diversified portfolio could be achieved from Hierarchical Clustering Based Asset Allocation compared to more traditional asset allocation strategies, with focus on left hand tail risk. Dissimilarity distance between assets were derived from the asset's correlation coefficients, which the dendrogram was based on. Allocation methods were applied on the dendrogram, both within and across clusters. Portfolios were tested on simulated and historical values. The model has room for improvements, but generally gives trustworthy results. Hierarchical clustering-based portfolios are truly diversified and achieve a statistically better risk-adjusted performances than commonly used portfolio optimization techniques. With focus on the left-hand tail risk, except for Inverse Volatility who had similar results. The two portfolios based on hierarchical clustering with regime shift outperformed the other portfolios.

Bibliography

[1] Barroso, P and Santa-Clara, P. "Managing the risk of momentum". In: (2012). eprint: https://www.cbs.dk/files/cbs.dk/momentumrisk.pdf(Online;accessed07-Feb-2021).

[2] Barroso, P and Santa-Clara, P. "Momentum Has Its Moments". In: *Journal of Financial Economics* 116.1 (2015), pp. 111–120.

[3] Bernardi, S., Leippold, M. and Lohre, H. "Maximum diversification strategies along commodity risk factors". In: *European Financial Management* 24.1 (2018), pp. 53–78.

[4] Deguest, R., Meucci, A. and Santangelo, A. "Risk budgeting and diversification based on optimized uncorrelated factors". In: *Risk* 11.29 (2015), pp. 70–75.

[5] DiCiccio, T. J. and Efron, B. "Bootstrap Confidence Intervals". In: *Statistical Science* 11.3 (1996), pp. 189–228.

[6] Fama, E. F., French, K. R. "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics* 33.1 (1993), pp. 3–56.

[7] French, K. R. "Current Research Returns". In: (2020). eprint: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html(Online;accessed13-Nov-2020).

[8] Hastie, T., Tibshirani, R. and Walther, G. "Estimating the number of clusters in a data set via the gap statistic". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2 (2001), pp. 411–423.

[9] Kolrep, M. Lohre, H., Radatz, E. and Rother, C. "Economic versus statistical clustering in multi-asset multi-factor strategies". In: *Risk & Reward* (2020), pp. 26–32. eprint: https://www.invesco.com/content/dam/invesco/apac/en/pdf/insights/risk-and-reward-q1-2020/multi-asset-20200206-risk-and-reward-Q1-2020.pdf(Online;accessed03-Dec-2020).

[10] Liew, L. "What is a Walk-Forward Optimization and How to Run It?" In: *AlgoTrading101* (2020). eprint: `https://algotrading101.com/learn/walk-forward-optimization/(Online;accessed07-Feb-2021)`.

[11] Lillo, F., Mantegna, R. N. and Tumminello, M. "Correlation, Hierarchies, and Networks in Financial Markets". In: *Journal of Economic Behavior and Organization* 75.1 (2010), pp. 40–58.

[12] Lohre, H., Opfer, H. and Ország, G. "Diversifying risk parity". In: *Journal of Risk* 16.5 (2015), pp. 53–79.

[13] Lopez de Prado, M. "Building diversified portfolios that outperform out of sample". In: *Journal of Portfolio Management* 42.4 (2016), pp. 59–69.

[14] Maillard, S., Roncalli, T. and Teiletche, J. "The properties of equally weighted risk contribution portfolios". In: *Journal of Portfolio Management* 36.4 (2010), pp. 60–70.

[15] Mantegna, R. "Hierarchical structure in financial markets". In: *The European Physical Journal B - Condensed Matter and Complex Systems* 11.1 (1999), pp. 193–197.

[16] Markowitz, H. "Portfolio Selection". In: *Journal of Finance* 7.1 (1952), pp. 77–91.

[17] Meucci, A. "Managing diversification". In: *Risk* (2009), pp. 35–40.

[18] Nystrup, P., Hansen, B. W., Larsen, H. O., Madsen, H. and Lindström, E. "Dynamic allocation or diversification: a regime-based approach to multiple assets". In: *Journal of Portfolio Management* 44.2 (2017), pp. 62–73.

[19] Papenbrock, J. "Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization". In: *Ph.D. thesis* (2011).

[20] Pažický, M. "Stock Price Simulation Using Bootstrap and Monte Carlo". In: *Scientific Annals of Economics and Business* 64.2 (2017), pp. 155–170.

[21] Prim, R. C. "Shortest Connection Networks And Some Generalizations". In: *Bell System Technical Journal* 36.6 (1957), pp. 1389–1401.

[22] Raffinot, T. "Hierarchical Clustering-Based Asset Allocation". In: *Journal of Portfolio Management* 44.2 (2017), pp. 89–99.

[23] Raffinot, T. "The Hierarchical Equal Risk Contribution Portfolio". In: (2018).

Appendix A

Data Selection

The chosen assets to create the multi asset multi factor universe are presented in Table A.1. Extract from Bloomberg terminal with their given Bloomberg Ticker. The selected asset to represent factor investments with equities and momentum and factor investments with equities and value, were extracted from constructed portfolios from French [7]. Based on Fama and French [6].

Table A.1: The selected assets in the multi asset multi factor universe, retrieved with Bloomberg terminal

Equities Developed Markets		
Description	Asset	Bloomberg Ticker
Swedish stock market	OMXS30	OMXS30B Index
European stock market	Stoxx Euro 600	SXXP Index
American stock market	S&P 500	SPXT Index
American stock market	Nasdaq	XNDX Index
American stock market	Russell 2000	RTY Index
London stock market	FTSE100	UKX Index
Tokyo stock market	Nikkei 225	NKY Index
Equities Emerging Market		
Description	Asset	Bloomberg Ticker
Emerging markets	MSCI EM	NDUEEGF Index
Shanghai stock market	China	SHCOMP Index
Brazilian stock market	Brazil	IBOV Index
Bombay stock market	India	SENSEX Index
Commodities		
Description	Asset	Bloomberg Ticker
Commodity market for gold	Gold	XAU Curncy
Commodity market for silver	Silver	XAG Curncy
Commodity market for copper	Copper	SPGSIC Index
Commodity market for oil	Oil	SPSIOP Index
Bonds		
Description	Asset	Bloomberg Ticker
US Government Bonds (10-30y maturity)	Rates USA	LUATTRUU Index
German Government Bonds (10-30y maturity)	Rates DEU	BCEG1T Index
Credits		
Description	Asset	Bloomberg Ticker
US credits investment grade	US Inv.grade	LBUSTRUU Index
US credits high yield	US High Yield	LF98TRUU Index
Factor Investments		
Description	Asset	Bloomberg Ticker
Factor investment equities and low volatility	EQ.Low vol	SP5LVI
Factor investment equities and size	EQ.Size	IJR US Equity
Factor investment commodities and momentum	Comdty.Mom	DBABCM Index
Factor investment rates and momentum	Rates.Mom	DBABRMB Index
Factor investment foreign exchange and momentum	FX.Mom	DBMOMEFF Index
Factor investment foreign exchange and value	FX.Value	DBPPPEFF Index

Appendix B

Portfolio weights

In Figure B.1, B.2 and B.3 are the given portfolio weights obtained from the different allocation methods. The portfolio weights from the allocation methods are based on the whole data set from 2000-05-25 to 2020-10-30. The MPT portfolios is obtained based on number of assets times 1000 of random portfolios. Demdrogram is derived on wards linkage criterion, except for HRP derived in single linkage. Optimal number of clusters is set to seven and α is set to 5% in equation 3.11.

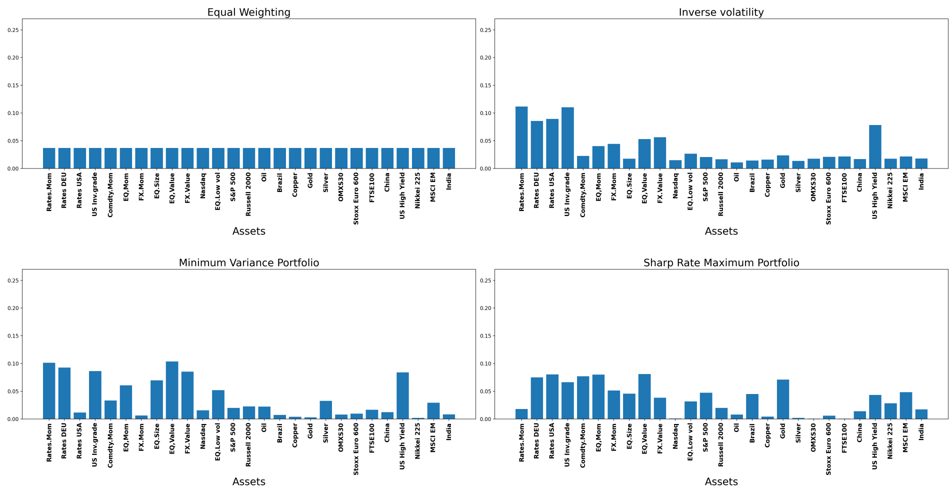

Figure B.1: Obtained portfolio weights from allocation methods equal weighting (top left), inverse volatility (top right), MVP (bottom left) and maximum sharpe ratio (bottom right)

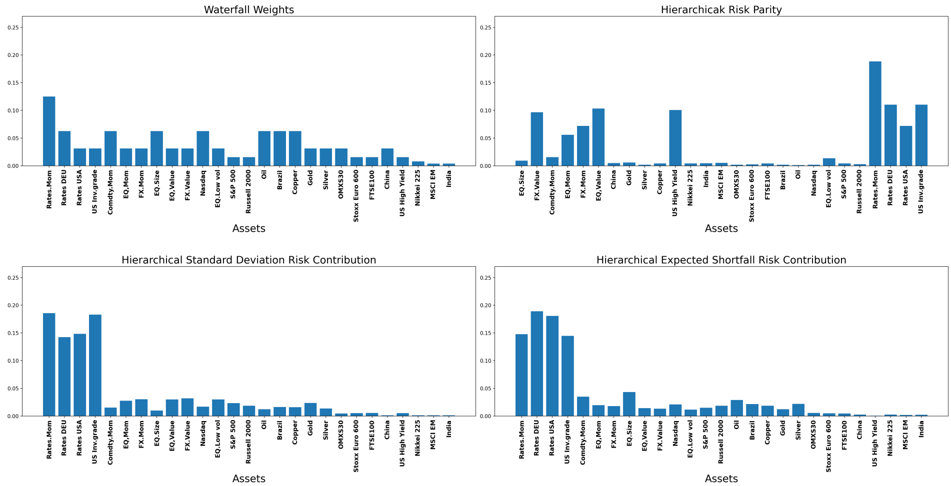

Figure B.2: Obtained portfolio weights from allocation methods waterfall weighting (top left), HPR (top right), hierarchical standard deviation risk contribution (bottom left) and hierarchical expected shortfall risk contribution (bottom right)

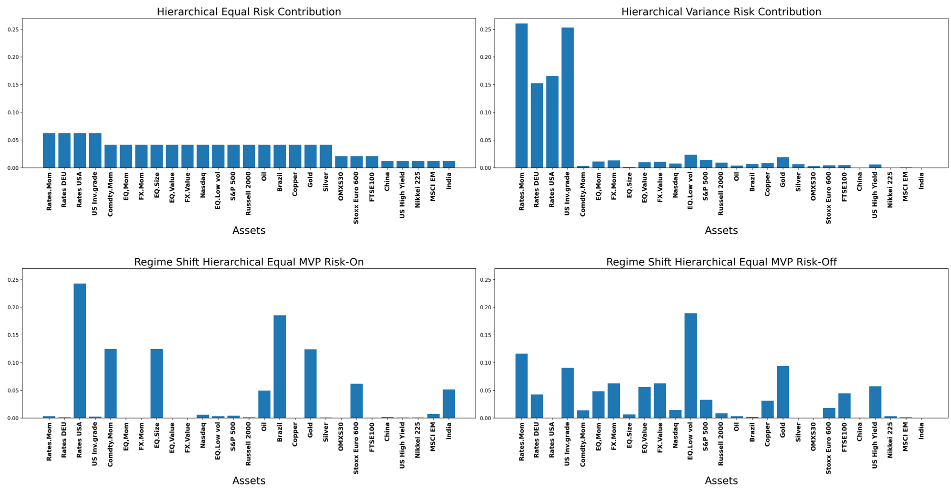

Figure B.3: Obtained portfolio weights from allocation methods hierarchical equal risk contribution (top left), hierarchical variance risk contribution (top right), regime shift portfolio based on hierarchical equal across cluster and maximum sharpe ratio within (bottom left) and regime shift portfolio based on hierarchical equal across cluster and MVP within (bottom right)

Appendix C

Correlation Matrices

In Figure C.1-C.5 are the obtained correlation matrices, reordering the assets accordingly to the dendrogram over different time periods from Figure 5.2-5.6

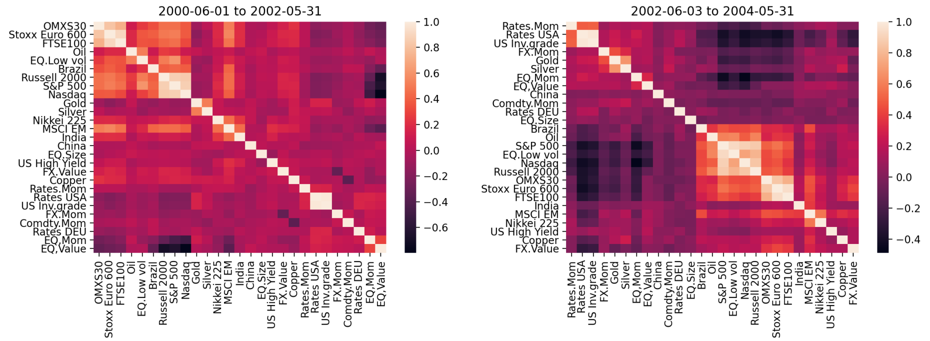

Figure C.1: Obtained correlation matrices, reordering assets according to dendrogram using Ward's methods as linkage criterion over historical data from 2000-05-25 to 2002-05-31 to the left and from 2002-06-03 to 2004-05-31 to the right

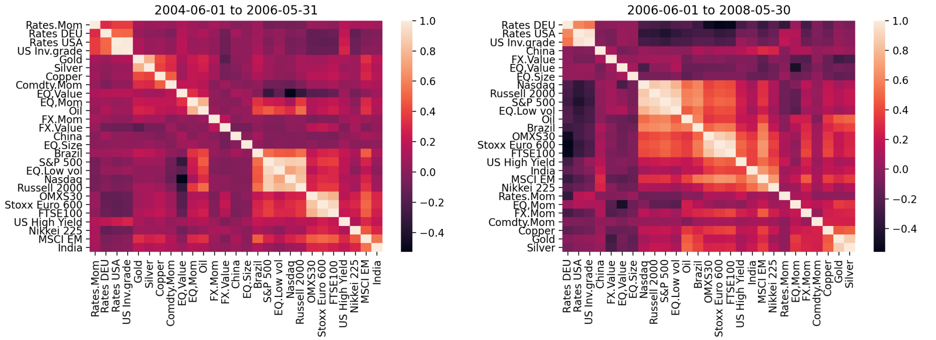

Figure C.2: Obtained correlation matrices, reordering assets according to dendrogram using Ward's methods as linkage criterion over historical data from 2004-06-01 to 2006-05-31 to the left and from 2006-06-01 to 2008-05-30 to the right

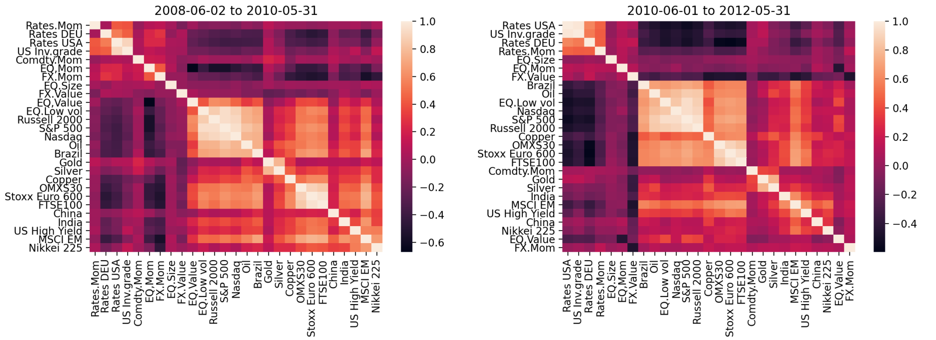

Figure C.3: Obtained correlation matrices, reordering assets according to dendrogram using Ward's methods as linkage criterion over historical data from 2008-06-02 to 2010-05-31 to the left and from 2010-06-01 to 2012-05-31 to the right

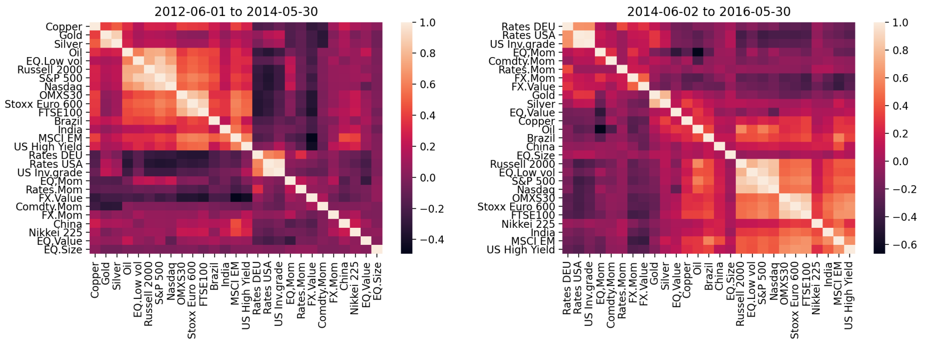

Figure C.4: Obtained correlation matrices, reordering assets according to dendrogram using Ward's methods as linkage criterion over historical data from 2012-06-01 to 2014-05-30 to the left and from 2014-06-02 to 2016-05-30 to the right

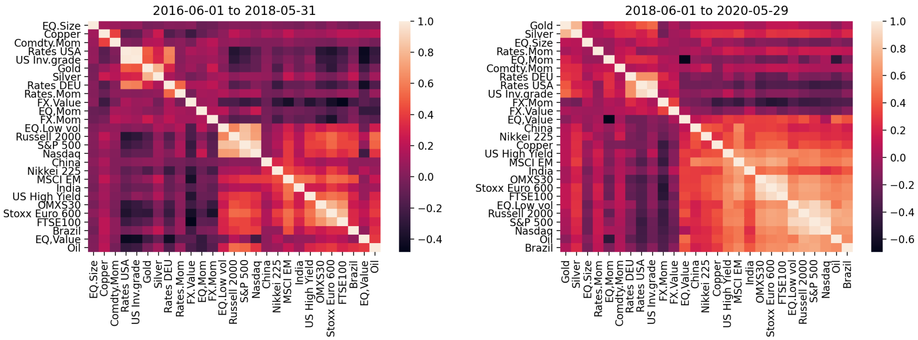

Figure C.5: Obtained correlation matrices, reordering assets according to dendrogram using Ward's methods as linkage criterion over historical data from 2016-06-01 to 2018-05-31 to the left and from 2018-06-01 to 2020-05-29 to the right