POLITECNICO DI TORINO

Master's Degree in Mathematical Engineering

Master's Degree Thesis

Cluster based portfolio optimization under uncertainty: Statistical and Robust approaches

Supervisors

Prof. Edoardo FADDA

Cristian URBINATI

Candidate

Michele NARESE

JULY 2025

Summary

Portfolio optimization refers to the process of determining the ideal fraction of an investment capital across a set of assets so that it meets specific requirements or constraints, while aiming for an objective such as minimizing risk or maximizing returns. Often in academic research, this problem is treated as a standalone task, with the portfolio of available securities assumed to be predefined and given as input.

In a realistic scenario, the first step towards an optimal allocation is the choice of that portfolio of securities from a set of possible investment opportunities. This thesis aims at studying in detail the two fundamental steps of portfolio management, namely portfolio selection and asset allocation. The first step is addressed through the application of several clustering algorithms in order to extract useful knowledge from a vast set of investment assets. Methods like K-means and hierarchical clustering have proven useful for many different tasks including data summarization, enhanced decision-making and ultimately, efficient portfolio selection. The resulting clusters have been analyzed in detail and a well-diversified portfolio has been built. Later, it served as a dataset for studying and comparing different portfolio optimization strategies.

Optimal asset allocation has been historically studied by Harry Markowitz when he first introduced the Mean Variance Portfolio: an optimization problem that aims at maximizing the expected return of the financial position while maintaining a desired level of risk, measured in volatility or variance of the resulting portfolio. However, this method suffers from several drawbacks especially during uncertain market conditions, and for this reason advanced mathematical methods have recently been developed to overcome the criticalities of Markowitz asset allocation.

This project is focused on two main approaches: hierarchical clustering-based allocation and robust optimization allocation. For what concerns the first category, the main contributions in the literature have been implemented and a different way of estimating the similarity matrix has been tested: this new similarity matrix is based on the concept of lower tail dependence coefficient estimated with two non-parametric, copula-based estimators.

The second approach, with the use of a quadratic-uncertainty set, consists of a

two-step optimization problem that can be efficiently solved when reformulated as second-order cone program.
Since both approaches are relatively new, the impact of the optimization parameters involved has been studied in detail with a particular emphasis on their performance during periods of great recession, subsequently different combinations of parameters have been selected, representing different risk-aversion levels of an ideal investor.
Finally, the chosen configurations have been tested in a simulation with real stock data and compared against several classical approaches and a real benchmark in order to assess the goodness of the developed methods.

Table of Contents

List of Tables vii

List of Figures viii

Acronyms xi

1 Introduction 1

1.1 Research Questions and Objectives 2

2 Theoretical Background 4

2.1 Classic Approaches for Portfolio Optimization 4

2.1.1 Mean Variance Portfolio 4

2.1.2 Global Minimum Variance Portfolio 5

2.2 Overcoming estimation errors on returns: hierarchical risk allocation and robust optimization 6

2.2.1 Robust Optimization 7

2.2.2 Clustering Techniques and their application to Portfolio Optimization 9

3 Asset allocation with Clustering methods 12

3.1 Hierarchical Risk Parity 12

3.2 Hierarchical Equal Risk Contribution 16

3.3 Hierarchical Allocation with Lower Tail Dependency Coefficient . . 18

3.3.1 Copulas and LTDC estimators 18

3.3.2 A general framework for LTDC-based allocation 21

4 Robust Portfolio Optimization 22

4.1 Problem Definition 22

4.2 Problem Reformulation 24

4.3 Hyperparametrs estimation 25

5 Clustering analysis of assets 28
5.1 Clustering strategies 28
5.1.1 K-means 29
5.1.2 Hierarchical Clustering 30
5.2 Experimental Setting 31
5.2.1 Datasets 31
5.2.2 Methodology 34
5.3 Results 37
5.3.1 S&P 500 stocks dataset 37
5.3.2 Multi-asset dataset 41
5.3.3 Final considerations 44

6 Simulations and Experimental Results 46
6.1 Framework 46
6.1.1 Dataset 47
6.1.2 Evaluation metrics 48
6.1.3 Simulations and Setting 50
6.1.4 Approaches tested 53
6.2 RO-MVP analysis and hyper-parameters choice 55
6.3 HERC-LTDC analysis and hyper-parameters choice 60
6.4 Real-life Investment Simulation 66

7 Conclusions 73
7.1 Personal Contribution 74
7.2 Further Research 74

Bibliography 75

List of Tables

4.1 Comparison of different choices for $\boldsymbol{\Omega_\mu}$. Source: Table 1 in [16]. . . 26

5.1 Features considered for clustering SP500 dataset 33
5.2 Features considered for clustering multi asset dataset. 35

6.1 Dataset composition. 49
6.2 Asset allocation methods tested. † This is indeed a simple heuristic that has proven to be quite hard to beat, for example the authors in [25] find that "Of the fourteen models of optimal portfolio choice [...], none is consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover." †† This repartition corresponds to a "Conservative Target Risk" in Morningstar's U.S. allocation categories. 54
6.3 Hyperparameters options for RO-MVP † even if γ isn't properly an optimization parameter but more of a value derived from the investor's characteristics, multiple choices have been tested to asses the influence of different values on the resulting portfolios. 55
6.4 Investment strategies chosen for RO-MVP approach 56
6.5 Results of comparison between MVP and their Robust counterpart (the best value for each metric is highlighted in bold). 60
6.6 Hyperparameters options for HERC-LTDC 61
6.7 Investment strategies chosen for HERC-LTDC approach 62
6.8 Results of comparison between MVP and hierarchical allocation methods (the best value for each metric is highlighted in bold). . . . 65
6.9 Results of the simulation performed between January 2022 and June 2025 (the best value for each metric is highlighted in bold, while underlined values are less than 5% far to the best value). 68

List of Figures

3.1 Dendrogram of a subset of multi-asset dataset, described in 5.2.1 . . 14
3.2 Quasi-diagonalization applied to a subset of the multi-asset dataset 14
3.3 Bisection based on the number of assets and bisection induced by the dendrogram. Source: Exhibit 3 in [9] 17

5.1 Different dendrograms produced by changing only linkage type on multi-asset dataset 30
5.2 Subset of S&P 500 sector composition 32
5.3 Multi asset dataset sector composition 34
5.4 Metrics for different values of K. As K-means is sensitive to initialization, several runs for each value of K are performed. 37
5.5 Adjusted Rand Index among pairs of clustering approaches on S&P500 dataset 39
5.6 Sector composition of clusters for S&P500 dataset 40
5.7 Clusters average return and volatility on S&P500 dataset 40
5.8 Clusters average Sharpe Ratio and ESG total risk on S&P500 dataset 41
5.9 Clusters average ROE and ROA on S&P500 dataset 41
5.10 Adjusted Rand Index among pairs of clustering approaches on multi asset dataset 42
5.11 Asset type composition of clusters for multi asset dataset 43
5.12 Clusters average return and volatility on multi asset dataset 44
5.13 Clusters average Sharpe Ratio and CVaR on multi asset dataset . . 44

6.1 Correlation heatmap of the selected investment portfolios 48
6.2 Maximum drawdowns by asset. Blue points indicate the two worst drawdowns for each asset, separated by at least 2 years. 51
6.3 Rolling 30 day annual volatility by asset. The two red points highlight highest volatility periods separated by at least 2 years. 52
6.4 Portfolio evolution over time for top 20 hyper-parameters combinations in RO-MVP optimization. 56

6.5 Portfolio optimization aggregated performance based on different hyper-parameters for RO-MVP optimization. 57
6.6 Portfolio optimization performance dispersion based on different hyper-parameters for RO-MVP optimization. 58
6.7 Portfolio cumulative returns for RO-MVP strategies and classical MVP optimization. 59
6.8 Portfolio evolution over time for top 20 hyper-parameters combinations in HERC-LTDC optimization. 61
6.9 Portfolio optimization performance dispersion based on different hyper-parameters for HERC-LTDC allocation. 63
6.10 Proportion of simulations in the top 10% of results by Sharpe ratio for each parameter for HERC-LTDC simulations. 64
6.11 Portfolio cumulative returns for different HERC strategies and HRP allocation. 65
6.12 Portfolio cumulative returns, all methods between 2022 and the end of 2023. 66
6.13 Portfolio cumulative returns, all methods between 2024 and January 2025. 67
6.14 Evolution over time of weights, all methods. 70
6.15 Total reallocation costs. 71
6.16 Monthly returns (mean values reported in red). 72

Acronyms

AR-GARCH
AutoRegressive - Generalized AutoRegressive Conditional Heteroskedasticity

ARI
Adjusted Rand Index

CAGR
Compounded Annual Growth Rate

CDaR
Conditional Drawdown at Risk

CSR
Conditional Spearman's Rho

CVaR
Conditional Value at Risk

EWP
Equal Weighting Portfolio

GMVP
Global Minimum Variance Portfolio

HCAA
Hierarchical Clustering-based Asset Allocation

HERC
Hierarchical Equal Risk Contribution

HRP

Hierarchical Risk Parity

IVarP

Inverse Variance Portfolio

LTDC

Lower Tail Dependency Coefficient

MVP

Mean Variance Portfolio

QP

Quadratic Program

RO

Robust Optimization

SOCP

Second-Order Cone Program

Chapter 1

Introduction

Financial portfolio management has been a central concern since commodity were first traded in late 13th-century Bruges. The possible reward of owning goods, securities or, in general, assets has been of increasing importance up to recent times. Today more than ever, this topic is of growing interest both socially and academically.

Portfolio management refers to the art of selecting and managing a collection of investments that are owned by a particular person or organization. This decision process can often be categorized in two subsequent steps:

1. Asset Selection, which involves identifying appropriate assets classes based on the objectives, market conditions and investors characteristics.
2. Capital Allocation, that is determining the optimal fraction of the investment budget to distribute across the investment portfolio.

Both phases have been widely studied and analyzed but still, they remain open problems given the complex nature of financial markets and the risk associated to them. Recent advances in machine learning and new optimization paradigms have opened the possibility to exploit the latest analysis techniques of dependence and interaction between different assets with unknown behavior.

This thesis project aims at studying how advanced mathematical and statistical models can help perform both steps in portfolio management: identify patterns in complex environments such as big investment universes, as knowledge extraction and pattern recognition are key for building diversified portfolios, and allocate the capital in order to minimize the risk associated and maximize its reward, especially during unstable market conditions. Given that this field of study is relatively recent, the main focus of this work will be on implementing and studying new methods starting from the criticalities and limitations of the approaches present in the literature, as well as evaluating how such methods can hedge losses during crises in real-world scenarios.

This thesis will be then structured in four main parts: first of all the theory behind portfolio optimization and asset allocation will be reviewed in order to give a comprehensive overview of the relevant concepts on which this work will be based on. Secondly, the methods that have been developed in this project will be presented and motivated from both a theoretical and a practical perspective. Finally, real-life experiments will be carried out encompassing both asset selection and portfolio optimization. The first objective will be achieved by applying advanced clustering techniques to different investment universes in order to shed some light on how valuable insights can be inferred using these methods. The final part will be dedicated to the practical analysis and evaluation of the proposed approaches by performing different simulations with real market data and comparing how these novel methods can hedge risk compared to standard and classical portfolio optimization models.

1.1 Research Questions and Objectives

This research project will be developed in a corporate setting, therefore it will serve a dual purpose by exploring how new mathematical models could improve existing ones, but at the same time keeping an eye on their practical implications and implementation. To this end, particular relevance will be given to providing a strong theoretical formulation which still maintains explainability and real-life application.
The *fil rouge* outlined already from the first page of this dissertation will then be motivated by two big sets of research questions, belonging to both steps of portfolio management.

First, we observe whether valuable information can be extracted by applying clustering algorithms to different datasets of assets. This has a dual role:

- Do clustering models help increasing context awareness and therefore improve the decision making process?
- Can we further exploit this knowledge for building well-diversified portfolios that hedge risk of large losses during periods of high market instability?

Then, attention will be given to actual capital allocation models. We will try to assess the quality of new approaches to a problem that has been intensively studied for almost a century now:

- How are the models parameters of the new approaches affecting capital allocation?
- Are the implemented methods superior in terms of risk mitigation during recession periods?

- How do such methods perform under stable market conditions?

Chapter 2

Theoretical Background

This chapter will introduce theoretical concepts crucial to understand the methods presented in the next chapters. It will include a general overview of the problem addressed, as well as how it has been classically formulated and solved. It will be explained why classical approaches may not always be ideal and propose methods that might overcome their criticalities.

2.1 Classic Approaches for Portfolio Optimization

Portfolio Optimization refers to the process of selecting assets or financial instruments aiming for an investment that is optimal with respect to a given metric, which usually takes into account returns (the reward for the investor) and risks (what threatens the investor).
It's exactly based on this intuition that Harry Markowitz wrote the seminal paper "Portfolio Selection" in 1952 [1]. He introduced the idea of risk-adverse investors, that should aim at finding their portfolio based on a combination of two objectives: risk and expected returns. The Vanilla Markowitz portfolio formulation (that will be referred as MVP from now on) identifies as risk measure the volatility or, equivalently, the variance of an investment.

2.1.1 Mean Variance Portfolio

Imagine an investor who has to choose between a given number of investment opportunities -let us say n-, then it will have to decide how much of the initial capital will be invested on each one of them. We will indicate as w_i the fraction of the capital invested on asset i, $i = 1, ..., n$. The investor will then estimate the expected return and variance of each one of the assets μ_i and σ_i^2 as well as the

covariance between each pair of assets σ_{ij}. The information will then be stored in vector-matrix form as $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}_{ii} = \sigma_i^2$ and $\boldsymbol{\Sigma}_{ij} = \sigma_{ij}$ for $i \neq j$.
The portfolio return X will then be the sum of each asset's return r_i, weighted by w_i, with:

Expected return

$$\mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^{n} w_i r_i] = \sum_{i=1}^{n} w_i \mathbf{E}[r_i] = \mathbf{w}'\boldsymbol{\mu}$$

Variance

$$Var(X) = Var(\sum_{i=1}^{n} w_i r_i) = \sum_{i=1}^{n} w_i^2 \boldsymbol{\Sigma}_i^2 + \sum_{i,j=1}^{n} w_i w_j \boldsymbol{\Sigma}_{ij} = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$$

Investors, when choosing the optimal portfolio, must face a trade-off between risk and returns, as there is a relevant positive correlation between these two metrics. This translates into a bi-objective optimization problem that can be formulated as QP (quadratic programming) if the variance $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ is used or SOCP (second-order cone programming) if the volatility $\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}$ is used.
One way to formulate the trade-off problem introduced in [1], that will be used as benchmark in Chapter 6, as

$$\begin{aligned} \max_{\mathbf{w}} \quad & \boldsymbol{\mu}'\mathbf{w} - \frac{\lambda}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0, \end{aligned} \tag{2.1}$$

where λ is a hyper-parameter controlling the risk-aversion of the investor, the first constraint ensures that all capital is invested and we restricted to long-only positions (one cannot short sell assets). The reason for this last modeling choice is in accord with the methods that will be implemented and studied in the next sections, in order to ensure a fair comparison.

2.1.2 Global Minimum Variance Portfolio

The global minimum variance portfolio (GMVP from now on) is a degenerate case of the MVP where $\lambda \to \infty$ and therefore when the investor cares only about minimizing the variance. As mentioned by Herik Hult in Risk and Portfolio Analysis [2], empirical studies suggest that if the estimates of mean and covariance are based only on historical data, then the GMVP is not necessarily a bad choice. Estimators of expected returns, even with a large amount of historical data, are often inaccurate. The GMVP is often a sensible choice when the estimation of $\boldsymbol{\mu}$ is uncertain, which is a key aspect considered in this thesis project. This approach

can therefore be considered a valid benchmark for the methods implemented and analyzed. Moreover, it doesn't require the hyper-parameter λ.
The optimization problem related to it is a convex QP

$$\begin{aligned} \min_{\mathbf{w}} \ & \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \\ \text{s.t.} \ & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0. \end{aligned} \tag{2.2}$$

2.2 Overcoming estimation errors on returns: hierarchical risk allocation and robust optimization

Markowitz's mean-variance portfolio has surely been a milestone in Portfolio Optimization theory. However, while it's great in theory, it poses in practice great problems and dangers so that in real scenarios it tends to perform poorly out of sample.

First of all, as already mentioned in 2.1.2, it has long been recognized that the sample estimation of $\boldsymbol{\mu}$ is extremely imprecise and the optimal solution $\mathbf{w}^*$ is highly sensitive to small changes in the estimated values of the population mean. The estimated covariance still contains errors, but overall its value tends to vary less over time and the degree of error is negligible compared to that of noise contained in the estimation of $\boldsymbol{\mu}$, in fact the authors in [3] found that the errors in expected returns estimates are roughly ten times as important as that in the covariance matrix.

Secondly, the *measure of risk* used in the MVP is the volatility (and seemingly variance) but one can argue that an increase in volatility doesn't always imply a higher risk: this measure penalizes both down-side risk and up-side risk. Considering the distribution of returns of assets, investors surely won't mind an increase in the up-side risk while, following MVP formulation, this will be penalized during the optimization procedure.

It is for these drawbacks that, especially when one has reasons to believe that the current estimates might be noisy (e.g., in a low data regime or if there's the belief that past performances of assets might not be good indicators of future performances) or when the distribution analyzed are not symmetrical (i.e., they are skewed or multi modal) that modern techniques have been developed to overcome such issues. The techniques of robust optimization and allocation through hierarchical clustering algorithms will be theoretically described in a general fashion in this section and they will be further developed in the following chapters.

2.2.1 Robust Optimization

In a general mathematical optimization problem there are two types of problem data: decision variables (here referred as $\boldsymbol{x} \in \mathbb{R}^n$) and problem parameters $\boldsymbol{\theta} \in \mathbb{R}^n$. The problem parameters are given externally as input to the problem and therefore the solution found in terms of decision variables is dependent on them. In MVP optimization, for example, $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
Consider a general uncertain optimization problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x};\boldsymbol{\theta}) \\ \text{s.t.} \quad & g_j(\boldsymbol{x};\boldsymbol{\theta}) \leq 0 \quad j \in \mathcal{J} \\ & \boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^n \end{aligned} \tag{2.3}$$

where $\mathcal{J}$ is the set of constraints and it has been made explicit the dependency of the function on the parameters $\boldsymbol{\theta}$. In practice $\boldsymbol{\theta}$ will be unknown so it has to be estimated as $\hat{\boldsymbol{\theta}}$ and therefore the solution found will be $\boldsymbol{x}^*(\hat{\boldsymbol{\theta}}) \neq \boldsymbol{x}^*(\boldsymbol{\theta})$. There are several techniques that address the issue of finding solutions that are robust to parameters error, the one presented is *worst-case robust optimization*.
In worst-case robust optimization parameters are not single point estimates but rather a set that reflects the uncertainty around their estimation:

$$\boldsymbol{\theta} \in \mathcal{U}_{\boldsymbol{\theta}} \tag{2.4}$$

The choice of the uncertainty set is paramount, as it should express the level of confidence we have in the estimated parameters. Typical choices are:
Box-uncertainty set

$$\mathcal{U}_{\boldsymbol{\theta}} = \{\boldsymbol{\theta} \mid |\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i| \leq \xi_i,\ i = 1, ...n\} \tag{2.5}$$

Elliptical or Quadratic uncertainty set

$$\mathcal{U}_{\boldsymbol{\theta}} = \{\boldsymbol{\theta} \mid (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\boldsymbol{\Omega}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \ \leq k^2\} \tag{2.6}$$

Where $\boldsymbol{\xi}$ and k are the size (or radius) of the uncertainty set and $\boldsymbol{\Omega}$ is the uncertainty matrix which defines the shape of the ellipsoid.
The box uncertainty set, while being the simplest and most intuitive way to express the uncertainty in inputs, it assumes that the estimation error of each parameter can be modeled independently from the others. Each parameter lies around his own confidence interval. On the other hand, by introducing quadratic uncertainty we may derive a set that allows for correlations among the parameters, this happens when $\boldsymbol{\Omega}^{-1}$ is not diagonal. In general, elliptical sets might remind of confidence regions in multivariate statistics, and in fact it can be obtained by assuming that $\boldsymbol{\theta}$ is normally distributed with mean vector $\hat{\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$.

The idea behind worst-case robust optimization is to treat input parameters as variables that needs to be optimized in a first stage of the optimization, aiming to find the worst possible realization within the uncertainty set and with respect to an objective function. Then in the second stage the problem tackled will be the original one fed with the results of the first-stage optimization regarding the input parameters.
We can now formulate the robust counterpart of problem 2.3:

$$\begin{aligned} &\min_{\boldsymbol{x}} \ \{\max_{\boldsymbol{\theta}\in\mathcal{U}_{\boldsymbol{\theta}}} \ f(\boldsymbol{x};\boldsymbol{\theta})\} \\ &\text{s.t.} \quad g_j(\boldsymbol{x};\boldsymbol{\theta}) \leq 0 \quad j \in \mathcal{J}, \ \boldsymbol{\theta} \in \mathcal{U}_{\boldsymbol{\theta}} \\ &\qquad \boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^n \end{aligned} \tag{2.7}$$

The key concept behind this approach is aiming for a solution that will perform reasonably well for every value of the uncertain parameters by taking an adversarial -someone might say pessimistic- view in what nature will select after we made our choice.

When facing robust optimization problems in the field of portfolio optimization, it is common to encounter non linear problems, such as Quadratic Programs (QP). For instance, the approaches presented in previous sections (namely 2.1.1 and 2.1.2) are indeed of that nature. Then it is useful to know that many robust QPs can be reformulated as SOCP problems. For example Goldfarb and Iyengar in [4] prove that the robust version of maximum Sharpe ratio, robust VaR and even the mean-variance optimal portfolio selection problem in 2.1.1 (where mean returns are uncertain, but the covariance is known and fixed), can be reformulated as an SOCP, this will be further developed in Chapter 4.

A general SOCP has the following form[1]:

$$\begin{aligned} &\min_{\boldsymbol{x}} \ \boldsymbol{f}'\boldsymbol{x} \\ &\text{s.t.} \quad ||\mathbf{A}_i\boldsymbol{x} + \boldsymbol{b}_i||_2 \leq \boldsymbol{c}_i^{'}\boldsymbol{x} + d_i \quad i \in \mathcal{I} \end{aligned} \tag{2.8}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the optimization variable, and the problem parameters are $\boldsymbol{f} \in \mathbb{R}^n$, $\mathbf{A}_i \in \mathbb{R}^{(n_i-1)\times n}$, $\boldsymbol{b}_i \in \mathbb{R}^{n_i-1}$, $\boldsymbol{c}_i \in \mathbb{R}^n$ and $d_i \in \mathbb{R}$.
The standard or unit second-order cone of dimension k is defined equivalently in the following ways:

$$\begin{aligned} \mathcal{L}^k &\doteq \left\{ \boldsymbol{z} \in \mathbb{R}^k : \sqrt{\sum_{i=2}^{k} z_i^2} \leq z_1 \right\} \\ \mathcal{L}^k &\doteq \left\{ \begin{pmatrix} t \\ \boldsymbol{u} \end{pmatrix} \ | \ \boldsymbol{u} \in \mathbb{R}^{k-1}, \ t \in \mathbb{R}, \ ||\boldsymbol{u}||_2 \leq t \right\} \end{aligned} \tag{2.9}$$

[1]see [5] for more details about SOCP

The set of points satisfying a second-Order cone constraint is the inverse image of the unit second-Order cone under an affine mapping:

$$||\mathbf{A}_i \boldsymbol{x} + \boldsymbol{b}_i||_2 \leq \boldsymbol{c}_i^{'} \boldsymbol{x} + d_i \iff \begin{pmatrix} \mathbf{A}_i \\ \boldsymbol{c}_i^{'} \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} \boldsymbol{b}_i \\ d_i \end{pmatrix} \in \mathcal{L}^{n_i} \tag{2.10}$$

and hence is convex. Thus, the SOCP is a convex programming problem as the objective function is linear (hence convex) and the constraints define a convex set. This is indeed relevant as convex problem can be efficiently solved.

2.2.2 Clustering Techniques and their application to Portfolio Optimization

These category of methods overcome the criticalities involved in the estimation of $\boldsymbol{\mu}$ by completely avoiding its computation or, as Lopez de Prado said in his paper Building Diversified Portfolios that outperform out-of-sample [6]: “Given that returns can rarely be forecasted with sufficient accuracy, many authors have opted for dropping them altogether and focus on the covariance matrix.”

Hierarchical Clustering portfolios are then built starting from the graph information of the assets (encoded in what is called a distance matrix) and this is why they are also referred to as “Graph-Based Portfolios” [7].
For this reason, Graph-based portfolios start with the computation of the distance matrix, and its choice and estimation is key as the dependency structure in the asset universe will be primarily defined based on this element. There are various methods to acquire and build this graph information.
The simplest way to obtain a distance matrix $\boldsymbol{D}$ is directly by applying an element-wise transformation to the correlation matrix of assets $\boldsymbol{R}$:

$$\boldsymbol{D}_{ij} = \sqrt{\frac{1}{2}(1 - \boldsymbol{R}_{ij})} \tag{2.11}$$

This transformation ensures that $\boldsymbol{D}$ is a proper metric space. A drawback of such a correlation-based distance matrix is that each element is derived by an element-wise transformation, ignoring the rest of the assets. A more holistic definition is to compute a new distance matrix $\tilde{\boldsymbol{D}}$ with elements containing the Euclidean distance between any two column-vectors:

$$\tilde{\boldsymbol{D}}_{ij} = ||\boldsymbol{D}_{\cdot,i} - \boldsymbol{D}_{\cdot,j}||_2 \tag{2.12}$$

This way $\tilde{\boldsymbol{D}}$ is a function of the entire asset universe instead of a single estimate unrelated with the rest of the assets: $\tilde{\boldsymbol{D}}_{ij}$ indicates the closeness of assets i and j with respect to the rest of the universe.
Distance matrices aren’t always built on correlation matrices, in Chapter 3 a method

based on Copulas and Lower Tail Dependency Coefficients will be presented and analyzed.

Once the distance matrix is created, it is used as input in a hierarchical clustering algorithm. This is a statistical data analysis technique employed in the fields of unsupervised machine learning, pattern recognition and of course financial markets. Based on the distance matrix and a way of computing similarities between groups of elements, this technique sequentially clusters items based on the derived distance. Hierarchical Clustering algorithms can be divided in two main paradigms: agglomerative (bottom up) and divisive (top-down). Bottom-up strategies begin with each item alone, as a singleton cluster and progressively merge elements or groups of elements reducing their number step by step until only one remains. In contrast, top-down strategies start with a single cluster containing all elements and recursively split them into smaller clusters. The tree structure obtained is called dendrogram and each level of the hierarchy represents a particular grouping of the data into disjoint clusters of observations.

The process of splitting or merging groups of elements requires a measure of similarity (or dissimilarity) among multiple elements. Different measures can produce different results. Here are reported different ways of computing such distances, also called *linkage type*:

- *single linkage*: the distance between two clusters is the minimum distance between any two points in the clusters, which is related to the so-called minimum spanning tree

- *complete linkage*: the distance between two clusters is the maximum of the distance between any two points in the clusters

- *average linkage*: the distance between two clusters is the average of the distance between any two points in the clusters

- *Ward's method*: the distance between two clusters is the increase of the squared error from when two clusters are merged, which is related to distances between centroids of clusters

The choice of the linkage method produces very different results and will be further analyzed in Chapter 5.

An important feature of hierarchical clustering algorithms, unlike other algorithms analyzed in the next Chapter such as k-means, is that they do not require specifying a fixed number of clusters in advance. One can analyze the resulting dendrogram and devise an appropriate number of clusters based on its shape: grouping data by fewer, bigger, clusters necessarily loses certain fine details, on the other hand representing data with more clusters may lead to overfit or discover spurious patterns that do not really exist. One way to determine an "optimal" number of clusters is by employing the *gap statistics index*[2]:

$$k^* = \underset{k=\ 1,\dots,n}{\arg\max}\, Gap_n(k) = \underset{k=\ 1,\dots,n}{\arg\max}\, \mathbb{E}_n^*[\log(W_k)] - \log(W_k), \tag{2.13}$$

where $W_k = \sum_{r=1}^{k} \left(\frac{1}{2|C_r|} \sum_{i,i' \in C_r} d_{ii'} \right)$ which is an average of intra-cluster distances (distance between elements of the same clusters) and $\mathbb{E}_n^*$ denotes the expected value of $\log(W_k)$ under a sample of size n for uniformly distributed data, a distribution with no obvious clustering (that can be considered as a "null" distribution used for reference).

Hierarchical Clustering-based Portfolios can be then characterized by the following features [7]:

1. *distance matrix*: used to measure the pairwise similarity among elements
2. *linkage type* employed in the hierarchical clustering process (e.g., single, complete, average, or Ward)
3. *clustering stopping criterion*: used to stop the hierarchical clustering process (e.g., all the way down to single-item clusters or early stopping based on the gap statistic)
4. *splitting criterion*: how the assets are recursively split during the allocation phase (i.e. dendrogram based or by simple bisection)
5. *intra-weight allocation*: how weights are computed within clusters
6. *inter-weight allocation*: how weights are computed across clusters

All these characteristics will be further analyzed in Chapter 3, where specific algorithms will be described and implemented. In general, once all parameters above are defined, the main approach is starting from the top of the dendrogram and allocating the capital in a top-down manner. Initially, all assets will have a unit weight of 1, every time a node in the dendrogram is traversed, the weight of the cluster will be split based on the inter-weight allocation parameter of the two sub-clusters until the clustering stopping criterion is met. Then the cluster weight will be allocated to single assets following the intra-weight allocation logic and, finally, normalizing the vector of weights.

[2] Introduced in [8] by Tibshirani et Al.

Chapter 3

Asset allocation with Clustering methods

Section 2.2.2 has been devoted to define a general approach for asset allocation based on hierarchical clustering techniques. This approaches rely on the concept of hierarchical structure found by performing clustering techniques from a similarity matrix. This should help unraveling complex dependency patterns in the data.
This Chapter is divided in three sections. The first two describe important allocation techniques that have been implemented: Hierarchical Risk Parity (HRP), a fundamental contribution to hierarchical allocation methods, and Hierarchical Equal Risk Contribution (HERC), a technique that has been built upon HRP starting from what the author defined as its "remarks".
Both methods, however, are based on a correlation similarity matrix, but this isn't the only possible technique: the author of HERC strategy, at the end of their paper [9], confirm the need of further experiments "Last but not least, this article opens the door for further research. Typical machine learning issues have to be investigated, such as the choice of the distance metric and the criteria used to select the number of clusters."

For this reason, Section 3.3 develops a different similarity metric based on asymptotic dependency. It describes how the matrix is derived and estimated and integrates such feature in a general framework of hierarchical clustering allocation based on HERC approach.

3.1 Hierarchical Risk Parity

Hierarchical Risk Parity approach has been introduced by Lopez de Prado in 2016 [6] to specifically address three major concerns of the problems introduced in Section 2.1.1, that are: instability, concentration and underperformance. These

issues has been already described in Section 2.2.
HRP uses modern mathematical fields (graph theory and machine learning techniques) to achieve diversification in assets porfolios. As highlighted in Section 2.2, hierarchical approaches don't require the estimation of expected returns. The only inputs needed are therefore the covariance matrix (for weight allocation) and the correlation matrix (for clustering).
HRP allocation is performed in three stages, where the first two serve as input for the final one which is the actual algorithm that creates the portfolio weights.

Tree Clustering

This stage involves creating a hierarchical tree from the correlation-based distance matrix 2.12. This allows building a hierarchical structure based on the relative distance (in terms of correlation of returns) between two assets with respect to the universe.
Based on such matrix, an agglomerative clustering based on *single linkage* criterion is performed. This type of linkage is strictly related to the so-called minimum spanning tree.
The result of the clustering algorithm is a dendrogram -an example can be found in Figure 3.1- that contains the hierarchical dependency structure of the asset universe. The output that will be used in the following step is a linkage matrix, a structured version encapsulating the dendrogram structure.

Quasi-diagonalization

This stage reorganizes the rows of the covariance matrix so that "closer" assets in terms of correlation are placed close to each other. This means that similar investment opportunities will be closer together than dissimilar ones. As shown in Figure 3.2.

Recursive Bisection

This last step performs asset allocation based on the derived quasi-diagonal covariance matrix. It splits recursively the assets in half based on the quasi-diagonalized matrix order and assigns each part a weight based on the inverse-variance allocation. Such allocation choice is justified as the IVarP weights are optimal (in terms of the GMVP described in 2.1.2) for a diagonal covariance matrix. Correlation

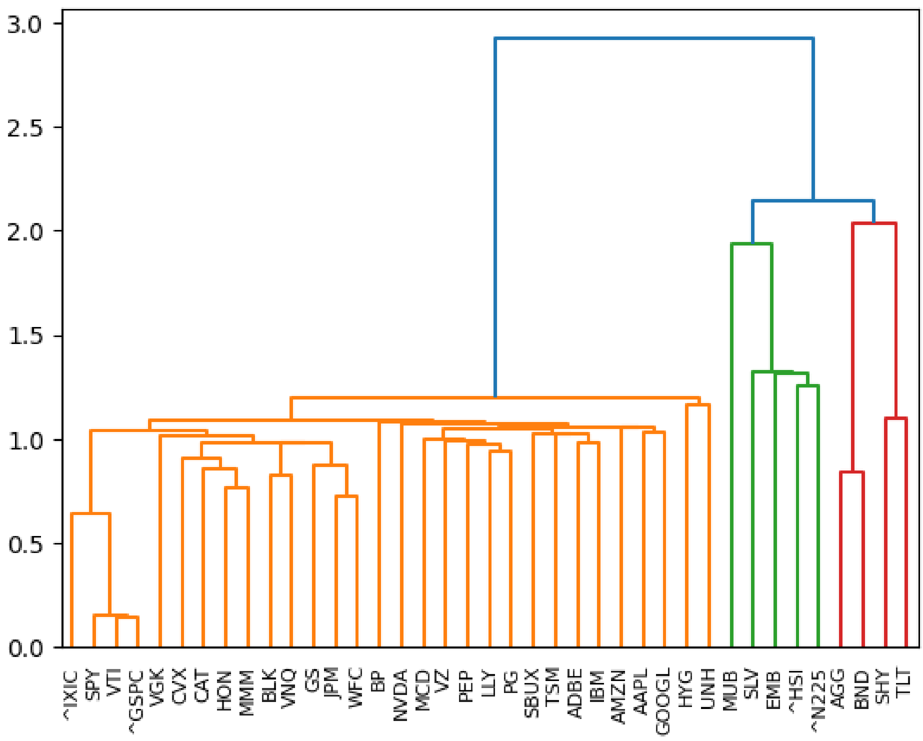
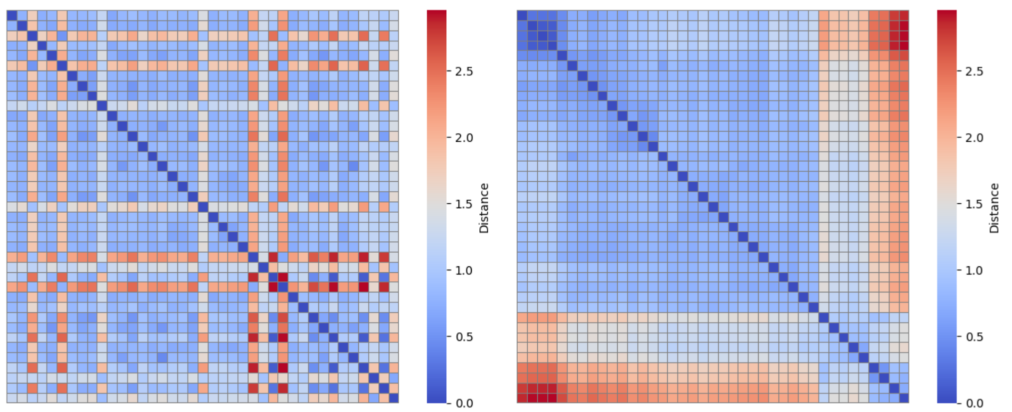


Figure 3.1: Dendrogram of a subset of multi-asset dataset, described in 5.2.1



(a) Distance matrix before quasi-diagonalization **(b)** Distance matrix after quasi-diagonalization

Figure 3.2: Quasi-diagonalization applied to a subset of the multi-asset dataset

between subset is therefore ignored as they should be low correlated thanks to the quasi-diagonalization applied before. The algorithm can be formalized as in

Algorithm 1.

Algorithm 1 HRP Algorithm

Initialize:
Set the list of items: $\mathcal{L} = \{\mathcal{L}_0\}$, with $\mathcal{L}_0 = \{n\}_{n=1}^{N}$ in the same order of the quasi-diagonal covariance matrix
Assign unit weights: $w_n = 1,\ \forall n = 1, \dots, N$
while $\exists \mathcal{L}_i \in \mathcal{L}$ s.t. $|\mathcal{L}_i| > 1$ **do**
  for all $\mathcal{L}_i \in \mathcal{L}$ s.t.$|\mathcal{L}_i| > 1$ **do**
    Bisect $\mathcal{L}_i$ into two subsets $\mathcal{L}_i^{(1)}$ and $\mathcal{L}_i^{(2)}$, such that:
      • $\mathcal{L}_i^{(1)} \cup \mathcal{L}_i^{(2)} = \mathcal{L}_i$
      • $|\mathcal{L}_i^{(1)}| = \left\lfloor \frac{1}{2} |\mathcal{L}_i| \right\rfloor$
      • Order is preserved
    for $j = 1, 2$ **do**
      Let $V_i^{(j)}$ be the covariance matrix of $\mathcal{L}_i^{(j)}$
      Compute weights: $\tilde{w}_i^{(j)} = \frac{\operatorname{diag}\left[V_i^{(j)}\right]^{-1} \cdot \mathbf{1}}{\operatorname{tr}\left[\operatorname{diag}\left[V_i^{(j)}\right]^{-1}\right]}$
      Compute variance: $\tilde{V}_i^{(j)} = \tilde{w}_i^{(j)\prime} V_i^{(j)} \tilde{w}_i^{(j)}$
    end for
    Compute split factor:
    $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}}$ where $0 \leq \alpha_i \leq 1$
    Update $\mathcal{L}$ by replacing $\mathcal{L}_i$ with $\mathcal{L}_i^{(1)}$ and $\mathcal{L}_i^{(2)}$
    Re-scale allocations: $w_n \leftarrow \alpha_i \cdot w_n,\ \forall n \in \mathcal{L}_i^{(1)}$
    Re-scale allocations: $w_n \leftarrow (1 - \alpha_i) \cdot w_n,\ \forall n \in \mathcal{L}_i^{(2)}$
  end for
end while
return $\{w_n\}_{n=1}^{N}$

Following the characterization presented in 2.2.2, Hierarchical Risk Parity allocation can be summarized as:

1. *distance matrix*: correlation-based, pairwise similarity with respect to the rest of the assets
2. *linkage type*: single linkage
3. *clustering stopping criterion*: all the way down to the leaves of the dendrogram
4. *splitting criterion*: simple bisection

5. *intra-weight allocation*: inverse-variance allocation

6. *inter-weight allocation*: inverse-variance allocation with N=2, therefore not considering correlation

3.2 Hierarchical Equal Risk Contribution

The Hierarchical Equal Risk Contribution allocation strategy builds on the fundamental notion of hierarchical allocation introduced by the Hierarchical Risk Parity approach but tries to improve several criticalities that have been highlighted by Thomas Raffinot in [9], an improvement of his previous work where he developed the "Hierarchical Clustering based Asset Allocation" (HCAA) [10].
The first remark comes from the usage of inverse-variance allocation assuming that clusters are uncorrelated, as in this special case the inverse-variance portfolio is the optimal solution of the minimum variance problem. "From a theoretical point of view, it seems more appropriate to make no assumptions on the correlation between clusters, thereby expressing the bisection in terms of ERC rather than in terms of minimum-variance between uncorrelated assets."
The second remark is related with the linkage criterion employed, single linkage has a strong connection with graph theory and the minimum spanning tree but it suffers from chaining: the distance between two clusters is equivalent to the closest pair of assets in each of them, irrespective to the rest. As we will see in Chapter 5, this may produce imbalanced clusters that can be too spread out.
Third, the splitting criterion used is simple bisection, which means that the derived dendrogram is only used for reordering the assets but its actual shape is not considered in the allocation phase, potentially splitting two assets close together because they are in the middle point, while from a logical point of view this is in contrast with the results obtained from clustering the assets, as showed in Figure 3.3.

The last remark comes from the clustering stopping criterion, as HRP performs all the way top-down allocation and doesn't consider the relevant number of clusters: "It may happen that you use more information than the clustering can provide with reliability. Intuitively, it can be seen as a form of overfitting, leading to potential bad results" [9].

Starting from these remarks, HERC algorithm performs allocation in four steps:

1. Hierarchical Clustering

2. Selection of the optimal number of clusters, based on the Gap index [8]

3. Top-Down recursive division into two parts based on the dendrogram and following an Equal Risk Contribution allocation, e.g. the weights are $\tilde{w}_1 = \frac{\mathcal{RC}_1}{\mathcal{RC}_1+\mathcal{RC}_2}$ and $\tilde{w}_2 = 1 - \tilde{w}_1$. Where $\mathcal{RC}_i$ is the risk contribution of cluster

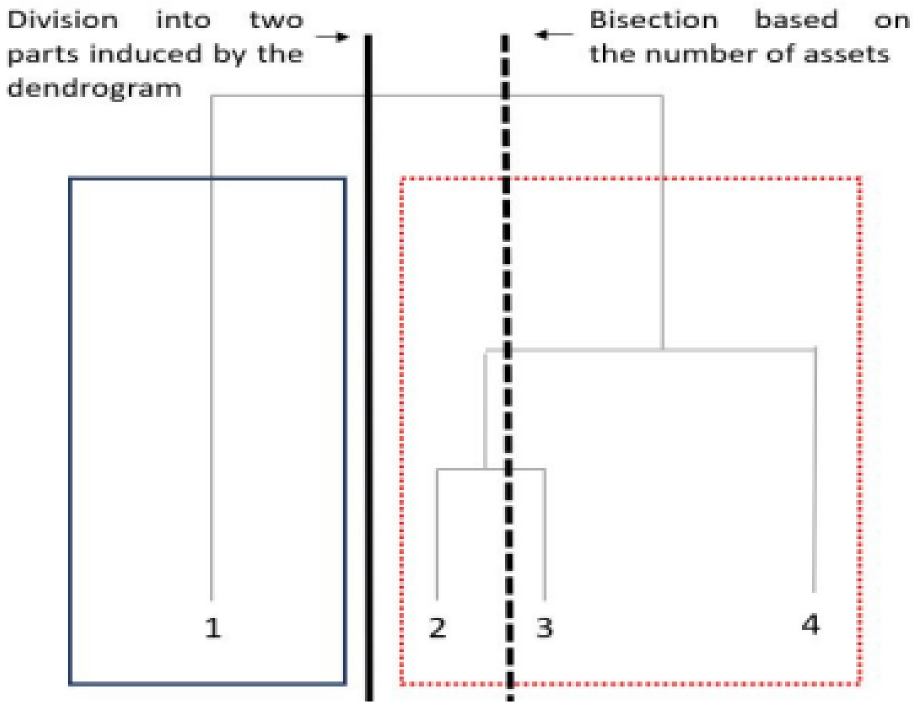


Figure 3.3: Bisection based on the number of assets and bisection induced by the dendrogram. Source: Exhibit 3 in [9]

i (which can be variance or other downside risk measures such as CVar or CDaR)

4. Naive Risk Parity within clusters (within the same cluster the correlation between assets should be elevated)

Concerning the risk measure $\mathcal{RC}_i$, this method allows for a wide variety of possibilities. As Raffinot mentions, variance is the common choice. The main drawback of it is that it it not a good measure for the risk of assets presenting significant tail risk (i.e. heavy tailed distributions). For this reason, HERC is extended to common downside risk measures such as conditional value at risk (CVaR) and Conditional Drawdown at Risk (CDaR) [1].

CVaR at level α is a risk measure that is obtained by taking the expected value of the losses in the left tail distribution (if we consider for example returns distribution) for values exceeding the VaR at level α, ie the $1 - \alpha$ quantile of the distribution in this case.

[1]Both measures applied to Portfolio Optimization are descibed in depth in [11]

CDaR can be seen as a particular case of CVaR where in this case the loss function is defined as the drawdown, i.e. the cumulative loss between a peak and another. In this case the risk measure is the average of all drawdowns exceeding a given threshold.

3.3 Hierarchical Allocation with Lower Tail Dependency Coefficient

The methods that have been presented in the previous sections aim at finding portfolio featuring a high level of diversification enhanced by the dependency structure derived from the hierarchical clustering algorithms. This dependency structure is highly sensitive to the distance matrix given as input, this matrix in the previous methods has always been correlation based.
As mentioned by Lohre et Al. in [12], “correlations can increase significantly during financial crises because of contagion effects [...], hence diversification by correlation clusters may fail to work when it is needed most”.
For this reason one might cluster assets based on their co-movements when large losses occur. This approach has been first proposed by De Luca and Zuccolotto [13].
Before diving into the implementation details, a section presenting concepts that will be employed is given.

3.3.1 Copulas and LTDC estimators

The similarity measure employed in this section will be based on the concept of Lower Tail Dependency Coefficient (LTDC), therefore a formal definition is provided.

Definition 3.3.1 *Let X and Y be two random variables and let $F_X(x) = \mathbb{P}(X \leq x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$ be their distribution functions. The lower tail dependence coefficient (LTDC) is defined as*

$$\lambda_L = \lim_{t \to 0^+} \mathbb{P}(X \leq F_X^{-1}(t) | Y \leq F_Y^{-1}(t)) \tag{3.1}$$

The LTDC is then defined for a pair of variables that we call *asymptotically independent* if $\lambda_L = 0$ and and *asymptotically dependent* if $0 < \lambda_L \leq 1$.
The LTDC strongly depends on the joint bivariate distribution that models the pair of variables and therefore choosing the right CDF is paramount. When dealing with a large number of variables, estimating each pair-wise joint distribution becomes hard in practice.

Copulas functions are widely employed in the field of finance and risk management as they allow to split the modeling of a joint distribution in two distinct steps. First one can estimate a univariate marginal distribution independent on the other variables, then a copula model is chosen in order to inject a specific dependency structure as a function of the obtained marginals. A copula C is a distribution function of a random vector $\mathbf{U}$ whose components are uniformly distributed in (0, 1)

$$\begin{aligned} C(u_1, ..., u_d) =& \mathbb{P}(U_1 \leq u_1, ..., U_d \leq u_d) \\ =& \mathbb{P}(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ =& C(F_1(x_1), ..., F_d(x_d)) \end{aligned} \tag{3.2}$$

When dealing with LTDC only, this proves to be beneficial since

$$\begin{aligned} \lambda_L &= \lim_{u \to 0^+} \mathbb{P}(X \leq F_X^{-1}(t) | Y \leq F_Y^{-1}(t)) \\ &= \lim_{u \to 0^+} \frac{\mathbb{P}(F_X(X) \leq t, F_Y(Y) \leq t)}{\mathbb{P}(F_Y(Y) \leq t)} \\ &= \lim_{u \to 0^+} \frac{\mathbb{P}(U_X \leq t, U_Y \leq t)}{\mathbb{P}(U_Y \leq t)} \\ &= \lim_{u \to 0^+} \frac{C(t,t)}{t} \end{aligned} \tag{3.3}$$

Hence the LTDC depends only on the copula function C and not on the marginals. λ_L can be then estimated in a parametric way, where a specific copula structure is assumed, its parameters are estimated (through maximum likelihood for example) and the LTDC is derived as a function of the estimated parameters (see De Luca and Zuccolotto [13]). Alternatively, a non-parametric estimation as proposed in [12] can be employed. This approach has the advantage of being agnostic to assumptions of the underlying copula, which is suitable when dealing with a large number of assets (as it is required to estimate $\frac{n(n-1)}{2}$ values and therefore devise that many dependency structures). For this reason, two non-parametric estimators will be employed in this project.

Standard non-parametric estimator Given a random sample $\mathbf{X} \in \mathbb{R}^{dxn}$ with n observations $j = 1, ..., n$ of a d dimensional vector $(i = 1, ..., d)$ X with joint distribution function F and copula C, the standard nonparametric estimator for a pair of variables $(X_i, X_{i'})$ of their LTDC is defined in [14] by

$$\hat{\lambda}_{L,n}(\frac{k}{n}) = \frac{\hat{C}_n(\frac{k}{n}, \frac{k}{n})}{\frac{k}{n}} \tag{3.4}$$

This is indeed obtained directly by substituting in 3.3 Copula C with its empirical counterpart, that can be estimated by

$$\hat{C}_n(u,v) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{\hat{U}_{1j,n} \le u\}} \mathbb{1}_{\{\hat{U}_{2j,n} \le v\}}, \quad (u,v) \in [0,1]^2 \tag{3.5}$$

Where $\hat{U}_{ij,n}$ is obtained through the empirical estimators of the marginal CDF

$$\hat{F}_{i,n}(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_{ij} \le x\}}, \quad i = 1, ..., d \text{ and } x \in \mathbb{R} \tag{3.6}$$

$$\hat{U}_{ij,n} = \hat{F}_{i,n}(X_{ij}) = \frac{1}{n}(\text{rank of } X_{ij} \text{ in } X_{i1}, ..., X_{in}) \tag{3.7}$$

The other clear difference between equation 3.3 and 3.4 is the presence of a parameter $k \in [1, ..., n]$. When $k/n \to 0$ we retrieve the original definition of LTDC. As stated by Garcin et Al in [14], in an empirical setting choosing a too low value for k would lead to a non-robust estimator while a higher value would depict some properties of the copula which are not specifically the ones of its lower tail. To this end, we have to find a trade-off between variance and bias when estimating this parameter.

LTDC based on a conditional version of Spearman's rho[2]

The multivariate Conditional version of Spearman's Rho (CSR) at level p is given by

$$\rho(p) = \frac{\int_{[0,p]^d} C(u)du - (\frac{p^2}{2})^d}{\frac{p^{d+1}}{d+1} - (\frac{p^2}{2})^d}, \quad 0 < p \le 1 \tag{3.8}$$

Schmid and Schmidt (2007) in [15] then define a multivariate LTDC ρ_L, a generalization of the bivariate LTDC λ_L, by

$$\rho_L = \lim_{p \to 0^+} \frac{d+1}{p^{d+1}} \int_{[0,p]^d} C(u)du \tag{3.9}$$

Following a non-parametric approach, the Copula C and the underlying marginals distributions are assumed unknown and substituted by their empirical counterparts in 3.6.

Equation 3.5 can easily be generalized for a d dimensional vector as

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d} \mathbb{1}_{\{\hat{U}_{ij,n} \le u_i\}}, \quad \mathbf{u} = (u_1, ..., u_d)' \in [0,1]^d \tag{3.10}$$

[2]This estimator has been used in [12]

This leads to the following non-parametric estimator for 3.8

$$\hat{\rho}_n(p) = \frac{\frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}(p-\hat{U}_{ij,n})^{+} - (\frac{p^2}{2})^d}{\frac{p^{d+1}}{d+1} - (\frac{p^2}{2})^d} \tag{3.11}$$

Finally, the estimator for 3.9 is

$$\rho_L = \hat{\rho}_n(\frac{k}{n}), \quad k = \{1, ..., n\} \tag{3.12}$$

where k corresponds to the same parameter as in 3.4.
In general, we will refer to the estimator in the above equation as $\hat{\lambda}_{L,n}(X_i, X_j)$ when treating a pair of random variables (i.e. d=2) at a fixed value of k.

3.3.2 A general framework for LTDC-based allocation

The LTDC-based approach proposed here can be seen as a general framework of hierarchical clustering allocation based on a distance matrix that is obtained with a proper transformation of the LTDC matrix $\mathbf{L}$, defined as $L_{ii} = 1,\ i = 1, ..., n$ and $L_{ij} = \hat{\lambda}_{L,n}(X_i, X_j)$ if $i \neq j$.
Once matrix $\mathbf{L}$ has been estimated, it is transformed into a dissimilarity matrix as suggested by De Luca and Zuccolotto in [13].

Proposition 3.3.1 *Let X_i and X_j be two random variables and $\hat{\lambda}_{L,n}(X_i, X_j)$ their estimated lower tail dependence coefficient, then the following function $d : \mathbb{R} \times \mathbb{R} \to [0, \infty)$, is a dissimilarity measure:*

$$d(X_i, X_j) = -\log(\hat{\lambda}_{L,n}(X_i, X_j)) \tag{3.13}$$

Before estimating the asymptotic correlation coefficients, Lohre et Al in [12] suggest to apply a univariate Student-t AR(1)-GARCH(1,1) model to every time series in order to remove autocorrelation and heteroskedasticity. This modeling step has been implemented and simulations will be performed both with and without applying the model to the returns series before estimating the LTDC coefficients.

When the similarity matrix is estimated, the algorithm can continue as described in 2.2.2. In this project the impact of the allocation parameters will be studied, so what has been described in this Chapter is a general paradigm that extends both HRP and HERC logic to advanced similarity measures and encompasses both approaches adopted by the previously described methods.

Chapter 4

Robust Portfolio Optimization

This Chapter will present the specific Robust Mean Variance Optimization problem (RO-MVP). In 2.2.1 the general Robust Optimization approach has been presented, as well as the general techniques associated.
The problem addressed will be formulated in the next section: it will encompass the problem definition and the uncertainty set employed. Section 4.2 will show how we can reformulate RO-MVP into a problem that can be efficiently solved by common solvers. Finally, the last section will be dedicated to giving practical estimates for the hyper-parameters involved in the optimization.

4.1 Problem Definition

We recall the standard MVP optimization problem with long-only positions introduced in Section 2.1.1

$$\begin{aligned} \max_{\mathbf{w}} \ & \hat{\boldsymbol{\mu}}'\mathbf{w} - \frac{\lambda}{2}\mathbf{w}'\hat{\boldsymbol{\Sigma}}\mathbf{w} \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \end{aligned} \tag{4.1}$$

Where it has specifically been made explicit the estimated returns vector $\hat{\boldsymbol{\mu}}$ as well as the estimated covariance $\hat{\boldsymbol{\Sigma}}$. For the reasons presented in Section 2.2, the parameters considered uncertain are the expected returns $\boldsymbol{\mu}$ while the covariance of assets is considered known. This modeling choice is consistent with common practices in the relevant literature.

The robust counterpart of problem 2.1 can be formulated as

$$\begin{aligned} \max_{\mathbf{w}} & \left\{ \min_{\boldsymbol{\mu} \in \mathcal{U}_\mu} \{\boldsymbol{\mu}'\mathbf{w}\} - \frac{\lambda}{2}\mathbf{w}'\hat{\boldsymbol{\Sigma}}\mathbf{w} \right\} \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \end{aligned} \tag{4.2}$$

Here the expected returns are treated as a parameter that lies in the uncertainty set $\mathcal{U}_\mu$, that can be defined in different ways. In this project the quadratic uncertainty set will be employed: on top of the reasons expressed in 2.2.1, there are other motivation coming from the literature[1]:

1. Ben-Tal and Nemirovski (1998) review both box and quadratic uncertainty sets and argue that the latter leads to a tractable robust counterpart of the convex optimization problem while the robust counterpart induced by box uncertainty set is only tractable in linear programming.

2. Pachamanova and Fabozzi (2016) in [17] point out that the box uncertainty set assumes that all assets will achieve their worst-case return at the same time and this assumption is not verified in practice. They suggest that it may be more practical to assume that not all assets attain their worst-case returns at the same time and more informative to take into account the variancecovariance structure of the expected returns as formulated with a quadratic uncertainty set.

3. Goldfarb and Iyengar (2003) in [4] demonstrated analytically that the quadratic uncertainty set is generated naturally from the estimation process using regression when the expected returns are estimated with a linear factor model, which is quite common in the finance industry.

For these reasons, the final robust optimization problem can be formulated in the following way:

$$\begin{aligned} \max_{\mathbf{w}} & \left\{ \min_{\boldsymbol{\mu}} \{\boldsymbol{\mu}'\mathbf{w}\} - \frac{\lambda}{2}\mathbf{w}'\hat{\boldsymbol{\Sigma}}\mathbf{w} \right\} \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \\ & (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \ \leq k^2 \end{aligned} \tag{4.3}$$

[1]All this remarks are summarized in [16]

4.2 Problem Reformulation

As it is common in Robust Optimization, problem 4.3 is a max-min optimization problem. It can therefore be solved in two steps.
The first step involves solving the inner problem and finding the worst-case returns lying in the uncertainty set. This problem is easily tackled by using KKT conditions and considering fixed the vector of weights (the variable associated with the outer problem). Using the Lagrangian

$$\mathcal{L}(\boldsymbol{\mu},\gamma) = \boldsymbol{\mu}'\mathbf{w} + \gamma\left[(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})'\boldsymbol{\Omega}_{\mu}^{-1}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}) - k^2\right] \tag{4.4}$$

Where $\gamma \geq 0$ is the multiplier associated with the uncertainty set constraint. As the feasible set is convex and the objective function linear in $\boldsymbol{\mu}$, at the optimum the inequality constraint will be satisfied with an equality. The stationarity conditions yield

$$\begin{aligned}
&\begin{cases}
\dfrac{d\,\mathcal{L}(\boldsymbol{\mu},\gamma)}{d\,\boldsymbol{\mu}} = \boldsymbol{w} + 2\gamma\,\boldsymbol{\Omega}_{\mu}^{-1}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}) = \mathbf{0} \\
\dfrac{d\,\mathcal{L}(\boldsymbol{\mu},\gamma)}{d\,\gamma} = (\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^{\top}\boldsymbol{\Omega}_{\mu}^{-1}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}) - k^2 = 0
\end{cases} \\
\Rightarrow\quad &\begin{cases}
\boldsymbol{\mu}-\hat{\boldsymbol{\mu}} = -\dfrac{1}{2\gamma}\,\boldsymbol{\Omega}_{\mu}\boldsymbol{w} \\
\left(-\dfrac{1}{2\gamma}\,\boldsymbol{\Omega}_{\mu}\boldsymbol{w}\right)^{\top}\boldsymbol{\Omega}_{\mu}^{-1}\left(-\dfrac{1}{2\gamma}\,\boldsymbol{\Omega}_{\mu}\boldsymbol{w}\right) = k^2
\end{cases}
\end{aligned} \tag{4.5}$$

Finally, this leads to

$$\frac{1}{2\gamma} = \frac{k}{\sqrt{\boldsymbol{w}'\boldsymbol{\Omega}_{\mu}\boldsymbol{w}}} \Rightarrow \boldsymbol{\mu}_{\text{opt}} = \hat{\boldsymbol{\mu}} - \frac{k}{\sqrt{\boldsymbol{w}'\boldsymbol{\Omega}_{\mu}\boldsymbol{w}}}\Omega_{\mu}\boldsymbol{w} \tag{4.6}$$

The overall problem becomes a convex optimization problem

$$\begin{aligned}
\max_{\mathbf{w}} \quad & \left\{\hat{\boldsymbol{\mu}}'\mathbf{w} - k\sqrt{\mathbf{w}'\boldsymbol{\Omega}_{\mu}\mathbf{w}} - \frac{\lambda}{2}\mathbf{w}'\hat{\boldsymbol{\Sigma}}\mathbf{w}\right\} \\
\text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\
& \mathbf{w} \geq 0
\end{aligned} \tag{4.7}$$

If $\boldsymbol{\Omega}_{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}} \in \mathbb{S}^n_{++}$, then they can be expressed as $\boldsymbol{\Omega}_{\boldsymbol{\mu}} = O'O$, $\hat{\boldsymbol{\Sigma}} = S'S$ by applying Cholesky decomposition, and the above problem becomes (switching to minimization)

$$\begin{aligned} \min_{\mathbf{w}} \; & -\hat{\boldsymbol{\mu}}'\mathbf{w} + k\,||O\mathbf{w}||_2 + \frac{\lambda}{2}\mathbf{w}'S'S\mathbf{w} \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \end{aligned} \tag{4.8}$$

This problem, as stated in 2.2.1, can be recast in SOCP form by introducing the auxiliary variables δ, ν and moving the SOC terms to the constraints

$$\begin{aligned} \min_{\mathbf{w},t,r} \; & -\hat{\boldsymbol{\mu}}'\mathbf{w} + k\,\nu + \frac{\lambda}{2}\delta \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \\ & ||O\mathbf{w}||_2 \leq \nu \\ & \sqrt{(1-\delta)^2 + (2S\mathbf{w})'(2S\mathbf{w})} \leq 1 + \delta \end{aligned} \tag{4.9}$$

Finally, it becomes

$$\begin{aligned} \min_{\mathbf{w},\nu,\delta} \quad & -\hat{\boldsymbol{\mu}}'\mathbf{w} + k\nu + \frac{\lambda}{2}\delta \\ \text{s.t.} \quad & \mathbf{1}'\mathbf{w} = 1 \\ & \mathbf{w} \geq 0 \\ & \begin{pmatrix} \nu \\ O\mathbf{w} \end{pmatrix} \in \mathcal{L}^{n+1} \\ & \begin{pmatrix} 1+\delta \\ 1-\delta \\ 2S\mathbf{w} \end{pmatrix} \in \mathcal{L}^{n+2} \end{aligned} \tag{4.10}$$

Where $\mathcal{L}^k$ denotes the k-dimensional second-order cone defined in 2.9.
This reformulation is in standard SOCP form, thus it can be solved very efficiently using interior point algorithms and it's handled directly by modern software such as CVXPY with Gurobi, the one used in this project.

4.3 Hyperparametrs estimation

Problem 4.10 introduces a practical problem: how do we characterize the uncertainty set numerically? That is, how can we determine a suitable value for k and what is the form the uncertainty matrix should have?
While λ should be derived from the optimizer risk aversion and therefore it has a "practical" meaning, choosing appropriate values for k and $\boldsymbol{\Omega}_{\boldsymbol{\mu}}$ is not as trivial.

Concerning the form of $\boldsymbol{\Omega_\mu}$, a simple choice could be setting $\boldsymbol{\Omega_\mu}$ as $\hat{\boldsymbol{\Sigma}}$, but there are several other possibilities that have been employed in the literature, namely:

1. $\boldsymbol{\Omega_\mu} = \text{diag}(\hat{\boldsymbol{\Sigma}})$
2. $\boldsymbol{\Omega_\mu} = \left(\text{diag}(\hat{\boldsymbol{\Sigma}})\right)^{\frac{1}{2}}$
3. $\boldsymbol{\Omega_\mu} = \mathbb{I}_n$

There isn't a unified view on what is the right choice, therefore in this project all four possibilities will be taken into account and the results will be analyzed in Chapter 6.
For example, Yin et Al in [16] present two criteria the uncertainty matrix should provide:

1. The ideal uncertainty matrix is expected to reduce the sensitivity to inputs by shrinking the original correlation coefficients towards zero
2. The ideal uncertainty matrix should keep the original volatilities unchanged.

Based on these desirable qualities, they prefer the diagonal matrix with variances on the main diagonal. A general overview is displayed in Table 4.1.

	Reducing Sensitivity	Preserving Volatilities
Case 1: $\boldsymbol{\Omega_\mu} = \hat{\boldsymbol{\Sigma}}$	No	Yes
Case 2: $\boldsymbol{\Omega_\mu} = \textbf{diag}(\hat{\boldsymbol{\Sigma}})$	Yes	Yes
Case 3: $\boldsymbol{\Omega_\mu} = \mathbb{I}_n$	Yes	No
Case 4: $\boldsymbol{\Omega_\mu} = \left(\textbf{diag}(\hat{\boldsymbol{\Sigma}})\right)^{\frac{1}{2}}$	Yes	No

Table 4.1: Comparison of different choices for $\boldsymbol{\Omega_\mu}$. Source: Table 1 in [16].

As the possible candidates for the uncertainty matrix have been presented, this leaves us in determining a suitable choice for the value of k. It represents the level of uncertainty to which we restrict our uncertainty set.
Two main approaches have been studied for determining this parameter: a statistical point of view and an optimization-focused one.

As mentioned in Section 2.2.1, quadratic uncertainty set has a strong mathematical connection with confidence regions in multivariate statistics. For this reason Fabozzi et Al in [17] propose that k represents the level of the scaled deviation of realized returns from the forecasts against which one wish to be protected.

In the case of expected asset returns being normally distributed, we can indicate as ω_q the level of confidence at which we would like to be sure that the true expected returns will fall in the ellipsoidal set around the estimates $\hat{\boldsymbol{\mu}}$. Then we can assign as k^2 the value of the ω_q-th quantile of a χ^2 distribution with degrees of freedom equal to the number of assets in the portfolio, that is

$$k^2 = \chi^2_n(\omega_q) \tag{4.11}$$

Normality assumption may not hold in practice due to heavy-tails or skewness in the returns distribution. If the expected returns are assumed to belong to any possible probability distribution, we can guarantee that the estimates will fall in the uncertainty set with probability at least ω_e by assigning

$$k^2 = \frac{1-\omega_e}{\omega_e} \tag{4.12}$$

Notice the difference in meaning between the two ωs: while the first indicates the level of confidence at which the true estimate will fall in the uncertainty set, the second one (ω_e) represent the error we would like to bound our estimates to.

On the other hand, Yin et Al in [16] argue that k is a key parameter in the optimization procedure and it should satisfy the first order condition at optimum. They find that k acts as a regularization term that shifts a modified version of the correlation matrix towards the identity matrix. This is relevant as Mean-Variance optimization can be reformulated in terms of estimated Sharpe ratios and correlation matrix, and the optimal solution requires the inversion of such matrix. The higher the value of k, the more the modified correlation matrix shifts towards the identity matrix and thus becomes less prone to instability issues caused by small eigenvalues. At the same time, a high value of k may distort correlations in a way that the resulting dependency structure is too far from reality. They propose a starting point for the calibration of k as half of the average of estimated Sharpe ratios that, empirically, has produced a fair trade-off between stability and loss of information. In this case

$$k = \frac{1}{2n}\sum_{i=1}^{n} \hat{SR}_i = \frac{1}{2n}\sum_{i=1}^{n} \frac{\hat{\mu}_i - r_f}{\hat{\sigma}_i} \tag{4.13}$$

Where $\hat{\mu}_i$ is the estimated expected return for asset i, $\hat{\sigma}_i$ its estimated volatility and r_f the risk-free rate.

Chapter 5

Clustering analysis of assets

In theoretical studies on portfolio management and optimization, attention is often focused on how to assign weights to a set of predefined investment opportunities. When facing real investment decisions, however, a key step is deciding which assets will take part in the optimization process: as presented by Tang in [18], even though more than half a century of studies on the matter have not reached any consensus as to what the optimal number of stocks should be, many works suggest that the benefit of diversification is mostly exhausted with a portfolio size of 10 – 15 assets. It is therefore of fundamental impact the preliminary process of selecting a subset of the available investment universe that will constitute the investment portfolio. Since this thesis work was developed in a corporate setting, this initial step will be taken into account and studied in detail.
This Chapter will be dedicated to the development of clustering algorithms for financial assets, serving a dual purpose: it enables advanced data mining and knowledge extraction from a complex environment such as the financial markets and useful for many different applications (as it will be mentioned in next sections), at the same time it may be useful in defining the assets to which the optimization algorithms will be applied.

5.1 Clustering strategies

Many different clustering algorithms have been developed in the literature of financial clustering: from classical approaches such as K-means to more "exotic" ones like SOM (Self-Organizing Maps).
This study will focus on two main methods: K-means and Hierarchical clustering. The motivations behind this choice are multiple and will be discussed in the following sections, but on top of everything there is the company need of producing a flexible framework that will help analyzing investment universes of many different types. Starting from this requirement, the methods developed will serve as a data analytics tool and therefore they will take into consideration many different contexts and objectives.

5.1.1 K-means

K-means is one of the most popular clustering approaches and it is widely employed especially in the financial domain. One of the main reasons for its popularity is that the grouping process is intuitive and easy for non-technical individuals to understand, therefore this method has good explainability and can lead to valuable insights. These characteristics are less effective in higher dimensionality as our ability to visualize results is limited to a small number of dimensions (3 or 4 at most): choosing fewer features or effective representations can lead to meaningful results, nevertheless more indicators might provide additional information valuable for the algorithm to split assets in distinct groups.
Suppose we have a collection of N samples $\boldsymbol{X} \in \mathbb{R}^{N \times P}$, each of dimension P. We wish to find K vectors points that can well represent our dataset $\boldsymbol{X}$. The idea then is to find a set of vectors $\{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K\} \in \mathbb{R}^P$ representing the clusters centers and assign each datapoint to one and only one cluster, such that the inter-point distances are small compared with the distances to points outside of the cluster.
The core problem of K-means algorithm as described above is described with the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K} \; & J \doteq \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \; \mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}_k) \\ \text{s.t.} \quad & \sum_{k=1}^{K} r_{n,k} = 1 \quad \forall n \end{aligned} \tag{5.1}$$

where commonly $\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}_k) = ||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2$.
In practice, K-means clustering is performed through an iterative algorithm. First of all, the set of the cluster centers $\{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K\}$ is randomly initialized. Then, until no clear improvements between subsequent steps are registered, an Expectation-Maximization algorithm is performed:

1. Minimization of J wrt $r_{n,k}$ - *Expectation* step

$$r_{n,k} = \begin{cases} 1 \text{ if } k = arg\min_k \mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}_k) \\ 0 \text{ otherwise} \end{cases} \tag{5.2}$$

2. Minimization of J wrt $\boldsymbol{\mu}_k$ - *Maximization* step
   If $\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}_k) = ||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2$

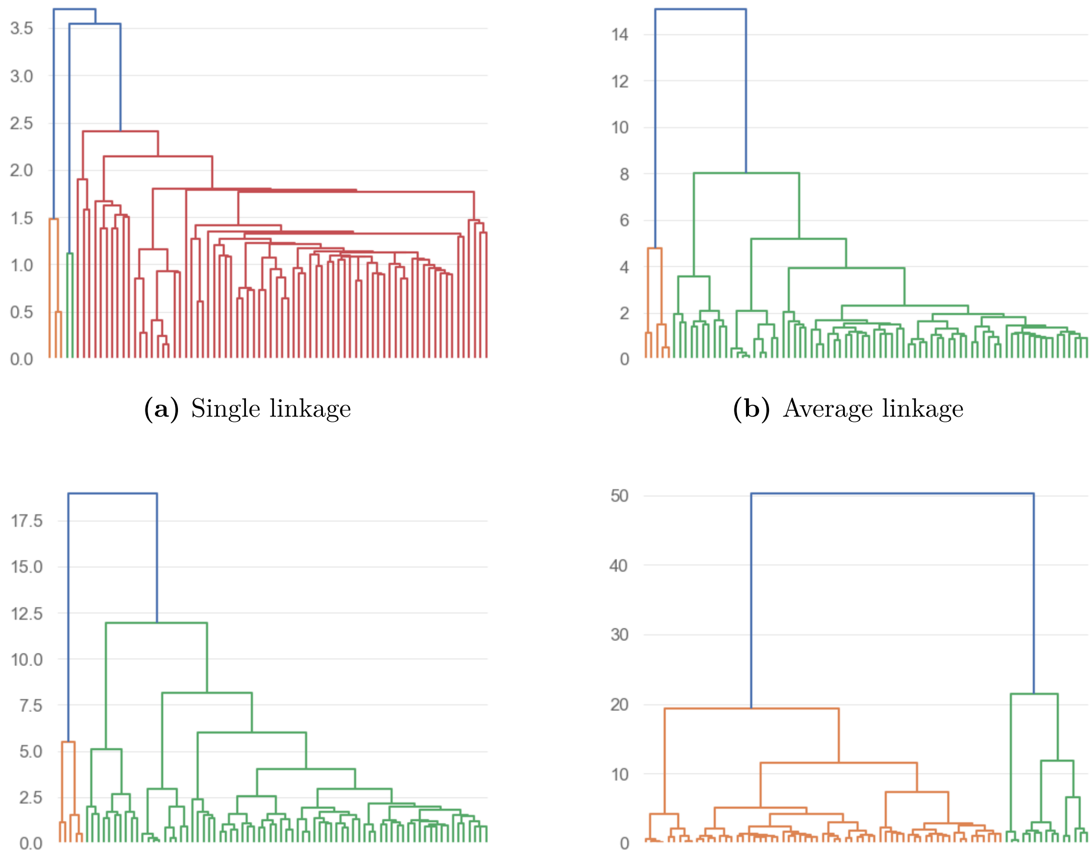
$$\boldsymbol{\mu}_k = \frac{\sum_n r_{n,k} \; \boldsymbol{x}_n}{\sum_n r_{n,k}} \tag{5.3}$$

A key aspect to take in consideration is that many indicators reflect a rather "static" point of view: many of them (e.g. ESG ratings or profitMargins) will not change value if the returns are shuffled within the same year. This aspect will be further analyzed in results Section.

5.1.2 Hierarchical Clustering

The other approach that has been implemented and tested is hierarchical clustering. This method differs significantly from K-means, as it doesn't rely on a set of features but instead requires a single square matrix as input, called *distance matrix*. All the necessary information for the algorithm must be encoded within such data structure which encapsulates all pairwise distances between each sample pair.
In this chapter, two approaches regarding *distance matrix* design will be used: as mentioned in Chapter 3, a popular choice is a correlation-based matrix as in 2.12. The other distance matrix will be instead be built based on LTDC coefficient, described as in 3.3.

Another significant modeling aspect is the *linkage type* (see 2.2.2). As we can see in Figure 5.1, changing its type produces very dissimilar results. In general, single linkage



(a) Single linkage

(b) Average linkage

(c) Complete linkage

(d) Ward linkage

Figure 5.1: Different dendrograms produced by changing only linkage type on multi-asset dataset

tends to produce imbalanced clusters due to the chaining effect: while this might be helpful is some optimization settings as HRP algorithm, the results obtained in terms of clusters are hard to understand from a human perspective and to justify. This behavior is gradually less pronounced when using Complete, Average, Ward linkage types. Ward's method, in particular, uses the Ward variance minimization algorithm [19] and usually produces compact, spherical clusters of similar size. For this reason it has been employed in this analysis and the results presented in 5.3 will be obtained with such linkage type.

There are other differences between K-means and Hierarchical clustering: even though it's not possible to use multiple indicators in the latter, distance matrices are often based on a "temporal" metric: correlation, for example, is built by taking in consideration the whole time series of returns, therefore its value depends on their time-evolution. Moreover, the input matrix already encapsulates a form of interaction between samples while indicators in K-means algorithm are estimated independently.
The last consideration will be devoted to a very different modeling choice: in K-means algorithm, the number of clusters has to be decided in advance while hierarchical clustering offers the possibility of deciding the appropriate level based on a graphical representation (the dendrogram).

5.2 Experimental Setting

This section is dedicated at presenting the experiments performed as well as the general methodology applied. The datasets and feature selection will be motivated and then the results obtained will be analyzed. Finally, the knowledge that can be extracted and valuable insights are presented in the last part of the chapter.

5.2.1 Datasets

The purpose of the analysis is to devise properties and behavior of clustering algorithms in a general and flexible setting, this is why clustering strategies will be performed on two very different datasets. This choice is not uncommon in the literature (see for example [9]), as the nature of a "financial universe" is highly dependent on the context, market conditions and investment preferences.
For these reasons, two very different datasets are chosen, both in number and nature of assets. This will help investigate general properties of the analyzed methods and context specific ones.

S&P 500 stocks dataset

As the name indicates, this dataset will be constituted as a subset of the famous Standard & Poor's 500 Index, a market-capitalization-weighted index of the 500 biggest traded companies in the U.S.
Only stocks with sufficient data in the time period considered are kept, this leaves us with 416 assets that are distributed sector-wise as in Figure 5.2. This dataset has been employed as it is composed of globally known industries, therefore it's a common investment or benchmark choice. Moreover, there is an abundance of information regarding each of

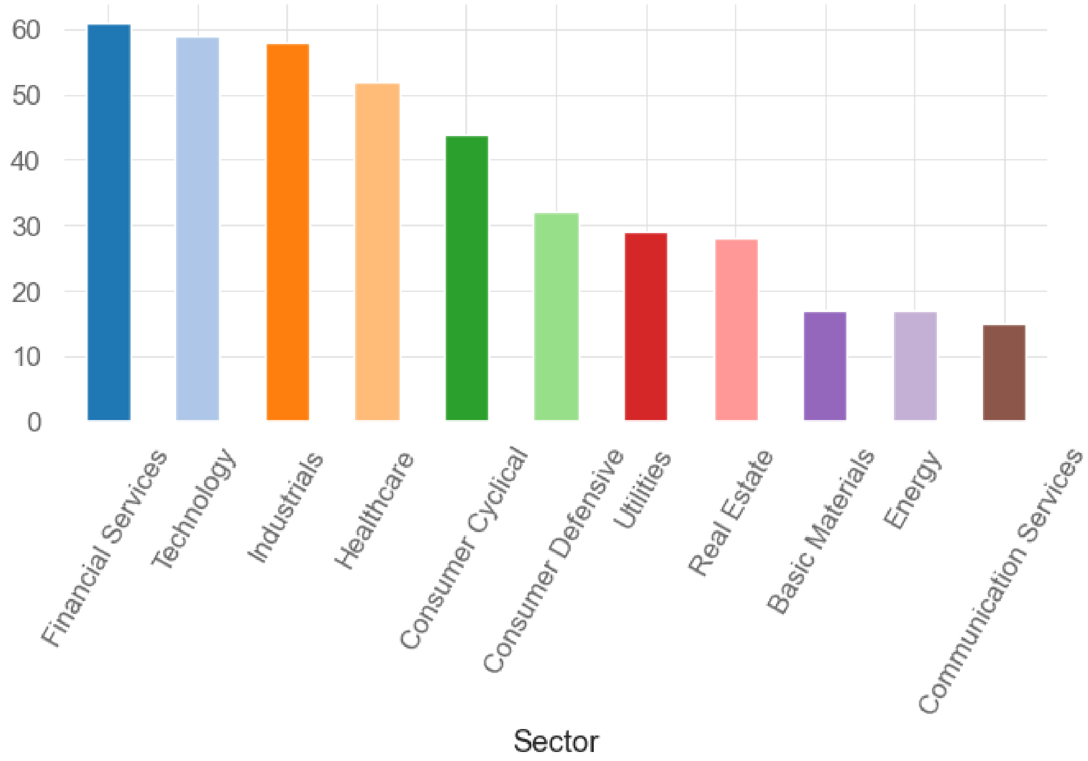


Figure 5.2: Subset of S&P 500 sector composition

these stocks which is a key component in clustering algorithms. As it will be shown in the next paragraph, this allows clustering techniques based on a wide variety of features: from economical (e.g ROE or ROA), statistical (e.g. volatility or skewness) or non-financial (ESG indicators). Selecting the right features is key in clustering algorithms, as the results will be based on such decision. Each indicator should be chosen when it is considered a meaningful characteristic for distinguishing a given element from another. At first, a wide variety of indicators is calculated and considered, each indicator is widely employed in financial data analysis and common in clustering of assets. This initial dataset will be used for analytics and comparison of the clusters devised, even though all indicators won't be used in every clustering algorithm. A summary of all 25 features that can be used is reported in Table 5.1, with a short description of each one of them.

On the other hand, this dataset is composed only by large-capitalization companies from the U.S, so one might argue that it might not contain an "intrinsic diversity", characteristic that would surely enhance clustering results. This is why a second test dataset has been employed.

Feature	Type	Description
Yavg_return	Statistical	Historical estimate of yearly expected return
Yavg_volatility	Statistical	Historical estimate of yearly volatility
industry	Market-Related	Sector in which the industry operates.
mkt_corr	Statistical/economical	Correlation with "market" benchmark
1Y_momentum	Statistical/economical	Last year return
Davg_span	Market-related	Average variation of daily prices
Davg_volume	Market-related	Average volume traded daily
Davg_Kurtosis	Statistical	Average daily returns kurtosis
Davg_Skewness	Statistical	Average daily returns skewness
Davg_eVaR	Statistical	Average VaR
Davg_eCVaR	Statistical	Average CVaR
Sharpe_ratio	Statistical/economical	Average Sharpe Ratio
totalEsg	Non-financial	ESG risk
environmentScore	Non-financial	Environmental risk
socialScore	Non-financial	Social risk
governanceScore	Non-financial	Governance risk
beta	Economical/Market-related	CAPM beta
ROA	Economical	Return On Assets
ROE	Economical	Return on Equity
est_ROI	Economical	Estimated Return on Investments
profitMargins	Economical	Margins of profit
P_B	Economical	Price to Book ratio
earningsGrowth	Economical	Growth rate for the most recent quarter.
revenuePerShare	Economical	Revenue the company generates per each outstanding share.
priceToSalesTrailing - 12Months	Economical	How much investors are paying for each dollar of the company's sales.

Table 5.1: Features considered for clustering SP500 dataset

multi-asset dataset

This dataset is constituted of asset classes (e.g. etfs, bonds, indexes or gold) that naturally present different risk-return characteristics. It counts 78 assets in total, with distribution as in Figure 5.3.

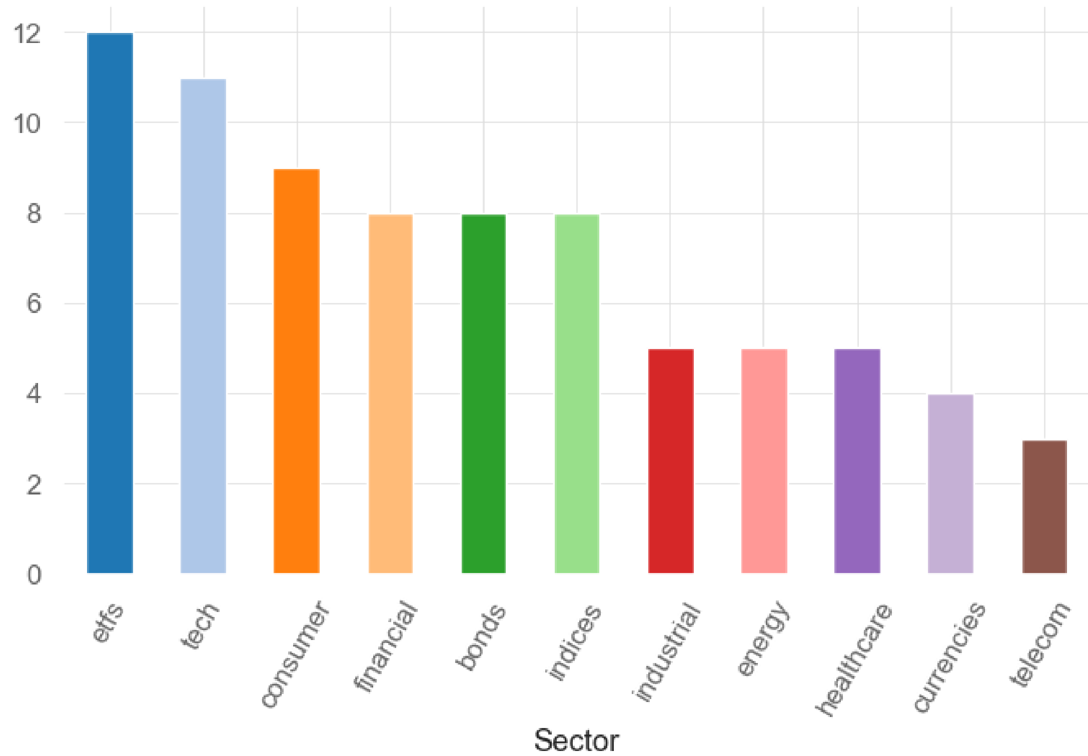


Figure 5.3: Multi asset dataset sector composition

This dataset presents elements that may be "intrinsically" more different and therefore achieve meaningful results. On the other hand, the available information is limited: the initial set of features considered is the same of the S&P dataset, but columns with not enough data are removed in order to avoid meaningless predictors. This leaves us with a total of 11 indicators, as shown in Table 5.2.

5.2.2 Methodology

All the information used in both clustering and performance analysis is obtained through **yfinance** in python: this library is built as a wrapper around Yahoo! Finance API, a powerful and rich tool that contains valuable financial data.
Since this analysis is usually performed at the start of the investment period, the time interval considered will be from January 2007 to December 2015. This is in agreement with what has been suggested by De Luca and Zuccolotto (2011) in [13]: an eight-year rolling window is required to obtain an accurate estimate for the LTDCs.
Once the starting data is retrieved, the standard data pre-processing procedure is performed:

Feature	Type	Description
Yavg_return	Statistical	Historical estimate of yearly expected return
Yavg_volatility	Statistical	Historical estimate of yearly volatility
industry	Market-Related	Sector/Class of the asset
mkt_corr	Statistical/economical	Correlation with “market” benchmark
Davg_span	Market-related	Average variation of daily prices
Davg_volume	Market-related	Average volume traded daily
Davg_Kurtosis	Statistical	Average daily returns kurtosis
Davg_Skewness	Statistical	Average daily returns skewness
Davg_eVaR	Statistical	Average daily Value-at-Risk (VaR)
Davg_eCVaR	Statistical	Average daily Conditional Value-at-Risk (CVaR)
Sharpe_ratio	Statistical/economical	Average Sharpe Ratio

Table 5.2: Features considered for clustering multi asset dataset.

1. Data Scaling: this is particularly helpful in K-means algorithm as it may become unstable when features have very different scales. This ensures the distance between different indicator has the same magnitude (either zero mean with unit standard deviation or with values between 0 and 1).
2. Outlier Detection: outliers can drastically influence K-means algorithms as they may outset the position of cluster centers: the standard distance metric (L2 norm or Euclidean distance) penalizes points assignment with the square of the distance, thus a cluster center may be “forced” to move towards a single datapoint that is significantly far from the rest of the dataset.
3. Missing Values handling
4. Categorical features transformed as numerical ones: K-means operates through a distance defined on vectors of real numbers so features like market-sector are then transformed into numerical values
5. Removal of redundant/highly correlated features: before any feature selection is applied, the correlation between features is calculated and the indicators showing high values in correlation are removed.

Once the dataset as well as the algorithms used have been presented and motivated, this still leaves us with crucial decisions, especially regarding k-means algorithm:

- Which features should be included in the clustering algorithm?
- What is the optimal number of clusters?

As is common for such research questions, there is no right answer. The most appropriate subset of indicators is highly dependent on different factors such as data availability, field-knowledge and, more importantly, the objective of the analysis. All the approaches tested will be then motivated and evaluated based on these parameters: explainability, target audience and possible use-context.

To this end, several subset of features have been tested on both datasets regarding k-means algorithm:

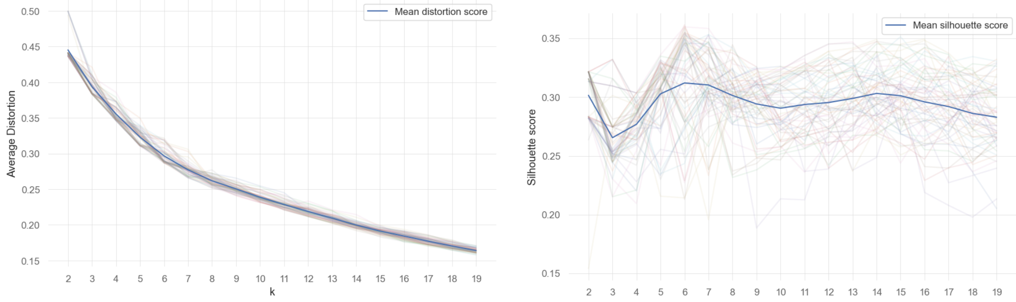
1. a basic version with: *Yavg_return*, *Yavg_volatility* and *industry*
2. a complete version with a linear projection of all features
3. a version including 5 technical indicators (only for S&P 500 dataset): *ROA*, *ROE*, *revenuePerShare*, *profitMargins* and *priceToSalesTrailing12Months*

One can motivate such modeling choice in the following way: the first selection is based on Modern Portfolio Theory, where Risk and Returns are key features for representing different asset investment opportunities and sector indicators help in providing additional information about the systematic risk an asset may have based on the sector it operates or on its nature.
The second approach is an attempt of remaining agnostic to any further assumption or feature selection, it is based on the assumption that each calculated feature is somehow significant in the process of dividing assets into different groups. Of course clustering on a lot of different features will make the optimization process harder and less explainable, for these reasons in this case PCA dimensionality reduction has been performed before running k-means algorithm in order to remove information redundancy and ease the optimization process.
Last approach is based on what Momeni et Al found in [20]: by interviewing with several of the capital market experts, the 5 profitability ratios reported above are selected as meaningful for stock categorization.

Once the first question has been addressed, we are left in determining the right number of clusters for K-means and hierarchical clustering algorithms. For the last method a common choice is to employ the *gap statistics* mentioned in 2.2.2, while for K-means there are different methods that are applied the same way: different runs with different clustering numbers are performed and some key metrics are collected. Once all runs are completed, the appropriate k is chosen based on such metrics. There is no single best choice, so several metrics are computed:

- Silhouette score [21]
- Distortion score (average squared distance between each data point and its assigned cluster center)

It is worth mentioning that a higher value of K will result in a lower distortion score as it takes into account only the tightness of the cluster, in the degenerate case of $K = N$ the distortion score would be 0 as each point would be the cluster center of itself. For this reason the optimal number of clusters based on distortion score is performed by considering the "elbow" method: it consists in identifying a value of K after which the decrease in the distortion score is only marginal. An example of the results can be observed in Figure 5.4



(a) Distortion Score **(b)** Silhouette Score

Figure 5.4: Metrics for different values of K. As K-means is sensitive to initialization, several runs for each value of K are performed.

5.3 Results

This section will present the results obtained from different clustering algorithms.
A key aspect to keep in mind is that in such settings there is no ground truth that can serve as a fair mean of comparison. Moreover, financial market is a technical and complex field therefore evaluating the goodness of a certain clustering algorithm may be difficult from a human perspective. For these reasons, the main focus of this section will be about comparing the associations produced with different clustering algorithms and devise common properties and differences.
Before diving into the analysis, it is worth mentioning that it is not guaranteed that all clustering algorithms will find the same optimal number of clusters. In this case, though, all methods were rather concordant about the number of clusters in each dataset. In order to ease comparison and improve visualization, it has been considered a common number of clusters in every dataset, based on all estimates provided by each method.

5.3.1 S&P 500 stocks dataset

The first analysis will be performed by evaluating the Adjusted Rand Index (ARI) for each pair of clustering approach: it is a common technique for evaluating the concordance of two data clustering. It is based on the Rand Index [22] but then "adjusted for chance". The adjusted Rand index is thus ensured to have a value close to 0.0 for random labeling independently of the number of clusters and samples and exactly 1.0 when clusters are identical (up to a permutation). The adjusted Rand index is bounded below by -0.5 for especially discordant clusters. Then clusters distributions of risk and performance will be presented and studied.

Clustering Similarity

Figure 5.5 depicts already a strong result: different clustering methods produce very different results. The pair of hierarchical algorithms has a non negligible similarity that is still rather close to 0 and two of the K-means subset of features share some similarity too. Apart from these cases, all other clustering choices are very far from each other. This result may come from two distinct reasons: either the elements of the dataset are inherently similar and therefore the clusters are unstable, or these clustering techniques are based on fundamentally different concepts of grouping data, or a combination of both. Analyzing the results produced from the multi asset dataset and performing further comparisons on the clusters composition will help shading some light on this matter.
Something that may surprise is the extremely low similarity between K-means performed with 5 indicators suggested by a pool of experts and all other approaches: this suggests that such method is radically different as those 5 indicators were employed in the agnostic k-means approach and still results are very different since the latter aligns more with the setting where only 3 standard indicators were employed.

Sector Distribution

Figure 5.6 shows the sector composition of each cluster divided by method. From this plot there are several observations that can be used for studying these approaches. One simple way to collect information about these results is by studying frequent associations of two or more sectors that are often grouped together. For example, K-means algorithms tend to group the industrial sector (at least part of it) with the healthcare industry, as well as technology with real estate. Hierarchical clustering instead tends to separate them. This might signify that, while there is a strong connection in terms of market and economical features, returns tend to move quite differently. This is a powerful insight: one could cluster investors based on their preferences and, if the characteristic of an investor match the one of these groups of companies (as we'll see later based on e.g. risk aversion or ESG commitment) then this association might come helpful as these firms share high indicators similarity but low correlation, especially in bad times.
Another key insight emerges by observing that there are some sectors (Energy, Utilities and part of financial services) consistently grouping together in all approaches. This is indeed important as we might argue that firms in such market sector are closely related under every metric analyzed, for this reason it might not be beneficial to include several elements of the same group as there wouldn't be such improvement in diversification. Other sectors instead -like Consumer cyclical, technology or industrials- are always split, meaning such segments are very diverse and therefore additional analyses are required before selecting a subset of assets of these sectors.
From this plot one can also devise how clusters are composed numerically: K-means tends to produce more balanced clusters, whereas hierarchical algorithms, even with Ward's linkage, create highly imbalanced groups.

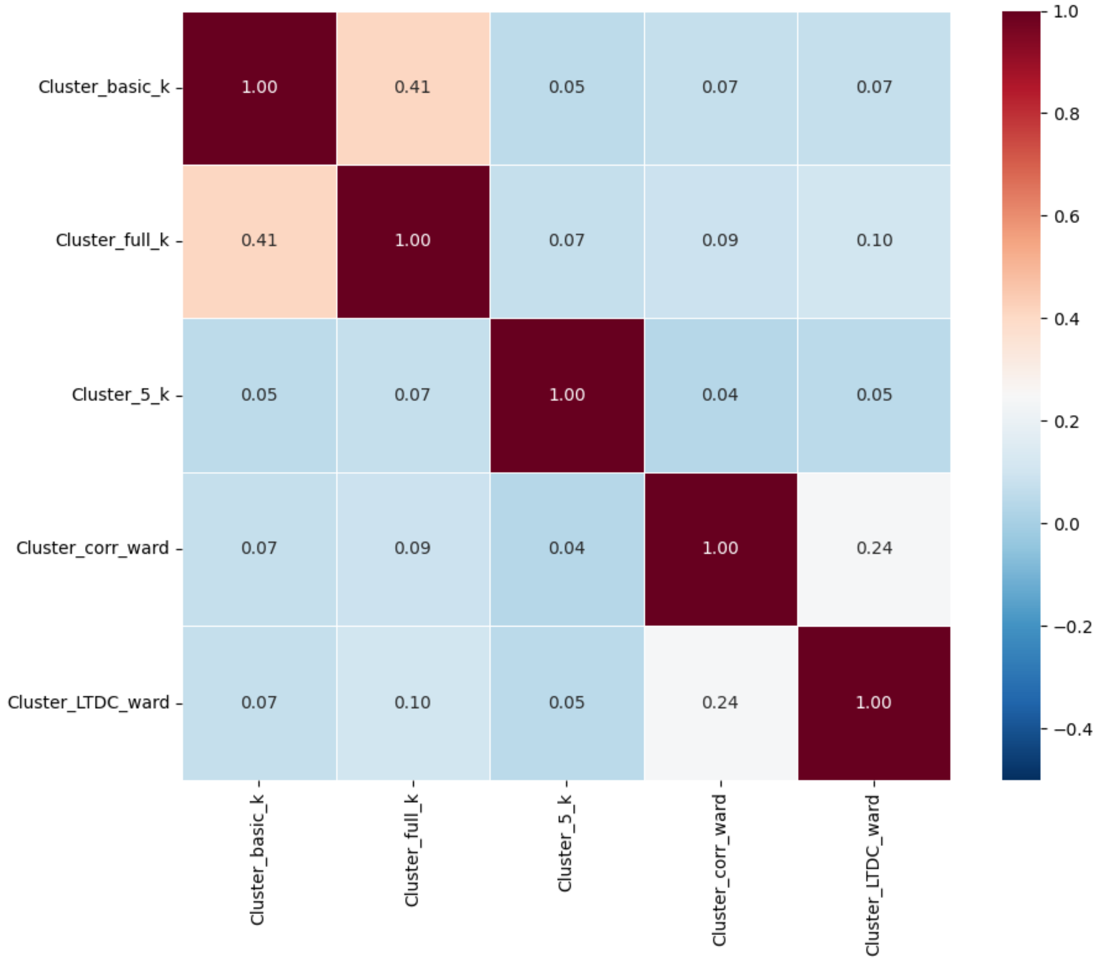


Figure 5.5: Adjusted Rand Index among pairs of clustering approaches on S&P500 dataset

Risk/Performance Analysis

Once sector composition has been analyzed, another key characterization of each cluster is assessing its risk compared to its returns. To this end, Figure 5.7 displays the average yearly return and volatility for all assets in the same cluster.
As we can expect, these two metrics are strongly positively correlated. First of all one can notice that the first two plots have a wider span of risk/returns performances if compared to the other 3 plots. This is indeed relevant as the “full kmeans” has these 2 indicators out of the full set of features. It suggests that these indicators are an important source of separation, highlighting their pivotal role in characterizing assets groups.
Last three plots do not employ expected returns and volatilities directly when grouping assets, but clusters produced with hierarchical methods exhibit a clear separation of

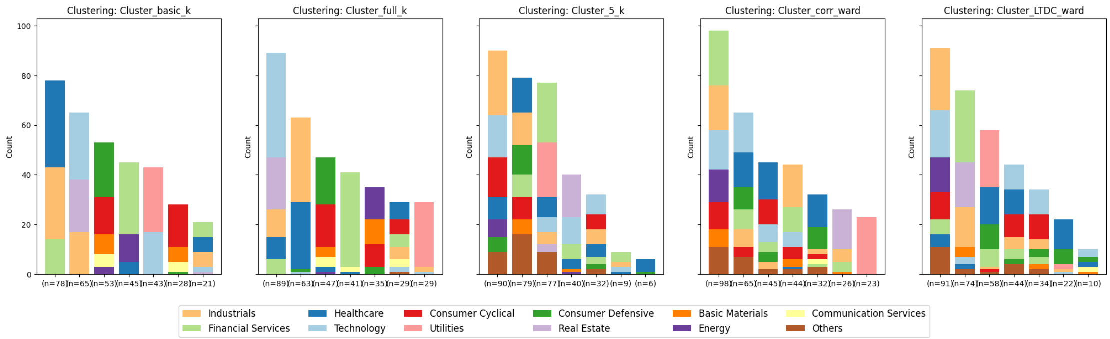


Figure 5.6: Sector composition of clusters for S&P500 dataset

returns and volatility, indicating that assets moving together in absolute terms tend also to share the same characteristics. On the contrary, clustering based on the 5 economical indicators produce rather homogeneous clusters based on these metrics: this signifies that such subset depicts a very different logic of associations.

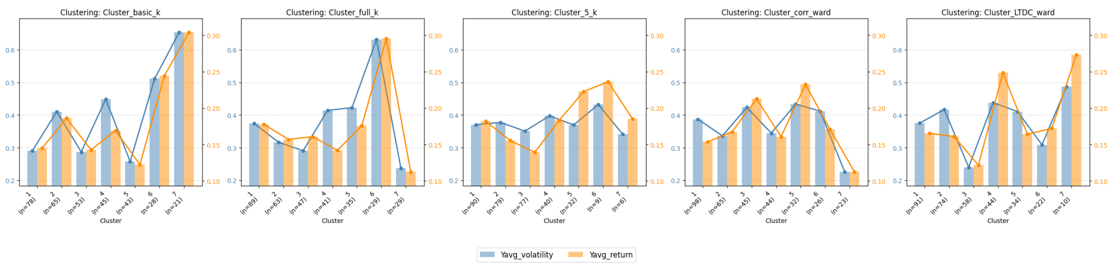


Figure 5.7: Clusters average return and volatility on S&P500 dataset

Figure 5.8 analyzes risk and performance based on different indicators: Sharpe Ratio and ESG risk. One thing that makes such plot different is the unclear relation between these two indicators. ESG ratings, in fact, capture some part of risk that is not directly encoded in simple returns and volatility, therefore SR isn't sufficient for quantifying how good an investment is, as this indicator accounts both for positive (surely wanted) volatility and negative (what's risky for real) volatility and doesn't take into account non-financial risk. It's surprising that, overall, all methods present a clear separation between clusters, meaning that these two metrics, while not enough in some cases, are crucial and effective for asset categorization.

This result is not obvious at all, since Figure 5.9 shows a very different situation: besides K-means applied with five indicators (two of which are directly Return On Equity and Return On Assets), all other clustering algorithms are rather homogeneous and not clearly separated. Once again this demonstrates how such metrics are not directly influential in describing assets within a general framework and suggests that clustering based on these

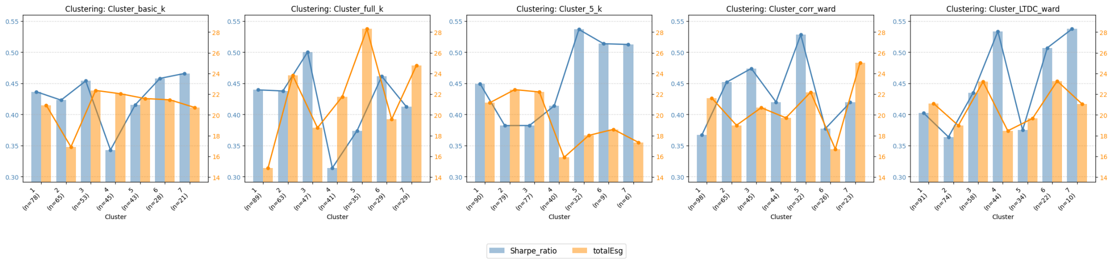


Figure 5.8: Clusters average Sharpe Ratio and ESG total risk on S&P500 dataset

indicators will likely yield very different groups: the use of this method clearly serves a distinct purpose.

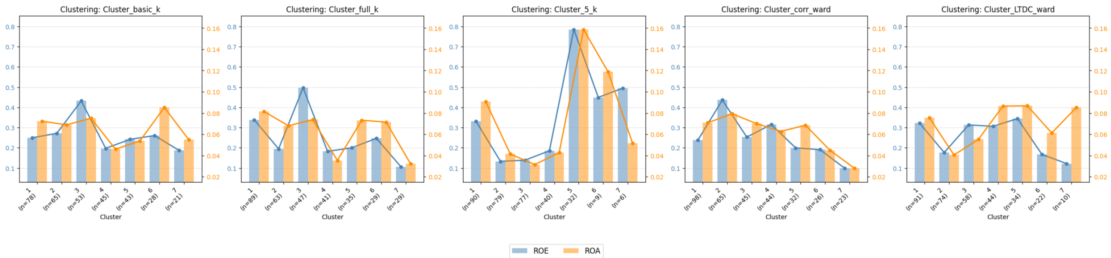


Figure 5.9: Clusters average ROE and ROA on S&P500 dataset

5.3.2 Multi-asset dataset

The same analysis performed with S&P500 dataset is now conducted on the multi-asset dataset. The cluster similarity will be studied in terms of the ARI, then the analysis will focus on sector composition and risk-return characteristics of clusters.

Clustering Similarity

An important property of clustering algorithms that was left on hold in the previous analysis was related to how close these methods could be considered. Figure 5.10 helps us in observing that, still, there is a significant difference among label assignments. At the same time, however, there is only one pair of methods that has a complete different assignment (even though it's not in contrast, it is equivalent to random labeling). Besides it, all other pairs have a non-negligible similarity. What is really surprising is the relatively high concordance between full K-means approach and hierarchical clustering based on correlation. This partially brings some evidence to our hypothesis in 5.3.1: similarity among different approaches tends to increase when assets are indeed more "distinguishable", but still there exist clear differences between them.

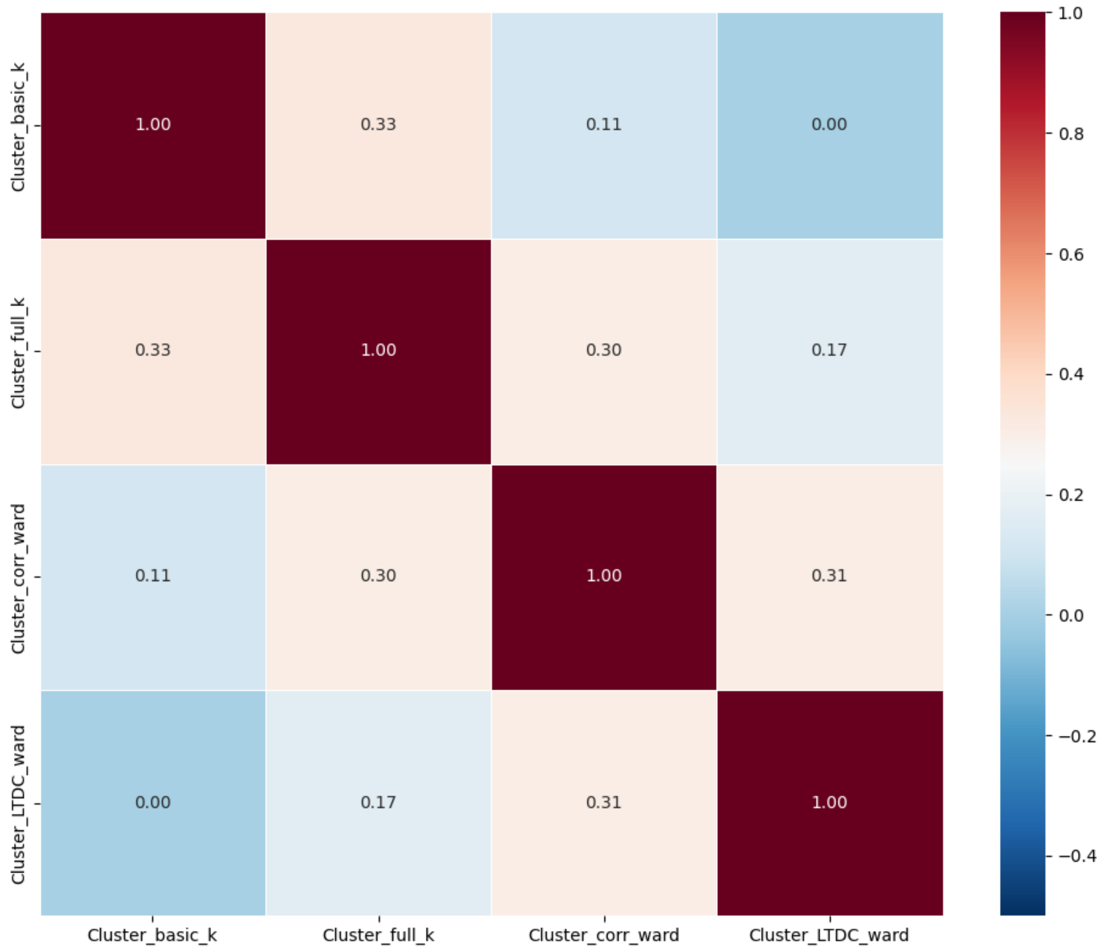


Figure 5.10: Adjusted Rand Index among pairs of clustering approaches on multi asset dataset

Sector Distribution

Similarly as in the previous analysis, Figure 5.11 displays each cluster composition based on asset classes. There are several similarities with what observed in Figure 5.6: k-means algorithms present rather balanced clusters in terms of numerosity and sectors are rather well split only in basic k-means. This plot suggests a clear link between the two central clustering assignments even though the two methods involved are different. Hierarchical clustering based on LTDC estimation, instead, result in very big clusters with mainly company based assets and then smaller clusters composed of bonds, etfs and currencies. This suggests that, during periods of crisis, similarity in terms of returns among firm based assets is very pronounced if compared to other types of financial securities, this insight will need further considerations if there is a particular need of diversification in unstable scenarios, suggesting that stock-only portfolios might not provided the diversification level needed during difficult periods.

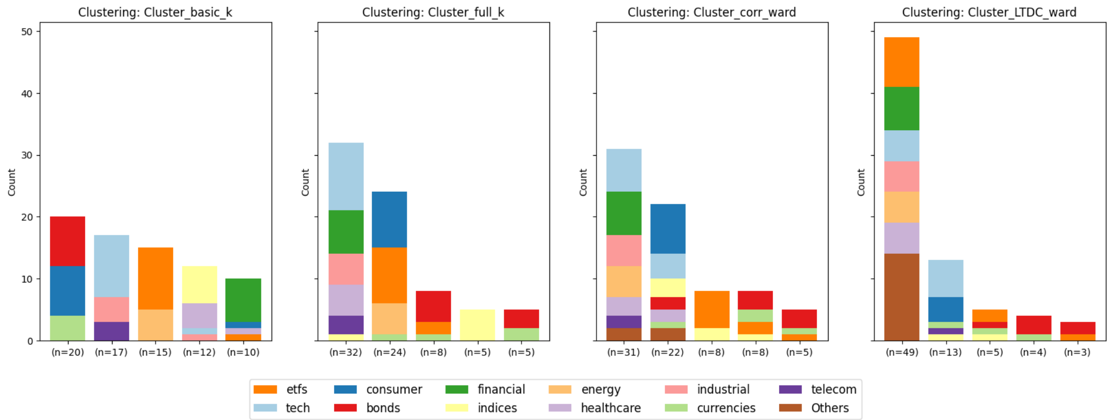


Figure 5.11: Asset type composition of clusters for multi asset dataset

Risk/Performance Analysis

The analysis continues by inspecting risk and return associated with each cluster formed. Figure 5.12 shows how different clusters are based on such metrics. Something that differs from Figure 5.7 is the correlation between expected returns and volatility: in this case it seems to be slightly lower than in the previous case. By using the same scale among all plots, we can also visualize the tradeoff between these two measures: in S&P 500 dataset, both bars were relatively closer to each other, suggesting a fair balance between risk and payoff. In this case, however, several clusters tend to clearly overperform (e.g. cluster 2 in basic k-means or LTDC-based hierarchical clustering) while others that are significantly underperforming (e.g. clusters 4 and 5 in full k-means). This is a powerful yet simple tool for performing a preliminary analysis and maybe rule out clusters that seem to be closely related and that have not performed well in the past.

In this dataset, as already mentioned, several indicators were not available. For this reason a different visualization is chosen in order to study different metrics on risk and performances. Figure 5.13 displays the average Sharpe Ratio, representing a common performance measure, and the average Expected Shortfall (CVaR), that measures tail risk, i.e. risk for heavy losses. First, there is no clear relationship between these two metrics, suggesting that these two indicators might both be important in assessing key characteristic of a group of assets. This is crucial in assets analysis as their returns aren't always normally distributed so it's important to investigate securities that may face very large losses even if their sharpe ratio is above average.

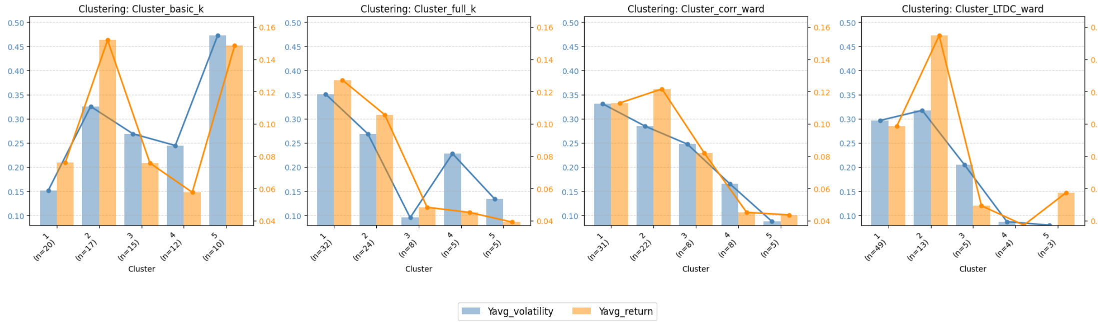


Figure 5.12: Clusters average return and volatility on multi asset dataset

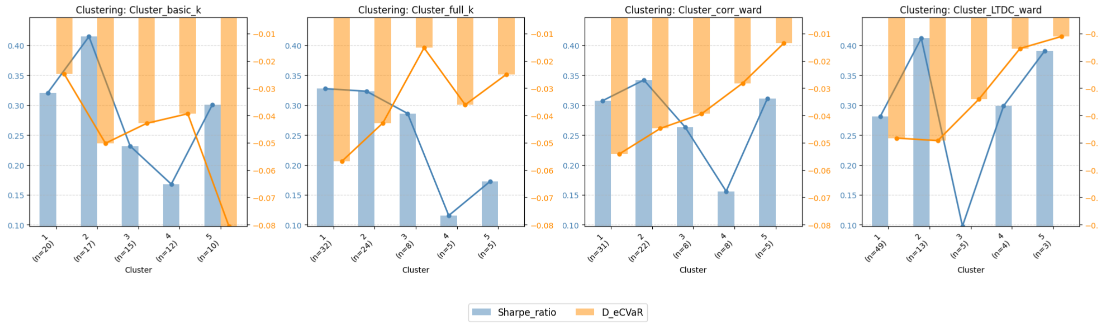


Figure 5.13: Clusters average Sharpe Ratio and CVaR on multi asset dataset

5.3.3 Final considerations

The analysis of asset clustering is a key component in portfolio construction and management and it is also a powerful data mining technique for extracting useful insights and making informed decisions, this is the reason why it has been included in this project. The approach that has been adopted has given the attention on flexibility and explainability of results and, while it is not the only possible way, it helps in achieving both goals. Once all the analysis has been performed, it becomes then relevant to summarize the knowledge extracted and categorize clustering techniques based on the following factors: explainability, target audience and possible use-context.
K-means algorithm resulted in balanced clusters, especially in its basic and full version: it achieves better explainability and it has to be considered when the main objective is to use clustering as an analytic tool. If the main focus is on asset analysis, dataset summarization or highlighting associations (for example between sectors based on complex patterns or simple risk and returns) then this method is surely the one to be preferred. Experimental tests reveal higher correlation between clusters than hierarchical approaches, thus the

main use of this method would be to suggest a given cluster or subset of it based on investor characteristics and then perform optimization only on that subset of assets. If we consider K-means based on the indicators suggested by a pool of experts, we find the results to be very far from all other methods. This indicates that associations based on these indicators are very different from other approaches. The use of this method might then be limited to experts of financial economy.
Hierarchical clusters, instead, tend to be harder to visualize from a human perspective. As they are widely employed when focus is given on the optimization that follow this first stage, then their use-context is highly related to it. Nevertheless, they may also be employed for asset recommendation in the presence of particular hedging/diversification needs, for example when a portfolio has already been constructed and only small changes are allowed. Another valuable application of these methods could be using them as backtesting for associations identified through other methods, this would help enforcing evidences for a specific type of investment decision.

Chapter 6

Simulations and Experimental Results

Once the analysis of the investment universe is conducted and sufficient information has been gathered, it is now time to use this acquired knowledge and construct a well diversified portfolio of assets that can perform well even during unstable periods. The first part of this Chapter will be devoted to explaining how the dataset has been chosen, how different allocation methods will be evaluated and compared, and how the simulations are conducted.
After that, a comprehensive analysis of the hyper-parameters involved in the methods developed will help in identifying which combinations seem to be more promising for a specific investor profile (e.g. more conservative or aggressive) or market condition.
Finally, an extensive simulation comparing various allocation strategies will help to assess the effectiveness of the implemented approaches against benchmarks and traditional allocation methods.

6.1 Framework

In the first part of the following section, it will be presented how a portfolio has been derived from the results of the clustering analysis and why this specific dataset of assets has been chosen.
Afterwards, this section will be dedicated to describing in detail how simulations have been structured and implemented in practice, including the motivations behind such modeling choices.
Finally, all approaches tested in the final simulation will be described in order to give a comprehensive overview of the experimental settings.

6.1.1 Dataset

It has already been mentioned in Chapter 5 how the choice of the investment portfolio is crucial, before even thinking of any allocation method. Evidences suggest that it is not necessary to hold a big portfolio (in number of assets) in order to obtain a good level of diversification. For this reason, 15 securities will be used in the following simulations. A common benchmark for all investments will be employed to compare with a real world allocation strategy how the implemented approaches have performed. To confirm such hypothesis, a portfolio composed of VT (Vanguard Total World Stock ETF)[1] -a global equity market index that well represents global stocks performances- and AGG (iShares Core U.S. Aggregate Bond ETF)[2] -an index representing bond market exposure across various maturities and credit qualities- will be employed as benchmark. This indexes contain respectively 9792 assets and over 11,000 individual bonds, way more securities than the ones employed in this simulation.

First of all, the investment universe needs to be chosen. To this end, we can exploit the previous clustering analysis in order to correctly choose a well diversified portfolio. The multi asset dataset presented in 5.2.1 will be used as our investment universe: the motivations behind such modeling choice are multiple

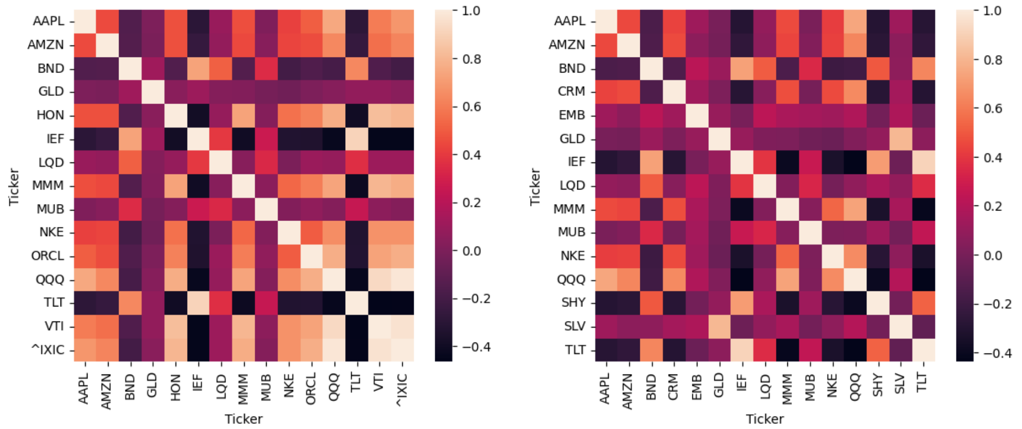
1. Overall, clustering methods have shown higher concordance than in the other dataset, suggesting better clustering performances overall
2. Assets are, on average, less correlated as they belong to different classes

Once the investment universe is chosen, we need to restrict ourselves to a subset of assets that will constitute our investment portfolio. We start by considering the clusters identified in Chapter 5. In the conclusions Section, we identified the hierarchical methods as the best candidates for optimization and diversification clustering. This is why, based on that groupings, three assets from each of the 5 clusters identified are chosen based on historical Sharpe ratio, estimated with historical data between 2007 and 2015 (the same window frame as for the clustering experiments). Figure 6.1 gives evidence to the effectiveness of hierarchical clustering methods.

By considering the whole dataset constituted of 74 assets, the average pairwise correlation between assets is 0.3322. When we instead focus only on the two subsets of 15 assets presented above, we observe that the average pairwise correlation decreases to 0.2015 when choosing based on correlation hierarchical clustering, while with LTDC based clustering the overall correlation drops to 0.1136. This highlights the effectiveness of such clustering techniques that help finding low correlated assets within a wide universe. For the following experiments, LTDC based subset will be chosen as dataset. It is important to notice, however, that the two considered subsets overlap greatly, this is why we restricted the analysis to this subset only.
Table 6.1 summarizes the assets employed in the following experiments.

[1] Look at `https://investor.vanguard.com/investment-products/etfs/profile/vt#portfolio-composition` for more information

[2] For further information go to `https://www.ishares.com/us/products/239458/ishares-core-total-us-bond-market-etf`



(a) Selection with correlation-based hierarchical clustering

(b) Selection with LTDC-based hierarchical clustering

Figure 6.1: Correlation heatmap of the selected investment portfolios

6.1.2 Evaluation metrics

Each method will be evaluated based on many different metrics in order to provide a wide understanding of its strengths and criticalities. The simplest indicators are indeed the final return obtained -that is normalized based on the time of the investment period as the compounded annual growth (CAGR)- and the realized volatility of the portfolio returns, where Y is the length of the simulations in years and $\bar{r}$ the mean return:

$$\text{CAGR} = \left(\prod_{t=1}^{T} (1 + r_t) \right)^{\frac{1}{Y}} - 1 \tag{6.1}$$

$$\hat{\sigma} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2} \tag{6.2}$$

A common metric that is used to encode both indicators is the Sharpe ratio, here calculated as[3]:

$$\mathcal{SR} = \frac{CAGR - r_f}{\hat{\sigma}} \tag{6.3}$$

When we asses the riskiness of an investment, there are other measures that might be more indicated if we want to evaluate other risks associated with the resulting asset allocation: the Max drawdown (MDD) and the Conditional Value At Risk (CVaR). The first is an indicator of what has been the biggest drop of our portfolio value, i.e. the

[3] In accordance with [9], $r_f = 0$ is assumed when calculating this and the following metrics that require such indicator.

Ticker	Type	Description
AAPL	Technology (Single Stock)	Apple Inc.
AMZN	Consumer (Single Stock)	Amazon.com Inc.
BND	Bond (ETF)	Vanguard Total Bond Market Index Fund.
CRM	Technology (Single Stock)	Salesforce, Inc.
EMB	Bond (ETF)	iShares J.P. Morgan USD Emerging Markets Bond ETF.
GLD	Precious Metals (ETF)	SPDR Gold Shares.
IEF	Bond (ETF)	iShares 7-10 Year Treasury Bond ETF.
LQD	Bond (ETF)	iShares iBoxx $ Investment Grade Corporate Bond ETF.
MMM	Industrial (Single Stock)	3M Company.
MUB	Bond (ETF)	iShares National Muni Bond ETF.
NKE	Consumer (Single Stock)	NIKE, Inc.
QQQ	US Large-Cap (ETF)	Invesco QQQ Trust.
SHY	Bond (ETF)	iShares 1-3 Year Treasury Bond ETF.
SLV	Precious Metals (ETF)	iShares Silver Trust.
TLT	Bond (ETF)	iShares 20+ Year Treasury Bond ETF.

Table 6.1: Dataset composition.

largest cumulative loss from peak to bottom. It can be calculated as

$$\mathcal{MDD} = \max_{t^*<t<T} \frac{V_t - V_{t*}}{V_{t*}} \tag{6.4}$$

As mentioned in Section 3.2, CVaR at level α is a risk measure that is obtained by taking the expected value of the losses in the left tail distribution (if we consider for example returns distribution) for values exceeding the VaR at level α, i.e. the $1-\alpha$ quantile of the distribution in this case. Considering as X the returns distribution and R_0 the percentage return of a risk-free asset[4] and assuming X is continue:

$$\text{CVaR}_\alpha = \mathbb{E}\left[L|L \geq VaR_\alpha(X)\right] = \mathbb{E}\left[L|L \geq F_L^{-1}(1-\alpha)\right] \qquad L = -\frac{X}{R_0} \tag{6.5}$$

[4]Check [2] Chapter 6 for more details

Another category of indicators is the one assessing the quality of the optimization approach in terms of stability and underlying level of diversification. For this reason, two metrics that will be calculated are the average turnover per reallocation ($\mathcal{TO}$) and the Sum of Squared Portfolio Weights[5] ($\mathcal{SSPW}$)

$$\mathcal{TO} = \frac{1}{T}\sum_{t=2}^{T}\frac{1}{N}\sum_{i=1}^{N}|w_{i,t} - w_{i,t-1}| \tag{6.6}$$

$$\mathcal{SSPW} = \frac{1}{T}\sum_{t=2}^{T}\sum_{i=1}^{N} w_{i,t}^2 \tag{6.7}$$

6.1.3 Simulations and Setting

This Thesis project, as already mentioned in Chapter 1, will serve a dual purpose with the following experiments. First of all, the two main approaches that have been implemented are relatively new, this leaves us room for new experiments and parameters studying. As stated in previous Sections or for example by Raffinot in his paper "The Hierarchical Equal Risk Contribution Portfolio" (2018): "Last but not least, this article opens the door for further research. Typical machine learning issues have to be investigated, such as the choice of the distance metric and the criteria used to select the number of clusters". Moreover, we have also brought up in Chapter 4 how there is no consensus on the hyper-parameters to use in the robust optimization problem. This motivates us in dedicating importance to studying the influence that different parameters have in the resulting portfolios. Since this work focuses on the allocation of assets during uncertain periods, this analysis will be conducted over a period containing great market instability as well as more stable phases, namely between January 2015 and December 2021. The reasons behind it can be observed in Figure 6.2: the biggest periods of recession can be identified as the years 2022-2023 first and 2020-2021 secondly. The analysis will then be performed so that includes at least one of those two periods, as well as some minor drawdonws in the years before.
This will help analyzing how different parameters influence the allocation type and identifying the best parameters combination for a given purpose.

The second objective of these experiments is to simulate how these methods behave in real investment scenarios. The following years have been identified in order to give a general overview of their behavior :

1. A period of a relatively static market. It has been identified as the interval from January 2024 and June 2025. This will shed some light on how these approaches perform in a standard investment setting (although recent events greatly destabilized global markets again)
2. An investment during a moment of great market instability. The period between January 2022 and December 2023 has been chosen, as it has been the worst period in

[5]See [23] for more information

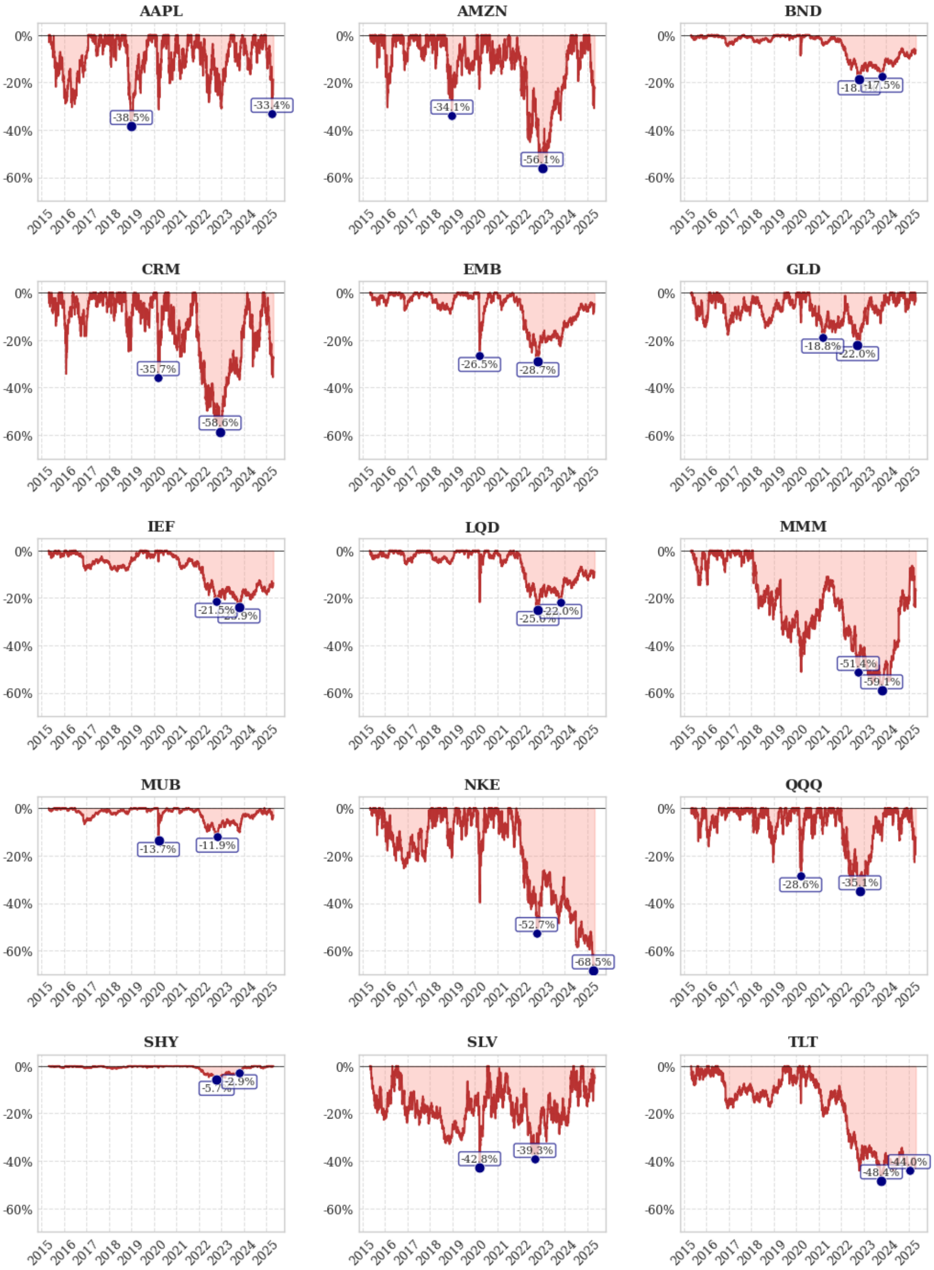


Figure 6.2: Maximum drawdowns by asset. Blue points indicate the two worst drawdowns for each asset, separated by at least 2 years.

terms of drawdown and could have been detected by analyzing the rolling volatility shown in Figure 6.3 of the previous years. This will show how these methods can mitigate losses and, therefore, hedge risk during uncertain periods.

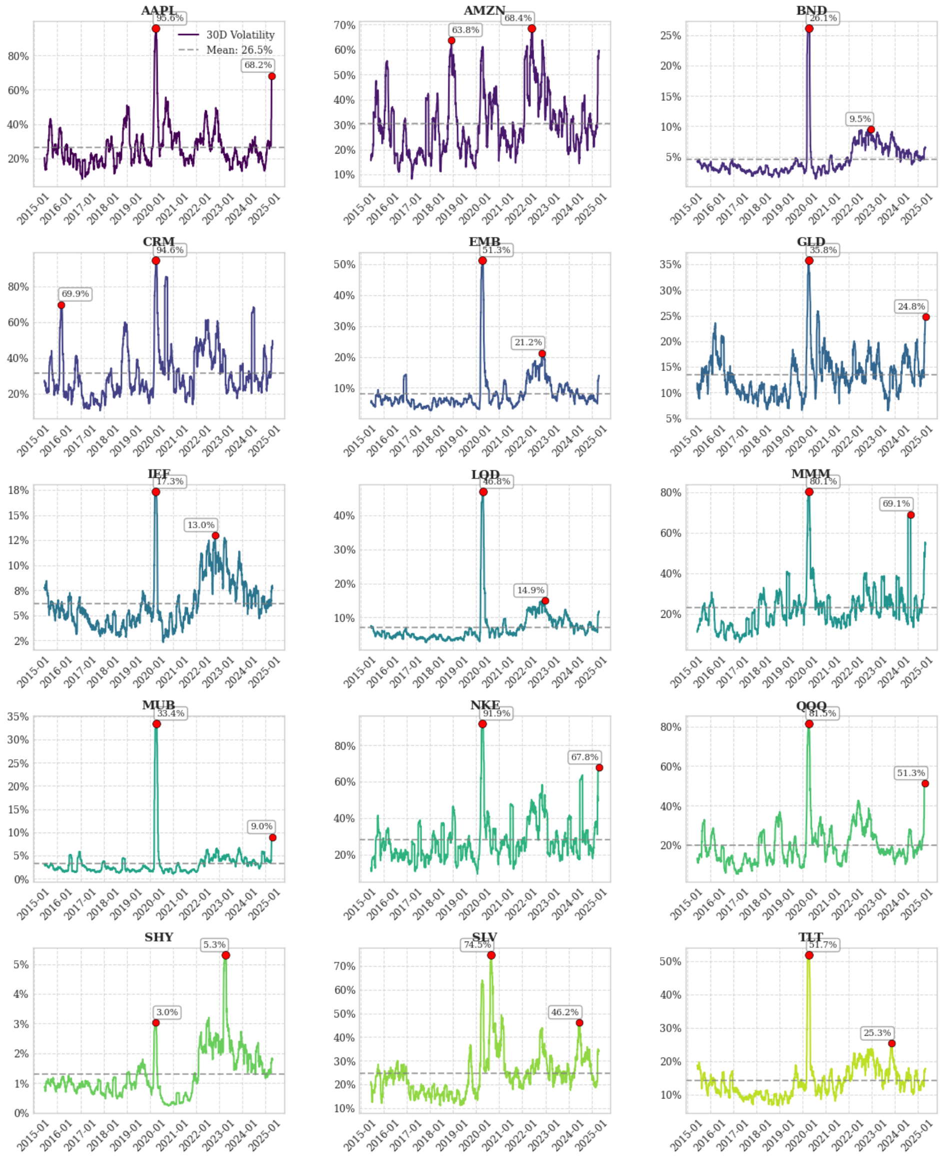


Figure 6.3: Rolling 30 day annual volatility by asset. The two red points highlight highest volatility periods separated by at least 2 years.

The key factors influencing the portfolio constitution are the periodic reallocation and the natural evolution of each asset price. For this reason, the simulations will be performed with the following temporal logic: at the start of day t, the portfolio weights might be re-adjusted based on the reallocation frequency chosen (more on it later in this paragraph). A valid set of weights at time t will have the following properties:

- $\sum_{i=1}^{N} \tilde{w}_{i,t} = 1$

- $\tilde{w}_{i,t} \geq 0 \quad i = 1, ..., N$
- transaction costs can be afforded, i.e

$$\tilde{V}_t = V_{t-1}\left(1 - \xi \sum_{i=1}^{N} |\tilde{w}_{i,t} - w_{i,t-1}|\right) > 0 \rightarrow \xi \sum_{i=1}^{N} |\tilde{w}_{i,t} - w_{i,t-1}| < 1$$

  where ξ is the transaction fee[6] and $\tilde{V}_t$ is the portfolio value at the start of day t after reallocating and V_{t-1} is the portfolio value at the end of day $t-1$.

Notice that $\tilde{V}_t$ and $\tilde{w_{i,t}}$ indicate an intermediate situation: once the portfolio weights are updated, a day passes and the market evolves, generating new returns for each asset $\boldsymbol{r}_t \in \mathbb{R}^N$. These returns influence the current portfolio in the following way:

$$V_t = (1 + r_{P,t})\tilde{V}_t = \tilde{V}_t \sum_{i=1}^{N}(1 + r_{i,t})\tilde{w}_{i,t} \tag{6.8}$$

Moreover, this will influence the weights of the portfolio

$$w_{i,t} = \frac{(1 + r_{i,t})\tilde{w}_{i,t}}{\sum_{i=1}^{N}(1 + r_{i,t})\tilde{w}_{i,t}} \tag{6.9}$$

Portfolios are therefore updated on a daily basis, with a rolling window of 2 years (504 days) for computing historical estimates of covariance and expected returns. Literature suggests[7] that estimates for lower tail dependency coefficients require instead an eight-year window size in order to achieve an accurate estimate. This suggestion will be examined and tested in the next Sections. Both the analysis and simulations will feature 2 types of reallocation frequency: yearly and every 4 months. The first reallocation frequency is in accord with common investing principles, while a more frequent portfolio adjustment might help preventing sudden market changes even at cost of higher transaction costs. This will allow for better comparison in terms of stability of the analyzed methods and recovery performances.

6.1.4 Approaches tested

As the setting of each simulation has been explained in detail, this paragraph will describe all tested methods that will be evaluated in the next chapters. Section 6.2 analyzes the approach described in Chapter 4, the impact of the involved parameters (e.g. the choice of k or $\boldsymbol{\Omega_\mu}$). The main source of comparison in that case will be the equivalent non robust MVP. In the end, a couple of "optimization parameters settings" will be chosen to be evaluated in the real simulations.

Section 6.3 will perform a similar analysis but based on hierarchical clustering allocation with the use of estimated LTDC as similarity matrix. Again, experiments on the influence

[6]In this project $\xi = 0.5\%$, for more details check [24] in Section "Actual Allocation"

[7]Check [13]

of the modeling parameters will be run. Such method will be evaluated in comparison with the other two main hierarchical allocation methods present in the literature: Hierarchical Risk Parity (HRP) and Hierarchical Equal Risk Contribution (HERC). Moreover, also the classic MVP will be compared to these hierarchical approaches in order to give a broader overview on the effectiveness of those methods. Again, the most promising parameters sets will be then chosen and further studied in the later Sections by testing them in real investment scenarios.

The last section will then be dedicated to the real simulation, various methods will be taken into account in order to ensure a fair and comprehensive analysis. Table 6.2 gives an overview of all methods tested and implemented along with their acronyms and a short description.

Method	Description
Equal-Weighted Portfolio (EWP)	Assigns an equal weight to each asset†.
Mean Variance Portfolio (MVP)	Mean Variance Portfolio with long only positions.
Global Minimum Variance Portfolio (GMVP)	Portfolio with the lowest possible variance.
Hierarchical Risk Parity (HRP)	Allocates assets based on hierarchical clustering to minimize risk concentration.
Hierarchical Equal Risk Contribution (HERC)	Method built on top of HRP.
RO-MVP	Robust Optimization of MVP allocation.
HERC-LTDC	Developed method that uses LTDC-based hierarchical structure.
Benchmark (30% VT + 70% AGG)	A reference portfolio†† used for performance comparison and assess overall diversification of the chosen portfolio.

Table 6.2: Asset allocation methods tested.
† This is indeed a simple heuristic that has proven to be quite hard to beat, for example the authors in [25] find that "Of the fourteen models of optimal portfolio choice [...], none is consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover."
†† This repartition corresponds to a "Conservative Target Risk" in Morningstar's U.S. allocation categories.

6.2 RO-MVP analysis and hyper-parameters choice

The first analysis performed evaluates the effect of different optimization parameters on the resulting portfolio allocation weights for the Robust Optimization model.
Table 6.3 shows all possible values tested for each parameter, with respect to the definitions given earlier in this project. A grid search is then performed by testing all combinations on the period mentioned above, and the results obtained are evaluated. The values of the risk aversion parameter γ have been chosen based on [26], where it is stated: "Most individuals have risk aversions between 1 and 10" concerning Mean-Variance utility, and many studies have directly estimated the average value from a pool of investors through surveys. The outcome of these studies are rather spread around 3 and 8. Since this optimization is carried in an uncertain setting, the value of 5 and 10 have been chosen as they represent the average risk aversion and the highest possible (rationally) one.

Hyper-Parameter	Values
$\boldsymbol{\Omega_\mu}$	$\{\hat{\boldsymbol{\Sigma}}, \text{diag}(\hat{\boldsymbol{\Sigma}}), \left(\text{diag}(\hat{\boldsymbol{\Sigma}})\right)^{\frac{1}{2}}, \mathbb{I}_n \}$
k	{SR, agn, norm}
$\gamma^\dagger$	{5, 10}
Reallocation frequency	{84, 252}

Table 6.3: Hyperparameters options for RO-MVP
† even if γ isn't properly an optimization parameter but more of a value derived from the investor's characteristics, multiple choices have been tested to asses the influence of different values on the resulting portfolios.

Figure 6.4 displays the cumulative returns of the portfolio over time from the top 20 combinations ordered by Sharpe ratio. It is quite evident how the parameters play a crucial role in the portfolio characteristics as there are three different type of portfolio obtained: a more conservative one, presenting low variability but also lower returns, a more aggressive investment style and a balanced allocation.

For this reason, 3 different combinations of hyper-parameters have been chosen as the most promising ones, shown in Table 6.3. Each combination has been selected as the one having the highest Sharpe ration among the ones belonging to the same cluster.

Before diving into the comparison with the classic MVP optimization, the effect of each parameter is studied in detail.
Figures 6.5 and 6.6 bring some key insights concerning the validity and influence of the parameters:

Choice of $\boldsymbol{\Omega_\mu}$ Overall, there isn't a particular choice that is right away preferable, but it is worth mentioning that $\hat{\boldsymbol{\Sigma}}$ and $\mathbb{I}_n$ have better mean performances. The first

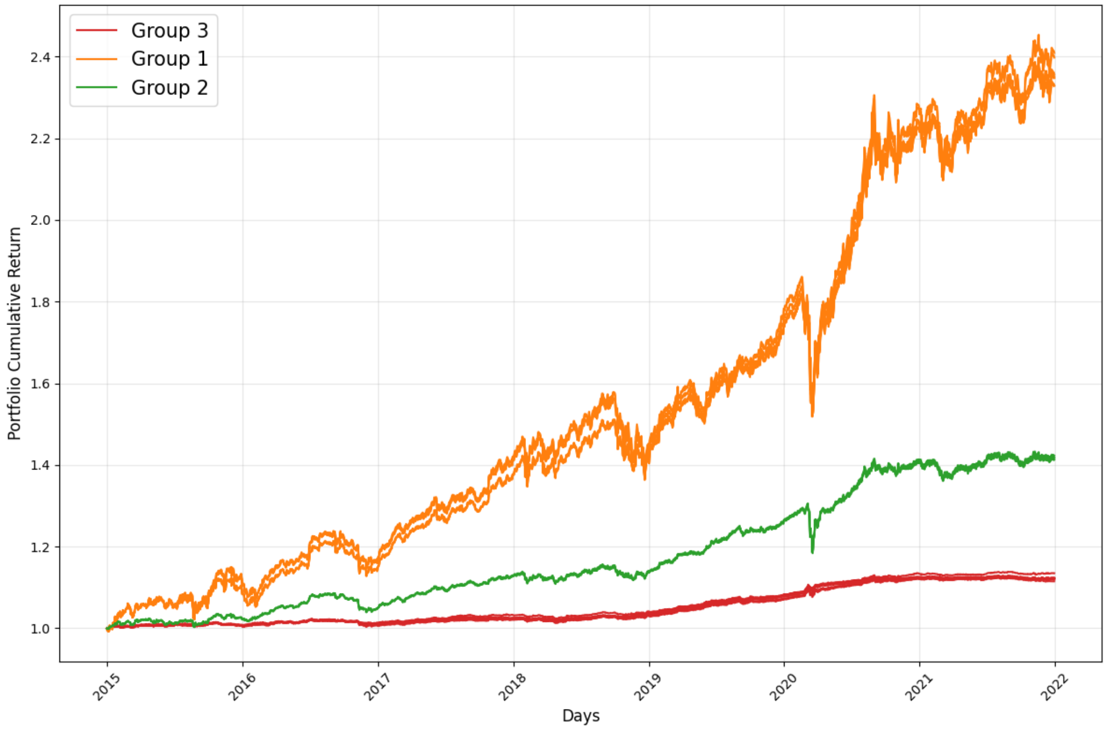


Figure 6.4: Portfolio evolution over time for top 20 hyper-parameters combinations in RO-MVP optimization.

Label	γ	$\boldsymbol{\Omega}_{\mu}$	k	Reallocation frequency
Conservative (RO-MVP-C)	5	$\hat{\boldsymbol{\Sigma}}$	agn	84
Balanced (RO-MVP-B)	5	$(\mathrm{diag}(\hat{\boldsymbol{\Sigma}}))^{\frac{1}{2}}$	norm	84
Aggressive (RO-MVP-A)	5	$\mathbb{I}_n$	norm	84

Table 6.4: Investment strategies chosen for RO-MVP approach

choice produces very different results but, overall, the majority of them present solutions that are able to hedge risk in terms of volatility and drawdowns even at cost of lower returns. The identity matrix, instead, produces results that tend to vary less and that generate substantially higher returns and higher volatility, even if in terms of drawdown the results are quite promising.

Choice of *Reallocation frequency* There is no clear effect on the obtained allocations, this was indeed expected since the method has been constructed to be less sensible to noise and inputs variability. Yearly reallocation strategies present, overall, a slightly lower drawdown and lower spread of results in terms of sharpe ratio, but this is not enough to prefer it over a four months reallocation strategy.

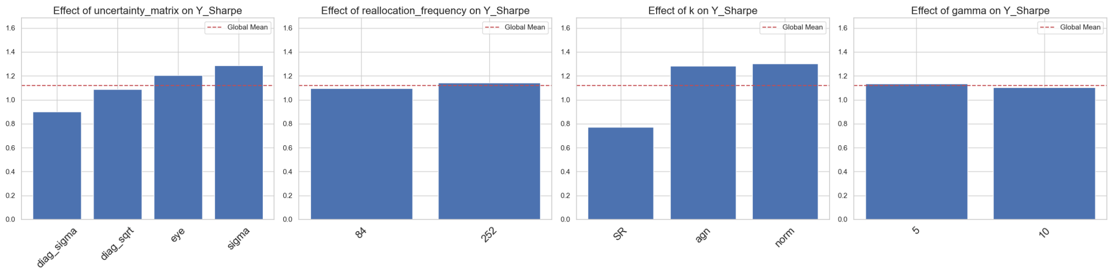


Figure 6.5: Portfolio optimization aggregated performance based on different hyper-parameters for RO-MVP optimization.

Choice of k In this case the choice of the uncertainty level based on the average sharpe ratio can be considered inferior to the statistical-based strategies. We can notice a Compounded Annual Return that is higher but a cost of an even higher risk both in terms of volatility and drawdown. Both "norm" and "agn" based values of k, instead, share similar characteristics and therefore there is no evidence to prefer one over the other.

Choice of γ The risk aversion parameter, similarly to the reallocation frequency, doesn't impact significantly the optimization solution. This means that, in terms of sensitivity, the optimization problem is highly more sensible to the Robust Optimization parameters rather than the ones derived from the classical MVP optimization. This indeed reflects the concept of Worst-Case Robust Optimization as we are preparing ourselves to the "worst" possible outcome independent on how we are categorized. Then the key modeling aspect is to characterize such "worst" possible scenario.

Once the analysis has been conducted, the combinations in Table 6.3 are compared against an investment portfolio obtained with the classic MVP optimization. Figure 6.7 highlights that classical approaches result in portfolios exhibiting less robustness during recession periods and higher returns during stable market conditions. Even when employing a high risk aversion coefficient of $\gamma = 10$, the resulting MVP is more exposed to drawdown risk and volatility than the aggressive configuration *RO-MVP-A*. This once again confirms that using standard approaches during uncertain conditions might expose the investor to larger losses and even, in the case of *RO-MVP-A* and *MVP-10*, lower returns.

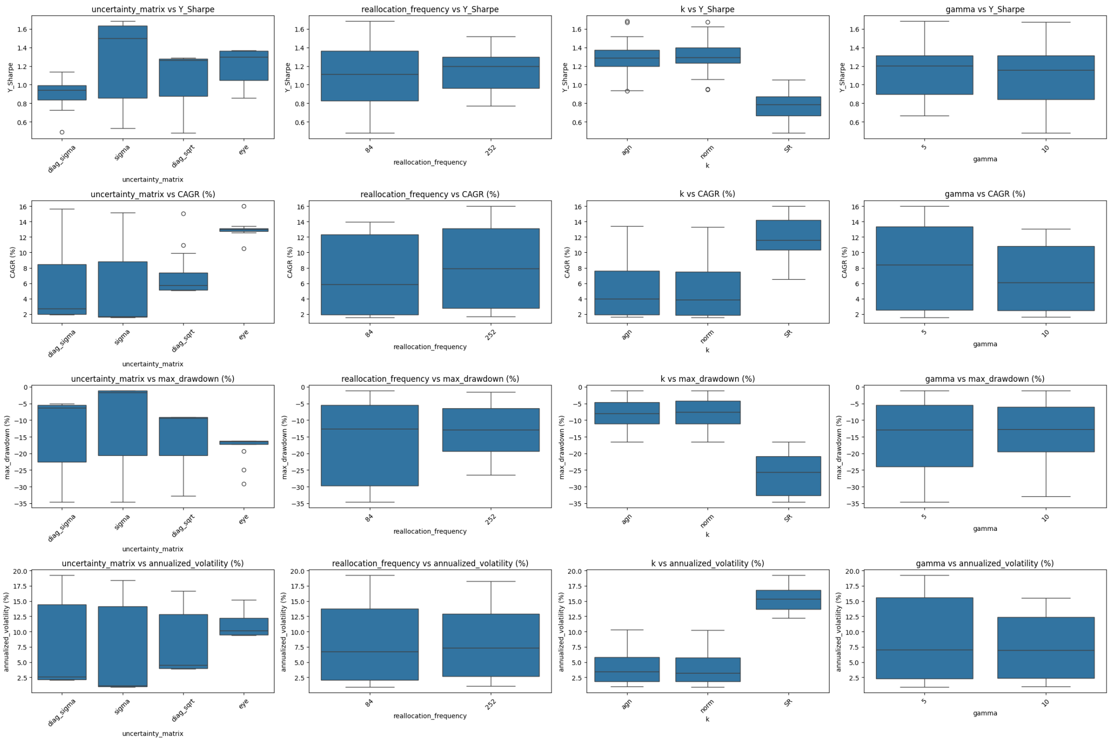


Figure 6.6: Portfolio optimization performance dispersion based on different hyper-parameters for RO-MVP optimization.

The plot also displays the cumulative return of IEF[8] fund: a portfolio constituted of 20 different U.S. treasury bonds with a maturity between 7 to 10 years (close to the analysis horizon performed). We can see that the the conservative approach, while maintaining stable returns and low volatility, it experiences a smaller growth than a bond portfolio. This might indicate an over-pessimistic approach, useful for protecting a capital invested during recession periods but too conservative for investments in a long time-horizon.
It is also worth mentioning that around 94% of all Robust Optimization combinations tested have had a similar or better performance in terms of Sharpe Ratio than both MVP optimization allocation strategies, the same proportion of simulations also achieved better results in terms of annualized volatility and 96% of all simulations presented a lower maximum drawdown. This is a crucial information as we can observe that, during both stable conditions and uncertain periods a more robust approach is key for obtaining better risk adjusted performances even without fine tuning the parameters. If one can also analyze the overall impact of each variable, then the gap becomes even wider.

[8]for detailed information, check `https://www.ishares.com/us/products/239456/ishares-710-year-treasury-bond-etf`

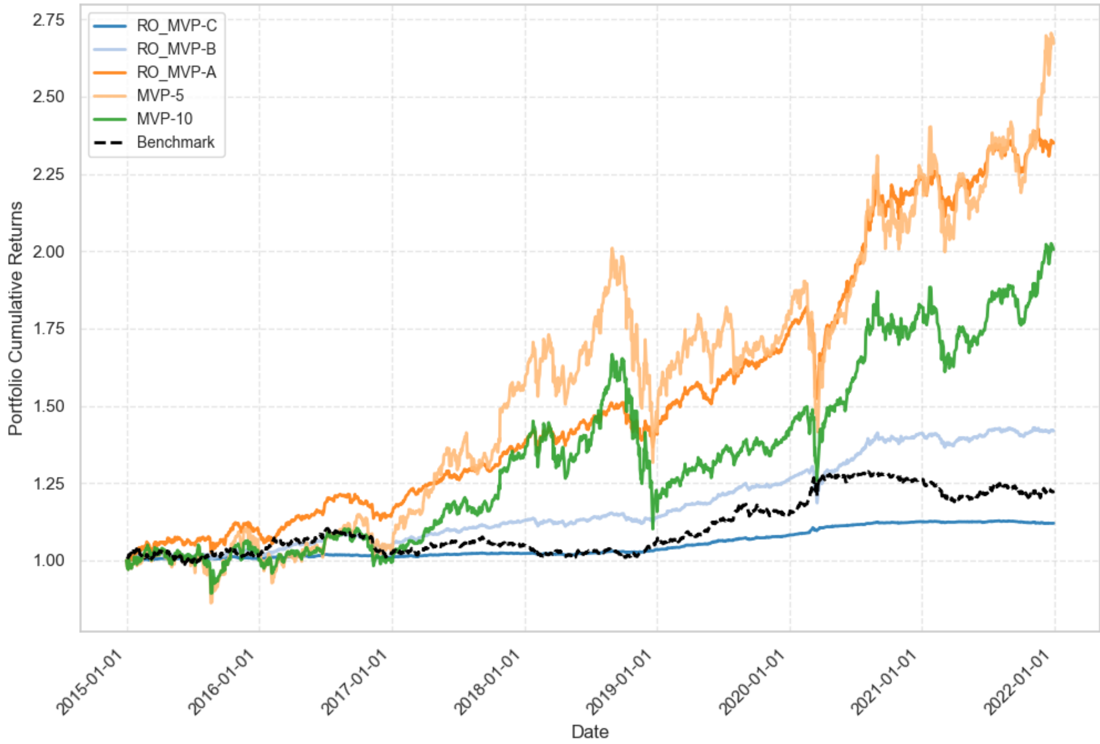


Figure 6.7: Portfolio cumulative returns for RO-MVP strategies and classical MVP optimization.

All metrics related to this analysis are reported in Table 6.5, where it is clear that classical approaches suffer from high volatility and drawdown. The resulting yearly return (compounded) does not justify the exposure to such larger losses. All risk metrics are in fact showing a substantial degradation in risk adjusted performances as, for example, the Sharpe ratio.

There is, however, a key aspect to consider: the conservative portfolio also presents the highest concentration of weights in a single asset, this indicates that in order to achieve a volatility below 1%, the portfolio have been concentrated in a single security. This might expose the method to idiosyncratic risk that has not been diversified by this allocation strategy. On the other hand, the aggressive approach results in a highly diversified portfolio very close to equal weighting. The balanced method, instead, achieves a decent tradeoff between those two extremes, maintaining a fair level of diversification (in terms of spread of weights) while differing substantially from the EWP.

Label	CAGR (%)	annualized volatility (%)	$\mathcal{SR}$	$\mathcal{TO}$ (%)	$\mathcal{SSPW}$ (%)	$\mathcal{MDD}$ (%)	$\text{CVaR}_{0.95}$
Conservative (RO-MVP-C)	1.64	**0.33**	**1.68**	**0.005**	93.41	**- 1.15**	**-0.127**
Aggressive (RO-MVP-A)	12.97	9.5	1.37	0.73	**6.7**	-16.31	-1.4
Balanced (RO-MVP-B)	5.13	4	1.29	0.70	20.08	-9.10	-0.57
MVP-5	**14.83**	22.2	0.67	7.97	50.5	- 34.50	-3.37
MVP-10	10.24	17	0.60	7.74	36.4	-33.88	-2.71

Table 6.5: Results of comparison between MVP and their Robust counterpart (the best value for each metric is highlighted in bold).

6.3 HERC-LTDC analysis and hyper-parameters choice

In this section a similar analysis to the one presented in the previous paragraph will be performed on hierarchical clustering methods, with a particular focus on lower tail-based approaches. There are multiple possibilities to be explored, given that a general hierarchical clustering method is fully characterized as described in 2.2.2. This study will mainly focus on three categories of parameters:

1. The ones related to the estimation of the similarity matrix: the estimator used for the pairwise LTDC evaluation, the threshold balancing the variance-bias tradeoff in the estimation and the window of estimation.
2. The parameters related to the allocation methods, such as the intra and across cluster weighting schemes.
3. The frequency of the reallocation.

A complete overview of the values tested can be found in Table 6.6. Once again a grid search is performed in order to evaluate the impact of each hyper-parameter as well as the interactions between multiple modeling choices.
Figure 6.8 displays the top 20 parameters combinations by Sharpe Ratio. Unlike what observed in Figure 6.4 for robust optimization portfolios, HERC-LTDC approaches can be clustered in two main groups: one featuring higher returns but at the same time non negligible drawdowns and the other performing a more conservative strategy.

Three main combinations have been chosen, however, in order to keep the same structure of the previous section. The best solution by Sharpe ratio has been chosen as the balanced option, the combination having the lowest possible volatility while maintaining a fair tradeoff with the realized returns is chosen for a conservative approach. Finally, the best combination in terms of Sharpe ratio belonging to the first group is selected as the aggressive combination. Table 6.6 summarizes these choices in terms of parameters chosen.

Hyper-Parameter	Values
LTDC-estimator	{CRS, std}
k	{0.05, 0.07}
Estimation window frame	{504, 2016}
Between cluster method (BCM)	{CDaR, CVaR, volatility}
Within cluster method (WCM)	{equal, volatility, minvar}
Reallocation frequency	{84, 252}

Table 6.6: Hyperparameters options for HERC-LTDC

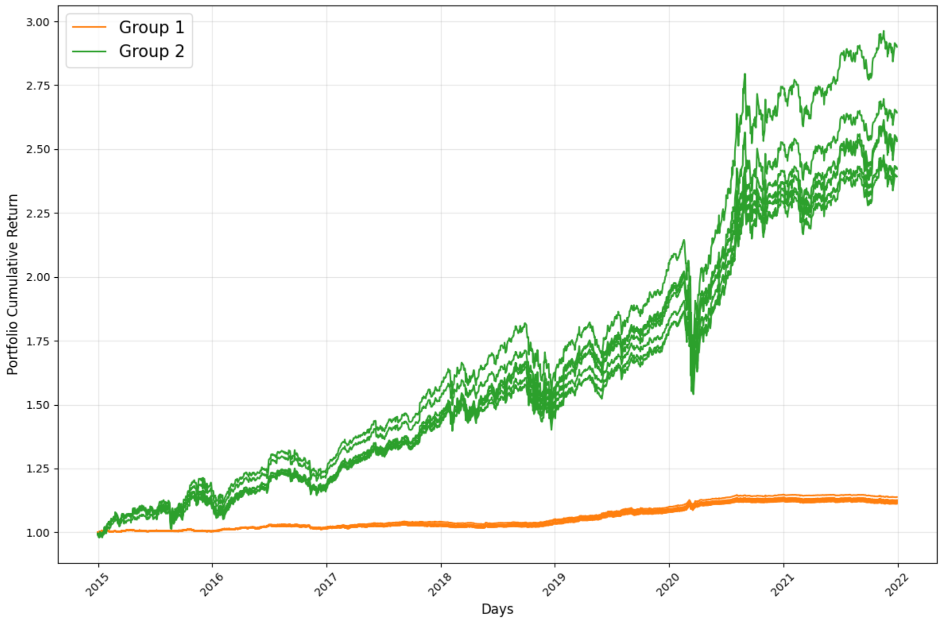


Figure 6.8: Portfolio evolution over time for top 20 hyper-parameters combinations in HERC-LTDC optimization.

Similarly to what has been done in the previous section, before diving into the comparison with the methods directly related to this approach, the influence of each parameter will be examined in detail.

Figures 6.9 and 6.10 depict a very counter-intuitive scenario: the first plot suggests that there is no parameter that, overall, tends to perform notably better than another in terms of Sharpe Ratio: there is no straightforward choice (as for the uncertainty set in the

Label	LTDC-estimator	k	Estimation wf	BCM	WCM	Reallocation frequency
Conservative (HERC-LTDC-C)	CSR	0.05	2016	volatility	minvar	252
Balanced (HERC-LTDC-B)	std	0.05	504	volatility	minvar	252
Aggressive (HERC-LTDC-A)	CSR	0.05	504	CVaR	equal	84

Table 6.7: Investment strategies chosen for HERC-LTDC approach

robust optimization analysis). The main parameters influencing the actual behavior of the portfolios constructed are both the BCM and the WCM of weighs allocation. Overall, volatility based allocation produces stable portfolios that suffer from low realized returns, while advanced risk measures produce more volatile solutions that obtain higher returns at the cost of higher expected losses. At the same time, the within cluster allocation method has also a big impact on the portfolio performance: equal weighting results in higher volatile portfolio with higher possible returns, while minimum variance allocation allows for more conservative asset allocations. Volatility-based WCM, finally, constitutes an in between solution in terms of risk/reward tradeoff.

Once again, the reallocation frequency as well as the window frame for the estimation of the similarity matrix do not play a crucial role in the optimization performance: this might indicate that the proposed method is robust in terms of the approximated parameters since their estimation does not vary with a larger amount of data or a small change in the observed input (like, for example, a shift of 4 months in the observed returns due to the rolling window-based estimation). In terms of the LTDC-estimator, the aggregated performances do not suggest to prefer one estimator over the other.

The second Figure shown below, however, highlights that analyzing the aggregated results per parameter might not be enough in this case, as there are certain parameters that tend to appear in the top 10% simulations sensibly more often than others.
This indicates that there are some types of interactions between different hyper-parameters that perform sensibly better if compared to the general performance trend. For example, volatility BCM allocation and minvar WCM are extremely more present in the top combinations, with a non-negligible gap if compared to the other solutions. From the results obtained, it is indeed the interaction between those two parameters that tend to produce highly robust portfolios particularly effective during uncertain periods.

We now proceed to study the proposed combinations of hyper-parameters by comparing them with the two main methods proposed in the literature: HRP and HERC with correlation based similarity matrix. For visualization purposes, also the MVP optimization results have been included.
To this end, the HERC portfolio has been built using the same parameters of the best performing simulation. Figure 6.11 plots the cumulative returns of each portfolio. We can clearly see that the HRP approach is comparable to the Balanced and Conservative methods, but features lower returns. Moreover, the correlation-based HERC allocation behaves in a similar way as the LTDC-based one, even though it produces slightly lower

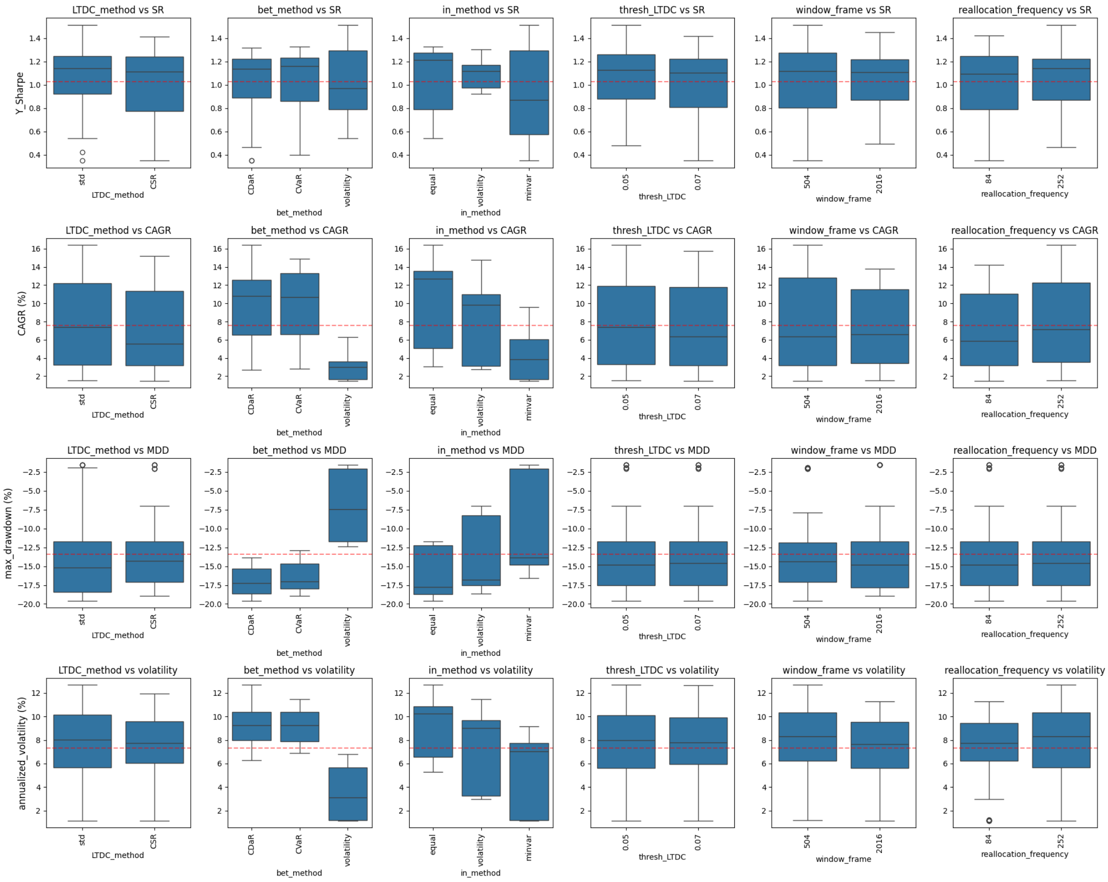


Figure 6.9: Portfolio optimization performance dispersion based on different hyper-parameters for HERC-LTDC allocation.

returns.
The aggressive approach differs significantly from the previous ones proposed, and it is more similar to what one would obtain from the classical Mean Variance Optimization portfolio in terms of final returns. It is important to note, however, how the hierarchical allocation method exhibits a much smaller volatility and drawdown risk, proving the effectiveness of this approach even during stable market condition as well as uncertain ones.
Again, all portfolios are also compared to a benchmark portfolio of bonds: we can see that this benchmark tends to oscillate more than the conservative and balanced approaches, but in the end it obtains higher returns. Once again we might think that these strategies might result too pessimistic during a long term investment, while still remaining suitable for loss hedging in the short term as for the years 2017 to 2019, when even what are considered "safe" investment classes experienced a stable recession period.

Table 6.8 reports the results of each allocation method concerning risk, returns and key performance indicators.

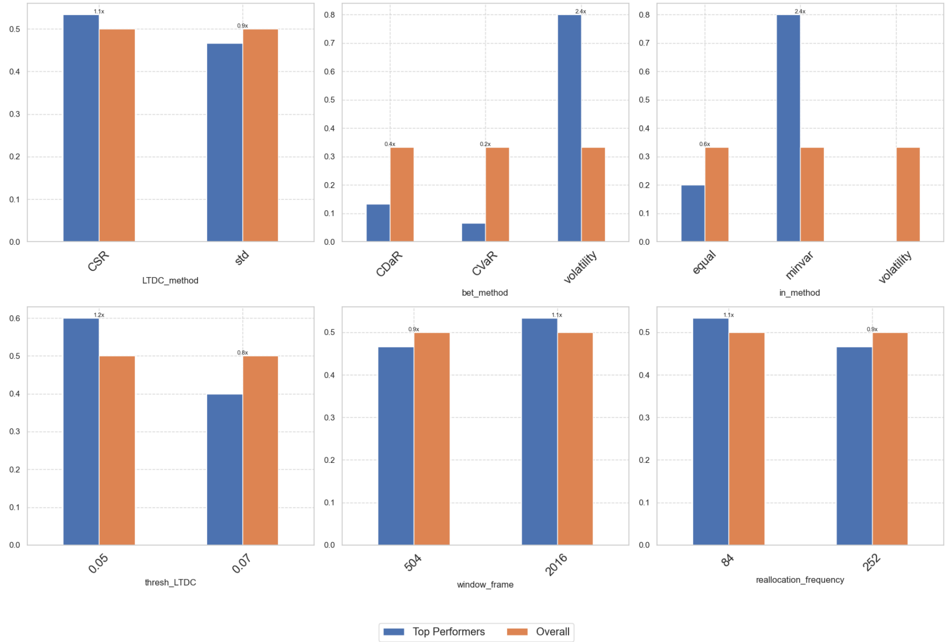


Figure 6.10: Proportion of simulations in the top 10% of results by Sharpe ratio for each parameter for HERC-LTDC simulations.

Once again, we can observe how advanced allocation methods are able to outperform classical ones on risk adjusted indicators. Moreover, tail-based allocation strategies seem to perform better than standard hierarchical allocation strategies. Both balanced and conservative methods present low volatility and drawdown risk, making them suitable for hedging possible losses or sudden market changes.
The Aggressive approach, instead, produces much higher returns but that also generate higher volatility. It is worth observing however, that the maximum drawdown and volatility is reduced by half when compared to the one of Mean Variance Portfolio, even when the yearly returns are similar.
Hierarchical methods seem more robust also from the point of view of average turnover $\mathcal{TO}$, as they reduce it at least by half if compared to the one of MVP. This means that these approaches tend to change less often the portfolio strategy as they are less sensible to inputs and noise. This is also important in terms of costs: the lower the turnover, the lower the impact of transaction fees on the portfolio value.
Similarly to what observed in the previous section, the price for smaller volatility isn't only lower expected returns: conservative methods concentrate their allocation mainly on one or two assets. This is indeed something that has to be considered as it represents a choice that is hardly humanly acceptable (low financial meaning) and inevitably carries some risks derived from the concentration of the portfolio wealth on few securities. It is worth mentioning, however, that in general those assets are indeed already somehow diversified since they are mainly ETF that incorporate multiple holdings. The main asset

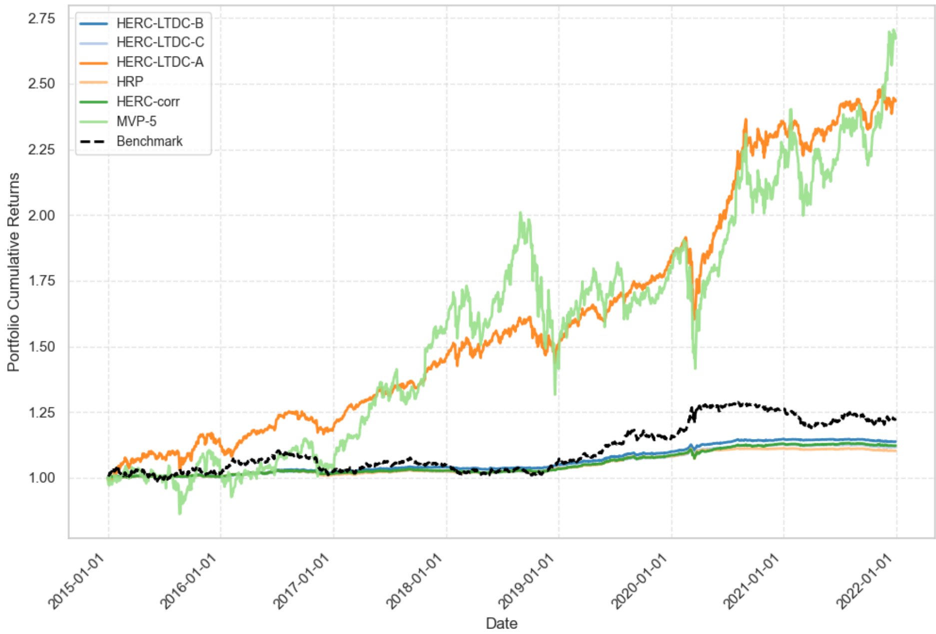


Figure 6.11: Portfolio cumulative returns for different HERC strategies and HRP allocation.

where the wealth is concentrated is SHY[9] that already contains several holdings.

Label	CAGR (%)	annualized volatility (%)	$\mathcal{SR}$	$\mathcal{TO}$ (%)	$\mathcal{SSPW}$ (%)	$\mathcal{MDD}$ (%)	$\text{CVaR}_{0.95}$
Balanced (HERC-LTDC-B)	1.87	1.23	**1.68**	1.04	80.57	-1.95	-0.171
Conservative (HERC-LTDC-C)	1.61	**1.14**	1.41	**0.50**	83.65	**-1.55**	**-0.154**
Aggressive (HERC-LTDC-A)	13.475	10.13	1.33	1.63	**7.091**	-16.31	-1.523
HERC-corr	1.66	1.36	1.21	1.19	77.79	-3.08	-0.188
HRP	1.41	1.45	0.97	1.13	75.66	-3.49	-0.186
MVP-5	**14.83**	22.2	0.67	7.97	50.5	- 34.50	-3.37

Table 6.8: Results of comparison between MVP and hierarchical allocation methods (the best value for each metric is highlighted in bold).

[9]check Table 6.1.

6.4 Real-life Investment Simulation

Once the main strategies have been outlined in the last two Sections, we now proceed on testing how these allocation approaches would have performed during the period between January 2022 and June 2025. We will compare all methods in Table 6.2 and evaluate them based on the proposed metrics.

Figure 6.12 displays the cumulative returns for the years 2022 and 2023. We can see that conservative methods, especially the ones studied in this project, are able to efficiently reduce the portfolio loss, behaving in a similar way and therefore producing similar returns. GMVP performs similarly to these approaches, proving to be a simple and valid alternative during uncertain market conditions.
Aggressive methods, instead, perform very similarly to the EWP, incurring in large losses and high oscillations. However, it should be noted that such methods present a lower recovery time compared to the benchmark and the MVP.
This leads to the belief that, while being exposed to higher risks if compared to more balanced strategies, they can adapt and effectively reallocate the capital based on the current situation, while the classical approach isn't able to recover at all and concludes this simulation period with a loss of over 17% of the starting capital.

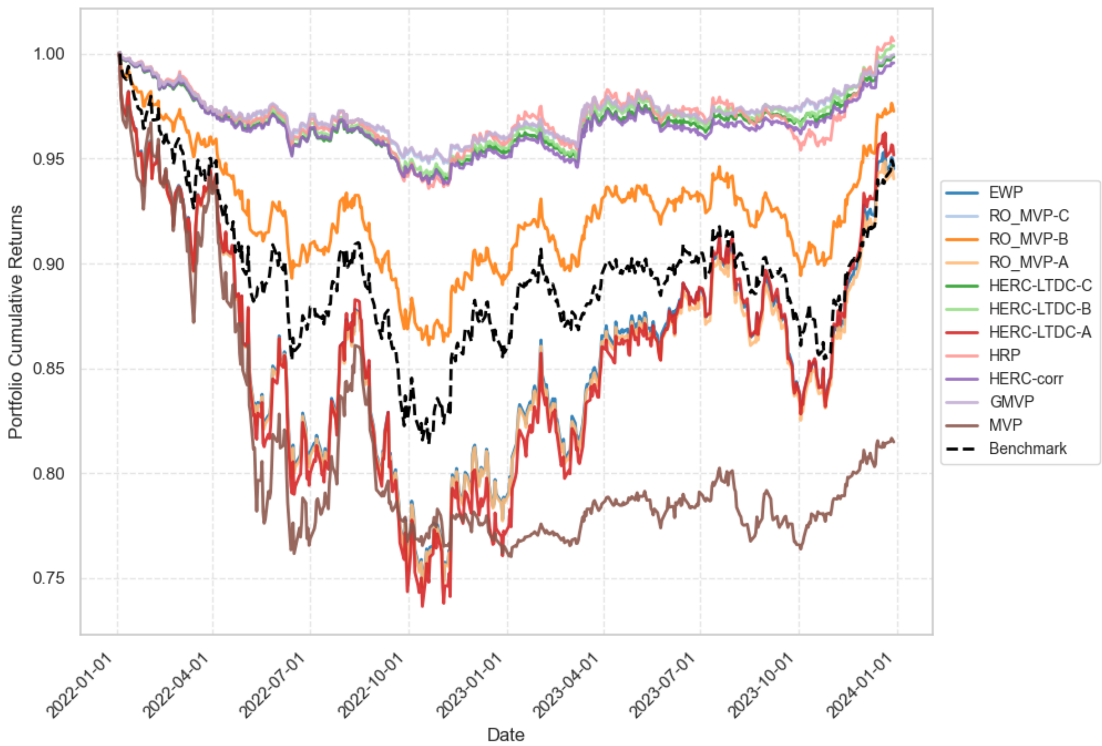


Figure 6.12: Portfolio cumulative returns, all methods between 2022 and the end of 2023.

When we move to the period between January 2024 and June 2025, instead, the situation is different. In Figure 6.13 we see that during 2024, a year of relatively upward trend, aggressive methods as well as MVP have produced higher returns. Once again, with the political and economical instability started in 2025, they then experienced large losses. Similarly to what has been observed for the years 2022 and 2023 concerning aggressive approaches, advanced allocation methods have had faster recovery time and also mitigated the overall drawdown compared to MVP.

Looking at the more conservative approaches, we can see that they tend to be more stable throughout the whole investment period. GMVP and HRP allocations, while being valid when big crises occur, they might result a bit too conservative when things are going well: in fact, they tend to under perform the other approaches by producing lower returns.
Balanced strategies -especially HERC, whether based on correlation or LTDC- are able to achieve a fair trade-off between risk and returns. They effectively limit drawdowns during the 2025 recession while still delivering strong performance throughout 2024.

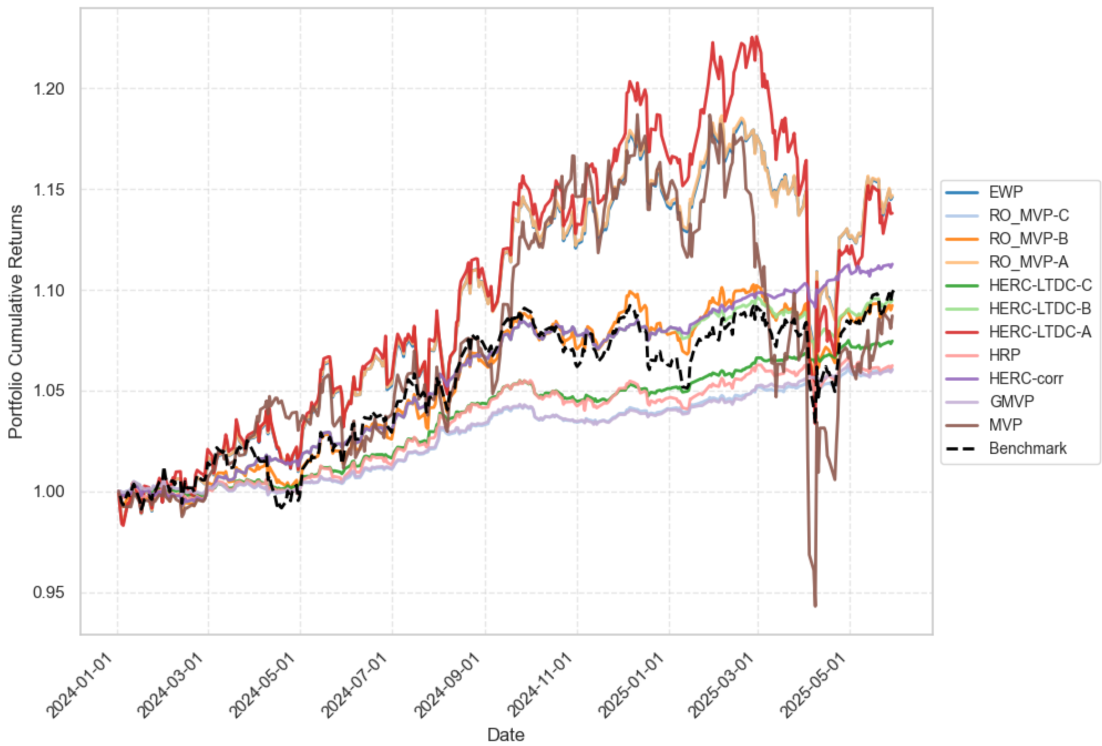


Figure 6.13: Portfolio cumulative returns, all methods between 2024 and January 2025.

Looking at the full picture described in Table 6.9, we can see that overall HERC allocation strategy has been the best in terms of Sharpe ratio. All strategies tested have strongly outperformed the classical MVP, showing all its weaknesses exposed in the first part of these thesis project. Most of the methods performed sensibly better than the benchmark proposed. This suggests that the choice of assets and the optimization methods used during uncertain periods is of great importance and still relatively unexplored.

Correlation based HERC has achieved the highest CAGR while containing the volatility and the risk associated. LTDC-based allocation (the Balanced and Conservative configurations), instead, has produced slightly lower returns with a similar volatility. However, the maximum drawdown is smaller for these last two approaches, proving that the use of the asymptotic dependency structure limits losses if compared to the standard approach.

GMVP and the conservative RO strategy have produced the lowest variance portfolios that also hedged against drawdown risk and CVaR. They behaved in a similar way: the main differences come from the $\mathcal{SSPW}$ metric, where both values are extremely high but the RO method produced a slightly more diversified portfolio. This aspect will be further analyzed later in this Section.

Finally, aggressive methods show all their criticalities during uncertain periods: they produced higher returns than conservative methods but at the price of an extremely high volatility (higher than MVP for the Aggressive HERC approach) and high risk of large losses.

Label	CAGR (%)	annualized volatility (%)	$\mathcal{SR}$	$\mathcal{TO}$ (%)	$\mathcal{SSPW}$ (%)	$\mathcal{MDD}$ (%)	$\text{CVaR}_{0.95}$
HERC-corr	**2.67**	2.9	**0.92**	4.93	78.29	-6.35	-0.37
HERC-LTDC-B	2.41	3.04	0.79	4.97	77.67	-6.06	-0.39
HERC-LTDC-C	1.70	2.5	0.68	3.98	87.66	-6.28	-0.33
GMVP	1.34	**2.26**	0.59	1.28	99.24	**-5.25**	**-0.30**
RO-MVP-C	1.34	<u>2.28</u>	0.59	1.39	97.1	<u>-5.33</u>	**-0.30**
HRP	1.56	3.21	0.49	2.33	53.23	-6.56	-0.41
RO-MVP-B	1.37	6.13	0.22	1.45	17.19	-14.15	-0.82
EWP	1.94	12.45	0.16	0.78	**6.71**	-25.56	-1.77
benchmark	1.2	8	0.15	—	—	-18.38	-1.09
RO-MVP-A	1.82	12.48	0.15	**0.68**	**6.71**	-25.65	-1.78
HERC-LTDC-A	1.91	14.6	0.13	1.70	7.28	-26.55	-2.1
MVP	-4.05	13.65	-0.3	11.52	44.92	-24.27	-2.16

Table 6.9: Results of the simulation performed between January 2022 and June 2025 (the best value for each metric is highlighted in bold, while underlined values are less than 5% far to the best value).

One key aspect for evaluating the practical use of these methods, relies on analyzing how different allocation strategies distribute the capital across the investment portfolio. Figure 6.14 displays the repartition of weights over time for each allocation method.
We can observe that the aggressive approaches highly follow the equal weighting scheme logic, resulting in diversified portfolios but exposing themselves to high oscillations.
If we consider instead more conservative approaches, we notice that the majority of

the capital is concentrated on a single asset: this is extremely evident for GMVP and RO-MVP-C, where the portfolio is basically composed by a single asset. This, as already mentioned in the previous sections, is something worth considering when building a portfolio, as the concentration of the capital in a single security (what is commonly referred to as *"putting all eggs in one basket"*) exposes the investor to different source of threat on top of market risk: concentration risk and, even if very small, the risk that the issuer (iShares in this case) fails to manage the fund properly or defaults.

When we observe the top performers, namely HERC-corr and HERC-LTDC-B, we see that for the first period they produce a highly concentrated portfolio similar to the methods mentioned before. As the simulation proceeds, they progressively increase their exposure to different asset classes in order to improve both returns and diversification. Finally, in 2025 when the market becomes unstable again, they reduce the volatility by concentrating again their capital on low volatile assets. From an economical point of view, it is worth observing that the main security after iShares 1-3 Year Treasury Bond ETF used is gold (GLD), a key component commonly used for hedging purposes.
This suggests that the information extracted from the historical data and encoded inside the dependency structures coming from the hierarchical clustering algorithms is indeed representative of reality.

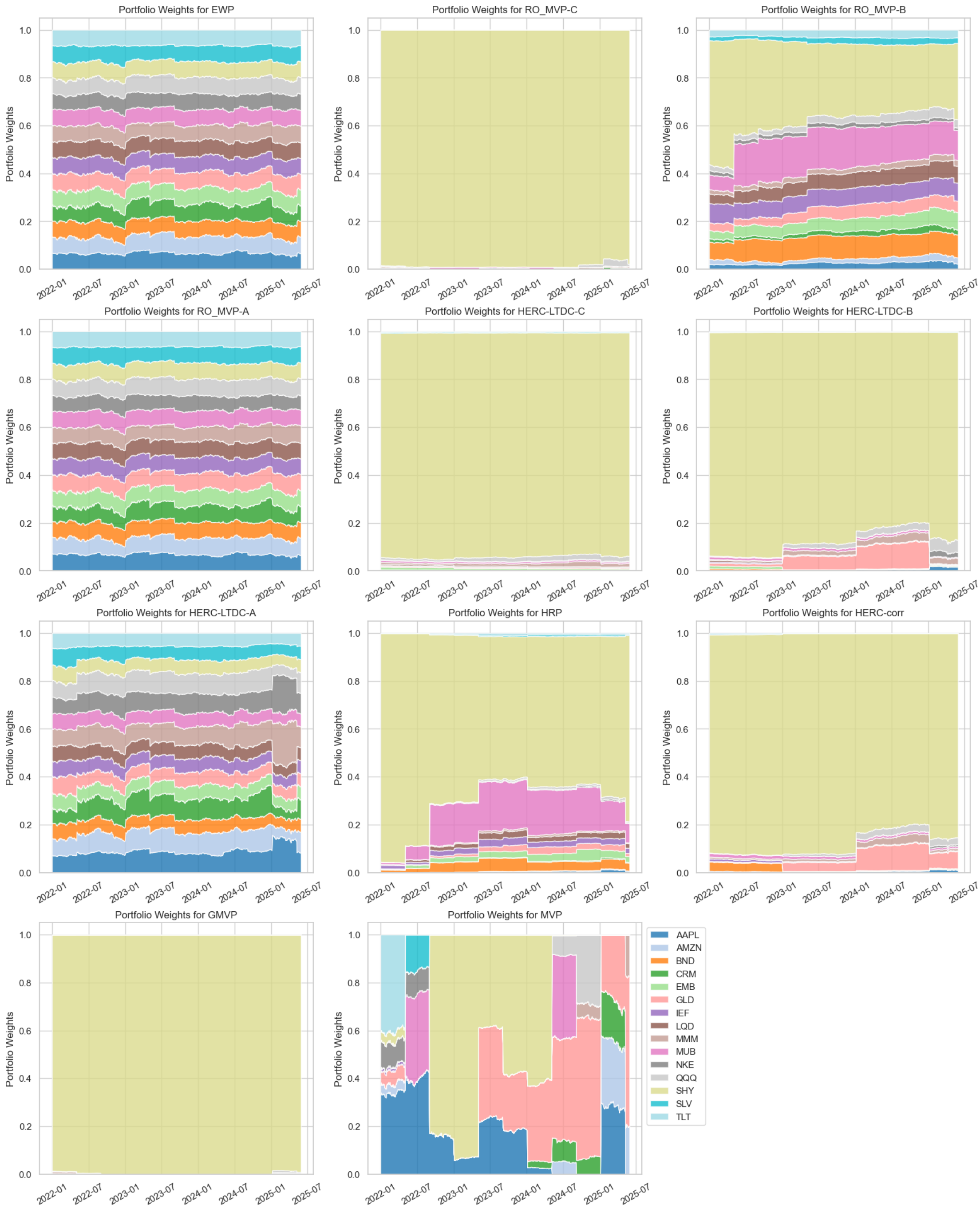


Figure 6.14: Evolution over time of weights, all methods.

If we look at the plot displaying MVP weights evolution over time, once again we have evidences supporting our initial claim: the main criticality of MVP allocation approach is its high sensibility to inputs, making the resulting portfolios particularly unstable and

variable.
This is a key component that negatively impacts both overall volatility and reallocation costs and it is in fact evident when we observe the total reallocation fees incurred by each allocation method, shown in Figure 6.15. It is clear that the MVP incurs high turnover and therefore high transaction costs if compared to the other methods. The EWP together with the RO-MVP-A incur in lower fees, while the rest of the proposed methods have similar reallocation costs.

In a practical scenario, it is quite common to characterize an asset or allocation strategy by its historical return distribution. This helps in identifying and visualizing how it is important to study how it portfolio is expected to vary from time to time. Figure 6.16 displays the distribution of monthly returns for each method. This plot helps to explain the practical differences between allocation strategies: there is a big difference between Aggressive methods, whose returns span on a wide range of values, and more Conservative ones, keeping returns more stable and within a narrower range.

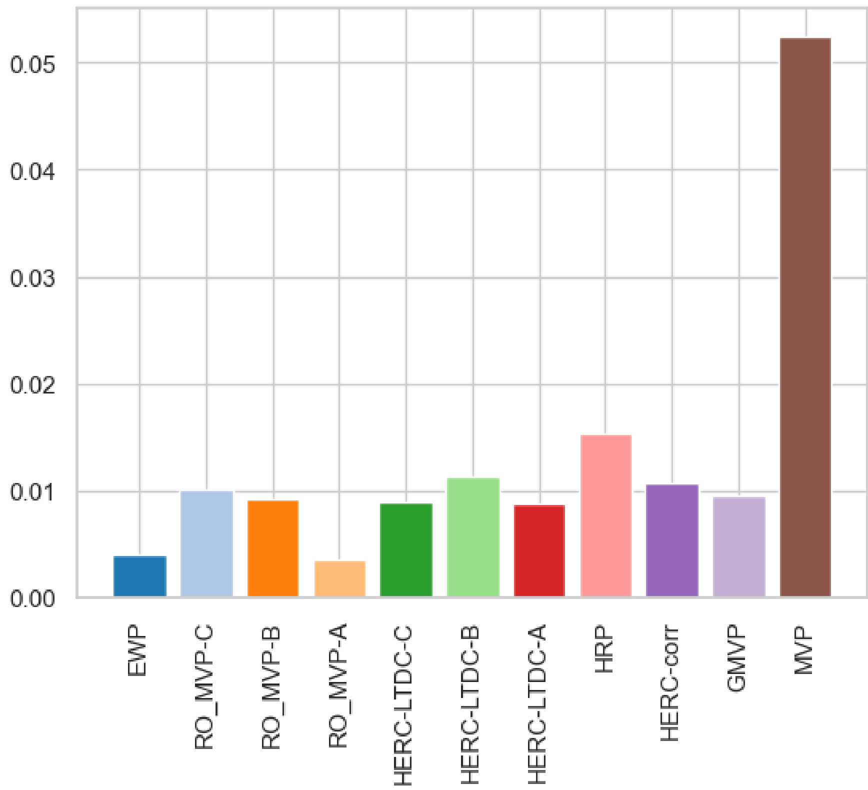


Figure 6.15: Total reallocation costs.

It is also quite interesting to notice how the benchmark, composed of two widely used ETFs in reality, behaves similarly to MVP in terms of returns dispersion, even if their mean return is notably different.

Finally, in terms of median returns we can notice how aggressive methods tend to achieve better performances even if the gap with balanced approaches is not sufficient to justify such increase in exchange for the higher level of risk.

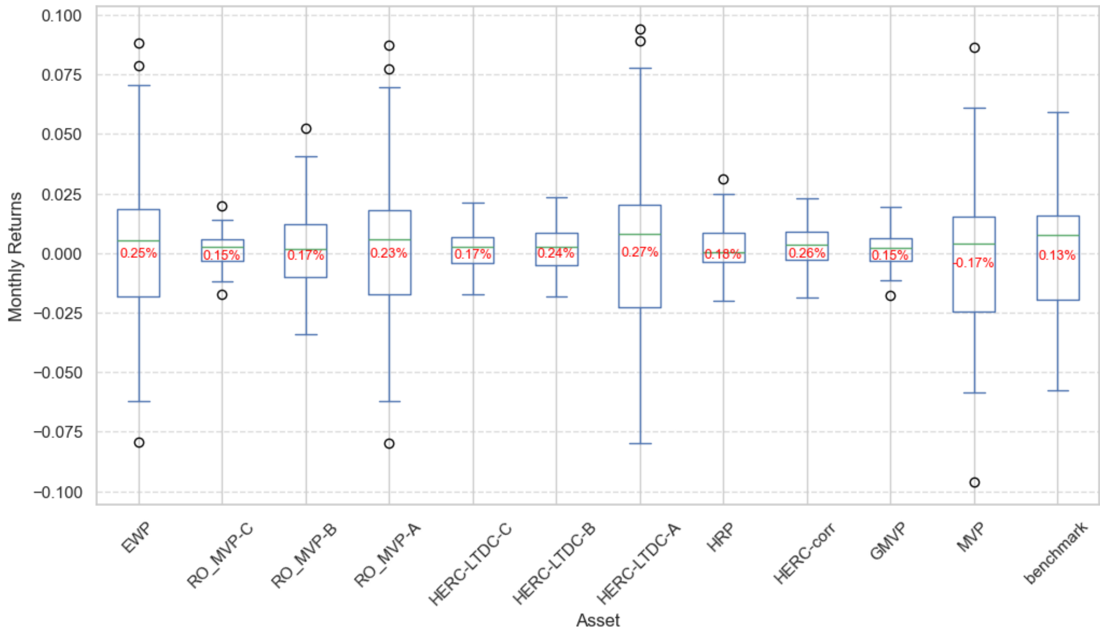


Figure 6.16: Monthly returns (mean values reported in red).

Chapter 7

Conclusions

This thesis aims to study in detail the complete decision and optimization process behind the construction and dynamic update of an asset portfolio, with a specific focus on uncertain conditions. For this reason, two main topics are analyzed and discussed: asset selection and capital allocation.

The first part of this work focuses on building a diversified portfolio and extracting insights by employing several clustering techniques on two datasets of assets. By considering the research questions and objectives stated in Section 1.1, we can affirm that these unsupervised learning methods helped unravel complex patterns that can be used for many applications, ranging from data summarization, asset analysis and, indeed, portfolio selection. This last use of clustering algorithms helps us create a well-diversified portfolio: it presents sensibly lower correlation between assets than the average correlation of the initial dataset and, at the same time, even a simple equal weighting strategy produces better risk-adjusted results when compared to the benchmark used in the simulations.

Once asset selection has been performed, the focus moves to the next step of portfolio management: capital allocation. Two main family of approaches are tested: hierarchical-based and robust optimization allocation. Although these methods may seem different from one another, they were developed from the same need: providing a stable allocation method that is less affected by noise and estimation errors. This characteristic is crucial during unstable periods, where the past performances of a security -in terms of historical returns and volatility- might not always be good indicators for future returns.

For this reason, different methods are implement and, given the recent nature of such approaches, the impact of the parameters involved is studied. The analysis outlined in Sections 6.2 and 6.3 highlights how influential the parameters choice is on the final portfolio risk and returns. This suggests that choosing the right combination based on the investor's characteristics is key even when considering robust approaches.

Finally, the results presented in Section 6.4 bring evidence supporting the view that these methods outperform the classical mean-variance portfolio as well as the benchmark employed during recession periods, providing at the same time stable performances in standard settings.

7.1 Personal Contribution

The techniques that have been implemented are mainly present in the recent literature of portfolio optimization: the hierarchical allocation based on the lower tail dependence coefficient for creating the similarity matrix had already been discussed in [12] and [13], the robust version of the mean-variance portfolio has been studied in [16] or [17]. The main contributions of this work lies in two key areas: the implementation of a general framework for hierarchical-based allocation and the analysis on the influence of the optimization parameters both for robust optimization methods and hierarchical approaches.

While [12] has performed some experiments on the use of LTDC coefficients in hierarchical settings, they did not implement it in HERC logic by introducing early stopping with the use of the gap index, as a short confrontation with the authors confirmed it. The final result is then an extension of the limitations exposed by Raffinot when he proposed the HERC allocation approach: "Last but not least, this article opens the door for further research. Typical machine learning issues have to be investigated, such as the choice of the distance metric and the criteria used to select the number of clusters."

At the same time, both approaches lacked of a comprehensive experimental analysis. This was even more evident for the robust optimization approach: [16] and [17] give different values and interpretations on the optimization parameters, without providing an explicit practical comparison on real data. For this reason, the analysis performed is crucial for assessing the impact and goodness of such parameters.

7.2 Further Research

Since this project focuses on real data experiments, one possible extension and validation approach for the proposed strategies may be to test the optimization results on simulated data to allow for multiple simulations, thus obtaining stronger evidence in statistical terms. In general, the lack of extensive simulations -especially for newly developed techniques- has been a consistent limitation in the literature, requiring more experiments both for the general performances of the methods outlined as for the role that each parameter plays in the allocation process, one of the main contributions of this thesis.

Bibliography

[1] Harry Markowitz. «Portfolio Selection». In: *Journal of Finance* 7.1 (Mar. 1952), pp. 77–91. DOI: j.1540-6261.1952.tb01525.. URL: https://ideas.repec.org/a/bla/jfinan/v7y1952i1p77-91.html (cit. on pp. 4, 5).

[2] Henrik Hult, Filip Lindskog, Ola Hammarlid, and Carl Rehn. *Risk and Portfolio Analysis: Principles and Methods.* Jan. 2012. ISBN: 978-1-4614-4102-1. DOI: 10.1007/978-1-4614-4103-8 (cit. on pp. 5, 49).

[3] Vijay Chopra and William Ziemba. «The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice». In: *Journal of Portfolio Management - J PORTFOLIO MANAGE* 19 (Dec. 1993), pp. 6–11. DOI: 10.3905/jpm.1993.409440 (cit. on p. 6).

[4] Donald Goldfarb and Garud Iyengar. «Robust Portfolio Selection Problems». In: *Math. Oper. Res.* 28 (Feb. 2003), pp. 1–38. DOI: 10.1287/moor.28.1.1.14260 (cit. on pp. 8, 23).

[5] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. «Applications of second-order cone programming». In: *Linear Algebra Appl.* 284.1–3 (1998), pp. 193–228 (cit. on p. 8).

[6] Marcos Lopez de Prado. «Building Diversified Portfolios that Outperform Out of Sample:» in: *The Journal of Portfolio Management* 42 (July 2016), pp. 59–69. DOI: 10.3905/jpm.2016.42.4.059 (cit. on pp. 9, 12).

[7] Daniel P. Palomar. *Portfolio Optimization: Theory and Application.* Cambridge University Press, 2025 (cit. on pp. 9, 11).

[8] Robert Tibshirani, Guenther Walther, and Trevor Hastie. «Estimating the Number of Clusters in a Data Set Via the Gap Statistic». In: *Journal of the Royal Statistical Society Series B* 63 (Feb. 2001), pp. 411–423. DOI: 10.1111/1467-9868.00293 (cit. on pp. 11, 16).

[9] Thomas Raffinot. «The Hierarchical Equal Risk Contribution Portfolio». In: (Aug. 2018) (cit. on pp. 12, 16, 17, 31, 48).

[10] Thomas Raffinot. «Hierarchical Clustering Based Asset Allocation». In: *The Journal of Portfolio Management* 44.2 (May 2018). Available at SSRN: https://ssrn.com/abstract=2840729 or http://dx.doi.org/10.2139/ssrn.2840729, pp. 89–99. DOI: 10.3905/jpm.2018.44.2.089 (cit. on p. 16).

[11] Alexei Chekhlov, Stan Uryasev, and MICHAEL ZABARANKIN. «Drawdown Measure in Portfolio Optimization». In: *International Journal of Theoretical and Applied Finance (IJTAF)* 08 (Jan. 2005), pp. 13–58. DOI: 10.2139/ssrn.544742 (cit. on p. 17).

[12] Harald Lohre, Carsten Rother, and Kilian Schäfer. «Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations». In: July 2020, pp. 329–368. ISBN: 9781119751182. DOI: 10.1002/9781119751182.ch9 (cit. on pp. 18–21, 74).

[13] Giovanni De Luca and Paola Zuccolotto. «A tail dependence-based dissimilarity measure for financial time series clustering». In: *Advances in Data Analysis and Classification* 5.4 (2011), pp. 323–340. DOI: 10.1007/s11634-011-0098-3 (cit. on pp. 18, 19, 21, 34, 53, 74).

[14] Matthieu Garcin and Maxime L. de Nicolas. *Nonparametric estimator of the tail dependence coefficient: balancing bias and variance.* https://hal.science/hal-03439526v4. HAL Id: hal-03439526v4. 2023 (cit. on pp. 19, 20).

[15] Friedrich Schmid and Rafael Schmidt. «Multivariate conditional versions of Spearman's rho and related measures of tail dependence». In: *Journal of Multivariate Analysis* 98.6 (2007), pp. 1123–1140. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2006.05.005. URL: https://www.sciencedirect.com/science/article/pii/S0047259X06000662 (cit. on p. 20).

[16] Chenyang Yin, Romain Perchet, and François Soupé. «A Practical Guide to Robust Portfolio Optimization». In: *SSRN Electronic Journal* (Jan. 2019). DOI: 10.2139/ssrn.3490680 (cit. on pp. 23, 26, 27, 74).

[17] Frank J. Fabozzi, Petter N. Kolm, Dessislava A. Pachamanova, and Sergio M. Focardi. «Robust Portfolio Optimization». In: *The Journal of Portfolio Management* 33.3 (2007) (cit. on pp. 23, 26, 74).

[18] Gordon Y. N. Tang. «How efficient is naive portfolio diversification? an educational note». In: *Omega* 32.2 (Apr. 2004), pp. 155–160. URL: https://ideas.repec.org/a/eee/jomega/v32y2004i2p155-160.html (cit. on p. 28).

[19] Joe H. Ward Jr. and. «Hierarchical Grouping to Optimize an Objective Function». In: *Journal of the American Statistical Association* 58.301 (1963), pp. 236–244. DOI: 10.1080/01621459.1963.10500845 (cit. on p. 31).

[20] Mansoor Momeni, Maryam Mohseni, and Mansour Soofi. «Clustering Stock Market Companies via K- Means Algorithm». In: *Kuwait chapter of Arabian Journal of Business & Management Review* 4 (2015), pp. 1–10. URL: https://api.semanticscholar.org/CorpusID:18778657 (cit. on p. 36).

[21] Peter J. Rousseeuw. «Silhouettes: A graphical aid to the interpretation and validation of cluster analysis». In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: https://www.sciencedirect.com/science/article/pii/0377042787901257 (cit. on p. 36).

[22] William M. Rand. «Objective Criteria for the Evaluation of Clustering Methods». In: *Journal of the American Statistical Association* 66.336 (1971), pp. 846–850. ISSN: 01621459, 1537274X. URL: http://www.jstor.org/stable/2284239 (visited on 04/15/2025) (cit. on p. 37).

[23] Alok Kumar. «Equity Portfolio Diversification». In: *Review of Finance* 12 (Nov. 2001). DOI: 10.1093/rof/rfn005 (cit. on p. 50).

[24] Rim El Bernoussi and Georg Michael Rockinger. *Rebalancing With Transaction Costs: Theory, Simulations, and Actual Data*. Available at SSRN: https://ssrn.com/abstract=3495461. 2019. DOI: 10.2139/ssrn.3495461. URL: https://ssrn.com/abstract=3495461 (cit. on p. 53).

[25] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. *1/N*. https://ssrn.com/abstract=911512. EFA 2006 Zurich Meetings. June 2006. DOI: 10.2139/ssrn.911512 (cit. on p. 54).

[26] Andrew Ang. *Asset Management: A Systematic Approach to Factor Investing*. OUP Catalogue 9780199959327. Oxford University Press, Dec. 2014. ISBN: ARRAY(0x82963a00). URL: https://ideas.repec.org/b/oxp/obooks/9780199959327.html (cit. on p. 55).