Journal of Contemporary Issues in Business and Government Vol. 30, No. 03, 2024
https://cibgp.com P-ISSN :2204-1990; E-ISSN: 1323-6903

QUANTITATIVE WEALTH AND INVESTMENT MANAGEMENT (QWIM) ADVANCED PORTFOLIO DIVERSIFICATION

[1]Jifan He, [2]Yipeng Li, [3]Ziyu Zhou

jh6208@nyu.edu, yl5312@nyu.edu, zz4118@nyu.edu

To Cite This Article: He, J. ., Li, Y. ., & Zhou, Z. . (2024). QUANTITATIVE WEALTH AND INVESTMENT MANAGEMENT (QWIM) ADVANCED PORTFOLIO DIVERSIFICATION. The Journal of Contemporary Issues in Business and Government, 30(3), 33–70. Retrieved from https://cibgp.com/au/index.php/1323-6903/article/view/2825

Received: 07/2024 **Published:08/2024**

ABSTRACT

Our study explores various Advanced Portfolio Diversification (APD) techniques, specifically Hierarchical Clustering (HC) with Hierarchical Principal Component Analysis (HPCA) and Dynamic Time Warping (DTW), to address the inherent estimation challenges associated with the traditional Mean-Variance (MV) Analysis framework. We find that these APD techniques significantly outperform the MV strategy in the longterm horizon across multiple risk-adjusted evaluation metrics. This superior performance is due to (1) the more diverse weight allocation of HC models and (2) the flexibility of HC models in selecting different risk measures. By utilizing advanced hierarchical clustering network approaches combined with DTW, these innovative methods refine the diversification process, mitigating most of the problems incurred by the MV framework, such as its strict assumptions and tendency to create portfolios concentrated in a few assets.

Contents

1. Introduction

Advanced Portfolio Diversification (APD) is a critical aspect of Quantitative Wealth and Investment Management (QWIM), focusing on improving the Sharpe ratio, reducing sector exposures and volatility, and mitigating skewness and tail correlations during market downturns. The Mean-Variance (MV) Analysis framework Markowitz (1989) is central to data-driven investment strategies, balancing risk against expected returns to construct optimal portfolios for efficient risk diversification without compromising expected gains. MV achieves this through an efficient frontier, representing asset combinations that optimize return for a given level of risk.

Despite its theoretical appeal, MV often underperforms in real-world applications, failing to meet expectations. It is criticized for its sensitivity to errors in estimating input parameters like expected asset returns and covariance matrices, where minor changes can drastically alter the optimal portfolio composition. This issue is compounded by the difficulties in accurately predicting returns and the need for extensive, reliable historical data, with uncertainties in returns impacting outcomes more significantly than those in the covariance matrix [Chen and Zhao (2003)]. MV's limitations include its assumption of normal asset return distributions, neglect of investor risk preferences, and a tendency to create overly concentrated portfolios vulnerable to input fluctuations [Tsiang (1989); Eichner (2008)].

This research explores APD strategies that overcome the traditional MV framework's shortcomings. By implementing sophisticated Hierarchical Clustering (HC) with Hierarchical Principal Component Analysis (HPCA) and Dynamic Time Warping (DTW), we introduce innovative approaches to portfolio optimization. These methods address common estimation challenges of MV, utilizing structured clustering and time series similarity algorithms to amend MV's weaknesses. Our comparative analysis shows that these advanced portfolio management strategies significantly enhance long-term performance and resilience, as reflected by various risk-adjusted metrics. HC models achieve this by providing broader diversification in weight allocation and greater flexibility in the choice of risk measures.

2. Literature Review

2.1. Criticism of MV Framework

The classic MV framework remains a cornerstone of investment management [Markowitz (1989)], yet it has faced increasing scrutiny. Despite its widespread use, many scholars have identified significant practical challenges with MV theory [Chen and Zhao (2003); Tsiang (1989); Eichner (2008)]. Yuan and Zhou (2022) highlight the difficulty of outperforming the naive 1/N diversification strategy, noting the biases in estimated investment strategies, particularly when the dimensionality is high relative to the sample size. Additionally, Kinlaw et al. (2023) critique the traditional reliance on correlation measures for evaluating an asset's diversification potential.

2.2. Advancements in Portfolio Diversification

In response, academic discourse has broadened to include a variety of advanced statistical methods aimed at refining diversification strategies. Notably, Chua et al. (2009) discuss fullscale optimization as a technique that potentially offers better diversification by managing conditional correlations more effectively. Similarly, Martellini and Milhau (2018) propose a factor-based framework that promises more effective measurement and management of diversification across multi-asset portfolios. Concurrently, Braga et al. (2023) introduce kurtosisbased risk metrics and a kurtosis-based risk parity strategy to distribute the responsibility for portfolio returns' dispersion more evenly among assets, providing an alternative perspective on APD.

2.3. Emergence of Hierarchical Methods

In recent decades, hierarchical methods for portfolio diversification have emerged to address the limitations of previous approaches. Pfitzinger and Katzke (2023) introduce a Convex Hierarchical Optimization framework (CHI) that aims to optimize portfolio diversification across distinct risk clusters. Fusai et al. (2020) advocate for an Equally Diversified Portfolio (EDP), laying the foundational principles for the Hierarchical Equal Risk Contribution (HERC) strategy. Building on this, Raffinot (2018) elaborates on the HERC approach, exploring its implementation and effectiveness within hierarchical asset allocation. Ibanez (2023) presents a diversified spectral portfolio using unsupervised learning methods, such as hierarchical agglomerative clustering, to achieve robust diversification. Further innovations include the Hierarchical Risk Parity (HRP) model by Jaeger et al. (2021), the HPCA strategy for modeling asset correlations by Avellaneda (2020); Serur and Avellaneda (2020), and the DTW approach by Lim and Ong (2021); Lim and Ng (2022), which offers a novel method for clustering assets based on their time-series patterns.

The maturation of HC strategies, combined with the integration of Machine Learning (ML), signifies a pivotal shift towards more sophisticated financial strategies, highlighting the sector's increasing reliance on computational power and algorithmic insights. The growing field of ML for portfolio diversification represents a significant evolution in financial strategy, driven primarily by variations in hierarchical clustering. Schwendner et al. (2021) emphasize the potential for enhanced performance through careful implementation, while noting that the definitive benefits and practicality of such approaches remain to be fully established. This discussion becomes particularly relevant in the context of ML, where Deep Reinforcement Learning (DRL), a key subset of ML, plays a crucial role in advancing portfolio diversification strategies, as extensively explored by Millea and Edalat (2022); Sood et al.

(2023). Despite DRL's considerable promise for portfolio optimization, it requires extensive computational resources and may lead to substantial initial expenses. Amidst this technological advance, the integration of advanced portfolio diversification methods with predictive analytics, particularly through Deep Learning (DL) models, has shown notable success. Ma et al. (2021) [1] validate the optimized results of advanced MV frameworks with Random Forest (RF), demonstrating the potential to significantly improve portfolio optimization. However, the main considerations for combining DL with portfolio optimization are time efficiency and the penalties during the training process, as sophisticated models utilize more resources and the machine must learn from significant penalties to learn how to make a profit.

2.4. Contributions of Our Work

Our research significantly advances portfolio diversification by innovatively applying HC families, including HPCA and DTW, to stable and long-term ETFs, overcoming the limitations of traditional MV optimization techniques. By leveraging cutting-edge advancements in structured clustering and temporal sequence models, our analysis navigates and rectifies the complexities inherent in conventional portfolio optimization. Our comprehensive comparative analysis demonstrates that these novel portfolio management strategies significantly enhance long-term performance and resilience across diverse risk-adjusted metrics. Consequently, our work constitutes a pivotal enhancement to APD strategies within the sphere of QWIM, setting new benchmarks for analytical depth and strategic innovation.

3. Models

In this section, we discussed the mathematical logic behind each model we chose. The main models we selected are the HRP model proposed by Lopez de Prado (2016) and the HERC model by Raffinot (2018), an improvement of HRP. The other models are further advancements of these two models by modifying the correlation or distance matrices in inputs [Avellaneda (2020); Lim and Ong (2021); Lim and Ng (2022)].

3.1. HC

This subsection explains the intuitions behind the HC models – HRP and HERC. Note that they are similar in the overall clustering process but different in some sub processes.

3.1.1. HRP

In 2016, Lopez de Prado (2016) unveiled the HRP algorithm, a new method for optimizing investment portfolios, for optimizing portfolios. HRP tackles previous mentioned problems

[1] Ma et al. (2021) validate the optimized results of advanced MV framework with Random Forest (RF).

of MV framework proposed by Markowitz (1989) by employing a hierarchy-based strategy and processes through three key steps:

1. HC – categorizes assets into hierarchical clusters, which happens to have the same name as the overall clustering model.
2. Quasi-Diagonalization – rearranges the covariance matrix to group similar assets closely.
3. Recursive Bisection – distributes weights among the portfolio's assets.

In addition to the benefits of demonstrating reduced sensitivity to market fluctuations and a decreased reliance on accurate predictions of asset returns, HRP optimization guarantees that within our investment strategy, assets are vying for a share of the portfolio's weight only among those within their cluster, resulting in the creation of a significantly more diversified portfolio. Another advantage of HRP is the flexibility of selection of risk measures in addition to variances of the portfolio, which will be discussed in the final step of Recursive Bisection.

- HC

Hierarchical clustering organizes our assets into groups based on specific criteria. This method ensures that assets within the same cluster are similar according to these criteria. The aim is to build a hierarchical tree that shows clusters at different levels. For example, a common way to measure similarity between data points is by using their Euclidean distance. Data points that are closer together are considered more similar. If we set a distance threshold, say 5, any data points within this distance are grouped together. To help visualize this concept, a dendrogram can be used to represent the hierarchical tree as shown in Figure 1 below:

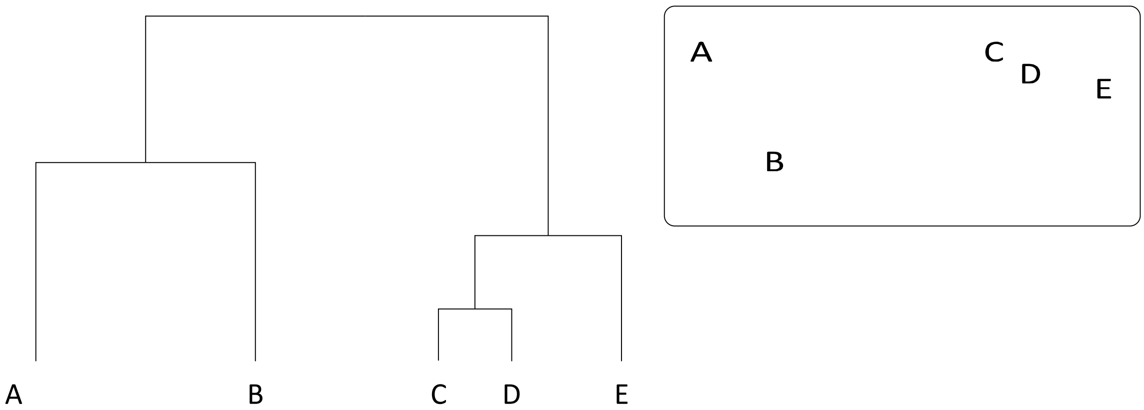


Figure 1: Dendrogram (left) and Cluster Representation (right)

First, data points C and D are the closest; thus, they are grouped together (the shorter y-distance linking C and D indicates their proximity). Next, E, being farther from the C-D cluster but closer than A or B, is grouped with C and D. Finally, A and B, being closer to each other, are grouped together. The y-distances among these points reflect their degrees of similarity. Specifically, the distance between the centroids of the A-B cluster and the C-D-E cluster is the largest, followed by the distance between A and B, then the distance between E and the centroid of C-D, and finally the distance between C and D.

Actually, we can observe that data points C, D, and E likely form a cluster, while data points A and B appear to be individual points, despite their Euclidean distances being relatively smaller than those within the C, D, E cluster (particularly between D and E). A and B can be clustered together if the threshold is set high enough, although the cluster density will be lower compared to the C, D, and E clusters.

To determine the closeness or similarity between two assets (or clusters) in the financial domain, we first need to obtain the correlations of the assets' returns, analogous to the Euclidean distances in the previous example. Given *N* assets, each with a duration of *T*, we form an $N \times T$ matrix *R* representing these assets' daily returns. After obtaining the covariance matrices, Σ, of our assets' daily returns ($N \times T$) matrix, using the formula (with minus one in the denominator because it is a sample dataset, not the whole population)

$$\Sigma = \frac{1}{T-1}(R-\bar{R})^T(R-\bar{R}) , \quad (1)$$

we can calculate the corresponding correlation matrix based on different criteria. There are various methods to calculate correlation coefficients ρ based on covariance matrices, with the most common ones being Spearman, Pearson, Kendall, and their absolute versions. Consequently, we can transform the correlation matrix into a distance matrix. Depending on the correlation method used, the corresponding distance matrix translation might vary. For the common methods—Spearman, Pearson, and Kendall—the distance matrix *D* is

$$D = \sqrt{0.5(1-\rho)}, \quad (2)$$

and for their absolute versions, the formula is

$$D = \sqrt{(1-|\rho|)}. \quad (3)$$

Based on the obtained distance matrix representing the distance between each asset, our algorithm can employ various criteria. Common methods to calculate distance or similarity include:

1. Single Linkage – the distance between two clusters is the shortest distance between any elements in each cluster.

2. Complete Linkage – the distance between two clusters is the longest distance between any elements in each cluster.

3. Average Linkage – the distance between two clusters is the average distance between elements across clusters.

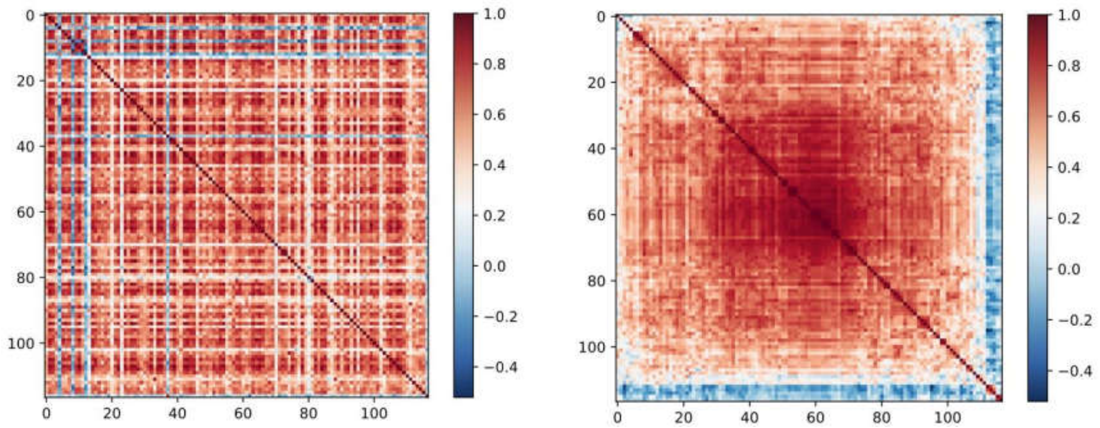
4. Ward Linkage – the distance between two clusters is the increase in the squared error when two clusters are merged.

These linkage methods are documented in the Python package *Riskfolio-Lib*[2]. For additional details on the various linkage methods available in *Riskfolio-Lib*, please refer to Appendix A.

- Quasi-Diagonalization

Following the organization of our assets into a hierarchical structure, the next phase of our process involves applying a quasi-diagonalization technique within our algorithm. Initially, we sorted our assets into a tree-like arrangement using a specific distance metric to evaluate their similarities. Now, we proceed to reorder the rows and columns of the asset covariance matrix based on the hierarchical structure obtained in Step 1 – HC. This reordering clusters more closely related assets and separates less related ones. Upon completion, the covariance matrix will be organized such that larger covariance values align along the diagonal, while smaller values are distributed around it. This resultant matrix, with off-diagonal elements closer to zero, is referred to as a quasi-diagonal covariance matrix. Figure 2 shows a template for comparing the original and quasi-diagonal covariance matrices provided by Millea and Edalat (2022).

In our analysis, we used a 1-year rolling window, advancing 1 month at a time from 1990 to 2023. Consequently, our quasi-diagonal matrices may differ each time. However, the figure below provides a blueprint for the structure of these quasi-diagonal matrices, with larger covariance values nearer to the diagonal.



[2] https://riskfolio-lib.readthedocs.io/en/latest/hcportfolio.html

Figure 2: Original correlation distance matrix (left) and after matrix seriation or quasidiagonalization (right) by Millea and Edalat (2022)

- Recursive Bisection

This subsection presents the final phase of weight distribution to the assets, leveraging the clustering performed in the preceding steps. The basic idea for recursive bisection is that we recursively calculate each cluster weight each time we bisect our dendrogram until we reach each asset (i.e., each cluster, in this case, is the assets themselves) based on our sorted covariance matrix from Step 2 - Quasi-Diagonalization.

First, we detail and explain the entire calculation process for Recursive Bisection:

1. First, we set the weights of all assets w_i to 1 (dummy variables).

2. Second, beginning at the root, we apply the following weights to each cluster C_j, j = 1,2, to calculate their volatilities, where Cov_{C_j} being the covariance matrix for cluster j, where j = 1,2[3]:

$$w_{C_j} = \frac{\mathrm{diag}(Cov_{C_j})^{-1}}{\mathrm{trace}(\mathrm{diag}(Cov_{C_j})^{-1})} \tag{4}$$

   This takes advantage of the principle that allocations based on inverse variance are the most effective when dealing with a diagonal covariance matrix.

3. Third, for each cluster C_j, j = 1,2, we calculate their newly updated corresponding variance $sigma_i$, where j = 1,2:

$$\sigma_j = w_{C_j}^T Cov_{C_j} w_{C_j} \tag{5}$$

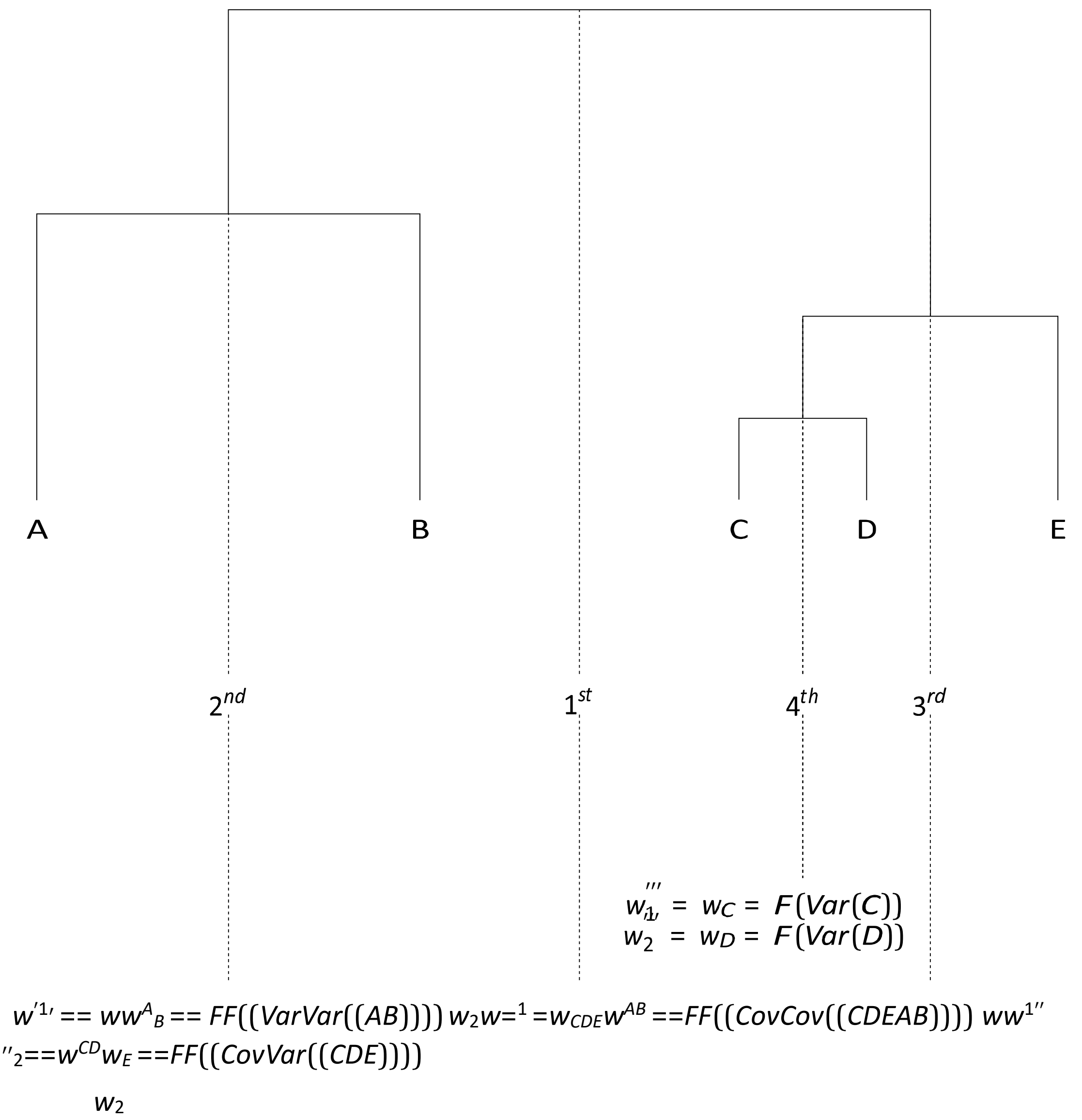
4. Finally, we update the weights and multiply each cluster weight α_j by each asset within it. Note that in this case, we only determine each cluster's weight α_j, j = 1,2:

$$w_i = \alpha_j w_i, \text{ with } \alpha_1 = 1 - \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad (6) \quad \alpha_2 = 1 - \alpha_1 \quad (7)$$

[3] We cut the tree obtained from the first step into halves each time, and the bisection rule is based on the number of assets. For instance, when we are at the root of the dendrogram tree, we cut the tree at its midpoint. So that the differences of the number of elements of each cluster each bisection time is no bigger than one, with one being the remainder effect.

We repeatedly perform Steps 2 to 4 until we reach each individual asset. It is important to note that w_{Cj} is the weight of each asset in cluster *j*, ignoring the covariance factor, while $sigma_j$ considers the covariance factor. As a result, we first calculate each asset's weight within each cluster, ignoring their interaction factors, and then for each cluster, we update the cluster's weights by considering these asset's weights and the whole cluster's covariance matrix (including off-diagonal covariance values). To elucidate this subsection comprehensively, we provide an intuitive example in Figure 3 below:

> Consider the five data points in Figure 1 as five financial assets. The dendrogram in Figure 1 is constructed based on their distance matrices, as detailed in subsection 1, rather than Euclidean distances. The entire dendrogram tree is recursively bisected until each asset is isolated. Each bisection results in two clusters, with we applying certain functions (refer to equations (4), (5), (6), and (7) for details) of each cluster's covariance matrix to determine the weight assigned to each cluster. This process is repeated until every asset is individually considered. It is crucial to note that after we reach and calculate each asset, the sum of weights equals one, and this is proven by Step 4, where $alpha_1 + alpha_2 = 1$.
>
> In Figure 3, the bisection process is clearly illustrated. We start by bisecting the top of the tree, creating two clusters: cluster 1 (assets A and B) and cluster 2 (assets C, D, and E). We then perform calculations, as shown in equations (4),
>
> (5), (6), and (7), to determine each cluster's weight. For clarity, we denote the series of calculations on each cluster's covariance matrix as capital *F* (if a cluster contains only one asset, its covariance matrix is the asset's variance). After completing the first bisection, obtaining w_1 and w_2 in Figure 3, we proceed to the second cut, targeting assets A and B separately and repeating the process. Next, we move to the second cluster from the initial cut, divide it into subclusters CD and E, and repeat the procedure. Finally, we divide the previous subcluster into sub-subclusters C and D and apply the same operations. After all bisections, we obtain a column vector $w^T = [w'_1, w'_2, w''_1, w''_2, w'''_2]$, containing the weights of all individual assets. It is crucial to initially set the weights of all assets to 1 to enable all recursions to function.



$$w_T = [w'_1, w'_2, w'''_1, w'''_2, w''_2] = [w_A, w_B, w_C, w_D, w_E]$$

Figure 3: Dendrogram for Recursive Bisection

Note that *Riskfolio-Lib* extends the risk metrics from only covariance initially to a variety of other powerful measures, such as Value at Risk (VaR), Conditional VaR (CVaR), Calmar Ratio, etc. In other words, C_i can be VaR, CVaR, or Calmar Ratio of each cluster. This further strengthens the performance power of the HRP model, with higher risk-adjusted measures in the majority of times (particularly during recent decades). However, If this is the case, equations (4), (5), (6), and (7) require slight adjustments. These adjustments will be discussed in the following model – HERC, which offers the flexibility to switch risk metrics and other performance measures, a significant advantage over HRP.

- Summarization

We now summarize the core of the HRP model as follows: First, we perform HC to create an inverted tree-like graph, or Dendrogram, to group our assets. Using this Dendrogram, we reorder the covariance matrix to group similar assets closer together and less similar ones further apart, resulting in a matrix with larger values along the diagonal and smaller values towards the edges. Finally, we recursively bisect the Dendrogram and determine each cluster's weight using the sorted covariance until we reach each individual asset.

3.1.2. HERC

Subsequent to Lopez de Prado (2016)'s work, in 2018, Raffinot (2018) introduced a novel algorithm in his paper, "The Hierarchical Equal Risk Contribution Portfolio." This algorithm, known as the HERC, combines and improves upon the machine learning techniques used in HCAA with the top-down recursive bisection approach characteristic of HRP. The HERC algorithm comprises four main stages:

1. HC – the same as the HRP model.
2. Determining optimal number of clusters – usually the default method is Gap statistic.
3. Hierarchical Recursive Bisection – the same as the HRP model except that for each time it bisects the tree graph, the HERC model aligns with the clustering results obtained in Step 1, and stops when the number of clusters k reaches the optimal number in Step 2.
4. Naive Risk Parity (NRP) – this applies the NRP method for assets in each cluster and multiplies NRP weights with cluster weights in which they reside.

Due to the fact that *Riskfolio-Lib* has already enhanced the HRP model by addressing criticisms related to its exclusive reliance on the variance metric as the sole criterion for risk assessment, HERC model mainly improves the HRP model in the following two ways:

1. Not Following the Dendrogram Structure – HRP deviates from the dendrogram structure, opting instead to bisect the tree according to the number of assets (i.e., the HRP method evenly cuts the tree each time, not considering the effect of hierarchical clustering in the first step). Figure 4 provides an excellent example of the differing bisections for these two models:

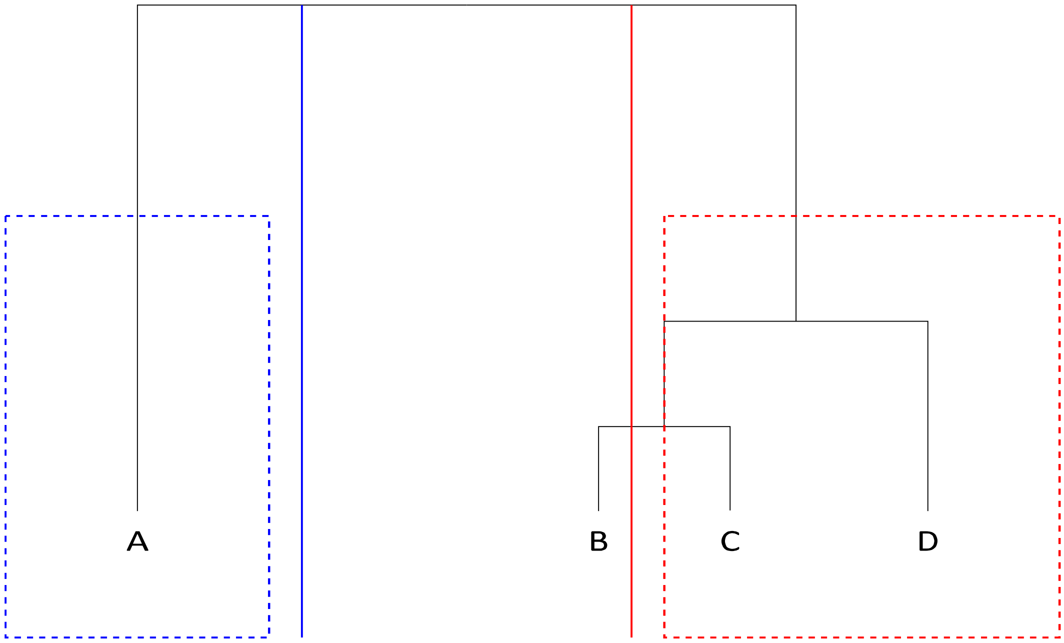


Figure 4: Dendrogram Bisections for HERC (blue left) and HRP (red right)

2. Identifying the Optimal Number of Clusters – HRP approach does not need to determine the optimal number of clusters k, but continually evenly bisects the tree until it reaches all of the individual elements in the tree. While it is convenient to skip the process of estimating the number of clusters, constructing such extensive trees slows down the algorithm significantly when dealing with very large datasets. In addition to this, allowing the tree to fully develop based on our data also creates a risk of overfitting. This means that minor inaccuracies in the data can cause substantial errors in estimating portfolio weights.

The HERC method fixes these problems of HRP by following the dendrogram structure and dives into the tree until it reaches the optimal number of clusters k derived from Step 2. Now, we only discuss the step(s) where the HERC model is significantly different from the HRP approach:

- Determining optimal number of clusters

HERC diverges from the conventional HRP algorithm at this point. Initially, the tree is fully developed to its maximum depth, after which it undergoes pruning to achieve the desired number of clusters. In this situation, we will apply the default method, which is Gap statistic. For example, imagine you have data with k clusters – $C_1, C_2, C_3, \ldots\ldots, C_k$. The sum of pairwise distances within a cluster, D_r, can be described as follows:

$$D_r = \sum_{i,j \in C_r} d_{ij} \tag{8}$$

where d_{ij} represents the Euclidean distance between two data points, i and j. From this, we compute the within-cluster sum of squares centered on the cluster means, denoted as W_k (cluster inertia).

$$W_k = \sum_{r=1}^{k} \frac{D_r}{2n_r} \tag{9}$$

where n_r represents the count of data points within the r^{th} cluster. The final step for Gap statistic is to find the optimal number of clusters k such that k maximizes the following equation:

$$\text{Gap}_n(k) = E_n^*[\log(W_k)] - \log(W_k) \tag{10}$$

where E^* denotes the expected value according to a certain reference distribution [4]. Note that we can select various methods to estimate the optimal number of clusters k, such as the Average Siloutte and Elbow methods. This can be achieved by inputting the parameter k in the *Riskfolio-Lib*'s *HCPortfolio.optimization* member function before directly applying our wanted approaches to determine the optimal cluster number k.

- Hierarchical Recursive Bisection

Having found the required number of clusters, this step calculates weights for each finally determined clusters. To illustrate this process more clearly, we refer back to Figure 3 for the 5 assets we used as a perfect example.

1. At the top of the tree, we have one big cluster and its weight is 1 (identical to Step 1 of Recursive Bisection in HRP).

2. We now descend through the dendrogram structure and successively assign weights at each level of the tree. At each point, the tree are always bisected into two sub-clusters, let's say $-C_1 = \{A,B\}$ and $C_2 = \{C,D,E\}$ – in Figure 3. The respective cluster weights, CW_{Cj}, where j = 1,2, are given by the following formulae:

$$CW_{C1} = \frac{RC_{C1}}{RC_{C1} + RC_{C2}} \tag{11}$$

$$CW_{C2} = 1 - CW_{C1} \tag{12}$$

[4] The standard selection is a uniform distribution across the data's range, presuming a structureless uniform distribution. For deeper research on the appropriate reference distribution, see Tibshirani et al. (2001).

RC_{C1} and RC_{C2} represent the risk contributions of clusters C_1 and C_2 respectively. To determine RC_{C1} and RC_{C2}, we must first calculate the risk contribution RC_i of each asset i within the clusters. Any traditional risk measure can be used to compute RC_i, with HERC currently supporting the following primary measures: Variance, Standard Deviation, Expected Shortfall (CVaR), and Conditional Drawdown at Risk (CDaR)[5]. The risk contribution RC for a cluster j is the sum of the risk contributions of all individual assets within that cluster. For instance, referring to Figure 3 and assuming $C_1 = \{A,B\}$ and $C_2 = \{C,D,E\}$, we get:

$$RC_{C1} = RC_1 + RC_2 \tag{13}$$

$$RC_{C2} = RC_3 + RC_4 + RC_5 \tag{14}$$

3. Recurse through the tree until all the clusters have been assigned weights.

- Naive Risk Parity (NRP)

The last step is to calculate the final asset weights. Continuing our previous example, let us calculate the weights of assets residing in C_2, i.e., assets 3, 4, and 5. Please note that the asset weights in C_1 follow the same approach.

1. The first step is to calculate the naive risk parity weights, W_{NRP}, which uses the inverserisk allocation to assign weights to assets in a cluster.

$$W_{NRPi} = \frac{\frac{1}{RC_i}}{\sum_{k=3}^{5} \frac{1}{RC_k}}, i \in \{3,4,5\} \tag{15}$$

2. Multiply the risk parity weights of assets with the weight of the cluster in which they reside.

$$W_{final i} = W_{NRPi} \times CW_{C2}, \quad i \in \{3,4,5\} \tag{16}$$

In this way, the final weights are calculated for assets in all the other clusters.

3.2. Improvements

This subsection discusses the modifications of HRP and HERC portfolio optimization methods: HPCA in the linear case versus DTW in the nonlinear case.

3.2.1. HPCA

The HPCA conducts the traditional PCA to each cluster we found in either an HRP or HERC situation previously. There are two main advantages of applying HPCA portfolio

[5] The *Riskfolio-Lib* Python package includes additional options. For more details, please refer to Footnote 2.

optimization, as shown by Avellaneda (2020); Serur and Avellaneda (2020). First of all, HPCA uses a parsimonious approach, meaning it does not rely on too many parameters. The user only needs to define the number of clusters or eigenvectors, which simplifies the model complexity. Second, they demonstrated that HPCA performs well across various markets (US, Europe, China, and Emerging Markets), indicating its robustness and adaptability. Recognizing this, it is a good idea to combine the PCA model with HC algorithms. First of all, it is important for us to have a quick review of PCA.

- PCA Revisited

PCA is a method used to reduce the dimensionality of data through linear transformations, making it valuable for exploratory data analysis, visualization, and data preprocessing. This technique reorients the original data into a new coordinate system where the axes (principal components) that exhibit the greatest variance are clearly distinguishable. Assuming our portfolio has a total of N individual assets, the mathematical model for implementing PCA for asset allocations is the Singular Value Decomposition (SVD) of the correlation matrix of the assets' daily returns, C:

$$C = UDV^T \tag{17}$$

where U, D, and V all being $N \times N$ matrices, D a diagonal matrix in addition to the square matrix, columns of U eigenvectors, and diagonal entries of D eigenvalues. Note that these matrices are in decreasing order, meaning the largest eignvalue is always the first diagonal entry in D, and the first column of U corresponds to the largest eignvalue, the second column of U corresponds to the second largest eignvalue, and so on. When the covariance matrix is positive definite [6], $V = U$, so we obtain:

$$C = UDU^T \tag{18}$$

The eigenvectors as columns obtained in U represent uncorrelated features of the original dataset, which could be correlation (or covariance) matrices in our case, and the eigenvalues as diagonal entries in D are the amount of information retained by each feature.

Although we do not apply the final step of PCA in our asset allocations, this step involves transforming the original dataset by multiplying the matrix of assets' daily returns R by the sub-matrix of$U$ containing the first k eigenvectors$U_{[:,1:k]}$ and naming the result of this matrix multiplication as R^*_k:

$$R^*_k = R \cdot U_{[:,1:k]} \tag{19}$$

[6] There are many definitions for positive semidefinite/definite matrix, but the most common one is: $x^T Sx \geq 0$ for any nonzero $n \times 1$ vector x, with = being the definition of positive semidefinite matrix.

- Correlation Update

After reviewing the concept of PCA, now it is the time to adapt to the HPCA scenario. As we indicated before, all of the improvement of HPCA (and DTW as well) is to update the covariance matrices, or more specifically, the correlation matrices. We first need to obtain our first eigenportfolio F_1 (i.e., the eigenportfolio [7] linked to the leading principal component) for each cluster j after finishing the hierarchical clustering process in either HRP or HERC case.

The formula for the first eigenportfolio for each cluster j, F_1^j, is:

$$F_1^j = \frac{1}{\sqrt{\lambda_1^j}} \sum_{i=1}^{N_j} \left(\mathrm{v}_{1,i}^j \circ R_i^j \right) \tag{20}$$

where λ_1^j is the first (i.e., (1,1)-st, containing the most significant information of assets' correlations) entry of the diagonal matrix D from the SVD of the correlation matrix of asset daily returns in cluster j, C^j, $\mathrm{v}_{1,i}^j$ (for i=1,2,…,N_j) is the i-th element of the first eigenvector corresponding to λ_1, $\circ$ denotes the element-wise multiplication (Hadamard product), and R_i^j is the matrix of the asset i's daily returns in cluster j. Note that $\sum_{i=1}^{N_j} \left(\mathbf{v}_{1,i}^j \circ R_i^j \right)$ effectively collapses the $N \times T$ matrix $\mathrm{v}_{1,i}^j \circ R_i^j$ into a T-element array by summing across the N_j tickers in cluster j for each time period t = 1,2,…,T. To better understand how equation (20) works, it is better to provide an example:

Let us assume we have the following daily returns $R^{(2,3)}$ for the two assets over three time periods:

$$R^{(2,3)} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{pmatrix} = \begin{pmatrix} 0.01 & 0.03 & -0.02 \\ 0.02 & 0.01 & 0.01 \end{pmatrix} \tag{21}$$

To calculate the corresponding first eigenportfolio $F_1^{(2,3)}$ of matrix $R^{(2,3)}$, we need to obtain the correlation matrix of $R^{(2,3)}$ using equation (1) and either Spearman, Pearson, or Kendall method. Further, let us assume the correlation matrix of $R^{(2,3)}$, $C^{(2,3)}$ is as follows:

[7] *ith* eigenportfolio is the portfolio created by the combinations of assets capturing the *ith* significant co-movement in their returns, adjusting the eigenvectors of the covariance matrix to different scales.

$$C = \begin{pmatrix} 1 & 0.53 \\ 0.53 & 1 \end{pmatrix} \tag{22}$$

And we can then obtain the first eigenvalue $v_1^{(2,3)}$ and eigenvalue $\lambda_1^{(2,3)}$ correspondingly:

$$\lambda_1^{(2,3)} = 1.53, \quad v_1^{(2,3)} = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix} \tag{23}$$

Then, we compute the element-wise multiplication and summation $v_1^j \circ R^j$:

$$\sum_{i=1}^{N_j} (\mathbf{v}_{1,i}^{(2,3)} \circ R^{(2,3)}) = (\mathbf{v}_{1,1}^{(2,3)} \cdot R_1^{(2,3)^T}) + (\mathbf{v}_{1,2}^{(2,3)} \cdot R_2^{(2,3)^T}) \tag{24}$$

where $R_1^{(2,3)} = [0.01, 0.03, -0.02]$ and $R_2^{(2,3)} = [0.02, 0.01, 0.01]$, and $v_{1,i}^{(2,3)} = 0.71$ for i = 1,2. Arriving at this step, we can finally compute $F_1^{(2,3)}$:

$$F_1^{(2,3)} = \frac{1}{\sqrt{\lambda_1^{(2,3)}}} \sum_{i=1}^{2} (\mathbf{v}_{1,i}^{(2,3)} \circ R_i^{(2,3)}) = \left(\frac{1}{\sqrt{1.53}}\right) \left[(0.71 \cdot R_1^{(2,3)^T}) + (0.71 \cdot R_2^{(2,3)^T})\right] \tag{25}$$

So, the $F_1^{(2,3)}$ that has been finally calculated is:

$$F_1^{(2,3)} = \frac{1}{\sqrt{1.53}} \begin{pmatrix} 0.71 \cdot 0.01 + 0.71 \cdot 0.02 \\ 0.71 \cdot 0.03 + 0.71 \cdot 0.01 \\ 0.71 \cdot (-0.02) + 0.71 \cdot 0.01 \end{pmatrix} = \begin{pmatrix} 0.0172 \\ 0.0229 \\ -0.0057 \end{pmatrix} \tag{26}$$

which is a T by 1 column vector, where T in our case is 3.

After obtaining the principal eigenportfolio F_1^j, we perform an ordinary linear regression of each asset i's daily returns in its corresponding cluster j, R_i^j, on F_1^j, ignoring the intercept term:

$$R_i^j = \beta_i^j F_1^j + \varepsilon_i^j, \qquad i = 1,2,\dots,N \tag{27}$$

where β_i^j and ε_i^j are the coefficient and the residuals in this linear regression. Assuming two assets i and k, and the clusters $j1$ and $j2$ each of them belongs to, the corresponding final (i,k)-th entry of the HPCA correlation matrix $\tilde{R}_{ik}$ is defined by:

1. If $j1 = j2$, $\tilde{R}_{ik} = R_{ik}$,
2. If $j1 \neq j2$, $\tilde{R}_{ik} = \beta_i^{j1}\beta_k^{j2}\mathrm{Corr}(F_1^{j1}, F_1^{j2})$.

in which $\mathrm{Corr}(F_1^{j1}, F_1^{j2})$ is the correlation coefficient between the first eigenportfolio of cluster $j1$, F_1^{j1}, and that of cluster $j2$, F_1^{j2}. In other words, we only update the correlation entries where the two assets it corresponds to are from different clusters. When the two assets are from the same cluster, we just left what the entry is – the entry obtained by the previously specified correlation method transforming the covariance matrix to the correlation matrix [8].

Finally, it's crucial to use a consistent method for calculating correlation coefficients. If we initially used the Pearson method, we must use it again to update the correlation matrix whenever two assets are in different clusters.

3.2.2. DTW

Combining DTW with HC optimization is the shining part of this research. DTW is a powerful algorithmic tool designed to find an optimal alignment between two given temporal sequences in a nonlinear case. In other words, DTW finds the minimum cost of alignment of a pair of time series. It is adept at measuring the similarity between sequences that may vary in speed or duration due to phase shifts or different operation rates. In finance, where market data are often non-linear and out of phase, DTW becomes instrumental in aligning time series data for better comparison and analysis.

DTW accomplishes this by 'warping' the time dimension of sequences to enable a point-bypoint comparison, essentially stretching or compressing the sequences as needed to identify patterns that would be missed by conventional linear analysis. This technique is invaluable in comparing financial time series data such as stock prices, trading volumes, or economic indicators that do not align perfectly over time due to various market conditions or external events affecting the assets differently. The essence of DTW is that is a dynamic programming approach updating the corresponding distance of two time series at each time step by its previous value adding the minimum adjustment to the current time step.

The formula of DTW in two-time series $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_m)$ is summarized as follows [9]:

$$\mathrm{DTW}(x, y) = \min_{\pi} \sqrt{\sum_{(i,j)\in\pi} d(x_i, y_j)^2} \tag{28}$$

where $\pi = [\pi_0, \ldots, \pi_K]$ is a path satisfying the following properties:

[8] The HPCA assumption is that for two assets i and k, $corr(\varepsilon_i, \varepsilon_k) = 0$ when $cluster(i) \neq cluster(k)$. While we did not visualize the plotting of residuals in this situation, the similar (in fact, slightly better) performance of HPCA version of HRP model compared with the traditional HRP model should convince the legitimacy of this assumption.

[9] https://tslearn.readthedocs.io/en/stable/user_guide/dtw.html

1. it is a list of index pairs $\pi_k = (i_k, j_k)$ with $0 \le i_k < n$ and $0 \le j_k < m$
2. $\pi_0 = (0,0)$ and $\pi_K = (n-1, m-1)$
3. for all $k > 0$, $\pi_k = (i_k, j_k)$ is related to $\pi_{k-1} = (i_{k-1}, j_{k-1})$ as follows:
   - $i_{k-1} \le i_k \le i_{k-1} + 1$
   - $j_{k-1} \le j_k \le j_{k-1} + 1$

In this context, a path represents the synchronization of time series data in a way that the Euclidean distance between the corresponding (that is, resampled) data points of the time series is as small as possible. In our situation, we calculated the DTW distance matrix based on the formula above, rescaled it from a range of 0 to 1, converted it to the correlation and thus the covariance matrices, and performed the Hierarchical Clustering model (i.e., HRP and HERC) based on this updated covariance matrix input. Feel free to refer to Appendix B for the detailed algorithm implementation for calculating the DTW distance matrix.

While the above formula seems complicated to understand and calculate, it has the following equivalent formula equivalently:

$$DTW[i,j] = \text{Dist}(i,j) + \min(DTW[i-1,j], DTW[i,j-1], DTW[i-1,j-1]) \quad (29)$$

with

$$DTW[0,0] = 0, \qquad DTW[0,i] = DTW[j,0] = \infty \text{ for } i = 1,2,\ldots,n \text{ and } j = 1,2,\ldots,m \quad (30)$$

A quick example is the continued example from equation (21). When we want to calculate the DTW distance, $DTW[1,1]$ for r_{11} and r_{21}, we first calculate the distance between r_{11} and r_{21}:

$$\text{Dist}[1,1] = (X[0]-Y[0])^2 = (0.01-0.02)^2 = 0.0001 \quad (31)$$

Then we calculate the corresponding $DTW[1,1]$ using equations (29) and (30):

$$DTW[1,1] = \text{Dist}[1,1] + \min(DTW[0,1], DTW[1,0], DTW[0,0]) \quad (32)$$

$$= 0.0001 + \min(\infty, \infty, 0) = 0.0001 \quad (33)$$

When it comes to APD, DTW is particularly useful for synchronizing disparate financial time series data. This allows investors to recognize leading and lagging indicators across different market segments. It also helps identify temporal relationships and correlations that are not immediately apparent, enabling portfolio managers to anticipate and adjust to market dynamics proactively.

Moreover, DTW can enhance risk management by detecting when assets or strategies begin to deviate from their historical patterns, signaling potential shifts in market trends or emerging risks. It can also contribute to better tailoring of the investment strategy to specific market regimes by recognizing the periods where particular strategies outperform others.

By integrating DTW into APD strategies, investment models can adapt to the market's rhythmic fluctuations and identify opportunities that a static model might overlook. This dynamic approach allows for a more responsive portfolio that can adjust in near-real-time to changing market conditions, mitigating risks and capturing growth more effectively, which is crucial for the construction of a robust and adaptable investment portfolio.

4. Implementations

4.1. Data Collections

The assets of our portfolios include a mixture of Indexes and ETFs. We carefully selected our datasets for the following reasons:

1. Act as dependable indicators for the most significant classes of assets and their subsets.
2. Be easily accessible for widespread use.
3. Exhibit a high degree of liquidity.
4. Provide data points on a daily basis.
5. Include a wide array of market environments, ideally dating back to the early 1990s.
6. Possess advantageous statistical qualities that facilitate easier modeling. In this regard, financial indices are generally preferable due to their more desirable statistical attributes when compared to the time series of individual stocks or bonds.

Based on these criteria, we finally selected 14 ideal Indexes and ETFs from Bloomberg for the purpose of research, as shown in Table 1:

Table 1: Daily Data Sets

Name	Description
BCOMTR	Bloomberg Commodity Index Total Return
LBUSTRUU	Bloomberg Barclays US Aggregate Bond Index
RU20INTR	Russell 2000 Total Return
S5COND	S&P 500 Consumer Discretionary Index
S5CONS	S&P 500 Consumer Staples Index
S5ENRS	S&P 500 Energy Index
S5FINL	S&P 500 Financials Sector GICS Level 1 Index
S5HLTH	S&P 500 Health Care Index
S5INDU	S&P 500 Industrials Index
S5INFT	S&P 500 Information Technology Index

S5MATR	S&P 500 Materials Index
S5TELS	S&P 500 Communication Services Index
S5UTIL	S&P 500 Utilities Index
SPXT	Proshares S&P 500 EX Technology ETF

In addition to the daily returns dataset, we also considered the importance of interest rates in constructing our portfolio. After careful selection, we chose the monthly data on the ”1-Year Treasury Constant Maturity Rate” because the rolling window is approximately 252 trading days (i.e., 1-year data), and we moved this window forward by 1 month. This data is available from Federal Reserve Economic Data[10].

4.2. Backtesting

We utilized a rolling window approach with the previous year's daily returns data to estimate portfolio weights, advancing by one month at a time from January 1st, 1990, to December 31st, 2022. This method involved using the past year's data to generate covariance, correlation, and distance matrices for the modified HC portfolio diversification method. Our portfolio, containing the research targets, was updated monthly. Besides the standard cumulative and daily returns performance metrics, we evaluated a range of additional criteria: Drawdowns, Maximum Drawdowns (Max DD), Annualized Returns, Annualized Excess Returns, Annualized Volatilities, Sharpe ratios, Sortino ratios, Calmar ratios, Historical VaR, and Historical CVaR.

For implementation, we directly used *Riskfolio-Lib* and *tslearn* for HC and DTW cases. For the HPCA model, we manually created a class named *HPCA* to generate updated correlation and covariance matrices, following the guidelines in Section 3.2.1. to obtain the final HPCA version. We excluded transaction fees in our monthly rebalancing to avoid introducing noise into our analysis.

5. Results

The results of the modified HC model are promising: in most cases, especially in recent decades, our proposed portfolio outperforms the traditional MV portfolio. After experimenting with various correlation methods, risk metrics, and linkage methods, we found that the absolute Kendall codependence method for the traditional HC model, Drawdowns for the risk metrics, and DBHT clustering for the linkage method [11] yielded the best results overall. Specifically, the optimal risk metrics for each model are: the Calmar Ratio for the classical HC model (i.e., HRP and HERC) and its DTW version (for stable performance), Average Drawdown of compounded cumulative returns for the HPCA version of the HRP method, and Relativistic Drawdown at Risk (RLDaR) for the DTW case (for high-return

[10] https://fred.stlouisfed.org/

[11] The DBHT linkage method only fails with the HPCA version of the HRP case. In this situation, the complete linkage method performs the best.

performance) [12]. These superior performances are due to the HC models' inherent advantages of more diversified investments and versatile risk metrics.

To better interpret our results, we divide our discussion into two subsections: Portfolio Returns and Other Performance Metrics. The first subsection includes three sub-subsections: Cumulative Returns Analysis, Daily Returns Analysis, and Drawdown Analysis.

5.1. Portfolio Returns

5.1.1. Cumulative Returns Analysis

Figure 5 elucidates the temporal progression of cumulative returns, delineating the ascent of investment values from the early stages of the 1990s to the peak of 2024. The graph serves as a synoptic display where each trajectory articulates the empirical yield of distinct investment strategies, against the "Base Portfolio" which anchors the normative benchmark.

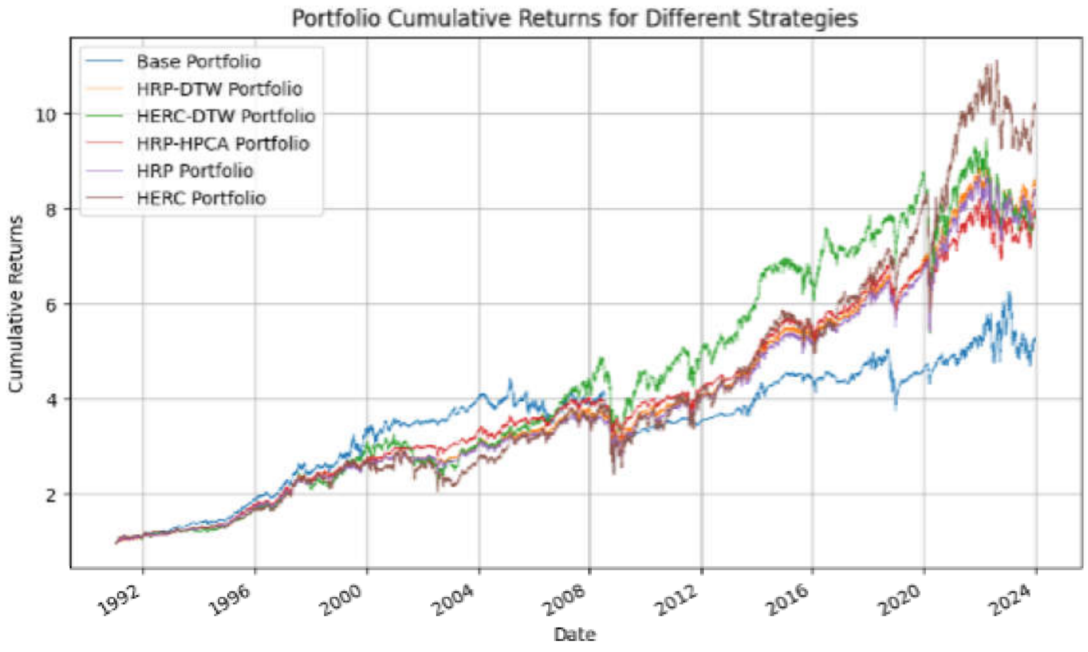


Figure 5: Cumulative Returns of Stability-Focused Portfolio Strategies

In this analytical framework, "Stable Solutions" carve out a path marked by decidedly less volatility when juxtaposed with traditional MV-optimized portfolios. This less tumultuous trajectory embodies a calculated approach to growth, reflective of an overarching strategic resilience to macroeconomic shifts and a keen navigation through the undercurrents of market inefficiencies. When set against the MV baseline, this discourse amplifies the intrinsic

[12] High-return performance requires more running time, occasionally fails to find an optimal solution, and may incur warnings of inaccurate estimates

aptitude of the APD framework for long-term capital accrual, underpinning its robustness and functional efficacy.

Moreover, the juxtaposition of different solution sets within the "Stable Solutions" suite, such as the HRP and HERC strategies, manifests a tempered volatility when contrasted with the DTW-enhanced equivalents. These strategies are anticipated to mirror market vicissitudes, their adaptive responsiveness illuminated during pivotal periods such as the turn of the millennium, the aftermath of the 2008 financial turmoil, and during the COVID-19 pandemic. In comparison with the "Base Portfolio," the APD strategies have exhibited adept maneuvering through economic tumult, marked by pronounced devaluations swiftly followed by rebounds, indicative of efficacious capital preservation and risk management protocols.

The HERC portfolio strategy, particularly post-2024, emerges as a paragon of resilience and strategic acuity, encapsulating the tenets of APD's investment philosophy. This portfolio, in tandem with its DTW-enhanced counterpart, has shown commendable economic recovery post-crisis. It maintained stability amidst the pandemic-induced financial oscillations, underscoring the portfolio's capacity for withstanding excessive market volatility.

Figure 6's depiction of the "High-Return Solutions" suggests a trajectory of marked ascensions, embodying the venturesome spirit intrinsic to such strategies. The aggressive inclines, especially noted in the HERC Portfolio, signal a more ambitious capital growth approach, potentially indicative of dynamic asset allocation strategies, which might include leveraging, derivatives trading, or high-frequency trading—practices generally associated with higher risk and reward prospects.

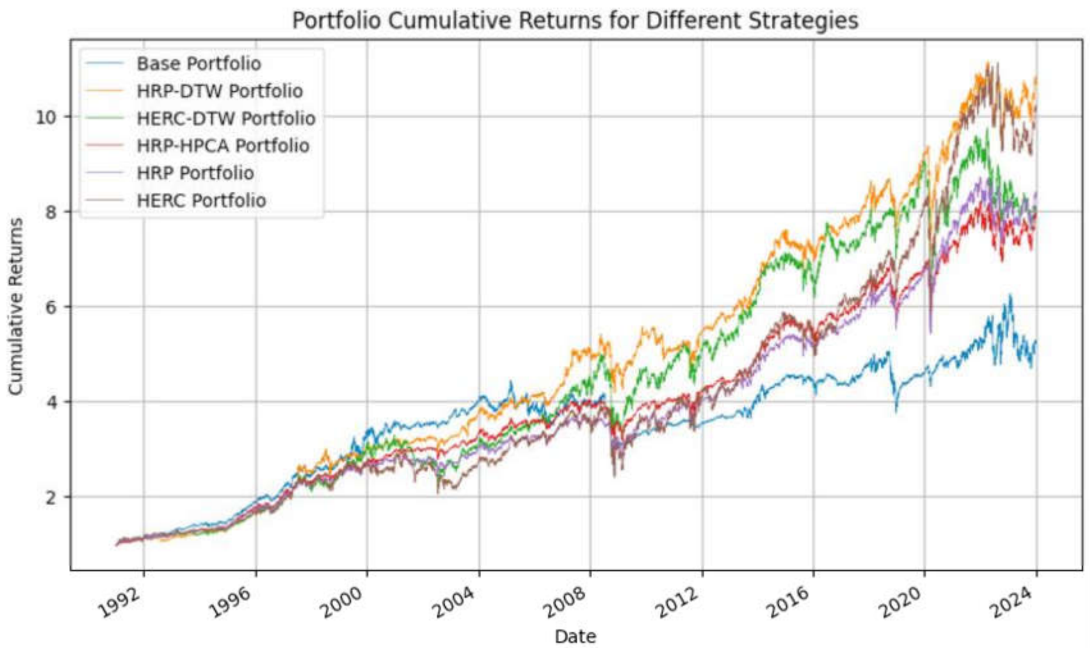


Figure 6: Cumulative Returns of High-Reward Investment Strategies

Across both ”Stable” and ”High-Return” solution spectrums, the APD strategies advocate for a balanced amalgamation of mitigated risk growth and potent capital enhancement. The strategic deployment of APD principles promises a controlled and progressive investment experience, offering solace to conservative but growth-aspiring investors.

5.1.2. Daily Returns Analysis

In our preceding analysis, we evaluated the cumulative returns over a given period and concluded that APD strategies exhibit stability in long-term growth, demonstrating resilience during catastrophic financial events such as economic crises with less impact and quicker recovery capabilities. Building upon this foundation, Figure 7 offers an interpretive view of the granular fluctuations in returns within the ”Stable Solutions,” dissecting the inherent volatility and risk dynamics characteristic of each investment strategy under the umbrella of stable approaches.

The data depicted across the entire temporal scope indicates that APD portfolios, which include the ”HRP-DTW Portfolio,” ”HERC-DTW Portfolio,” and ”HRP-HPCA Portfolio,” have maintained a strict dispersion range of returns. This denotes the application of finelytuned mitigation strategies against the market’s frequently unpredictable fluctuations, leveraging hierarchical clustering, principal component analysis, and temporal alignment algorithms to ensure a disciplined dispersion range, thereby dampening stochastic market noise. In particular, the temporal alignment strategy facilitated by the DTW method has further insulated these portfolios from sporadic and anomalous market movements that typically induce greater volatility in daily returns. In contrast, the baseline ”Base Portfolio” (i.e., the MV method of weights’ determination) exhibits a broader volatility of returns with more pronounced peaks and troughs, indicating a greater sensitivity to transient market anomalies.

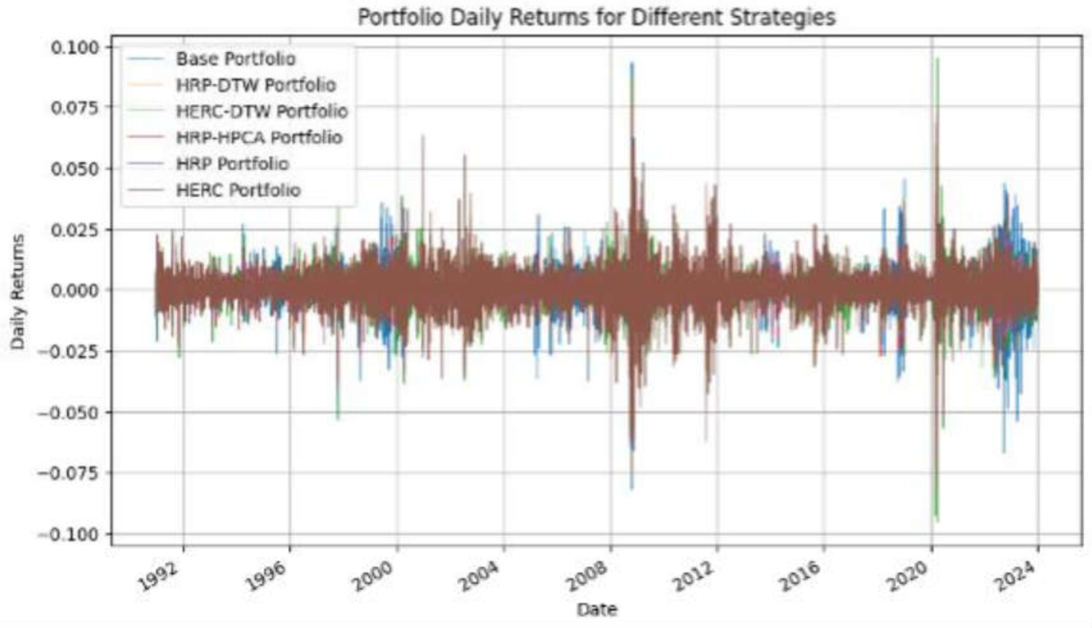


Figure 7: Portfolio Daily Return of Stable-Return Investment Strategies

Figure 8's "High-Return Solutions" display a wider range of daily returns, in line with their risk-receptive stance aimed at achieving greater financial gains. APD strategies consistently provide a stable and stringent range of return dispersion. Such precision in risk control, especially within the "Stable Solutions," contrasts starkly with the broader variances seen in "High-Return Solutions," indicating a greater tolerance for market upheaval in pursuit of superior financial outcomes.

Together, these analyses offer a nuanced understanding of the risk-return profiles inherent to each investment strategy. They demonstrate that APD strategies furnish a more meticulous market navigation, delivering a level of controlled performance not consistently achievable by traditional MV optimization. The APD framework, with its strategic emphasis on risk reduction and capital enhancement, assures a more predictable and less volatile journey towards wealth accumulation, appealing to investors across the spectrum of risk appetites.

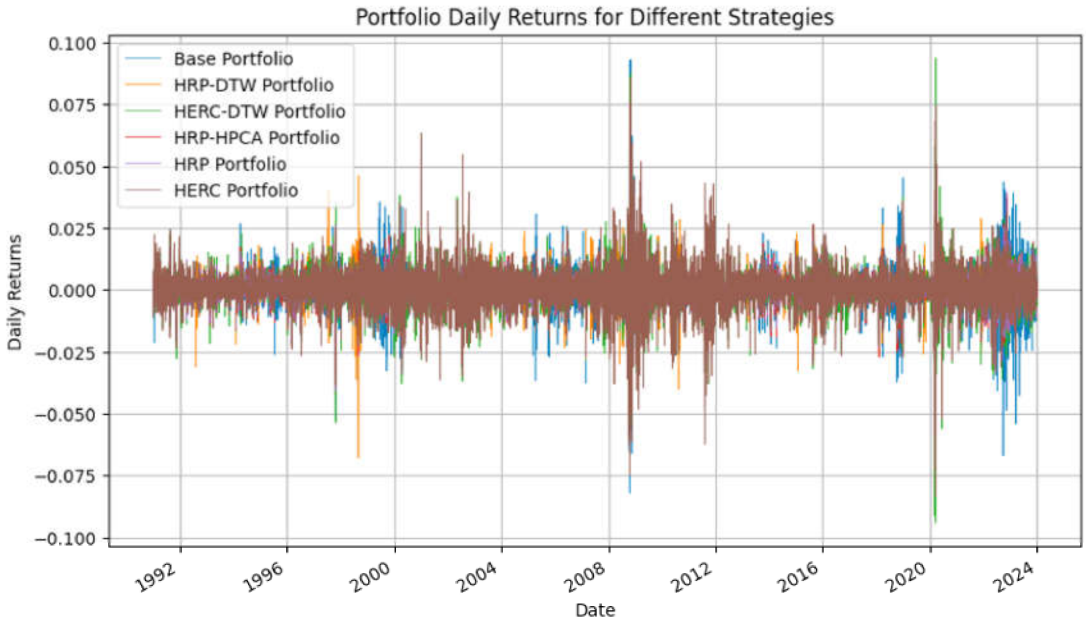


Figure 8: Portfolio Daily Return of High-Return Investment Strategies

5.1.3. Drawdown Analysis

Figure 9 provides a visual representation of portfolio performance under stress, showing the extent of losses each strategy sustained from peak to trough. The chart specifically illustrates the ”stable solution” and its ability to reduce losses during market downturns, demonstrating the effectiveness of the APD strategy.

The volatility curve reveals the maximum depth of decline experienced by each strategy, indicating the retracement from the peak. The ”Stable Solution” includes strategies such as the ”HRP Portfolio” and the ”HERC Portfolio”, which have significantly smaller drawdowns than their ”Basic Portfolio” and DTW-enhanced counterparts. This demonstrates the APD strategy’s robustness in terms of capital preservation and its ability to maintain portfolio valuations during volatile market conditions.

The “HRP-DTW” and “HERC-DTW” strategies show a slight increase in drawdowns, suggesting that while DTW integration helps with timing and potentially identifying market changes, it may not always imply sensitivity to extreme market moves decline. However, the ”HRP-HPCA portfolio” integrating hierarchical principal component analysis often exhibits superior resilience, highlighting the utility of this approach in distilling and responding to

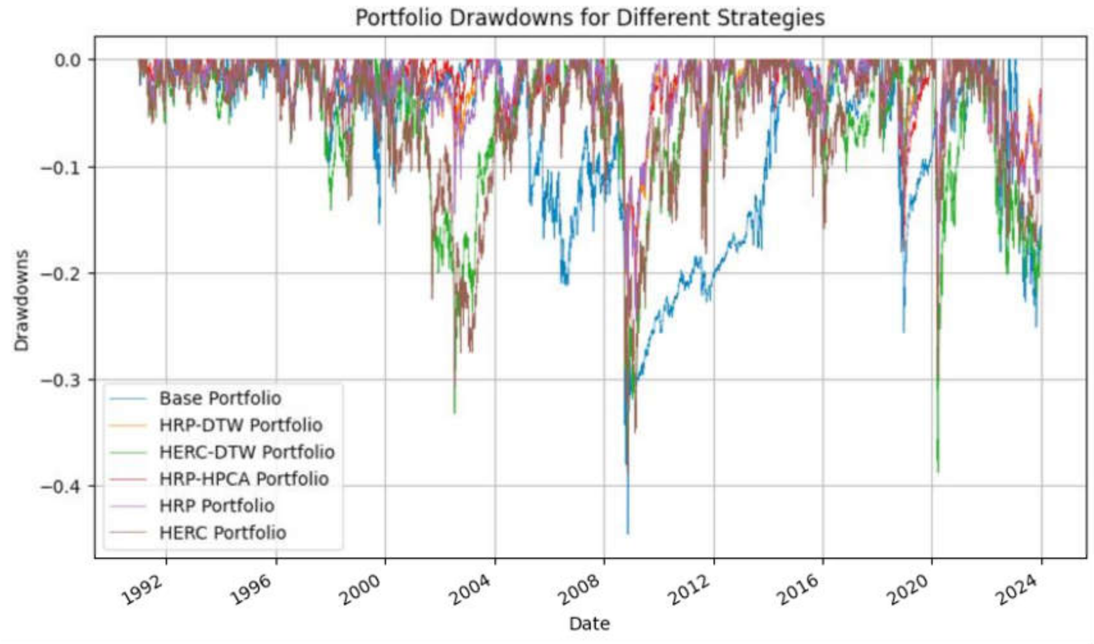


Figure 9: Portfolio Drawdowns of Stable-Return Investment Strategies

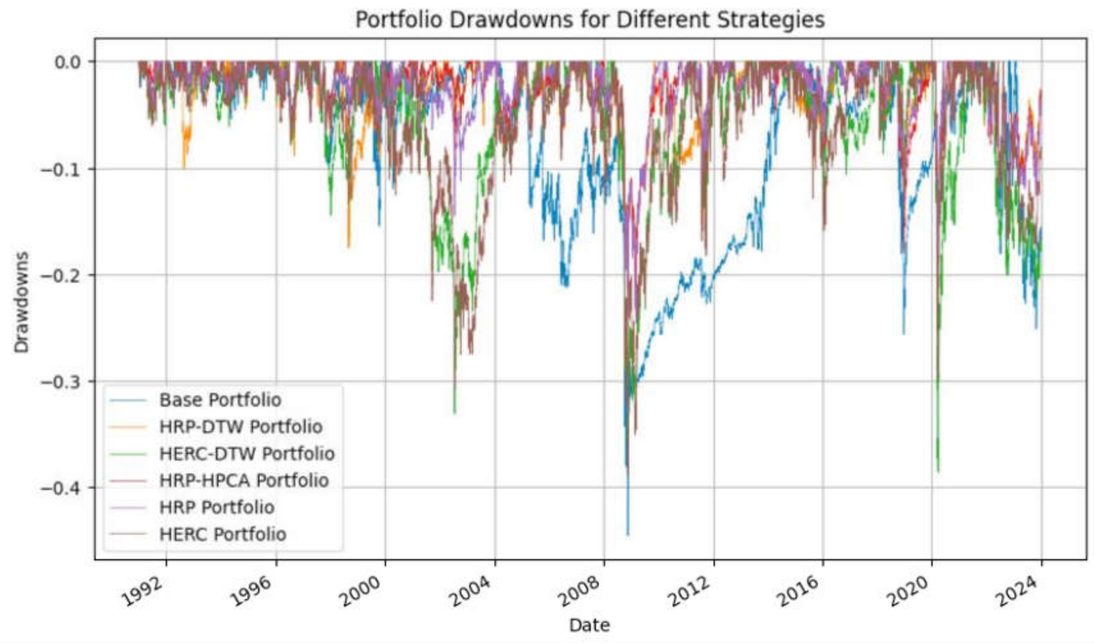


Figure 10: Portfolio Drawdowns of High-Return Investment Strategies core market drivers while mitigating the impact of peripheral fluctuations.

In Figure 10, the retracement value fluctuation under high returns is similar to that under steady state, except that due to the influence of DTW, the retracement elasticity of HRP increases.

By comparing the drawdowns of these strategies to a traditional ”underlying portfolio,” APD’s ability to respond to and recover from financial turbulence is significantly enhanced. This attribute is invaluable to investors with risk-averse tendencies and gives us confidence that the strategic deployment of APD principles can lead to a more controlled, less chaotic investing experience.

5.2. Other Performance Metrics

In the comparative assessment of portfolio strategies under the ambit of APD and traditional MV optimization, the key performance metrics reveal a distinct demarcation in risk propensity and endurance. As Table 2, the APD strategies, embodied by the likes of ”HRP-DTW,” ”HRP-HPCA,” and ”HERC,” demonstrate a notable reduction in Max DD when juxtaposed against the baseline MV strategy, the ”Base Portfolio.” This reduced Max DD underscores the APD strategies’ enhanced resilience to market downturns, thereby endorsing their efficacy in mitigating the severity of potential losses.

Table 2: Stable-Return Solutions Metrics

	MV	*HRP-DTW*	*HERC-DTW*	*HRP-HPCA*	*HRP*	*HERC*
Max DD	−0.45	−0.25	−0.39	−0.22	−0.25	−0.39
RET	0.01	0.02	0.02	0.02	0.02	0.03
SD	0.12	0.09	0.13	0.09	0.09	0.14
Sharpe	0.05	0.25	0.15	0.23	0.24	0.20
Sortino	0.08	0.40	0.23	0.36	0.37	0.32
Calmar	0.13	0.29	0.18	0.32	0.29	0.21
VaR (5%)	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
VaR (1%)	−0.02	−0.02	−0.02	−0.02	−0.02	−0.03
CVaR (5%)	−0.02	−0.01	−0.02	−0.01	−0.01	−0.02
CVaR (1%)	−0.03	−0.02	−0.04	−0.02	−0.02	−0.04

The Sharpe and Sortino ratios, quintessential gauges of risk-adjusted performance, show a marked improvement for APD strategies, thereby substantiating their superiority in delivering higher returns per unit of total and negative risk, respectively. Such improvement in risk-adjusted returns is a testament to the APD’s sophisticated risk management and optimization processes that refine the investment landscape by taming volatility and tempering drawdowns.

The Calmar ratio, which marries the annualized rate of return to the maximum drawdown, illuminates the prudence of the APD strategies. Elevated Calmar ratios in the context of ”Stable Solutions” such as ”HRP” and ”HRP-HPCA” portfolios highlight their proficiency in securing commendable returns despite facing adverse market movements.

VaR and CVaR further punctuate the risk narrative, quantifying the prospective loss within a defined confidence level. APD strategies exhibit lower VaR and CVaR figures, indicative of a systematic approach to risk containment, where the potential for tail-end losses is prudently restrained.

For the ”High-Return Solutions” within the realm of APD, the performance metrics delineate a compelling case for their risk-reward profile when contrasted against the MV optimization. The metrics, as Table 3 shows, unfold a narrative of deliberate risk engagement to harness superior gains. Max DD for these high-return portfolios indicates a strategic acceptance of

Table 3: High-Return Solutions Metrics

	MV	*HRP-DTW*	*HERC-DTW*	*HRP-HPCA*	*HRP*	*HERC*
Max DD	−0.45	−0.23	−0.39	−0.22	−0.25	−0.39
RET	0.01	0.03	0.02	0.02	0.02	0.03
SD	0.12	0.10	0.13	0.09	0.09	0.14
Sharpe	0.05	0.31	0.16	0.23	0.24	0.20
Sortino	0.08	0.49	0.24	0.36	0.37	0.32
Calmar	0.13	0.35	0.18	0.32	0.29	0.21
VaR (5%)	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
VaR (1%)	−0.02	−0.02	−0.02	−0.02	−0.02	−0.03
CVaR (5%)	−0.02	−0.02	−0.02	−0.01	−0.01	−0.02
CVaR (1%)	−0.03	−0.03	−0.04	−0.02	−0.02	−0.04

deeper troughs in exchange for higher returns, as evidenced by the escalated return (RET) figures. The HRP-DTW strategy, for instance, with a Max DD of -0.233964, demonstrates a significant reduction in adverse excursions compared to the baseline MV, highlighting the nuanced risk management of APD strategies while seeking amplified returns.

The Sharpe and Sortino ratios for these high-return strategies, such as HRP-DTW with a Sharpe ratio of 0.312291 and a Sortino ratio of 0.388919, suggest that the additional risk incurred is being adequately rewarded by the excess returns over the risk-free rate. These enhanced ratios signify that the strategies not only navigate the markets more proficiently but also utilize risk in a manner that proportionately increases the potential for returns.

The Calmar ratios of these portfolios, particularly for HRP-DTW at 0.346863, reaffirm the APD strategies' strength in sustaining growth over long periods despite market setbacks, providing a clear illustration of sustained performance against the backdrop of the deepest drawdowns experienced.

Furthermore, the VaR and CVaR metrics reinforce the risk posture of these strategies. Lower VaR and CVaR values for HRP-DTW, at -0.009365 and -0.015305 respectively, compared to the baseline MV, exhibit a strategic risk containment where the likelihood and mean of potential losses are kept within manageable bounds despite the aggressive pursuit of higher returns.

The high-return APD strategies exemplify a bolder, yet calculated approach to portfolio management, where heightened returns are pursued without overlooking the implications of risk. These strategies advocate for an investment ethos where the acceptance of larger fluctuations is integral to achieving outsized financial outcomes, aligning with the aspirations of investors who possess a more robust appetite for risk in their pursuit of wealth maximization.

In the interplay of APD and MV optimizations, it becomes apparent that the former offers a more robust platform for investors, characterized by lower volatility and drawdowns without significantly compromising on return potential. These metrics coalesce to portray a holistic view of the investment strategies, where APD not only transcends the traditional frameworks in terms of return metrics but also provides a more nuanced approach to risk management, affirming its suitability for investors with varying risk appetites and investment horizons.

6. Conclusions

This study's comprehensive analysis of APD strategies represents a significant advancement in portfolio optimization. By integrating HPCA and DTW with HC portfolio diversification, our models surpass the traditional MV optimization approach. Empirical investigation and quantitative metrics reveal that APD strategies significantly enhance portfolio robustness, reduce volatility, and mitigate risk while delivering competitive returns. This high performance is attributed to two main factors: (1) more diverse weight allocation and (2) multiple risk minimization methods beyond just volatility.

For further implication, the findings of this paper advocate for the adoption of APD strategies by investors who prioritize both capital appreciation and prudent risk management. APD strategies are adaptable to various investor profiles, from risk-averse individuals to those seeking aggressive growth. The integration of advanced statistical models within APD establishes it as a formidable approach in modern finance, offering a structured yet flexible framework for portfolio construction that aligns with today's complex market dynamics.

In conclusion, as the financial industry evolves, APD principles provide a solid foundation for navigating the multifaceted investment environment. Investors and portfolio managers are encouraged to consider APD as a means to achieve a more nuanced, informed, and dynamic portfolio management process, capable of responding to the unpredictable nature of global financial markets and the ever-changing investment horizon.

7. Future Improvements

As we look ahead, the continuous refinement of APD strategies remains pivotal in adapting to the evolving dynamics of financial markets. The promising results obtained thus far serve as a foundation upon which future research and practical enhancements can be constructed. To sustain the momentum of progress and address the challenges uncovered during the analysis, the following suggestions for future research and practical implications are proposed:

1. Data Enrichment – Incorporating a broader spectrum of data, including alternative and unstructured data sources (such as macroeconomic indicators, sentiment analysis, and ESG factors), could provide deeper insights into asset behavior and market dynamics.

2. Algorithmic Advancements – Exploring cutting-edge machine learning algorithms and artificial intelligence techniques could improve the predictive accuracy of the models. This could involve the use of reinforcement learning for dynamic strategy adjustments and neural networks for pattern recognition within financial time series.

3. Model Hybridization – Combining the strengths of different models may yield a more robust framework. For instance, integrating DTW with machine learning classifiers can enhance the detection of regime shifts and asset class behaviors, leading to more responsive portfolio adjustments.

4. Portfolio Customization – Tailoring APD strategies to different market conditions and investor preferences can enhance their applicability. Customizing strategies for various market environments and risk appetites can make them more versatile and effective.

5. Performance Benchmarking – Establishing comprehensive benchmarks to evaluate APD strategies against other advanced portfolio optimization techniques can provide clearer insights into their relative performance and areas for improvement.

The roadmap for future improvements must be navigated with a commitment to rigorous research, innovative thinking, and adherence to sound investment principles. As the financial landscape becomes increasingly complex, the APD framework’s ability to evolve and incorporate new methodologies will be critical to maintaining its relevance and efficacy in portfolio management. Despite these suggestions, our research significantly contributes to the evolving academic field. Specifically, our research has shown the profitability of combining DTW with the HC model to outperform the traditional HC method under various portfolio performance metrics. In terms of practical application, our strategies better mimic the real world by incorporating fluctuating interest rates into our asset allocations. Furthermore, our approaches and results are legitimate due to the proofs of these models separately by previous researchers and the beneficial properties[13] of the research targets we selected for testing our strategies' performance.

Appendix A

- Single Linkage (SL)

Single Linkage (SL) is a method of Hierarchical Clustering. Starting with a matrix that measures the distances, this approach initially gives each object its distinct cluster. In subsequent steps, it consistently combines the two closest clusters. This process continues until all objects are grouped into a single cluster. Throughout this process, the measure used to determine the distance between two clusters $d_{A,B}$ is updated to reflect the smallest distance between any two elements in each cluster.

$$d_{A,B} = \min_{a \in A, b \in B} D(a,b) \tag{34}$$

- Complete Linkage (CL)

Complete Linkage (CL) is a variant of SL, with the distance between two clusters $d_{A,B}$ is the largest distance between any two elements in each cluster:

$$d_{A,B} = \max_{a \in A, b \in B} D(a,b) \tag{35}$$

- Median Linkage (ML)

Median Linkage (ML) is another variant of SL, with the distance between two clusters $d_{A,B}$ is the median distance between any two elements in each cluster:

$$d_{A,B} = \operatorname*{median}_{a \in A, b \in B} D(a,b) \tag{36}$$

- Average Linkage (AL)

[13] The assets in our portfolio have a long life, are highly convertible, and serve as reliable measures or benchmarks for the main categories of investments (such as stocks, bonds, commodities, etc.) and their more specific divisions.

Average Linkage (AL) extends SL, with the distance between two clusters $d_{A,B}$ is the mean distance between any two elements in each cluster:

$$d_{A,B} = \underset{a\in A, b\in B}{\text{mean}}\, D(a,b) \tag{37}$$

AL has another name: UPGMA (Unweighted Pair Group Method with Arithmetic Mean). The UPGMA algorithm generates a rooted tree, also known as a dendrogram, which visualizes the relationships delineated in a pairwise similarity matrix or a dissimilarity matrix. With each iteration, it merges the two closest clusters into a larger cluster.

Thus, the distance between any two clusters A and B, $d_{(A\cup B),X}$, which have cardinalities (i.e., number of elements in each cluster) $|A|$ and $|B|$ respectively, is computed as the mean of all distances $d(x,y)$ between the object pairs x in A and y in B, thereby establishing the average distance between the members of each cluster. Put another way, during each step of clustering, the new distance that is determined between the merged clusters $A\cup B$ and another cluster X is calculated using a weighted average of the distances $d_{A,X}$ and $d_{B,X}$:

$$d_{(A\cup B),X} = \frac{1}{|A|\cdot|B|}\sum_{x\in A}\sum_{y\in B} d(x,y) = \frac{|A|\cdot d_{A,X} + |B|\cdot d_{B,X}}{|A|+|B|} \tag{38}$$

- Weighted Linkage (WL)

Weighted Linkage (WL) is a variation of AL, with the distance between two clusters is the weighted average distance (i.e., Weighted Pair Group Method with Arithmetic Mean, WPGMA) between any two elements in each cluster.

The WPGMA algorithm is a method for constructing a phylogenetic tree, known as a dendrogram, which represents the relationships indicated by a pairwise distance matrix or a similarity matrix. During the process, the algorithm pairs the two closest clusters, denoted by A and B, into a single, larger cluster represented by $A\cup B$. The distance of this new cluster to another cluster, denoted by $d_{(A\cup B),X}$, is calculated as the arithmetic mean of the average distances from X to both A and X to B:

$$d_{(A\cup B),X} = \frac{d_{A,X} + d_{B,X}}{2} \tag{39}$$

- Centroid Linkage (CL)

Centroid Linkage (CL) includes calculating the average position of each cluster (i.e., centroid), $d_{A,B}$, and then measuring the distance between these central points:

$$d_{A,B} = d(\bar{A},\bar{B}) \tag{40}$$

- Ward Linkage (Ward's Method)

Ward Linkage (Ward's Method) is an alternative to SL. It calculates the increase in the total sum of squared errors. Essentially, the minimum variance criterion of Ward's aims to keep the variance within each cluster as low as possible. In other words, we want to find two clusters *A* and *B* such that when they are combined, the increase in the total sum of squared within-cluster distances is least. And the distance between two clusters $d_{A,B}$ is the increase in overall sum of squared within-cluster distances:

$$d_{A,B} = \Delta(\mathrm{SS_{within}}) = \mathrm{SS_{within}}(T) - (\mathrm{SS_{within}}(A) + \mathrm{SS_{within}}(B)) \tag{41}$$

where:

- $\Delta(\mathrm{SS_{within}})$ is the increase in the sum of squares within clusters due to merging.
- $\mathrm{SS_{within}}(A)$ and $\mathrm{SS_{within}}(B)$ are the sum of squares within the individual clusters *A* and *B*, respectively.
- $\mathrm{SS_{within}}(T)$ is the sum of squares within the new cluster T, formed by merging *A* and *B*.

- Direct Bubble Hierarchical Tree Linkage (DBHT Clustering)

The DBHT (Dynamic Branching Hierarchical Tree) clustering method is a graph-theoretic approach to extracting clusters and hierarchies from complex datasets. It does so deterministically and without prior information, which distinguishes it from many clustering methods that may require prior information or supervision. The exact steps to perform DBHT Clustering are complex, but the general idea of DBHT Clustering is to form a bubble tree and select the CL method based on this bubble tree to form clusters. You can see a more detailed explanation by Song et al. (2012).

The traditional linkage algorithms (i.e., ranging from SL to Ward's method) begin by organizing distances between elements (such as stocks) from smallest to largest. They then create a dendrogram, which groups subsets based on these minimal distances, and determine the clusters by selecting a specific number of them. In contrast, the DBHT method inverts this sequence: it first identifies clusters through topological analysis of a planar graph [14], then establishes a hierarchy within and among these clusters. Thus, DBHT differs in the type of information it uses and its overall procedural strategy.

[14] Planar Maximally Filtered Graph (PMFG); the PMFG is defined as a type of graph that consists of a set of vertices *V*, edges *E*, weights *W* assigned to *E*, and a set of distances *D* associated with *E*.

Appendix B

For two time series X and Y, $n = len(X)$ and $m = len(Y)$. The DTW algorithm has a time complexity of $O(nm)$, capable of determining the precise optimal solution for this issue. The pseudo-code in Python is as follows [15]:

```
def dtw(x, y):
# Initialization
    for i = 1..n
        for j = 1..m
            C[i, j] = inf

    C[0, 0] = 0.

    # Main loop
    for i = 1..n
        for j = 1..m
            dist = d(x_i, y_j)**2
            C[i, j] = dist + min(C[i-1, j], C[i, j-1], C[i-1, j-1])

    return sqrt(C[n, m])
```

Here, $d(x_i,y_j) = |x_i - y_j|$, and in our case, $n = m$. DTW is a Dynamic Programming (DP) technique where each entry at row i and column j is updated based on the distance between asset x at time i and asset y at time j and the minimum of both these assets' previous time steps.

References

[1] Avellaneda, M. (2020). Hierarchical pca and applications to portfolio management. *Revista mexicana de econom´ıa y finanzas*, 15(1):1–16.

[2] Braga, M. D., Nava, C. R., and Zoia, M. G. (2023). Kurtosis-based risk parity: methodology and portfolio effects. *Quantitative Finance*, 23(3):453–469.

[3] Chen, Z.-p. and Zhao, C.-e. (2003). Sensitivity to estimation errors in mean-variance models. *Acta Mathematicae Applicatae Sinica*, 19(2):255–266.

[4] Chua, D. B., Kritzman, M., and Page, S. (2009). The myth of diversification. *The Journal of Portfolio Management*, 36(1):26–35.

[5] Eichner, T. (2008). Mean variance vulnerability. *Management Science*, 54(3):586–593.

[6] Fusai, G., Mignacca, D., Nardon, A., and Human, B. (2020). Equally diversified or equally weighted? *Available at SSRN 3628585*.

[7] Ibanez, F. A. (2023). Diversified spectral portfolios: An unsupervised learning approach to diversification. *Journal of Financial Data Science*, 5(2).

[8] Jaeger, M., Krugel, S., Marinelli, D., Papenbrock, J., and Schwendner, P. (2021). Under-¨ standing machine learning for diversified portfolio construction by explainable ai. *Markus Jaeger, Stephan Krugel, Dimitri Marinelli, Jochen*

[15] This pseudo-code can be found here: https://tslearn.readthedocs.io/en/stable/user_guide/ dtw.html.

Papenbrock and Peter Schwendner. Inter-¨pretable Machine Learning for Diversified Portfolio Construction. The Journal of Financial Data Science Summer.

[9] Kinlaw, W. B., Kritzman, M., and Turkington, D. (2023). Co-occurrence: A new perspective on portfolio diversification.

[10] Lim, T. and Ng, H. (2022). Financial portfolio management based on shaped-based unsupervised machine learning: A dynamic time warping baycenter averaging approach to international markets and periods of downside event risks. *The Journal of Investing*.

[11] Lim, T. and Ong, C. S. (2021). Portfolio diversification using shape-based clustering. *Journal of Financial Data Science*, 3(1):111.

[12] Lopez de Prado, M. (2016). Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management*.

[13] Ma, Y., Han, R., and Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, 165:113973.

[14] Markowitz, H. M. (1989). Mean—variance analysis. *Finance*, pages 194–198.

[15] Martellini, L. and Milhau, V. (2018). Proverbial baskets are uncorrelated risk factors! a factor-based framework for measuring and managing diversification in multi-asset investment solutions. *Journal of Portfolio Management*, 44(2):8–22.

[16] Millea, A. and Edalat, A. (2022). Using deep reinforcement learning with hierarchical risk parity for portfolio optimization. *International Journal of Financial Studies*, 11(1):10.

[17] Pfitzinger, J. and Katzke, N. (2023). Perspectives on optimal hierarchical portfolio selection. *Perspectives on Optimal Hierarchical Portfolio Selection (April 4, 2023)*.

[18] Raffinot, T. (2018). The hierarchical equal risk contribution portfolio. *Available at SSRN 3237540*.

[19] Schwendner, P., Papenbrock, J., Jaeger, M., and Krugel, S. (2021). Adaptive seriational risk¨ parity and other extensions for heuristic portfolio construction using machine learning and graph theory. *The Journal of Financial Data Science*, 3(4):65–83.

[20] Serur, J. A. and Avellaneda, M. (2020). Hierarchical pca and modeling asset correlations. *Available at SSRN 3903460*.

[21] Song, W.-M., Di Matteo, T., and Aste, T. (2012). Hierarchical information clustering by means of topologically embedded graphs. *PloS one*, 7(3):e31929.

[22] Sood, S., Papasotiriou, K., Vaiciulis, M., and Balch, T. (2023). Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization. *FinPlan 2023*, page 21.

[23] Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

[24] Tsiang, S.-C. (1989). The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money. In *Finance Constraints and the Theory of Money*, pages 221–248. Elsevier.

[25] Yuan, M. and Zhou, G. (2022). Why naive diversification is not so naive, and how to beat it? *Journal of Financial and Quantitative Analysis*, pages 1–32.