Stevens Institute of Technology

Financial Analytics

Portfolio Diversification in QWIM

Author:
Khooshi Ajmera
Abhimanyu Sheth
Tirthkumar Patel

Supervisor:
Dr. Cristian Homescu
Under the Guidance of:
Professor Ionut Florescu

December 18, 2023

Abstract

In this study, we analyze portfolio diversification using data from 50 stocks across five sectors: Tech, Pharma, FMCG, Banking, and Auto. We employed correlation matrix and network analysis, and then we assess inter-stock and inter-sector correlations. We further apply hierarchical and K-means clustering to identify distinct stock clusters. The crux of our methodology involves Mean-Variance Optimization (MVO) and Global Mean-Variance, focusing on effective weight allocation in portfolios. Our findings reveal that the MVO approach not only enhances portfolio diversification but also yields favorable returns. This comprehensive approach, combining analytical techniques and optimization strategies, offers a robust framework for effective portfolio management in diverse market sectors.

Keywords: portfolio diversification, correlation matrix, network analysis, clustering, Mean-Variance Optimization (MVO), Global Mean-Variance, weight allocation, market sectors.

Introduction

In the ever-evolving landscape of financial markets, the strategic allocation of investments across various asset classes stands as a cornerstone of effective portfolio management. The concept of diversification, a fundamental risk management technique, has long been hailed as the maxim of prudent investing. This project delves into the realm of portfolio diversification, with a focus on equities, by examining 50 stocks from five key sectors: Technology (Tech), Pharmaceuticals (Pharma), Fast-Moving Consumer Goods (FMCG), Banking, and Automobile (Auto). The goal is to construct an optimized portfolio that not only withstands market volatilities but also maximizes returns.

The rationale behind selecting these diverse sectors is rooted in their distinct market behaviors and growth trajectories. The Tech sector, known for its innovation-driven growth, contrasts with the stable and demand-driven nature of FMCG. Pharma presents a unique blend of research-intensive advancements and regulatory impacts, while Banking and Auto sectors reflect the broader economic cycles. This diversity provides fertile ground for exploring correlations and divergences in stock performances, which is pivotal in crafting a well-balanced portfolio.

Our methodology embarks with a detailed Correlation Matrix and Network Analysis. This initial phase is critical in identifying the degree of linear relationship between the stocks and across sectors. By mapping these correlations, we aim to understand how stocks move in tandem or in opposition under varying market conditions. A high degree of correlation might suggest a redundancy in risk management, while low correlation could imply a potential for risk mitigation through diversification.

Following the correlation study, we implement Hierarchical Clustering and K-Means Clustering. These clustering techniques serve as powerful tools in segregating stocks into distinct groups based on similar characteristics. Hierarchical Clustering offers a visual insight into the groupings, forming a dendrogram that illustrates how each stock is linked. K-Means Clustering, on the other hand, partitions the stocks into clusters based on predefined criteria, providing a clear segmentation for further analysis.

The subsequent phase of our project involves the application of Mean-Variance Optimization (MVO) and Global Mean-Variance. These optimization models are central to modern portfolio theory, introduced by Harry Markowitz. MVO is designed to find the optimal balance between risk and return by allocating assets in a way that maximizes expected return for a given level of risk. The Global Mean-Variance extends this concept to a global investment horizon, allowing for a more comprehensive risk-return analysis.

Our application of these models is not merely theoretical; it is grounded in empirical data and real-world market conditions. By calculating the optimal weightage for each stock, we aim to construct a portfolio that not only promises higher returns but also aligns with the investor’s risk tolerance. Our analysis culminates in the discovery that portfolios created using the MVO method are not just theoretically sound but also practically viable, offering substantial returns.

This project, therefore, stands at the intersection of theoretical financial modeling and practical investment strategy. It exemplifies how data-driven approaches, coupled with advanced statistical and optimization techniques, can lead to more informed and effective portfolio management. By demystifying the complexities of stock correlations, clustering methodologies, and optimization models, we aim to provide a comprehensive guide for investors seeking to navigate the multifaceted world of portfolio diversification.

Research Objective

- Sector Analysis: Investigate the dynamics and correlations among stocks within five key sectors: Technology, Pharmaceuticals, FMCG, Banking, and Automotive.
- Correlation and Network Analysis: Utilize correlation matrices and network analysis to explore inter-stock and inter-sector relationships.
- Clustering Techniques: Implement hierarchical and K-means clustering to identify and analyze distinct stock clusters.
- Optimization Strategies: Apply Mean-Variance Optimization (MVO) and Global Mean-Variance methods to determine optimal stock weight allocations in portfolios.
- Performance Evaluation: Evaluate the effectiveness of MVO in enhancing portfolio diversification and generating superior investment returns compared to traditional methods.
- Benchmark Comparison: Compare the performance of the optimized portfolio against the S&P 500 to assess the relative performance. This involves tracking the portfolio’s alpha, which represents its excess returns over the benchmark, as well as analyzing the beta, which measures the portfolio’s volatility in relation to the S&P 500. This is crucial for evaluating the true effectiveness of any investment strategy, as outperforming a broad market index like the S&P 500 is often a goal of active portfolio management.

Data

We sourced comprehensive data through the Bloomberg Terminal, focusing on a meticulously selected set of 50 stocks. This data, spanning from January 2000 to September 23, 2023, encompassed daily price fluctuations, offering a rich historical perspective. Our selection was strategically diversified across five key sectors: Technology, Pharmaceuticals, Banking, Automobile, and Fast-Moving Consumer Goods (FMCG), with an equal representation of 10 stocks from each sector. This diverse yet focused approach allowed us to capture the nuanced performance trends and sector-specific dynamics, thereby laying a robust foundation for our in-depth analysis and strategic decision-making in portfolio management.

Literature Review on Portfolio Diversification Measures

The topic of portfolio diversification and its measurement has attracted significant research attention, with a diverse range of methodologies emerging in recent years. We figured that traditional methods like Parmentier and Loïc's rely on simple metrics like standard deviation, while others, like Serur and Avellaneda's hierarchical PCA, delve deeper into asset correlations and risk dependencies. Network-based approaches, spearheaded by works like Jing and Rocha's on cryptocurrency portfolios, have gained traction due to their ability to visualize and assess complex interrelationships between assets. Marti et al. offer a comprehensive review of this trend, highlighting the growing adoption of networks and clustering techniques in financial markets.

Tang et al.'s "Asset Selection via Correlation Blockmodel Clustering" and Akansu et al.'s "Quant investing in cluster portfolios" exemplify this, while Giudici et al. extend the application to robot advisory portfolios. Lim and Ong's shape-based clustering opens up another avenue for diversification analysis. Meanwhile, Olmo and Peralta and Zareei champion network-based portfolio selection, advocating for an emphasis on asset centrality and interconnectedness. Raffinot's hierarchical clustering techniques, including the "Hierarchical Equal Risk Contribution Portfolio," further refine these network-based methods. Beyond traditional and network-based approaches, machine learning is making its mark on portfolio diversification. Works like Snow's on asset weight optimization and Jaeger et al.'s on interpretable machine learning models showcase the potential of this technology for constructing diversified portfolios. In line with this, Li et al.'s research on interpretable machine learning for equities and Perrin and Roncalli's exploration of machine learning optimization algorithms further support the integration of AI into diversification strategies. Finally, studies like Kim et al.'s on personalized goal-based investing and Das et al.'s on dynamic optimization for multi-goals wealth management highlight the growing focus on tailoring diversification strategies to individual investor needs and preferences. This diverse landscape of research underscores the importance of continuously refining and adapting our understanding of portfolio diversification in the face of evolving market dynamics and technological advancements.

Methodology

Our study employs a systematic approach to portfolio diversification, leveraging a range of analytical techniques to analyze data from 50 stocks across five key sectors: Technology, Pharmaceuticals, FMCG, Banking, and Automotive.

Correlation Matrix and Network Analysis: We began by constructing a correlation matrix for the selected stocks. This matrix served as the foundation for our network analysis, enabling us to visualize and understand the intricate relationships between the stocks and within sectors. The network diagrams provided insights into how stock prices move in relation to one another, highlighting potential diversification opportunities.

Hierarchical Clustering: To group similar stocks, we implemented hierarchical clustering. This method allowed us to determine natural groupings based on stock price movements and sectoral characteristics. We utilized dendrograms to visually interpret these clusters, providing a hierarchical view of stock relationships.

K-means Clustering: Complementing hierarchical clustering, we applied K-means clustering. This technique helped us to categorize stocks into distinct clusters based on their characteristics. We iteratively determined the optimal number of clusters, ensuring the most effective and meaningful classification of the stocks.

Mean-Variance Optimization (MVO): For portfolio weight allocation, we utilized MVO. This process involved calculating expected returns, variances, and covariances for the portfolio. MVO allowed us to find the optimal balance between risk and return, maximizing the portfolio's efficiency.

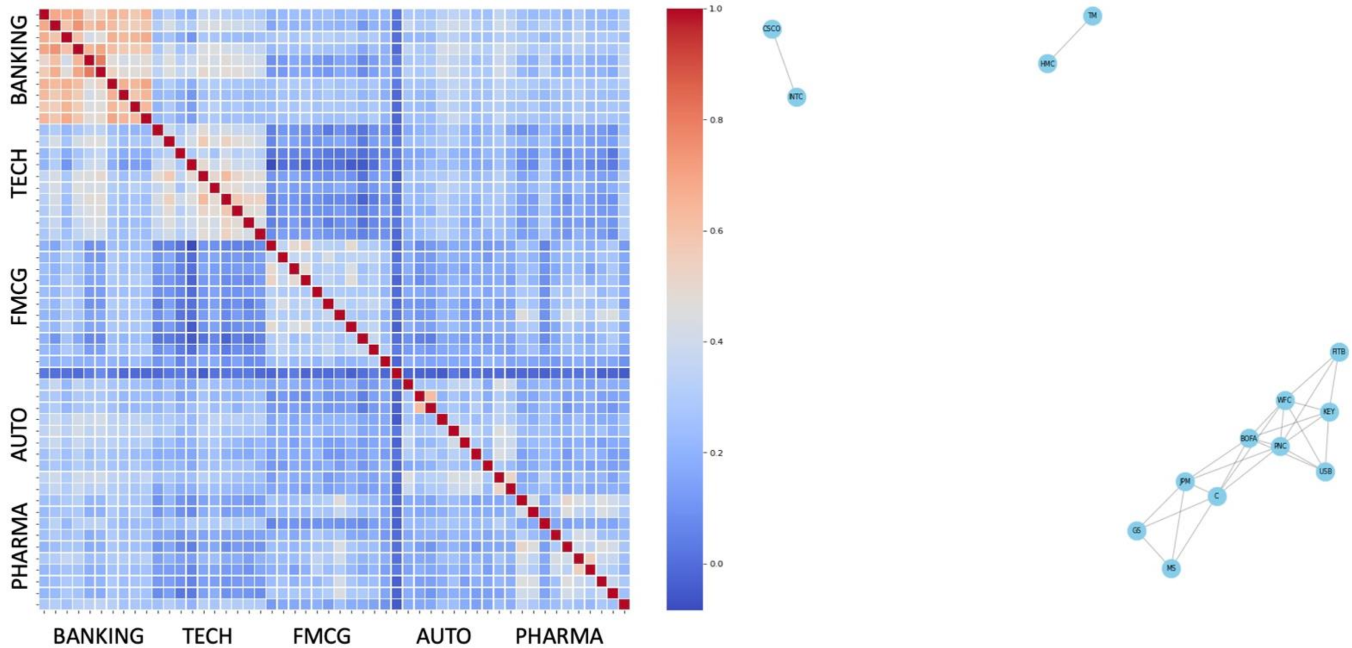
Global Minimum Variance (GMV): Alongside MVO, we employed the GMV strategy. This method focused on minimizing the portfolio's overall variance, offering an alternative perspective on weight allocation compared to MVO. We analyzed and compared the outcomes of both MVO and GMV to determine their effectiveness in portfolio diversification.

Throughout our study, we employed various statistical tools and software, ensuring accurate and efficient analysis. Our methodology was chosen to provide a comprehensive understanding of stock correlations, sector dynamics, and optimal portfolio construction, aligning with our goal of achieving enhanced portfolio diversification and improved investment returns

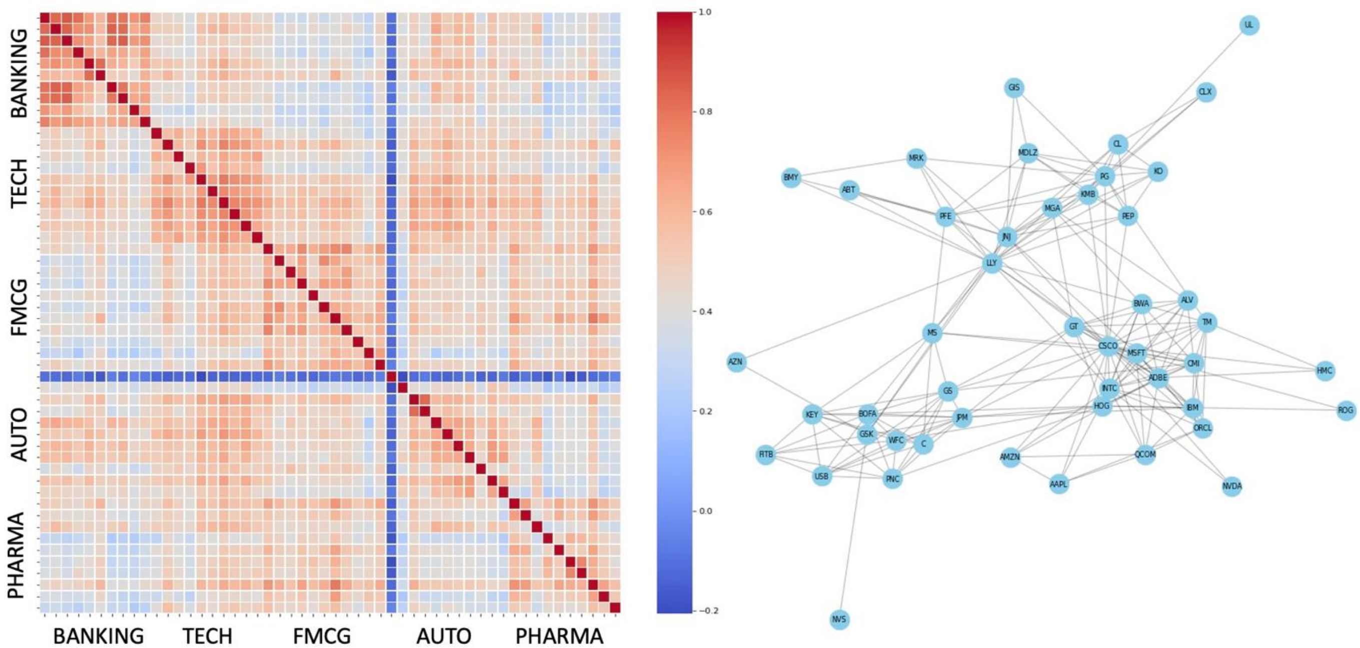
Correlation Heatmaps

We generated a Correlation Matrix Heatmap using Python to visually assess the relationships between stocks. This heatmap, created with seaborn and matplotlib libraries, displayed correlation coefficients, offering an intuitive understanding of how stock prices in different sectors are interrelated. It served as a crucial tool for identifying potential diversification strategies.

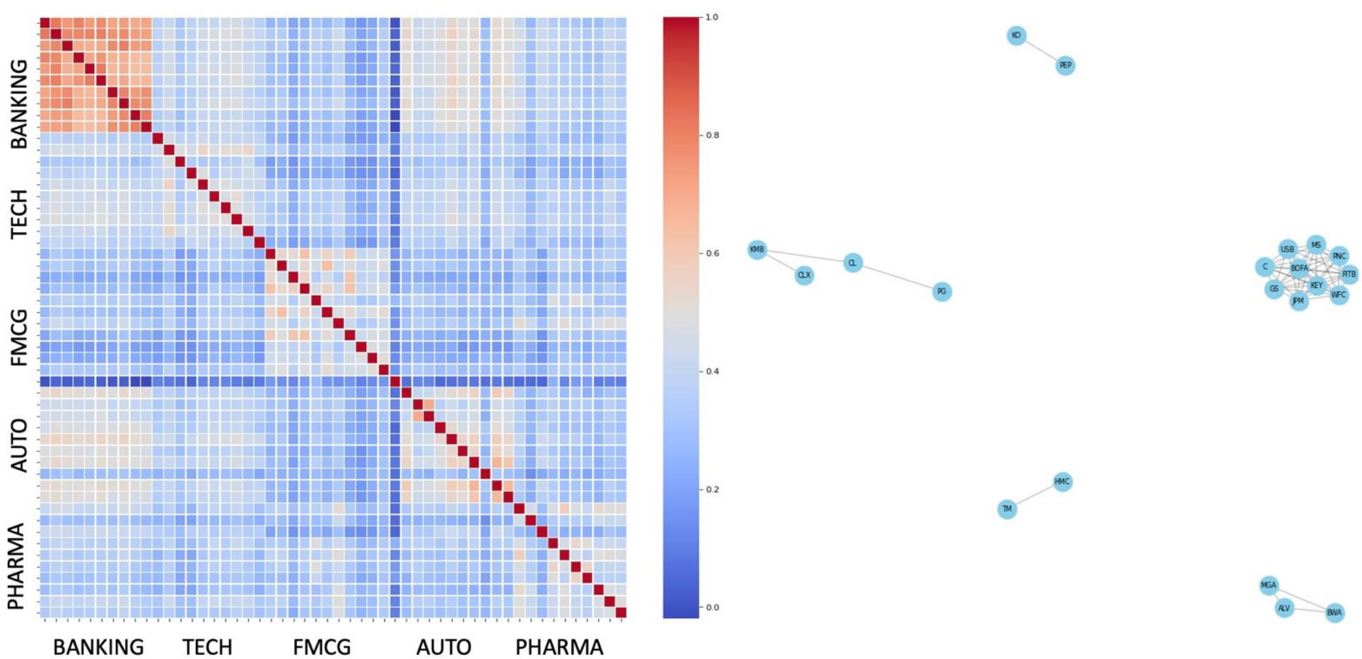
In addition, we utilized Python to create Network Graphs, effectively illustrating the interconnectedness between stocks across sectors. These graphs, developed using libraries like NetworkX and Matplotlib, provided a dynamic visualization of stock relationships, highlighting the strength and nature of correlations. This approach was instrumental in identifying key linkages and potential cluster formations within our dataset.



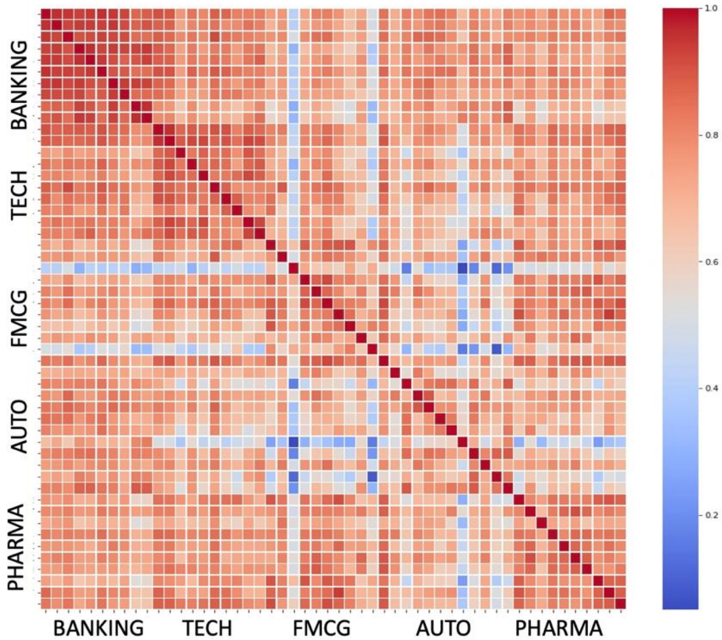
2000-01-01 to 2007-11-30: This period, preceding the global financial crisis, shows a relatively stable inter-sectoral correlation. The heatmap indicates moderate to high correlations within sectors, suggesting a clustered market behavior typical of pre-crisis periods. The network graph illustrates this interconnectedness, underlining the importance of diversification across sectors to mitigate systemic risks. The financial sector stocks (MS, GS, PM, BOFA, PNC) are the most tightly coupled cluster in the graph. This suggests that these stocks are likely to move in the same direction, regardless of what is happening in the rest of the market. The technology sector stocks (CSCO, INTC) are also tightly coupled, but they are less correlated with the financial sector stocks. This suggests that the technology sector may be more resilient to economic downturns than the financial sector. The auto sector stocks (HMC and TM) are relatively isolated from the other stocks in the graph. This suggests that they may be good stocks to own for diversification purposes.

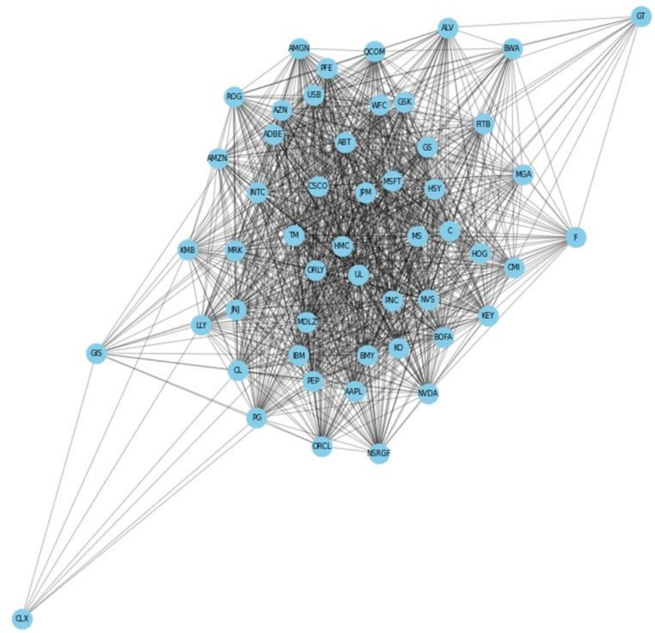


2007-12-01 to 2009-06-30: Encompassing the financial crisis, this phase exhibits heightened correlations, especially in the Banking sector, reflective of the turmoil in financial markets. The network graph becomes more interconnected, indicating a convergence of market movements, a common phenomenon in times of financial distress. The stock tickers with the highest average daily return during the crisis include NVS and UL. These stocks were relatively immune to the crisis and were able to generate strong returns for investors. The stock tickers with the lowest average daily return during the crisis include banking and IT stocks. These stocks were more closely tied to the overall market and were therefore more affected by the crisis.

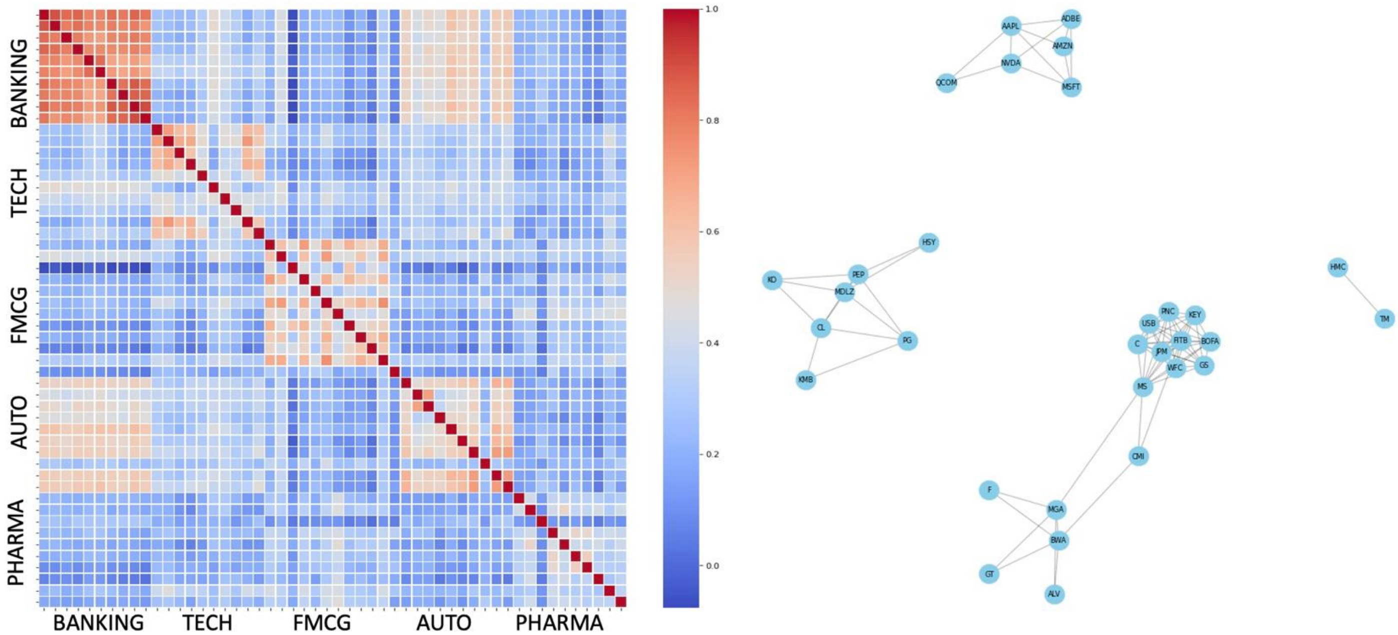


2009-07-01 to 2020-02-19: Post-crisis recovery and growth are marked by a diversification in correlations. Sectors like Tech and Pharma begin to show divergent patterns from traditional sectors like Banking and Auto, as indicated in the heatmap. This period underscores the evolving nature of sectoral dynamics in a recovering economy. The Networkgraph suggests that whatever the scenario maybe, banking sector stocks are always correlated. But if you closely analyze, you will realize that correlation among the stocks from different sectors is not that much, hence for portfolio diversification, we can take stocks from 5 different sectors.





2020-02-20 to 2020-04-10: Capturing the initial shock of the COVID-19 pandemic, the correlations spike across all sectors, as seen in the densely connected network graph. This short but intense period highlights how external, non-financial crises can lead to a systemic increase in market correlations. It can be inferred from this network graph that almost all the stocks from different sectors are tightly coupled with each other except for CLX which demonstrates that since this company deals more in hygiene equipment's, hence it was able to resist somehow as compared to other stocks.



2020-04-13 to 2023-10-26: The most recent period shows a new normal with varied correlations, possibly due to the ongoing adjustments in the global economy post-pandemic. The heatmap and network graph display a more complex interplay between sectors, suggesting evolving opportunities for portfolio diversification. There are several smaller clusters and individual nodes on the periphery of the graph, such as "TM," "F," "BWA," and "HMC." These stocks appear to be less correlated with the central cluster and might represent different sectors or have unique factors influencing their price movements. Stocks like "PG," "MDLZ," and "PEP" on the lower left are somewhat isolated from the rest of the graph, suggesting that their daily return prices do not correlate strongly with those of the other stocks on the graph. They may be influenced by different industry dynamics or have unique company-specific news affecting their stock prices.

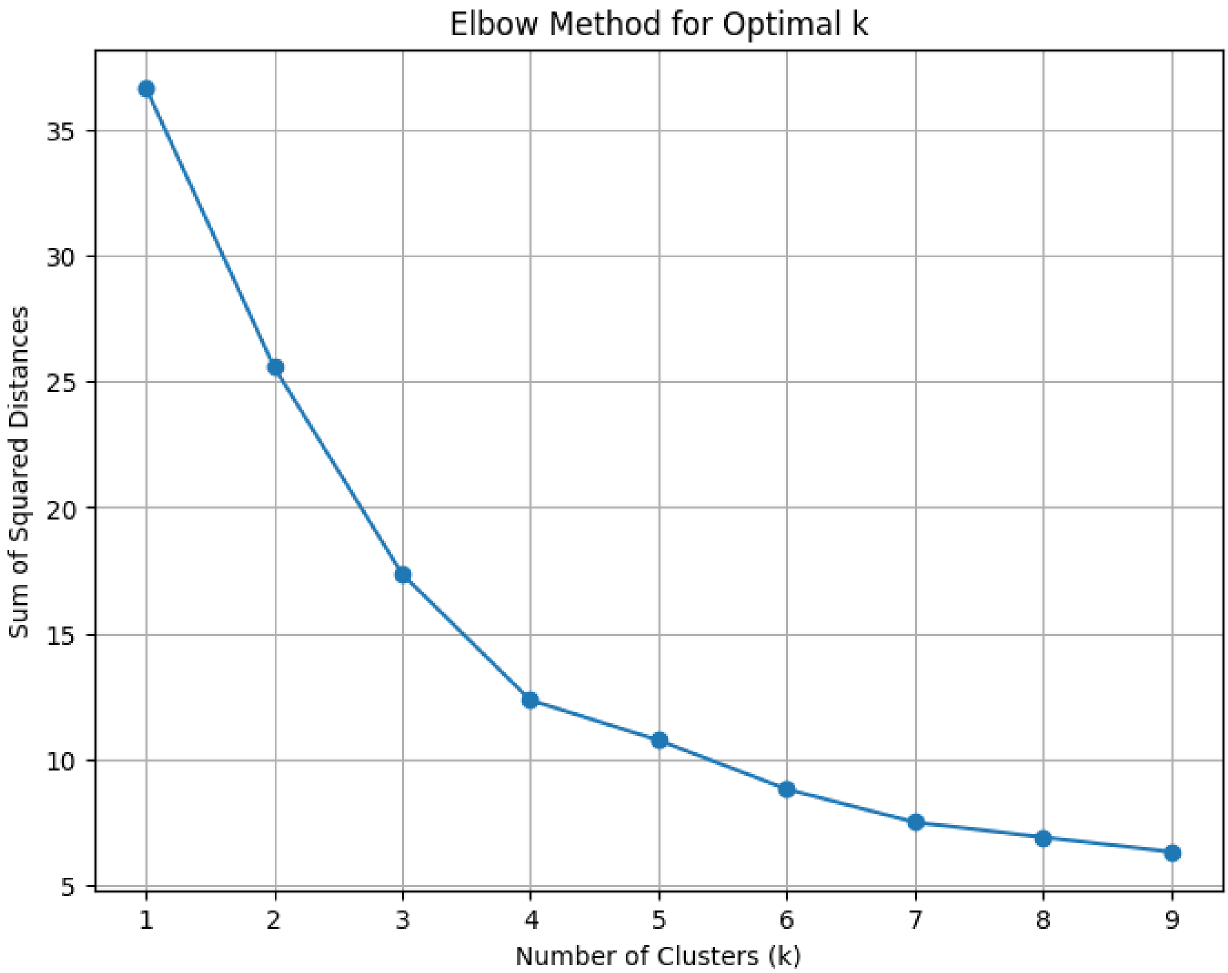
Overall, these heatmaps and network graphs serve as crucial tools in understanding market dynamics and the importance of diversification across different economic cycles. They illustrate how correlations between sectors evolve over time, influenced by global events and economic conditions, underlining the significance of adaptive portfolio strategies in managing risk and maximizing returns.

K-means Clustering

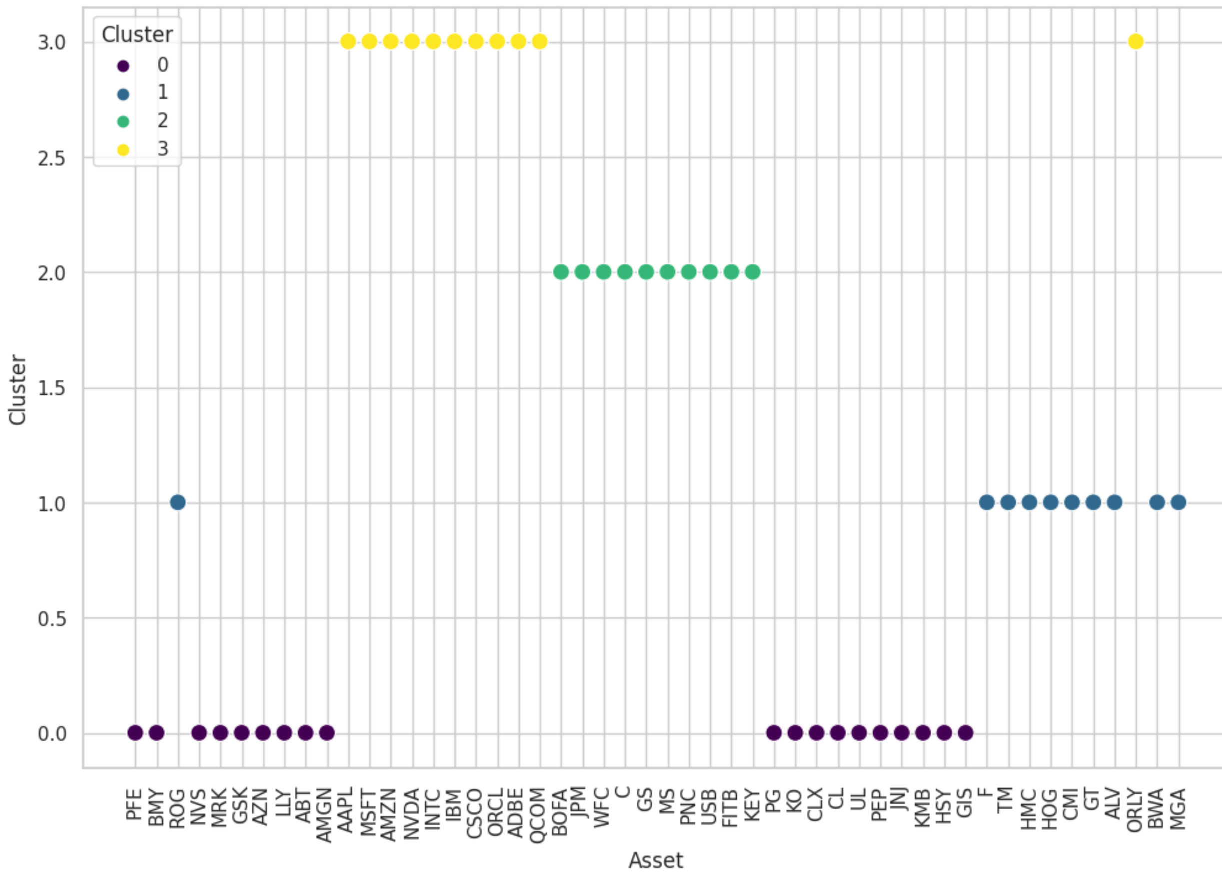
We used K-means clustering, which is a centroid-based algorithm, which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. To determine the optimal number of clusters for a given dataset we used the Elbow method. The Elbow method is employed as a heuristic to determine the point at which the within-cluster sum of squares (WSS) begins to diminish, indicating the optimal cluster count. The K-means algorithm partitions a dataset into K distinct non-overlapping subsets or clusters without any cluster-internal structure. Points in a cluster are often in proximity based on certain distance metrics. The Elbow method is a visual tool used in conjunction with the K-means algorithm. It involves plotting the

WSS as a function of the number of clusters (K) and identifying the number of clusters where the change in WSS begins to level off (the "elbow").

During the entire period: To figure out if the number of clusters is changing, we performed the Elbow Method and K-Means Clustering over the entire training period (2000-2020) and during the Banking Crisis (2007-2009)

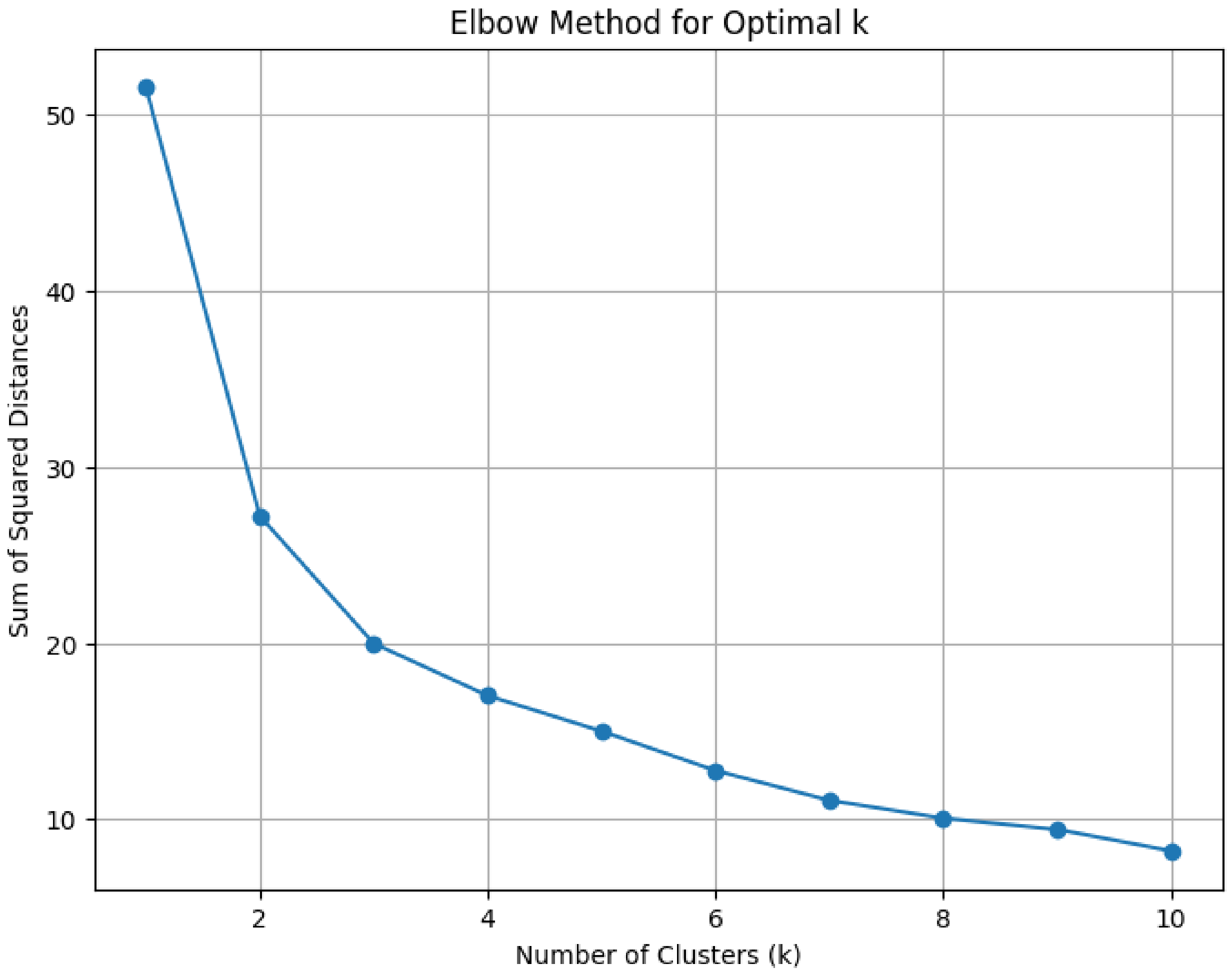


Upon applying the K-means algorithm to the entire training dataset and using the Elbow method, a plot of the WSS against the number of clusters was generated. The graph indicates a significant reduction in WSS as the number of clusters increases from 1 to 4. However, beyond 4 clusters, the reduction in WSS is marginal, suggesting that additional clusters do not capture substantial additional variance within the data. The Elbow method graph provided suggests that the optimal number of clusters for the dataset is 4. This is inferred from the observation that the elbow point is located at 4, where the rate of decrease in WSS no longer justifies the complexity of adding more clusters. The choice of 4 clusters is thus a trade-off between maximizing variance explained and maintaining model simplicity.

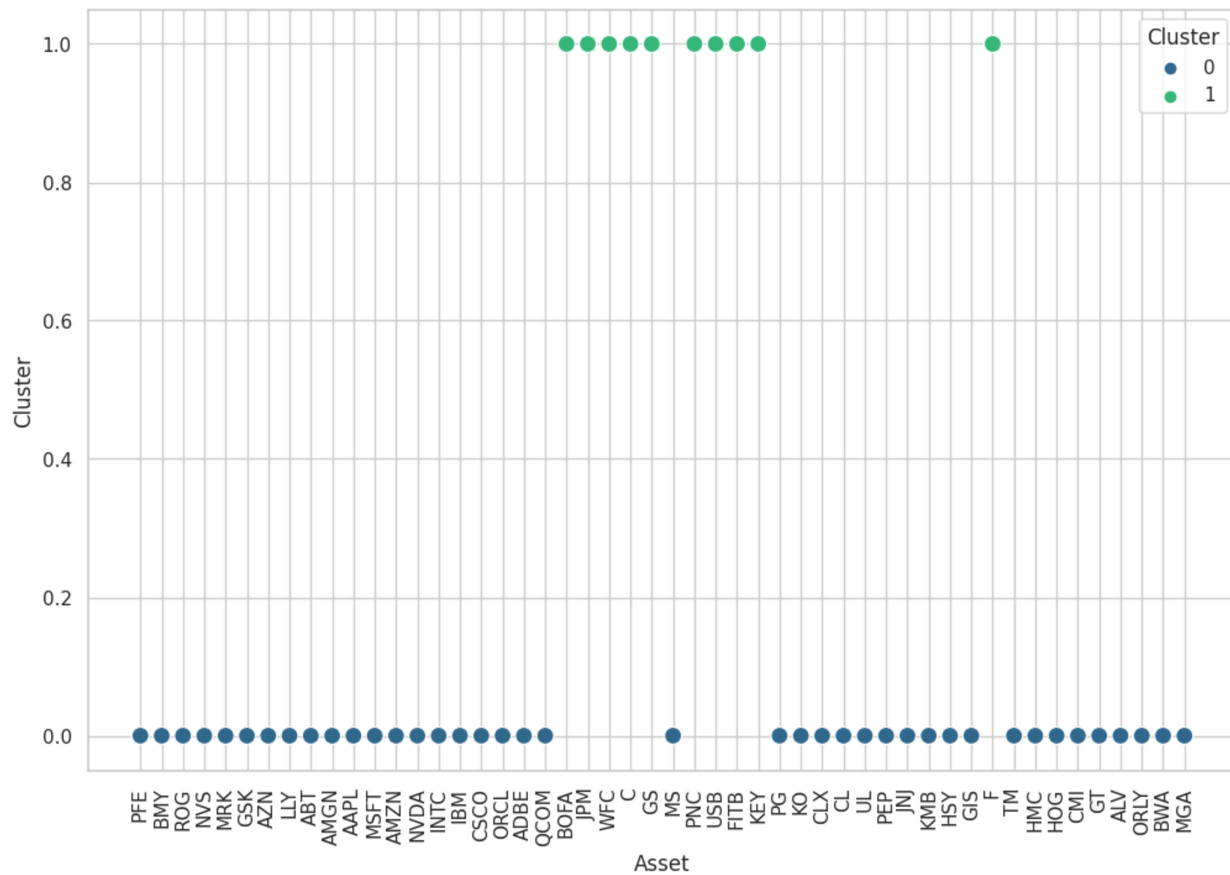


The scatter plot above displays the results of the K-means clustering analysis. In this plot, individual points represent assets that have been grouped into clusters based on their similarities across certain dimensions or features not shown here. The assets are plotted along the x-axis, while the y-axis corresponds to the cluster each asset has been assigned to. Each cluster is represented by a different color. The graph shows that the purple cluster contains a significant number of assets. The fact that these assets are grouped together suggests that they share similar characteristics which are distinct from those in other clusters. It's the largest cluster in terms of the number of assets and it consists of Pharma and FMCG Stocks. All other sectors have their own cluster. For example, Technology is in yellow, and Banking is in Blue. It is interesting to note that Rogers Corp belongs to the Pharma Sector but is in the same cluster as Auto Sector (in Blue).

During the banking crisis period: To figure out if the number of clusters is changing, we performed the Elbow Method and K-Means Clustering during the Banking Crisis (2007-2009)



Upon analysis of the dataset from the banking crisis period using the K-means algorithm and the Elbow method, it was observed that the optimal number of clusters is 2. This conclusion was reached by examining the within-cluster sum of squares (WSS), which showed a notable decrease when the number of clusters increased to 2. Further increments in cluster numbers beyond 2 resulted in only marginal reductions in WSS, indicating that additional clusters would not significantly improve the variance captured by the model.



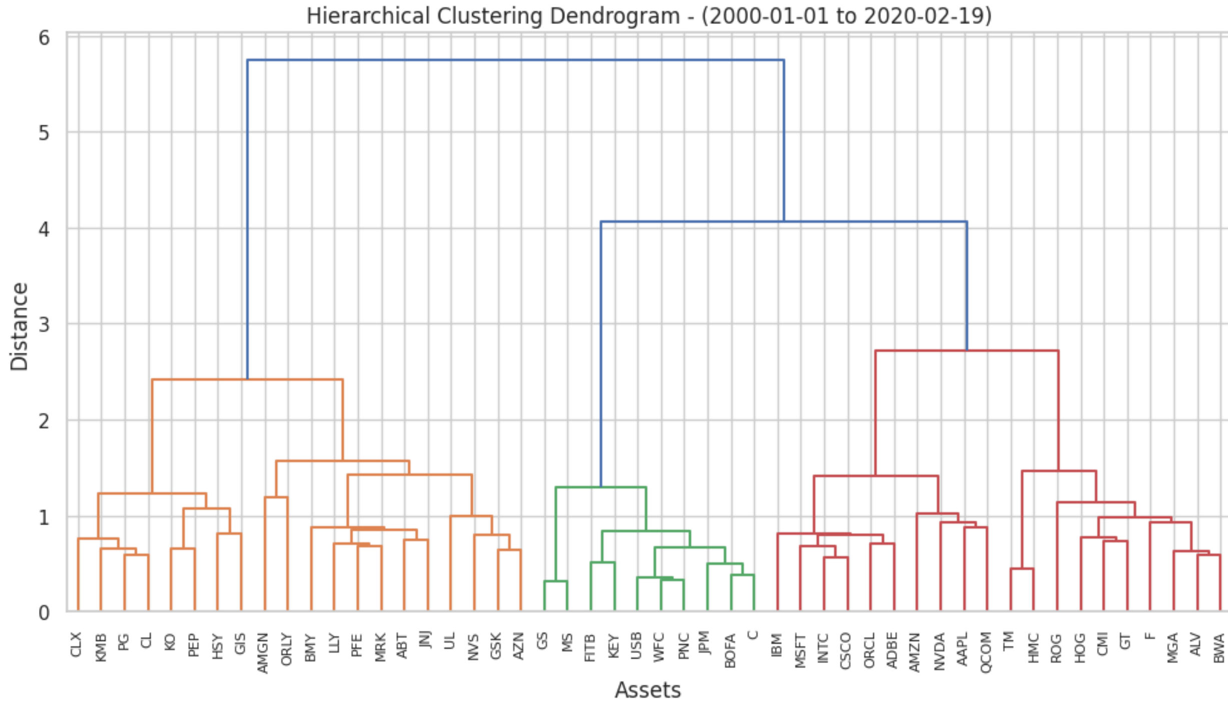
The scatterplot illustrates the clustering of various stocks into two distinct groups using the K-means algorithm during the banking crisis. The blue dots represent stocks from four different sectors: technology, automotive, pharmaceuticals, and fast-moving consumer goods (FMCG). The green dots, on the other hand, represent stocks from the banking sector. This segregation into its own cluster suggests that the banking sector stocks were distinctly different in performance or characteristics from the stocks of the other four sectors during the crisis. The clustering results suggest that during the banking crisis, stocks from the technology, automotive, pharmaceutical, and FMCG sectors performed similarly enough to be grouped into the same cluster by the K-means algorithm. This could be indicative of a broader market trend where these sectors, despite their different market dynamics, were similarly affected by the economic environment of the crisis. In contrast, the banking sector formed its own separate cluster, implying that the factors affecting the banking sector were unique to it during the crisis.

Hierarchical Clustering

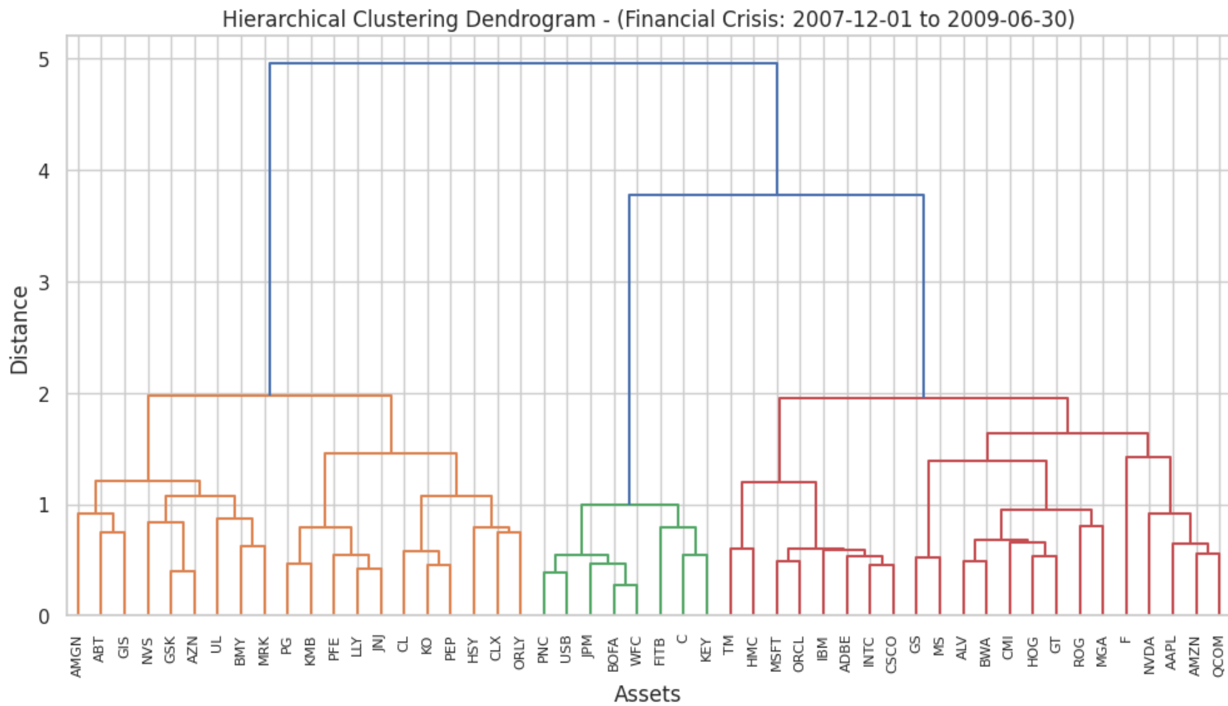
Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. In general, the goal is to create a tree of clusters called a dendrogram, where similar objects are grouped in nearby branches. There are two main strategies for hierarchical clustering:

Agglomerative (Bottom-Up): This approach starts by treating each data point as a single cluster. Then, iteratively, the closest pairs of clusters are merged until all points have been merged into a single cluster that contains all data points.

Divisive (Top-Down): This approach starts with all the data points in a single cluster. At each step, the cluster is divided into smaller clusters, until eventually each data point is in its own cluster.



We decided to use Agglomerative Approach as it is more computationally efficient. The provided graph is a dendrogram resulting from an agglomerative hierarchical clustering analysis of financial assets between January 1, 2000, and February 19, 2020. In this analysis, each asset is initially considered as a separate cluster, and clusters are successively merged based on their similarity, typically calculated using a distance metric such as Euclidean distance. The dendrogram visually represents the similarity between assets, with the vertical lines indicating individual clusters and the height of the horizontal joins reflecting the distance at which clusters are merged. Lower horizontal joins correspond to clusters with higher similarity.



The graph shows the dendrogram resulting from an agglomerative hierarchical clustering analysis. It visualizes how different assets (likely stocks, given their ticker symbols) are grouped based on their similarity, which is typically calculated using a distance metric like Euclidean distance. Each vertical line represents a cluster, with the height of the horizontal connections indicating the distance at which clusters are merged. The color coding may represent different levels of hierarchy, with closer clusters in the same color indicating higher similarity. For example, clusters that merge at the lower distance levels (closer to the bottom of the graph) are more like each other than those that merge at higher levels. If we compare the hierarchical dendrogram created for the entire period and the banking crisis period, we can see the banking period shows more similarity due to them merging at shorter distances.

Score Comparison

To figure out which clustering method performs better, we calculated three different scores for the training and the banking crisis periods.

For the training period:
K-Means Clustering:
Silhouette Score : 0.35
Calinski Harabasz Score: 34.63
Davies Bouldin Score: 1.05

Hierarchical Clustering:
Silhouette Score : 0.28
Calinski Harabasz Score: 20.21
Davies Bouldin Score: 1.39

Based on the provided scores, K-means clustering outperforms hierarchical clustering for the dataset in question. The Silhouette Score for K-means is higher at 0.35 compared to 0.28 for hierarchical, indicating better cohesion and separation of clusters in K-means. The Calinski-Harabasz Score, which assesses cluster validity based on the mean between and within-cluster variance, is significantly higher for K-means (34.63) than hierarchical (20.21), suggesting K-means clusters are more densely packed and better separated. Lastly, the Davies-Bouldin Score, where lower values represent clusters that are further apart, is lower for K-means (1.05) than hierarchical (1.39), indicating a more favorable clustering structure. Therefore, according to these scores, K-means is the superior clustering method for the data provided.

For the banking crisis period:
K-Means Clustering:
Silhouette Score : 0.34
Calinski Harabasz Score: 34.36
Davies Bouldin Score: 1.01

Hierarchical Clustering:
Silhouette Score : 0.33
Calinski Harabasz Score: 32.71
Davies Bouldin Score: 1.16

In this comparison, the scores indicate that K-means clustering, and hierarchical clustering perform similarly on the dataset, with K-means having a slight edge. The Silhouette Scores are almost equivalent, with K-means at 0.34 and hierarchical at 0.33, suggesting that both methods have similar levels of cluster tightness and separation. The Calinski-Harabasz Score, which is higher for clusters that are well-defined and separated, is marginally better for K-means (34.36) compared to hierarchical clustering (32.71). Lastly, the Davies-Bouldin Score, which should be lower for better clustering quality, is again slightly better for K-means (1.01) than hierarchical (1.16). Overall, while the differences are minor, K-means clustering appears to have a slight advantage over hierarchical clustering according to these metrics.

Market Regimes:

Market regimes are clusters of persistent market conditions that affect the behavior and performance of financial markets and assets. They can be influenced by various factors, such as macroeconomic trends, investor sentiment, volatility, liquidity, and risk appetite. Market regimes can have different characteristics, such as duration, frequency, and magnitude, and can change over time due to shocks or transitions.

Following is the list of events that we have taken to study market regimes:

Financial Crisis 2008: The global financial crisis of 2008-2009 was triggered by the collapse of the U.S. housing market and the subsequent failure of several major financial institutions. It resulted in a severe contraction of liquidity and credit, a sharp decline in asset prices, and a deep recession in many countries. The market regime during this period was characterized by high volatility, low confidence, and risk aversion. Investors fled to safe-haven assets, such as U.S.

Treasuries and gold, and avoided riskier assets, such as stocks and emerging markets. The crisis also prompted unprecedented policy responses from central banks and governments, such as quantitative easing, fiscal stimulus, and regulatory reforms.

Covid – 19: The coronavirus pandemic of 2020-2021 caused a major disruption to the global economy and society, as lockdowns, travel restrictions, and social distancing measures were imposed to contain the spread of the virus. The market regime during this period was marked by high uncertainty, low growth, and low inflation. Investors initially reacted to the pandemic by selling off risky assets and seeking liquidity, causing a sharp drop in equity and bond markets in March 2020. However, the markets recovered quickly, thanks to the massive monetary and fiscal support from central banks and governments, as well as the development of vaccines and treatments. The pandemic also accelerated some existing trends, such as digitalization, e-commerce, and green transition, creating new opportunities and challenges for different sectors and companies.

Government Bond Hike Rates: Government bond yields reflect the cost of borrowing for sovereign issuers, as well as the expectations of future inflation and economic growth. Government bond yields can rise or fall due to various factors, such as changes in monetary policy, fiscal policy, supply and demand, and market sentiment. The market regime associated with rising government bond yields depends on the underlying causes and implications of the yield movements. For example, if government bond yields rise due to higher inflation expectations, that could signal a regime of higher growth and higher risk appetite, which could benefit cyclical and value stocks, but hurt growth and defensive stocks. On the other hand, if government bond yields rise due to higher real interest rates, that could signal a regime of tighter monetary policy and lower liquidity, which could weigh on both equity and bond markets.

For better understanding, we have devised one training period and two testing periods:

- **Training period: 2000-2018**
- **Testing period 1: 2019-2020**
- **Testing period 2: 2021-2023**

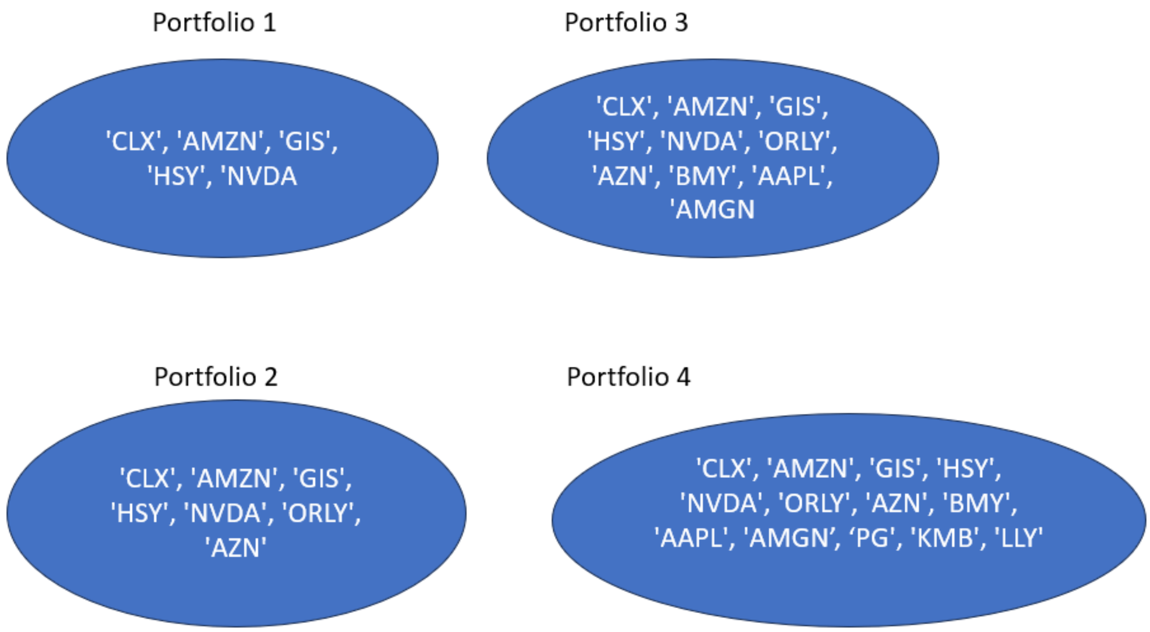
Portfolios Constructed Based on Clustering and Network Analysis

Clustering and network analysis are two methods that can be used to group assets based on their similarity and dependence structure. Clustering aims to minimize the dissimilarity within each group and maximize the dissimilarity between different groups. Network analysis represents the assets as nodes and their correlations as edges and identifies the most connected or influential nodes in the network.

Portfolios constructed based on clustering and network analysis can have several advantages over traditional portfolios, such as:

- Reducing the dimensionality and complexity of the portfolio optimization problem, by selecting representative assets from each cluster or network.

- Enhancing the diversification and stability of the portfolio, by avoiding highly correlated or redundant assets and capturing the dynamic changes in the market structure.
- Improving the risk-adjusted performance of the portfolio, by exploiting the information embedded in the clustering and network patterns.



Computing Weights for Portfolios

A. GMV
B. MVO

A. Global Minimum Variance

It refers to the portfolio with the lowest possible variance (or standard deviation) of returns among all possible portfolios of a set of assets. In other words, the Global Minimum Variance portfolio is the portfolio that offers the highest level of diversification, given a set of assets, to minimize overall portfolio risk. We have used the following steps to compute the weights for assets in the Global Minimum Variance portfolio:

1. Covariance and Variance Matrix: Compute the covariance matrix for the returns of the assets in the portfolio. The diagonal elements of this matrix represent the variances of individual assets, and off-diagonal elements represent the variances between asset pairs.

2. Inverse of Covariance Matrix: Calculate the inverse of the covariance matrix. This matrix is denoted as $\sum^{-1}$.

3. Vector of Ones: Create a vector of ones, denoted as **1**, with a length equal to the number of assets

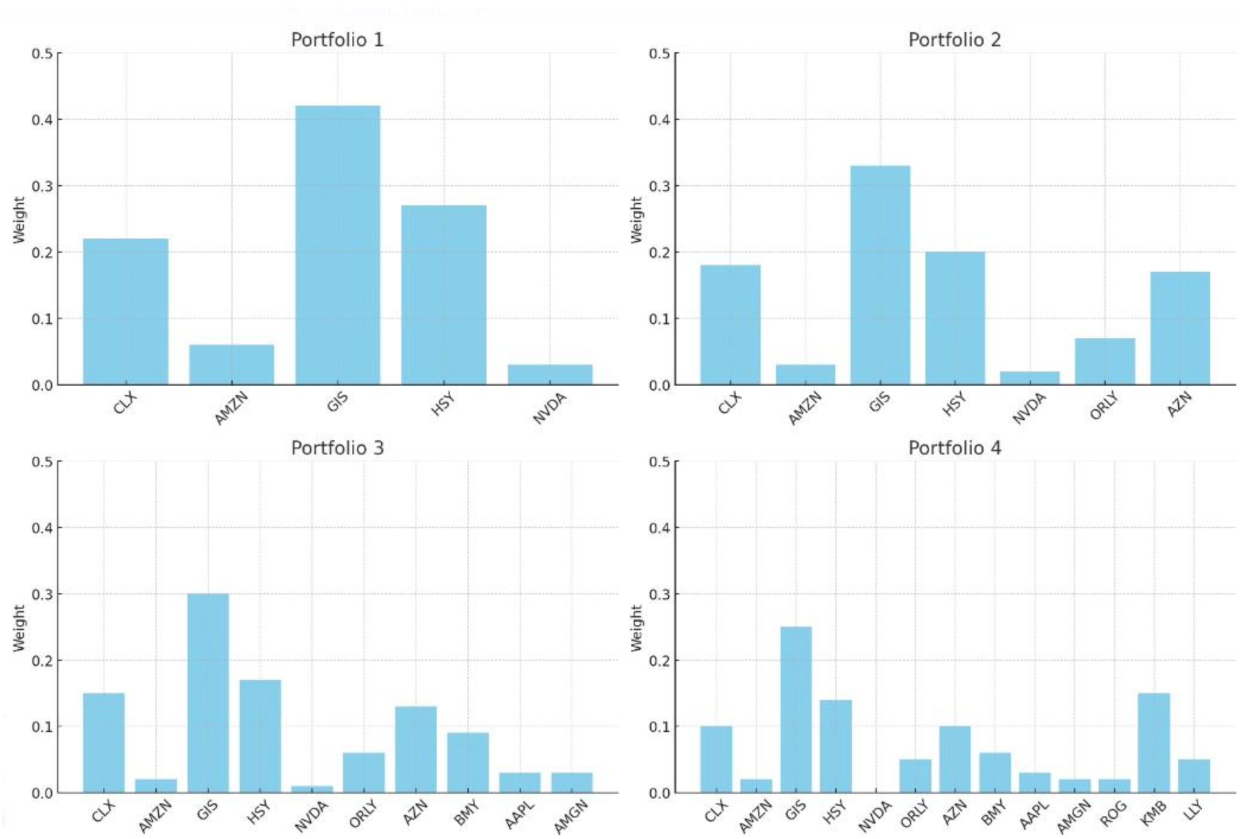
4. Calculation of GMV Weights: The weights for the Global Minimum Variance portfolio, denoted as $\mathbf{w}_{\text{GMV}}$.

   Here

$$\mathbf{w}_{\text{GMV}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$$

- $\mathbf{w}_{\text{GMV}}$ is the vector of weights for the GMV portfolio.
- $\sum^{-1}$ is the inverse of a covariance matrix.
- **1** is the vector of ones.
- 1^T represents the transpose of the vector of ones.
- $1^T\sum^{-1}$ is a scaler.

The intuition behind this formula is to find the weights that minimize the portfolio variance. The denominator ensures that the weights sum to 1. Investors use the Global Minimum Variance portfolio as a reference point in portfolio construction, and it plays a crucial role in the development of the Capital Market Line and the efficient frontier in Modern Portfolio Theory.



B. Mean-Variance Optimization

Mean-variance optimization (MVO), also known as modern portfolio theory (MPT), is a framework for constructing an investment portfolio that aims to maximize expected return for a given level of risk, or equivalently, minimize risk for a given level of expected return. The basic idea behind MVO is to find the optimal allocation of assets that provides the best trade-off between risk and return.

1. Calculate Mean Returns and Covariance Matrix:

   - Mean Returns: $\bar{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t}$ where $R_{i,t}$ is the return of asset i at time t, and T is the number of observations.

   - Covariance Matrix: $\Sigma = \frac{1}{T}(X - \bar{X})^T(X - \bar{X})$ where X is the matrix of asset returns and $\bar{X}$ is the mean vector.

2. Number of Assets: Count the number of assets in the portfolio, denoted as N.

3. Define Variables:
   w_i : Weight of asset i in the portfolio.

4. Define Expected Return and Portfolio Variance:
   - Expected Portfolio Return: $E(R_p) = \sum_{i=1}^{N} w_i \bar{R}_i$
   - $\mathrm{Var}(R_p) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}$ where σij is the covariance between the returns of assets i and j .

5. Define the Objective Function (Setting the Objective function to Maximize Sharpe Ratio):

   - Sharpe Ratio:

$$S = \frac{E(R_p) - R_f}{\sqrt{Var(R_p)}}$$

   - Objective Function: Maximize S or equivalently maximize

     $E(R_p) - \lambda Var(R_p)$, here λ is the risk aversion parameter.

6. Define Constraints:

   - Fully invested constraint: $\sum_{i=1}^{N} \omega_i = 1$

- Minimum weight constraint: $\omega_i \geq$ min_weight , where min_weight is a predefined minimum weight for each asset.

7. Formulate the problem:

Maximize the objective function subject to the constraints:

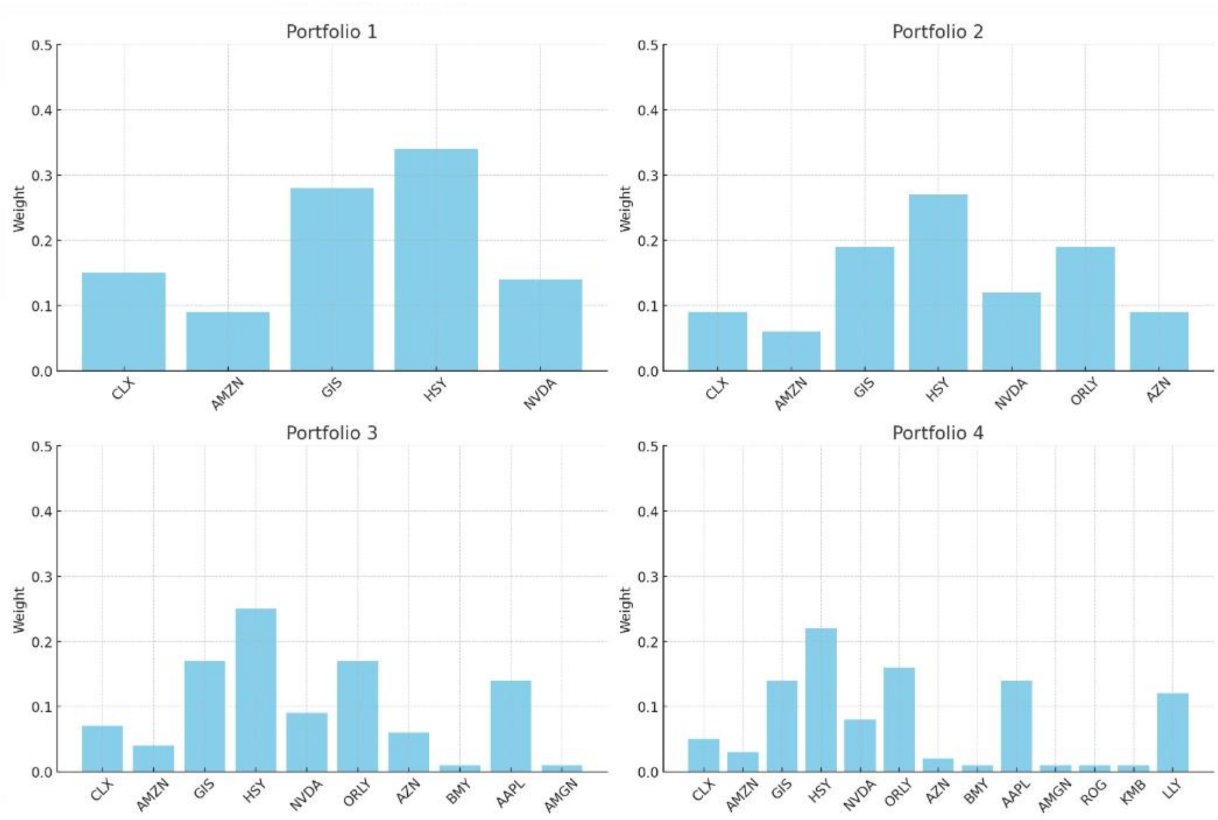
$$\max_w E(R_p) - \lambda Var(R_p) \text{ s.t. } \sum_{i=1}^{N} \omega_i = 1 \text{ and } \omega_i \geq \text{min\_weight}.$$

8. Solve the Optimization Problem: Utilize optimization techniques (e.g., quadratic programming) to find the optimal weights that maximize the objective function under the given constraints.

9. Normalize weights:

Normalize the obtained weights:

$$w_i = \frac{w_i}{\sum_{i=1}^{N} w_i}$$



Portfolio Returns:

Portfolio return is a measure that represents the expected performance of a portfolio, considering the individual returns of the assets within the portfolio and their respective weights. The portfolio return is a weighted average of the returns of the individual assets, where the weights reflect the proportion of the total investment allocated to each asset. The formula for calculating the portfolio return $E(Rp)$ is given by:

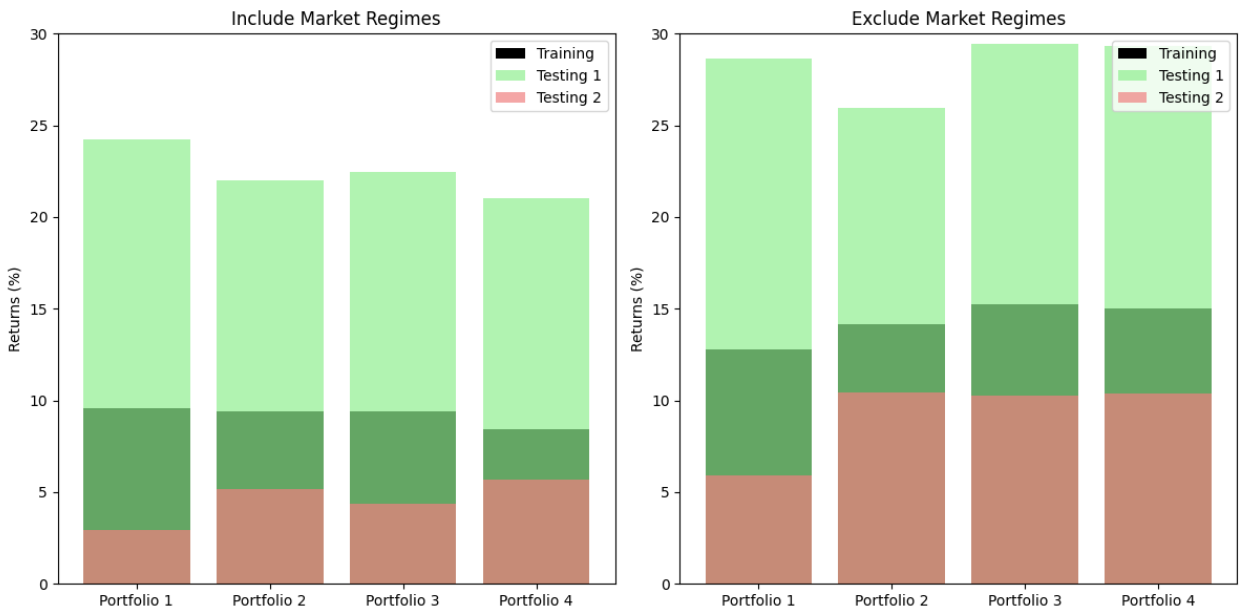
$$E(R_p) = \sum_{i=1}^{N} w_i \bar{R}_i$$

N is the number of assets in the portfolio.

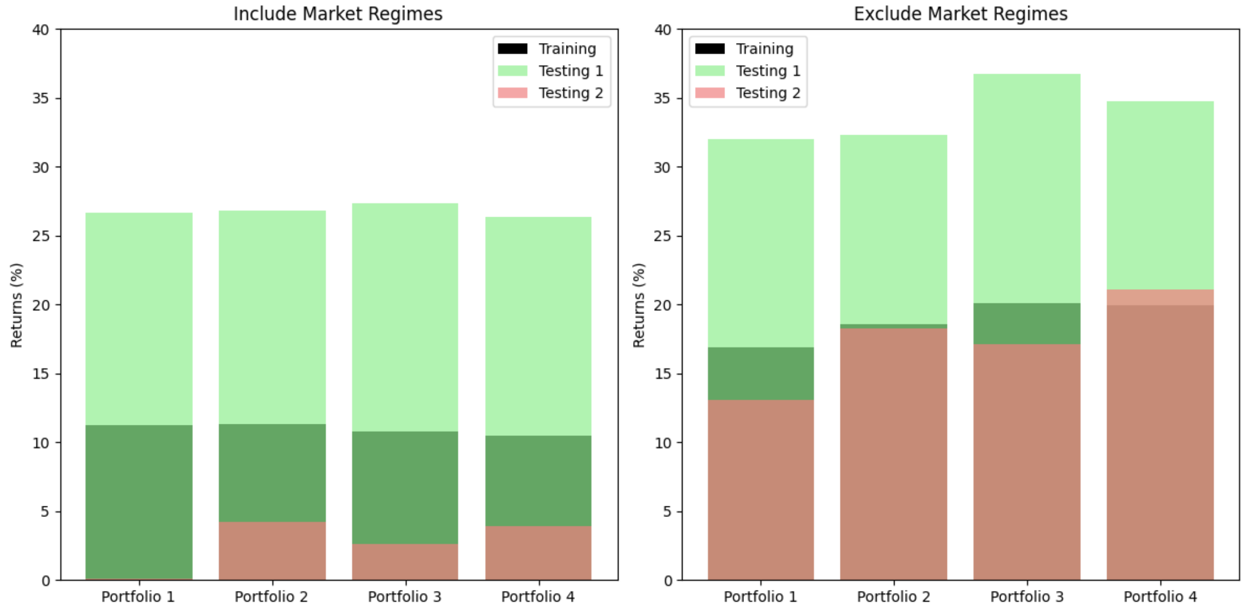
ω_i is the weight of asset i in the portfolio (proportion of total investment)

$\bar{R}_i$ is the mean return of asset i.

A. GMV Based Portfolio Returns



B. MVO Based Portfolio Returns



- Market regimes are periods when the stock market behaves in a certain way, such as rising, falling, or being volatile. Market regimes can be influenced by various factors, such as economic conditions, political events, investor sentiment, and technological innovations.

- COVID-19, the financial crisis of 2008, and government bond rate hikes are examples of events that can cause market regime changes, as they affect the supply and demand of goods and services, the availability and cost of credit, and the expectations and confidence of investors.

- According to the charts, we can say that the stock market experienced a regime change in 2020, as the COVID-19 pandemic triggered a sharp selloff in March, followed by a rapid recovery in the subsequent months, driven by fiscal and monetary stimulus, vaccine development, and digital acceleration. This recovery was uneven across sectors and regions, as some industries and countries were more resilient and adaptable to the pandemic than others.

- The 2019-2020 testing 1 period may have generated high returns for some investors who bought stocks at low prices in March 2020 and sold them at higher prices later in the year. However, this strategy also involved high risks, as the market was very uncertain and volatile during that time. Moreover, the past performance of the market does not guarantee future results, as new events and factors may affect the market regime in the future.

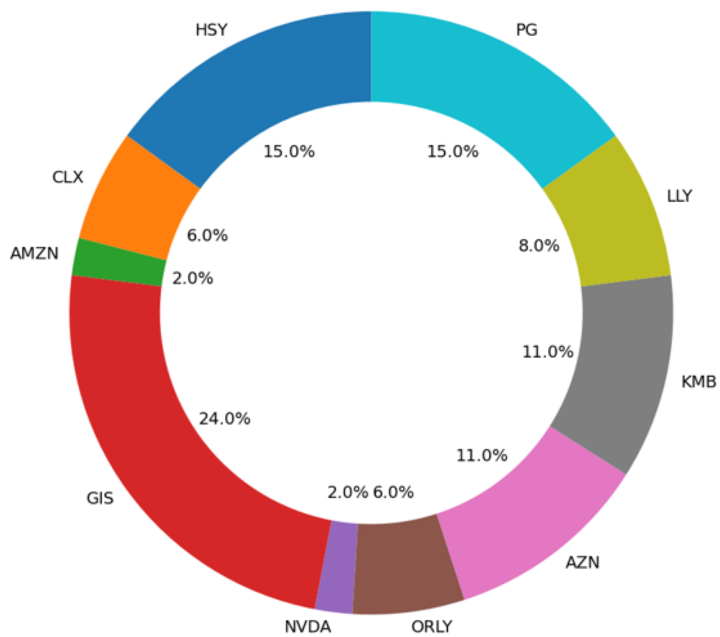
- Moreover, training has a higher return than testing 2 period because it considers 18 years and testing 2 period was severely affected due to government bond hike rates and Russia-Ukraine war.

Portfolios Considered:

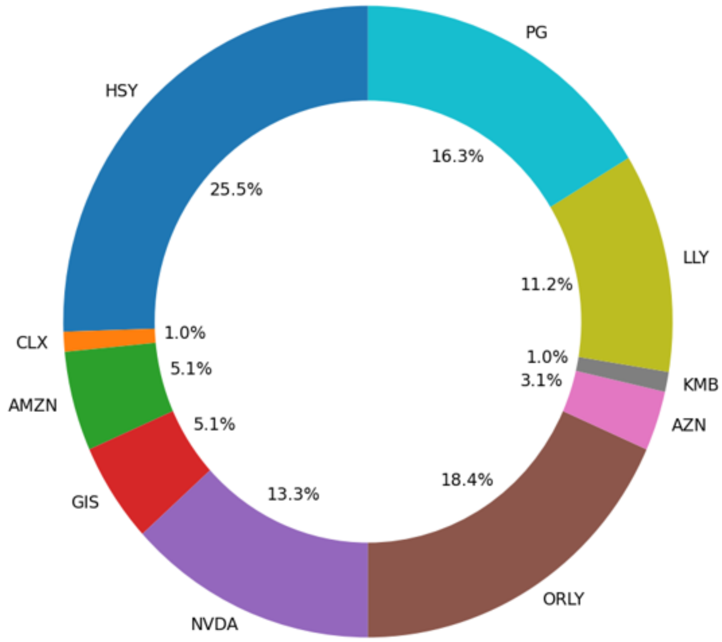
We have considered the portfolio with 10 stocks for final comparison with benchmark instead of the portfolios with 5, 7 or 13. With 10 stocks, there is efficient diversification to mitigate risk, and managing the portfolio becomes more practical for investors and fund managers. The diminishing marginal benefit of adding more stocks beyond 10 is considered, optimizing the trade-off between risk reduction and portfolio management efforts. This choice aligns well with benchmark indices that often consist of a moderate number of securities. A portfolio of 10 stocks allows for focused risk management and in-depth analysis of each security, enhancing overall decision-making. Resource constraints, including time and research capabilities, make a 10-stock portfolio more feasible for many investors. Comparing this portfolio to a benchmark simplifies the evaluation process, providing a clear metric for performance assessment. Ultimately, the selection of 10 stocks strikes a balance between diversification, practicality, and effective benchmark comparison in the dynamic landscape of investment.

'CLX', 'AMZN', 'GIS', 'HSY', 'NVDA', 'ORLY', 'AZN', 'BMY', 'AAPL', 'AMGN

A. GMV Weights



B. MVO Weights



Performance Statistics

1. **Treynor Ratio:** The Treynor Ratio measures the excess return per unit of systematic risk (beta). It is particularly useful for evaluating the performance of an investment or portfolio in relation to its exposure to market risk. A higher Treynor Ratio indicates better risk-adjusted performance.

$$Treynor\ Ratio = \frac{R_p - R_f}{\beta_p}$$

2. **Information Ratio**: The Information Ratio assesses the active return of a portfolio per unit of active risk (tracking error) relative to a benchmark. It helps evaluate the value added by active management. A higher Information Ratio indicates more efficient active management.

$$Information\ Ratio = \frac{R_p - R_b}{\sigma_p}$$

3. **Sharpe Ratio:** The Sharpe Ratio measures the excess return per unit of total risk (standard deviation). It evaluates the risk-adjusted performance of a portfolio. A higher Sharpe Ratio indicates better risk-adjusted returns.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

4. **Sortino Ratio:** The Sortino Ratio is like the Sharpe Ratio but considers only the downside risk (volatility of returns below a target or minimum acceptable return). It provides a measure of risk-adjusted performance focusing on downside volatility.

$$Sortino\ Ratio = \frac{R_p - R_f}{\sigma_{downside}}$$

5. **Calmar ratio:** The Calmar Ratio assesses the risk-adjusted performance by comparing the average annual return to the maximum drawdown. It is commonly used in the evaluation of hedge funds and trend-following strategies.

$$Calmar\ Ratio = \frac{R_p - R_f}{\text{Maximum Drawdown}}$$

6. **Jensen's alpha:** Jensen's Alpha measures the excess return of a portfolio compared to its expected return based on its beta and the market risk premium. A positive alpha indicates that the portfolio outperformed its expected return, considering systematic risk.

$$Jensen's\ Alpha = R_p - [R_f + \beta_p \cdot (R_m - R_f)]$$

Results and Analysis

A) Treynor Ratio: 0.13
The Treynor Ratio of 0.13 indicates that the portfolio or investment has generated a positive excess return relative to the risk-free rate (or benchmark) per unit of systematic risk (beta). A higher Treynor Ratio suggests better risk-adjusted performance.

B) Information Ratio: 0.05
The Information Ratio of 0.05 suggests that the portfolio's active return, beyond the benchmark, is positive relative to its active risk (tracking error). While positive, the small value indicates a modest level of outperformance for the risk taken.

C) Sharpe Ratio: 1.03
The Sharpe Ratio of 1.03 signifies that the portfolio has delivered a strong excess return per unit of total risk (standard deviation). A Sharpe Ratio above 1 is generally considered good, indicating attractive risk-adjusted returns.

D) Sortino Ratio: 0.12
The Sortino Ratio of 0.12 measures the excess return of the portfolio relative to its downside risk (volatility of returns below a target or minimum acceptable return). A positive Sortino Ratio suggests that the portfolio is delivering returns with a focus on minimizing downside risk.

E) Calmar Ratio: 0.18
The Calmar Ratio of 0.18 assesses the risk-adjusted performance by comparing the average annual return to the maximum drawdown. A Calmar Ratio of 0.18 indicates that the portfolio has provided a positive average return relative to its maximum drawdown.

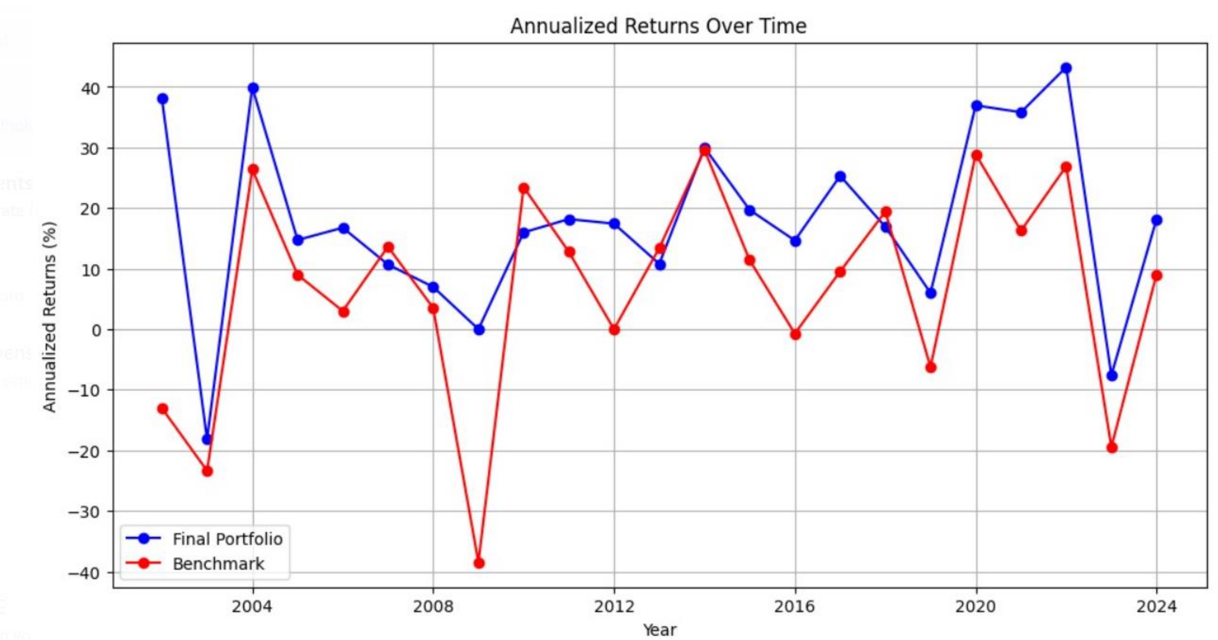
F) Jensen's Alpha: 0.08
Jensen's Alpha of 0.08 suggests that the portfolio has outperformed its expected return based on its beta and the market risk premium. A positive alpha indicates that the portfolio's performance exceeded what would be predicted by its level of systematic risk.

In summary, these ratios collectively provide insights into the risk-adjusted returns and performance of the portfolio. The positive values across most ratios indicate that the portfolio has generally performed well, considering the level of risk taken. Investors often use a combination of these metrics to comprehensively assess and compare different investment opportunities.

Comparison with the Benchmark

A. Annualized returns over time

For the final comparison purpose, we have taken S&P500 as a benchmark. Now let's understand how our portfolio performed better as compared to the benchmark:



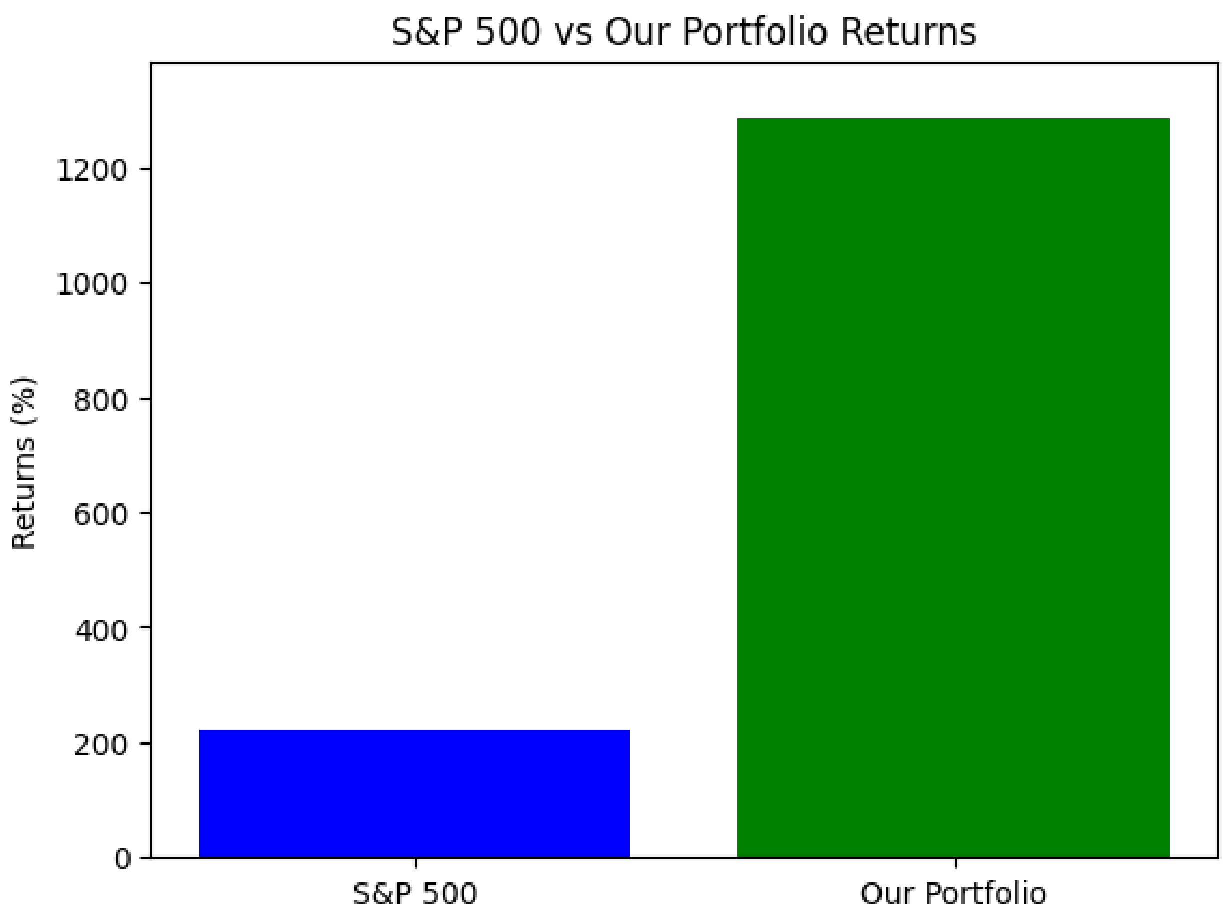
- The line graph comparing the annualized returns of "Our Portfolio" and a "Benchmark" from the years 2000 to 2023 provides valuable insights into the relative performance of these investments over time. Annualized returns, representing the average returns earned each year over a specific period, serve as a crucial metric for assessing the financial performance and attractiveness of investment portfolios.

- Upon examination of the graph, it is evident that "Our Portfolio" exhibits a generally superior performance compared to the "Benchmark" across the entire timeline, particularly in recent years. The upward trend in the line representing "Our Portfolio" suggests that the investment has consistently generated higher average returns on an annual basis. This outperformance may be attributed to effective investment strategies, superior stock selection, or other factors contributing to the portfolio's success.

- Conversely, the line representing the "Benchmark" displays fluctuations and variability in annualized returns throughout the observed period. These fluctuations indicate that the benchmark's performance has been subject to market dynamics, economic conditions, and potential shifts in the underlying index or asset composition. The fact that the benchmark's returns are not consistently higher than those of "Our Portfolio" underscores the effectiveness of the latter's investment strategy.

- The significant fluctuations in both lines over the years imply that market conditions and investment strategies have experienced changes. These fluctuations could be influenced by economic events, policy changes, industry trends, or other external factors affecting the financial markets. Investors and analysts need to delve deeper into these fluctuations to understand the specific drivers behind the performance disparities between "Our Portfolio" and the "Benchmark."

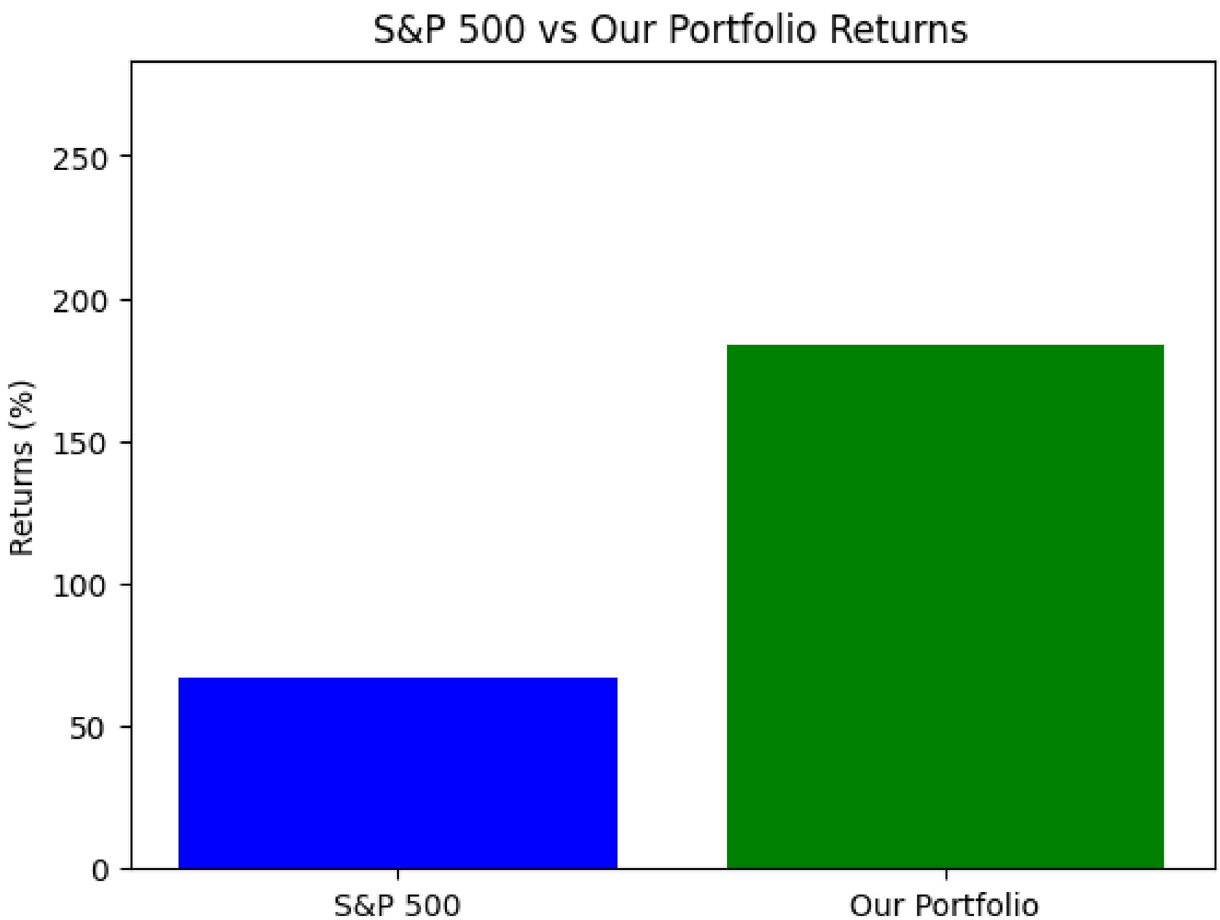
B. Absolute Returns over the Entire Period



According to the bar chart, “Our Portfolio” has significantly higher returns than the S&P 500, which is a widely used benchmark for the US stock market. The S&P 500 has a return of about 200%, while “Our Portfolio” has a return of about 1200%. This means that if you invested $100 in the S&P 500 at the beginning of the period, you would have $300 at the end of the period.

However, if you invested $100 in “Our Portfolio” in the beginning of the period, you would have $1300 at the end of the period.

C. Absolute Returns over the Testing Period



- The graph indicates that “Our Portfolio” has significantly outperformed the S&P 500 in terms of returns. The green bar representing “Our Portfolio” is almost four times higher than the blue bar representing the S&P 500.

- This suggests that “Our Portfolio” has either taken more risk, diversified better, or timed the market more effectively than the S&P 500. Alternatively, it could also mean that the graph is based on a short or selective time frame that does not reflect the long-term performance of both strategies.

References:

Parmentier, Loïc. Measures of Portfolio' Diversification.

Juan A. Serur and Marco Avellaneda. Hierarchical PCA and Modeling Asset Correlations

Jing and Rocha ("A network-based strategy of price correlations for optimal cryptocurrency portfolios," 2023)

Marti et al. ("A review of two decades of correlations, hierarchies, networks and clustering in financial markets")

Tang et al. ("Asset Selection via Correlation Blockmodel Clustering," 2021)

Akansu et al. ("Quant investing in cluster portfolios," 2021)

Giudici et al. ("Network models to improve robot advisory portfolios," 2022)

Lim and Ong ("Portfolio Diversification Using Shape-Based Clustering, " 2021)

Olmo ("Optimal portfolio allocation and asset centrality revisited, " 2021)

Peralta and Zareei ("A network approach to portfolio selection," 2016)

Raffinot ("Hierarchical Clustering-Based Asset Allocation, " 201)

Raffinot ("The Hierarchical Equal Risk Contribution Portfolio," 2018)

Sakurai et al. ("Correlation Diversified Passive Portfolio Strategy Based on Permutation of Assets")

Snow ("Machine Learning in Asset Management - Part 2: Portfolio Construction - Weight Optimization, 2020”)

Giudici et al. ("Network models to improve robot advisory portfolios, " 2022)

Jaeger et al. ("Interpretable Machine Learning for Diversified Portfolio Construction, " 2021)

Jaeger et al. ("Adaptive Serial Risk Parity and other Extensions for Heuristic Portfolio Construction using Machine Learning and Graph Theory," 2021)

Li et al. ("Investable and Interpretable Machine Learning for Equities, " 2022)

Perrin and Roncalli ("Machine Learning Optimization Algorithms & Portfolio Allocation, " 2020)

Schwendner et al. ("Adaptive Serial Risk Parity and Other Extensions for Heuristic Portfolio Construction Using Machine Learning and Graph Theory, " 2021)

Kim et al. ("Personalized goal-based investing via multi-stage stochastic goal programming, " 2020)

Das et al. ("Dynamic optimization for multi-goals wealth management, " 2022

Cong et al. ("AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI, " 2022)