南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY | 统计与数据科学系 Department of Statistics and Data Science

STA320 Project Proposal Report American Stocks hirearchical Risk Based Portfolio Optimization

Members:

12110633 Vincent William Hadiasali
2111925 Nadia Donita
12111313 Zhuang Rongqin

Completion Date: July 31, 2025

1 Brief Description

In the realm of portfolio optimization, achieving an optimal balance between risk and return remains one of the most fundamental challenges for investors. Traditional methods often rely on simplistic strategies that fail to fully account for complex correlations between assets. This limitation often leads to suboptimal portfolio constructions, especially in environments with high volatility and tail risks. Recent advancements in hierarchical portfolio optimization seek to overcome these challenges by incorporating machine learning techniques and hierarchical clustering approaches. In this project, we want to try to optimize the portfolio filled with American stocks and beat the S&P500 Index benchmark as much as possible. Not only does this topic involve a novel application, but also comparison models. Furthermore, it involves an application of the Minimum-variance Markowitz portfolio theorem in backtesting to compare with the mentioned modern portfolio optimization method. To assist with the project, finance and portfolio related libraries are necessary.

2 Background Knowledge about HERC Portfolio Optimization

The "Hierarchical Equal Risk Contribution Portfolio" (HERC) approach, as proposed by Raffinot (2018), builds upon two a combination of two key methods in modern optimization, Hierarchical Risk Parity (HRP) and Hierarchical Clustering-Based Asset Allocation (HCAA).

HRP uses minimum spanning tree and recursive bisection to diversify risk. HCAA uses hierarchical clustering to allocate capital across and within clusters. HERC method involves hierarchical clustering from HCAA and recursive allocation from HRP. It allocates weights using equal risk contribution to ensure that each clusters contribute equally to the portfolio. This method improves upon traditional risk measures by incorporating downside risk metrics such as Conditional Value at Risk (CVaR) and Conditional Drawdown at Risk (CDaR) which better capture extreme losses. The allocation process in HERC first clusters the assets, determines the optimal number of clusters using the Gap statistic, then applies a top-down recursive bisection based on the dendrogram shape, assigning weights according to the risk contribution of each cluster. Within each cluster, a naive risk parity (equal-weight) approach is used, assuming high correlation among assets. Overall. HERC provides a risk-aware and robust approach, especially in environments with fat-tailed returns and higher drawdown risks. This ensures that the portfolio is balanced not only in terms of assets but also in risk exposure.

This project seeks to implement and evaluate the performance of the HERC portfolio optimization strategy, using both multi-asset and individual stock datasets, to demonstrate its ability to achieve superior diversification and manage risk effectively.

3 Methodology

3.1 Data Collection

The data collection steps are as follow:

1. With a specific stock index chosen, pick a number of stocks involved in the index. It is impossible to consider all stocks as not only it is cumbersome to download every single stock, but also the percentage allocation might be tiny for every stock. So, we choose 20 stocks.

2. After the number of stocks chosen, have a solid reason of why each stock is chosen. We think it is better to have a diversified industries in the portfolio.

3. Download all stocks and S&P 500 datas. Additionally, plot the time series stock data to confirm that it is consistent with the time series plot in Google. The S&P 500 Index daily data is downloaded from a verified trading software called Wind Financial Terminal. While for the chosen stocks, we use a website called Investing.com as the data instead of Wind Financial Terminal, Python

and R Quantitative Finance related libraries as Vincent confirmed that using those source libraries led to a different stock price pattern from Google Search due to decimal format place fault as he look over the Excel.

4. Download the risk free rate, which is U.S. Treasury Bonds, from the official website of U.S. Department of Treasury and choose coupon equivalent and the most common portfolio analysis interval which is 13 weeks. Note that the rate in table is annualized and in percentage form.

3.2 Data Analysis

The data analysis process is given below.

1. Returns are computed for each asset using various frequencies (daily, weekly, monthly, etc.). The compute_returns function is used to calculate returns depending on the specified frequency for each stock. In this project, we selected daily.

2. Observe the correlation heatmap of all stocks, if we have a lot of positive correlated stocks. Reduce the size as positive correlation means that there are additional risk aside from individual risk of portfolio.

3. After obtaining the returns data, the next step is to calculate the Agglomerative Coefficient (AC), which is used to measure the quality of hierarchical clustering.
   (a) This is done using different linkage methods: single, average, complete, and ward. The method that maximizes the agglomerative coefficient is selected as the best linkage method.
   (b) This step helps in selecting the optimal way to cluster assets based on their returns.

4. Decide the optimal number of clusters (optimal_k) by calculating the silhouette width for different values of k.
   (a) The silhouette method evaluates how well each asset fits into its assigned cluster. A higher silhouette width indicates better clustering quality.
   (b) The number of clusters (optimal_k) is selected based on the maximum silhouette width.

5. With the optimal linkage method and the number of clusters determined, the hierarchical clustering is performed using the hclust function.
   (a) The returns of assets are clustered hierarchically, and the data is split into optimal_k clusters.
   (b) A dendrogram is plotted to visualize the asset hierarchy, and branches are colored to indicate different clusters.

6. After clustering the assets, the next step is to assign weights to each cluster. Two types of inter-cluster weights can be chosen:
   - Risk-based weighting: This method assigns weights to each cluster based on the risk (standard deviation) of the individual assets within the cluster.
   - Equally-weighted: This method assigns equal weights to all clusters, regardless of their individual risk profiles.

7. Within each cluster, the assets are weighted based on their risk contributions, ensuring that more volatile assets receive lower weights. The naive risk parity method is used to compute these intra-cluster weights, where each asset's weight is inversely proportional to its risk.

8. The final portfolio weights are calculated by combining the intra-cluster weights with the inter-cluster weights. The portfolio is then constructed by allocating weights to each asset based on its cluster's risk profile and the cluster's weight.

9. The final portfolio allocation is plotted along with the hirearchical result.
10. The performance of the optimized portfolio is evaluated using metrics like the Sharpe Ratio. The Sharpe Ratio measures the risk-adjusted return of the portfolio. A higher Sharpe ratio indicates better performance relative to risk.

4 Exploratory Data Analysis

The correlation heatmap of all stocks and S&P500 index benchmark is given below

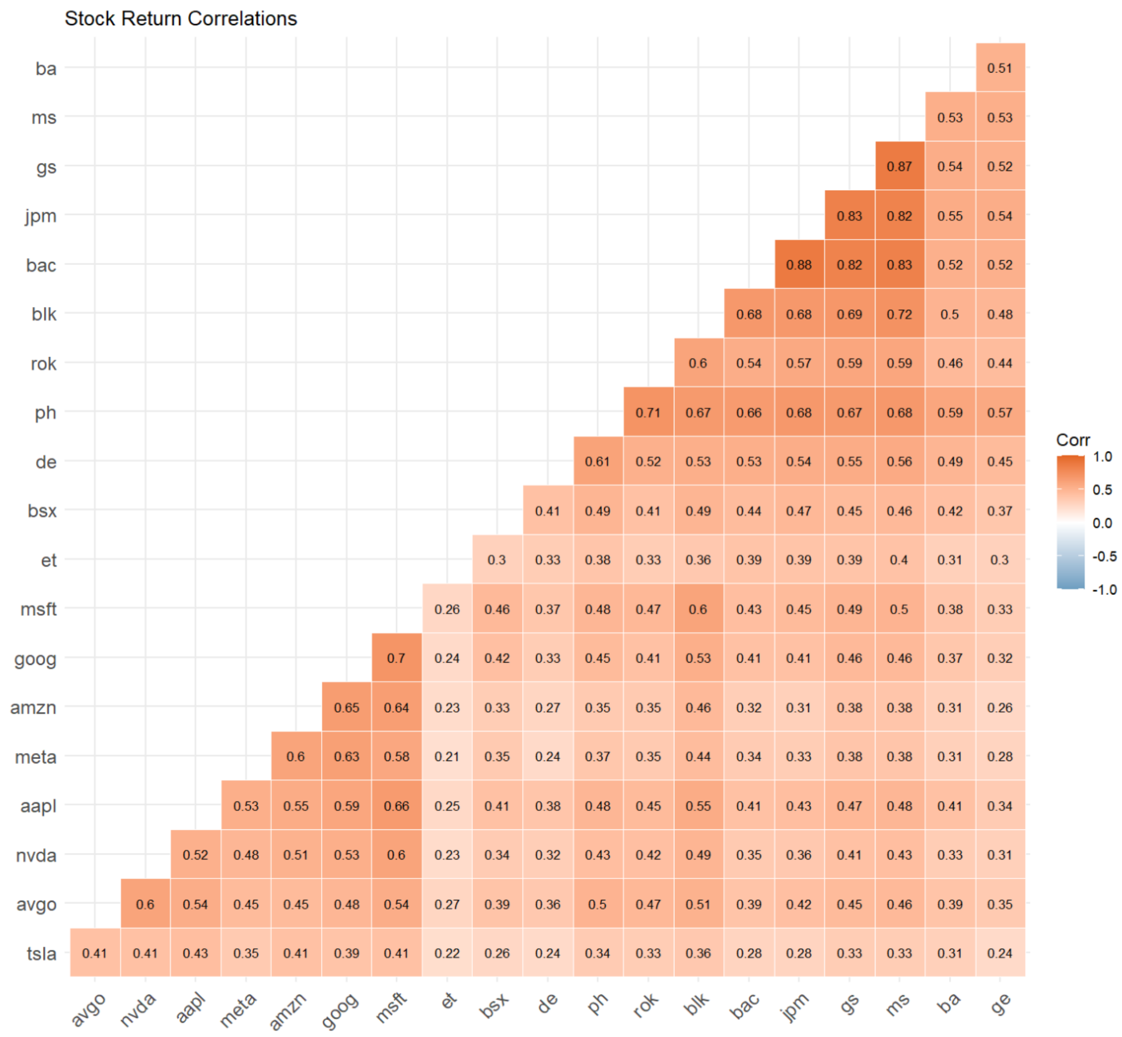

Figure 1: Heatmap of daily return from all stocks

Based on the overall heatmap result, investing in redundant amount of stocks will result in unnecessary additional risk. So, I decided to keep 9 stocks which are NVDA, TSLA, AVGO, AMZN, META, APPL, MSFT, BSX and GOOGL tickers if it were for investment that tries to reap as many benefit as possible.

5 Results

5.1 Hierarchical Clustering Weight distribution

According to the optimal k result, the outpput is 2.

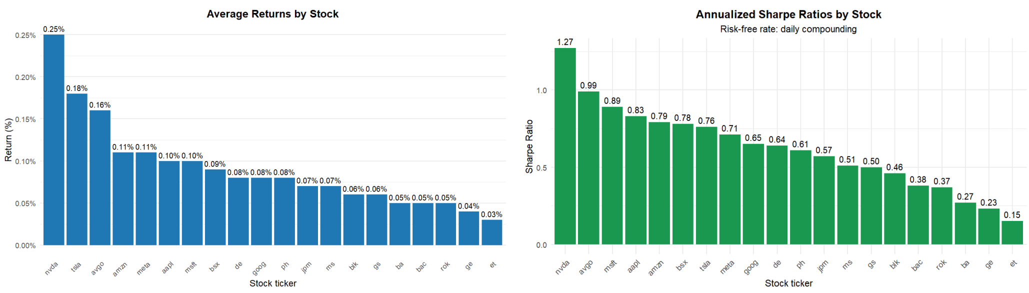

Figure 2: A pair of barplot showing the mean return of returns from all stocks and Annualized Sharpe Ratio both based on daily rate

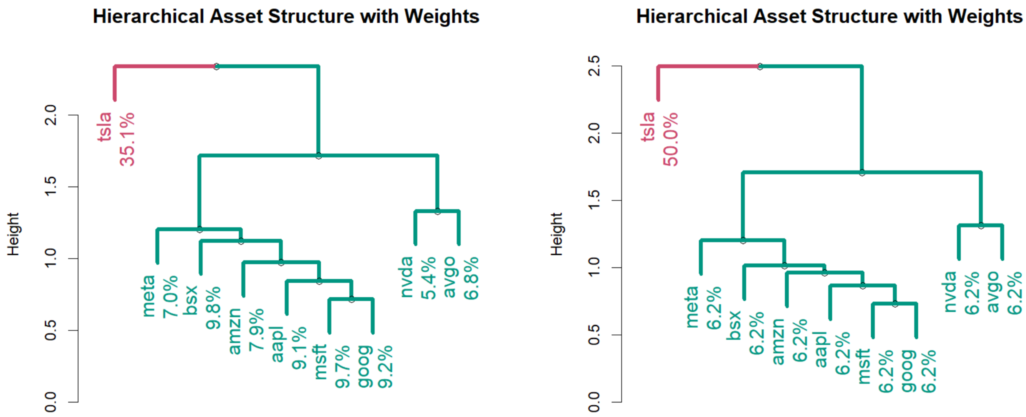

Figure 3: hierarchical portfolio optimization where left side is risk-weighted and right side is equally-weighted

In the risk-weighted portfolio, the weight distribution is based on the individual assets' risk, with assets that have lower risk receiving higher weights, and those with higher risk receiving lower weights. For example, Tesla (TSLA) has the highest weight at 35.1% because its risk is considered higher. Other assets that belongs to cluster 2, such as Meta (META) and Nvidia (NVDA), receive smaller allocations due to their risk profile. This method aims to minimize the overall portfolio risk by balancing risk contributions across clusters.

In the equally weighted portfolio, each asset receives the same weight, regardless of its risk or return. In this case, all assets are given a weight of 6.25%. This approach assumes that no asset is more important than another in terms of risk or reward, and aims to keep the portfolio balanced.

5.2 Investment result

	Benchmark	Hierarchical Risk Weighted	Hierarchical Equal Weighted
Annualized Sharpe Ratio	0.58	1.02	0.88
Total Return (%)	219.79	1891.79	1487.06
Annualized Return (%)	11.10	31.10	28.43
Annualized Volatility (%)	17.89	28.87	32.07

Table 1: A table showing the metric result of the hierarchical clustering result

The table presents the investment results of three different portfolio strategies compared to a benchmark. The key metrics shown are the Sharpe Ratio, Total Return, Annualized Return, and Annualized Volatility.

1. Sharpe Ratio: The benchmark has a Sharpe ratio of 0.58, which suggests a moderate return relative to risk. The Hierarchical Weighted portfolio has a higher Sharpe ratio of 1.02, indicating a better risk-adjusted return. The Hierarchical Equal Weighted portfolio has a Sharpe ratio of 0.88, which is lower than the Hierarchical Weighted portfolio but still higher than the benchmark.

2. Total Return: The total return for the Hierarchical Weighted portfolio is 1891.79%, significantly higher than the benchmark's return of 219.79%. The Hierarchical Equal Weighted portfolio shows a total return of 1487.06%, which is still much higher than the benchmark but lower than the Hierarchical Weighted portfolio.

3. Annualized Return: The annualized return for the benchmark is 11.10%, whereas the Hierarchical Weighted portfolio shows an annualized return of 31.10%. The Hierarchical Equal Weighted portfolio has an annualized return of 28.43%, which is still strong but not as high as the Hierarchical Weighted portfolio.

4. Annualized Volatility: Volatility measures the risk associated with the portfolios. The benchmark has a volatility of 17.89%, while the Hierarchical Weighted portfolio has higher volatility at 28.87%, reflecting the higher return. The Hierarchical Equal Weighted portfolio shows a volatility of 32.07%, indicating it is the riskiest among the three strategies.

In conclusion, we can say that amongst the two hirearchical optimization technique, the risk weighted is the winner. Additionally, hirearchical-based portfolio optimization can create abundant profit compared to the benchmark. But note that this comes with a price of having more volatile result.

5.3 Backtesting

In this section, we include all of the 20 stocks as this section is just purely for comparison. Backtesting involves training and testing data where training data is a zone for the model to learn the pattern of the data and testing data is a simulation zone for the model performance. The time interval for training data is January 1, 2022 until May 21, 2024 and the time interval for testing data last from May 21, 2024 until May 20, 2025.

In the two graphs provided here, we observe the performance of a HERC-based portfolio compared to the S&P 500 index and a risk-free asset over a period from May 2024 to May 2025.

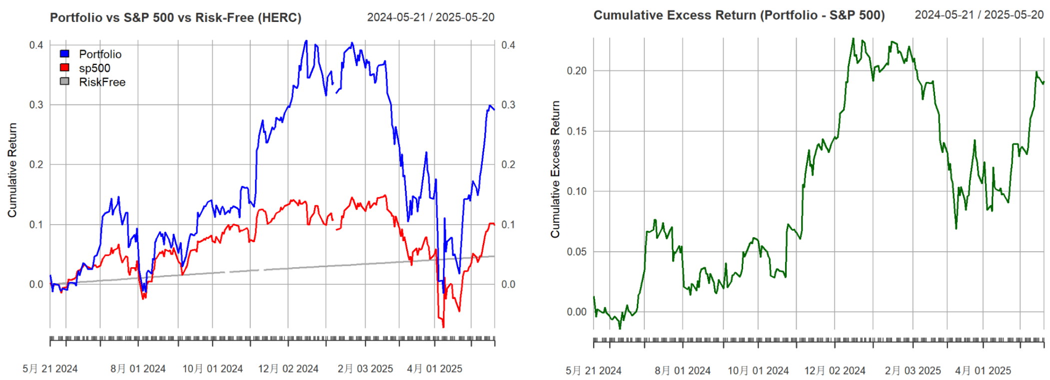

Figure 4: A pair of images showing the performance of HERC in backtesting. Cumulative Return is in decimal form.

The first graph compares the cumulative return of the portfolio, the S&P 500 index, and a risk-free asset. The portfolio's return (shown in blue) is higher than both the S&P 500 (red) and the risk-free asset (grey) throughout most of the observed period. However, there are notable fluctuations. The portfolio underperforms the S&P 500 at the beginning of just the first month but surpasses it after the middle of 2024, leading to a strong upward trend in cumulative returns by May 2025.

The second graph illustrates the cumulative excess return of a portfolio compared to the S&P 500 index. It starts at zero as we assumed to start investing at that date and fluctuates throughout the period. The portfolio experiences several sharp increases and decreases in value, especially between late 2024 and early 2025. The graph's significant rise towards the end of the period indicates that the portfolio outperformed the S&P 500 in the latter stages of the time frame. This suggests that, after a period of underperformance, the portfolio managed to generate superior returns.

One particularly interesting observation is the ability of HERC portfolio to outperform the S&P 500 after a period of underperformance. This suggests that the strategy behind HERC may have been designed to capitalize on certain market conditions or adjustments that were not as effective for the S&P 500, especially in the last months of 2024 and into 2025. The portfolio's strong return towards the end of the period underscores the volatility and potential high rewards of active portfolio management compared to a more passive approach like investing in the S&P 500 or a risk-free asset.

The reported information ratio of HERC backtesting is 1.27. This is generally considered to be a good one as it is above 1.

The performance of a portfolio based on Markowitz compared to the S&P 500 index and a risk-free asset over a period from May 2024 to May 2025 are given by this two plots.

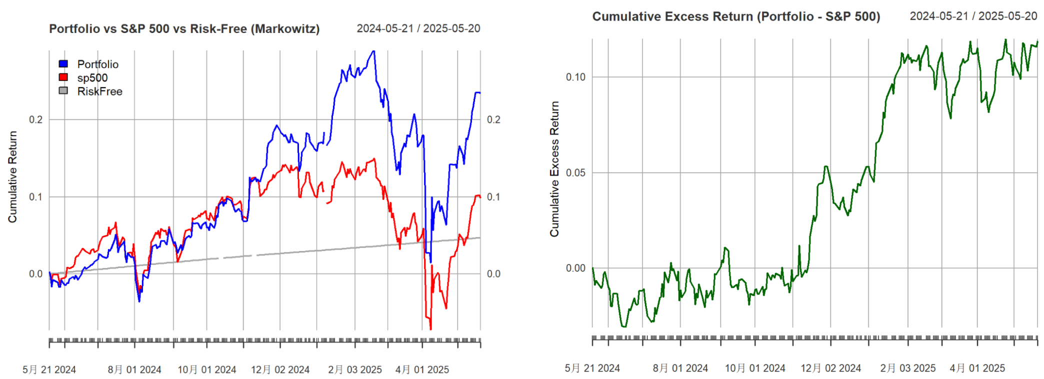

Figure 5: A pair of images showing the performance of Markowitz in backtesting. Cumulative Return is in decimal form.

Unlike HERC, this method yields a worse result where it takes 6 months to truly become a superior from the benchmark which is something that is not ideal for those who wants to have high frequency trading. However, notice that if it was winning against the benchmark, it was relatively have a more stable cumulative excess return compared to that of HERC since the Markowitz Cumulative Excess Return stabilizes around 10% starting from the second quarter of 2025.

Surprisingly, The reported information ratio for Markowitz is around 1.61 which is higher than that of HERC. Therefore, HERC is preferred to Markowitz.

6 Discussion

This report is aimed at building a hirearchical-based portfolio optimization called Hierarchical Equal Risk Contribution Portfolio in R from scratch based on a mentioned paper. The algorithm involves selecting the optimal way to cluster assets based on their returns by selecting the clustering type that maximizes the agglomerative coefficient (AC). Then, the optimal number of clusters (optimal_k) is extracted so that hierarchical clustering can be done. All available type of inter-cluster weights can be chosen are compared to see which one is better and how far can they beat the benchmark. After seeing the investment result, we proceed to do backtest to simulate how profitable can it be by using training and testing data splitting. It was discovered in the backtesting that the hirearchical based optimization performs twice better than benchmark. It can be concluded that this method of portfolio optimization is a big success compared to a traditional optimization method and benchmark. With this fascinating finding, we hope that hirearchical based optimization should be included as a standard in financial investment curriculum as stocks are usually positive correlated most of the time in real case which contradicts the assumption of Markowitz.

7 Appendix

7.1 Financial Background Knowledge

7.1.1 Investment metrics

- Compounded return

  $$r = (1 + r_{\text{annual}})^{\frac{1}{n}} - 1$$

  which is used for converting risk free where n represents the number of compounding periods within a year.

- Sharpe Ratio: a standard measurement of the attraction of a portfolio (from reward-tovolatility) by the ratio of its risk premium to the SD of its excess returns.

  $$\textbf{Sharpe Ratio of stock } i = \frac{\text{Excess Return}}{\text{Risk}} = \frac{E(r_i) - r_f}{\sigma}$$

- Information Ratio

  $$\text{Information Ratio} = \frac{\text{Average Excess Return}}{\text{Standard Deviation of Excess Return}}$$

  where the excess is the difference between the portfolio and benchmark durin testing.

7.2 About the stocks selection

The selection of the 20 stocks are generally based on popularity prospect, industry types prospect and average analyst.

Name	Industry	Pick Reasoning
Apple	Technology	Innovative thinker with a strong track record
Amazon	Finance	Excellent analytical skills and market knowledge
Broadcom Inc	Healthcare	Strong Buy by average analyst rating
NVIDIA	Technology	Frontier in AI and Strong Buy by average analyst rating
Microsoft	Technology	Frontier in AI and Strong Buy by average analyst rating
The Boeing Company	Industrials (Aerospace)	The number one airline manufacturer in the world
Bank of America Corporation	Financial Services	Leading banking institution with strong fundamentals
Blackrock	Financial Services	A company that owns the world
Boston Scientific	Healthcare (Medical Devices)	Strong Buy by average analyst rating
GE	Aerospace	Strong Buy by average analyst rating
DE	Industrial	Has a YTD Return of more than 20%
Energy Transfer LP	Energy	Strong Buy by average analyst rating
Google	Technology	Dominant in search and digital advertising
JP Morgan Chase	Banks	Global financial services leader
META	Technology	Social media and metaverse pioneer
Morgan Stanley	Financial Services	Top-tier investment banking firm
Rockwell Automation, Inc	Industrial	Strong Buy by average analyst rating
TESLA	Automotive - Industrial	EV market leader and innovator

Table 2: Stock Analysis: Companies, Industries, and Investment Rationales

8 Reference

1. U.S. Department of the Treasury. (2022). Daily treasury bill rates. U.S. Department of the Treasury. https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_bill_rates&field_tdr_date_value=2022

2. Raffinot, T. (2018). The Hierarchical Equal Risk Contribution Portfolio. SSRN. https://ssrn.com/abstract=3237540

3. Bodie, Z., Kane, A., & Marcus, A. J. (2024). Investments (13th ed.). McGraw-Hill.