Data Science Project, Master Quantitative Economics
Lecturer: MADALINA OLTEANU, ARTHUR THOMAS

PORFOLIO OPTIMIZATION

Duy-Tung TRAN
January 2025

Table of Contents

1. Introduction

Background

Portfolio Optimization is one of the cornerstones of modern financial theory and aims to achieve the most optimal trade-off between risk and return in segregating investment strategies. Investors seek a minimum risk approach while achieving a desired return on investment by allocating capital strategically among a diversified pool of assets. This perspective is based on Modern Portfolio Theory (MPT), developed by Harry Markowitz, which emphasizes the risk-return tradeoff and the creation of the efficient frontier, a collection of portfolios that provide the best expected return for a specific risk level.

Problem Statement

When facing the challenging problem of choosing the best set of assets from a large candidate set to obtain the best portfolio performance, investors are often presented with thousands of different options. This report centers on an optimization problem in a portfolio consisting of 30 stocks across several sectors. A popular metric used, being based on risk adjusted returns, is the Sharpe Ratio, which will be one of the goals of the solution.

This study aims to:

- Develop a data-driven model for portfolio optimization.
- Utilize historical stock data to compute annualized returns and volatility (risk).
- Maximize the Sharpe Ratio under practical constraints, including limitations on individual stock weights and the prohibition of short selling.

The analysis is conducted using stock price data spanning from January 2010 to 2020, sourced from Yahoo Finance. The study imposes the following constraints:

- Individual stock weights are capped at 20% of the portfolio.
- Short selling is not permitted.

2. Data Description

The dataset consists of historical daily closing prices for 30 stocks, spanning multiple sectors. The data was sourced from **Yahoo Finance**, covering the period from **2010 to 2020**, capturing long-term market trends while minimizing the impact of short-term fluctuations.

Data Characteristics

The dataset contains the following key attributes:

- **Stock Tickers**: A diverse selection of 30 stocks, including funds and equities representing a wide range of industries:
    - **Technology**: AAPL, MSFT, ADBE, CSCO

  - **Financials**: JPM, BAC, GS, AXP
  - **Consumer Staples**: PG, KO, PEP
  - **Energy**: XOM, CVX
  - **Healthcare**: JNJ, PFE, MRK
  - **Consumer Discretionary**: AMZN, NKE, TSLA
  - **Communication Services**: T, VZ, TMUS
  - **Industrials**: CAT, DE
  - **Materials**: APD, LIN
  - **Real Estate**: PLD, AMT, EQIX
- **Daily Closing Prices**: Prices are adjusted for stock splits and dividends, ensuring accuracy and consistency by representing the final traded price each day.

Data Preprocessing

1. Missing Data Handling

Missing values in the dataset were addressed using **forward-fill** and **backward-fill** methods, ensuring continuity and consistency in the time series data.

2. Return Calculation

- **Daily Returns (r_t)**: Calculated using the formula: $r_t = \frac{P_t}{P_{t-1}} - 1$

where P_t is the price at time t and P_{t-1} is the price at time $t-1$.

- **Annualized Returns**: Derived by multiplying the mean daily return by 252, the approximate number of trading days in a year (except for weekends and holidays).

3. Variance-Covariance Matrix

A critical component for portfolio optimization, the **variance-covariance matrix** quantifies the relationships between asset returns. It is calculated as:

$$\mathrm{Cov}(X) = \frac{1}{T-1}\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})^T$$

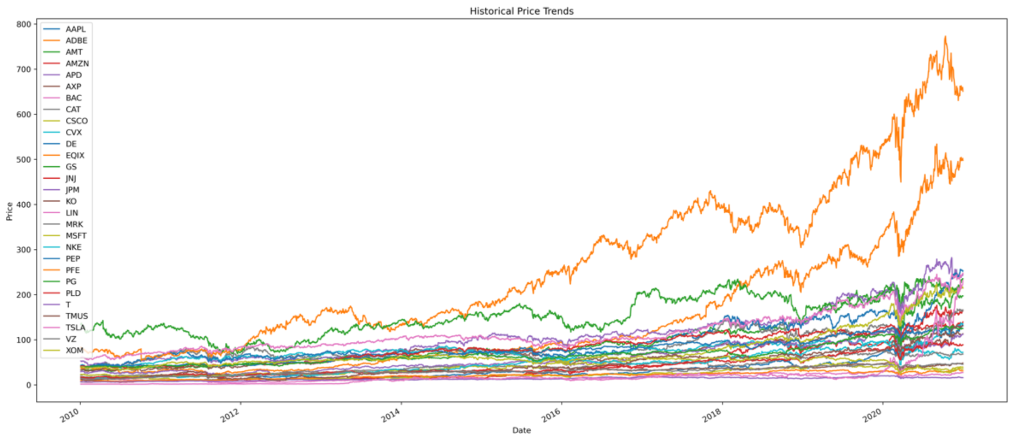
Descriptive Statistics

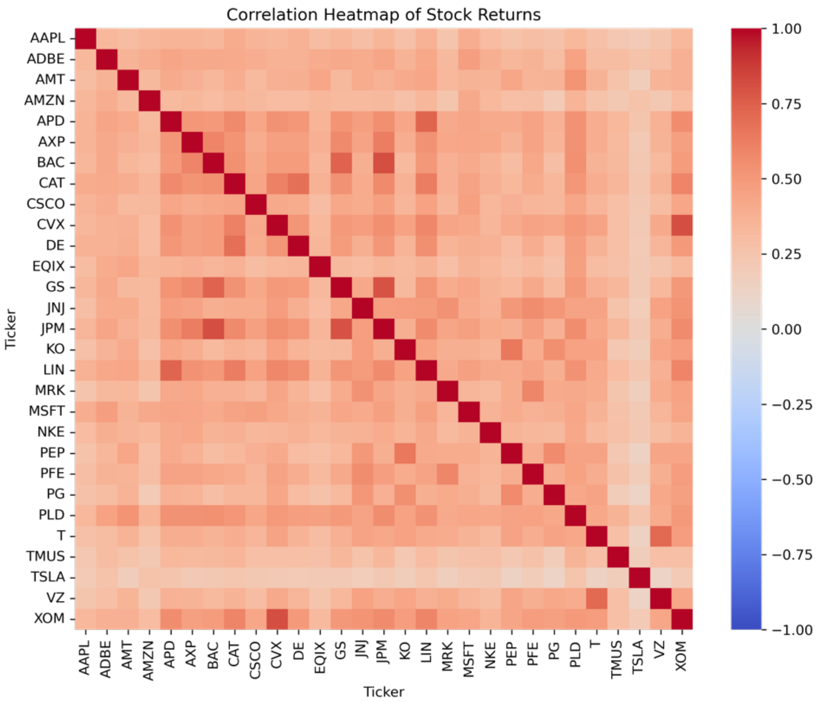
To understand the behavior of individual stocks, summary statistics were generated, including:

- **Mean Annualized Return**: The average returns each stock generates over a year.
- **Annualized Standard Deviation (Risk)**: The volatility of the stock's returns.
- **Minimum and Maximum Returns**: Historical extremes, providing a sense of potential risk and reward.

A table summarizing these statistics for all 30 stocks is included below:

	Mean Return (Annualized)	Std Dev (Annualized)	Min Return	Max Return
AAPL	0.314	0.283	-0.129	0.12
ADBE	0.283	0.304	-0.19	0.177
AMT	0.191	0.235	-0.152	0.122
AMZN	0.342	0.317	-0.127	0.157
APD	0.168	0.236	-0.126	0.137
AXP	0.155	0.291	-0.148	0.219
BAC	0.135	0.357	-0.203	0.178
CAT	0.173	0.292	-0.143	0.103
CSCO	0.116	0.271	-0.162	0.16
CVX	0.083	0.272	-0.221	0.227
DE	0.202	0.277	-0.136	0.135
EQIX	0.233	0.297	-0.331	0.133
GS	0.095	0.294	-0.128	0.176
JNJ	0.125	0.17	-0.1	0.08
JPM	0.165	0.29	-0.15	0.18
KO	0.105	0.175	-0.097	0.065
LIN	0.15	0.213	-0.103	0.117
MRK	0.127	0.207	-0.089	0.104
MSFT	0.234	0.254	-0.147	0.142
NKE	0.242	0.258	-0.117	0.152
PEP	0.125	0.175	-0.114	0.129
PFE	0.123	0.205	-0.077	0.09
PG	0.12	0.172	-0.087	0.12
PLD	0.194	0.275	-0.173	0.118
T	0.093	0.192	-0.092	0.1
TMUS	0.285	0.383	-0.366	0.368
TSLA	0.605	0.551	-0.211	0.244
VZ	0.119	0.176	-0.066	0.077
XOM	0.019	0.238	-0.122	0.127





Observations on the Data Table and Charts

- TSLA (Tesla) shows the largest expected return (60.5%), however also the largest risk (55.1%). TSLA and AMZN has a big amount of volatility meaning that the prices of these assets fluctuate sharply.
- Consumer staples (PG, KO, PEP) tend to have stocks that display lower volatility, indicating more stable performance. Stocks such as XOM and CVX demonstrate more stable yet less remarkable growth (1.9% and 8.3% respectively).
- As from historical price trend chart, TSLA and AMZN exhibit steep upward trends, confirming their role as growth stocks with exponential returns over time.
- Stocks across different sectors exhibit relatively low correlations (e.g., between technology and healthcare), confirming potential diversification benefits.
- The majority of stocks are correlated positively especially for the same sector (ex: JPM, BAC, GS in financials).
- The tradeoff is evident, with higher-return stocks generally exhibiting higher risk, aligning with Modern Portfolio Theory's principles.

3. Methodology

3.1 Theoretical Framework

Portfolio optimization aims to identify the optimal allocation of assets that maximizes expected returns while minimizing associated risks. The theoretical foundation is rooted in **Modern Portfolio Theory (MPT)**, introduced by Harry Markowitz, which provides a quantitative framework for constructing an efficient portfolio. The following core concepts underpin the methodology:

Expected Portfolio Return

The expected return of a portfolio is calculated as the weighted sum of the expected returns of individual assets. Mathematically, this can be expressed as:

$$R_p = \sum_{i=1}^{N} w_i R_i$$

- R_p: Expected portfolio return.
- w_i: Weight of asset i in the portfolio.
- R_i: Expected return of asset i.

Portfolio Risk (Volatility)

The portfolio's overall risk or volatility is determined by the variance-covariance matrix of asset returns. The risk is computed as:

$$\sigma_p = \sqrt{\mathbf{w}^{\mathbf{T}} \Sigma \mathbf{w}}$$

- σ_p: Portfolio standard deviation (risk).

- $\mathbf{w}$: Vector of asset weights $(w_1, w_2, \dots, w_N)$.
- Σ: Variance-covariance matrix of asset returns.

The variance-covariance matrix quantifies the relationships between assets, allowing the model to account for diversification effects. Assets with lower or negative correlations help reduce portfolio risk.

Optimization Objective

The optimization problem is framed as maximizing the Sharpe Ratio, a metric that measures risk-adjusted return:

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

- R_f: Risk-free rate of return.
- R_p: Portfolio return.
- σ_p: Portfolio risk.

We define the optimization problem mathematically as: $\max\limits_{\mathbf{w}} \frac{\mathbf{w}^{\mathbf{T}}\mathbf{R} - R_f}{\sqrt{\mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w}}}$

or equivalently (for computational simplicity): $\min\limits_{\mathbf{w}} - \frac{\mathbf{w}^{\mathbf{T}}\mathbf{R} - R_f}{\sqrt{\mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w}}}$

Constraints

The optimization is subject to the following practical constraints:

1. **Weight Sum Constraint**: $\sum_{i=1}^{N} w_i = 1$
2. **Non-Negativity Constraint**: No short selling is allowed: $w_i \geq 0 \; \forall i$
3. **Weight Cap**: Individual weights are capped at 20%: (e.g., $w_i \leq 0.2 \; \forall i$).

This limits the weight of any single asset to 20% of the portfolio. This constraint prevents over-concentration in a few assets, promoting diversification.

3.2 Implementation

- Data Preparation: Annualized returns and the variance-covariance matrix.
- Define the Optimization Problem: Use additional functions from SciPy minimize function subject to (i.e. Sharpe Ratio).
- Generate Efficient Frontier: Optimize portfolios using weight variations to trace the risk-return curve.
- Find the Best Portfolio: Porfolio with the highest Sharpe Ratio.

4. Results and Analysis

4.1 In-Sample Optimization

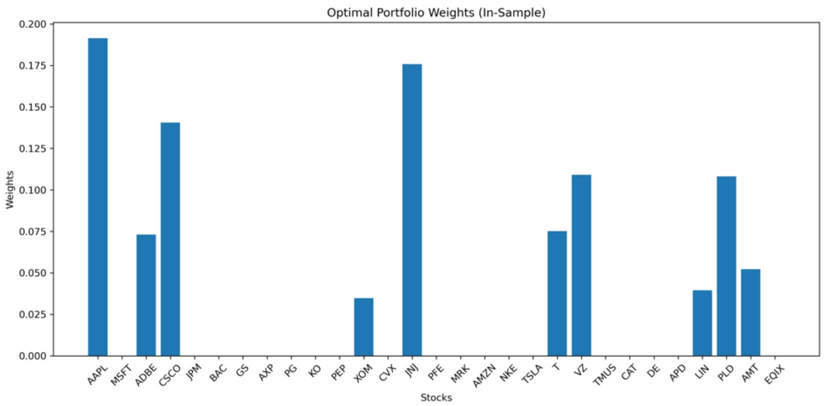
The in-sample (split day at 1/1/2018) optimization process involved maximizing the Sharpe Ratio while adhering to practical constraints. These constraints included prohibiting short selling and capping individual stock weights at a maximum of 20%. The results of the optimized portfolio are summarized below:

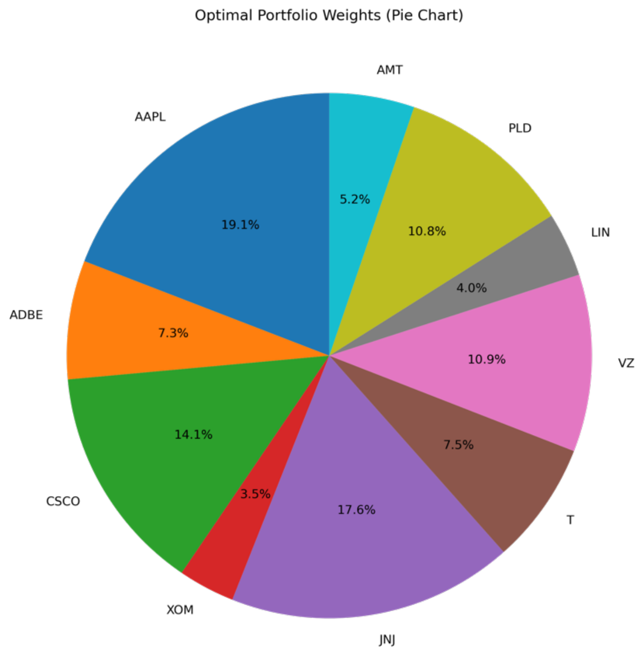
Optimized Portfolio (In-Sample)

- **Sharpe Ratio**: 1.3991
- **Expected Portfolio Return**: 23.53%
- **Portfolio Risk (Standard Deviation)**: 15.39%

These results indicate a well-balanced portfolio that effectively maximizes risk-adjusted returns within the specified constraints.

Stock	Weight (%)
AAPL	19.1
ADBE	7.3
CSCO	14.09
XOM	3.5
JNJ	17.59
T	7.5
VZ	10.9
LIN	4.0
PLD	10.8
AMT	5.2





Observations on the Optimal Portfolio Weights

1. Allocation Summary

- With a weighting of 19.1%, AAPL (Apple) is the highest weight, meaning it contributes the most to weighted expected return and risk of the portfolio, followed by JNJ (Johnson & Johnson) with its weight of 17.58%,
- Cisco (CSCO), Prologis (PLD) hold 14.05% and 10.82% to offset potential growth with combat against the risks.
- Their weights between 7.31% and 10.91% almost confirm their role in diversification for VZ (Verizon), T (AT&T) and ADBE (Adobe).
- The final two positions, XOM (Exxon Mobil) and LIN (Linde), have low weights (3.48% and 3.96%, respectively) as a result of lower expected returns or correlation with other assets.
- MSFT, AMZN, TSLA gets 0. This indicates that, for the framework defined above, they aren't economic in terms of increasing the Sharpe Ratio.

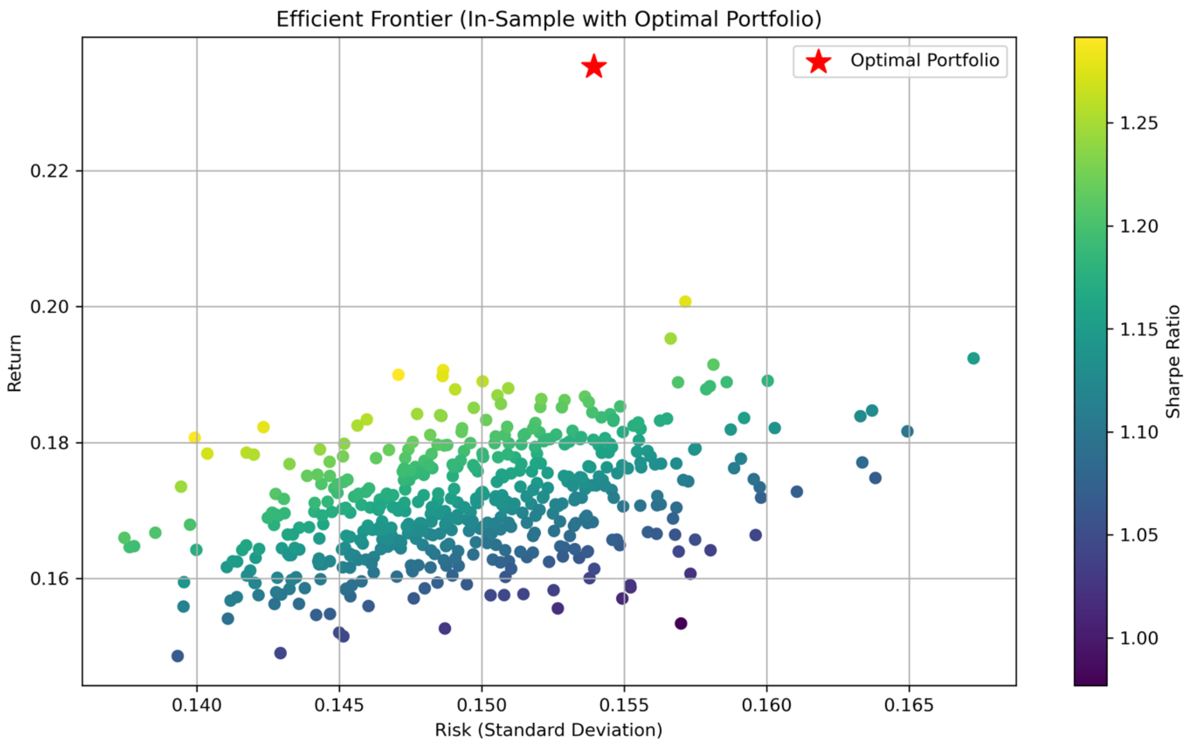
2. Diversification

- The portfolio achieves diversification by including stocks across multiple sectors:
    - **Technology**: AAPL, CSCO, ADBE.
    - **Healthcare**: JNJ.
    - **Telecommunication Services**: T, VZ.
    - **Real Estate**: PLD, AMT.
    - **Energy and Materials**: XOM, LIN.
- The absence of stocks from financials, consumer staples, and consumer discretionary sectors suggests that these sectors either provide insufficient risk-adjusted returns or exhibit higher correlation with already included assets.

3. Risk-Return Tradeoff

AAPL and JNJ are the most dominant portfolio characteristics, presumably because of their strong risk-adjusted profile. But such concentration risks could also amplify risks in particular sectors (e.g., technology and healthcare).

An efficient frontier was built by simulating 500 random portfolios with different asset allocations. The risk (standard deviation) and returns for each of the portfolios was calculated. The above data was plotted to show the tradeoff of risk and return and create the efficient frontier. As shown by the red star on the plot, the best portfolio has the best Sharpe Ratio, which is a best risk-return profile.

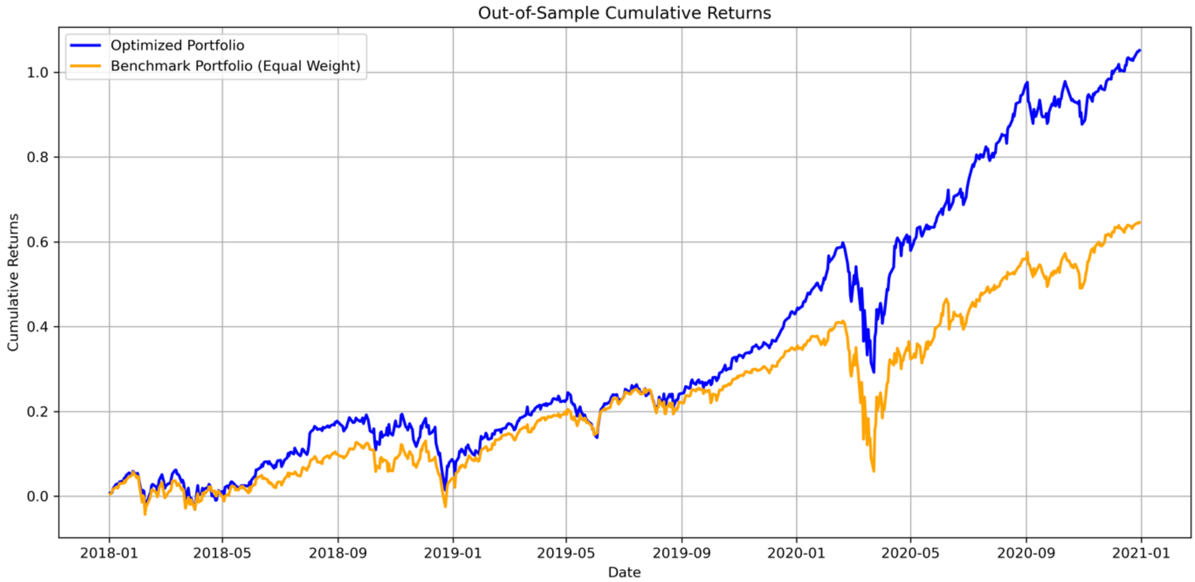


- The optimal portfolio identified by the red star is located on the efficient frontier boundary with the maximum Sharpe Ratio. This verifies that the portfolio has an optimal risk-return trade-off under decided constraints.
- The efficient portfolio gives a return of 23.53% and standard deviation of 15.39% emphasizing that compared to 500 simulated portfolios this portfolio gives a better risk-adjusted performance.
- The color bar indicates the Sharpe Ratio of each portfolio. More optimal portfolios also naturally will tend to have higher Sharpe Ratios, as seen in the coverage over the transition to brighter colors.
- Investors can ensure that their portfolio achieves the highest possible return, given the risk they are prepared to accept, by choosing a portfolio on the efficient frontier, and in particular, a portfolio near the optimal point.

4.2 Out-of-Sample Performance

To assess the robustness and practical effectiveness of the optimized portfolio, its performance was evaluated on out-of-sample data. This approach ensures that the portfolio's optimization process generalizes well to unseen market conditions. Key performance metrics are summarized as follows:

Out-of-Sample Performance



- Optimized Portfolio Sharpe Ratios: 1.404
- Benchmark Portfolio Sharpe Ratios: 0.8430

The optimized Sharpe Ratio of the optimized portfolio is significantly greater than the benchmark portfolio (all weights equal). This outcome signifies that the optimized portfolio is more efficient in terms of risk-adjusted returns, as it generates increased returns for every unit of risk undertaken.

- Optimized Porfolio Cumulative Returns: 1.0511(~105.1%)
- Benchmark Portfolio Cumulative Returns: 0.646 (~64.6%)

Cumulative returns of optimized portfolio are much greater than returns of benchmark portfolio. Optimizing the portfolio construction process was also evident in the portfolio of stocks performing better than its benchmark, which supports the innovative aspect of this study.

Key Insights

The optimized portfolio is obviously superior to the benchmark portfolio when it comes to risk-adjusted performance (Sharpe Ratio) and cumulative returns.

The varied Sharpe Ratios highlight the significance of specifying portfolio weights according to risk-return tradeoffs instead of equal weights.

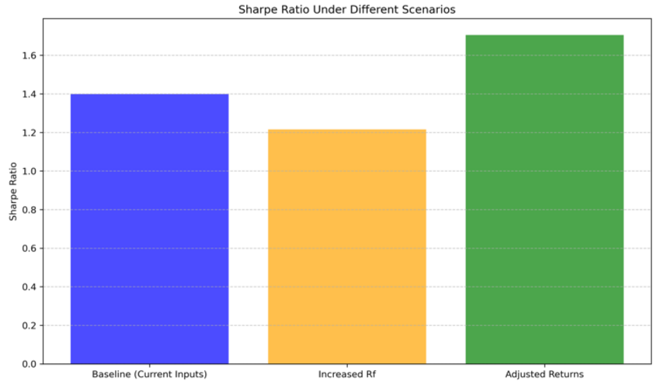
The robustness of the optimization process to in-sample conditions is also demonstrated by the yield on out-of-sample data. Such robustness cultivates confidence that it can apply meaningfully in real-world outcomes.

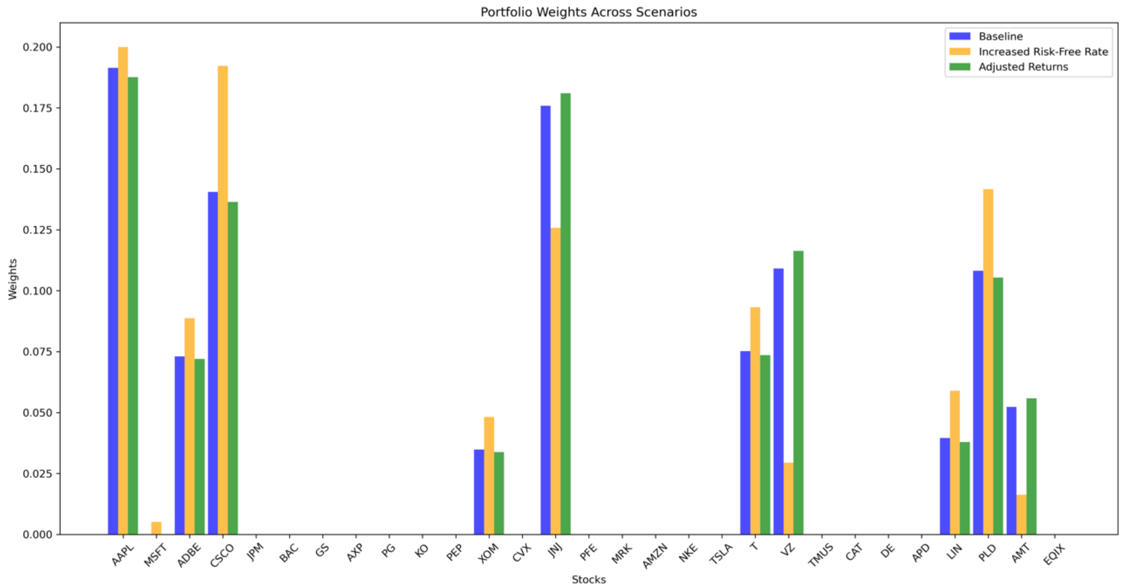
Optimizing investor portfolios yield better returns at the same or lower risk levels versus that of an equal weight benchmark.

4.3 Sensitivity Analysis

To evaluate the robustness and flexibility of the portfolio optimization model, sensitivity analysis was conducted. This analysis explores how variations in key assumptions impact portfolio performance, providing insights into the dependency of results on the model's inputs.

Scenario	Sharpe Ratio	Expected Return (%)	Risk (%)
Current Inputs	**1.398**	23.532	15.392
Increased Rf	**1.215**	26.426	17.622
Adjusted Returns	**1.704**	27.914	15.200





Impact of the Risk-Free Rate

- Increasing the risk-free rate Rf reduced the **Sharpe Ratio** from **1.3989** to **1.2158**, illustrating the sensitivity of risk-adjusted performance to changes in the baseline Rf.
- This result highlights the critical role of the risk-free rate in determining portfolio efficiency, as a higher Rf increases the benchmark for risk-adjusted returns, effectively penalizing portfolios with higher risk.

Changes in Expected Returns

- Adjusting expected returns (e.g., scaling them upward) significantly influenced portfolio outcomes:
  - The **Sharpe Ratio** increased to **1.7048**, reflecting improved risk-adjusted performance.
  - The **Expected Return** rose to **27.91%**, while the **Risk** slightly decreased to **15.20%**.
- These findings demonstrate the portfolio's strong dependency on the accuracy of expected return estimates. Higher expected returns can lead to more favorable

outcomes, but such assumptions must be grounded in realistic projections to avoid over-optimistic results.

5. Discussion

5.1 Interpretation of Results

Based on the constraints defined, the portfolio optimization process was able to find an efficient portfolio maximizing the Sharpe Ratio. The results show some key takeaways, including:

Optimal Portfolio: The optimized portfolio strikes a good balance of risk and return, returning a Sharpe Ratio of 1.404 versus the baseline's 0.843. This highlights the ability of its optimization framework in driving risk-adjusted return.

Diversification: Using weights given by an optimization, we can successfully puzzle together a combination of red points, ensuring they are less interactive so that we can enjoy now a different portfolio which in turn reduce management cost (feasibility of managing 10 over 30 stocks).

5.2 Limitations

While the model's results are promising there are several limitations that deserve mentioning:

Reliance on Past Data: The optimizer uses past returns and covariances as proxies for future behavior. But financial markets are evolving, and past performance is not always a predictor of future results.

Stationarity Assumption: It assumes stationarity of asset returns over time. However, market conditions/economic shocks/regime changes can break this assumption and yield differences in portfolio performance (as COVID-19 in March 2020 shows unexpected downturn of the whole market which affected negatively investors and funds at that time).

6. Conclusion

6.1 Summary of Findings

In this project we aimed to optimize a portfolio of 30 diverse stocks for maximum Sharpe Ratio under specific constraints. The analysis made several key findings, including:

The Sharpe Ratio was 1.404 for the optimized portfolio, vastly superior to that of the benchmark portfolio. Thus, its respective yearly returrn and risk are 23.53% and 15.39%

By using weights across different stocks with low correlations, unsystematic risk was successfully reduced through effective diversification.

6.2 Future Work

Non linear model ARMA-GARCH can be first intergrated in the analysis to predict returns. Then updated variance-covariance matrix derived from the GARCH model to optimize portfolio weights dynamically.

Machine learning predictive algorithms present a solid approach to estimate expected returns and volatilities. These methods also improve the resilience of return predictions and risk assessment by integrating macroeconomic indicators and sentiment analysis. That is, via and consistent with APT or the extended "AIPT" (see Antoine Didisheim, Shikun Ke, Bryan Kelly and Semyon Malamud, "The Surprising Dominance of Large Factor Models," 9/2024).