PORTFOLIO ANALYSIS AND OPTIMIZATION

Dibya Ranjan Dash and Sahil (Roll Number: XXXXX)

Term Project Report
Trimester 7

B.Sc. (Honours) Data Science and Artificial Intelligence
Indian Institute of Technology Guwahati, India
your-email@op.iitg.ac.in

ABSTRACT

This report presents a portfolio construction and optimization project using selected Indian stocks and a benchmark index. The objective is to determine optimal capital allocation that maximizes expected returns while effectively managing risk, which is essential for informed investment decision-making. Historical stock and benchmark data were collected from the Yahoo Finance API and preprocessed to address missing values and outliers. Key portfolio metrics, including returns, volatility, and Sharpe ratios, were computed, and random portfolio simulations were conducted to analyze the risk–return trade-off.

Modern Portfolio Theory and Markowitz mean–variance optimization were employed to identify optimal portfolios, including the Global Minimum Variance Portfolio (GMVP) and the Tangency Portfolio corresponding to the maximum Sharpe ratio. The efficient frontier, Capital Market Line (CML), and Security Market Line (SML) were analyzed to evaluate portfolio efficiency and systematic risk exposure.

The results indicate that optimized portfolios outperform equal-weighted strategies and the benchmark index. In particular, the tangency portfolio achieves superior risk-adjusted returns, while the GMVP effectively minimizes overall portfolio risk. The study demonstrates a structured and reproducible framework for portfolio optimization and risk-aware investment decision-making.

***Index Terms*—** Portfolio Optimization, Modern Portfolio Theory, Efficient Frontier, Risk Analysis, Financial Data Science

1. INTRODUCTION

Portfolio optimization is a fundamental aspect of investment management, focused on combining assets in a manner that maximizes expected returns while effectively controlling risk. Achieving this balance requires careful analysis of historical financial data and strategic capital allocation to mitigate the impact of market volatility. Modern Portfolio Theory (MPT) provides a quantitative framework for constructing diversified portfolios along the efficient frontier, enabling investors to understand the trade-off between risk and return and make informed investment decisions.

In this project, selected Indian stocks and a benchmark index are analyzed to demonstrate the advantages of data-driven portfolio construction. Historical price data were obtained from the Yahoo Finance API and preprocessed to address missing values and outliers. Key portfolio performance metrics, including returns, volatility, and Sharpe ratios, were computed, and random portfolio simulations were performed to explore the risk–return space. Optimal portfolios, such as the Global Minimum Variance Portfolio (GMVP) and the Tangency Portfolio corresponding to the maximum Sharpe ratio, were identified. Additionally, the Capital Market Line (CML) and Security Market Line (SML) were analyzed to assess systematic risk and overall portfolio performance.

2. PROBLEM STATEMENT AND OBJECTIVES

2.1. Problem Statement

Investors must allocate capital across multiple financial assets to maximize expected returns while effectively managing risk. In the absence of a structured and quantitative approach, investment portfolios may suffer from poor diversification, excessive volatility, and inefficient performance. This project addresses the problem of constructing an optimal equity portfolio by systematically balancing expected return and risk using historical market data.

The study analyzes selected Indian stocks drawn from diverse industry sectors, along with the Nifty 50 benchmark index, over a specified time horizon. Portfolio construction is performed under practical investment constraints: asset weights are restricted to be non-negative, the total allocation sums to one, and short selling is not permitted. The analysis assumes that historical price movements and inter-asset correlations provide meaningful information for portfolio optimization.

This problem is highly relevant to real-world investment decision-making, as both individual and institutional investors increasingly rely on quantitative methods for efficient capital allocation. Addressing this problem illustrates the role of diversification in risk reduction and demonstrates how optimized portfolios can outperform simple or naive allocation strategies.

2.2. Objectives

The specific objectives of this project are as follows:

1. To collect and preprocess historical stock price and benchmark index data for portfolio analysis.
2. To compute key portfolio performance metrics, including expected return, volatility, and Sharpe ratio.
3. To analyze the impact of diversification through inter-asset correlations.
4. To simulate multiple portfolio combinations in order to examine the risk–return trade-off.

5. To design and implement optimized portfolios, including the Global Minimum Variance Portfolio (GMVP) and the Tangency Portfolio corresponding to the maximum Sharpe ratio.
6. To evaluate and compare the performance of optimized portfolios against equal-weighted portfolios and the benchmark index.

3. METHODOLOGY / APPROACH

This section describes the methodology adopted to construct and evaluate optimized stock portfolios. The approach follows a systematic workflow, beginning with data collection and preprocessing and culminating in portfolio optimization, performance evaluation, and result visualization. All steps are designed to be reproducible using consistent data sources, parameters, and computational tools.

3.1. Overall Workflow

The overall workflow of the project consists of the following stages:

Data Collection: Historical adjusted closing prices for selected Indian stocks and the benchmark index (Nifty 50) were collected using the Yahoo Finance API over a five-year period.

Data Preprocessing: The raw price data were cleaned to handle missing values and potential outliers, ensuring consistency and reliability for subsequent analysis.

Return and Risk Estimation: Daily asset returns were computed from adjusted prices. Statistical measures, including mean returns and the variance–covariance matrix, were estimated to capture individual asset behavior as well as inter-asset co-movement.

Portfolio Simulation: A large number of portfolios were generated by assigning different combinations of asset weights under realistic investment constraints in order to explore the risk–return space.

Portfolio Optimization: Optimization techniques were applied to identify key portfolios, including the Global Minimum Variance Portfolio (GMVP) and the Tangency Portfolio corresponding to the maximum Sharpe ratio.

Performance Evaluation and Visualization: The optimized portfolios were evaluated and compared against equal-weighted portfolios and the benchmark index using risk–return metrics and visualization tools such as the efficient frontier, Capital Market Line (CML), and Security Market Line (SML).

Overall, the workflow starts from raw market data and concludes with quantitative insights into portfolio performance and risk characteristics.

3.2. Technical Details

3.2.1. Data and Tools

Historical price data were obtained using the Yahoo Finance API. The entire analysis was implemented in Python, utilizing libraries such as NumPy, Pandas, SciPy, Matplotlib, and `yfinance`. Configuration parameters, including asset selection and analysis period, were managed through a configuration file (`config.yaml`) to enhance flexibility and reproducibility.

3.2.2. Return and Risk Computation

Asset returns were calculated as logarithmic daily returns. Expected returns were estimated as the mean of historical returns, while portfolio risk was quantified using the variance–covariance matrix of asset returns.

3.2.3. Portfolio Performance Metrics

For a given portfolio weight vector w, the expected portfolio return R_p and volatility σ_p were computed as

$$R_p = w^\top \mu, \tag{1}$$

$$\sigma_p = \sqrt{w^\top \Sigma w}, \tag{2}$$

where μ denotes the vector of expected asset returns and Σ represents the covariance matrix.

Risk-adjusted performance was measured using the Sharpe ratio, defined as

$$S = \frac{R_p - R_f}{\sigma_p}, \tag{3}$$

where R_f denotes the risk-free rate.

3.2.4. Optimization Method

Portfolio optimization was carried out using numerical optimization techniques under the following constraints:

- The sum of portfolio weights equals one.
- All portfolio weights are non-negative, prohibiting short selling.

The Global Minimum Variance Portfolio (GMVP) was obtained by minimizing portfolio variance, while the Tangency Portfolio was identified by maximizing the Sharpe ratio.

4. EXPERIMENTS AND RESULTS

This section describes the experimental setup used in the project and presents the obtained results. All results are reported objectively, without interpretation or conclusions.

4.1. Experimental Setup

The experiments were conducted using historical stock price data for five Indian equities, namely HCL Technologies, Axis Bank, Reliance Industries, Bharti Airtel, and Sun Pharma, along with the Nifty 50 index as the benchmark. The dataset consists of five years of daily adjusted closing prices, resulting in more than 1,200 trading observations per asset.

Data collection was performed using the Yahoo Finance API. Prior to analysis, the data were preprocessed to handle missing values using forward-filling and to remove extreme outliers through Z-score-based filtering. Daily logarithmic returns were computed from the cleaned price series and subsequently used to estimate annualized expected returns and the variance–covariance matrix.

All experiments were implemented in Python using libraries such as NumPy, Pandas, SciPy, Matplotlib, and Seaborn. Portfolio optimization was performed using the Sequential Least Squares Programming (SLSQP) algorithm. A risk-free rate of 4% was assumed for Sharpe ratio calculations. The experiments were executed on a standard personal computer using a Python-based computational environment.

The evaluation metrics used in this study include expected annual return, annualized volatility, and the Sharpe ratio.

4.2. Results

Table 1 presents the expected annual return and volatility of individual stocks derived from historical data.

Table 1. Expected Annual Return and Volatility of Individual Stocks

Ticker	Expected Annual Return	Annual Volatility
BHARTIARTL.NS	0.260951	0.223657
SUNPHARMA.NS	0.225563	0.206967
HCLTECH.NS	0.160695	0.237393
AXISBANK.NS	0.150084	0.248419
RELIANCE.NS	0.096513	0.225106

Figure 1 shows the correlation matrix of daily log returns for the selected stocks.

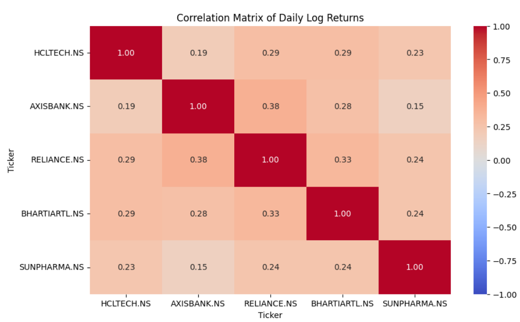

Fig. 1. Correlation matrix of daily log returns

Figure 2 displays the simulated random portfolios along with the efficient frontier.

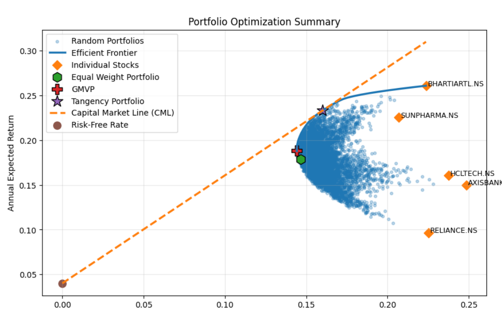

Fig. 2. Simulated random portfolios with efficient frontier and Capital Market Line

Table 2 reports the asset weight allocations for the Equal Weight Portfolio, GMVP, and Tangency Portfolio, along with their performance metrics.

Table 2. Portfolio Weights and Performance Metrics

Ticker	Weight	Allocation (%)
Equal Weight Portfolio		
HCLTECH.NS	0.2	20.0
AXISBANK.NS	0.2	20.0
RELIANCE.NS	0.2	20.0
BHARTIARTL.NS	0.2	20.0
SUNPHARMA.NS	0.2	20.0
Return: 0.1788 — Volatility: 0.1466 — Sharpe Ratio: 0.9467		
GMVP		
SUNPHARMA.NS	0.3170	31.7030
BHARTIARTL.NS	0.1888	18.8766
HCLTECH.NS	0.1761	17.6130
AXISBANK.NS	0.1638	16.3821
RELIANCE.NS	0.1543	15.4253
Return: 0.1885 — Volatility: 0.1442 — Sharpe Ratio: 1.0303		
Tangency Portfolio		
BHARTIARTL.NS	0.4529	45.2871
SUNPHARMA.NS	0.4287	42.8724
HCLTECH.NS	0.0603	6.0299
AXISBANK.NS	0.0581	5.8105
RELIANCE.NS	0.0000	0.0000
Return: 0.2333 — Volatility: 0.1601 — Sharpe Ratio: 1.2075		

Table 3 compares the expected return, volatility, and Sharpe ratio of the optimized portfolios with the benchmark index.

Table 3. Comparison of Optimized Portfolios with Benchmark

	Expected Return	Volatility (Risk)	Sharpe Ratio
Benchmark	0.1199	0.1391	0.5746
Eq. Weight Portfolio	0.1788	0.1466	0.9467
GMVP	0.1885	0.1442	1.0303
Tangency Portfolio	0.2333	0.1601	1.2075

Figure 3 illustrates the growth of one unit of investment over time.

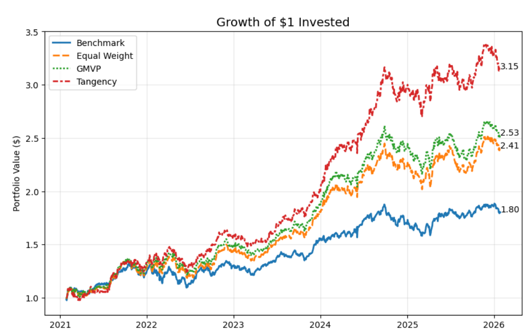

Fig. 3. Growth of one unit of investment over time for benchmark and portfolios

Figure 4 presents the Security Market Line.

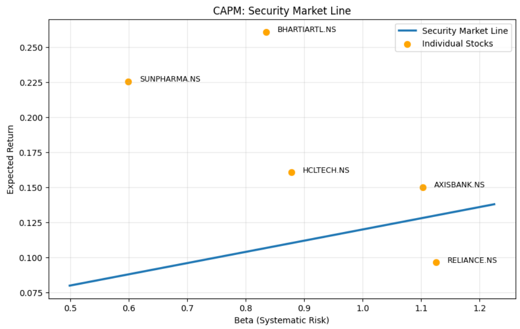

Fig. 4. Security Market Line showing beta vs. expected return for individual stocks

Table 4 lists the top five stocks with the highest beta values.

Table 4. Top Five Stocks with Highest Beta Values

Ticker	Beta	Expected Return
RELIANCE.NS	1.1253	0.0965
AXISBANK.NS	1.1026	0.1501
HCLTECH.NS	0.8780	0.1607
BHARTIARTL.NS	0.8344	0.2610
SUNPHARMA.NS	0.5986	0.2256

All tables and figures are referenced consistently and summarize the quantitative outcomes of the experiments.

5. DISCUSSION

This section interprets and analyzes the results presented in the previous section. The discussion connects the observations to the project objectives and highlights the strengths, trade-offs, and limitations of the proposed portfolio optimization approach.

5.1. Performance of Different Portfolios

The results in Table 1 show clear differences in expected annual returns and volatility across individual stocks. BHARTIARTL.NS and SUNPHARMA.NS exhibit relatively high expected returns, while AXISBANK.NS and HCLTECH.NS show higher volatility, indicating greater risk. These differences justify the need for portfolio-level optimization rather than relying on individual assets.

A clear performance difference is observed between the benchmark, equal-weight portfolio, and optimized portfolios. The benchmark index has the lowest Sharpe ratio, indicating weaker risk-adjusted performance compared to all constructed portfolios (Table 3). This suggests that passive exposure to the market is less efficient than systematic portfolio construction using historical data.

The equal-weight portfolio performs better than the benchmark, even without optimization. This improvement highlights the benefit of spreading investment evenly across multiple stocks instead of concentrating capital in a single index. However, its performance is still lower than the optimized portfolios (Table 3).

The Global Minimum Variance Portfolio (GMVP) achieves the lowest volatility among all portfolios (Table 3). This happens because the optimization focuses on minimizing overall risk by selecting assets with favorable correlations. The correlation matrix confirms that the selected stocks are not perfectly correlated, which helps reduce total portfolio risk (Figure 1). The GMVP slightly improves returns compared to the equal-weight portfolio while keeping risk low, making it suitable for risk-averse investors.

The Tangency Portfolio shows the highest Sharpe ratio and expected return (Table 3). This portfolio lies on the Capital Market Line and represents the best possible trade-off between risk and return for the chosen risk-free rate (Figure 2). The higher allocation to certain stocks indicates that the optimization favors assets with better return-to-risk characteristics rather than equal allocation. This explains why some stocks receive very low or zero weights (Table 2).

5.2. Risk-Free Rate, Capital Market Line, and Efficient Frontier

The inclusion of the risk-free rate allows the construction of the Capital Market Line (CML). The CML shows how investors can combine a risk-free asset with the tangency portfolio to achieve higher expected returns for a given level of risk. Portfolios lying below the CML are inefficient, while the tangency portfolio represents the optimal risky portfolio (Figure 2).

The efficient frontier highlights portfolios that offer the highest expected return for each level of volatility. Portfolios below the frontier are dominated by better alternatives. The random portfolio simulation helps visualize the full risk–return space and confirms that optimized portfolios lie on or near the frontier, validating the optimization process (Figure 2).

5.3. Security Market Line and Beta Analysis

The Security Market Line (SML) explains the relationship between systematic risk (beta) and expected return (Figure 4). Stocks with higher beta values are expected to deliver higher returns as compensation for higher market risk. The beta table shows that stocks like RELIANCE.NS and AXISBANK.NS have higher sensitivity to market movements, while stocks such as SUNPHARMA.NS have lower beta values (Table 4).

However, the observed expected returns do not always align perfectly with the SML. This deviation can be due to estimation errors, short sample periods, or market inefficiencies. This highlights a limitation of relying only on historical data and CAPM assumptions.

5.4. Trade-Offs Observed

A clear trade-off is observed between risk and return. Portfolios with higher expected returns also show higher volatility. While the tangency portfolio offers superior risk-adjusted performance, it also concentrates investments in fewer assets, which may increase exposure to specific sectors. In contrast, the GMVP provides better stability but sacrifices some return potential.

Another trade-off is between model simplicity and realism. The assumption of no short selling and fully invested portfolios makes the results practical, but it also limits the solution space.

5.5. Limitations of the Approach

This study relies entirely on historical price data, assuming that past behavior will continue in the future. Market conditions, economic events, and structural changes are not explicitly modeled. Transaction costs, taxes, and liquidity constraints are also ignored, which may affect real-world performance.

Additionally, expected returns and covariances are estimated from a limited dataset, which can introduce estimation risk. The results may change with a different time period, asset selection, or risk-free rate.

5.6. Connection to Objectives

All stated objectives were successfully achieved. Historical data was collected and processed, portfolio metrics were computed, diversification effects were analyzed, and optimized portfolios were constructed and evaluated. The comparison with the benchmark confirms that systematic portfolio optimization can improve investment outcomes.

Overall, the results demonstrate the practical value of quantitative portfolio analysis and provide a strong foundation for further extensions and real-world applications.

6. CONCLUSION AND FUTURE WORK

This project focused on the analysis and optimization of a stock portfolio using historical market data and quantitative methods. The study demonstrated how portfolio risk and return can be measured, compared, and improved through systematic portfolio construction rather than relying on individual asset selection or passive market exposure.

One of the key outcomes of this work is the comparison between different portfolio strategies. The results show that diversification alone improves performance over the benchmark, while optimized portfolios further enhance risk-adjusted returns. The Global Minimum Variance Portfolio effectively reduced overall risk, making it suitable for conservative investment strategies. In contrast, the Tangency Portfolio achieved the highest Sharpe ratio, highlighting its ability to deliver superior risk-adjusted performance when combined with a risk-free asset. The use of the efficient frontier, Capital Market Line, and Security Market Line provided clear insights into the trade-offs between risk, return, and systematic market exposure.

This project also strengthened practical skills in data preprocessing, financial metrics computation, portfolio optimization, and result interpretation. It reinforced the importance of quantitative analysis in informed investment decision-making and highlighted the limitations of relying solely on historical data and simplified assumptions.

Future work can extend this study by including a larger set of stocks, different market periods, or multiple benchmark indices. Additional improvements may involve incorporating transaction costs, short selling, dynamic rebalancing, or alternative risk measures. More advanced models, such as multi-factor models or real-time portfolio updates, can also be explored to improve real-world applicability.

7. ARTIFACTS AND DEMONSTRATIONS

This section provides external resources that support the implementation and experimental results presented in this report. All artifacts are designed to ensure reproducibility and transparency of the project work.

7.1. Source Code Repository

The complete source code for data ingestion, preprocessing, portfolio analysis, optimization, and visualization is available in a public repository. The repository includes well-documented Python scripts, configuration files, and Jupyter notebooks used to generate all results and figures in this report.

Link: GitHub Repository

7.2. Project Documentation / README

Detailed instructions on how to set up the environment, run the code, and reproduce the experiments are provided in the project documentation. This resource explains the workflow, dependencies, and configuration options used in the study.

Link: Included in the code repository

7.3. Notebook Demonstrations

Jupyter notebooks are provided to demonstrate exploratory data analysis, portfolio simulations, efficient frontier construction, and visualization of results such as the Capital Market Line and Security Market Line. These notebooks allow step-by-step reproduction of the experiments.

Link: Included in the code repository

7.4. Video Demonstration

A short video walkthrough explaining the project structure, execution steps, and key outputs is provided to visually demonstrate system behavior.

Link: Project Demonstration Video

8. REFERENCES

[1] Dr. Dipankar Mondal, *DA361 Financial Valuation and Portfolio Analytics*, Indian Institute of Technology Guwahati (IITG), Course Material, 2025.

[2] Francis, J., & Kim, K., *Modern Portfolio Theory: Foundations, Analysis, and New Developments*, Wiley Finance, Hoboken, NJ, 2013.

[3] Python Software Foundation, *Python Language Reference, Version 3.12*, 2023. Available: `https://www.python.org/`

[4] Ran, S., *yfinance: Yahoo! Finance market data downloader*, Python Package, 2019. Available: `https://pypi.org/project/yfinance/`

[5] *Jupyter Notebooks – a publishing format for reproducible computational workflows*, Positioning and Power in Academic Publishing: Players, Agents and Agendas, 2016.