QuantPulse: Short-term Portfolio Optimizer

Ayushman Anupam (MDS202411)
Boda Surya Venkata Jyothi Sowmya (MDS202413)

Supervisor: Prof. M Chandramouli

Data Science
Chennai Mathematical Institute

Contents

1 Abstract

This project develops a dynamic, prediction-driven framework for short-term portfolio optimisation, addressing the limitations of static return and risk estimation methods. Instead of relying on long-term historical averages, expected returns are derived from model-predicted future prices generated through a rolling XGBoost forecasting system. Both return and risk are computed adaptively using recent data windows or decay-weighted schemes, enabling the framework to respond to regime shifts and rapidly changing market conditions.

The study introduces three approaches for determining short-term, risk-tolerated return targets: a finance-based heuristic, a Bayesian formulation, and a machine-learning–oriented method. These estimates are integrated with multiple allocation strategies—including Efficient Frontier, Bayesian portfolio optimisation, and HERC—within a rolling optimisation pipeline that continuously updates portfolio weights.

Experimental results demonstrate that the proposed dynamic framework provides more realistic short-term risk–return estimates and achieves superior performance compared to traditional static methods and market baselines. The findings highlight the effectiveness of combining predictive modelling with adaptive risk assessment for short-horizon portfolio construction.

2 Acknowledgement

We express our sincere gratitude to all those who supported and guided us throughout this project.

First and foremost, we are deeply thankful to **Prof. M Chandramouli**, our project guide, for his valuable insights, continuous support, and encouragement.

We also extend our thanks to the faculty and staff of the **Department of Data Science** at **Chennai Mathematical Institute**, for providing a conducive environment for research and learning. We are grateful to our peers and friends for their help, motivation, and meaningful discussions during the project.

3 Introduction

Financial markets evolve quickly due to macroeconomic news, sector trends, and changes in investor sentiment. In short-term trading, movements in returns and risk happen within hours or even minutes. Traditional long-horizon portfolio methods, such as the Markowitz mean–variance model, rely on long historical windows and therefore fail to represent the current market regime. This mismatch often results in inaccurate estimates and reduced performance for short-term decisions.

3.1 Problem Statement

Short-term portfolio optimisation faces three key issues:

1. **Rapidly changing returns.** Market conditions shift frequently, and long-term averages do not adapt fast enough, leading to misleading return estimates.
2. **Time-varying risk.** Volatility responds quickly to news, liquidity changes, and macroeconomic events. Static volatility assumptions fail to reflect these short-term fluctuations.
3. **No short-term risk-conditioned return.** Existing portfolio models are designed for long-term investment horizons and do not provide a clear way to estimate a short-term return target that respects the trader's risk tolerance.

These limitations cause traditional optimisation models to overlook recent patterns and underestimate the actual risk present in short-horizon trading.

3.2 Solution Approach

We develop a concise two-stage framework that brings together predictive modelling and rolling portfolio optimisation for short-term decision-making.

3.2.1 Stage 1: Rolling Short-Term Price Prediction

A rolling XGBoost model is used to forecast short-term prices, updating continuously as new observations become available. Expected returns are calculated from **predicted** prices, making them more responsive and representative of the current market environment.

3.2.2 Stage 2: Dynamic Portfolio Construction

At each step, risk and return are recomputed using:

- a fixed *lookback window* to capture recent behaviour, or
- a decay-weighted method that gives more weight to the latest data.

Portfolios are constructed using:

- **Efficient Frontier** — focuses on maximising expected return,
- **Bayesian Optimisation** — accounts for uncertainty in estimates,
- **HERC** — improves diversification and controls volatility.

Short-Term Risk-Tolerated Return

Existing literature doesn't define short-term risk-return measure, we introduce three new methods:

1. a finance-based heuristic for practical short-term use,
2. a Bayesian approach that incorporates parameter uncertainty,
3. a machine-learning method that adjusts targets based on model predictions.

Overall, this framework integrates predictive signals, adaptive risk estimation, and rolling optimisation to construct portfolios that better reflect real-time market behaviour and support more informed short-term trading.

4 Data Overview and Pre-processing

This section summarises the data used in the project, the selected assets, the structure of the raw dataset, and the pre-processing steps completed before modelling.

4.1 Data Sources and Frequency

All stock and index data were collected using the Yahoo Finance API at a frequency of one hour. Since the Indian stock market operates for approximately seven hours per day, each trading day provides around seven data points. Each asset is downloaded using its corresponding *ticker symbol*, which uniquely identifies the stock or index on the exchange.

The data for each asset are stored in separate files and updated regularly. These files serve as the foundation for model training, prediction, and portfolio optimisation.

Assets Considered

The project includes major Indian indices and large-cap stocks across sectors. Table 1 lists all assets used in the analysis.

Indices	**Stocks**
`^NSEI` : NIFTY_50	`RELIANCE.NS` : RELIANCE INDUSTRIES LTD.
`^NSEBANK` : NIFTY_BANK	`TCS.NS` : TATA CONSULTANCY SERVICES
`^CNXIT` : NIFTY_IT	`SUNPHARMA.NS` : SUN PHARMACEUTICAL INDUSTRIES
`^CNXPHARMA` : NIFTY_PHARMA	`ICICIBANK.NS` : ICICI BANK LTD.
`^CNXFMCG` : NIFTY_FMCG	`INFY.NS` : INFOSYS LTD.
`^CNXAUTO` : NIFTY_AUTO	`SBIN.NS` : STATE BANK OF INDIA
`^CNXMETAL` : NIFTY_METAL	`BHARTIARTL.NS` : BHARTI AIRTEL LTD.
`^CNXREALTY` : NIFTY_REALTY	`ITC.NS` : ITC LTD.
`^CNXENERGY` : NIFTY_ENERGY	`LT.NS` : LARSEN & TOUBRO LTD.
`NIFTY_FIN_SERVICE.NS` : NIFTY_FIN_SERVICE	`HINDUNILVR.NS` : HINDUSTAN UNILEVER LTD.

Table 1: Indices and Stocks Used in the Analysis

4.2 Raw Data Structure and Data Cleaning

The following fields are downloaded from Yahoo Finance: **Datetime, Open, High, Low, Close, Volume**. Only **Datetime** and **Close** are used for modelling short-term price movements.

The data undergo the following steps before modelling:

- Removal of duplicate timestamps
- Forward filling or interpolation of missing values
- Ensuring consistent hourly intervals
- Aligning timestamps across assets

These steps ensure the dataset is complete and synchronised for prediction and optimisation.

5 Data Analysis and Feature Engineering

5.1 Statistical Analysis of Returns

We first examine the statistical properties of hourly returns using normality and stationarity tests. Normality is assessed using the **Shapiro–Wilk** and **Jarque–Bera** tests, while stationarity is evaluated using the **Augmented Dickey–Fuller (ADF)** and **KPSS** tests. Across the five representative assets studied, returns consistently exhibit non-normal behaviour but remain stationary, making them suitable for modelling.

Asset	Shapiro p	JB p	ADF p	KPSS p	Conclusion
NIFTY_50	5.3×10^{-50}	0.0	1.2×10^{-29}	0.1	Not Normal, Stationary
NIFTY_BANK	2.3×10^{-47}	0.0	5.4×10^{-30}	0.1	Not Normal, Stationary
NIFTY_IT	9.1×10^{-49}	0.0	0.0	0.1	Not Normal, Stationary
NIFTY_PHARMA	4.7×10^{-47}	0.0	0.0	0.1	Not Normal, Stationary
RELIANCE	1.0×10^{-45}	0.0	0.0	0.1	Not Normal, Stationary

Table 2: Normality and stationarity test results for sample assets

5.2 Correlation Structure

We analyse the cross-asset relationships using a correlation heatmap. Most assets show strong positive correlation, which highlights the importance of diversification. Highly correlated assets can increase portfolio volatility, making correlation analysis essential for constructing risk-balanced portfolios.

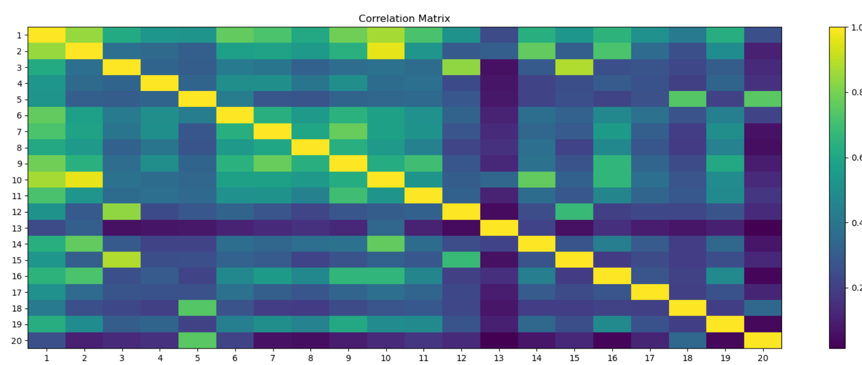

Figure 1: Correlation Heat-Map

5.3 Trend, Risk and Return Behaviour

Price trends indicate clear short- and medium-term variations across sectors. The risk–return plot illustrates that assets differ substantially in volatility and expected return, and these values evolve over time. Although high-volatility assets often provide higher return, the relationship is not stable, reinforcing the need for dynamic modelling.

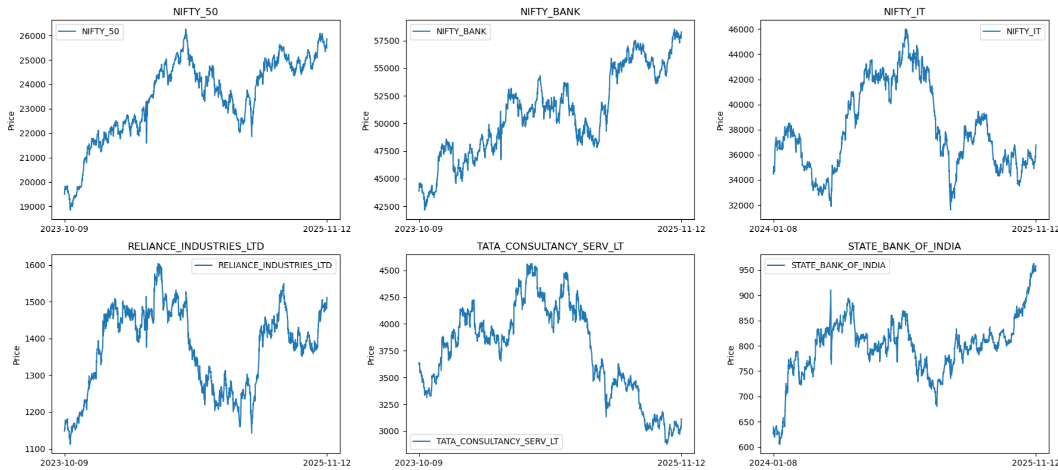

Figure 2: Price of Some Stocks and Indices

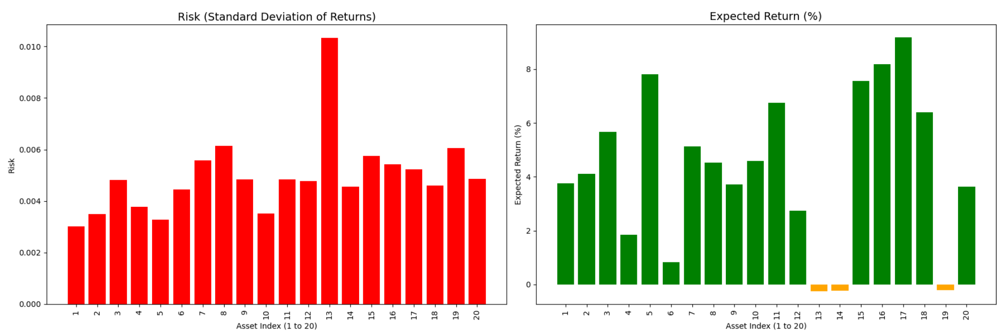

Figure 3: Risk-Return Plot

5.4 Feature Engineering

Only time and close price were available initially. To improve predictive performance, we constructed several lag-based and seasonal features that help the model learn short-term dynamics and recurring patterns. These include:

- past 8 hours of price data (recent movement),
- price at the same hour over the past 7 days (weekly pattern),
- price on the same date over the past month (monthly trend),
- previous year's price on the same date (yearly seasonality).

These engineered features capture both short-term behaviour and longer cyclical patterns that affect intraday price movements.

Creation of Prediction Files

For each asset, we create a dedicated file named `[stock_name]_prediction`, containing: Time, Actual Price, Predicted Price, Return.

The first two fields are available immediately, while predicted price and return are populated during the modelling process. This structure ensures clean separation between original market data and model-generated values and simplifies subsequent portfolio optimisation.

6 Future Price Prediction Using XGBoost

To generate one-hour-ahead price forecasts for all stocks and indices, a rolling prediction framework is implemented. This ensures the model learns from the most recent data and produces timely predictions suitable for short-term portfolio decisions.

6.1 Rolling Prediction Framework

For each asset, the workflow is as follows:

1. **Load Data:** The historical lag-feature dataset and the prediction file are loaded to identify missing predicted prices.
2. **Identify Missing Points:** Rows with empty predicted price fields are located. The earliest missing timestamp initiates the rolling loop.
3. **Rolling-Window Training:** At each step, the model trains on the most recent 2000 data points, focusing on evolving market patterns.
4. **One-Step Prediction:** The next one-hour price is predicted and saved after every 20 steps to maintain consistency.

6.2 XGBoost Model Configuration and Evaluation

The model is an XGBoost regressor with the following hyper-parameters:

```
model = XGBRegressor(
    n_estimators=650, random_state=42,
    colsample_bytree=0.8, learning_rate=0.06,
    max_depth=6, n_jobs=-1
)
```

Predictions are evaluated using:

- **MSE (%)** — squared percentage error
- **MAPE (%)** — average percentage deviation from actual prices

6.3 Prediction Results

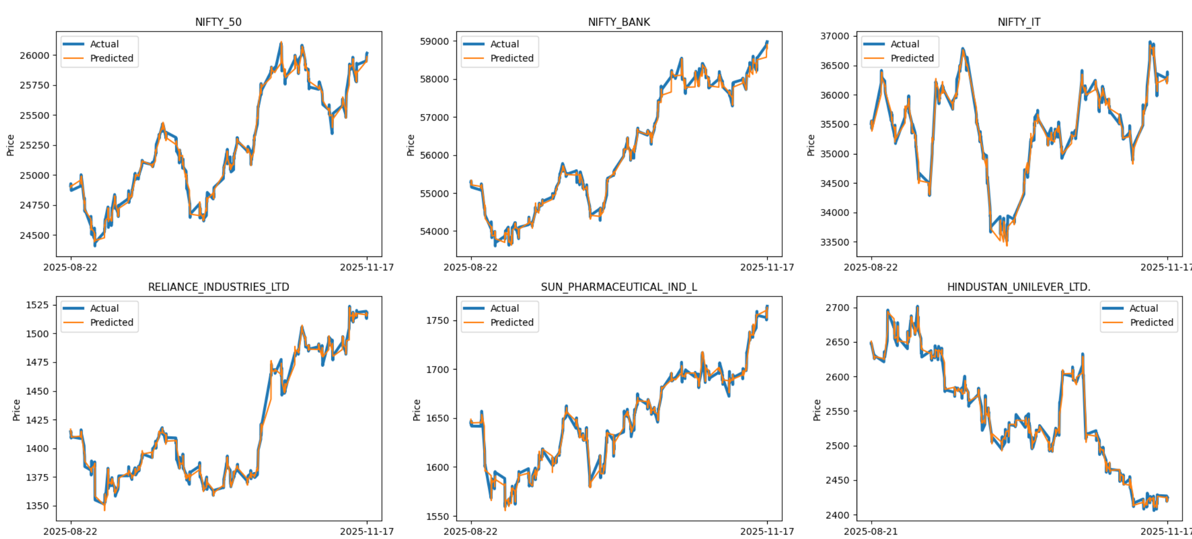

Figure 4: Actual vs. Predicted Prices for Sample Assets

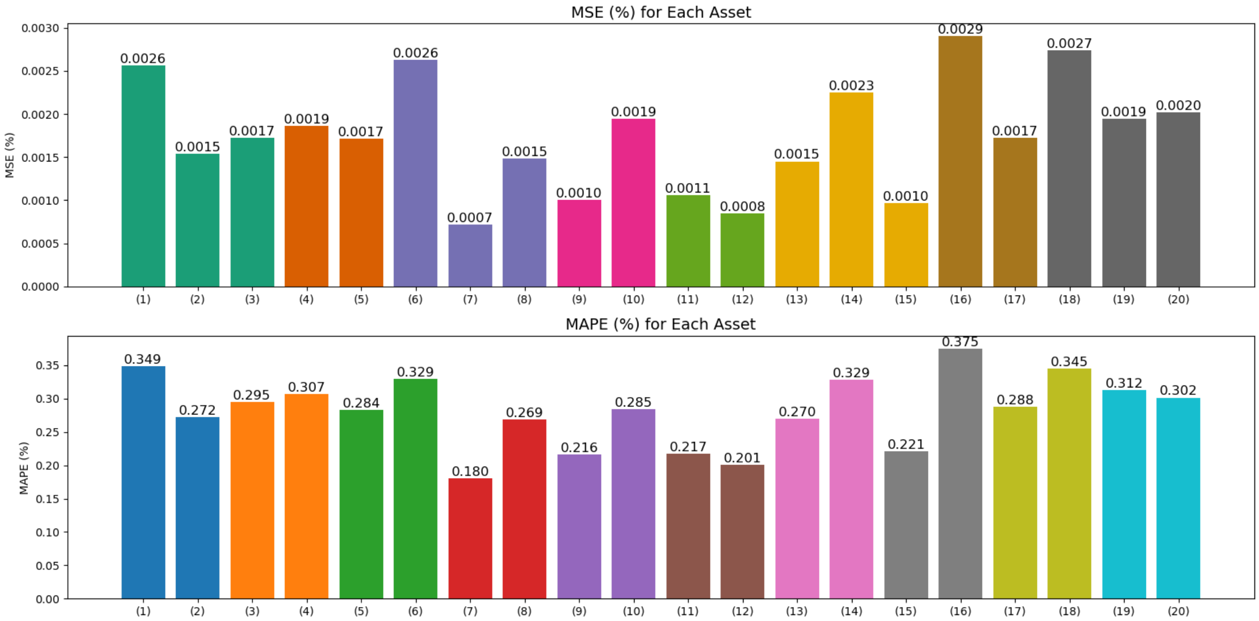

Figure 5: MSE(%) and MAPE(%) Across All Assets

Key Observations

- MSE values are between **0.0007–0.0025**, indicating stable predictions.
- MAPE ranges from **0.18%–0.37%**, showing predicted prices closely follow actual values.
- Volatile sectors (Realty, Metal, Pharma) maintain low errors.
- Major indices (NIFTY 50, FMCG, IT) have the lowest errors due to consistent patterns.

The XGBoost-based rolling framework accurately captures short-term price movements across all assets. These predictions provide reliable input for subsequent portfolio optimisation, ensuring timely and robust decision-making.

7 Portfolio Construction

In this project, we construct portfolios using three different optimisation frameworks, each offering a unique balance between risk and return. The methods used are: (i) Efficient Frontier (Mean–Variance Optimisation), (ii) Bayesian Portfolio Method, and (iii) Hierarchical Equal Risk Contribution (HERC). These three methods allow us to compare return-focused, balanced, and risk-focused portfolio strategies using the predicted asset returns.

7.1 Efficient Frontier (Mean-Variance Optimisation)

The Efficient Frontier, introduced by Markowitz, is a classical optimisation approach that identifies portfolios that maximise return for a given level of risk, or equivalently minimise risk for a given return. In our project, the expected returns are not historical averages but **model-predicted short-term returns**, which makes the method suitable for intraday or near-term trading. Key Components of this method include:

- **Predicted Returns:** The vector of expected returns $\boldsymbol{\mu}$ is derived from the machine learning prediction model.
- **Dynamic Covariance Matrix:** We compute Σ using only recent data to capture the latest volatility patterns instead of relying solely on long historical windows.
- **Optimisation Objective:** The model solves a quadratic optimisation problem that maximises the Sharpe ratio:

$$\text{Sharpe}(\mathbf{w}) = \frac{\mathbf{w}^\top \boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}$$

  subject to:

$$\sum_i w_i = 1, \qquad 0 \leq w_i \leq 1.$$

This approach typically results in weight vectors that aim for the highest risk-adjusted returns based on the predicted market behaviour.

7.2 Bayesian Portfolio Method

The Bayesian Portfolio Method incorporates prediction uncertainty by blending model-predicted returns with prior beliefs. This helps stabilise portfolio weights, particularly when short-term predicted returns are noisy. Key Components of this method include:

- **Posterior Expected Return:** The predicted return $\hat{\mu}$ is combined with a prior mean μ_0 to obtain a posterior estimate:

$$\mu_{\text{post}} = \left(\Sigma^{-1} + \tau I\right)^{-1} \left(\Sigma^{-1}\hat{\mu} + \tau \mu_0\right).$$

- **Posterior Covariance:**

$$\Sigma_{\text{post}} = \left(\Sigma^{-1} + \tau I\right)^{-1}.$$

- **Optimisation:** We maximise $\mathbf{w}^\top \mu_{\text{post}}$ subject to variance, weight sum, and long-only constraints.

This method provides smoother and more reliable weight allocations by reducing the influence of extreme short-term predictions.

7.3 Hierarchical Equal Risk Contribution (HERC)

HERC is a risk-based portfolio construction method that does not rely on expected returns. It instead uses hierarchical clustering to group correlated assets and then allocates risk equally across clusters. Key Components of this method include:

- **Correlation-Based Clustering:** Assets are organised into clusters using hierarchical clustering based on their correlation structure.
- **Risk Allocation by Cluster:** Risk is first allocated across clusters and then within each cluster, resulting in a diversified and balanced portfolio.
- **No Dependence on Predicted Returns:** This makes HERC particularly suitable during high volatility or when predictions are unstable.

HERC avoids matrix inversion and reduces concentration risk, making it robust for short-term decision-making.

7.4 Comparison of Methods

Efficient Frontier

- **Best for Return**
- Uses predicted returns to try to earn higher profit.
- Works well when the model predictions are strong.
- **Trade-off:** Can take more risk to chase higher returns.

Bayesian Method

- **Balanced: Return + Risk**
- Gives a safer version of the predicted return.
- Good when you want reasonable return without taking too much risk.
- **Trade-off:** Returns may be slightly lower, but risk is reduced.

HERC (Risk Parity)

- **Best for Lower Risk**
- Focuses on spreading risk evenly across assets.
- Useful when market is uncertain or predictions are unreliable.
- **Trade-off:** Doesn't chase high return; aims for safety.

8 Results

This section summarises the performance of the three portfolio construction methods—Efficient Frontier, Bayesian, and HERC—over different investment periods. Portfolios are evaluated based on predicted returns, volatility, and Value at Risk (VaR).

Evaluation Inputs Each portfolio is constructed using three inputs:

- **Investment date (apply_date):** Start of the portfolio.
- **Withdrawal date (withdraw_date):** End of the position.
- **Lookback period (k):** Number of past days used for rolling optimisation.

8.1 Portfolio Performance Examples

Example 01:

Parameter	Efficient Frontier	Bayesian	HERC
Apply Date	2025-04-06	2025-04-06	2025-04-06
Withdraw Date	2025-11-06	2025-11-06	2025-11-06
Lookback Period	40 days	40 days	40 days
Return	49.26%	50.86%	20.56%
Avg Volatility	0.00414	0.00401	0.00419
Parametric VaR (95%)	0.6189%	0.6128%	0.5872%

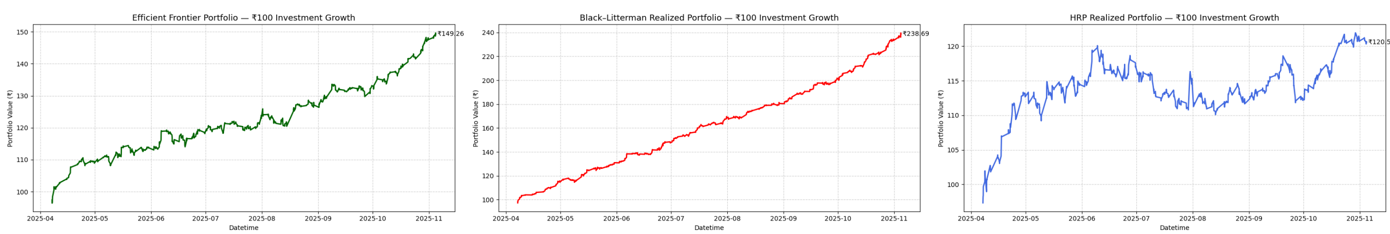

Figure 6: Efficient Frontier, Bayesian, and HERC returns (Example 01)

Example 02:

Parameter	Efficient Frontier	Bayesian	HERC
Apply Date	2025-10-06	2025-10-06	2025-10-06
Withdraw Date	2025-11-06	2025-11-06	2025-11-06
Lookback Period	10 days	10 days	10 days
Return	15.05%	11.15%	4.10%
Avg Volatility	0.00271	0.00259	0.00327
Parametric VaR (95%)	0.3664%	0.3521%	0.5112%

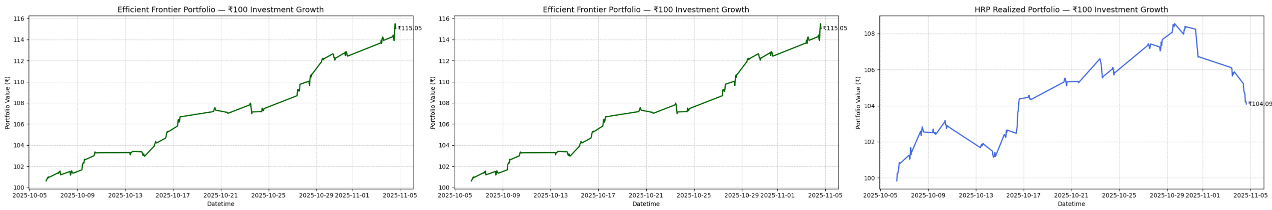

Figure 7: Efficient Frontier, Bayesian, and HERC returns (Example 02)

8.2 Overall Results Interpretation

Monthly Return Comparison

Period	Efficient Frontier	Bayesian	HERC	Market
7 Months (Ex 01)	5.87%	6.05%	2.71%	1.02%
1 Month (Ex 02)	15.05%	11.15%	4.10%	1.02%

The two evaluation periods show a consistent pattern across all three portfolio construction methods. The Efficient Frontier generally achieves the highest returns because it directly exploits the model-predicted short-term returns. However, this also leads to slightly higher volatility, as the method tends to allocate more weight to assets with strong predicted momentum.

The Bayesian portfolio provides a more balanced outcome. By shrinking predicted returns toward a prior mean, it reduces sensitivity to noisy or extreme predictions. As a result, it achieves returns close to the Efficient Frontier but with noticeably smoother volatility and lower Value at Risk (VaR). This makes the Bayesian approach more stable, especially in periods where prediction uncertainty is high.

HERC, being a pure risk-based method, produces the most conservative performance. Its returns are lower because it does not use predicted returns, but the volatility remains consistently low due to equal-risk allocation across hierarchical clusters. This method performs best in uncertain or turbulent market phases where prediction-based strategies may fluctuate.

Overall, the results indicate a clear trade-off:

- **Efficient Frontier**: Highest return, moderate risk.
- **Bayesian**: Strong return with the best stability.
- **HERC**: Lowest return, lowest risk.

These findings show that incorporating short-term predictions improves performance meaningfully, while Bayesian shrinkage and HERC diversification help manage risk during volatile market conditions.

9 Industrial Applications

The portfolio optimisation and short-term return forecasting framework developed in this study has several real-world applications across the financial industry. Some major use-cases include:

- **Short-term trading and intraday strategies**
  Predicted 1-hour returns enable traders to make fast, data-driven decisions during volatile periods. This is highly useful for high-frequency and intraday strategies where timing is crucial.
- **Algorithmic trading systems**
  The model can be integrated into automated trading bots that generate buy/sell signals and rebalance portfolios dynamically in response to market movements.
- **Risk management and hedging**
  By incorporating recent volatility and covariance dynamics, firms can identify riskier assets quickly and hedge market exposure more effectively.
- **Portfolio advisory and robo-advisory platforms**
  Investment platforms can use these methods to offer personalised portfolios tailored to user preferences—high return, balanced, or low risk.
- **Market surveillance and stress testing**
  Short-term prediction behaviour helps in detecting market regime changes and monitoring risk spikes during uncertain periods.

10 Limitations

Although the proposed framework provides meaningful insights into short-term portfolio performance, it relies on several simplifying assumptions that may not hold in real-world markets:

- **Instant trade execution**
  The model assumes trades occur instantly at the predicted price, ignoring delays, slippage, and execution time.
- **No transaction costs**
  Brokerage fees, STT, taxes, and other trading charges are ignored. These can significantly impact frequent short-term trades.
- **No market impact**
  It is assumed that trades do not affect market prices. While reasonable for small positions, large institutional trades may move prices.
- **No liquidity constraints**
  The model assumes that assets can always be bought or sold in any quantity. During low-volume periods, this may not be feasible.

11 Conclusion

This project demonstrates that combining short-term return forecasting with advanced portfolio optimisation techniques can significantly improve risk-adjusted performance. The Efficient Frontier approach provides higher returns when predictions are strong, the Bayesian method ensures stability under uncertainty, and the HERC method offers robustness in highly volatile conditions. By using predicted returns, dynamic covariance estimation, and risk-based clustering, the framework successfully adapts to fast-changing market behaviour. Overall, the project highlights the importance of integrating predictive modelling with portfolio construction for intraday and short-term trading applications.

12 Future Work

There are several directions for future improvements:

- **Incorporating transaction costs and slippage**
  Including realistic trading costs will make performance metrics more accurate.
- **Using nonlinear models for return prediction**
  LSTM, Transformers, and GARCH-based hybrid models may improve short-term predictive power.
- **Regime detection models**
  Identifying bull/bear/sideways regimes can help adjust risk levels more effectively.
- **Live deployment and backtesting engine**
  Implementing the model in a live paper-trading environment will validate real-world feasibility.
- **Handling liquidity and market depth**
  Future models can incorporate order book data to adjust weights based on actual trading volume.

13 References

- "yahoo-finance Python API," ReadTheDocs. `https://python-yahoofinance.readthedocs.io/en/latest/api.html`
- "XGBoost: Extreme Gradient Boosting," GeeksforGeeks. `https://www.geeksforgeeks.org/machine-learning/xgboost/`
- "Efficient Frontier — Modern Portfolio Theory," Corporate Finance Institute (CFI). `https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/efficient-frontier/`
- "Bayesian Machine Learning and Portfolio Applications," DataRobot Blog. `https://www.datarobot.com/blog/bayesian-machine-learning/`
- "Hierarchical Clustering Portfolio using Riskfolio-Lib," ReadTheDocs. `https://riskfolio-lib.readthedocs.io/en/latest/hcportfolio.html`

14 Work Contribution Summary

Problem Selection

- Both team members contributed equally to selecting the problem statement through discussions with the supervisor.
- **Supervisor:** Prof. M. Chandramouli

Code Contributions

- **Ayushman**
  - Data gathering and preprocessing pipeline
  - Portfolio construction (Efficient Frontier, Bayesian, HERC)
- **Sowmya**
  - Exploratory Data Analysis (EDA)
  - Short-term price prediction for each stock (XGBoost rolling framework)

Report and Presentation Contributions

- **Ayushman**
  - Problem statement and solution approach
  - Data gathering section
  - Portfolio construction section
  - Results (part), Industrial Applications, and Limitations
- **Sowmya**
  - Introduction, Abstract, and Acknowledgement
  - Exploratory Data Analysis (EDA)
  - Prediction methodology and implementation
  - Results (part), Conclusion, and Future Work