

Portfolio Simulation System

Technical Report

Abstract

This report describes the development of a dynamic portfolio simulation system that integrates PCA-based fundamental analysis with LSTM-driven technical price forecasting. The system extracts key financial features, such as P/E ratios, ROE, and dividend yield, from daily data (January 2015 to April 2025) for 29 stocks. Principal Component Analysis condenses these metrics into a composite quality score, while LSTM neural networks, trained on technical indicators such as MA5, RSI, and MACD, predict short-term price movements. The final stock recommendation is determined by combining these signals with a volatility penalty to manage risk. Backtesting over a ten-year period demonstrates that a 14-stock, semiannually rebalanced portfolio can deliver an annual return of approximately 25.5% with a Sharpe ratio of 1.10. The significance of this integrated system lies in its ability to offer strong, data-driven decision support for investors across varied market conditions.

1. Introduction

Within the volatile financial markets of today, clear and effective decision-making is critical. This report presents a hybrid stock recommendation engine that addresses investor needs by synergistically combining traditional fundamental analysis with modern machine learning techniques. The project's goals are to:

- Combine key fundamental indicators using PCA to generate a single quality score.
- Use LSTM models to forecast short-term price trends based on technical indicators.
- Integrate both approaches via a combined score that also accounts for stock volatility and sector diversification. Data is sourced from Yahoo Finance for 29 stocks spanning various sectors from January 2015 to April 2025. Details on the methodology, portfolio simulation, performance evaluation, and recommendations for further improvement are laid out below..

2. Methodology and Analysis

2.1 Data Preparation

- Sources: Daily Yahoo Finance data, including fundamental figures and historical prices for 29 stocks.
- Cleaning & Feature Engineering:
 - Missing fundamental values are imputed using the median.
 - Technical indicators (MA5, RSI, MACD) are computed.

- Fundamentals are normalized and reduced to a composite score via PCA, where the first principal component explains approximately 53% of the variance.
- **Data Splitting:** The dataset is partitioned chronologically into 80% training and 20% validation segments.

2.2 Modeling Approach

- **Baseline Model:** A 5-day moving average predictor establishes an RMSE benchmark (≈ 17).
- **LSTM Price Forecasting:**
 - Models are trained using a 5-day input sequence with batch size 32 and 10 epochs.
 - For example, the AAPL model's RMSE improved from 10.47 (baseline) to 3.75.
 - Directional accuracy hovers around 50%, contributing to the overall signal despite its modest level (close to random guess).
- **Hybrid Signal Integration:**
 - The Final Combined Score is Calculated by assigning weights of 50% to the PCA-derived fundamental score, 30% to the inverse LSTM RMSE, and 20% to the directional accuracy, minus a 10% penalty for high volatility.
 - This integrated approach ensures selections are based on both potential profitability and risk management.

2.3 Portfolio Simulation and Evaluation

- **Simulation Setup:**
 - Portfolios were simulated by varying stock composition, rebalancing frequency (semiannual, quarterly, annual, etc.), and investment time horizon.
 - A selected 14-stock portfolio configuration demonstrated an annual return of approximately 25.53% and a Sharpe ratio of 1.10.
- **Risk and Diversification Measures:**
 - Investment criteria enforces sector diversification limits and reduces exposure to high-volatility stocks.
 - Weighting by inverse volatility (risk-parity) further refines the portfolio balance.
- **Technical Performance:**

- The LSTM models reduce RMSE by around 65% compared to the simple moving average.
- PCA effectively minimizes co-linearity among fundamental factors, emphasizing quality attributes such as low P/E ratios and high ROE/dividend yield.

3. Conclusions and Recommendations

The hybrid portfolio simulation system represents a strong base for a data-driven tool purposed for stock selection and portfolio simulation that seamlessly integrates fundamental quality and technical forecasting. Key conclusions:

- The combined PCA-LSTM framework delivers strong predictive power and superior risk-adjusted performance.
- Backtesting results validate the system's capacity to outperform traditional benchmarks like the S&P 500.
- The use of a volatility penalty and sector diversification rules enhances the scope and malleability of the portfolio under varied market conditions.

Recommendations for Future Work:

- **Macroeconomic Enhancements:** Incorporate interest rates, CPI, employment, and PMI to capture broader economic trends.
- **Real-Time Data Integration:** Enable intraday model updates to improve responsiveness and allow future price predictions.
- **Advanced Optimization:** Explore hyperparameter tuning via grid, random, or Bayesian search methods.
- **Sentiment Analysis:** Supplement quantitative signals with qualitative insights from news and social media (Text Mining).

Overall, this dynamic simulation system provides valuable decision-support for investors, facilitating a strategic approach to stock selection and portfolio management.

Visualizations.

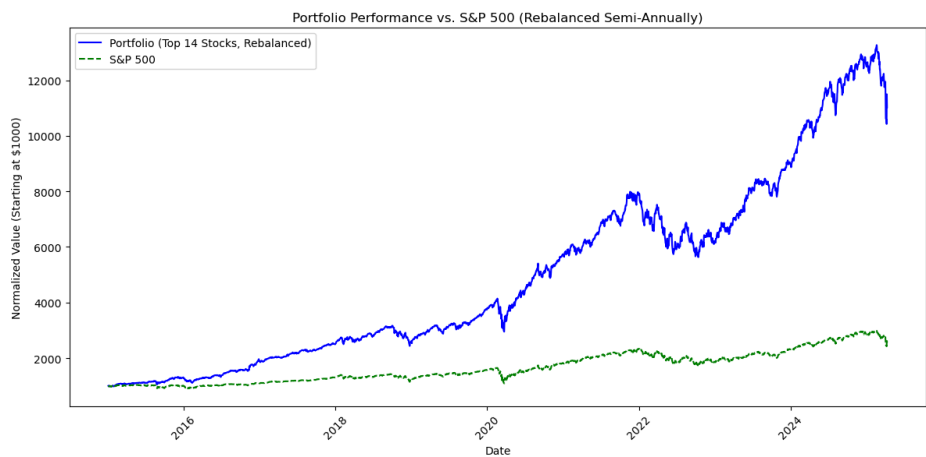


Fig.1

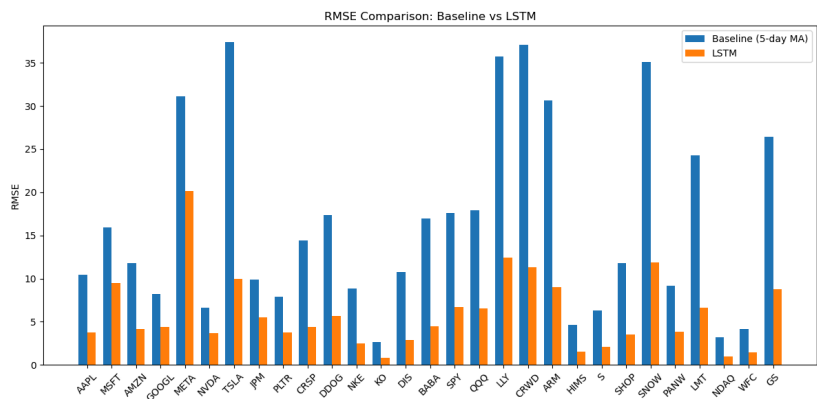


Fig.2

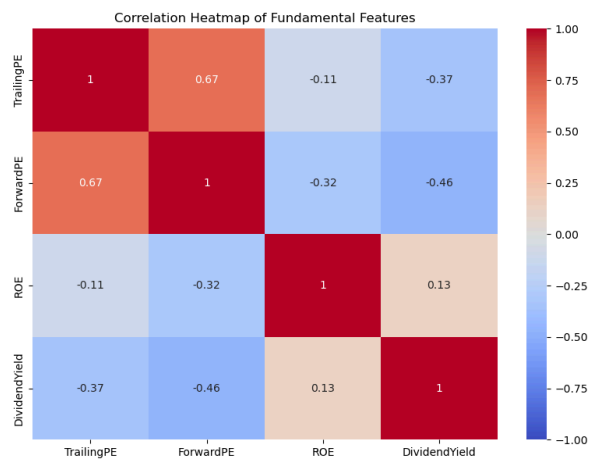


Fig.3

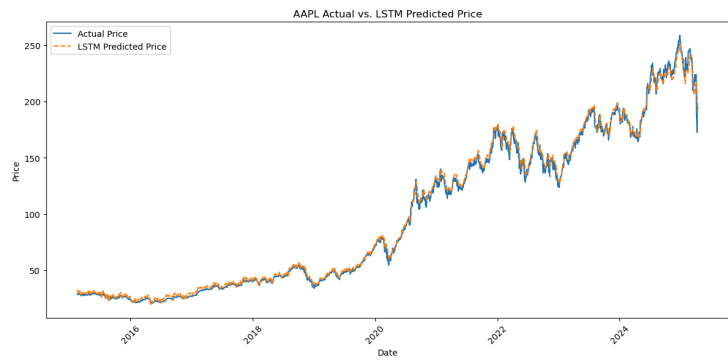


Fig.4