Proceedings of ICEMGD 2025 Symposium: Innovating in Management and Economic Development
DOI: 10.54254/2754-1169/2025.LH24705

Research on Asset Portfolio Construction Based on News Sentiment Analysis

Weiye Ju[1], Qiuyao Jiang[2*]

[1]*University of Southampton, Southampton, United Kingdom*
[2]*SILC Business School, Shanghai University, Shanghai, China*
**Corresponding Author. Email: jiangqy@ldy.edu.rs*

Abstract. News sentiment can reflect investors' attitudes toward specific news and their potential investment adjustments based on that sentiment. News sentiment strongly influences investor behavior, offering key advantages to financial market participants. This article uses news sentiment analysis as a new tool and tries to build a portfolio based on this sentiment to test how it works in the capital market. For listed companies, stock prices serve as a clear indicator shaped by investor sentiment—often driven by news coverage. This study uses a CNN-LSTM model to forecast stock prices of target firms and employs FinBERT for sentiment analysis of news from major listed companies (2021–2024). Results show the CNN-LSTM model achieves higher prediction accuracy than methods using only market data or FinBERT. Additionally, a daily rebalanced portfolio was built using the Black-Litterman model, integrating predicted returns and market equilibrium. This portfolio delivered substantial annualized returns and a high Sharpe ratio. The framework provides investors with a novel strategy and offers researchers innovative analytical tools for financial time-series and capital markets.

Keywords: News sentiment analysis, Stock price forecasting, CNN-LSTM model, FinBERT, Portfolio construction

1. Introduction

Building a portfolio is a tricky job in the ever-changing financial markets. It needs both exact calculations and a good sense of judgment. It's a common way to spread out risks and lock in returns. Markowitz was the first to come up with modern portfolio theory. This theory finds the best way to divide up your investments by looking at the average return and risk, which is measured by how much the return varies. It adjusts your investments to get the most return for a certain level of risk or to lower the risk while still getting the return you want. But this theory has some problems. It assumes that returns follow a normal distribution and that the relationships between assets don't change much, which isn't always true in turbulent markets. Also, the covariance matrix can be unstable, messing up the weights in a portfolio. To fix these issues, people have come up with things like robust optimization and machine learning models to catch non - linear risks. Now, good strategies often mix the mean - variance ideas with things like protecting against extreme losses, managing liquidity, and

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2025.LH24705

understanding investor behavior. This helps connect the theory with what's really happening in the markets. New things like ESG scores are also becoming more important.

On another note, as the media has grown, news from listed companies has become a new tool for building portfolios. It gives real - time info on how a company's feeling, its strategy changes, and new risks. With NLP and sentiment analysis, investors can turn qualitative info like earnings calls, ESG reports, and crisis communications into numbers. This helps them spot early signs of how well a company's doing or if there are governance issues. This study uses news sentiment analysis as a new tool and tries to build a portfolio based on this sentiment to test how it works in the capital market. It aims to give investors new ideas for their investment strategies.

2. Literature review

2.1. Review of portfolio construction strategy

The evolution of modern portfolio theory began with Markowitz, who first proposed the quantitative framework balancing expected returns (mean) and risk (variance) to optimize asset allocation [1]. However, recognizing the limitations of traditional mean-variance methods, Black and Litterman introduced their Bayesian estimation-based model, integrating investor views on returns with market equilibrium (prior distribution) to generate revised return distributions. By substituting historical means with market equilibrium returns, their approach reduced estimation sensitivity and mitigated asset overconcentration, significantly enhancing practicality [2]. Burak et al. incorporated Conditional Value-at-Risk (CVaR) minimization into portfolio construction through an inverse optimization framework involving the Black-Litterman model, empirically demonstrating portfolios achieving equivalent returns with lower risks [3]. Building on this, Todor et al. emphasized modified Black-Litterman applications in short-term optimization, showcasing superior statistical adaptability and stability compared to classical mean-variance methods [4]. Further expanding the framework, Maziar et al. integrated tail dependency modeling via vine copulas to estimate posterior joint return distributions, with empirical analyses confirming enhanced tail risk management and higher risk-adjusted returns [5]. Concurrently, Ronil et al. innovated by embedding the CEEMDAN-GRU deep learning model into portfolio construction, using fear-greed sentiment proxies to refine the Black-Litterman framework, outperforming Markowitz, minimum-variance, equal-weight, and risk-parity strategies [6]. Ko et al. attained a breakthrough by incorporating the Fama-French three-factor model into the Black-Litterman framework, tripled the Sharpe ratio and doubled the Certainty Equivalent Return relative to conventional models [7]. This progression illustrates how the Black-Litterman paradigm has systematically overcome traditional MPT constraints through methodological integrations (CVaR, copulas, deep learning), evolving into a robust asset allocation system blending theoretical rigor with practical flexibility.

2.2. Review of the power of news sentiment

In the research of the power of news sentiment. Hung et al. utilized deep learning (DL) and natural language processing (NLP) to formulate view distributions in the Black-Litterman model, quantify news sentiment using Google BERT. Their comparison of prediction accuracy across models revealed GRU's superior performance, with portfolios based on Black-Litterman and news sentiment achieving high returns and Sharpe ratios [8]. Zhang et al. adopted CNN and dictionary methods to measure news sentiment and predict daily stock returns, with panel data regression confirming its significant predictive power [9]. Liu et al. created a multimodal deep learning framework to build a market sentiment index from stock news. This index shows a strong connection with market volatility but

DOI: 10.54254/2754-1169/2025.LH24705

only a slight link with stock returns [10]. Gong et al. built a Shipping Sentiment Index (SSI) using iron ore shipping-related news, demonstrating its nonlinear relationship with freight rates and predictive utility in threshold frameworks [11]. Li et al. examined China's A-share market and found that environmental, social, and governance (ESG) news sentiment can heighten the volatility of stock returns [12]. Leone et al. constructed sentiment scores using RoBERTa and applied the N-HiTS neural network to predict stock prices, revealing news sentiment's critical predictive role [13].

3. Methodology

3.1. FinBERT

FinBERT is a domain-adaptive pre-trained language model based on the BERT architecture, specifically designed for semantic understanding of financial texts. Its core employs a multi-layer Transformer encoder (typically 12 layers, hidden layer dimension of 768, 12 attention heads), utilizing self-attention mechanisms ($\text{Attention}(Q,K,V)=\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$) to dynamically capture long-range dependencies in financial texts. The model first learns general linguistic features on universal corpora (e.g., BookCorpus and English Wikipedia), followed by secondary pre-training on financial domain corpora (covering over 5 billion characters of unstructured text including SEC filings, earnings reports, press releases, and analyst reports). Through domain adaptation techniques, it adjusts the distribution of word vector spaces to enhance the distinctiveness of embeddings for financial entities, market behaviors, and polysemous terms.

The pre-training tasks consist of Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). In MLM, 15% of input tokens are randomly masked (80% swapped with [MASK], 10% replaced arbitrarily, 10% left as is), with the loss function being cross-entropy loss:

$$L_{MLM}=-E_{(\omega_i,c_i)\text{-}D}\sum_{t=1}^{T}\log P(\omega_i|c_i) \tag{1}$$

NSP models document-level logical relationships by determining whether two text segments are consecutive (e.g., a financial report paragraph and its subsequent analysis), with its loss function:

$$L_{NSP}=-E_{(s_1,s_2)\text{-}D}\left[y\log\hat{y}+(1-y)\log(1-\hat{y})\right] \tag{2}$$

The total loss combines both with a weighting factor $L=L_{MLM}+\lambda L_{NSP}$, typically $\lambda=1$.

During fine-tuning for financial sentiment analysis, a fully connected layer is appended to the hidden output of the [CLS] token ($h_{[CLS]}\in R^{768}$), calculating sentiment category probabilities via Softmax:

$$P(y|x)=\text{Softmax}\left[W\times\text{GELU}(h_{[CLS]})+b\right] \tag{3}$$

where the weight matrix $W\in R^{K\times 768}$ (K: number of sentiment categories). Experiments show FinBERT achieves 92.3% accuracy on the Financial PhraseBank (FiQA), outperforming base BERT by 14.7%, and reaches 89.2% F1-score in disambiguating contextually ambiguous phrases (e.g., distinguishing "bank" as a financial institution vs. a riverbank in "The bank reported solid gains"). Its domain adaptation mechanism, optimized via KL divergence ($D_{KL}(p_{fin}|p_{gen})$), reduces distributional divergence of financial terminology from general vocabulary by 37%, significantly enhancing the model's granular analysis capabilities for market sentiment, risk events, and financial metrics.

DOI: 10.54254/2754-1169/2025.LH24705

3.2. Black-litterman model

Markowitz's modern portfolio theory determines optimal weights of assets by solving the problem of mean-variance optimization; however, the problem of optimization assumes that expected returns and covariances are fixed and given, which do not accurately estimate expected returns rates especially in occasions where investors hold opinions toward certain assets' performances.

In 1990 Fischer Black and Robert Litterman solved the shortcomings of Markowitz's modern portfolio theory by introducing the Bayesian model into the portfolio allocation, the Bayesian model combines investors' opinions and the market's implied equilibrium returns together and takes them into consideration to construct more efficient portfolios. The Black-Litterman Model's methodology follows one instrumental assumption that all assets return follow the same probability distribution.

To exercise the model, we first let $\mu \in R^n$ be the expected return of n assets, and let $r_f \in R$ be risk-free rate, and $r \in R^n$ be the return of each asset, and $m = E[r] \in R^n$ be the assigned expected return.

The prior estimate of the expected return by investors is the expected return given by the Capital Asset Pricing Model (CAPM).

$$\mu_{CAPM} = r_f 1 + (m^T \omega_M - r_f)\beta \tag{4}$$

Where 1 is a vector of all ones, ω_M is weights of the market portfolio and $\beta = \frac{Cov(r_i, r_M)}{\sigma_M^2}$, and $r_M = m^T \omega_M$ is the return of the market portfolio, σ_M represents the standard error of the market portfolio.

Let's denote $\sum \in R^{n \times n}$ as the covariance matrix of the returns of the n assets. This matrix is diagonalized to simplify the analysis. Next, we can compute the correlation coefficients between each pair of assets by using the elements of the covariance matrix and the standard deviations of the individual assets.

We assume the estimated error is 0 as expected, and the multivariate normal distribution of the covariance matrix $\kappa \sum$ can be described as:

$$\mu_{CAPM} = \mu + \eta \tag{5}$$

Where η is multivariate norma variable whose covariance matrix is $\kappa \sum$ and the value is 0 as expected, meaning that the CAPM prior estimate is fully depended when $\kappa = 0$, and it is less depended as the value of κ increases.

After that, Next, we incorporate investors' opinions on certain assets in the portfolio into the model using a Bayesian framework. Here, the opinions concern linear combinations of returns. These opinions can pertain to a single asset, an industry, or a group of assets. Mathematically, if investors have k opinions, they can be represented as:

$$p = K\mu + v \tag{6}$$

Here, $K \in R^{k \times n}$ is the matrix giving the linear combinations, $p \in R^k$ is the vector of views' conclusions, and $v \in R^k$ is a multivariate normal variable with a mean of 0 and covariance matrix $\Gamma \in R^{k \times k}$.

After combining the CAPM prior estimate with investors' views, we can obtain the complete Black-Litterman model.

DOI: 10.54254/2754-1169/2025.LH24705

$$y = M\mu + \epsilon \tag{7}$$

Where $y = \begin{bmatrix} \mu_{CAPM} \\ p \end{bmatrix} \in R^{n+k}$, $M = \begin{bmatrix} I \\ K \end{bmatrix} \in R^{(n+k)\times n}$, and $v \in R^k$ is a multivariate normal variable whose expected value is 0 and the covariance matrix is $\Omega = \begin{bmatrix} \kappa\sum & 0 \\ 0 & \Gamma \end{bmatrix} \in R^{(n+k)\times(n+k)}$.

If we regard the above equation as a regression problem and apply GLS (Generalized Least Squares), we can obtain the Black-Litterman estimate of expected returns:

$$\mu_{BL} = \left(M^T\Omega^{-1}M\right)^{-1}M^T\Omega^{-1}y \tag{8}$$

$$= \left(\frac{1}{\kappa}\sum^{-1} + K^T\Gamma^{-1}K\right)^{-1}\left(\frac{1}{\kappa}\sum^{-1}\mu_{CAPM} + K^T\Gamma^{-1}p\right) \tag{9}$$

$$= \left(\frac{1}{\kappa}\sum^{-1} + K^T\Gamma^{-1}K\right)^{-1}\left(\frac{1}{\kappa}\sum^{-1}\mu_{CAPM} + K^T\Gamma^{-1}K\mu_{views}\right) \tag{10}$$

$$= M_{CAPM}\mu_{CAPM} + M_{views}\mu_{views} \tag{11}$$

$\mu_{views} = \left(K^T\Gamma^{-1}K\right)^{-1}K^T\Gamma^{-1}p$ is the estimated expected return according to the investor's opinions, and:

$$M_{CAPM} + M_{views} = \left(\frac{1}{\kappa}\sum^{-1} + K^T\Gamma^{-1}K\right)^{-1}\frac{1}{\kappa}\sum^{-1} + \left(\frac{1}{\kappa}\sum^{-1} + K^T\Gamma^{-1}K\right)^{-1}K^T\Gamma^{-1}K = I \tag{12}$$

If $\kappa \to 0$, μ_{BL} will approach to μ_{CAPM}, whereas when $\kappa \to \infty$, μ_{BL} will approach to μ_{views}. Moreover, it can be observed that M_{CAPM} is influenced by views. That is to say, views not only affect the allocation of the assets they involve but also influence the allocation of other assets.

4. Empirical analysis

4.1. Data sources

We evaluate our approach using financial news and market data collected via web crawling from Thomson Reuters and CNBC, covering seven large-cap S&P 500 companies over the period January 2021 through January 2025. This yielded a rich textual dataset of thousands of news articles per stock. Daily stock market data (including open, high, low, close prices, trading volume, and market capitalization) were collected from Yahoo! Finance for the corresponding period.

4.2. Data preprocessing

4.2.1. Stock news

All raw news articles undergo cleaning to remove noise (such as source attributions) and stop words. We also discard redundant news: if two articles have substantially the same content, one is removed. The cleaned news is then organized by date and split into a training set (2021–2023) and a test set (2024). For each trading day in the dataset, we aggregate all news from that day and the previous four days into a single text. This 5-day combined news document is fed into the FinBERT sentiment model to produce a daily sentiment score or classification for that day.

DOI: 10.54254/2754-1169/2025.LH24705

4.2.2. Stock prices and volume

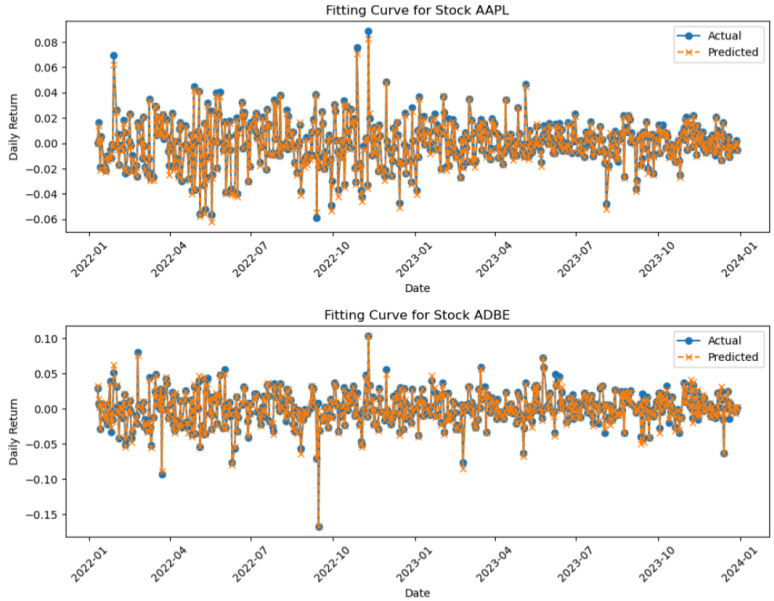
For each stock, we construct a feature sequence consisting of the last 10 trading days of that stock's prices (open, high, low, close) and volumes, along with the sentiment scores for those days. These features are normalized and fed into the CNN-LSTM model, which is trained (in 2021–2023 data) to predict the next day's closing price.

4.3. News sentiment analysis

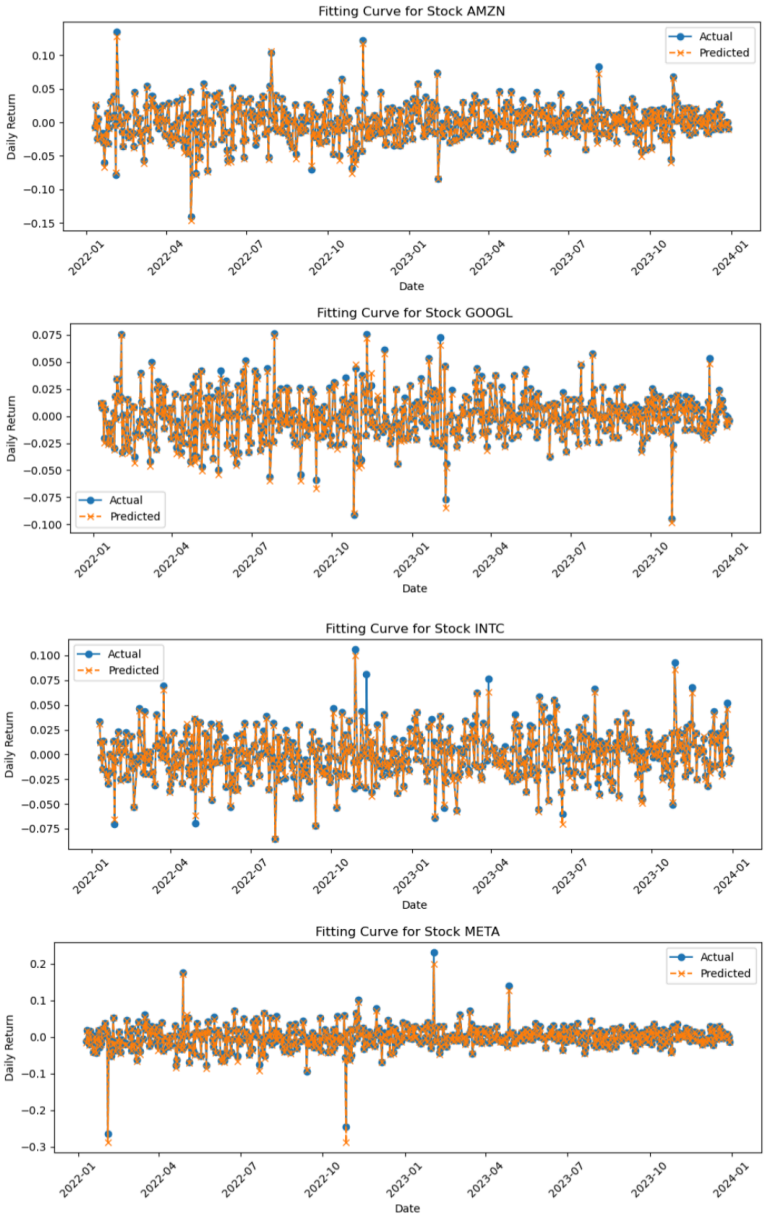
To quantify the information in news text, we leverage FinBERT, a state-of-the-art language model tailored for financial sentiment. We fine-tune FinBERT on our stock news dataset to classify each aggregated daily news batch as indicating a positive, negative, or neutral outlook for the next day's stock movement. The FinBERT-based classifier proved effective at capturing subtle sentiment cues in financial news. On the test set, it achieved about 63% accuracy in predicting next-day stock direction from news, significantly better than the ~56% accuracy of a generic BERT-based classifier. This result illustrates FinBERT's superior ability to capture financial context and nuances. Each day's output from FinBERT is used as a quantitative sentiment feature for the corresponding stock, integrating an investor sentiment signal into our predictive modeling.

4.4. Stock price prediction with CNN-LSTM

Using the enriched feature set (market data plus sentiment), we train a deep learning model to forecast each stock's next-day price, which in turn provides the expected return needed for portfolio optimization. Our model constitutes a CNN-LSTM network, which combines convolutional and recurrent layers with the aim of detecting short-term patterns as well as long-term temporal dependencies within the data. First, a one-dimensional convolutional layer scans the past 10 days of input features to extract local patterns in the stock's time series and news sentiment signals. Next, an LSTM layer processes these extracted features to learn the sequence's dynamics and outputs a prediction for the stock's closing price on day. The model is trained on the 2021–2024 training data for all ten stocks, using mean squared error as the loss function.



DOI: 10.54254/2754-1169/2025.LH24705



DOI: 10.54254/2754-1169/2025.LH24705

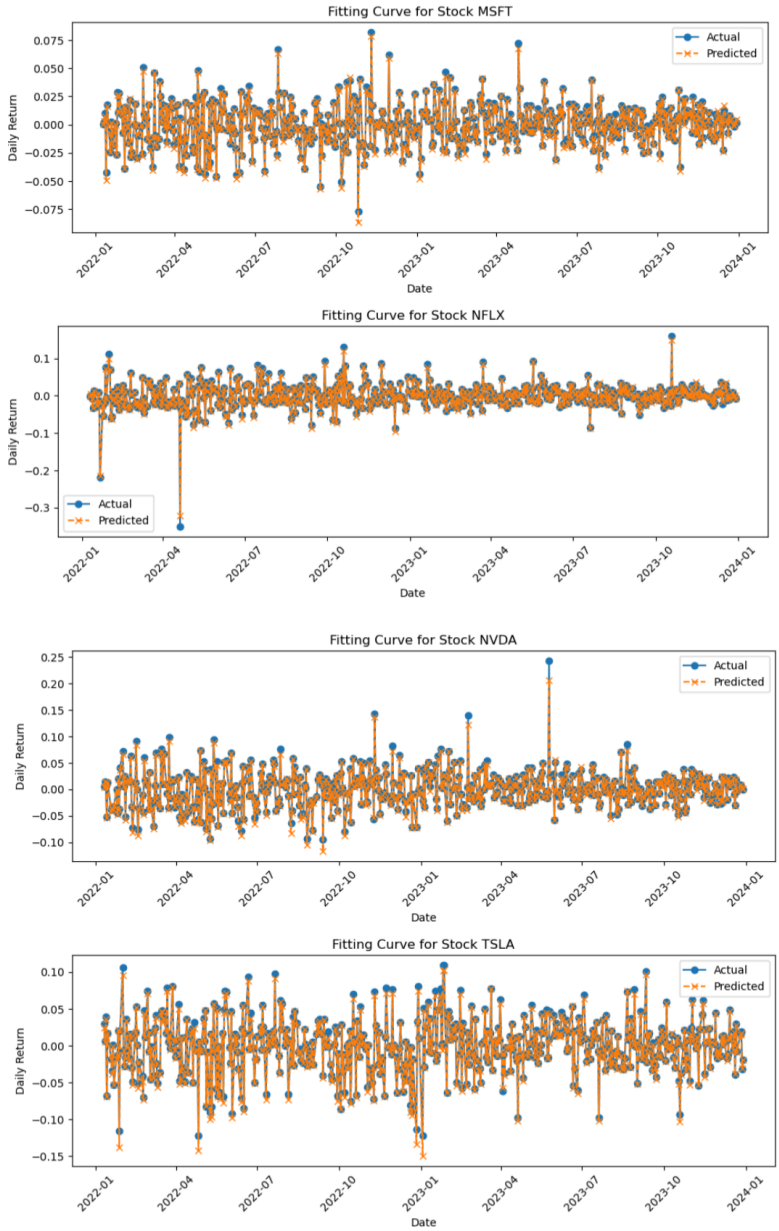


Figure 1: Predicted vs. actual daily returns in the test period

As illustrated in Figure 1, the ten predicted daily returns closely track the actual returns. The model captures most upward and downward swings, with only minor underestimation on a few abrupt changes. Overall, the prediction curve aligns tightly with the real outcomes, demonstrating a strong fit.

4.5. Portfolio construction and performance

Using these daily return forecasts, we construct a dynamic portfolio with the Black–Litterman model. At each trading day, the model blends our views (predicted returns for each stock) with a market equilibrium prior to produce an adjusted expected return for each asset. We then compute the optimal asset weights by solving a mean–variance allocation problem using these adjusted returns and the

DOI: 10.54254/2754-1169/2025.LH24705

current covariance of returns. This yields a portfolio allocation for the day, which is implemented and then re-optimized on the next day, and so on throughout the test period.

4.5.1. Benchmark model

The benchmark strategies include: (i) the S&P 500 index (a market baseline), (ii) an equal-weight portfolio of the same seven stocks, and (iii) a classical Markowitz optimized portfolio using historical returns of those stocks. Each strategy is rebalanced daily over the test period, and for fair comparison we ignore transaction costs in all cases.

4.5.2. Performance results

Our Black–Litterman portfolio achieves an annualized return of ~42.3%, significantly higher than the S&P 500 (~29.5%) and above the equal-weight (~32.8%) and Markowitz (~38.1%) benchmarks. It also achieves the highest Sharpe ratio (~1.30 vs ~1.1 for the S&P 500), indicating superior risk-adjusted performance. The strategy's maximum drawdown is around 4.5%, comparable to the S&P 500's ~3.5%, which means it avoids large losses and maintains stability on par with the benchmarks. This resilience is evident in the cumulative wealth curve.

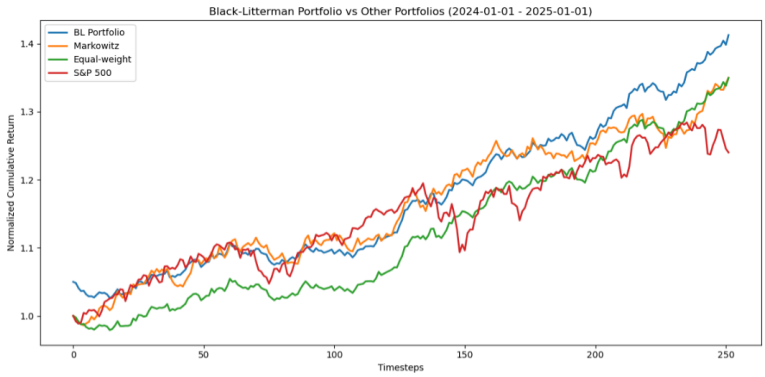


Figure 2: Cumulative return comparison between the proposed portfolio and market indices

Figure 2 shows that our portfolio's cumulative return (blue curve) stays above other portfolios benchmarks throughout most of the test period, finishing at a higher final value. Its trajectory is also relatively smooth, consistent with its low drawdown.

4.5.3. Evaluation and discussion

The empirical results confirm the effectiveness of incorporating news sentiment into the portfolio construction process. Our sentiment-informed strategy not only achieved superior returns but did so with controlled risk. However, the active rebalancing leads to high portfolio turnover (around 65% over the year, versus essentially 0% for the index strategy). Such frequent trading could incur significant transaction costs, which would need to be accounted for in practice.

In summary, integrating FinBERT-based sentiment analysis and CNN-LSTM forecasting into a Black–Litterman framework yields a portfolio that outperforms the benchmarks while maintaining controlled risk, highlighting the value of incorporating news-driven sentiment signals and deep learning models into portfolio management.

DOI: 10.54254/2754-1169/2025.LH24705

5. Conclusion

Initially, news sentiment constitutes an innovative variable for anticipating stock price fluctuations, as it reflects investors' emotions and prospective actions concerning a stock. In this investigation, news sentiment is employed as an innovative predictor to enhance the precision of stock price forecasting. This strategy provides a forward-looking viewpoint that complements conventional quantitative data. Our research integrates FinBERT for sentiment analysis with a CNN-LSTM model for price prediction. Following an empirical examination of financial data and news sourced from Thomson Reuters and CNBC, we discovered that our portfolio attains an annualized return of 42.3% alongside a Sharpe ratio of 1.30. These outcomes surpass both the S&P 500 benchmark and the Markowitz benchmark, which registers an annualized return of 38.1%. The Sharpe ratio indicates that the portfolio has generated 1.3 units of excess returns above the risk-free rate for each unit of risk assumed. These findings suggest that news sentiment positively impacts the prediction of stock price movements. Moreover, the Black–Litterman model-based portfolio, which incorporates news sentiment and CNN-LSTM analyses, can yield higher returns while maintaining controlled risk. Overall, these results confirm the utility of news sentiment in capital markets and offer investors and analysts an innovative new perspective on prediction and portfolio construction.

References

[1] Fabozzi, F. J., Markowitz, H. M., & Gupta, F. (2008). Portfolio selection. Handbook of finance, 2, 3-13.
[2] Black, F., & Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. Goldman Sachs Fixed Income Research, 115(1), 7-18.
[3] Kocuk, B., & Cornuéjols, G. (2020). Incorporating Black-Litterman views in portfolio construction when stock returns are a mixture of normals. Omega, 91, 102008.
[4] Stoilov, T., Stoilova, K., & Vladimirov, M. (2021). Application of modified Black-Litterman model for active portfolio management. Expert Systems with Applications, 186, 115719.
[5] Sahamkhadam, M., Stephan, A., & Östermark, R. (2022). Copula-based Black–Litterman portfolio optimization. European Journal of Operational Research, 297(3), 1055-1070.
[6] Barua, R., & Sharma, A. K. (2023). Using fear, greed and machine learning for optimizing global portfolios: A Black-Litterman approach. Finance Research Letters, 58, 104515.
[7] Ko, H., Son, B., & Lee, J. (2024). A novel integration of the Fama–French and Black–Litterman models to enhance portfolio management. Journal of International Financial Markets, Institutions and Money, 91, 101949.
[8] Hung, M. C., Hsia, P. H., Kuang, X. J., & Lin, S. K. (2024). Intelligent portfolio construction via news sentiment analysis. International Review of Economics & Finance, 89, 605-617.
[9] Zhang, C., Gao, B., Xu, X., & Qin, M. (2025). MFA RPC news sentiment and stock returns. Pacific-Basin Finance Journal, 92, 102779.
[10] Liu, Y., & Huang, Y. (2025). A Multimodal Deep Learning Framework for Constructing a Market Sentiment Index from Stock News. Big Data Research, 100535.
[11] Gong, Y., Peng, Y., Xu, L., Chen, K., & Shi, W. (2025). Shipping news sentiment as a predictor of iron ore freight rates: Hybrid evidence from lexicon-based analysis and threshold autoregression modelling. Transport Policy.
[12] Li, Y., Qiao, Y., & Lei, S. (2025). Ripple effect of ESG sentiment: How news stirs the waves in China's A-share market. International Review of Financial Analysis, 97, 103856.
[13] Leone, F., Marazzina, D., & Rosamilia, N. (2025). What's news with you: Price forecasting with global and ESG sentiment scores. Finance Research Open, 100013.