



Iterative Deep Learning Approach to Active Portfolio Management with Sentiment Factors

Javier Orlando Pantoja Robayo¹ · Julián Alberto Alemán Muñoz¹ · Diego F. Tellez-Falla²

Accepted: 9 August 2024
© The Author(s) 2024

Abstract

We suggest using deep learning networks to create expert opinions as part of an iterative active portfolio management process. These opinions would be based on posts from the X platform and the fundamentals of stocks listed in the S&P 500 index. Expert views are integral to active portfolio management, as proposed by Black–Litterman. The method we propose addresses the original subjectivity of the opinions by incorporating innovation and accuracy to generate views using analytical techniques. We utilize daily data from 2010 to 2022 for stocks from the S&P 500 and daily posts from Twitter API v2, collected under a research account license spanning the same period. We found that incorporating sentiment factors with machine learning techniques into the view generation process of the Black–Litterman model improves optimal portfolio allocation. Empirically, our results notably outperform the S&P 500 market when considering the annualized alpha.

Keywords Black–litterman · Active portfolio · Tweets · Sentiments · LSTM

JEL Classification G12 · G11

✉ Diego F. Tellez-Falla
dftellezf@eafit.edu.co

Javier Orlando Pantoja Robayo
jpantoja@eafit.edu.co

Julián Alberto Alemán Muñoz
jaalemanm@eafit.edu.co

¹ School of Finance, Economics and Government, Universidad EAFIT, Medellín, Colombia

² School of Finance, Economics and Government, Universidad EAFIT, Carrera 49 7 Sur 50, Medellín, Colombia

1 Introduction

Active and passive portfolio management complement each other to beat the market. In general, passive management is considered less uncertain than active management, held by historical results and time diversification. However, when considering business cycles, active management is more tractable from a forward-looking point of view making it a complement to obtaining the strategic aim in the long term (Lumholdt et al., 2018). The challenge is creating a model to understand these business cycles and incorporate the information in portfolio optimization.

Fisher Black and Robert Litterman (BL) propose a model for active portfolio management incorporating experts' views in the mean–variance optimization process (Black & Litterman, 1990). When managers know specific information about stock yields, the expected returns will be changed in the same direction according to the correlation between stocks⁴. Empirical evidence seems to show the power of BL to outperform other portfolio models and market indexes. Bessler et al. (2017) also show how BL outperforms mean–variance (MV), Bayes stain (BS), and equality weighed optimized (EW) portfolio.

The definition of experts' views in BL model is subjective and does not have any restrictions on how to construct them. This fact represents an opportunity to incorporate innovative and accurate technics to generate views either in analytical methods or text mining technics⁵. In this paper we propose an iterative active portfolio management process that uses deep learning networks to construct the views for a BL model using posts from X platform and fundamental variables in S&P 500. We use this information to train an LSTM network and compare the performance with the S&P 500 index and the classical BL model using Bloomberg analysts' recommendations.

Hochreiter and Schmidhuber (1997) proposed long-short-term memory (LSTM) network to solve the problem of vanishing gradients using specific memory cells and gates that incorporate good predictions for long-dependent patterns. Forecasting models that use LSTM are increasingly demanding because they allow us to model complex relationships that analytical models cannot unless we consider strong assumptions in many of these. Recent authors have used LSTM in financial applications. Liu (2019) shows how deep learning techniques exceed the GARCH to predict the volatility in S&P and APPL. In both cases, LSTM is the preferred choice with high computation power. Yıldırım et al. (2021) used an LSTM neural network to predict the direction of the Forex market.

Another benefit of neural networks in forecasting is the use of innovative sources like social media. The theories about overreacting and underreacting mention that investors overreact to consecutive positive patterns of benefits in the long term; the sentimental analysis proof of this behavior shows patterns that the arbitrageurs could use to beat the market. (Chari et al., 2017); Liu et al. (2021) apply this method to predict the close prices of the SSE 50 improving the accuracy in the models based only in historical data and social media platforms.

To test our setting, we collect data from two main sources. First, we use Bloomberg to collect daily information from S&P500 stocks spanning the period

2010–2022. Particularly, we are interested in stock price, the price to book, price to equity and price to sales ratio as for return on assets, total revenue and ISM Manufacturing index. Second, we use Twitter API v2 to download daily posts (tweets) to capture sentiment in the market. Our method allows us to construct portfolios and report risk metrics as Jensen alpha (Jensen, 1968), Treynor (Treynor, 1965) and Sharpe ratio (Sharpe, 1966).

We contribute to the literature showing empirical evidence of the use of sentiment information via X posts and fundamentals in asset allocations. In our empirical setup the views generated with LSTM allow us to build portfolios that overcome in return standard BL portfolios. In particular, the portfolios created report better Jensen alpha up to thirty one percent annualized.

This study differs from previous ones regarding LSTM and portfolio optimization (e.g., Colasanto et al., 2022; Rezaei et al., 2021) by selecting other fundamental variables, sentiment measures, the benchmark to measure the performance and the iterative proposal to active portfolio management. Furthermore, we extend the work of Sul et al. (2017) on sentiment analysis using transformers to extract sentiment and include fundamental variables.

The rest of the paper is organized as follows. The second section briefly shows the literature review regarding market efficiency and factors, LSTM and forecasting, sentimental analysis, and the BL model. The data, the general scheme, and the details of the construction of the active management process are shown in the third section. The fourth section contains the results, and finally, the fifth section offers the conclusions.

2 Literature Review

In 1952, Markowitz published the seminal article "Portfolio Selection," widely regarded as the foundation of modern portfolio theory. Portfolio models serve as tools designed to assist portfolio managers and investors in determining the allocation of assets within a fund or portfolio. Markowitz's ideas have had a profound impact on portfolio theory and have, in theory, stood the test of time. However, in practical portfolio management, the adoption of Markowitz's model has yet to match its academic influence. Many fund and portfolio managers find the portfolio compositions generated by the Markowitz model counterintuitive (Black & Litterman, 1992; Michaud, 1989). The challenges encountered in applying the Markowitz model prompted Fisher Black and Robert Litterman in the early 1990s to develop a new model. This model, commonly known as the Black-Litterman model (hereafter referred to as the B-L model), builds upon Markowitz's framework to address some of its practical limitations.

Unlike the optimization process in the Markowitz model, which begins from the null portfolio, the B-L model's optimization starts from what Black and Litterman term the equilibrium portfolio—often interpreted as the benchmark weights of assets in the portfolio. "Bets" or deviations from the equilibrium portfolio are then made on assets for which the investor has formed views. The manager assigns Each view a confidence level, indicating the degree of certainty

associated with that particular view. The confidence level influences the extent to which the weight of a specific asset in the B-L portfolio deviates from the weights of the equilibrium portfolio. Importantly, these opinions are considered only for particular assets, while others are left unaffected. The magnitude of these investments, relative to the equilibrium portfolio weights, is determined by both the user-specified confidence levels and a parameter known as the weight-on-views, which dictates the influence of the investor views compared to the market equilibrium.

A management portfolio is well-known for its use of two techniques. In one way, passive or strategic management argues that the returns are positive in the long term (this is called the buy-and-hold strategy). On the other hand, active management or management searches for buy-and-sell plans based on anomalies present in the short term. Why do anomalies exist?

The efficient market hypothesis (EM) proposed by Paul Samuelson and Eugene Fama in the 1960s and condensed in Fama (1970) states that prices intrinsically incorporate all the information in the market, including the exogenous information related to the firm. This potent form of EM eliminates any effort to beat the market. The semi-strong form of EM says that although prices reflect all public information about the firm, investors can take privileged information into account to beat the market. Lastly, the weak form of efficiency concerns the price reflecting all historical information; in this form, historical information could be considered public information. The underlying idea behind an efficient market is that prices follow a random walk; it means the prices from yesterday are not related to today's prices.

However, the EM theory is contradicted by several observations. Shiller (1987) explains the volatility of the S&P from 1871 to 1986, starting with an efficient market assumption and using three factors: change of dividends, change of real interest rate, and inter-period marginal rate substitution. Other examples include intertemporal capital asset pricing proposed by Merton and the arbitrage pricing theory. Fama and French (1993) proposed the three factors model that consists of the market risk premium (RMRF), market value factor, that is, small market capitalization minus significant market capitalization (SMB), and books-to-market ratio corresponding to the firms with high book to market ratio minus low book to market ratio (HML). Based on the three-factor model, Carhart (1997) proposes a new factor to extend the model using the momentum factor. With new evidence and inspired by her previous three-factor model, Fama and French (2015) propose a five-factor model covering most factors not included in the last work, adding robust and weak profitability (RMW) and the difference between the returns on diversified portfolios of the stocks of low and high investment firms (CMA). All these factors were evidence of anomalies in the EM. However, Pesaran (2010) mentioned that the findings may need to be more conclusive due to data mining. Later, Lo (2004) proposed the adaptive market hypothesis and incorporated behavior to explain how the market allows certain anomalies and dynamically adjusts inspired by a biological framework.

2.1 Sentiment and Behavioral Finance

Most anomalies are frequently found by practitioners looking for a new explanation for this evidence about market inefficiency, most of them based on the irrational behavior of some agents at some point, supported by psychology and sociology. Sentiment analysis connects with this theory because the agent makes decisions based on state and emotions; these states could be the mood a message represents in a social network.

The noise theory in the market tries to explain how mispricing appears. Splitting the market between arbitragers that adjust the market to guarantee that the asset has the fundamental value and the noise traders that react to expectation or news; the mispricing arises when the reaction in noise traders is massively in the same direction and shifts the demand (Shleifer & Summers, 1990).

Sul et al. (2017) analyzed two point five million X posts and found a relation between information reported in the tweets and the returns of S&P 500 firms. The authors propose a strategy that obtains, in theory, eleven to fifteen percent excess of returns annually. Piñeiro-Chousa et al. (2016) also show a connection between fundamental variables, like price over earnings, and market capitalization. The evidence highlights the importance of including sentiment analysis jointly with the most recent fundamental variables to gain more accuracy in returns forecasting, especially in extreme market situations. (Li et al., 2017).

2.2 Active Management Portfolios and Black–Litterman Model

The Black–Litterman model was initially introduced by Fischer Black and Robert Litterman of Goldman Sachs in an internal Goldman Sachs Fixed Income Research Note, as referenced in Black and Litterman (1990).

This research note was subsequently expanded into a paper officially published in the *Journal of Fixed Income* in 1991, as detailed in Black and Litterman (1991). While this paper offers a comprehensive overview of the model's features, it only presents some of the formulas used in the model. Another internal Goldman Sachs Fixed Income Research Note was issued in the same year, documented in Black and Litterman (1991). This paper was later expanded and formally published in the *Financial Analysts Journal* (FAJ) by Black and Litterman (1992), which FAJ republished in the mid-1990s. Copies of the FAJ article are readily accessible online and provide a rationale for the methodology and some insights into its derivation. However, not all formulas or a complete derivation are included. Furthermore, the article features a complex work example based on global equilibrium, with additional details on the methods required to address this issue provided by Litterman (2003). Unfortunately, reproducing their results can be challenging due to the complexity of integrating these two problems.

2.2.1 Deep Learning and LSTM in Asset Pricing

At this point, the factors or anomalies are considered linearly related. However, there is evidence that at least some of them present no linearity; for instance, the size factor in the German market is analyzed from 2000 to 2008 by Amel-Zadeh

(2011), finding that the size effect is conditional on the previous performance of the firm. Another aspect of the factors is the dynamic change through time, as shown by Daniel and Moskowitz (2016), showing how the winners' decile underperformed the loser portfolio during March and May in 2009.

Deep learning is a form of artificial intelligence consisting of multiple processing layers that can learn and map relational and non-relational data to output linear and nonlinear results. Moritz and Zimmermann (2016) argue that traditional linear frameworks only utilize some relevant information in the data, while machine learning approaches are more powerful. Li and Ma (2010) define machine learning as a computational model based on biological neural networks composed of basic units called neurons that simulate the electrical communication in the human brain. A perceptron takes input values and weighs them, while an activation function regulates the output within a specific range, often from 0 to 1 (Graupe, 2013). When there is feedback between neurons, it forms a recurrent neural network (RNN). To recognize patterns over time, Hochreiter and Schmidhuber (1997) propose Long Short-Term Memory (LSTM) networks, which address long-term dependencies that can describe seasonal patterns and other time-based relationships.

LSTM is the most frequently used method to predict stocks and forex markets. Hu et al. (2021) found thirty-eight studies related to LSTM eighty-eight between 2015 and 2021. They compared them to other deep learning methods like CNN, RNN, and others and obtained from the digital bibliography and library Project (DBLP).

Akita et al., (2016, June) use LSTM to predict the stock prices in the ten stocks in the NIKKEI 225 market and use the news as fundamental data to train the model in joint with historical returns to measure the performance using experiments on the textual data and the LSTM network and conclude: "LSTM was capable of capturing time-series influence of input data than other models and considering the companies in the same industry was effective for stock price prediction". Jiang et al. (2019) use LSTM to predict the stock prices of the ten stocks in the NIKKEI 225 market and use the news as fundamental data to train the model in joint with historical returns. They measure the performance using experiments on the textual data and the LSTM network and conclude: "LSTM was capable of capturing time-series influence of input data than other models and considering the companies in the same industry was effective for stock price prediction". Jiang et al. (2019) used an LSTM-MPL hybrid framework to forecast the NYSE, AMEX, NASDAQ, and TAQ markets between 1993 to 2017. He used two models in three tests. They were first based on 80 days before predicting one day using a 2-layer LSTM. The Second predicts the following day using the same NN before but selecting 17 industries based on Fama and French recommendations. The last uses an LSTM-MPL hybrid neural network; the latter test outperforms the other ones in turnover.

Minami et al. (2018) predicted the stock price of Tsugami Corporation, a manufacturing company, using historical stocks, press releases, and business and financial information; they observed that only using the press release obtained the minimum prediction error.

2.3 Related Works

The earliest effort may be traced to the work of Palomba (2008), who constructed a multivariate GARCH model to estimate the views in a BL model combining information from volatility and personal information. Albarrak et al. (2020) combine technical and sentiment analysis to generate views in the BL model. They use sentiment indicators from the American Association of Individual Investors instead of social media sentiment indicators, and they use a random forest model to generate the views.

Recent works suggest the use of hybrid methods of machine learning to improve Forecasting. Rezaei et al. (2021) implement a deep hybrid learning (CEEMD-CNN-LSTM) to generate analyst recommendations and incorporate them into the BL model. Our research was inspired initially by this study. However, it differed from theirs in adding the X posts and fundamentals variables to improve the accuracy of the forecasting recommendations, the use of sentimental data, and the process of selecting assets. Barua and Sharma (2022) constructed another hybrid (CNN-BiLSTM) with eighteen technical indicators and historical returns from ten sectors in the MSCI Asia Pacific index to generate recommendations that were used in a time-variant BL model that yields 23.3 of expected annual excess of return for 24 months, from April 2020 to April 2022.

Before concluding this section, we would like to briefly mention that Colasanto et al. (2022) used a pre-trained FinBert to extract news sentiments, improve stocks' Forecasting, and incorporate them in a BL model. They use the trend in the sentiment news joint with Monte Carlo simulation to generate the possible path in the Forecasting; they improve the Sharpe ratio from 1.14 to 1.07 when comparing the BL using views from simulation vs. maintaining the same quantity of assets.

3 Iterative Deep Learning Process

In what follows, we introduce the iterative proposal used in this work to build portfolios using deep learning and compare these with the ones obtained in a more traditional form. We run a five-step process starting with the selection of assets. We apply the combinatory formula to get all combinations of assets from the S&P 500 composition in June 2022, filtering by the assets with complete data (historical returns, X posts, and fundamental values from January 2010 to June 2022). These groups of assets could be discrete, in ten, fifteen, or twenty securities, for instance, and each group generates C subgroups described by (1).

$$C = \frac{n!}{r!(n-r)!} = \binom{n}{r} \quad (1)$$

where n is the number of assets in the S&P500 to choose from, and we choose r of them, with no repetition, and order does not matter. In the empirical case, when r is

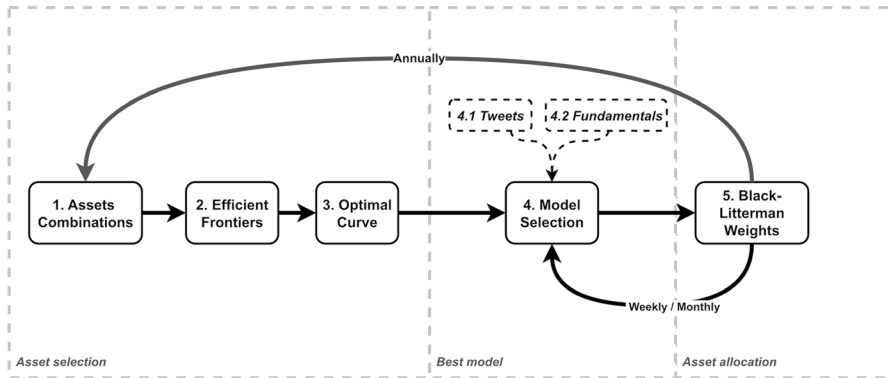


Fig. 1 Deep learning is a five-step iterative process for an active management portfolio. This process describes the five steps to an active management portfolio; the arrows indicate the direction of the process; in stage five (Black-Litterman Weights), the model proposes to iterate in the model selection or new asset combination instead. The big dotted boxes are macro processes

10, C is 2.51×10^{20} approximately; this is a vast number, and it proposes a sample of them to look for an approximation solution.

Once we have the assets, our second step consists of finding the efficient frontier.¹ By each subgroup. In finance literature, the efficient frontier represents the combination of assets based on the expected return and risk. Portfolio theory states that investors either maximize portfolio return, given a level of risk, or minimize risk, given an expected return (Fig. 1).

The third step in the iterative learning process is measuring the area under the curve for all portfolios created in the previous steps. The area under the curve summarizes all chances for the combinations of returns and risk. Thus, a more extensive area is preferred over a small area. Figure 2 illustrates this idea. In that case, portfolio A is preferred over B, and B over C.

Considering more than the return-risk optimization by the most significant area for the efficient frontier is needed to make a better-informed decision for the investors; they need to add diversification. In that sense, each curve of the efficient frontier has a matrix of correlation assets; then, the sum of all items in this matrix is a scalar that summarizes the correlation grade of all possibilities.

This work proposes the ratio between the normalized area under the curve of efficient portfolio curves and the normalized summarization of the matrix's correlation (φ) for each curve as a metric to select the best asset selection.

As mentioned by portfolio theory, each portfolio is associated with an optimal Sharpe ratio. To avoid excess risk, this work proposes bounding the efficient frontier area until the 75th percentile of optimal Sharpe ratios.

After obtaining the asset selection, we use past returns, fundamentals, and the extracted summary of the sentiments to get the views generator model in the fourth

¹ This efficient frontier is created using last year's historical data.

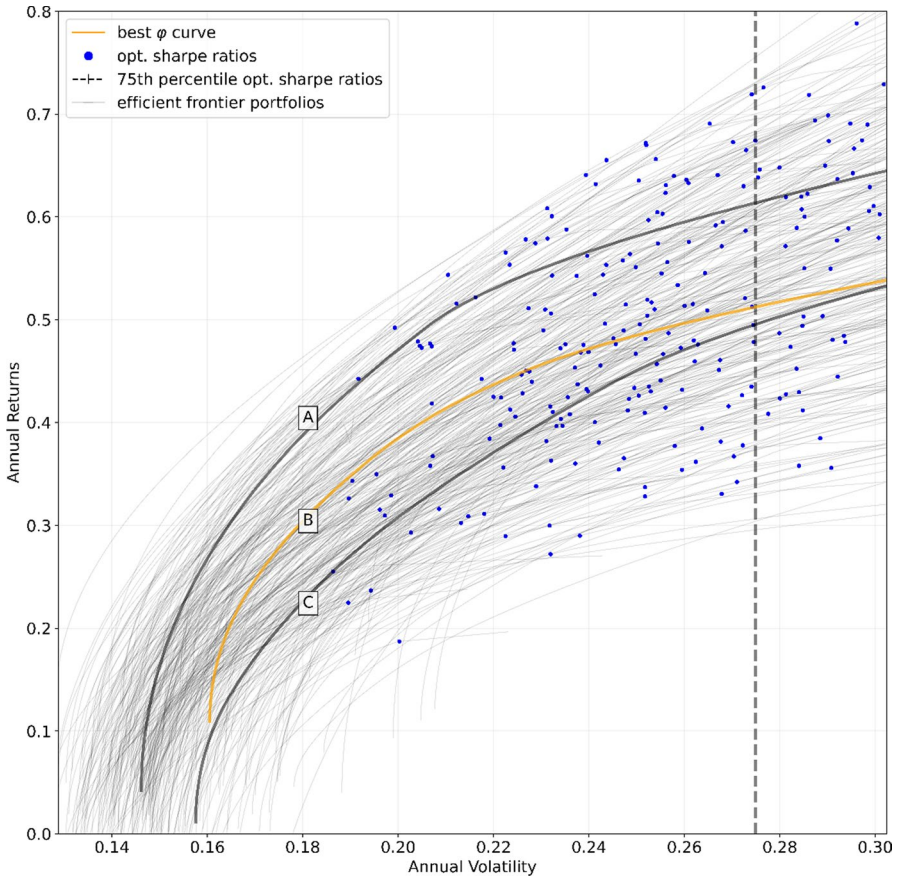


Fig. 2 The best efficient frontier curve into asset selection

step. Data is split into two groups, in sample and out of sample; the out sample is predicted in t steps using the best model obtained in a rolling forecasting origin, and the process is iterative for the next $(t+1)$ period.² The best model with low validation loss was selected using the Mean Absolute Percentage Error (MAPE) metric.

In our final step, we implement the BL model. For this purpose, the market cap is the joined market cap of the sum of the asset's selection; we calculate the excess return with the ninety-day treasury bills. This process gets the weights through the BL process but uses the view from the predicting model explained before. Then, we compare the BL results using the views of the recommendations.³ Bloomberg (BL-Bloomberg) versus the recommendations obtained from

² This process implies that the models in each period may be different.

³ We use the best_target_price variable in the historical table from Bloomberg.

the best model gathered in the five-step and compared their Jensen alpha, Treynor and Sharpe ratio. The general iterative process is summarized in Fig. 1.

4 Data and methods

For our empirical analysis, we use two sources. First, we use Bloomberg to collect daily data of S&P 500 stocks spanning the period 2010–2022. We are mainly interested in stock prices, price-to-book, price-to-equity and price-to-sales ratio, return on assets, revenue and ISM Manufacturing index. Second, we download daily X posts from Twitter API v2 through a research account license spanning the same period.

The X posts are filtered by the cashtag (\$) symbol that contains investment information. We extract the sentimental category of each X posts for each stock using a pre-trained network called FinBert (Araci, 2019); for each X post is obtained a class between negative, neutral, or positive; these categories are summarized by stock and by the period of analysis; finally, this summary is the input to the LSTM model.

To capture the real impact of the X post in the market, we consider all X posts in the slot time before the close trading hour on the day of trading and after the close hour in the previous market day; for the sake of simplicity, we only use the new york stock exchange (NYSE) calendar to all S&P500 stocks analyzed to consider the open and close hours in trading days. This analysis is based on the techniques proposed by Sul et al. (2017), except that this work does not exclude repeated X posts for each stock.

5 Deep Learning Forecasting

The first thing to do when using an LSTM network is to split the data into chunks to predict periods; the network processes each chunk; we try several sizes of the fragment, for instance, predicting one week with a chunk of 1 week before, predicting one week with four weeks before and one week with eight weeks earlier. The Forecasting is performed for each firm simultaneously in a multivariate multi-parallel LSTM series (Brownlee, 2018).

Next, the network is constructed with a few neurons and increasing layers, and neurons by layer, to select the optimal architecture. Each experiment is tested using a validation method that consists of separate data into train and test, with 80% to train and 20% to test. According to the results, we select the highest accuracy in terms of mean absolute percentage error (MAPE) in the validation data. To avoid overfitting in the LSTM, we apply three well-known techniques (Kamara et al., 2020): dropout layers, testing multiple configurations of neurons and hidden layers, and stopping based on validation loss.

6 Black–Litterman Portfolio

Fisher Black and Robert Litterman proposed the BL model in 1991 to Goldman Sachs; it starts from the equilibrium returns obtained from the current portfolio market.

$$\Pi_{eq} = \delta \Sigma w_{mkt} \tag{2}$$

where Π_{eq} stands for equilibrium returns, δ is the risk aversion, Σ is the covariance matrix between assets, and w_{hat} is the weight of each asset in the portfolio. After finding the more precise forecasting returns, we denote vector P containing them. The next step is using them in absolute shape, denoted by the vector P ; their relationship is named Q .

$$QxP = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} \tag{3}$$

The covariance matrix is obtained using equation (4).

$$\Omega = \text{diag}(P(\tau\Sigma)P^T) \tag{4}$$

Then, we have all the necessary to calculate expected returns and the covariance matrix using the BL formulas that Idzorek (2007) detailed.

$$\mu_{BL} = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P] [(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}Q] \tag{5}$$

$$\Sigma_{BL} = \Sigma + ((\tau\Sigma)^{-1} + P'\Omega^{-1}P)^{-1} \tag{6}$$

Where τ is a scalar factor to the covariance matrix, we assume this value is near zero and inversely proportional to the observations mentioned in Idzorek (2007). With these posterior results, we obtain the optimal weights, which in the absence of restrictions, have a closed form, in other cases, are commonly used numerical solutions.

$$U = w'\mu_{BL} - \left(\frac{\delta}{2}\right)w'\Sigma_{BL}w \tag{7}$$

$$\mu_{BL} - \Sigma_{BL}w = \frac{dU}{dW} = 0 \tag{8}$$

$$w^* = (\delta\Sigma_{BL})^{-1} \mu_{BL} \tag{9}$$

7 Portfolio Performance and Comparison

The portfolios obtained by BL-X are compared using three risk metrics. First, we use the Sharpe ratio, also known as the reward-to-variability ratio (Sharpe, 1966). The basic idea is to measure the amount of excess return (return of portfolio minus risk-free rate) per unit of risk (standard deviation of the portfolio excess return). We also use the Treynor ratio, the reward-to-volatility ratio (Treynor, 1965). Unlike the Sharpe ratio, Treynor uses the denominator systematic risk (beta) to measure the sensitivity of portfolio return to changes in return for the overall market. Finally, we use Jensen alpha (Jensen, 1968). This is a risk-adjusted performance measure that shows the average return on a given portfolio is above or below that predicted by capital asset pricing model portfolio performance metrics are described in equations (10), (11), and (12), respectively.

$$\text{Sharpe} = \frac{\overline{R}_p - R_f}{\sigma_p} \quad (10)$$

$$\text{Treynor} = \frac{\overline{R}_p - R_f}{\beta_p} \quad (11)$$

$$\text{Jensen's alpha} = \overline{R}_p - \left[R_f + (\overline{R}_m - R_f) \beta_p \right] \quad (12)$$

8 Analysis and Results

According to the iterative deep learning process, this work presents the results grouped into the three extensive macro processes; the first one corresponds to asset selection, the model's selection is shown in the second one, and the third one will present the BL process and results of the comparative benchmark.

8.1 Asset Selection

To create optimal portfolios, we use the following procedure: First, we collect weekly historical returns to assets from SP500 from 2010 to 2022. We generate a hundred random combinatory groups by ten, fifteen, and twenty assets without replacement.

For each combinatory (300 subgroups), we estimate the efficient frontier and Sharpe Ratio. Sharpe Ratios are separated in four groups according to risk level, obtaining four quartiles. We drop Sharp Ratios located in the last quartile and

Table 1 Descriptive stats for stocks in the asset selection step

	PARA	ADP	HPQ	CMCSA	FMC	ORLY	CLX	MOH	WAT	NOC
mean	0.00263	- 0.00004	- 0.00026	0.00169	0.00020	- 0.00013	0.00121	- 0.00016	0.00062	- 0.00074
std	0.03030	0.01540	0.02437	0.01603	0.01900	0.01856	0.02071	0.01938	0.01791	0.01679
min	- 0.13305	- 0.03753	- 0.12856	- 0.04699	- 0.11206	- 0.04780	- 0.05688	- 0.06520	- 0.07509	- 0.07344
max	0.21670	0.09835	0.07517	0.07879	0.08263	0.14441	0.16922	0.05424	0.04922	0.08228

Mean, standard deviation, minimum and maximum for the nine assets selected in the portfolio from S&P500 in January 2018

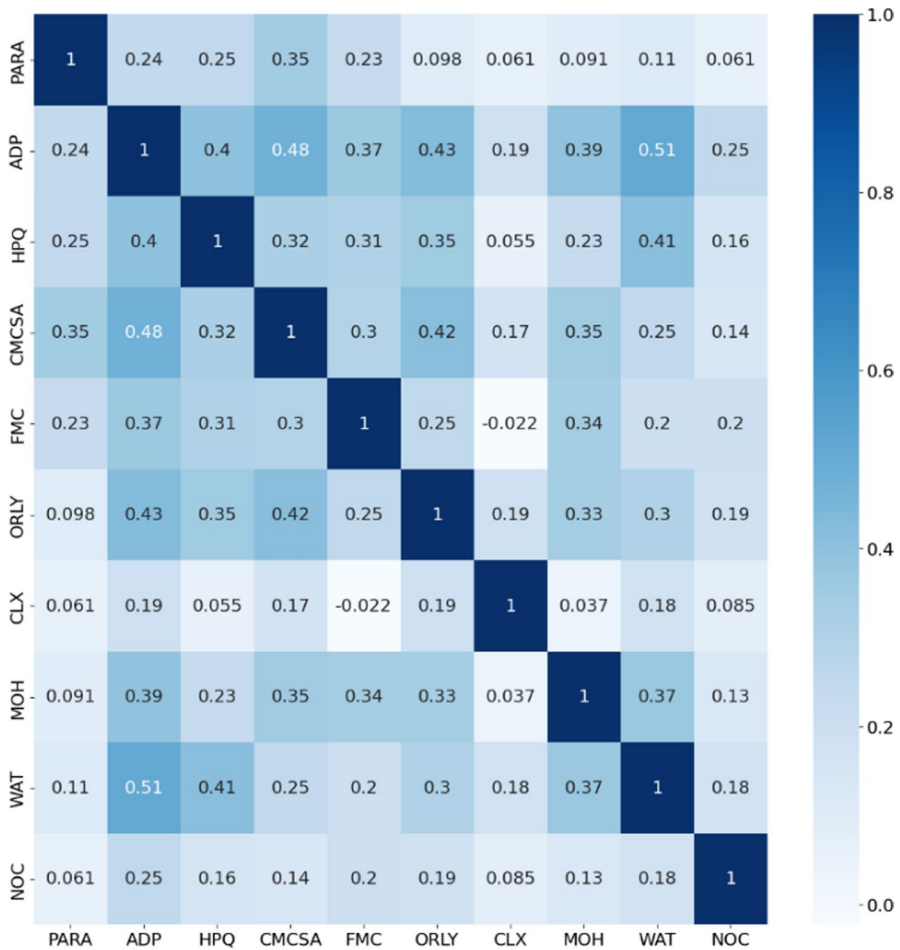


Fig. 3 Correlation matrix to the best asset selection for ten assets from 431 in S&P500. Matrix correlation between assets selecting for the portfolio selecting assets step, the cells with the most intense color indicate a high positive correlation and more clear cells represent the most intense negative correlation; the securities are in the lowest and left side of the matrix, the diagonal of this matrix indicate the auto-correlation that is expected in one

estimate a new index composed of the quotient between Sharpe Ratio and correlation level to adjust for diversification.

Table 1 and Fig. 3 report some descriptive statistics and correlation coefficients for the asset in the optimal curve, respectively. Table 1 shows that the returns are centered around zero, with a similar standard deviation. The correlation matrix in Fig. 3 lets us know that the optimal curve allows us to obtain a well-uncorrelated portfolio; all coefficients are under 0.6, This asset selection generally opens the possibilities (optimal curve) of a good ratio return over risk by each correlation point.

Table 2 Validation results for each model in each week

Week	Neurons	Dropvalues	Periods	MSE	MAE	MAPE
1	[80]	[0.05]	300	0.010195	0.082359	13.56012
2	[80]	[0.05]	300	0.010704	0.08044	13.28088
3	[80]	[0.05]	300	0.014502	0.090325	15.08989
4	[80]	[0.05]	300	0.014743	0.091367	15.16484
5	[100]	[0.05]	270	0.013477	0.092046	14.35724
6	[80]	[0.05]	300	0.010533	0.079591	12.93514
7	[80]	[0.05]	300	0.013085	0.089486	14.26327
8	[80]	[0.05]	300	0.015514	0.101317	16.35983
9	[80]	[0.05]	300	0.012079	0.086484	13.38963
10	[80]	[0.05]	300	0.009004	0.07573	11.75328
11	[80]	[0.05]	300	0.013866	0.096075	15.06203
12	[80]	[0.05]	300	0.012765	0.092064	13.49025
13	[80]	[0.05]	300	0.013291	0.088998	13.5488
14	[80]	[0.05]	300	0.013448	0.089754	13.69475
15	[80]	[0.05]	300	0.007598	0.070045	10.95446

The columns “neuron layers” and dropouts layers describe each layer’s number of neurons and the dropout value, respectively. The column period refers to the periods before that are used to forecast the next period. MSE, MAE, and MAPE metrics are mean square error, mean absolute error, and mean absolute percentage error, respectively

8.2 Model Selection

Sentiment Analysis. In practice, only seven assets had information from all periods from 2010; therefore, we reduced the selection from ten to seven.

Using the pre-trained Bert neural network, we obtain the sentiment of each X post in three categories: positive, negative, and neutral, with 29.408, 511.192, and 139.592, respectively, for all nine assets from 2010 to 2022.

Models. Three kinds of hyper-parameters (neurons, historical periods, and dropout layers) and ten models were tested in each period (15 periods); to obtain a statistical result, we tested each model ten times, one for each period. The best model in each period was selected with the lowest mean absolute percentage error (MAPE) on average; Table 2 shows us the best model for each period and the parameters. With these criteria, the best model for each period corresponds to 1 layer, and the historical period preferred was 300; thus, historical data has a high impact over the following periods (Fig. 4).

Forecasting. The models were trained and validated with historical data from January 2010 to December 2017. The next fifteen weeks were forecasted using the model selection process, which slid the training and validation data when a new period was predicted. Figure 5 shows the Forecasting obtained with the BL with X posts, Bloomberg analyst recommendations, and the actual price. We observe that some stocks have low performance but are compensated with high performance in other investments, as seen in Table 3

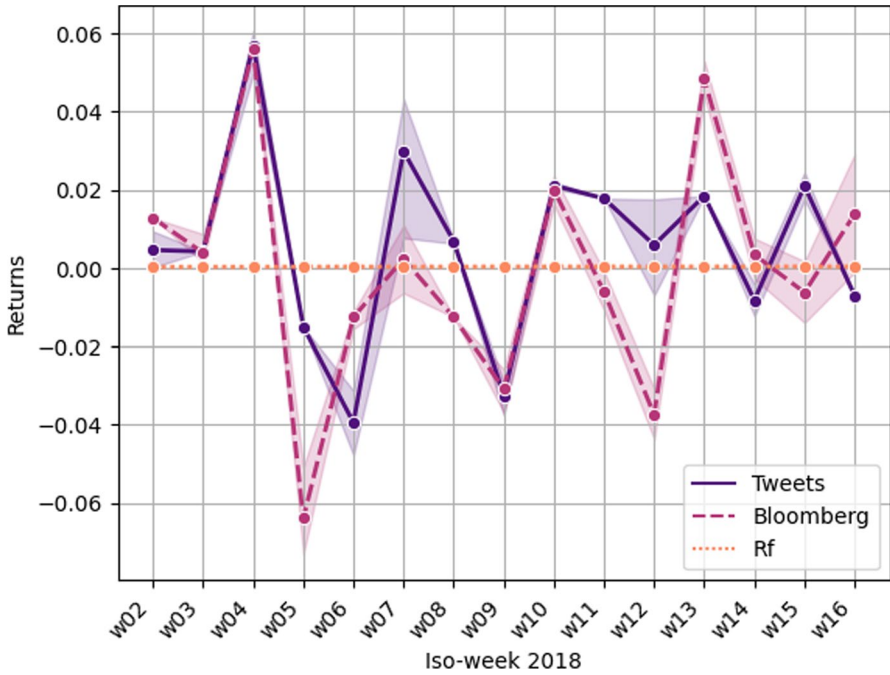


Fig. 4 Returns of BL with Bloomberg views returns for BL with tweets views, and risk-free rate for each period, each point is measured ten times, and the shaped area shows each measure’s deviation. The evolution of the portfolio returns generated by the BL-Bloomberg based on the Bloomberg best price and portfolio return generated by the BL-tweets based on the tweets for each asset for the first 15 weeks of 2018 based on the data from January 2010

Table 3 Metrics out of sample for predictions to nine stocks in S&P500

Stock	BI-Bloomberg		BI-Tweets		Relative Tweets-Bloomberg	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
CMCSA	10.31047	0.284893	10.585788	0.277775	0.275321	- 0.007118
CLX	8.816067	0.069692	45.117514	0.342072	36.301447	0.27238
NOC	12.44273	0.036975	180.14731	0.526473	167.704573	0.489498
ORLY	35.9466	0.149104	150.86088	0.614872	114.914279	0.465768
FMC	14.22902	0.201075	32.565362	0.444485	18.336342	0.24341
PARA	16.1204	0.304621	10.19523	0.188531	- 5.92517	- 0.11609
MOH	5.948333	0.074655	28.549958	0.357275	22.601624	0.28262

Forecasting metrics for nine stocks in S&P500 in out sample for multiple models in each period taken from January 2018 to April 2018, the relative MAE and MAPE is the difference between the MAE and MAPE for BL-Bloomberg minus BL-Tweets, respectively

To analyze the results of Table 4 presenting the values of the Treynor, Sharpe, Alpha, and Beta indices for two different portfolios (BI-Bloomberg and BI-Tweets), we can observe the following:

Table 4 Portfolio performance metrics for BL-Bloomberg, BL-tweets. In all cases, the empirical evidence seems to suggest a better performance of the portfolios built with

	Treynor	Sharpe	Alpha	Beta
BL-Bloomberg	- 0.099116	- 0.227597	- 0.041839	1.024927
BL-Tweets	1.66701	1.562485	0.319119	0.343963

Treynor, Sharpe, and Jensen’s alpha annualized for portfolio comparison in BL with Bloomberg target price views and BL model that constructs the views from an LSTM deep learning model with tweets

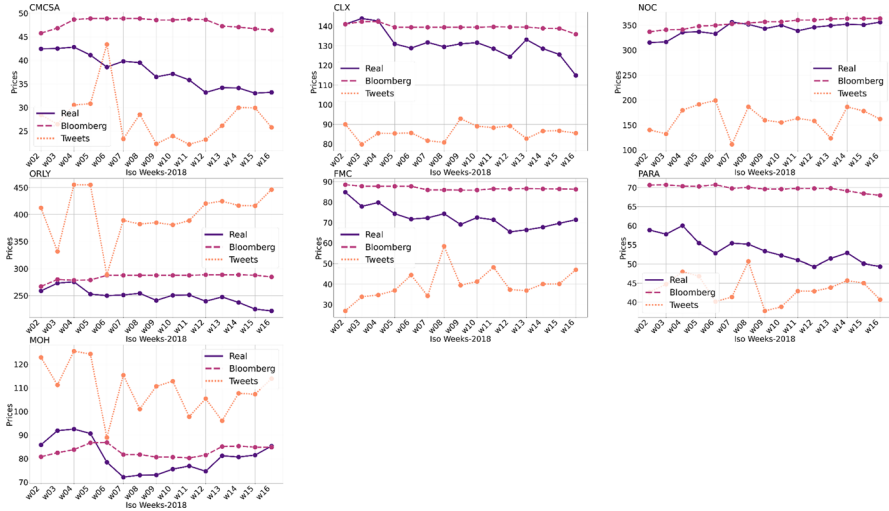


Fig. 5 Forecasting of BL model using bloomberg target prices like views and BL model using tweets in an LSTM network to generate views against the actual price for each stock. Each subgraph shows the time series in Iso-week format of the prices predicted by the actual value, the best price target from Bloomberg, and the value predicted by the model using tweets –

8.3 BI-Bloomberg

- **Treynor Ratio: -0.099116**
 - This negative ratio indicates that the portfolio has not been effective in generating risk-adjusted returns compared to the market. A negative Treynor suggests that the portfolio has performed below the risk-free rate, given its beta.
- **Sharpe Ratio: -0.227597**
 - Like the Treynor Ratio, a negative Sharpe Ratio indicates that the portfolio has not provided returns above the risk-free rate after adjusting for volatility. This is a sign of underperformance.
- **Alpha: -0.041839**

- A negative Alpha suggests that the portfolio has underperformed expectations given its level of risk (beta). This indicates that the portfolio has not added value compared to its benchmark.
- **Beta: 1.024927**
 - A beta of approximately 1 indicates that the portfolio is sensitive to the market, similar to that of the market itself. This means the portfolio is expected to move in sync with the market, though slightly more volatile.

8.4 BI-Tweets

- **Treynor Ratio: 1.66701**
 - A positive and high Treynor Ratio suggests that the portfolio has effectively generated risk-adjusted returns compared to the market. This indicates good risk management and superior performance relative to market risk.
- **Sharpe Ratio: 1.562485**
 - A positive and high Sharpe Ratio indicates that the portfolio has generated returns above the risk-free rate after adjusting for volatility. This suggests strong performance and sound risk management.
- **Alpha: 0.319119**
 - A positive Alpha indicates that the portfolio has performed above expectations, given its level of risk (beta). This suggests that the portfolio has added value compared to its benchmark.
- **Beta: 0.343963**
 - A beta less than 1 indicates that the portfolio is less volatile than the market. This suggests that the portfolio is less sensitive to market movements and has a more conservative risk profile.

8.5 Comparison Between BI-Bloomberg and BI-Tweets

- **BI-Bloomberg** shows inferior performance across all indices. The negative Treynor and Sharpe Ratios and a negative Alpha indicate that this portfolio has been ineffective in generating risk-adjusted returns and has not added value compared to its benchmark. The beta close to 1 suggests that this portfolio has a market sensitivity similar to the market but with unsatisfactory performance.
- **BI-Tweets** shows superior performance across all indices. The positive Treynor and Sharpe Ratios and a positive Alpha indicate that this portfolio has been effective in generating risk-adjusted returns and has added value compared to its benchmark. The beta less than 1 suggests that this portfolio has a more conservative risk profile and is less volatile than the market.

In summary, the *BL-Tweets* portfolio performs significantly better than the *BL-Bloomberg* portfolio regarding risk-adjusted returns and value addition.

In the finance literature, fund managers use Jensen's alpha to measure the effectiveness of active management. If a particular portfolio mirrors the market index, its beta equals one, and theoretically, there is no way to outperform the market. Alpha captures the ability to achieve returns above or below the expected return. This ability includes factors such as information management. Our empirical results suggest that BL-X portfolios report, on average, a significantly higher alpha (31%) than BL-Bloomberg portfolios. We interpret this as evidence that using recurrent neural networks like LSTM and fundamentals and sentiment data (X posts) can enhance the Forecasting of better-performing assets (higher returns).

Additionally, the portfolio beta for BL-X deviates more from one compared to BL-Bloomberg. This implies that, without additional information, the best strategy for a fund manager is to mimic the market index. Figure 4 shows a graph of the evolution of weekly returns for BL-X and BL-Bloomberg.

9 Software Environment and Performance

This study was developed using Python on the Google Colab platform. Most of the code was developed from scratch except for the use of the code for efficient frontiers generation and BL model analysis provided by Martinsky (2022). Tweet sentiments were extracted using a spark cluster provided by the pay Google Cloud Platform (GCP). Each period of analysis takes around 1 h of computational time using version Pro+ of Colab, which includes an A100-SXM4-40 GB graphing processing unit (GPU).

10 Conclusions

The asset selection process for the portfolio composition can be seen as a combinatorial problem that admits an approximation solution using random sampling to select a few options from the vast population of combinations. Another significant contribution of this work is the proposal of φ metric to select the optimal curve to select assets; It contains the best returns, risk, and diversification tradeoff between all efficient curves. Constructing a process to select assets methodically to generate tactical active portfolio management that considers risk, returns, and diversification is generally possible.

In selecting the best forecasting model, the number of historical periods impacts the performance; the most extensive slot of time is preferred over the small. The models could be changed in each iteration; this will be indicated by adjustments in the underlying structures between the assets and their relationships. Features work could be interesting in analyses of this change in the models and its association with market dynamics.

This work has provided evidence for using deep learning artificial intelligence techniques like LSTM for Forecasting and constructing recommendations to

implement views in the BL model and notably beat the S&P500 market by 31.91% alpha annualized, namely, discounting systematic risk and free risk rate.

There are obvious limitations with the amount of time invested in computation for the training and validation in each iteration. However, it is a proposal for features works that apply big data techniques to scale in parallel with training and testing, like the map-reduce paradigm, to train many models and perform real-time analyses of tweets or news. Another limitation of this study is that it uses the classical version of the BL model, assuming a normal distribution of returns, and there is no covariance matrix model for the views. Further studies could incorporate fewer assumptions in their models.

Our main contribution is to show empirical evidence about using sentimental information (tweets) systematically joined with fundamental variables in the asset allocations conducted to generate alpha in the S&P500 market in out-sample periods.

Funding Open Access funding provided by Colombia Consortium. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have not disclosed any conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. In: *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, (pp. 1–6)
- Albarrak, M. S., Elnahass, M., Papagiannidis, S., & Salama, A. (2020). The effect of Twitter dissemination on the cost of equity: A big data approach. *International Journal of Information Management*, *50*, 1–16.
- Amel-Zadeh, A. (2011). The return of the size anomaly: Evidence from the German stock market. *European Financial Management*, *17*, 145–182.
- Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint [arXiv:1908.10063](https://arxiv.org/abs/1908.10063)
- Barua, R., & Sharma, A. K. (2022). Dynamic Black Litterman portfolios with views derived via CNN-BiLSTM predictions. *Finance Research Letters*, *49*, 103111.
- Bessler, W., Opfer, H., & Wolff, D. (2017). Multi-asset portfolio optimization and out-of-sample performance: Evaluating Black–Litterman, mean-variance, and naïve diversification approaches. *The European Journal of Finance*, *23*, 1–30.

- Black, F., & Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research*, 115(1), 7–18.
- Black, F., & Litterman, R. (1991). Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income*, 1(2), 7–18. <https://doi.org/10.3905/jfi.1991.408013>
- Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43. <https://doi.org/10.2469/faj.v48.n5.28>
- Brownlee, J. (2018). Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. *Machine Learning Mastery*
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52, 57–82.
- Chari, S., Hegde-Desai, P., & Borde, N. (2017). A literature review on short term overreaction generated by news sentiment in the stock market
- Colasanto, F., Grilli, L., Santoro, D., & Villani, G. (2022). BERT's sentiment score for portfolio optimization: A fine-tuned view in Black and Litterman model. *Neural Computing and Applications*, 34(20), 1–15.
- Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122, 221–247.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25, 383–417.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116, 1–22.
- Graupe, D. (2013). *Principles of artificial neural networks (Vol. 7)*. World Scientific
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780.
- Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4, 9.
- Idzorek, T. (2007). A step-by-step guide to the Black-Litterman model: Incorporating user-specified confidence levels. *Forecasting expected returns in the financial markets* (pp. 17–38). Amsterdam: Elsevier.
- Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23, 389–416.
- Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132, 126–149.
- Kamara, A. F., Chen, E., Liu, Q., & Pan, Z. (2020). A hybrid neural network for predicting days on market a measure of liquidity in real estate industry. *Knowledge-Based Systems*, 208, 106417.
- Li, H., Guo, Y., & Park, S. Y. (2017). Asymmetric relationship between investors' sentiment and stock returns: Evidence from a quantile non-causality test. *International Review of Finance*, 17, 617–626.
- Li, Y., & Ma, W. (2010). Applications of artificial neural networks in financial economics: a survey. In: *2010 International symposium on computational intelligence and design*, (Vol. 1, pp. 211–214)
- Litterman, B. (2003). *Modern investment management: An equilibrium approach*. Wiley.
- Liu, Y. (2019). Novel volatility forecasting using deep learning—long short term memory recurrent neural networks. *Expert Systems with Applications*, 132, 99–109.
- Liu, K., Zhou, J., & Dong, D. (2021). Improving stock price prediction using the long short-term memory model combined with online social networks. *Journal of Behavioral and Experimental Finance*, 30, 100507.
- Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30, 15–29.
- Lumholdt, H., Lumholdt, H., & Weis. (2018). *Strategic and Tactical Asset Allocation*. Springer
- Martinsky, O. (2022, Nov 13). *Portfolio Optimization II : Black-Litterman model*. Retrieved from www.quantandfinancial.com/2013/08/black-litterman.html
- Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31–42. <https://doi.org/10.2469/faj.v45.n1.31>
- Minami, S., et al. (2018). Predicting equity price with corporate action events using LSTM-RNN. *Journal of Mathematical Finance*, 8, 58.
- Moritz, B., & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. Available at SSRN 2740751
- Palomba, G. (2008). Multivariate GARCH models and the Black-Litterman approach for tracking error constrained portfolios: An empirical analysis. *Global Business and Economics Review*, 10, 379–413.

- Pesaran, M. H., et al., (2010). Predictability of asset returns and the efficient market hypothesis. Tech. rep., CESifo
- Piñeiro-Chousa, J. R., López-Cabarcos, M. Á., & Pérez-Pico, A. M. (2016). Examining the influence of stock market variables on microblogging sentiment. *Journal of Business Research*, 69, 2087–2092.
- Rezaei, H., Faaljou, H., & Mansourfar, G. (2021). Intelligent asset allocation using predictions of deep frequency decomposition. *Expert Systems with Applications*, 186, 115715.
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39, 119–138.
- Shiller, R. J. (1987). The volatility of stock market prices. *Science*, 235, 33–37.
- Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives*, 4, 19–33.
- Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48, 454–488.
- Treynor, J. (1965). How to rate management of investment funds
- Yıldırım, D. C., Toroslu, I. H., & Fiore, U. (2021). Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation*, 7, 1–36.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.