Proceedings of the 3rd International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/176/2025.22100

# *Machine Learning Approaches to Stock Index Prediction*

**Shuo Wang**

*Shenzhen College of International Education, Shenzhen, China*
*s22312.wang@stu.scie.com.cn*

***Abstract:*** In response to the stock market's volatile nature, this research examines stock index forecasting evolution from traditional econometric models to advanced machine learning techniques. Market volatility, influenced by economic conditions, investor sentiment, and market interconnectedness, often renders linear models inadequate. While fundamental, conventional methods like multiple regression and ARMA face limitations with non-linear, noisy data, prompting development of machine learning approaches such as BP neural networks, SVM, and attention-enhanced CNN-LSTM models. These advanced techniques better capture data complexity, significantly improving prediction accuracy. The study explores both macro factors (economic linkages) and micro elements (herding behavior, loss aversion), alongside innovations like social media sentiment analysis that incorporate emotional and behavioral insights. Despite progress, challenges remain in balancing model complexity with accuracy and overcoming traditional statistical constraints in non-linear environments. This review emphasizes the necessity for integrated, fuzzy prediction models that consider multiple influences, with potential applications extending to other time series like commodity prices. These findings underscore the need for flexible, accurate forecasting methodologies to help authorities and investors navigate the unpredictable financial landscape.

***Keywords:*** Deep learning, Index prediction, Factor analysis, Bibliometric analysis, Investor sentiment.

## 1. Introduction

Nowadays stock forecasting has been increasingly challenging due to the complications of the stock market. As the stock market itself is complicated, many factors such as national policies, national economic conditions, external emergencies and investor sentiment will have a significant impact on the stock market, which makes the trend of the stock market extremely difficult to predict. Stock market prediction has been the core focus in the field of quantitative investment, and it is also an important problem troubling investors [1]. Many scholars at home and abroad have also paid a lot of energy to study the relevant forecasting methods, aiming to analyze the historical data related to the stock market. In the early research on the stock market, traditional econometric models based on mathematical statistics were mainly used to forecast the stock price, common ones include the multiple regression model, hidden Markov model, auto-regressive moving average (ARMA) model, etc. However, since the stock price data is non-linear, non-stationary and contains noise data, the stock price data can be used to predict the stock price. However, econometric models are mostly applicable to stationary series, so the use of econometric models to forecast often can not get good

DOI: 10.54254/2754-1169/176/2025.22100

prediction results. With the rapid development of modern informatization and digitalization, computer-related technologies are becoming more and more complete, human intelligence, machine learning and other fields have entered people's vision, and the field of quantitative investment has gradually risen. Scholars have begun to combine econometric models with machine learning-related algorithms to build prediction models. Common models such as neural network BP and support vector machine (SVM) can better mine the information in the data and improve the accuracy of the prediction [2-3].

The stock market has a dynamic and complex nonlinear environment, and its future trend has always been considered difficult to predict. However, predicting the trend of the stock market is the core field to research: on the one hand, it can provide a reference for the government and regulatory agencies, help them understand the market development situation, conduct macro-control in advance to avoid major risks, and promote the healthy development of the market economy; On the other hand, for investment institutions and individual investors, it can also provide an important reference to help judge the timing of buying and selling. In the previous research on stock index prediction, many researchers focused on the construction and optimization of prediction model, and improved the prediction accuracy by adjusting parameters and model structure, but often neglected the analysis from the perspective of influencing factors of stock index. The traditional stock index prediction model usually relies on the volume, transaction amount, opening price, closing price, maximum price, minimum price and other indicators, and the relationship between these factors may not show an obvious linear form. If the input data feature is too single, the prediction accuracy will be limited. However, too many features may lead to excessive model complexity. Therefore, the stock index prediction model needs to scientifically select the input characteristics. In addition, the effectiveness of a single prediction model is often limited, and integrated methods often provide better prediction performance.

This paper aims to conduct bibliometric analysis and review the research in the forecast system of stock index rise and fall.

## 2. Influencing Factors

### 2.1. Economic

According to the economic market hypothesis, the linkage of stock markets between countries stems from the common changes in the economic fundamentals of each country and is mainly transmitted through international trade and capital flow. When a country's economy changes, countries with close trade and investment ties are the first to be affected, and the stock market is sensitive to such changes. At this point, the volatility of stock markets increased the interconnectedness between countries. In addition, adjustments by international investors can exacerbate stock market reactions. However, when macro fundamentals are not closely linked, the stock market can still move in sync, which is often attributed to irrational expectations or consistent expectations of investors. Irrational expectation means that investors follow others in making decisions when information is asymmetrical, while consistent expectation means that investors make the same decisions due to similar information and social background, which leads to stock market linkage.

### 2.2. Micro Factors

Herding refers to the phenomenon that individuals are influenced by group behavior and tend to conform to the majority. In the financial market, investor sentiment is often affected by the external environment and internal decision-making bias, and the herd effect is often accompanied by emotional contagion in the process of information transmission. Due to differences in investors' ability to access information, many prefer to follow others' investment decisions rather than make

DOI: 10.54254/2754-1169/176/2025.22100

independent judgments based on their own analysis. In the stock market, herding induces overconfidence through collective similar behavior, leading investors to chase gains and losses. Positive emotions generally precede a rise in stock prices and negative emotions a decrease in stock prices. Simply said, the herd effect boosts investor mood, which drives more people to follow like-minded approaches, hence raising market volatility. Proposed by Kahneman and Tversky [3], loss aversion theory holds that people naturally want to avoid losses rather than chase profits. This psychological bias drives investors to exhibit illogical emotional reactions in the stock market, including too pessimistic or optimistic attitudes, therefore aggravating market volatility. Investors may act aggressively or conservatively in response to possible losses, therefore influencing market supply and demand and increasing stock price volatility. Ultimately, loss aversion theory shows how much emotional impact shapes stock price volatility. Rational expectations' systematic bias hypothesis claims that cognitive restrictions and emotional elements influence investors' market expectations, hence generating systematic bias in expectations. This bias might cause investors to make illogical assessments of the future trajectory of the market, therefore aggravating stock price volatility. According to the hypothesis, the development and modification of market expectations depend much on investor mood, which influences stock price swings. In the first stage of the market, the deviation makes it difficult for investors to form consistent expectations based on objective information, which raises the uncertainty of the market and causes the aggravation of stock price volatility. Deviation and the "herding effect" help investor mood converge in the process of expectation adjustment; so, the propagation of irrational expectations increases market volatility even more.

Studies abound demonstrating how frequently academics use text data like Weibo, Twitter, financial news and stock comments in financial text analysis. With social media data like Weibo becoming available, Nisar et al. contend that sentiment analysis over an extended period of time can help forecast market movements on Twitter [2]. Maqbool et al. used NLTK and NLP libraries to conduct sentiment analysis on stock news and studied the influence of emotion polarity on stock prediction [4]. Bao et al. used convolutional neural networks (CNN) to conduct supervised learning on different types of stock news and extract news event features for stock prediction [5]. Yuan Zheng combined the text CNN and BiLSTM to conduct sentiment analysis on stock investor comments and found that the fusion model of emotion factors was more effective in stock price prediction [6]. Deveikyte et al. analyzed stock bar comments by SVM sentiment classifier and found that the market sentiment was similar to the trend change of FTSE100 [7]. Although many studies believe that text such as microblogs or Twitter can better extract investor sentiment, the dynamic relationship between stocks in an industry is often ignored in the stock price analysis. Through the improved cross-modal Transformer fusion model and multi-graph convolutional attention network, information of different modes can be fused more effectively and the accuracy of stock price prediction can be improved.

## 3. Stock Index Forecasting Method

### 3.1. Attention Mechanism-based Approach

Although traditional time series methods can predict stock prices, they rely on certain assumptions that are not always applicable in real data and therefore have limited effect in practical applications. In contrast, machine learning methods do not rely on these assumptions and are able to learn data features, adapt to non-linear data, and make predictions more effectively. Many classical machine learning models (such as support vector machines, random forests, etc.) play an important role in financial timing prediction. According to Yi et al., the Attention mechanism is integrated into the CNN-LSTM model, the data dependence in time series is captured by Attention, the long sequence information is extracted, and the sampling prediction is made based on the probability density

DOI: 10.54254/2754-1169/176/2025.22100

function, and the point prediction and interval prediction of stock price is finally realized [8]. The experiment shows that the CNN-LSTM model with Attention mechanism is superior to other benchmark models in comprehensive performance and can effectively improve the multi-step prediction accuracy of SSE index closing price. The introduction of the attention mechanism into the CNN-LSTM model can directly generate multi-step predictions in the prediction process. Compared with recurrent neural networks, the attention mechanism can extract the information of the entire sequence and capture long-distance dependencies, thus retaining the original data better. Experiments show that the model outperforms the benchmark model in predicting the closing price of Shanghai Composite Index.

In recent years, attention mechanisms have been gradually applied to time series prediction because they can learn spatio-temporal dynamics and assign different weights to different attributes. Spatial attention mechanism pays attention to the influence of attributes on prediction results. The model combining BiLSTM and spatial attention mechanism proposed by CHEN et al can achieve higher prediction accuracy by increasing the weight of key features. LSTM is unable to capture long-term dependency effectively due to the problem of training instability and gradient disappearance, but the temporal attention mechanism alleviates this problem by weighting implicit states and can dynamically select temporal dependencies. Shih et al. proposed a time-pattern-based attention mechanism that can learn the interdependence between variables in different time steps [9]. LSTM effectively alleviates the gradient disappearance problem through its unique gate structure (forgetting gate, input gate, output gate), but it is still difficult to capture the different contributions of different time points and different input characteristics to the result. The attention mechanism simulates the human brain mechanism, optimizes the allocation of computing resources and improves the model performance by calculating the attention weight. Yang proposed a novel spatiotemporal attention mechanism that combines spatial and temporal attention mechanisms to capture dynamic spatiotemporal correlations in the stock market. In the spatial dimension, the spatial attention mechanism adaptively captures the dynamic correlation between nodes, and the weights are used to measure the influence of nodes on the prediction in the time step [10]. BiLSTM stores time information and controls its changes through a gate mechanism, but may underestimate the influence of earlier states in long time series predictions. By measuring the importance of hidden states in different time Windows and combining them with the temporal attention mechanism, the neural network can pay more attention to valuable information. However, RNN and LSTM models are not obvious enough to learn the influence of some policy factors. For example, from July 2020 to January 2022, the global economy declined due to the impact of the pandemic, and the stock market was also affected. In order to cope with the economic impact brought by the COVID-19 epidemic, China issued a series of market rescue policies. The model is not enough to capture the influencing factors of emergencies.

## 4. Existing Limitations and Future Outlook

Since the fluctuation of the stock index is affected by a variety of nonlinear factors, the prediction model of the stock index should have the following characteristics: strong nonlinear data processing ability, automatic adjustment and self-learning ability, the ability to deal with multiple factors and indicators, and can take into account both quantitative and non-quantitative indicators. Although neural networks (such as BP and RBF neural networks) are better than statistical models in stock index prediction, they still have defects, such as slow convergence and easy fall into local extremes. The prediction accuracy of the model still needs to be further improved. This paper studies the fuzzy prediction model of the stock index considering the linkage effect of investor sentiment and the stock market, and further discusses the analysis of other influencing factors in the future, and tries to integrate more factors into the model to improve the accuracy and stability of stock index prediction.

DOI: 10.54254/2754-1169/176/2025.22100

This paper mainly focuses on the prediction of the stock index, and in the future, the application of the model can be extended to analyze the influence mechanism in other fields, such as commodity prices, and other time series in real life.

Since stock index fluctuations are affected by numerous nonlinear factors, prediction models should possess strong nonlinear data processing capabilities, self-learning abilities, and capacity to handle multiple factors including both quantitative and non-quantitative indicators. While neural networks outperform statistical models, they still face challenges like slow convergence and local extrema issues [6]. Future research should focus on fuzzy prediction models incorporating investor sentiment linkage effects and other influencing factors to improve accuracy and stability. The application of these models could extend beyond stock indices to other time series analysis domains such as commodity prices [7]. Looking ahead, including more varied inputs and extending relevance to other fields promises to significantly increase model usability and accuracy in an ever-changing financial environment.

## 5. Conclusions

Stock market forecasting remains exceptionally challenging due to inherent market complexity and unpredictability. Traditional econometric models like multiple regression, hidden Markov models, and ARMA struggle with the non-linear, non-stationary nature of stock data, limiting their real-world effectiveness. These limitations have driven advances in machine learning approaches, with CNN-LSTM models featuring attention mechanisms, SVMs, and BP neural networks demonstrating superior performance in pattern recognition and prediction accuracy. This paper highlights the importance of examining both macro factors (economic fundamentals, international market linkages) and micro elements (investor sentiment, herding behavior, loss aversion). Sentiment analysis from social media and financial news has further enhanced forecasting by incorporating psychological dimensions of market behavior. The integration of econometric foundations with machine learning and multi-factor analysis offers a comprehensive strategy for addressing market nonlinearity. Improved multi-step prediction accuracy for indices like the Shanghai Composite Index demonstrates the effectiveness of attention-based mechanisms and influencing-factor approaches. These developments provide institutions and investors with robust tools for informed decision-making, risk mitigation, and market stability promotion.

## References


[1] *Maganioti, A.E., Chrissanthi, H.D., Charalabos, P.C., Andreas, R.D., George, P.N. and Christos, C.N. (2010) Cointegration of Event-Related Potential (ERP) Signals in Experiments with Different Electromagnetic Field (EMF) Conditions. Health, 2, 400-406.*

[2] *Bootorabi, F., Haapasalo, J., Smith, E., Haapasalo, H. and Parkkila, S. (2011) Carbonic Anhydrase VII—A Potential Prognostic Marker in Gliomas. Health, 3, 6-12.*

[3] *Glendinning, I. (2013). Comparison of policies for Academic Integrity in Higher Education across the European Union. Retrieved from http://ketlib.lib.unipi.gr/xmlui/bitstream/handle%20European%20Union.pdf?sequence=2*

[4] *Yi, S., Liu, H., Chen, T., Zhang, J., & Fan, Y. (2023). A deep LSTM-CNN based on self-attention mechanism with input data reduction for short-term load forecasting. IET Generation, Transmission & Distribution, 17(4), 931-938. https://doi.org/10.1049/gtd2.12763*

[5] *Nisar, T. M., & Yeung, M. (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. The Journal of Finance and Data Science, 4(2), 101-119. https://doi.org/10.1016/j.jfds.2017.11.002*

[6] *Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263-292. http://links.jstor.org/sici?sici=0012-9682%28197903%2947%3A2%3C263%3APTAAOD%3E2.0.CO%3B2-3*

[7] *Maqbool, J., Aggarwal, P., Kaur, R., et al. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach. Procedia Computer Science, 218, 1067-1078. https://doi.org/10.1016/j.procs.2023.01.086*

[8] *Bao, W., Cao, Y., Yang, Y., Che, H., Huang, J., & Wen, S. (2025). Data-driven stock forecasting models based on neural networks: A review. Information Fusion, 113, 102616. https://doi.org/10.1016/j.inffus.2024.102616*

[9] *Zheng, X. (2023). Stock price prediction based on CNN-BiLSTM utilizing sentiment analysis and a two-layer attention mechanism. Advances in Economics Management and Political Sciences, 47(1), 40-49. https://doi.org/10.54254/2754-1169/47/20230369*

[10] *Deveikyte, J., Geman, H., & Piccari, C. (2022). A sentiment analysis approach to the prediction of market volatility. Frontiers in Artificial Intelligence, 5, AI in Finance. https://doi.org/10.3389/frai.2022.836809*