Proceedings of the 3rd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/150/2024.19306

An Analysis of the Application and Impact Mechanisms of Machine Learning in the Field of Asset Pricing

Motong Li[1,a,*]

[1]*Fudan University, Shanghai, 200433, China*
a. 23300680085@m.fudan.edu.cn
**corresponding author*

Abstract: Issues related to asset pricing and return forecasting have consistently been prominent research topics in both the financial industry and academia. With the rapid advancement of science, technology, and artificial intelligence, machine learning has garnered extensive attention from scholars for its robust self-learning and adaptive capabilities, as well as its exceptional advantages in large-scale data processing. This paper provides a detailed overview of the application of emerging machine learning methods in asset pricing and offers an in-depth analysis of machine learning's contributions to this field based on existing research. This research finds that machine learning has introduced profound changes and progressive innovations in asset pricing models. Additionally, this paper examines the limitations of machine learning methods, objectively highlighting current shortcomings such as poor interpretability and model overfitting. Finally, this paper proposes future directions for improving machine learning models to advance the financial field further.

Keywords: Machine Learning, Asset Pricing, Return Forecasting.

1. Introduction

Asset pricing involves the uncertain valuation of financial assets to determine their fair price and ultimately achieve the maximum expected rate of return. In investment and trading, both asset pricing and forecasting future price trends are enduring research interests and the focus on innovation for industry professionals, as well as a topic of widespread public concern. The most established approach in asset pricing is the factor pricing model. Starting with the Capital Asset Pricing Model (CAPM), which centers on market factors, the field has evolved through the three-factor model (adding scale and value factors to explain stock excess returns), the four-factor model (introducing the momentum effect), and the five-factor model (incorporating profitability and investment factors). Traditional quantitative model development generally aligns with the goal of identifying market anomalies unexplained by existing models. When parameter adjustments fail to address an issue, new factors are introduced, resulting in the publication of updated models.

The financial market is a complex, time-sensitive system, particularly in the case of China's A-share market, which has long been a focal point of public and professional interest. With the explosive growth of transaction data in recent years, market factors have expanded in scope, creating a high-dimensional data environment. Moreover, substantial amounts of noise and previously unexplained anomalies exist within the market. Studies in both domestic and international literature have observed non-linear characteristics in financial transaction data, including long-term memory, fat tails, and

DOI: 10.54254/2754-1169/150/2024.19306

fractal patterns. These new dynamics indicate that the traditional asset pricing methods, based on linear measurement and modeling, face significant challenges and require innovative adaptation.

With the rapid development of Internet big data in recent years, machine learning—a product of the new era—has been applied innovatively in fields such as biomedicine, offering promising new directions for breakthroughs in asset pricing. Financial transaction data is vast, and AI possesses strong capabilities for large-scale data processing and storage. Numerous factors influence securities market prices, and machine learning provides powerful self-organizing, self-learning, and adaptive abilities. Most importantly, the nonlinear capabilities of machine learning methods can significantly enhance the accuracy of traditional multi-factor models in predicting nonlinear time series, addressing the limitations of traditional financial models and accelerating the innovation and development of quantitative financial models in asset pricing.

This paper will detail the current applications of machine learning in asset pricing, analyzing its potential impacts and underlying mechanisms. Our aim is to bring attention to the advantages and prospects of machine learning in quantitative finance while highlighting its potential value in advancing research directions in asset pricing.

2. Overview of Machine Learning in Financial Asset Pricing

In recent years, emerging machine learning models have excelled at handling the vast amount of financial time series data, which are inherently nonlinear, non-equilibrium, high-dimensional, and noisy. These models are able to accurately capture hidden features within the data. Research shows that quantitative models integrating innovative machine-learning methods outperform traditional models. Although the application of machine learning in asset pricing is still in its early stages, the benefits of such cross-disciplinary integration are evident in the existing applications of several popular machine learning methods.

2.1. Support Vector Machines

The support vector machine (SVM), originally proposed by Vapnik and Alexey in 1963, has gained widespread recognition and application in machine learning due to its unique classification capabilities and efficient algorithmic implementation. The core principle of SVM is to find a hyperplane that separates different types of data points while maximizing the minimum distance, or "margin," between the points and the hyperplane. The hyperplane with the largest margin is known as the optimal hyperplane, typically found by solving the Lagrange multiplier and dual optimization problems.

Currently, the SVM algorithm is applied in fields such as text classification, image recognition, and bioinformatics. It has also been adopted in asset pricing, where financial researchers have made notable progress. For example, Endri et al. [1] addressed financial conditions and stock return issues in the Islamic stock market by developing an SVM-based early warning model to predict delisting events through asset pricing metrics and various financial indicators. To address nonlinear and nonparametric financial time series issues, Lai and Liu [2] implemented a wavelet-based SVM platform and constructed three related algorithms, achieving optimal performance with the least-squares SVM (LSSVM). Kurani et al. [3] observed that SVM's simplified decision boundary helps mitigate overfitting, achieving an accuracy rate of around 60%–70%. They further noted that integrating SVM with other methods, such as genetic algorithms, could improve pricing performance.

Additionally, as research deepens into the innovative integration of machine learning in asset pricing, an increasing number of hybrid models leveraging complementary algorithmic strengths are emerging. Matilda et al. [4] employed a Hybrid Crawler Search Lévy (HRSR) algorithm to enhance the performance of the traditional SVM model. This SVM-HRSR approach predicts stock market trends with lower error rates by leveraging predictive patterns. Santoso et al. [5] noted that an integrated model combining the Gaussian mixture model with SVM achieved an advanced Sharpe ratio during testing and has been used to forecast stock price trends in the ASCII.JK market. Furthermore, rigorous business studies have shown that the EMD-LSSVM model, which incorporates empirical mode decomposition, the least squares method, and SVM with optimized parameters, offers superior predictive accuracy compared to models like GS [6].

The SVM model has demonstrated robust classification performance on small sample sets in the financial market. By constructing and selecting appropriate kernel functions, it can effectively address nonlinear issues, making it well-suited for in-depth asset pricing applications. However, SVM faces limitations in training efficiency with large datasets and is sensitive to parameter choices, such as kernel function parameters. Future research may enhance SVM's capability to handle large-scale data, multi-class classification challenges, and integration with other learning models.

2.2. Random Forests

Random forests are an ensemble technique based on decision trees, which construct multiple decision trees to improve overall model accuracy. By averaging the results of multiple trees, random forests enhance predictive accuracy and help control overfitting. Since their introduction by Breiman in 2001, random forests have become one of the most popular algorithms in machine learning, widely used for classification and regression tasks. Stock price fluctuations and return rate predictions are problems that random forests are particularly well-suited to analyze.

Due to their strong predictive performance and flexibility, random forests have been applied across various fields, including healthcare, ecology, and environmental monitoring. Recently, they have also gained traction in financial asset pricing, with initial research yielding promising results. Bin et al. [7] proposed three autoregressive machine learning models based on stock market data during the COVID-19 pandemic. After comparing experimental data, they recommended using the AR-RF model for data sequences with fewer observations to support investment decision-making and financial policy formulation. Tan et al. [8], using Chinese stock market data, selected two feature spaces for price trend prediction, assessing the robustness of the RF model and verifying the potential for significant excess returns in these paradigms. Basak et al. [9] concentrated on the classification problem of stock price prediction, employing random forests and gradient-boosted decision trees to create a framework for predicting stock price direction, which successfully demonstrated efficiency. Moreover, their selection of indicators and features contributed to improved accuracy in medium- and long-term stock price predictions. Ravinder and Lokesh [10] highlighted the practical application of intelligent machine-learning tools in high-frequency stock markets. They proposed an ensemble model that incorporates deep learning, gradient boosting machines, and distributed random forests, confirming its high accuracy in stock price prediction.

Random forests integrate multiple decision trees, generally yielding more accurate results than a single decision tree. By enhancing model diversity, they also reduce susceptibility to overfitting. However, this model often incurs high computational costs, and as a “black-box” model, it lacks strong interpretability. Addressing these limitations is a key focus for the future development and research of random forest models in financial applications, such as securities yield prediction.

2.3. Neural Networks

Neural networks are computational models designed to mimic the structure and function of neurons in the human brain, making them useful for solving various machine learning and artificial intelligence problems. The model comprises numerous nodes, or neurons, typically arranged in layers: an input layer, hidden layers, and an output layer. There are several types of neural networks,

DOI: 10.54254/2754-1169/150/2024.19306

including the commonly used feedforward neural networks, convolutional neural networks (suitable for processing image data), and recurrent neural networks (suitable for sequential data processing), among others. Neural networks provide numerous modern conveniences in fields such as image recognition, natural language processing, and speech recognition.

Neural networks were introduced to the financial field earlier than random forests and support vector machines, and they play an active role in asset pricing model research. Xiaohan et al. [11] combined meta-path attention mechanisms with graph neural networks to predict stock market volatility by classifying multi-source heterogeneous graph data, demonstrating that this approach effectively captures the unique characteristics of stock market data and the implicit semantic information within relationships. Lei [12] developed a three-layer RNN neural network to analyze convergence speed and predict stock prices on the Shanghai Stock Exchange, noting that neural networks are feasible and reasonable tools for stock price forecasting. Xiangwei and Xin [13] identified challenges with the standard backpropagation neural network, such as slow convergence speed, susceptibility to local optimization, and difficulty in determining network parameters and structure. They addressed these by incorporating the LM algorithm, enhancing generalization ability, and adopting a superior normalization method, resulting in improved predictive accuracy. Zhang et al. [14] developed a granular computing-based neural network (GNN) stock prediction model, comparing its performance against the BP algorithm on the same dataset and finding that GNN exhibited a lower average error rate. Ma and Han [15] focused on the application of deep neural networks for stock trading strategies. By analyzing historical data from the A-share trading market using the sliding window method, they derived a trading strategy that demonstrated high accuracy in predicting returns, low volatility, and suitability for deep neural networks.

Neural networks, valued for their strong nonlinear fitting and generalization capabilities, have been applied to financial asset pricing research from an early stage. Their flexibility, scalability, and support for multi-task learning offer a broader optimization space than other machine learning models. While challenges such as overfitting risk and limited interpretability persist, neural networks still show significant promise in the area of price prediction.

2.4. Deep Learning

Deep learning is a subfield of machine learning that traces its development back to basic neural network models. It employs multi-layer neural networks to simulate human learning, solving complex problems by learning multi-level representations and abstractions of data.

Often regarded as an advanced version of basic neural networks, deep learning has been recognized by industry insiders and researchers for its unique advantages in handling the large, high-dimensional financial data of the modern era. Among deep learning models, CNN and LSTM have garnered widespread attention due to their adoption of end-to-end data processing methods. Particularly in asset pricing, where streamlined modeling techniques are urgently needed, these models hold significant research value. Ronil and Anil K. [16] found in their sample data analysis, based on the indices of the top ten industries in the Asia-Pacific region, that the out-of-sample forecasting ability of the integrated CNN-BiLSTM model's closing price is stronger than that of the two models used separately. Additionally, the portfolio results obtained by combining the classic Black-Litterman model exhibited better financial efficiency and diversity. Man et al. [17] introduced a new similarity measurement method called logical weighted dynamic time warping (LWDTW) to improve the efficiency of cluster analysis, and further implemented a cluster-based forecasting framework. They combined three deep learning models and ultimately introduced a real sample dataset of US stocks, achieving the best forecasting results by combining LWDTW clustering and LSTM models. Agrawal et al. [18] argued that there is a certain degree of non-correlation between the attributes of stock data. They constructed an evolutionary deep learning model (EDLM) that uses stock trend indicators (STIs)

DOI: 10.54254/2754-1169/150/2024.19306

to identify stock price trends. This model significantly improved upon the original deep learning model, a point they verified through data analysis. Eapen et al. [19] began with basic CNN and GRU models, proposing several variants of multi-pipeline and single-pipeline deep learning models based on different CNN kernel sizes and GRU unit numbers. They established a new type of deep learning model, which, after training on data, showed significant improvements in both forecasting performance and model implementation time.

Compared to neural networks with relatively limited learning capabilities that can only capture shallow features of data, deep learning models extend the neural network concept. Their multiple hidden layers allow them to learn higher-level features of data, making them more adept at handling complex problems.

2.5. Summary

From a practical standpoint, machine learning methods, represented by popular models such as SVM and RF, have demonstrated superior performance compared to traditional econometric models in most empirical studies after being introduced to the field of asset pricing. Whether considering the original models themselves or innovative machine learning models that have undergone parameter optimization and integration of multiple methods, their inherent ability to handle unstructured data and extract deep underlying features not only gives machine learning a unique advantage in the pricing field but also introduces a new way of thinking for the transformation of quantitative finance, especially when facing financial markets characterized by uncertainties and noisy data.

3. An Analysis of the Impact Mechanisms of Machine Learning Methods in Financial Asset Pricing

Although the application of machine learning methods in asset pricing is still in its early stages, it has already injected fresh impetus into the traditional research paradigm within the field. The new developmental path it offers breaks away from the research process of traditional multi-factor pricing models, shifting the focus from market phenomenon research to data feature processing patterns. The intelligent advantages and characteristics of machine learning perfectly align with and address the main developmental issues faced by traditional econometric models. However, they also bring about new challenges and prospects for concern.

3.1. Analysis of the Promoting Effect of Machine Learning Methods on Asset Pricing

On the one hand, machine learning promotes the internal principles and operational models of pricing, enabling them to better adapt to market development trends.

At the current stage, the biggest challenge faced by traditional financial econometric models is that their asset pricing results are based on a classic linear relationship with pricing factors. However, as non-linear characteristics such as long memory, fat tails, and fractals in financial markets have been successively identified, and with the rapid development of online media and other factors leading to the explosive growth of unstructured financial big data, the entire market transaction data is showing a diversified development trend in both volume and features. This trend has gradually made traditional pricing models unsuitable for the new era of financial markets. At this point, machine learning methods, with their inherent non-linear characteristics, strong capabilities for processing and storing large-scale data, and exceptional self-organization, self-learning, and self-adaptive abilities, can address the current challenges. They significantly promote the innovative development of models in the field of asset pricing and effectively predict future asset returns. For example, Luyang et al. [20] successfully applied generative adversarial networks and recurrent long short-term memory networks—non-linear asset pricing models—to capture a large amount of macroeconomic and firm-

DOI: 10.54254/2754-1169/150/2024.19306

specific information, a set of small hidden economic state processes, and the non-linear effects of conditional variables, generating a novel model with out-of-sample performance that surpassed all other benchmark methods.

On the other hand, machine learning simplifies the modeling methods to some extent in selecting pricing factors.

In the secondary market, numerous factors influence stock prices and returns. Within the framework of mainstream multi-factor models, the number of influencing factors identified through various methods is also large, leading to the notion of a "factor zoo" [21]. Traditional pricing models require experts and scholars to study economic laws based on the characteristics of the model and data, select a subset of pricing factors, and determine the suitability of factor types. This choice significantly impacts the final output of the pricing model. In contrast, machine learning methods, such as CNN and LSTM models, adopt an end-to-end data processing approach. One simply feeds the data into the AI, which determines the number and type of pricing factors suitable for the model based on the data's characteristics and then produces the output results. This eliminates the need for manual intervention, representing a significant improvement in both the accuracy of price prediction and the simplification of model construction. For instance, Heaton et al. [22] found that financial trading prediction problems often involve large datasets with complex data interactions, which are difficult or impossible to explicitly specify in complete theoretical economic models. However, deep learning methods can start with the data itself, detect and utilize interactions within the data, and produce superior predictive outcomes.

Additionally, the integration of machine learning can help address other issues beyond the classification prediction of stock price movements and the regression prediction of returns, such as managing the high-dimensional features of financial market data. From the perspective of data dimensionality reduction, through large-scale data training and analysis, machine learning can rank the importance of each factor in multi-factor models used for predicting returns. In other words, machine learning can further refine and supplement conclusions and patterns in the field of asset pricing based on the inherent patterns in the data itself.

3.2. Analysis of the Adverse Effects of Machine Learning Methods on Asset Pricing

First, the process of establishing models using machine learning methods is unknowable and highly non-interpretable, which makes it difficult to provide specific explanations using economic theory.

Machine learning methods based on end-to-end processing only require human intervention at the input stage for data importation; the process of model establishment is autonomously completed by the machine learning algorithms. Such a modeling approach, while highly convenient, lacks strong interpretability and is considered a "black-box" model. As a result, the output of research or scientific achievements is more difficult to explain and justify through rigorous economic principles. This raises questions about the roles of human expertise and artificial intelligence in the asset pricing process. Cao et al. [23] have addressed this issue. They found that when institutional and professional knowledge is crucial, such as in cases involving intangible assets and financial distress, humans win in the "human vs. machine" competition; however, when information is transparent but voluminous, artificial intelligence prevails. They also highlighted the additional value of human knowledge in asset pricing, which can greatly reduce extreme errors and prevent AI from focusing too much on the data itself, neglecting the study of patterns. Therefore, in the initial stages of machine learning's application in the field of quantitative finance, while focusing on the significant non-linear and other advantages of machine learning, one should not overlook the interpretation and evaluation of models that involve economic principles in traditional pricing models due to human intervention. Otherwise, it will increase ambiguity in the theoretical explanation and hinder development.

DOI: 10.54254/2754-1169/150/2024.19306

Secondly, machine learning has the drawback of being overly data-oriented throughout the entire modeling and forecasting process, which hinders the incorporation of models that involve certain non-quantifiable indicators.

Traditional econometric asset pricing research focuses primarily on the study of market patterns, while machine learning emphasizes improving data processing, enhancing algorithms, extracting features deeply, and exploring the relationships between them. Machine learning methods based on feature processing rely heavily on extensive data training, enabling AI to autonomously discover patterns within the data, which is grounded in vast transaction data. However, such an approach inevitably overlooks the consideration of objective factors that are difficult to capture through data and are not easily quantifiable. Scholars have found that models based on neural network methods have often failed to predict asset prices over the same period. Despite being among the most advanced technologies today, relying on powerful computers and algorithms, neural networks cannot disregard basic economic principles. If data is processed without regard for these principles, even the most sophisticated models and algorithms cannot yield accurate results.

Furthermore, many existing machine learning models suffer from overfitting, meaning they become too tailored to the characteristics of the data, which leads to issues with prediction accuracy and other outcomes. Due to its non-linear characteristics, machine learning inevitably increases model complexity, but the model must still align with the input data. For a given dataset, models that are either too simple or too complex are unsuitable. Continuous adjustments are often required to achieve the optimal model, striking a balance between accurately reflecting the relationships between variables and preventing overfitting.

4. Conclusion

4.1. Conclusions of the Study

This article focuses on the innovative introduction and development of emerging machine-learning methods in the field of asset pricing. It provides a detailed introduction to the current popular applications of machine learning methods within the field and further delves into and contemplates the various impacts and mechanisms that machine learning may have in this area. The goal is to raise awareness of the advantages and prospects of machine learning in the field of asset quantification, establish a basic cognitive framework for the current application status, and reveal its potential value in promoting innovative directions in asset pricing research, as well as enriching research methods and approaches within the field.

After discussion, this article finds that machine learning brings about a profound change in the pricing model of the asset pricing field. Traditional econometric models are mostly linear in nature, whereas, in today's era of financial big data explosion, machine learning methods that handle non-linear data characteristics have emerged. These methods can also deal with high-dimensional, large-volume market transaction data. Traditional econometric asset pricing research is rooted in market research, with modeling usually requiring manual summarization of patterns and factor selection. However, as the financial market becomes increasingly complex with numerous influencing factors and various anomalies, both feature-based and end-to-end machine learning models can achieve better results in a more effective way by focusing on the data itself.

Although the machine learning model is more suitable for financial transaction prediction today, it inevitably brings many new problems. The development of machine learning technology in the asset pricing field under big data is still in its infancy, and most of the existing literature simply applies existing algorithms to different data sources, revealing the fact that machine learning models perform outstandingly in various aspects. However, there is a lack of in-depth exploration and analysis of more valuable issues, such as its inherent "black box" nature, which is also the research direction currently lacking.

4.2. Future Prospects

In response to the deficiencies reflected in the preliminary application of machine learning models in the pricing field, future research can be conducted and explored in the following areas:

To begin with, factor adjustment and parameter optimization of existing models. Consider adding innovative factors that machine learning models may overlook during data analysis, such as investor sentiment and industry correlations, which are not easily discernible from existing data. Attention should also be given to the allocation of their weights in the model. Additionally, adjust the relevant parameters in the model through multiple runs to improve the model's prediction accuracy.

Besides, the integration of multiple models. Machine learning and traditional econometric models include many different types of algorithms. By integrating multiple models, they can learn from each other, maximize prediction efficiency through data training, and utilize their respective advantages to complement each other's shortcomings, thus preventing the most common problem of overfitting in machine learning.

Last but not least, the improvement of model interpretability. Especially for machine learning research based on end-to-end processing, the model itself produced by AI, as well as the determination of its parameters—including the presentation of prediction results—are difficult to study and accurately explain from an economic perspective. This is also one of the biggest challenges that machine learning needs to overcome in future developments in the pricing field.

References

[1] Endri, E., Kasmir, K., & Dewi, A. S. (2020). Delisting Sharia stock prediction model based on financial information: Support Vector Machine. Decision Science Letters, 207-214.

[2] Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models. (2014). Computer Science and Information Technology, 2(1).

[3] Kurani, A., Doshi, P., Vakharia, A., et al. (2021). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1), 1-26.

[4] Matilda, P. S., S. P., P. V., et al. (2023). HRSR-SVM: Hybrid reptile search remora-based support vector machine for forecasting stock price movement. International Journal of Information Technology, 15(6), 3127-3134.

[5] Santoso, M., Sutjiadi, R., & Lim, R. (2018). Indonesian stock prediction using support vector machine (SVM). MATEC Web of Conferences, 16401031.

[6] Support Vector Machines; Studies from School of Business Yield New Data on Support Vector Machines (A Hybrid Least Square Support Vector Machine Model with Parameters Optimization for Stock Forecasting) [J]. Journal of Robotics & Machine Learning, 2015.

[7] Bin, A. O., Shuai, H., A. A. S., et al. (2022). Stock market forecasting using the random forest and deep neural network models before and during the COVID-19 period. Frontiers in Environmental Science, 10.

[8] Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. Heliyon, 5(8), e02310.

[9] Basak, S., Kar, S., Saha, S., et al. (2018). Predicting the direction of stock market prices using tree-based classifiers. North American Journal of Economics and Finance, 47, 552-567.

[10] Ravinder, K., & Lokesh, K. S. (2022). An ensemble of random forest gradient boosting machines and deep learning methods for stock price prediction. Journal of Information Technology Research (JITR), 15(1), 1-19.

[11] Xiaohan, L., Jun, W., Jinghua, T., et al. (2022). A graph neural network-based stock forecasting method utilizing multi-source heterogeneous data fusion. Multimedia Tools and Applications, 81(30), 21-23.

[12] Lei, L. (2022). Stock prediction and analysis based on RNN neural network. SHS Web of Conferences, 151.

[13] Xiangwei, L., & Xin, M. (2012). Based on BP neural network stock prediction. Journal of Curriculum and Teaching, 1(1), 45.

[14] Zhang, Y., Akkaladevi, S., Vachtsevanos, G., et al. (2002). Granular neural web agents for stock prediction. Soft Computing, 6(5), 406-413.

DOI: 10.54254/2754-1169/150/2024.19306

[15] Ma, Y., & Han, R. (2018). Research on stock trading strategy based on deep neural network, International Conference on Control, Automation and Systems (ICCAS), PyeongChang, Korea (South), 2018, pp. 92-96.
[16] Ronil, B., & K. A. S. (2022). Dynamic Black Litterman portfolios with views derived via CNN-BiLSTM predictions. Finance Research Letters, 49.
[17] Man, L., Ye, Z., Yuxin, S., et al. (2022). Clustering-enhanced stock price prediction using deep learning. World Wide Web, 26(1), 21-26.
[18] Agrawal, M., Shukla, K. P., Nair, R., et al. (2022). Stock prediction based on technical indicators using deep learning model. Computers, Materials & Continua, 70(1), 287-304.
[19] Eapen, J., Verma, A., & Bein, D. (2021). Improved big data stock index prediction using deep learning with CNN and GRU. International Journal of Big Data Intelligence, 7(4).
[20] Luyang, C., Markus, P., & Jason, Z. (2023). Deep Learning in Asset Pricing. Management Science. 70(2):714-750.
[21] Harvey, C. R., & Liu, Y. (2019). A census of the factor zoo. SS-RN, 3341728.
[22] Heaton, B. J., Polson, G. N., & Witte, H. J. (2016). Deep learning in finance. CoRR, abs/1602.06561.
[23] Cao, S., Jiang, W., Wang, J., et al. (2024). From man vs. machine to man + machine: The art and AI of stock analyses. Journal of Financial Economics, 160, 103910-103910.