

Intelligent Finance: Emerging Applications and Challenges of Machine Learning in Asset Pricing

Yimin Jiang^{1,a,*}

¹Hefei University of Technology, Hefei City, Anhui Province, 230000, China

a. jym13530150684@163.com

**corresponding author*

Abstract: This paper explores in depth the impact of big data and machine learning on asset pricing. By addressing the limitations of traditional asset pricing models, it analyzes the challenges and opportunities brought by big data to the financial markets. Traditional models, due to their linear assumptions, struggle to handle the complexity of high-dimensional data. However, machine learning techniques, particularly deep learning, overcome this issue with their powerful data processing capabilities, thereby enhancing the model's external validation capacity. This paper summarizes various methods of applying modern machine learning in asset pricing, including feature engineering and end-to-end deep learning, and discusses the adaptive mechanisms of AI technologies to the complexity of asset pricing. Through case analysis, the article also emphasizes the significant effects of AI in improving predictive accuracy and market efficiency. This paper provides a new perspective for asset pricing research and offers practical pathways for the intelligence and dynamic adaptability of financial markets.

Keywords: Asset pricing, Machine learning, Artificial intelligence, Big data

1. Introduction

In modern finance, asset pricing has always been one of the core areas of research. It primarily analyzes the relationship between asset returns and risks through traditional models such as the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT). These classical models rely on linear regression and a limited set of key factors, offering concise and clear mathematical expressions, and have been widely applied in the financial field. However, with the acceleration of globalization and the rapid development of technology, financial markets have become increasingly complex, and the limitations of traditional models have become more evident, as they cannot fully adapt to the changes in today's financial environment.

The emergence of big data, machine learning, and artificial intelligence technologies has provided new perspectives and methodologies for asset pricing and corporate decision-making. These technologies can significantly improve the accuracy of predictions and the effectiveness of decision-making strategies by analyzing complex nonlinear relationships and processing high-dimensional data.

The objective of this paper is to explore how to leverage these advanced machine learning technologies in conjunction with traditional asset pricing models to enhance the accuracy of financial market predictions and the performance of factor investment strategies. Specifically, we categorize

machine learning models into two main types: feature engineering-based methods and end-to-end deep learning methods, and analyze their applications in specific market contexts, as well as their performance in identifying risk premium signals and handling complex data.

Through these in-depth analyses, we aim to reveal the potential of technologies such as machine learning in enhancing predictive accuracy, optimizing portfolios, and explaining financial market behavior, while further exploring the underlying economic theoretical mechanisms. This research not only offers a new research perspective for academia but also provides important guidance for the formulation of strategies in actual financial markets.

2. Limitations of Traditional Asset Pricing Models and Challenges of Big Data

2.1. Limitations of Traditional Asset Pricing Models

In the field of empirical asset pricing, factor pricing models such as the CAPM, APT, and Fama-French models are traditional mainstream tools. The CAPM model, proposed by Sharpe and Lintner, explains asset excess returns based on the market factor [1, 2]. However, its single-factor structure fails to capture the diversity of cross-sectional returns in the market. To address this, researchers have introduced additional factors, leading to the development of multi-factor models. The three-factor model by Fama and French explains stock excess returns through market, size, and value factors [3]. While effective in the U.S. market, it requires momentum, profitability, and investment factors to improve predictive power. In the Chinese market, multi-factor models are also widely used. However, the unique structure and institutional changes of the Chinese stock market affect the stability of these models. Shah et al. [4] found several limitations of traditional asset pricing models in the Chinese market. First, the models fail to account for the special market phenomenon of small-cap companies serving as shell resources, limiting their applicability. Second, traditional models predominantly use the book-to-market ratio, but in the Chinese market, the earnings-to-price ratio more effectively captures value effects. Moreover, the Fama-French model fails to explain various market anomalies, such as profitability and volatility, in the Chinese market, leaving significant unexplained annualized excess returns. This suggests that the model needs to be adjusted for specific markets to improve its explanatory power and effectiveness.

2.2. New Challenges in the Era of Big Data

The era of big data has provided abundant informational resources for asset pricing and corporate decision-making, but it has also brought challenges, particularly in financial markets, where the volume and variety of data are rapidly increasing. This includes structured financial data and unstructured social media sentiment data. This high-dimensional data provides a more comprehensive view of market trends and economic signals. However, traditional linear asset pricing models, due to their reliance on linear assumptions and a limited number of factors, struggle to handle complex data, leading to overfitting and reduced predictive capability. Asset price fluctuations are influenced not only by known factors but also by unseen factors and market changes, exhibiting nonlinear characteristics. To address this challenge, scholars have intensified research into market anomaly factors. Qiao pointed out that among 231 market anomalies, 41 had a significant impact, particularly in the areas of trading friction and value-growth factors [5]. Company fundamentals have expanded the measurement dimensions of market anomalies, becoming an important resource for asset pricing.

However, the vast amounts of multidimensional data generated by modern financial markets also present the challenge of “dimensionality disaster,” which increases computational difficulty and estimation instability, while raising the risks of overfitting and sensitivity to anomalous samples. This, in turn, reduces the model’s generalizability and out-of-sample performance. As a result, financial research has shifted towards using machine learning and artificial intelligence technologies,

incorporating dimensionality reduction, regularization, and feature selection algorithms to optimize high-dimensional pricing models. This enhances predictive power while maintaining the simplicity and interpretability of the models, driving the intelligent development of financial markets.

3. The Impact of Machine Learning on Asset Pricing and Its Mechanisms

Asset pricing evaluates the future value of financial instruments in uncertain contexts, influenced by fundamentals, risk, and market sentiment, and determined by supply and demand relationships. Scholars have explored its patterns from various perspectives, including the random walk theory, efficient market hypothesis, and behavioral finance. The random walk theory posits that market responses are unpredictable, while the efficient market hypothesis suggests that markets can be classified into weak, semi-strong, and strong forms, where stock prices reflect all available information. However, the reversal effect, momentum effect, and size effect challenge the validity of this hypothesis [6]. Behavioral finance emphasizes the impact of investor behavior on prices. The complex and dynamic nature of financial markets has been extensively studied, with analysis methods including both fundamental and technical analysis, leading to the development of the “factor zoo” [4, 7].

The increasing complexity and noise in financial data have made predictions more difficult, rendering traditional linear methods unsuitable. In this context, deep learning and AI technologies, with their multi-layered neural network structures, offer new solutions by automatically learning complex features and nonlinear relationships within the data. Machine learning has made significant breakthroughs in handling big data and is widely applied across various fields, including finance. Its efficiency, adaptability, and ability to process large datasets make it widely applicable in asset pricing. From a mathematical perspective, it maps variable spaces to approximate the true function curve. Compared to traditional methods, nonlinear prediction models can more accurately forecast stock returns, assist in constructing efficient investment strategies, and improve both prediction accuracy and strategy performance.

4. Machine Learning Methods and Categories in Asset Pricing

Asset pricing models hold significant importance in finance, primarily including risk-return trade-off models and price prediction models. Both rely on high-quality structured and unstructured data. Structured data, such as stock prices and trading volumes, help analyze the drivers of price changes, while unstructured data, such as social media text and corporate reports, reflect market sentiment and company performance [8].

In data processing, machine learning models are divided into two categories: feature engineering and end-to-end deep learning. Feature engineering involves dimensionality reduction and feature extraction using techniques such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and Independent Component Analysis (ICA). Additionally, Bayesian algorithms, Markov methods, Support Vector Machines (SVM), decision trees, and random forests are widely used in asset pricing to handle uncertainty and time series prediction. Particle swarm optimization and genetic algorithms are employed for model parameter tuning and strategy optimization. Transfer learning and ensemble learning show potential in the integration of diverse data.

End-to-end deep learning methods, on the other hand, automate feature extraction by using Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) to process time series and image data. Reinforcement learning optimizes dynamic asset allocation, while text analysis and natural language processing enhance information extraction capabilities. Ultimately, model fusion improves prediction accuracy and stability.

5. Machine Learning Methods Based on Feature Engineering

Feature engineering is crucial for improving model prediction accuracy. By combining domain-specific expertise, feature engineering can extract key input features from raw data while minimizing the loss of valuable information, thereby supporting subsequent model applications. The quality of data features directly sets the upper limit of machine learning capabilities. The tasks involved in feature engineering include: data cleaning, noise removal, handling outliers, addressing data imbalance issues, data normalization, discretization, and filling in missing data.

5.1. High-Dimensional Data Dimensionality Reduction Algorithms

In the context of asset pricing, machine learning methods in feature engineering have significantly improved the robustness and interpretability of models through dimensionality reduction and signal decomposition. Techniques like Principal Component Analysis (PCA) simplify the data structure by converting high-dimensional data into linearly uncorrelated principal components. Lettau and Pelger enhanced the identification of weak factors and improved pricing accuracy in high-frequency data environments through Risk Premium PCA (RP-PCA), combining it with Arbitrage Pricing Theory [9]. Kelly et al. used Instrumental PCA (IPCA) to optimize factor load estimates using observable features, boosting the model's predictive ability [10]. Onatski and Ait-Sahalia and Xiu provided new methods for evaluating the performance of principal component estimators and new techniques for covariance matrix decomposition, enhancing systemic risk identification capabilities [11, 12].

Singular Value Decomposition (SVD) decomposes any matrix into the product of three special matrices, showcasing its powerful data processing ability. Gu and Shao used SVD and its improved methods to provide effective tools for market forecasting and short-term asset pricing [13]. Wang introduced adjustable robust estimators to enhance SVD robustness, especially improving data processing efficiency under conditions with outliers [14].

Independent Component Analysis (ICA) estimates the independent non-Gaussian signals of the original data, optimizing investment strategies and risk management. Back and Weigend showed that ICA, in stock portfolio analysis, could identify the key drivers of drastic stock price changes [15]. Taken together, these methods offer more precise market dynamics analysis tools for asset pricing, helping investors optimize investment decisions and enhancing both return potential and risk management capabilities.

5.2. High-to-Low Dimensional Conversion Algorithms

In financial market analysis, the complexity of high-dimensional data and the vast amount of data are simplified and optimized using nonlinear dimensionality reduction techniques such as autoencoders and Support Vector Machines (SVM). Autoencoders create a "bottleneck layer" through deep neural networks to achieve nonlinear compression and reconstruction of data. Gu et al. [16] introduced Conditional Autoencoders, applying this technique to asset pricing and capturing the complex nonlinear relationships between asset characteristics and factor exposures, thereby improving portfolio performance. Suimon et al. used autoencoders to analyze the Japanese government bond yield curve and optimize bond pricing [17]. Huang employed Support Vector Regression (SVR) combined with genetic algorithms for stock selection, improving the efficiency of portfolio construction [18]. On the other hand, research shows that SVM performs excellently in small sample environments. Abedin et al. combined multilayer perceptrons with SVM to optimize credit scoring and bankruptcy prediction [19]. Additionally, SVM has demonstrated its high generalization performance and nonlinear processing capabilities in predicting the NIKKEI 225 index and the trends of the six major Asian stock markets.

These studies demonstrate that through nonlinear dimensionality reduction and feature selection, autoencoders and Support Vector Machines provide powerful tools for identifying complex patterns in financial market data, significantly improving the accuracy and efficiency of asset pricing and offering forward-looking decision support for investors. These methods combine the strengths of novel algorithms and traditional approaches, providing a stronger foundation for market trend analysis and investment decision-making.

5.3. Mathematical Statistical Algorithms

This paper explores the impact of Bayesian classification and Markov algorithms, based on mathematical statistics, in asset pricing. Bayesian classification algorithms address the uncertainty in financial data using Bayes' theorem, enhancing predictive capabilities. Fulop and Yu applied Bayesian learning methods to dynamically adjust asset prices, improving market warning capabilities [20]. Turner used Bayesian analysis to reveal how investor confidence in the Capital Asset Pricing Model (CAPM) affects their investment strategies [21]. Busse and Irvine applied Bayesian methods to improve the accuracy of mutual fund performance forecasts, enhancing financial decision-making information [22].

Markov algorithms, which describe state transition mechanisms, perform excellently in time series analysis. Geweke and Amisano introduced the Hierarchical Markov Normal Mixture Model (HMNM), improving the prediction of asset return volatility, particularly excelling in value-at-risk assessment [23]. Psaradakis et al. presented the Markov Error Correction Model (MEC), which effectively adapts to asset price adjustment processes under different economic environments [24]. These models leverage the "memoryless" characteristic of Markov processes to efficiently capture price dynamics, improving the precision and reliability of market predictions. Markov algorithms, with their robustness and interpretability, provide precise tools for asset pricing, optimizing risk management and market response strategies.

5.4. Traditional Machine Learning Classification Algorithms

Decision trees divide the feature space in an intuitive manner but are prone to overfitting and need pruning to enhance generalization ability. Random forests, as an ensemble algorithm, improve model robustness and accuracy through the combination of multiple decorrelated decision trees, making them especially suitable for handling the complexity and uncertainty of financial markets. Research by Moritz and Zimmermann [25] indicates that random forests can effectively identify key variables, overcoming the limitations of traditional portfolio ranking methods in large datasets, thereby enhancing the accuracy of variable selection. NTI et al. [26] combined random forests with Long Short-Term Memory (LSTM) neural networks to improve the prediction accuracy of macroeconomic variables' impact on liquidity in the Ghanaian stock market, demonstrating its superiority in precisely ranking key variables. Krauss et al. [27] highlighted that random forests show excellent statistical arbitrage potential on S&P 500 index constituents by identifying noise and generating trading signals, particularly during market turbulence. Random forests excel in feature selection and noise resistance, ensuring robust predictions in different market environments, making them well-suited for times of financial crises or market volatility. By combining with other deep learning models and ensemble learning, random forests provide more robust asset pricing strategies, helping investors navigate complex market conditions and improve prediction accuracy.

5.5. Heuristic Algorithms

This paper discusses the application and impact of Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) in asset pricing. PSO, inspired by bird foraging behavior, seeks the optimal solution

in the search space through the sharing of information. Its advantages include simple implementation and rapid convergence, and it can be combined with deep learning techniques to enhance data processing capabilities. Research by Thakkar and Chaudhari shows that PSO exhibits exceptional global optimization capability and adaptability in optimizing stock portfolios and predicting price trends [28]. On the other hand, GA is based on biological evolution theory, optimizing the solution set through selection, crossover, and mutation, making it suitable for capturing nonlinear features of the market. Rather et al. [29] constructed a hybrid forecasting model combining ARIMA, ESM, and RNN, using GA to optimize the weights, making significant progress in handling volatility and nonlinear characteristics of financial data, highlighting the potential of multi-model integration in asset pricing.

The combination of PSO and GA demonstrates an enhanced effect in the analysis of complex financial data. PSO's rapid convergence and GA's strong global search capability complement each other, improving the model's robustness and adaptability to dynamic markets, providing smarter and more accurate analytical tools for asset pricing. This integrated approach not only optimizes model parameters but also significantly improves the flexibility and accuracy of investment decisions.

6. End-to-End Deep Learning Methods

End-to-end methods demonstrate significant advantages in data processing by automatically extracting features from raw data without requiring deep prior knowledge. By decomposing complex deep features into simple hierarchical feature representations, they achieve effective matrix representations of multiple data sources.

The merging of feature vectors into a unified input enriches the data sources and enhances model adaptability. Deep learning is renowned for its ability to extract complex features and perform nonlinear fitting, and through end-to-end models, the design is simplified, making it especially advantageous in financial big data and high-frequency data analysis.

In the face of the ever-expanding scale of financial data, end-to-end methods support automatic learning and optimization of the CNN decision-making process. Next, we will explore several typical end-to-end machine learning methods, including CNN, LSTM, reinforcement learning.

6.1. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN), as a powerful structured feedforward neural network, extract features and capture complex nonlinear relationships through convolutional layers. They are widely applied in the financial field for asset pricing. For example, Gudelek et al. [30] used CNN to predict the price volatility of Exchange-Traded Funds (ETFs). By generating image snapshots of market activities and inputting them into the CNN, they automatically learned the complex price patterns in the market. This approach integrated trend, momentum, and fundamental analysis indicators, creating a diversified dataset that significantly improved the model's generalization ability, demonstrating CNN's potential in financial time series forecasting. On the other hand, the CNN framework designed by Hoseinzadeh and Haratizadeh employed three-dimensional tensor analysis to examine the correlations between different financial markets [31]. This method revealed latent inter-market relationships, enabling the consideration of information from other markets when predicting one, thus improving prediction accuracy and reliability. The framework provides effective tools to help investors make more informed decisions in complex market environments. Additionally, the LSTM-CNN feature fusion model proposed by Kim et al. [32] combined the strengths of LSTM and CNN, enhancing the accuracy of stock price prediction. The model leveraged LSTM's ability to process time series data and CNN's advantages in extracting image features to analyze complex patterns in stock charts. Through feature fusion, it significantly reduced prediction errors, demonstrating the

potential of ensemble learning in asset pricing. Overall, these studies indicate that CNN holds significant advantages in capturing complex financial patterns and optimizing asset pricing, playing a key role in advancing financial market forecasting.

6.2. Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory (LSTM) networks address the vanishing gradient problem in traditional Recurrent Neural Networks (RNN) when processing long sequences of non-stationary data by introducing memory components and three key gate structures (forget gate, input gate, and output gate). This enhances the ability to capture short-term fluctuations and long-term trends. In the financial field, LSTM effectively handles complex time series data, improving the accuracy of price predictions and risk management. The online LSTM model proposed by Borovkova and Tsiamas [33], which integrates numerous technical analysis indicators and utilizes LSTM's long-term dependencies along with an online weighting mechanism, successfully improved the prediction accuracy of high-frequency stock market data, outperforming traditional models. This model structure enhanced the adaptability and predictive flexibility of LSTM in dynamic markets. Yildiz et al. [34] demonstrated the innovative application of LSTM in predicting monthly stock closing prices by combining strategies such as smart-beta portfolios. By leveraging LSTM's ability to remember both short- and long-term data, precise predictions were made, providing a basis for investment decisions, optimizing risk management, and enhancing investment returns. These research findings suggest that LSTM holds significant advantages in capturing complex financial patterns and its application in asset pricing. However, LSTM also faces challenges, such as the complexity of its internal structure, which makes interpretation difficult, and issues with adapting to dynamic market changes. Consequently, researchers are exploring the integration of new methods like reinforcement learning to improve the reliability of the model and its adaptability to market changes. Overall, LSTM's application in financial forecasting has provided new directions for asset pricing theory and has contributed to the robustness of economic decision-making.

6.3. Reinforcement Learning

Reinforcement learning optimizes strategies to maximize cumulative rewards through the interaction between an agent and a dynamic environment. Deep Reinforcement Learning (DRL), which combines deep learning techniques, is capable of handling high-dimensional state spaces, significantly enhancing asset pricing and investment decisions in the financial field. Chen and Hsieh [35] used the three-parameter Roth-Erev model to reveal how the heterogeneity of traders' psychological traits influences market strategies and levels of return, highlighting the application of reinforcement learning in understanding the diversity of market behavior. Cao et al. [36] optimized derivative hedging strategies through reinforcement learning, achieving efficient management of hedging costs and volatility, demonstrating adaptability in complex market environments. The DeRecv model developed by Cao and Zhai [37] combined deep learning with econometrics to accurately predict market price trends and provide refined analytical tools for trading behavior. Lee et al. [38] combined Deep Q Networks and Convolutional Neural Networks to process stock images, demonstrating exceptional forecasting ability in multinational markets and proving the potential of reinforcement learning in global financial markets. Overall, reinforcement learning surpasses traditional models by directly optimizing investment returns and addressing market uncertainty. Despite challenges such as high data requirements and complex training, its dynamic decision-making ability and flexible adaptation to market changes have had a transformative impact on asset pricing.

6.4. Summary

End-to-end deep learning has demonstrated significant advantages in asset pricing. By automating feature extraction and refining complex features from multi-source heterogeneous data, it eliminates the need for extensive domain expertise, enhancing model adaptability and application breadth. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, in particular, effectively capture complex patterns and dynamic changes in market data through nonlinear function fitting and deep feature analysis. These characteristics make them especially well-suited for handling big financial data and high-frequency data, providing accurate predictions. Compared to traditional statistical methods, end-to-end deep learning models convert complex financial data into easy-to-understand representations by expressing features layer by layer, reducing data dimensionality complexity and enhancing information processing efficiency.

Moreover, this approach excels in multi-model integration and the combination of different data features, providing a comprehensive solution for asset pricing. Although research is still in its early stages, facing challenges such as data diversity and inconsistent evaluation standards, the growing application trend and technological innovation demonstrate its development potential and profound impact on financial market research, laying the technological foundation for intelligent financial decision-making and automated market analysis.

7. Impact of AI on Asset Pricing and Its Mechanisms

Artificial Intelligence (AI) is revolutionizing asset pricing methods by analyzing historical market data, macroeconomic indicators, news events, and social media sentiment to uncover potential pricing signals, thereby improving the accuracy of asset price predictions. Donata and Fabrizio [39] explored the application of AI in high-frequency trading, combining simplified AI methods to enhance prediction accuracy and profitability, and using reinforcement learning to improve the effectiveness of trading strategies. The research by Barboza et al. [40] shows that generative AI significantly increases the accuracy of financial asset valuation, performing excellently in routine predictions. AI enhances risk management and market efficiency by automating trading and utilizing behavioral finance analysis tools, which improve market response speed and pricing efficiency. However, the application of AI faces challenges such as model transparency and data privacy, requiring a balance between innovation and market stability to ensure its sustainable application.

8. Conclusion

The potential of machine learning in asset pricing is immense, and its deep integration with modern information technology could lead to significant innovations. The future research directions and their potential challenges are as follows:

(1) **Model Scalability:** Although deep learning algorithms such as CNN and LSTM have been successful in image and natural language processing, they still have limitations in asset pricing applications. This is mainly due to their complexity, which may lead to the extraction of too many deep features that may not effectively reflect the essential principles of financial markets. Therefore, these models need to be optimized and adjusted to better adapt to the characteristics of financial data, thereby playing a larger role in asset pricing.

(2) **Data Availability and Integration Challenges:** The data available in financial markets have a relatively limited time span, particularly non-structured and short-sample data, such as social media content, which poses an obstacle to extensive machine learning training. Methods like transfer learning may offer solutions by integrating multiple data sources to enhance model effectiveness. Additionally, models need to consider real-world trading constraints, such as liquidity and transaction costs, to ensure their reliability and adaptability in practical applications.

(3) Model Interpretability: While improving model accuracy, maintaining its interpretability is crucial. Traditional methods, such as linear regression, are easy to interpret but cannot handle complex data relationships, while deep learning models can capture deep relationships but often lack transparency. Therefore, combining technologies like knowledge graphs to enhance model interpretability and make the reasoning process more transparent will be an important direction to increase the trustworthiness of investment decisions.

(4) Model Generalization and Adaptability: The dynamic and uncertain nature of financial markets places high demands on the generalization ability of machine learning models. As information dissemination accelerates, models not only need to quickly adapt to market changes but also must have robust adaptability to cope with market stage transitions. This involves how to effectively extract key features from data and form robust out-of-sample predictive capabilities, which is a core challenge in model development and application.

Through in-depth exploration of these research directions and technological innovations, the application of machine learning in asset pricing holds vast potential, and will bring revolutionary changes to financial market analysis and decision-making.

References

- [1] Sharpe, W. F. (1964). *Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19(3), 425–442.
- [2] Lintner, J. (1965). *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*. *The Review of Economics and Statistics*, 47(1), 13–37.
- [3] Fama, E. F., & French, K. R. (1993). *Common risk factors in the returns on stocks and bonds*. *Journal of Financial Economics*, 33(1), 3–56.
- [4] Shah, D., Isah, H., & Zulkernine, F. (2019). *Stock market analysis: A review and taxonomy of prediction techniques*. *International Journal of Financial Studies*, 7(2), 26. <https://www.mdpi.com/2227-7072/7/2/26>
- [5] Qiao, F. (2019). *Replicating anomalies in China*. SSRN. <https://ssrn.com/abstract=3263990>
- [6] Lo, A. W., & Mackinlay, A. C. (1988). *Stock market prices do not follow random walks: Evidence from a simple specification test*. *The Review of Financial Studies*, 1(1), 41–66.
- [7] Harvey, C. R., & Liu, Y. (2019). *A census of the factor zoo*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3341728
- [8] Ming, F., & Stephen, T. (2021). *A machine learning based asset pricing factor model comparison on anomaly portfolios*. *Economics Letters*, 204.
- [9] Lettau, M., & Pelger, M. (2020). *Factors that fit the time series and cross-section of stock returns*. *The Review of Financial Studies*, 33(5), 2274–2325.
- [10] Kelly, B. T., Pruitt, S., & Su, Y. (2019). *Characteristics are covariances: A unified model of risk and return*. *Journal of Financial Economics*, 134(3), 501–524.
- [11] Onatski, A. (2012). *Asymptotics of the principal components estimator of large factor models with weakly influential factors*. *Journal of Econometrics*, 168(2), 244–258.
- [12] Aït-Sahalia, Y., & Xiu, D. (2017). *Using principal component analysis to estimate a high-dimensional factor model with high-frequency data*. *Journal of Econometrics*, 201(2), 384–399.
- [13] Gu, R., & Shao, Y. (2016). *How long the singular value decomposed entropy predicts the stock market? Evidence from the Dow Jones Industrial Average Index*. *Physica A: Statistical Mechanics and Its Applications*, 453, 150–161.
- [14] Wang, D. (2017). *Adjustable robust singular value decomposition: Design, analysis, and application to finance*. *Data*, 2(3), 29.
- [15] Back, A. D., & Weigend, A. S. (1997). *A first application of independent component analysis to extracting structure from stock returns*. *International Journal of Neural Systems*, 8(4), 473–484.
- [16] Gu, S., Kelly, B., & Xiu, D. (2020). *Autoencoder asset pricing models*. *Journal of Econometrics*. <https://sciencedirect.com/science/article/pii/S0304407620301998>
- [17] Suimon, Y., Sakaji, H., Izumi, K., et al. (2020). *Autoencoder-based three-factor model for the yield curve of Japanese government bonds and a trading strategy*. *Journal of Risk and Financial Management*, 13(4), 82.
- [18] Huang, C. F. (2012). *A hybrid stock selection model using genetic algorithms and support vector regression*. *Applied Soft Computing*, 12(2), 807–818.

- [19] Chen, W. H., Shih, J. Y., & Wu, S. (2006). Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *International Journal of Electronic Finance*, 1(1), 49-67.
- [20] Fulop, A., & Yu, J. (2017). Bayesian analysis of bubbles in asset prices. *Econometrics*, 5(4), 47. <https://www.mdpi.com/2225-1146/5/4/47>.
- [21] Turner, J. A. (2010). Momentum portfolios and the capital asset pricing model: A Bayesian approach. *Quarterly Journal of Finance and Accounting*, 43, 43-59. <https://www.jstor.org/stable/23074629>
- [22] Busse, J. A., & Irvine, P. J. (2006). Bayesian alphas and mutual fund persistence. *The Journal of Finance*, 61(5), 2251-2288.
- [23] Geweke, J., & Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1), 1-29.
- [24] Psaradakis, Z., Sola, M., & Spagnolo, F. (2004). On Markov error correction models, with an application to stock prices and dividends. *Journal of Applied Econometrics*, 19(1), 69-88.
- [25] Moritz, B., & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740751.
- [26] Nti, K. O., Adekoya, A., & Weyori, B. (2019). Random forest based feature selection of macroeconomic variables for stock market prediction. *American Journal of Applied Sciences*, 16(7), 200-212.
- [27] Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.
- [28] Thakkar, A., & Chaudhari, K. (2020). A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization. *Archives of Computational Methods in Engineering*, 1-32. <https://link.springer.53yu.com/>.
- [29] Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, 42(6), 3234-3241.
- [30] Gudelek, M. U., Boluk, S. A., & Ozbayoglu, A. M. (2017). A deep learning-based stock trading model with 2-D CNN trend detection. In *2017 IEEE Symposium Series on Computational Intelligence* (pp. 1-8). IEEE.
- [31] Hoseinzade, E., & Haratizadeh, S. (2018). CNNPred: CNN-based stock market prediction using several data sources. *arXiv*, 1810.08923.
- [32] Kim, T., & Kim, H. Y. (2019). Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PLOS ONE*, 14(2), e0212320. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0212320>.
- [33] Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting*, 38(6), 600-619.
- [34] Yildiz, Z. C., & Yildiz, S. B. (2020). A portfolio construction framework using LSTM-based stock market forecasting. *International Journal of Finance & Economics*. <https://online-library.wiley.com/doi/abs/10.1002/ijfe.2277>.
- [35] Chen, S. H., & Hsieh, Y. L. (2011). Reinforcement learning in experimental asset markets. *Eastern Economic Journal*, 37(1), 109-133.
- [36] Cao, J., Chen, J., Hull, J., et al. (2021). Deep hedging of derivatives using reinforcement learning. *arXiv*, 2013.16409.
- [37] Cao, Y., & Zhai, J. (2020). Estimating price impact via deep reinforcement learning. *International Journal of Finance & Economics*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijfe.2353>.
- [38] Lee, J., Kim, R., Koh, Y., et al. (2019). Global stock market prediction based on stock chart images using deep Q-network. *IEEE Access*, 7, 167260-167277.
- [39] Donata, P., & Fabrizio, C. (2021). Artificial intelligence methods applied to financial assets price forecasting in trading contexts with low (intraday) and very low (high-frequency) time frames. *Strategic Change*, 30(3).
- [40] Barboza, F., Silva, N. G., & Fiorucci, A. J. (2023). A review of artificial intelligence quality in forecasting asset prices. *Journal of Forecasting*, 42(7)./F.2024.04.199.