Financial Innovation

Dynamic Factor-Informed Reinforcement Learning for Enhancing Portfolio Optimization

--Manuscript Draft--

Manuscript Number:	FINI-D-24-00878R1
Full Title:	Dynamic Factor-Informed Reinforcement Learning for Enhancing Portfolio Optimization
Article Type:	Methodology
Funding Information:	National Research Foundation of Korea (RS-2023-00220762) — Prof. Ha Young Kim National Research Foundation of Korea (2023R1A2C200337911) — Prof. Ha Young Kim
Abstract:	Portfolio optimization is essential for investors seeking to manage risk, diversify assets, and maximize returns. Although recent studies have primarily focused on enhancing technical aspects such as model architecture through the application of deep learning or reinforcement learning, knowledge of factor portfolios, grounded in modern portfolio theory, remains paramount. Therefore, to effectively utilize the knowledge of factor investment strategies, we propose a novel hybrid portfolio investment method that integrates reinforcement learning with dynamic factors, called the dynamic factor portfolio model. The dynamic factors encompass five fundamental factors: size, value, beta, investment, and quality. The proposed model comprises two modules: a dynamic factor module that calculates a score based on factors reflecting the macro market and a price score module that calculates a score based on prices expressing the relationship between assets and their future value. With dynamic factor-informed knowledge, the proposed model can make portfolio decisions adaptively based on market conditions. Through comprehensive experiments, we validate the effectiveness of the dynamic factor module and demonstrate that the proposed model outperforms both traditional portfolio investment strategies and existing reinforcement learning-based strategies. Moreover, the proposed model offers interpretability by identifying critical factors across varying market scenarios, thereby enhancing portfolio management practices.
Corresponding Author:	Ha Young Kim, Ph.D. Yonsei University Seoul, KOREA, REPUBLIC OF
Corresponding Author E-Mail:	HAYOUNG.KIM@YONSEI.AC.KR
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Yonsei University
Corresponding Author's Secondary Institution:	
First Author:	Sungsoo Kim
First Author Secondary Information:	
Order of Authors:	Sungsoo Kim Hyukjae Kwon Jisang Yoon Ha Young Kim, Ph.D.
Order of Authors Secondary Information:	
Author Comments:	Dear Gang Kou, PhD Financial Innovation I hope this message finds you well. First, I would like to sincerely thank you for facilitating the review process for our manuscript. Peer feedback is invaluable, and we deeply appreciate the time and effort reviewers dedicate to evaluating submissions.

	However, I would like to express some concerns regarding the feedback provided by Reviewer #2. While we fully acknowledge that critical feedback is essential for improving the quality of a manuscript, the nature of the comments suggests a fundamental misunderstanding of the manuscript's content, which may stem from a lack of familiarity with the specific intersection of financial theory, deep learning, and reinforcement learning that our research explores. Reviewer #2 openly admits in their review: "With all my respect to the authors, I sadly must admit that this paper is completely uncomprehensible to me... I have no possibility to understand the remaining part of the manuscript, which I read and did not fully understand." This admission raises concerns about the reviewer's ability to provide a fair and comprehensive assessment. The reviewer seems unfamiliar with core concepts in our work, such as factor investing, the use of deep learning in financial applications, and the theoretical underpinnings of reinforcement learning. For instance, the reviewer notes: "The same problem arises with the subsequent section, 'Dynamic Factor Module.' Here the problem seems to be the computation of the factor importance weights calculated from five factors, which should be value, size, beta, quality, and investment. To be frank, I do not understand the meaning of them but only can have a vague idea of what they may represent. What is the exact meaning of value? Could you define the size? What do you mean by quality?" These concepts—value, size, and quality—are standard in financial literature and are defined by widely accepted methodologies such as Fama-French factor models. This lack of familiarity significantly undermines the reviewer's ability to engage critically with the manuscript's methodology. Additionally, Reviewer #2 expresses confusion regarding standard reinforcement learning assumptions, such as the use of a Markovian framework for state variables, and suggests highly complex probabilistic modeling requirements that go beyond the scope of most reinforcement learning applications in financial portfolio management. This further indicates a disconnect between the reviewer's expertise and the research domain. Finally, the review includes broad statements such as: "The logical steps are not clear. The authors use a lot of acronyms to say they are doing something that is not explained at all... In summary, it is a kind of black box-producing result that requires an act of faith to believe in it." These comments are subjective and lack specific, actionable feedback. While constructive criticism is crucial, remarks such as these, without clear references to specific issues or actionable suggestions, provide limited value in improving the manuscript. Given these points, I would respectfully request reconsideration of Reviewer #2's evaluation. A review by someone with expertise in both financial theory and advanced machine learning methodologies (including deep learning and reinforcement learning) would provide more meaningful feedback. This would ensure a fair and constructive review process, ultimately benefiting the manuscript and the journal. Thank you for your understanding and for your continued efforts in managing this process. I am happy to address any additional concerns or provide further clarifications regarding our manuscript. Best regards, Sungsoo Kim; Hyukjae Kwon; Jisang Yoon; Ha Young Kim Graduate School of Information, Yonsei University
Response to Reviewers:	We kindly request all reviewers to refer to the supplementary document titled "Detailed Response to Reviewer Comments", which provides detailed explanations of the revisions made in response to your valuable feedback. This document includes

comprehensive information on updates to figures, tables, and manuscript text, along with justifications for each change. We sincerely hope this additional document offers clarity and addresses your insightful comments thoroughly.

1.Responses to the Reviewer #1
Comment #1
How does the dynamic factor scoring model adapt to sudden market shocks or unprecedented events that were not present in the training data?

Response #1)
Thank you for your insightful comment regarding the dynamic factor scoring model's adaptability to sudden market shocks or unprecedented events that were not part of its training data. To address this concern, we used the COVID-19 pandemic as a case study to evaluate the proposed model's performance. Specifically, we analyzed its behavior during two critical periods: the sharp market downturn from February 1, 2020, to March 31, 2020, and the subsequent recovery phase from April 1, 2020, to June 30, 2020.
During the market downturn, when the NASDAQ index declined by 30.11%, the DFPM exhibited exceptional adaptability. As summarized in Table 1, DFPM achieved the highest fAPV (1.1390), highest Sharpe ratio (0.1722), and competitive MDD (12.83%). In comparison, benchmark strategies such as MVO, Uniform Allocation, Factor Separate, and Factor Mix exhibited significantly higher drawdowns (ranging from 9.11% to 11.32%) and lower Sharpe ratios, with some even producing negative values. These results underscore DFPM's ability to dynamically adjust portfolio compositions through real-time factor updates, ensuring stability during extreme volatility.
The subsequent recovery period provided additional insights into DFPM's robustness. Between April 1, 2020, and June 30, 2020, the NASDAQ index rebounded by 27.34%. As shown in Table 2, DFPM outperformed the benchmark strategies, achieving the highest fAPV (1.1200) and Sharpe ratio (3.4969), with an MDD of 0, indicating comprehensive risk management. In contrast, traditional models, such as MVO and Uniform Allocation exhibited inferior metrics, while factor-based models, such as Factor Separate and Factor Mix lacked the dynamic adaptability exhibited by DFPM.
These findings validate DFPM's ability to adapt to unprecedented market events by leveraging dynamic factor scoring and real-time updates, thereby establishing it as a robust tool for portfolio optimization under volatile conditions. To highlight these results, an experimental section was included in the revised manuscript. This section provides a comprehensive analysis of DFPM's performance during both the COVID-19 market downturn and recovery phases, emphasizing its practical effectiveness in navigating dynamic and unpredictable market environments.

Revised manuscript: Cross-Validation Results Analysis section 2-5 paragraphs (page 35)

Comment #2
What are the limitations of using the TA-LSTM architecture in the dynamic factor module, and how might these limitations be addressed in future research?

Response #2)
Thank you for your valuable feedback on the limitations of the Temporal Attention-LSTM (TA-LSTM) architecture in the DFM. We acknowledge the challenges associated with TA-LSTM, particularly its computational demands and susceptibility to overfitting. The combined processing of temporal dependencies and attention scores in the TA-LSTM model increases computational complexity, particularly when processing large-scale datasets. To address this issue, future research could explore efficient attention mechanisms or dimensionality reduction techniques to optimize resource usage without compromising model performance. The model's complexity can lead to overfitting, especially in volatile markets. To mitigate this issue, we plan to incorporate regularization techniques and ensemble methods to improve model generalizability across diverse market conditions. These adjustments will enhance TA-LSTM's scalability and adaptability, further reinforcing its application in financial portfolio models. The conclusion section of the manuscript has been revised to incorporate these insights and outline the proposed future directions for improving the TA-LSTM architecture in the DFM.

Revised manuscript: Conclusion section fifth paragraph (page 42)

Comment #3
What are the practical challenges of implementing this model in real-time trading environments, and how can these challenges be mitigated?

Response #3)
Thank you for your insightful comment on the practical challenges of implementing DFPM in real-time trading environments. We acknowledge the importance of addressing these challenges to ensure the model's applicability and effectiveness. Transaction costs and slippage are significant challenges that affect real-world trading performance. While transaction costs are explicitly incorporated into the model, slippage—particularly in volatile or illiquid markets—remains a factor that can erode returns. To address this issue, future research will integrate slippage models into the training process and evaluate the DFPM under simulated market conditions that account for execution risks, thereby improving its robustness in real-world applications. Another key challenge is adapting to sudden market shocks in a highly dynamic and unpredictable financial landscape. Although the DFPM's dynamic factor scoring provides an inherent advantage in adjusting to changing market conditions, extreme market events can render past patterns less relevant. Future research will explore incorporating online learning methods and more frequent retraining to enhance the model's ability to respond effectively to abrupt market shifts. Finally, model complexity and interpretability pose practical concerns in real-time environments. Although DFPM's dual-module structure is robust, it increases the difficulty of interpreting investment decisions in real time. Future research will employ advanced interpretability methods, such as Shapley Additive Explanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME), to provide greater transparency into the model's decision-making processes. This approach will allow investors to better understand the contributions of factors and price components, fostering trust and usability in practical applications. Addressing these challenges will be a key focus of future research to ensure that DFPM evolves into a more robust, adaptive, and practical tool for real-time trading. The conclusion section of the revised manuscript has been updated to incorporate these practical challenges and outline the proposed directions for future research. Thank you for your valuable feedback, which has helped shape these considerations.

Revised manuscript: Conclusion section fifth paragraph (page 42)

Comment #4
How does the proposed model compare with other hybrid models that might integrate different machine learning techniques (e.g., reinforcement learning combined with genetic algorithms)?

Response #4)
Thank you for your comment regarding the comparison of the proposed DFPM to other hybrid models that integrate machine learning techniques. DFPM stands out by combining dynamic factor modeling with reinforcement learning (RL), achieving a unique balance of adaptability, robustness, and interpretability that traditional machine learning-based hybrid models cannot achieve.
Traditional hybrid models often integrate machine learning techniques, such as support vector machines (SVM), decision trees, and DL models, with financial data to predict asset prices or optimize portfolios. Although these models excel at leveraging complex patterns in historical data, they often lack the ability to dynamically adapt to real-time market conditions. For instance, supervised learning models typically rely on static relationships between input features and outcomes, which may lose relevance during periods of market volatility or structural changes. In contrast, DFPM dynamically integrates financial factors—size, value, and quality—within an RL framework, enabling continuous learning and real-time portfolio adjustments in response to changing market conditions. In contrast to traditional hybrid models that treat financial factors as static inputs, DFPM uses a DFM to compute real-time factor scores that reflect evolving market dynamics. This capability ensures that DFPM is aligned with well-established investment principles while adapting to current market trends.
Compared to hybrid models that integrate machine learning with heuristic approaches, such as mean-variance optimization or risk-parity strategies, DFPM advances portfolio

optimization by leveraging RL learning to dynamically weight and integrate both macroeconomic trends (via the DFM) and asset-specific price information (via the PSM). This approach allows DFPM to achieve a balance between risk management and return optimization, which is a challenge that static models or supervised learning-based approaches often struggle to maintain. As shown in Table 8 of the manuscript, the proposed DFPM consistently outperformed traditional hybrid models across various metrics. For example, DFPM achieved a significantly lower MDD of 6.78% compared to 8.29% for an advanced supervised learning-based model, a higher Sharpe ratio of 2.4479 compared to 2.0067, and a superior factor-adjusted portfolio value (fAPV) of 2.2090 compared to 1.9270. These results underscore DFPM's ability to adapt to dynamic market conditions while maintaining robust performance, which distinguishes it from models that rely solely on static or supervised learning methods.
By integrating machine learning techniques with dynamic factor modeling and RL, the proposed DFPM addresses the limitations of traditional hybrid approaches, providing a more versatile and practical solution for portfolio optimization in dynamic market environments. This discussion has been included in the revised manuscript.

Revised manuscript: Related work section first paragraph (page 6)

Comment #5
Are there specific market conditions or asset classes where the proposed model performs particularly well or poorly? If so, what might explain these variations in performance?

Response #5)
Thank you for your comment regarding DFPM's performance across different market conditions and asset classes. Rather than excelling in specific scenarios, DFPM demonstrates consistent performance across various environments due to its dynamic adaptability. The model's dual-module structure—integrating the DFM and PSM—allows it to balance long-term macroeconomic trends with short-term price dynamics. During market downturns, DFPM prioritizes stability-oriented factors, such as quality and size, while in recovery- or bullish markets, it shifts its focus to growth-oriented factors, such as value and investment. The empirical results indicate that DFPM delivered robust performance under both volatile and stable conditions, consistently outperformed benchmarks, and achieved stable risk-adjusted returns. This adaptability makes DFPM a reliable and versatile tool for portfolio managers seeking robust optimization tools across diverse market scenarios.

2.Responses to the Reviewer #2
Comment #1
The logical steps are not clear. The authors use a lot of acronyms to say they are doing something that is not explained at all. I have serious doubts that a reader can benefit from reading this manuscript. In summary, it is a kind of black box-producing result that requires an act of faith to believe in it.
For example, the section called "Dynamic Factor Portfolio Model" is a kind of storytelling, and after reading it, I did not understand anything about the DFPM methodology. The process is illustrated in Figure 1, and no comment is made to this figure and there is no possibility of understanding it.

Response #1)
Thank you for highlighting the lack of explanation for Figure 1. We have addressed this issue by revising the caption and adding a detailed paragraph in the revised manuscript that explicitly describes the process depicted in the figure. The revised manuscript guides readers through each component of Figure 1, explaining the interaction between the DFM and PSM and their contribution to the overall portfolio optimization process. This explanation ensures that Figure 1 is no longer standalone but is fully integrated into the manuscript's narrative, providing readers with a comprehensive understanding of its relevance to the DFPM methodology. We appreciate your feedback, which has significantly improved the coherence and clarity of this section.

Revised manuscript: Dynamic factor section 1-3 paragraphs (page 8)

Comment #2

The same problem arises with the subsequent section, "Dynamic Factor Module." Here the problem seems to be the computation of the factor importance weights calculated from five factors, which should be value, size, beta, quality, and investment. To be frank, I do not understand the meaning of them but only can have a vague idea of what they may represent. What is the exact meaning of value? Could you define the size? What do you mean by quality? In this section, you introduce, without any motivation or explanation, some formulas. Hence, after reading this section, I did not understand the idea or the methods.

Response #2)
Thank you for emphasizing the need for clearer definitions and explanations in the “Dynamic Factor Module” section. We appreciate your constructive feedback regarding the computation of factor importance weights and the interpretation of the five factors: value, size, beta, quality, and investment. In the revised manuscript, we have provided detailed clarifications for these factors:
Value: Represents the relative valuation of an asset based on fundamental metrics, such as price-to-earnings (P/E) or price-to-book (P/B) ratios. A lower P/E or P/B ratio generally indicates undervaluation.
Size: Reflects the market capitalization of an asset, distinguishing between small-cap and large-cap stocks. Smaller assets are often associated with higher growth potential but increased risk.
Beta: Indicates an asset's sensitivity to market movements, reflecting its systematic risk. A beta greater than 1 indicates higher volatility than the market, while a beta less than 1 indicates lower volatility.
Quality: Encompasses financial health indicators, such as profitability, earnings stability, and debt levels. High-quality stocks tend to perform well during market downturns due to their resilience.
Investment: Captures the asset's growth in capital expenditure or reinvestment rates, often linked to future growth potential.
We have included a detailed explanation of the motivation behind each factor and described their use in computing importance weights within the DFM. The model assigns weights dynamically based on the prevailing market regime and macroeconomic conditions, enabling effective portfolio allocation adjustments.
Additionally, we acknowledge that some formulas in the original manuscript did not provide sufficient explanation or context. In the revised manuscript, we have elaborated on the rationale behind each formula, detailing its role in the computation of factor importance weights and its contribution to the dynamic scoring process. These updates make the "Dynamic Factor Module" section more accessible and comprehensible.

Revised manuscript: Dynamic factor module section first paragraph (page 10)

Comment #3
Section "Model optimization": The state variables are prices of assets, factors, macroeconomic data. So, do you have a probabilistic description of the joint distribution of all these variables conditionally on their values one step before? This seems already a very complex task. The description of the dynamic of a single financial asset has been recognized to be non-Markovian by a large literature depending on the time scale considered. Thus, how can you consider the description of such a huge economic and financial system as a Markovian model? What is the fitting distribution (depending on the state of the system)? How do you estimate it? Why, you are considering that this overall process is Markovian? All these aspects must be explained, checked and verified statistically and not given as a fact.

Response #3)
Thank you for highlighting the critical points regarding the probabilistic modeling and assumptions of the Markovian property in the “Model Optimization” section. We acknowledge the inherent complexity of describing the dynamics of financial and economic systems, particularly given the well-documented non-Markovian nature of financial assets in the literature.
In the DFPM context, the joint probability distribution of all state variables was not explicitly modeled. Instead, dynamic factor modeling was employed to capture temporal dependencies and interactions between variables without relying on a fully explicit probabilistic framework for the entire system. DFPM uses factor scores derived

from macroeconomic and financial data to dynamically adjust portfolio allocations, indirectly reflecting the complex dependencies among state variables. Regarding the Markovian assumption, we emphasize that many studies on portfolio optimization through reinforcement learning adopt the Markovian assumption as a simplified framework (e.g., Xie et al., 2020; Deng et al., 2017). Although financial systems often exhibit non-Markovian behavior, these studies have demonstrated through extensive experiments that the Markovian assumption can yield effective results. Similarly, the DFPM employs the Markovian framework to model state transitions and integrates a dynamic factor scoring mechanism to incorporate real-time updates and adapt to changing market conditions. This combination ensures the practical validity of the reinforcement-learning-based approach, as evidenced by the multiple experimental results. To address concerns regarding fitting distributions and estimation methods, the proposed DFPM does not rely on a fixed probabilistic distribution of the system's state. Instead, it continuously learns factor weights and adjusts allocations based on observed data, ensuring responsiveness to current market conditions.

References

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), 653–664. https://doi.org/10.1109/TNNLS.2016.2522401

Xie, S., Ding, L., Wang, J., & Pan, W. (2020). Deep reinforcement learning for dynamic portfolio optimization. Quantitative Finance, 20(4), 541–552. https://doi.org/10.1080/14697688.2019.1675466

Mastrogiovanni, M. (2024). Reinforcement learning applied to dynamic portfolio management: An article review. SSRN. https://doi.org/10.2139/ssrn.4981709

Response #4)

Thank you for your valuable suggestion. We agree that including a concise summary of the empirical and experimental results in the introduction would enhance the paper's clarity and impact. In response to your feedback, we have revised the introduction to include a summary of the key findings from our empirical and experimental analyses. This summary highlights the performance improvements achieved by the proposed model compared to traditional portfolio strategies and other reinforcement learning approaches, focusing on metrics such as risk-adjusted returns and stability under various market conditions. By integrating these results into the introduction, we aim to provide readers with a clear understanding of the practical benefits and significance of our approach from the outset.

Revised manuscript: Introduction section eighth paragraph (page 4)

Comment #5

Some of the references are unpublished manuscripts that were written some years ago. These studies have not received a review report and have not been validated by peers; hence, they have limited validity that should still be checked.

Response #5)

Thank you for your valuable feedback regarding the inclusion of unpublished manuscripts. After reviewing the references, we have taken the following actions to address your concerns:

The following two references, which were unpublished manuscripts, have been removed from the manuscript:

Zhang Z, Zohren S, Roberts S (2020): "Deep learning for portfolio optimization." arXiv preprint arXiv:2005.13665.

Jiang Z, Xu D, Liang J (2017): "A deep reinforcement learning framework for the financial portfolio management problem." arXiv preprint arXiv:1706.10059.

To improve the credibility and reliability of the references, we replaced two other unpublished manuscripts with peer-reviewed literature:

Huang K, Wang F (2020): "Comprehensive Review on Attention Mechanism in Neural Networks." has been replaced with Zhang T, Wang Q, Zhang P, Wang Y (2022): "A review on the attention mechanism of deep learning." Neurocomputing, 509:58–78. https://doi.org/10.1016/j.neucom.2022.07.013.

Bahdanau D, Cho K, Bengio Y (2014): "Neural machine translation by jointly learning to align and translate." has been replaced with Vaswani A, Shazeer N, Parmar N,

Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017): "Attention is all you need." Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 5998–6008.
These changes ensure that the references included in the manuscript are peer-reviewed and contribute to enhancing the overall validity and reliability of the study.

3.Responses to the Reviewer #3
Comment #1
The introduction of this paper talks about investment theory and deep learning respectively, but does not mention how to connect the two. Although it mentions the importance of portfolio optimization and briefly outlines the modern portfolio theory (MPT), capital asset pricing model (CAPM) and arbitrage pricing theory (APT), it does not fully connect these theories with the combination of deep learning (DL) and reinforcement learning (RL). The article simply mentions the application of deep learning and reinforcement learning in the financial field in recent years but does not explain why they need to be combined with traditional factor investment strategies.

Response #1)
We sincerely appreciate the reviewer's insightful feedback on strengthening the connection between traditional investment theories and the integration of DL and RL into our model. This feedback has been invaluable in refining our manuscript to present a clearer and more compelling relationship.
In response, we have revised the introduction to explicitly address the foundational principles of traditional investment theories, such as the MPT, the CAPM, and APT, in understanding risk-return trade-offs. Although these theories have significantly shaped the field of portfolio management, they are inherently limited by their static nature and simplified assumptions, which limit their ability to capture the dynamic complexities of modern financial markets. Our approach builds on these traditional frameworks by leveraging the factors they inspired—such as size, quality, value, investment, and volatility—as the basis for our model. These factors are regarded as essential drivers of portfolio performance and deeply rooted in financial theory. However, rather than treating these factors as static or deterministic, our model employs DL and RL to dynamically score and adjust these factors based on real-time market conditions. The combination of traditional factor insights with DL and RL creates a hybrid approach that balances stability and adaptability. Traditional factor investing provides a stable foundation for portfolio construction, while DL and RL enable the model to process complex, high-dimensional data and adapt to evolving market environments. This combination allows the DFPM to overcome the limitations of traditional methods, offering a flexible and robust solution capable of capturing short-term market dynamics and long-term investment opportunities. To highlight the importance of this integration, we have revised the introduction to explicitly articulate the manner in which DL and RL complement traditional factor-based strategies. This revision highlights DL's ability to process large-scale data, which enhances the utility of traditional factors, and the RL's decision-making capabilities, which allow the model to learn optimal portfolio strategies dynamically in response to market changes. We have revised the manuscript to address this critical gap by clearly articulating the integration of traditional financial theories with advanced machine-learning techniques. These revisions address the critical gap noted by the reviewer and ensure that our approach is more accessible and comprehensible to the reader.

Revised manuscript: Introduction section 2-4 paragraphs (page 2)

Comment #2
The paper compares 1. DFPM with traditional strategies; 2. DFPM with other strategies based on reinforcement learning; 3. Comparison with other research results and so on. This part is complete. However, there is no economic intuition involved. It is recommended to explain the reasons of these comparisons before analyzing the comparative results.

Response #2)
Thank you for highlighting the importance of providing economic intuition behind our comparative analysis. We agree that a clear rationale is essential for understanding the relevance and implications of DFPM's performance across various strategies. To address this issue, we have revised the analysis section to emphasize the reasoning

behind our comparisons. Specifically, we aim to demonstrate that DFPM’s dynamic factor-based scoring and adaptive RL approach outperforms traditional strategies that rely on static factor assumptions by effectively adapting to changing market conditions. Furthermore, by comparing the proposed DFPM with other RL-based approaches and state-of-the-art research results, we demonstrate its unique ability to integrate traditional factor insights into modern machine learning techniques. This integration allows the proposed DFPM to capture both market conditions and asset-specific characteristics more effectively, thereby positioning it as a hybrid framework that bridges the gap between traditional financial theories and advanced machine learning methods. These revisions have been added to the revised manuscript to provide a clearer economic context, enabling readers to better interpret the comparative results and recognize DFPM's practical value.

Revised manuscript: Baselines section 1-2 paragraphs (page 22)

4.Responses to the Reviewer #4
Comment #1
The details of integrating reinforcement learning with DFMs should be discussed. The combination of reinforcement learning for the PSM with DFMs is both innovative and highly relevant to current portfolio management research. While the methodology is well-constructed, the complexity of implementing the model, especially the integration of multiple components such as DFM and PSM, should be explained in greater detail. This will help readers better understand the steps involved.

Response #1)
Thank you for your valuable feedback on providing additional details regarding the integration of RL with the DFM and the PSM. We appreciate your emphasis on clarifying this integration because it is fundamental to understanding DFPM’s contributions to portfolio management research. In response to your comment. We have significantly expanded the methodology section to provide a detailed explanation of the integration process.

Revised manuscript: Dynamic factor portfolio model section 1-3 paragraphs (page 4), Integrated score module section first paragraph (page 13)

Comment #2
The macroeconomic factors used in the empirical analysis are insufficient. For example, variables related to economic growth, such as GDP growth or the Chicago Fed National Activity Index (CFNAI), should be included. These macroeconomic variables should be in real-time rather than final revised data to avoid forward looking bias in real investment decisions. Additionally, the "Index standard deviation" is seldom used and may not be particularly relevant.

Response #2)
Thank you for highlighting the importance of enriching macroeconomic factors in our empirical analysis. In response to your suggestion, we have incorporated GDP growth and the Chicago Fed National Activity Index (CFNAI) alongside the existing "Index standard deviation" feature. To ensure the analysis remains applicable to real-world investment decisions, all macroeconomic variables were sourced in real-time rather than relying on final revised data, thereby eliminating forward-looking bias.
As detailed in Table 3, the inclusion of GDP growth and CFNAI alongside the "Index standard deviation" produced the best results across both the Nasdaq 100 and Dow Jones indices. For the Nasdaq 100, this feature set achieved the highest fAPV (3.4961) and Sharpe ratio (2.6698). Similarly, for Dow Jones, the model had an fAPV of 2.2090 and a Sharpe ratio of 2.4479. Using GDP growth and CFNAI alone resulted in relatively lower performance metrics. For instance, the Nasdaq 100 achieved an fAPV of 1.7120 and a Sharpe ratio of 1.4012, while the Dow Jones achieved an fAPV of 1.5332 and a Sharpe ratio of 1.6198. The integration of these factors with the “Index standard deviation” demonstrated significant synergies, improving returns and risk-adjusted metrics. These findings highlight the importance of retaining the “Index standard deviation” because it effectively captures market volatility and is critical in enhancing the dynamic factor scoring process.
These improvements in macroeconomic feature selection contributed to enhanced final model performance. The inclusion of updated macroeconomic factors enriched the

dynamic factor scoring process within the DFPM, resulting in superior portfolio optimization outcomes. The revised manuscript reflects these results in both the macroeconomic analysis (Section: Data) and the overall model performance evaluation (Section: Experimental Results). Given the substantial updates and changes to the Experimental section, we request that you refer to the revised manuscript for a comprehensive overview of the updated experiments and their outcomes.

Revised manuscript: Data section fifth paragraph (page 21)

Comment #3
Reinforcement learning and other machine learning methods often face criticism for their lack of model interpretability. It is recommended to provide more explanations on how the model can be interpreted. For example, factor importance could be illustrated not only for the Nasdaq 100 in Figure 11 but also for the Dow Jones index. Additionally, it would be helpful to distinguish the contributions of dynamic factors versus reinforcement learning to the overall portfolio performance.

Response #3)
Thank you for highlighting the importance of model interpretability. In response to your comment, we extended the factor importance analysis to include the Dow Jones index and incorporated the results into the revised manuscript. The analysis revealed that the Nasdaq 100, with its growth-oriented and volatile composition, exhibited high variability in factor importance. Factors such as Quality, Beta, and Investment exhibited significant shifts across different market phases, particularly during transitions between bearish and bullish trends. This dynamic behavior underscores DFPM’s ability to adapt effectively to rapidly changing market conditions. In contrast, the Dow Jones, which consists primarily of large-cap, stable companies, exhibited more consistent factor trends. Factors such as Value and Size dominated across market conditions, reflecting the macroeconomic stability of the index. This comparison highlights DFPM's ability to balance dynamic adaptability for volatile indices, such as the Nasdaq 100, with stable performance for indices, such as the Dow Jones.
Additionally, as detailed in the original manuscript’s section, “The Effect of the Dynamic Factors” (Table 5), we evaluated the impact of dynamic factor scoring by applying it to traditional factor portfolio strategies, such as Factor Separate and Factor Mix. These experiments highlighted the advantages of dynamically updating factor scores in response to changing market conditions. The results demonstrated that dynamic factor scoring improves risk-adjusted returns (Sharpe ratio) and portfolio value (fAPV) across both the Nasdaq 100 and Dow Jones indices, albeit with variations in effectiveness based on market characteristics. Specifically, the Nasdaq 100 exhibited significant improvements in capturing short-term market movements, while the Dow Jones exhibited more stable, incremental benefits due to its less volatile nature.
In future work, we plan to explore advanced interpretability methods, including SHAP and LIME, to provide a more granular understanding of the contributions of individual factors and the RL process. These methods will enhance transparency and offer actionable insights for both researchers and practitioners.

Revised manuscript: Analysis of factor importance section first paragraph (page 37)

Comment #4
The details of the hyperparameters listed in Table 3 need to be clarified. It would be beneficial to explain why these specific hyperparameters were chosen. Additionally, it is important to address whether any training and validation samples were used, and if so, also describe the cross-validation techniques employed to determine the optimal hyperparameters. Providing details on the methods used to select these hyperparameters will enhance the transparency and robustness of the model development process.

Response #4)
Thank you for your suggestion regarding clarifying the hyperparameters in Table 3. We have revised the manuscript to include additional details on the selection and validation processes to ensure greater transparency and robustness of the model development process. The hyperparameters listed in Table 3 were determined through an extensive hyperparameter search process using grid and random search techniques over predefined ranges. The objective was to identify the optimal combination of

hyperparameters that maximized model performance on the validation set using key metrics, such as Sharpe ratio and fAPV, as evaluation criteria. We employed a rolling-window approach for training and validation to ensure robust evaluation. This method involved iteratively training the model on a fixed-length training window and validating it in the subsequent period. The rolling-window cross-validation approach is particularly well-suited for time-series data because it preserves temporal dependencies and avoids data leakage from future observations, which could otherwise bias the results. This approach allowed the model to be tested systematically across multiple validation sets, providing a reliable assessment of its generalizability. For each hyperparameter combination, performance metrics were averaged across all validation windows to ensure that the selected hyperparameters were not overfitted to specific periods or market conditions. This systematic approach enhances the robustness and stability of the model's performance under diverse scenarios, and it ensures that hyperparameters are selected objectively and systematically, thereby improving the model's interpretability and reliability. The revised manuscript now includes these details in the "implementation and hyperparameter" section and incorporates a dedicated section discussing additional cross-validation experiments. These updates address your concerns and provide a more comprehensive evaluation of the model's performance.

Revised manuscript: Implementation and hyperparameters section 1-3 paragraphs (page 24), Cross-Validation Results Analysis section second paragraph (page 34)

5.Responses to the Reviewer #5
Comment #1:
In Table 2, how did you choose the sample periods, e.g. 2008-2019 for Dow Jones and 08/1984-11/2022 for Nasdaq 100?

Response #1)
Thank you for your comment regarding the selection of sample periods in Table 2. The sample periods for the Dow Jones (2008-2019) and Nasdaq 100 (08/1984-11/2022) were selected to align with major market phases and comprehensively analyze both long-term trends and recent market behaviors.
For Dow Jones, the 2008-2019 period was selected to capture the effects of the 2008 financial crisis, the subsequent recovery, and the stable market phase leading up to 2019. This period highlights the model's performance across economic cycles and high-volatility periods. Additionally, this specific timeframe was chosen to ensure consistency and comparability with other benchmark models evaluated during similar periods, facilitating a fair assessment of DFPM's relative performance. The Nasdaq 100 sample period spans 1984 to 2022, covering a broader timeline to encompass various market cycles, including the dot-com bubble, the 2008 financial crisis, and recent episodes of market volatility. This extended timeline facilitates an evaluation of the model's robustness and adaptability across various economic events, providing a more comprehensive view of its long-term effectiveness. These sample periods provide a well-rounded view of the model's performance under diverse market conditions while enabling meaningful comparisons with other benchmark models. We appreciate your insightful comment, which enabled us to clarify this choice.

Comment #2:
There is not a detailed description of how the results in Table 5 is obtained. How do you "apply dynamic factor scoring to traditional factor portfolio investment methodologies" (Line 43-46 of Page 22)?

Response #2)
Thank you for your comment regarding the methodology behind the results in Table 5 and the application of dynamic factor scoring from the DFM module to traditional factor portfolio investment strategies. In response to your comment, we conducted experiments by integrating dynamic factor scoring into two traditional factor models—the Factor Separate and Factor Mix Models—to evaluate the effectiveness of the DFM module in established investment frameworks.
In the Factor Separate Model, each factor score is applied independently, with stocks receiving weights based solely on whether their individual factor scores exceed a specific threshold. The incorporation of dynamic factor scoring allows the model to adjust factor weightings over time in response to changing market conditions. This dynamic adjustment enables the Factor Separate Model to capture shifts in factor

relevance, thereby improving selection by prioritizing the most impactful factors at each period. In the Factor Mix Model, dynamic factor scores were aggregated into a composite score, ranking stocks based on the sum of all relevant factor scores. Investments were then allocated equally to stocks having composite scores above the median. Dynamic factor scoring allows the model to adjust the relative importance of each factor in real-time, reflecting current market conditions and enabling more flexible allocation decisions based on the latest market insights. In both models, the DFM module recalculated factor scores at each time step, incorporating recent market data to ensure factor weightings were aligned with evolving market conditions. This dynamic approach provides a framework for evaluating the enhancements introduced by the DFM module when applied to traditional factor investment strategies. The results in Table 5 reflect these improvements. These updates have been added to the revised manuscript to enhance the clarity of our methodology and the application of dynamic factor scoring in traditional investment frameworks.

Revised manuscript: The effect of the dynamic factors section second paragraph (page 27)

Comment #3:
On Page 17 before Figure 6, the authors write "This indicates that higher ranks correspond to higher investment potential. As demonstrated in the example, if the factor rank increases, it tends to indicate a rise in the actual stock prices." This statement is not obvious in Figure 6, particularly in the early sample period where the price difference is not distinctive. A better way is to show the log return as opposed to price trend.

Response #3)
Thank you for your suggestion regarding the use of log returns instead of the price trend in Figure 6. Although we acknowledge that log returns provide a more standardized view of stock performance, we deliberately selected the price trend to align with our objective of illustrating the relationship between factor ranks and investment potential over time. The price trend directly reflects actual stock price movements, which are the primary focus of investors in practice. We acknowledge that the early sample period may exhibit less distinct price differences, reflecting the overall variability in market conditions during that period. To address this issue, we have revised the accompanying text to clarify this point and emphasize the price trend's reflection of the long-term alignment between factor rank changes and stock price performance. Additionally, for comparison, we have included a supplementary figure (Figure 1), which presents a log-return-based visualization, alongside the existing figure (Figure 2), which illustrates the stock price trend, allowing readers to examine the differences and understand the influence of visualization choice on result interpretation.

Revised manuscript: Data section fourth paragraph (page 20)

Comment #4:
What period of data are used to calculate beta in Eq (12)? Notice that as a valid investment strategy, for the investment decision at time t, only data before (including) t can be used.

Response #4)
Thank you for your comment regarding the period of data used to calculate beta in Eq. (12) and the importance of timing for valid investment strategies. We agree that to support effective decision-making at any time t, beta calculations must rely solely on data up to and including time t to avoid forward-looking bias. In our approach, beta is calculated using a rolling three-month period of daily returns for both the index and stock data. This approach allows the beta to effectively capture recent market conditions and short-term dynamics. Specifically, we computed the covariance of the stock's returns with the market's returns over this three-month window, alongside the variance of the market's returns. This method ensures that only historical data up to time t are used, which aligns with real-time decision-making practices and enhances the model's validity as an investment strategy. Thank you for this observation, which allowed us to clarify our methodology further.

Revised manuscript: Model optimization section 2-3 paragraphs (page 16)

Comment #5:
Panel (D) of Figure 4 seems to be the CPI index, not CPI return. Although it is a bit obvious, Panel (C) Interest rate differential is not clearly defined.

Response #5)
Thank you for highlighting the distinction in Figure 4, Panel (D), regarding the CPI measure and the clarity of the interest rate differential in Panel (C). We confirm that Panel (D) represents the CPI Index. Panel (C) correctly represents the interest rate differential as an index, calculated as the difference between the 10-year Treasury bond rate and the 1-year Treasury rate. This measure captures the slope of the yield curve and serves as an important indicator of market expectations for economic growth and monetary policy. We have clarified this definition in the manuscript and have ensured that it is reflected accurately in the caption for Figure 4.

Revised manuscript: Data section first paragraph (page 17), Data section figure 4 (page 18)

Comment #6:
Repeated sentence on line 53-59 of Page 14.

Response #6)
Thank you for highlighting the repetition between lines 53 and 59 on Page 14. We have reviewed this section and removed the redundant sentence to improve the overall readability of the manuscript.

Revised manuscript: Data section first paragraph (page 17)

Comment #7:
On Line 27 of Page 19, there is a missing citation after "Described in".

Response #7)
Thank you for noticing the missing citation on Line 27 of Page 19 after "Described in." We have reviewed this section and added the appropriate reference to ensure proper attribution of the source material.

Revised manuscript: Baselines section first paragraph (page 23)

Comment #8:
In Figure 2, it is better to add other letters in addition to x_1, ..., x_K, e.g. those in Eq (1)-(3).

Response #8)
Thank you for your valuable suggestion to include additional notations from Eqs. (1)–(3) in Figure 2. In response to your feedback, we have updated Figure 2 to incorporate key notations, such as 〖h〗_k, 〖e〗_k and c to ensure closer alignment with the equations and to enhance the clarity of the illustration. By integrating these critical elements, the revised figure provides a more comprehensive visual representation of the Temporal Attention-LSTM (TA-LSTM) framework, thereby facilitating a clearer understanding of the relationship between the figure and the equations.

Revised manuscript: Dynamic factor module section figure 2 (page 10)

Comment #9:
In Eq (6), what is the definition of w_i, i.e. investment ratio? Is it the market value of asset i divided by the total market value of the portfolio?

Response #9)
Thank you for your comment. In Eq (6), w_i represents the investment ratio for asset i, which is defined as the market value of asset i divided by the total market value of the portfolio:
w_i=Market Value of Asset i/(Total Market Value of the Porfolio)

This definition ensures that w_i represents the relative weight of each asset, with the sum of w_i across all assets equal to 1. This clarification has been added to the revised manuscript.

Revised manuscript: Integrated score module section first paragraph (page 14)

Comment #10:
In the table "Algorithm 1", in step 8, the definition of S is the portfolio return, not Sharpe ratio.

Response #10)
Thank you for highlighting the discrepancy in the definition of S in step 8 of “Algorithm 1.” Upon careful review, we confirm that S represents the portfolio return, not the Sharpe ratio. The manuscript has been updated to reflect this definition and ensure consistency throughout. We appreciate your careful review and valuable feedback, which has helped improve the accuracy of this study.

Revised manuscript: Algorithm 1 (page 16)

Comment #11:
In Figure 6, the unit of Factor rank is missing (purple line).

Response #11)
Thank you for your observation regarding the unit of the Factor rank (purple line) in Figure 6. Upon careful review, we confirm that the Factor rank (purple line) is a unitless, normalized ranking metric rather than an absolute value. To maintain visual consistency and facilitate direct comparisons, the blue line (stock price trend) and purple line were scaled and aligned on the same axis. This approach ensures that both trends are effectively compared, highlighting their relative relationships over time.

Comment #12:
On Line 55-56 of Page 17, "market index volatility" is not in the model (e.g. Table 1 does not have market index volatility).

Response #12)
Thank you for highlighting the reference to “market index volatility” on Lines 55-56 of Page 17. To ensure clarity and consistency with the model parameters listed in Table 1, we have revised this term to “index standard deviation.” This adjustment aligns the terminology with the metrics used in our analysis, thereby maintaining accuracy and coherence throughout the manuscript.

Revised manuscript: Data section fifth paragraph (page 21)

Comment #13:
On Line 37-38, "which are representative US stock markets" does not make sense.

Response #13)
Thank you for identifying the ambiguity in the phrase “which are representative US stock markets” on Lines 37-38. To improve clarity, we have removed this phrase from the revised manuscript.

Revised manuscript: Data section sixth paragraph (page 22)

6.Responses to the Reviewer #6
Comment #1:
Further clarify how DFM handles changes in factor importance under different market conditions (Page 10).

Response #1)
Thank you for your comment regarding the DFM adjustment of factor importance across varying market conditions. The DFM dynamically recalculates factor scores by continuously analyzing recent market data, allowing it to adapt to evolving market conditions. Specifically, the DFM employs a rolling window approach that recalculates each factor’s weight based on its recent performance, enabling the model to capture

	shifts in factor relevance as market conditions evolve. This adaptive mechanism allows the DFM to prioritize the most relevant factors for current market conditions, thereby enhancing the robustness of portfolio decisions. This explanation has been included in the manuscript to clarify the DFM's functionality and its role in dynamic portfolio optimization. Revised manuscript: Dynamic factor portfolio model section 1-3 paragraphs (page 8) Comment #2: Has the DFM's prediction results been validated through cross-validation or leave-one-out methods (Page 10)? Response #2) Thank you for raising this point. The prediction results of the DFM were validated using a rolling cross-validation approach, which is particularly effective for time-series data (Bergmeir & Benítez, 2012). In this method, the dataset was divided into overlapping training and validation sets, where each training set comprised a fixed-length rolling window, and the subsequent period was...
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a Thematic Series?	No

Title : Dynamic Factor-Informed Reinforcement Learning for Enhancing Portfolio Optimization

Sungsoo Kim+, Hyukjae Kwon+, Jisang Yoon, Ha Young Kim*

Graduate School of Information, Yonsei University, Yonsei-ro 50, Seodaemun-gu, Seoul, 03722, Republic of Korea; kss8421@yonsei.ac.kr (SK); kwonhj1015@yonsei.ac.kr (HK);

sangjiyoon@yonsei.ac.kr (JY); hayoung.kim@yonsei.ac.kr (HYK)

+ These authors contributed equally

* Corresponding author: Ha Young Kim

Tel: +82-2-2123-4194 / Fax: +82-2-2123-8654 / E-mail: hayoung.kim@yonsei.ac.kr

Authors' information

Sungsoo Kim is currently pursuing an integrated master's and doctoral degree in information system at the Graduate School of Information at Yonsei University, Korea. He earned his bachelor's degree in computer science from Ajou University in February 2016. From 2016 to 2019, he worked in IBK Systems, contributing to the development of banking systems. His research interests encompass Reinforcement Learning, Explainable AI, Financial Data Analysis, and Deep Learning.

Hyuk-Jae Kwon received a bachelor's degree in information management and system from Peking University in 2022. He is pursuing a master's degree at the Graduate School of Information, Yonsei University in Korea since 2023. His current research interests include data science, computational finance, contrastive learning and time series data analysis.

Jisang Yoon is currently employed as an AI investment product developer at Qraft Technologies. He obtained his bachelor's degree from Ajou University, Korea, in February 2020, followed by a master's degree from Yonsei University in February 2022. His research interests encompass portfolio strategies, financial AI, and reinforcement learning.

Ha Young Kim is an Associate Professor at Graduate School of Information and department of artificial intelligence, Yonsei University, Republic of Korea. She received her Ph.D. degree at

Department of Mathematics, Purdue University, United States. From 2011 to 2016, she was a researcher in Samsung Advanced Institute of Technology (SAIT) of Samsung Electronics, Korea, working on various recognition systems with deep learning. Her primary research areas are deep learning and computer vision. She has published in leading journals, including Information Fusion, Computers in Human Behavior, Applied Soft Computing, Expert Systems with Applications, Stochastic and Dynamics, Computers in Biology and Medicine, PLoS ONE, Automation in Construction, Journal of Computing in Civil Engineering, Sustainability, and Annals of Finance. She is the inventor of 8 patents and 13 patent applications.

Declarations

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

Not applicable.

Funding

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2023R1A2C200337911 and No. RS-2023-00220762).

Authors' contributions

Sungsoo Kim: Writing–Original Draft, Methodology, Formal Analysis, Software, Data Curation, Investigation, Visualization. Hyukjae Kwon: Writing–Original Draft, Methodology, Formal Analysis, Software, Data Curation, Investigation, Visualization. Jisang Yoon: Software, Writing–Review & Editing, Formal Analysis, Methodology, Data Curation, Investigation. Ha Young Kim: Conceptualization, Methodology, Formal Analysis, Investigation, Writing–Review & Editing, Supervision, Project Administration, Funding Acquisition, Resources. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Abstract

Portfolio optimization is essential for investors seeking to manage risk, diversify assets, and maximize returns. Although recent studies have primarily focused on enhancing technical aspects such as model architecture through the application of deep learning or reinforcement learning, knowledge of factor portfolios, grounded in modern portfolio theory, remains paramount. Therefore, to effectively utilize the knowledge of factor investment strategies, we propose a novel hybrid portfolio investment method that integrates reinforcement learning with dynamic factors, called the dynamic factor portfolio model. The dynamic factors encompass five fundamental factors: size, value, beta, investment, and quality. The proposed model comprises two modules: a dynamic factor module that calculates a score based on factors reflecting the macro market and a price score module that calculates a score based on prices expressing the relationship between assets and their future value. With dynamic factor-informed knowledge, the proposed model can make portfolio decisions adaptively based on market conditions. Through comprehensive experiments, we validate the effectiveness of the dynamic factor module and demonstrate that the proposed model outperforms both traditional portfolio investment strategies and existing reinforcement learning-based strategies. Moreover, the proposed model offers interpretability by identifying critical factors across varying market scenarios, thereby enhancing portfolio management practices.

Keyword: Portfolio Optimization; Dynamic Factors; Factor Investing; Deep Learning; Reinforcement Learning; Factor Interpretability

Introduction

Portfolio optimization has emerged as a critical concern for investors seeking to manage risk and generate returns. Since the inception of Modern Portfolio Theory (MPT) as the cornerstone of the modern portfolio (Markowitz 1952), research on investing and managing multiple assets has garnered significant interest in the financial academic community. Consequently, ongoing research has focused on effective portfolio investment methods. Sharpe (1964), Lintner (1965), and Mossin (1966) developed the capital asset pricing model (CAPM), which compensates investors for assuming systematic risk, i.e., market risk, alongside total risk. Moreover, various portfolio investment strategies have been proposed that leverage a diverse set of factors that influence asset prices.

Risk factors such as asset undervaluation and profit realization during asset price declines prompted by an increase in the required rate of return have been considered (L'Her et al. 2002). Consequently, factor portfolio strategies have been designed based on market anomalies, using these factors to selectively choose stocks. Although incorporating all risk factors into an investment portfolio can enhance diversification, anticipating specific returns is challenging because of inherent uncertainty in financial markets (Jensen 1969). Furthermore, actual returns on real assets are affected by various factors, including macroeconomic conditions and company-specific factors. **Traditional factor strategies often face challenges in highly volatile and noisy markets because they rely on static models that lack the adaptability required to respond to rapid market shifts (Bertola 1988). Although these theories form a strong foundation, they may not fully capture the complexity of modern, dynamically evolving markets characterized by data irregularities and rapid fluctuations. These limitations highlight the need for adaptive models that more effectively adapt to shifting market conditions and manage real-world financial complexities more effectively. Integrating advanced techniques, such as Deep Learning (DL) and Reinforcement Learning (RL), with traditional factor strategies offers a more flexible solution, enabling dynamic adjustments to factor importance as market conditions change.**

Advancements in DL and increased computational power have driven significant progress in the financial industry, enabling successful applications ranging from exchange rate prediction to

comprehensive financial analysis (Li et al., 2022; Roberts et al., 2023; Htun et al., 2024; Wang et al., 2024). Building on these achievements, researchers have increasingly leveraged DL and RL techniques for portfolio optimization (Heaton et al., 2017; Wang et al., 2021; Ma et al., 2021; Li et al., 2021). These approaches have demonstrated significant promise, consistently outperformed traditional financial portfolio methods and highlighted the transformative impact of machine learning in modern portfolio management.

DL techniques excel at uncovering complex patterns in financial data, and RL provides a robust framework for dynamic decision-making under uncertainty. However, financial time-series data are inherently irregular and often characterized by substantial noise and heightened volatility due to diverse market influences, which complicate accurate modeling (Sewell, 2011; Engle, 2004). Consequently, models relying solely on data-driven approaches without incorporating financial domain expertise may struggle to consistently produce reliable features, particularly under fluctuating market conditions. When integrated with traditional factor strategies, DL and RL offer a more adaptive approach by combining stable relationships based on financial theories with the flexibility to respond to real-time market changes. This hybrid approach captures static factor-based insights and dynamic market behaviors, effectively addressing the complexities of irregular time series and volatile financial environments.

To address the complexities of portfolio optimization, this study proposes a Dynamic Factor Portfolio Model (DFPM) that integrates traditional factor investment methodologies with RL techniques. This integration enables DFPM to respond dynamically to shifting market conditions while leveraging the predictive capabilities of key investment factors, thereby enhancing portfolio performance. The DFPM framework comprises two primary components: the Dynamic Factor Module (DFM) and the Price Score Module (PSM), each contributing complementary insights for robust portfolio allocation.

The DFM calculates adaptive scores for established investment factors, including size, value, beta, quality, and investment. These factors are derived from factor investment strategies and dynamically adjusted by the DFM based on prevailing market conditions, capturing the evolving importance of each factor. This adaptability allows the DFPM to integrate fundamental investment insights while making it responsive to broad economic trends that influence portfolio allocation. The PSM complements this by analyzing asset price data, assessing both inter-asset correlations and individual price patterns, thereby providing real-time price signals and stock-level insights.

Within the RL framework, DFM and PSM outputs are combined to train the DFPM with portfolio weights optimized for the Sharpe ratio, thereby achieving a dynamic balance between risk and return. By embedding factor-based investment strategies in the RL framework, DFPM adapts its portfolio decisions based on a comprehensive perspective that considers both macroeconomic trends and individual asset behaviors. This integration of factor investment insights with adaptive RL learning broadens the information base for decision-making and enhances DFPM's ability to optimize portfolios in diverse market conditions, making it highly applicable to real-world financial applications.

We conduct experiments employing various portfolio selection methods and reward objectives to optimize the model. The experimental results demonstrate the robustness and adaptability of the proposed DFPM. By leveraging dynamic factors and integrating macroeconomic and price data, DFPM consistently outperformed traditional portfolio strategies and State-of-the-Art (SOTA) RL methods in risk-adjusted metrics, such as the Sharpe ratio and fAPV. Its ability to dynamically adjust to market conditions enhanced its performance during volatility periods, with reduced drawdowns and optimized portfolio allocations. The analysis also highlighted the critical role of factors such as value, size, and momentum in driving investment decisions, validating DFPM’s ability to capture market dynamics and deliver practical advantages in portfolio management. The main contributions of this paper are as follows:

• We propose a novel RL-based portfolio optimization framework, DFPM, informed by the factor portfolio strategy aimed at enhancing profitability. DFPM incorporates a module that dynamically scores the five major factors based on market conditions. To the best of our knowledge, this is the first framework that directly incorporates these five key factor indicators into an RL model for portfolio optimization.

• The proposed DFPM is an explainable RL framework, allowing for the assessment of factor importance under varying market conditions. This provides valuable insights for portfolio factor investment strategy.

• Through extensive experiments, we demonstrate the effectiveness of dynamic factors and show that DFPM outperforms traditional portfolio methods on the NASDAQ100 and Dow Jones indices. In addition, we reveal that DFPM surpasses recent SOTA RL models on the Dow Jones.

Literature review

Traditional portfolio theory

MPT is fundamentally rooted in the pioneering work of Harry Markowitz in 1952. This work, which is pivotal to the field of finance, laid the foundation for the mean-variance portfolio model. Markowitz introduced key statistical concepts, such as the mean (average) and standard deviation, to quantify the return and risk inherent in stocks. He also advanced the assessment of risk within stock portfolios by considering covariance, thereby establishing a framework for optimizing asset allocation. This framework allows investors to either minimize risk or maximize return while adhering to a predetermined level of risk. Markowitz's contributions constitute the bedrock on which all subsequent portfolio theories have been constructed.

Expanding on the mean-variance portfolio model, two pivotal theories emerged: CAPM and the Arbitrage Pricing Theory (APT) (Ross 1976). Within a capital market equilibrium framework, CAPM defines the equilibrium return of an asset. CAPM enables the calculation of an asset's equilibrium return by assessing its individual risk in relation to the market, referred to as systematic risk, and defining it as a risk premium. In contrast, APT is a theoretical model based on the absence of arbitrage opportunities. It enables asset valuation through multiple risk factors and shares some conceptual overlap with CAPM. APT serves as the foundation for factor-based portfolio investment strategies that depend on diverse risk factors. Simultaneously, the introduction of risk-parity strategies equalizes the risk contributions of individual stocks. These strategies encompass a spectrum of approaches, including inverse volatility (Asness and Liew 2011), equal risk contribution (Maillard et al. 2010), and maximum diversification (Choueifaty and Coignard 2008) strategies, each designed to optimize portfolio risk management. Concurrently, the landscape of factor-based portfolio investment strategies now includes numerous risk factors. A fundamental premise of a factor-based portfolio investment strategy is that an asset's return embodies a premium related to a specific risk factor. Consequently, these strategies prioritize investments in stocks undervalued in relation to this risk factor. An example of such a factor is the market risk factor, which denotes the systematic risk of an asset and is encapsulated in the CAPM beta coefficient.

The Fama-French three-factor model (Fama and Kenneth 1993) significantly enhanced the explanatory power of pricing models by incorporating value and size factors in conjunction with the CAPM. This model demonstrated that portfolios comprising value stocks with a high book-to-market ratio and small-cap stocks with a low market capitalization exhibit superior returns. Carhart (1997) introduced the momentum factor, another crucial risk factor, emphasizing its ability to capture the persistence of past returns into the future. In another study, Fama and French (2015) introduced the concept of profitability, often referred to as quality, using profitability metrics such as net income to capital as risk factors. Their research demonstrated that companies with a high profitability factor tend to have higher expected returns. Similarly, the investment factor (Fama and French 2015) challenges the assumption that companies prioritizing aggressive capital growth outperform others, indicating no clear return advantage. Finally, the low volatility factor (Baker et al. 2011) posits that stocks with lower volatility exhibit higher expected returns. Collectively, these foundational theories and risk factors have not only shaped the landscape of modern portfolio management but also paved the way for the development of sophisticated investment strategies that continue to evolve in the pursuit of optimal risk-adjusted returns.

Related works

The rapid advancement of artificial intelligence has led to an increasing application of DL- or RL-based methodologies in solving portfolio optimization problems. Portfolio optimization studies have primarily focused on enhancing DL or RL models. Yang et al. (2020) back tested Dow Jones 30 stocks using an ensemble trading strategy that incorporates three actor-critic-based algorithms: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Deep Deterministic Policy Gadient (DDPG). The results demonstrated higher profits than the existing algorithms. In the quest for an improved portfolio management solution, Wu et al. (2021) developed a portfolio management system that leverages neural networks, specifically Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), for predicting future stock prices. These predictions were used as inputs for the RL model. Also, Wang et al. (2021) proposed a novel ensemble portfolio optimization framework, integrating a hybrid variant of mode decomposition-bidirectional long short-term memory DL model and RL-based model, applied to a wide range of commodity assets for predicting future returns. Sun et al. (2024) improved the architecture of the PPO agent by introducing a GraphSAGE-based feature extractor to capture complex non-Euclidean relationships between market indices, industry indices, and stocks**. Despite these advancements, many of these methods struggle to adapt to rapidly changing market conditions, integrate domain-specific financial insights, and maintain interpretability, which are crucial for practical portfolio management.**

However, alternative approaches to portfolio optimization problems involve integrating DL algorithms with different methods. For instance, Chen et al. (2019) introduced a fuzzy grouping genetic algorithm that incorporates the concept of collective intelligence. This algorithm employs fuzzy logic to dynamically adjust parameters during the evolution process to discover a suitable and diversified group stock portfolio. Furthermore, in a subsequent study, Chen et al. (2021) devised a grouping trading strategy portfolio. This portfolio is formed by strategic groups using the grouping genetic algorithm, and its performance is enhanced through the proposal of a trading algorithm designed to mitigate losses. **Although these methods effectively optimize structural configurations, they struggle to adapt dynamically to real-time market fluctuations. This limitation arises from their reliance on static rules, which hinders flexibility in rapidly changing market environments.**

An alternative approach involves research that integrates existing finance theories with DL or RL. Jang and Seong (2023) combined modern portfolio theory and RL and solved a multi-mode problem using Tucker decomposition. However, this study lacks the use of the most important factors in the financial market in the RL model. In contrast to previous studies, this study incorporates a broader range of essential factors used in traditional financial factor methods into models using RL methodology. By combining the dynamic factor scoring model with stock price information, the model achieves increased stability and higher profits.

Methodology

Dynamic factor portfolio model

The DFPM is formulated as a DL-based RL model, distinguishing itself from traditional RL approaches through its adaptive engagement with the financial environment, learning of actions, and autonomous strategy optimization. **By leveraging DL methodologies, DFPM can effectively manage complex datasets and scenarios, building on the core principles of RL—reward systems and learning agents—that are fundamental for optimal decision-making in portfolio management.**

To account for macroeconomic conditions and the unique characteristics of individual stock prices, DFPM incorporates two dedicated modules: DFM and PSM. The DFM computes scores based on key factors—size, value, beta, quality, and investment—that represent prevailing market conditions. This factor-based analysis allows the DFPM to dynamically adjust portfolio allocations in response to broader economic trends. Conversely, the PSM consolidates stock price data, evaluating both inter-stock relationships and individual price patterns within a portfolio. This module provides a granular perspective on stock-level dynamics, thereby enhancing the model's ability to capture real-time price fluctuations.

To integrate these modules, DFPM employs an RL framework that dynamically weights the outputs of DFM and PSM based on their relevance to current market conditions. Fig. 1 presents an

overview of this process, which highlights the DFPM's dual-module structure and its contribution to deriving optimal portfolio strategies within an RL framework. This integrative approach allows DFPM to achieve adaptive, balanced portfolio allocations that maximize the Sharpe ratio while remaining robust to changing market conditions.

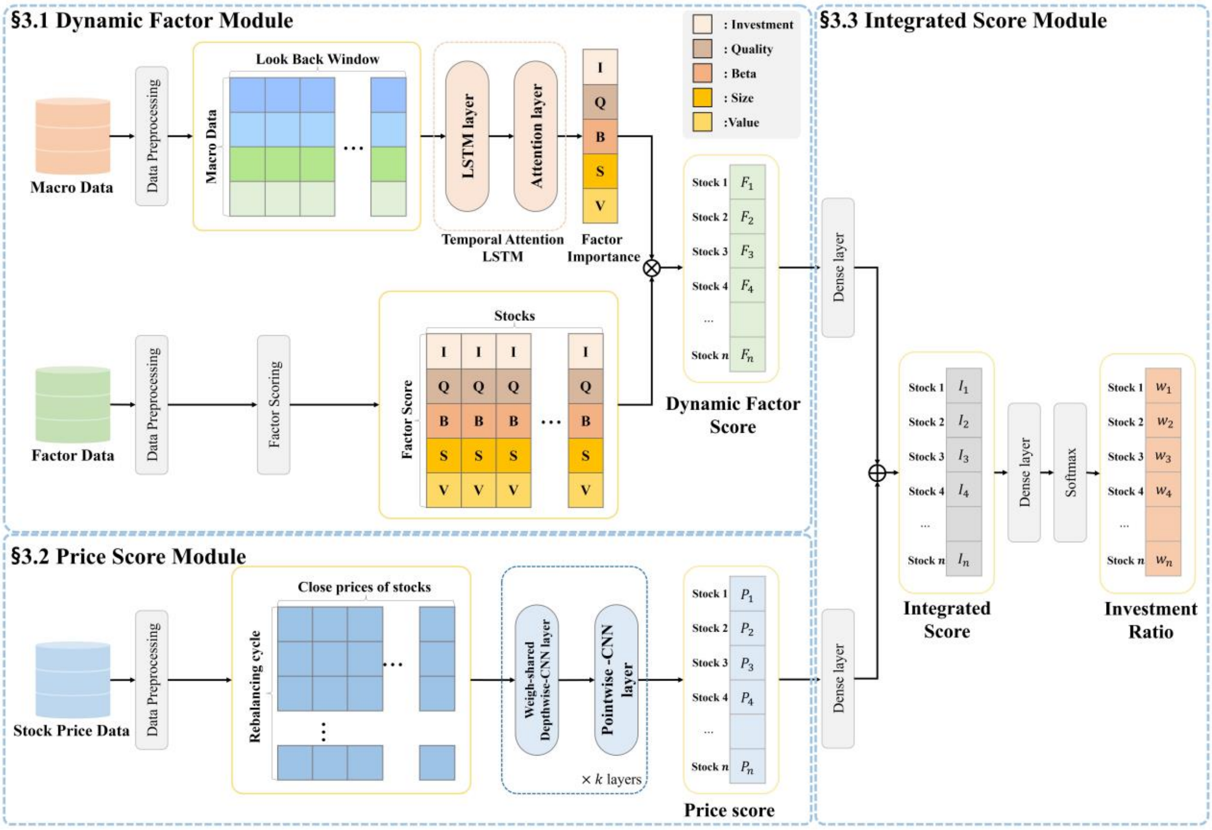


Fig. 1 The proposed DFPM Framework

Dynamic factor module

Given the highly unstable and erratic nature of financial markets, studying dynamic factors based on market conditions is essential to enhance portfolio returns. Therefore, DFM leverages macro market data to calculate factor importance weights (M), which indicate the impact of each risk factor on portfolio performance. By integrating factor importance weights M and factor data, the DFM learns dynamic factor scores that represent asset-specific values exclusively calculated from five factors reflecting market conditions. These five factors are defined as follows:

- Value: Represents the relative valuation of an asset based on fundamental metrics, such as price-to-earnings (P/E) or price-to-book (P/B) ratios. Lower ratios indicate a higher potential for undervaluation.

- Size: Refers to the market capitalization of an asset, distinguishing between small-cap and large-cap stocks. Small-cap stocks often exhibit higher growth potential but are associated with increased risk.
- Beta: Measures an asset's sensitivity to overall market movements, with a beta greater than 1 indicating higher volatility than the market and less than 1 indicating lower volatility.
- Quality: Includes indicators such as profitability, earnings stability, and financial health. High-quality stocks are generally more resilient during market downturns.
- Investment: Captures growth in capital expenditures or reinvestment rates, linked to the asset's potential for future growth.

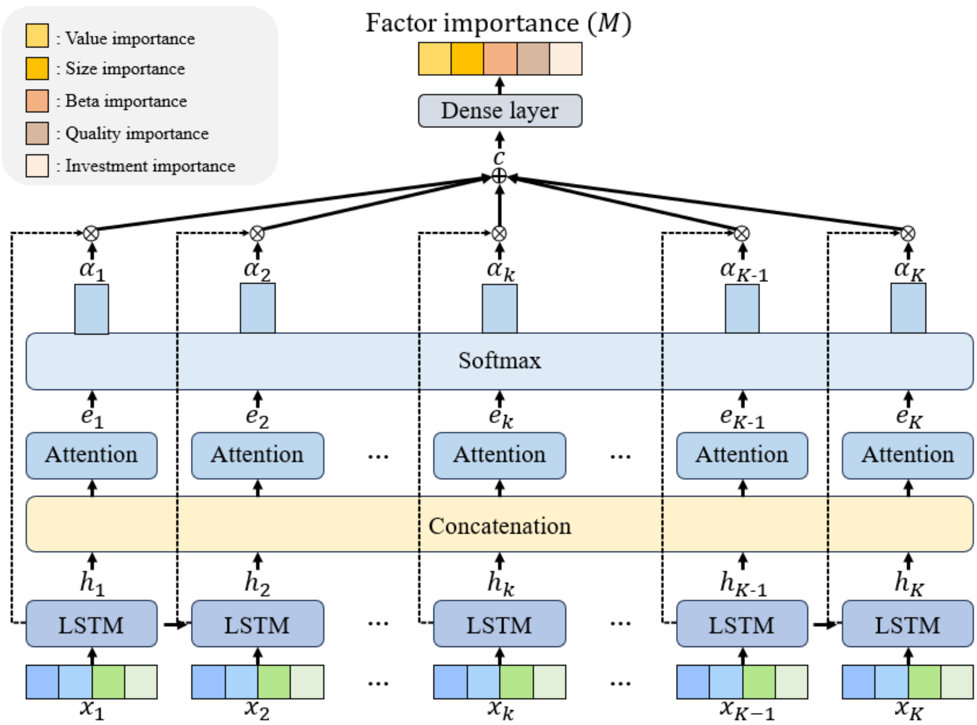


Fig. 2 TA-LSTM architecture

Previous studies (Wang et al. 2019; Ta et al. 2020; Wang et al, 2021) have used Long Short-term Memory (LSTM) (Hochreiter and Schmidhuber 1997) to extract information from time series data in portfolio management optimization. However, LSTM encounters challenges in handling long-term dependencies as the data length increases, and it may struggle to learn contextual information about market conditions. Furthermore, recent transformer-based models present a potential risk of overfitting in small datasets and exhibit limitations in capturing temporal relations among a continuous set of points (Zhang et al. 2022). To address these problems, we propose an architecture called Temporal Attention-LSTM (TA-LSTM), which combines an LSTM model with an attention mechanism for DFM. The final output value of the TA-LSTM is obtained by calculating the attention score as a weighted sum of the hidden state at each time step and the value of the last time step. **The attention mechanism in the TA-LSTM enables the model to assign different weights to each time step, allowing it to focus on periods that are more relevant to specific market conditions. This selective emphasis on key time intervals gives TA-LSTM a distinct advantage over regular LSTM, which assigns equal priority to all time steps.** Fig. 2 shows the architecture of the TA-LSTM. The attention mechanism used in TA-LSTM follows the concat-based function, which was introduced by Vaswani et al. (2017J). The context vector c represents the output of the TA-LSTM. The overall formula is as follows:

$$e_k = W_a^T * tanh(W_b * [h_k; h_K] + W_c x_k) \tag{1}$$

$$\alpha_k = \frac{exp(e_k)}{\sum_{k'=1}^{K} exp(e_{k}')}, \tag{2}$$

$$c = \sum_{k=1}^{K} \alpha_k \, h_k \tag{3}$$

Here, k and K denote the time step index and the total number of time steps in the input data, respectively. h_i represents the ith hidden state of the LSTM, and the learnable weights are denoted by $W_a \in R^A$, $W_b \in R^{A \times 2H}$, *and* $W_c \in R^{A \times P}$. For the parameters of the LSTM model, the hidden state dimension (H) was set to 32, the input data dimension (P), which represents the number of macroeconomic variables, was set to 4, and the look back window size (A) was set to 18. The symbol $*$ signifies the multiplication operation. The attention score is represented by e_k, and α_k denotes the normalized attention weight of kth hidden state. The context vector (c) is the attention value, which is the weighted sum of the attention weights for the input information at each timestep.

Subsequently, the computed context vector c is passed through the dense layer, where the $tanh$ function is employed as the activation function to ensure that the values of the vector are confined within the range of -1 and 1. **This feature is particularly valuable for reflecting scenarios in which certain factors either enhance or hinder portfolio performance under varying market conditions. For instance, a negative score indicates a factor's reduced relevance in a given market scenario, while a positive score highlights its increased importance.** Consequently, the outcome comprises the factor importance weights encapsulated in vector M. The rationale behind selecting the tanh activation function is to interpret the importance of factors in both positive and negative directions.

In this module, the vector M obtained is multiplied by the factor data containing the factor score information for each stock. This factor data is represented by the matrix $D \in R^{n \times m}$, where m denotes the number of investment assets and n represents the number of factors.

$$V = M * D. \tag{4}$$

The DFM dynamically recalculates factor importance weights using a rolling window approach to further strengthen portfolio decisions in response to market fluctuations. This approach updates each factor's weight based on recent performance data, enabling the DFM to adaptively prioritize factors that are most relevant to current market conditions, thereby enhancing the robustness of portfolio decisions.

Price score module

The DFM calculates values solely based on asset factor information. To incorporate asset price information into portfolio optimization, a PSM is integrated into the proposed framework. Effective portfolio construction hinges on extracting price features that encompass not only the historical movements of individual assets but also crucial inter-asset correlations (Akita et al. 2019). Therefore, the architecture of the PSM leverages weight-shared depthwise convolution to extract historical price information and pointwise convolution to capture inter-asset relationships. Therefore, the PSM architecture leverages weight-shared depthwise convolution to extract temporal price information and pointwise convolution to capture cross-asset relationships.

The PSM first applies weight-shared depthwise convolution to each stock's historical price data, allowing temporal information from individual price movements to be processed separately. This weight-sharing mechanism applies the same filter across all stocks, thereby enabling the module to capture common patterns in price movements across assets. This step reduces computational complexity and minimizes overfitting. Additionally, this approach is particularly effective for identifying correlations between price trends across different assets.

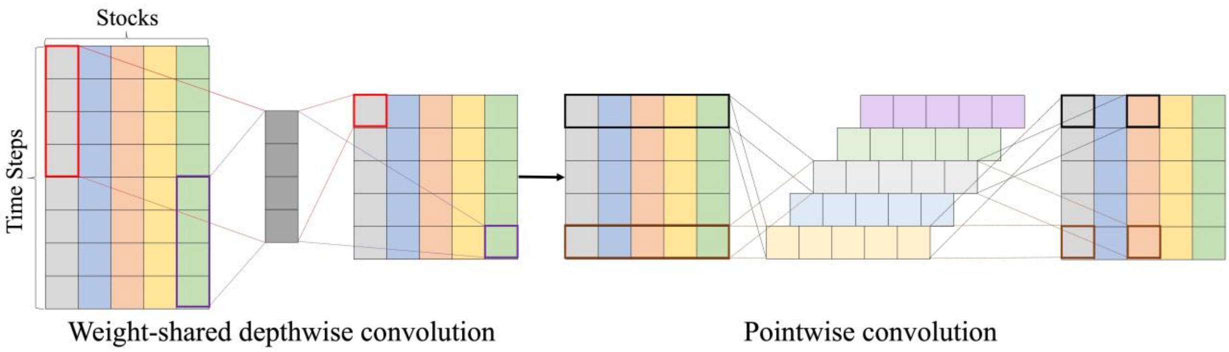


Fig. 3 Weight-shared depthwise separable convolution

Following depth-wise convolution, pointwise convolution is applied to aggregate the inter-asset information. However, in contrast to conventional pointwise convolution, our design maintains the number of stocks while compressing historical price information by preserving the channels. To achieve this, we match the number of filters in the pointwise convolution to the number of stocks. The process continues with weight-shared depthwise separable convolution until the time dimension is reduced to 1. **Consequently, the PSM captures cross-sectional relationships and interactions among different assets, effectively incorporating correlation data into the model. This layered approach of depth-wise and pointwise convolution enables the PSM to extract essential temporal features and inter-stock correlations, which are critical for constructing robust portfolios.**

Integrated score module

The integrated score module calculates the stock weight score using the two scores obtained from DFM and PSM. **This module translates factor and price scores into normalized investment weights, ensuring a balanced and data-driven approach to portfolio allocation.** The stock weight score represents a comprehensive value score for each asset, incorporating both price and factor information. **The stock weight score is a comprehensive valuation of each asset, integrating factor scores (FS) from the DFM and price scores (PS) from the PSM. The module computes the stock weight score as follows:**

$$Stock\ Weight\ Score = W_p^T \tanh(W_F FS + W_T PS), \tag{5}$$

where:

- **$W_p \in R^m$: Learnable weight vector for final aggregation, where m denotes the number of assets.**
- **$W_F,\ W_T \in R^{s \times m}$: Learnable weight matrices that dynamically adjust the influence of FS and PS, with s representing the rebalancing period.**
- **FS : Factor score vector from the DFM, reflecting macroeconomic trends.**
- **PS: Price score vector from the PSM, capturing short-term price dynamics.**

Combining FS and PS allows the model to consider both broad market trends and asset-specific characteristics, generating a comprehensive score for each stock. The portfolio weights are derived by applying Boltzmann's (1957) SoftMax function to the calculated stock weight score as follows:

$$w_i = SoftMax(Stock\ Weight\ Score), \tag{6}$$

where w_i represents the investment ratio for the ith stock, which is defined as the proportion of the portfolio allocated to that stock. The SoftMax function ensures that the weights are probabilistic and sum to 1, thereby facilitating an interpretable and flexible portfolio allocation.

Model optimization

This study is rooted in the foundational assumption of the Markov Decision Process (MDP) by Bellman (1957), which is a fundamental concept in RL. MDP is defined as $MDP = < S, A, P, R >$, where S represents the state space, A represents the action space, P represents the probability density function, and R represents the reward. When an agent performs an action $a_t \in A$ at state $s_t \in S$. The state transition distribution to the next state s_{t+1} is $s_{t+1} \sim P(s_{t+1} | s_t, a_t)$. The reward received by the agent at time t is $r_t = R(s_t, a_t, s_{t+1})$. A policy refers to a set of strategies or rules that dictate the actions an agent should take in a particular state. The agent learns a policy π to compute $a = \pi(s) \in A$ that maximizes the expectation of the sum of all discounted rewards:

$$J = E_{p(\tau)}[\sum_{t=1}^{T} \gamma^{t-1} r_t], \tag{7}$$

where $p(\tau)$ represents a probability distribution across time steps, and γ denotes the discount factor. In this study, the action is represented by the variable w_i, as shown in Equation (5). Furthermore, state space S corresponds to the prices of assets, factors, and macroeconomic data. The agent receives $s_t \in$ S, and then chooses an action a_t involving the distribution of portfolio items. The rewards for the proposed model can be set to various metrics, such as the rate of return or Sharpe ratio. Based on the experiments (see Table. 5), we adopted the Sharpe ratio as the primary reward metric to maximize portfolio profitability.

The portfolio strategy employed in this study revolves around creating a long-only portfolio. This approach involves executing trades solely with equity capital and investing in risk-free assets without resorting to borrowing for leverage. Although long-short portfolio management, which involves purchasing and selling overvalued and undervalued assets, may be more suitable for assessing the effectiveness of strategies using market anomalies, its practicality is often impeded by restrictions on short selling. Consequently, we opt for a long-only portfolio management strategy that is more accessible to the average investor. Regarding the definition of the investment ratio at t as W_t, the subsequent reward value r_{t+1} and portfolio returns C are given by:

$$r_{t+1} = \left(W_t \frac{P_{t+1}}{P_t} - \lambda C\right), \tag{8}$$

$$C = \{(W_t - w_{t-1})P_t\}, \tag{9}$$

where P_t denotes the price vector of the stocks bought at time t , C represents the transaction cost, and the weight of this transaction cost is denoted by λ. Therefore, the objective J that the agent aims to maximize is expressed as follows:

$$J = E_{p(\tau)}[H_T], \tag{10}$$

$$H_T = \frac{A_T}{V_T}, \tag{11}$$

$$A_T = \frac{\sum_{t=1}^{T} r_t}{T}, \tag{12}$$

$$V_T = \sqrt{\frac{\sum_{t=1}^{T}(r_t - A_T)^2}{T}}, \tag{13}$$

where H_T represents the Sharpe index, which is a risk-adjusted return proposed by Sharpe (1964). **Here, A_T represents the average return over the investment horizon T, and V_T denotes the standard deviation of returns during the same period, capturing the portfolio's risk.** The mean and variance expressions for calculating the Sharpe index are equal to A_T and V_T in the above equation.

Only historical data up to and including time *t* are employed to calculate beta, preventing forward-looking bias and aligning with real-time investment decision-making practices. Beta is computed using a rolling three-month window of daily returns for both the market index and individual stocks, effectively capturing recent market conditions and short-term dynamics. Specifically, for each calculation, the covariance between a stock's returns and the market's returns over this three-month period is divided by the variance of the market's returns during the same period. This methodology ensures that all calculations supporting H_T adhere to valid investment strategies and reflect real-time decision-making requirements. By optimizing the portfolio to maximize H_T, the model dynamically balances the trade-off between risk and return, providing a robust framework for portfolio management.

Algorithm 1. DFPM training algorithm

1: Set number of stocks I, Market macro data m, Stock price data s, Factor data D, Monthly stock's return data r, Dynamic factor module F, Price score module P

2: Initialize model parameter θ

3: for epoch = 1 to n do

4: Calculate factor importance weights (M) and dynamic factor scores (FS)

$$M = F(m)$$

$$FS = M * D$$

5: Calculate price scores $PS = P(s)$

6: Integrate dynamic factor scores and price scores to calculate Stock Weight Score WS

$$WS = W_p^T tan\,h(W_F FS + W_T PS)$$

7: The investment ratio of ith stock $w_i = SoftMax(WS)$

8: The portfolio return $r = \sum_{i=1}^{I} w_i * r_i$

9: Compute Average Portfolio Return $A_T = \frac{\sum_{t=1}^{T} r_t}{T}$

10: Compute Portfolio Return Volatility $V_T = \sqrt{\frac{\sum_{t=1}^{T}(r_t - A_T)^2}{T}}$

11: Calculate portfolio's Sharpe ratio $S = \frac{A_T}{V_T}$

12: Calculate objective function $J = E_{p(\tau)}[S]$

13: Update parameters $\theta\_(t+1) = \theta\_t + \eta \nabla J$

14: end for

Data

This study leverages three data categories: factors, macro market indicators, and stock prices. These data sources reflect stock prices and macroeconomic information, employing DFM and PSM. Table 3 provides an overview of the variables used in our study. Specifically, we obtained U.S. Consumer Price Index (CPI) data, 10-year treasury bond rate data, and 1-year treasury bond rate data from Federal Reserve Economic Data (FRED), while the remaining data were acquired from Wharton Research Data Services (WRDS). To address any missing values within the dataset, we employed the backfill method. **Additionally, the interest rate differential, a key macroeconomic variable in this study, is calculated as the difference between 10-year and 1-year treasury bond rates.** Fig. 4 shows the trend of macro market variables. These figures represent the flow of macroeconomic data over the data period. The units for the NASDAQ Index and the Dow Jones Index are in thousands and the return of the two indices experienced a temporary decline due to the impact of the Dot-com bubble, subprime mortgage crisis, COVID-19 pandemic, but overall, they have shown an upward trend.

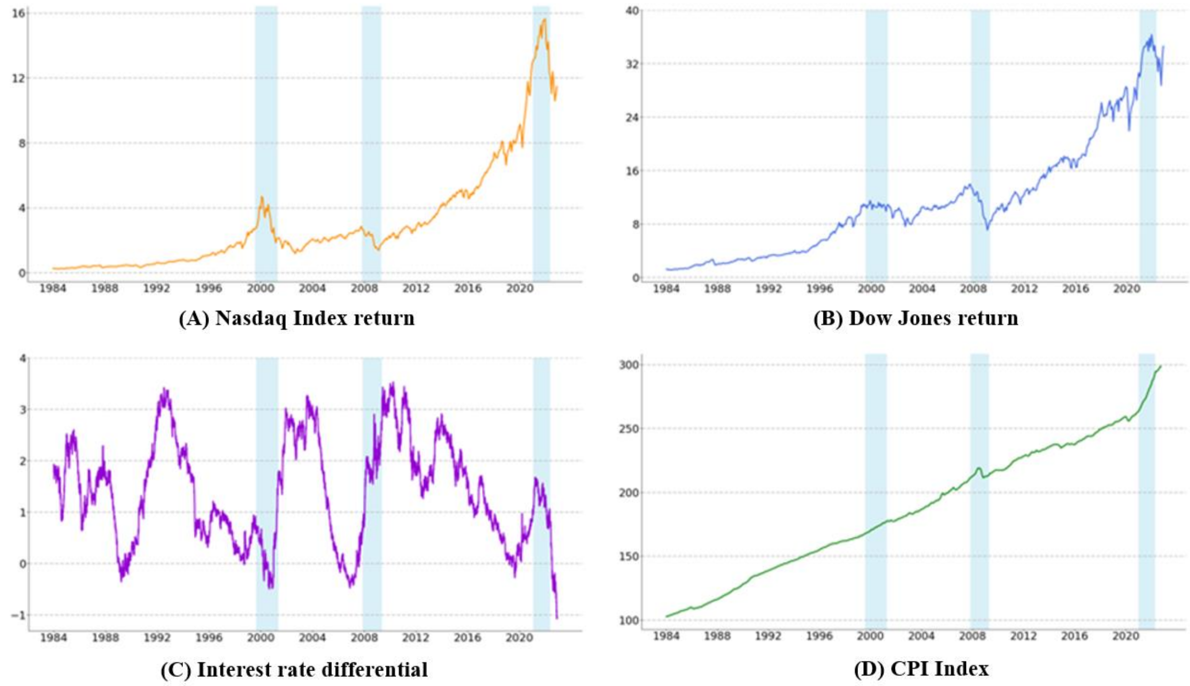


Fig. 4 Macro market variables from financial market

As previously mentioned, this study incorporates risk factors including size, value, beta, investment, and quality. These risk factors form the basis of the five-factor asset pricing model proposed by Fama et al. (2015), which is renowned for its enhanced explanatory power for returns compared with the traditional three-factor model. The beta factor is calculated using the daily returns r_m from the index data and the daily returns r_i from the stock data, employing the equation proposed by Sharpe (1964):

$$\beta_{i,m} = \frac{Cov(r_m, r_i)}{Var(r_m)} \tag{12}$$

We use market capitalization to represent the size factor. This is calculated by multiplying the number of outstanding shares by the stock price. For the value factor, we directly use book-to-market ratio data. These quality and investment factors are represented by a company's return on equity (ROE) and return on assets (ROA), respectively. Fama et al. (2015) initially derived these factors from financial statement items.

However, subsequent studies by Paliienko et al. (2022) and Hapsari et al. (2017) used financial ratios for ease of calculation and high accessibility, using ROE for quality and ROA for investment. This study adopts the latter approach for consistency and simplicity.

Risk factor data exhibit significant variation in scale and distribution across instruments, potentially resulting in unstable learning. Fig. 5 shows the distribution and skewness of each risk factor. The x-axis of this figure shows the values of each risk factor and the y-axis shows the frequency of these values, which shows that there is skewness in the data distribution. To address this issue, this study employs a scoring method similar to existing factor portfolio strategies for preprocessing. Specifically, we calculated the ranking of the risk factors for each day. Subsequently, we divided the ranking of stocks by the total number of stocks, converting the ranking into a value between 0 to 1.

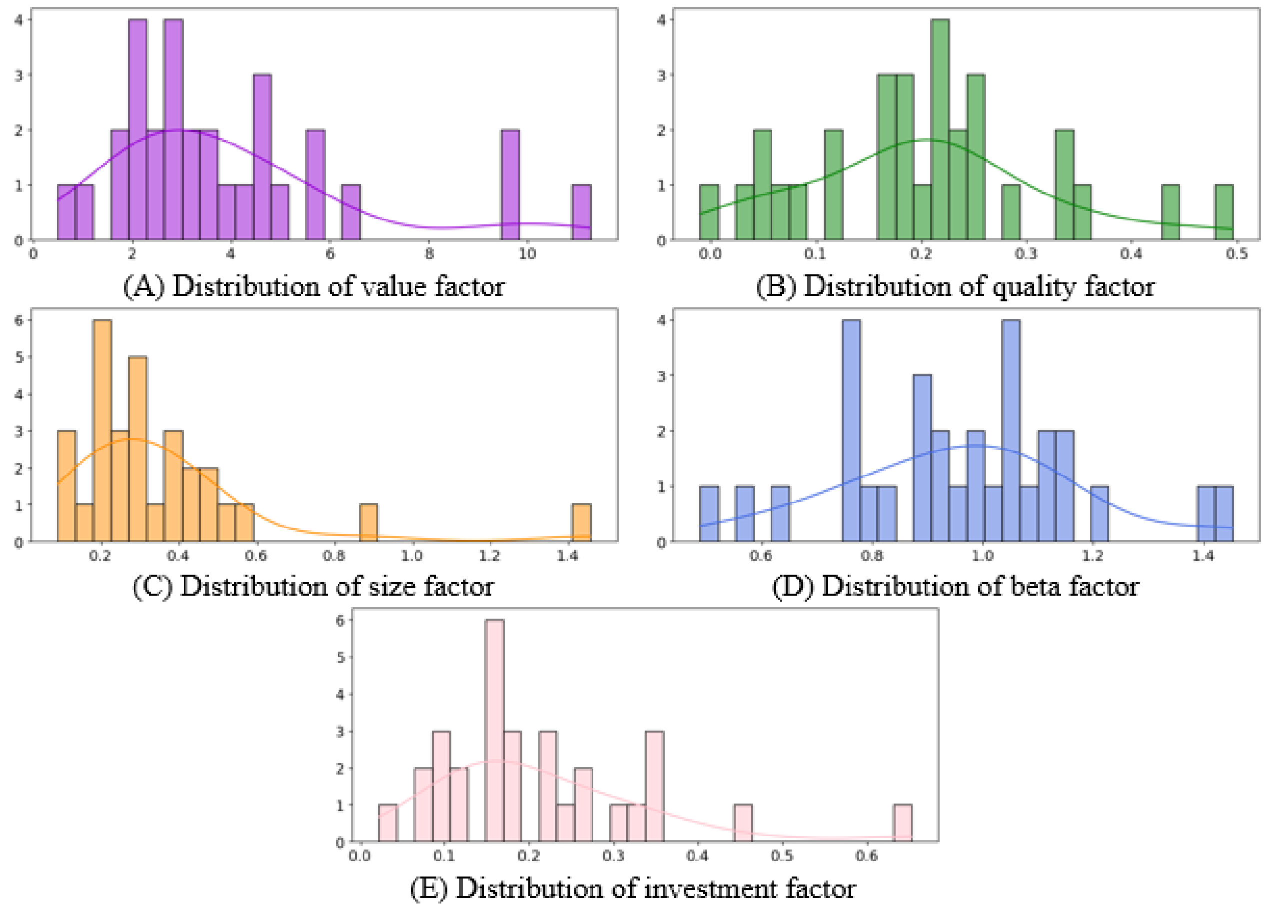
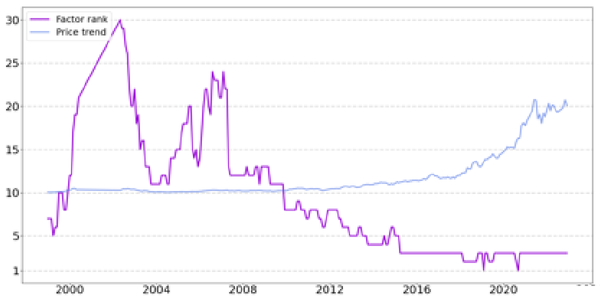


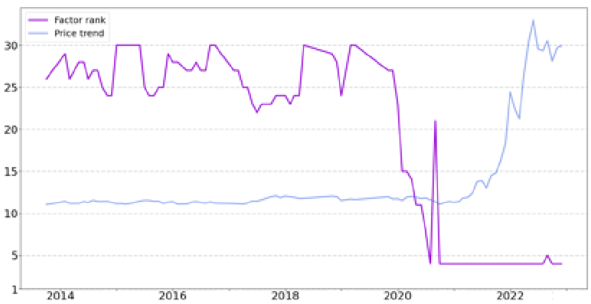
Fig. 5 The distribution and skewness of the risk factors before preprocessing

Following this, we subtracted the ranking into a value ranging from 0 and 1, where 1 signifies the highest-ranked risk factor and 0 the lowest. Subsequently, the risk factor data were standardized again to ensure that the higher-ranked risk factors were closer to 1. In a typical factor investing strategy, adjustments were made to quality and investment factors to approach 1 for larger values. Conversely, for size, beta, and value factors, scores were determined by subtracting their values from 1, assigning higher scores to lower values. The resulting calculated factor scores were then subjected to normalization. To ensure accurate comparisons in the Dow Jones data, we fixed the number of stocks at 29. In contrast, the Nasdaq 100 data includes 210 stocks in the data period, and we dynamically selected the 30 stocks with the highest sum of factor scores for each month. Fig. 6 shows the relationship between the actual stock price trend and the factor rank calculated using our method. The figure illustrates the variation in factor rankings for a sample stock, selected from a set of 30 Nasdaq-listed stocks using our proposed methodology. This indicates that higher ranks correspond to higher investment potential. **Although the early sample period exhibited less pronounced price differences due to the variability in market conditions at that time, the overall trend underscores the consistent relationship between increasing factor ranks and corresponding stock price performance. This correlation highlights the practical value of factor rankings as a predictive tool for stock performance.**

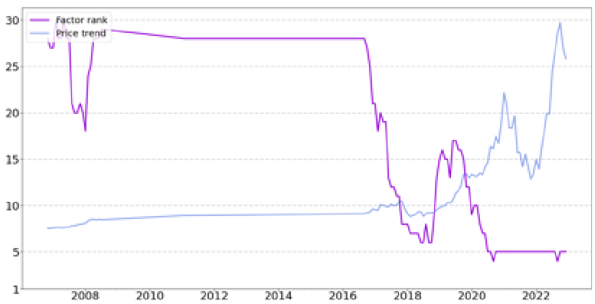
(A) Factor rank of Apple



(B) Factor rank of Amazon



(C) Factor rank of Tesla



(D) Factor rank of Nvidia

Fig. 6 The comparison between the stock price trend and the factor

Macro market data encompass market index prices, the **market index standard deviation**, long-term and short-term interest rate differentials, **the U.S. CPI, the Chicago Fed National Activity Index (CFNAI), and GDP**. The long-term and short-term interest rate differential, defined as the difference between the 10-year Treasury constant maturity representative and the 1-year Treasury constant maturity, is calculated using the respective rates obtained from FRED**. The CFNAI, a monthly index designed to gauge overall economic activity and related inflationary pressure, provides an additional macroeconomic perspective, while GDP growth reflects broader economic trends and helps contextualize market dynamics. Stock and Watson (2003) demonstrated that inflation indicators, such as the CPI, have significant predictive power in terms of economic activity and market trends.**

Table 1. Input variable description.

Categories	Variable	Values corresponding to variables
Factors	Beta	Daily returns of index, Daily returns of stocks
	Size	Number of outstanding shares, Stock prices
	Value	Book-to-market ratio
	Quality	Enterprise ROE
	Investment	Enterprise ROA
Macro market	Index return	Index prices
	Cpi return	U.S. consumer price index
	Interest rate differential	10-year treasury bond rate, 1-year treasury bond rate
	Index standard deviation	**Index prices**
	Chicago Fed National Activity Index	**Chicago Fed National Activity Index**
	GDP growth	**GDP growth**
Stock price	Daily close return	Stock's close price

To evaluate the proposed DFPM model, we conducted experiments on the Nasdaq 100 and Dow Jones datasets. For the Dow Jones dataset, in line with previous studies (Chen et al. 2019; Yang et al. 2020; Wu et al. 2021; Chen et al. 2021; Jang et al. 2023), we partitioned it into a train dataset and a test dataset. For the Nasdaq 100 dataset, we allocated 90% for training, 10% for validation included in the training dataset, and the remaining 10% for testing. Table 2 presents a description of the employed datasets.

Table 2. Input variable description.

Market	Dataset	Start	End	#month
Dow Jones	Train	01/01/2008	31/12/2016	108
	Test	01/01/2017	31/12/2019	48
	All	01/01/2008	31/12/2019	156
Nasdaq 100	Train	01/08/1984	31/01/2019	414
	Test	01/02/2019	30/11/2022	46
	All	01/08/1984	30/11/2022	460

Experiment

Baselines

To comprehensively evaluate the performance of the proposed DFPM model, we benchmarked it against both traditional portfolio strategies and recent RL-based methodologies. This dual comparison highlights DFPM's advantages, demonstrating its ability to bridge traditional financial principles with advanced machine learning techniques.

Traditional portfolio strategies often rely on static methods that lack the flexibility to dynamically adapt to changing market conditions. The comparison of DFPM with these methods underscores the benefits of its dynamic factor scoring and RL-driven adaptability, which enable the model to respond effectively to real-time market shifts. Four traditional strategies were selected as baselines for this evaluation:

- Uniform portfolio: This strategy allocates an equal amount of investment capital to each asset or security within the portfolio.

- The mean-variance optimization portfolio: As described in MPT by Markowitz (1952), this strategy uses modern portfolio theory to construct an investment portfolio aiming to maximize expected returns for a given level of risk or to minimize risk for a targeted level of expected return. The resulting portfolio allocation balances the trade-off between risk and return.
- Factor separate portfolio: This strategy involves assigning equal weights to stocks based on the number of factors with scores greater than zero for each stock.
- Factor mix portfolio: This strategy sums up all factor scores for each stock to obtain a comprehensive factor ranking value and equally invests in stocks with a ranking higher than half.

To further highlight the DFPM's unique contributions, we compared it to recent RL-based methodologies. This analysis demonstrated DFPM's competitive edge in combining macroeconomic trends with asset-specific characteristics. Additionally, the results demonstrate DFPM's capacity to align established financial principles with the flexibility of RL, delivering superior portfolio performance across diverse market conditions.

The following RL-based strategies were used as baselines:

- Chen et al. (2019) proposed a fuzzy GGA-based approach to speed up the evolutionary process for diverse group stock portfolio optimization.
- Yang et al. (2020) introduced the ensemble strategy, which automatically selects the best-performing agent among PPO, A2C, and DDPG based on the Sharpe ratio.
- Chen et al. (2021) presented an approach using a group genetic algorithm to procure a portfolio comprising group trading strategies while simultaneously identifying optimal stop-loss and take-profit thresholds.
- Wu et al. (2021) developed a Portfolio Management System (PMS) using RL with two neural networks (CNN and RNN).
- Jang et al. (2023) proposed a method that integrates RL and traditional financial theory.

Implementation and hyperparameters

All experiments were performed using an NVIDIA RTX 3090 GPU and executed on Python 3.8.18 with PyTorch 1.7.1 and CUDA 11.0. The hyperparameters used for model training are outlined in Table 3. The portfolio rebalancing cycle is set at 20 days, indicating a periodic review of stock price information approximately one month prior to the rebalancing. **To align with real-world trading practices while emphasizing dynamic factor allocation, a transaction cost of 0.3% was applied exclusively to selling transactions (Jang et al., 2023). This adjustment provides a more realistic evaluation of the model's performance by incorporating transaction costs into the trading process.**

The hyperparameters were determined through an extensive search process, as described in the methodology. Both grid and random searches were employed across predefined ranges for each parameter, with a rolling window approach used for training and validation to preserve temporal dependencies in the time-series data. These hyperparameters were fine-tuned to optimize key metrics, such as the Sharpe ratio and fAPV, to ensure robustness and generalizability across different market conditions. These settings reflect the rigorous and systematic approach to achieving optimal model performance under realistic trading conditions. The results confirm the robustness of the selected hyperparameters and their generalizability across various scenarios.

In addition, early stopping was incorporated into the training process. Early stopping was implemented based on the validation loss, with a patience parameter of 10 epochs, which implies that training was terminated if no improvement was observed in the validation set for 10 consecutive epochs. This approach ensures optimal generalizability without overfitting the training data.

Table 3. Hyperparameters settings used for training.

Hyperparameter	Value
DFM	
Size of input features	4
Length of sequence (month)	18
No. of hidden layers	1

Size of each hidden layer	32
Size of output	5
Batch size	108
Learning rate	0.005
Optimizer	Adam
Epochs	100
PSM	
Size of input features	29
No. of hidden layers	3
Kernel size of each hidden layer	[8, 7, 7]
Activation layer	ReLU
Size of output	29
Batch size	32
Stride	1
Padding	0
Dilation	1
Learning rate	0.005
Optimizer	Adam
Epochs	100

Evaluation metrics

Various metrics are employed to evaluate the performance of portfolio optimization. The most intuitive measure is the annualized return, which quantifies the model's earnings over a year. However, this metric has the disadvantage of overlooking portfolio risk. Consequently, we use the Sharpe ratio, which is a risk-adjusted return, and it is defined as follows:

$$Sharpe\ Ratio = \frac{E[\rho_t - \rho_F]}{std(\rho_t - \rho_F)}, \quad (13)$$

where ρ_t represents the return of a portfolio and ρ_F denotes the return of the risk-free asset. The Sharpe ratio indicates that a higher ratio corresponds to a lower level of risk relative to higher returns. However, it encounters the challenge of treating upward and downward movements equally. Therefore, to provide a more comprehensive assessment, we also used the Maximum Drawdown (MDD) (Magdon-Ismail et al.

2004). MDD measures the maximum decline in portfolio value from its peak to its trough, serving as an indicator of portfolio volatility. In contrast to the Sharpe ratio, MDD does not consider the portfolio's return, but it effectively gauges stability. The formula for calculating MDD is as follows:

$$MDD_t = max(\frac{V_{max}-V_t}{V_{max}} \times 100\%), \tag{14}$$

where V_{max} represents the historical maximum value of the investment income and V_t denotes the value of the investment income at time t. Furthermore, we evaluate the portfolio using the Final Accumulative Portfolio Value ($fAPV$) (Jang et al. 2023), which is defined as follows:

$$fAPV = \frac{v_t}{v_0}, \tag{15}$$

where v_t is the portfolio value at time t.

Experiment Result

The effect of the dynamic factors

We empirically investigate the effects of dynamic factors, which is the core idea of this study. First, given that DFPM integrates the DFM and PSM modules, it facilitates individual assessments by selectively omitting each module. This enables the examination of the effect of the DFM module that reflects dynamic factors. Table 4 presents the results, demonstrating superior performance of the integrated DFPM model across three evaluation metrics compared with models using either module individually, in both the Nasdaq 100 and Dow Jones markets. These findings support our hypothesis that integrating price and factor information is effective. In the Nasdaq 100 market, the standalone PSM model outperformed the standalone DFM model. Conversely, in the Dow Jones index, the standalone PSM model did not outperform the single DFM model. This performance disparity can be attributed to the relatively limited information available on stock relationships in the Dow Jones market, which may be attributed to the less diverse price trends exhibited by stocks within the Dow Jones market compared with those in the Nasdaq market.

To further explore the impact of the DFM module, which incorporates dynamic factor scoring, we conducted experiments by applying dynamic factor scoring to traditional factor portfolio investment methodologies. These approaches include both factor separate and factor mix models. **In the factor-separate model, each factor score is applied independently, with stocks weighted based on individual factor scores. Dynamic factor scoring allows weights to adjust over time, responding to market changes and highlighting relevant factors during each period. In the factor mix model, dynamic factor scores are aggregated into a composite score, with investments allocated to stocks having scores above the median. This adaptive scoring approach allows the model to reflect up-to-date market insights by dynamically adjusting the weight of each factor.**

The summarized results are presented in Table 5. Across both the Nasdaq 100 and Dow Jones markets, the application of dynamic factor scoring consistently demonstrated effectiveness across three key performance indicators. **In the Nasdaq 100 market, the fAPV and Sharpe ratio improved across both the Factor Separate and Factor Mix models. However, in the Dow Jones market, while dynamic factor scoring led to enhancements in the fAPV and Sharpe ratios, there was a slight increase in MDD.** This outcome aligns with the Dow Jones market's relative stability during that period. By adopting the Sharpe ratio as an objective that considers both returns and risks, we confirmed a slight decrease in MDD performance. Therefore, our findings support the advantages of our proposed DFM when applying dynamic factor scoring to the two traditional factor models.

Table 4. Comparison of results with individual modules.

Dataset	Model	MDD	Sharpe ratio	fAPV
Nasdaq 100	Only_Factor (DFM)	26.04	1.2556	1.5981
	Only_Price (PSM)	19.77	1.2987	1.6143
	DFPM (DFM+PSM)	19.05	2.7100	3.6014
Dow Jones	Only_Factor (DFM)	7.88	1.4552	1.3953
	Only_Price (PSM)	10.11	1.7712	1.4013
	DFPM (DFM+PSM)	6.78	2.4479	2.2090

Table 5. Results of applying dynamic factor scoring to traditional factor portfolio strategies.

Dataset	Model	MDD	Sharpe ratio	fAPV
Nasdaq 100	Factor separate	19.35	1.1862	1.5425
	+ Dynamic factor scoring	15.25	1.3821	1.6508
	Factor mix	19.39	1.1845	1.5781
	+ Dynamic factor scoring	17.96	1.4809	1.7924
Dow Jones	Factor separate	9.14	1.4666	1.4097
	+ Dynamic factor scoring	10.44	1.6387	1.4606
	Factor mix	8.86	1.1668	1.3033
	+ Dynamic factor scoring	12.69	1.2683	1.3865

Comparison with traditional portfolio strategies

We compared the performance of our proposed method with that of traditional portfolio strategies, evaluating its effectiveness in the Nasdaq 100 and Dow Jones markets. The detailed experimental results for the Nasdaq 100 are presented in Table 6, and the Dow Jones experimental results are presented in Table 7. Focusing on the Nasdaq 100 results, our proposed model exhibited superior performance in terms of the Sharpe ratio, fAPV and MDD compared with established traditional methodologies. Comparing the proposed model with the Nasdaq index, we observed a 134.33% increase in the Sharpe ratio and an impressive 126.02% increase in fAPV. MDD also decreased by about 0.7%, showing the stability of the model.

Notably, when analyzing the return trend depicted in Fig. 7, our model demonstrated a more rapid increase during upward trends compared with traditional portfolio strategies. In addition, it maintained stability without a significant decline during downward trends.

When evaluating the results of the Dow Jones market, the proposed model demonstrated superior performance across all three indicators. Compared with the Dow Jones index, MDD decreased by 37.31%, and the Sharpe ratio and fAPV increased by 57.39% and 34.09%, respectively. Analyzing the rate of return trend in Fig. 8, it initially exhibited a similar rate of return but rose more rapidly than other models from 2019. Therefore, the proposed model serves as a baseline for traditional portfolio strategies in both markets.

Table 6. Results of comparative experiments using traditional portfolio management methods in the Nasdaq 100.

Model	**MDD**	**Sharpe ratio**	**fAPV**
Nasdaq index	32.40	1.1393	1.5468
MVO	18.72	1.4050	1.6730
Uniform	19.74	1.2463	1.5849
Factor separate	19.35	1.1862	1.5425
Factor mix	19.39	1.1845	1.5781
DFPM (Ours)	18.58	2.6698	3.4961

Table 7. Results of comparative experiments using traditional portfolio management methods in the Dow Jones.

Model	**MDD**	**Sharpe ratio**	**fAPV**
Dow Jones index	18.17	1.5553	1.6467
MVO	9.12	1.3001	1.3785
Uniform	9.88	1.4152	1.4375
Factor separate	9.14	1.4666	1.4097
Factor mix	8.86	1.1668	1.3033
DFPM (Ours)	6.78	2.4479	2.2090

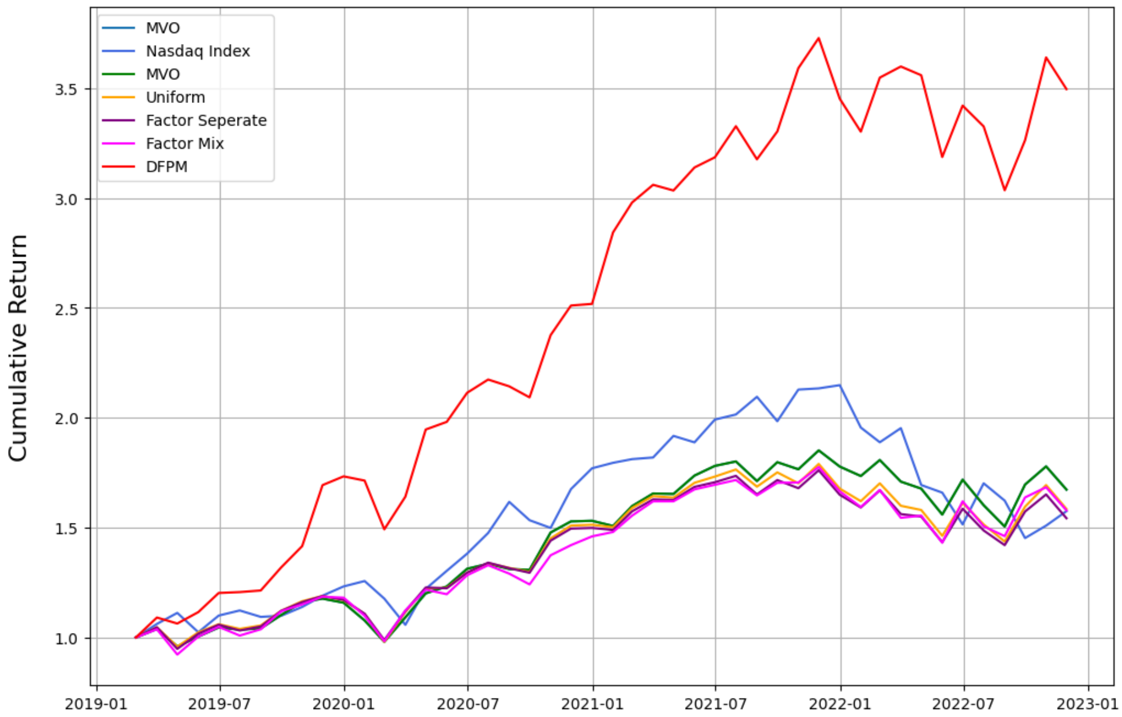


Fig. 7 Comparison of return trends during the testing period in the Nasdaq 100

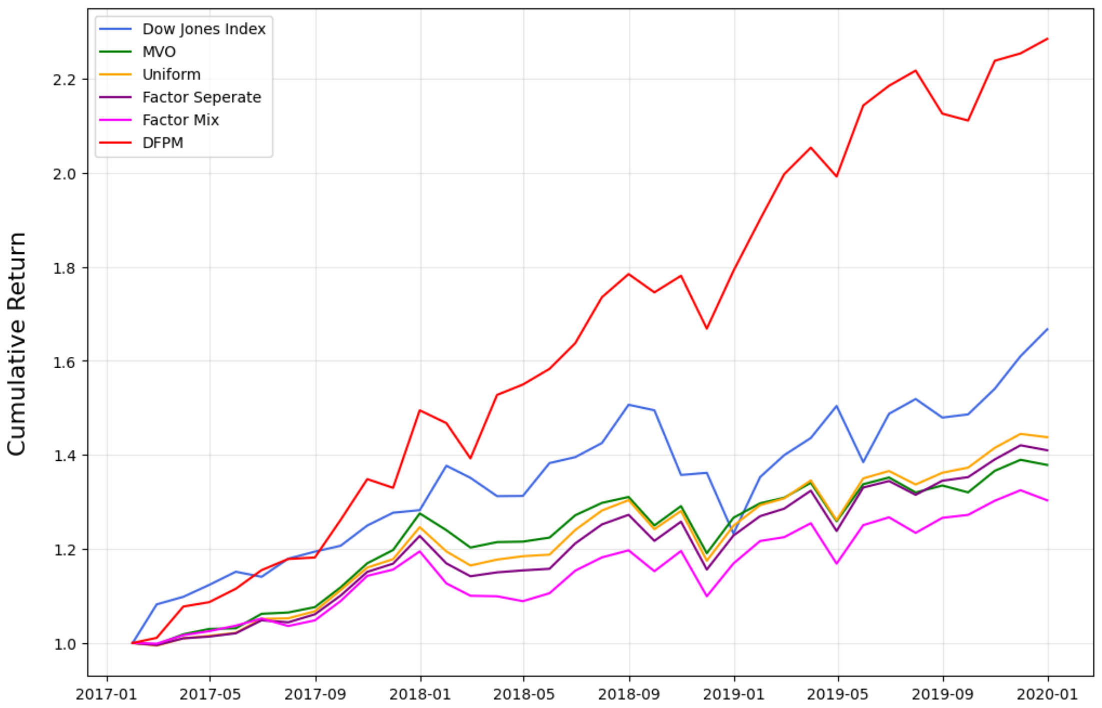


Fig. 8 Comparison of return trends during the testing period in the Dow Jones

Comparison with state-of-the-art RL methods

To demonstrate the superiority of the proposed DFPM model, we compared its performance with the most recent RL methods that optimize portfolios using Dow Jones data. The portfolio performance results are presented in Table 8. The proposed model significantly outperformed the SOTA RL methods in both Sharpe ratio and fAPV, while demonstrating superior performance in terms of MDD. In particular, the DFPM model achieved a **21.98**% improvement in Sharpe ratio and a **14.63**% improvement in fAPV compared to the method proposed by Jang et al. (2023). Jang's method, which incorporates modern portfolio theory, previously exhibited superior performance among earlier RL models. Moreover, the reduction of MDD by **22.27**% underscores the model's effectiveness in terms of risk management. The dynamic scoring of the five indicators and the efficient fusion of price information within the DFPM model are key contributors to its superior performance. Consequently, our model stands out as a superior choice, excelling in both profitability and stability across all dimensions, thereby affirming the excellence of the proposed model.

Table 8. Results of the comparative experiment using the SOTA RL methods in Dow Jones.

Model	MDD	Sharpe ratio	fAPV
Chen et al. (2019)	12.02	1.2794	1.5027
Yang et al. (2020)	15.45	0.7338	1.2502
Wu et al. (2021)	9.58	1.6403	1.5960
Chen et al. (2021)	10.56	1.4643	1.5861
Jang et al. (2023)	8.29	2.0067	1.9270
DFPM (Ours)	**6.78**	**2.4479**	**2.2090**

Comparisons of portfolio selection methods

When building a portfolio from the Nasdaq100, we selected the top 30 based on the sum of scores of all factors. This method is referred to as DFPM-Mix. To analyze performance based on portfolio selection methods, we compared our method with single-factor approaches such as DFPM-Size, DFPM-Beta, DFPM-Quality, DFPM-Investment, and DFPM-Value. Table 9 displays the experimental results. Among the tested models, the DFPM-Mix model achieved the highest Sharpe ratio and fAPV. The DFPM Beta model yielded superior MDD results. **Conversely, the DFPM-Investment model exhibited the highest MDD, indicating increased risk exposure**. Analyzing the trends of the Sharpe ratio and fAPV in Fig. 9 reveals that during the test period, the DFPM-Mix model and DFPM-Quality exhibited a significantly higher rate of increase in the bull market compared with the other models. Conversely, during the bearish trend when the Nasdaq index experiences a decline, the **DFPM-Size** model demonstrates remarkable stability, while the DFPM-Mix model exhibits some volatility but maintains a consistent level of return. Therefore, supported by empirical evidence, the DFPM-Mix model, which constructs a portfolio by leveraging all relevant factors, emerged as the optimal approach for the Nasdaq 100 dataset.

Table 9. Test results based on portfolio selection methods on Nasdaq 100.

Model	MDD	Sharpe ratio	fAPV
DFPM-Size	19.94	1.8734	2.4160
DFPM-Beta	10.91	1.6483	1.9467
DFPM-Quality	29.92	1.9505	3.0197
DFPM-Investment	33.54	1.5324	1.9847
DFPM-Value	20.51	0.9443	1.4765
DFPM-Mix	18.58	2.6698	3.4961

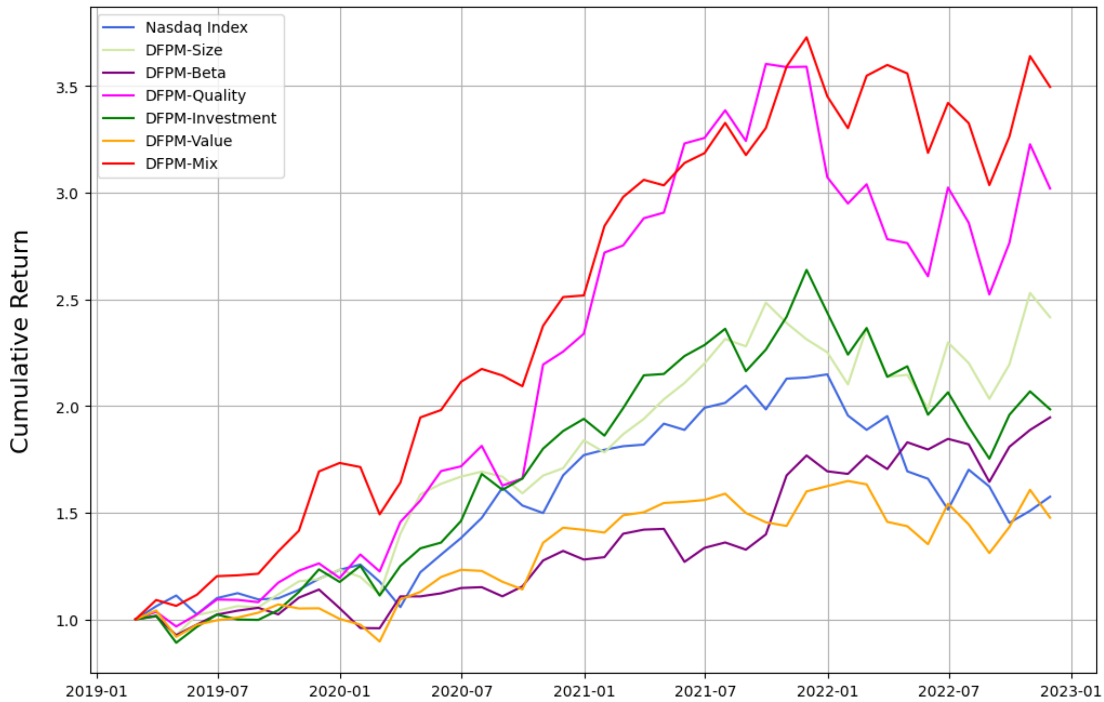


Fig. 9 Trend of returns based on portfolio selection methods in the Nasdaq 100

Comparison according to the reward objective

We conducted performance comparison experiments with the DFPM model using the Nasdaq 100 dataset, in line with the reward objective. The objective functions for the reward included three commonly employed components: return, standard deviation, and Sharpe ratio. The experimental results are presented in Table 10. When the reward was measured by the Sharpe ratio, the proposed model achieved similar performance in MDD but demonstrated significantly superior results in the other two metrics: Sharpe ratio and fAPV. In addition, analyzing the return trend shown in Fig. 10, the DFPM-Sharpe model demonstrates a substantial surge starting in 2020 and consistently sustains superior returns compared with the other two models. This stable and commendable performance is attributed to the Sharpe ratio's ability to consider both the risk and profit of the portfolio. Given the clear superiority of our model across all three metrics, the reward objective formula in this study is based on the Sharpe ratio.

Table 10. Test results according to reward objective.

Model	MDD	Sharpe ratio	fAPV
DFPM-Return	18.16	1.3919	1.7303
DFPM-Std	19.29	1.2591	1.5891
DFPM-Sharpe	18.58	2.6698	3.4961

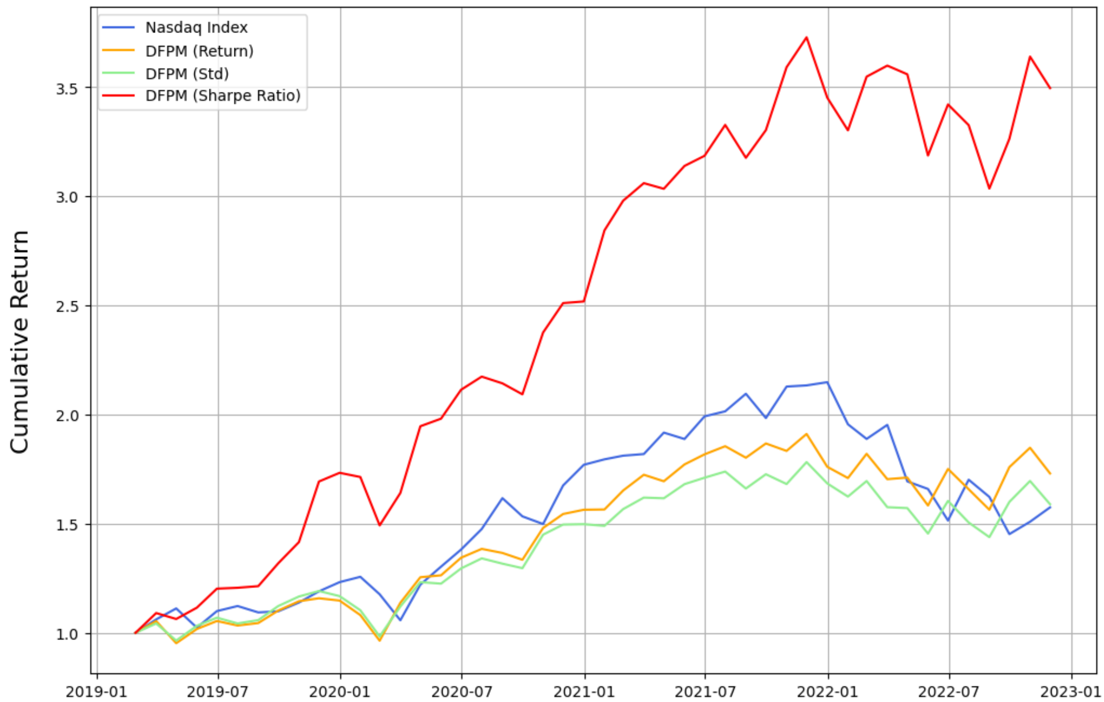


Fig. 10 Trend of returns based on the reward objective in Nasdaq100

Cross-Validation Results Analysis

To evaluate the robustness and generalizability of the proposed DFPM, cross-validation experiments were conducted across three overlapping test periods, comparing its performance against the Nasdaq 100 index using key metrics: MDD, Sharpe ratio, and fAPV. The results in Table. 11 demonstrate DFPM's strong risk-adjusted performance across diverse market conditions. In the first period (2016–2019), DFPM achieved a significantly lower MDD and a higher Sharpe ratio than the Nasdaq 100, indicating effective risk management and superior portfolio optimization. During the second period (2017–2020), although DFPM's MDD was higher, it maintained a competitive Sharpe ratio and significantly outperformed the index for fAPV. In the final period (2018–2021), DFPM exhibited a lower MDD of 17.73% while achieving an improved Sharpe ratio and fAPV, demonstrating adaptability to changing market conditions. These results highlight DFPM's ability to dynamically adjust factor importance and optimize portfolios under varying market conditions, consistently achieving higher fAPV and competitive Sharpe ratios. Although the higher MDD during 2017–2020 indicates room for improved volatility management, DFPM's overall performance validates its robustness and practical applicability as a portfolio optimization tool.

Table 11. Cross-Validation results: Comparison of DFPM and Nasdaq 100 index.

Test peorid	Model	MDD	Sharpe	fAPV
2016.02.01	Nasdaq 100 index	18.17	2.4744	1.6708
~2019.11.30	DFPM	9.81	2.6772	2.4715
2017.02.01	Nasdaq 100 index	12.13	3.0428	1.8470
~2020.11.30	DFPM	18.51	2.8732	3.1241
2018.02.01	Nasdaq 100 index	18.17	2.1610	1.8153
~2021.11.30	DFPM	17.73	2.3062	2.1610

Market Condition Analysis: DFPM Performance During Extreme Market Phases

The adaptability and robustness of the DFPM were rigorously tested under two contrasting market conditions: the sharp downturn during the COVID-19 pandemic (February–March 2020) and the subsequent recovery phase (April–June 2020). These periods served as critical benchmarks for evaluating the model's ability to adjust dynamically to extreme volatility and rapid market changes.

During the COVID-19 pandemic-induced market downturn, the NASDAQ index experienced a historic decline of 30.11% between February 1, 2020, and March 31, 2020. This period of severe disruptions and heightened market volatility provided an ideal scenario for assessing DFPM's ability to adapt to sudden market shocks. As presented in Table 12, DFPM exhibited remarkable stability, achieving the highest fAPV, highest Sharpe ratio, and competitive MDD. In contrast, benchmark strategies such as MVO, Uniform Allocation, Factor Separate, and Factor Mix experienced significantly higher drawdowns and lower Sharpe ratios, with some benchmarks even recording negative values. These results validate DFPM's ability to dynamically adjust its portfolio composition through real-time factor updates, underscoring its robustness and effectiveness in volatile market environments.

Following this sharp decline, the NASDAQ index experienced a strong recovery, gaining 27.34% between April 1, 2020, and June 30, 2020. Table 13 compares the performance of DFPM during the recovery phase to that of benchmark strategies. The proposed DFPM consistently outperformed all benchmarks, achieving the highest fAPV, highest Sharpe ratio, and zero MDD. In contrast, traditional models, such as MVO and Uniform Allocation, exhibited higher drawdowns and weaker performance metrics, while factor-based models, such as Factor Separate and Factor Mix, exhibited limited dynamic responsiveness. DFPM's superior performance during this phase demonstrates its ability to optimize portfolio allocations under rapidly improving market conditions.

DFPM consistently exhibited adaptability and robustness throughout these two extreme market phases. During the downturn, DFPM effectively mitigated losses and maintained stability by leveraging real-time factor updates. In the recovery phase, DFPM delivered superior performance metrics, highlighting its practical utility as an adaptive and reliable portfolio optimization tool. This analysis underscores DFPM's ability to navigate both unprecedented downturns and rapid recoveries, further solidifying its potential as an effective dynamic portfolio management approach.

Table 12. Model performance during the COVID-19 NASDAQ downturn (Feb–Mar 2020, -30.11%).

Model	fAPV	Share ratio	MDD
MVO	1.0246	0.0489	9.11
Uniform	1.0101	0.0302	11.32
Factor seperate	0.9744	-0.0365	10.72
Factor mix	0.9704	-0.0400	10.45
DFPM (Ours)	1.1390	0.1722	12.83

Table 13. Model performance during the COVID-19 NASDAQ recovery (Apr–Jun 2020, +27.34%).

Model	fAPV	Share ratio	MDD
MVO	1.1082	2.0245	2.91
Uniform	1.0971	2.4493	2.62
Factor seperate	1.0952	3.0603	2.02
Factor mix	1.0945	2.6241	1.60
DFPM (Ours)	1.1200	3.4969	0.00

Analysis of factor importance

The proposed model facilitates the extraction of factor importance, enabling factor-based interpretation of the results. Fig. 11 illustrates the factor importance results for the Nasdaq 100, and Fig. 12 shows the corresponding results for the Dow Jones. Factor importance analysis revealed distinct differences between the Nasdaq 100 and Dow Jones indices, reflecting their contrasting market dynamics. The Nasdaq 100 exhibited high variability in factor importance, with factors such as Quality, Beta, and Investment exhibiting significant shifts, particularly between the bearish and bullish phases. This dynamic behavior aligns with the Nasdaq's volatile, growth-oriented composition. Fig. 12 shows that Dow Jones exhibited more stable factor trends, with Value and Size consistently dominating across market conditions. This stability reflects Dow Jones's composition of large-cap, resilient companies that are less sensitive to short-term market fluctuations. Although the Nasdaq 100 benefitted significantly from DFPM's ability to dynamically adjust factor importance in response to rapid market changes, the Dow Jones required less frequent adjustments, emphasizing long-term macroeconomic stability. These findings highlight the DFPM's adaptability in capturing the Nasdaq's dynamic behavior while maintaining robust performance in the steadier Dow Jones market.

Leveraging Nasdaq 100 data that span various market phases, we analyzed changes in factor importance based on market conditions. Fig. 13 presents the corresponding portfolio allocation outcomes for Nasdaq 100. When analyzing market conditions, particularly during the bear market in early 2020 triggered by the COVID-19 pandemic (Phase Bear), the size factor exhibited relatively lower significance, indicating that smaller-sized stocks were overvalued. This observation supports financial theories that propose that smaller-sized stocks may harbor risk in adverse market conditions (Fama and French 1992; Fama and

French 1993; Jegadeesh et al. 1993; Lakonishok et al. 1994). In addition, the importance of value and investment factors increased during this period. Conversely, during the bullish market from mid-2021 to early 2022 (Phase Bull), the significance of value and investment factors diminished. Furthermore, the size factor gained greater importance, and the beta factor was relatively more significant than that in other periods. These observations align with financial factor theories (Fama and French 1992; Fama and French 1993; Jegadeesh et al. 1993), indicating that during a market upswing, the value and investment factors are devalued, whereas the smaller size and higher beta factors are highly regarded. The portfolio allocation results in Fig. 11 provide insights into the impact of these factors on portfolio composition across various market phases. Furthermore, Table 14 presents the top 10 stock lists based on average rankings for each factor during phases bear and bull.

Through a comprehensive analysis, it becomes clear that during Phase Bear, the DFPM model exhibited favored stocks with higher average value factor rankings. Notably, this preference was observed in the selection of equities, including IDEXX Laboratories Inc (IDXX), CDW Corporation (CDW), and Copart (CPRT). Furthermore, there was a distinct preference to allocate higher investments to stocks with elevated average investment factor rankings, such as Incyte Corporation (INCY) and CDW. Conversely, during phase bull, the DFPM model exhibited a preference for investing in stocks with higher average size factor rankings and investment factor rankings. Notably, it favored Lululemon Athletica Inc (LULU) and IDXX, both of which had elevated average size and investment factor rankings. In addition, a consistent trend emerged: higher investments were allocated to stocks with elevated size factor rankings, such as AMGN and Monster Beverage Corporation (MNST). This dynamic shift in investment focus based on market conditions highlights instances in which the model adeptly captured market opportunities and strategically allocated resources based on key factors.

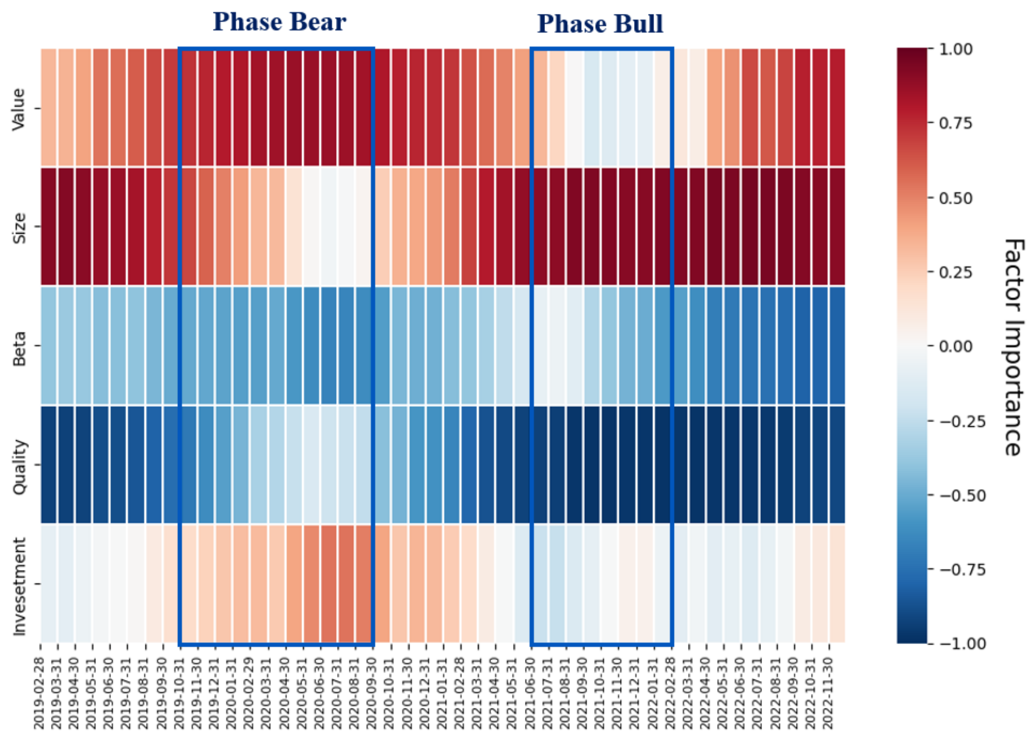


Fig. 11 Factor importance results for the Nasdaq 100

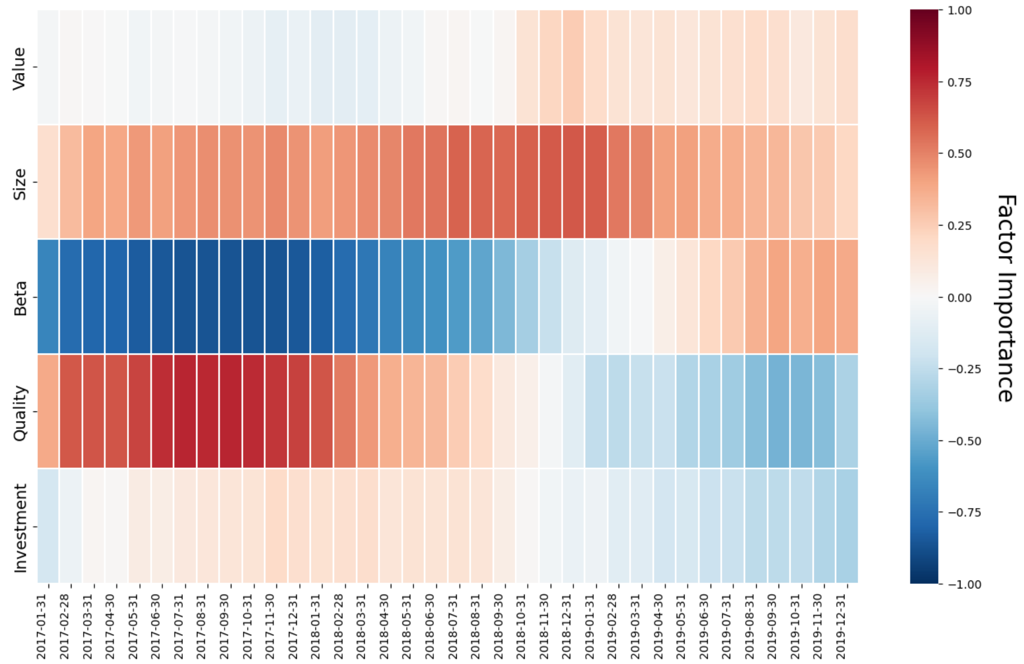


Fig. 12 Factor importance results for the Dow Jones

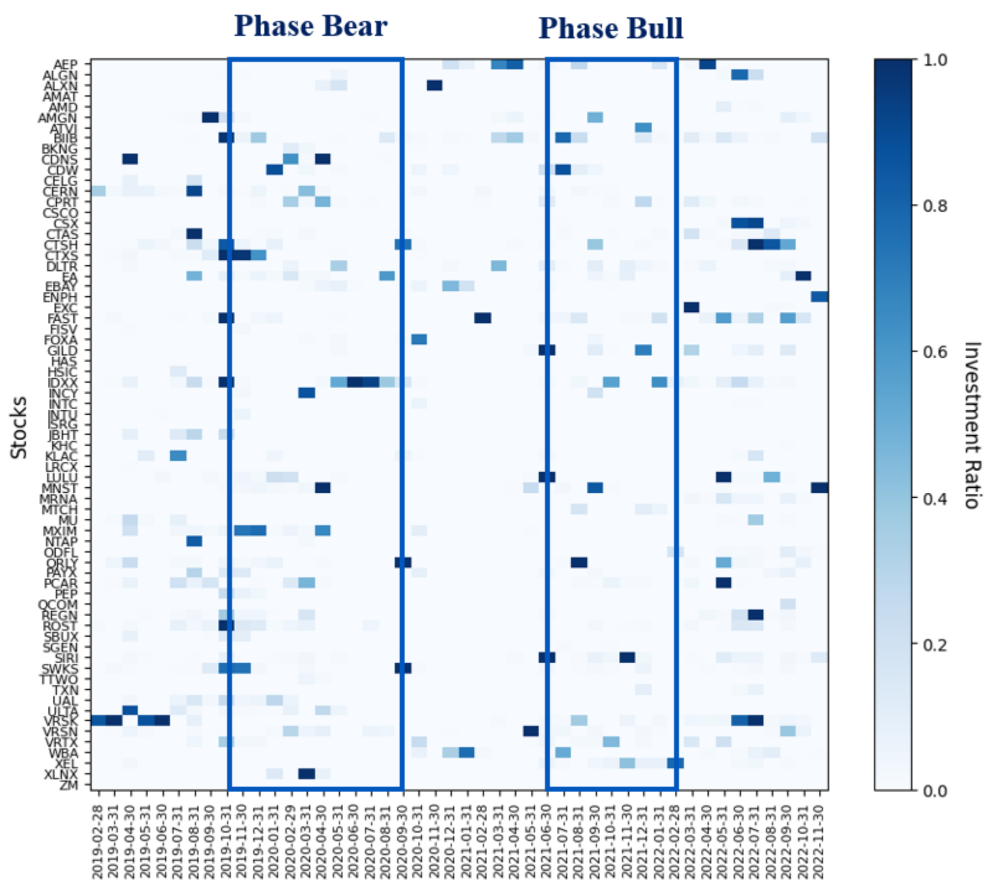


Fig. 13 Portfolio allocation results for the Nasdaq 100

Table 14. Top 10 stock lists of average rankings for each factor during phases.

Phase	Ranking	Value	Size	Beta	Quality	Investment
	1	SIRI	FISV	KLAC	CERN	WBA
	2	**IDXX**	REGN	SWKS	WBA	DLTR
	3	ORLY	BIIB	UAL	DLTR	UAL
	4	CTXS	MNST	CDW	CTSH	EA
	5	LULU	WBA	XLNX	VRSK	PCAR
Bear	6	**CDW**	EA	MXIM	VRSN	**INCY**
	7	ROST	ROST	ULTA	INCY	CTXS
	8	VRSK	LULU	PAYX	SWKS	**CDW**
	9	MXIM	CTSH	LULU	PCAR	PAYX
	10	**CPRT**	ORLY	VRSN	REGN	VRSK
Bull	1	ORLY	**AMGN**	KLAC	AEP	AEP

2	IDXX	GILD	SWKS	XEL	XEL
3	SIRI	CSX	**LULU**	CSX	PCAR
4	CDW	REGN	CPRT	CERN	FOXA
5	LULU	VRTX	**IDXX**	PCAR	DLTR
6	AMGN	**IDXX**	CDW	CTSH	CSX
7	PAYX	KLAC	CTAS	INCY	BIIB
8	KLAC	**MNST**	EBAY	BIIB	INCY
9	FAST	AEP	VRSN	DLTR	CTSH
10	CTAS	**LULU**	CTSH	VRSN	EBAY

*****The stocks mentioned in the text are boldfaced

Conclusion

In this study, we proposed the DFPM, a dynamic factor allocation portfolio RL model that integrates dynamic factor scores and stock price scores to improve portfolio performance. The proposed model leverages macroeconomic variables to understand market phase information and determine factor importance, resulting in superior results compared with conventional strategies and recent RL models. This finding validates the effectiveness of integrating traditional factor strategy knowledge into RL.

The effectiveness of the DFPM model was demonstrated through comprehensive experiments. The DFPM model, which integrates both the DFM and PSM modules, outperformed models utilizing only an individual module by a large margin across multiple evaluation metrics in both the Nasdaq 100 and Dow Jones markets. Additionally, applying dynamic factor scoring to traditional factor portfolio methods further confirmed its effectiveness, although some performance variations were observed due to the focus on optimizing the Sharpe ratio.

When compared with traditional portfolio strategies, the DFPM model consistently outperformed in both the Nasdaq 100 and Dow Jones markets, showing marked improvements in Sharpe ratio and fAPV. It demonstrated enhanced growth during bullish trends and stability during bearish periods. Specifically, in the Dow Jones market, the DFPM significantly reduced MDD while increasing the Sharpe ratio and fAPV.

Furthermore, the DFPM exhibited superior performance compared to recent RL methods, particularly in terms of Sharpe ratio, fAPV, and MDD. The dynamic scoring of factors and effective fusion of price information were critical to this success. The model's ability to dynamically adjust factor importance based on market conditions provided deep insights into market behaviors and portfolio optimization, aligning well with established financial theories and enhancing the model's practical applicability.

Despite its promising results, this study has certain limitations that offer opportunities for future research. Although the TA-LSTM architecture in the DFM captures long-term dependencies, it encounters challenges such as computational demands and risks of overfitting. Future studies could explore the use of efficient attention mechanisms, dimensionality reduction techniques, and regularization methods to address these issues. Although transaction costs were incorporated into the analysis, slippage—a critical factor in real-world trading performance—was not explicitly considered. Integrating slippage models and assessing the DFPM under simulated market conditions can improve its robustness and practical applicability. Additionally, the study is restricted to five factors and long-only portfolios. Expanding the factor set to include volatility, momentum, and dividend yield and validating the model across diverse markets can enhance its generalizability. Adapting to sudden market shocks remains another challenge; the incorporation of online learning methods and frequent retraining could further improve the model's responsiveness to abrupt market changes. Finally, although the dual-module structure of the proposed DFPM is effective, it poses interpretability challenges. Advanced methods, such as Shapley Additive Explanations (SHAP) by Lundberg and Lee (2017), which provides consistent and fair feature importance measures, and Local Interpretable Model-agnostic Explanations (LIME) by Ribeiro and Guestrin (2016), which generates local explanations for individual predictions, can enhance transparency and build confidence in real-world applications. Addressing these limitations will enable DFPM to evolve into a more robust, adaptive, and practical tool for portfolio optimization, offering valuable insights to the financial industry.

Abbreviations

CAPM	Capital Asset Pricing Model
DL	Deep Learning
RL	Reinforcement Learning
DFPM	Dynamic Factor Portfolio Model
DFM	Dynamic Factor Module
PSM	Price Score Module
SOTA	State-of-the-art
MPT	Modern Portfolio Theory
APT	Arbitrage Pricing Theory
PPO	Proximal Policy Optimization
A2C	Advantage Actor Critic
DDPG	Deep Deterministic Policy Gradient
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-term Memory
TA-LSTM	Temporal Attention-LSTM
MDP	Markov Decision Process
U.S.	United States of America
CPI	Consumer Price Index

FRED	Federal Reserve Economic Data
WRDS	Wharton Research Data Services
ROE	Return on Equity
ROA	Return on Assets
MVO	Mean-variance Optimization
MDD	Maximum Drawdown
fAPV	Final Accumulative Portfolio Value
IDX	IDEXX Laboratories Inc
CDW	CDW Corporation
CPRT	Copart
INCY	Incyte Corporation
LULU	Lululemon Athletica Inc
MNST	Monster Beverage Corporation

References

Markowitz H (1952) Portfolio Selection. Journal of Finance 7(1):77-91. https://doi.org/10.2307/2975974

Sharpe WF (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance 19(3):425-442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In Stochastic optimization models in finance:131-155. https://doi.org/10.1016/B978-0-12-780850-5.50018-6

Mossin J (1966) Equilibrium in a Capital Asset Market. Econometrica: Journal of the econometric society:768-783.

L'Her, Jean-Francois, Oumar S, Mohamed YT (2002) Country, industry, and risk factor loadings in portfolio management. Journal of Portfolio Management 28(4):70.

Jensen MC (1969) Risk, the pricing of capital assets, and the evaluation of investment portfolios. The Journal of business 42(2):167-247.

Bertola G. (1988) Adjustment costs and dynamic factor demands: investment and employment under uncertainty (Doctoral dissertation, Massachusetts Institute of Technology).

Li Y, Jiang S, Li X, Wang S (2022) Hybrid data decomposition-based deep learning for bitcoin prediction and algorithm trading. Financial Innovation 8(1):31. https://doi.org/10.1186/s40854-022-00336-7

Roberts P, Hamilton EL (2023) Portfolio management: A new direction in public sector strategic management research and practice. Public Administration Review 83(4):947-959. https://doi.org/10.1111/puar.13633

Htun HH, Biehl M, Petkov N (2024) Forecasting relative returns for S&P 500 stocks using machine learning. Financial Innovation 10(1):118. https://doi.org/10.1186/s40854-024-00644-0

Wang J, Wang S, Lv M, Jiang H. (2024) Forecasting VaR and ES by using deep quantile regression, GANs-based scenario generation, and heterogeneous market hypothesis. Financial Innovation 10(1): 36. https://doi.org/10.1186/s40854-023-00564-5

Heaton JB, Polson NG, Witte JH (2017) Deep learning for finance: deep portfolios. Applied Stochastic Models in Business and Industry 33(1):3-12. https://doi.org/10.1002/asmb.2209

Wang Z, Huang B, Tu S, Zhang K, Xu L (2021) DeepTrader: a deep reinforcement learning approach for risk-return balanced portfolio management with market conditions Embedding. In Proceedings of the AAAI conference on artificial intelligence 35(1):643-650.

Ma Y, Han R, Wang W (2021) Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications 165:113973. https://doi.org/10.1016/j.eswa.2020.113973

Li Y, Jiang S, Wei Y, Wang S (2021) Take Bitcoin into your portfolio: a novel ensemble portfolio optimization framework for broad commodity assets. Financial Innovation 7(1):63. https://doi.org/10.1186/s40854-021-00281-x

Chen CH, Chiang BY, Hong T P, Wang DC, Lin JCW, Gankhuyag M (2019). A fuzzy GGA-based approach to speed up the evolutionary process for diverse group stock portfolio optimization. Journal of Intelligent & Fuzzy Systems 37(6):7465-7479. https://doi.org/10.3233/JIFS-179354

Chen CH, Chen YH, Diaz VG, Lin JCW (2021) An intelligent trading mechanism based on the group trading strategy portfolio to reduce massive loss by the grouping genetic algorithm. Electronic Commerce Research 23(1):3-42. https://doi.org/10.1007/s10660-021-09467-y

Wu ME, Syu JH, Lin JCW, Ho JM (2021) Portfolio management system in equity market neutral using reinforcement learning. Applied Intelligence:1-13. https://doi.org/10.1007/s10489-021-02262-0

Yang H, Liu XY, Zhong S, Walid A. (2020) Deep reinforcement learning for automated stock trading: An ensemble strategy. In Proceedings of the first ACM international conference on AI in finance:1-8.

Sun Q, Wei X, Yang X. (2024) GraphSAGE with deep reinforcement learning for financial portfolio optimization. Expert Systems with Applications 238:122027. https://doi.org/10.1016/j.eswa.2023.122027

Sewell M (2011) Characterization of financial time series. Rn 11(01):01.

Engle R (2004) Risk and volatility: Econometric models and financial practice. American economic review 94(3):405-420. https://doi.org/10.1257/0002828041464597

Ross SA (1976) The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13(3):341-360. https://doi.org/10.1142/9789814417358_0001

Asness C, Liew J (2011) The Volatility Effect: Lower Risk Without Lower Return. Journal of portfolio management:102-113.

Maillard S, Roncalli T, Teïletche J (2010) The properties of equally weighted risk contribution portfolios. The Journal of Portfolio Management 36(4):60-70.

Choueifaty Y, Coignard Y (2008) Toward maximum diversification. The Journal of Portfolio Management 35(1):40-51.

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1):3-56. https://doi.org/10.1016/0304-405X(93)90023-5

Carhart MM (1997) On Persistence in Mutual Fund Performance. The Journal of Finance 52:57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Fama EF, French KR (2015) A five-factor asset pricing model. Journal of Financial Economics 116(1):1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Baker M, Bradley B, Wurgler J (2011) Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. Financial Analysts Journal 67(1):40-54. https://doi.org/10.2469/faj.v67.n1.4

Jang J, Seong N (2023) Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. Expert Systems with Applications 218:119556. https://doi.org/10.1016/j.eswa.2023.119556

Wang J, Zhang Y, Tang K, Wu J, Xiong Z (2019) Alphastock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining:1900-1908.

Ta VD, Liu CM., Tadesse DA (2020) Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. Applied Sciences 10(2):437. https://doi.org/10.3390/app10020437

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735-1780.

Zhang T, Wang Q, Zhang P, Wang Y (2022) A review on the attention mechanism of deep learning. Neurocomputing 509:58–78. https://doi.org/10.1016/j.neucom.2022.07.013

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NIPS

2017), pp. 5998–6008.

Akita R, Yasuda Y, Uchida Y (2019) Time series analysis on financial markets using deep learning. International Journal of Economics, Commerce and Management 7(1):23-34.

Boltzmann L (1868) Studien uber das gleichgewicht der lebenden kraft. Wissenschafiliche Abhandlungen 1:49-96.

Bellman R (1957) A Markovian decision process. Journal of Mathematics and Mechanics 6:679-684.

Paliienko O, Naumenkova S, Mishchenko S (2020) An empirical investigation of the Fama-French five-factor model. Investment Management and Financial Innovations 17(1):143–155.

Hapsari C, Wasistha GH (2017) Portfolio formation using the Fama-French five-factor model with modification of a profitability variable: An empirical study on the Indonesian stock exchange. In Competition and Cooperation in Economics and Business. Routledge:83-88

Magdon-Ismail M, Atiya AF, Pratap A (2004) On the maximum drawdown of a Brownian motion. Journal of Applied Probability 41(1):147-161.

Fama EF, French KR (1992) The cross-section of expected stock returns. The Journal of Finance 47(2):427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance 48(1):65-91.

Lakonishok J, Shleifer A, Vishny RW (1994) Contrarian investment, extrapolation, and risk. The Journal of Finance 49(5):1541-1578. https://doi.org/10.1111/j.1540-6261.1994.tb04772.x

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778