Journal of Trends in Financial and Economics
ISSN: 3007-696X
DOI: https://doi.org/10.61784/jtfe3044

THE APPLICATION OF DEEP REINFORCEMENT LEARNING IN ASSET ALLOCATION: A THEORETICAL FRAMEWORK AND EMPIRICAL ANALYSIS

ZiLin Zhou
La Salle College Preparatory High School, Pasadena 91107, United States.
Corresponding Email: zilinzhou08@gmail.com

Abstract: Asset allocation is a fundamental challenge in investment management, traditionally addressed through models such as mean-variance optimization. However, dynamic market environments and multi-period investment horizons limit the effectiveness of static methods. In recent years, the emergence of deep reinforcement learning (DRL) has provided a powerful tool for addressing complex sequential decision-making problems in finance. This paper conducts a comprehensive academic analysis of the application of DRL in asset allocation. First, we introduce the asset allocation problem and its challenges, then review the basic concepts of DRL and its relevance to financial decision-making. Next, we propose a theoretical framework for transforming the asset allocation problem into a Markov decision process and describe in detail how DRL agents learn optimal investment strategies under various assumptions and structures within this framework. Subsequently, through a review of foreign academic literature, this paper examines existing findings on the application of DRL in asset allocation from a qualitative perspective, including the superior performance of DRL strategies relative to traditional methods in certain scenarios and cautionary results where DRL remains competitive even under simple benchmarks. We discuss the current limitations of DRL methods, high transaction costs, and potential directions for improvement and future research priorities. The study concludes that while DRL holds great potential for enhancing asset allocation theory and practice, several key practical challenges must be addressed before its full potential can be realized.
Keywords: Asset allocation; Deep Reinforcement Learning (DRL); Markov decision process

1 INTRODUCTION

Asset allocation is the process of distributing capital across different asset classes and is a core component of portfolio management. Its objective is to achieve the optimal balance between expected returns and risk. The modern portfolio theory proposed by Markowitz (1991) first provided a rigorous quantitative framework for this problem. Markowitz's mean-variance model aims to maximize expected returns given a certain level of risk, and this model remains the foundation of finance to this day[1]. However, Markowitz's method is essentially a single-period model, assuming that asset returns and investor preferences remain unchanged throughout the decision-making period. In practice, investors face multi-period decision-making and evolving markets[1]. Merton (1975) extended portfolio theory to continuous-time dynamic environments, introducing the concept of dynamic asset allocation[2]. Merton's framework emphasizes that optimal asset weights need to adjust over time as conditions change; his research demonstrated that solutions to multi-period asset allocation problems may differ significantly from those of single-period problems[2].

The dynamic characteristics of real-world markets—including changes in correlations, transitions between market states, and the presence of transaction costs—add complexity to asset allocation. Traditional analytical solutions for multi-period portfolio optimization often require strong assumptions or simplifications to maintain solvability. This has sparked interest in numerical computation and algorithmic methods to address more realistic complex scenarios. Reinforcement learning (RL), a branch of machine learning, enables agents to learn sequential decisions through trial-and-error feedback, offering a promising framework for addressing such problems. Unlike traditional optimization, RL can, in principle, learn strategies adaptively through interaction with the environment without requiring an analytical solution to the problem, thereby potentially addressing the intractable aspects of investor optimization problems.

This paper examines the application of DRL in asset allocation from both theoretical and empirical perspectives. We first outline the core concepts of DRL and its role in financial decision-making. Then, we construct a theoretical framework for the application of DRL in asset allocation, discussing the definition of states, actions, and rewards, as well as model assumptions. Next, we review the qualitative empirical findings of existing academic research on the application of DRL in asset allocation, summarizing the research approaches and main findings. Finally, we discuss the limitations of current methods and propose directions for future improvements and research. Through a comprehensive literature review, we aim to clarify how and to what extent DRL can advance the practice and theory of asset allocation.

2 OVERVIEW OF DEEP REINFORCEMENT LEARNING AND ITS APPLICATION IN FINANCIAL DECISION-MAKING

Reinforcement learning is a machine learning paradigm in which an agent learns optimal policies by interacting with an environment and receiving cumulative rewards[3]. Formally, such problems are often modeled as Markov decision

processes (MDPs), consisting of states, actions, state transition probabilities, and reward functions. Unlike supervised learning, which relies on labeled data for learning, RL learns from the outcomes of actions: the agent receives reward signals based on the actions taken at each step and adjusts its strategy accordingly to gradually improve the long-term cumulative reward. In a financial context, the RL agent can be viewed as a trading or allocation strategy, where the state represents market information, the action corresponds to portfolio adjustments, and the reward is a measure of investment performance. This framework is naturally suited to sequential decision-making problems, making it well-suited to multi-period asset allocation, where investment decisions at different time points interact with each other.

The application of DRL in financial decision-making has been explored in multiple domains, including portfolio allocation, trading (market timing), option pricing, and order execution. Specifically in asset portfolio management, numerous studies have shown that DRL agents outperform some heuristic or static strategies due to their ability to time decisions or exploit complex dependencies. DRL algorithms applied to learning trading strategies include deep Q-networks, policy gradient methods, and actor-critic architectures. These algorithms must address unique challenges in financial settings, including the low signal-to-noise ratio in financial data and risk management requirements. However, the flexibility of RL frameworks—such as the ability to incorporate risk considerations directly into the reward function or as constraints—offers significant opportunities for innovation in asset allocation methods.

3 THEORETICAL FRAMEWORK: APPLICATION OF DRL IN ASSET ALLOCATION

Applying deep reinforcement learning to asset allocation requires transforming the problem into a RL form. This involves defining the environment, state space, action space, reward function, and learning mechanism:

State (S): The state should contain the information needed to make the optimal allocation decision. In an asset allocation environment, the state at time t can include recent asset prices or returns, macroeconomic indicators, volatility estimates, and any other relevant market characteristics. For example, the state can be a vector containing the returns of each asset over the past N days, plus some macroeconomic variables or market sentiment indicators. Some frameworks distinguish between fully observed and partially observed states; in finance, we typically face partial observation, but we construct the state vector based on the available information. Recent research often uses deep networks such as CNNs or recurrent neural networks (RNNs) to process historical price data and extract useful features as the state representation for DRL agents.

Action (A): An action refers to the portfolio decision made by the agent. In simplified cases, actions can be discrete choices. However, in real asset allocation, actions are more naturally represented as a continuous vector—the investment weights of each asset. Many DRL methods treat actions as continuous variables. This typically requires the use of policy gradients or actor-critic algorithms. The action space may be constrained: for example, weights may be required to be non-negative and sum to 1, or leverage levels may be restricted. These constraints can be satisfied in implementation by transforming the network outputs.

Reward (R): The definition of reward is a key design element that influences the learned strategy. In portfolio management, a natural reward is the change in portfolio value over each period. For example, after executing an action at time t, the portfolio return from t to t+1 can serve as the reward at time t. Thus, the cumulative reward over a complete episode corresponds to total return or compound growth. Some studies adopt risk-adjusted rewards: for example, penalizing volatility or maximum drawdown in the reward function to implicitly consider risk. It is worth noting that Benhamou et al. (2024) provide theoretical insights showing that if an RL agent is myopic and its reward is determined solely by the first and second moments of returns, the converged strategy of the agent is the Markowitz mean-variance portfolio[4]. In other words, when the reward is appropriately chosen, the classical Markowitz portfolio can be viewed as a special case of RL one-step optimization. By extending the reward to multiple periods and incorporating more information, agents can theoretically find strategies that outperform the static Markowitz solution[1].

Environmental dynamics: In RL, the state transition rules of the environment do not require prior knowledge by the agent but are critical for the simulation process during training. In asset allocation, the environment is essentially the market. For model-free DRL, we do not need an explicit market model; instead, we can sample state transitions using historical or simulated data. A common approach is to use rolling windows of historical time series as training episodes. Each episode may correspond to a fixed time interval, during which the agent "virtually trades." State transitions are determined by the actual market trends: the asset price in the next state is obtained from actual observations. In simulation-based methods, a generative model can also be used to generate synthetic asset paths for training to increase the number of training samples. However, if the simulation generator is not accurate enough, the intelligent agent may learn to exploit biases in the simulated environment rather than effective patterns in the real market, introducing model risk.

Assumptions: When modeling asset allocation as RL, a key assumption is the Markov property—that is, future state transitions depend only on the current state and action, and are independent of more distant history. Financial markets are not strictly Markovian, but we select appropriate state representations to satisfy the Markov condition as much as possible. Another assumption is that the statistical properties of the environment are stationary during training: the agent typically assumes that the data distribution experienced during training and execution remains consistent. However, in reality, markets evolve, meaning that an RL strategy trained during one period may perform poorly when market conditions change, unless the strategy can adapt on its own. Some theoretical frameworks partially address non-stationarity by explicitly incorporating time or state indicators into the state space, such as incorporating labels

representing the macroeconomic environment.

Learning algorithms: Under the above assumptions, various algorithms can be used to train DRL agents. Since the action space for asset allocation is mostly continuous, policy gradient methods are commonly used. The agent's policy is represented by a neural network that maps states to actions. In each training step, the agent takes an action in a given state, observes the reward and the next state, and then updates the network parameters to improve the expected future cumulative reward. The objective function is typically to maximize the expected cumulative return. Some implementations use a discount factor to reduce the weight of future rewards, although in investment, a more natural objective is often the undiscounted total return or terminal wealth.

The above theoretical framework essentially transforms asset allocation into a “game” scenario: the agent “plays” against the market, with the goal of maximizing its own wealth. If training is successful, the resulting strategy can exhibit complex behavioral patterns, such as timing and dynamic rebalancing. These behaviors are not pre-programmed through human rules but are autonomously discovered by the agent through learning if they can improve rewards. For example, even without explicitly instructing the agent on momentum or hedging strategies, it may independently learn to increase allocations to strong assets during upward trends or hold more risk-averse assets during periods of high correlation based on reward feedback.

On this basic framework, researchers have proposed several improvements. One is to incorporate transaction costs into the environment and rewards. Transaction costs can be modeled by imposing penalties on portfolio rebalancing behavior. This is crucial because strategies that appear profitable when costs are ignored may become unfeasible once costs are factored in. Another improvement is allowing the setting of risk aversion levels: for example, using logarithmic utility or mean-variance utility functions as rewards. Adjusting the reward function yields a range of strategies with different risk preferences, analogous to points on the efficient frontier. Jiang, Olmo, and Atwi (2025) explicitly incorporated risk aversion parameters into the DRL framework and demonstrated how strategy aggressiveness varies with different parameter settings[5].

In summary, the theoretical framework for modeling asset allocation as a DRL problem is as follows: an agent observes market states and decides portfolio weights to maximize cumulative returns. Within this framework, traditional strategies can be viewed as special cases. The flexibility of this framework allows for the incorporation of numerous real-world factors, making DRL a powerful theoretical tool for searching for improved allocation strategies in complex environments.

4 EMPIRICAL ANALYSIS: A REVIEW OF RESEARCH ON THE APPLICATION OF DRL IN ASSET ALLOCATION

Utilizing alternative data and cross-sectional information: Some studies have incorporated data beyond prices into the DRL framework. Aboussalah, Xu, and Lee (2021) explored the value of cross-sectional learning methods. In their paper published in Quantitative Finance, the agents learn not only from the time series of a single portfolio but also from the overall market cross-sectional data of numerous assets, thereby enabling the strategy to be more generalizable across a wider range of market conditions and assets[6]. They found that this cross-sectional training improved performance, indicating that information extracted from a broader market can benefit allocation agents. Compared to agents trained solely on the historical data of a single asset, agents trained using cross-sectional data demonstrated superior risk-adjusted performance in terms of returns.

Chen and Ge (2021) proposed a “learning-based strategy” for portfolio selection[7]. At its core, they introduced an algorithm in the International Journal of Economics and Finance that can adapt to changing market conditions. Their research emphasizes that a data-driven strategy is more resilient than static models under changing market conditions. Although the paper details are somewhat vague here, the key conclusion is that machine learning/DRL-driven strategies can continue to learn from new data, adjust themselves when traditional models fail, and remain effective[7].

Incorporating market sentiment and macroeconomic context: Financial markets are not only influenced by historical prices but also driven by investor sentiment and macroeconomic news. Recognizing this, Wei et al. (2021) incorporated asymmetric investor sentiment as part of the state in their RL portfolio model[8]. By using sentiment indicators, their agents could anticipate market changes that could not be predicted solely based on price history. Their research in the Journal of Expert Systems and Applications showed that DRL agents using sentiment data achieved better performance, especially during periods of market stress or euphoria: the agents were able to reduce positions before a decline and increase positions during a recovery. This result highlights the flexibility of DRL in integrating diverse data types; in principle, RL agents can learn to interpret sentiment indicators and price trends comprehensively to make better allocation decisions.

Macroeconomic variables or regime indicators are also important contextual information. Some studies provide agents with contextual data beyond asset prices, such as interest rates, volatility indices, or macroeconomic cycle labels. Benhamou et al. found that providing this additional information improves the performance of DRL models relative to traditional mean-variance strategies. In fact, it has been demonstrated that agents can utilize macroeconomic environment data to adjust their strategies. However, they also note that this comes with increased complexity—the more information the agent considers and the more forward-looking it is, the more challenging the training process becomes, though the potential benefits are greater.

Performance comparison with traditional methods: A common theme in many empirical studies is that DRL-driven asset allocation strategies often outperform traditional strategies in backtesting. For example, Jiang, Olmo, and Atwi (2025)

designed a DRL agent combining CNN and WaveNet components to handle high-dimensional portfolios and multi-period problems[5]. They tested the strategy under various market conditions, risk aversion levels, and with real transaction costs factored in. The results showed that DRL strategies outperformed multiple benchmark methods in terms of both returns and adaptability. This suggests that, under conditions of ample training data and carefully designed models, DRL can identify trading patterns and portfolio adjustments overlooked by static models, thereby achieving a better risk-return tradeoff.

Challenges and Differentiating Results: While many reports are positive, some important studies have highlighted potential shortcomings of DRL. Kruthof and Müller (2025) provide a cautionary example. These authors conducted a rigorous evaluation of state-of-the-art DRL algorithms (Soft Actor-Critic, SAC) in the Financial Research Letters, using a sample spanning seven stock markets and a total of 300 years. Interestingly, they found that in a no-transaction-cost scenario, DRL agents did exhibit some timing ability—outperforming market benchmarks during certain periods—but overall, they did not systematically outperform a simple 1/N equal-weight strategy. More notably, when introducing a mild transaction cost of 0.1%, the DRL strategy, due to its high turnover rate, resulted in negative net returns and performed worse than the 1/N strategy. The latter, which rarely adjusts its portfolio, is largely unaffected by transaction costs and thus significantly outperforms the DRL strategy when costs are considered. They also examined return distributions and tail risk metrics and found no consistent advantage for the DRL agents. This study highlights that a complex AI strategy does not necessarily outperform simple rules, especially when real-world frictions are incorporated. Its conclusions emphasize that without careful design, DRL strategies may lose their advantage due to over-trading or overfitting to noise, and require cost awareness and rigorous validation to enhance their practicality.

In summary, the empirical literature to date paints an encouraging but cautious picture. Many studies have documented performance improvements of DRL asset allocation strategies relative to traditional methods, which are attributed to DRL's ability to leverage complex data and continuously optimize strategies. However, the best results often come from carefully designed models that incorporate domain knowledge. At the same time, reliable evaluation is crucial—simple DRL models, if mishandled, may suffer from overfitting or over-trading issues, as revealed by some rigorous tests. Therefore, it is necessary to gain a deeper understanding of these limitations, which we will further discuss and propose future improvement directions in the next section.

5 CONCLUSION AND DISCUSSION: LIMITATIONS, POTENTIAL, AND FUTURE DIRECTIONS

The application of deep reinforcement learning to asset allocation remains an emerging field with significant potential but also clear limitations. Based on the theoretical framework and empirical findings discussed above, we summarize several key issues and propose possible directions for future research.

5.1 Limitations and Challenges

Market non-stationarity: Financial markets are non-stationary—their statistical properties change over time. DRL algorithms typically assume that the statistical properties of the environment remain constant during training. Strategies learned under one market regime may perform poorly when conditions change. This introduces the risk of overfitting to historical data. Many published DRL strategies perform well in-sample or during specific backtesting periods but may not generalize to new data or different market environments. Hambly et al. (2023) review that fat-tailed return distributions and regime switching pose fundamental challenges for RL in financial settings. To mitigate this issue, future methods can incorporate regime detection techniques or adopt meta-learning to enable agents to adjust themselves when detecting new regimes. Continuous learning frameworks or periodically retraining agents when new data arrives may be necessary to maintain the effectiveness of strategies.

Sample efficiency and data scarcity: Unlike games, financial data is constrained by historical length—markets have only a limited number of past years available for learning, and underlying processes are complex and variable. DRL algorithms typically require a large number of training samples to converge to a good strategy, which is problematic in finance because each episode is unique and not independently and identically distributed. Researchers have attempted to expand data through bootstrapping or simulation, but simulation data must be used cautiously; if the simulation model is overly simplified, the agent may learn strategies that rely more on the characteristics of the simulation environment rather than effective signals from the real market. This requires the development of RL algorithms with higher sample efficiency or algorithms that can incorporate prior knowledge. Model-based RL is a potential direction: the agent first learns a model of market dynamics and then uses this model for planning to find optimal strategies with fewer real samples. However, learning an accurate market model itself is also quite challenging.

Exploration and exploitation in real trading: In typical RL, exploration is necessary to discover the optimal strategy. However, in real trading, exploration means intentionally taking suboptimal actions to learn environmental characteristics, which is unacceptable to investors. Therefore, in practice, DRL often relies on offline training followed by online execution of learned strategies. This means that the agent may not have encountered scenarios that perfectly match those it will face, and it cannot “trial and error” extensively in real trading without incurring financial losses. Future research could explore safe exploration methods or construct highly realistic market simulation environments to allow agents to practice and refine strategies without directly exposing real capital to risk. For example, generative adversarial networks could be used to simulate realistic market scenarios for agent training, better preparing them for real-world trading environments.

High Turnover Rate and Transaction Costs: As emphasized by Kruthof and Müller (2025), DRL strategies may suffer from over-trading issues. An unconstrained RL agent may pursue every perceived short-term opportunity, frequently rebalancing its portfolio[9]. This can lead to extremely high turnover rates, which, when accounting for transaction costs, may erode or even offset all gross returns. High turnover rates may also introduce tax burdens and other frictions that may not be adequately modeled in research. For practical applications, it is essential to incorporate transaction cost constraints or penalties into DRL strategies during training. Future research could integrate more refined cost models into RL environments. Additionally, regularization or adding costs to actions in the reward function could encourage agents to avoid frequent portfolio rebalancing unless the expected return is significant.

Risk Management and Tail Risk: Many DRL implementations optimize for average returns, but in investing, downside risk is a primary concern. Agents may unintentionally learn strategies that yield high average returns but expose the portfolio to rare, massive losses. Traditional portfolio management often sets risk limits or uses risk measures such as VaR. Current DRL frameworks may not address these concerns unless explicitly incorporated through reward functions or constraints. This requires the development of risk-sensitive DRL that incorporates risk elements into the agent's objectives. Some approaches may include optimizing CVaR in the objective or using multi-objective RL to optimize both return and risk metrics simultaneously. It is also necessary to ensure that agents adhere to risk limits even under extreme market conditions. This is critical for the robustness of strategy implementation in practice.

Interpretability and trust: Deep learning models are typically “black boxes.” In the financial sector, a completely unexplainable strategy is difficult to trust, especially when fund managers or investment committees need to understand and trust the decision-making process, and regulators require explanations of trading logic. Unlike the Markowitz model, which explicitly shows the trade-offs between risk and return, the rationale behind the investment decisions output by DRL strategies is difficult to explain in human language. This opacity may hinder its application. Therefore, the application of explainable artificial intelligence (XAI) in finance has become an important direction. Cong et al. (2022) incorporated explainability into DRL models in their AlphaPortfolio study, using techniques to make the decision-making basis of agents more transparent[10]. Future research can focus on extracting rules or important factors from trained DRL agents—for example, using sensitivity analysis to determine which input features have the greatest impact on agent decisions. In summary, improving the explainability of DRL decisions while maintaining its flexibility is crucial for its adoption in real-world investment.

5.2 Potential and Future Directions

Despite the challenges, DRL holds significant potential for asset allocation. With increasing computational power and access to more data, DRL methods are likely to continue improving. We anticipate the following directions for enhancing DRL in asset allocation:

Benchmarking and Reproducibility: The academic community will benefit from establishing standard benchmark data and environments for portfolio management tasks. This will enable fair comparisons between different algorithms and drive progress. Consistent evaluation protocols need to be developed, including testing under multiple market scenarios and considering transaction costs. As emphasized by Kruthof and Müller (2025), rigorous and robust validation protocols are essential[9]. Through community-agreed benchmarks, researchers can better identify which improvements truly lead to performance gains.

Contextualization and meta-reinforcement learning: Making DRL agents more “context-aware” can improve their robustness. For example, contextual reinforcement learning involves providing agents with information about the current market state. Agents can learn different sub-strategies for different contexts. Another example is meta-learning algorithms, which enable agents to learn how to quickly adjust their learning process when encountering new environments—essentially “learning how to learn.” This is analogous to an investment strategy that knows when its conventional methods may fail and adjusts its behavior accordingly. Through meta-learning, an agent might adapt to entirely new market characteristics after observing only a small amount of new data, which would be highly valuable for handling sudden shifts in market structure.

Multi-agent reinforcement learning: Markets are composed of numerous interacting participants. Single-agent RL frameworks treat the market as part of the environment, but an interesting extension involves including multiple RL agents in the model that compete against each other or interact with models of other agents. Multi-agent RL can be used to simulate a market where some participants are also RL-driven, observing how they co-evolve. Although complex, this may reveal some equilibrium behaviors or demonstrate how RL agents can exploit or coordinate with other strategies. It can also be used to characterize the game-theoretic aspects of markets. Additionally, multi-agent frameworks can simulate investor-environment interactions, such as agents interacting with market makers or with several typical trading strategies, which may help in understanding strategy robustness.

Integration with traditional methods: Rather than viewing DRL as an alternative to traditional portfolio optimization, it is more productive to consider their integration. For example, a hierarchical approach could be adopted: RL agents determine high-level allocations, while traditional methods handle more granular-level allocations. Conversely, traditional models could guide the exploration of RL agents. This hybrid approach may combine the strengths of both—the theoretical robustness of classical models and the adaptability of RL.

Improved training algorithms: From an algorithmic perspective, methodological innovations in the field of reinforcement learning can also be applied to financial scenarios. For example, more stable training methods, more effective exploration strategies, or distributed RL are all worth exploring. For example, distributed RL can be used to

characterize the entire distribution of portfolio returns, thereby adapting to risk management objectives. Another example is safety-constrained reinforcement learning, a hot topic in recent years that aims to ensure that RL agents adhere to safety constraints during training. When mapped to asset allocation, this means ensuring that risks do not exceed certain thresholds or losses do not exceed certain thresholds during training, which helps to obtain strategies that are optimized under constraints and is beneficial for financial applications.

Actual deployment and feedback: Finally, an important development step is to deploy DRL strategies in real trading and obtain feedback. Real-world performance can inform research: Analyzing cases of strategy success and failure can help improve models. The growing interest in artificial intelligence within the financial industry may spur more collaboration between academia and industry, providing data, computing power, and expertise to advance the application of DRL in asset allocation. With real-market validation and data, researchers can calibrate models to better align with practical needs.

5.3 Conclusions

In summary, applying deep reinforcement learning to asset allocation provides a rich and flexible framework for this field, addressing the uncertainty and dynamics that traditional methods struggle with from a sequential decision-making perspective. Modeling asset allocation as an MDP enables us to unify classical methods with modern machine learning techniques, revealing that traditional solutions are merely special cases within the broader RL paradigm. Empirical studies provide encouraging evidence: carefully designed DRL agents can adapt to complex market patterns and sometimes outperform static or short-sighted strategies. These agents are capable of processing large information sets—including price history, asset interrelationships, and even sentiment data—and continuously adjusting their decisions as the environment changes, thereby endowing asset allocation decisions with context sensitivity and continuous optimization characteristics.

However, the current state of research also cautions us to remain cautious. Some of the successes reported in the literature may reflect learning from historical data and may not hold true in new data or different market environments. Issues such as high turnover rates, sensitivity to hyperparameters, and the need for large amounts of training data mean that a DRL strategy may require careful calibration and risk control before being deployed in real-money applications. Some studies have found that once real-world frictions are considered, a complex DRL algorithm does not outperform a simple equal-weight strategy—a thought-provoking reminder that in finance, strategy complexity does not automatically guarantee success; the true test lies in its robust generalization performance and ability to navigate market cycles.

Ongoing developments in this field hold promise for addressing many of the aforementioned challenges. As researchers integrate more financial domain knowledge into DRL frameworks and design algorithms specifically tailored to financial scenarios, we can anticipate more reliable and interpretable outcomes. The intersection of finance and reinforcement learning is highly promising: in the future, DRL-driven asset allocation systems may be able to adapt in real-time to market changes, strictly adhere to predefined risk profiles, and provide human managers with insights previously difficult to obtain.

Overall, the application of deep reinforcement learning in asset allocation represents a prime example of interdisciplinary innovation at the intersection of finance and artificial intelligence—blending financial theory, economic principles, and cutting-edge AI technologies. Despite ongoing challenges, the existing theoretical foundations and empirical evidence suggest that, with further refinement, DRL has the potential to emerge as a powerful tool for portfolio management, driving more adaptive and effective investment strategies in the coming years.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Markowitz H M. Foundations of portfolio theory. The journal of finance, 1991, 46(2): 469-477.
[2] Merton R C. Optimum consumption and portfolio rules in a continuous-time model. In Stochastic optimization models in finance. Academic Press, 1975: 621-661.
[3] Sutton R S, Barto A G. Reinforcement learning: An introduction. Cambridge: MIT press, 1998, 1(1): 9-11.
[4] Benhamou E, Guez B, Ohana J J. Deep Reinforcement Learning: Extending Traditional Financial Portfolio Methods. Available at SSRN, 2024.
[5] Jiang Y, Olmo J, Atwi M. High-dimensional multi-period portfolio allocation using deep reinforcement learning. International Review of Economics & Finance, 20025, 98: 103996.
[6] Aboussalah A M, Xu Z, Lee C G. What is the value of the cross-sectional approach to deep reinforcement learning? Quantitative Finance, 2021, 22(6): 1091–1111.
[7] Chen S, Ge L. A learning-based strategy for portfolio selection. International Review of Economics & Finance, 2021, 71: 936-942.
[8] Wei J, Yang Y X, Jiang M, et al. Dynamic multi-period sparse portfolio selection model with asymmetric investors' sentiments. Expert Systems with Applications, 2021, 177: 114945.

[9] Kruthof G, Müller S. Can deep reinforcement learning beat 1N. Finance Research Letters, 2025: 106866.
[10] Cong L W, Tang K, Wang J, et al. AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. Available at SSRN, 2021: 3554486.