Dynamic Asset Allocation Strategies Based on Reinforcement Learning

Chengyue Shi

Faculty of Science, University of Auckland, Auckland CBD 1010, New Zealand
cshi929@aucklanduni.ac.nz

Abstract. Inherent non-stationary, non-linear, and sudden risk shock characteristics of financial markets often cause decision biases and model failures for traditional dynamic asset allocation models, which depend on steady-state returns. Reinforcement learning, because it can do adaptive sequential decision making, has become a key tool to deal with such problems. This paper addresses the main shortcomings of traditional models by constructing a reinforcement learning dynamic asset allocation model based on the PPO algorithm that incorporates multiple market features and transaction costs. Empirical testing was carried out with the daily data of domestic asset classes between 2010 and 2023. From the empirical results, the study observes that this strategy achieves an annualized return of 14.82% and a maximum drawdown of 11.68%, demonstrating strong profit and risk-control performance. Strategy Sharpe Ratio and Calmar Ratio are 1.23 and 1.27, respectively, much better than the two traditional benchmarks of equal weight and mean variance. It shows a good adaptability under different market situations, and multi-dimensional robustness tests also support the validity of the conclusion. This research improves the theoretical framework of dynamic asset allocation and gives asset management market players a smart allocation instrument that is both dependable and useful.

Keywords: Dynamic Asset Allocation Strategies, Reinforcement Learning, Risk Control.

1 Introduction

In such a complicated and changeable financial market environment, the effective allocation of capital becomes a fundamental issue for the asset management industry. Traditional dynamic asset allocation strategies, such as the Mean-Variance model and the Black-Litterman model, assume that the distribution of asset returns is stationary and that the historical covariance matrix can be estimated statically. But financial markets are always changing, not constant, and they follow no clear pattern, so there are big risks all of a sudden. The frequent occurrence of black swan events causes the traditional model to experience model failure, over-trading, and the accumulation of prediction errors, making it unable to adapt to the constantly changing market environment [1]. Recently, reinforcement learning, which mainly relies on

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_43

continuously improving decision-making through interactions with the market environment without needing to explicitly model the environment's dynamics, has emerged as an important tool for addressing the challenges associated with dynamic asset allocation. Based on this, this paper takes dynamic asset allocation as the research object and explores how reinforcement learning can break through the limitations of traditional methods to build an adaptive asset allocation strategy that has both return elasticity and risk controllability.

The current research has already proven the application of reinforcement learning in finance decisions. Relevant studies have used reinforcement learning to improve trading execution strategies and confirmed that the algorithm works well in high-dimensional financial decision-making situations [2]. With the advancement of deep learning technology, deep reinforcement learning has become mainstream in this area. Related research built a deep reinforcement learning trading model in a multi-asset setting, showing it outperforms conventional approaches [3]. Related recent studies keep going deeper. A risk-aware Dynamic Portfolio Model based on the PPO algorithm has been established, and it has been proven that the algorithm performs well under real-world constraints [4]. Deep reinforcement learning's application framework in portfolio management was comprehensively examined, and appropriate circumstances for various algorithms were verified [5]. The reinforcement learning asset allocation model was improved in terms of strategy rebalance and risk-return tradeoff [6]. Relevant studies have also made the research system of reinforcement learning in asset allocation more complete from the aspects of bond portfolio management and adapting to non-short-selling markets [7].

The existing literature usually verifies the feasibility of reinforcement learning algorithms or evaluates the performance of strategies in one market environment, but there are still some research gaps in the identification of multi-dimensional market states and the detailed modeling of liquidity and transaction costs. According to this, this paper builds a reinforcement learning dynamic asset allocation framework that combines market sentiment recognition and transaction cost constraints. Using daily data of domestic asset classes from 2010 to 2023 as the research sample, the PPO algorithm is used to design a strategy model that can make real-time position adjustments. Through adding risk penalty terms and market state adaptation methods, the model in this paper greatly decreases the portfolio's biggest drawdown and transaction resistance, but keeps return elasticity. Theoretically, this paper extends the scope of application for reinforcement learning within intricate financial systems and augments the theoretical study framework concerning dynamic asset allocation. In practice, it gives asset management companies and individual investors a smart, strong, and real-life asset allocation instrument that follows actual trading rules.

2 Applicability of Reinforcement Learning

The core of dynamic asset allocation is to dynamically adjust the allocation weights of various assets based on real-time updated market information, aiming to maximize portfolio returns while controlling risk. This decision-making process faces three core

challenges: Firstly, there's a high degree of non-linearity between financial asset prices and their influencing factors, so traditional linear models find it hard to grasp the complicated interactions among features. Secondly, market frictions such as transaction costs and liquidity constraints will greatly reduce the return of the strategy, and a static model cannot achieve dynamic optimization of rebalance action. Lastly, return and risk balance runs throughout the whole decision-making procedure. Models need to have the capacity for global optimization of multi-period sequential decisions, not just shortsighted choices aimed at maximizing one-period returns. The traditional asset allocation models fail to deal with all of these problems at once. But reinforcement learning can fit the dynamic asset allocation decision logic by continuously interacting with the environment to improve its own strategies.

Deep Q-Networks and Proximal Policy Optimization are the most commonly used reinforcement learning algorithms in finance. DQN algorithms, which are based on value function iterations, are appropriate for discrete action space situations. They use deep neural networks to approximate the Q-value function so as to make the best choice among various discrete actions. But they are limited when it comes to allocating weights continuously for assets and making fine-tuned positions hard. PPO algorithms belong to the policy gradient family, and they can directly optimize policy functions in continuous action spaces. By clipping the objective function to control the size of the policy update step, PPO solves the instability and low sample efficiency problems of traditional policy gradient algorithms, and it is suitable for continuous asset weight allocation tasks. It has become the go-to algorithm for dynamic asset allocation.

Reinforcement Learning plus Dynamic Asset Allocation has 3 advantages compared to conventional ways. Firstly, reinforcement learning does not require any assumptions regarding the form of the asset return distribution. It learns the decision rules based on the market data alone, removing the bias caused by wrong model specification and adapting to the non-normal and non-stationary return characteristics of financial markets. Second, reinforcement learning makes a temporal decision structure according to the Markov decision process, turning asset allocation into a multi-period sequence decision problem. And so it's able to consider the impact of current decisions on future income over time, and sidestep the shortsighted decision-making issues that older static models face. Lastly, with the flexible reward function design, reinforcement learning can directly incorporate realistic investment constraints such as risk aversion level, transaction cost, lack of liquidity, and prohibition of short selling. It changes the objective from maximizing return to maximizing risk-adjusted return, making it more aligned with real-world investments [8].

3 Design of Dynamic Asset Allocation Strategy

3.1 Core Elements

State space is the main carrier for the reinforcement learning agent to understand the market environment. The paper sets up a many-sided, all-inclusive state space system,

particularly with three main kinds of features: First, there are some general economic conditions such as the yearly change rate of GDP, the rate of increase in consumer prices and producer prices, the interest rate on ten-year government bonds, the purchasing managers' index, and the growth rate of the money supply, these all show the big picture of the economy and how much money is out there. Second, asset price features such as daily closing price, log return, 20-day rolling vol, trade volume, turnover rate, and momentum indicator for each underlying asset, which show the price-volume dynamic and trend of the assets themselves. Third, market sentiment features such as the VIX panic index, stock-bond yield ratio, growth rate of margin financing and securities lending balance, and news sentiment indicators representing changes in market sentiment and risk appetite [9]. To eliminate the difference in dimensions between different features and accelerate the convergence of the model, all features are normalized using the Z-score method.

Action space is the set of all possible actions that an agent can perform. This paper uses a continuous action space, which is suitable for the PPO algorithm. The model outputs a weight vector with a length corresponding to the number of assets. Every element in this vector is between 0 and 1, which represents the proportion of one type of asset. Meanwhile, a SoftMax activation function is applied to normalize the output vector so that the total allocation weights of all assets add up to 1, and each asset's weight is no less than zero. It satisfies the no-short-selling trading constraints present in mainstream markets such as the A-share market, enabling fine-tuning and continuous adjustment of asset weights.

The reward function is the main objective for reinforcement learning policy improvement. It decides the convergence direction and end result of the strategy directly. This paper designed a ternary reward function that takes into account the returns, risks, and transaction costs by referring to the most commonly used design framework in the current research. The formula is as model (1):

$$R_t = r_{p,t} - \lambda_{tc} \cdot TC_t - \lambda_{risk} \cdot Risk_t \tag{1}$$

Where $r_{p,t}$ is the immediate reward at time t, $r_{p,t}$ is the log return of the investment portfolio at time t. TC_t is the transaction cost at time t, it is calculated as the sum of the absolute changes in asset weight before and after rebalance multiplied by the bidirectional unit transaction cost, indicating the return erosion due to the rebalance action. $Risk_t$ is the risk indicator of the portfolio at time t, measured by downside volatility, it only includes the negative return fluctuation of the portfolio, which is closer to the actual risk perception of investors, is the transaction cost penalty coefficient, and λ_{risk} is the risk aversion coefficient. These two coefficients can be changed to fit different risk preferences and trading conditions.

3.2 Training Method

The main algorithm applied in this paper is the Proximal Policy Optimization algorithm. Its main benefit is that it restricts how much policies can change with a clipping process. This mechanism resolves the issues faced by traditional policy gradient algorithms, such as the challenge of setting the right step size and being

prone to large policy fluctuations or even collapse, and greatly reduces the algorithm's computational complexity, making it more stable and efficient in terms of samples. This objective function sets a limit on the largest step size of the policy update by taking the minimum value, guaranteeing the stability of the policy iteration.

The core parameter settings for the training process in this paper refer to common standards in existing research and are as follows: Learning rate is $3e^{-4}$, Adam optimizer is used for parameter update. Discount factor γ=0.99, which gives the model a higher weight to long-term reward as compared to short-term fluctuation. Generalized Advantage Estimation parameter λ is set to 0.95 to balance the bias and variance of the advantage estimate. Batch size is 64, training epoch per iteration is 10, total training iterations is 1000. Experience replay buffer size is set to 100,000, which stores samples of the agent's interactions with its environment, allowing for sample reuse and better training stability. Early stopping is also set up: Training will stop automatically if the Sharpe ratio on the validation set doesn't get better for 50 consecutive iterations to avoid overfitting.

4 Empirical Analysis

4.1 Sample Data and Benchmark Strategies

This paper selects daily trading data from January 1, 2010, to December 31, 2023, as the research sample. The underlying assets cover the CSI 300 Index, CSI 500 Index, China Bond Total Return Index, Shanghai Gold Spot AU9999, and Nanhua Commodity Index, comprehensively covering the three core asset classes: Equities, Bonds, and Commodities. All data are from the Wind database. After cleaning the missing values and outliers, there is a total of 3,405 valid trading days.

The entire sample is divided into a training set and a test set: 2010-2018 is the training set, which is used for learning the parameters of the model and policy iteration. 2019-2023 is the test set, which is used to test the generalization ability of the model. This period covers a variety of market conditions, including the epidemic shock and the Federal Reserve's rapid interest rate hikes, which enable a full test of the effectiveness of this method across different market scenarios. Two major traditional strategies in the asset management field are selected as the reference benchmarks. First, the equal-weight strategy, where all assets have the same weight and the portfolio is rebalanced every six months. Secondly, the mean-variance strategy that aims at minimizing the portfolio's variance by estimating the covariance matrix from the past 12 months' worth of data and rebalancing the portfolio once per month.

This paper constructs a strategy evaluation system from four dimensions: First, profitability metrics, using annualized return to evaluate the strategy's long-term profit. Second, risk control metrics choose the maximum drawdown and downside volatility as indicators for the strategy's extreme downside risk and negative return fluctuation characteristics. Third, risk-return metrics use the Sharpe ratio and Calmar ratio to measure the strategy's return for each unit of risk taken. Fourth, trading

efficiency metrics, with annualized turnover rate and transaction cost loss, to measure the degree of return erosion caused by rebalancing behavior.

4.2 Empirical Results

In the test period, the core performance indicators of the three strategies are compared in Table 1. Generally speaking, the PPO reinforcement learning strategy that was created in this paper performs much better than the two conventional benchmark strategies on various aspects such as returns, risks, and efficiency, which aligns with the results from other relevant studies.

Table 1. Comparison of core performance metrics for each strategy.

Performance metric	Reinforcement learning strategy	Equal weight strategy	Mean-variance strategy
Annualized return	14.82%	7.95%	9.76%
Maximum Drawdown	11.68%	26.34%	22.17%
Annualized Volatility	10.82%	10.57%	11.16%
Downside volatility	4.21%	8.32%	7.95%
Sharpe ratio	1.23	0.61	0.74
Calmar ratio	1.27	0.30	0.44
Annualized turnover rate	2.13 times	1.00 times	2.87 times

In terms of profitability, the annualized return of the reinforcement learning strategy is 14.82%, which is 86.4% higher than the equal-weight strategy and 51.8% higher than the mean-variance strategy, indicating a high profit level. Risk Control: The maximum drawdown of the reinforcement learning strategy is 11.68%, which is less than half of the two traditional strategies. Downside volatility also goes down from more than 8% in old ways to 4.21%. During major market events such as the COVID-19 shock and large-scale market declines, it can rapidly lower risk exposure via dynamic position changes, showing a risk resistance that outpaces conventional models.

Risk-return metric, the Sharpe ratio of the reinforcement learning strategy is 1.23, which is twice as much as the equal weight strategy and 1.66 times as much as the mean variance strategy. Calmar ratio is 1.27, which greatly exceeds the two traditional standards, proving that this strategy has a considerable edge in risk-adjusted returns. Trading efficiency-wise, due to the transaction cost penalty included in the reward function, the reinforcement learning strategy's annualized turnover rate is kept under 2.13 times, less than the monthly rebalance mean variance strategy. While doing dynamic rebalancing, it can also prevent the loss of returns caused by transaction friction. It is consistent with the results obtained by Su et al. for non-short-selling markets.

Furthermore, the study will segment the market states within the test set according to the trend characteristics of the CSI 300 Index, which includes Bull Market, Range-Bound Market, and Bear Market. The maximum drawdown of the reinforcement learning strategy is only 8.72%, compared with over 20% for the other two traditional strategies, showing great resilience. The reinforcement learning strategy has an annualized return of 10.25% in the range-bound market, which is much higher than the 4.18% of the equal-weight strategy and the 5.63% of the mean-variance strategy, indicating that it can take advantage of structural opportunities during small fluctuations. In a bull market, the strategy can also follow the market trend well, and its profitability is not inferior to that of traditional trend-following strategies. These results also demonstrate that the reinforcement learning strategy developed herein performs effectively and is applicable across diverse markets.

5 Robustness Tests

To confirm the validity of the study's findings regarding the reinforcement learning approach and prevent accidental empirical outcomes due to parameters or market conditions, robustness tests were carried out along two main dimensions: Parameter Sensitivity and Transaction Cost Sensitivity, which are standard practices in earlier works. Also, a robustness test was performed with different sample intervals. Firstly, the parameter sensitivity test of the risk aversion coefficient λ_{risk} in the reward function was carried out. This is one of the most important parameters affecting the risk-return characteristics of the strategy. The base value of this coefficient was set at 0.5, and λ_{risk} was adjusted to 0.1, 0.3, 0.7, and 1.0 to observe its impact. The test results indicate that as λ_{risk} increases, the annualized return of the strategy declines from 18.26% to 11.34%, and the maximum drawdown falls from 16.82% to 7.53%, exhibiting a regular pattern. Most importantly, the Sharpe Ratio was always above 1.1 under all conditions, way better than normal benchmarks, indicating that the coefficient just alters risk preference but not fundamental efficiency, so it shows good parameter stability.

Secondly, it simulated three different bilateral transaction cost situations to check how well the strategy would work under those circumstances. When the transaction costs went up to 1.0%, the annualized net return for the reinforcement learning strategy only decreased by 16.4 percentage points, whereas the mean-variance strategy saw a drop of 45.1%. This strategy reduces turnover on its own because it has a transaction cost penalty built into its reward function, so returns do not get eroded as much. Another test with 2015-2018 as the test period, replacing the test interval, shows that the reinforcement learning strategy still has an annualized return over 12%, a maximum drawdown below 15%, and a Sharpe ratio above 1.0, which further confirms the robustness and reliability of the research outcomes.

6 Conclusion

To tackle the main shortcoming of conventional dynamic asset allocation models, which depend too much on previous return distribution assumptions and have poor adaptability in nonlinear and nonstationary financial markets, this paper creates a dynamic asset allocation model using the PPO reinforcement learning algorithm. The model has a multi-state space, a continuous action space for allocating assets' weights, and a reward function that balances returns, risks, and transaction costs. It can dynamically allocate assets among different asset classes. Empirical results demonstrate that the proposed strategy achieves a higher annualized return than the traditional mean-variance and equal-weight strategies during the test period, with the maximum drawdown controlled at 12% or below. The Sharpe ratio and Calmar ratio are significantly improved, and the strategy presents strong adaptability across bull, sideways, and bear markets.

The paper's core contributions are twofold: It extends the application scope of reinforcement learning in asset allocation, improving the theoretical framework of dynamic asset allocation. It incorporates real-world investment restrictions into the model, giving investors a practical and sturdy smart instrument. This research has its own limitations, such as a lack of adaptability to small sample extreme black swan events, and a need to improve on integrating high-frequency and low-frequency data. Future work would include adding high-frequency and textual sentiment data along with meta-learning to improve the model's robustness and generalization.

References

1. Moody, J., Saffell, M.: Learning to Trade via Direct Reinforcement. IEEE Transactions on Neural Networks 12(4), 875–889 (2001).
2. Cheng, L.C., Sun, J.S.: Multiagent-Based Deep Reinforcement Learning Framework for Multi-Asset Adaptive Trading and Portfolio Management. Neurocomputing 594, 127800 (2024).
3. Jang, J., Seong, N.Y.: Deep Reinforcement Learning for Stock Portfolio Optimization by Connecting with Modern Portfolio Theory. Expert Systems with Applications 218, 119556 (2023).
4. Wang, J.F., Li, Y.M., Tang, W.Q., et al.: Portfolio Management Based on a Reinforcement Learning Framework. Journal of Forecasting 43(7), 2792–2808 (2024).
5. Yu, J., Chang, K.C.: Smart Tangency Portfolio: Deep Reinforcement Learning for Dynamic Rebalancing and Risk-Return Trade-Off. International Journal of Financial Studies 13(4), 227 (2025).
6. Nunes, M., Gerding, E., Mcgroarty, F., et al.: Reinforcement Learning for Bond Portfolio Management: An Actor-Critic Approach. The European Journal of Finance, 1–28 (2025).
7. Su, R., Chi, C., Tu, S., et al.: A Deep Reinforcement Learning Approach for Portfolio Management in Non-Short-Selling Market. IET Signal Processing (1), 5399392 (2024).
8. Yang, Y., Wang, T., Fu, Y., et al.: Portfolio Management Based on Value Distribution Reinforcement Learning Algorithm. Frontiers in Artificial Intelligence 8, 1709493 (2026).
9. Xue, P., Ye, Y.: Attention-Enhanced Reinforcement Learning for Dynamic Portfolio Optimization. Intelligent Systems with Applications 29, 200622 (2026).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

