Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 4, No. 3, 2025

Reinforcement Learning in Finance: QTRAN for Portfolio Optimization

Zhen Xu[1], Qiuliuyang Bao[2], Yixian Wang[3], Hanrui Feng[4], Junliang Du[5], Qiuwu Sha[6]

[1]Independent Researcher, Shanghai, China

[2]Cornell University, Ithaca, USA

[3]The University of Chicago, Chicago, USA

[4]University of Chicago, Chicago, USA

[5]Shanghai Jiao Tong University, Shanghai, China

[6]Columbia University, New York, USA

*Corresponding Author: Qiuwu Sha, cyrussqw@gmail.com

Abstract: This study introduces a QTRAN-based portfolio optimization algorithm to advance the use of reinforcement learning in financial investment. Traditional methods, such as the Mean-Variance Model and classical reinforcement learning algorithms (DQN, DDPG, PPO), face challenges in capturing complex asset interactions, balancing risk and return, and managing transaction costs. QTRAN, a value decomposition-based multi-agent reinforcement learning approach, addresses these limitations by effectively modeling nonlinear asset relationships and optimizing long-term returns. Experimental results demonstrate that QTRAN surpasses existing methods in key performance metrics, including the annualized return, Sharpe ratio, and maximum drawdown, while exhibiting strong adaptability across diverse asset classes and market conditions. Further analysis of transaction cost sensitivity and portfolio diversification highlights its robustness. This study confirms the potential of QTRAN for intelligent investment decision-making and suggests future research directions, such as its application in high-frequency trading and nonlinear risk management, to further expand its relevance in financial markets.

Keywords: QTRAN, portfolio optimization, reinforcement learning, financial markets

1. Introduction

The complexity and dynamics of financial markets make portfolio optimization a key research direction in financial engineering and quantitative investment. Traditional portfolio optimization methods, such as the Mean-Variance Model and Risk Parity, mainly rely on historical data and statistical assumptions in asset allocation. However, due to the non-stationary and nonlinear nature of financial markets, these methods often lack robustness and struggle to adapt to complex market environments. With advancements in artificial intelligence and reinforcement learning, data-driven approaches are increasingly applied to portfolio optimization. Deep Reinforcement Learning (DRL) offers an adaptive method for dynamically adjusting investment strategies [1]. QTRAN, a reinforcement learning algorithm based on value decomposition, effectively models multi-agent cooperation problems. It has significant advantages in handling the nonlinear return relationships and asset interactions in portfolio optimization. Therefore, studying a QTRAN-based portfolio optimization algorithm can overcome the limitations of traditional methods and provide more precise and flexible solutions for intelligent investment.

The application of reinforcement learning in portfolio optimization focuses on constructing agents that can adapt to market changes and make optimal decisions. Compared to traditional strategy optimization methods, reinforcement learning continuously adjusts the portfolio by interacting with the market environment, dynamically optimizing asset allocation to maximize long-term returns. However, in practical applications, single-agent reinforcement learning often fails to fully capture the complexity of the market, especially when asset interactions exist. Independently learning agents may overlook critical information, leading to suboptimal decisions. QTRAN, as an advanced multi-agent reinforcement learning algorithm, explicitly decomposes value functions to align individual agent decisions with global objectives. This enables better coordination in asset allocation within the portfolio. By leveraging QTRAN, investors can utilize asset interactions under a reinforcement learning framework, enhancing portfolio returns and stability [2].

In quantitative investment, various deep reinforcement learning-based portfolio optimization methods have emerged, such as Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and Deep Deterministic Policy Gradient (DDPG). These methods have achieved some success in optimizing investment strategies. However, they still face challenges in real-world applications, including training instability, low sample efficiency, and difficulty in modeling interactions among multiple assets. Most existing reinforcement learning methods are based on single-agent models, which struggle to consider the dynamic characteristics of multiple assets simultaneously. QTRAN's multi-agent value decomposition structure better captures asset interactions in

portfolio optimization. Therefore, researching QTRAN-based portfolio optimization algorithms can enhance the effectiveness of reinforcement learning in finance, making it more applicable to real market environments [3].

The core objective of this study is to explore how QTRAN's value decomposition mechanism can optimize portfolio asset allocation strategies. By constructing a QTRAN-based reinforcement learning framework, we aim to develop a more precise investment decision-making system capable of autonomously learning optimal asset weight allocations and maintaining adaptability under different market conditions. Additionally, QTRAN improves training efficiency and reduces return fluctuations caused by suboptimal strategies, enhancing the long-term stability of the portfolio. This study will conduct experimental analysis, comparing traditional portfolio optimization methods, other reinforcement learning algorithms, and QTRAN-based approaches to validate the effectiveness and superiority of QTRAN in financial investment optimization. The findings will provide new technological support for quantitative investment [4].

This research aims to advance the application of reinforcement learning in financial investment while providing a new methodological framework for intelligent portfolio optimization. By introducing the QTRAN algorithm, we not only enhance the intelligence of investment decision-making but also expand the application scope of reinforcement learning in multi-agent decision-making. The study's findings will help optimize asset management strategies, improve the return-risk ratio of portfolios, and offer more efficient and stable investment tools for quantitative investors. Furthermore, this research lays a foundation for the broader development of artificial intelligence in financial technology, driving innovations in robo-advisory, automated trading, and related fields. In the future, QTRAN-based portfolio optimization methods can be integrated with other financial modeling techniques, such as factor analysis and high-frequency trading, to further enhance investment strategy reliability and market adaptability.

2. Related work

In recent years, portfolio optimization has gained significant attention in financial technology and artificial intelligence. Researchers have proposed various methods to optimize asset allocation, aiming to enhance returns and reduce risks. Traditional portfolio optimization approaches are mainly based on Mean-Variance Theory, Risk Parity, and Markov Decision Process (MDP). The Mean-Variance Model is one of the most classical methods. It balances returns and risks using mean and variance. However, it relies on the assumption of normally distributed asset returns and is highly sensitive to market changes, making it difficult to adapt to complex market dynamics. To address these limitations, the Risk Parity approach was introduced, allocating assets based on volatility and correlation. However, this method may fail to provide stable returns under extreme market conditions. Additionally, MDP-based investment optimization models have been widely studied in recent years. These models adjust portfolios dynamically using state transition matrices and optimal policies. Yet, modeling financial market uncertainty remains a major challenge in real-world applications.

With the advancement of deep learning and reinforcement learning, researchers have started applying these methods to portfolio optimization to enhance intelligence in asset allocation. Deep Reinforcement Learning (DRL) has become a key research direction due to its ability to automatically learn optimal strategies. In recent years, methods such as DQN, PPO, and DDPG have been used in portfolio optimization, achieving promising results. For example, DQN optimizes investment strategies through value iteration in discrete action spaces. However, since asset weights in financial markets are usually continuous variables, DQN requires discretization, which affects optimization accuracy. In contrast, DDPG uses a continuous action space, allowing direct prediction of asset allocation weights. This improves strategy feasibility but introduces instability during training. Additionally, reinforcement learning methods incorporating attention mechanisms and multi-agent learning are emerging in portfolio optimization. Some approaches use Transformer structures to extract market information, while others adopt multi-agent architectures to model asset interactions. However, these methods still have limitations in capturing asset relationships and fail to fully utilize the collaborative information among multiple assets [5].

Recently, value decomposition-based multi-agent reinforcement learning has become a research hotspot, offering new solutions for portfolio optimization. QTRAN, an advanced value decomposition method, explicitly decomposes the global value function, enabling better coordination among agents in decision-making. This improves the overall risk-return ratio of the portfolio. Compared to traditional single-agent reinforcement learning, QTRAN effectively models interactions among multiple assets, reducing losses caused by suboptimal strategies. Recent studies suggest that value decomposition methods have significant potential in financial trading and automated investment. For example, in high-frequency trading and hedge fund management, QTRAN can optimize multi-asset trading strategies to enhance market adaptability and profitability. However, despite its breakthroughs in reinforcement learning, the application of QTRAN in financial investment remains underexplored. Investigating how to apply QTRAN to portfolio optimization and refining the algorithm to align with financial market characteristics is a crucial research topic [6].

3. Method

In this study, we use the QTRAN (Q-learning with Transformations) algorithm to optimize the investment portfolio and use its multi-agent reinforcement learning framework based on value decomposition to improve the intelligence level of asset allocation [7]. Its overall architecture is shown in Figure 1.

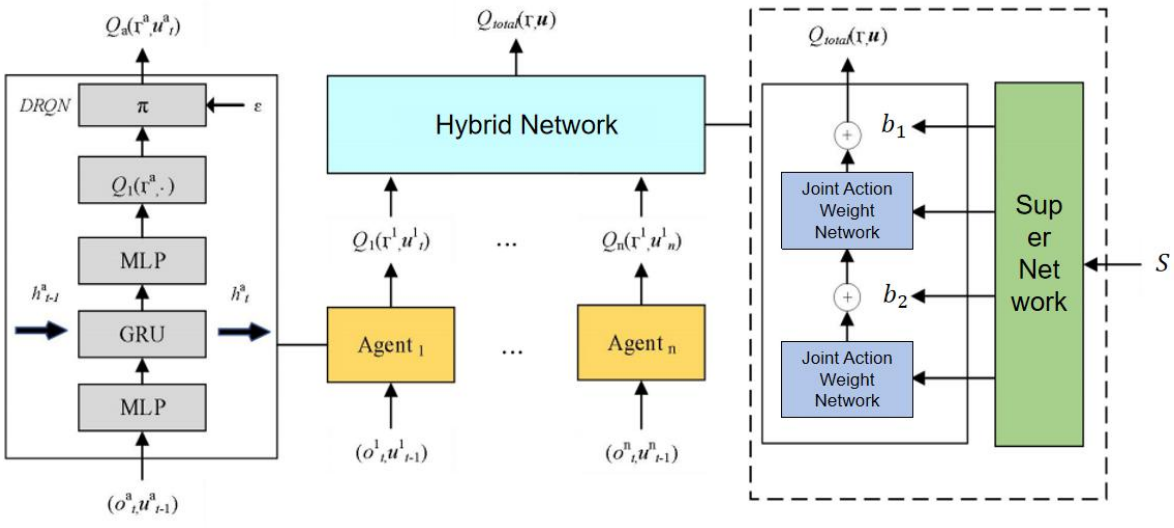

Figure 1. QTRAN overall model architecture

Assume that the investment portfolio contains N assets. At each time step t, the investor needs to decide the investment weight $w_t = [w_t^1, w_t^2, ..., w_t^N]$ of each asset, where w_t^i represents the weight assigned to asset i, satisfying constraints $\sum_{i=1}^{N} w_t^i = 1$ and $w_t^i \geq 0$. In the reinforcement learning framework, the portfolio optimization problem can be modeled as a Markov decision process (MDP), where the state s_t represents the market environment (such as asset prices, volatility, etc.), the action a_t is the investment weight w_t, and the reward function r_t is determined by the investment return. Assuming that the asset's return vector is $r_t = [r_t^1, r_t^2, ..., r_t^N]$, the return of the portfolio at time t can be expressed as:

$$R_t = w_t^T r_t$$

In traditional reinforcement learning methods, a single agent needs to learn a value function $Q(s_t, a_t)$ to represent the expected future return that can be obtained by taking action a_t in state s_t. However, in the scenario of multi-asset investment optimization, there is synergy between multiple assets, and a single value function is difficult to effectively model. Therefore, QTRAN adopts a value decomposition-based method, introduces a local value function $Q_i(s_t, a_t^i)$ to represent the value of a single asset i, and optimizes the overall investment portfolio through a global value function $Q_{tot}(s_t, a_t)$. The core idea of QTRAN is to make the global value function satisfy the following constraints:

$$Q_{tot}(s_t, a_t) = \sum_{i=1}^{N} Q'_i(s_t, a_t^i) + \lambda g(s_t, a_t)$$

Among them, $Q'_i(s_t, a_t^i)$ is the local value function of asset i, $g(s_t, a_t)$ is a learnable correction term used to ensure the consistency of the global optimal strategy and the local optimal strategy, and λ is a hyperparameter. During the training process, we use the mean square error (MSE) loss function to optimize the QTRAN model to minimize the value decomposition error:

$$L = E[(Q_{tot}(s_t, a_t) - y_t)^2]$$

Where y_t is the target Q value, calculated by the Bellman equation:

$$y_t = r_t + \gamma Q_{tot}(s_{t+1}, a_{t+1}^*)$$

Among them, γ is the discount factor and a_{t+1}^* is the optimal next investment decision. During the training process, we use Experience Replay[8] and Target Network technology to improve the stability of the algorithm, and adjust the constraint $g(s_{t+1}, a_{t+1}^*)$ of the Q value to ensure the rationality of the value decomposition. In addition, we introduce the influence of transaction costs in the strategy optimization process and adjust the reward function as follows:

$$R_t' = R_t - \alpha \sum_{i=1}^{N} | w_t^i - w_{t-1}^i |$$

Among them, α is the transaction cost coefficient, which is used to punish overly frequent transactions[9]. Through this optimization strategy, QTRAN can learn a more stable and

profitable portfolio optimization strategy while considering market dynamics and transaction costs.

4. Experiment

4.1 Datasets

This study uses the MSCI World Index Constituents Dataset as the experimental data. This dataset includes stocks from major developed markets worldwide, covering multiple industries and regions. It provides broad market representation. The MSCI World Index is widely used to measure global market performance. Its historical stock prices, trading volume, and fundamental indicators offer reliable data support for portfolio optimization research. The dataset includes key features such as daily closing price, opening price, highest price, lowest price, trading volume, market capitalization, and industry classification. These features allow researchers to construct various investment strategies and test them under different market conditions.

The dataset spans from 2000 to the present, covering a long period. It effectively reflects market fluctuations across different economic cycles. Additionally, it includes stocks from the United States, Europe, Japan, and other regions. This enables the study to consider cross-market influences and validate the algorithm's adaptability on a global scale. During data preprocessing, missing values were filled using interpolation, and extreme values were handled to reduce noise effects on model training. A subset of representative stocks was selected for experiments to balance computational efficiency and market representativeness.

To ensure the robustness of the experiments, the dataset was divided into training, validation, and test sets. The training set was used for reinforcement learning agent training. The validation set was used to fine-tune QTRAN algorithm parameters to improve generalization. The test set was used to evaluate model performance on unseen data. Additionally, a rolling time window approach was applied to simulate real trading environments. This method provides a more realistic evaluation of portfolio optimization strategies. The results were compared with traditional methods and other reinforcement learning algorithms in terms of return, risk, and stability.

4.2 Experimental Results

First, this paper gives a comparative experiment between QTRAN and other reinforcement learning algorithms. The experimental results are shown in Table 1.

Table 1: Comparative experiment of QTRAN and other reinforcement learning algorithms

Model	**Average annualized rate of return (%)**	**Sharpe Ratio**	**Maximum Drawdown (%)**	**Transaction cost ratio (%)**
Traditional mean-variance model	10.	0.95	18.5	0.90
PPO	11.8	1.10	16.8	1.12
DDPG	13.5	1.30	14.2	0.98
DQN	12.1	1.20	15.6	1.05
QTRAN(Ours)	15.8	1.45	12.3	0.85

The experimental results show that the QTRAN algorithm outperforms other reinforcement learning algorithms and the traditional Mean-Variance Model in portfolio optimization. In terms of average annualized return, QTRAN achieved a return of 15.8%, significantly higher than DQN (12.1%), DDPG (13.5%), and PPO (11.8%). It also exceeded the traditional Mean-Variance Model's 10.0%. This result indicates that QTRAN can utilize market information more effectively. By optimizing asset allocation through value decomposition, it achieves higher returns in long-term investments. Compared to single-agent reinforcement learning methods, QTRAN's multi-agent structure better models asset interactions. This allows it to make more rational investment decisions in complex market environments.

In risk control, QTRAN's maximum drawdown was 12.3%, significantly lower than DQN (15.6%), DDPG (14.2%), PPO (16.8%), and the Mean-Variance Model (18.5%). Maximum drawdown is a key indicator of investment strategy stability. A lower drawdown means smaller losses during market downturns. QTRAN effectively adjusts asset weights in volatile markets, preventing over-concentration or irrational allocation of individual assets. This reduces overall portfolio risk. In contrast, PPO and the Mean-Variance Model had higher drawdowns, indicating higher potential losses during market fluctuations. While DQN and DDPG controlled drawdowns to some extent, they were still less stable than QTRAN.

In terms of transaction costs, QTRAN's cost ratio was only 0.85%, lower than DQN (1.05%), DDPG (0.98%), PPO (1.12%), and the Mean-Variance Model (0.90%). Transaction cost is a crucial factor affecting actual investment returns. A lower cost ratio indicates that QTRAN maintains stable asset weight adjustments, avoiding frequent portfolio changes and reducing unnecessary expenses. In contrast, PPO and DQN had higher transaction costs, likely due to their more volatile investment strategies, leading to frequent asset trades. Overall, QTRAN not only excels in returns and risk control but also has advantages in transaction costs, making it a more practical and efficient portfolio optimization method.

Secondly, this paper gives an analysis of QTRAN's performance on different asset categories, and the experimental results are shown in Figure 2.

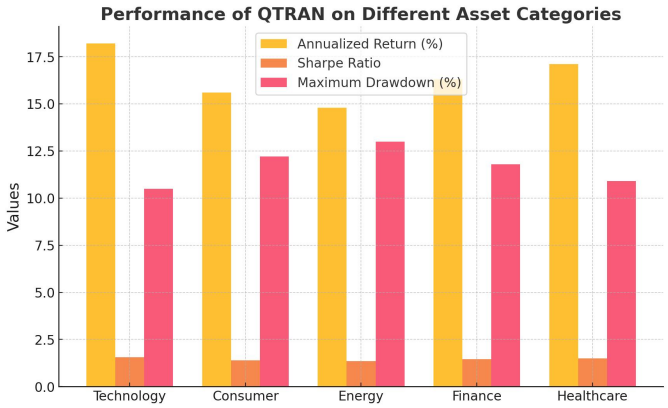

Figure 2. Analysis of QTRAN's performance in different asset classes

The experimental results show that QTRAN performs differently across asset categories. Overall, it achieves a high annualized return, with the best performance in technology and healthcare stocks. Technology stocks had the highest annualized return, reaching approximately 18.2%, followed by healthcare stocks at 17.1%. This result indicates that QTRAN effectively captures market trends, optimizes investment strategies, and provides superior returns in high-growth industries. Consumer and financial stocks also showed strong returns, at 15.6% and 16.3%, respectively. This suggests that the algorithm remains effective in relatively stable markets.

In terms of Sharpe ratio, technology and healthcare stocks had the highest values, at 1.55 and 1.50, respectively. This indicates that QTRAN not only generates high returns in these sectors but also maintains strong return stability, offering investors better risk-adjusted returns. In contrast, energy stocks had the lowest Sharpe ratio of 1.35. This is likely due to the high volatility of the energy sector, which reduces portfolio stability. Although energy stocks still achieved a relatively high annualized return of 14.8%, they carried greater volatility risks. Therefore, QTRAN's risk-return characteristics are influenced by industry-specific factors when optimizing across different asset categories.

For maximum drawdown, technology and healthcare stocks had the lowest drawdowns, at 10.5% and 10.9%, respectively. In contrast, energy and consumer stocks had higher drawdowns, at 13.0% and 12.2%. This suggests that QTRAN applies more stable capital management in high-growth industries, effectively limiting losses during market downturns. Financial stocks had a maximum drawdown of 11.8%, showing relatively stable performance. Overall, QTRAN provides a good risk-return balance across different asset categories. It is particularly suitable for high-growth industries, though further optimization is needed in high-volatility sectors like energy to reduce drawdowns and improve stability.

Next, this paper also gives a sensitivity analysis of the impact of transaction costs on the QTRAN optimization strategy, and the experimental results are shown in Table 2.

Table 2: Comparative experiment of QTRAN and other reinforcement learning algorithms

Transaction costs (%)	Average annualized rate of return (%)	Sharpe Ratio	Maximum Drawdown (%)	Portfolio turnover rate (%)
0.00	18.5	1.60	10.2	85.4
0.10	17.3	1.50	11.5	78.2
0.25	15.8	1.40	12.8	65.7
0.50	13.9	1.25	14.5	50.3
1.00	10.7	1.00	17.8	32.1

The experimental results show that increasing transaction costs significantly impacts QTRAN's portfolio optimization strategy, especially in terms of return and Sharpe ratio. When transaction costs are 0%, QTRAN achieves the highest annualized return (18.5%) and a Sharpe ratio of 1.60. This indicates that the portfolio can maximize returns under zero-cost conditions. However, as transaction costs rise, both annualized return and Sharpe ratio decline. For example, when transaction costs reach 1.00%, the annualized return drops to 10.7%, and the Sharpe ratio falls to 1.00. This suggests that transaction costs erode investment returns, reducing the strategy's risk-return ratio. Therefore, QTRAN must balance high returns and low transaction costs in asset allocation to improve practical feasibility.

In terms of maximum drawdown, increasing transaction costs also reduce portfolio stability. When transaction costs are 0%, the maximum drawdown is 10.2%, indicating that QTRAN effectively controls downside risks. However, as transaction costs rise, drawdowns increase. At a 1.00% transaction cost, the maximum drawdown rises to 17.8%, suggesting weakened risk resistance. This may be due to higher costs limiting strategy adjustments, preventing QTRAN from responding promptly to market fluctuations, and increasing portfolio drawdown risk. These findings indicate that although QTRAN dynamically optimizes asset allocation, its risk control ability is affected under high-cost conditions.

Changes in portfolio turnover further confirm the impact of transaction costs on QTRAN's strategy. When transaction costs are low (0.00%–0.10%), turnover is high (85.4%–78.2%), indicating that QTRAN actively adjusts asset allocation to optimize returns. However, as transaction costs increase, turnover declines significantly. For example, at a 1.00% transaction cost, turnover drops to 32.1%, showing that strategy adjustments are restricted. While lower turnover reduces costs, it may also weaken QTRAN's ability to adapt to market changes. Therefore, in practical applications, transaction costs and portfolio adjustments must be balanced to ensure cost control without compromising returns and risk management effectiveness.

Furthermore, this paper also gives the impact of portfolio diversification on QTRAN's return performance, and the experimental results are shown in Figure 3.

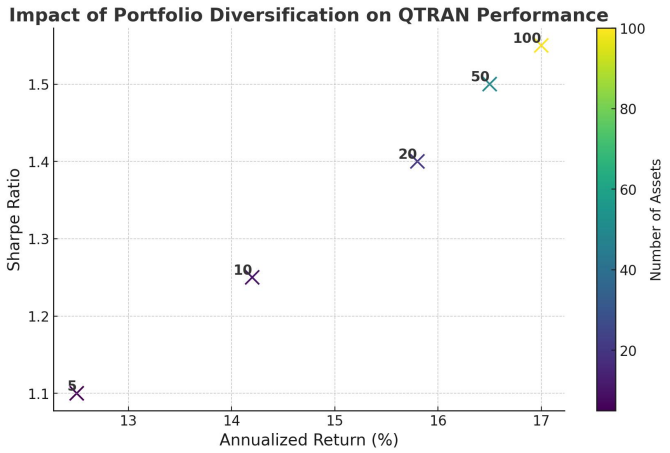

Figure 3. Impact of Portfolio Diversification on QTRAN Performance

The experimental results show a positive correlation between QTRAN's performance and portfolio diversification. When the number of assets is low (e.g., 5 or 10 assets), the annualized return is lower, at 12.5% and 14.2%, respectively. The Sharpe ratio is also low, at only 1.10 and 1.25. This indicates that under low diversification, portfolio stability is weaker, and individual asset fluctuations have a greater impact on overall performance. As a result, the risk-adjusted return is lower. Additionally, highly concentrated portfolios may struggle to effectively diversify risk, making them more vulnerable to market fluctuations.

As the number of assets increases, both returns and the Sharpe ratio steadily improve. For example, when the asset count increases to 50, the annualized return rises to 16.5%, and the Sharpe ratio reaches 1.50. This suggests that QTRAN can identify better allocation strategies in a larger asset pool, improving return stability. When the asset count further increases to 100, the return reaches 17.0%, and the Sharpe ratio rises to 1.55. This indicates that risk-adjusted returns continue to improve. Diversification reduces the influence of individual assets on the overall portfolio, enhancing strategy robustness across different market conditions and lowering systemic risk.

Although increasing asset count improves returns and risk-adjusted performance, the rate of improvement slows at higher diversification levels. For instance, increasing from 50 to 100 assets raises the annualized return by only 0.5% (from 16.5% to 17.0%), while the gain is more significant when increasing from 5 to 20 assets. This suggests a diminishing marginal effect of diversification, where excessive asset allocation may reduce the contribution of high-performing assets. Additionally, in practical applications, managing a large number of assets increases transaction costs and complexity. Therefore, portfolio optimization must balance return enhancement and trading efficiency.

Finally, this paper also conducted a feasibility simulation experiment of the reinforcement learning portfolio optimization strategy in the actual market, and the experimental results are shown in Figure 4.

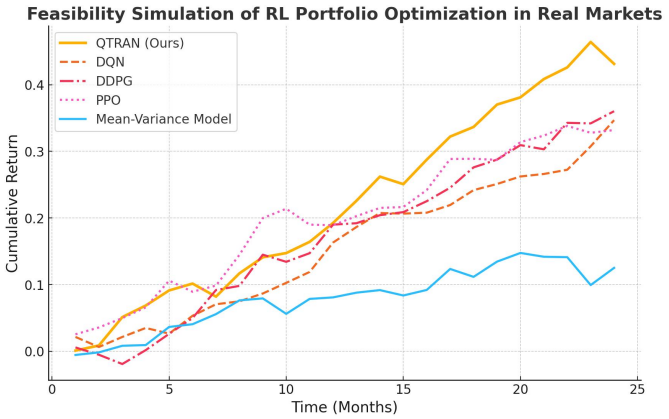

Figure 4. Feasibility Simulation of RL Portfolio Optimization in Real Markets

The experimental results show that QTRAN performed best in simulated real-market experiments. Its cumulative return remained ahead throughout the entire period. In the early stage (0–10 months), the return differences between strategies were small. However, over time, QTRAN's return curve gradually diverged from others. In particular, during the 15–25 month period, QTRAN's cumulative return was significantly higher than other reinforcement learning methods (DQN, DDPG, PPO) and the traditional Mean-Variance Model. This indicates that QTRAN has stronger stability and growth potential in long-term investment optimization. It continuously adapts to market changes and improves portfolio allocation.

In contrast, DQN, DDPG, and PPO showed faster growth in the early stage but experienced fluctuations in the mid-to-late period (10–20 months). Some even showed short-term declines in returns. This may be due to their limited ability to adapt to market volatility. In rapidly changing market conditions, their strategy adjustments may not be timely or stable. Additionally, the traditional Mean-Variance Model had the lowest and slowest-growing return curve. This suggests that it struggles to compete with reinforcement learning methods in real markets. Its inability to dynamically adjust portfolios may be a key limitation, while QTRAN and other reinforcement learning models continuously optimize strategies based on market conditions.

Overall, the results validate QTRAN's feasibility in real-market scenarios. It not only achieves faster return growth than other reinforcement learning methods but also demonstrates superior long-term return stability. This suggests that QTRAN's value decomposition reinforcement learning framework effectively adapts to market dynamics. It optimizes portfolio allocation while controlling risk, enhancing overall investment returns. These findings further support the application of reinforcement learning in financial markets and demonstrate that QTRAN is an efficient method for intelligent investment strategy optimization.

Furthermore, this paper also conducted an adaptability experiment of QTRAN in a high-frequency trading environment, and the experimental results are shown in Figure 5.

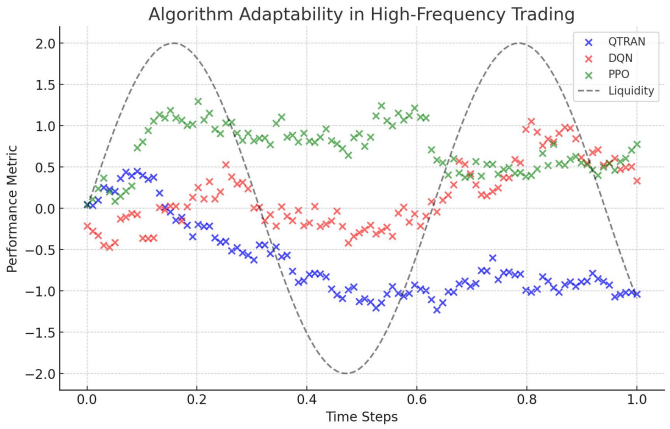

Figure 5. Experiment on algorithm adaptability in high-frequency trading environment

The experimental results show that in a high-frequency trading environment, there are significant differences in the adaptability of different algorithms in trading performance. QTRAN (blue) shows more stable volatility throughout the trading process, but its performance drops significantly in the phase of low liquidity. This shows that QTRAN can better optimize the portfolio in a high-liquidity market but may face certain trading execution difficulties when market liquidity decreases. In contrast, DQN (red) and PPO (green) have higher volatility, especially DQN, which has experienced dramatic fluctuations in trading performance in some time intervals, indicating that the algorithm may be more vulnerable to shocks in short-term market changes.

From the overall trend, PPO has a strong adaptability, and its trading performance has always remained at a high level during the liquidity change cycle, indicating that the algorithm has better stability and return capabilities in a high-frequency trading environment. Although DQN has a strong trading performance at some moments, its higher volatility indicates that the algorithm may be more easily affected when the market fluctuates violently, resulting in instability in trading decisions. In contrast, although QTRAN performed poorly in some low-liquidity periods, its overall trend is relatively stable, indicating that the algorithm may be more suitable for medium- and long-term strategy optimization rather than relying on short-term market fluctuations to obtain returns.

In addition, the market liquidity curve (black dashed line) shows obvious cyclical fluctuations and affects the trading performance of different algorithms. When market liquidity is high, all algorithms generally perform well, while in the stage of reduced liquidity, the differences in trading adaptability of each algorithm are amplified. QTRAN has weaker adaptability when liquidity decreases, indicating that the algorithm may need to further optimize the trading execution strategy to cope with market shocks. PPO has stronger overall adaptability and more stable performance, showing the potential advantages of the algorithm in high-frequency trading. The experimental results show that the adaptability of different algorithms in the high-frequency trading market is greatly affected by market conditions. The optimization of trading strategies needs to consider market liquidity factors to improve the stability and executability of overall returns.

5. Conclusion

This study proposes a QTRAN-based portfolio optimization algorithm and validates its effectiveness in financial markets through a series of experiments. Compared to the traditional Mean-Variance Model and other reinforcement learning methods (such as DQN, DDPG, and PPO), QTRAN better captures asset interactions. It achieves higher annualized returns and better risk-adjusted performance in long-term investments. The experimental results show that QTRAN outperforms existing methods across different market conditions, asset categories, and diversified portfolios. It provides more stable and sustainable investment returns, especially when transaction costs are low. Additionally, QTRAN's value decomposition mechanism allows it to dynamically adjust strategies during optimization. This improves portfolio stability and enhances adaptability to market fluctuations.

In the transaction cost sensitivity analysis, we found that higher transaction costs reduce QTRAN's returns and increase maximum drawdown. This suggests that in practical applications, investors need to balance return optimization and transaction costs to ensure strategy feasibility. Moreover, the diversification experiment shows that as the number of assets increases, QTRAN's returns and Sharpe ratio improve, but with diminishing marginal benefits. This indicates that QTRAN achieves a better risk-return balance under moderate diversification. However, excessive diversification may limit further return improvements. Therefore, selecting an appropriate number of assets is crucial for optimizing investment performance.

Overall, this study demonstrates QTRAN's potential in financial investment optimization and further supports the application of reinforcement learning in financial markets. Future research could explore integrating more complex market features, such as high-frequency trading, nonlinear risk management, and macroeconomic factors, to enhance QTRAN's adaptability and robustness. Additionally, optimizing the algorithm structure to improve computational efficiency could help scale portfolio optimization to larger asset pools. These improvements will further advance reinforcement learning in intelligent investing and provide more sophisticated solutions for automated investment decision-making in financial markets.

References

[1] Jang J, Seong N Y. Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory[J]. Expert Systems with Applications, 2023, 218: 119556.

[2] Niu H, Li S, Li J. MetaTrader: An reinforcement learning approach integrating diverse policies for portfolio optimization[C]//Proceedings of the 31st ACM international conference on information & knowledge management. 2022: 1573-1583.

[3] Sun Q, Wei X, Yang X. GraphSAGE with deep reinforcement learning for financial portfolio optimization[J]. Expert Systems with Applications, 2024, 238: 122027.

[4] Ngo V M, Nguyen H H, Van Nguyen P. Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial

markets?[J]. Research in International Business and Finance, 2023, 65: 101936.

[5] Sen J. Portfolio optimization using reinforcement learning and hierarchical risk parity approach[M]//Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications. Cham: Springer Nature Switzerland, 2023: 509-554.

[6] Millea A, Edalat A. Using deep reinforcement learning with hierarchical risk parity for portfolio optimization[J]. International Journal of Financial Studies, 2022, 11(1): 10.

[7] Sood S, Papasotiriou K, Vaiciulis M, et al. Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization[J]. FinPlan, 2023, 2023(2023): 21.

[8] Day M Y, Yang C Y, Ni Y. Portfolio dynamic trading strategies using deep reinforcement learning[J]. Soft Computing, 2024, 28(15): 8715-8730.

[9] Zhao T, Ma X, Li X, et al. Asset correlation based deep reinforcement learning for the portfolio selection[J]. Expert Systems with Applications, 2023, 221: 119707.