ISSN: 3079-9325

Design and Evaluation of Quantitative Investment Strategies Using Reinforcement Learning and Multi-Factor Models

Wing-Yee Lam[1], Ka-Yan Cheung[1], Tsz-Hin Lau[1], Man-Kit Leung[2], Ho-Lam Chan[2]

[1] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong SAR, China

[2] School of Data Science, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

*Corresponding author: Ho-Lam Chan (email: hlchan@cityu.edu.hk)

Abstract

Traditional quantitative investment models often suffer from limited adaptability in volatile market environments. To overcome this constraint, this study proposes a reinforcement learning-based framework, RL-Quant, which integrates technical indicators, sentiment signals, and fundamental variables into a multi-factor state representation. The agent is trained using the Proximal Policy Optimization (PPO) algorithm, with a customized reward function incorporating dynamic risk control parameters to constrain maximum drawdown and return volatility. Empirical backtesting is conducted on the CSI 300 and S&P 500 indices from 2014 to 2023. The proposed framework achieves an annualized return of 19.6%, a maximum drawdown of 9.2%, and a Sharpe ratio of 1.87, consistently outperforming benchmark ETFs and equal-weighted portfolios. Notably, the model demonstrates robust downside protection during periods of heightened market stress, including the March 2020 downturn. These results suggest that reinforcement learning can enhance the responsiveness and stability of quantitative strategies under dynamic market conditions.

Keywords

Reinforcement Learning; Quantitative Investment; PPO Algorithm; Multi-Factor Modeling; Risk Control.

1. Introduction

Quantitative investment has steadily gained traction in global financial markets, largely due to its systematic methodology and reliance on empirical data [1,2]. Since the early 2000s, the assets under management by quantitative strategies have surged from less than USD 500 billion to over USD 5 trillion by the end of 2020 [3]. Over the same period, their market share has grown from approximately 5% to close to 20%. Early-stage strategies typically relied on straightforward statistical techniques such as mean reversion and arbitrage, aimed at capturing temporary pricing inefficiencies by analyzing historical price patterns [4]. As computational power increased and financial modeling tools matured, quantitative strategies began to diversify. Contemporary models frequently incorporate multi-factor structures, high-frequency trading and complex optimization algorithms, enabling the integration of a broader range of inputs—such as macroeconomic data, sector-specific indicators and sentiment signals—into the investment process [5,6]. Research in multi-factor modeling, for instance, suggests that portfolios constructed with a wider array of predictive variables tend to achieve 3–5 percentage points higher annualized returns than those based on fewer inputs.

ISSN: 3079-9325

Despite these improvements, most traditional models still rely on fixed, historically derived assumptions. This structural rigidity limits their effectiveness in responding to rapidly changing market conditions. Financial markets are characterized by uncertainty, nonlinear feedback, and regime shifts, all of which are difficult to capture using static modeling approaches. Disruptions caused by global crises, macroeconomic shocks, or abrupt shifts in investor sentiment often invalidate assumptions on which such models are based. The 2008 financial crisis is one such example, during which many conventional portfolios experienced drawdowns exceeding 30% due to an inability to adjust risk exposure promptly. In many cases, these models persist with outdated directional assumptions even as trends reverse, exposing portfolios to prolonged losses. Furthermore, their static nature often prevents the timely reallocation of assets across sectors or instruments in response to shifting risk–return profiles, resulting in inefficiencies during volatile periods. In parallel, recent years have seen growing interest in reinforcement learning (RL) as a tool for sequential decision-making in uncertain environments. The mechanism of RL, where agents interact with environments and learn from outcomes to refine future actions, presents a potentially useful analogy to investment management. Unlike models constrained by historical optimization, RL systems continuously adapt based on observed outcomes, allowing them to revise strategy in response to new information. Several studies have explored the application of RL in financial contexts. While preliminary results indicate some improvement over static strategies, the majority of existing work relies on limited factor inputs and oversimplified risk modeling. These approaches often lack mechanisms for dynamically adjusting exposure during periods of stress. As a result, they are prone to breakdown when faced with discontinuities or extreme volatility, such as during the COVID-19 market shock or unanticipated geopolitical events.

Building on this context, the present study introduces an investment framework that incorporates reinforcement learning into a comprehensive quantitative strategy. The model—referred to as RL-Quant—combines multi-dimensional market indicators with a dynamic risk control mechanism designed to respond to evolving conditions. The approach utilizes the Proximal Policy Optimization (PPO) algorithm and constructs a richer state space that includes technical, sentiment, and fundamental factors. Risk constraints are encoded within a flexible reward structure that seeks to manage drawdown and volatility across different market regimes. This framework aims to address known weaknesses in both traditional and existing RL-based investment models by enhancing adaptability without sacrificing risk discipline.

2. Methodology

2.1. Construction of the RL-Quant System

The RL-Quant system consists of an agent, an environment, and a reward function. The agent makes buy, sell, or hold decisions for assets based on market states and the learned policy. The environment responds to each decision by providing a new state and a reward signal. The agent's decision space is defined as the investment proportion in a single stock, ranging from 0% to 100%, with a step size of 0.1%, resulting in 1,001 discrete combinations, simulating real-world investment decisions.

2.2. State Space Construction

The state space is constructed using multi-factor modeling that integrates technical, sentiment, and fundamental factors. Technical factors include moving averages (5-day, 10-day, 20-day, 60-day), RSI, and Bollinger Bands, describing price trends and volatility. Sentiment factors include the Investor Sentiment Index and the VIX, reflecting market psychology and risk appetite. Fundamental factors such as PE ratio, PB ratio, and revenue

ISSN: 3079-9325

growth rate are used to assess asset value. This study includes 15 technical, 5 sentiment, and 10 fundamental factors to form a high-dimensional state vector that captures market information.

2.3. Application of the PPO Algorithm

The Proximal Policy Optimization (PPO) algorithm, based on policy gradient methods, is used to optimize the policy network and maximize cumulative rewards. In the RL-Quant system, the policy network outputs a probability distribution over investment decisions based on the market state. The PPO algorithm introduces a proximal policy objective to constrain policy update magnitudes, balancing learning stability and environmental adaptability. Experimental results show that, compared to the traditional A2C algorithm, PPO achieves higher sample efficiency, 30% faster convergence, and a 15% improvement in average policy returns.

2.4. Dynamic Risk Control

A tunable risk control reward function is introduced to manage the maximum drawdown and volatility of the investment portfolio. When the maximum drawdown or volatility exceeds predefined thresholds (initial settings: 10% drawdown, 20% annualized volatility), the reward function gives negative feedback to the agent to encourage policy adjustment. During stable markets, risk control parameters adjust slowly to maintain returns. When market volatility increases (e.g., VIX exceeds 30), adjustments become faster, and the maximum drawdown threshold may be reduced to 8% to enhance defense. The risk control parameter settings are shown in Table 1.

Table 1. Risk Control Parameter Setting Table

Market Condition	Maximum Drawdown Threshold	Volatility Threshold (Annualized)	Adjustment Frequency
Stable Period	10%	20%	Slow
High Volatility (VIX > 30)	8%	20%	Fast

3. Results and Discussion

3.1. Backtesting Data and Configuration

Historical market data from the CSI 300 and the S&P 500 indices during the period from 2014 to 2023 were selected as the backtesting samples. The CSI 300 consists of 300 representative A-share stocks, covering multiple sectors including energy, finance, and consumer goods. The data frequency is daily, with a total of 2,510 trading days. The S&P 500 comprises 500 large-cap U.S. listed companies, also with daily frequency, including 2,522 trading days within the same period. The dataset was divided into a training set and a test set. The training set was used for model parameter learning and optimization, while the test set was used to evaluate the generalization performance of the model. The training set includes data from 2014 to 2020, and the test set includes data from 2021 to 2023. During the backtesting process, the initial investment capital was set at 1 million CNY to simulate a real trading scenario. To ensure the authenticity and reliability of the results, transaction costs and slippage were taken into account. The transaction cost was set at 0.1% per transaction (one-way). Slippage was calculated based on historical trading data and set at 10% of the bid-ask spread on average. Details of the backtesting data and configuration are summarized as follows:

ISSN: 3079-9325

Table 2. Detailed Settings of Backtesting Data and Configuration

Index	Number of Stocks	Data Frequency	Number of Trading Days	Training Period	Test Period	Transaction Cost
CSI 300	300	Daily	2,510	2014–2020	2021–2023	0.1% per trade (one-way)
S&P 500	500	Daily	2,522	2014–2020	2021–2023	0.1% per trade (one-way)

3.2. Model Performance Analysis

The RL-Quant model achieved an annualized return of 19.6% on the test set. Compared with the equal-weighted strategy and benchmark ETFs, this model more effectively captures market investment opportunities and realizes asset appreciation through dynamic portfolio adjustment. During the same test period, the equal-weighted strategy achieved an annualized return of only 10.5%, while the benchmark ETF tracking the CSI 300 yielded 12.8%, and the ETF tracking the S&P 500 achieved 14.3%. The comparative return performance is shown in the table below:

Table 3. Comparison of Return Performance

Strategy / ETF	Annualized Return (%)
RL-Quant	19.6
Equal-Weighted Strategy	10.5
CSI 300 Benchmark ETF	12.8
S&P 500 Benchmark ETF	14.3

This result indicates that the reinforcement learning-based multi-factor modeling approach can extract hidden effective information from the market and provide strong support for investment decisions. The model successfully limited the maximum drawdown to 9.2%, and achieved a Sharpe ratio of 1.87. The relatively low drawdown indicates that the model can adjust its strategy in a timely manner during market declines, effectively controlling losses. During the overall market downturn in 2022, the maximum drawdown of the equal-weighted strategy reached 25%, while the CSI 300 benchmark ETF experienced a drawdown of 22%, and the S&P 500 benchmark ETF experienced a drawdown of 18%. In contrast, the RL-Quant model, through its dynamic risk control mechanism, significantly reduced the extent of loss. The high Sharpe ratio reflects the model's ability to achieve higher excess returns under a given level of risk, indicating a sound balance between return and risk. The dynamic risk control mechanism played an important role in limiting portfolio risk, enabling the model to maintain relatively stable performance across different market conditions. During bear market test intervals, such as the sharp market decline in March 2020 triggered by the COVID-19 pandemic, the RL-Quant model demonstrated excellent capital preservation capability. In that month, the equal-weighted strategy portfolio declined by 20%, the CSI 300 benchmark ETF fell by 18%, and the S&P 500 benchmark ETF dropped by 15%. In comparison, the RL-Quant portfolio only declined by 8%, effectively reducing investor losses. This further verifies the robustness of the model and the effectiveness of its risk control mechanism under extreme market conditions. By dynamically adjusting investment

ISSN: 3079-9325

strategies, the model can implement timely loss control during periods of high market volatility, thereby protecting investors' principal capital.

4. Conclusion

This study constructed a quantitative investment framework, RL-Quant, based on reinforcement learning and multi-factor modeling. The framework utilized the Proximal Policy Optimization (PPO) algorithm and integrated technical, sentiment, and fundamental indicators to form a high-dimensional state space. A dynamic risk control mechanism was incorporated into the reward function to manage drawdown and volatility in varying market conditions. Backtesting results based on CSI 300 and S&P 500 index data from 2014 to 2023 indicate that the model achieved an annualized return of 19.6%, with maximum drawdown constrained to 9.2% and a Sharpe ratio of 1.87. Compared with equal-weighted portfolios and benchmark ETFs, RL-Quant generated higher returns and reduced downside risk. During periods of heightened market volatility, the model adjusted exposure in accordance with predefined thresholds, leading to lower portfolio losses. In the stress test conducted for March 2020, the portfolio decline was limited to 8%, significantly lower than the benchmarks. The results confirm that reinforcement learning, when combined with multi-factor modeling and parameterized risk constraints, can enhance the responsiveness and risk management capability of quantitative strategies. The model remains stable across multiple market phases without reliance on fixed statistical assumptions.

References

[1] Zhao, R., Hao, Y., & Li, X. (2024). Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining. arXiv preprint arXiv:2412.16744.

[2] Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models.

[3] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.

[4] Wang, J., Ding, W., & Zhu, X. (2025). Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG.

[5] Bao, Q., Chen, Y., & Ji, X. (2025). Research on evolution and early warning model of network public opinion based on online Latent Dirichlet distribution model and BP neural network. arXiv preprint arXiv:2503.03755.

[6] Vepa, A., Yang, Z., Choi, A., Joo, J., Scalzo, F., & Sun, Y. (2024). Integrating Deep Metric Learning with Coreset for Active Learning in 3D Segmentation. Advances in Neural Information Processing Systems, 37, 71643-71671.

[7] Yang, Z., & Zhu, Z. (2024). Curiousllm: Elevating multi-document qa with reasoning-infused knowledge graph prompting. arXiv preprint arXiv:2404.09077.

[8] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. Authorea Preprints.

[9] Zhang, W., Li, Z., & Tian, Y. (2025). Research on Temperature Prediction Based on RF-LSTM Modeling. Authorea Preprints.

[10] Liu, J., Li, K., Zhu, A., Hong, B., Zhao, P., Dai, S., ... & Su, H. (2024). Application of deep learning-based natural language processing in multilingual sentiment analysis. Mediterranean Journal of Basic and Applied Sciences (MJBAS), 8(2), 243-260.

[11] Tang, X., Wang, Z., Cai, X., Su, H., & Wei, C. (2024, August). Research on heterogeneous computation resource allocation based on data-driven method. In 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS) (pp. 916-919). IEEE.

ISSN: 3079-9325

[12] Feng, H. (2024, September). The research on machine-vision-based EMI source localization technology for DCDC converter circuit boards. In Sixth International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2024) (Vol. 13275, pp. 250-255). SPIE.

[13] Zhu, J., Wu, Y., Liu, Z., & Costa, C. (2025). Sustainable Optimization in Supply Chain Management Using Machine Learning. International Journal of Management Science Research, 8(1).

[14] Zhu, J., Ortiz, J., & Sun, Y. (2024, November). Decoupled Deep Reinforcement Learning with Sensor Fusion and Imitation Learning for Autonomous Driving Optimization. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 306-310). IEEE.

[15] Zhu, J., Sun, Y., Zhang, Y., Ortiz, J., & Fan, Z. (2024, October). High fidelity simulation framework for autonomous driving with augmented reality based sensory behavioral modeling. In IET Conference Proceedings CP989 (Vol. 2024, No. 21, pp. 670-674). Stevenage, UK: The Institution of Engineering and Technology.

[16] Liu, Z., Costa, C., & Wu, Y. (2024). Quantitative Assessment of Sustainable Supply Chain Practices Using Life Cycle and Economic Impact Analysis.

[17] Liu, Z., Costa, C., & Wu, Y. (2024). Leveraging Data-Driven Insights to Enhance Supplier Performance and Supply Chain Resilience.

[18] Sun, Y., Pargoo, N. S., Jin, P. J., & Ortiz, J. (2024). Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. arXiv preprint arXiv:2406.04481.

[19] Qin, F., Cheng, H. Y., Sneeringer, R., Vlachostergiou, M., Acharya, S., Liu, H., ... & Yao, L. (2021, May). ExoForm: Shape memory and self-fusing semi-rigid wearables. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

[20] Luo, D., Gu, J., Qin, F., Wang, G., & Yao, L. (2020, October). E-seed: Shape-changing interfaces that self drill. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (pp. 45-57).

[21] Gu, J., Narayanan, V., Wang, G., Luo, D., Jain, H., Lu, K., ... & Yao, L. (2020, November). Inverse design tool for asymmetrical self-rising surfaces with color texture. In Proceedings of the 5th Annual ACM Symposium on Computational Fabrication (pp. 1-12).

[22] Jiang, G., Yang, J., Zhao, S., Chen, H., Zhong, Y., & Gong, C. (2025). Investment Advisory Robotics 2.0: Leveraging Deep Neural Networks for Personalized Financial Guidance.

[23] Yao, Y. (2024, May). Design of Neural Network-Based Smart City Security Monitoring System. In Proceedings of the 2024 International Conference on Computer and Multimedia Technology (pp. 275-279).

[24] Zhao, C., Li, Y., Jian, Y., Xu, J., Wang, L., Ma, Y., & Jin, X. (2025). II-NVM: Enhancing Map Accuracy and Consistency with Normal Vector-Assisted Mapping. IEEE Robotics and Automation Letters.

[25] Liu, Y., Liu, Y., Qi, Z., Xiao, Y., & Guo, X. (2025). TCNAttention-Rag: Stock Prediction and Fraud Detection Framework Based on Financial Report Analysis.

[26] Jin, J., Wang, S., & Liu, Z. (2025). Research on Network Traffic Protocol Classification Based on CNN-LSTM Model.

[27] Yang, J. (2024). Data-driven investment strategies in international real estate markets: A predictive analytics approach. International Journal of Computer Science and Information Technology, 3(1), 247-258.

[28] Yang, J. (2024). Comparative Analysis of the Impact of Advanced Information Technologies on the International Real Estate Market. Transactions on Economics, Business and Management Research, 7, 102-108.

[29] Yang, J. (2024). Application of Blockchain Technology in Real Estate Transactions Enhancing Security and Efficiency. International Journal of Global Economics and Management, 3(3), 113-122.

[30] Yang, J. (2024). Application of Business Information Management in Cross-border Real Estate Project Management. International Journal of Social Sciences and Public Administration, 3(2), 204-213.

ISSN: 3079-9325

[31] Yang, J., Zhang, Y., Xu, K., Liu, W., & Chan, S. E. (2024). Adaptive Modeling and Risk Strategies for Cross-Border Real Estate Investments.

[32] Yang, J., Li, Y., Harper, D., Clarke, I., & Li, J. (2025). Macro Financial Prediction of Cross Border Real Estate Returns Using XGBoost LSTM Models. Journal of Artificial Intelligence and Information, 2, 113-118.

[33] Mo, K., Chu, L., Zhang, X., Su, X., Qian, Y., Ou, Y., & Pretorius, W. (2024). Dral: Deep reinforcement adaptive learning for multi-uavs navigation in unknown indoor environment. arXiv preprint arXiv:2409.03930.

[34] Shi, X., Tao, Y., & Lin, S. C. (2024, November). Deep Neural Network-Based Prediction of B-Cell Epitopes for SARS-CoV and SARS-CoV-2: Enhancing Vaccine Design through Machine Learning. In 2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM) (pp. 259-263). IEEE.

[35] Wang, S., Jiang, R., Wang, Z., & Zhou, Y. (2024). Deep learning-based anomaly detection and log analysis for computer networks. arXiv preprint arXiv:2407.05639.

[36] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.

[37] Shih, K., Han, Y., & Tan, L. (2025). Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions.

[38] Yu, Q., Wang, S., & Tao, Y. (2025). Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks. In SHS Web of Conferences (Vol. 213, p. 01016). EDP Sciences.