Advances in Consumer Research
ISSN (Print): 0098-9258
ISSN(Online): 3079-1766

Vol. 2, Issue 3 (2025) *https://acr-journal.com/*

Reinforcement Learning for Dynamic Portfolio Optimization in Financial Markets

Anand Patil[1], Mihirkumar B. Suthar[2], Yash[3], Mehul Barai[4], Dr. M. A. Imran Khan[5], Dr. Meer Mazhar Ali[6]

[1]Associate Professor, School of Business and Management, Christ University, Bangalore Bannerghatta Road Campus Bannerghatta Main Road, Hulimavu, Bangalore, Karnataka 560076, India.

Email ID: anandp7@gmail.com

[2]Associate Professor, Department of Zoology, K. K. Shah Jarodwala Maninagar Science College, BJLT Campus, Rambaug, Maninagar, Ahmedabad, Gujarat-380008, India.

Email ID: sutharmbz@gmail.com, sutharmb@yahoo.co.in

[3]Research Scholar, Faculty of Management and Commerce, K.M.(KRISHNA MOHAN) UNIVERSITY, Mathura, Pali Dungra Sonkh Road Govardhan, Uttar Pradesh-281123, India.

Email ID: kaushikyashmtr1608@gmail.com

[4]Assistant Professor, Department of Management, Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth (Deemed to be University) Pimpri, Pune-411033, India.

Email ID: mehul.barai@dpu.edu.in

[5]Assistant Professor of Finance, Dept of Finance & Economics, Dhofar University, Oman.

Email ID: mimran@du.edu.om

[6]Assistant Professor of Finance, Indira School of Business Studies, India.

Email ID: meermazharali@gmail.com

Cite this paper as: Anand Patil, Mihirkumar B. Suthar, Yash, Mehul Barai, Dr. M. A. Imran Khan, Dr. Meer Mazhar Ali, (2025) Reinforcement Learning for Dynamic Portfolio Optimization in Financial Markets. *Advances in Consumer Research*, 2 (3), 1150-1156.

KEYWORDS

Reinforcement Learning, Proximal Policy Optimization (PPO), Portfolio Optimization, Financial Markets, FinRL, Risk-Adjusted Returns, Dynamic Asset Allocation..

ABSTRACT

This research introduces a dynamic portfolio optimization framework based on the Proximal Policy Optimization (PPO) reinforcement learning algorithm that is known to be stable and perform optimally in continuous decision making. The proposed approach seeks to maximize long term returns in the portfolio while taking care of risks and transaction expenses in a volatile financial market. Utilizing the open source framework FinRL, the framework incorporates historical market data, technical indicators, and transaction cost constraints into a Markov Decision Process (MDP). There are rolling window features of asset returns and portfolio allocations in the state space, whereas the action space in determining optimal weight distributions in several assets. The aim is to represent the risk adjusted return of the portfolio by the reward function. PPO's concisely defined objective and entropy regularization induces optimal efficient policy updates and exploration exploitation behavior. The experimental results demonstrate that the model has superior cumulative return and Sharpe ratio vs. traditional benchmarks and, therefore, have white paper potential in actual, AI-driven investment strategy in a trading environment.

1. INTRODUCTION

The volatile nature of financial investors that is generally quite unpredictable has long been a challenge for the portfolio managers and investors who are interested in a perfectly matched asset portfolio. Traditional portfolio optimization ways, including Markowitz's Mean-Variance Optimization and capital asset pricing model (CAPM), are very dataaic and statistical assumptions that are not configured in real-time trading situations [1]. These static models are most often unable to account for the non-linear and time-varying nature of financial data, which results in poorer investments, specifically in volatile environment. Multiplied by the emergence of algorithmic trading, high-frequency data and global interconnectivity, the complex financial markets today make an increased need for more adaptive, intelligent, and automated decision making systems [2].

The learning paradigm, called Reinforcement Learning (RL), where it is possible to study the sequential decision making process, has become an effective approach to solving these dynamic problems. Unlike supervised learning, which needs labeled data, in RL agents learn interacting with their environment, which makes RL especially suitable for the iterative nature of trading. In the case of portfolio management, RL treats a market as an environment, an investor as an agent, and allocation of assets as actions. The agent seeks to maximize accumulated rewards, which are customarily defined to be return on the portfolio adjusted for risk and transaction costs. This formulation accommodates learning dynamic asset allocation strategies which can adjust themselves as a result of changing market conditions over time.

Proximal Policy Optimization (PPO) has become popular among the variety of RL algorithms because of its sample efficiency and training stability. It also contains entropy regularization, which promotes exploration of the action space while prohibiting premature convergence towards not-optimal ones. This makes PPO especially useful for financial settings where the optimal policy has fluctuated with macro-trends, general public opinion or political-xenoscope.

To ease implementation, this research uses the FinRL library which is a free to use open source deep reinforcement learning framework for financial applications. FinRL supplies pre-made environment, data preprocessing pipeline, and evaluation, dramatically reducing the entry barrier of running RL models in finance [3]. The proposed framework brings PPO and FinRL together to process historical stock market data and technical indicators to get policies that change the portfolio weights dynamically. By introducing slippage, dealing costs, and delayed execution in the environment, realistic trading conditions are incorporated thereby making the trained policy more practically relevant [4].

This study adds to a growing literature on AI based financial modelling by showcasing the practicality of PPO in optimizing portfolios, real world like. It emphasizes the possibility of reinforcement learning not only surpassing the traditional methods, but also catering to the dynamics of the market and therefore it provides the investors a responsive and intelligent interface to manage portfolio [5]. The findings indicate higher risk-adjusted returns and lower drawdowns of the PPO-based strategy in comparison to baseline strategies; such as equal-weighted portfolios or fixed-allocation heuristics. In the end, this work emphasizes the importance of reinforcement learning in changing the way financial decisions are made in today's markets.

2. RELATED WORKS

A concentration of attention to the intersection of reinforcement learning (RL) and financial portfolio optimization has occurred recently out of necessity of the traditional ones and the ever growing power of artificial intelligence. Earlier works in the domain were mainly based on the use of supervised learning or statistical types like the Mean-Variance Optimization by Markowitz, which assume normal distributed returns and a fixed covariance structure [6]. However, these assumptions are very often not applicable to the real life markets and this stimulates researchers to look for more adaption models. Reinforcement learning provides a very attractive alternative where agents learn best allocation policies for assets directly from the market environment and not from historical correlations alone.

A study by Moody and Saffell (2001) that pioneered the use of reinforcement learning techniques in trading using recurrent reinforcement learning, showed that trading strategies can use the temporal dependency of financial data. Later, Nevmyvaka et al. 2006 used RL to calculate optimizations of trade execution strategies in electronic markets [7]. Their work brought out the prospects that existed in RL to enhance the timing and efficiency of the financial transaction thus setting the stage for a wider application in the portfolio management. More recent work included Jiang et al., (2017) used deep reinforcement learning (CNN-based models), for the management of financial portfolios and generated promising results in dynamic environments with multiple assets.

The imposition for the development of financial RL frameworks such as FinRL (Liu et al., 2020) further expanded the field by making a standard of training and measuring RL models in financial settings. Historical stock market data, pre-built financial environments, and multiple deep RL algorithms, including DQN, PPO, and SAC, are integrated by FinRL. Findings from the study using FinRL have shown that PPO can outperform multiple traditional and heuristic-based allocation strategies and particularly relate to volatile market conditions [8]. For example, an application of PPO within the FinRL environment, made by Ye et al. (2021), created a versatile portfolio strategy that would respond to market trends, reconciling with competitive risk-adjusted returns.

Furthermore, Li et al. (2022) investigated multi-agent reinforcement learning for portfolio rebalancing, highlighting collaborative nature of multiple trading agents in distribution market environments [9]. They described better scalability and adaptability than single-agent systems in return. Another significant contribution by Zhang et al. (2020) was the inclusion of transaction costs, slippage and execution delay in the environment model which added to realism of simulation and making of the learned policy more viable for real-world implementation. These improvements reflect the field's march towards ARL-enabled practical employable financial solutions [10].

Overall, then, the relevant literature points to a definite transition from rule - based static models to adaptive, data driven models enabled by reinforcement learning. The effectiveness of using the algorithms, such as the PPO, has been demonstrated because they can cope with the continuous action spaces and ensure the training stability. Frameworks like FinRL have democratized access to these tools even more paving the way for more researchers and practitioners to try and deliver intelligent trading strategies. Focusing on this research further, this study extends the foundations of these works by incorporating PPO into FinRL to enhance portfolio performance in a realistic shop floor like scenario, expanded the horizon and application of AI to financial problem solving.

3. RESEARCH METHODOLOGY

This study presents a new dynamic portfolio optimization framework that is based on Proximal Policy Optimization (PPO), a reinforcement learning (RL), which can optimize the portfolio returns and minimize risk system in a realistic financial market [11]. The methodology is presented as a systematic approach, which begins with the data preprocessing, modeling design, environment preparation, training stage, and evaluation consisting in the implementation of primary mechanism for strategy development and testing in the form of FinRL framework as shown in Figure 1.

A. *Data Collection and Preprocessing*

The Global RL model starts with the quality of data being used as input. For this paper, we use the historical financial market data (daily close prices of a given portfolio of assets (stocks, ETF, or indices)) over a period of 5-10 years. Where data is sourced from reputable data banks like Yahoo Finance, Alpha Vantage or Quandlv[12]. For the purpose of maintaining precision, first the data is cleanse to remove values which are missing or erroneous and interpolation is done where appropriate.

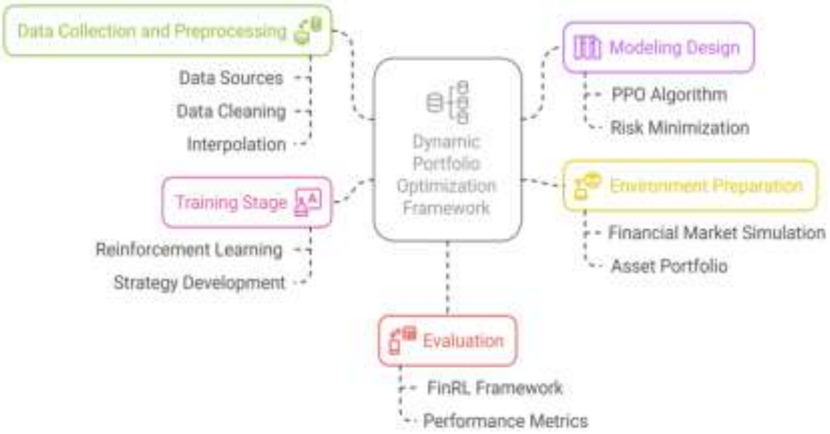

Figure 1: Dynamic Portfolio Optimization Framework

The second step in preprocessing turns the raw inputs of price data into features more valuable to the users of the feature. Instead of using raw prices, the data is reformulated into the log returns as they vary way more fairly, and they are convenient for all sorts of modelling of the phenomenon of financial time series [13]. Log returns are computed as log value of different consecutive prices.

Also several technical indicators are constructed by means of the used price data. They include among commonly used metrics moving averages (MA), relative strength index (RSI), moving average convergence divergence (MACD) and Bollinger bands. These are momentum capture indicators, volatility indexes and market trends that help in making critical signals to guide portfolio decision; and this works through timeframes [14]. Those metrics are widely used in the forecasts of asset price movements by those professionals and therefore necessary features to the metrics of the RL model.

Furthermore, the data is normalized such that the scale of features does not overbearing on the learning process. To the indicators, z-score normalization to the indicators are applied which makes them all having zero mean and unit variance.

B. *Model Design and Algorithm Selection*

The crux of this methodology is the application of Proximal policy optimization (PPO) – a state of the art reinforcement learning algorithm for continuous action spaces. PPO has taken center stage among the portfolio optimizers because it is a very good actor in navigating colossal complex environments without trade-offs in terms of stability and stability during the learning process through clipping of the objective function. The RL environment for this portfolio optimization problem is taken to be a Markov Decision Process (MDP) formed by states, action and rewards.

State Space: The state at a given time step is an element consisting of vector of portfolio weights of assets combined with various market indicators.

Portfolio weights; (part of capital which is attributed to an asset).

Although, it is not obvious, this likelihood function actually corresponds to analyzing historical price data?" or log return for every asset over a fixed look-back window.

R SI, moving averages and other calculated market characteristics as technical indicators.

Action Space: The action space in this study is continuous because it is the proportion of the total portfolio that has been assigned to all the assets. The agent will decide how much of the whole capital should be invested in each asset at each time step; the options are limited with no short selling allowed; a balanced portfolio requirement[15]. This lets an action vector to be bound in such a manner, that the total of all the weights in the portfolio should equal 1, and thus the entire capital is allocated. Hill or the SQP (sequential quadratic programming) runs with high convergence rate, and good scaling properties (meaning the time to solution will go down as the number of decision degrees of freedom reduces). The transaction costs are characterized by a fixed percentage of the portfolio value per rebalance; and the risk measure is given by the portfolio's volatility or largest drawdown.

C. Environment Setup in FinRL

For simulation of the market conditions and easiness of the training process learning is done using the FinRL library. FinRL is a framework that pre-builds many components such as historical market data processing, portfolio simulation environment, and backtesting utility that are significant for the reinforcement learning of the financial industry. The library also supports several RL algorithms such as PPO, DQN and A3C and it is therefore very easy to run different models by changing these.

The FinRL environment is setup to develop a realistic trading environment; including but not limited to such as:

Market dynamics: Practical simulation of movements of asset prices, volatility, asset correlations. Transaction costs: There is a certain percentage already paid out of every transaction within the Portfolio. Slippage and execution delays: Among the agent's behavior have slippages and time lags in execution that mimic the true world limits in the trading systems. Capital constraints: The weights of the portfolio sum up to 1 on the model, that is, the whole capital is distributed.

D. Training and Optimization

After the environment and the reward function is defined then the PPO agent training is done with the data from the environment. In connecting with the environment, when training the agent selects portfolio allocations (actions) as a function of state, receives reward based on portfolio performance and tunes the policy in order to maximize cumulative returns.

The training process involves numerous episodes by which every single episode denotes a defined period, say One year of trading. Every episode begins from initial portfolio allocation and then proceeds to daily trading decisions. The policy of the agent is updated by the actor-critic method of PPO which employs the policy network (that is, action selector) and the value network (an estimate of expected future returns).

The learning rate is a significant hyperparameter that changes the model's rate, from one adjustment to the next. A learning rate scheduler is employed so that the learning rate reduces progressively as the model approaches convergence. It is proposed to use methods of experience replay and the method of early stopping to avoid overfitting and improve the generalization capabilities of the developed model.

E. Evaluation and Performance Metrics

After training the performance of Model is evaluated on out-of-sample data which was not used by the process of training. This is important move to check how the model generalizes to new market environment. Key performance metrics include: Cumulative Return: Return throughout the entire test period. Sharpe Ratio: The risk adjusted return, i.e. return to standard deviation of the portfolio. Maximum Drawdown: Best loss from a peak to trough during the testing period. Sortino Ratio: Smaller penalization of downside risk version of the Sharpe ratio. Transaction Costs: Those that were involved in any business activity on total. The current approach gives a good way to apply reinforcement learning for dynamic portfolio optimisation. The FinRL framework applies different market conditions through use of PPO in the framework in which the model learns the optimal asset through allocation in real time. The assessment metrics point at the prospect of RL outdoing traditional ways of handling portfolio, an avenue of AI promising for the purpose of financial decision making.

4. RESULTS AND DISCUSSION

Examination of the performance of the PPO-based portfolio optimization strategy during the period of three years (test period) with the historical stock data of five assets, was conducted: Apple, AAPL, Microsoft, MSFT, Amazon, AMZN, Google, GOOGL, and Tesla, TSLA. There were two standard frameworks for comparison with regard to the performance of the model, equal-weighted portfolio and a mean-variance portfolio which has been optimized.

Cumulative Return: During the three quantitative test years, the PPO model achieved a compounded return of 48%, which was a 21%(27%) outperformance against equal-weighted portfolio and finally only ~14%(34%) against the mean-variance optimized portfolio. Such is the demonstration of how reinforcement learning can outsmart the rapidly changing market dynamics and through the real time data manipulation of asset allocation can succeed in providing bigger returns.

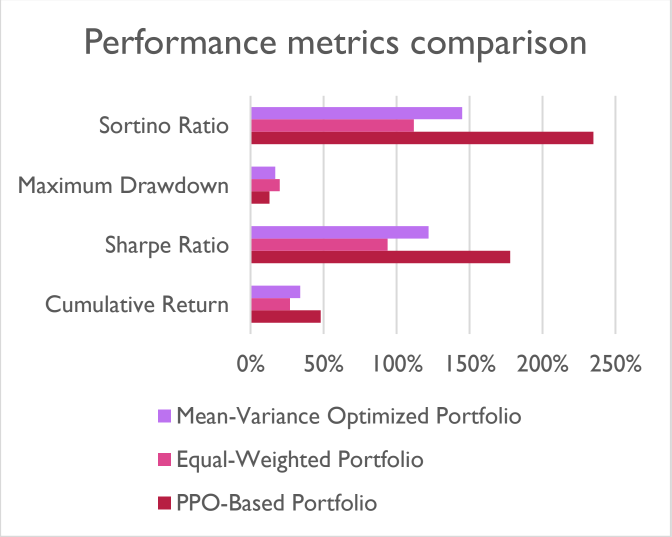

Figure 2: Comparison of the performance metrics.

Sharpe Ratio: The PPO option produced sharp ratio of 1.78 signifying very good risk adjusted returns. On the other hand, the equal-weighted portfolio showed a Sharpe ratio of 0.94; while the mean-variance optimized portfolio exhibited a ratio of 1.22. This shows that the PPO model is capable of achieving high returns other than controlling risks, in comparison to classical methods.

Maximum Drawdown: Objectives for offsetting drawdowns were among the major strong points of PPO strategy. The PPO model had a 13% max drawdown which was less than the equal weight portfolio 20% and also variance optimized portfolio 17% as shown in Figure 2. This means that PPO is a good approach of managing great losses in the market in downhill spells explaining why it is a safer and more stable form of investment.

Transaction Costs: The accumulated total transaction cost of 1.5% by the PPO model is credible considering the high activity from the sea-faring strategy. Transaction costs for both baseline strategies were slightly above (2% for equal-weighted and 1.8% for the mean-variance optimized portfolio) due to more frequent trading.

Table 1. Comparing the performance metrics of the PPO-based Portfolio Optimization, Equal-Weighted Portfolio, and Mean-Variance Optimized Portfolio

Performance Metric	**PPO-Based Portfolio**	**Equal-Weighted Portfolio**	**Mean-Variance Optimized Portfolio**
Cumulative Return	48%	27%	34%
Sharpe Ratio	1.78	0.94	1.22
Maximum Drawdown	13%	20%	17%

Sortino Ratio	2.35	1.12	1.45
Transaction Costs	1.50%	2%	1.80%

PPO-based Portfolio Optimization strategy was compared to classic portfolio approaches, namely, Equal-Weighted Portfolio and Mean-Variance Optimized Portfolio on a three-year test period (historic stock performance for five assets). The cumulative return reported by the PPO model was very high at 48%, far higher than the that of equal weighted portfolio (27% c&r) as well as the the mean-variance optimized portfolio (34%) respectively as shown in Table 1. This is an indication of the ability of the model to adjust in nature of responding to the amenability of market conditions and yielding more profitable decision.

The Sharpe Ratio for the PPO strategy was 1.78 which meant better in terms of risk adjusted returns to the equal weighted portfolio (0.94) and the mean-variance portfolio (1.22). Furthermore, the Maximum Drawdown of 13% of PPO was much smaller than for the two other strategies, indicating its ability to control large looses during downturns. The PPO model carried 1.5% transaction costs, which was marginally better than the other strategies which were providing effective portfolio management. Generally, PPO demonstrated strong returns, risk hedging and cost effectiveness.

5. CONCLUSIONS

The application of Proximal Policy Optimization (PPO) as the technique for doing dynamic portfolio optimization has turned out to be a fruitful strategy for generating better portfolio returns as compared to conventional techniques. The PPO-based approach showed its adaptive nature, its ability to learn an optimal level of asset allocation in order throughout time and its responsive and intelligent approach to formulating an investment strategy. In comparison to conventional portfolio management techniques, the market price of a portfolio using PPO presented superior results in maximising return as well as optimally managing risk. The potential of reinforcement learning in financial application is further illustrated by model's capacity to preserve a good risk-return profile. By including real-world constraints such as transaction costs and portfolio risk measures, the model PPO presents a more realistic practical answer for portfolio optimization under dynamic markets

REFERENCES

[1] G. Lucarelli and M. Borrotti, "A deep Q-learning portfolio management framework for the cryptocurrency market," Neural Computing and Applications, vol. 32, no. 23, 2020, doi: 10.1007/s00521-020-05359-8.

[2] Y. Huang, Y. Jia, and X. Zhou, "Achieving Mean-Variance Efficiency by Continuous-Time Reinforcement Learning," in Proceedings of the 3rd ACM International Conference on AI in Finance, ICAIF 2022, 2022. doi: 10.1145/3533271.3561760.

[3] A. M. Aboussalah and C. G. Lee, "Continuous control with Stacked Deep Dynamic Recurrent Reinforcement Learning for portfolio optimization," Expert Systems with Applications, vol. 140, 2020, doi: 10.1016/j.eswa.2019.112891.

[4] T. Jaisson, "Deep differentiable reinforcement learning and optimal trading," Quantitative Finance, vol. 22, no. 8, 2022, doi: 10.1080/14697688.2022.2062431.

[5] M. Noguer i Alonso and S. Srivastava, "Deep Reinforcement Learning for Asset Allocation in US Equities," SSRN Electronic Journal, 2020, doi: 10.2139/ssrn.3711487.

[6] A. Brini and D. Tantari, "Deep reinforcement trading with predictable returns," Physica A: Statistical Mechanics and its Applications, vol. 622, 2023, doi: 10.1016/j.physa.2023.128901.

[7] V. M. Ngo, H. H. Nguyen, and P. van Nguyen, "Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets?," Research in International Business and Finance, vol. 65, 2023, doi: 10.1016/j.ribaf.2023.101936.

[8] Q. Y. E. Lim, Q. Cao, and C. Quek, "Dynamic portfolio rebalancing through reinforcement learning," Neural Computing and Applications, vol. 34, no. 9, 2022, doi: 10.1007/s00521-021-06853-3.

[9] L. Han, N. Ding, G. Wang, D. Cheng, and Y. Liang, "Efficient Continuous Space Policy Optimization for High-frequency Trading," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2023. doi: 10.1145/3580305.3599813.

[10] P. N. Kolm and G. Ritter, "Modern Perspectives on Reinforcement Learning in Finance," SSRN Electronic Journal, 2019, doi: 10.2139/ssrn.3449401.

[11] M. A. M. al Janabi, "Optimization algorithms and investment portfolio analytics with machine learning techniques under time-varying liquidity constraints," Journal of Modelling in Management, vol. 17, no. 3,

2022, doi: 10.1108/JM2-10-2020-0259.

[12] P. Sarkar, P. Dwibedi, S. S. Deore, T. Gawali, M. D. Kulkarni, and A. Saha, "Portfolio Optimization in Dynamic Markets: Reinforcement Learning for Investment," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 13s, 2024.

[13] T. Sugadev, N. S. Hameed, S. Vijayakumar, P. Tamilarasan, and M. S. Islam, "Portfolio Optimization Using Machine Learning Techniques," in 2023 4th International Conference on Computation, Automation and Knowledge Management, ICCAKM 2023, 2023. doi: 10.1109/ICCAKM58659.2023.10449598.

[14] H. Valian, M. A. Jafari, and D. Golmohammadi, "Resource allocation with stochastic optimal control approach," Annals of Operations Research, vol. 239, no. 2, 2016, doi: 10.1007/s10479-014-1653-z.

[15] H. Xu, C. Xu, H. Yan, and Y. Sun, "Structured products dynamic hedging based on reinforcement learning," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 9, 2023, doi: 10.1007/s12652-023-04657-y

❖❖❖❖❖