Dynamic Asset Allocation with Markowitz Theory and Proximal Policy Optimization

Seoyun Baek

August 17, 2024

Abstract

This study is significant in that it integrates the traditional portfolio proposed by Markowitz with a portfolio based on reinforcement learning. This integration allows for the optimization and maximization of performance by addressing the shortcomings of the two models. The application of reinforcement learning contributes to the dynamic adaptation of Markowitz's theory-based portfolio to rapidly changing market conditions. Furthermore, the Professional Policy Optimization (PPO) algorithm improves the stability of this process through the clipping technique. The portfolio is designed by combining Markowitz's portfolio with the portfolio based on reinforcement learning when determining the proportion of asset distribution at each point in the purchase or sale decision. This resulted in a cumulative return of 168% of Markowitz's portfolio and 336% of the portfolio based on reinforcement learning when back-testing based on market data from the previous year. Furthermore, it was demonstrated that the potential risks associated with high returns are effectively managed by outperforming the two models in terms of the Sharp ratio and Sortino ratio, which represent risk-to-risk performance.

1 Introduction

In the context of portfolio optimization, it is of paramount importance to ascertain the proportion of assets to be distributed in order to achieve the greatest possible performance. Previously, Markowitz's proposed Mean-Variance Optimization (MVO) or the traditional asset distribution weight determination model designed based on it were the primary means of determining asset distribution proportions. However, this approach was not conducive to dynamic responsiveness to changing market conditions in real-time or the ability to grasp nonlinear relationships between financial data, as the majority of these models were constructed using linear assumptions.

In recent times, as artificial intelligence has demonstrated remarkable proficiency in data analysis and pattern recognition, the technology has been proposed as a potential solution to the aforementioned issues. In particular, reinforcement learning has the advantage of enabling the model to adapt rapidly even in the event of changes in market conditions through interaction with a given environment, such as financial markets. When the model based on this was tested using actual market data, it demonstrated superior performance compared to a traditional portfolio.

The sole use of reinforcement learning for determining the proportion of asset distribution inevitably entails an inevitable loss due to the inherent instability of the initial learning stage. Furthermore, the Epsilon Decay phenomenon introduces a potential issue wherein the performance of the model may be influenced by the initial exploration process, particularly due to the inherent randomness involved.

Accordingly, this study sought to address the deficiencies of the two portfolios by developing a hybrid portfolio that combined the distribution weight of the Markowitz portfolio with the distribution weight of the portfolio based on reinforcement learning. This hybrid portfolio was designed to facilitate decision-making by the reinforcement learning model, specifically in determining the distribution weight of assets. Furthermore, the introduction of Proximal Policy Optimization (PPO) enabled the model to update the policy in a stable manner through the clipping technique.

2 Related Works

2.1 Portfolio using Reinforcement Learning only

A number of studies have been conducted with the objective of constructing a portfolio that is based exclusively on reinforcement learning. Moreover, studies have been conducted with the objective of enhancing the efficacy of reinforcement learning models through the integration of adversarial deep reinforcement learning techniques [1]. Moreover, studies have been conducted in which graph neural networks are applied to value-based algorithms, such as the Deep Q-Network (DQN) [2]. However, in the case of these studies, the performance of the model is significantly influenced by the initial exploration process. Furthermore, despite these efforts, the model still exhibits limitations due to the high probability of underperformance during the adaptation process to market conditions.

2.2 Portfolio using Integration of DDPG and Markowitz Theory

Zheng et al. [3] have conducted studies that integrate Markowitz Theory with a reinforcement learning model based on DDPG. However, the DDPG algorithm is highly susceptible to noise, which limits its efficacy in financial market scenarios where environmental fluctuations are prevalent. Furthermore, the DDPG algorithm necessitates a multitude of hyperparameters, rendering the process of tuning the model exceedingly intricate. Moreover, Araújo et al. [4] conducted a study that integrated the two portfolios using knowledge distilling. However, this approach resulted in a portfolio that was overly influenced by the Markowitz Theory, thereby limiting its ability to respond promptly to real-time market data.

3 Methodology

3.1 Data Collection

Stocks from the S&P 500 were downloaded over a ten-year period, with the downloaded data subsequently accumulated.

3.2 Reinforcement Learning-based Baseline Model

Once the portfolio balance, distribution weight, and rate of return were established as a state at the present time, an action was devised to allow for the adjustment of the portfolio weight at a ratio between -10% and 10% for each asset. In the case of rewards, they were distributed according to the rate of return of the current portfolio. If the portfolio balance increased, a positive reward was given in an amount equal to the rate of return. Conversely, if the portfolio balance decreased, a negative reward was given.

A policy-based algorithm was selected as the reinforcement learning algorithm due to the inherent volatility of asset prices and returns, which necessitates continuous adjustment of the proportion of assets. Additionally, the financial market is characterized by high volatility, which can rapidly result in significant losses. To address this, the PPO algorithm was employed to facilitate stable policy convergence through the clipping technique.

3.3 Integration with Markowitz Thoery

In the baseline model, the portfolio weight was determined through a reinforcement learning algorithm in the process of taking an action at each step. However, in this case, the portfolio's performance was markedly diminished at the outset of the learning process, and there was also an issue with the agent's inability to adapt to the environment due to the fact that the learning was occurring in the incorrect direction.

Consequently, the stability of the model was ensured by employing the ensemble technique to take action with the weighted average of the portfolio weight derived from the Markowitz theory's mean-variance optimization and the portfolio weight determined through the reinforcement learning algorithm.

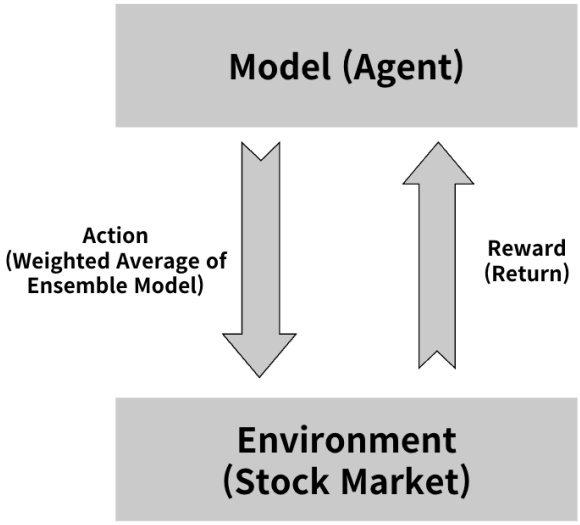

Figure 1: Workflow of Reinfocement Learning-Markowitz Theory Integrated Model

The Markowitz theory initially determines the proportion of asset distribution for the model, after which the reinforcement learning algorithm is weighted by 0.01 at each step, beginning at a 1:1 ratio. This compensates for the instability of reinforcement learning in the early stages of learning and enables dynamic adaptation to the market through reinforcement learning, with the objective of maximizing performance.

4 Experiment

4.1 Backtesting Environment

To assess the efficacy of the portfolio, a backtest was conducted using actual market data from August 1, 2023 to August 1, 2024. Furthermore, the value of the portfolio could be quantified based on the rate of return over time when configuring the environment.

4.2 Result and Analysis

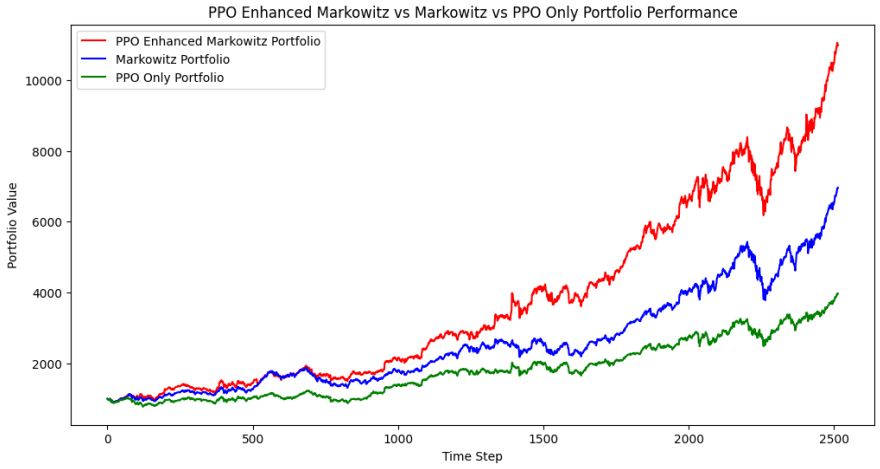

Figure 2: Portfolio Value Changes Over Time Step

In this study, we conducted a comparative analysis of the value changes over time of three portfolios: the portfolio introduced in this study (red graph), the portfolio using Markowitz theory only (blue graph), and the portfolio using reinforcement learning only (green graph).

The S&P 500 stocks were classified into three categories based on volatility: high, medium, and low. This allowed for the selection of stocks from each category to form a portfolio. The implementation of Markowitz theory resulted in an increased proportion of low-volatile stocks and a stable improvement in portfolio value. However, the rate of return exhibited a decline in the latter half of the period, indicating an inability to adapt to changes in market conditions.

The reinforcement learning portfolio exhibited the least favorable performance among the three portfolios, with a higher volatility than that of the Markowitz portfolio. Although performance improved gradually over time, high performance was not recorded due to low returns in the early stages.

In the case of the portfolio introduced in this study, stable returns were recorded from the beginning through asset allocation based on Markowitz's theory. Furthermore, the rate of return gradually accelerated, reaching a record high portfolio value. This indicates that Markowitz's stable portfolio optimization and the dynamic adaptability of reinforcement learning algorithms based on PPOs are effectively combined to reflect market changes.

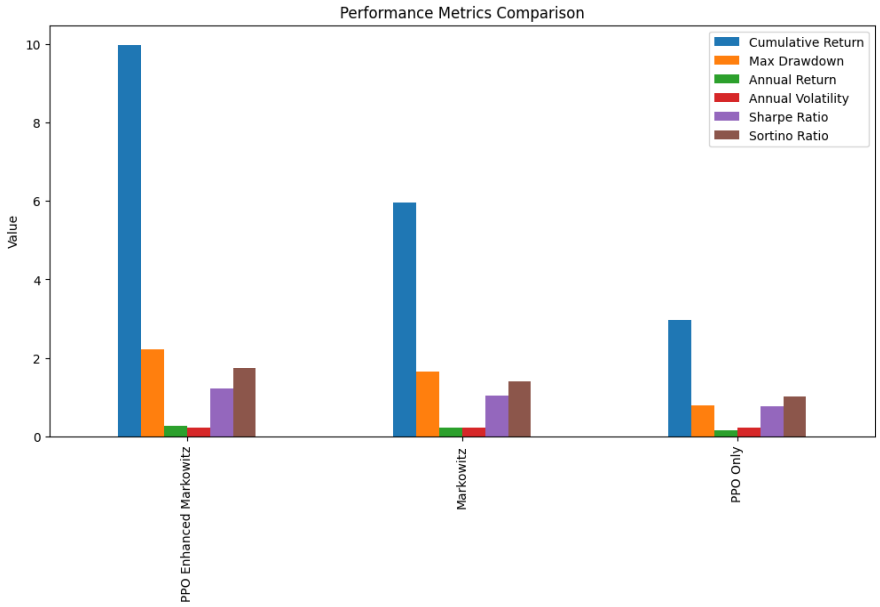

Figure 3: Comparison of Performance Metrics Across Different Portfolio Strategies

Metrics	**PPO Enhanced Markowitz**	**Markowitz**	**PPO Only**
Cumulative Return	9.9836	5.9599	2.9685
Max Drawdown	2.2110	1.6512	0.7817
Annual Return	0.2640	0.2165	0.1606
Annual Volatility	0.2184	0.2093	0.2115
Sharpe Ratio	1.2089	1.0343	0.7595
Sortino Ratio	1.7354	1.3966	1.0155

Table 1: Specific Values of Figure 3

To facilitate a comparison of the three portfolios in terms of their respective performance, the cumulative and annual returns, maximum losses, Sharpe's index, and Sortino's index were represented in the form of a visual representation as illustrated in the aforementioned figure. With regard to cumulative returns, which are typically employed as a means of assessing the value of a portfolio, the portfolio under consideration in this study yielded returns of 168% in comparison to the Markowitz portfolio and 336% in comparison to the reinforcement learning portfolio based on PPO. Additionally, the maximum loss is also the highest among the three models. However, it is noteworthy that the risk

associated with high returns is effectively managed, as evidenced by the portfolio's performance in the Sharpe and Sortino indices, which provide a measure of performance against risk.

5 Conclusion

The objective of this study was to enhance dynamic adaptability to the market and achieve stability by complementing the shortcomings of a portfolio based on Markowitz's theory and a portfolio based on a reinforcement learning algorithm. The portfolio introduced in this study demonstrated superior performance compared to a portfolio employing a single strategy in terms of portfolio valuation indicators, including cumulative returns. Additionally, the portfolio exhibited exemplary risk management capabilities.

This study is notable for its presentation of a novel portfolio optimization strategy, designed to address the dynamic and volatile nature of the financial market. However, the issue of data dependence and complexity inherent to machine learning-based models, including reinforcement learning, represents a significant challenge that requires further investigation. Consequently, future research should aim to enhance the versatility of reinforcement learning models across diverse market environments and pursue additional risk management improvements.

References

[1] Jiang, Z., Xu, D., and Liang, J. (2017). Adversarial Deep Reinforcement Learning in Portfolio Management.

[2] Wang, Y., and Ni, Z. (2019). GraphSAGE with Deep Reinforcement Learning for Financial Portfolio.

[3] Zheng, G., and Yu, Y. (2021). Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning.

[4] Araújo, T., Dias, R., and Leão, P. (2021). Markowitz Meets Bellman: Knowledge-distilled Reinforcement Learning for Portfolio Management.