Master's Thesis

Portfolio Selection via Deep Reinforcement Learning: Comparative Analysis with Classical Strategies

Jae Yeon Park

Department of Mathematical Sciences

2024

Portfolio Selection via Deep Reinforcement Learning: Comparative Analysis with Classical Strategies

Jae Yeon Park

Department of Mathematical Sciences

Portfolio Selection via Deep Reinforcement Learning: Comparative Analysis with Traditional Strategies

A thesis/dissertation submitted to

in partial fulfillment of the
requirements for the degree of
Master of Science

Jae Yeon Park

06.12.2024 of submission

Approved by

Advisor

Selection via Deep Reinforcement Learning: Comparative Analysis with Traditional Strategies

Jae Yeon Park

This certifies that the thesis/dissertation of Jae Yeon Park is approved.

06.12.2024 of submission

Signature

Advisor:

Signature

Signature

One Stopping time, another Martingale

Abstract

Portfolio management involves the strategic allocation of assets to achieve investment objectives while minimizing risk. This study explores portfolio management by comparing traditional Mean-Variance Optimization (MVO), the Black-Litterman Model, and Deep Reinforcement Learning (DRL) for optimizing portfolio allocations. MVO utilizes the Ledoit-Wolf shrinkage method to estimate covariance matrices and applies Efficient Frontier techniques for optimization. The Black-Litterman Model extends MVO by incorporating investor views into the equilibrium market returns derived from the Capital Asset Pricing Model (CAPM), offering a more balanced approach. In contrast, Proximal Policy Optimization (PPO) is used as the DRL method to dynamically adjust portfolio weights. Empirical analysis, based on backtesting with historical market data, shows that the DRL approach significantly outperforms both the MVO and the Black-Litterman Model across various performance metrics, including cumulative return, annual return, volatility, and Sharpe ratio. These results highlight the potential of DRL, particularly PPO strategies, as robust tools for dynamic and adaptive portfolio management, capable of achieving superior returns and effectively managing risks effectively in modern financial markets.

Contents

Chapter 1

Introduction

Overview

Portfolio management is a critical component of financial services, focusing on allocating funds across assets to achieve uncorrelated returns while minimizing risk and costs. Modern Portfolio Theory (MPT), introduced by Harry Markowitz, has historically guided portfolio optimization by balancing risk and return [1]. Recent advances in machine learning, particularly Deep Reinforcement Learning (DRL), provide new methods for optimizing portfolio allocations by learning from historical data and adapting to market changes. This paper compares traditional Mean-Variance Optimization (MVO), the Black-Litterman Model, and DRL to evaluate their effectiveness in portfolio management, highlighting the strengths and limitations of each approach through backtesting.

This research builds on the paper "Deep Reinforcement Learning for Optimal Portfolio Allocation: A Comparative Study with Mean-Variance Optimization" by Sood et al. from J.P. Morgan AI Research and the Oxford-Man Institute of Quantitative Finance [2]. By conducting a series of backtests, this study compares the effectiveness of classical MVO against model-free DRL methods. It highlights the strengths and limitations of each approach, providing insights into their real-world application. Further experiments under varied conditions delve deeper into the comparative effectiveness of MVO and DRL strategies, enhancing traditional methods with advanced machine learning insights.

Summary of Contents

- **Chapter 2:** Theoretical Principles - Overview of Mean-Variance Optimization (MVO), the Black-Litterman Model, and Deep Reinforcement Learning (DRL).
- **Chapter 3:** Algorithmic Implementation - Methods and frameworks for implementing MVO, Black-Litterman, and DRL approaches.
- **Chapter 4:** Experimental Analysis - Results of backtesting these strategies on U.S. and South Korean markets, with comparative performance analysis.
- **Chapter 5:** Conclusion - Summarizes findings and suggests future research directions.

Chapter 2

Theoretical Principles

Portfolio optimization can be divided into two phases: the formation of beliefs about future security performance and the subsequent construction of portfolios based on those beliefs. Modern Portfolio Theory (MPT), introduced by Harry Markowitz in 1952, provides a framework for achieving maximum expected return for a given level of risk through diversification. Mean-Variance Optimization (MVO) uses historical data to estimate returns and covariances with the goal of maximizing Sharpe Ratio through optimal portfolio allocation. The Black-Litterman model incorporates investor sentiment into equilibrium market returns to create more personalized portfolios. Deep Reinforcement Learning (DRL) represents the confluence of reinforcement learning (RL) and deep learning, enabling the discovery of optimal portfolio strategies through algorithms such as DQN and PPO. Each method can be evaluated through back-testing on historical data to assess its effectiveness in optimizing portfolio allocations and managing risk.

I Classical Strategies

Mean-Variance Optimization (MVO)

Mean-Variance Optimization (MVO), pioneered by Harry Markowitz in his influential 1952 paper "Portfolio Selection," is considered a pillar of modern portfolio theory. MVO offers a quantitative approach to building portfolios that balance risk and return. It allows investors to construct portfolios that aim for the highest possible return within a given risk tolerance, or conversely, minimize risk for a desired level of return. This methodology has become a cornerstone of investment management and finance, influencing a wide range of portfolio construction and risk mitigation strategies [1, 3–6].

Optimization Framework

MVO provides a mathematical framework for allocating investment capital among various assets. This framework helps investors achieve their goals, such as maximizing return for a given level of risk or achieving a targeted return with minimal risk exposure. Here, risk is typically measured by the portfolio's volatility, which reflects the fluctuations in its returns.

The MVO framework is based on several key components:

The expected return of a portfolio, $E(R_p)$, is the weighted sum of the expected returns of the individual assets within the portfolio. It is mathematically expressed as:

$$E(R_p) = \sum_{i=1}^{n} w_i E(R_i)$$

where w_i is the weight of asset i in the portfolio, $E(R_i)$ is the expected return of asset i, and n is the total number of assets in the portfolio. The sum of the weights w_i must equal 1, indicating full investment.

The risk of the portfolio, measured by its variance σ_p^2, is given by:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}$$

where σ_{ij} is the covariance between the returns of assets i and j. The covariance σ_{ij} captures how the returns of the assets move together, and the summation includes all pairs of assets in the portfolio.

Portfolios on the efficient frontier are considered optimal because they provide the best return for their level of risk. The shape of the efficient frontier is typically an upward-sloping curve, illustrating the trade-off between risk and return. Investors select a portfolio on this frontier based on their individual risk tolerance and return expectations.

By introducing a risk-free asset into the analysis, the Capital Market Line (CML) can be constructed, which represents the set of portfolios that optimally combine the risk-free asset and risky assets to provide the best possible risk-return trade-off. The slope of the CML is determined by the Sharpe ratio [7], which is calculated as follows:

$$S = \frac{E(R_p) - R_f}{\sigma_p}$$

where $E(R_p)$ is the expected return of the portfolio, R_f is the risk-free rate, and σ_p is the standard deviation of the portfolio's excess return. The numerator $E(R_p) - R_f$ represents the excess return of the portfolio over the risk-free rate, while the denominator σ_p indicates the risk (volatility) of the portfolio. This ratio provides a standardized measure of the risk-adjusted return of an investment.

Optimization Problem

The optimization problem in MVO involves finding the portfolio weights w_i that minimize the portfolio variance for a given expected return, or equivalently, maximize the expected return for a given level of variance. This can be formulated as a quadratic programming problem:

$$\min \sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}$$

subject to:

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} w_i E(R_i) = E(R_p)$$

where $E(R_p)$ is the expected return of the portfolio, R_f is the risk-free rate, σ_p is the standard deviation of the portfolio, and w_i are the portfolio weights.

Black-Litterman Model

The Black-Litterman model, developed by Fischer Black and Robert Litterman at Goldman Sachs in 1990, addresses the limitations of traditional Mean-Variance Optimization (MVO) by integrating subjective views with market equilibrium returns. This model enhances portfolio construction by incorporating investor insights alongside the equilibrium returns derived from the Capital Asset Pricing Model (CAPM) [8, 9]. By blending equilibrium market returns with specific investor views, the model provides a balanced approach to portfolio optimization, addressing estimation errors and extreme portfolio weights often found in traditional MVO.

Optimization Framework

The Black-Litterman framework includes two key components:

Equilibrium Returns: Uses the CAPM to derive baseline excess returns from the market portfolio:

$$\Pi = \delta \Sigma w_m$$

where Π represents equilibrium excess returns, δ is a risk aversion coefficient, Σ is the covariance matrix of returns, and w_m is the market capitalization weights [8].

Incorporating Views: Adjusts these equilibrium returns based on specific investor views, resulting in a new return vector:

$$E(R) = \Pi + \tau \Sigma P^T (\Omega + \tau P \Sigma P^T)^{-1} (Q - P\Pi)$$

where τ is a scaling factor, P identifies the assets in the views, Ω is the uncertainty in the views, and Q represents investor views on returns [8, 9].

Optimization Problem

The optimization problem in the Black-Litterman model involves adjusting the equilibrium returns Π to incorporate investor views, resulting in a new vector of expected returns $E(R)$. This new return vector is then used in a Mean-Variance Optimization framework to determine the optimal portfolio weights w_i. The objective is to find the weights that maximize the Sharpe ratio:

$$\max \frac{E(R_p) - R_f}{\sigma_p}$$

subject to:

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} w_i E(R_i) = E(R_p)$$

where $E(R_p)$ is the expected return of the portfolio, R_f is the risk-free rate, σ_p is the portfolio's standard deviation, and w_i are the portfolio weights [8].

By incorporating investor views, the Black-Litterman model enhances the stability and robustness of the portfolio optimization process, mitigating the impact of estimation errors and extreme weights. This systematic approach addresses several limitations inherent in traditional MVO methods, offering a personalized yet systematic approach to portfolio optimization [9, 10].

II Deep Reinforcement Learning (DRL)

Deep Reinforcement Learning (DRL) combines reinforcement learning (RL) with deep learning to create powerful algorithms capable of learning complex behaviors from high-dimensional inputs. This combination allows DRL algorithms to perform exceptionally well in tasks such as game playing, robotics, and financial trading.

Fundamentals of Reinforcement Learning and Deep Learning

Reinforcement Learning (RL) Basics

Reinforcement Learning (RL) involves an agent learning to make decisions by interacting with an environment to maximize cumulative rewards. This process is typically modeled using a Markov Decision Process (MDP). An MDP is characterized by a tuple (S, A, P, R, γ), where S represents all possible states the agent can encounter, and A denotes all possible actions the agent can take. The transition probabilities $P(s'|s,a)$ describe the likelihood of moving from state s to state s' after taking action a. The rewards $R(s,a,s')$ quantify the immediate gain from the transition s to s' via action a. The discount factor γ, a value between 0 and 1, determines the present value of future rewards, balancing immediate and future gains. The objective in RL is to learn a policy $\pi(s)$ that maximizes the expected sum of discounted rewards:

$$E\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1})\right]$$

[11, 12].

Deep Learning Basics

Neural networks are an integral part of Deep Learning and are inspired by the structure and function of the human brain. A neural network is composed of an input layer, several hidden layers, and an output layer. The input layer receives data, the hidden layers process the data through multiple transformations, and the output layer generates the final result. Learning in neural networks involves adjusting the weights of connections between neurons to minimize a loss function. This is achieved through backpropagation and optimization algorithms like stochastic gradient descent (SGD) [13, 14].

DRL Algorithms: Combining RL and Deep Learning

Deep Q-Networks (DQN)

Deep Q-Networks (DQN) integrate Q-learning with deep neural networks to approximate the Q-value function, which predicts the expected reward for actions taken in a given state. This allows the agent to handle high-dimensional state spaces. The Q-learning algorithm is off-policy, updating the Q-value function $Q(s,a)$ using the rule:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right)$$

where α is the learning rate, r is the reward, and γ is the discount factor. Key features include experience replay, which stores past transitions (s,a,r,s') in a replay buffer to break correlation between updates, and target networks, which stabilize training by using a separate network to compute target Q-values:

$$\text{Loss} = \left(r + \gamma \max_{a'} Q_{\text{target}}(s',a') - Q(s,a) \right)^2$$

[15].

Algorithm 1 Training DQN Agent

Initialize the Q-network $Q(s,a;\theta)$ with random weights θ.
Initialize target network $Q_{\text{target}}(s,a;\theta^-)$ with weights $\theta^- = \theta$.
for each episode **do**
  Initialize state s_0.
  for each step in the episode **do**
    With probability ε, select a random action a; otherwise, select $a = \arg\max_a Q(s,a;\theta)$.
    Execute action a, observe reward r and next state s'.
    Store the transition (s,a,r,s') in a replay buffer.
    Sample a mini-batch of transitions (s_j,a_j,r_j,s'_j) from the replay buffer.
    Compute target Q-value for each transition:

$$y_j = \begin{cases} r_j & \text{if } s'_j \text{ is terminal} \\ r_j + \gamma \max_{a'} Q_{\text{target}}(s'_j,a';\theta^-) & \text{otherwise} \end{cases}$$

    Perform a gradient descent step on the loss function:

$$\text{Loss} = \frac{1}{N} \sum_j (y_j - Q(s_j,a_j;\theta))^2$$

    Update Q-network weights θ using stochastic gradient descent.
    Periodically update target network weights $\theta^- = \theta$.
  end for
end for

Asynchronous Advantage Actor-Critic (A3C)

A3C is an on-policy algorithm that uses multiple worker agents to interact with the environment in parallel. Each worker maintains its own copy of the environment and network parameters, and asynchronously updates a global model. This approach improves learning speed and exploration. In the actor-critic framework, the actor updates the policy and the critic evaluates it by estimating the value function. The advantage function $A(s,a)$ measures the difference between the expected return of action a in state s and the average return of all actions in s:

$$A(s,a) = Q(s,a) - V(s)$$

where $Q(s,a)$ is the action-value function, and $V(s)$ is the state-value function [16].

Algorithm 2 Training DRL Agent with A3C

1: Initialize global network parameters θ.
2: Spawn N worker agents with local parameters θ'.
3: **for** each worker **do**
4:     Initialize local parameters $\theta_i' = \theta$.
5:     **while** not done **do**
6:         Collect a trajectory $\{s_t, a_t, r_t, s_{t+1}\}$ by interacting with the environment.
7:         Compute the advantage $A(s_t, a_t) = \sum_{t'} \gamma^{t'-t} r_{t'} - V(s_t; \theta_v)$.
8:         Compute the actor loss:

$$L_{\text{actor}} = -\log \pi(a_t|s_t;\theta) A(s_t, a_t)$$

9:         Compute the critic loss:

$$L_{\text{critic}} = \left(\sum_{t'} \gamma^{t'-t} r_{t'} - V(s_t; \theta_v) \right)^2$$

10:         Compute gradients $\nabla_{\theta'} L_{\text{total}}$ where $L_{\text{total}} = L_{\text{actor}} + \beta L_{\text{critic}}$.
11:         Update global parameters asynchronously using:

$$\theta \leftarrow \theta - \eta \nabla_{\theta'} L_{\text{total}}$$

12:         Synchronize local parameters $\theta_i' = \theta$.
13:     **end while**
14: **end for**

Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) is an on-policy Deep Reinforcement Learning (DRL) algorithm designed to provide stable and reliable policy updates. It optimizes the policy directly by computing the gradient of the expected reward with respect to the policy parameters.

PPO uses a policy network π_θ, a neural network parameterized by θ that maps states to a probability distribution over actions. The value function $V_\phi(s)$, parameterized by ϕ, estimates the expected return from a given state s. This helps reduce the variance of the policy gradient estimates by providing a baseline for advantage estimation. PPO uses a clipped objective function to stabilize training by preventing large updates, given by:

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_t \left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right)\right]$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ is the probability ratio of taking action a_t under the new and old policies, and $\hat{A}_t$ is the advantage function. The advantage function $\hat{A}_t = R_t - V_\phi(s_t)$ measures how much better an action a_t is compared to the expected return from state s_t.

The optimization problem in PPO involves finding the policy parameters θ that maximize the expected sum of discounted rewards:

$$\max_\theta \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_t\right]$$

subject to:

$$\text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}, 1-\varepsilon, 1+\varepsilon\right)$$

where γ is the discount factor, and ε is the clipping parameter.

To implement the PPO strategy, the policy network π_θ is initialized, and the current policy is run to collect trajectories of states, actions, and rewards. The advantage estimates $\hat{A}_t$ are then calculated using the value function. The policy is updated by maximizing the clipped surrogate objective using stochastic gradient ascent. This process is repeated to continually improve the policy. [17]

Algorithm 3 Training DRL Agent with PPO

Initialize the policy network π_θ and the value function V_ϕ.
for each iteration **do**
  Collect trajectories by running the current policy π_θ.
  Compute rewards R_t for each trajectory.
  Estimate the advantage $\hat{A}_t$ using:

$$\hat{A}_t = R_t - V_\phi(S_t)$$

  Compute the probability ratio:

$$r_t(\theta) = \frac{\pi_\theta(a_t|S_t)}{\pi_{\theta_{\text{old}}}(a_t|S_t)}$$

  Define the clipped objective function:

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right)\right]$$

  Update the policy network π_θ by maximizing the clipped objective function:

$$\theta \leftarrow \theta + \alpha \nabla_\theta L^{\text{CLIP}}(\theta)$$

  Update the value function V_ϕ by minimizing the loss:

$$\phi \leftarrow \phi - \beta \nabla_\phi \left(V_\phi(S_t) - R_t\right)^2$$

end for

Chapter 3

Algorithmic Implementation

I Mean-Variance Optimization (MVO)

MVO Approach

In this study, MVO is used as a baseline method for comparison with the model-free Deep Reinforcement Learning (DRL) approach. To ensure a fair comparison, consistent training and operating parameters are maintained. Specifically, the MVO approach employs a 60-day lookback period to derive the mean and covariance estimates for asset returns, ensuring that the portfolio is regularly updated with current market data and adapts to changing conditions.

The optimization problem in MVO is to find the portfolio weights w_i that minimize the portfolio variance for a given expected return or equivalently maximize the expected return for a given level of variance. This is formulated as a convex optimization problem:

$$\min_{\mathbf{w}} \mathbf{w}^T \Sigma \mathbf{w}$$

subject to:

$$\mathbf{w}^T \mu \geq \mu^*, \quad \sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

Here, $\mathbf{w}$ is the weight vector for the assets, Σ is the covariance matrix, and μ^* is the desired return. This optimization puzzle can be effectively solved using convex optimization techniques.

The implementation framework for MVO includes the use of a 60-day historical data window to determine the average returns and covariances associated with asset returns. Expected returns are computed as sample averages over the look-back period, while the covariance matrix is estimated using the Ledoit-Wolf shrinkage method to reduce estimation error and ensure positive semidefinite matrices. The portfolio optimization problem is formulated to maximize the Sharpe Ratio, defined as the ratio of expected excess returns to portfolio volatility. Weights are recalculated daily based on the latest data, ensuring that the portfolio adapts to changing market conditions.

The MVO approach requires no training; it simply relies on the historical 60-day window to determine portfolio weights. For example, an MVO backtest initiated in January 2012 would use historical data from October 2011. By using MVO, we aim to establish a benchmark for evaluating the effectiveness of the DRL approach in optimizing portfolio allocations. The comparison will highlight the strengths and limitations of DRL in a real-world financial environment.

The rolling MVO implementation involves a series of steps that are repeated over a defined rolling window, ensuring that the portfolio is updated daily based on the most recent data. This allows for continuous adjustment and optimization of the portfolio, providing a dynamic approach to asset allocation.

MVO Algorithm

The Mean-Variance Optimization (MVO) strategy is implemented using the `pypfopt` library [18]. The following is a detailed explanation of the MVO process, focusing on the key steps and methodology.

Environment Setup

The environment setup for MVO involves downloading historical price data for the selected assets using the `yfinance` library. This data includes daily adjusted close prices for the specified date range. Simple returns are then calculated as the percentage change in price from one period to the next using the pandas library's `pct_change` function. This forms the basis for estimating expected returns and risk [19].

Experimental Framework

The MVO framework handles the calculation of expected returns and the covariance matrix. The expected returns are estimated using the `mean_historical_return` function from the `pypfopt` library, which calculates the mean of historical returns over the lookback period. The covariance matrix is estimated using the Ledoit-Wolf shrinkage method to reduce estimation error and to ensure that the matrix is non-singular and positive semidefinite. The `EfficientFrontier` class from `pypfopt` constructs the efficient frontier and solves for the portfolio that maximizes the Sharpe ratio. This optimization process involves solving a quadratic programming problem to find the optimal weights that maximize the Sharpe ratio, defined as the ratio of expected excess returns to portfolio volatility.

Backtesting

Backtesting simulates the performance of the optimized portfolio over time using historical price data and the calculated portfolio weights to compute portfolio values and returns. Each backtest starts with an all-cash portfolio allocation. The historical price data and the calculated portfolio weights are used to compute the portfolio values and returns. This process helps evaluate the effectiveness of the MVO strategy in achieving an optimal balance between risk and return. The optimized portfolio weights are used to backtest the portfolio, simulating its performance over time and providing insight into the effectiveness of the strategy in a real financial environment.

Algorithm 4 Rolling MVO Implementation

Initialize directories for saving results.
Download historical price data for the selected assets using `yfinance`.
Calculate simple returns $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ from the price data.
Define a 60-day lookback period and risk-free rate for calculations.
Initialize data structures to store weights, expected returns, covariance matrices, Sharpe ratios, and changes in portfolio weights.
for each rolling window starting from the lookback period to the end of the data **do**
    Extract data for the current window.
    Compute expected returns $\mu = \frac{1}{N}\sum_{t=1}^{N} R_t$ using `mean_historical_return` from `pypfopt`.
    Estimate the covariance matrix Σ using the Ledoit-Wolf shrinkage method:

$$\Sigma = \delta \mathbf{F} + (1-\delta)\mathbf{S}$$

    where $\mathbf{F}$ is the prior (structured estimator), $\mathbf{S}$ is the sample covariance matrix, and δ is the shrinkage parameter.
    Create an `EfficientFrontier` object with the computed returns and covariance matrix.
    Solve the optimization problem to maximize the Sharpe ratio:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T(\mu - R_f)}{\sqrt{\mathbf{w}^T \Sigma}}$$

    subject to:

$$\sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

    Store the optimized weights, expected returns, covariance matrix, and Sharpe ratios.
    Compute the change in portfolio weights $\Delta \mathbf{w}$ from the previous period:

$$\Delta \mathbf{w} = \mathbf{w}_t - \mathbf{w}_{t-1}$$

end for
Backtest the portfolio using the stored weights to compute portfolio values over time:

$$\text{Portfolio Value}_t = \sum_{i=1}^{n} w_i P_{i,t}$$

Plot and save the portfolio value, daily returns, daily asset weights, Sharpe ratios, and changes in portfolio weights.

II Black-Litterman Model

Black-Litterman Model Approach

The Black-Litterman Model is a sophisticated portfolio optimization method that integrates traditional Modern Portfolio Theory (MPT) with subjective investor views to create balanced and personalized portfolio strategies. The Black-Litterman Model is employed as an alternative portfolio optimization method for comparison with Deep Reinforcement Learning (DRL). This model enhances traditional Mean-Variance Optimization (MVO) by integrating investor views with market equilibrium returns. To ensure a fair comparison with DRL, the Black-Litterman model is also implemented using a 60-day rolling optimization window. This means that the portfolio is updated daily based on the most recent 60 days of data. By continuously updating the portfolio, this approach adapts to changing market conditions and maintains relevance with current data.

To optimize the portfolio in the Black-Litterman model is to maximize the expected return while considering the investor's views:

$$\max_{\mathbf{w}}(\mathbf{w}^T \mathbf{E}(\mathbf{R}) - \lambda \mathbf{w}^T \Sigma \mathbf{w})$$

subject to:

$$\sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

where $\mathbf{w}$ is the weight vector for the assets, $\mathbf{E}(\mathbf{R})$ is the adjusted expected return vector incorporating investor views, Σ is the covariance matrix, and λ is a risk aversion coefficient.

Incorporating views in the Black-Litterman model involves setting the expected growth rate over the 60-day lookback period as the investor's view. This allows the model to adjust the equilibrium returns according to recent market performance, leading to more dynamic and responsive portfolio optimization.

Black-Litterman Model Algorithm

Environment Setup

The implementation starts by collecting historical price data for selected assets using the `yfinance` library. This data includes daily adjusted close prices, and simple returns are calculated as the percentage change in price from one period to the next, forming the basis for estimating expected returns and risks [19].

Experimental Framework

Equilibrium returns are calculated using the Capital Asset Pricing Model (CAPM) to derive baseline excess returns from the market portfolio. The library utilized includes `pypfopt` functions such as `black_litterman`, `risk_models`, `expected_returns`, and `BlackLittermanModel` [20].

The optimization problem is formulated to maximize the Sharpe ratio while adhering to portfolio constraints. The Efficient Frontier is constructed using the `EfficientFrontier` class from `pypfopt`. The optimization is performed using a rolling 60-day window to ensure the portfolio adapts to the latest market data. Each iteration involves calculating past returns, estimating the covariance matrix with the Ledoit-Wolf method, and incorporating investor views to adjust equilibrium returns. The adjusted returns are then used to optimize the portfolio, with weights recalculated daily.

Backtesting

Backtesting simulates the performance of the optimized portfolio over time, starting with an all-cash portfolio allocation. Daily portfolio values and returns are calculated based on the optimized weights. This process evaluates the effectiveness of the Black-Litterman strategy in balancing risk and return under real market conditions. The backtest results, including portfolio values, daily returns, and portfolio weights, provide insights into the strategy's performance over the specified period.

Algorithm 5 Black-Litterman Model Implementation

Initialize directories for saving results.
Download historical price data for the selected assets using `yfinance`.
Calculate simple returns $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ from the price data.
Define a 60-day lookback period and risk-free rate for calculations.
Initialize data structures to store weights, expected returns, covariance matrices, Sharpe ratios, and changes in portfolio weights.
for each rolling window starting from the lookback period to the end of the data **do**
    Extract data for the current window.
    Compute expected returns $\mu = \frac{1}{N}\sum_{t=1}^{N} R_t$ using `mean_historical_return` from `pypfopt`.
    Estimate the covariance matrix Σ using the Ledoit-Wolf shrinkage method:

$$\Sigma = \delta \mathbf{F} + (1-\delta)\mathbf{S}$$

    where $\mathbf{F}$ is the prior (structured estimator), $\mathbf{S}$ is the sample covariance matrix, and δ is the shrinkage parameter.
    Estimate market-implied risk aversion coefficient using `black_litterman.market_implied_risk_aversion`.
    Create a `BlackLittermanModel` object with the covariance matrix and investor views.
    Compute adjusted expected returns using the `bl_returns` method.
    Create an `EfficientFrontier` object with adjusted returns and covariance matrix.
    Solve the optimization problem to maximize the Sharpe ratio:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T(\mathbf{E}(\mathbf{R}) - R_f)}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}}$$

    subject to:

$$\sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

    Store the optimized weights, expected returns, covariance matrix, and Sharpe ratios.
    Compute the change in portfolio weights $\Delta\mathbf{w}$ from the previous period:

$$\Delta\mathbf{w} = \mathbf{w}_t - \mathbf{w}_{t-1}$$

end for
Backtest the portfolio using the stored weights to compute portfolio values over time:

$$\text{Portfolio Value}_t = \sum_{i=1}^{n} w_i P_{i,t}$$

Plot and save the portfolio value, daily returns, daily asset weights, Sharpe ratios, and changes in portfolio weights.

III Deep Reinforcement Learning (DRL)

DRL Approach for Portfolio Management

In this portfolio optimization framework, we model the environment to interact with a reinforcement learning (RL) agent by simulating the market with historical data. The environment processes the agent's actions, rebalances the portfolio, and provides updated states and rewards. Here is the detailed algorithmic process:

The process starts the environment updating the portfolio value by applying the newly acquired weights w_i to the returns of each asset. This is followed by a rebalancing step where the environment adjusts the portfolio weights by multiplying them by the current portfolio value. After this rebalancing, the environment generates the next state representation $S_{t+1}$and transitions to the next time step $t+1$. Within this new time step, the environment calculates the updated portfolio value based on the returns R_{t+1} derived from the most recent asset prices P_{t+1}. Finally, the environment calculates the Differential Sharpe Ratio D_t and passes this value back to the reinforcement learning agent.

The calculation of the new portfolio value is given by:

$$\text{new\_portfolio\_value} = \sum_{i=1}^{n} w_i \times (1 + R_{i,t}) \times \text{current\_portfolio\_value}$$

Where, w_i are the portfolio weights for asset i, $R_{i,t}$ are the returns of asset i at time t, and n is the number of assets in the portfolio.

This process ensures that the portfolio is updated to reflect the latest market conditions, providing accurate feedback to the agent for learning and optimization of investment strategies.

Actions

For portfolio allocation across N assets, an agent assigns portfolio weights $\mathbf{w} = [w_1, \ldots, w_N]$ such that the total weight sums to 1, i.e., $\sum_{i=1}^{N} w_i = 1$. Each weight w_i is in the range $[0, 1]$, indicating a non-leveraged, long-only position. A weight of 0 means no investment in that asset, while a weight of 1 means the entire portfolio is invested in that asset. Although some models allow for short selling (weights less than 0) and leverage (weights greater than 1), this particular implementation restricts itself to long-only, unleveraged positions. These restrictions are enforced by normalizing the agent's action outcomes using the softmax function.

In this implementation, the dimensionality of the action space is reduced by one because the weights must sum to 1. Instead of representing all N dimensions, only $N-1$ dimensions are used, simplifying the action space while still satisfying the portfolio constraints. This is achieved by normalizing the action array and then appending the final weight so that the total sum equals 1.

To adjust the action vector, the actions are normalized as follows: Convert the action outputs to a numpy array, normalize the action values to ensure that they sum to 1, and append the final weight as $1 - \sum_{i=1}^{N-1} w_i$ to maintain the constraint that the total weight is 1.

This approach efficiently manages portfolio weights within the defined constraints, facilitating the implementation of reinforcement learning algorithms for portfolio optimization. The action space for N

assets is inherently reduced to $N-1$ dimensions because the sum of all weights must be 1. Therefore, representing the action space in $N-1$ dimensions simplifies the problem. By normalizing the initial action outputs and then appending the last weight to ensure the sum is 1, this method ensures a valid portfolio allocation. This process involves normalizing the action values so that their sum is 1, and taking an additional step to ensure that the sum of all the weights remains exactly 1 by appending the final weight as the difference from 1. This normalization guarantees that the agent's actions result in a valid, fully invested portfolio that satisfies the constraints of unleveraged, long-only positions.

States

The agent's perception of the market environment at a given time step (t), represented by the state vector (S_t), is structured to capture a comprehensive view of market conditions and is a matrix with dimensions $(n+1)\times T$. Here, n denotes the number of tradable securities, and the additional element accounts for cash c. The state matrix S_t is composed of several key elements:

First, the portfolio allocation vector $\mathbf{w}$ at the beginning of timestep t is represented as:

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n & w_c \end{bmatrix}$$

where w_i is the allocation for asset i, and w_c is the allocation for cash. This vector may differ slightly from the previous timestep due to the conversion of continuous weights into actual investable units (whole shares) and rebalancing to ensure a total allocation of 1.

Next, the historical returns of each asset over a lookback window of length T are included. The one-period simple return on an asset is defined as $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$, where P_t is the asset's price at time t. Daily simple returns are calculated using the closing prices at the end of each day. The historical returns matrix is:

$$\begin{bmatrix} r_{1,t-1} & r_{1,t-2} & \cdots & r_{1,t-T+1} \\ r_{2,t-1} & r_{2,t-2} & \cdots & r_{2,t-T+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,t-1} & r_{n,t-2} & \cdots & r_{n,t-T+1} \end{bmatrix}$$

where $r_{i,t-k}$ represents the historical return for asset i at time $t-k$. The lookback window size T can vary depending on the experiment, with typical values being 60, 30, or 10 days.

Additionally, three market volatility measures at time t are included in the last row: vol20_t (the standard deviation of daily returns over a rolling 20-day period), vol20/vol60_t (the ratio of vol20 to vol60, indicating short-term versus long-term volatility trends), and either VIX_t (the VIX index representing market expectations of S&P 500 volatility) or VKOSPI_t (the VKOSPI index representing market expectations of KOSPI volatility) [21]. These measures are normalized by subtracting the mean and dividing by the standard deviation, using an expanding look-back window to avoid forward-looking bias.

Reward

In portfolio management, optimizing risk-adjusted returns requires a reward function that supports this objective. The traditional Sharpe ratio measures risk-adjusted returns over a period of time T, making it unsuitable for online learning. Instead, the Differential Sharpe Ratio (DSR) evaluates risk-adjusted returns at each timestep t, providing a real-time basis for reinforcement learning. The DSR measures the change in Sharpe Ratio over time, taking into account incremental updates to returns and variances:

$$D_t = \frac{\partial S_t}{\partial t}$$

To derive this, we define R_t as the return at time t, μ_t as the exponential moving average of returns up to time t, and σ_t^2 as the exponential moving average of squared returns up to time t.

The incremental updates are given by:

$$\Delta R_t = R_t - \mu_{t-1} \quad \text{and} \quad \Delta\sigma_t^2 = R_t^2 - \sigma_{t-1}^2$$

Then, the DSR is expressed as:

$$D_t = \frac{\sigma_{t-1}\Delta\mu_t - \frac{1}{2}\mu_{t-1}\Delta\sigma_t^2}{(\sigma_{t-1}^2 - \mu_{t-1}^2)^{3/2}}$$

where:

$$\Delta\mu_t = \frac{R_t - \mu_{t-1}}{t} \quad \text{and} \quad \Delta\sigma_t^2 = \frac{R_t^2 - \sigma_{t-1}^2}{t}$$

Thus, the DSR simplifies to:

$$D_t = \frac{\sigma_{t-1}(R_t - \mu_{t-1}) - \frac{1}{2}\mu_{t-1}(R_t^2 - \sigma_{t-1}^2)}{(\sigma_{t-1}^2 - \mu_{t-1}^2)^{3/2}}$$

In the implementation, arrays **A** and **B** are used to keep track of the first and second moments of the returns, respectively. At each timestep t, the return R_t is calculated, and **A** and **B** are updated using differential updates. The DSR is calculated using the updated values, and the reward at each time step is the sum of the DSR values across all assets. This approach ensures that the agent's reward is directly tied to the risk-adjusted performance of the portfolio, encouraging optimization of DSR rather than just raw returns.

The calculation steps can be summarized as follows:

$$D_t = \frac{B_{t-1}\Delta A_t - \frac{1}{2}A_{t-1}\Delta B_t}{(B_{t-1} - A_{t-1}^2)^{3/2}}$$

with

$$\Delta A_t = R_t - A_{t-1}$$

$$\Delta B_t = R_t^2 - B_{t-1}$$

$$A_t = A_{t-1} + \eta\Delta A_t$$

$$B_t = B_{t-1} + \eta\Delta B_t$$

Here, η is a small value, typically $\approx \frac{1}{252}$ assuming 252 trading days in a year. This ensures that the reinforcement learning agent continuously evaluates and optimizes risk-adjusted returns, making it robust for dynamic portfolio management.

By using the DSR, the reinforcement learning agent can continuously assess and optimize the risk-adjusted returns, making this method robust for dynamic portfolio management [22].

In some experiments, the reward was given as a simple return value. This was done to evaluate the performance based on the format of the reward, and the results will be analyzed later in the analysis.

DRL Algorithm

The Proximal Policy Optimization (PPO) algorithm from `StableBaselines3` is used to train the DRL agent. To increase training efficiency and robustness, vectorized environment wrappers, specifically the Vectorized Sub-Process Vectorized Environment (`Sub-ProcVecEnv`), are used to facilitate multiprocessing across independent instances of the environment. This setup allows the collection of experience rollouts in parallel, leveraging the power of multiprocessing.

A custom trading environment is created using the `gym` library to simulate the trading process. This environment processes market data and executes trades based on the agent's actions, providing a realistic framework for the DRL agent to learn and optimize its trading strategy [23, 24].

Environment Setup

Setting up the environment for PPO involves downloading historical price data for the selected assets using the `yfinance` library. This data includes daily adjusted closing prices for the specified date range. Simple returns are calculated as the percentage change in price from one period to the next using the `pct_change` function from the Pandas library. These simple returns are the basis for estimating expected returns and risk. In addition, volatility metrics such as the 20-day rolling standard deviation of daily returns and the ratio of 20-day to 60-day rolling standard deviations are calculated. These metrics are standardized using an expanding window to ensure comparability over time and to prevent information leakage.

This comprehensive setup ensures that the DRL agent operates in a realistic and dynamic trading environment, allowing it to learn and adapt effectively to market conditions.

Experimental Framework

To overcome the limitations of limited and time-varying financial data, a multi-year testing strategy is implemented for the DRL framework. The entire dataset is segmented into 12 different groups, each spanning seven years. A one-year time lag is applied between each group. Within each group, a five-year window is dedicated to training the DRL agents, followed by a one-year window for validation and another year for out-of-sample testing.

Training Process

During each training period, initial training starts with five agents, each with a unique random seed, using data from 2006 to 2011. The `set_random_seed` function is used to ensure varied initialization. The performance of each agent is periodically evaluated using validation data. At the end of each training round, the agent with the highest total reward during the validation period is selected. The performance of this agent is further validated using data from the sixth year to ensure generalizability and to avoid overfitting. The top-performing agent is then tested on out-of-sample data from the seventh year, providing an unbiased assessment of real-world performance. This process is iterated over multiple periods until the final validation period in 2022. As a result, this multi-year testing approach generates a total of

60 agents (5 agents per window over 12 windows) and enables 12 corresponding backtests for comprehensive evaluation. This robust process of training, validation and testing ensures the development of effective trading strategies.

The PPO algorithm is used to train the agent, with key hyperparameters such as discount factor, learning rate, and clip range tuned for optimal performance. Multiple agents are trained with different random seeds to ensure robustness. Each round of training consists of total 6 million timesteps: Total groups $\times$ Agents per group $\times$ Timesteps per agent $= 12 \times 5 \times 100,000 = 6,000,000$Timesteps. To ensure that the agent gains enough experience in different environments, the size of the rollout buffer is configured as $n\_steps = 252 \times 3 \times n\_envs$. This setting ensures that a sufficient number of historical interactions are stored for effective learning. The batch size is 1260 (252 trading days per year $\times$ 5 years), ensuring that each batch covers multiple trading days. The number of epochs (*n_epochs*) is set to 16, which allows the model to adequately optimize the surrogate loss function. The discount factor γ is 0.9, which balances immediate and future rewards. A hyperparameter called GAE Lambda is set to 0.9. This parameter plays a critical role in the Generalized Advantage Estimator (GAE) by balancing the tradeoff between bias and variance in the estimated advantage values. The clip range is 0.25, which is used to clip the probability ratios to stabilize training. The number of environments (*n_envs*) is set to 10, i.e. 10 environments are run in parallel to collect different experiences and to speed up training. The number of steps per environment (n_steps) is set to 756 to ensure that enough experience is collected in all environments. The learning rate, which controls how much the agent adjusts its policy based on new experience, starts at a higher value of 3×10^{-4} and gradually decays to 1×10^{-5} as training progresses. This decay helps the agent converge to an optimal policy while preventing overfitting to the training data.

Validation Process

The validation process is used to identify the most effective agent from a group of trained candidates. This is achieved by evaluating their performance on a dedicated validation dataset. In the first round of training, five agents are created with different random seeds (initialization). These agents are trained on data from 2006 to 2011. To monitor their progress, their performance is evaluated periodically against the 2011 validation data. The agent that achieves the highest average reward per episode during validation is considered the top performer. This top agent's policy then becomes the basis (seed policy) for training the next group of agents in the next time window. This iterative process of training, validation, and selection using a seed policy from the previous champion continues until the final validation period in 2020. As a result, this multi-stage approach yields a total of 50 agents (5 agents per window trained over 10 windows) and provides 10 corresponding backtests for comprehensive evaluation.

Backtesting Process

Backtesting serves as a critical tool to assess how the DRL strategy performs in real-world scenarios. This process involves simulating the optimized portfolio's behavior over time using historical price data. The calculated portfolio weights are applied to this data to determine the portfolio's value and returns at each point in time. By analyzing these results, we can gauge the effectiveness of the DRL strategy in achieving the desired balance between risk and reward.

Each backtest commences with a virtual portfolio allocated entirely to cash. Daily portfolio values and returns are then calculated based on the specific portfolio weights generated by the trained DRL agent's strategy. This allows us to monitor the portfolio's performance over time. This testing phase involves accumulating and recording portfolio values, daily returns, and portfolio weights to provide a comprehensive evaluation of the agent's effectiveness.

Algorithm 6 Training DRL Agent with PPO for Portfolio Optimization

Initialize the policy network π_θ with random weights.

Input: Historical price data, lookback period T, training timesteps, number of environments n_envs, steps per environment n_steps, batch size, epochs n_epochs, discount factor γ, GAE lambda, clip range, learning rate.

for each sliding window group **do**

  Split data into training (5 years), validation (1 year), and testing (1 year) sets.

  for each training round **do**

    Initialize 5 agents with different seeds.

    for each agent **do**

      Train the agent using PPO:

- Collect trajectories by running policy π_θ.
- Compute advantages $\hat{A}_t$:
$$\hat{A}_t = \sum_{t'=t}^{T} (\gamma^{t'-t} R_{t'} - V_\phi(s_t))$$
- Update the policy by maximizing the clipped surrogate objective using gradient ascent:
$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t \left[\min\left(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right)\right]$$
- Update the value function:
$$V_\phi = \hat{\mathbb{E}}_t \left[(R_t - V_\phi(s_t))^2\right]$$

    end for

    Evaluate agents' performance on validation set.

    Choose the top-performing agent based on its average reward per episode during validation.

    Save the best agent for the next window.

  end for

end for

Backtesting: Backtest the trained agent on the testing set. Initialize portfolio with all-cash. Calculate daily portfolio values, returns, and weights based on the agent's actions:

$$\text{new\_portfolio\_value} = \sum_{i=1}^{n} w_i \times (1 + R_{i,t}) \times \text{current\_portfolio\_value}$$

Log Results: Store daily portfolio values, returns, and portfolio weights.

Chapter 4

Experimental Analysis

I Preliminary Metrics

Below is the detailed summary of portfolio performance metrics with their mathematical formulas and explanations, presented in a continuous narrative format: Cumulative Return measures the total percentage change in portfolio value over the investment period, indicating overall gain or loss. It is calculated as:

$$CR = \left(\frac{V_{\text{end}}}{V_{\text{start}}}\right) - 1$$

Geometric Mean Return reflects the average compounded return per period, accounting for compounding effects over multiple periods. It is calculated as:

$$GMR = (1 + CR)^{\frac{1}{T}} - 1$$

Annualized Return projects the geometric mean return to an annual scale, showing the expected yearly return. It is given by:

$$AR = (1 + GMR)^{252} - 1$$

Annual Volatility represents the standard deviation of daily returns annualized, indicating the risk or variability of returns over a year. It is computed as:

$$AV = \sigma_{\text{daily}} \times \sqrt{252}$$

Sharpe Ratio measures the risk-adjusted return by comparing the portfolio's excess return over the risk-free rate to its volatility. The formula is:

$$SR = \frac{\mu_{\text{daily}}}{\sigma_{\text{daily}}} \times \sqrt{252}$$

Calmar Ratio evaluates return per unit of drawdown risk, indicating how well the portfolio compensates for its largest observed loss. It is defined as:

$$CR = \frac{AR}{|\text{Max Drawdown}|}$$

Stability assesses the consistency of portfolio returns, with lower values indicating more stable performance. It is calculated as:

$$St = \frac{\sigma_{\text{value}}}{\mu_{\text{value}}}$$

Max Drawdown captures the maximum observed loss from a peak to a trough, showing the worst-case downside risk. The calculation is:

$$MD = \frac{\max(V_{\text{cummax}} - V)}{V_{\text{cummax}}}$$

Omega Ratio compares the sum of positive returns to the absolute sum of negative returns, reflecting the likelihood of gains versus losses. It is expressed as:

$$OR = \frac{\sum_{R>0} R}{|\sum_{R<0} R|}$$

Sortino Ratio is similar to the Sharpe ratio but focuses on downside risk, measuring risk-adjusted return relative to negative volatility. It is given by:

$$Sortino = \frac{\mu_{\text{daily}}}{\sigma_{\text{downside}}} \times \sqrt{252}$$

Skew indicates the asymmetry of return distribution. Positive skew suggests more frequent small gains and fewer large losses. It is measured as:

$$Skew = \text{skew}(R)$$

Kurtosis measures the "tailedness" of the return distribution. High kurtosis indicates frequent extreme returns. It is calculated as:

$$Kurt = \text{kurtosis}(R)$$

Tail Ratio compares the magnitude of extreme positive returns to extreme negative returns, providing insight into tail risks. The formula is:

$$TR = \frac{\sum_{R>Q_{0.95}} R}{|\sum_{R<Q_{0.05}} R|}$$

Value at Risk (VaR) estimates the maximum loss expected with a 95

$$VaR = Q_{0.05}$$

Delta Portfolio Weights Average measures the average daily change in portfolio weights, indicating the frequency and magnitude of portfolio rebalancing. It reflects the stability and transaction frequency of the portfolio management strategy. It is calculated as:

$$\Delta p_w = \sum_{t=1}^{T} \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}|$$

II U.S. Market Analysis

For the U.S. market, historical price data sourced from the S&P 500 sector indices, the VIX index, and the S&P 500 index covers the period from 2006 to 2023. This data, extracted from Yahoo Finance, includes daily adjusted close prices. The S&P 500 sector indices used in the analysis are: Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financials (XLF), Technology (XLK), Utilities (XLU), and Energy (XLE). This comprehensive dataset enables a robust implementation of the Mean-Variance Optimization (MVO) strategy, ensuring accurate risk-return assessments and optimal portfolio construction.

The initial portfolio started with a full-cash position of $100,000, providing a basis for comparing the performance of different portfolio strategies. This approach allows for a clear assessment of how each strategy grows or preserves the initial investment over time.

Classical Strategies

The traditional Mean-Variance Optimization (MVO) approach shows mixed performance and is particularly vulnerable in volatile years. For example, in 2020, the MVO approach had a cumulative return of -0.1638 and a maximum drawdown of 0.4099, indicating significant sensitivity to market shocks. Both the PPO and Black-Litterman models outperformed the MVO approach in such volatile conditions.

The Black-Litterman model integrates investor views, achieving a cumulative return of 0.1502 and an annual return of 0.1496. Although the returns of this model are competitive, it has a higher volatility of 0.1819 compared to PPO strategies. The Sharpe ratio of 0.9670 and the Calmar ratio of 4.4815 indicate balanced risk management, albeit with higher volatility than PPO strategies. Notably, the Black-Litterman model still outperforms the S&P 500 and traditional MVO methods in terms of cumulative and annual returns.

In contrast, the Black-Litterman model performed strongly in years with significant market movements, such as in 2023, achieving a cumulative return of 0.4540 and an annual return of 0.4496, with a Sharpe ratio of 2.2122. This demonstrates the model's resilience and the value of incorporating investor views for balanced portfolio construction, outperforming both the S&P 500 and MVO strategies.

PPO Strategies

The PPO Lookback 60 strategy has shown notable outperformance across various measures from 2012 to 2023. With an average cumulative return of 0.1685 and an annual return of 0.1679, it significantly outperforms the S&P 500, which had a cumulative return of 0.1250 and an annual return of 0.1252. These returns indicate that the PPO Lookback 60 strategy has been highly effective in generating significant long-term and annual growth, outperforming the S&P 500, the Black-Litterman model and the MVO approach.

The Sharpe Ratio of 1.1530 for the PPO Lookback 60 strategy indicates a robust risk-adjusted return, suggesting that the strategy delivers substantial returns relative to the risk taken, again outperforming the S&P 500, the Black-Litterman model and traditional MVO approaches. This is further underscored by the Calmar ratio of 5.7894, highlighting effective drawdown management. A maximum drawdown of 0.1138 indicates limited losses during market downturns, highlighting the strategy's resilience and effective risk management compared to the S&P 500, the Black-Litterman model and MVO.

In comparison, the PPO Lookback 10 strategy, despite having a higher Sharpe ratio of 1.2449 and a Calmar ratio of 5.6133, indicates more frequent portfolio rebalancing with a Δp_w value of 0.1713. While this results in superior risk-adjusted performance, it also implies higher transaction costs due to more frequent rebalancing. Nevertheless, this strategy still outperforms the S&P 500, the Black-Litterman model, and traditional MVO approaches.

Overall, PPO strategies with 60- and 120-day look-back periods tend to offer better stability and risk management. These strategies consistently outperform the S&P 500, the Black-Litterman model and traditional MVO strategies, with higher cumulative returns and superior drawdown management.

PPO NoVIX Strategies

Excluding VIX data in the PPO NoVIX Lookback 60 strategy resulted in even higher returns, with a cumulative return of 0.1771 and an annual return of 0.1770. This indicates an improvement over the corresponding PPO strategy with VIX data, suggesting that excluding VIX can improve performance. The cumulative and annual returns of the PPO NoVIX Lookback 60 strategy outperform both the S&P 500 and traditional MVO methodologies, as well as the Black-Litterman model.

The Sharpe ratio for the PPO NoVIX Lookback 60 strategy was 1.2512, reflecting an excellent risk-adjusted return. The maximum drawdown of 0.1134 also demonstrates effective drawdown management. Overall, excluding VIX data results in higher cumulative and annual returns while maintaining strong risk-adjusted performance and effective drawdown control, outperforming the S&P 500, the Black-Litterman model and the MVO strategies.

PPO Reward to Return Strategies

An analysis of various metrics from 2012 through 2023 shows that the PPO Return-LB60 strategy achieved a cumulative return of 0.1177 and an annual return of 0.1177, which, while lower than the PPO Lookback strategies, still exceeded the cumulative return of 0.0616 for the traditional MVO approach. The PPO NoVIX Return-LB60 strategy further improved its performance with a cumulative return of 0.1220 and an annual return of 0.1220, outperforming the S&P 500's cumulative return of 0.1250 and the Black-Litterman model's 0.1502.

The Sharpe ratios for PPO Return-LB60 and PPO NoVIX Return-LB60 were 0.9733 and 0.9855, respectively, competitive with the Black-Litterman model's Sharpe ratio of 0.9670 and significantly better than the MVO approach's 0.6044. The Calmar ratios for these strategies also indicate effective drawdown management, with values of 4.6530 for PPO Return-LB60 and 4.5512 for PPO NoVIX Return-LB60, outperforming the MVO's 2.8817 and closely matching the S&P 500's 5.1759.

In addition, the maximum drawdowns for PPO Return-LB60 and PPO NoVIX Return-LB60 were 0.1169 and 0.1168, respectively, lower than the MVO's 0.1379 and the Black-Litterman model's 0.1320, indicating better risk management. The Sortino ratios of 1.3924 and 1.3820 for PPO Return-LB60 and

PPO NoVIX Return-LB60 further highlight their superior downside risk management compared to the MVO approach's 0.9306 and the Black-Litterman model's 1.4188.

The final portfolio values for the PPO strategies show significant growth. For example, the PPO Lookback 60 strategy ended with a final portfolio value of $601,915.35, significantly higher than the S&P 500's $373,500.85, the Black-Litterman model's $498,448.37, and the MVO approach's $194,396.62. This reflects the superior growth and compounding effect of the PPO strategies over the period analyzed.

Significant Findings

The analysis reveals common trends, such as higher Sharpe ratios in years of low volatility. For example, in 2017, the S&P 500 had an annual volatility of 0.0669 and a Sharpe ratio of 2.6994. Conversely, higher volatility years, such as 2020, generally corresponded to lower performance efficiency across all strategies, with lower Sharpe ratios and higher drawdowns. Notably, the PPO and Black-Litterman strategies consistently outperformed the SP 500 and traditional MVO approaches in terms of risk-adjusted returns and drawdown management.

Within the PPO strategies, the PPO Lookback 10 strategy often had higher Sharpe ratios but also higher Δp_w values, indicating more frequent portfolio rebalancing and potentially higher transaction costs. The PPO Lookback 60 and 120 strategies provided a balance between return and risk management, with the PPO Lookback 60 showing an average Calmar ratio of 5.7894 from 2012 to 2023, indicating effective drawdown management and outperforming both the S&P 500 and MVO approaches.

Overall, the Black-Litterman model has consistently outperformed MVO in terms of cumulative and annual returns, particularly during periods of significant market movement. This underscores the benefit of integrating investor views and market equilibrium considerations into portfolio construction, which consistently outperformed the S&P 500 and MVO strategies.

Conclusion

The detailed analysis of U.S. market metrics shows that PPO strategies generally outperform traditional approaches such as MVO and the S&P 500. The Black-Litterman model provides a competitive advantage by effectively incorporating investor insights and demonstrating resilience under varying market conditions. The adaptive and dynamic nature of these strategies is critical to achieving optimal portfolio performance, consistently outperforming the SP 500 and traditional MVO strategies. Extensive tables in the appendix provide detailed annual metrics to support these findings.

Table 4.1: Metrics Average (2012-2023) for PPO Strategies

Metric	PPO Lookback 10	PPO Lookback 30	PPO Lookback 60	PPO Lookback 120	S&P 500
Cum. Return	0.1574	0.1390	0.1685	0.1474	0.1250
Geo. Mean Return	0.0005	0.0005	0.0006	0.0005	0.0004
Ann. Return	0.1557	0.1374	0.1679	0.1468	0.1252
Ann. Volatility	0.1608	0.1570	0.1623	0.1569	0.1539
Sharpe Ratio	1.2449	1.1598	1.1530	1.1557	1.0912
Calmar Ratio	5.6133	5.4804	5.7894	5.7218	5.1759
Stability	0.0569	0.0514	0.0580	0.0558	0.0529
Max Drawdown	0.1227	0.1208	0.1138	0.1190	0.1182
Omega Ratio	1.2585	1.2376	1.2274	1.2294	1.2273
Sortino Ratio	1.7094	1.6646	1.6831	1.6743	1.5038
Skew	-0.2904	-0.2359	-0.2339	-0.2075	-0.2928
Kurtosis	2.0266	2.1241	1.8585	2.0044	2.1848
Tail Ratio	0.9520	0.9612	0.9652	0.9731	0.9345
VaR	-0.0158	-0.0158	-0.0164	-0.0157	-0.0158
Δp_w	0.1713	0.1418	0.0779	0.0378	-
Final Portfolio Value	527,869.83	360,901.10	601,915.35	488,799.34	373,500.85

Table 4.2: Metrics Average (2012-2023) for PPO NoVIX Strategies

Metric	PPO NoVIX LB 10	PPO NoVIX LB 30	PPO NoVIX LB 60	PPO NoVIX LB 120	S&P 500
Cum. Return	0.1560	0.1349	0.1771	0.1340	0.1250
Geo. Mean Return	0.0005	0.0005	0.0006	0.0005	0.0004
Ann. Return	0.1579	0.1343	0.1770	0.1340	0.1252
Ann. Volatility	0.1581	0.1590	0.1576	0.1568	0.1539
Sharpe Ratio	1.2657	1.0774	1.2512	1.1210	1.0912
Calmar Ratio	6.1189	5.1306	5.5682	5.6153	5.1759
Stability	0.0549	0.0525	0.0559	0.0546	0.0529
Max Drawdown	0.1216	0.1148	0.1134	0.1220	0.1182
Omega Ratio	1.2569	1.2123	1.2544	1.2306	1.2273
Sortino Ratio	1.7875	1.5503	1.7643	1.6244	1.5038
Skew	-0.3418	-0.2403	-0.3599	-0.2373	-0.2928
Kurtosis	1.8963	1.8966	2.1217	1.8137	2.1848
Tail Ratio	0.9299	0.9537	0.9441	0.9701	0.9345
VaR	-0.0156	-0.0158	-0.0152	-0.0157	-0.0158
Δp_w	0.1303	0.0669	0.0836	0.0831	-
Final Portfolio Value	531,738.84	412,158.62	646,796.76	403,697.03	373,500.85

Table 4.3: Metrics Average (2012-2023) for Different Strategies

Metric	PPO Return-LB60	PPO NoVIX Return-LB60	Mean-Variance Optimization	BLack-Litterman	S&P 500
Cum. Return	0.1177	0.1220	0.0616	0.1502	0.1250
Geo. Mean Return	0.0004	0.0004	0.0002	0.0005	0.0004
Ann. Return	0.1177	0.1220	0.0612	0.1496	0.1252
Ann. Volatility	0.1595	0.1564	0.1773	0.1819	0.1539
Sharpe Ratio	0.9733	0.9855	0.6044	0.9670	1.0912
Calmar Ratio	4.6530	4.5512	2.8817	4.4815	5.1759
Stability	0.0554	0.0492	0.0527	0.0590	0.0529
Max Drawdown	0.1169	0.1168	0.1379	0.1320	0.1182
Omega Ratio	1.2031	1.1969	1.1144	1.1874	1.2273
Sortino Ratio	1.3924	1.3820	0.9306	1.4188	1.5038
Skew	-0.2571	-0.2840	-0.2442	-0.2626	-0.2928
Kurtosis	1.7708	1.9893	2.6612	2.8357	2.1848
Tail Ratio	0.9628	0.9328	0.9362	0.9644	0.9345
VaR	-0.0161	-0.0160	-0.0175	-0.0180	-0.0158
Δp_w	0.0394	0.0282	0.2990	0.3743	-
Final Portfolio Value	326,126.20	380,771.82	194,396.62	498,448.37	373,500.85

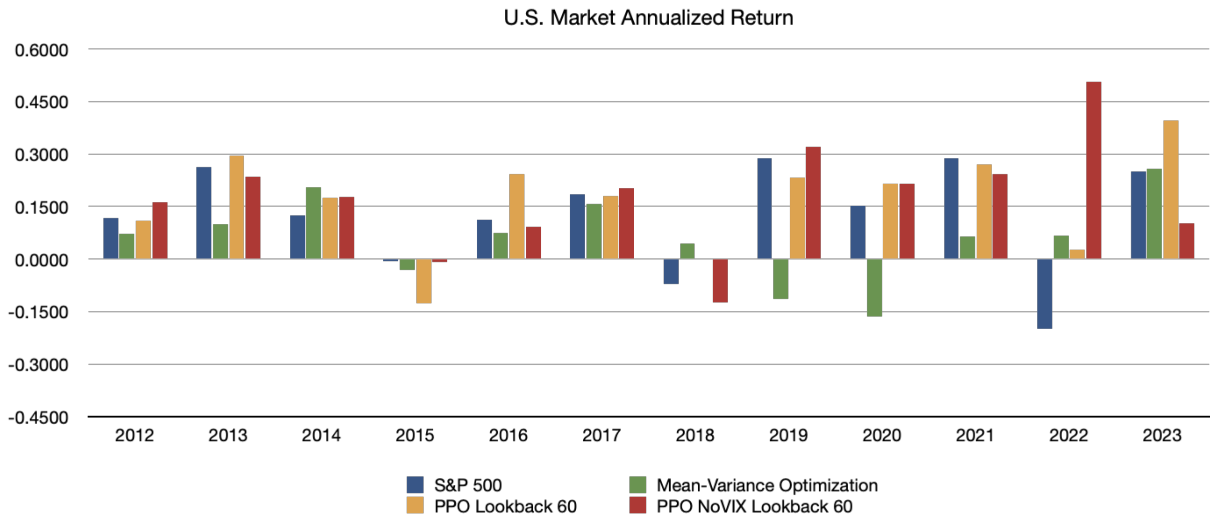

Figure 1: U.S. Market Annualized Return

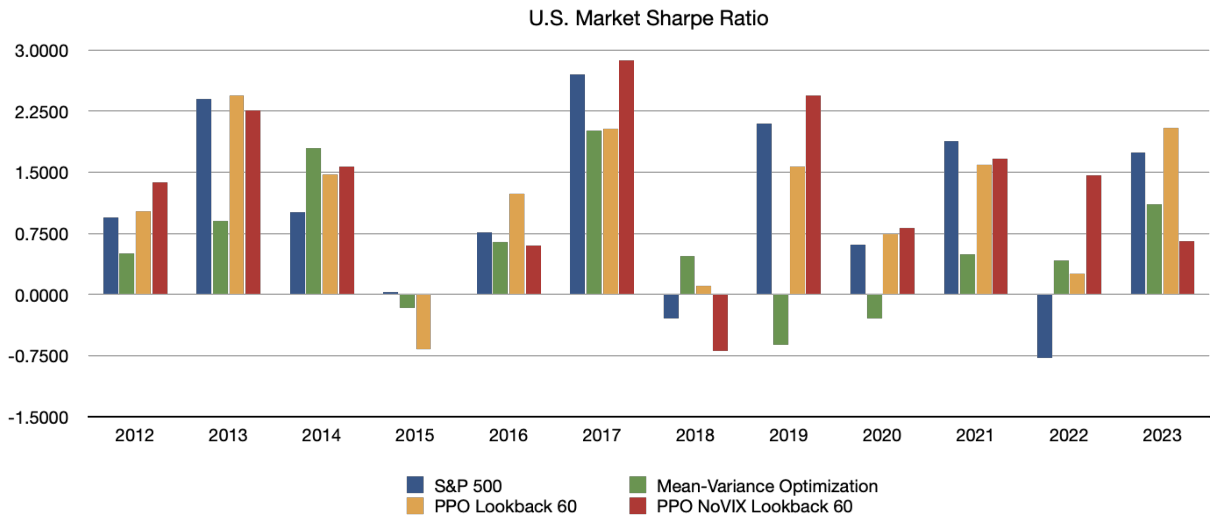

Figure 2: U.S. Market Sharpe Ratio

III South Korean Market Analysis

For the South Korean market, a portfolio was constructed using representative stocks from various sectors. The selected stocks are Samsung Electronics Co., Ltd. (005930.KS) for manufacturing, SK hynix Inc. (000660.KS) for electrical and electronics, KB Financial Group Inc. (105560.KS) for financials, Hyundai Motor Company (005380.KS) for transportation, NAVER Corporation (035420.KS) for services, LG Chem, Ltd. (051910.KS) for chemicals, Yuhan Corporation (000100.KS) for pharmaceuticals, Posco International Corporation (047050.KS) for distribution, HANMI Semiconductor Co., Ltd. (042700.KS) for machinery, POSCO Holdings Inc. (005490.KS) for steel and metals, Samsung Fire & Marine Insurance Co., Ltd. (000810.KS) for insurance, HMM Co., Ltd. (011200.KS) for transportation and warehouse, CJ Corporation (001040.KS) for food and beverages, SK Telecom Co., Ltd. (017670.KS) for telecommunications, and Mirae Asset Securities Co., Ltd. (006800.KS) for securities.

The initial portfolio starts with a full-cash position of 100,000,000 KRW, which serves as a baseline for evaluating the performance of different portfolio strategies. This setup allows for a clear comparison of how effectively each strategy grows or preserves the initial investment over time.

Classical Strategies

The Mean-Variance Optimization MVO strategy had a cumulative return of 0.1023, indicating a moderate total return over the period. The annual return of 0.0996 reflects a compound annual growth rate of nearly 10%, which is considered reasonable in traditional finance. However, the strategy's annual volatility of 0.2571 indicates a relatively high level of risk, suggesting that returns are subject to significant volatility. The Sharpe Ratio of 0.3021, although positive, shows only moderate risk-adjusted returns, indicating the strategy's limited ability to generate returns relative to its risk. With a Calmar ratio of 2.3498, this strategy efficiently manages risk relative to its returns, as a Calmar ratio above 2 is generally favorable. Its stability score of 0.1007 and maximum drawdown of 0.2261 further highlight its performance consistency and risk tolerance. Other metrics such as Omega Ratio 1.0760, Sortino Ratio 0.5659 and VaR -0.0235 support the conclusion that the MVO is effective but carries significant risk.

The Black-Litterman strategy had the lowest performance among the strategies analyzed, with a cumulative return of 0.0597 and an annual return of 0.0583, indicating modest total and annual growth. The annual volatility of 0.2612 is similar to that of the MVO strategy, indicating high risk. The Sharpe ratio of 0.1536 is relatively low, indicating weaker risk-adjusted returns. Its Calmar ratio of 0.3463 is significantly lower than that of MVO, indicating less efficient risk management. Stability 0.1028 and maximum drawdown 0.2418 are comparable to MVO, but the strategy's omega ratio 1.0447 and Sortino ratio 0.2312 confirm its lower performance.

PPO Strategies

Among the PPO strategies, the PPO Lookback 30 strategy stands out with the highest cumulative return of 0.1784. This substantial cumulative return is complemented by an annual return of 0.1834, indicating robust annual growth. The annual volatility of 0.2150, although significant, is lower than that of the classic strategies, suggesting a better risk profile. With a Sharpe ratio of 0.7562, this strategy shows superior risk-adjusted returns compared to MVO and Black-Litterman. The Calmar ratio of 4.6106 shows exceptional performance in managing drawdowns, with a maximum drawdown of 0.1957. The stability of 0.0912 and an Omega ratio of 1.1598 indicate consistent and favorable performance. Other metrics such as Sortino Ratio 1.2678, Skewness 0.1971 and VaR -0.0209 further underscore its robust risk-adjusted performance.

PPO NoVKOSPI Strategies

The PPO NoVKOSPI 30 strategy performed particularly well, with a cumulative return of 0.1619 and an annual return of 0.1665. This strategy's annual volatility of 0.2036 is slightly lower than that of the PPO Lookback 30, indicating slightly lower risk. Its Sharpe ratio of 0.6238 reflects strong risk-adjusted returns, although slightly below the PPO Lookback 30. The Calmar ratio of 3.9708 shows solid drawdown management with a maximum drawdown of 0.1673, which is lower than the PPO Lookback 30. The Stability Ratio of 0.0767 and the Omega Ratio of 1.1270 further demonstrate the strategy's reliable performance. The Sortino ratio of 1.0661 coupled with skewness of -0.0960 and VaR of -0.0200 indicates that this strategy effectively balances risk and return.

The lower volatility and drawdown compared to PPO Lookback 30 suggest that the PPO NoVKOSPI 30 strategy manages risk more efficiently while still providing substantial returns. The Sharpe and Calmar ratios indicate a strong ability to generate returns on a risk-adjusted basis. Overall, the stability and Omega ratios confirm the consistency and reliability of this strategy, making it a robust choice for investors seeking a balanced approach between risk and return.

PPO Reward to Return Strategies

Looking at the PPO Return-LB60 and PPO NoVKOSPI Return-LB60 strategies, their performance metrics show a significant improvement over classical methods and the KOSPI index. The PPO Return-LB60 shows a cumulative return of 0.1391, which is modest but still competitive given its lower annual volatility of 0.2151. The annual return of 0.1434 and a Sharpe ratio of 0.5456 indicate better risk-adjusted returns compared to traditional strategies. The Calmar ratio of 2.5039 and the stability of 0.0803 indicate efficient risk management and consistent performance. The maximum drawdown of 0.1849 is comparable to PPO Lookback strategies, demonstrating robust resilience.

Similarly, the PPO NoVKOSPI Return-LB60 strategy has a cumulative return of 0.1058 and an annual return of 0.1099, closely tracking the MVO but with superior risk metrics. Its annualized volatility of 0.2162 and Sharpe ratio of 0.4328 indicate strong risk-adjusted performance. Its Calmar ratio of 2.0014, stability of 0.0759 and maximum drawdown of 0.1819 underscore effective risk management. The Omega ratio of 1.0986 and the Sortino ratio of 0.6340 further emphasize the strategy's favorable risk/reward balance.

Significant Findings

Despite the overall growth in the equity market, these strategies demonstrate that higher returns can be achieved through customized approaches such as PPO Lookback and NoVKOSPI strategies. These strategies show superior performance in cumulative returns, Sharpe ratios, and Calmar ratios, indicating robust risk-adjusted returns and resilience during market drawdowns. This finding is critical because it demonstrates that advanced portfolio optimization techniques can generate higher returns than simply following market trends. In particular, the PPO strategies show that customized approaches outperform classical strategies and the KOSPI Index in terms of both returns and risk management.

In the context of the South Korean market, several key findings stand out. The PPO strategies, particularly the Lookback and NoVKOSPI variants, consistently outperform traditional methods such as Mean-Variance Optimization and the Black-Litterman model. This shows that dynamic and adaptive strategies that take into account recent market trends and selectively exclude volatile components can better capture upside potential and manage risk. In addition, the lower volatility and lower drawdowns of PPO strategies indicate superior risk management, providing more stable returns even during market downturns. These findings underscore the value of advanced, customized portfolio optimization techniques in delivering better risk-adjusted returns and managing drawdowns more effectively than traditional methods.

Conclusion

Analysis of historical data for the South Korean market shows that while the KOSPI may not be a market for high returns, customized approaches such as PPO lookback and NoVKOSPI strategies can deliver higher returns. The detailed analysis shows that advanced portfolio optimization techniques, such as PPO strategies, deliver better risk-adjusted returns and effectively manage drawdowns. The Black-Litterman model, while complex, offers competitive performance by incorporating investor views, making it a valuable strategy for optimizing portfolio allocations. The data underscores the importance of dynamic asset allocation and risk management in achieving optimal portfolio performance under varying market conditions.

Table 4.4: Metrics Average (2012-2023) for PPO Strategies

Metric	PPO Lookback 10	PPO Lookback 30	PPO Lookback 60	PPO Lookback 120	KOSPI
Cum. Return	0.1658	0.1784	0.1391	0.1714	0.0450
Geo. Mean Return	0.0006	0.0006	0.0004	0.0006	0.0001
Ann. Return	0.1751	0.1834	0.1434	0.1786	0.0463
Ann. Volatility	0.2004	0.2150	0.2151	0.2000	0.1483
Sharpe Ratio	0.7558	0.7562	0.5456	0.7391	0.3072
Calmar Ratio	4.0147	4.6106	2.5039	3.8778	1.5237
Stability	0.0715	0.0912	0.0803	0.0838	0.0493
Max Drawdown	0.1601	0.1957	0.1849	0.1712	0.1533
Omega Ratio	1.1501	1.1598	1.1167	1.1472	1.0712
Sortino Ratio	1.2611	1.2678	0.8626	1.1846	0.4452
Skew	0.1440	0.1971	0.2125	0.0750	-0.1208
Kurtosis	1.7920	2.1064	3.6888	1.7547	1.8812
Tail Ratio	1.1096	1.1062	1.0644	1.0951	0.9385
VaR	-0.0190	-0.0209	-0.0209	-0.0190	-0.0154
Δp_w	0.1377	0.1013	0.0550	0.0190	-
Final Portfolio Value	556,593,359	505,958,662	372,696,087	497,472,699	135,827,664

Table 4.5: Metrics Average (2012-2023) for PPO NoVKOSPI Strategies

Metric	PPO NoVKOSPI LB10	PPO NoVKOSPI LB30	PPO NoVKOSPI LB60	PPO NoVKOSPI LB120	KOSPI
Cum. Return	0.1437	0.1619	0.1058	0.1248	0.0450
Geo. Mean Return	0.0005	0.0005	0.0003	0.0004	0.0001
Ann. Return	0.1538	0.1665	0.1099	0.1309	0.0463
Ann. Volatility	0.2285	0.2036	0.2162	0.2202	0.1483
Sharpe Ratio	0.5615	0.6238	0.4328	0.5608	0.3072
Calmar Ratio	3.8615	3.9708	2.0014	3.1718	1.5237
Stability	0.0838	0.0767	0.0759	0.0799	0.0493
Max Drawdown	0.1964	0.1673	0.1819	0.1985	0.1533
Omega Ratio	1.1224	1.1270	1.0986	1.1144	1.0712
Sortino Ratio	1.0463	1.0661	0.6340	0.9534	0.4452
Skew	0.4438	-0.0960	0.0222	0.0036	-0.1208
Kurtosis	3.3543	1.7090	1.3813	2.1140	1.8812
Tail Ratio	1.1722	1.0228	1.0245	1.0636	0.9385
VaR	-0.0216	-0.0200	-0.0217	-0.0209	-0.0154
Δp_w	0.4808	0.4400	0.6200	1.1338	-
Final Portfolio Value	375,437,413	408,909,190	240,076,052	288,281,059	135,827,664

Table 4.6: Metrics Average (2012-2023) for Different Strategies

Metric	PPO Return-LB60	PPO NoVKOSPI Return-LB60	Mean-Variance Optimization	Black-Litterman	KOSPI
Cum. Return	0.0667	0.0974	0.1023	0.0597	0.0450
Geo. Mean Return	0.0002	0.0003	0.0003	0.0001	0.0001
Ann. Return	0.0690	0.1002	0.0996	0.0583	0.0463
Ann. Volatility	0.2018	0.2095	0.2571	0.2612	0.1483
Sharpe Ratio	0.4141	0.4644	0.3021	0.1536	0.3072
Calmar Ratio	1.9443	2.1512	2.3498	0.3463	1.5237
Stability	0.0655	0.0650	0.1007	0.1028	0.0493
Max Drawdown	0.1939	0.1889	0.2261	0.2418	0.1533
Omega Ratio	1.0807	1.0929	1.0760	1.0447	1.0712
Sortino Ratio	0.6610	0.7190	0.5659	0.2312	0.4452
Skew	-0.0473	0.0879	-0.0624	0.0517	-0.1208
Kurtosis	1.1066	2.5789	4.7986	3.1079	1.8812
Tail Ratio	1.0161	1.0435	1.0768	1.0316	0.9385
VaR	-0.0205	-0.0204	-0.0235	-0.0245	-0.0154
Δp_w	0.1392	1.4255	0.2723	0.2945	-
Final Portfolio Value	192,721,477	255,788,422	197,822,628	122,265,282.4	135,827,664

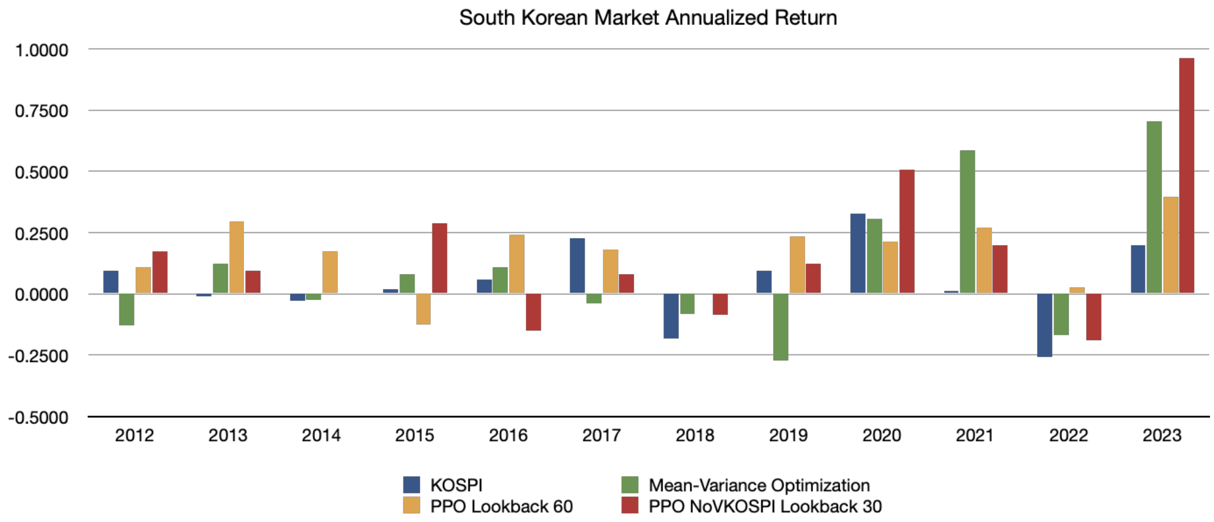

Figure 3: South Korean Market Annualized Return

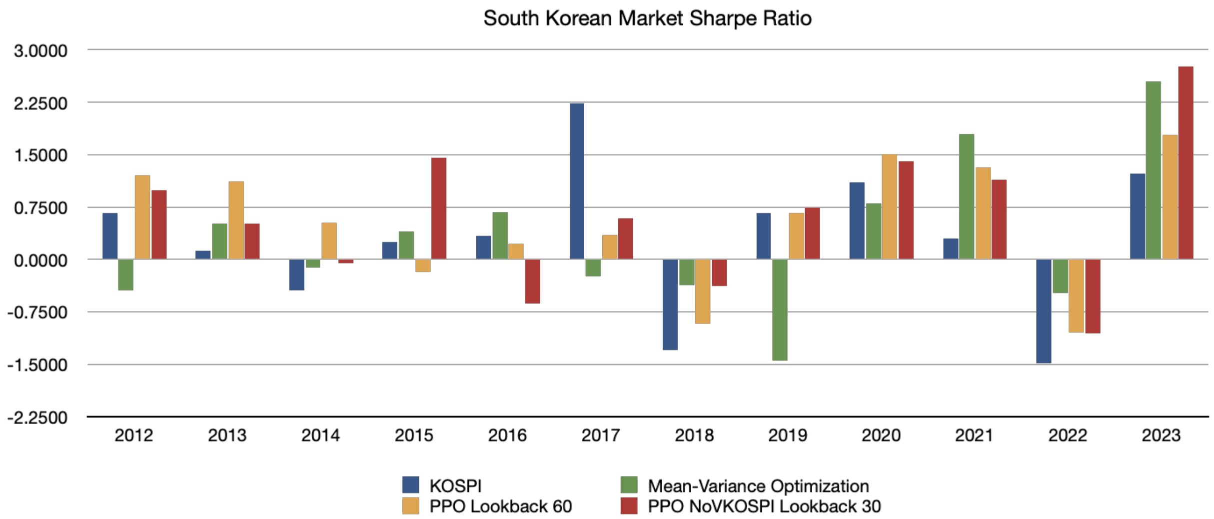

Figure 4: South Korean Market Sharpe Ratio

Chapter 5

Conclusion

The comparative analysis highlights the superior effectiveness of Deep Reinforcement Learning (DRL) in portfolio optimization, consistently outperforming the traditional Mean-Variance Optimization (MVO) approach across several performance metrics, including cumulative return, annualized return, volatility, and Sharpe ratio. This exceptional performance underscores the potential of DRL as a dynamic and adaptive tool for robust portfolio management.

Despite the challenges of identifying specific trends or patterns across different variables, DRL methods have generally shown strong performance. The impact of these variables can vary significantly depending on the specific market conditions or asset classes being considered. Therefore, it is critical to tailor the parameters and approaches for DRL-based portfolio optimization to the unique characteristics of the market or assets under consideration. This requires thorough testing and validation to identify the most effective strategies and parameters for each unique scenario.

Analysis of different market strategies provides detailed insights into investment performance under different models and conditions. In particular, PPO strategies demonstrated overall adaptability, effectively balancing responsiveness and stability. Specifically, PPO with a 60-day or 120-day look-back period, combined with volatility indices, optimized the risk/return trade-off, as indicated by low Δp_w values. This suggests that PPO can be an effective asset allocation method in real markets where transaction costs are present.

Future work will focus on several key areas to improve the robustness and applicability of DRL in portfolio optimization. One important direction is to incorporate transaction costs into the model, a critical factor in real-world trading that can significantly affect net returns. In addition, the investigation of regime switching models that can adapt to different market conditions and regimes will be a priority. These models can potentially improve the resilience and performance of DRL-based strategies in different market environments, further establishing DRL as a powerful tool in financial decision making.

Overall, the results indicate that DRL, particularly PPO strategies, is a promising approach to portfolio optimization that can generate superior returns and effectively manage risk. This conclusion underscores the potential of DRL to revolutionize dynamic and adaptive portfolio management.

Appendix

I Detailed U.S. Market Metrics

Table 5.1: S&P 500 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1168	0.2639	0.1239	-0.0069	0.1124	0.1842	-0.0701	0.2871	0.1529	0.2879	-0.1995	0.2473
Geometric Mean Return	0.0004	0.0009	0.0005	-0.0000	0.0004	0.0007	-0.0003	0.0010	0.0006	0.0010	-0.0009	0.0009
Annualized Return	0.1178	0.2639	0.1239	-0.0069	0.1124	0.1849	-0.0704	0.2871	0.1523	0.2879	-0.2002	0.2495
Annual Volatility	0.1270	0.1107	0.1137	0.1549	0.1310	0.0669	0.1705	0.1247	0.3443	0.1310	0.2417	0.1309
Sharpe Ratio	0.9433	2.3989	1.0056	0.0302	0.7610	2.6994	-0.2939	2.0974	0.6091	1.8843	-0.7771	1.7363
Calmar Ratio	4.7788	10.5518	5.4251	-0.1758	3.1285	10.1742	-1.7170	9.6430	1.2707	11.2132	-4.6313	12.4494
Stability	0.0338	0.0604	0.0413	0.0266	0.0484	0.0447	0.0366	0.0517	0.0992	0.0673	0.0712	0.0538
Max Drawdown	0.0962	0.0520	0.0712	0.1235	0.0826	0.0249	0.1978	0.0622	0.3058	0.0493	0.2543	0.0986
Omega Ratio	1.1761	1.4841	1.1883	1.0052	1.1474	1.6259	0.9479	1.4480	1.1311	1.3678	0.8826	1.3238
Sortino Ratio	1.4453	3.4273	1.3169	0.0435	0.9944	3.6193	-0.3595	2.5887	0.6902	2.6734	-1.2419	2.8484
Skew	0.0697	-0.3328	-0.3954	-0.1686	-0.3789	-0.4425	-0.4254	-0.5842	-0.5535	-0.3393	0.0451	-0.0083
Kurtosis	0.9034	1.4185	1.3358	1.9266	2.3351	2.9354	3.1394	3.2925	8.0313	0.6981	0.3878	-0.1866
Tail Ratio	1.0914	0.9678	0.8666	0.9005	0.9321	0.9999	0.7688	0.8372	0.9351	0.9155	0.8842	1.1152
VaR	-0.0126	-0.0118	-0.0122	-0.0152	-0.0128	-0.0054	-0.0207	-0.0121	-0.0336	-0.0131	-0.0264	-0.0138

Table 5.2: PPO Lookback 10 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1047	0.2602	0.2056	0.0929	0.1443	0.2199	-0.1479	0.3505	0.1997	0.3718	-0.1126	0.1995
Geometric Mean Return	0.0004	0.0009	0.0007	0.0004	0.0005	0.0008	-0.0006	0.0012	0.0007	0.0013	-0.0005	0.0007
Annualized Return	0.1025	0.2579	0.2038	0.0921	0.1431	0.2161	-0.1458	0.3473	0.1989	0.3683	-0.1113	0.1952
Annual Volatility	0.1416	0.1105	0.1150	0.1429	0.1334	0.0702	0.1856	0.1168	0.3824	0.1666	0.2224	0.1424
Sharpe Ratio	0.7906	2.3417	1.5943	0.6711	0.9963	2.9812	-0.7159	2.6401	0.7178	1.9189	-0.4228	1.4257
Calmar Ratio	3.4873	9.3509	8.8158	2.4337	5.0274	11.1684	-3.6172	11.3111	1.5726	11.8538	-2.5367	8.4922
Stability	0.0384	0.0600	0.0536	0.0264	0.0585	0.0489	0.0453	0.0631	0.1204	0.0791	0.0472	0.0421
Max Drawdown	0.1223	0.0586	0.0495	0.1015	0.0830	0.0321	0.2447	0.0678	0.3252	0.0668	0.1990	0.1218
Omega Ratio	1.1420	1.4755	1.3110	1.1191	1.1890	1.6941	0.8819	1.5814	1.1535	1.3662	0.9328	1.2553
Sortino Ratio	1.2179	3.1920	2.1173	0.9864	1.3646	3.8906	-0.9361	3.3026	0.8237	2.9185	-0.6619	2.2977
Skew	-0.0165	-0.5568	-0.3974	-0.2303	-0.2451	-0.5301	-0.1257	-0.5615	-0.5747	-0.1758	0.0212	-0.0924
Kurtosis	0.8166	1.9188	1.2829	1.5281	1.4401	2.7749	3.2384	2.9637	7.1824	0.4802	0.7491	-0.0560
Tail Ratio	1.0554	0.9396	0.9044	0.9543	0.9573	0.9978	0.8005	0.9014	0.9297	1.0240	0.9210	1.0385
VaR	-0.0142	-0.0098	-0.0122	-0.0147	-0.0138	-0.0057	-0.0205	-0.0101	-0.0358	-0.0164	-0.0218	-0.0147
Δp_w	0.1460	0.1545	0.1474	0.1633	0.3535	0.1580	0.1413	0.1316	0.2017	0.1746	0.1440	0.1395

Table 5.3: PPO Lookback 30 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.2180	0.3254	0.0925	0.0082	0.1677	0.2115	-0.2067	0.2033	0.1275	0.2734	0.0433	0.2040
Geometric Mean Return	0.0008	0.0011	0.0004	0.0000	0.0006	0.0008	-0.0009	0.0007	0.0005	0.0010	0.0002	0.0007
Annualized Return	0.2142	0.3210	0.0921	0.0082	0.1663	0.2087	-0.2038	0.1998	0.1275	0.2722	0.0427	0.2004
Annual Volatility	0.1481	0.1051	0.1192	0.1573	0.1435	0.0689	0.1857	0.1190	0.3676	0.1272	0.2061	0.1363
Sharpe Ratio	1.5570	2.7093	0.7025	0.1186	1.0292	2.9418	-1.0832	1.8480	0.5317	1.8500	0.3255	1.3869
Calmar Ratio	7.5077	13.8313	4.0419	0.2049	4.7689	12.4625	-5.6951	6.8307	0.9694	10.5071	1.0928	9.2427
Stability	0.0480	0.0742	0.0353	0.0266	0.0615	0.0468	0.0466	0.0356	0.0974	0.0631	0.0430	0.0384
Max Drawdown	0.1103	0.0474	0.0886	0.1225	0.0680	0.0207	0.2898	0.0721	0.3243	0.0597	0.1497	0.0964
Omega Ratio	1.2879	1.5594	1.1261	1.0205	1.1962	1.6615	0.8275	1.3833	1.1177	1.3688	1.0544	1.2477
Sortino Ratio	2.5657	3.8318	0.9411	0.1684	1.3875	4.0301	-1.4697	2.3277	0.6020	2.7236	0.5103	2.3564
Skew	0.0192	-0.4145	-0.3750	-0.2006	-0.3335	-0.4414	0.0783	-0.6026	-0.4443	-0.0658	-0.0765	0.0263
Kurtosis	0.4381	1.0077	1.0332	1.8338	1.7232	2.0643	3.2409	2.8845	9.6885	0.9657	0.4809	0.1288
Tail Ratio	1.1188	0.9400	0.8830	0.8892	0.9371	0.9971	0.8070	0.8667	0.9372	1.0897	0.9881	1.0804
VaR	-0.0132	-0.0108	-0.0139	-0.0167	-0.0153	-0.0053	-0.0206	-0.0113	-0.0334	-0.0139	-0.0214	-0.0136
Δp_w	0.1421	0.1254	0.0726	0.1673	0.1714	0.0940	0.1838	0.1693	0.2194	0.1078	0.0718	0.1772

Table 5.4: PPO Lookback 60 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.1111	0.2950	0.1748	-0.1262	0.2405	0.1804	0.0007	0.2332	0.2133	0.2693	0.0277	0.4023
Geometric Mean Return	0.0004	0.0010	0.0006	-0.0005	0.0009	0.0007	0.0000	0.0008	0.0008	0.0009	0.0001	0.0013
Annualized Return	0.1092	0.2950	0.1748	-0.1262	0.2416	0.1796	0.0007	0.2332	0.2142	0.2705	0.0277	0.3948
Annual Volatility	0.1236	0.1187	0.1075	0.1761	0.1786	0.0868	0.1293	0.1458	0.3541	0.1469	0.2101	0.1704
Sharpe Ratio	1.0204	2.4420	1.4719	-0.6757	1.2399	2.0331	0.1022	1.5692	0.7372	1.5936	0.2586	2.0433
Calmar Ratio	4.7939	11.5497	7.7298	-2.8113	6.8062	8.0827	0.0208	7.4722	1.7675	9.2118	0.6338	14.2150
Stability	0.0305	0.0684	0.0511	0.0485	0.0796	0.0441	0.0366	0.0402	0.1163	0.0640	0.0407	0.0762
Max Drawdown	0.0878	0.0571	0.0532	0.1788	0.0885	0.0375	0.1243	0.0856	0.2872	0.0737	0.1592	0.1330
Omega Ratio	1.1935	1.4849	1.2846	0.8923	1.2383	1.4121	1.0177	1.3072	1.1595	1.3028	1.0440	1.3915
Sortino Ratio	1.5659	3.4998	1.9351	-0.9683	1.7901	3.3260	0.1312	2.0813	0.8339	2.3483	0.3978	3.2556
Skew	0.1162	-0.4262	-0.4318	-0.2424	-0.1253	0.0566	-0.5972	-0.4400	-0.5435	-0.1145	0.0186	-0.0774
Kurtosis	0.9105	1.0054	1.2878	1.5415	0.9206	1.4550	2.4841	2.1932	8.0166	1.4869	1.1424	-0.1425
Tail Ratio	1.1117	0.9333	0.9039	0.8666	0.9793	1.1934	0.7676	0.8413	0.9284	0.9909	0.9564	1.1098
VaR	-0.0120	-0.0115	-0.0117	-0.0181	-0.0215	-0.0065	-0.0133	-0.0147	-0.0334	-0.0148	-0.0221	-0.0176
Δp_w	0.0873	0.0701	0.0589	0.1193	0.1264	0.0665	0.0699	0.1079	0.0438	0.0692	0.0964	0.0193

Table 5.5: PPO Lookback 120 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.1331	0.2711	0.1686	0.0162	0.1821	0.2099	-0.0745	0.2638	0.1340	0.2260	-0.0802	0.3193
Geometric Mean Return	0.0005	0.0010	0.0006	0.0001	0.0007	0.0008	-0.0003	0.0009	0.0005	0.0008	-0.0003	0.0011
Annualized Return	0.1309	0.2711	0.1686	0.0162	0.1821	0.2090	-0.0739	0.2638	0.1346	0.2260	-0.0799	0.3135
Annual Volatility	0.1320	0.1189	0.1149	0.1598	0.1588	0.0978	0.1587	0.1102	0.3624	0.1070	0.2167	0.1457
Sharpe Ratio	1.0429	2.3054	1.3408	0.1828	1.0483	2.0909	-0.3564	2.1895	0.5430	1.8294	-0.2593	1.9112
Calmar Ratio	4.3338	10.9022	6.8791	0.4033	4.6475	9.7819	-1.8798	10.6618	1.0847	8.8113	-1.9211	14.9573
Stability	0.0346	0.0599	0.0495	0.0319	0.0703	0.0604	0.0352	0.0529	0.0935	0.0521	0.0579	0.0716
Max Drawdown	0.1019	0.0576	0.0573	0.1447	0.1016	0.0561	0.1894	0.0540	0.3285	0.0481	0.2004	0.0886
Omega Ratio	1.1928	1.4564	1.2539	1.0316	1.1992	1.4431	0.9381	1.4469	1.1193	1.3556	0.9591	1.3573
Sortino Ratio	1.5428	3.5977	1.7913	0.2684	1.4831	3.1529	-0.4536	2.8854	0.6278	2.4214	-0.4292	3.2031
Skew	-0.1270	-0.0735	-0.4221	-0.1409	-0.1842	-0.0131	-0.2400	-0.4765	-0.3055	-0.5716	0.0523	0.0127
Kurtosis	1.0167	1.0638	1.2988	1.5252	1.4775	1.7844	3.5378	2.2683	8.3246	1.4630	0.4010	-0.1086
Tail Ratio	1.0008	1.0684	0.8929	0.9440	0.9991	1.1180	0.7971	0.8513	0.9783	0.8585	0.9922	1.1766
VaR	-0.0136	-0.0110	-0.0110	-0.0160	-0.0166	-0.0077	-0.0201	-0.0116	-0.0310	-0.0116	-0.0237	-0.0143
Δp_w	0.0363	0.0317	0.0343	0.0296	0.0191	0.0480	0.0535	0.0219	0.0245	0.0117	0.0481	0.0951

Table 5.6: PPO NoVIX Lookback 10 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Annualized Return	0.1612	0.2668	0.1498	0.0449	0.2094	0.2355	-0.1684	0.3224	0.0051	0.4176	0.1119	0.1383
Geometric Mean Return	0.0006	0.0009	0.0006	0.0002	0.0008	0.0008	-0.0007	0.0011	0.0000	0.0014	0.0004	0.0005
Cumulative Return	0.1578	0.2644	0.1485	0.0445	0.2076	0.2313	-0.1660	0.3195	0.0051	0.4137	0.1105	0.1354
Annual Volatility	0.1102	0.1048	0.1201	0.1595	0.1328	0.0700	0.1880	0.1473	0.3513	0.1664	0.2023	0.1446
Sharpe Ratio	1.3698	2.5621	1.1452	0.3439	1.4104	3.1511	-0.8262	2.0023	0.2008	2.1287	0.6492	1.0507
Calmar Ratio	6.9770	10.3984	6.6345	1.0639	6.7756	13.0400	-3.8757	10.4897	0.0423	13.5294	2.5680	5.7842
Stability	0.0423	0.0607	0.0453	0.0285	0.0636	0.0546	0.0495	0.0568	0.0925	0.0838	0.0450	0.0367
Max Drawdown	0.0726	0.0508	0.0827	0.1195	0.0682	0.0183	0.2540	0.0746	0.3811	0.0589	0.1561	0.1228
Omega Ratio	1.2557	1.5311	1.2138	1.0600	1.2695	1.7262	0.8676	1.4163	1.0415	1.4007	1.1121	1.1886
Sortino Ratio	2.1742	3.6097	1.4815	0.4805	1.8867	4.2587	-1.0814	2.5139	0.2320	3.3139	0.9604	1.6203
Skew	0.0140	-0.3152	-0.5064	-0.3555	-0.4470	-0.4403	-0.2969	-0.5424	-0.6529	-0.2846	-0.2403	-0.0344
Kurtosis	0.7893	1.8056	1.3435	1.8647	1.4533	2.2644	2.2785	2.2273	7.2893	0.1831	0.7454	0.5109
Tail Ratio	1.0949	0.9705	0.8166	0.8666	0.9046	1.0516	0.7620	0.8490	0.9035	0.9667	0.9496	1.0230
VaR	-0.0112	-0.0102	-0.0132	-0.0162	-0.0139	-0.0052	-0.0214	-0.0149	-0.0292	-0.0159	-0.0208	-0.0151
Δp_w	0.0507	0.1579	0.1625	0.1254	0.0526	0.1152	0.1584	0.1922	0.1886	0.0437	0.1377	0.1786

Table 5.7: PPO NoVIX Lookback 30 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Annualized Return	0.1082	0.3667	0.1383	-0.0031	0.1707	0.1225	-0.0885	0.2536	0.1631	0.2225	-0.0701	0.2279
Geometric Mean Return	0.0004	0.0012	0.0005	-0.0000	0.0006	0.0005	-0.0004	0.0009	0.0006	0.0008	-0.0003	0.0008
Cumulative Return	0.1082	0.3701	0.1394	-0.0031	0.1714	0.1225	-0.0885	0.2558	0.1652	0.2244	-0.0701	0.2239
Annual Volatility	0.1222	0.1189	0.1194	0.1431	0.1591	0.0646	0.1768	0.1038	0.3565	0.1168	0.2589	0.1676
Sharpe Ratio	1.0033	2.8710	1.0605	0.0483	0.9583	1.9032	-0.3837	2.1491	0.5878	1.6615	-0.0848	1.1542
Calmar Ratio	4.2060	14.7135	5.4631	-0.0793	3.7387	9.3086	-2.1082	10.6664	1.4033	8.8356	-1.4869	6.9061
Stability	0.0335	0.0795	0.0441	0.0216	0.0697	0.0342	0.0415	0.0471	0.0988	0.0560	0.0482	0.0553
Max Drawdown	0.0862	0.0448	0.0861	0.1028	0.1051	0.0290	0.2241	0.0436	0.3187	0.0536	0.1917	0.0919
Omega Ratio	1.1794	1.6002	1.1950	1.0081	1.1815	1.3685	0.9342	1.4476	1.1283	1.3183	0.9865	1.2005
Sortino Ratio	1.6228	4.3083	1.4671	0.0717	1.2914	2.7604	-0.4846	2.7776	0.7023	2.2979	-0.1377	1.9263
Skew	0.0188	-0.2019	-0.2924	-0.2117	-0.3583	-0.2553	-0.3952	-0.4976	-0.1977	-0.4069	-0.0065	-0.0785
Kurtosis	0.8206	0.9939	1.3335	1.6044	2.3687	0.8022	3.0129	2.2910	7.8544	1.1037	0.3252	0.2484
Tail Ratio	1.0819	1.0567	0.8821	0.9356	0.9437	0.9634	0.7632	0.8566	0.9904	0.9189	0.9491	1.1027
VaR	-0.0116	-0.0111	-0.0126	-0.0147	-0.0173	-0.0071	-0.0206	-0.0100	-0.0292	-0.0125	-0.0270	-0.0161
Δp_w	0.1297	0.1013	0.1188	0.1442	0.0488	0.0218	0.0385	0.0375	0.0605	0.0405	0.0338	0.0270

Table 5.8: PPO NoVIX Lookback 60 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.1644	0.2356	0.1764	-0.0094	0.0905	0.2043	-0.1251	0.3213	0.2153	0.2424	0.5057	0.1043
Geometric Mean Return	0.0006	0.0008	0.0006	-0.0000	0.0003	0.0007	-0.0005	0.0011	0.0008	0.0009	0.0016	0.0004
Annualized Return	0.1616	0.2356	0.1764	-0.0094	0.0909	0.2034	-0.1242	0.3213	0.2162	0.2435	0.5057	0.1026
Annual Volatility	0.1281	0.1074	0.1005	0.1573	0.1326	0.0677	0.1679	0.1152	0.3016	0.1241	0.3406	0.1483
Sharpe Ratio	1.3789	2.2594	1.5737	-0.0002	0.5966	2.8660	-0.6951	2.4377	0.8157	1.6626	1.4614	0.6575
Calmar Ratio	6.1883	8.0968	8.8323	-0.2581	2.5076	11.7603	-2.9623	11.1810	2.0834	9.5300	6.2715	3.5880
Stability	0.0406	0.0517	0.0469	0.0269	0.0476	0.0507	0.0391	0.0620	0.1103	0.0581	0.1066	0.0297
Max Drawdown	0.0903	0.0539	0.0480	0.1280	0.0850	0.0283	0.2183	0.0490	0.2623	0.0529	0.2420	0.1024
Omega Ratio	1.2674	1.4539	1.3018	0.9999	1.1126	1.6260	0.8842	1.5350	1.1800	1.3203	1.2629	1.1091
Sortino Ratio	2.0931	3.1461	2.1673	-0.0002	0.7804	4.3992	-0.8364	2.9905	0.9042	2.3042	2.1404	1.0833
Skew	0.0344	-0.4907	-0.3370	-0.0611	-0.4172	-0.1170	-0.6720	-0.6082	-0.8054	-0.3521	-0.4413	-0.0507
Kurtosis	0.9457	2.2482	1.0430	1.0171	2.2624	1.9093	2.5515	3.4010	8.3752	1.1024	0.5988	0.0062
Tail Ratio	1.0875	0.9120	0.9594	0.9611	0.9101	1.1941	0.6800	0.8618	0.9253	0.9343	0.8763	1.0268
VaR	-0.0123	-0.0091	-0.0098	-0.0160	-0.0136	-0.0056	-0.0192	-0.0110	-0.0269	-0.0129	-0.0316	-0.0151
Δp_w	0.0734	0.0351	0.1874	0.1054	0.0268	0.0709	0.1022	0.0485	0.0790	0.0890	0.1103	0.0756

Table 5.9: PPO NoVIX Lookback 120 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.0846	0.3941	0.2478	-0.0268	0.1596	0.2275	-0.1195	0.3069	0.1087	0.2940	-0.0017	-0.0670
Geometric Mean Return	0.0003	0.0013	0.0009	-0.0001	0.0006	0.0008	-0.0005	0.0011	0.0004	0.0010	-0.0000	-0.0003
Annualized Return	0.0832	0.3941	0.2478	-0.0268	0.1596	0.2265	-0.1186	0.3069	0.1092	0.2940	-0.0017	-0.0660
Annual Volatility	0.1112	0.1206	0.1001	0.1560	0.1492	0.0805	0.1851	0.1142	0.3641	0.1368	0.1957	0.1680
Sharpe Ratio	0.8426	2.9808	2.1377	-0.0966	0.9767	2.6611	-0.5492	2.4138	0.4502	1.8793	0.0860	-0.3309
Calmar Ratio	4.0200	15.3446	14.1995	-0.6358	4.1801	12.6119	-2.7074	10.6346	0.9932	10.3595	-0.0402	-1.5764
Stability	0.0261	0.0846	0.0557	0.0289	0.0642	0.0463	0.0426	0.0567	0.1016	0.0632	0.0442	0.0417
Max Drawdown	0.0825	0.0453	0.0469	0.1328	0.0878	0.0331	0.2256	0.0560	0.3386	0.0528	0.1751	0.1869
Omega Ratio	1.1511	1.6358	1.4231	0.9838	1.1848	1.5423	0.9078	1.5272	1.0929	1.3578	1.0142	0.9463
Sortino Ratio	1.3420	4.3824	3.1906	-0.1421	1.3309	4.2040	-0.7138	2.9396	0.5352	2.7791	0.1340	-0.4885
Skew	0.0721	-0.2489	-0.1070	-0.2163	-0.3536	-0.1726	-0.3813	-0.6564	-0.4281	-0.2593	-0.0496	-0.0462
Kurtosis	0.6536	0.9908	0.6855	1.6399	1.6645	1.0299	2.3421	3.2286	6.7669	0.6414	0.6494	1.4715
Tail Ratio	1.1112	1.0262	1.1084	0.8986	0.9215	1.1095	0.7731	0.8404	0.9278	0.9572	0.9780	0.9893
VaR	-0.0108	-0.0112	-0.0102	-0.0161	-0.0164	-0.0066	-0.0212	-0.0116	-0.0307	-0.0144	-0.0210	-0.0181
Δp_w	0.0571	0.0776	0.1911	0.1295	0.0387	0.0213	0.2212	0.0662	0.0883	0.0399	0.0492	0.0171

Table 5.10: PPO Lookback 60 Using Daily Return as Reward Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.0460	0.3809	0.0702	-0.0504	0.0983	0.1996	-0.0888	0.2789	0.2433	0.3389	-0.1781	0.0731
Geometric Mean Return	0.0002	0.0013	0.0003	-0.0002	0.0004	0.0007	-0.0004	0.0010	0.0009	0.0012	-0.0008	0.0003
Annualized Return	0.0453	0.3809	0.0702	-0.0504	0.0987	0.1987	-0.0881	0.2789	0.2444	0.3404	-0.1781	0.0719
Annual Volatility	0.1268	0.1170	0.1316	0.1480	0.1181	0.0663	0.1578	0.0982	0.3609	0.1606	0.2480	0.1807
Sharpe Ratio	0.5180	2.9971	0.5389	-0.2530	0.7236	2.8790	-0.4669	2.4255	0.8105	1.8288	-0.6466	0.3247
Calmar Ratio	1.6541	14.9409	2.7062	-1.2700	3.1820	11.9842	-2.2794	15.8481	1.9145	9.4829	-4.1407	1.8128
Stability	0.0331	0.0807	0.0337	0.0324	0.0408	0.0483	0.0328	0.0601	0.1226	0.0807	0.0620	0.0372
Max Drawdown	0.0960	0.0444	0.0848	0.1383	0.0641	0.0306	0.1870	0.0341	0.2876	0.0684	0.2406	0.1269
Omega Ratio	1.0887	1.6294	1.0915	0.9594	1.1344	1.6355	0.9215	1.4969	1.1770	1.3493	0.9011	1.0529
Sortino Ratio	0.8215	4.4165	0.7390	-0.3782	0.9648	4.3074	-0.5913	3.5119	0.9178	2.5589	-1.0883	0.5286
Skew	-0.0209	-0.3190	-0.4791	-0.2753	-0.4090	-0.1093	-0.4027	-0.2158	-0.5735	-0.4402	0.2075	-0.0479
Kurtosis	0.6848	1.0582	0.4350	1.1015	1.9951	2.0974	2.6287	0.8455	8.3753	0.9319	0.6130	0.4828
Tail Ratio	1.0365	1.0114	0.8300	0.9062	0.9457	1.1895	0.7609	1.0237	0.9269	0.8754	0.9830	1.0644
VaR	-0.0120	-0.0107	-0.0148	-0.0145	-0.0123	-0.0052	-0.0186	-0.0103	-0.0345	-0.0165	-0.0264	-0.0175
Δp_w	0.0286	0.0717	0.0448	0.0164	0.0320	0.0053	0.0624	0.0267	0.0231	0.0934	0.0417	0.0268

Table 5.11: PPO NoVIX Lookback 60 Using Daily Return as Reward Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1402	0.2802	0.1006	-0.0062	0.0779	0.1936	-0.0605	0.2635	0.1831	0.3021	-0.0243	0.0142
Geometric Mean Return	0.0005	0.0010	0.0004	-0.0000	0.0003	0.0007	-0.0002	0.0009	0.0007	0.0010	-0.0001	0.0001
Annualized Return	0.1378	0.2802	0.1006	-0.0062	0.0782	0.1927	-0.0601	0.2635	0.1839	0.3034	-0.0243	0.0140
Annual Volatility	0.1368	0.1159	0.1292	0.1481	0.1254	0.0711	0.1652	0.1166	0.3602	0.1563	0.2093	0.1431
Sharpe Ratio	1.1553	2.4244	0.7326	0.0447	0.5368	2.5986	-0.2606	2.0361	0.6685	1.7115	0.0132	0.1653
Calmar Ratio	5.1667	10.6326	3.9387	-0.1555	2.4125	13.7356	-1.5228	9.4763	1.5181	9.5081	-0.5783	0.4826
Stability	0.0414	0.0657	0.0405	0.0244	0.0440	0.0410	0.0349	0.0474	0.1088	0.0724	0.0394	0.0312
Max Drawdown	0.0997	0.0533	0.0920	0.1165	0.0793	0.0234	0.1979	0.0492	0.3227	0.0706	0.1621	0.1346
Omega Ratio	1.2148	1.4975	1.1329	1.0075	1.1001	1.5400	0.9540	1.4168	1.1468	1.3226	1.0022	1.0277
Sortino Ratio	1.7516	3.3691	0.9672	0.0657	0.7094	3.9697	-0.3158	2.6623	0.7792	2.3572	0.0208	0.2480
Skew	-0.0742	-0.3866	-0.4305	-0.2255	-0.3810	-0.1029	-0.4852	-0.4729	-0.3260	-0.4690	-0.0415	-0.0121
Kurtosis	0.7114	1.5283	1.2405	1.5469	2.0198	0.7134	3.0978	2.7644	8.0960	0.7135	0.6223	0.8172
Tail Ratio	1.0189	0.9328	0.8215	0.9132	0.9059	1.1709	0.7484	0.8414	0.9762	0.8981	0.9718	0.9944
VaR	-0.0140	-0.0123	-0.0139	-0.0150	-0.0126	-0.0070	-0.0199	-0.0109	-0.0326	-0.0162	-0.0222	-0.0159
Δp_w	0.0226	0.0528	0.0343	0.0134	0.0127	0.0195	0.0190	0.0153	0.0435	0.0410	0.0393	0.0250

Table 5.12: Mean-Variance Optimization Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.0724	0.1006	0.2056	-0.0299	0.0747	0.1571	0.0449	-0.1130	-0.1638	0.0641	0.0671	0.2594
Geometric Mean Return	0.0003	0.0004	0.0007	-0.0001	0.0003	0.0006	0.0002	-0.0005	-0.0007	0.0002	0.0003	0.0009
Annualized Return	0.0718	0.1006	0.2056	-0.0299	0.0747	0.1564	0.0448	-0.1130	-0.1644	0.0641	0.0668	0.2571
Annual Volatility	0.1651	0.1234	0.1037	0.1659	0.1531	0.0746	0.1241	0.1628	0.3761	0.2030	0.2595	0.2166
Sharpe Ratio	0.5076	0.9004	1.7981	-0.1618	0.6388	2.0071	0.4690	-0.6219	-0.2959	0.4898	0.4121	1.1098
Calmar Ratio	1.7689	3.4141	9.9589	-0.7194	1.9434	10.8333	1.7264	-2.6566	-1.2621	1.3228	1.3154	6.9359
Stability	0.0323	0.0521	0.0617	0.0325	0.0321	0.0430	0.0306	0.0400	0.1202	0.0483	0.0642	0.0754
Max Drawdown	0.1155	0.0708	0.0725	0.1614	0.0574	0.0455	0.1050	0.2002	0.4099	0.0979	0.2275	0.0910
Omega Ratio	1.0976	1.1579	1.3415	0.9728	1.1152	1.4080	1.0826	0.8950	0.9359	1.0866	1.0705	1.2095
Sortino Ratio	0.6662	1.2216	2.7184	-0.2229	0.8329	2.9463	0.6470	-0.7643	-0.3023	0.7403	0.6313	2.0529
Skew	-0.2151	-0.6110	-0.1815	-0.2316	-0.5464	0.0263	-0.3279	-0.6130	-1.1650	-0.0047	-0.0523	0.9919
Kurtosis	2.6188	1.1195	0.5801	2.7024	1.5919	2.2548	1.0671	2.4465	10.7353	1.6891	0.3644	4.7644
Tail Ratio	0.9496	0.7960	1.0107	0.8648	0.8169	1.0676	0.8609	0.7157	0.7985	1.0152	1.0069	1.3310
VaR	-0.0176	-0.0125	-0.0106	-0.0152	-0.0176	-0.0074	-0.0139	-0.0189	-0.0310	-0.0190	-0.0271	-0.0192
Δp_w	0.3486	0.3962	0.2138	0.4110	0.2235	0.2205	0.2710	0.3471	0.3458	0.3030	0.2732	0.2347

Table 5.13: Black-Litterman Model Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1398	0.2400	0.2348	-0.0876	0.1690	0.0278	-0.0227	0.1708	-0.0419	0.2439	0.2747	0.4540
Geometric Mean Return	0.0005	0.0009	0.0008	-0.0004	0.0006	0.0001	-0.0001	0.0006	-0.0002	0.0009	0.0010	0.0015
Annualized Return	0.1386	0.2400	0.2348	-0.0876	0.1690	0.0277	-0.0226	0.1708	-0.0421	0.2439	0.2735	0.4496
Annual Volatility	0.1582	0.1229	0.1194	0.1690	0.1740	0.0845	0.1812	0.1354	0.4017	0.1689	0.2918	0.1754
Sharpe Ratio	1.0395	2.0132	1.7492	-0.4549	0.8797	0.4984	0.0319	1.2375	0.1203	1.3159	0.9612	2.2122
Calmar Ratio	3.6393	9.8123	9.5357	-2.1954	3.6232	1.1279	-0.5492	3.3847	-0.3080	7.6302	3.6215	14.4562
Stability	0.0518	0.0505	0.0718	0.0389	0.0554	0.0184	0.0283	0.0368	0.1014	0.0547	0.0930	0.1074
Max Drawdown	0.1147	0.0596	0.0493	0.1756	0.0998	0.0763	0.1525	0.1050	0.4059	0.0791	0.1768	0.0894
Omega Ratio	1.1899	1.3845	1.3426	0.9237	1.1758	1.0883	1.0058	1.2574	1.0243	1.2421	1.1714	1.4428
Sortino Ratio	1.6259	2.9709	2.5254	-0.6247	1.1600	0.6627	0.0430	1.4315	0.1377	2.0501	1.3497	3.6935
Skew	-0.0106	-0.3302	-0.1649	-0.2290	-0.1202	-0.6601	0.0849	-0.8473	-0.5994	0.0318	-0.4769	0.1702
Kurtosis	1.0173	0.7400	1.0807	2.1473	3.5802	2.3589	4.4493	8.0339	8.2183	0.6955	1.0768	0.6303
Tail Ratio	1.0725	0.9330	1.0256	0.8695	0.9863	0.8289	0.9034	0.7988	0.9113	1.1253	0.8531	1.2648
VaR	-0.0161	-0.0114	-0.0124	-0.0166	-0.0167	-0.0085	-0.0193	-0.0135	-0.0412	-0.0165	-0.0276	-0.0159
Δp_w	0.4540	0.3686	0.3598	0.5256	0.3047	0.4172	0.3342	0.4313	0.4062	0.2923	0.2799	0.3174

II Detailed South Korean Market Metrics

Table 5.14: KOSPI Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Annualized Return	0.0950	-0.0099	-0.0270	0.0184	0.0575	0.2288	-0.1821	0.0958	0.3270	0.0115	-0.2589	0.1991
Geometric Mean Return	0.0004	-0.0000	-0.0001	0.0001	0.0002	0.0008	-0.0008	0.0004	0.0011	0.0000	-0.0012	0.0007
Cumulative Return	0.0935	-0.0097	-0.0262	0.0181	0.0561	0.2178	-0.1769	0.0934	0.3210	0.0113	-0.2519	0.1930
Annual Volatility	0.1545	0.1232	0.1015	0.1269	0.1222	0.0942	0.1436	0.1256	0.2826	0.1641	0.1872	0.1539
Sharpe Ratio	0.6671	0.1203	-0.4438	0.2521	0.3351	2.2345	-1.2918	0.6660	1.1055	0.3022	-1.4848	1.2245
Calmar Ratio	2.7931	-0.4955	-1.5639	0.7460	1.8596	11.6455	-4.1022	3.1453	3.8955	0.3788	-7.3572	7.3398
Stability	0.0361	0.0293	0.0227	0.0341	0.0251	0.0685	0.0669	0.0376	0.1240	0.0345	0.0769	0.0355
Max Drawdown	0.1366	0.1216	0.0889	0.1581	0.0533	0.0515	0.2318	0.1507	0.2818	0.1411	0.2789	0.1459
Omega Ratio	1.1185	1.0201	0.9293	1.0437	1.0618	1.4652	0.8046	1.1187	1.2252	1.0508	0.7885	1.2284
Sortino Ratio	1.0194	0.1971	-0.6458	0.3708	0.4172	3.1690	-1.6745	0.8780	1.3851	0.4950	-2.2912	2.0223
Skew	-0.0380	0.2820	-0.1791	0.0076	-0.6363	-0.2350	-0.6527	-0.5575	-0.0917	0.2029	-0.1034	0.5519
Kurtosis	1.0642	1.0484	0.5973	1.2071	2.4777	1.6718	2.7965	1.0517	5.3769	1.1226	0.0605	4.0998
Tail Ratio	1.0622	1.0820	0.8981	1.0030	0.8194	0.9483	0.7031	0.8103	0.9177	1.0828	0.8572	1.0782
VaR	-0.0147	-0.0131	-0.0113	-0.0139	-0.0121	-0.0099	-0.0155	-0.0137	-0.0298	-0.0159	-0.0203	-0.0151

Table 5.15: PPO Lookback 10 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1054	0.1640	0.1123	0.2756	0.0819	0.1401	-0.2362	0.1223	0.7421	0.3139	-0.0741	0.2417
Geometric Mean Return	0.0004	0.0006	0.0004	0.0010	0.0003	0.0006	-0.0011	0.0005	0.0023	0.0011	-0.0003	0.0009
Annualized Return	0.1100	0.1737	0.1193	0.2871	0.0861	0.1510	-0.2500	0.1300	0.7868	0.3288	-0.0780	0.2564
Annual Volatility	0.1628	0.1871	0.1647	0.2343	0.1966	0.1813	0.1836	0.1391	0.3562	0.2473	0.1878	0.1637
Sharpe Ratio	0.6650	1.0123	0.7136	1.2593	0.5825	0.8729	-1.6248	0.9779	1.8916	1.4478	-0.2950	1.5670
Calmar Ratio	2.9291	5.7125	3.4131	8.4907	2.3725	3.5670	-6.5319	4.9004	7.8438	8.3341	-1.9763	9.1212
Stability	0.0435	0.0676	0.0687	0.0767	0.0365	0.0615	0.0749	0.0296	0.2165	0.0901	0.0477	0.0452
Max Drawdown	0.1706	0.1368	0.1217	0.1704	0.1108	0.1300	0.2617	0.1191	0.2273	0.1697	0.1772	0.1260
Omega Ratio	1.1145	1.1956	1.1228	1.2336	1.1015	1.1553	0.7653	1.1721	1.4025	1.2867	0.9516	1.2999
Sortino Ratio	1.0115	1.9028	1.0888	2.3103	0.9043	1.3152	-2.4688	1.4328	2.5216	2.6895	-0.4264	2.8517
Skew	-0.1797	1.4022	-0.2136	0.5020	0.0325	-0.1848	0.0185	-0.2929	-0.3708	0.7525	-0.1667	0.4291
Kurtosis	0.6256	8.3119	0.8031	0.7748	0.7651	1.1937	0.9464	0.6800	4.0834	1.9539	0.5577	0.8086
Tail Ratio	0.9687	1.4022	0.9347	1.3957	1.0084	0.9602	0.8892	0.8857	1.0790	1.5152	0.9410	1.3349
VaR	-0.0173	-0.0163	-0.0154	-0.0212	-0.0215	-0.0176	-0.0203	-0.0139	-0.0297	-0.0205	-0.0204	-0.0140
Δp_w	0.1216	0.1123	0.1228	0.1679	0.1209	0.1344	0.1336	0.1332	0.2264	0.1619	0.1425	0.0745

Table 5.16: PPO Lookback 30 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	-0.0311	0.0667	0.0292	-0.0835	0.4624	0.4707	-0.1528	0.1829	0.4530	0.4484	-0.1825	0.4773
Geometric Mean Return	-0.0001	0.0003	0.0001	-0.0003	0.0015	0.0016	-0.0007	0.0007	0.0015	0.0015	-0.0008	0.0016
Annualized Return	-0.0314	0.0675	0.0298	-0.0838	0.4714	0.4968	-0.1562	0.1861	0.4574	0.4505	-0.1865	0.4988
Annual Volatility	0.1835	0.3250	0.1455	0.1839	0.2204	0.1408	0.1770	0.1650	0.3369	0.2302	0.2393	0.2327
Sharpe Ratio	-0.1365	0.4436	0.1791	-0.3295	1.7517	2.9462	-0.8596	0.9795	1.2383	1.7272	-0.7059	1.8404
Calmar Ratio	-0.9090	1.0370	1.2502	-2.4433	13.3556	21.4956	-3.6017	5.4358	4.3496	9.1810	-2.8310	9.0068
Stability	0.0903	0.1150	0.0582	0.0682	0.0948	0.1178	0.0564	0.0426	0.1470	0.1162	0.0666	0.1207
Max Drawdown	0.2771	0.2332	0.1141	0.2574	0.0679	0.0843	0.2352	0.1429	0.2938	0.1817	0.2837	0.1768
Omega Ratio	0.9786	1.0832	1.0287	0.9478	1.3380	1.6302	0.8687	1.1751	1.2569	1.3331	0.8866	1.3907
Sortino Ratio	-0.2183	0.7724	0.2866	-0.5445	3.0826	4.9555	-1.2495	1.5769	1.5876	2.8119	-1.0536	3.2065
Skew	-0.0904	0.7736	-0.0896	0.1580	0.4218	0.1661	-0.1934	0.0757	-0.0782	0.1062	-0.0752	1.1909
Kurtosis	-0.0710	2.8253	-0.1267	0.6148	1.2971	0.9028	1.1927	0.8863	6.0922	0.9808	2.1868	8.4955
Tail Ratio	0.9322	1.3987	0.9504	1.0548	1.3449	1.2270	0.8468	1.0687	0.9942	1.1522	0.9696	1.3355
VaR	-0.0196	-0.0278	-0.0162	-0.0188	-0.0208	-0.0113	-0.0208	-0.0161	-0.0335	-0.0226	-0.0225	-0.0205
Δp_w	0.2617	0.0565	0.0871	0.0804	0.1260	0.0958	0.0861	0.1224	0.1003	0.0788	0.0589	0.0615

Table 5.17: PPO Lookback 60 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.2352	0.2667	0.0790	-0.0537	0.0439	0.0333	-0.1868	0.1139	0.5305	0.2372	-0.2189	0.5883
Geometric Mean Return	0.0009	0.0010	0.0003	-0.0002	0.0002	0.0001	-0.0009	0.0004	0.0017	0.0009	-0.0010	0.0019
Annualized Return	0.2416	0.2753	0.0820	-0.0545	0.0448	0.0352	-0.1966	0.1178	0.5437	0.2414	-0.2252	0.6158
Annual Volatility	0.1886	0.2700	0.1565	0.1791	0.1872	0.1609	0.2225	0.1799	0.3158	0.2213	0.2147	0.2853
Sharpe Ratio	1.2054	1.1165	0.5258	-0.1829	0.2173	0.3461	-0.9143	0.6675	1.5091	1.3178	-1.0471	1.7867
Calmar Ratio	6.3981	2.8968	1.9635	-1.8086	1.0345	0.9094	-3.5240	3.4112	5.6917	8.2018	-5.5988	10.4707
Stability	0.0634	0.0924	0.0737	0.0386	0.0271	0.0473	0.0677	0.0608	0.1661	0.0999	0.0963	0.1306
Max Drawdown	0.1257	0.1295	0.1203	0.1713	0.0937	0.1590	0.2515	0.2218	0.2470	0.2288	0.2915	0.1782
Omega Ratio	1.2082	1.2572	1.0930	0.9710	1.0370	1.0599	0.8604	1.1161	1.3260	1.2407	0.8458	1.3856
Sortino Ratio	1.9623	1.8517	0.7155	-0.2758	0.3148	0.5106	-1.4134	0.9952	1.8297	2.5005	-1.7922	3.1522
Skew	-0.1818	2.0607	-0.3863	-0.1292	-0.2467	-0.1719	-0.0213	-0.1695	-0.1322	0.5016	0.1688	1.2578
Kurtosis	0.2199	22.6437	2.1286	0.3625	1.2114	1.8946	0.9314	0.3222	6.1274	0.6187	0.1115	7.6938
Tail Ratio	0.9685	1.4918	0.8473	0.8836	0.9249	0.9515	0.9811	0.9716	0.9354	1.4200	1.0065	1.3902
VaR	-0.0184	-0.0192	-0.0171	-0.0205	-0.0200	-0.0141	-0.0243	-0.0181	-0.0328	-0.0199	-0.0225	-0.0239
Δp_w	0.0505	0.0700	0.0838	0.0363	0.0430	0.0378	0.0616	0.0630	0.0609	0.0553	0.0613	0.0361

Table 5.18: PPO Lookback 120 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.2532	0.1500	-0.0303	0.1480	0.1041	0.2158	-0.1398	-0.0071	0.8220	0.1968	-0.1531	0.4973
Geometric Mean Return	0.0009	0.0006	-0.0001	0.0006	0.0004	0.0008	-0.0006	-0.0000	0.0025	0.0007	-0.0007	0.0017
Annualized Return	0.2601	0.1553	-0.0318	0.1532	0.1077	0.2299	-0.1485	-0.0074	0.8581	0.2039	-0.1601	0.5225
Annual Volatility	0.1675	0.1805	0.1426	0.2102	0.1631	0.1470	0.2069	0.1360	0.3639	0.2349	0.2470	0.2000
Sharpe Ratio	1.4360	0.8655	-0.1643	0.8150	0.5316	1.5441	-0.6265	0.1816	1.8762	0.8676	-0.6548	2.1971
Calmar Ratio	8.9206	5.1148	-0.7318	4.5603	3.1896	6.7140	-3.9457	-0.2748	7.9109	4.3191	-3.4048	14.1614
Stability	0.0680	0.0839	0.0447	0.0729	0.0305	0.0643	0.0527	0.0445	0.2514	0.1059	0.0883	0.0988
Max Drawdown	0.1315	0.1012	0.1443	0.1568	0.0733	0.1101	0.2517	0.1662	0.2527	0.2713	0.2696	0.1260
Omega Ratio	1.2575	1.1680	0.9735	1.1387	1.0894	1.2905	0.9000	1.0292	1.4205	1.1529	0.8994	1.4467
Sortino Ratio	2.5426	1.5895	-0.2308	1.3480	0.8349	2.2179	-1.0346	0.2934	2.1500	1.4222	-1.0537	4.1357
Skew	0.1932	1.3688	-0.5340	0.0242	-0.0980	-0.3703	0.4158	-0.0555	-0.7195	0.1093	0.0127	0.5531
Kurtosis	0.5117	6.9891	1.7845	-0.0511	0.9215	1.2463	1.5480	0.3721	5.6209	0.4503	0.2400	1.4229
Tail Ratio	1.2015	1.4229	0.8686	1.0790	1.0269	0.9490	1.1206	1.0088	0.8658	1.1864	0.9569	1.4542
VaR	-0.0155	-0.0171	-0.0142	-0.0231	-0.0154	-0.0139	-0.0200	-0.0137	-0.0327	-0.0235	-0.0227	-0.0163
Δp_w	0.0180	0.0252	0.0290	0.0123	0.0204	0.0236	0.0082	0.0051	0.0198	0.0322	0.0272	0.0070

Table 5.19: PPO NoVKOSPI Lookback 10 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.2126	0.7639	-0.0309	-0.2389	0.0146	0.3491	-0.0739	0.0182	0.3718	0.2101	-0.1175	0.2453
Geometric Mean Return	0.0008	0.0024	-0.0001	-0.0011	0.0001	0.0013	-0.0003	0.0001	0.0013	0.0008	-0.0005	0.0009
Annualized Return	0.2223	0.8192	-0.0327	-0.2465	0.0153	0.3787	-0.0787	0.0192	0.3918	0.2197	-0.1235	0.2603
Annual Volatility	0.1818	0.2649	0.1703	0.1799	0.3233	0.1846	0.2075	0.1900	0.3096	0.2331	0.2483	0.2492
Sharpe Ratio	1.1322	2.3973	-0.2940	-1.4547	0.1851	1.7711	-0.2408	0.1863	1.2674	1.0802	-0.3951	1.1035
Calmar Ratio	7.9155	26.4797	-0.9889	-8.4194	0.1655	9.0452	-1.8215	0.4249	5.0477	5.4887	-2.7150	5.7157
Stability	0.0679	0.1771	0.0320	0.0943	0.0504	0.0938	0.0396	0.0480	0.1568	0.0875	0.0949	0.0639
Max Drawdown	0.1519	0.0956	0.1504	0.2692	0.1967	0.1235	0.1702	0.2200	0.2609	0.2150	0.2891	0.2144
Omega Ratio	1.2010	1.5535	0.9524	0.7889	1.0348	1.3503	0.9567	1.0321	1.2620	1.1936	0.9360	1.2078
Sortino Ratio	1.8423	5.5011	-0.4269	-2.4236	0.2864	3.0186	-0.3858	0.2787	1.6474	1.9229	-0.6106	1.9055
Skew	0.0137	2.1454	-0.2310	0.3563	0.5927	0.5071	1.1611	-0.0859	-0.0639	0.3110	0.0392	0.5803
Kurtosis	0.0945	11.6413	0.5592	0.9472	6.5934	3.2742	7.5337	1.8816	3.8506	0.5247	0.5368	2.8139
Tail Ratio	1.0493	1.9129	0.8462	1.0521	1.1877	1.3315	1.1722	0.9754	0.9934	1.2888	0.9897	1.2670
VaR	-0.0192	-0.0193	-0.0186	-0.0202	-0.0257	-0.0158	-0.0217	-0.0187	-0.0315	-0.0221	-0.0243	-0.0224
Δp_w	0.3814	0.7447	0.3644	0.6560	0.4926	0.3138	0.3878	0.4791	0.7491	0.5819	0.4456	0.1733

Table 5.20: PPO NoVKOSPI Lookback 30 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.1695	0.0918	-0.0026	0.2878	-0.1486	0.0748	-0.0841	0.1225	0.5020	0.1993	-0.1849	0.9156
Geometric Mean Return	0.0006	0.0004	-0.0000	0.0010	-0.0006	0.0003	-0.0004	0.0005	0.0016	0.0007	-0.0008	0.0027
Annualized Return	0.1717	0.0929	-0.0027	0.2891	-0.1508	0.0784	-0.0861	0.1245	0.5069	0.2002	-0.1890	0.9623
Annual Volatility	0.1785	0.2020	0.1507	0.1925	0.2592	0.1642	0.1914	0.1526	0.3309	0.1904	0.1794	0.2520
Sharpe Ratio	0.9941	0.5117	-0.0580	1.4522	-0.6253	0.5872	-0.3735	0.7440	1.4030	1.1418	-1.0571	2.7653
Calmar Ratio	6.1882	2.7521	-0.0774	8.1681	-2.1124	1.7599	-1.8675	3.6800	4.7435	6.7171	-5.2284	22.9269
Stability	0.0480	0.0716	0.0370	0.0900	0.0497	0.0310	0.0473	0.0458	0.1714	0.0605	0.0760	0.1923
Max Drawdown	0.1561	0.1202	0.1123	0.1387	0.1927	0.1326	0.1808	0.1747	0.2799	0.1064	0.2452	0.1683
Omega Ratio	1.1735	1.0865	0.9909	1.2563	0.8900	1.1000	0.9404	1.1271	1.3020	1.1994	0.8390	1.6193
Sortino Ratio	1.6928	0.8075	-0.0882	2.3368	-0.7898	0.9296	-0.5567	1.1799	1.5246	2.0644	-1.6248	5.3165
Skew	0.1328	-0.0492	-0.3013	-0.1817	-0.4175	-0.1258	-0.1812	-0.0986	-1.0651	0.1631	0.0101	0.9620
Kurtosis	0.1035	0.2756	0.3683	-0.0925	5.6957	1.0988	0.9146	0.7630	7.0379	-0.0748	0.5901	3.8274
Tail Ratio	1.1281	1.0011	0.8518	0.9895	0.7926	1.0419	0.9092	1.0135	0.8071	1.2013	0.9290	1.6090
VaR	-0.0169	-0.0192	-0.0161	-0.0196	-0.0265	-0.0151	-0.0224	-0.0144	-0.0318	-0.0187	-0.0194	-0.0199
Δp_w	0.2890	0.0681	0.3993	0.7907	0.9162	0.2461	0.5011	0.7698	0.4828	0.2487	0.2621	0.3058

Table 5.21: PPO NoVKOSPI Lookback 60 Results Metrics by Year

Metric	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Return	0.1946	0.0065	-0.0787	-0.2091	0.1559	0.2860	-0.2705	0.2306	0.6224	0.0581	-0.2098	0.4840
Geometric Mean Return	0.0007	0.0000	-0.0003	-0.0009	0.0006	0.0011	-0.0013	0.0009	0.0020	0.0002	-0.0010	0.0016
Annualized Return	0.1998	0.0066	-0.0815	-0.2121	0.1593	0.3037	-0.2839	0.2390	0.6384	0.0590	-0.2159	0.5059
Annual Volatility	0.1925	0.1685	0.2478	0.1999	0.2197	0.1710	0.2139	0.1660	0.3873	0.2343	0.2011	0.1922
Sharpe Ratio	1.1008	0.2267	-0.3152	-1.0634	0.6644	1.6876	-1.4437	1.1828	1.5118	0.5073	-1.0809	2.2153
Calmar Ratio	5.5655	0.2500	-1.3092	-6.7757	4.0963	8.5243	-7.6219	6.8910	4.7120	1.6323	-5.2877	13.3399
Stability	0.0535	0.0308	0.0537	0.0776	0.0463	0.0914	0.0780	0.0318	0.2253	0.0511	0.0716	0.0996
Max Drawdown	0.1937	0.1450	0.2109	0.2784	0.0810	0.1203	0.2948	0.0894	0.2410	0.1605	0.2517	0.1162
Omega Ratio	1.1950	1.0361	0.9481	0.8473	1.1215	1.3231	0.7911	1.2183	1.3259	1.0864	0.8418	1.4483
Sortino Ratio	1.8684	0.3907	-0.4614	-1.7422	1.0794	2.4619	-2.2918	1.6973	1.7947	0.8308	-1.8052	3.7853
Skew	0.1830	0.0222	-0.1615	-0.0525	0.6900	-0.1299	0.0867	-0.2403	-0.5889	0.1391	0.0759	0.2422
Kurtosis	0.4794	-0.2876	1.2520	-0.3239	4.3221	1.9090	0.1414	1.0837	6.3101	0.2458	0.1373	1.3066
Tail Ratio	1.1848	1.0169	0.9355	0.8910	1.1360	0.9718	0.9315	0.9561	0.8977	1.1140	0.9566	1.3026
VaR	-0.0187	-0.0166	-0.0245	-0.0238	-0.0209	-0.0161	-0.0231	-0.0172	-0.0379	-0.0250	-0.0203	-0.0158
Δp_w	0.4593	0.2874	0.6099	1.1542	1.3550	0.8902	0.4872	0.9610	0.4282	0.1858	0.3342	0.2882

Table 5.22: PPO NoVKOSPI Lookback 120 Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.2469	0.0875	0.0934	-0.1112	-0.1954	0.3199	-0.1405	0.1106	0.4636	0.2154	-0.1989	0.6064
Geometric Mean Return	0.0009	0.0003	0.0004	-0.0005	-0.0009	0.0012	-0.0006	0.0004	0.0016	0.0008	-0.0009	0.0020
Annualized Return	0.2536	0.0905	0.0983	-0.1146	-0.2011	0.3416	-0.1493	0.1159	0.4819	0.2232	-0.2077	0.6381
Annual Volatility	0.1886	0.2312	0.1597	0.2117	0.2502	0.1482	0.2212	0.1694	0.3357	0.2173	0.2461	0.2630
Sharpe Ratio	1.2537	0.4293	0.6415	-0.4540	-0.8305	2.0685	-0.5893	0.8569	1.3019	0.9700	-0.8852	1.9672
Calmar Ratio	8.2624	1.0810	3.7071	-2.8855	-2.6107	15.5568	-3.0367	2.6609	4.4181	5.9682	-4.4515	9.3919
Stability	0.0725	0.0622	0.0686	0.0713	0.0481	0.0702	0.0695	0.0514	0.1862	0.0848	0.0662	0.1080
Max Drawdown	0.1318	0.1723	0.1267	0.2442	0.2255	0.0585	0.2848	0.2203	0.2820	0.1935	0.2818	0.1607
Omega Ratio	1.2219	1.0782	1.1077	0.9297	0.8621	1.4005	0.9033	1.1504	1.2665	1.1715	0.8619	1.4187
Sortino Ratio	2.1841	0.5992	1.0554	-0.7807	-1.1060	3.9433	-0.8641	1.2460	1.6391	1.6508	-1.3558	3.2295
Skew	0.1280	-0.3564	0.0058	0.2779	-0.4212	0.4350	-0.0716	-0.3727	-0.3815	0.1839	0.0535	0.5628
Kurtosis	0.0692	5.3042	0.1064	1.4375	4.6630	0.6340	1.2055	1.0902	5.4241	0.3420	0.9035	4.1886
Tail Ratio	1.1261	1.0620	1.0737	1.0162	0.7953	1.4387	0.9125	0.8869	0.9325	1.1393	0.9542	1.4253
VaR	-0.0183	-0.0210	-0.0153	-0.0198	-0.0257	-0.0126	-0.0230	-0.0177	-0.0319	-0.0208	-0.0237	-0.0211
Δp_w	1.1240	1.0803	0.6041	1.0000	1.2477	1.2677	1.3085	1.3481	1.0814	1.2527	1.2398	1.0522

Table 5.23: PPO Lookback 60 Using Daily Return as Reward Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.1394	-0.0632	0.2253	-0.0393	0.0643	0.0543	-0.1963	0.0042	0.1454	0.0727	-0.0930	0.4864
Geometric Mean Return	0.0005	-0.0003	0.0008	-0.0002	0.0003	0.0002	-0.0009	0.0000	0.0005	0.0003	-0.0004	0.0016
Annualized Return	0.1430	-0.0649	0.2345	-0.0400	0.0657	0.0574	-0.2066	0.0044	0.1485	0.0739	-0.0959	0.5084
Annual Volatility	0.1522	0.1921	0.1270	0.1791	0.2439	0.1640	0.2012	0.1838	0.2850	0.2254	0.2467	0.2213
Sharpe Ratio	0.9358	-0.0815	1.5502	-0.0724	0.3505	0.4032	-0.9650	-0.0792	0.5731	0.6074	-0.2124	1.9591
Calmar Ratio	4.4717	-1.7183	8.6174	-0.9627	1.4287	1.2007	-4.4744	0.1098	1.7778	1.7835	-2.0259	13.1230
Stability	0.0383	0.0411	0.0771	0.0455	0.0389	0.0439	0.0743	0.0771	0.0898	0.0609	0.0864	0.1131
Max Drawdown	0.1260	0.2034	0.0537	0.1889	0.1554	0.0987	0.2810	0.2491	0.3344	0.2074	0.2800	0.1490
Omega Ratio	1.1694	0.9871	1.2889	0.9886	1.0599	1.0686	0.8532	0.9874	1.1100	1.1061	0.9659	1.3839
Sortino Ratio	1.3515	-0.1357	2.2126	-0.1130	0.5496	0.6269	-1.3979	-0.1166	0.7443	1.0163	-0.3529	3.5468
Skew	-0.2363	0.1166	-0.3624	-0.1880	0.0949	-0.1489	-0.2494	-0.3559	-0.1768	0.2934	0.1605	0.4845
Kurtosis	0.8888	0.4646	1.0188	0.4105	0.6899	1.5051	0.6787	0.5154	4.5057	0.7542	0.2620	1.5860
Tail Ratio	0.9453	1.0444	0.9184	0.8866	1.0913	1.0599	0.8619	0.8239	0.9406	1.1994	1.0887	1.3327
VaR	-0.0160	-0.0197	-0.0128	-0.0204	-0.0233	-0.0151	-0.0221	-0.0194	-0.0295	-0.0232	-0.0248	-0.0193
Δp_w	0.1476	0.1183	0.1748	0.1769	0.1119	0.1121	0.1426	0.1161	0.1794	0.1459	0.1080	0.1372

Table 5.24: PPO NoVKOSPI Lookback 60 Using Daily Return as Reward Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Cumulative Return	0.2507	0.0144	0.0769	-0.1227	0.0855	-0.0227	-0.1296	0.2576	0.3472	0.2565	-0.2197	0.3744
Geometric Mean Return	0.0009	0.0001	0.0003	-0.0005	0.0003	-0.0001	-0.0006	0.0009	0.0012	0.0009	-0.0010	0.0013
Annualized Return	0.2575	0.0148	0.0798	-0.1245	0.0873	-0.0239	-0.1367	0.2671	0.3554	0.2611	-0.2260	0.3907
Annual Volatility	0.1883	0.1698	0.1604	0.2222	0.2460	0.1542	0.2039	0.1801	0.3356	0.2070	0.2203	0.2260
Sharpe Ratio	1.2912	0.2981	0.4135	-0.4767	0.3403	0.0296	-0.5824	1.2689	1.0287	1.3191	-0.9416	1.5840
Calmar Ratio	7.9709	0.4692	2.3411	-2.2248	1.5139	-0.7821	-3.5630	6.7838	2.6174	6.9080	-5.4282	9.2085
Stability	0.0580	0.0375	0.0486	0.0607	0.0374	0.0338	0.0816	0.0400	0.1321	0.1010	0.0693	0.0797
Max Drawdown	0.1441	0.1535	0.1096	0.2403	0.1206	0.1477	0.2650	0.1322	0.3084	0.1758	0.2711	0.1979
Omega Ratio	1.2354	1.0499	1.0674	0.9233	1.0648	1.0048	0.9073	1.2392	1.2141	1.2370	0.8543	1.3167
Sortino Ratio	2.1807	0.5001	0.6475	-0.7863	0.5351	0.0429	-0.8717	1.9493	1.1707	2.2198	-1.4621	2.5025
Skew	0.2112	0.1557	-0.2593	0.4593	1.1910	-0.3231	0.0375	0.0835	-1.0948	0.0767	0.1284	0.3882
Kurtosis	0.5807	0.6023	0.3510	3.6183	10.2679	0.2991	0.7751	1.6263	9.4388	0.5149	0.7293	2.1431
Tail Ratio	1.2051	1.1328	0.8931	1.0495	1.0815	0.8695	0.9656	1.1024	0.8379	1.1181	1.0076	1.2589
VaR	-0.0185	-0.0164	-0.0152	-0.0195	-0.0261	-0.0160	-0.0211	-0.0177	-0.0293	-0.0200	-0.0238	-0.0215
Δp_w	1.4819	1.4125	1.4661	1.4784	1.4353	1.4301	1.4258	1.3422	1.3664	1.4095	1.4381	1.4198

Table 5.25: Mean-Variance Optimization Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Annualized Return	-0.1282	0.1216	-0.0247	0.0817	0.1097	-0.0381	-0.0813	-0.2725	0.3072	0.5856	-0.1683	0.7028
Geometric Mean Return	-0.0006	0.0005	-0.0001	0.0003	0.0004	-0.0002	-0.0003	-0.0013	0.0011	0.0019	-0.0008	0.0022
Cumulative Return	-0.1301	0.1242	-0.0254	0.0827	0.1125	-0.0398	-0.0838	-0.2782	0.3128	0.5975	-0.1733	0.7289
Annual Volatility	0.2453	0.3215	0.1745	0.2838	0.2206	0.1602	0.1918	0.2376	0.4301	0.2815	0.3105	0.2273
Sharpe Ratio	-0.4482	0.5092	-0.1214	0.4023	0.6742	-0.2392	-0.3608	-1.4499	0.7975	1.7939	-0.4848	2.5520
Calmar Ratio	-1.9203	0.7158	-0.6558	1.4912	2.4060	-1.2283	-1.9742	-5.4792	1.7600	14.1231	-2.5997	21.5590
Stability	0.0922	0.0611	0.0478	0.1151	0.0457	0.0320	0.0649	0.0953	0.1711	0.2154	0.0732	0.1940
Max Drawdown	0.2961	0.2278	0.1197	0.3061	0.1439	0.1163	0.2260	0.3791	0.3331	0.1976	0.3050	0.0621
Omega Ratio	0.9264	1.1198	0.9806	1.0748	1.1236	0.9605	0.9426	0.7824	1.1661	1.3475	0.9189	1.5687
Sortino Ratio	-0.6754	0.5811	-0.1843	0.6389	0.9851	-0.3521	-0.5562	-1.9620	0.8864	3.5282	-0.7614	4.6623
Skew	0.0431	-1.1412	-0.2008	0.5253	0.0662	-0.0140	-0.0419	-0.3227	-1.3949	0.6548	0.3859	0.6910
Kurtosis	2.5668	31.3580	0.3173	2.9421	2.5680	1.2995	0.5489	0.9532	9.2561	1.3334	2.6041	1.8363
Tail Ratio	0.9887	1.0539	0.8814	1.2418	1.0109	0.9905	1.0291	0.7633	0.8333	1.5152	1.1094	1.5035
VaR	-0.0229	-0.0200	-0.0192	-0.0254	-0.0203	-0.0168	-0.0200	-0.0285	-0.0368	-0.0252	-0.0285	-0.0186
Δp_w	0.2602	0.2465	0.2511	0.2151	0.2413	0.2235	0.3345	0.3093	0.2741	0.2588	0.4005	0.2525

Table 5.26: Black-Litterman Model Results Metrics by Year

Metric	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Annualized Return	-0.0833	0.1337	-0.1567	0.1972	0.1050	0.0502	-0.3803	0.0387	0.6901	0.0217	-0.3334	0.4171
Geometric Mean Return	-0.0004	0.0005	-0.0007	0.0007	0.0004	0.0002	-0.0020	0.0002	0.0021	0.0001	-0.0017	0.0014
Cumulative Return	-0.0846	0.1366	-0.1608	0.1997	0.1076	0.0525	-0.3900	0.0397	0.7045	0.0220	-0.3422	0.4313
Annual Volatility	0.2097	0.2125	0.2169	0.2896	0.2255	0.1674	0.2798	0.2031	0.5052	0.2632	0.3144	0.2469
Sharpe Ratio	-0.2889	0.7928	-0.7218	0.7806	0.5017	0.3699	-1.6717	0.1036	1.2670	0.3982	-1.2300	1.5423
Calmar Ratio	-1.9329	2.4870	-2.6979	2.1295	2.4927	1.7859	-6.5741	0.8363	3.5313	0.3790	-4.1492	5.8683
Stability	0.0860	0.0358	0.0745	0.0892	0.0510	0.0288	0.1791	0.0539	0.2303	0.1268	0.1312	0.1467
Max Drawdown	0.2456	0.1491	0.2891	0.2148	0.1366	0.1037	0.4305	0.1617	0.2917	0.3420	0.3846	0.1528
Omega Ratio	0.9550	1.1441	0.8860	1.1456	1.0872	1.0613	0.7508	1.0173	1.2904	1.0699	0.8048	1.3244
Sortino Ratio	-0.4812	1.2361	-0.9955	1.0796	0.7432	0.6492	-2.4479	0.1592	1.6240	0.6500	-1.7798	2.3376
Skew	0.0319	0.2240	-0.4322	-0.4190	-0.0713	0.1911	0.1139	-0.0157	0.1479	0.2389	0.0267	0.5847
Kurtosis	0.3099	2.8504	1.6545	3.0574	0.8947	0.2418	2.6055	1.4231	12.9348	1.7984	2.6407	6.8829
Tail Ratio	0.9766	1.1081	0.8484	0.9669	0.9964	1.1321	0.8714	0.9314	1.1518	1.1955	0.9496	1.2512
VaR	-0.0208	-0.0207	-0.0211	-0.0287	-0.0251	-0.0161	-0.0287	-0.0190	-0.0387	-0.0214	-0.0339	-0.0202
Δp_w	0.2683	0.2456	0.3095	0.2169	0.2435	0.2479	0.3848	0.3160	0.3216	0.3193	0.3967	0.2639

References

[1] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[2] S. Sood, K. Papasotiriou, M. Vaiciulis, and T. Balch, "Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization," *J.P. Morgan AI Research*, 2023.

[3] A. C. Logue, "Harry markowitz and modern portfolio theory," *Britannica*.

[4] G. Cornuejols and R. Tuncel, *Optimization methods in finance*. Cambridge University Press, 2006, vol. 5.

[5] E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*. Wiley, 2009.

[6] R. C. Gibson, *Asset Allocation: Balancing Financial Risk*. McGraw-Hill, 2008.

[7] W. F. Sharpe, "The sharpe ratio," *The Journal of Portfolio Management*, 1994.

[8] F. Black and R. Litterman, "Combining investor views with market equilibrium," *The Journal of Fixed Income*, 1991.

[9] T. Idzorek, "A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels," in *Forecasting Expected Returns in the Financial Markets*. Elsevier Ltd, 2007, pp. 17–38.

[10] J. Walters, "The black-litterman model in detail," *SSRN Electronic Journal*, 2014, (February 2007):1–65.

[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[12] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[15] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[16] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1928–1937, 2016.

[17] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[18] R. Martin, "Pyportfolioopt: Portfolio optimization in python," 2018. [Online]. Available: https://pyportfolioopt.readthedocs.io/en/latest/MeanVariance.html

[19] R. Aroussi, "yfinance: Yahoo! finance market data downloader," 2019. [Online]. Available: https://pypi.org/project/yfinance/

[20] "Black-litterman model," https://pyportfolioopt.readthedocs.io/en/latest/BlackLitterman.html.

[21] KRX Information Data System, "Vkospi indices," http://data.krx.co.kr/contents/MDC/MDI/mdiLoader/index.cmd?menuId=MDC0201010105, 2023.

[22] J. Moody and M. Saffell, "Learning to trade via direct reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 875–889, 2001.

[23] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, *Stable Baselines3 Documentation: Proximal Policy Optimization*, 2020, https://stable-baselines3.readthedocs.io/en/master/modules/ppo.html.

[24] Z. Zhang, S. Zohren, and S. Roberts, "Deep reinforcement learning for trading," *arXiv preprint arXiv:1911.10107*, 2019.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Jin Hyuk Choi, whose caring guidance and insightful teaching have been instrumental in bringing me to this point. His mentorship has provided me with invaluable experiences and opportunities that have shaped my academic journey.

I would also like to express my sincere gratitude to the members of my thesis committee, Professor Bongsoo Jang and Professor Pilwon Kim, for their invaluable feedback and constructive criticism throughout the course of my research.

My sincere thanks to my research lab members, Heeyoung Kwon, Sungjin Kim, and Jongkyung Choi, for their continued support and companionship. Our time together has been both meaningful and enjoyable, and their encouragement has been a source of strength.

I am also grateful for the wonderful relationships I have made at UNIST. Special thanks to Taehoon Kim, Gangmin Choe, Youngjin Shim, Heejin Choi, Joonseo Lee, Joonhyuk Choi, Minji Lee, Jeongsik Yoon, and Jaesung Jeon. Their friendship and support have made my life in the Master's program truly enriching and enjoyable.

Finally, I am deeply grateful to my family, Jeongyong Park, Hyejeong Jeong, and Jihoon Park, for their endless support, encouragement, and love. Their unwavering belief in me has been a constant source of motivation and comfort.

To all of you mentioned and to everyone reading this, for your contributions and support throughout this journey, I express my heartfelt gratitude.