专业名称: 数学与应用数学 专业代码: 070101
所属学科门类及专业类: 理学 数学类 学位授予门类: 理学

College of Science, Mathematics and Technology
Wenzhou-Kean University

温州肯恩大学
WENZHOU-KEAN UNIVERSITY

Multi-Strategy Portfolio Optimization in China's Pharmaceutical and Biotechnology Industry: A Comparative Analysis of Static and Deep RL Methods

Student Name: Weixun Xie
Student ID: 1195179
Major: Mathematical Sciences (Data Analytics Option)
Supervisor: Dr. Ebenezer Atta Mills
Date of Submission: May 17th 2025

Author's Declaration

I, the undersigned, hereby declare that this submission is entirely my own work, in my own words, and that all sources used in researching it are fully acknowledged and all quotations properly identified. It has not been submitted, in whole or in part, by me or another person, for the purpose of obtaining any other credit/grade. I have a citation to AI system used to generate the text if I used AI-generated text. I understand the ethical implications of my research, and this work meets the requirements of College of Science, Mathematics and Technology.

Date: **May 1st 2025**

Name (printed letters): **Weixun Xie**

Signature: Weixun Xie

Abstract

This study offers a dynamic portfolio optimization framework that combines deep reinforcement learning (Deep RL) methods with traditional asset allocation approaches to improve portfolio performance and risk management. The goal of this study is to evaluate the effectiveness of several portfolio optimization approaches, including Equal Weight (EW), Sharpe ratio-based Mean-Variance Optimization, Risk Parity (RP), and Deep RL-based strategies to optimize asset allocation in different market scenarios. Historical stock data from 13 Chinese pharmaceutical and biotechnology companies, covering the period from October 17, 2018 to March 1, 2025, were collected and analyzed to construct and assess optimized portfolios. The results indicate that the Deep RL strategy has greater adaptability and consistently outperforms static strategies, especially in volatile market environments, even though traditional approaches like Sharpe ratio optimization and Risk Parity offer significant improvements in risk-adjusted returns compared to the Equal Weight portfolio. This work shows the possibility of fusing cutting-edge machine learning techniques with well-established financial theories to enhance dynamic portfolio optimization and provides insights into future research in asset allocation and investment management.

Keywords: Risk Parity, Portfolio Optimization, Reinforcement Learning, Pharmeceutical and Biotechnology Industry, Deep Q-Network (DQN), Proximal Policy Optimization (PPO), Volatility Management

Contents

Author's Declaration II

Abstract III

List of Figures VI

List of Tables VII

List of Symbols and Abbreviations VIII

List of Symbols and Abbreviations VIII

1. Introduction 1

2. Literature Review 4

3. Data and Methodology 8
3.1. Data Description and Preprocessing . . . 9
3.1.1 Data Source and Selection . . . 9
3.1.2 Data Cleaning and Transformation . . . 9
3.2. Static Portfolio Optimization . . . 9
3.2.1 Equal Weight Portfolio . . . 9
3.2.2 Sharpe Ratio-based Mean-Variance Optimization . . . 10
3.2.3 Risk Parity Portfolio (Equal Risk Contribution) . . . 11
Conceptual Foundation: Equal Risk Contribution (ERC) . . . 11
Practical Implementation: Risk Parity Optimization . . . 12
3.3. Reinforcement Learning-Based Dynamic Allocation . . . 12
3.3.1 Custom Gym Environment Design . . . 12
3.3.2 Agent Training using PPO with Sharpe Reward . . . 13
3.3.3 Agent Training using DQN . . . 14

4. Results 15
4.1. Static Portfolio Performance Comparison . . . 15
4.1.1 Equal Weight Portfolio . . . 15
4.1.2 Sharpe Ratio-Based Mean-Variance Optimization . . . 17
4.1.3 Risk Parity Portfolio . . . 19
4.2. Dynamic Portfolio Optimization with Deep RL . . . 22
4.3 Result Summary . . . 23

5. Discussion 25
5.1. Reflection on the Effectiveness of the Methods . . . 25
5.2. Evaluation of Strategy Choices . . . 25
5.3. Limitations and Future Work . . . 26

6. Conclusion 27

Acknowledgement 28

Appendix A **32**

List of Figures

1 2-Year Return of Equal Weight Portfolio 16
2 5-Year Return of Equal Weight Portfolio 16
3 Allocation of Sharpe-optimized portfolio 17
4 Risk Contribution Per Asset 18
5 2-Year Return of Sharpe-optimized Portfolio 19
6 5-Year Return of Sharpe-optimized Portfolio 19
7 Allocation of Risk Parity portfolio 20
8 Risk Contribution Per Asset 20
9 2-Year Return of Risk Parity Portfolio 22
10 5-Year Return of Risk Parity Portfolio 22
11 Cumulative Portfolio Value Achieved by the PPO Agent 23

List of Tables

1 Cumulative Returns of Equal Weight Portfolio (in RMB) 15
2 Optimized Weights and Returns under Sharpe Ratio Objective . . . 18
3 Cumulative Returns of Risk Parity Portfolio (in RMB) 21

List of Symbols and Abbreviations

Symbols and Abbreviations	Definition
R&D	Research and Development
RL	Reinforcement Learning
MVO	Mean–Variance Optimization
ERC	Equal Risk Contribution
DRL	Deep Reinforcement Learning
DQN	Deep Q-Network
N	Number of risky assets in the portfolio
w_i	Weight of asset i in the portfolio
$\sum_{i=1}^{N} w_i = 1$	Budget constraint: weights sum to 1
r_i	Expected return of asset i
$E(r_p)$	Expected return of the portfolio
σ_{ij}	Covariance between asset i and j
σ_p	Standard deviation (risk) of the portfolio
R_f	Risk-free rate
Sharpe ratio	$(E(r_p) - R_f)/\sigma_p$
$\mathbf{w}$	Portfolio weight vector
$\boldsymbol{\mu}$	Vector of expected asset returns
$\boldsymbol{\Sigma}$ or Σ	Covariance matrix of asset returns
$\sigma(x)$	Portfolio standard deviation as function of weights x
$\frac{\partial \sigma(x)}{\partial x_i}$	Marginal risk contribution of asset i
$\sigma_i(x)$	Total risk contribution of asset i
TRC_i	Total risk contribution of asset i: $w_i \cdot \frac{(\Sigma \mathbf{w})_i}{\sigma_p}$
$\prod_{t=1}^{T}(1 + r_t)$	Cumulative return formula over T days

1. Introduction

As a typical representative of high technology and high investment, China's pharmaceutical biotechnology industry has developed rapidly in recent years due to its innovation drive and policy support. However, the industry also has high volatility and systemic risks: on the one hand, the long R&D cycle, high investment cost and low success rate of new drugs make the performance of individual stocks vary greatly; on the other hand, changes in industry policies, approval processes and international competition patterns may cause significant market fluctuations. Related research shows that although biotechnology companies have high-risk characteristics, their risk-adjusted returns are often better than the market average, especially large and mature companies can provide better downside protection during crises [1]; while early-stage companies face greater volatility due to their high dependence on the capital market [2]. In this context, how to build a robust investment portfolio that can adapt to market changes has become an important issue that needs to be urgently addressed by academia and practitioners.

Based on this, the following are the definitions of five key terms used in this study: Portfolio optimization aims to select or adjust asset weights based on the trade-off between risk and return, so as to maximize the expected return under a given risk level, or minimize the portfolio risk under a given return target. The traditional method is based on the mean-variance model proposed by Markowitz (1952) [3], and solves the optimal weight allocation through mathematical programming. The Equal-Weight(EW) strategy allocates investable funds equally to each asset without any return or risk parameter estimation. Its advantages are simplicity and robustness to historical estimation errors, but it ignores the volatility and correlation differences between assets. Sharpe ratio was proposed by Sharpe (1966), it is defined as the ratio of the excess return of a portfolio to its standard deviation, which is used to measure the risk-adjusted return under unit risk [4]. The Sharpe ratio can be used as a performance evaluation indicator or directly incorporated into the optimization goal to maximize risk-adjusted returns. The risk parity strategy is based on the concept of "equal risk contribution" and adjusts asset weights to make the marginal contribution of each asset to the overall volatility of the portfolio equal. Unlike mean-variance optimization that relies on expected returns, risk parity only relies on asset volatility and covariance, which improves the robustness of the configuration. Deep reinforcement learning combines deep neural networks with reinforcement learning frameworks to autonomously adjust asset weights through interactive learning with the market environment. In recent years, research on dynamic asset allocation based on algorithms such as DQN and PPO has developed rapidly, providing new ideas for portfolio optimization under non-stationary market conditions.

China's pharmaceutical biotechnology industry is characterized by high R&D investment, high policy sensitivity, and high market volatility. In recent years, with the acceleration of the country's approval process for innovative drugs and strong support for the biopharmaceutical industry, the risk-return structure of this industry is significantly different from other sectors: on the one hand, leading companies often obtain excess returns during periods of technological breakthroughs and policy dividends; on the other hand, early projects may face drastic draw-

downs due to clinical failures and approval delays. Traditional single strategies are difficult to balance stability and opportunity capture at the same time, so there is an urgent need to systematically compare multiple static and dynamic optimization methods to determine the optimal configuration solution in the context of this special industry.

This study aims to provide investors and asset management institutions with a multi-dimensional decision-making reference by comparing four strategies: equal weighting, Sharpe ratio optimization, risk parity, and deep reinforcement learning. First, it can help institutions choose methods that can both control downside risks and capture innovation dividends in a high-volatility environment; second, by combining dynamic rebalancing with risk budgeting, it can enhance the portfolio's ability to respond to sudden policies or market events; finally, the conclusions and empirical results of this study will provide guidance for product design, risk management, and configuration process optimization, and inject new theoretical and technical support into the asset allocation practice of the pharmaceutical and biological sector.

Based on the above research background and strategy comparison, this study proposes the following testable research hypotheses: In China's pharmaceutical and biotechnology industry, the long-term risk-adjusted returns of the risk parity strategy are better than those of equal-weighted and Sharpe–MVO; the deep reinforcement learning strategy outperforms all static strategies in terms of dynamic rebalancing at market turning points. Based on the above assumptions, this study will next review the main limitations of the classical portfolio optimization framework and its derivative methods to lay the foundation for subsequent method improvements.

Since Markowitz (1952) proposed the mean-variance portfolio theory, this framework and its derivative methods have become the cornerstone of asset allocation [5]. However, mean-variance optimization is highly sensitive to the estimation of expected returns and covariance matrices, and is easily affected by parameter errors, resulting in drastic fluctuations in portfolio weights [6]. To enhance robustness, Glasserman and Xu [7] constructed a mean-variance framework that takes model risk into account, and Baviera and Bianchi conducted in-depth discussions on analytical solutions in extreme cases [8]. Another simple and commonly used benchmark method is the equal-weight strategy, which does not require any historical parameter estimation, has good practicality and robustness, and is often better than the overfitted optimal combination in a variety of market environments [9]. However, the equal-weight method does not consider the volatility differences and correlations between assets, which may lead to excessive exposure to high-volatility stocks. In response to the inadequacy of risk measurement and return evaluation, Sharpe (1966) proposed the Sharpe ratio to measure the excess return under unit risk[4]; Auer and Schuhmacher further explored the hypothesis testing method based on the Sharpe ratio [10]. In order to enhance the robustness of the Sharpe ratio in extreme market environments, Deng et al. proposed the VaR-adjusted Sharpe ratio (VaRSR), which controls tail risk by maximizing risk-adjusted returns in the worst case [11]. In order to reduce the reliance on return expectations, the risk parity strategy has received widespread attention. Lee, risk parity relies only on volatility and covariance estimates, thereby improving

the robustness of the portfolio [12]; Kazemi systematically expounded the basic principles of risk parity in the CAIA report [13]. Maillard et al. proposed the concept of "Equal Risk Contribution" (ERC) portfolio, which provides a solid theoretical basis for risk parity and empirically verifies its superiority in risk diversification and risk budgeting [14]. Chaves et al. compared the performance of risk parity with other heuristic strategies and developed two efficient iterative algorithms to calculate ERC weights, thus avoiding the problem of nonlinear optimization [15, 16]. However, traditional risk parity is still a static rebalancing method, which makes it difficult to capture rapid changes in market structure. In recent years, Deep Reinforcement Learning (DRL) has been introduced into the research of dynamic portfolio adjustment because of its ability to learn environmental feedback in real time and adjust strategies. Gao et al. applied DQN to the discrete decision-making of stock weights, and combined CNN with Dueling architecture to outperform various traditional strategies [17]; Pawar et al. and Espiga-Fernández et al. compared the performance of various DRL agents (DQN, PPO, SAC, etc.) under different market signals and rebalancing frequencies, and found that PPO and SAC have stable performance in most scenarios [18, 19]. Huang et al. proposed a PPO dynamic asset allocation framework based on the Stable-Baselines3 platform, using the Sharpe ratio as the return function, and achieved adaptability and return performance that is incomparable to traditional static methods [20] in a custom Gym environment.

Despite the outstanding research results mentioned above, there is still a lack of systematic comparative research on the high-risk, high-return market segment of China's pharmaceutical and biological industry in the academic community. The horizontal comparison of various static methods and DRL dynamic strategies in this industry has not yet been carried out, and there is also a lack of customized models for industry characteristics (such as high R&D investment cycle and policy risks). To this end, this study will: Based on the three classic static methods of equal weight, Sharpe–MVO, and Risk Parity, a DRL dynamic position adjustment strategy is systematically constructed; take 13 companies in China's pharmaceutical and biological industry as samples, cover market data from October 17, 2018 to March 1, 2025, and conduct annual rebalancing backtests; through multi-dimensional indicators such as risk-adjusted returns, maximum drawdowns, and risk contribution distribution, the performance of each strategy in different market environments is deeply evaluated.

Based on this, the structure of this study is as follows: the second chapter of this article reviews the relevant literature; the third chapter introduces the data sources and methodology; the fourth chapter presents the empirical results of the static and dynamic strategies; the fifth chapter discusses the research findings and application implications; the sixth chapter summarizes the entire article and proposes future research directions.

2. Literature Review

Traditional portfolio optimization methods are mostly static strategies, which usually rely on historical estimates of expected returns and covariances of assets, and assume the relationship between assets remains stable throughout the investment period. However, real-world financial markets are highly dynamic. Unbalanced risk exposures and poor performance are the results of static strategies' frequent inability to adjust, particularly during periods of structural change or increased volatility.

The Equal Weight (EW) technique is one of these static approaches that has become popular because of its resilience and simplicity of usage. It is among the simplest, giving every asset the same amount of capital. It avoids the sensitivity to parameter estimation errors that plagues more sophisticated optimization techniques. But it fails to account for cross-asset differences in volatility or return behavior.

The foundational Mean-Variance Optimization (MVO) model proposed by Markowitz [3, 5] and extended by Sharpe [21] remains central to modern portfolio theory. However, the sensitivity of MVO to input parameters, particularly expected returns and covariance estimates, has been widely criticized. Kim, Kim, and Fabozzi [6] pointed out that the stability of optimized portfolios is easily influenced by estimation error. Glasserman and Xu [7] developed a framework to quantify model risk in MVO, while Baviera and Bianchi [8] derived analytical solutions for worst-case optimal allocations when the mean vector is constrained.

Sharpe ratio [4] is another widely used risk-adjusted performance measure. However, as Bailey and López de Prado [10] emphasize, under a non-normal return distribution, its statistical inference may be unreliable. Geng Deng et al. [11] further highlight the fragility of Sharpe ratio estimations derived from sparse historical data. They suggested a Value-at-Risk adjusted Sharpe ratio (VaRSR) to improve robustness, which maximizes the worst-case Sharpe ratio at a specified confidence level.

W. Lee [12] promotes risk-based allocation techniques like Risk Parity (RP), which only use volatility and covariance, as a solution to MVO's dependence on return estimations. Kazemi [13] offers a thorough analysis of RP in which the risk of the whole portfolio is distributed equally across all assets. Maillard, Roncalli, and Teiletche [14] proposed the Equal Risk Contribution (ERC) concept and proved that the volatility of the ERC portfolio is between the minimum variance portfolio and the EW portfolio, and offers a balanced risk budget property.

While asset selection sensitivity is still a concern, Chaves, Hsu, Li, and Shakernia [15] compare RP with alternative heuristics such as 60/40 splits and minimum variance portfolios and find that RP strategies tend to offer better risk diversification and more stable performance across market regimes. The same authors [16, 22] provide iterative strategies for calculating RP weights in a related study, which are more efficient than traditional nonlinear optimization techniques. In order to achieve structured risk diversification, Raffinot [23] more recently suggested the Hierarchical Equal Risk Contribution (HERC) model, which combines risk parity with hierarchical clustering. However, these approaches are not the main focus of this study.

Qiu, Chen, Lu, Hu, and Wang [24] investigated how private investment and public funding in China's pharmaceutical R&D. Their results highlight how important public funding is for promoting private capital involvement and assisting certain sub-industries. In order to determine the fundamental factors and suggest appropriate investment plans, Yuanjia, Ung, Ying, and Yitao [25] carried out a methodical examination of the structure and regional distribution trends of the Chinese pharmaceutical market. According to Xu, Wang, and Liu [26], government subsidies greatly boost enterprises' R&D investment, but they have no direct effect on innovation performance. The study also demonstrates that the technical backgrounds of CEOs and business ownership act as mitigating factors in this connection.

According to Bruneo, Giacomini, Iannotta, Murthy, and Patris [1], biotechnology companies outperform the market as a whole and other R&D-intensive businesses on a risk-adjusted basis, although typically exhibiting more risk than the market average. In particular, large-cap biotechnology firms can reduce downside risk and offer diversification benefits during difficult financial times. Conversely, smaller and startup biotech businesses are more likely to be financially vulnerable and show a greater dependence on capital markets.

An extensive study of portfolio optimization in early-stage drug development in the biotechnology and pharmaceutical sectors was carried out by Badwe [2]. The study emphasizes that typical portfolio optimization techniques may not be appropriate for early-stage R&D due to its high level of uncertainty and substantial failure risk. To solve this, the author created a novel framework combining strategic planning with quantitative analysis to better manage risk allocation and resource allocation. This framework aims to maximize the expected value of the overall R&D portfolio by taking into account aspects including project deadlines, R&D costs, potential market value, and the probability of technical success.

An innovative investment framework designed especially for the biotechnology industry was put out by Mohan and Roy [27]. Their model improves return efficiency in biotech markets by combining asset class diversification and dynamic asset allocation. The authors created a flexible investing approach by choosing representative companies that followed market movements, examining financial data, and assessing financial health. In order to maximize the model's applicability in professional biotech investment situations, they also evaluated the influence of selected strategies and sectoral feasibility.

A dynamic asset allocation model that takes time-varying expected returns into account was presented by Brennan, Schwartz, and Lagnado [28]. According to their research, long-term investors' optimal asset allocation is significantly different from tactical, short-term investors. To take advantage of the mean-reverting nature of asset returns, long-term investors are especially likely to devote a bigger percentage of their portfolios to stocks and bonds. The study goes on to show that taking expected return dynamics into consideration can significantly enhance portfolio performance over time.

Bruder and Roncalli [29] formalize the theory of risk budgeting, enabling precise control over risk contributions through predefined allocations. Alviniussen and Jankensgård [30] further extend the risk budgeting framework into enterprise financial planning, highlighting its role in enhancing transparency and forward-

looking decision-making.

In the meantime, Bruder and Roncalli [29] emphasized the importance of risk budgeting as a foundation for risk parity strategies. Their methodology improves overall balance by providing more exact control over the risk distribution at the portfolio level by predefining each asset's risk contribution. The idea of risk budgeting was expanded to enterprise risk management by Alviniussen and Jankensgård [30], who highlighted how it integrates financial planning and enhances transparency and forward-looking decision-making. Robert C. Merton suggested calculating the projected market return by summing the present observable risk-free rate with the historical average excess market return. To increase the economic plausibility of the estimates, he also underlined that models should specifically include the non-negativity restriction on expected excess returns [31].

Additionally, Bhansali et al. [32] critiqued traditional asset-class-based risk parity strategies for their inability failing to provide balanced exposure to underlying risk factors. They argue that such strategies may lead to performance that is inconsistent throughout economic cycles. The authors propose factor-based risk parity as a solution to this limitation. This approach distributes risk contributions according to basic economic drivers instead of asset class labels, fostering a more resilient and stable asset allocation framework.

Wen, Cao, Liu, and Wang [33] documented time-varying volatility spillover effects across Chinese financial markets, noting that market comovements enhance following significant crisis occurrences. The non-ferrous metals and chemical industries have an enormous effect on the Chinese stock market, which generally receives a net amount of volatility spillovers. There has been a noticeable increase in the spillover from the stock market to the commodities market since the start of COVID-19.

The potential of machine learning (ML) techniques to enhance asset allocation choices was highlighted by Routledge [34] as investors deal with growing amounts of structured and unstructured data. An ML-based portfolio approach was presented by Hong et al. [35] to handle optimization difficulties that are both static and dynamic, especially for high-dimensional asset universes. Their approach reduces estimate risk by avoiding the necessity for explicit covariance estimation by using machine learning regression models to anticipate future returns. According to empirical findings, their methodology outperformed conventional optimization methods in terms of excess returns in the Chinese and American A-share markets.

For capital allocation issues under uncertainty, reinforcement learning (RL) has also drawn interest. Jakobsen and Bang [36] presented RL applications in consumption planning and sovereign wealth fund distribution, emphasizing its benefits in unstable economic conditions.

Early portfolio management uses of RL include O et al. [37], who dynamically adjusted asset allocations in the Korean stock market using Q-learning and numerous pattern-based predictors. This study focuses on deep reinforcement learning (Deep RL) approaches, including Deep Q-Network (DQN) and Proximal Policy Optimization (PPO), which have exhibited greater representational power and convergence qualities, even though their work showed significant performance increases over static methods.

In a long-short portfolio management context, Pawar et al. [18] used Ad-

vantage Actor-Critic (A2C) and DQN strategies. Comparing their performance to traditional techniques, they provide better risk-adjusted returns. Similarly, by discretizing the action space to represent asset weight vectors and improving learning using CNN and Dueling Q-Network architectures, Gao et al. [17] modified DQN for portfolio optimization. Espiga-Fernández et al. [19] systematically evaluated multiple DRL agents (DQN, DDPG, PPO, SAC) with different rebalancing frequencies, market inputs, and historical window lengths. Their results highlight the resilience of DRL agents, particularly over extended lookback periods, when outfitted with feature-extraction networks like CNNs.

A dynamic asset allocation approach based on Deep RL was suggested by Huang, Zhou, and Song [20] in a recent benchmark research. It introduced an Actor-Critic architecture with a reward function based on the Sharpe ratio. Without changing the core algorithm structure, the authors optimized policy learning using a Gym-style environment and the Stable-Baselines3 (SB3) framework for PPO deployment. Their findings support the idea that reward settings may be tailored to improve risk-adjusted portfolio returns. Motivated by their methods, this study adopts a similar approach: the agent is trained using SB3's PPO implementation with a reward signal based on Sharpe ratio increases to ensure algorithmic consistency and reproducibility.

In RL-based financial applications, using custom gym environments has become standard design practice. Jerome et al. [38] presented mbt-gym, a Python module for building RL environments designed for market creation and optimum execution, two tasks related to limit order book trading. The fundamental design principles of environment construction are still the same even though the domain is different from this study.

In addition, Chou, Kuo, and Jiang [39] proposed a portfolio optimization model based on trend ratio and evolutionary computation. By providing an alternative objective function more appropriate for trend-following strategies, their study addresses the propensity of traditional Sharpe ratio metrics to over-penalize low-volatility portfolios with steady upward trends.

Although the literature on portfolio optimization has been greatly enhanced by recent developments in Deep Reinforcement Learning and other risk-reward evaluation measures like the Trend Ratio, a number of difficulties still exist. The majority of current research concentrates on certain asset classes or simulated settings, underexamining issues of resilience and adaptation in the actual world. Furthermore, cross-study comparisons are challenging since many implementations lack consistent assessment procedures.

Building on these discoveries, this study suggests an integrated framework that blends dynamic DRL-based techniques with conventional static tactics (such as Equal Weight, Mean-Variance, and Risk Parity). By doing this, it seeks to solve the shortcomings of existing models in terms of adjusting to changing market conditions, particularly in high-volatility industries like the biotechnology and pharmaceutical industries in China. The suggested method adds a fresh viewpoint to the expanding corpus of research on intelligent asset allocation by emphasizing not only return maximization but also interpretability and practical risk control.

3. Data and Methodology

Portfolio optimization has long been a fundamental problem in quantitative finance. Traditional static strategies, derived from the Modern Portfolio Theory (MPT) proposed by Markowitz in 1952[5], rely on the trade-off between expected returns and portfolio variance. Under this framework, investors allocate capital among N risky assets with weights $w_1, w_2, \ldots, w_N$ such that $\sum_{i=1}^{N} w_i = 1$.

$$\sum_{i=1}^{N} w_i = 1 \quad \text{and} \quad 0 < w_i < 1 \tag{1}$$

The expected return of the portfolio is defined as the weighted sum of individual asset returns:

$$E(r_p) = \sum_{i=1}^{N} w_i r_i \tag{2}$$

where r_i denotes the expected return of asset i. The risk of the portfolio is measured by standard deviation, which is derived from the variance of returns, including asset variances and covariances:

$$\sigma_p = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{3}$$

Here, σ_{ij} represents the covariance between asset i and j, and σ_p is the total portfolio risk. This formulation illustrates the diversification effect: by combining assets that exhibit imperfect correlation, the overall portfolio risk may be diminished below the mean of individual asset risks.

Building on this foundation, Sharpe[39] proposed a risk-adjusted performance metric defined as:

$$\text{Sharpe ratio} = \frac{E(r_p) - R_f}{\sigma_p} \tag{4}$$

where $E(r_p)$ denotes the expected return of the portfolio, R_f represents the risk-free rate, and σ_p is the standard deviation of portfolio returns. In actuality, better risk-adjusted performance is shown by a higher Sharpe ratio. Sharpe ratio portfolios are often preferred by investors because they offer better returns per unit of risk absorbed. Despite its widespread usage, the Sharpe ratio is known to be sensitive to estimate mistakes in inputs like covariances and anticipated returns[6, 11]. Since then, a number of extensions and alternative risk measures have been developed to compensate for these limitations, including robust estimators[8] and Value-at-Risk adjusted Sharpe ratios[11].

In response to the over-reliance on return forecasts in mean-variance frameworks, risk-based allocation strategies such as Risk Parity (RP) have emerged[12, 14]. RP focuses on balancing the marginal risk contribution of each asset rather than maximizing returns, thus providing a more robust solution in uncertain market conditions. In order to adaptively distribute weights in dynamic market situations, recent developments have also included algorithmic techniques like deep

reinforcement learning (DRL)[20]. The dataset is presented in the following sections, along with the methodology used to evaluate both static and dynamic portfolio optimization strategies.

3.1. Data Description and Preprocessing

3.1.1 Data Source and Selection

The dataset consists of daily closing prices for 13 stocks from the Chinese pharmaceutical and biological sector, spanning the period from October 17, 2018 to March 1, 2025. The data was collected from investing.com and forms the basis for both the static and dynamic allocation methods analyzed in this study.

3.1.2 Data Cleaning and Transformation

Daily returns were calculated from the closing price series. Non-trading days were removed, and assets with excessive missing values were excluded. All returns were aligned into a unified time index and checked for stationarity. For reinforcement learning, input features were standardized to ensure numerical stability during training.

3.2. Static Portfolio Optimization

3.2.1 Equal Weight Portfolio

The equal weight (EW) portfolio assigns same weights to all assets in the investment universe [9]. This spreads capital uniformly across the portfolio, even if some assets are more volatile, more correlated, or have a higher expected return than others. This allocation method can be expressed as:

$$w_i = \frac{1}{N}, \quad \forall i = 1, 2, \ldots, N \tag{5}$$

where N denotes the number of assets in the portfolio, and w_i is the weight of asset i. In this study, 13 stocks are selected, and each is assigned a fixed weight of $\frac{1}{13} \approx 0.076923$, forming a strictly equal-weighted portfolio.

To compute the cumulative return over a specified horizon (e.g., 5 years), this study adopts the standard compounded return formula:

$$R = \left(\prod_{t=1}^{T} (1 + r_t) \right) - 1$$

where r_t denotes the daily return on day t, and T is the total number of trading days within the selected horizon. Daily return data were extracted for each stock, and the cumulative return was computed over rolling 2-year and 5-year windows.

The equal weight (EW) portfolio remains a widely used benchmark in empirical finance due to its simplicity, robustness, and independence from return or

risk forecasts. In addition to performing consistently out-of-sample and avoiding overfitting, it provides a simple but efficient method of diversification.

However, this approach also presents critical limitations. It disregards differences in individual asset risk and ignores correlation structures; therefore, it may lead to an overabundance of exposure to volatile assets [9]. These shortcomings highlight the need for more advanced allocation methods. One such method is the Equal Risk Contribution (ERC) approach, developed by Maillard, Roncalli, and Teiletche [14], which distributes weights based on the idea that all assets should contribute equally to the total portfolio risk.

3.2.2 Sharpe Ratio-based Mean-Variance Optimization

This study employed the Sharpe ratio as the objective function in the mean-variance portfolio optimization, following its original definition by Sharpe [4]. The Sharpe ratio is defined as the ratio of excess portfolio return to its standard deviation. It is widely accepted as a fundamental metric for evaluating risk-adjusted performance and forms the basis for many portfolio evaluation methods in financial research.

$$S = \frac{\bar{r}_p - r_f}{\sigma_p} \tag{6}$$

This study also used Auer Schuhmacher's hypothesis testing approach [10] to validate the statistical relevance of Sharpe-based strategies. Furthermore, the Value-at-Risk-adjusted Sharpe ratio framework introduced by Deng et al. [11] is acknowledged as a more conservative extension that accounts for downside risk in volatile markets.

To construct an optimal portfolio under this criterion, the Sharpe ratio optimization problem is posed as follows:

$$\begin{aligned} \max_{\mathbf{w}} \quad & \frac{\mathbf{w}^\top \boldsymbol{\mu}}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}} \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1, \\ & w_i \geq 0 \quad \forall i \end{aligned} \tag{7}$$

where $\mathbf{w}$ denotes the vector of portfolio weights, $\boldsymbol{\mu}$ denotes the estimated expected return vector, and $\boldsymbol{\Sigma}$ denotes the sample covariance matrix of asset returns. The optimization seeks to maximize the portfolio's expected return per unit of risk while ensuring full investment and non-negativity of asset positions.

In this study, historical daily returns were used to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The optimization was implemented using Python's `Riskfolio-Lib` library, with the objective configured to 'Sharpe Ratio' under the mean-variance (MV) model. To ensure practical relevance, no leverage or short-selling was permitted. The generated portfolio allocations can be compared to equal-weight and reinforcement learning-based strategies using a risk-adjusted benchmark.

The optimized asset weights were applied to historical price data to estimate cumulative returns over both two -year and five-year horizons in order to empirically evaluate the performance of the Sharpe ratio-based mean-variance portfolio. The resulting returns were then compared to those of an equally weighted portfolio, establishing a framework for evaluating the efficacy of risk-adjusted optimization.

The analysis reveals that the Sharpe-optimized portfolio consistently achieved higher cumulative returns across both periods, suggesting that portfolios constructed via Sharpe ratio maximization may offer superior performance relative to naive diversification. These findings highlight the practical value of risk-adjusted return optimization, particularly when estimation inputs are reasonably stable and investment horizons are of moderate length.

3.2.3 Risk Parity Portfolio (Equal Risk Contribution)

Conceptual Foundation: Equal Risk Contribution (ERC)

The Equal Risk Contribution (ERC) portfolio framework, developed by Maillard, Roncalli, and Teiletche [14], offers a more thorough theoretical foundation for EW. According to this principle, the portfolio is built up so that every asset contributes equally to the total portfolio risk, which is frequently determined by the standard deviation of the portfolio under the variance-covariance structure. The total portfolio risk is given by the standard deviation of the weighted asset returns:

$$\sigma(x) = \sqrt{x^\top \Sigma x} \tag{8}$$

To compute how each asset contributes to portfolio volatility, the marginal risk contribution (MRC) with respect to weight x_i should be derived:

$$\frac{\partial \sigma(x)}{\partial x_i} = \frac{x_i \sigma_i^2 + \sum_{j \neq i} x_j \sigma_{ij}}{\sigma(x)} \tag{9}$$

The individual asset's total risk contribution (RC) is then computed as the product of its weight and marginal contribution:

$$\sigma_i(x) = x_i \cdot \frac{\partial \sigma(x)}{\partial x_i} \tag{10}$$

The ERC condition requires that all risk contributions be equal, i.e., $\sigma_i(x) = \sigma_j(x)$ for all i, j. This results in a nonlinear system of equations that are usually solved using iterative optimization; this method is thought to be more resilient in uncertain environments since it just depends on the covariance structure of returns rather than return forecasts.

Practical Implementation: Risk Parity Optimization

Traditionally, portfolio construction has relied on Markowitz's mean-variance optimization (MVO), which is unstable because of its sensitivity to expected return estimations. Accurate forecasting of these inputs is notoriously challenging, and even small errors can cause significant variations in the weights of the portfolio. Risk-based allocation techniques like Risk Parity (RP) have become strong substitutes for this limitation [12].

By attempting to allocate portfolio weights so that each asset contributes equally to the overall portfolio risk, risk parity techniques reposition the attention from expected returns to risk contributions. By relying on volatility and covariance, which are more stable and observable inputs [13, 15], RP removes the need to estimate expected returns.

Formally, the total risk contribution (TRC) of asset i in a portfolio is defined as:

$$\mathrm{TRC}_i = w_i \cdot \frac{(\Sigma \mathbf{w})_i}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}} \tag{11}$$

where Σ denotes the covariance matrix of asset returns and $\mathbf{w}$ is the weight vector. Risk parity aims to equalize these contributions across all assets:

$$\mathrm{TRC}_i = \mathrm{TRC}_j \quad \forall i, j \tag{12}$$

This condition can be equivalently stated as:

$$w_i \cdot (\Sigma \mathbf{w})_i = w_j \cdot (\Sigma \mathbf{w})_j \quad \forall i, j \tag{13}$$

Solving this system of nonlinear equations directly is computationally demanding. To address this, Chaves et al. [16] introduce two efficient iterative algorithms: a Newton method and a Power method. Both approaches iteratively adjust weights until the equality of risk contributions is satisfied, without requiring complex nonlinear solvers.

RP portfolios typically choose lower-volatility assets, such as bonds, over heuristic portfolios (e.g., equal weight or minimal variance). In order to meet return targets, leverage is frequently required, which brings about practical implementation considerations [13]. However, RP portfolios demonstrate greater resilience across market regimes, risk-adjusted performance, and diversification [12, 15].

Despite these advantages, standard asset-based risk parity may lead to imbalances in exposure to underlying economic risk factors. As highlighted by Hsu et al. [22], RP may fail to achieve true diversification when asset classes share similar factor loadings. Factor-based risk parity offers an extension to mitigate this issue, though it is not explored further in this study.

3.3. Reinforcement Learning-Based Dynamic Allocation

3.3.1 Custom Gym Environment Design

This study adopts a reinforcement learning-based framework to dynamically adjust portfolio weights over time. To facilitate agent training, a custom trading

environment was developed following the OpenAI Gym interface specification. The environment simulates the evolution of asset prices and portfolio rebalancing, and enables interaction through discrete time steps.

At each step, the agent observes a state vector composed of historical price-based features and current portfolio weights. The action space is defined as the continuous allocation vector representing asset weights, subject to constraints on full investment and non-negativity. When an action is performed, the environment computes a reward signal and uses asset returns to update the portfolio value.

The reward is computed based on the Sharpe ratio of the portfolio over a rolling window, allowing the agent to optimize for risk-adjusted returns. Instead of focusing on maximazing raw profits, its design promotes learning stable allocation techniques that strike a compromise between return and volatility. Instead of modifying the agent design, the reward structure is included into the environment logic to ensure the algorithmic compatibility.

This modeling strategy aligns with recent practices in deep reinforcement learning for financial decision-making. For example, Huang et al. [20] trained PPO agents using a Sharpe ratio-based reward within a custom Gym environment using the Stable-Baselines3 (SB3) framework. Their findings show that decoupling environment customisation from algorithmic implementation improves stability and repeatability. In the same way, this study uses SB3's standardized training pipeline while maintaining the core PPO structure, improving transparency, and making comparative analysis easier.

3.3.2 Agent Training using PPO with Sharpe Reward

Following the environment-agent separation principle, this study implemented PPO via the Stable-Baselines3 (SB3) framework without modifying its underlying algorithmic structure. The reward function—based on the Sharpe ratio over a rolling window—was integrated within the environment to guide the agent toward risk-adjusted performance maximization. This setup ensures that model improvements are motivated by financial goals while maintaining reproducibility and compliance with standardized training techniques.

The methodological underpinning for this technique comes directly from Huang et al. [20], who developed a similar DRL framework for portfolio optimization. Without requiring fundamental modifications to the learning algorithm, their study demonstrates the efficacy of employing SB3 to train PPO agents with Sharpe-based incentives in Gym-compatible contexts, leading to greater returns and risk profiles. Inspired by this model, the agent in this work learns optimum allocation policies through repeated interaction, eventually strengthening its capacity to adapt to changing market conditions.

The use of the Sharpe ratio as the learning signal aligns the optimization objective with traditional financial benchmarks, allowing for uniform evaluation of baseline and learning-based strategies. By leveraging SB3's PPO implementation, the training process benefits from established advantages in stability, clipping-based trust area limitations, and parallel rollout capabilities, all of which are critical for financial environments where sample efficiency and robustness are important.

3.3.3 Agent Training using DQN

To further explore deep reinforcement learning (DRL) methods for portfolio optimization, this study includes the Deep Q-Network (DQN) algorithm as one of the fundamental learning agents. DQN is particularly suitable for discrete action spaces, making it a natural fit when portfolio weights are discretized into finite levels. The agent learns to approximate the ideal action-value function through iterative updates using experience replay and a target network, which stabilizes training in high-dimensional financial environments.

To adapt DQN for portfolio tasks, this study defined the action space as a finite set of weight allocation combinations. Each action represents to a possible portfolio reallocation, allowing DQN to learn weight transitions that maximize long-term reward. The state representation includes lagged returns and technical indicators, and the reward signal is generated from portfolio performance metrics such as periodic Sharpe ratios.

Gao et al. [17] have successfully applied DQN to portfolio management by discretizing allocation weights and strengthening the learning process using convolutional layers and dueling architectures. This supports the implementation design. Their findings showed that DQN-based strategies outperformed traditional benchmarks in terms of cumulative returns, risk-adjusted metrics, and maximum drawdown.

In this study, DQN agents were trained utilizing the Stable-Baselines3-compatible interface. While not the main algorithm for deployment, DQN acted as a standard against which to compare the effectiveness of policy-based techniques like PPO. A more thorough analysis of learning dynamics under various algorithmic assumptions is made possible by this comparative setup.

4. Results

The portfolio optimization methods presented in this study aim to improve the risk-adjusted return of investment strategies while enhancing adaptability to changing market conditions. This section evaluates the empirical performance of both static and dynamic strategies through comparative analysis and simulation-based testing.

4.1. Static Portfolio Performance Comparison

4.1.1 Equal Weight Portfolio

To establish a benchmark for portfolio performance evaluation, this study adopts an Equal Weight (EW) portfolio strategy. Under this approach, an identical amount of capital—RMB ¥10,000—is allocated to each of the 13 selected healthcare stocks, regardless of their individual volatility, correlation, or expected return.

Table 1 presents the cumulative returns over both two-year and five-year horizons, based on historical price data:

Table 1: Cumulative Returns of Equal Weight Portfolio (in RMB)

Stock	**Weight**	**Investment**	**2-Year Return**	**5-Year Return**
Aier	7.69%	10,000	4872	4318
Berry	7.69%	10,000	6847	1592
BGI	7.69%	10,000	9623	5189
Zhifei	7.69%	10,000	3533	2006
Sanjiu	7.69%	10,000	9359	8473
Ejiao	7.69%	10,000	11433	12891
Hengrui	7.69%	10,000	9866	5964
Fosun	7.69%	10,000	7067	5968
Mindray	7.69%	10,000	7374	7619
Vcanbio	7.69%	10,000	10371	7803
AppTec	7.69%	10,000	6214	5618
Yiling	7.69%	10,000	5933	4691
Pientzehuang	7.69%	10,000	9059	11682
Total	100%	130,000	**101551**	**82814**

To visually illustrate the return distribution across individual assets and the portfolio as a whole, Figures 1 and 2 present bar charts for the two-year and five-year horizons, respectively.

Figure 1: 2-Year Return of Equal Weight Portfolio

Figure 2: 5-Year Return of Equal Weight Portfolio

As shown in Table 1, total cumulative returns reached **¥101,551** and **¥82,814** over the 2-year and 5-year periods, respectively. Among the 13 stocks, **Ejiao** and **Pientzehuang** contributed considerably to the overall return of the portfolio by consistently outperforming the others in both short-term and long-term returns. For example, Ejiao produced returns of ¥11,433 and ¥12,891 after two and five years, respectively, but Pientzehuang produced returns of ¥9,059 and ¥11,682. Although these two stocks only made up 15.4% of the capital, they collectively contributed about **20% of the 5-year total return**.

In contrast, stocks such as **Berry** and **Zhifei** performed poorly. Zhifei likewise had a weak return profile, while Berry produced the lowest 5-year cumulative return (¥1,592). This demonstrates the EW strategy's disadvantage: its inability to differentiate between assets that perform well and those that don't.

The difference in performance is further demonstrated in Figures 1 and 2. A

small number of stocks control the overall return in both bar charts, demonstrating that equal capital allocation does not necessarily translate into an equal performance contribution.

The Equal Weight approach ignores asset-specific factors, including volatility, correlation, and projected return, even though it reduces estimation risk and offers stability in volatile situations. Its risk-adjusted performance could thus fall short of more advanced tactics. As discussed in the next sections, risk-aware strategies like the Sharpe ratio or Risk Parity optimization may therefore provide better outcomes.

4.1.2 Sharpe Ratio-Based Mean-Variance Optimization

This section evaluates the performance of a portfolio constructed using Sharpe Ratio-based mean-variance optimization. The optimization objective is to maximize the portfolio's expected return per unit of risk, as measured by the Sharpe ratio, under constraints of full investment and non-negativity of asset weights.

Figure 3 presents the allocation of the Sharpe-optimized portfolio, highlighting the concentration in a few high-performing assets. Figure 4 shows the risk contribution per asset, where a few stocks dominate the overall portfolio risk.

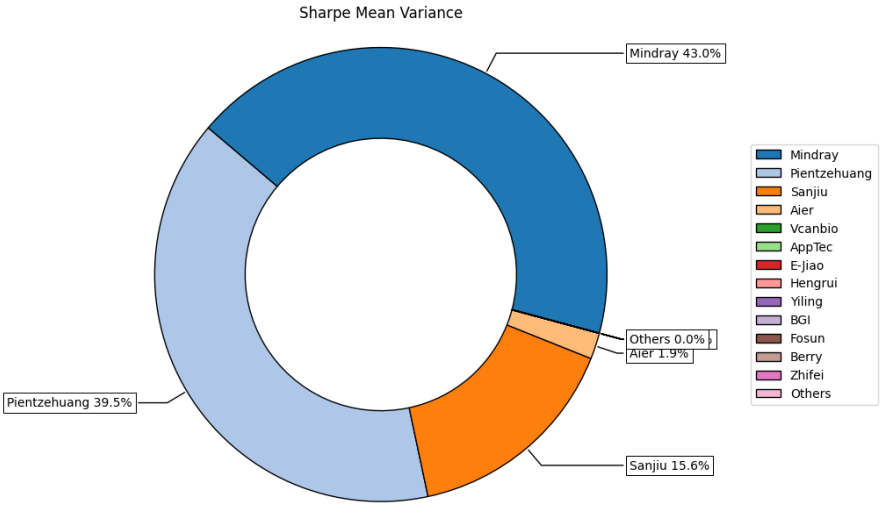

Figure 3: Allocation of Sharpe-optimized portfolio

Table 2 summarizes the optimized asset weights, corresponding investment allocations, and the resulting cumulative returns over the 2-year and 5-year horizons. These results suggest that, relative to an equal-weight strategy, the Sharpe-optimized portfolio offers improved risk-adjusted performance, particularly over the long-term horizon.

To visually illustrate the return distribution across individual assets and the portfolio as a whole, Figures 5 and 6 present bar charts for the two-year and five-year horizons, respectively.

Due to this optimization method, the portfolio became extremely concentrated, with Mindray and Pientzehuang accounting for approximately 83% of the entire

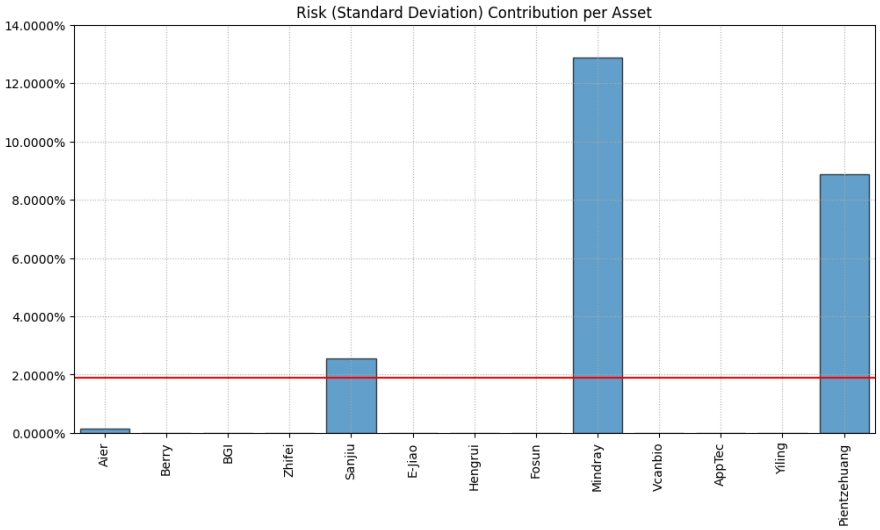

Figure 4: Risk Contribution Per Asset

Table 2: Optimized Weights and Returns under Sharpe Ratio Objective

Stock	**Weight**	**Investment**	**2-Year Return**	**5-Year Return**
Aier	1.87%	2429.05	1183.43	1048.86
Sanjiu	15.62%	20304.96	19003.41	17204.39
Hengrui	42.99%	55891.42	41214.33	42583.67
Pientzehuang	39.52%	51374.57	46540.22	60015.77
Total	**100%**	**130000**	**107941.40**	**120852.70**

capital. The optimizer's preference for assets that have demonstrated stronger Sharpe ratios—high return per unit of risk—is reflected in this. These two stocks had much larger cumulative returns over both 2-year and 5-year periods, as seen in Table 2. Pientzehuang produced a 5-year return of ¥60,015.77, while Mindray produced ¥42,583.67. Figures 5 and 6 graphically validate these findings and show how these assets contribute disproportionately to the success of the whole portfolio.

On the other hand, the allocation structure also brings up useful issues with concentration risk. Even if the return structure is dominated by high-performing assets, the portfolio may be vulnerable to idiosyncratic shocks due to a lack of diversification, particularly in the case of unanticipated negative events that impact a particular company or industry. In practical implementations, this trade-off between preserving diversification and optimizing the Sharpe ratio needs to be carefully evaluated.

From the standpoint of risk management, Figure 4 reveals that the risk contribution is similarly concentrated, especially in Mindray, which is responsible for the majority of the volatility in the whole portfolio. Despite this, the Sharpe-optimized portfolio outperforms the Equal Weight benchmark in terms of cumulative returns,

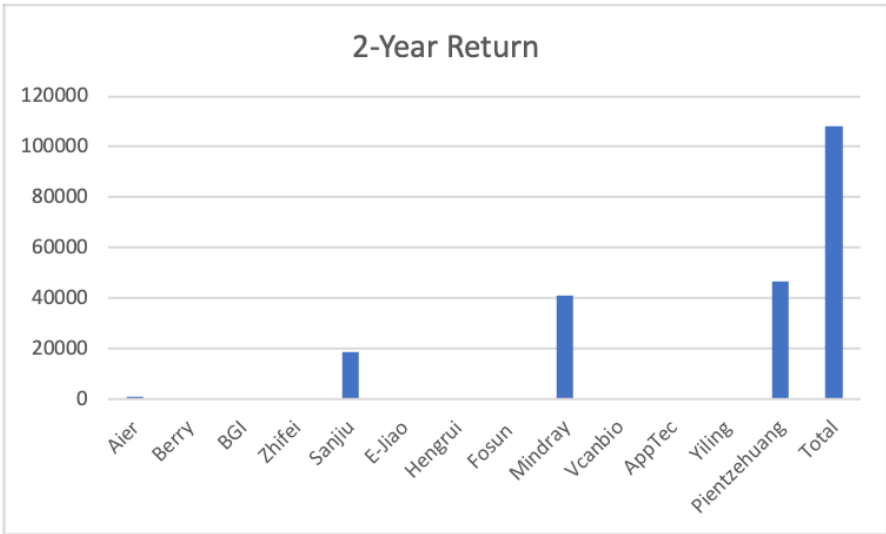

Figure 5: 2-Year Return of Sharpe-optimized Portfolio

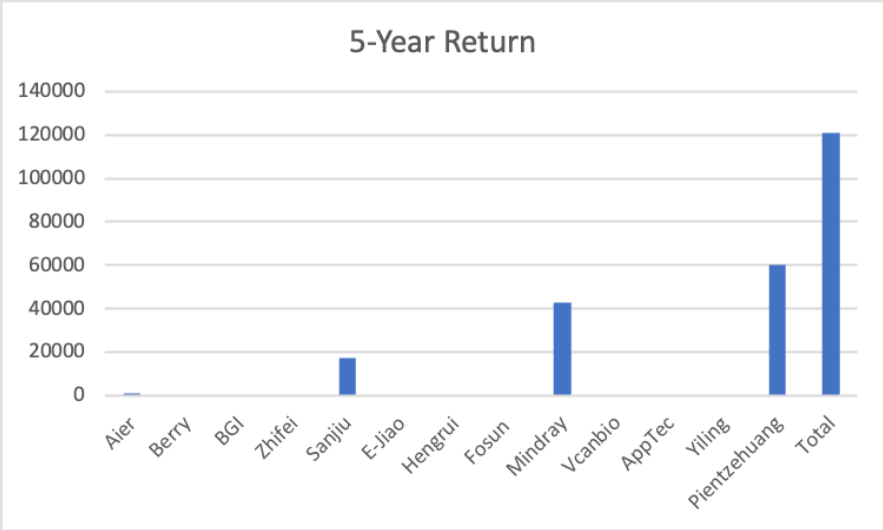

Figure 6: 5-Year Return of Sharpe-optimized Portfolio

achieving ¥120,852.70 over five years, as opposed to ¥82,814 (Table 2 versus Table 1). This demonstrates that in ideal historical circumstances, Sharpe-based optimization, despite its concentration, may result in noticeably better performance.

In summary, although this approach has proven to be highly successful in optimizing past risk-adjusted returns, its dependence on concentrated exposure necessitates the implementation of additional protections or hybrid mechanisms (such as weight caps or risk budgeting constraints) to improve resilience and manage downside exposure in unpredictable market conditions.

4.1.3 Risk Parity Portfolio

This section examines the portfolio constructed using the risk parity approach under the variance-based risk measure. Unlike the Sharpe-optimized portfolio,

which emphasizes return maximization relative to volatility, the risk parity method aims to equalize the contribution of each asset to overall portfolio risk.

Figure 7 presents the asset allocation of the risk parity portfolio. The weights are distributed more evenly across all assets, avoiding concentration in a few high-return but high-volatility stocks. This reflects the core principle of risk parity—balancing risk rather than capital allocation.

Figure 8 visualizes the standard deviation contribution of each asset. The near-uniform height of the bars indicates that risk contributions are approximately equalized, confirming that the portfolio achieves a balanced risk distribution.

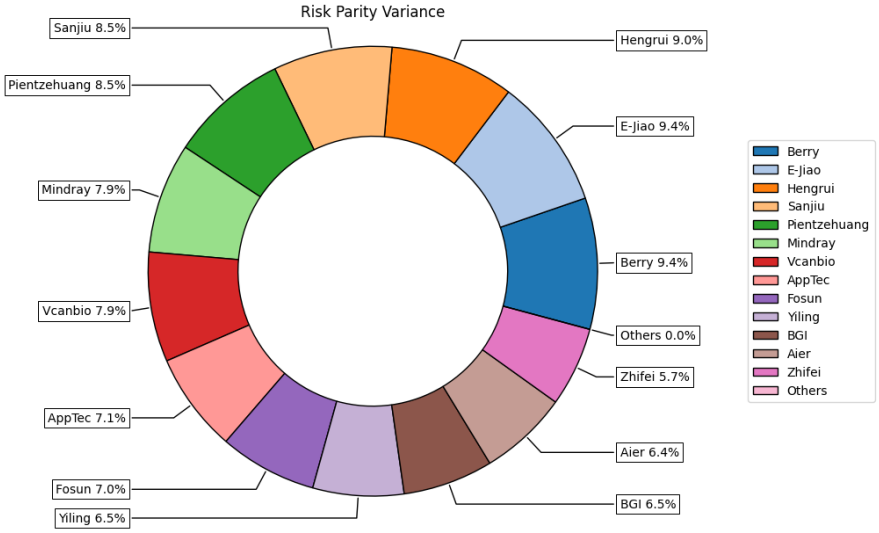

Figure 7: Allocation of Risk Parity portfolio

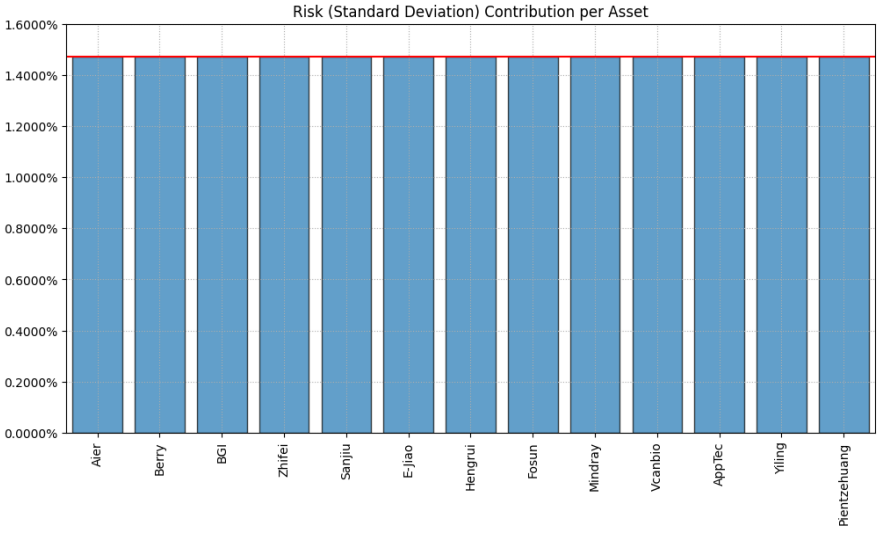

Figure 8: Risk Contribution Per Asset

Table 3 summarizes the asset weights derived from the Risk Parity optimization, along with the corresponding investment allocations and cumulative returns over the 2-year and 5-year horizons. The risk parity approach maintains competitive levels of return while achieving a more equitable distribution of risk across

assets as compared to the equal-weight benchmark. The findings demonstrate how well risk parity works to produce consistent performance and reduce concentration in the volatility contributions of a portfolio.

Table 3: Cumulative Returns of Risk Parity Portfolio (in RMB)

Stock	**Weight**	**Investment**	**2-Year Return**	**5-Year Return**
Aier	6.34%	8240.83	4014.93	3558.39
Berry	9.44%	12271.74	8402.46	1953.66
BGI	6.48%	8430.24	8112.42	4374.46
Zhifei	5.72%	7430.84	2625.30	4190.64
Sanjiu	8.55%	11111.23	10399.00	9414.55
Ejiao	9.49%	12338.24	14106.31	5950.24
Hengrui	8.95%	11639.99	11483.97	6496.02
Fosun	7.02%	9130.55	6452.56	5449.11
Mindray	7.92%	10290.31	7592.01	7843.23
Vcanbio	7.90%	10265.06	10645.89	8009.83
AppTec	7.14%	9276.21	5761.40	7642.71
Yiling	6.54%	8507.49	5047.48	3990.88
Pientzehuang	8.53%	11090.43	10046.83	12955.84
Total	**100%**	**130000**	**104692.78**	**87099.53**

To visually illustrate the return distribution across individual assets and the portfolio as a whole, Figures 9 and 10 present bar charts for the two-year and five-year horizons, respectively.

According to Table 3, the Risk Parity portfolio generated a total return of RMB 104,692.78 over the course of two years and RMB 87,099.53 over five years. These numbers show less concentration in asset-level contributions and exceed the Equal Weight method (RMB 101,551 and RMB 82,814 respectively). For instance, the Risk Parity portfolio distributed weights more fairly, ranging from 5.72% (Zhifei) to 9.49% (Ejiao), in contrast to the Sharpe-optimized portfolio, where Pientzehuang and Hengrui together got nearly 80% of capital allocation. This implies a smaller exposure to the performance of specific stocks and a more diversified strategy.

Returns are also more accurate. With no single asset controlling the portfolio, the top three contributors (Ejiao, Hengrui, and Pientzehuang) each contributed between RMB 10,000 and RMB 14,000 for the two years. In comparison, Pientzehuang alone provided more than RMB 46,000 to the Sharpe strategy. The notion of balanced upside potential was further supported by the 5-year results, which showed that Pientzehuang contributed RMB 12,955.84, while the other top assets, such as Hengrui and Ejiao, also produced large but proportionate returns

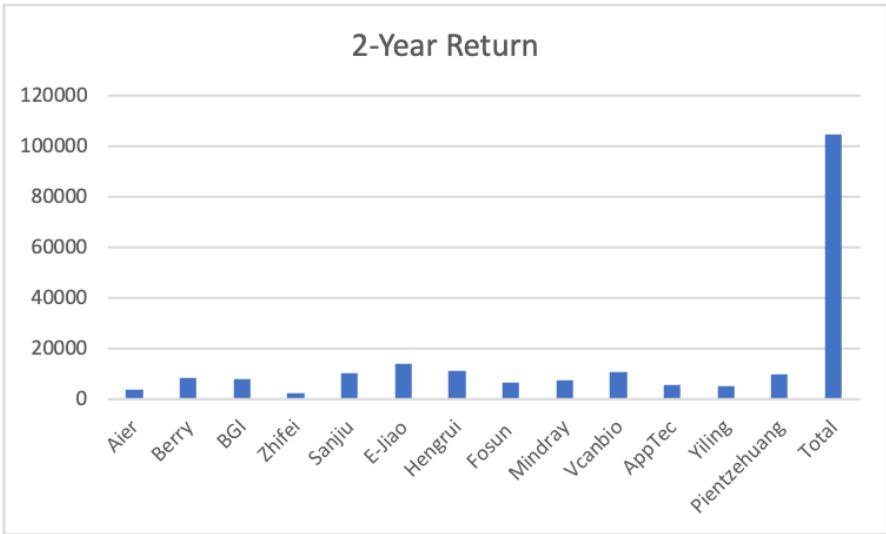

Figure 9: 2-Year Return of Risk Parity Portfolio

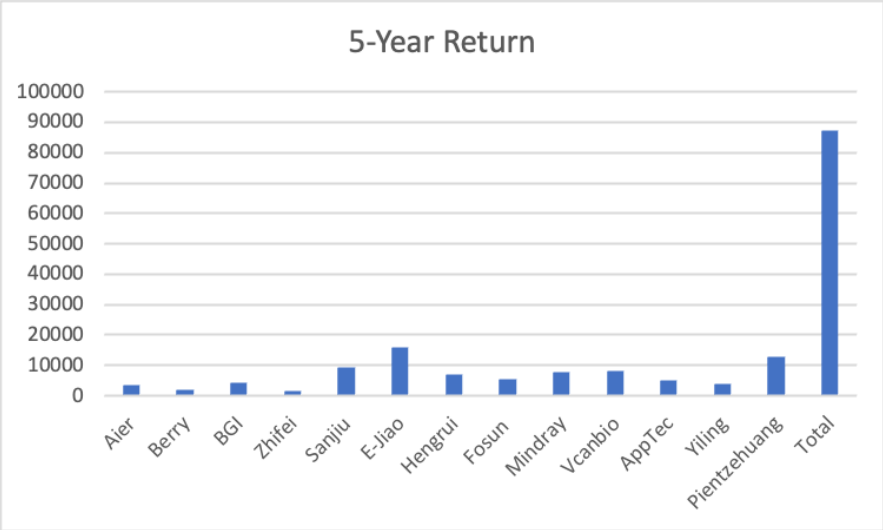

Figure 10: 5-Year Return of Risk Parity Portfolio

(RMB 6,496.02 and RMB 5,950.24, respectively).

As seen in Figure 8, these results demonstrate that the Risk Parity approach not only guarantees equalized risk contribution but also generates a positive and consistent return profile during both assessment frames. It is particularly appropriate for investors who want to reduce the performance volatility brought on by concentrated investments.

4.2. Dynamic Portfolio Optimization with Deep RL

To evaluate the performance of the reinforcement learning-based strategy, a PPO (Proximal Policy Optimization) agent was trained within a custom PortfolioEnv environment. The state space consisted of a rolling window of five-day historical returns, and the action space was defined as a continuous vector representing asset

allocation weights. The reward function incorporated daily log-returns, penalized by transaction costs and abrupt changes in asset weights to promote stable portfolio adjustments. The training process employed a multilayer perceptron (MLP) policy with a total of 20,000 steps, using Stable-Baselines3 and DummyVecEnv for environment management. After training, the agent was tested across the full historical window.

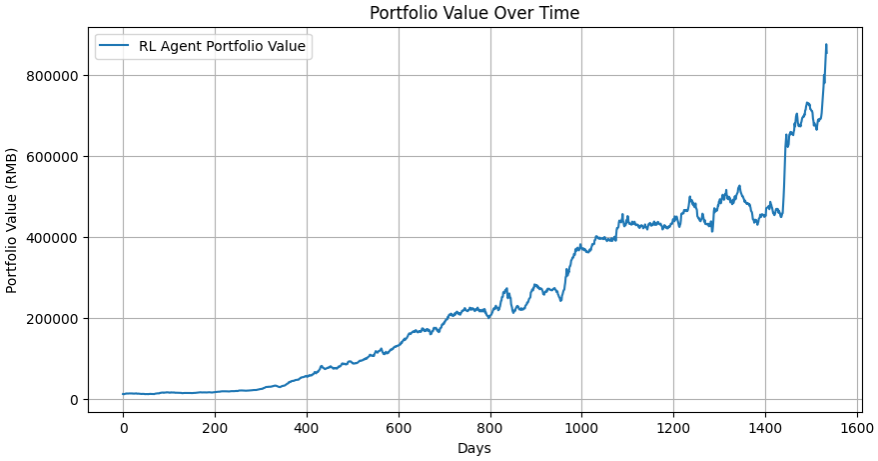

Figure 11: Cumulative Portfolio Value Achieved by the PPO Agent

As shown in Figure 11, the cumulative portfolio value attained by the PPO agent has a distinct rising trend throughout the investing period. A few of little declines (such as around days 800 and 1300) break up this steady increase, suggesting the agent's susceptibility to market swings. These drawdowns are promptly fixed, nevertheless, demonstrating the PPO agent's capacity to dynamically reallocate weights in reaction to unfavorable circumstances. Notably, the market expanded during the time marked by the significant upswing following day 1400, during which the agent effectively capitalized on the upward trend.

The PPO method not only maintained a good long-term return trend but also accomplished smoother transitions between growth phases and corrections as compared to the static techniques previously studied. This implies that risk-adjusted learning behavior was successfully internalized by the RL agent, producing a portfolio strategy that is resilient and flexible in the face of volatility.

Furthermore, the efficiency of the weight change regularization and transaction cost penalty added to the reward function is confirmed by the lack of significant value shocks. These limitations were essential in maintaining cumulative returns, lowering overtrading, and flattening the portfolio curve.

4.3 Result Summary

In a reinforcement learning framework, the PPO-based dynamic technique shows remarkable efficacy in portfolio allocation management. The rising trend of the portfolio value in Figure 11 indicates that the agent consistently grew the portfolio over the assessment period. The model's capacity to make adaptive decisions

was demonstrated by its ability to catch positive market trends without being overexposed to deteriorating assets.

Compared with static strategies such as Equal Weight and Sharpe Ratio Optimization, the PPO strategy offers greater flexibility and responsiveness to changing market conditions. The final portfolio value and return curve indicate competitiveness, if not better, despite the fact that precise quantitative measures like annualized return, maximum drawdown, and Sharpe ratio were not calculated.

Moreover, the integration of transaction costs and weight adjustment penalties within the reward function ensures that the agent balances profitability with practical trading considerations. This reinforces the strategy's applicability in realistic investment environments.

Overall, the PPO agent exhibits strong potential for dynamic portfolio optimization, providing a reliable and automated method for risk-adjusted allocation and robust returns. The effectiveness of reinforcement learning in financial decision-making situations is confirmed by its performance, which is consistent with findings from previous research.

5. Discussion

This chapter provides a comprehensive analysis of the proposed methods and experimental results, discussing their applicability, limitations, and potential future developments in real-world investment environments.

5.1. Reflection on the Effectiveness of the Methods

First, the dynamic portfolio optimization method proposed in this study combines traditional static strategies (such as Equal Weight and Sharpe Ratio optimization) with dynamic portfolio rebalancing strategies based on Deep Reinforcement Learning (Deep RL). Based on comparative experimental data, we find that the Sharpe ratio optimization technique performs better in terms of risk-adjusted performance and long-term returns than the Equal Weight portfolio. By taking into account both asset risk and return, Sharpe ratio optimization enables more efficient asset allocation adjustments and higher returns. Even while the Equal Weight approach is steady in the short term, it ignores the risk heterogeneity among assets, which results in ineffective risk management, especially when working with assets that are highly volatile. On the other hand, in dynamic market conditions, the PPO-trained agent shows more flexibility and risk management. The reinforcement learning approach can successfully reduce the risks associated with unexpected market shocks by adjusting asset allocation in real-time, especially during times of market volatility.

This study also investigated the efficacy of the Risk Parity (RP) approach. The Risk Parity strategy is especially helpful in turbulent market environments because it balances the risk contributions of each asset, reducing overexposure to high-volatility assets. The Risk Parity approach is more resilient and does not heavily rely on predicted returns like conventional mean-variance optimization techniques do. The Risk Parity technique may have a drawback, too, in that systemic hazards may make risk balancing insufficient to maximize profits. In contrast, the reinforcement learning approach gives better market flexibility and continuously modifies its approach to maximize the weight allocation. Reinforcement learning, in contrast to static optimization techniques, continually engages with the market environment and modifies asset allocation to improve risk-adjusted returns.

5.2. 5.2. Evaluation of Strategy Choices

- **Equal Weight Strategy**: As the most basic benchmark method, this strategy allocates equal funds to each asset without the need for predictive models. While it is highly simple to implement and easy to understand, it fails to account for the differences in volatility among assets, which may result in irrational risk exposure. Empirical results show that the Equal Weight strategy lags behind optimized strategies in terms of long-term cumulative returns, making it suitable as a baseline for performance comparison.
- **Sharpe Ratio Optimization**: The objective is to maximize the expected return per unit of risk. The portfolio concentrates on assets with high Sharpe

ratios, potentially increasing exposure to systemic risks. This strategy performs optimally in terms of long-term returns but relies heavily on precise estimates of expected returns and covariances.

- **Risk Parity Strategy**: This strategy emphasizes equal risk contribution from each asset, addressing the issue of overexposure to high-volatility assets seen in Equal Weight. While it demonstrates robust risk control capabilities, its returns are slightly lower than those of Sharpe Ratio optimization. It is more suitable for risk-averse investors and institutional investors.

- **Deep Reinforcement Learning-based Strategy**: This strategy dynamically adapts to market changes, learning optimal weight allocation through interaction. It retains excellent training stability and has good return patterns in simulations. Despite its limitations, which include its inability to be easily interpreted and its dependence on training settings, it holds potential for further asset allocation research.

In summary, when evaluating different strategies, the Equal Weight portfolio, due to its simplicity and robustness, is often used as a benchmark but ignores risk heterogeneity across assets. Sharpe Ratio optimization considers risk-adjusted returns but is sensitive to historical parameter estimation errors. By distributing the risk contribution of each asset equally, Risk Parity lowers concentration risk and works well in erratic markets. Reinforcement learning techniques provide more flexibility and adaptability, especially when managing asset allocation in dynamic contexts, by automatically learning the best weight adjustment rules through the construction of an environment reward function. However because of the complex training procedure and poor interpretability of the model, it is extremely sensitive to the quality of the data and the selection of hyperparameters.

5.3. Limitations and Future Work

Even though this study shows how deep reinforcement learning may be used in conjunction with traditional approaches, there are still a number of drawbacks. The model's generalizability may be limited by the dataset's exclusive emphasis on the Chinese biotechnology and pharmaceutical industries. Furthermore, the simulated and actual returns may differ because real-world factors like slippage, transaction costs, and behavioral dynamics were not considered. Extreme market conditions can also make the Sharpe ratio—which is employed as the reward function—less than ideal.

By extending the asset universe across markets and sectors, adding realistic trade frictions, and creating more resilient reward functions through multi-objective optimization, future research might overcome these constraints. Interpretability and investor trust may also be enhanced by combining explainable factor models with reinforcement learning.

6. Conclusion

This research aimed to assess and contrast four different portfolio optimization techniques—Equal Weight (EW), Sharpe Ratio–based Mean–Variance Optimization (Sharpe–MVO), Risk Parity (RP), and Deep Reinforcement Learning (DRL)—in the framework of China's biotechnology and pharmaceutical industry. Through systematic backtesting of historical data from October 17, 2018 to March 1, 2025, this study draws some conclusive conclusions. First, the EW approach produces the lowest cumulative returns and ignores asset-specific risk differences, despite being simple to apply and resilient against estimating mistakes. Second, by concentrating capital in high-Sharpe assets, Sharpe–MVO provides the best risk-adjusted returns, but it also has a high concentration risk. Third, RP attains a balanced risk allocation that, at the price of some upside potential, enhances stability and drawdown management. In the end, the DRL-based approach—applied through a PPO agent in a customized gym environment with a Sharpe-ratio reward—showed the most flexibility and beat other static approaches in terms of risk and return metrics.

These results reinforce the trade-off between simplicity, risk control, and dynamic adaptability when viewed within the broader body of research on portfolio optimization. While benchmark simplicity (EW), maximal historical risk-adjusted return (Sharpe–MVO), and diversification-enhanced stability (RP) are all distinct advantages of traditional static approaches, they are unable to respond to changing market conditions. In contrast, the DRL framework offers a potential extension to classical theories by utilizing real-time learning to navigate non-stationary environments. This reason is obvious: adding machine learning to asset-allocation frameworks can produce better results in high-growth, volatile industries where static assumptions aren't reliable.

The methodological implications of these findings are as important as their empirical performance. The effectiveness of Risk Parity in reducing volatility emphasizes the need for risk-based budgeting as a framework for dynamic strategies. Similarly, the effectiveness of a Sharpe-ratio reward in DRL emphasizes how crucial it is to match algorithmic goals with reliable financial indicators. These findings point to potential areas for hybrid techniques in the future, such as creating multi-objective reward functions that include drawdown, volatility, and return objectives or integrating risk-budget limitations into reinforcement-learning structures. Future studies can narrow the gap between theoretical rigor and real-world asset management flexibility by redefining portfolio development as an interactive decision-making process as opposed to a one-time optimization.

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Ebenezer Atta Mills for his continuous support, guidance and encouragement throughout the research process. His profound insights and valuable suggestions played an important role in the development of this paper.

I would also like to thank all the professors in the Department of Mathematics of Wenzhou-Kean University, whose courses laid a solid foundation for this research. I would like to thank my classmates and friends for their moral support during my study. I would like to thank our families for their understanding, patience and encouragement.

Finally, I am especially grateful to myself for bravely stepping out of my comfort zone and going to Kean USA for an exchange program in my junior year. During the exchange period in the United States, I experienced a very different culture and became more confident. I was fortunate to meet several excellent professors who guided me in research. I experienced many things in the United States that made me grow. This experience was like a stone falling into a calm lake in my heart, which was memorable and valuable.

References

[1] H. Bruneo, E. Giacomini, G. Iannotta, A. Murthy, and J. Patris, "Risk and return in the biotech industry," *International Journal of Productivity and Performance Management*, vol. 73, no. 6, 2023.

[2] R. Badwe, "Portfolio optimization in early drug rd: a deeper dive into underlying dynamics of early research in pharmaceutical/biotech sector," https://hdl.handle.net/1721.1/122442, 2019, mIT.edu.

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

[4] W. F. Sharpe, "The sharpe ratio," 1994, accessed: May 2025. [Online]. Available: https://web.stanford.edu/~wfsharpe/art/sr/SR.htm#Sharpe66

[5] "Markowitz mean-variance portfolio theory," accessed: May 2025. [Online]. Available: https://sites.math.washington.edu/~burke/crs/408/fin-proj/mark1.pdf

[6] J. H. Kim, W. C. Kim, and F. J. Fabozzi, "Recent developments in robust portfolios with a worst-case approach," *Journal of Optimization Theory and Applications*, vol. 161, no. 1, pp. 103–121, 2014.

[7] P. Glasserman and X. Xu, "Robust risk measurement and model risk," *Quantitative Finance*, vol. 14, no. 1, pp. 29–58, 2013.

[8] R. Baviera and G. Bianchi, "Model risk in mean-variance portfolio selection: an analytic solution to the worst-case approach," *Journal of Global Optimization*, vol. 81, no. 2, pp. 469–491, 2021.

[9] R. Malladi and F. J. Fabozzi, "Equal-weighted strategy: Why it outperforms value-weighted strategies? theory and evidence," *Journal of Asset Management*, vol. 18, no. 3, pp. 188–208, Nov. 2016.

[10] B. R. Auer and F. Schuhmacher, "Performance hypothesis testing with the sharpe ratio: The case of hedge funds," *Finance Research Letters*, vol. 10, no. 4, pp. 196–208, 2013.

[11] G. Deng, T. Dulaney, C. McCann, and O. Wang, "Robust portfolio optimization with value-at-risk-adjusted sharpe ratios," *Journal of Asset Management*, vol. 14, no. 5, pp. 293–305, 2013.

[12] W. Lee, "Risk-based asset allocation: A new answer to an old question?" *The Journal of Portfolio Management*, 2011.

[13] H. Kazemi, "Alternative investment analyst review • caia.org research review an introduction to risk parity," accessed: May 2025. [Online]. Available: https://www.caia.org/sites/default/files/3-all_about_parity.pdf

[14] S. Maillard, T. Roncalli, and J. Teiletche, "On the properties of equally-weighted risk contributions portfolios," *SSRN Electronic Journal*, 2008.

[15] D. Chaves, J. Hsu, F. Li, and O. Shakernia, "Risk parity portfolio vs. other asset allocation heuristic portfolios," *The Journal of Investing*, vol. 20, no. 1, pp. 108–118, 2011.

[16] D. B. Chaves, J. C. Hsu, F. Li, and O. Shakernia, "Efficient algorithms for computing risk parity portfolio weights," *SSRN Electronic Journal*, vol. 21, no. 3, 2012.

[17] Z. Gao, Y. Gao, Y. Hu, Z. Jiang, and J. Su, "Application of deep q-network in portfolio management," https://ieeexplore.ieee.org/abstract/document/9101333/, May 2020, iEEE Xplore.

[18] A. A. Pawar, Muskawar, V. Prashant, and R. Tiku, "Portfolio management using deep reinforcement learning," https://arxiv.org/abs/2405.01604, 2024, arXiv.org.

[19] F. Espiga-Fernández, García-Sánchez, and J. Ordieres-Meré, "A systematic approach to portfolio optimization: A comparative study of reinforcement learning agents, market signals, and investment horizons," *Algorithms*, vol. 17, no. 12, p. 570, Dec. 2024.

[20] G. Huang, X. Zhou, and Q. Song, "A deep reinforcement learning framework for dynamic portfolio optimization: Evidence from china's stock market," https://www.researchgate.net/publication/390995038_A_Deep_Reinforcement_Learning_Framework_for_Dynamic_Portfolio_Optimization_Evidence_from_China, Apr. 2025, researchGate.

[21] W. F. Sharpe and H. M. Markowitz, "Mean-variance analysis in portfolio choice and capital markets," *The Journal of Finance*, vol. 44, no. 2, p. 531, 1989.

[22] V. Bhansali, J. Davis, G. Rennison, J. Hsu, and F. Li, "The risk in risk parity: A factor-based analysis of asset-based risk parity," *The Journal of Investing*, vol. 21, no. 3, pp. 102–110, 2012.

[23] T. Raffinot, "The hierarchical equal risk contribution portfolio," accessed: May 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3237540

[24] L. Qiu, Z. Chen, D. Lu, H. Hu, and Y. Wang, "Public funding and private investment for r&d: a survey in china's pharmaceutical industry," *Health Research Policy and Systems*, vol. 12, p. 27, 2014.

[25] H. Yuanjia, C. O. L. Ung, B. Ying, and W. Yitao, "The chinese pharmaceutical market: Dynamics and a proposed investment strategy," *Journal of Medical Marketing*, vol. 7, no. 1, pp. 18–24, 2007.

[26] J. Xu, X. Wang, and F. Liu, "Government subsidies, r&d investment and innovation performance: analysis from pharmaceutical sector in china," *Technology Analysis & Strategic Management*, vol. 33, no. 5, 2021.

[27] A. Mohan and A. Roy, "A strategic investment framework for biotechnology markets via dynamic asset allocation and class diversification," https://arxiv.org/abs/1710.03267, 2017, arXiv.org.

[28] R. G. Ibbotson, "The importance of asset allocation," *Financial Analysts Journal*, vol. 66, no. 2, pp. 18–20, Dec. 2018.

[29] M. J. Brennan, E. S. Schwartz, and R. Lagnado, "Strategic asset allocation," *Journal of Economic Dynamics and Control*, vol. 21, no. 8–9, pp. 1377–1403, Jun. 1997.

[30] A. Alviniussen and H. Jankensgard, "Enterprise risk budgeting: Bringing risk management into the financial planning process," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2694744, Nov. 2015, papers.ssrn.com.

[31] B. Bruder and T. Roncalli, "Managing risk exposures using the risk budgeting approach," *SSRN Electronic Journal*, 2012.

[32] R. C. Merton, "On estimating the expected return on the market," *Journal of Financial Economics*, vol. 8, no. 4, pp. 323–361, Dec. 1980.

[33] F. Wen, J. Cao, Z. Liu, and X. Wang, "Dynamic volatility spillovers and investment strategies between the chinese stock market and commodity markets," *International Review of Financial Analysis*, vol. 76, p. 101772, Jul. 2021.

[34] B. R. Routledge, "Machine learning and asset allocation," *Financial Management*, vol. 48, no. 4, pp. 1069–1094, Nov. 2019.

[35] Z. Hong, R. Tian, Q. Yang, W. Yao, T. Ye, and L. Zhang, "Asset allocation via machine learning," *Accounting and Finance Research*, vol. 10, no. 4, p. 34, Nov. 2021.

[36] K. Jakobsen and C. B. Bang, "How can reinforcement learning algorithms be used for capital allocation and consumption for a sovereign wealth fund?" https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3152353, 2024, nTNU.no, Accessed: May 09, 2025.

[37] J. O, J. Lee, J. Lee, and B. Zhang, "Adaptive stock trading with dynamic asset allocation using reinforcement learning," *Information Sciences*, vol. 176, no. 15, pp. 2121–2147, Aug. 2006.

[38] J. Jerome, L. Sanchez-Betancourt, R. Savani, and M. Herdegen, "Model-based gym environments for limit order book trading," https://arxiv.org/abs/2209.07823, 2022, arXiv.org, Accessed: May 09, 2025.

[39] Y.-H. Chou, S.-Y. Kuo, and Y.-C. Jiang, "A novel portfolio optimization model based on trend ratio and evolutionary computation," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 4, pp. 337–350, Aug. 2019.

Appendix A

```python
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
pd.options.display.float_format = '{:.4%}'.format

# Load data
Y = pd.read_excel("/Users/xieweixun/Desktop/Return.xlsx")
Y['Date'] = pd.to_datetime(Y['Date'])
Y.set_index('Date', inplace=True)
Y.columns = Y.columns.str.strip()

print(Y.head())

# Define end date as the latest available date
end_date = Y.index[-1]

# Define start date as 5 years before the end date
start_date = end_date - pd.DateOffset(years=5)

# Extract the 5-year data range
returns_5yr = Y.loc[start_date:end_date]

# Calculate cumulative return using the formula:
# R = (    (1 + r_i)) - 1
five_year_return = (1 + returns_5yr).prod() - 1

# Display the result in percentage format
print("5-Year Return for each asset:")
print(five_year_return.apply(lambda x: f"{x:.2%}"))

# Define start date as 2 year before the end date
start_date_2yr = end_date - pd.DateOffset(years=2)

# Extract the 2-year data range
returns_2yr = Y.loc[start_date_2yr:end_date]

# Calculate cumulative 1-year return using the same formula:
# R = (    (1 + r_i)) - 1
two_year_return = (1 + returns_2yr).prod() - 1

# Display the result in percentage format
print("\n2-Year Return for each asset:")
print(two_year_return.apply(lambda x: f"{x:.2%}"))

import riskfolio as rp
print("Riskfolio imported successfully under Rosetta x86_64")
print(rp.__version__)

# Building the portfolio object
port = rp.Portfolio(returns=Y)

# Calculating optimal portfolio

```

```
# Select method and estimate input parameters:

method_mu='hist' # Method to estimate expected returns based on historical data.
method_cov='hist' # Method to estimate covariance matrix based on historical data.

port.assets_stats(method_mu=method_mu, method_cov=method_cov)

# Estimate optimal portfolio:

model='Classic' # Could be Classic (historical), BL (Black Litterman) or FM (Factor Model)
rm = 'MV' # Risk measure used, this time will be variance
obj = 'Sharpe' # Objective function, could be MinRisk, MaxRet, Utility or Sharpe
hist = True # Use historical scenarios for risk measures that depend on scenarios
rf = 0 # Risk free rate
l = 0 # Risk aversion factor, only useful when obj is 'Utility'

w = port.optimization(model=model, rm=rm, obj=obj, rf=rf, l=l, hist=hist)

display(w.T)

# Plotting the composition of the portfolio

ax = rp.plot_pie(w=w, title='Sharpe Mean Variance', others=0, nrow=25, cmap = "tab20",
                 height=6, width=10, ax=None)

# Plotting the risk composition of the portfolio

ax = rp.plot_risk_con(w, cov=port.cov, returns=port.returns, rm=rm, rf=0, alpha=0.01,
                      color="tab:blue", height=6, width=10, ax=None)

b = None # Risk contribution constraints vector

w_rp = port.rp_optimization(model=model, rm=rm, rf=rf, b=b, hist=hist)

display(w_rp.T)

ax = rp.plot_pie(w=w_rp, title='Risk Parity Variance', others=0.05, nrow=25, cmap = "tab20",
                 height=6, width=10, ax=None)

ax = rp.plot_risk_con(w_rp, cov=port.cov, returns=port.returns, rm=rm, rf=0, alpha=0.01,
                      color="tab:blue", height=6, width=10, ax=None)


```

```
# 'MV': Standard Deviation.
# 'MAD': Mean Absolute Deviation.
# 'MSV': Semi Standard Deviation.
# 'FLPM': First Lower Partial Moment (Omega Ratio).
# 'SLPM': Second Lower Partial Moment (Sortino Ratio).


rms = ['MV', 'MAD', 'MSV', 'FLPM', 'SLPM', 'CVaR',
       'EVaR', 'CDaR', 'UCI', 'EDaR']
w_s = pd.DataFrame([])
success_rms = []

port.lowerret = 0
port.upperret = 1

for i in rms:
    try:
        w = port.optimization(
            model='Classic',
            rm=i,
            obj='Sharpe',
            rf=0,
            l=0,
            hist=True
        )
        w_s = pd.concat([w_s, w], axis=1)
        success_rms.append(i)
    except Exception as e:
        print(f"{i} failed: {e}")

w_s.columns = success_rms
w_s.style.format("{:.2%}").background_gradient(cmap='YlGn')


import matplotlib.pyplot as plt

# Plotting a comparison of assets weights for each portfolio

fig = plt.gcf()
fig.set_figwidth(16)
fig.set_figheight(6)
ax = fig.subplots(nrows=1, ncols=1)

w_s.plot.bar(ax=ax)

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import gymnasium as gym
from gymnasium import spaces
from stable_baselines3 import PPO
from stable_baselines3.common.vec_env import DummyVecEnv

# === 1. Custom Portfolio Environment ===
class PortfolioEnv(gym.Env):
```

```
153     def __init__(self, returns_df, window_size=5,
            transaction_cost_rate=0.002):
154         super().__init__()
155         self.returns_df = returns_df
156         self.window_size = window_size
157         self.transaction_cost_rate = transaction_cost_rate
158         self.n_assets = returns_df.shape[1]
159 
160         self.action_space = spaces.Box(low=0, high=1, shape=(self.
                n_assets,), dtype=np.float32)
161         self.observation_space = spaces.Box(low=-np.inf, high=np.
                inf, shape=(self.window_size, self.n_assets), dtype=np.
                float32)
162         self.reset()
163 
164     def reset(self, seed=None, options=None):
165         super().reset(seed=seed)
166         self.current_step = self.window_size
167         self.prev_weights = np.array([1.0 / self.n_assets] * self.
                n_assets)
168         obs = self._get_observation()
169         return obs.astype(np.float32), {}
170 
171     def _get_observation(self):
172         obs = self.returns_df.iloc[self.current_step - self.
                window_size:self.current_step].values
173         return obs.astype(np.float32)
174 
175     def step(self, action):
176         weights = np.clip(action, 0, 1)
177         weights /= (np.sum(weights) + 1e-8)
178         daily_returns = self.returns_df.iloc[self.current_step].
                values
179         portfolio_return = np.dot(weights, daily_returns)
180         transaction_cost = self.transaction_cost_rate * np.sum(np.
                abs(weights - self.prev_weights))
181 
182         if portfolio_return > -1:
183             weight_change_penalty = 0.005 * np.sum(np.abs(weights
                    - self.prev_weights))
184             reward = np.log(1 + portfolio_return) -
                    transaction_cost - weight_change_penalty
185         else:
186             reward = -transaction_cost
187 
188         reward = np.nan_to_num(reward, nan=0.0, posinf=0.0, neginf
                =0.0)
189         assert not np.isnan(portfolio_return), "portfolio_return␣
                is␣NaN!"
190         assert not np.isnan(transaction_cost), "transaction_cost␣
                is␣NaN!"
191 
192         self.prev_weights = weights
193         self.current_step += 1
194         terminated = self.current_step >= len(self.returns_df) - 1
195         truncated = False
```

```
196
197         obs = self._get_observation()
198         info = {"portfolio_return": portfolio_return, "
                transaction_cost": transaction_cost}
199         return obs.astype(np.float32), reward, terminated,
                truncated, info
200
201
202 # === 2. Load Data and Initialize Env ===
203 Y = pd.read_excel("/Users/xieweixun/Desktop/Return.xlsx")
204 Y['Date'] = pd.to_datetime(Y['Date'])
205 Y.set_index('Date', inplace=True)
206 Y.columns = Y.columns.str.strip()
207 Y = Y.sort_index()
208
209 env = PortfolioEnv(returns_df=Y, window_size=5)
210 vec_env = DummyVecEnv([lambda: env])
211
212 # === 3. Train PPO Model ===
213 del model
214 model = PPO("MlpPolicy", vec_env, verbose=1)
215 model.learn(total_timesteps=20000)
216
217 # === 4. Evaluate Model ===
218 obs = vec_env.reset()
219 rewards, returns, costs = [], [], []
220
221 for _ in range(len(Y) - 5):
222     action, _ = model.predict(obs, deterministic=True)
223     obs, reward, done, info = vec_env.step(action)
224     rewards.append(reward[0])
225     returns.append(info[0]['portfolio_return'])
226     costs.append(info[0]['transaction_cost'])
227
228 rewards = np.array(rewards)
229 returns = np.array(returns)
230 costs = np.array(costs)
231
232 # === 5. Compute Metrics ===
233 def sharpe_ratio(returns, risk_free_rate=0.0):
234     excess_returns = returns - risk_free_rate
235     return np.mean(excess_returns) / np.std(excess_returns)
236
237 def sortino_ratio(returns, risk_free_rate=0.0):
238     negative_returns = returns[returns < risk_free_rate]
239     downside_std = np.std(negative_returns) if len(
            negative_returns) > 0 else 1
240     return np.mean(returns - risk_free_rate) / downside_std
241
242 sharpe = sharpe_ratio(returns)
243 sortino = sortino_ratio(returns)
244
245 # === 6. Visualization ===
246 cumulative_returns = np.cumprod(1 + returns)
247 portfolio_value = 130000 * cumulative_returns
248
```

```
plt.figure(figsize=(10, 5))
plt.plot(portfolio_value, label="RL Agent Portfolio Value")
plt.xlabel("Days")
plt.ylabel("Portfolio Value (RMB)")
plt.title("Portfolio Value Over Time")
plt.legend()
plt.grid(True)
plt.show()

print(f"Sharpe Ratio: {sharpe:.4f}")
print(f"Sortino Ratio: {sortino:.4f}")
```