Portfolio Optimization Using Machine Learning Method and Monte Carlo Simulation

Yiwen Hu *

Duke Kunshan University, Kunshan, Jiangsu 215316, China

* Corresponding Author Email: yh369@duke.edu

Abstract. Investment portfolio optimization is a crucial aspect of quantitative finance, aiming to maximize returns and minimize risks. This study focuses on optimizing investment allocation among five selected stocks (BAIC Blue Valley, BYD, Chang' an Automobile, Kweichow Moutai, Sunwoda) over 100 days using Monte Carlo simulation and machine learning methods. This study uses the sliding window approach with the linear regression model to predict future returns and calculate the variance-covariance matrix to determine the optimal portfolio weights, leading to two key strategies: maximizing the Sharpe ratio and minimizing the risk. The results reveal that the Max Sharpe Ratio portfolio achieved a cumulative return of 0.0909, significantly outperforming the CSI 300 index's return of -0.0348. Additionally, the Min Risk portfolio exceeded the market index with a cumulative return of 0.0062. These findings demonstrate that both the Max Sharpe Ratio and Min Risk strategies are effective in achieving a balance between risk and return and surpassing the market, offering valuable insights for investors seeking optimized portfolio allocations.

Keywords: Portfolio Optimization; Monte Carlo Simulation; Machine Learning; Linear Regression; Risk Management; Sharpe Ratio.

1. Introduction

Portfolio optimization is a fundamental concept in financial management that plays a crucial role in the strategic allocation of assets. The primary objective is to maxim-ize returns while minimizing associated risks, thereby achieving a favorable risk-return trade-off. This concept is deeply rooted in modern portfolio theory, introduced by Harry Markowitz in the 1950s, which revolutionized the way investors approach asset allocation. Markowitz's theory emphasizes the importance of diversification, suggesting that a well-diversified portfolio can significantly reduce risk without sacri-ficing potential returns [1-2].

In the contemporary financial landscape, portfolio optimization is more relevant than ever due to the increasing complexity of global markets and the proliferation of investment opportunities. Investors are constantly seeking strategies that can navigate volatile market conditions, adapt to economic shifts, and leverage emerging trends.

Effective portfolio management is vital for achieving long-term financial goals, whether for individual investors, institutional funds, or corporate entities. The essence of portfolio optimization lies in its ability to balance the trade-offs between risk and return, ensuring that investors are adequately compensated for the risks they under-take. Traditional methods of portfolio optimization, which often rely on static models and historical data, may not be sufficient in the face of dynamic market conditions and evolving economic factors. The integration of advanced quantitative techniques, such as Monte Carlo simulations and machine learning methods, has further enhanced the precision and effectiveness of portfolio optimization. Monte Carlo simulations and machine learning methods offer a more sophisticated approach by incorporating randomness and uncertainty into the optimization process. Monte Carlo simulations generate various possible outcomes based on historical data and probabilistic scenari-os, allowing for a more comprehensive evaluation of potential investment strategies. Machine learning methods, particularly linear regression, enable modeling relationships between stock prices and their predictors, enhancing the accuracy of return pre-dictions and improving the overall optimization process [3].

The goal of this study lies in the application of advanced quantitative techniques to predict future stock performance and optimize portfolio allocation. This study focuses on five stocks: BAIC BluePark, BYD, Chang' an Automobile, Kweichow Moutai, and Sunwoda. These stocks were selected due to their significant roles and strong performances in their respective industries, making them valuable components for a diversified investment portfolio. By leveraging historical stock prices, the linear re-gression model in the sliding window approach is employed to forecast future returns. This method involves using a fixed-size window of past data to make predictions for subsequent periods, thus continuously updating the model with the latest available information [4-5]. This dynamic approach helps in capturing the evolving market trends and stock behaviors more accurately.

To the best of our knowledge, this paper makes the following contributions to the literature. The first step is to collect the daily stock price return data for five repre-sentative companies from various industries: electric vehicles (EV), automotive manu-facturing, consumer goods, and energy storage solutions. This diverse selection en-sures a comprehensive study of the stock performance across different sectors, thereby enhancing the robustness of the portfolio optimization process. Secondly, to ensure the accuracy and consistency of the financial time series data, rigorous data-cleaning techniques are applied. This step is crucial for maintaining the integrity of the analy-sis and ensuring that the results are reliable and actionable for investors. Third, it is important to utilize a combination of Monte Carlo simulations and machine learning methods, specifically linear regression, to forecast future stock returns and determine the optimal portfolio weights. This hybrid approach leverages the strengths of both techniques, providing a more nuanced and dynamic optimization strategy compared to traditional methods. This study's significance lies in its innovative methodology and practical implications. Integrating Monte Carlo simulations with machine learning offers a comprehensive framework that adapts to changing market conditions and investor preferences. The results of this study provide valuable insights for investors looking to enhance their investment strategies through advanced quantitative tech-niques. Moreover, this study contributes to the broader field of financial management by demonstrating the practical applications of these methods in optimizing portfolios composed of stocks from diverse sectors. The findings can help investors make more informed decisions, achieve better risk management, and ultimately improve their investment outcomes. This research underscores the importance of a systematic and data-driven approach to portfolio optimization, highlighting the benefits of leveraging modern analytical tools in financial decision-making.

2. Data Collection

The data for this study consists of the daily stock prices of five selected stocks over a period of 100 days. The data was sourced from the Investing official website (https://cn.investing.com), a reliable and comprehensive financial database. The da-taset includes the following columns: Close, Open, High, Low, Volume, and Change. The time range for the data collection spans from February 8, 2024, to July 12, 2024.

The five stocks selected for this study are: BAIC BluePark, BYD, Chang' an Au-tomobile, Kweichow Moutai, and Sunwoda. These stocks were chosen based on their relevance and performance in the Chinese market, representing different sectors. BAIC BluePark is a prominent player in the electric vehicle (EV) sector, reflecting the grow-ing trend towards sustainable transportation. BYD is an innovative company in both the EV and battery manufacturing industries, known for its technological advance-ments and rapid expansion. Chang' an Automobile represents a key manufacturer in the traditional automotive industry with a strong market presence. Kweichow Moutai plays the role of a leading producer of premium spirits, representing the consumer goods sector with high brand value. What's more, Sunwoda is a significant company in the battery industry, contributing to energy storage solutions critical for various technological applications. These stocks were selected to ensure a diversified portfolio that spans multiple sectors, thereby minimizing sector-specific risks and enhancing overall portfolio stability.

To provide a detailed overview of the selected stocks, Table 1 presents the descrip-tive statistics, including the minimum, maximum, mean, and variance of daily re-turns for each stock over 100 days.

Table 1. Three Scheme comparing

	BAIC BluePark	BYD	Chang' an Automobile	Kweichow Moutai	Sunwoda.
Min	-0.0908	-0.0770	-0.0672	-0.0338	-0.0843
Max	0.1043	0.0359	0.1114	0.0320	0.0564
Mean	-0.0082	-0.0035	-0.0007	0.0015	-0.0009
Variance	0.0016	0.0004	0.0008	0.0001	0.0006

The mean daily returns for the stocks range from -0.0082 to 0.0015, indicating both negative and positive average performances over the period. The variance of the daily returns, which measures the dispersion of returns, varies between 0.0001 and 0.0016, highlighting different levels of volatility among the stocks. For instance, Chang' an Automobile shows the highest maximum daily return of 0.1114 and a relatively high variance of 0.0008, suggesting significant fluctuations in its stock price. In contrast, Kweichow Moutai, known for its premium brand, exhibits a bal-anced profile with a mean return of 0.0020 and the lowest variance of 0.0001, indicat-ing consistent performance with moderate volatility. Analyzing the data of five repre-sentative stocks offers insights into the performance and risk profiles of the selected stocks, which are crucial for the subsequent portfolio optimization process. The di-verse characteristics of these stocks make them suitable candidates for constructing a well-balanced and diversified portfolio, aiming to maximize returns while minimizing risk.

3. Methods

3.1. Linear Regression Model

Linear regression is used in the sliding window approach to predict future stock re-turns based on historical data. This method is particularly applicable due to its sim-plicity and interpretability, making it suitable for financial time-series forecasting. The advantages of linear regression include simplicity, efficiency, and stability. It is straightforward to implement and interpret, providing a clear relationship between predictors and the target variable. Compared to more complex models, it requires fewer computational resources, making it feasible for large datasets. Simultaneously, it tends to be less sensitive to noise, which can be advantageous in financial time series where noise is prevalent.

The linear regression model used in the sliding window approach is expressed as:

$$R_{t+h} = \beta_0 + \beta_1 R_{t-1} + \beta_2 R_{t-2} + \cdots + \beta_p R_{t-p} + \epsilon_t \tag{1}$$

where R_{t+h} is the predicted return for time $t + h$, β_0 is the intercept term, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the lagged values, $R_{t-1}, R_{t-2}, \ldots, R_{t-p}$ are the lagged return values, and ϵ_t is the error term at time t [3].

3.2. Sliding Window Approach

In the sliding window approach, a fixed-size window of historical data is used to train the linear regression model and make predictions for the next period. The process involves the following steps. Firstly, according to the number of days of historical data used to predict future returns, it is easy to define a window size. In this study, the window size is 70 days, which means that the past 70 days of data are used as training data to predict the next 30 days' returns. Second, within each window, the linear regression model introduced above is trained using the historical return data. The model learns the relationship between the lagged returns and the future returns based on the historical data within the window. Once trained, the model is used to predict future returns. The most recent window of data is used for this purpose, allowing us to update predictions as new data becomes available. Then, after making predictions, the window is moved forward by one day, and the process is repeated until predictions cover the entire forecast period [4-5].

This approach allows us to dynamically adjust predictions based on the most recent data, capturing trends and patterns that might change over time. The linear regression model's coefficients are recalibrated with each window, helping to account for shifts in market conditions.

3.3. Portfolio Theories

In the portfolio optimization part, several theories will be used to build profitable portfolios.

After getting the predictions of the returns for the 71-100 days based on the sliding window approach, the predicted returns were used to compute the variance-covariance matrix (Σ) to capture the relationships between the stocks [6-7]. It consists of the variances and covariances of returns for a set of assets in a portfolio. Mathematically, it is represented as

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \tag{2}$$

where σ_{ij} represents the covariance between the stock i and stock j [7].

Then, to get a more accurate result, 1,000,000 Monte Carlo simulations are sufficient to generate random weights for the five stocks. Each portfolio's weights were normalized to ensure the total equaled one. Eventually, to identify the portfolios, the concepts of maximum Sharpe ratio and minimum variance are used here. The Sharpe ratio measures the risk-adjusted return of a portfolio. It allows investors to compare the return of an investment relative to its risk. A higher Sharpe ratio indicates a more favorable risk-adjusted return, making it a valuable metric for investors aiming to maximize returns while controlling for risk [8-9]. It effectively balances the trade-off between risk and reward. The formula for the Sharpe ratio (S) is

$$S = \frac{E(R_p) - R_f}{\sigma_p} \tag{3}$$

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. What's more, the minimum variance portfolio aims to minimize the portfolio's overall risk (variance). It is designed to provide the lowest possible risk for a given set of assets [9]. This strategy is particularly appealing to risk-averse investors who prioritize stability and seek to minimize exposure to volatility. By focusing on reducing risk, this approach aims to create a more stable investment portfolio [9]. The formula for the portfolio variance $({\sigma_p}^2)$ is

$${\sigma_p}^2 = w^T \Sigma\, w \tag{4}$$

where w is the vector of portfolio weights, and Σ is the variance-covariance matrix of asset return.

4. Summary

4.1. Linear Regression Model

The prediction results are shown in Figure 1, which compares the actual return rates with the predicted return rates for each of the five stocks. Stock 1 represents BAIC Blue Valley, stock 2 for BYD, stock 3 for Chang' an Automobile, stock 4 for Kweichow Moutai, and stock 5 for Sunwoda.

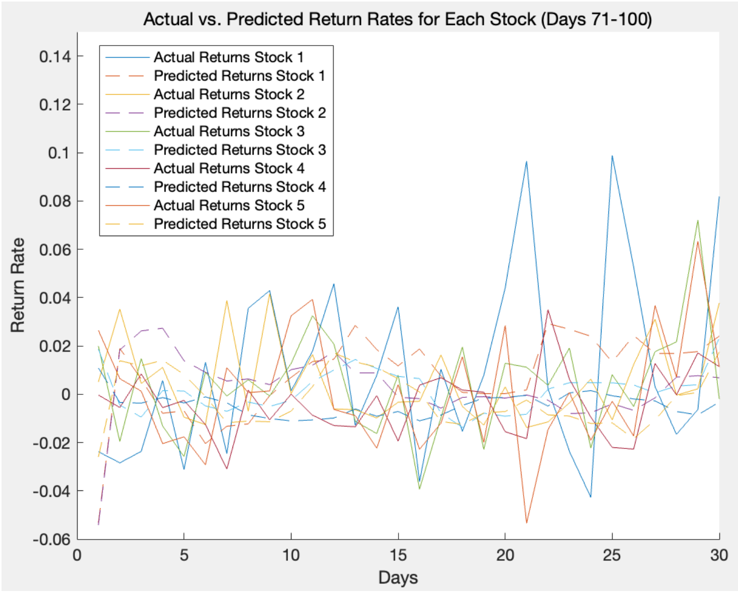

Fig. 1. Actual vs. Predicted Return Rates of these five stocks

Figure 1 displays actual return rates (solid lines) and predicted return rates (dashed lines) for each stock. The predictions generally follow the trends of actual returns, particularly for stocks with consistent historical performance. However, deviations are more noticeable for stocks with higher volatility, such as stock 1, BAIC Blue Valley, indicating the challenges in forecasting their returns accurately. The model shows a relatively good fit for stocks with more stable historical performance, such as stock 2, BYD, which has a relatively low variance.

4.2. Efficient Frontier

The results are summarized in Figure 2, which shows the scatter plot of the simulated portfolios.

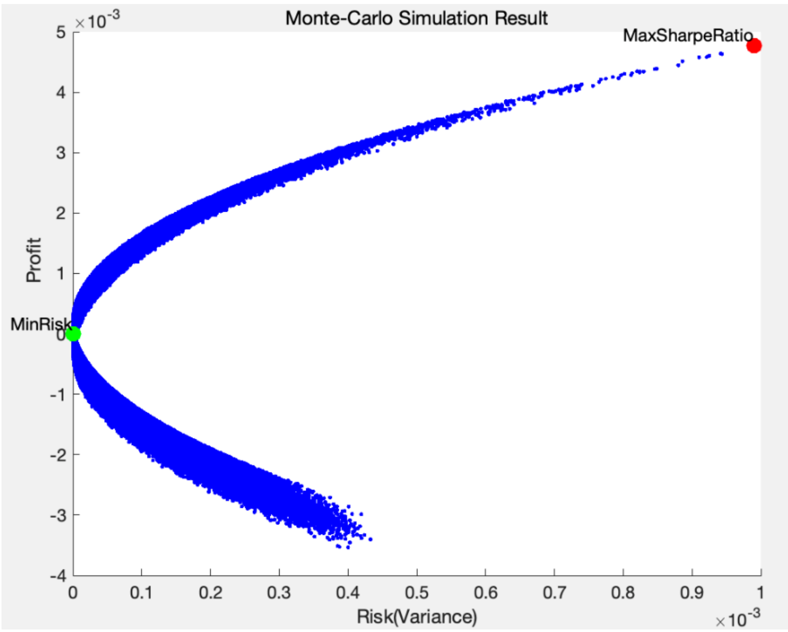

Fig. 2. Portfolio Optimization Results

Figure 2 shows a scatter plot of the portfolios generated by Monte Carlo simulations, with risk (variance) on the x-axis and return on the y-axis. The red point represents the portfolio with the maximum Sharpe ratio, while the green point indicates the portfolio with the minimum variance. This visualization helps in understanding the trade-offs between risk and return, with the optimal portfolios clearly identified [10-11]. The Max Sharpe Ratio portfolio achieved a risk of 0.98, which is close to 1, and a return of 4.9×10^{-5}. In contrast, the Min Risk portfolio had a risk of 0, but its return was also close to 0. This indicates that while the Min Risk portfolio effectively minimized volatility, it did not provide substantial returns, whereas the Max Sharpe Ratio portfolio achieved a balance between risk and return.

4.3. Portfolio Weights and Performance

The optimized portfolios identified through the simulations are detailed below in Table 2.

Table 1. Portfolio Weights and Cumulative Return

Portfolio	Weight of BAIC Blue Valley	Weight of BYD	Weight of Chang' an Automobile	Weight of Kweichow Moutai	Weight of Sunwoda	Cumulative Return
Max Sharpe Ratio	0.0096	0.0133	0.9076	0.0341	0.0353	0.096557
Min Risk	0.2009	0.0698	0.0878	0.6036	0.0379	0.005390

Comparing the portfolio returns to the CSI 300 can help us assess whether they are outperforming or underperforming the broader market.

The Cumulative Return under the CSI 300 Index is -0.034810. From Figure 3, both the Max Sharpe Ratio and Min Risk portfolios significantly outperform the CSI 300 index. The Max Sharpe Ratio portfolio excels in risk-adjusted returns, while the Min Risk portfolio focuses on minimizing volatility.

Fig. 3. The performance of Max Sharpe Ratio and Min Risk portfolios and CSI 300

Figure 3 shows that the Max Sharpe Ratio performs better than the Min Risk portfolio, and both of them provide better returns compared to the market benchmark, demonstrating the effectiveness of advanced portfolio optimization techniques in achieving superior financial outcomes. The Max Sharpe Ratio portfolio shows a robust performance in maximizing returns relative to risk, while the Min Risk portfolio offers a stable and low-risk investment option. This comparison highlights the benefits of using sophisticated financial models to enhance portfolio management strategies.

5. Conclusion

This study demonstrates the significant potential of employing Monte Carlo simu-lations and linear regression method for portfolio optimization in the context of five selected stocks: BAIC BluePark, BYD, Chang' an Automobile, Kweichow Moutai, and Sunwoda. Using a sliding window approach to predict stock prices helps calcu-late some values, such as expected returns, variances, and covariances, which con-structs a variance-covariance matrix. This facilitated the generation of optimized port-folios with the maximum Sharpe ratio and minimum risk through extensive Monte Carlo simulations.

The optimized portfolios not only outperformed the CSI 300 index, but also showcased the robustness of the applied methods in achieving superior financial out-comes. The Max Sharpe Ratio portfolio, in particular, excelled in delivering higher risk-adjusted returns, while the Min Risk portfolio demonstrated effective volatility minimization. These findings underscore the efficacy of advanced portfolio optimiza-tion techniques in enhancing investment performance. Overall, this study highlights the strategic importance of portfolio optimization in financial management and its practical applicability in maximizing investment returns while controlling for risk.

However, it is important to note some limitations of this study. The focus on just five stocks, while significant within their sectors, may not fully represent broader market dynamics. What's more, relying on historical data assumes that past trends predict future performance, which might not always hold true in volatile markets. Additionally, the analysis does not account for transaction costs, taxes, or other prac-tical considerations that investors face, which could impact the real-world applicability of the results. Future research could benefit from including a more diverse set of assets and exploring advanced predictive models to improve the robustness of the optimization outcomes.

References

[1] Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. Applied Sciences, 2020, 10(2): 437. DOI: https://doi.org/10.3390/app10020437.

[2] Mittal S, Bhattacharya S, Mandal S. Characteristics analysis of behavioural port-folio theory in the Markowitz portfolio theory framework[J]. Managerial Finance, 2022, 48(2): 277-288. DOI: https://doi.org/10.1108/MF-05-2021-0208.

[3] James G, Witten D, Hastie T, et al. Linear regression. In: An Introduction to Sta-tistical Learning: With Applications in Python. Cham: Springer International Publishing, 2023: 69-134.

[4] Khan I am, Akber A, Xu Y. Sliding Window Regression Based Short-Term Load Forecasting of a Multi-Area Power System[J]. Ithaca: Cornell University Library, arXiv.org, 2019. DOI:10.48550/arxiv.1905.08111

[5] Norwawi N M. Sliding window time series forecasting with multilayer perceptron and multiregression of COVID-19 outbreak in Malaysia[J]. Data Science for COVID-19, 2021: 547-564. DOI: 10.1016/B978-0-12-824536-1.00025-3.

[6] Poornima B G, Reddy Y V. An Analysis of Portfolio VaR: Variance-Covariance Ap-proach[J]. IUP Journal of Applied Finance, 2017, 23(3).

[7] Albuquerque P H M, de Moraes Souza J G, Kimura H. Artificial intelligence in portfo-lio formation and forecast: Using different variance-covariance matrices[J]. Commu-nications in Statistics - Theory and Methods, 2021, 52(12): 4229–4246. DOI:10.1080/03610926.2021.1987472.

[8] Pav S E. The Sharpe Ratio: Statistics and Applications[M]. Chapman and Hall/CRC, 2021.

[9] Vinzelberg A, Auer B R. A comparison of minimum variance and maximum Sharpe ra-tio portfolios for mainstream investors[J]. The Journal of Risk Finance, 2022, 23(1): 55-84.

[10] Boubaker S, Le T D Q, Manita R, et al. The trade-off frontier for ESG and Sharpe ratio: a bootstrapped double-frontier data envelopment analysis[J]. Ann Oper Res, 2023. https://doi.org/10.1007/s10479-023-05506-z.

[11] Siswanah E. Comparative analysis of mean variance efficient frontier and resampled efficient frontier for optimal stock portfolio formation[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020, 846(1): 012065.