International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 9, No. 2, 2023

Portfolio Optimization of Stocks – Python-Based Stock Analysis

Aibo Zhang[1, *]

[1] School of Business, Hunan University of Science and Technology, Xiangtan 411201, China
[*] Corresponding Author

Abstract: With the development of big data, blockchain artificial intelligence, and other technologies, the development of the financial industry also plays a great role in promoting the development of digital finance is also developing rapidly, the huge amount of financial data, the laws behind, randomness, the complexity have increased the difficulty of processing our data. The financial industry is also increasingly in need of data processing talents. For financial data such as: intra-day high-frequency data and stock price and volume data processing, Python has the points of fast calculation speed, open source, and excellent data visualization. In this paper, financial data analysis work based on the Python platform, six FTSE A50 constituent stocks of different industries are selected from the Chinese stock market, namely China Merchants Bank, SAIC Group, Haitong Securities, Capital Mining, China Unicom and Poly Development for financial data analysis. The optimal portfolio with the largest Sharpe ratio and the optimal portfolio with the smallest variance is obtained empirically by Python, and optimized by Monte Carlo simulation, and their expected returns, standard deviations, and Sharpe ratios are compared and analyzed, and finally, the effective boundaries of the asset portfolios are given. The importance of Markowitz's portfolio theory in financial risk management is further illustrated through empirical analysis.

Keywords: Stock price changes, Monte Carlo simulation, Sharpe ratio, Portfolio optimization.

1. Introduction

With the development of financial big data, the huge volume of data has put forward higher requirements for data storage, data analysis, and computational techniques [1], which undoubtedly increases the difficulty of portfolio management-related research. Lngkvist et al. propose that traditional statistical methods are not suitable for analyzing complex, high-dimensional, noisy financial market data series because the statistical analysis is often based on a large number of constraints with conditions for modeling, and these assumptions often do not hold completely in real life, so the model results are limited by the preconditions [2]. Since entering the twenty-first century, the trend of economic globalization has become more obvious, and a huge amount of financial data has been generated in economic life.

A huge amount of financial data has been generated in the economic life. It is of great research significance to obtain effective analysis data among the vast amount of financial data. By analyzing financial data, it is possible to quantitatively evaluate financial investments such as stocks and make correct investment decisions. Emerging computer technology provides new tools for data collection and analysis, and the Python language can realize the screening and analytical processing of financial data through data mining, thus enabling investors to more accurately grasp the operation of financial markets. In the financial activities of the modern economy and society, the securities market has an extremely important position. In terms of financial data, securities trading activities such as stocks can provide rich data with diversity and openness for data mining. In this paper, we use the Internet environment, financial securities market stock, and other trading data as the source of analysis data, and develop financial data analysis tools using the Python language to analyze and process financial data.

2. Related Theories

2.1. Markowitz Portfolio Theory

Harry Markowitz's paper "Portfolio Selection" uses probability theory and solving quadratic programming to solve portfolio selection problems is the birth of modern asset portfolio management theory [3]. Under the assumption that investors are risk averse and seek to maximize expected returns, Markowitz's portfolio theory consists of both the mean-variance model and the efficient frontier theory. The theory proposes to quantify the expected return and investment risk of a portfolio in terms of mean and variance, respectively, and the goal of investment decision is to find the portfolio with the lowest investment risk at the same level of return or the highest return at the same level of risk, and the portfolio located on the efficient frontier satisfies both conditions.

2.1.1. Mean-variance model

Assuming that the investor invests in a portfolio consisting of n risky assets in a single investment period, and r_i denotes the expected rate of return on the ith asset, the expected rate of return on the asset portfolio is:

$$E(r_p) = \sum_{i=1}^{n} x_i E(r_i) \qquad (1)$$

where denotes the investment weight of the ith asset. Denoting the variance of the ith asset by σ_i^2, the variance of the portfolio consisting of n assets is:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i\ x_j cov(r_i, r_j) = \sum_{i=1}^{n} x_i^2\ \sigma_i^2 \sum_{i=1}^{n} \sum_{j=1}^{n} x_i\, x_j \rho_{ij} \sigma_i \sigma_j \qquad (2)$$

Where i , j denote different assets, is the covariance between asset i and asset j, which i measures the linkage of two asset returns. ρ_{ij} is the correlation coefficient between asset i and asset j, which can be used to compare the magnitude of the correlation between two assets. σ_i and σ_j denote the standard deviation of asset i and asset j, respectively. According to equation (2), it can be seen that the risk of the portfolio mainly depends on the investment weight of each asset, the correlation coefficient between different securities, and the standard deviation of each asset. Therefore, we should prefer to choose assets with lower variance and lower correlation coefficient between two assets to build a portfolio and thus reduce the investment risk. In practice, the sample means and sample variance of past return data is usually used to estimate future returns and risks.

2.1.2. The effective boundary of the asset portfolio

All possible combinations of assets in a portfolio constitute the feasible set, which is shaped like a left convex solid region. The feasible set can be divided into two parts centered on the minimum variance asset portfolio point, where the portfolio located at the upper half of the borderline satisfies the conditions of both minimum risks at a given level of return and maximum return at the same level of risk, which is called the efficient boundary of the portfolio.

2.1.3. Sharpe ratio

In 1966, Sharpe proposed the Sharpe ratio as a risk-adjusted measure of fund performance[4]. The Sharpe ratio is the ratio of the portfolio's excess expected return to the overall standard deviation and is calculated as:

$$S_p = \frac{E(r_p)-R_f}{\sigma_P} \tag{3}$$

Where S_p denotes the Sharpe ratio, σ_P is the overall standard deviation of the portfolio, and R_f is the risk-free rate. indicates how much excess return the asset portfolio can have for each additional unit of risk taken. S_p integrates return and risk, and this indicator will also be used as a criterion for judging portfolio merit in the empirical study section of this paper.

2.2. Research Content

The main problem of this paper is to illustrate an example of the application of Markowitz portfolio theory to optimal portfolio selection in the stock market using the Python language, which offers great convenience for solving financial problems with its complete data analysis suite and program packages that can retrieve financial data such as stock prices. Using Python statements to solve the optimal portfolio can effectively solve the problems such as the complicated calculation of covariance and correlation coefficients among the assets. It is shown that the optimal portfolio can be computed quickly in Python, and the new technique is valuable for the application of Markowitz's portfolio theory in the Chinese financial market.

3. Empirical Analysis

3.1. Data Selection

This paper selects data related to the constituent stocks of FTSE A50 in different sectors and performs data processing based on Python.

3.1.1. Industry of the selected company

In this paper, we mainly use Python software to obtain stock data, perform relevant processing on stock data, and draw relevant conclusions. The research objects are selected from stocks of different industries in the Chinese stock market, as shown in Table 1.

Companies in the financial, automobile, securities, industrial metal, communication services, and real estate industries were selected to diversify risk by selecting stocks in different industries. In this paper, we mainly study the stock portfolio under the same market, so we did not select stocks from different countries to diversify risk, and according to the papers of Wu, Shi-Nong [5] and Yang, Ji-Ping [6], stocks were selected from 4-6 that can play an adequate effect of risk diversification, so one stock was selected from each industry, and a total of 6 stocks were selected for analysis.

Table 1. Industry of the selected company

Industry / Company	Finance	Automotive	Securities	Industrial Metals	Communication	Property
China Merchants Bank	√					
SAIC Group		√				
Haitong Securities			√			
Capital Mining				√		
China Unicom					√	
Poly Development						√

3.1.2. Selected company financial data

Companies with different betas are selected and the relevant financial data for the six companies are given in Table 2, along with their industry and market capitalization and beta.

Table 2. Selected company financial data

	Industry	Market Value	Beta	P/E ratio	Revenue	Earnings per share	Gross margin TTM	Net margin TTM
600036.SH	Finance	1.28T	1.04	11.60	232.51B	4.38	NA	35.05%
600104. SH	Automotive	242.05B	0.74	10.14	796.18B	2.07	12.56%	4.56%
600837.SH	Securities	159.91B	1.40	11.30	44.86B	1.08	NA	34.66%
601899.SH	Industrial Metals	267.78B	1.67	20.17	210.04B	0.50	12.99%	8.19%
600050.SH	Communication	122.73B	0.92	19.07	322.97B	0.21	24.63%	4.52%
600048.SH	Property	172.73B	0.90	5.86	264.19B	2.46	23.64%	15.57%

3.1.3. Statistical description of the selected data

We use the stock closing price data of these six companies from October 1, 2020, to October 1, 2021, to correlate and obtain the return data, which are described in Table 3.

It is possible to obtain 241 return data per stock for six stocks, with standard deviations of 0.02131, 0.02410, 0.01669, 0.03252, 0.01229, and 0.02088 for China Merchants Bank, SAIC, Haitong Securities, Capital Mining, China Unicom, and Poly Development, respectively.

Table 3. Stock Return Data Description

	600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
Size	241.00000	241.00000	241.00000	241.00000	241.00000	241.00000
Min	-0.06350	-0.07752	-0.08788	-0.07906	-0.03526	-0.06867
Max	0.06854	0.09531	0.05697	0.09546	0.03965	0.08098
Mean	0.00138	-0.00015	-0.00071	0.00201	-0.00075	-0.00056
Std	0.02131	0.02410	0.01669	0.03252	0.01229	0.02088
Skew	0.11926	0.76435	-0.53533	0.05982	0.41421	0.93885
Kurt	0.62413	2.65021	5.51804	0.19442	1.06234	2.50878

3.2. Statistical tests

Using Python's data crawling function we can easily obtain the closing prices of these six stocks for 242 trading days between 2020-10-01 and 2021-10-01 and plot the images as in Figure 1.

The trend of their stock prices can be known in Figure 1. During the period from October 1, 2020, to October 1, 2021, China Merchants Bank and Haitong Securities oscillate higher, while the remaining four stocks oscillate slightly lower.

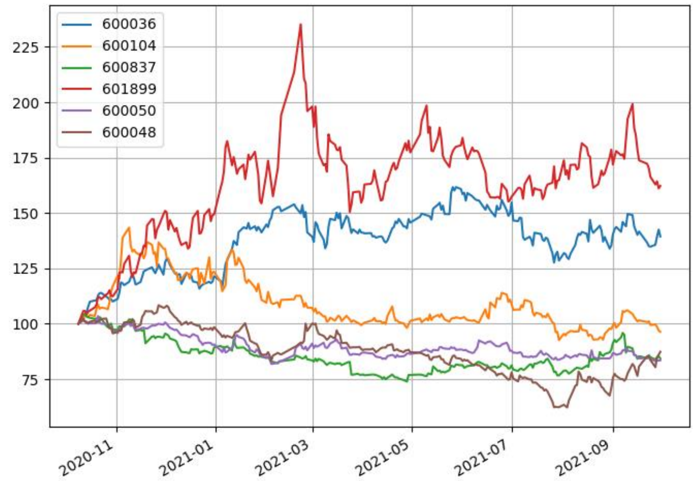

Figure 1. Closing price chart

Then, after calculating their log returns, the distribution of each stock's return is counted to plot the frequency of return, as in Figure 2.

From the log return frequency chart, we can observe that Poly Development deviates more from the center of the normal distribution compared to the other five stocks, and the oscillation is greater while China Merchants Bank has a higher frequency of small declines and a higher frequency of moderate increases, and the remaining four stocks have a more balanced normal distribution chart.

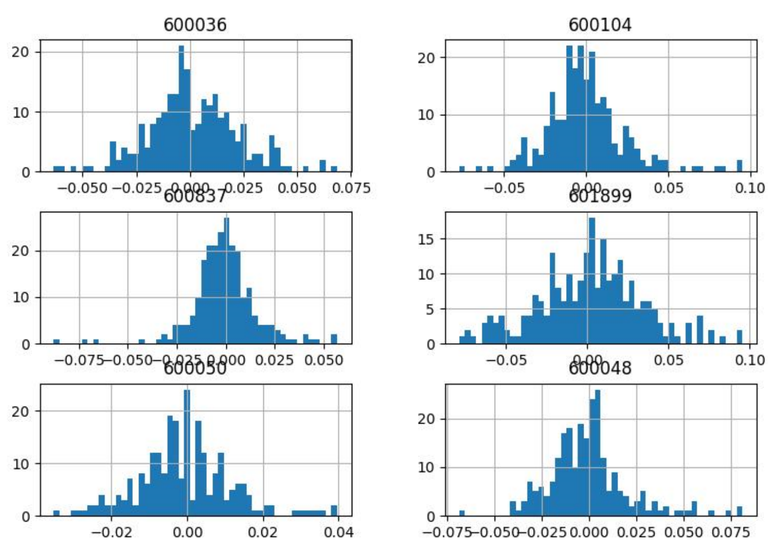

Figure 2. Yield frequency chart

A QQ plot (Quantile-quantile plot) is used to examine whether the log returns (log returns) of the six stocks conform to normality.

The Q-Q plots of the log-returns of the six stocks in Figure 3 show visually that the quantile values of the China Merchants Bank, Capital Mining, and China Unicom samples are on a straight line and roughly normally distributed, while SAIC, Haitong Securities, and Poly Development are not on a straight line, so all three stocks do not conform to a normal distribution. And there are many values far below and far above the straight line on the left and right sides, respectively. In other words, this time series information shows the "thick tail phenomenon", i.e. there are more positive and negative outliers observed in the frequency distribution than the normal distribution should show.

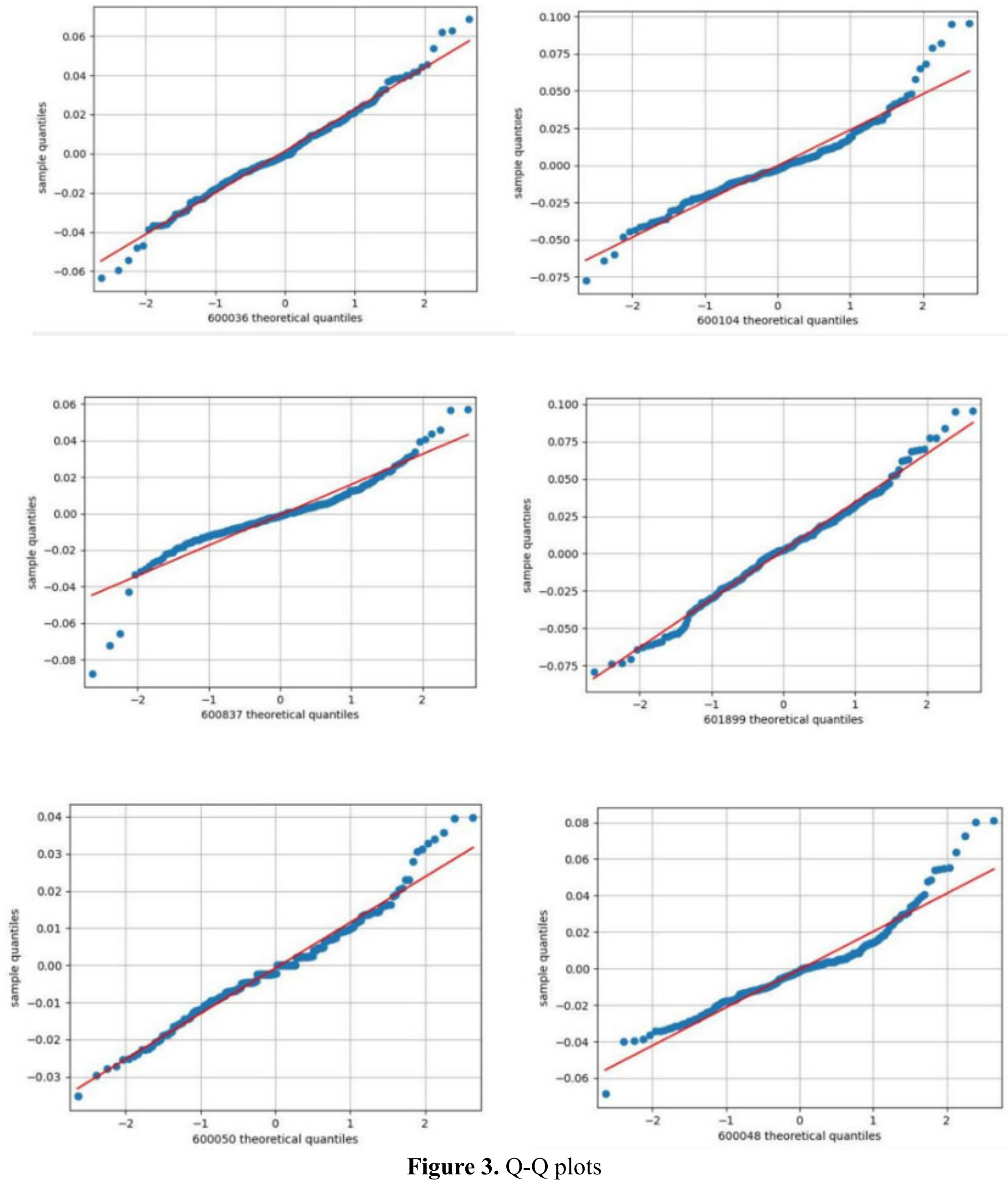

Figure 3. Q-Q plots

3.3. Optimal Portfolio Calculation

The next step in portfolio optimization is to perform the analysis of optimal portfolio proportions based on mean-variance theory (MPT), using the logarithmic rate of return described above.

We obtained the results as follows and plotted the results in a table, as in Table 4, Table 5, and Table 6.

Through the analysis we get the annualized return of each stock, and the covariance matrix, the covariance matrix is a very important step, through the calculation we derive the weight proportion of this portfolio position, and we allocate the position to invest according to the investment proportion can be a good diversification of risk, namely China Merchants Bank: 0.02877822, SAIC: 0.0226514, Haitong Securities: 0.2383445, Zijin Mining: 0.0863758, China Unicom: 0.45464201, and Poly Development: 0.16920798. and achieve an expected profit of -0.099152569. The expected variance of the portfolio is 0.031126121818943937 and the standard deviation of the portfolio, i.e. the expected volatility, is 0.17642596696332413. It can be found that the variance and standard deviation of the portfolio are well reduced by the portfolio, and the reasons for the negative expected return are: China's blue-chip stocks are in the market from October 1, 2020, to October 1, 2021, The collective downside from October 1, 2020, to October 1, 2021, reaching a state of low valuation. (Davis Double Kill)

Table 4. Annualized Yield

	600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
Annualized Yield	0.347652	-0.038732	-0.178515	0.505856	-0.188096	-0.141302

Table 5. Covariance matrix

Covariance matrix	600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
600036	0.114426	0.032692	0.041946	0.046187	0.016893	0.046095
600104	0.032692	0.146383	0.018800	0.051759	0.016027	0.012747
600837	0.041946	0.018800	0.070171	0.023818	0.011737	0.023379
601899	0.046187	0.051759	0.023818	0.266573	0.023958	0.020369
600050	0.016893	0.016027	0.011737	0.023958	0.038074	0.022795
600048	0.046095	0.012747	0.023379	0.020369	0.022795	0.109837

Table 6. Expected optimal portfolio

	600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
Investment Weights	0.02877822	0.0226514	0.2383445	0.0863758	0.45464201	0.16920798
Expected return	-0.099152569					
Expected portfolio variance	0.031126122					
Expected volatility	0.176425967					

The following Monte Carlo simulation is used to calculate the expected portfolio return and variance[7-8].

The obtained results are shown in Figure 4. The horizontal axis is the expected variance, the vertical axis is the expected return, and the color represents the Sharpe ratio. It can be seen that the expected variance in the top left position is relatively small, while the expected return is relatively large, and the optimal portfolio is distributed in this position. The following portfolio optimization is carried out to calculate the return and volatility according to the weights.

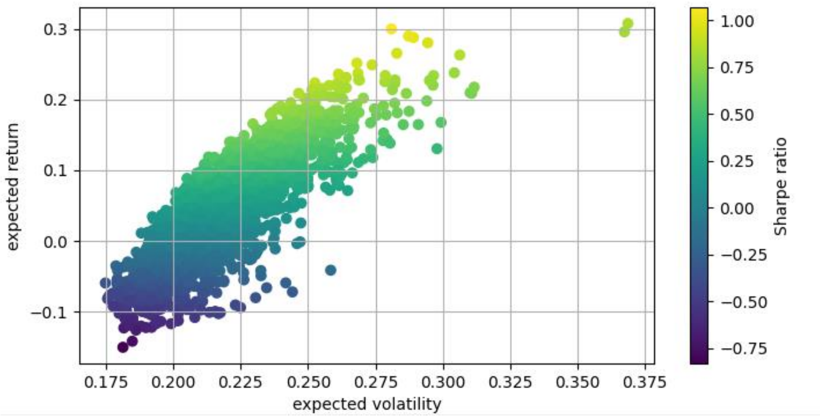

Figure 4. Expected portfolio return and variance

Further, by the Monte Carlo simulation algorithm, we obtain the results in the case of maximum Sharpe index, as shown in Table 7, and Table 8.

We can find that with the maximum Sharpe ratio, we can obtain a return of up to 99.6% in that time interval with a portfolio of that ratio i.e. China Merchants Bank: 0.623, SAIC Group: 0, Haitong Securities: 0, Zijin Mining: 0.377, China Unicom: 0, Poly Development: 0. At this time the volatility of the portfolio is 0.535 and the Sharpe ratio is 1.86.

Table 7. Optimal portfolio weights in the case of maximum Sharpe ratio

600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
0.6230	0	0	0.3770	0	0

Table 8. Optimal portfolio performance with maximum Sharpe ratio

Expected rate of return	Volatility	Optimal Sharpe Index
0.9960	0.5350	1.8600

Further calculate the optimal portfolio performance in the minimum volatility scenario, as shown in Table 9, Table 10.

With minimum volatility, we can still obtain a return of up to 28.8% in that time interval with a portfolio of that ratio, i.e. China Merchants Bank: 0.017, SAIC: 0.087, Haitong Securities: 0.248, Zijin Mining: 0.007, China Unicom: 0.576, Poly Development: 0.066, when the volatility of the portfolio is At this time, the portfolio is defensive (Sharpe ratio is less than 1), with less volatility, but still with a higher return.

Table 9. Optimal portfolio weights for minimum volatility case

600036.SH	600104.SH	600837.SH	601899.SH	600050.SH	600048.SH
0.0170	0.0870	0.2480	0.0070	0.5760	0.0660

Table 10. Optimal portfolio performance with minimum volatility

Expected rate of return	Volatility	Optimal Sharpe Index
0.2880	0.4370	0.6600

3.4. Effective boundary measurement

Based on the above measurement results, the effective boundaries are measured and Figure 5 and Figure 6 are obtained[9-10].

In Figure 5, the yellow * on the left side indicates a portfolio with minimum variance and volatility for a given level of return, and the other red * indicates a portfolio with maximum Sharpe Index.

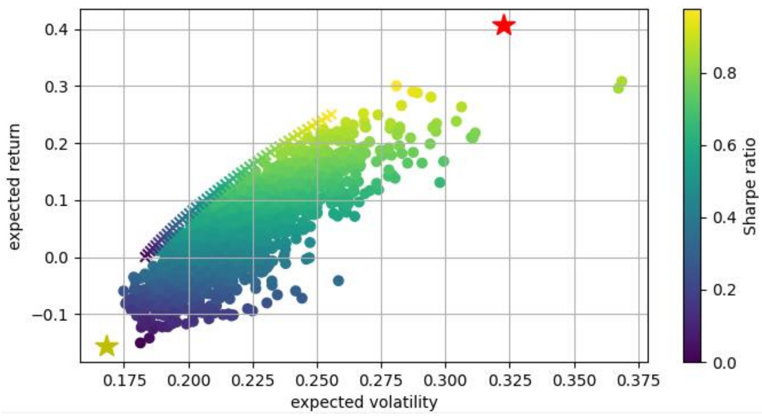

Figure 5. Maximum Sharpe Index and Minimum Volatility Portfolio

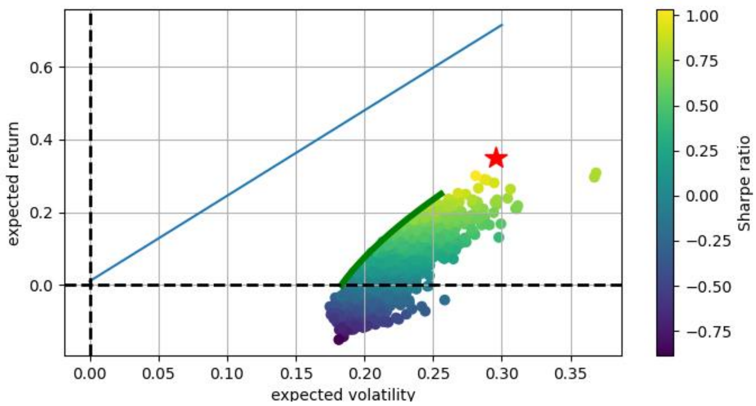

Figure 6. CML

In Figure 6 we can see the CML lines.

4. Conclusion

The experimental results show that using Markowitz portfolio theory, we can find the portfolio with the lowest risk or the highest Sharpe ratio and the efficient frontier in a portfolio consisting of multiple assets. Python largely facilitates the calculation of expected return and portfolio variance in Markowitz portfolio theory, and can quickly find the optimal portfolio, which is of great value for the application of Markowitz portfolio theory in the Chinese financial market. Combining traditional financial theories and emerging programming languages to solve financial problems more quickly and efficiently should be valued by financial analysts and emerging technical talents.

Acknowledgment

Research on the path and risk of digital finance to unleash the consumption potential of residents" (Grant No. CX20221078).

References

[1] N.J. Huang, M.Z. Yu: Advances in Research on the Impact of Machine Learning on Economics Research, Journal of Economic Dynamics, Vol. 12 (2018) No.7, p.115－129.

[2] M. Lngkvist, L. Karlsson, A. Loutfi: A Review of Unsupervised Feature Learning and Deep Learning for Time-Series Modeling, Pattern Recognition Letters, Vol. 48 (2014) No.42, p.11-24.

[3] H.M. Markowitz, H. Harry: Portfolio Selection: Efficient Diversification of Investment, Journal of the Institute of Actuaries, Vol. 119 (1992) No.1, p.243-265.

[4] W.F. Sharpe: The Sharpe Ratio, Streetwise–the Best of the Journal of Portfolio Management, Vol. 10 (1998) No.3, p.169-185.

[5] S.N. Wu, B. Chen: A Theoretical and Empirical Study of Risk Measurement Methods and Financial Asset Allocation Models, Journal of Economic Research, Vol. 12 (1999) No.9, p.30－38.

[6] J.P. Yang, L.J. Zhang: Further Study on the Relationship between Portfolio Size and Risk Diversification in Shanghai Stock Market, Journal of Systems Engineering Theory and Practice, Vol. 12 (2005) No.10, p.21－28.

[7] S.S. Zhu, R. Li, X.Y. Zhou, S.Y. Wang: On Portfolio and Financial Optimization: an Analysis and Reflection on Theoretical Research and Practice, Journal of Management Science, Vol. 6 (2004) No.6, p.1－12.

[8] S.M. Li, P. Xu: Markowitz Portfolio Theory Model Application Study, Journal of Economic Sciences, Vol. 6 (2000) No.1, p.42－51.

[9] J.X. Wu: A Theoretical and Empirical Study of the Effectiveness of ROE on Portfolio Construction - Stock Selection Criteria Based on the Core Concept of "Moat", Journal of Financial Economics, Vol. 12 (2021) No.9, p.59－66.

[10] L.P. Sun: An Empirical Study of Markowitz's Portfolio Theory Based on Python, Times Financial Journal, Vol. 36 (2020) No.25, p.46－67+50.