Stock Portfolio Optimization with Support of Python

Nguyen Thu Thuy[1], *, Vu Thi Thu Huong[1], Hoang Thi Thu Ha[1]

1Department of Digital Economy, Faculty of Mathematical Economics, Thuongmai University

DOI: https://doi.org/10.56293/IJMSSSR.2024.4913

IJMSSSR 2024
VOLUME 6
ISSUE 2 MARCH - APRIL **ISSN: 2582 - 0265**

Abstract: Securities business activities are understood as the activities of organizations, individuals, securities companies, stock brokers and other intermediary financial institutions in the stock market to perform operations. Securities trading aims to make a profit. A common task in portfolio management is to find the optimal investment portfolio according to the return and risk approach, that is, for the same level of risk, portfolio whose highest return gives is the optimal investment portfolio. Similarly, corresponding to the same predetermined rate of return, the portfolio with the lowest level of risk is the optimal investment portfolio. There are several popular data analysis tools today, including Python. This article will demonstrate the application of Python tools in data analysis to determine optimal stock investment portfolios.

Keywords: Portfolio Optimization, Python.

1. INTRODUCTION

An investment portfolio is a shopping cart containing many different investment products, such as stocks, bonds, fund certificates, gold, real estate, etc. Any investment portfolio has a level of risk. The more diverse an investment portfolio is, the lower the risk it is considered to be, since the risk is spread evenly across the market, helping investors minimize losses when one part of the market drops and another part goes up. In each investment portfolio, the allocation of capital to each product also affects the return and risk of the portfolio. Normally, there are two problems for investors: (i) For a given portfolio return, the investor will find capital allocation for assets to determine the risk level of the portfolio which is the lowest; (ii) For a given portfolio risk level, investors will find capital allocation to assets so that the portfolio's return is highest. From there, portfolio management is determining the appropriate capital density for the assets in the investment portfolio to achieve the set investment goals.

Portfolio management is a complex strategy that requires a lot of information, this article provides a reference for investors, with a theoretical perspective on returns and risks of investment portfolio. Data analytic tools in general and stock data analytic in particular provide solutions and support investors in making decisions. Among them, Python is a high-level programming language. Different fields like Web Development, Machine Learning, Business Applications, CAD Applications, Data Science, and Data Visualization use this software. Python focuses on readability. It has a large developer community. This extensive library makes the programmer's job easier in extracting data from the Internet, handling cumbersome computational tasks on big data and represent information in graphical form.

This article will demonstrate the application of Python tools in data analysis to determine optimal stock investment portfolios. The structure of the article is divided into 5 sections: Section 2 presents some theories about investment portfolios, Section 3 shows an overview of the research, Section 4 is devoted to the research results and Section 5 is for the conclusion.

2. PORTFOLIO THEORY

This section presents an overview of portfolio theory, optimal portfolios, and efficient frontiers. On that basis, the Python tool application supports finding the optimal investment portfolio and building an efficiency frontier for an investment portfolio on the Vietnamese stock market.

2.1. Investment portfolio

An investment portfolio is a group of assets such as stocks, bonds or some other assets such as real estate, derivative assets, ... held by an investor or an organization. The most basic purpose of building and managing an investment portfolio is diversification to avoid too large losses. Some commonly used characteristics of investment portfolios are the return (yield) and risk (volatility) of the portfolio. Portfolio P of N risky assets, with $w = (w_1, w_2, \ldots, w_N)$ is the vector of weights of portfolio assets P, $\sum_{i=1}^{N} w_i = 1$.

If call r_i is the return on the asset i ($i = 1, \ldots, N$) then the return of the portfolio P, denoted as r_P is calculated as follows:

$$r_p = \sum_{i=1}^{N} w_i r_i = (w, r)\ , \qquad (2.1)$$

where w and r are the N-asset asset density and return matrices.
Then, the expected return of portfolio P is:

$$E(r_p) = w_1E(r_1) + w_2E(r_2) + \ldots + w_nE(r_n) = \sum_{i=1}^{n} w_i E(r_i)\ . \qquad (2.2)$$

The risk variance of portfolio P is: $\sigma_P^2 = w'Vw$; and the volatility of portfolio P is:

$$\sigma_P = \sqrt{w'Vw}\ , \qquad (2.3)$$

where V is the covariance matrix of the assets i, j, with i, j = 1, …, N. On that basis, we can calculate the covariance between portfolio P and asset k:

$$Cov(r_k, r_p) = \sigma_{kP} = \sum_{i=1}^{N} w_i cov(r_k, r_i) = \sum_{i=1}^{N} w_i \sigma_{ki}\ . \qquad (2.4)$$

and the covariance of the two portfolios P and Q is:

$$Cov(r_P, r_Q) = \sum_{i=1, j=1}^{N} w_i u_j\, cov(r_i, r_j),$$

in which, u_j is the proportion of asset j in portfolio Q.

Portfolio management is building a portfolio of securities and investment assets that best meets the investor's needs and monitoring and adjusting this portfolio to achieve goals proposed by the investor. One of the basic contents in the process of managing a stock investment portfolio is to quantify the relationship between risk and expected return of that portfolio.

2.2. Expected benefits of investors

Whether investors spend money to buy risky assets with return r depends on the characteristics of their benefit function u(r) to determine their risk attitude (risk aversion, risk preference, or risk neutral). Assuming F(r) is the probability distribution function of return r, we have the expected benefit according to the investor's return:

$$U(r) = \int_{-\infty}^{+\infty} u(r)dF(r) \qquad (2.5)$$

If the return function U(r) has a quadratic form, the investor's expected benefit depends only on the mean and variance of the return: U(r) = U(E(r); V(r)). In many cases, people use an expected benefit function of the form:

$$U(r) = E(r) - 0.5A.\ \sigma^2\ ; \qquad A > 0.$$

The above equation shows that the investor's expected utility U(r) increases with the expected return of the asset's return and decreases with the variance of the return. When r has a normal distribution, if we know the expectation and variance, we can completely determine the probability distribution of r, and it can be shown: If investors want high expected returns, the volatility (risk) will also be high, investors who want low volatility must accept low expected profits.

2.3. Model for determining marginal portfolio

Given categories $P:(w_1, w_2, ..., w_N)\ (= W')$ so that:

$$\frac{1}{2}W'.V.W \to \text{Min}$$

$$\begin{cases} \sum_{i=1}^{N} w_i \bar{r}_i = \bar{r}_P \\ \sum_{i=1}^{N} w_i = 1 \end{cases}$$

We can represent the model as:

Identify categories $P:(w_1, w_2, ..., w_N)\ (= W')$ so that:

$$\frac{1}{2}W'.V.W \to \text{Min} \qquad (2.6)$$

$$\begin{cases} (W', \bar{r}) = \bar{r}_P & (2.7) \\ (W', [1]) = 1 & (2.8) \end{cases}$$

The optimization problem (2.6) – (2.8) is a global planning problem and because the set of solutions is compact, it always has a unique solution - denoted by $P(\bar{r}_P)$: $W(\bar{r}_P)$.

Solving the problem entirely using the Lagrange method, we get:

$$W(\bar{r}_P) = g + \bar{r}_P h \qquad (2.9)$$

Where $g = \frac{1}{D}\left[C\left(V^{-1}[1]\right) - B\left(V^{-1}\bar{r}\right)\right]$,

$h = \frac{1}{D}\left[A\left(V^{-1}\bar{r}\right) - B\left(V^{-1}[1]\right)\right]$.

Note that the quantities A, B, C, D (and hence the vectors g, h) depend only on the matrix V and the vector of expected returns of risky assets $\bar{r}$ and both of these factors are determined by the market. In other words, every investor has A, B, C, D, g and h in common.

2.4. Marginal portfolio

For each given expected return level $\bar{r}_P$. According to formula (2.9), we always uniquely identify the category $P(\bar{r}_P)$: $W(\bar{r}_P)$. Category $P(\bar{r}_P)$ called the marginal portfolio (corresponding to $\bar{r}_P$).

For the frontier category $P(\bar{r}_P)$ we have:

+ Expected return: $\bar{r}_P$

+ Variance: $\sigma_P^2(\bar{r}_P) = \dfrac{A\bar{r}_P^2 - 2B\bar{r}_P + C}{D}$ (2.10)

+ Fluctuation: $\sigma_P\left(\bar{r}_P\right)=\sqrt{\dfrac{A\bar{r}_P^{2}-2B\bar{r}_P+C}{D}}$ (2.11)

- Give $\bar{r}_P \in (-\infty,+\infty)$, set of categories $P\left(\bar{r}_P\right)$ The corresponding set is called the marginal porfolio. From the formula for determining the marginal porfolio (2.9) if:
- Give $\bar{r}_P = 0$ we have $P(0)=g$, So we can consider g as a marginal portfolio with an expected return of 0.
- Give $\bar{r}_P = 1$ we have $P(1)=g+h$, So we can consider (g + h) as a marginal portfolio with an expected return of 1.

Then the formula determines the marginal porfolio $W\left(\bar{r}_P\right)$ can be written as:

$$W\left(\bar{r}_P\right)=P(0)+\bar{r}_P\left[P(1)-P(0)\right] \quad (2.12)$$

2.5. Efficiency frontier

Marginal porfolio $P\left(\bar{r}_P\right)$ has an expected return greater than the expected return of the portfolio. The marginal portfolio has the smallest variance (Minimum Variance Portfolio – MVP) called the *effective portfolio*. So if $\bar{r}_P > B/A$ then $P\left(\bar{r}_P\right)$ is an effective portfolio. The other categories are called ineffective categories. The set of efficient portfolios and MVP is called the efficient frontier (corresponding to the group of risky assets considered).

The model for determining the efficient frontier will be the following optimization problem:
Identify categories $P:(w_1,w_2,....,w_N)(=W')$ so that:

$$\frac{1}{2}W'.V.W \rightarrow Min \quad (2.13)$$

$$\left(W',\bar{r}\right)=\bar{r}_P \quad (2.14)$$

$$\left(W',[1]\right)=1 \quad (2.15)$$

$$\bar{r}_P \geq \frac{B}{A} \quad (2.16)$$

For A and B are determined according to the formula:

$$[1]'V^{-1}[1]\equiv A$$

$$\bar{r}'V^{-1}\bar{r}\equiv C$$

$$\bar{r}'V^{-1}[1]\equiv B$$

and $\bar{r}_P$ is the parameter.

3. LITERATURE REVIEW

Portfolio optimization has always been a complex task in finance and management. Portfolio optimization deals with selection of portfolios in a situation of volatile market. Gunjan & Bhattacharyya (2023) reviewed different classical, statistical and intelligent approaches applied for portfolio optimization and management. A comparative study of different techniques were presented in this paper.

Agarwal & Muppalaneni (2022) asserted that if one invested in one stock, the risk would be more; however, one reduced the risk by diversifying the portfolio. To diversify the risk, investors need a strong portfolio of fundamentally strong stocks. Efficient Frontier was one of useful tool for considering stock porfolio optimization. Pla-Santamaria & Bravo (2013) created efficient funds appealing to a sector of bank clients, the objective of minimizing downside risk is relevant to managers of funds offered by the banks. The authors applied the mean-semivariance efficient frontier model to select portfolio of stocks when the investor was especially interested in the constrained minimization of downside risk measured by the portfolio semivariance. The mean-semivariance efficient frontier model was applied to an actual case of portfolio choice from Dow Jones stocks with daily prices observed over the period 2005–2009.

Mehrjoo et al. (2014) reviewed the concept of Efficient Frontier (EF), an important inadequacy of the Variance based models for deriving EFs and the high necessity for applying another risk measure, the traditional risk measure of Variance was replaced with Lower Partial Moment (LPM) of the first order. One part of the paper was devoted to a methodology for deriving EF on the basis of the new model. Then the model superiority over the old one is shown and finally shape of the new EFs under different situations is investigated. At last, it is concluded that application of LPM of the first order in financial models in the phase of deriving EF is completely wise and justifiable.

One can see that Efficient Frontier is useful in considering stock portfolio optimazation. Refering to tools, such as Python, in supporting stock portfolio optimazation problem, Hatemi-J & Mustafa (2023) presented a software component produced in Python by the authors that constructs the optimal portfolio using two alternative approaches. The first standard approach relies on obtaining the optimal weights via the minimization of the portfolio's variance as a measure of risk. The second approach as alternative method mixes return and risk explicitly in the objective function and gives the needed portion of investment for each asset that results in the highest possible return per unit of risk. The current technique is coherent with the rational expectation that investors combine return and risk once decisions on investment are made.

In agreement of that, Zhang (2023) considered Python as the points of fast calculation speed, open source, and excellent data visualization. In Zhang (2023), 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.

With a similar idea as Zhang (2023), in this paper, the authors employed Python as an useful tool in constructing efficient frontier of a Vietnamese stock portfolio through simulation with 1000 random weight for portfolio.

4. EMPIRICAL RESULTS

Research data on the Vietnamese stock market is taken from one of the world's top 4 financial information portals, according to Alexa rankings, https://www.investing.com/. The selected investment portfolio includes 2 stocks of Joint Stock Commercial Bank for Foreign Trade of Vietnam, stock code VCB and SSI Securities Joint Stock Company, stock code SSI, in the period from January 4, 2021 to September 29, 2023. These are two of many stock codes that are considered the most worthy of investment in 2024[1]. The following chart in Figure 4.1 provides a preliminary view of the prices of these two stocks. The price series of stocks, after being brought to the same relative benchmark for ease of comparison, is plotted using Python as follows.

[1] https://topi.vn/danh-sach-ma-chung-khoan-dang-dau-tu-nhat.html

Figure 4.1. Graph of price series of stocks by day from 4/1/2021 to 29/9/2023

Source: Authors draw with Python 3

The return series of stocks are calculated according to the formula:

$$R_t = \log\left(\frac{P_t}{P_{t-1}}\right),$$ with P_t is the value of the index at time *t*.

Descriptive statistics are given in Table 4.1.

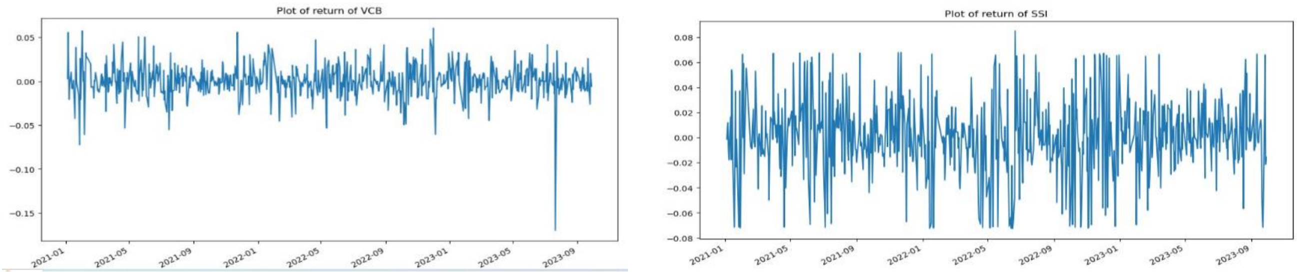
Table 4.1. Statistical description of daily stock index return from 4th Jan 2021 to 29th Sep 2023

	RVCB	RSSI
Mean	0.000194	0.000666
Maximum	0.060625	0.084850
Minimum	-0.170178	-0.072550
Std. Dev.	0.018649	0.030927
Observations	682	682

Source: Authors calculate using Python 3

In general, the average value of both return series is quite close to 0. The average return of VCB stock is 0.0194% and the average return of SSI stock is 0.0666%. Table 4.1 also shows that the average value of the returns is also relatively small compared to the standard deviation of these series. The average value of the VCB return series varies from – 17% to 6.06%, while the standard deviation of the VCB return series is 1.86%. The average value of the SSI return series varies from – 7.25% to 8.48%, and the standard deviation of the SSI return series is 3.09%. The return series graph of the two stocks is shown in Figure 4.2.

Figure 4.2. Graph of daily stock index return series from 4th Jan 2021 to 29th Sep 2023



Source: Authors draw with Python 3

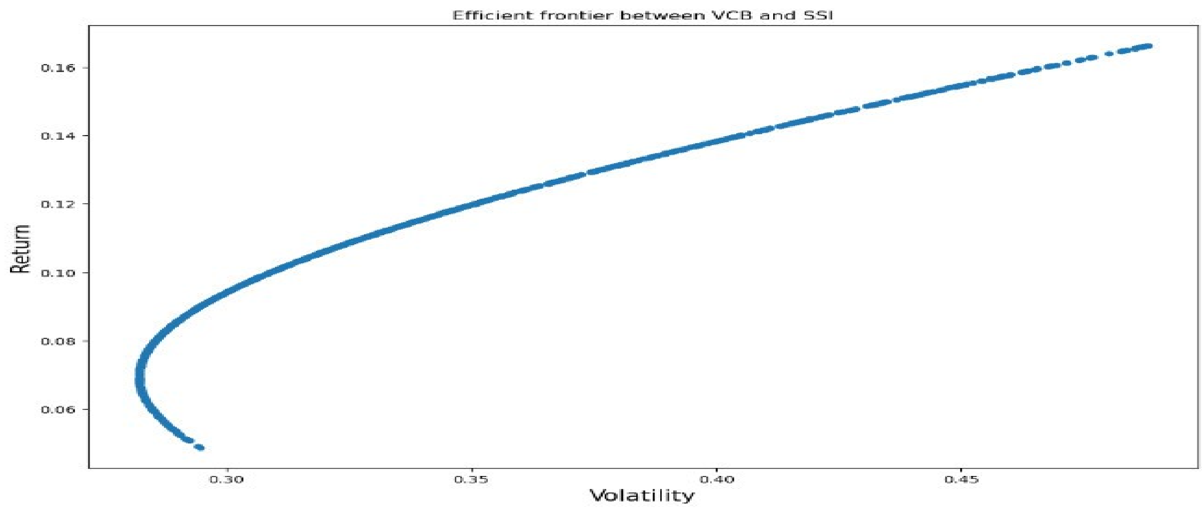
When we create an investment portfolio consisting of 2 stocks VCB and SSI, for each pair of investment proportions in 2 stocks, we can calculate the average return of the portfolio according to formula (2.2) and calculate the volatility of the investment portfolio's return according to formula (2.3). Now to find the optimal portfolio, we will use simulation to create 1000 pairs of weights, for each pair of weights, we will also calculate the average return and volatility of the portfolio. , from there we come to solve the following problem: (i) For each given volatility (accepted by investors), which investment portfolio has the highest rate of return; (ii) For each given rate of return (chosen by the investor), which investment portfolio has the lowest volatility. That is, the highest possible Sharp ratio. As a reminder, the Sharp ratio is a measure of how much profit is earned per unit of risk when investing in an asset or investing according to a business strategy. The Sharpe ratio was developed by William F. Sharpe and is used to help investors understand an investment's return relative to its risk. This ratio is the average return earned in excess of the risk-free return per unit of risk. Formula to calculate Sharp ratio:

$$Sharp\ ratio = \frac{R_P - R_f}{\sigma_P}$$

In which, R_p is the rate of return of the investment portfolio; R_f is the risk-free rate of return; σ_p is the standard deviation of the portfolio's excess return.

To perform that simulation, we use the random generator function in Python. Then use the average return and volatility of the portfolio for the 1,000 different weighting pairs to draw the portfolio's efficient frontier, as shown in Figure 4.3.

Figure 4.3. Efficient frontier of investment portfolio including VCB and SSI in the period from 4th Jan 2021 to 29th Sep 2023



Source: Authors draw with Python 3

The points on the efficient frontier are the points that describe the optimal investment portfolio. Observing the efficient frontier, we see that suppose at a point on the efficient frontier, which corresponds to the vertical axis, the rate of return is 14%, and corresponding to the horizontal axis, the volatility is 41%. That is, if we accept a risk level of 41%, we can get a portfolio return of 14%, but we do not choose portfolios with the same risk level that have a return rate of profitability that is lower than 14%. It means that you should not choose points below the efficient frontier. Similarly, if we want the return rate of our investment portfolio to be 14%, then we only accept a risk level of 41%, but we do not choose portfolios with the same return rate that have higher level of risk. In other words, with the portfolio return of 14%, the risk which is higher than 41% - lying below the efficient frontier - should not be chosen.

5. CONCLUSION

In this article, the authors illustrated the selection of the optimal investment portfolio by simulating 1000 weight pairs for a portfolio of 2 stocks, and received the results as in Section 4.

For a given group of N risky assets, there always exists an efficiency frontier described by the solution set (according to parameters $\bar{r}_P$) of the problem (2.13) – (2.16). The symbol for this solution set is P(Eff). If the investor is risk-averse with the expected benefit function $U(P) = U(\bar{r}_P, \sigma_P^2)$ When choosing a group of N risky assets to invest in, you will choose an efficient portfolio (choose a point on the efficient frontier) by solving the problem:

$$\begin{aligned} &\mathrm{Max} U(P) \\ &P \in P(\mathrm{Eff}) \end{aligned}$$

We can analyze the above problem to see that at the optimal portfolio P* the level curve of the expected benefit function U(P) will touch the efficiency frontier. Thus, the efficiency frontier can be considered as the optimal portfolio "curve" for investors to choose. Depends on the level of return trade-off $\bar{r}_P$ and risks σ_P^2 Different investors will choose different efficient portfolios. If more risky assets are added to the original group of N assets (N' > N), the efficiency frontier will remain the same but will shift to the left because the problem of determining the efficiency frontier of N assets is the field private case of N' properties, where $w_i = 0$ where i is the complementary property. Then along with the expected return $\bar{r}_P$. An efficient portfolio of N' assets will have less risk. However, adding additional assets will make the calculation process more complicated.

In the case of prohibiting short selling of assets, to determine the MVP portfolio and the efficient portfolio we only need to add conditions $w_i \geq 0, i = 1 \div N$. However, determining the efficiency frontier will be more complicated because we do not have explicit analytical formulas to calculate the characteristics. The efficiency frontier in the case of banning short selling will be lower and less "curved" than in the case of allowing short selling. Then the same expected return $\bar{r}_P$ An efficient portfolio in the presence of short selling will be less risky. In reality, when encountering this situation (like the Vietnamese stock market), determining the effective portfolio of a group of risky assets can be done in two ways: (i) Establishing and solving the planning problem corresponding omnidirectional convexity (with conditions $w_i \geq 0, i = 1 \div N$) or (ii) Use the Single Index Model. These problems will be implemented by the authors in subsequent studies.

Acknowledgement: this research is supported by Thuongmai University, Hanoi, Vietnam (Grant number: CS23.62), titled "Research on main tools in business data analytics and application in teaching." Introduction to Business Analysis".

REFERENCES

1. Agarwal, S., Muppalaneni, N.B. Portfolio optimization in stocks using mean–variance optimization and the efficient frontier. *Int. j. inf. tecnol.* 14, 2917–2926 (2022). https://doi.org/10.1007/s41870-022-01052-2.
2. Gunjan, A., Bhattacharyya, S. A brief review of portfolio optimization techniques. *Artif Intell Rev* 56, 3847–3886 (2023). https://doi.org/10.1007/s10462-022-10273-7.
3. Hatemi-J, A. & Mustafa, A. A Python Module for Selecting the Number of Assets in Optimal Portfolios via Two Alternative Techniques, *2023 9th International Conference on Optimization and Applications (ICOA)*, AbuDhabi, United Arab Emirates, 2023, pp. 1-6, doi: 10.1109/ICOA58279.2023.10308823.
4. Mehrjoo, Sh.; Jasemi, M. & Mahmoudi, A. A new methodology for deriving the efficient frontier of stocks portfolios: An advanced risk-return model. *Journal of AI and Data Mining*. Vol. 2, No. 2, 2014, 113-123.
5. Pla-Santamaria, D., Bravo, M. Portfolio optimization based on downside risk: a mean-semivariance efficient frontier from Dow Jones blue chips. *Ann Oper Res* 205, 189–201 (2013). https://doi.org/10.1007/s10479-012-1243-x.
6. Zhang, A. Portfolio Optimization of Stocks – Python-Based Stock Analysis. *International Journal of Education and Humanities*, Vol. 9, No. 2, 2023. 32-38.

