

Portfolio Design Based on Bigdata Analysis and Sharpe Ratio Optimal model

Zhirong Piao^{1, a, †}, Guohua Wu^{2, *, †}, Lu Zhou^{3, b, †}

¹Faculty of Business and Economics University of Melbourne, Australia

²Faculty of Business and Economics Monash University, Melbourne, Australia

³Bachelor of Commerce the University of Queensland, Australia

*Corresponding author: gwuu0008@student.monash.edu.au, ^apiaoz@student.unimelb.edu.au,

^bl.zhou4@uqconnect.edu.au

[†]These authors contributed equally

Abstract. Sharpe ratio is a classic index that can comprehensively consider both return and risk. At the same time, it is only used to investigate the asset of the final venture capital portfolio. Therefore, Sharpe ratio is an important research topic in the financial field. This paper is based on big data analysis and minimum variance model. Thereinto, the companies selected in these five industries are AVIC Electrical Systems Co., Ltd., Archer Daniels Midland Company, Bank of America Corporation, Eastman Chemical Company and China XD Plastics Company Limited. Sharpe ratio is a representation of unit risk and return on investment. The greater the Sharpe ratio, the higher the unit risk return of this portfolio. Secondly, the data we use comes from Yahoo Finance, in which we select the year from 2015 to 2021, because the data of these seven years are more accurate. Meanwhile, the tool we use is RStudio, because RStudio is very excellent in graphics and charts. According to the analysis, these results indicate that one can find the portfolio with the highest risk return rate in these seven years. These results shed light on portfolio design based on bigdata analysis in terms of optimal approaches.

Keywords: Bigdata Analysis; Sharpe Ratio; Risk Management; Optimal; Variance Minimum; Portfolio.

1. Introduction

The Sharpe ratio proposed by Nobel laureate William F. Sharpe aims to help investors comprehend the return of an investment compared to its risk [1]. With the development of economy and information technology, it is widely used in finance and economics field. Simultaneously, a lot of investors use it to help make investment decision, e.g., the Sharpe ratio is useful in making asset allocation decisions comparing portfolio. [2]. In addition, the Sharpe ratio is observably applied in fund-ranking [3].

Generally, most of the investors are risk averse. On this basis, in order to reduce the risk, the portfolios which could well diversify the risk are suitable for them. The main investment logic behind it is that do not put all eggs in the same basket, i.e., hold some high volatility assets and low risk at the same time. They can diversify the risk by buying portfolio. However, the way to choose a profitable portfolio is needed to discuss. This article will use Sharpe ratio to show how to design a reasonable portfolio. There are five different assets which are electromechanical industry, food processing company, bank, chemical field and plastics material in this portfolio. Sharpe ratio method is used to choose the best portfolio. Meanwhile, Sharpe ratio has been proved to be useful in portfolio. Leung and Wong states that the multivariate Sharpe ratio as a tool could help investors analysis their portfolio performances and make a good decision in their investment [4].

Additionally, Bodnar and Zabolotsky just consider risk in investment is not good in investment field [5], the different kind ratios (Treynor ratio, Sordino ratio and Sharpe ratio) are the most famous suggestions in investment industry rather than VaR. Vinod shows that it is a very famous tool for comparing portfolios by using the Sharpe ratio which is expected return divided its standard deviation [6].

However, there are some limitations for using the Sharpe ratio. Before we use Sharpe ratio, we need assume those five assets are normal distribution, but some special situation cannot be display well. Ziemba shows that the Sharpe ratio is based on mean-variance theory, and therefore it is logical only for quadratic preference or normal distributions [7]. Thus, we need assume the five assets are normal distributions before we design the portfolio. Nevertheless, some problems might be pointed out by non-normal return [8]. Meanwhile, some manager could make a bet when using Sharpe ratio to measure return and risk without legitimately analyzing [9]. In this article, we will use R to carry out analysis and hope to help a rational investor to find a profitable portfolio.

The rest part of the paper is organized as follows. The Sec. II will introduce the data origination and evaluation methods. Subsequently, the results will be demonstrated and discussed and the limitations of the analysis will be presented in the meantime. Eventually, a brief summary will be given in Sec. IV.

2. Data and Method

The data, which contains electromechanical industry, agricultural field, bank, chemical industry and plastic industry is chosen from Yahoo Finance [10]. The data range is 1st January 2015 to 31st December 2021. In electromechanical industry, AVIC Electromechanical Systems Co., Ltd (SZ), which was established in 2000 and came into market in Shenzhen Stock Exchange in 2004, is selected as a sample [11]. This company mainly runs motor vehicle parts and aviation product and parts. Archer-Daniels-Midland (ADM), an American food processing and commodities trading company, was found in 1902 and was listed on New York Stock Exchange in 1924. The sample of bank industry is Bank of American Corporation (BAC) that mainly do investment business and financial services. It was found in 1784 and appeared on New York Stock Exchange in 1973. Eastman Chemical Company (EMN) which was listed on New York Stock Exchange in 1993 primarily run chemicals, fibers and plastics. China XD Plastics Company Limited (CXDC) manufacture engineering plastic, biodegradable plastic and other advanced synthetic plastics.

R is the main tool to analysis the data and design portfolio in this research. The whole sample size is 30. Meanwhile, the range of each assets' weight is -3 to 4. The main method of choosing the best portfolio is Sharpe ratio that could assistant investors understand the return of the investment per risk. There is an assumption that investment returns are normally distributed before using Sharpe ratio to choose portfolio [1]. The mathematic description of Sharpe ration can be given as:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \quad (1)$$

Here, R_p represents the return of portfolio, R_f denotes for risk-free rate, and σ_p is the standard deviation of portfolio's excess return. The criterion of choosing the best portfolio is the highest Sharpe ratio. Meanwhile, the highest Sharpe ratio is the highest slope in return and risk graph. Hence, we can quickly pick the best portfolio by comparing different slope.

3. Results and Discussion

The Fig. 1 demonstrates the result of the portfolio design in a standard deviation and mean space, with five risky assets. The entirety of the green region in this figure refers to the opportunity set after the portfolio design procedure. The upper segment of this region is the efficient frontier sought by rational investors and consists of all possible portfolios which stand out in a mean-variance analysis. When there are only two assets involved in the portfolio design, the efficient frontier is shaped like a parabola with its axle parallel to the horizontal (i.e., standard deviation) axis. As more assets are added into the portfolio until there are five assets being considered eventually, the initial parabola is recombined with each of the subsequent assets to result in such a shape in this diagram, which is fundamentally a synthesis of various parabolic curves.

The points in Fig. 1 at the middle-left corner of the opportunity set refer to the five risky assets being analyzed and included in the portfolio. It is noticeable that these initial assets take account of only a tiny proportion of the larger green opportunity set. As the weights of assets in the portfolio design are chosen arbitrarily to be between -3 and 4, when this wide range of weights are applied to the five assets, there are potentially a variety of available portfolios that can be strategically designed, and hence the resultant opportunity set is much more extended than the compact area involving only the five single assets. As a result, investors can have not only conservative investment opportunities by investing in only the five assets, but also some extreme investment strategies which lead to higher return as well as higher risk. Based on the formula mentioned before, we can compute the Sharpe ratios for each of those assets and portfolios designed, and part of the results are displayed in Table I, which are the Sharpe ratios of the five initial risky assets and are displayed here for the purpose of demonstrating how these ratios can be calculated for assets and portfolios. Furthermore, we can adopt a graphical representation to understand Sharpe ratios of the portfolios visually, as shown in Figure 2.

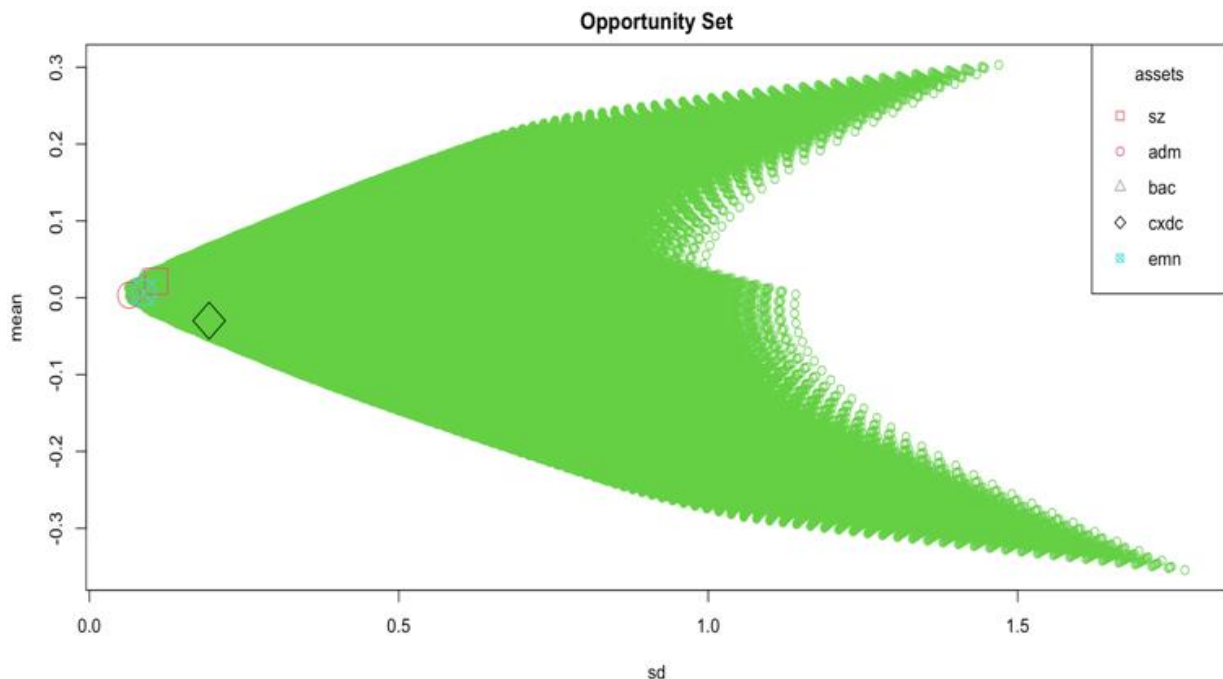


Figure 1. Opportunity Set from the Five-Asset Portfolio Design

Table 1. The Sharp ratio of five assets Between 2015 and 2021.

	SZ	ADM	BAC	CXDC	EMN
Value	0.198	0.053	0.133	-0.154	0.074

A key assumption here is that the riskless rate R_f is chosen to be zero in the calculation. Thus, the vertical axis in the above plot, mean divided by standard deviation, is simply the axis for all Sharpe ratios with their indices displayed on the horizontal axis.

To achieve a point with the maximum possible Sharpe ratio, the numerator of the ratio (mean) should be as large as possible while the denominator (standard deviation) should be as small as possible. However, as the mean return of a portfolio always increases simultaneously in the risk of the portfolio, there is an ambiguous effect on how to have the largest Sharpe ratio with the mean being large and the standard deviation being relatively small. After some empirical experiment by plotting each of the Sharpe ratios as in this diagram, one prominent phenomenon observed is that the magnitude of the Sharpe ratio reaches its maximum at some particular point. This is indeed the optimal portfolio with the highest Sharpe ratio wished by average investors, and we can simply retrieve the mean and standard deviation of this portfolio, as well as the weights of the five individual

assets in this portfolio, by identifying the index of this portfolio among all resultant portfolios in the portfolio design matrix. The weights of the five individual assets in this optimality portfolio are displayed in Table II.

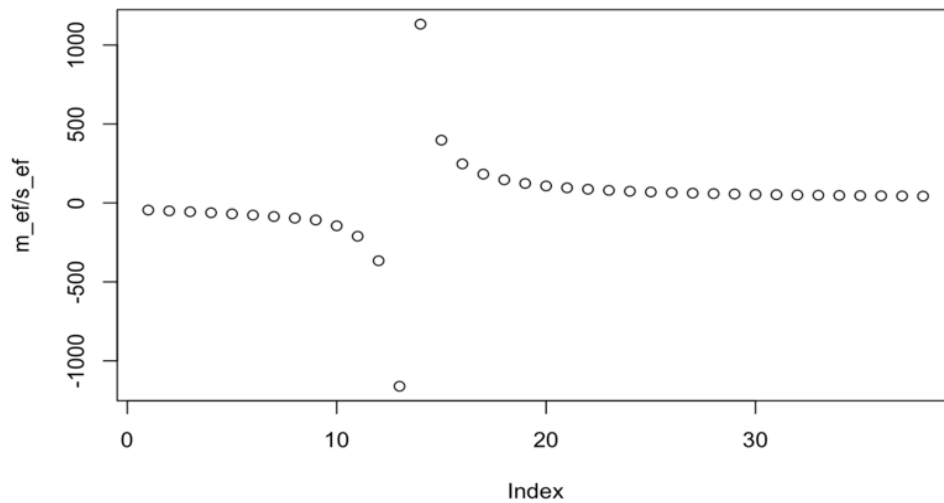


Figure 2. Sharpe Ratios of the Portfolios Designed.

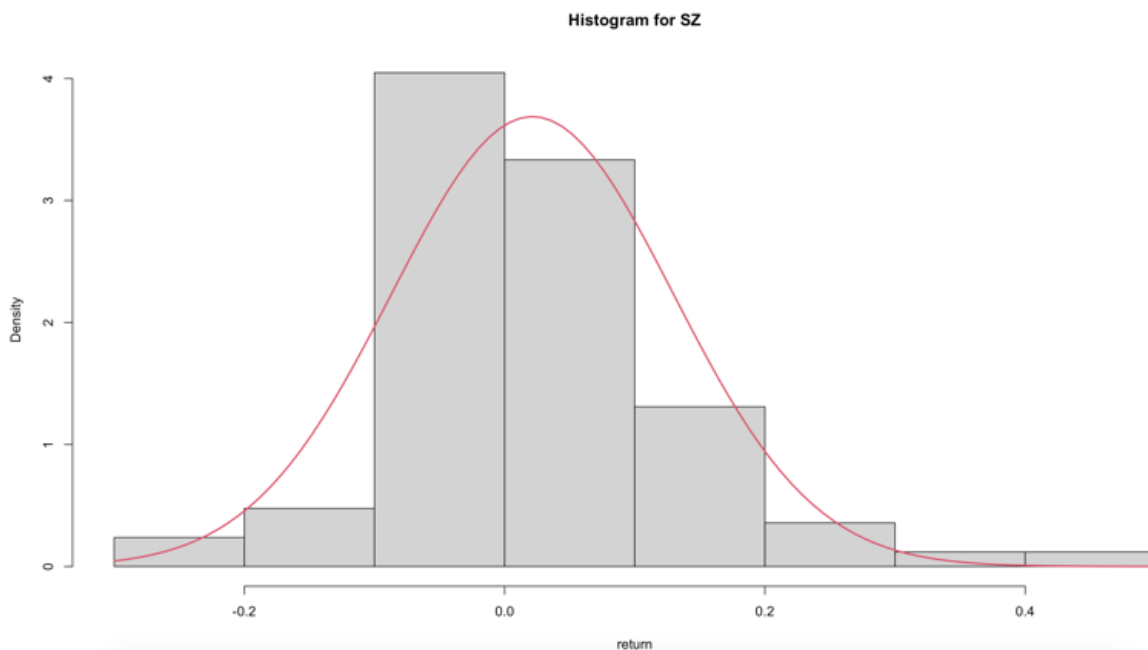


Figure 3. Histogram of AVIC NPV Systems Co., Ltd.

Table 2. The weights of the five assets when the sharpe ratio is optimal (highest)

	SZ	ADM	BAC	CXDC	EMN
Value	-3.000	-3.000	-3.000	2.793	7.207

There is also another possible dimension to cope with this Sharpe ratio issue. This numerical ratio can also be viewed as the slope of the opportunity set in the risk-return diagram, since its computation is to divide the variable on the vertical axis (mean) by the variable on the horizontal axis (standard deviation), which is exactly the definition of the gradient of a curve. By drawing some lines tangent to the opportunity set, identifying the particular tangent point with a maximum slope is equivalent to finding the particular portfolio with the highest Sharpe ratio.

According to Fig. 3, the histogram of AVIC NPV Systems Co., Ltd. is a normal distribution. First, the company's data are continuous random variables. Secondly, they are about the symmetrical

distribution centered on the average: 0.021 (total number of data / total value of data). As shown in Fig. 4, the histogram of Eastman Chemical Company is a symmetrical distribution centered on an average of 0.065 (total data / total data value). On the other hand, the company's data are continuous random variables. Therefore, its histogram is normal distribution. Fig. 5 shows the normal distribution of China XD Plastics Company Limited. Meanwhile, the average trading value of China XD Plastics Company Limited is -0.030. Fig. 6 presented that the histogram of Bank of America Corporation is normally distributed in the data of continuous random variables. And they are symmetrically distributed with an average of 0.012 as the center. As shown in Fig. 7, Archer-Daniels-Midland Company is a symmetrical normal distribution centered on the average value of 0.003. In the meantime, the company's data is also a continuous random variable.

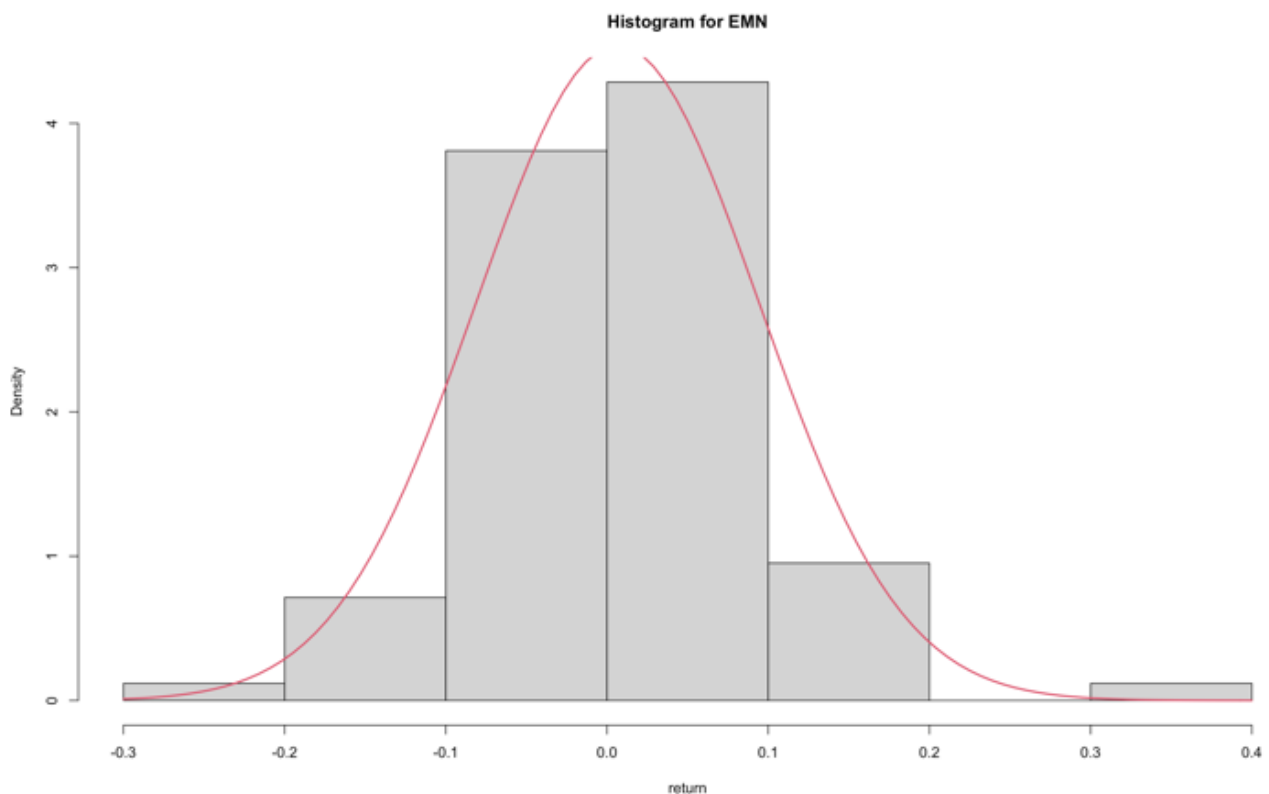


Figure 4. Histogram of Eastman Chemical Company.

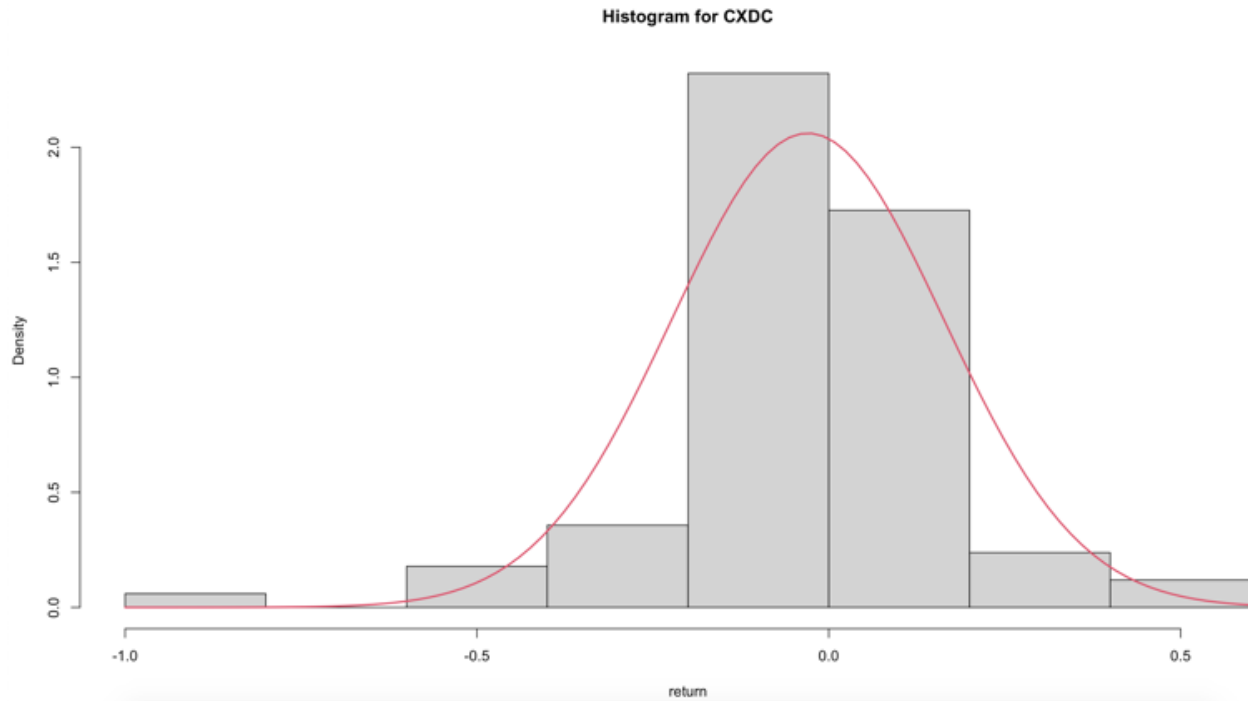


Figure 5. Histogram of China XD Plastics Company Limited.

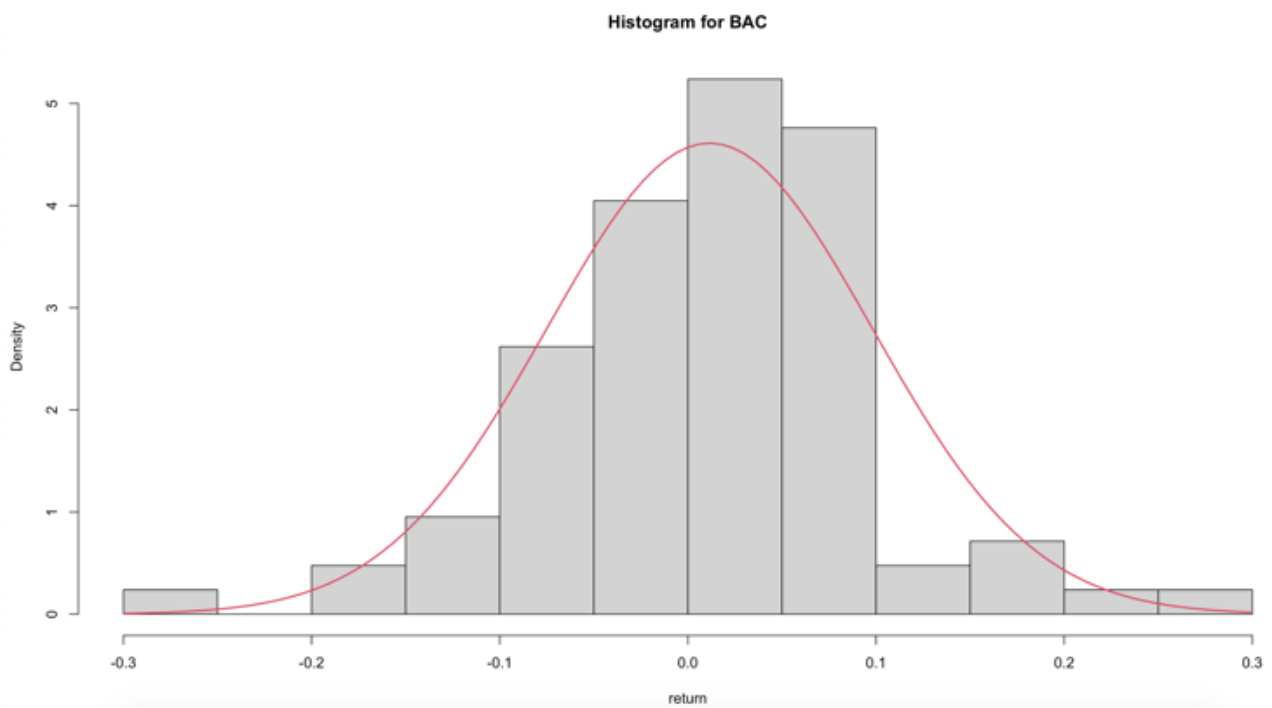


Figure 6. Histogram of Bank of America Corporation.

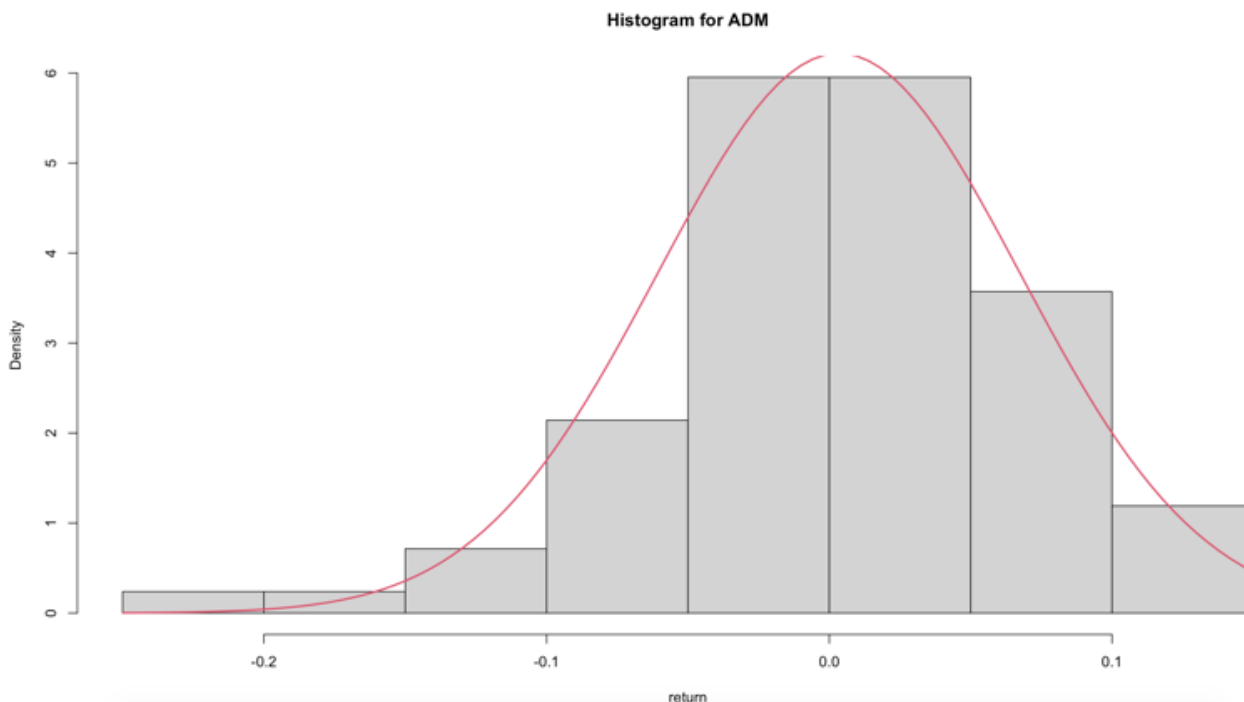


Figure 7. Histogram of Archer-Daniels-Midland Company.

According to Fig. 8, the correlation coefficient between Eastman Chemical Company and Bank of America Corporation is the largest, which is 0.679, indicating that there is a strong correlation between the two assets. In other words, they have a strong linear relationship. On the other hand, the correlation coefficient between China XD plastics company limited and Eastman Chemical Company is very small, only 0.138. This means that the linear relationship between the two assets is very weak.

As shown in Fig. 1, when the weight of assets is selected between -3 and 4, the five assets are combined to obtain the green area in the figure. At the same time, the red dots represent each of the five assets. On the other hand, it can also be found that these assets are close to the effective frontier in the figure. According to Table I, the Sharpe ratio is highest for AVIC NPV Systems Co., Ltd., with a result of 0.198. In other words, for AVIC NPV Systems Co., Ltd. the return per unit of risk is the greatest. It is worth mentioning that Sharpe ratio of China XD Plastics Company Limited. Because the result is negative, the result is -0.154. In other words, the return on this asset is lower than the risk-free rate of return. Fig. 2 presented the set of these optimal portfolios on the boundary. The Sharpe ratio first decreases and reaches the lowest point, then it suddenly rises to the highest point and finally decreases. On the one hand, because the Sharpe ratio of China XD Plastics Company Limited is negative, it means that the risk of this asset is greater than the rate of return. On the other hand, the Sharpe ratios of other assets are positive. This is the reason why the Sharpe ratio suddenly rises to the maximum. Based on the calculations, when the weight of the asset is set between any -3 and 4, the maximum Sharpe ratio of the five asset portfolios is 1132.135. According to Table II, when Sharpe ratio is the highest, the weights of these five asset portfolios are: -3.000, -3.000, -3.000, 2.793, 7.207.

Definitely, these evaluations have some drawbacks. Specifically, there are mainly three kinds of defects that should be addressed in the further study. Firstly, a fundamental assumption for the entire research to be appropriate is the normality assumption. If the returns of the risky assets are not normally distributed, then the methodology will not result in any reasonable conclusion. Thus, the research is limited to being applied to normally distributed data sets, and hence the check of normality is a crucial prerequisite before conducting any subsequent analytical tasks.

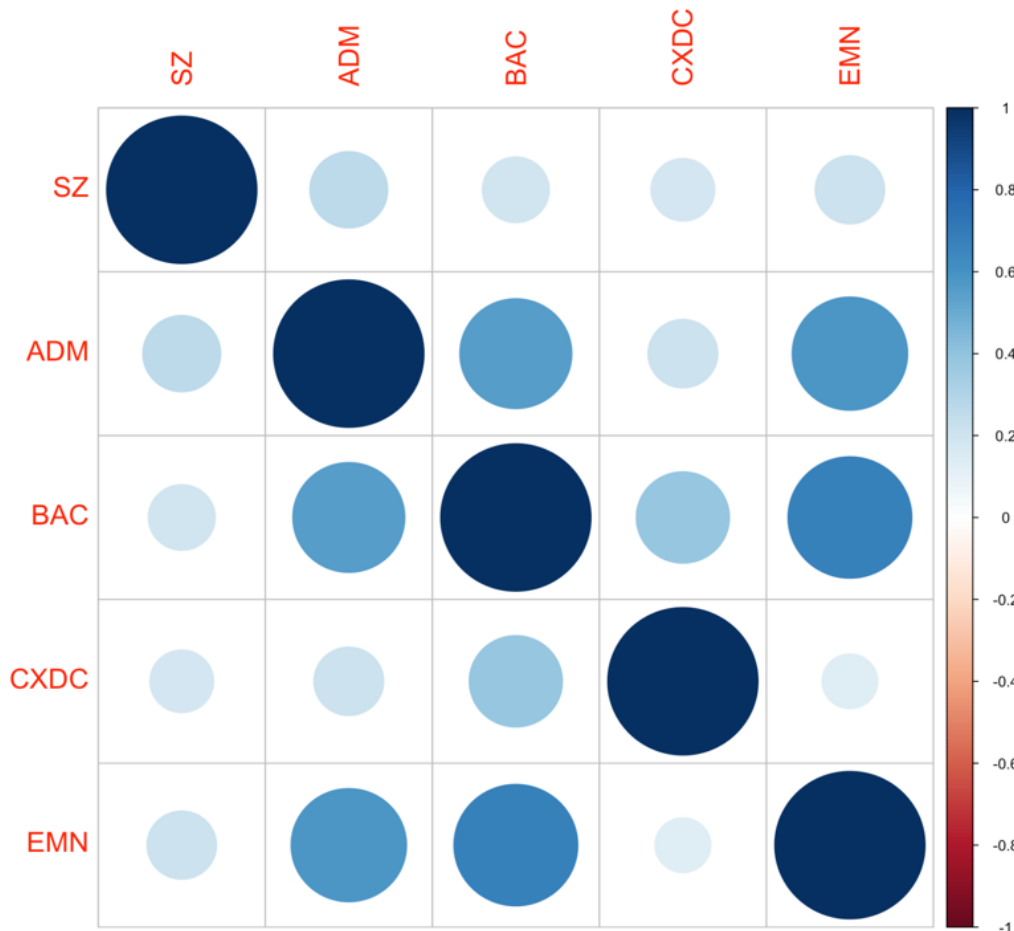


Figure 8. The correlation coefficient between five assets.

Secondly, according to the methodology, we applied all weights in the interval between -3 and 4. Notice here that a negative weight means the investors are taking a short position and are short-selling the particular asset with this negative weight. However, in reality, short-selling an asset might be restricted in the financial and investment market, so in some times, a minimum available weight of zero should be adopted in a portfolio analysis task. Similarly, a maximum weight of one is more relevant where investors only intend to invest a maximum of 100% in any particular asset, which will make more practical sense. In this case, the weights we chose are mainly for theoretical purposes for demonstrating more possible combinations of assets and a variety of differentiated portfolios. Furthermore, the plot of the opportunity set as shown in Fig/ 1, as well as the behavior of the Sharpe ratios as displayed in Fig. 2, are also better presented with these weights between -3 and 4. In this case, the key shape, feature and trend of these graphs are much clearer and easier to interpret.

Lastly, the methodology in this portfolio analysis research project is subject to some technical limitations. In reality, with a huge financial market and thousands of market participants, the possible number of assets to be considered in the portfolio design is much larger than five, which is the number of risky assets that we used in this research. To design portfolios with more and more assets, it will require a computer with much more storage capacity and much higher running speed, in order to implement the relevant block of code as well as to display necessary graphs. A numerical computer that is designed solely for the purpose of dealing with financial issues should be a much better alternative to complete this research. As one is unable to have access to such a computer, one has to reduce the analysis to the scope of five assets and less portfolios. Nevertheless, the results of the limited analysis are still representative enough to give useful insight on the investment market and on how investors can make optimal investing decision with a Sharpe ratio application.

4. Conclusion

In summary, this paper provides thorough analysis, explanation and interpretation about the wide financial market and the behavior of investors participating in this market. In particular, by standing on a Sharpe ratio perspective, the research demonstrates how investors can choose their optimal investment portfolio with the maximum Sharpe ratio, among a large number of portfolios configured with five initial assets. The research firstly shows the definition of Sharpe ratio and the methodology of calculating this ratio. This theoretical introduction is then followed by a real-world application, where the assets of five enterprises are chosen to design portfolios. Graphical approaches have been used to illustrate the portfolio design process, including a representation of the opportunity set and efficient frontier, as well as another pictorial representation of the Sharpe ratios associated with those portfolios designed. Simultaneously, Sharpe ratios and pertinent weights are tabulated to give readers better visual sense and more intuition. Eventually, the research reaches an intuitive conclusion that given the monotonic relationship between risk and return, Sharpe ratio does reach a maximum at a certain level, which leads to the optimal portfolio sought by investors.

Although mean-variance approach has been a prevalent methodology in the past decades in terms of investment analysis, there are still many possible mechanisms to determine the meaning of an optimal portfolio. This can be a portfolio with the highest return, or another one with the lowest risk (i.e., minimal variance portfolio). Sharpe ratio is indeed a comprehensive cardinal value, which combines both return and risk, to figure out the best available portfolio for investors. Although Sharpe ratio still has its drawbacks and limitations, it is definitely a straightforward method that can help junior investors make financial decisions according to their individual investment demand. In terms of the future research prospects of portfolio design and the Sharpe ratio analysis criterion, the general methodology is still indicative and applies to other research papers. Nonetheless, more research can still be carried out by considering different financial markets other than the American market, and by incorporating a larger amount of assets from a variety of differentiated industries. Additionally, some computational methods such as machine learning and data mining can be researched into, which will help address the current issue of computational limitations and facilitate the portfolio design and analysis process. Overall, these results offer a guideline for portfolio construction in terms of Sharpe ratio optimal.

References

- [1] J. Fernando, "Sharpe Ratio," Investopedia, 10 January 2022, retrieved from <https://www.investopedia.com/terms/s/sharperatio.asp>.
- [2] S. Christie, "Is the Sharpe Ratio Useful in Asset Allocation," Macquarie Applied Finance Centre Research Paper, vol. 1, May 2005, pp. 48.
- [3] U. Homm and C. Pigorsch, "Beyond the Sharpe ratio: An application of the Aumann-Serrano index to performance measurement," *Journal of Banking & Finance*, vol. 36(8), August 2012, pp. 2274 - 2284.
- [4] P. L. Leung, and W.K. Wong, "On testing the equality of the multiple Sharpe Ratios, with application on the evaluation of iShares," June 2006, pp.32.
- [5] T. Bodnar, and T. Zabolotsky, "How risky is the optimal portfolio which maximizes the Sharpe ratio," *AstA Advances in Statistical Analysis*, vol.101, May 2016, pp.1 - 28.
- [6] M. R. Morey, and H. D. Vinod, "A double Sharpe ratio," *Advances in Investment Analysis and Portfolio Management*, vol. 8, 2001, pp. 57 - 65.
- [7] W. T. Ziemba, "The Symmetric Downside-Risk Sharpe Ratio," *The Journal of Portfolio Management* Fall, vol. 32 (1), October 2005, pp. 108 - 122.
- [8] G. N. Gregoriou, and J. P. Gueyie, "Risk-Adjusted Performance of Funds of Hedge Funds Using a Modified Sharpe Ratio," *The Journal of Wealth Management* Winter, vol. 6(3), October 2003, pp. 77 - 83.
- [9] R. B. Spurgin, "How to Game Your Sharpe Ratio," *The Journal of Alternative Investments* Winter, vol. 4 (3), December 2001, pp. 38 - 46.

[10] Yahoo Finance, retrieved from <https://au.finance.yahoo.com>

[11] Avic Electromechanical systems co., ltd (China), EMIS, retrieved from <https://www.emis.com>.