

Portfolio Construction Based on Bigdata Analysis in Terms of Minimum Variance Model

Ziyu Guo

Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, China

19307100108@fudan.edu.cn

Abstract. Contemporarily, A-share market keeps being in a trough, hence investors (especially the retail investors) seek ways to expand profit and reduce loss when speculating in stock market, but sometimes fail because of the lack of knowledge in investment. To remedy this problem, this paper introduces a complete process from the concepts of bigdata analysis to figure out investment decisions, including evaluating single stocks and choosing stocks to construct portfolios. The objectives of portfolio construction used in this paper are minimum variance and maximum Sharpe Ratio. The whole article presents an example of making investment decisions in A-share market. In the particular situation, two best portfolios which minimizes variance and maximizes Sharpe Ratio respectively are constructed among the training of abundant data capacity. It is also observed that with different goal of investment the best choice for investors could be greatly different. Investment strategy given in this paper may be helpful for investors, which offers a guideline for constructing portfolio under some objectives.

Keywords: Portfolio Design; Bigdata Analysis; Time-series Analysis; Minimum Variance Model.

1. Introduction

Among past several years, one has witnessed a turbulent situation of A-share market and the data samples of stocks in Chinese market have been boosted larger than an order or magnitude compared to a decade ago. Not only did great declines take place in both Shanghai and Shenzhen board, risk of involving in the stock market has also increased for Chinese investors [1], especially the retail investors who have fewer knowledge about efficient investment compared with institutions [2, 3]. Researches have already shown that Chinese individual investors keep performing terrible in stock market, for the fact that they have no right investment philosophy, know little about current market and each industry. Most time individual investors take the strategy of blindly following the big dealers, what's worse is that they always dream of being instant millionaire by speculation on stocks [4, 5]. As a result, useful and technical investment strategies are necessary for retail investors, which is the research object of this paper.

Seeking to address the issue of retail investors keeping suffering losses in the operation of stock market, some scholars have offered suggestions, e.g., learning more professional financial knowledge from public lectures or online media, studying about current situation of stock market even the whole financial market, understanding industries which they aim to invest in, keeping calm and avoiding being impulse when making decisions [6-7]. However, there is almost no studies about how to effectively take quantitative measurement to evaluate stocks and to construct portfolios for individual investors, especially lacks ones using empirical analysis method to clearly demonstrate these scientific and appropriate investment strategies. Fortunately, academic field of investment has been fully-fledged in the past century by those great financial scientists. 1952, Markowitz first brought up the Mean-Variance model which is known as MV model, considering a portfolio from both return and risk and making a trade between the two issues [8]. 1966, Sharpe went a further step to construct an index which combines a portfolio's expected return and risk to evaluate a portfolio, called Sharpe Ratio [9]. In the following decades, numbers of new theories and models have been put forward, for instance, Value-at-Risk [10] and Conditional Value-at-Risk [11, 12] as a complementary to MV model have been widely used in institutional investment. Indexes (e.g., Sortino Ratio and Maximum Withdrawal Rate) are also considerable in reality. In a word, there are many kinds of method for

investors to consider in an investment decision, making investment decisions including portfolio design has been a part of financial management lessons [13]. Besides these theories, application of computer programming in investment has also been developed mature [14]. Rmetrics, a powerful R package to design and evaluate portfolios could be extremely useful in processing and calculating large volume data [15, 16]. Unfortunately, most of Chinese individual investors couldn't use these excellent tools to make correct decisions, expanding their return in the instantaneously changing stock market.

This paper seeks to give a systematic methodology for individual investment, including both evaluating single stock and structuring a well-performed portfolio, aiming to introduce the basic strategy of investment, thus help Chinese retail investors earn more than before in the future. The rest part of the paper is organized as follows. The second part introduces both data and method used in the following subsections: data part is composed of the source of data and concrete information which makes it possible to reproduce the research result; method part mainly focus on mathematic descriptions of the models. The third part posts the results of this research, giving explanations of the result at the same time. The fourth part discusses the implications of those results; deficiencies and inadequacies of this research will also be shared. The fifth part of the paper comes to the conclusion, apart from the professional quantitative result, some pragmatic suggestions will be given at the same moment.

2. Data and Method

2.1 Data

This paper collects stock trading data from NetEase Finance (<https://money.163.com>) since it is free and easy to download data. Instead of randomly choosing stocks to analyze, 10 industry sectors are selected and from each sector there are 3 leading shares taken to be used. List of industries and stocks chosen are demonstrated in Table 1 and 2, where stocks are noted by their ticker symbols.

Table. 1 Industries and Stocks Chosen I

Sector	Power Equipment	Computer Application	Auto Parts	Chemicals	Special Equipment
Stock1	300750	601360	000338	600309	600031
Stock2	601012	000938	600741	603260	000157
Stock3	600406	688111	600660	000691	000425

Table. 2 Industries and Stocks Chosen II

Sector	Media	Architecture Decoration	Chemical Pharmaceutical	Semi-conductor	General Equipment
Stock1	002027	601668	600276	688981	000039
Stock2	300413	601390	600196	002938	002444
Stock3	002602	601186	000963	603501	300861

Table. 3 Name and Sticker Symbol of Chosen Stocks

Name of Stock	Sticker Symbol
Focus Media (fzcm)	002027
Fuyao Glass (fybl)	600660
Fosun Pharma (fxyy)	600196
Hengrui Medicine (hryy)	600276
East China Medicine (hdy)	000963
Huayu Automobile (hyqc)	600741
Superstar Technology (jxkj)	002444
Sany Heavy Industry (syzg)	600031
Weichai Power (wcdl)	000338
Asia Pacific Industry (ytsy)	000691
Chinese Architecture (zgjz)	601668
Chinese Railway Construction (zgtj)	601186
CIMC Group (zjzt)	000039
Zoomlion (zlk)	000157

With the aim of evaluating each stock from analyzing its long-term performance, it is decided to collect 6-year's data (from 2016-01-28 to 2022-01-28) of each stock. Unfortunately, not all the 30 stocks are able to provide data of 6 years, so this paper just kept those ones which satisfies the condition of time to market. Through further data checking it is observed that there are data of stocks which contains missing data, thus here those incomplete data are removed out of consideration. At last, there exists only 14 stocks that are able to provide complete trading data during the required period, from 2016-01-28 to 2022-01-28, 1462 trading days in total. The list of their names and ticker symbols is in Table III (letters in parenthesis stand for abbreviations of stocks' Chinese names).

Besides the stocks' trading data, situation of the whole A-share market is also of vital importance, as well as the risk-free asset. As a consequence, CSI300 trading data are downloaded to represent the whole A-share market and data of 5-year treasury bond price are collected to represent risk-free asset.

2.2 Method

This section mainly contains three parts. At first there's a short introduction of how to process the data, in order to get the return of each stock. Secondly, it comes forward to the part of method for evaluating a single stock. The last part is about constructing and evaluating stock portfolio.

The risk-free rate is obtained by calculating the mean value of 5-year treasury bond's rate of return during the selected 6 years. For each stock and CSI300, its return of each day is calculated according to its closing price of each day, with the following formula:

$$r_t = 100\% \times \frac{p_t - p_0}{p_0} \quad (1)$$

Here, r_t represents the return of day t , p_t represents the closing price of day t , p_0 stands for the closing price of the starting day (2016-01-28).

With the calculated return data and risk-free rate, one can draw a line graph to directly demonstrate the general situation of the stocks, at the same moment some quantitative indexes of each stock which are useful to evaluate a single stock is also obtained. The index and formulas used are as follows.

(1) Expected Return

Expected return of a stock was estimated by mean value of each day's return, which can be described as:

$$r_e = \bar{r} = \frac{\sum_{t=1}^n r_t}{n} \quad (2)$$

Here, r_e stands for expected return, r_t stands for each day's return and n stands for the total amount of trading days of which data are collected (these mathematical variables stand for the same meaning when appearing in the following formulas).

(2) Risk

Risk of a stock was estimated by standard deviation of all days' return, which can be evaluated as follows:

$$Std = \sqrt{var} = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (r_t - r_e)^2} \quad (3)$$

(3) Sharpe Ratio

Sharpe Ratio of a stock was calculated by the following formula, which can be obtained as:

$$SharpeRatio = \frac{r_e - r_f}{Std} \quad (4)$$

Where r_f stands for risk-free rate.

(4) Beta Coefficient

Beta coefficient of a stock was calculated by the following formula, which can be given as:

$$\beta_a = \frac{Cov(r_a, r_m)}{\sigma_m^2} \quad (5)$$

Here, r_a stands for Stock A's expected return, r_m stands for the whole market's expected return and σ_m stands for standard deviation of the whole market's return (these mathematical variables stand for the same meaning when appearing in the following formulas).

(5) Return calculated by CAPM

According to Capital Asset Pricing Model, return of a stock, already known beta coefficient, could be calculated by the following formula:

$$r_a = r_f + \beta_a(r_m - r_f) \quad (6)$$

Here, β_a stands for Beta coefficient of Stock A.

By the evaluating of each stock, one is able to choose several from all the 14 stocks, aiming to construct well-performed portfolios. Not only are features of single stocks taken into account, but relationships between the stocks are also considerable, for the reason that by choosing stocks correlation among which are relatively weak to construct a portfolio, investing in portfolio instead of single stock can down the risk of investment to a great extent. Here, the equation to calculate correlation between two stocks is given as follows:

$$Corr(a, b) = \frac{Cov(a, b)}{Std(a)Std(b)} \quad (7)$$

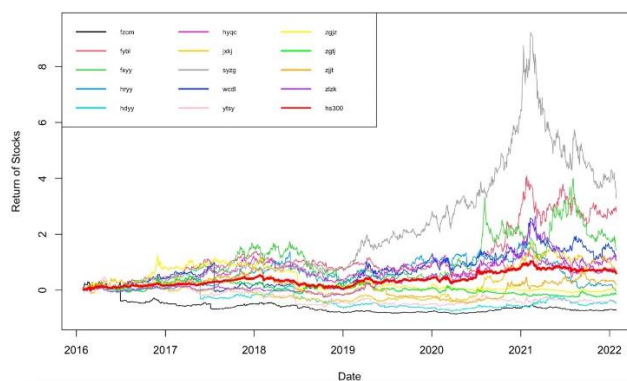
Here, $Corr(a, b)$ stands for the correlation coefficient between return of Stock A and Stock B, $Cov(a, b)$ stands for covariance between Stock A and Stock B, $Std(a)$ and $Std(b)$ stands for standard deviation of Stock A and Stock B, respectively.

Expected return, risk, return per unit of risk are all important indexes to evaluation a portfolio, but this paper only tried to do portfolio design from the view of risk and the Sharpe ratio referring to the fact that there's no short selling in Chinese financial market. Formulas used to calculate a portfolio's risk and Sharpe Ratio are the same as those used to calculate risk and Sharpe Ratio of single stock.

3. Results

3.1 General Performance by Graphing

Fig. 1 is the line graph which demonstrates general return of each stock's fluctuation and trend of the whole A-share market (in the graph, legends are the abbreviations of stocks' Chinese names, shown in Table 2).



3.2 Risk-free rate and Quantitative index of single stock

The risk-free rate calculated by the data mentioned above is about 0.03. There are 5 quantitative indexes calculated in this paper, as mentioned in the second part. In this section the results are together shown in Table 4.

Table. 4 Quantitative Indexes of Stocks

Sticker Symbol	Expected Return	Risk	Sharpe Ratio	Beta	CAPM Return
002027	-0.597	0.237	-2.648	-0.360	-0.100
600660	1.228	0.926	1.293	3.510	1.303
600196	1.057	0.789	1.302	2.788	1.042
600276	0.596	0.431	1.314	0.954	0.376
000963	-0.316	0.291	-1.190	-0.538	-0.165
600741	0.674	0.342	1.885	1.112	0.434
002444	0.154	0.541	0.229	1.756	0.667
600031	2.138	2.023	1.042	7.808	2.862
000338	0.700	0.563	1.189	2.047	0.773
000691	-0.252	0.251	-1.125	-0.381	-0.108
601668	0.303	0.365	0.747	-0.512	-0.155
601186	0.055	0.172	0.146	-0.473	-0.141
000039	0.111	0.283	0.283	0.224	0.112
000157	0.387	0.547	0.651	2.028	0.766

Through the analysis about single stock, it can be observed that some expected returns of the chosen stocks are lower than the risk-free rate, even negative, estimated by their performance in the recent 6 years. Their sticker symbols are 002027, 000963, 000691. Looking at each stock's return estimated by CAPM, one found that except those 3 stocks whose expected return are negative there's another one stock which may bring negative return. Sticker symbol of that stock is 601186. For there may be extremely probable danger occurred in investing in these 4 stocks (whose sticker symbols are 002027, 000963, 000691, 601186), it is determined to remove them out of consideration in further research about portfolio.

3.3 Portfolio Design

The correlation coefficients between each two stocks are calculated as illustrated in Fig. 2. According to the result, one can observe that there are some pairs of stocks which mutually have strong correlating relationship (the standard of strong correlated relationship is set as absolute value of correlation coefficient reaching 0.75). These pairwise relationships are listed in Table 5.

To make it easier to work out with portfolio with smaller risk, the most satisfactory figure is to construct a stock basket which contains highly-correlated pairs of stocks as few as possible. Meanwhile, it is also better that stocks in the basket have better performance than the unselected ones. Along the above principle, 5 stocks are eventually selected, their sticker symbols are 601668 (zgiz), 002444 (jxkj), 600196 (fxxy), 600660 (fybl), 600031 (syzg). Obtaining the above achievement, task of portfolio design shall begin, which is the most important of this research.



Fig. 2 Correlation Heatmap Between Each Two Stocks

Table. 5 Strong Correlations Between Stocks

Stock-A	Stock-B	Correlation Coefficient
000039 (zjzt)	002444 (jxkj)	0.82
600741 (hyqc)	000157 (zlzk)	0.92
600741 (hyqc)	600031 (syzg)	0.86
000157 (zlzk)	600031 (syzg)	0.79
000338 (wcdl)	600031 (syzg)	0.84
600741 (hyqc)	600276 (hryy)	0.75
600031 (syzg)	600276 (hryy)	0.84

As mentioned before, it is meaningless to discuss the expected return of the portfolio because Chinese market bans short selling. Therefore, one focuses on risk and Sharpe Ratio of portfolios. To work out with the portfolio which has the minimal risk, the R package ‘fPortfolio’ is utilized, whose method for use has been introduced detailed in *Portfolio Optimization with R/Rmetrics*. Weight of each stock in the optimal portfolio with the minimal standard deviation and some indexes of the portfolio are attached in Table 6.

Table. 6 Features of Portfolio with Minimal Variance

Sharpe Ratio	Expected Return	Risk	Weight of Each Stock				
1.106	0.306	0.250	fybl	fxyy	jxkj	syzg	zgjz
			0	0	0.291	0.025	0.684

To work out with the portfolio which has the maximum Sharpe Ratio, a paragraph of R codes introduced in *Basic R for Finance* is used. Weights of each chosen stock in the best portfolio which brings the maximum return per unit of risk and some indexes of the portfolio are attached in Table 7.

Table. 7 Features of Portfolio with Maximum Sharpe Ratio

Sharpe Ratio	Expected Return	Risk	Weight of Each Stock				
1.935	0.620	0.305	fybl	fxyy	jxkj	syzg	zgjz
			0.042	0.088	0	0.115	0.755

4. Discussion

Based on the analysis, the best portfolio is obtained based on minimum variance and maximum Sharpe Ratio respectively. It is implied that with different goal of investment, one will reach different

best choice. It should be noted that this paper isn't aiming to work out a best investment decision for investors to follow, but desiring to offer a thinking way of speculating in stock market which contains evaluating and choosing single stocks and finding the best portfolio under different goals.

There are some defects in this paper, mainly in 3 aspects. Firstly, this paper doesn't take currency inflation into account, thus accuracy of calculation may be affected. Secondly, risk-free rate used in this paper is the mean value of data for 5-year treasury bond's interest rate. However, it still needs to discuss whether 5-year treasury bond is the best resource to obtain information about risk-free rate. Moreover, risk-free rate keeps changing as return rate of stocks changing, using a constant value to be substituted in the formulas may also reduce accuracy of the calculation. Thirdly, in Chinese stock market, there are cases that stock exchange operates special treatment to a specific stock for its recent abnormal performance, whereas this paper doesn't give consideration to this kind of events. Although there are some defects, this paper still makes sense in helping investors making investment decisions.

5. Conclusion

In summary, this paper investigates making investment decision in A-share market, based on minimal variance and maximum Sharpe Ratio. The whole article presents a complete process of making investment decision from choosing stocks to building portfolio, hoping to offer some ideas to Chinese investors who keep being cut leek most of the time. There can be different best choices when the objectives of optimization are different, which has been proved in this paper. In the future, it is anticipated that there can be more articles doing research related to portfolio designing, especially in A-share market. Nevertheless, this paper only takes two objectives of optimization into consideration, future studies may consider the impacts of value-at risk, Sortino Ratio, conditional value-at-risk, etc. In a word, it is usually risky to make investment, especially speculating in stock market. However, with scientific method and correct principle of judge, investors are able to reduce their loss and expand their profit. Overall, these results shed light on portfolio designing for investors.

References

- [1] L. Wo, "2022 Financial Market Expectation (Part 1)," J. Dynamic Analysis of Stock Market, vol. 25, 2022, pp. 12-14.
- [2] H. Song, "Characteristics of Retail Investment Transactions and Their Impact on the Market," J. Shanghai Business, vol. 10, 2021, pp. 180-18.
- [3] S. Li and M. Jiang, and W. Xie, "Analysis of Stock Selection Strategy from the Perspective of Retail Investors," J. China's Collective Economy, vol. 30, 2021, pp. 90-91.
- [4] W. Liu, "A-share Investment Strategy of Small and Medium-sized Retail Investors in the New Era," J. Cooperative Economy and Science and Technology, vol. 22, 2018, pp. 60-61.
- [5] C. Sun and Q. Wang, "Research on Stock Market Investment Strategy from the Perspective of Retail Investors," J. Financial Circles (Academic), vol. 5, 2010, pp. 104-106.
- [6] Z. Zhou, "Research on the Strategy of Individual Investment in A- Shares," Guangxi University, 2013.
- [7] M. Zheng, "Discussion on Stock Selection Strategy of Retail Investors," J. New West (Next Issue), vol. 12, 2018.
- [8] H. Markowitz, "Portfolio Selection," J. The Journal of Finance, vol. 7 (1), 1952, pp. 77-91.
- [9] W. F. Sharpe, "Mutual Fund Performance," J. The Journal of Business, vol. 39(1), 1966, pp. 119-138.
- [10] J. Philippe, "Value at Risk, The new Benchmark for Controlling Derivatives Risk," M. Irwin Professional Pub, 2000.
- [11] R. T. Rockafellar and S. Uryasev, "Optimization of Conditional Value-at-risk," J. Risk, vol.2 (3), 2000, pp. 21-42.
- [12] R. T. Rockafellar and S. Uryasev, "Conditional-value-at-risk for general loss distributions," J. Banking Finance, vol.26 (7), 2002, pp. 1443-1471.
- [13] A. Stephen, W. Randolph, and J. Jeffrey, "Corporate Finance," M. McGraw-Hill Companies, 2002.

- [14] H. Chen and B. Yu, and A. Zhou, "Experimental Design of Modern Portfolio Theory and Application of R Software," J. Journal of Heihe University, vol. 12 (5), 2021, pp. 111-114.
- [15] Würtz, Diethelm, et al., "Portfolio optimization with R/Rmetrics," Rmetrics, Rmetrics Association and Finance Online Publishing, Zurich, 2009.
- [16] W. Diethelm, S. Tobias, C. William, C. Yohan, E. Andrew, "Basic R for Finance", Rmetrics Association & Finance Online Publishing, Zurich, 2010.