Investment Portfolio Optimization

Qing Wei
A53309131
qiwei@ucsd.edu

Sijia Yang
A53314776
siy004@eng.ucsd.edu

Danqi Yuan
A53296724
dayuan@eng.ucsd.edu

Qianyu Zhao
A53309512
qiz019@ucsd.edu

Task Assignment

We studied and worked together. Each of us is involved in each step during the whole work. And while writing this paper, we divided it into four parts and each of us is responsible for almost the same amount of work.

I. Introduction

Our group is interested in maximizing the investment gain by applying convex optimization tools that we learned along the course. The motivation for this project is that nowadays, the majority of people want to make money from investment, but most of them lack financial knowledge or risk management background to help them make the appropriate investment decision. Therefore, our group comes up with the idea of helping those people get the most out of their investment without knowing much about the finance market. The problem that we are trying to solve is to get the maximum investment's expected return within the appropriate exposure of risk. The client determines how much risk he/she wants to bear and our algorithm suggests client the weighted combination of the investment (stock, fund, and etc) portfolio that yields to maximum return of the investment. The investment portfolio optimization problem has a broad applications in financial market, especially in asset management, retire investment field and etc.

Contributions: Based on the Modern Portfolio Theory [1] (MPT), we implement and validate it on a real-word data set and get a reasonable result. Besides, we also use a novel way to select stocks. In our way, according to the history data, We can possibly choose stocks that have lower risks but higher returns.

Report Organization: Our group discusses about related work follows the introduction section, and then our group states the problem in the convex optimization form (primal, dual and KKT conditions). In IV section, we describe the dataset used for our problem and the approaches we experimented for our project. In V section, we report the experiment results, conclusions, and possible future works. Lastly, we include references in VI section.

II. Related Work

Nowadays, AI and ML have been a hot topic, and it turns out that people have already tried classic methods such as unsupervised, supervised machine learning approaches, reinforcement learning agents and some more exotic options for investment portfolio management.

A. Markowitz efficient frontier

Markowitz efficient frontier is proposed by Harry Markowitz with his dissertation on ”Portfolio Selection” [1], a paper which earned him the Nobel Prize in Economics. In the paper, he proposed Modern Portfolio Theory (MPT). He demonstrated that a diversified portfolio is less volatile than the total sum of its individual parts. Prior to MPT, investors mainly focused on investing the individual stocks. Markowitz suggests that the volatility decreases if looking at the entire portfolio instead of just focusing on one investment. In addition, Investors must determine the level of diversification that suits them. The efficient frontier graph could represent all possible combinations of risky securities for an optimal level of return given a particular level of risk.

B. Unsupervised Learning

The idea behind the unsupervised learning method is to group assets into ”cluster” based on their profitability and allocate more funds on the most predictive ones. Several models are proposed for unsupervised learning task such as Eigenportfolios, which used principle component analysis to decompose data into the first principle component representing an approximation of the market, and the second components giving uncorrelated to the market strategies. Another model is autoencoder risk, which is based on the neural networks. The autoencoder allows non-linear dimensionality reduction, which is an alternative for PCA.

C. Supervised Learning

One drawback of the above approaches is that we could not guarantee the future assets will move the same as in the past. Therefore, it would be better if we could predict the future and make it as our allocation weights. In the paper ”Fuzzy time-series based on Fibonacci sequence for stock price forecasting” [4], it proposes a new model, which incorporates the concept of the Fibonacci sequence, and other framework and models, and it turns out that this combine model surpasses conventional models in accuracy.

III. Statement of the Problem

A. Primal formulation

In the Investment Portfolio optimization problem, first assumes that we have n different assets, each of which has some amount of money and will has its own investment. Suppose that $w_i, i = 0, 1, 2, ...n$, is the fraction of money that we choose the investment for each assets. And thus we can know that $\mathbf{1}^T * w = 1$. Typically,$w \geq 0$. When $w_i \leq 0$, it means a short position in asset i or means that we borrow money for now and have to return it later.

Our model is based on a period of time, so suppose that the price at the start of the period is p_i, and the price at the end of the period is p_i'. Then the return rate $r_i = \frac{p_i' - p_i}{p_i}$. Therefore, the portfolio return rate $\mathbf{R} = r^T w$.

In the model, r is supposed to be a random variable whose mean value $\mu = \mathbf{E}R$ and covariance matrix $\Sigma = \mathbf{E}(r-\mu)(r-\mu)^T$. Also, we can have that $\mathbf{E}R = \mu^t w$ and $var(R) = w^T \Sigma w$. $\mathbf{E}R$ represents the mean return of the portfolio and the $var(R)$ represents the risk of the portfolio.

To optimization the portfolio, we need to maximize the mean return value and minimize the risk at the same time.

As a result, we can state the problem like below:

$$\begin{array}{c} \text{maximize } \mu^T w - \gamma w^T \Sigma w \\ \text{subject to } \mathbf{1}^T w = 1, w \in W, \end{array}$$

where $w \in \mathbf{R}^n$ is the optimization variable and W is a set of allowed portfolios, and γ is parameter that describe how one can bare the risk and is determined by the client.

B. Dual formulation

Its dual problem is:

$$\begin{array}{c} \text{maximize } \frac{1}{4\lambda}(\mu^T + 1^T\lambda)^T \Sigma^(-1)(\mu^T + 1^T\lambda) - \lambda \\ \text{subject to } \lambda \geq 0 \end{array}$$

C. KKT conditions

The KKT conditions are:

$$\frac{\partial f}{\partial w} + \lambda \frac{\partial h}{\partial w} = 0$$

$$\text{-}2\gamma\Sigma w + \mu^T + 1^T\lambda = 0$$

$$\text{w} = \frac{1}{2\gamma}\Sigma^{-1}(\mu^T + 1^T\gamma)$$

$$\text{h(w) = 0, where h(w) = } 1^T w - 1$$

IV. Approach

A. Dataset

We selected a dataset from Kaggle, a dataset publication platform to train and evaluate our model. The dataset, New York Stock Exchange [2], contains S&P 500 companies' historical prices with fundamental data. The NYSE dataset spans from 2010 to the end of 2016, and the prices were fetched from Yahoo Finance, NASDAQ Financials, extended by some fields from EDGAR SEC databases.

In addition to the ticker symbol, open and close prices on each date of those stocks, the GICS sector information of each stock provided in the NYSE database is also especially useful for us.

In this project, we used the data in 2015 for model training, and the data in 2016 for model validation.

B. Data Processing

1) Sharpe Ratio: The Sharpe ratio, developed by William F. Sharpe, is a common metric for measure the risk-adjusted return [3]. The ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk.

$$\begin{aligned} \text{Sharpe Ratio} &= \frac{R_p - R_f}{\sigma_p} \\ \textbf{where}: R_p &= \text{return of portfolio} \\ R_f &= \text{risk-free rate} \\ \sigma_p &= \text{standard deviation of return} \end{aligned}$$

In the following sections, we will use the Sharpe ratio to select a limited number of stocks from the 500 companies in our database and to evaluate our portfolio to get the best risk aversion parameter γ in our model.

2) GICS Sectors: The Global Industry Classification Standard (GICS), developed by MSCI and S&P is a widely-used standard for categorize companies. We analyzed the stock prices for the companies in each of the ten sectors: Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Telecommunications Services, Utilities.

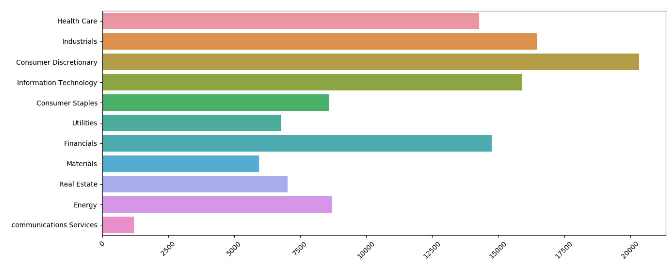

Fig. 1. Number of Companies in Each GICS Sector in S&P 500 Companies

Fig. 2. Covariance Matrix of Each Sector Based on Average Investment Return

We found that most of the companies fall into the Consumer Discretionary category, and the Communication Service category has fewest companies (see Fig. 1). In addition, when we calculated the covariance matrix of each sector based on the average return of the companies in it, some sectors have shown high correlations, such as Industries and Materials (see Fig. 2). Higher correlations often means higher probability that two stocks are moving in the same direction. This information may used in different investing strategies. When building a diversified portfolio, investors seek negatively correlated stocks so that each when one of the stock decreases in price, the negatively-correlated stock is likely to rise, thus reducing the risk of catastrophic losses in the portfolio. On the contrary, risk takers would love to seek for positively correlated stocks for higher expected return but with higher risks.

3) Stock Selection: To select a limited number of stock instead of evaluating all companies, we calculated the Sharpe ratio of every company using their historical data, and we decided to select 2 stocks with the highest risk-adjusted return from each sector in order to diversify our portfolio and reduce the risk.

C. Model Implementation

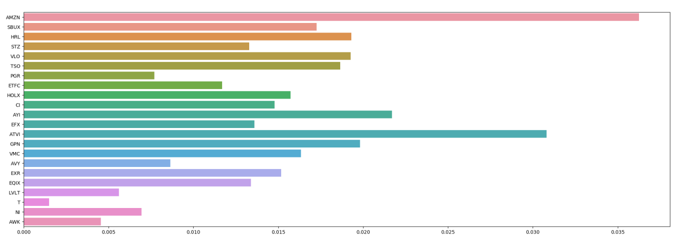

Fig. 3. Expected Return μ of the Selected Stocks

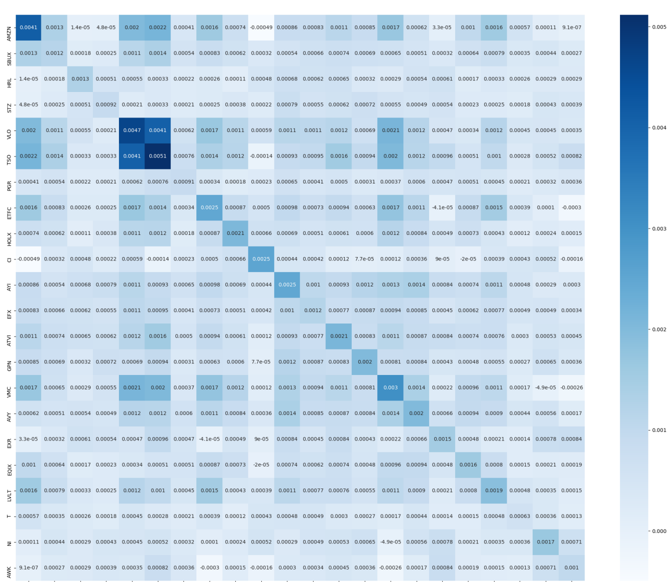

Fig. 4. Covariance Matrix Σ of the Selected Stocks

To solve the convex problem defined in Section III, we need to calculate the expected return $\mu = \mathbf{E}r$ (see Fig. 3) and the covariance matrix $\Sigma = \mathbf{E}(r-\mu)(r-\mu)^T$ (see Fig. 4) of the selected stocks. It is clearly illustrated in the figures that AMZN and ATVI have the highest expected return in the historical dataset, whereas PGR, STZ and T have the lowest risk.

For the risk aversion parameter γ, we consider it as a parameter spanned between 0.01 and 100. By solving the problem with different γ in this range, we can calculate the Sharpe ratio of each primal return and risk, and obtain a portfolio with the best return-risk trade-off.

The convex problem solver is implemented with CVXPY, which is a Python-embedded modeling language for convex optimization problems. [5]. The solver function takes μ, Σ and γ as the input, and returns the optimal portfolio weight, and the corresponding return and risk data on the trainset. The code for the solver is as follows.

```
import numpy as np
import cvxpy as cp

def solve(mu, Sigma, gamma):
    # mu: nx1 numpy array
    # Sigma: nxn numpy array
    # gamma>0: risk aversion parameter
    n=np.shape(mu)[0]
    w=cp.Variable(n)
    ret=mu.T * w
    risk=cp.quad_form(w, Sigma)
    prob=cp.Problem(cp.Maximize(ret-gamma*risk),
                    [cp.sum(w)==1, w>=0])
    prob.solve()
    return_data=ret.value
    risk_data=cp.sqrt(risk).value
    weight=w.value
    return return_data, risk_data, weight
```

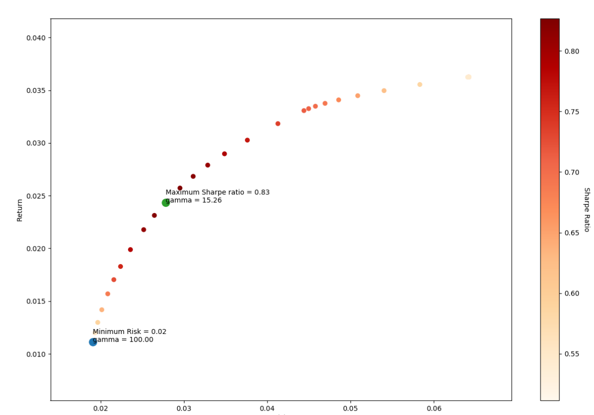

Fig. 5. Return-Risk value on different γs

The above figure Fig. 5 demonstrated the average return and risk on different choices of γ. It is obvious that the larger γ is, the lower the average return as well as the standard variance. To reach a trade-off between return and risk, we can choose the point with the highest Sharpe ratio, where $\gamma = 15.26$.

D. Experiments and Tests

We use the data after 2016 to do the test and validation.

First, we calculate each stock's return mean. The mean represents how much return this stock can provide for the investors. We figure out this data by getting each stock's return in each period first. And then for each stock, we calculate its return mean using these returns from each period.

Second, we use the σ and μ calculated from the 'C.Model Implementation' part and the same stock selection method mentioned in the previous part to calculate the weights. We spanned the risk aversion parameter γ from 0.01 to 20. For each γ, we can get a relevant weight composition.

Third, We combines the return mean of each stock, which we already had in the first step, and the weight of each stock, which we get in the second step, to calculate the final return the investors can have under each circumstance.

As a result, we can see the relationship between the final return and the gamma, as shown in Fig. 6. According to the graph, we can see that as the γ increases, which suggests the investor hope lower risks, the return also goes down. It's consistent with our common sense —— high return always comes with high risk, and low risk may lead to low return. Also, from the result, we can see that our model can really give investors different investment suggestions according to their risk taking abilities.

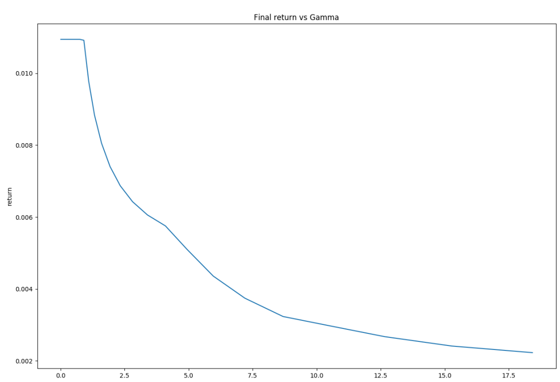

Fig. 6. The Rlationship of Final Return and Gamma

V. Conclusion

Investment, as a common approach to help us stay ahead of inflation and increase the value of our money, is studied and researched by many scholars. In our project, we designed and implemented a portfolio builder inspired by the nobel-winning approach Modern Portfolio Theory [1] (MPT). Based on MPT, we additionally devised a novel way of stock selection. By choosing companies from different GICS sectors, we diversified the portfolio and reduce the risk of total losses in the portfolio. To achieve better performance, from each sector, we only choose stocks with the most outstanding return.

We used real-world data to train and validate our model, and the experimental results on the test set proved our hypothesis that the larger γ will result in less risk but also lower return. Return-risk trade-off is evaluated and we calculated the Sharpe ratio and searched for the appropriate γ for the overall best investment portfolio.

VI. Possible Future Work

Currently, we searched the γ parameter to get in highest Sharpe ratio. However, for different investors, the investing strategy and risk taking ability are quite distinct, too. One solution is impossible to apply for everyone. Our possible future work will be designing a questionnaire and identifying the risk aversion parameter γ for each user based on their replies on the survey.

References

[1] H. Markowitz, "Portfolio Selection." The Journal of Finance, Vol. 7, No. 1, pp. 77-91. 1952

[2] D. Gawlik. "New York Stock Exchange." Retrieved from https://www.kaggle.com/dgawlik/nyse/metadata. March 2017.

[3] W. Sharpe, "The sharpe ratio." Journal of portfolio management 21, no. 1, pp. 49-58. 1994

[4] C.Tai-Liang, C.Ching-Hsue, and T.HiaJong,"Fuzzy time-series based on Fibonacci sequence for stock price forecasting,"Physica A: Statistical Mechanics and its Applications, vol. 380, pp. 377-390, July 2007.

[5] S. Diamond, S. Boyd, "CVXPY: A Python-Embedded Modeling Language for Convex Optimization." Journal of Machine Learning Research, Vol. 17, No. 83, pp. 1-5. 2016.

[6] S. Boyd, E. Busseti, et al. "Multi-Period Trading via Convex Optimization," Foundations and Trends in Optimization, Vol. 3, No. 1, pp. 1-76. August 2017.

[7] Tabak, Daniel; Kuo, Benjamin C. (1971). Optimal Control by Mathematical Programming. Englewood Cliffs, NJ: Prentice-Hall. pp. 19–20. ISBN 0-13-638106-5.

[8] Kuhn, H. W.; Tucker, A. W. (1951). "Nonlinear Programming". Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. pp. 481–492. MR 0047303.