MGT 6203 GROUP PROJECT FINAL REPORT

Optimal Portfolio Using S&P500 Stocks

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-80

Team 80:

Anoop Kumar
Daniel Webster
Harsh Ambasta
Hieu Nguyen
Sarah Boyette

Table of Contents

Background:

As of 2023, 61% of Americans invest in the stock market in the United States, and globally, approximately 20%-25% of people invest in the market. With nearly a quarter of the world's population investing, the use of data analytics for optimizing and mitigating risk to the owner's equity has increased substantially. Although nobody can predict the future of any stock in the stock market, leveraging big data and modeling the future performance of any given stock can help investors lower their risk and increase their returns.

The Dow Jones Industrial Average, S&P 500, and the Nasdaq Composite Index are the three most popular stock indexes. Investors and data analysts can use these market indexes to gauge market movements and use the information to make investment decisions regarding their portfolios. For the continuation of this project, our team will be utilizing the S&P 500 Index.

Share of adults investing money in the stock market in the United States from 1999 to 2023

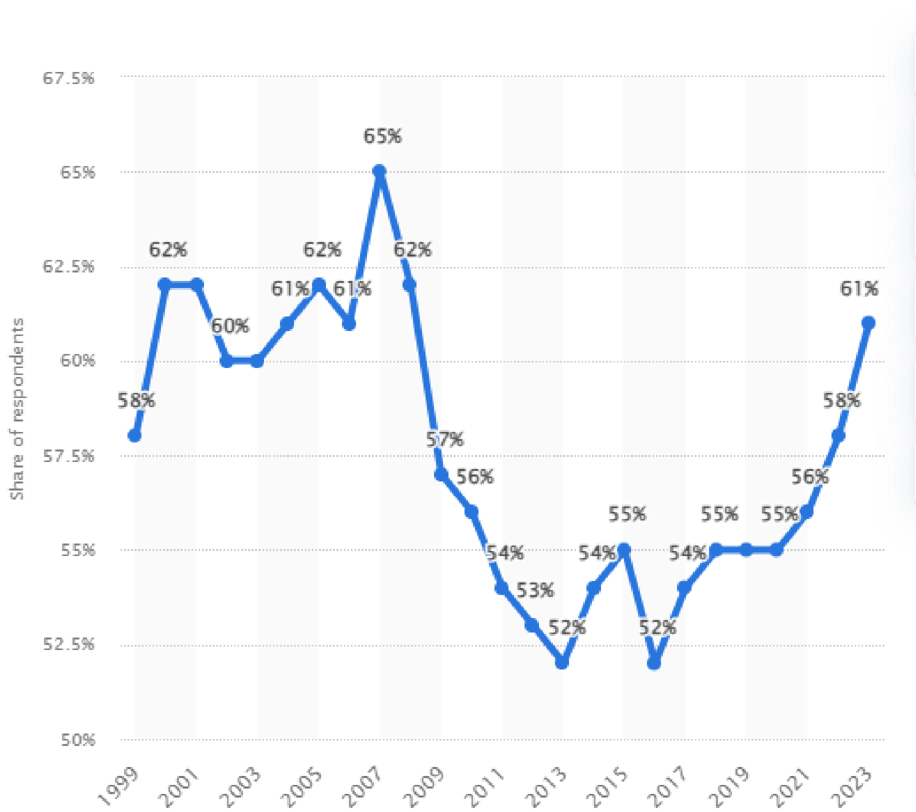


FIGURE 1: PERCENTAGE OF ADULTS WHO INVEST IN THE STOCK MARKET IN THE UNITED STATES

Objective/Approach:

Given the possible risks that investors face when investing their money in the stock market, our objective is to build the optimal portfolio that maximizes return on investment while minimizing risk. First, we will take the mean and variance of the individual stocks. Next, we will make the optimal portfolio using optimization programming to identify weights that will produce an optimal Sharpe ratio. Our data cleaning and portfolio construction will be done in Python using various libraries such as Pandas and Scipy. Once we have found the optimal weights for the best Sharpe Ratio, we will create non-optimal portfolios to compare to the optimal portfolio and plot the portfolios on an Efficient Frontier scatterplot.

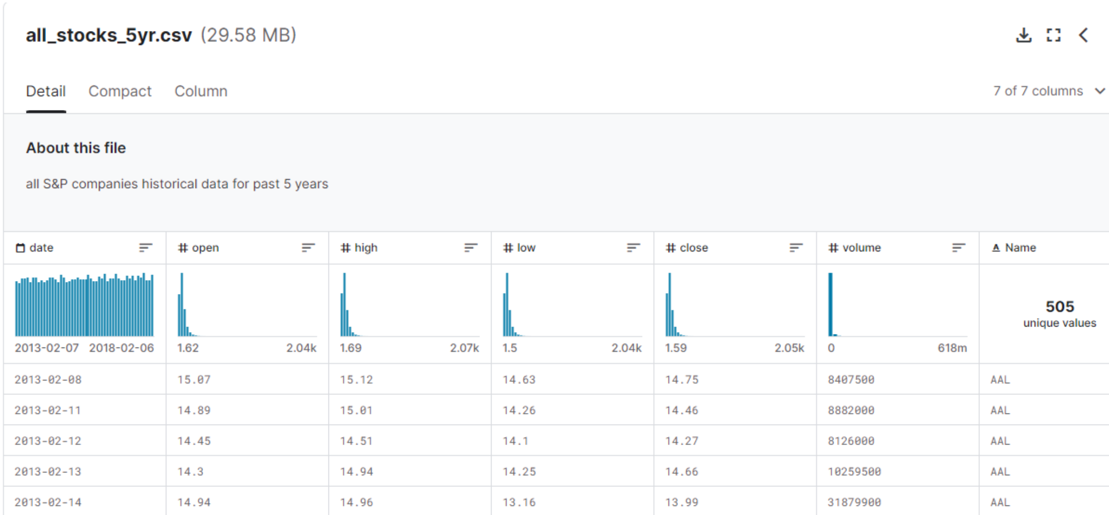
Business Justification:

Investors are constantly seeking to gain the best returns possible on their equity. Some methods are active (time-consuming) while others are passive (not time-consuming). By building an optimal portfolio, investors can passively gain returns on investments with the maximum possible return for risk. Building a passive, indexed portfolio is perfect for individuals who aren't interested in the stock market but want to build wealth for the future such as children's education or retirement.

With the SP500 constituting almost 80% of the total market capital at $37T, it has given a CAGR of ~12% since its inception. Its strong growth history makes it one of the most open platforms to grow your wealth. With our project, we are trying to minimize the losses incurred while doing so based on data-driven decisions.

Data Overview:

Historical S&P500 data was taken from the following Kaggle database: https://shorturl.at/vDGIK; also crucial for our analysis is broad S&P500 data taken from https://shorturl.at/htAKY and the daily treasury par yield curve rates from https://shorturl.at/ptAT9 as a source for a risk-free rate. The data was then cleaned and explored to ensure no anomalies or outliers were present. Once the data was reviewed and cleaned, each ticker's average return, compounded return, and standard deviation were calculated using Python. Moving forward, we will use these calculated returns as the input for an optimization algorithm to determine the weight of each stock and the number of stocks that minimize risk while maximizing the portfolio's return. As a comparison, we will also create many randomly weighted portfolios to confirm that the model's output is optimal. Finally, we will plot the expected return and expected standard deviation to visualize the performance of the optimal portfolio relative to the random portfolios.



all_stocks_5yr.csv (29.58 MB)

Detail Compact Column 7 of 7 columns

About this file

all S&P companies historical data for past 5 years

date	open	high	low	close	volume	Name
2013-02-07 – 2018-02-06	1.62 – 2.04k	1.69 – 2.07k	1.5 – 2.04k	1.59 – 2.05k	0 – 618m	505 unique values
2013-02-08	15.07	15.12	14.63	14.75	8407500	AAL
2013-02-11	14.89	15.01	14.26	14.46	8882000	AAL
2013-02-12	14.45	14.51	14.1	14.27	8126000	AAL
2013-02-13	14.3	14.94	14.25	14.66	10259500	AAL
2013-02-14	14.94	14.96	13.16	13.99	31879900	AAL

FIGURE 2: HISTORICAL S&P500 DATA USED FOR OPTIMAL PORTFOLIO CALCULATION.

Data Cleaning:

To pull all the necessary data together for computing the various returns, three data frames, “stocks,” “offer,” and “sp500,” are loaded from downloaded CSV files. The first step in the data cleaning is to match the dates of all the data frames and convert them to the same data type(numeric). The date column is then refactored to be the datetime index, daily returns are calculated, and new monthly returns for each data frame are calculated. On inspection, it is found that some months do not have returns for the whole month, so February 2013 and February 2018 are dropped from the dataset. Similarly, 29 of the 505 tickers in “stocks” don’t span the full date ranges needed, so they are also dropped from the dataset. Indexes were added to the new monthly returns data frames for ease of data manipulation. Once the data was cleaned and all missing and outlying data was removed, we proceeded with the return calculations (monthly compounded and average).

1	Ticker	Average Return	Compounded Return	Standard Deviation
2	AAL	0.026134127365342368	2.484745762711868	0.10813346939027328
3	AAPL	0.016264787299868466	1.3512177580754003	0.0661445245154208
4	AAP	0.008907954370605743	0.393282636248417	0.0848273525628736
5	ABBV	0.02104786323456991	2.1343448275862067	0.0664417315743053
6	ABC	0.013641682065382311	1.0093836638942197	0.0656323560462702
7	ABT	0.01033418552012215	0.7050276082534137	0.05579360804716142
8	ACN	0.013366034466378459	1.1163552039285212	0.04522134794467881
9	ADBE	0.028067467951814244	3.916666666666666	0.05851127735674123
10	ADI	0.012141353099082649	0.8676148796498893	0.06187216597293315

FIGURE 3: AVERAGE AND COMPOUNDED RETURNS OF S&P500 STOCKS

Methodology:

After the data had been cleaned and all arbitrary values were removed, the team began calculating the optimal portfolio. As stated above, the individual stocks' mean and variances were calculated first. The variance matrix is used to compute the portfolio risk, otherwise known as the portfolio volatility.

	AAL	AAPL	AAP	ABBV	ABC	ABT
1	0.011692847	0.000980127	0.003780113	0.001813175	0.002768457	0.002224327
2	0.000980127	0.004375098	8.41836E-05	0.000811234	0.000540346	0.001957769
3	0.003780113	8.41836E-05	0.00719568	0.001808679	0.001825944	0.001013047
4	0.001813175	0.000811234	0.001808679	0.004414504	0.001447492	0.002180681
5	0.002768457	0.000540346	0.001825944	0.001447492	0.004307606	0.001800607
6	0.002224327	0.001957769	0.001013047	0.002180681	0.001800607	0.003112927
7	0.00106311	0.001326188	0.000811599	0.001104141	0.000344878	0.001113863
8	0.001124919	0.001412462	0.00055943	0.00125393	0.000431248	0.001466191

Figure 4: Mean and Covariance of Individual Stocks

The next step was to determine the weights of each stock in the portfolio to maximize the return and minimize the risk to the investor. We chose not to include shorting in our portfolio, so all weights were constrained to zero and one. Had we included shorting, the constraints would have allowed for negative weights in the optimization. As this is a convex nonlinear problem, we used Scipy's minimize solver in order to find the smallest negative Sharpe ratio (in accordance with the principle of duality, this would give us the weights of the largest Sharpe ratio).

$$-SR = -\frac{W_X^T R_X - R_f}{\sqrt{W_X^T (W_X K_{XX})}}$$

Figure 5: Optimization Function; where W_X is the vector of optimal weights, R_X is the vector of mean returns, R_F is the risk free rate 0.0252, and K_{XX} is the covariance matrix of mean returns.

Out[4]:

Ticker	Stock Weight %
NVDA	35.773129
STZ	30.228812
FB	13.811315
CNC	9.176033
NFLX	6.310404
LUV	3.226672
AVGO	1.473636
NUE	0.000000
NTRS	0.000000
NTAP	0.000000

Figure 6: Optimal Stock Weights

Here are the first ten stock weights in our portfolio; however, moving forward, we only used stocks that had a nonzero optimal weight for constructing random portfolios. When attempting to construct random portfolios using the entire set of tickers, or even just a subset of 7-10 random portfolios, almost all of them would fall considerably short of the optimal portfolio. Creating enough portfolios to visualize the efficient frontier would be computationally infeasible with the number of tickers in our dataset, hence why an optimization approach is preferable for our purposes.

Results

The weight optimization program aims to produce a Sharpe ratio of at least 1 or greater. Using a for loop, we could iterate through many different portfolios to find the one with the greatest Sharpe ratio. We then plotted the Efficient Frontier to visualize all the portfolios that maximize the expected return at each given risk level. The X on the Efficient Frontier shows the optimal portfolio to maximize the investor's return and minimize their risk.

```
Largest Random Sharpe Ratio: 1.2446
Optimal Portfolio Sharpe Ratio: 1.2581
```

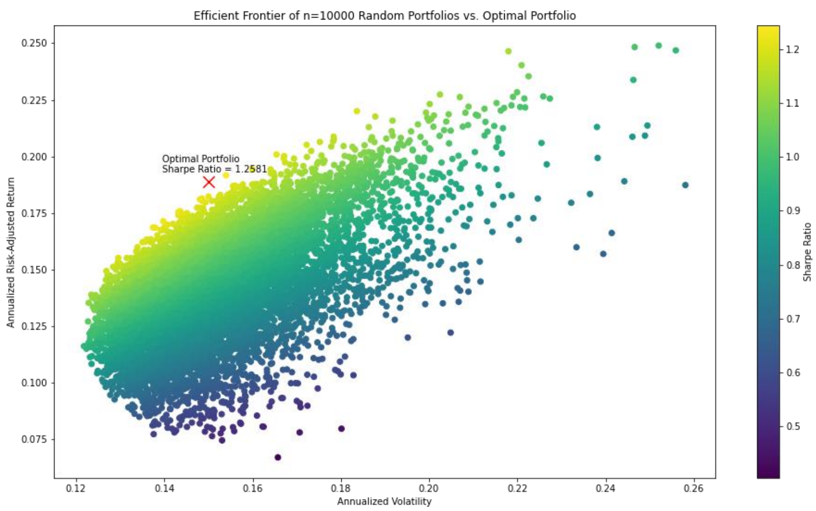


FIGURE 7: EFFICIENT FRONTIER AND CODE OUTPUT OF OPTIMAL AND BEST RANDOM PORTFOLIO SHARPE RATIOS

Conclusion

Using optimization programming, we were able to find the most optimal portfolio by maximizing the expected return while minimizing the risk. The optimal portfolio had 7 different stocks out of the 500 in the SP500 with an expected return of 4.09% and volatility of 4.33%. While the SP500 usually has an expected return of around 7-8%, it has a much higher volatility, which may not be ideal for some investors. Although we were able to find the portfolio with the optimal Sharpe Ratio, this portfolio may not be the optimal portfolio for all investors. People and organizations have different objectives and risk tolerance and thus their ideal portfolio will be different from each other. Many other better structured portfolios can be found near the top/left of the Efficient Frontier, which may better suit investors better than the portfolio with the best Sharpe Ratio.

The data used to calculate individual stock returns and volatilities was only between 2013-2018, which is a span of 5 years. Depending on the period and the length of the returns, the stocks in the portfolio would be different. While it is possible to obtain individual SP500 data for a longer period of time, this would require more time and effort to collect and clean the data rather than implementing the methods for portfolio construction. Regardless of the data, the methodology of the portfolio construction is applicable across periods and security datasets.

Overall, the portfolio with the optimal Sharpe Ratio is structured for individual investors who are looking for the best return to risk for their capital without the need of actively using the capital for returns. This portfolio has a better return and risk than most other random portfolios and individual stocks, and it is not complex to construct.

Lessons Learned

Throughout this term long project, we learned the below lessons:

- This project helped us to explore business avenues of stock investment and portfolio construction along with analytical skills to optimize them for best results.
- We learnt about the implementation of Sharpe's ratio and its interaction with volatility and returns.
- The initial dataset required us to perform data due diligence and cleaning to ensure that the model and analysis don't run into anomalies.
- This project challenged us to use Python's Scipy library for solving the Sharpe ratio optimization equation.
- We also learn to use comparison analysis by comparing our portfolio with SP500 to evaluate and comment on the model's effectiveness.
- Finally, all of this was able to come to fruition as we performed effective team communication to ensure that the project's requirements and progress are met within the timeline.

Works Cited

Daily Treasury Par Yield Curve Rates. U.S. Department of the Treasury. (2023, November 29). https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value=2018

Nugent, C. (2018, February 10). *S&P 500 stock data*. Kaggle. https://www.kaggle.com/datasets/camnugent/sandp500

S&P 500 historical rates (SPX). Investing.com. (n.d.). https://www.investing.com/indices/us-spx-500-historical-data

Statista Research Department. (2023, September 18). Share of Americans investing in stocks 2023. Statista. https://www.statista.com/statistics/270034/percentage-of-us-adults-to-have-money-invested-in-the-stock-market/

Yeh, IC., Liu, YC. Discovering optimal weights in weighted-scoring stock-picking models: a mixture design approach. *Financ Innov* **6**, 41 (2020). https://doi.org/10.1186/s40854-020-00209-x