Capstone Project 1
Portfolio Optimization

Introduction and Objective

1. Get the portfolio optimization
Assuming I would like to invest in technology stocks such as Apple, Cisco, IBM, Amazon, Microsoft, Visa, how can I maximize my return with specific risk? In this project, I will use Monte Carlo Simulation and Efficient Frontier to get the portfolio optimization.

2. Apply statistical reference to get Alpha (α), Beta (β), Value at Risk (VaR) for the optimal portfolio
After optimizing my portfolio, I will calculate Alpha (α), Beta (β), Value at Risk (VaR) for the portfolio.

Target Audience and Why

My target audience will be investors who are interested in trading stocks. With the amount of funds investors want to invest, they can understand how much money to put in each stock to maximize their return. In addition, they can have an overview for the return and Value at Risk of their portfolios.

Dataset Acquisition

Data was scraped from Yahoo Finance for the stocks from 1/1/1998 to 6/30/2016. The dataset comprises “closing prices”, “High”, “Low”, and “Volume” of all stocks.

Data Wrangling

Steps required to clean and modify the data into a necessary format for analysis.

1. Clean Data: Include only “Adj Close” for each stock
2. Make sure the index is datetime.

Final Dataset:

All stocks have the same time frame and dataset. Since Visa(V) was listed starting from 2008, the data is less than other stocks.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 3523 entries, 2006-01-03 to 2019-12-31
Data columns (total 6 columns):
AAPL    3523 non-null float64
AMZN    3523 non-null float64
CSCO    3523 non-null float64
IBM     3523 non-null float64
MSFT    3523 non-null float64
V       2968 non-null float64
dtypes: float64(6)
memory usage: 192.7 KB
```

Exploratory Data Analysis

From the six technology stocks, we can see that Amazon stock prices increased a lot in the last 15 years. In order to look deeper into the stocks relationship, I have the mean daily return and correlation for each stock.

Stock Price History

Stock Price Description

Symbols	AAPL	AMZN	CSCO	IBM	MSFT	V
count	3523.000000	3523.000000	3523.000000	3523.000000	3523.000000	2968.000000
mean	78.254635	501.532649	23.279711	113.639400	42.995563	61.670939
std	62.867671	563.043778	10.427005	31.288053	33.007738	47.451121
min	6.266412	26.070000	10.444414	49.635036	11.666746	9.748704
25%	22.592848	89.005001	16.264757	85.885769	21.216453	19.593692
50%	65.542587	257.750000	19.925550	126.290695	25.902451	50.029625
75%	111.237354	727.190002	25.660475	138.109840	51.197996	81.144733
max	292.163818	2039.510010	56.656013	163.324295	158.086731	188.799973

Mean Daily Return

```
Symbols
AAPL    0.001181
AMZN    0.001333
CSCO    0.000521
IBM     0.000339
MSFT    0.000730
V       0.001070
dtype: float64
```

Correlation of Daily Return

Symbols	AAPL	AMZN	CSCO	IBM	MSFT	V
Symbols						
AAPL	1.000000	0.434518	0.471804	0.446622	0.463115	0.459190
AMZN	0.434518	1.000000	0.429882	0.405514	0.485681	0.463976
CSCO	0.471804	0.429882	1.000000	0.548351	0.575420	0.478322
IBM	0.446622	0.405514	0.548351	1.000000	0.527498	0.452460
MSFT	0.463115	0.485681	0.575420	0.527498	1.000000	0.470037
V	0.459190	0.463976	0.478322	0.452460	0.470037	1.000000

Log Returns vs Arithmetic Returns

Here comes the problem. Should I use arithmetic returns or log returns to build up my model?Since Most technical analysis techniques require detrending or normalizing the time series, using log returns is a nice way to do that.

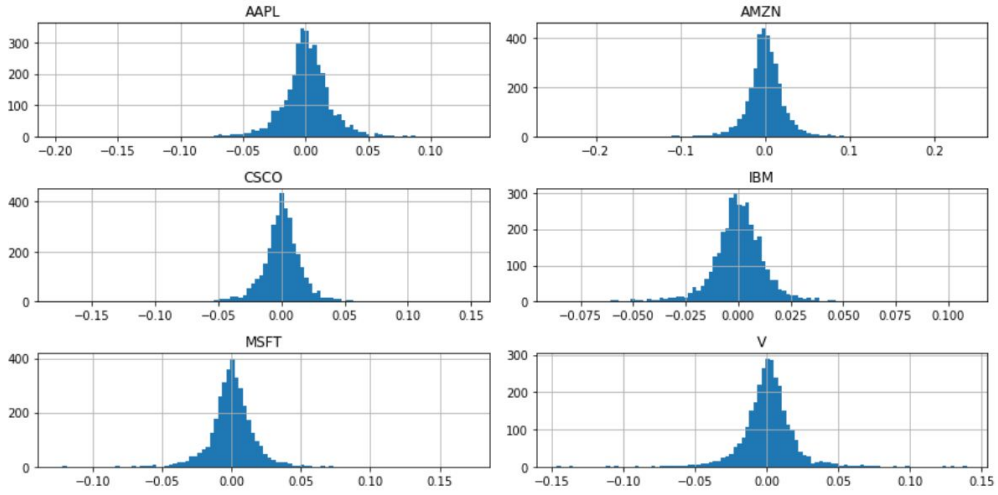
For this project, these two returns won't make a big difference. We will just use log return for the project.

From below log return histograms, we can see that all the stock log returns follow normal distribution. The log mean return is about 0 for all stocks.

Log Mean Return

```
Symbols
AAPL    0.000980
AMZN    0.001039
CSCO    0.000358
IBM     0.000246
MSFT    0.000590
V       0.000951
dtype: float64
```

Log Mean Return Histograms



Daily Covariance of Log Mean Return

Symbols	AAPL	AMZN	CSCO	IBM	MSFT	V
Symbols						
AAPL	0.000401	0.000212	0.000171	0.000122	0.000155	0.000159
AMZN	0.000212	0.000583	0.000189	0.000134	0.000196	0.000199
CSCO	0.000171	0.000189	0.000326	0.000135	0.000173	0.000159
IBM	0.000122	0.000134	0.000135	0.000186	0.000120	0.000116
MSFT	0.000155	0.000196	0.000173	0.000120	0.000278	0.000147
V	0.000159	0.000199	0.000159	0.000116	0.000147	0.000338

Daily Covariance of Log Mean Return

Symbols	AAPL	AMZN	CSCO	IBM	MSFT	V
Symbols						
AAPL	0.101088	0.053493	0.043013	0.030687	0.039030	0.040179
AMZN	0.053493	0.146914	0.047719	0.033891	0.049500	0.050121
CSCO	0.043013	0.047719	0.082263	0.033937	0.043576	0.039944
IBM	0.030687	0.033891	0.033937	0.046755	0.030219	0.029110
MSFT	0.039030	0.049500	0.043576	0.030219	0.070166	0.037139
V	0.040179	0.050121	0.039944	0.029110	0.037139	0.085071

Monte Carlo Simulation

What is Monte Carlo Simulation?

Monte Carlo simulation performs risk analysis by building models of possible results by substituting a range of values—a probability distribution—for any factor that has inherent uncertainty. It then calculates results over and over, each time using a different set of random values from the probability functions. Depending upon the number of uncertainties and the ranges specified for them, a Monte Carlo simulation could involve thousands or tens of thousands of recalculations before it is complete. Monte Carlo simulation produces distributions of possible outcome values.

How can we use Monte Carlo to find the optimal portfolio?

In order to start with Monte Carlo Simulation, I will first do the single run for some random allocation, which randomly assigns a weight to each security in our portfolio, then calculate its mean daily return and standard deviation of daily return. Below are the results for single run allocation.

Creating Random Weights
[0.51639863, 0.57066759, 0.02847423, 0.17152166, 0.68527698, 0.83389686]

Rebalance Random Weights to sum to 1.0
[0.18401825, 0.20335695, 0.01014677, 0.06112161, 0.24419792, 0.2971585]

Expected Portfolio Return
0.21096813950216284

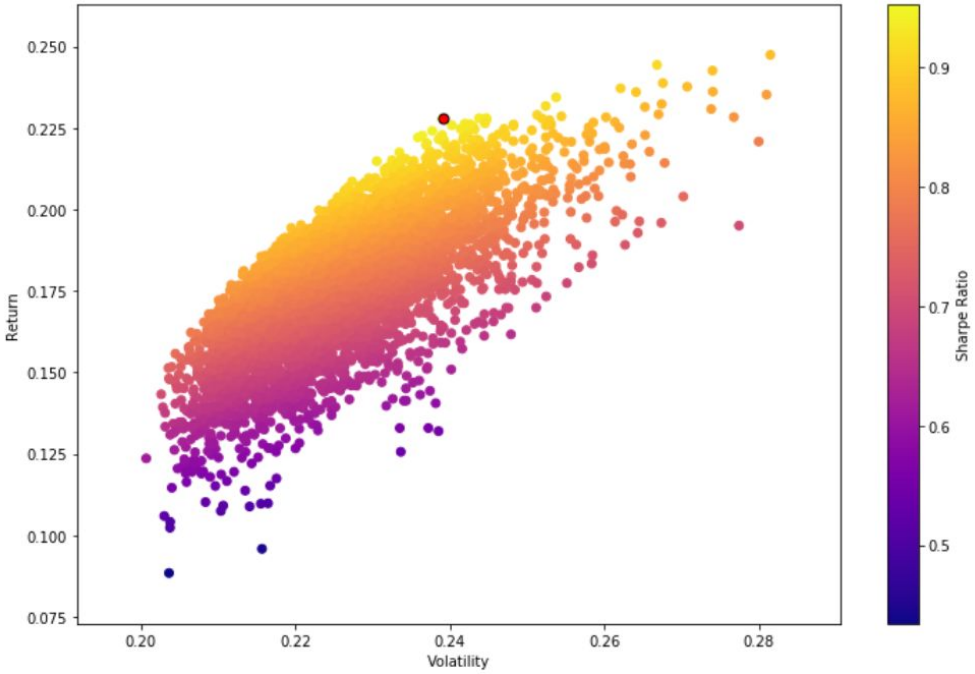
Expected Volatility
0.23237084573135564

Sharpe Ratio
0.9078941845659222

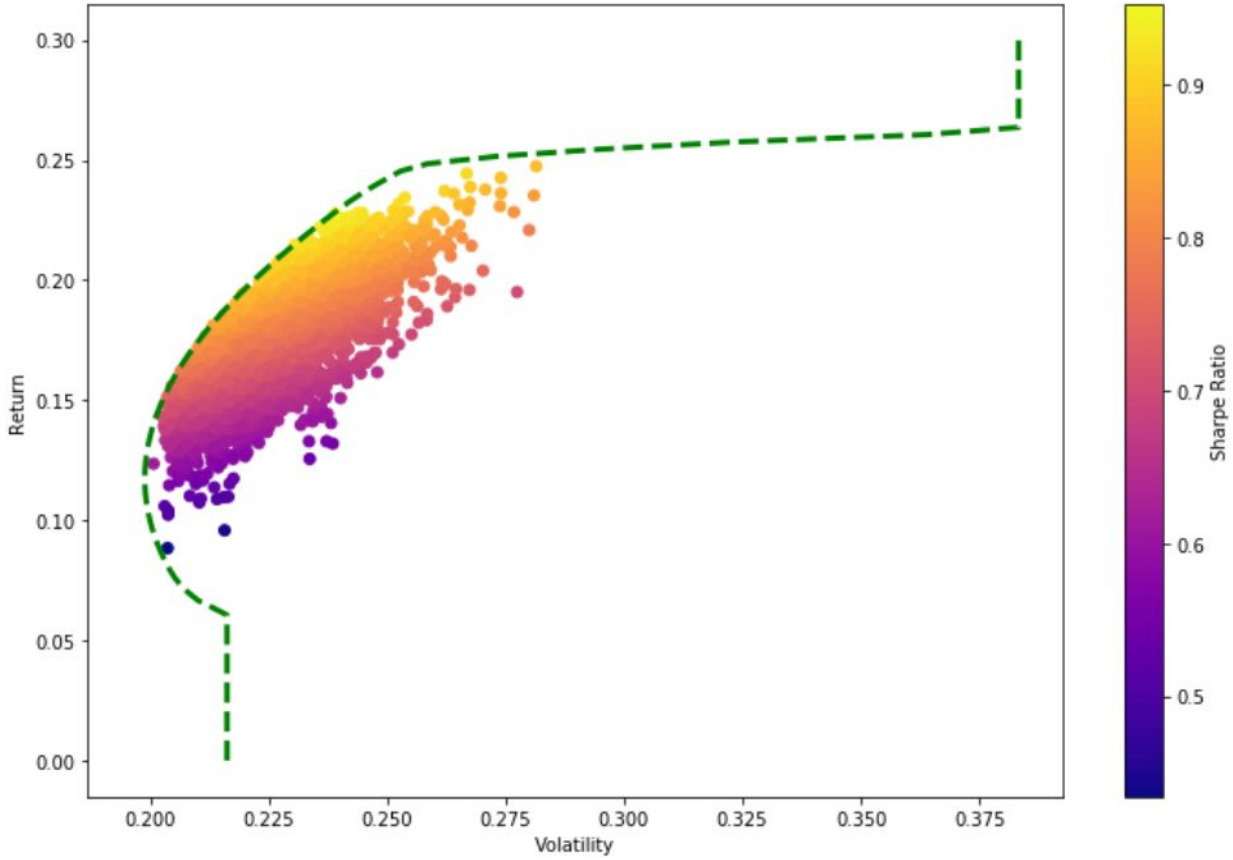
Multiple Run - Repeat single run for thousands of times

Next, I will repeat the single run for thousands of times and then plot all the dots with the specific volatility and return. Here are the steps of how I run multiple times.

Let's say the number of possible positions (or number of actual portfolios) is 5000.
Step 1: Get the maximum Sharpe Ratio and it's index location from the 5000 portfolios
Step 2: Pass the index location to the model
Step 3: Put a red dot for the max Sharpe ratio (That is, find the max Sharpe return and max Sharpe Volatility.)



Efficient Frontier:



Statistical Reference

From the above, we learn how to get the efficient frontier for our portfolios. Next, I will calculate alpha (α), beta (β), and Value at Risk (VR⊓) for the portfolios by using statistical methods.

Beta & Capital Asset Pricing Model (CAPM)

First, I will discuss alpha and beta for the Capital Asset Pricing Model (CAPM) Model. The goal of the CAPM formula is to evaluate whether a stock is fairly valued when its risk and the time value of money are compared to its expected return. The formula for calculating the expected return of an asset given its risk is as follows:

$$ER_i = R_f + \beta_i(ER_m - R_f)$$

Where:

$ER_i = expected\ retun\ of\ investment$

$R_f = risk - free\ rate$

$\beta_i = beta\ of\ the\ investment$

$(ER_m - R_f) = market\ risk\ premium$

The beta of a potential investment is a measure of how much risk the investment will add to a portfolio that looks like the market. If a stock is riskier than the market, it will have a beta greater than one. If a stock has a beta of less than one, the formula assumes it will reduce the risk of a portfolio.

What is Alpha?

Alpha is a measure of the performance of an investment as compared to a suitable market index, such as the S&P 500. An alpha of one (the baseline value is zero) shows that the return on the investment during a specified time frame outperformed the overall market average by 1%. A negative alpha number reflects an investment that is underperforming as compared to the market average.

Alpha is one of five standard performance ratios that are commonly used to evaluate individual stocks or an investment portfolio, with the other four being beta, standard deviation, R-squared, and the Sharpe ratio. Alpha is usually a single number (e.g., 1 or 4), and is expressed as a percentage that reflects how an investment performed relative to a benchmark index.

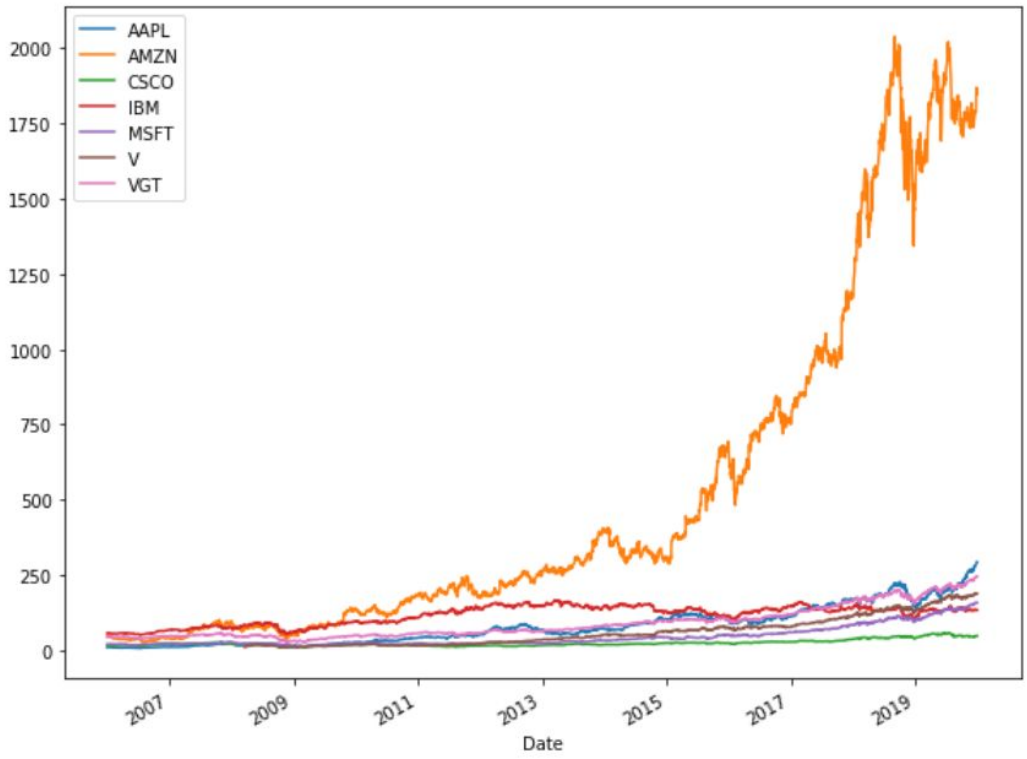
A positive alpha of 5 (+5) means that the portfolio's return exceeded the benchmark index's performance by 5%. An alpha of negative 5 (-5) indicates that the portfolio underperformed the benchmark index by 5%. An alpha of zero means that the investment earned a return that matched the overall market return as reflected by the selected benchmark index.

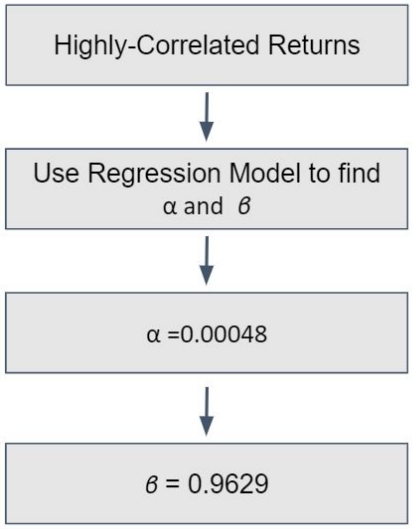
The alpha of a portfolio is the excess return it produces compared to the index. Investors in mutual funds or ETFs often look for a fund with a high alpha in hopes of getting a superior return on investment (ROI).

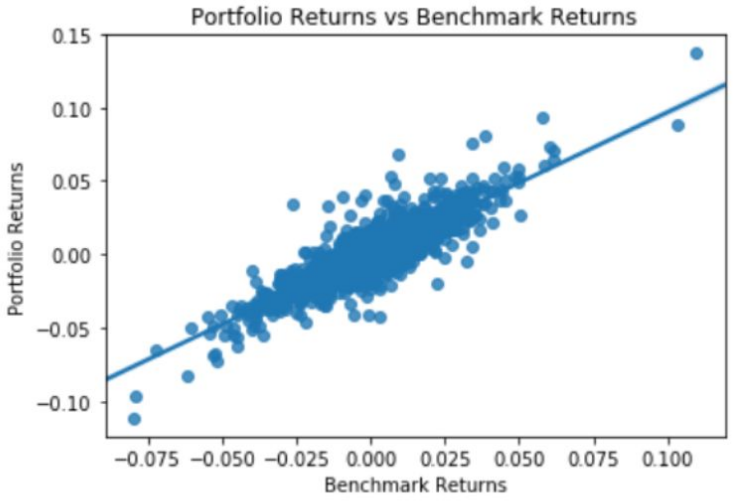
The alpha ratio is often used along with the beta coefficient, which is a measure of the volatility of an investment. The two ratios are both used in the Capital Assets Pricing Model (CAPM) to analyze a portfolio of investments and assess its theoretical performance.

Calculate Beta (β) and Alpha (α)

Step 1: Calculate the portfolio daily return
Step 2: Calculate the benchmark return
Step 3: Check the correlation for the benchmark and portfolio
Step 4: Since our portfolio returns are highly correlated to the benchmark returns, we can use the regression model to calculate the portfolio beta and the portfolio alpha.
Step 5: Use the linear regression model to calculate the alpha and the beta







Value at Risk (VaR)

Value at risk (VaR) is a statistic used to try and quantify the level of financial risk within a firm or portfolio over a specified time frame. VaR provides an estimate of the maximum loss from a given position or portfolio over a period of time, and you can calculate it across various confidence levels.

Estimating the risk of a portfolio is important to long-term capital growth and risk management, particularly within larger firms or institutions. VaR is typically framed as something like this: "We have a portfolio VaR of 250,000 USD over the next month at 95% confidence" This means that, with 95% confidence, we can say that the portfolio's loss will not exceed 250,000 USD in a month.

How is VaR calculated?

There are two main ways to calculate VaR: Using Monte Carlo simulation or Using the variance-covariance method.

I will focus on using variance-covariance method. In short, the variance-covariance method looks at historical price movements (standard deviation, mean price) of a given equity or portfolio of equities over a specified lookback period, and then uses probability theory to calculate the maximum loss within your specified confidence interval.

VaR calculation Assumptions

1. Normal distribution of returns - VaR assumes the returns of the portfolio are normally distributed. This is of course not realistic for most assets, but allows us to develop a baseline using a much more simplistic calculation. (Modifications can be made to VaR to account for different distributions, but here we'll focus on the standard VaR calculation)

2. Standard market conditions - Like many financial instruments, VaR is best used for considering loss in standard markets, and is not well-suited for extreme/outlier events.

To calculate the VaR of a portfolio:

Step 1: Calculate periodic returns of the stocks in the portfolio
Step 2: Create a covariance matrix based on the returns
Step 3: Calculate the portfolio mean and standard deviation (weighted based on investment levels of each stock in portfolio)
Step 4: Calculate the inverse of the normal cumulative distribution (PPF) with a specified confidence interval, standard deviation, and mean
Step 5: Estimate the value at risk (VaR) for the portfolio by subtracting the initial investment from the calculation in step (4)

Portfolio VaR

Confidence Level	Value at Risk (VaR)
90%	$981,773
95%	$976,295
99%	$966,019

Assume Our Initial Investment is $1M USD:
With 95% confidence level, the maximum loss will be $976,295.
With 5% probability, the minimum loss will be $976,295.

Conclusion

From the stock performance, Amazon stock price increased from $47 to $2000 per share while other stocks prices increase to $200-250. However, the optimal weights shows only 14% for Amazon in our portfolio.

Given the risk and volatility of investing in stocks, the optimal portfolio would choose the highest expected return under the defined level of risk