S&P 500 or STOCKS

Cuong Ly

November 6 2021

Abstract

Researchers have shown that in the long term, grouped stocks outperform individual stocks. However, in recent years, growth stock such as Netflix (NFLX) has earned significant returns for investors. What benefit investors? How should we invest our money? We will analyze and compare the individual stocks and grouped stocks by calculating the expected return and risk of the investment.

1 Research Question

The main question is between investing in individual stocks or putting money in the S&P500 which is more profitable? Investors cannot purchase the S&P500 directly. Instead, they buy ETF. Exchange-traded fund or ETF is a type of security that tracks the index and can be purchased or sold as a single stock share. A well-known ETF is SPY, which tracks the S&P 500 index, which is the weighted capitalization market value of the biggest five hundred companies in the U.S. Given time and financial constraints as graduate students, correctly determine the time horizon and where to invest are the keys to financial stability. I hypothesize that if investors try to make quick cash or they have the knowledge or technology ability, they should invest in individual stocks. In contrast, uneducated or risk-averse investors are recommended to put their money in ETF.

2 Background

Buying ETFs is boring. We cannot get rich if we only stick to “average” return. Return is in single digit, a little more than putting money in saving accounts. Given the risk and time, invest in index funds may not make you richer. In contrast, with enough preparation and research, one can invest in individual stocks and earn some nifty profit. And of course, the risk is also greater than index funds. In this project, we hopefully can find the answer to this dilemma. This topic is applicable and appropriate to college and especially graduate students. As we soon graduate from school and get jobs, prepare for the future is never too late. In general, scholars and legendary investor like Warren Buffet advise people to put their savings in index funds.

3 Data Source

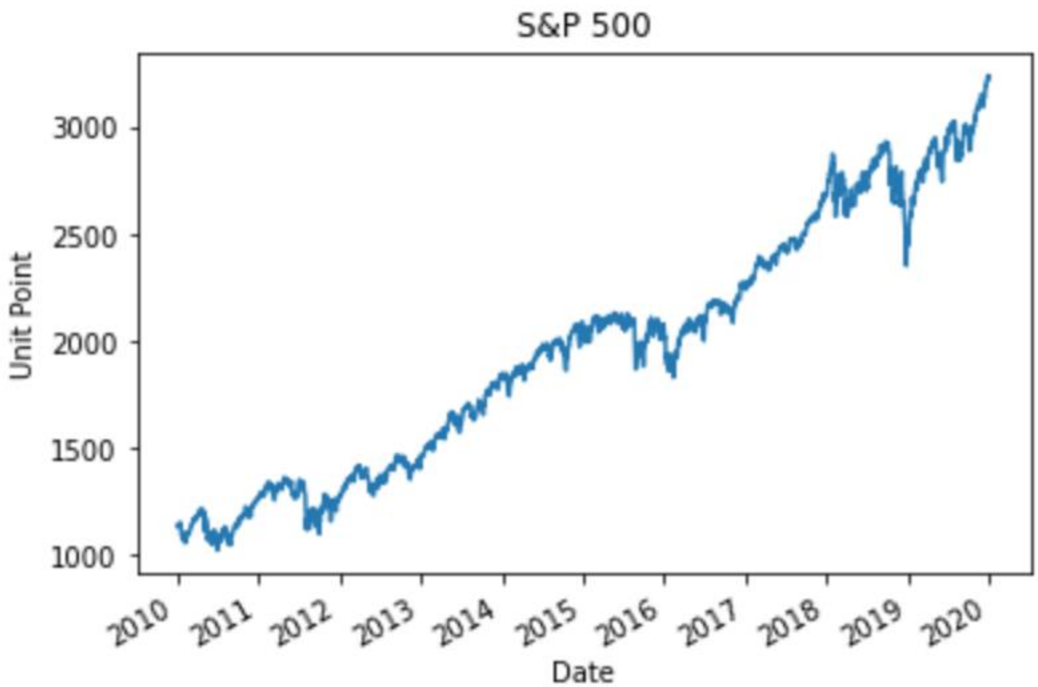
Yahoo Finance is a reliable source to get historical data. However, the site stops using the API and many programs that relied on it stop working. To solve this issue, the Python community created “yfinance” to download the data from yahoo finance. Yahoo Finance has a wide range of stocks with data dated back thirty years. When we first download the data, there are eight columns (variables) in the data set: date, open, high, low, close, and volume, dividends, and stock splits. However, we are only interested in two columns: date, adjusted close. Therefore, we will use pandas to create a new data frame. The date is a single trading day, except weekends and holidays.

4 Treatment

We will compare four stocks versus S&P 500 index fund. First, we need to download the data using yfinance. The next step is to eliminate empty rows, columns using the “drop” function. Since the data is daily, we need to group the data monthly to calculate the return. What we are interested in is the adjusted close price at the end of the day of the data. We will calculate the return based on the closing price. We will use the pct-change() function to calculate the return. For visualization, we will use matplotlib for graphing.

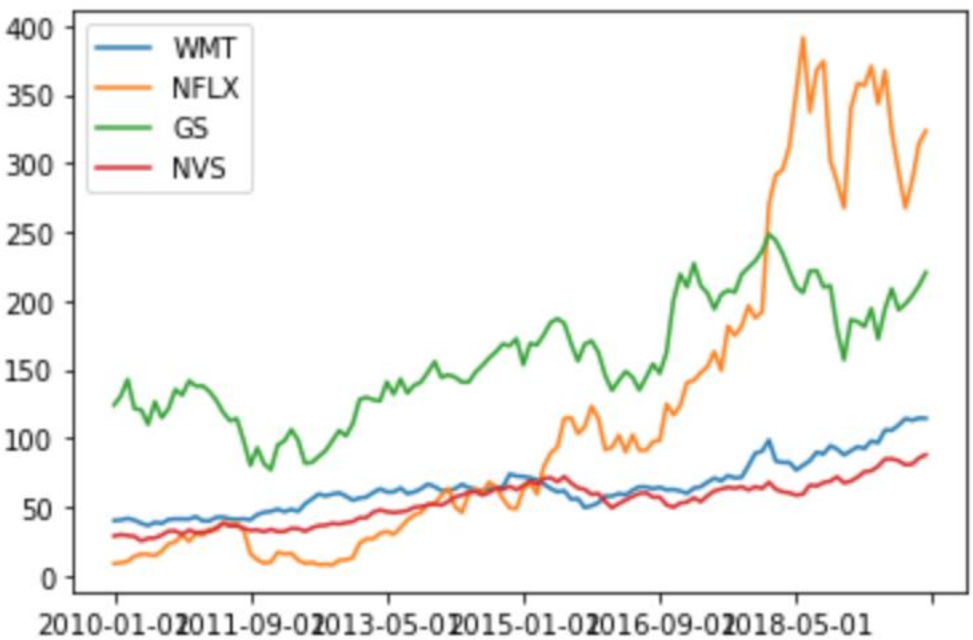
We look at the data from January 1st 2010 to January 1st 2021. That is a 10-year period. There was not much catastrophic events that cause a big wave in the market such as the Covid-19 that create chaos in March 2020. There are 2516 data points during the period, sufficient to test and make some conclusions. First, we need to get the data for S&P 500. We can do that by using yfinance module. There are multiples in the data frame. We need to eliminate the null values. After doing that, there are 2303 value points. There is currently one column, which is the data point of the index of that day. The data in the dataset is daily.

5 Exploratory Data Analysis



As we can see, the graph is non-stationary, as we should expect in finance. The line grows linearly. However, by eyeballing, we can point out there are three "chaos" during the 10 year period namely: mid 2012, 2016 and mid 2019. Bear these timelines in mind because later we will compare these to our portfolio timelines.

Below is a graph of our portfolio. As we can observe, Netflix (NFLX) earned investors nifty return, while the other three stocks grew constantly at a slow rate over 10 years.



The goal of this project is whether leaving money in S&P 500 index or actively invest in a group of stocks. We have computed the yearly S&P 500 return. The next step is finding the historical data of the stocks. I chose four stocks, in four different sectors Walmart (WMT) - Retail, Netflix (NFLX) -Technology, Goldman Sachs (GS) - Financial Services, and Novartis (NVS) – Healthcare. Each stock is equivalent to 25% of our portfolio. We use Yahoo Financials to download historical data. We are only interested in monthly adjusted close. Below is the first five data head of the dataset:

	WMT	NFLX	GS	NVS
2010-01-01	40.151688	8.892857	124.231789	28.989222
2010-02-01	40.632637	9.435714	130.605484	29.958588
2010-03-01	41.782402	10.534286	142.853775	29.297899
2010-04-01	40.536327	14.128571	121.563393	28.684473
2010-05-01	38.208736	15.878571	120.776443	25.390129

6 Testable Hypothesis and Model Description

We compute the expected monthly return using the CAPM model. Then we compare these with the actual return. We also find the return and risk for portfolio. Find the best optimized portfolio.

7 Model Output and Interpretation of Results

We calculate monthly return for each stock.

	WMT	NFLX	GS	NVS
2010-02-01	0.011978	0.061044	0.051305	0.033439
2010-03-01	0.028297	0.116427	0.093781	-0.022053
2010-04-01	-0.029823	0.341199	-0.149036	-0.020938
2010-05-01	-0.057420	0.123862	-0.006474	-0.114848
2010-06-01	-0.043734	-0.022492	-0.087770	0.073539

This does not tell us much how each stock return. So we compute the mean of each stock return.

```
WMT     0.010003
NFLX    0.044474
GS      0.007860
NVS     0.010512
```

It confirms from the graph earlier that Netflix performed best among stocks in our portfolio, about 4.4% every month.

As a next step, we can calculate the weighted monthly returns based on our portfolio allocation. Remember that we allocate a weight of 25% each to Walmart, Netflix, Goldman Sachs, and Novartis. Multiply the monthly return with the 25% weighted portfolio and then sum the result. The monthly return for our portfolio is about 1.8%. Annualize monthly return by this formula: [((1 + R)^12) - 1] x 100 where R is monthly return. For example, R in our case is 0.018. Using formula above, our yearly return is we get yearly return for the portfolio: 23.87%. Thus, over the 10 year period, on average, we get 23.87% return per year. However, is that number a good return? To answer that question, we do two things: look at how much risk we took to earn that 23.87% and compare the return of our portfolio to the benchmark return, which is S&P 500 in this case.

First, we look at risk by calculating portfolio standard deviation. Little background, portfolio standard deviation refers to the volatility of the portfolio. We calculate it using portfolio weight and covariance matrix and the transpose of portfolio weight. The formula for that is

$$\sigma_P^2 = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \mathbf{w}'\Sigma\mathbf{w}$$

Therefore, we need to calculate the covariance matrix since the portfolio weight is given. First, we calculate the monthly covariance and then we annualize it by multiplying by 12.

	WMT	NFLX	GS	NVS
WMT	0.027951	-0.005821	0.006048	0.007460
NFLX	-0.005821	0.359329	0.051869	0.010241
GS	0.006048	0.051869	0.073592	0.007394
NVS	0.007460	0.010241	0.007394	0.027303

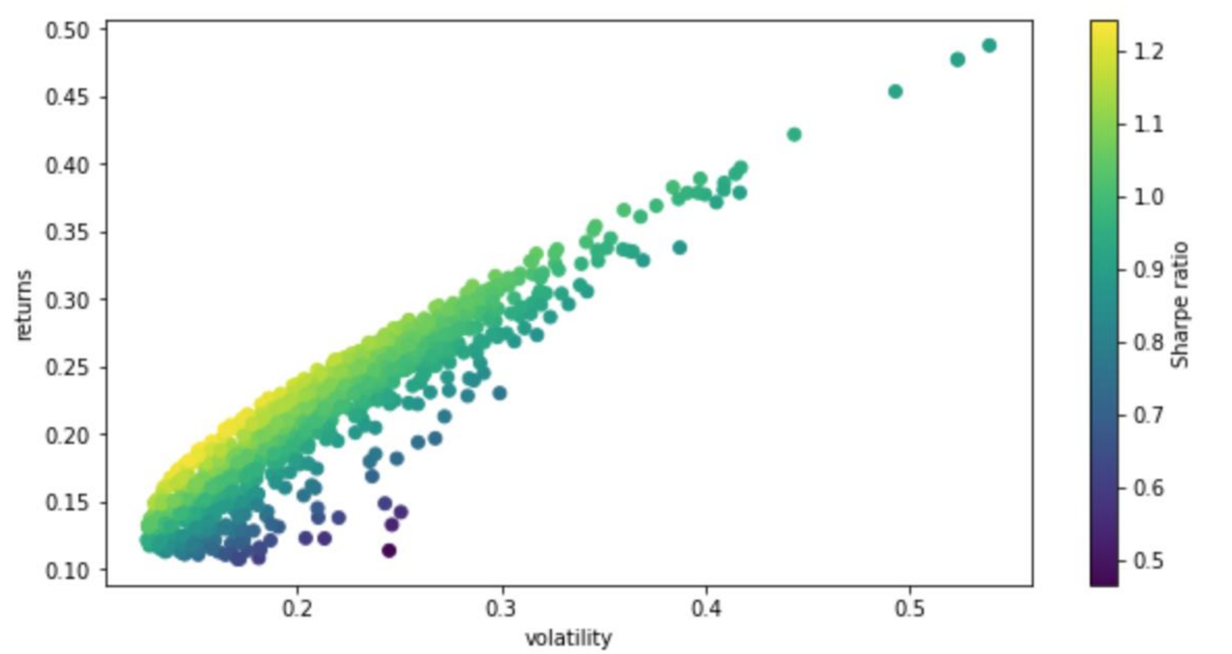
Having covariance matrix, we can calculate portfolio risk using the formula above with little help from numpy. The risk of our portfolio is about 20%. Okay, so the return is 23.87% annually in exchange for 20% risk. Let's see how these metric compared to the overall market.

In the previous calculation, the portfolio allocation is 25% evenly. I did it just to illustrate the portfolio. But how do we know it is the best allocation? We find the best portfolio allocation. We may have investors pursuing different objectives when optimizing their portfolio. For instance, young investors may prefer to find portfolios maximizing expected return. While older investors could aim to find portfolio minimizing the risk. To find an optimized portfolio allocation, one of the approaches is using Sharpe Ratio. Holding risk constant, the higher Sharpe Ratio earns a better portfolio return. We will generate 1000 portfolio with different level of allocation, calculate portfolio risk, portfolio return and Sharpe Ratio. The best portfolio has highest return, lowest portfolio risk and highest Sharpe Ratio.

We start by defining empty lists where we will append the calculated portfolio returns, risk and, Sharpe Ratio for each of the random portfolios. The calculation will happen in a for loop. Processed to calculate portfolio variance as we did in the 25% weight case. Also, we calculate Sharpe Ratio using this formula:

$$S_a = \frac{E[R_a - R_b]}{\sigma_a}$$

Where R_a is the annualized return, R_b is the risk free rate, and σ_a is the portfolio standard deviation. Now that we have created 1000 random portfolios, we can visualize them using matplotlib



As we can see from the graph, the best portfolio is the one with highest Sharpe Ratio (in the yellow color). Any portfolio that have in the neighborhood of 20% return and 15% risk is considered great portfolio. Previously, we used 25% equally weight and achieved 23.87% return with 20% risk. By looking at the graph, the 25% weight is “greener” and locate to upper right of “yellow”. We can do better and find an optimized allocation.

We can find the answer to that question by transforming our data into a Pandas DataFrame and performing some basic queries.

	Port Returns	Port Risk	Sharpe Ratio	Portfolio Weights
0	0.170947	0.187695	0.910774	[0.09708863498343116, 0.15047562586013344, 0.42688579409316413, 0.32554994506327134]
1	0.183521	0.177679	1.03288	[0.3266854451819343, 0.17475976645492639, 0.32907102129744864, 0.1694837670656907]
2	0.232431	0.27431	0.847331	[0.08989802770058927, 0.30637041085167444, 0.5536156857891772, 0.05011587565855928]
3	0.133907	0.1329	1.00758	[0.2406921995518703, 0.046522895704468555, 0.1847027633655259, 0.5280821413781353]
4	0.217403	0.185382	1.17273	[0.3329002577869336, 0.24429830192899232, 0.12533989123187966, 0.29746154905219446]

As an interpretation, looking at portfolio weights, we see that each row represents a different portfolio. For example, row 1 contains a portfolio with 9.7% weight in WMT, 15% in NFLX, 42.7% in GS, 32.5% in NVS. Now, we are ready to use Pandas methods such as idmax and idmin. They will allow us to find out which portfolio has the highest returns and Sharpe Ratio and minimum risk

```
Port Returns                                                                    0.19846
Port Risk                                                                      0.159293
Sharpe Ratio                                                                    1.24588
Portfolio Weights    [0.47057214784227974, 0.1927046265773679, 0.0286475974917604, 0.308075628088592]
Name: 534, dtype: object
Port Returns                                                                       0.133398
Port Risk                                                                          0.126533
Sharpe Ratio                                                                        1.05425
Portfolio Weights    [0.4497773212108164, 0.039717872526725255, 0.09443107127785845, 0.41607373498459993]
```

We print out the results for highest Sharpe Ratio and the portfolio with the lowest risk. Looking at the result, if investors like to go “all-in”, the recommended portfolio weights are 47% weight in WMT, 19% in NFLX, 2.8% in GS, 30.8% in NVS. This portfolio earns nearly 20% in return with 15% risk. In contrast, for risk-averse investor

Next, we also use yfinance to download the S&P500 data. Below is the first five of dataset.

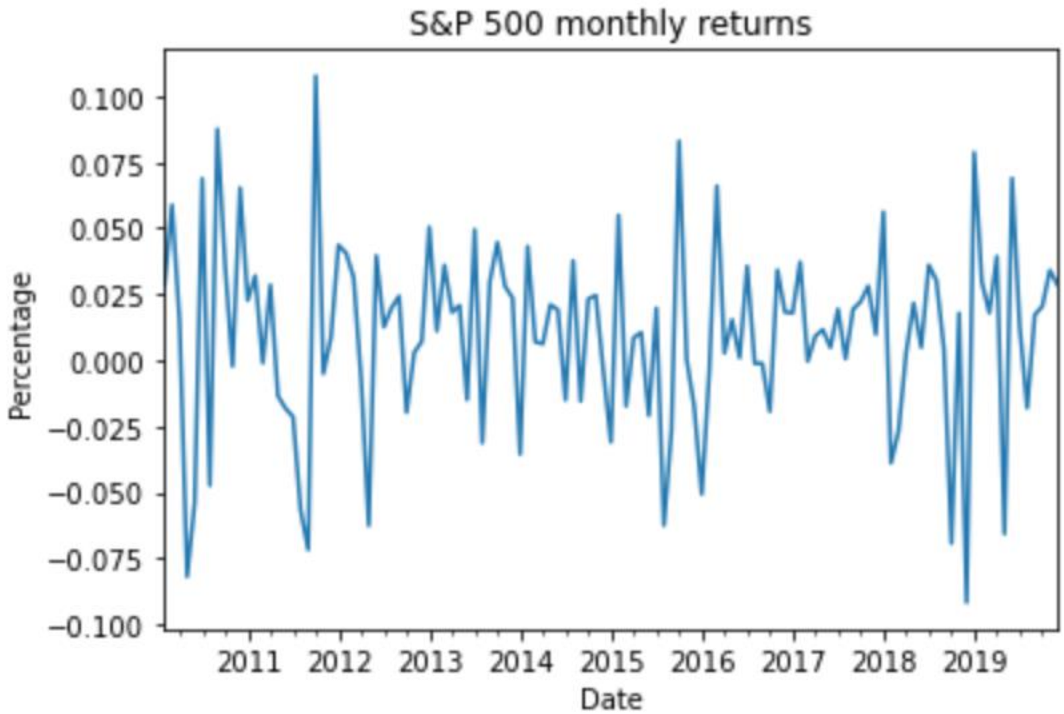
Date	Open	High	Low	Close	Adj Close	Volume
2010-01-04	1116.560059	1133.869995	1116.560059	1132.989990	1132.989990	3991400000
2010-01-05	1132.660034	1136.630005	1129.660034	1136.520020	1136.520020	2491020000
2010-01-06	1135.709961	1139.189941	1133.949951	1137.140015	1137.140015	4972660000
2010-01-07	1136.270020	1142.459961	1131.319946	1141.689941	1141.689941	5270680000
2010-01-08	1140.520020	1145.390015	1136.219971	1144.979980	1144.979980	4389590000

However, we are only interested in adj close, fortunately, we can use pandas to do that by using the square bracket []. We plot the data to have a general view

We also restructure the data to monthly frequency by using the last data point of that month. Calculate return by pct_change() and drop any null values

```
Date
2010-02-28    0.028514
2010-03-31    0.058796
2010-04-30    0.014759
2010-05-31   -0.081976
2010-06-30   -0.053882
```

A visualization of the monthly return. An interesting observation is the market was very volatile around August 2011, November 2015 and the beginning of 2019. For example, monthly return was down 7.5% in June compared to May, then one or two months later, it was up 10%. This confirmed our finding from earlier from the S&P 500 graph. We will do some comparison with our portfolio later to see how our portfolio performed during those volatile time.



The risk of S&P 500 is calculated using standard deviation is about 3.59% yearly. The expected monthly return, on average is about 0.99%. Thus, that should be 12.54% yearly. To sum up, if investors invest in only S&P 500, they would expect to receive 12.54% per year in exchange for 3.59% in risk. In comparison to our portfolio, clearly, S&P 500 is a safer investment. However, it also has a lower rate of return.

Now that we know expected return and risk of the market (S&P500) and risk and return of our portfolio. One may ask instead of investing in a portfolio, can we invest in one individual stock? How is it compared to the portfolio and the market. We can do just that using the famous CAPM model. Briefly about the model, this model represents the relationship between the expected return on a risky asset and the market risk

$$E(r_i) = r_f + \beta_i(E(r_m) - r_f)$$

Here, $E(r_i)$ denotes the expected return on asset i, r_f is the risk-free rate (such as a government bond), E(r_m) is the expected return on the market, and β is the beta coefficient.

To make thing simple, in current times, the risk-free rate is so low that we assume it is equal to zero.

Beta can be interpreted as how the asset's sensitivity compared to the general market. We can derive Beta coefficient from taking the covariance of asset and market, divides by the variance of market

$$\beta = \frac{cov(R_i, R_m)}{var(R_m)}$$

Walmart (WMT)

We start off by working with Walmart data. After downloading data and using statsmodels package, specifically OLS regression, here is our result

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    WMT   R-squared:                       0.096
Model:                            OLS   Adj. R-squared:                  0.088
Method:                 Least Squares   F-statistic:                     12.45
Date:                Wed, 15 Dec 2021   Prob (F-statistic):           0.000599
Time:                        13:07:59   Log-Likelihood:                 199.55
No. Observations:                 119   AIC:                            -395.1
Df Residuals:                     117   BIC:                            -389.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0059      0.004      1.351      0.179      -0.003       0.014
market         0.4132      0.117      3.528      0.001       0.181       0.645
==============================================================================
Omnibus:                        8.958   Durbin-Watson:                   2.035
Prob(Omnibus):                  0.011   Jarque-Bera (JB):               19.839
Skew:                           0.010   Prob(JB):                     4.92e-05
Kurtosis:                       5.000   Cond. No.                         28.0
==============================================================================
```

These results indicate that the beta (denoted as market here) is equal to 0.41, which means that Walmart's returns are 41% less volatile than the market, meaning if the market rises 1, Walmart stock rises 0.41 on average. The value of the intercept is relatively small and statistically insignificant at the 5% significance level. Now that we have β, we can go ahead and calculate expected return on Walmart stock. We previously calculate expected return on market is 12.54%.

Expected return on Walmart stock = 0 + 0.41 * (0.1254 – 0) = 0.051414

Thus, expected return on Walmart stock is approximately 5%, which is much lower compared to the market overall or our portfolio.

Netflix (NFLX)

We proceed to calculate Netflix beta

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   NFLX   R-squared:                       0.062
Model:                            OLS   Adj. R-squared:                  0.054
Method:                 Least Squares   F-statistic:                     7.772
Date:                Wed, 15 Dec 2021   Prob (F-statistic):            0.00619
Time:                        13:01:36   Log-Likelihood:                 44.226
No. Observations:                 119   AIC:                            -84.45
Df Residuals:                     117   BIC:                            -78.89
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0325      0.016      2.030      0.045       0.001       0.064
market         1.2044      0.432      2.788      0.006       0.349       2.060
==============================================================================
Omnibus:                       31.893   Durbin-Watson:                   1.827
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               86.527
Skew:                           0.971   Prob(JB):                     1.63e-19
Kurtosis:                       6.699   Cond. No.                         28.0
==============================================================================
```

These results indicate that the beta is equal to 1.2044, which means that Netflix's returns are 120% more volatile than the market, meaning if the market rises 1, Netflix stock rises 1.2 on average.

Expected return on Netflix stock = 0 + 1.2044 * (0.1254 – 0) = 0.15103

Thus, expected return on Netflix stock is approximately 15%.

Goldman Sachs (GS)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     GS   R-squared:                       0.451
Model:                            OLS   Adj. R-squared:                  0.446
Method:                 Least Squares   F-statistic:                     96.13
Date:                Wed, 15 Dec 2021   Prob (F-statistic):           6.30e-17
Time:                        13:35:59   Log-Likelihood:                 170.80
No. Observations:                 119   AIC:                            -337.6
Df Residuals:                     117   BIC:                            -332.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0067      0.006     -1.210      0.229      -0.018       0.004
market         1.4620      0.149      9.804      0.000       1.167       1.757
==============================================================================
Omnibus:                        9.063   Durbin-Watson:                   2.115
Prob(Omnibus):                  0.011   Jarque-Bera (JB):               11.971
Skew:                           0.408   Prob(JB):                      0.00251
Kurtosis:                       4.322   Cond. No.                         28.0
==============================================================================
```

These results indicate that the beta is equal to 1.4620, which means that Goldman's returns are 146% more volatile than the market, meaning if the market rises 1, Goldman stock rises 1.2 on average.

Expected return on Goldman stock = 0 + 1.4620 * (0.1254 – 0) = 0.1833

Thus, expected return on Goldman stock is approximately 18.3%.

Novartis (NVS)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    NVS   R-squared:                       0.193
Model:                            OLS   Adj. R-squared:                  0.186
Method:                 Least Squares   F-statistic:                     27.90
Date:                Wed, 15 Dec 2021   Prob (F-statistic):           5.99e-07
Time:                        13:44:32   Log-Likelihood:                 208.11
No. Observations:                 119   AIC:                            -412.2
Df Residuals:                     117   BIC:                            -406.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0042      0.004      1.034      0.303      -0.004       0.012
market         0.5756      0.109      5.282      0.000       0.360       0.791
==============================================================================
Omnibus:                        1.079   Durbin-Watson:                   2.181
Prob(Omnibus):                  0.583   Jarque-Bera (JB):                1.022
Skew:                           0.047   Prob(JB):                        0.600
Kurtosis:                       2.556   Cond. No.                         28.0
==============================================================================
```

These results indicate that the beta is equal to 0.5756, which means that Novartis's returns are 57% less volatile than the market, meaning if the market rises 1, Goldman stock rises 0.5756 on average.

Expected return on Novartis stock = 0 + 0.5756 * (0.1254 – 0) = 0.0721

Thus, expected return on Novartis stock is approximately 7.2%.

8 Conclusion and Discussion

Asset	Expected Return (%)	Actual Return (%)	Risk (%)
Market (S&P 500)		12.54	3.59
Portfolio	23.87		20
Walmart (WMT)	5.1	12.62	4.1
Netflix (NFLX)	15	68.56	20
Goldman Sachs (GS)	18	9.85	46
Novartis (NVS)	7.2	12.58	5.7

As we can see from the summary table, some observations. The risk of the market is the lowest as we expect. Netflix is considered a growth stock. It outperforms the other stock in our portfolio. One may think that why don't we just invest in an individual stock, let's say Netflix for example, then we may earn a big return. The problem is we are not always sure Netflix always rise. Overall, if someone prefer to play it say, the data recommends sticking with S&P 500.

9 Appendix

```
# Download S&P500 data using yfinance,
sp500 = yf.download('^GSPC', start=START_DATE,end=END_DATE)
sp500.head()
```

```
[*********************100%***********************]  1 of 1 completed
```

Date	Open	High	Low	Close	Adj Close	Volume
2010-01-04	1116.560059	1133.869995	1116.560059	1132.989990	1132.989990	3991400000
2010-01-05	1132.660034	1136.630005	1129.660034	1136.520020	1136.520020	2491020000
2010-01-06	1135.709961	1139.189941	1133.949951	1137.140015	1137.140015	4972660000
2010-01-07	1136.270020	1142.459961	1131.319946	1141.689941	1141.689941	5270680000
2010-01-08	1140.520020	1145.390015	1136.219971	1144.979980	1144.979980	4389590000

Portfolio consists of four stocks: Walmart (WMT), Netflix (NFLX), Goldman Sachs (GS) and Novartis (NVS)

[ ]:
```
stocks = ['WMT', 'NFLX', 'GS','NVS']
# Import stock data from Yahoo Finance
yahoo_financials = YahooFinancials(stocks)
data = yahoo_financials.get_historical_price_data(start_date='2010-01-01', end_date='2020-01-01', time_interval='monthly')
adj_close = pd.DataFrame({a: {x['formatted_date']: x['adjclose'] for x in data[a]['prices']} for a in stocks})
```

[ ]:
```
adj_close.head()
```

[ ]:

	WMT	NFLX	GS	NVS
2010-01-01	40.359360	8.892857	124.231819	30.977274
2010-02-01	40.842773	9.435714	130.605530	32.013130
2010-03-01	41.998482	10.534286	142.853760	31.307131
2010-04-01	40.746346	14.128571	121.563438	30.518614
2010-05-01	38.406700	15.878571	120.776428	27.013620