Team_11_Assignment

November 26, 2021

[3]:
```python
from IPython.display import display, Math, Latex

import pandas as pd
import numpy as np
import numpy_financial as npf
import yfinance as yf
import matplotlib.pyplot as plt
from datetime import datetime
import concurrent.futures
```

0.1 Group Assignment

0.1.1 Team Number: 11

0.1.2 Team Member Names: Richard Yang, William Zhang, Soham Basu

0.1.3 Team Strategy Chosen: SAFE

[4]:
```python
# Extract all the tickers from the csv file
# Use loop to append tickers into a empty list called tickers_lst one by one
givenTickers = pd.read_csv('Tickers.csv')
tickers_lst = []
for i in range(len(givenTickers)):
    tickers_lst.append(givenTickers.iloc[i, 0])


# Function outputs true if the stock is in USD, otherwise false
# The parameter is a string representing the ticker name
def currencyUSD(stock):
    if stock.info['currency'] == 'USD':
        return True
    else:
        return False

# Function outputs true if the stock has a daily average volume of at least␣
↪10000 shares in a given time period, otherwise false
# The first parameter is a string representing the ticker name, the second and␣
↪third parameter specify the time period
def volume10000(stock, start_date, end_date):
```

```
    hist = stock.history(start=start_date, end=end_date)
    if hist['Volume'].mean() >= 10000:
        return True
    else:
        return False
```

[5]:
```
# Create an empty list filterTickers, then filter out stocks that meet the␣
 ↪following requirements:
# 1. It is a US listed stock
# 2. Has a daily average volume of at least 10000 shares from July 02, 2021 to␣
 ↪October 22, 2021
filterTickers = []
start_date = '2021-07-02'
end_date = '2021-10-22'

# The parameter is a string
# This function will firstly check if the ticker is valid or not(we use .
 ↪info['regularMarketPrice'] to do that)
# Once the ticker is valid, the function will check whether the stock meet BOTH␣
 ↪two requirements stated above
# If the ticker meets all the requriements, the function will append this␣
 ↪ticker into "filterTicker" list
def filterStocks(ticker):
    stock = yf.Ticker(ticker)
    if stock.info['regularMarketPrice'] is not None:
        if currencyUSD(stock) & volume10000(stock, start_date, end_date):
            filterTickers.append(ticker)

for ticker in tickers_lst:
    filterStocks(ticker)
```

Volatility/Standard Deviation:

Weight: 45%

Volatility maps how much a stock deviates over a certain period. The stocks that have lower volatility deviate less. The less a stock deviates, the safer it is. The reason volatility has such a high weighting in our overall score is because of its ability to weed out the riskiest stocks from our portfolio by assigning them a low score. We were also aiming to find a balance between volatility in the short-term and long-term. The pandemic has had a major impact on stock prices over the last 2 years, which is why we also wanted to look at historical data over the last 10 years to get a bigger picture. That is why our final volatility calculates a weighted average of the volatility over the last 2 years, worth 25%, and the volatility over the last 10 years, worth 75% of the final volatility.

Sharpe Ratio:

Weight: 40%

Sharpe ratio is a way to measure how much return an investor can get per unit of risk. It takes both the average return and standard deviation into account. This allows us to look at whether the stock is worth buying. One potential dilemma with the sharpe ratio is that a stock with a low return and volatility has the same sharpe ratio as a stock with a high return and high volatility. Since we are trying to build a portfolio of safe stocks, it is important that we get rid of most stocks with high volatilities. In our portfolio, we can do this by giving standard deviation a high weighting, allowing it to weed out the riskiest of stocks. Once again, in order to get a balance between sharpe ratio in the short-term and sharpe ratio in the long term, we calculated mean return using a weighted average of the return over the last 2 years, worth 25%, and the return over the last 10 years, worth 75%.

Covariance:

Weight: 10%

We choose stocks which have high covariance with S&P 500. This is to find stocks that basically follow the market trend. That means the stocks do not have high non-systematic risk. Since the market index like S&P 500 normally has low volatility, stocks chosen based on this criterion will be safe.

Moving Price Average:

Weight: 5%

The reason for looking at the Two Hundred Moving Price Average is simply to get another indicator to see how safe the stock is. The closer a stock is to its moving price average, the more likely that it is a safer stock as it likely deviates less than other, more riskier stocks. Of course, a really risky stock could just happen to be very close to its moving price average; however, when combined with other indicators such as volatility and covariance, moving price average is a good way of finding safe stocks.

We wanted to use a variety of indicators such as volatility, sharpe ratio, covariance with the S&P 500 index, and moving price average. In addition, the data used to calculate many of these indicators aimed to strike a balance between the short-term stock history and long-term stock history. This helps to diversify our indicators and allows us to pick better stocks. Say for example, a stock gets a poor correlation with the S&P 500. If other factors such as volatility and moving price average show that the stock is safe, then it has a high chance of still being included in our portfolio. We decided the weights for each indicator based on how well each indicator can predict how safe a stock is. Multiple portfolios have been generated using different weights for the indicators, and this was found to be the optimal weighting.

[6]:
```
# uses factors such as a stock's volatility/standard deviation, sharpe ratio,␣
↪covariance, and moving price average
# to assign each stock a score
# the higher the score, the safer the stock is and the better it is for our␣
↪portfolio
def calculateScore(stockInfo, stockHist_2, stockHist_10, price, indexHist):

    # finds volatility (standard deviation) using the daily log returns in␣
↪terms of percentage,
    # and finds the volatility as an annual percentage
    stockHist_2['Log returns'] = np.log(stockHist_2['Close'] /␣
↪stockHist_2['Close'].shift())
```

```python
    volatility_2 = stockHist_2['Log returns'].std() * 252 ** 0.5
    stockHist_10['Log returns'] = np.log(stockHist_10['Close'] / stockHist_10['Close'].shift())
    volatility_10 = stockHist_10['Log returns'].std() * 252 ** 0.5
    # calculates a weighted average for volatility based on the average volatility over the last 2 years
    # and over the last 10 years
    # volatility over last 2 years: 25%, volatility over last 10 years: 75%
    volatility = ((volatility_2 * 0.25) + (volatility_10 * 0.75))

    # calculates the average return of the stock
    meanReturn_2 = stockHist_2['Log returns'].mean() * 252
    meanReturn_10 = stockHist_10['Log returns'].mean() * 252
    # finds a weighted average for return based on the average return over the last 2 years and
    # over the last 10 years
    # return over last 2 years: 25%, return over last 10 years: 75%
    meanReturn = (meanReturn_2 * 0.25) + (meanReturn_10 * 0.75)
    # uses mean return and volatility to calculate sharpe ratio
    # assumes risk-free rate of return is 0
    sharpeRatio = meanReturn / volatility

    # finds covariance by comparing the stock to the S&P 500 index using data over the last 10 years
    # organizes price data
    prices = pd.DataFrame(stockHist_10['Close'])
    prices.columns = ['Stock']
    prices['index'] = indexHist['Close']
    # calculates stock's correlation to S&P 500 index
    covariance = prices.corr()
    corr = float(covariance.iloc[0, 1])

    # calculates the stock's percent difference from its Two Hundred Day Moving Average to its current price
    price_average = stockInfo["twoHundredDayAverage"]
    percent_difference = (price - price_average) / price_average

    # uses all 4 factors to assign a final score to each ticker:
    # Volatility: 45%
    # Sharpe Ratio: 40%
    # Correlation: 10%
    # Moving Day Price Average: 5%
    score = ((1 - volatility) * 0.45) + (sharpeRatio * 0.4) + (corr * 0.1) + ((1 - percent_difference) * 0.05)
    return score
```

[ ]:
```
score_lst = []
price_lst = []

# gets data for each ticker and S&P 500 in order to calculate a score for each␣
↪stock
def getStockData(ticker):
    # finds stock data over last 2 and 10 years respectively
    stock = yf.Ticker(ticker)
    stockInfo = stock.info
    stockHist_2 = stock.history(period="2y")
    stockHist_10 = stock.history(period="10y")
    # gets the stock's current price and appends it to a list of prices
    price = stock.history(start="2021-11-26", end="2021-11-27")
    price_lst.append(price["Close"])
    # finds data for S&P 500 index over last 10 years
    index = yf.Ticker('^GSPC')
    indexHist = index.history(period="10y")  # Time period used to calculate␣
↪the covariance
    # calculates a final score for each stock and appends it to a list of scores
    score_lst.append(calculateScore(stockInfo, stockHist_2, stockHist_10,␣
↪price, indexHist))

for ticker in filterTickers:
    getStockData(ticker)
```

[7]:
```
# Already get one list of tickers(filterTickers), and one list of␣
↪score(score_lst), and one list of current prices of those stocks
# We want to find the 20 stocks with the highest score, and find their␣
↪corresponding current price and give them the weight
# create one list to store the stock with highest score to the stock with the␣
↪20th highest score (chosen_tickers)
# the weight should be (6.9%+6.7%+ ...3.3%+3.1%=100%), corresponding to stocks␣
↪of highest score to stocks of lowest score(among 20 stocks we choose)
# In list "chosen_weights", there will be a list of numbers from 0.069, to 0.
↪031(20 numbers)
# Then get the corresponding current price for those 20 stocks

chosen_tickers = []
chosen_weights = []
chosen_prices = []
chosen_scores = []
maxWeight=0.069

for i in range(20):
    max_score = max(score_lst)
    max_index = score_lst.index(max_score)
```

```
    chosen_tickers.append(filterTickers[max_index])
    chosen_weights.append(maxWeight)
    chosen_prices.append(price_lst[max_index])
    chosen_scores.append(score_lst[max_index])
    score_lst.pop(max_index)
    filterTickers.pop(max_index)
    price_lst.pop(max_index)
    maxWeight -= 0.002
    maxWeight = round(maxWeight, 3)
```

[8]:
```
# Starting value of $100,000
initial_investment=100000
# Create a dataframe called FinalPortfolio by using the list chosen_tickers
FinalPortfolio = pd.DataFrame(chosen_tickers)
# Change the column name
FinalPortfolio.columns=["Ticker"]
# Add another column, the current price of each stock, into the dataframe
FinalPortfolio["Price"]=chosen_prices
chosen_weights=list(chosen_weights)
FinalPortfolio['Shares']=0
FinalPortfolio
for i in range(len(FinalPortfolio['Price'])):
    FinalPortfolio["Shares"].iloc[i]=initial_investment/FinalPortfolio["Price"].
 ↪iloc[i]*chosen_weights[i]
FinalPortfolio['Value']=FinalPortfolio["Shares"]*FinalPortfolio["Price"]
FinalPortfolio['Weight']=chosen_weights
FinalPortfolio.index = np.arange(1,len(FinalPortfolio)+1)
```

```
C:\Users\A\anaconda3\lib\site-packages\pandas\core\indexing.py:1637:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  self._setitem_single_block(indexer, value, name)
```

[9]:
```
#Creating variable for total portfoliio weight to confirm the total weight is␣
 ↪100
totalWeight=FinalPortfolio["Weight"].sum()
print("Your total overall portfolio weighting is " + str(totalWeight))
#creating variable for overall portfolio value to confirm it is $100000
totalValue=FinalPortfolio['Value'].sum()
print("Your total overall value weighting is $" + str(totalValue))
```

```
Your total overall portfolio weighting is 1.0
Your total overall value weighting is $100000.0
```

[10]:
```
######## Creating the portfolio
# The input is a DataFrame containing columns Ticker, Price, Shares, Value,␣
↪Weight(eg FinalPortfolio)
# The function outputs another DataFrame containing only the portfolio value␣
↪created by the input stocks and back in 3 years
# We make the initial investment to be $100,000 three years ago, and track the␣
↪portfolio value change back in the past 3 years
def portfoliovalue(finalportfolio):
    PortfolioHistory=pd.DataFrame()
    PortfolioValue=pd.DataFrame()
    shares=0
    for i in range(len(finalportfolio)):
        stock = yf.Ticker(finalportfolio.iloc[i, 0])
        stockHist_3=stock.history(period='3y').resample('MS').first()
        shares=initial_investment/stockHist_3.Close[0]*chosen_weights[i]
        PortfolioHistory[finalportfolio.iloc[i, 0]] = stockHist_3.Close*shares
    PortfolioValue["Value"] = PortfolioHistory.sum(axis=1)
    return PortfolioValue
```

[11]:
```
# Find the 10 stocks with the lowest score and create a dataframe for it(the␣
↪same as FinalPortfolio)

min_chosen_tickers = []
min_chosen_prices = []
min_chosen_weight = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1]

for i in range(10):
    min_score = min(score_lst)
    min_index = score_lst.index(min_score)
    min_chosen_tickers.append(filterTickers[min_index])
    min_chosen_prices.append(price_lst[min_index])
    score_lst.pop(min_index)
    filterTickers.pop(min_index)
    price_lst.pop(min_index)

# Starting value of $100,000
initial_investment=100000
# Create a dataframe called riskyPortfolio by using the list chosen_tickers
riskyPortfolio = pd.DataFrame(min_chosen_tickers)
# Change the column name
riskyPortfolio.columns=["Ticker"]
# Add another column, the current price of each stock, into the dataframe
riskyPortfolio["Price"]=min_chosen_prices
riskyPortfolio['Shares']=0
min_chosen_weight = list(min_chosen_weight)
for i in range(len(riskyPortfolio['Price'])):
```

```
    riskyPortfolio["Shares"].iloc[i]=initial_investment/riskyPortfolio["Price"].
↪iloc[i]*min_chosen_weight[i]
riskyPortfolio['Value']=riskyPortfolio["Shares"]*riskyPortfolio["Price"]
riskyPortfolio['Weight']=min_chosen_weight
riskyPortfolio.index = np.arange(1,len(riskyPortfolio)+1)
```

Our portfolio, consisting of the stocks with the highest 20 scores, focuses on maintaining a low volatility/standard deviation as that indicator makes up 45% of our portfolio and has further influence in the calculation of the sharpe ratio. Therefore, as the data proves, our portfolio has a much lower volatility than the ten worst performing stocks in the list of tickers, calculated based off our four indicators for safe stocks. In addition our sharp ratio is much higher for our Final Portfolio compared to the portfolio of the stocks with the ten lowest scores due to factoring the sharpe ratio values into our model as can be seen below.

[12]:
```
Standard_deviation_FinalPortfolio = portfoliovalue(FinalPortfolio).pct_change().
↪std()
Sharpe_ratio_FinalPortfolio = portfoliovalue(FinalPortfolio).pct_change().
↪mean() / portfoliovalue(FinalPortfolio).pct_change().std()
print("For the FinalPortfolio, the standard deviation␣
↪is",float(Standard_deviation_FinalPortfolio),", and the Sharpe ratio is",␣
↪float(Sharpe_ratio_FinalPortfolio),".")
```

```
The standard deviation is 0.0526273510181542 , and the Sharp ratio is
0.3402303536713587 .
```

[23]:
```
Standard_deviation_riskyPortfolio = portfoliovalue(riskyPortfolio).pct_change().
↪std()
Sharpe_ratio_riskyPortfolio = portfoliovalue(riskyPortfolio).pct_change().
↪mean() / portfoliovalue(riskyPortfolio).pct_change().std()
print("For a riskier portfolio with equal weightings for the stocks with the 10␣
↪lowest scores, the standard deviation␣
↪is",float(Standard_deviation_riskyPortfolio), ", and the Sharpe ratio is",␣
↪float(Sharpe_ratio_riskyPortfolio),".")
```

```
For a riskier portfolio with equal weightings for the stocks with the 10 lowest
scores, the standard deviation is 0.06535541862787961 , and the Sharpe ratio is
0.2782827399715921 .
```

We can also check the standard deviation and Sharpe ratio of S&P 500 as benchmarks.

[24]:
```
SP500 = yf.Ticker('^GSPC')
SP500hist = SP500.history(period='3y').resample('MS').first()

Standard_deviation_SP500hist = SP500hist.Close.pct_change().std()
Sharpe_ratio_SP500hist = SP500hist.Close.pct_change().mean()/SP500hist.Close.
↪pct_change().std()
print("For the S&P 500, the standard deviation␣
↪is",float(Standard_deviation_SP500hist),", and the Sharpe ratio is",␣
↪float(Sharpe_ratio_SP500hist),".")
```

For the S&P 500, the standard deviation is 0.060465000861257195 , and the Sharpe ratio is 0.2816680973291308 .

[15]:
```
#calculating the beta
MarketIndex='^GSPC' #This is the symbol yfinance uses for the S&P 500
Ticker2 = yf.Ticker(MarketIndex)

#putting start and end date for getting s and p 500
start_date = '2017-01-01'
end_date = '2021-11-26'

MarketIndex_hist = Ticker2.history(start=start_date, end=end_date)
MarketIndex_hist= MarketIndex_hist.resample('MS').first()
#putting value/price of s and p 500 and final portfolio in one dataframe
prices = pd.DataFrame(portfoliovalue(FinalPortfolio)['Value'])
prices.columns = ["Final Portfolio"]
prices[MarketIndex] = MarketIndex_hist['Close']
prices.drop(index=prices.index[0], inplace=True)

#calculating percentage change and dropping the first value
monthly_returns=prices.resample('MS').ffill().pct_change()
monthly_returns.drop(index=monthly_returns.index[0], inplace=True)

#calculating the market variance
MarketVar= monthly_returns[MarketIndex].var()


#calculating beta by taking a covariance
Beta=monthly_returns.cov()/MarketVar
print("Beta:")
print(Beta)


print('The Final Portfolio Beta is: ', Beta.iat[0,1])
print('The Beta of the Market is: ', Beta.iat[1,1])
```

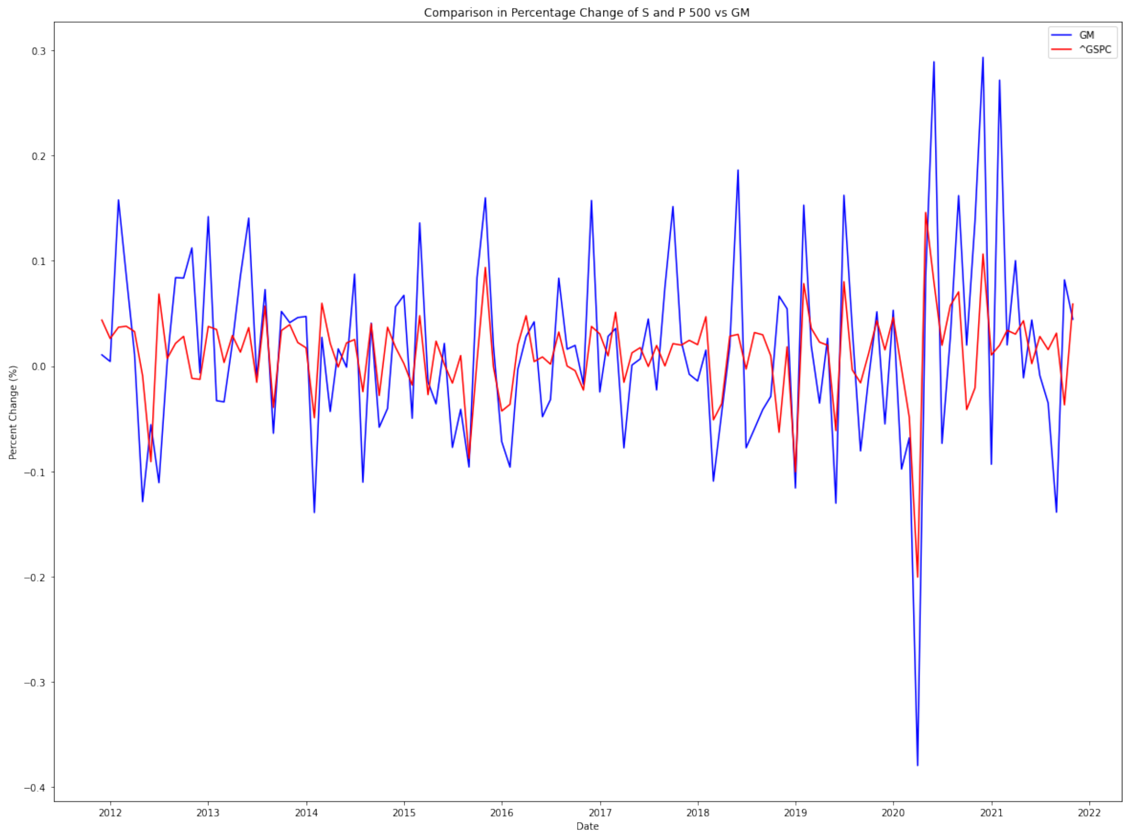
```
Beta:
                 Final Portfolio     ^GSPC
Final Portfolio         0.758194  0.843748
^GSPC                   0.843748  1.000000
The Final Portfolio Beta is:  0.8437479439035648
The Beta of the Market is:  1.0
```

Graph for correlation:

This is an example of a stock and its correlation with the S and P 500 in our portfolio. Based on its movements, it seems to be positively correlated with the S and P 500 although not completely. Stocks with similar trajectories to the S and P 500's movements may suggest that they are safer and less volatile as it has similar trends with the index.

[16]:

```
# Example of correlation between a stock in our portfolio and S&P 500
Stock1=chosen_tickers[0]
Stock2='^GSPC'
Ticker1 = yf.Ticker(Stock1)
Ticker2 = yf.Ticker(Stock2)

# Look at past 10 years
start_date = '2011-11-26'
end_date = '2021-11-26'

#get stock history
Stock1_hist = Ticker1.history(start=start_date, end=end_date)
Stock2_hist = Ticker2.history(start=start_date, end=end_date)

#get close prices
prices = pd.DataFrame(Stock1_hist['Close'])
prices.columns = [Stock1]
prices[Stock2] = Stock2_hist['Close']
prices.head()

#Calculate the monthly returns from the price dataFrame.
monthly_returns=prices.resample('MS').first().pct_change()
monthly_returns.drop(index=monthly_returns.index[0], inplace=True)
monthly_returns.head()

#plot graph
plt.figure(figsize=(20,15))

#plot the lines
plt.plot(monthly_returns.index,monthly_returns[Stock1], color='b', label=Stock1)
plt.plot(monthly_returns.index,monthly_returns[Stock2], color='r', label=Stock2)

#legend
plt.legend(loc='best')

#titles and labels
plt.title("Comparison in Percentage Change of S and P 500 vs " + str(Stock1))
plt.xlabel('Date')
plt.ylabel('Percent Change (%)')

plt.show()


# create new cell
print('Correlation:')
print(monthly_returns.corr())
```



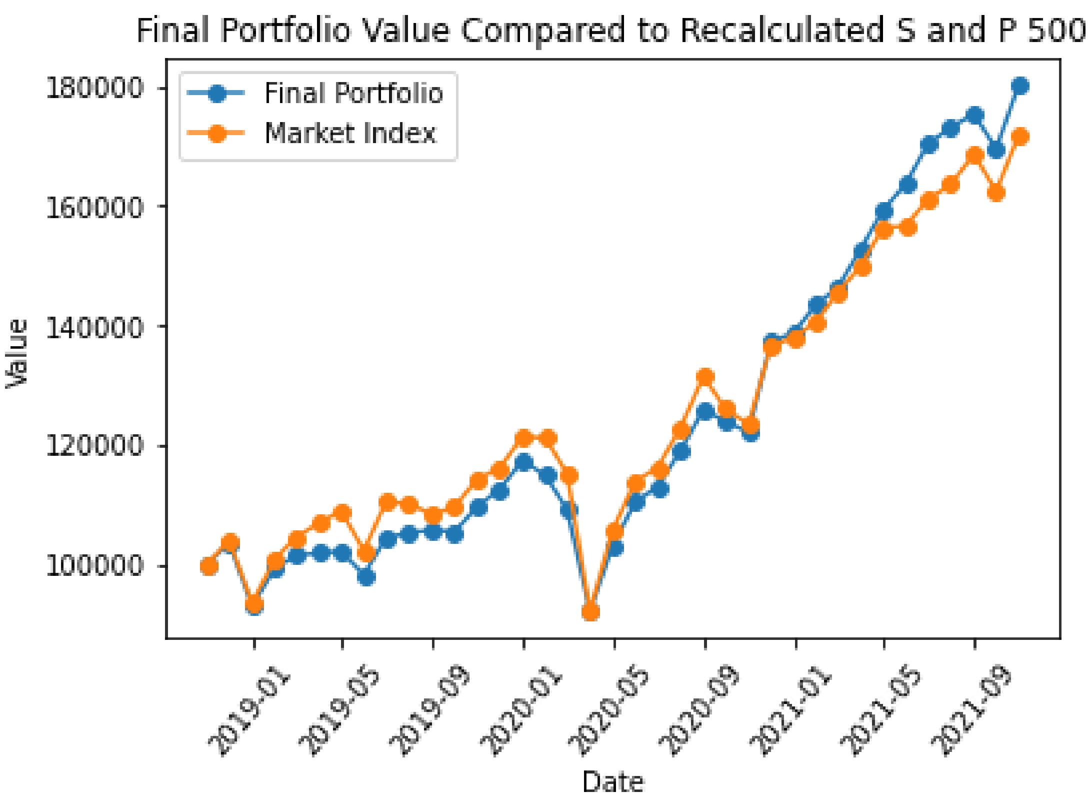

```
Correlation:
             GM     ^GSPC
GM     1.000000  0.625369
^GSPC  0.625369  1.000000
```

Graph for overall portfolio:

As you can see our portfolio generally has had a similar trend compared to the S and P 500 in the past 3 years, suggesting that our portfolio is quite safe with most of non-systematic risk removed. It supports our constructed models in creating a safe portfolio. Generally the growth is overall quite smooth with very little abrupt movements.

[18]:
```
#graphing our final portfolio vs S and P 500 vs a portfolio containing the 10␣
↪stocks with the lowest scores

#Market index value recalculated with initial investment
MarketIndex_hist=Ticker2.history(period='3y').resample('MS').first()
MarketIndexShares=initial_investment/MarketIndex_hist.Close[0]
#History for recalculated shares of market index
MarketIndexNewHist=pd.DataFrame(MarketIndex_hist.Close*MarketIndexShares)
MarketIndexNewHist.columns=["Value"]
```

```
#creating the graphs
plt.plot(portfoliovalue(FinalPortfolio).index,␣
 ↪portfoliovalue(FinalPortfolio)['Value'], marker='o', label='Final Portfolio')
plt.plot(MarketIndexNewHist.index, MarketIndexNewHist['Value'], marker='o',␣
 ↪label='Market Index')
#labels
plt.title('Final Portfolio Value Compared to Recalculated S and P 500')
plt.xlabel('Date')
plt.ylabel('Value')
plt.xticks(rotation=50)
#legend
plt.legend(loc='best')
plt.show()
```



[27]:
```
FinalPortfolio
```

[27]:

	Ticker	Price	Shares	Value	Weight
1	GM	529.37	13.034362	6900.0	0.069
2	MRK	341.27	19.632549	6700.0	0.067
3	CVS	3014.18	2.156474	6500.0	0.065
4	LLY	261.33	24.107450	6300.0	0.063

5	MS	87.60	69.634703	6100.0	0.061
6	AMZN	370.78	15.912401	5900.0	0.059
7	ACN	3696.06	1.542183	5700.0	0.057
8	UNH	207.47	26.509857	5500.0	0.055
9	PFE	51.41	103.092784	5300.0	0.053
10	TGT	252.05	20.234081	5100.0	0.051
11	BIIB	922.77	5.310099	4900.0	0.049
12	AXP	117.07	40.146921	4700.0	0.047
13	TXN	193.47	23.259420	4500.0	0.045
14	UPS	449.47	9.566823	4300.0	0.043
15	PEP	163.42	25.088728	4100.0	0.041
16	PG	147.12	26.508972	3900.0	0.039
17	OXY	171.35	21.593230	3700.0	0.037
18	USB	240.24	14.568765	3500.0	0.035
19	NEE	97.68	33.783784	3300.0	0.033
20	CL	52.78	58.734369	3100.0	0.031

[26]:
```
Stocks=pd.DataFrame(FinalPortfolio["Ticker"])
Stocks["Shares"]=FinalPortfolio["Shares"]
Stocks
```

[26]:

	Ticker	Shares
1	GM	13.034362
2	MRK	19.632549
3	CVS	2.156474
4	LLY	24.107450
5	MS	69.634703
6	AMZN	15.912401
7	ACN	1.542183
8	UNH	26.509857
9	PFE	103.092784
10	TGT	20.234081
11	BIIB	5.310099
12	AXP	40.146921
13	TXN	23.259420
14	UPS	9.566823
15	PEP	25.088728
16	PG	26.508972
17	OXY	21.593230
18	USB	14.568765
19	NEE	33.783784
20	CL	58.734369

[ ]:
```
portfolio_saved_file = Stocks.to_csv('Stocks_Group_11.csv')
portfolio_saved_file
```

0.2 Contribution Declaration

The following team members made a meaningful contribution to this assignment:

Richard Yang, Soham Basu, William Zhang

```
[ ]:
```