BMGT 431

5/10/2023

FINAL REPORT

PORTFOLIO ANALYSIS

Rohan Pandit, Ekow Andoh, Nancy Jarman

INTRODUCTION

The use of advanced algorithms, predictive analytics and other machine learning techniques has enabled financial institutions to extract insights from large and complex data sets that were previously impossible or too tedious to analyze manually. The combination of machine learning and finance has become pivotal for financial firms to improve their ability to identify and handle risks, realize higher returns and grow steadily.

The machine learning concepts introduced in this class are both practical and broad which makes them ideal for a variety of use cases. There were many methods to use for an analysis on stocks but our group ultimately decided to use k-means clustering and time series analysis to analyze the performance of a portfolio of stocks with different weights over a ten year period. The goal of this project was to use various machine learning techniques to categorize stocks in a portfolio with k-means clustering, predict its performance in the near future with time series analysis and evaluate performance with established financial market metrics.

DATA

In this assignment we used closing stock prices for General Electric, IBM, Walmart, Simon Property Group, Morgan Stanley, Disney, Exxon, Chevron, JP Morgan, Johnson & Johnson, T-Mobile and Verizon on every trading day from 2009 to 2019. The data was extracted from a built-in Yahoo Finance API which has been the standard for stock market information for a good amount of time. We ultimately decided on this time frame because it was the end of the 2008 financial crisis and a time of steady growth and recovery. A dataset like this is ideal for forecasting and analysis as it likely contains less outliers and noisy data.

The stock data normally is formatted in an extensible time series which is not conducive for any type analysis apart from time series analysis. Thus, transforming the dataset into a dataframe was the first course of action. The original dataset had the dates as the row names for each day and for every stock so we removed it from the dataframe to finally make it well prepared for processing.

To put the daily stock data into perspective, we created a new dataset and calculated the average daily returns of each stick over the ten year period, the average yearly returns for each stock and the average yearly variance in stock returns. These calculations would give us an idea of how well the individual stocks have been doing over time and their volatility over the ten year period. Once these calculations have been applied it is important to standardize all the values before performing any kind of analysis to ensure that each value has the same impact regardless of original measurement scale.These calculations and metrics are essential to any type of investor to increase earnings and optimize their portfolios as much as possible.

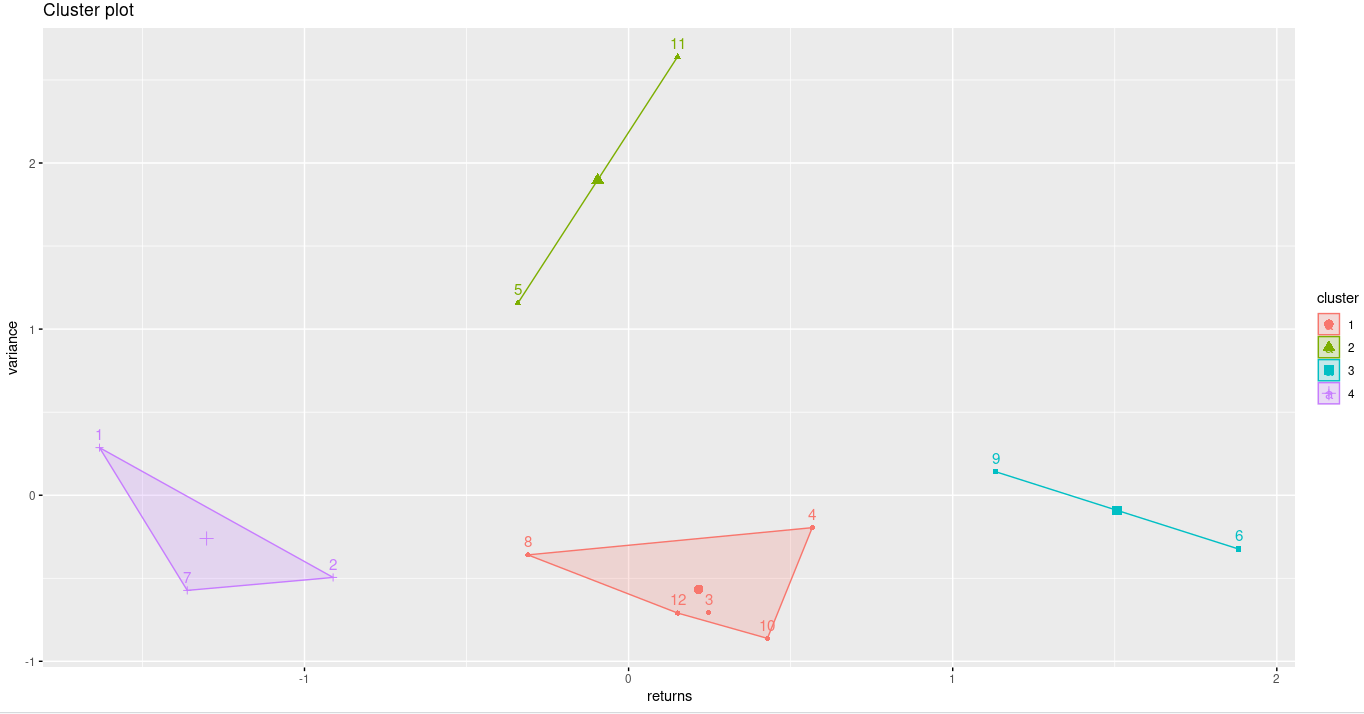
METHODS

K-Means Clustering

In order to identify the performance/ behavior of the stocks as it relates to the average yearly returns and the average yearly variance in stock returns we used a k-means clustering algorithm to group them by their properties. Our “iter.max” parameter was set to 10 which instructs the algorithm to reiterate 10 times to find the absolute best number of clusters for our dataset. With 12 rows, the number of iterations need not to be very large as it would only cause the algorithm to be computationally intensive and require a lot of RAM to run. Our “nstart” parameter was set to 5 which therefore runs the algorithm with different centroids five times to find the lowest total within-cluster sum of squares. A higher “nstart” value means a more stable clustering output and for our data five is just enough.

To run the clustering algorithm, we decided to use the silhouette method instead of the elbow method that we are all familiar with. The silhouette method is advantageous when the data points are of different sizes and density because it measures the accuracy of each cluster based on the difference or similarity of each datapoint to its own cluster compared to other clusters. Because we are looking at the annual returns and variance of each stock which are vastly different from each other this clustering method is the most conducive. Not to mention it is also much more robust compared to the elbow method for this particular dataset.

Once the data points were grouped into clusters we thought it would be useful to visualize the clusters and realize the behavior and performance of these stocks over the decade. The “fviz” package allowed us to do this with relative ease and quite neatly as well. Below is an exhibit of the clustered stocks along with a table mapping each number to a stock and the cluster it belongs to.



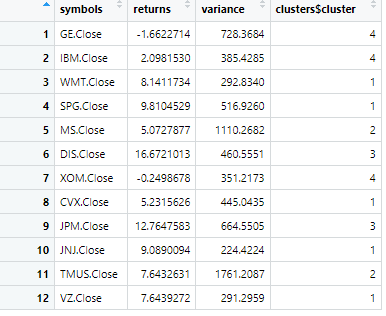


Exhibit 1: K-Means Clustering and Mapping Table for Stocks

Performance Analytics

Once the stocks have been grouped by their variance and returns we will not calculate the performance of the portfolio as a whole and possibly draw some conclusions from the results we generate. To do this we used the “Performance Analytics” package in R find some key metrics to evaluate the portfolio performance. We first assign weights to each stock (ie. how much is invested in each stock) and then import the S&P 500 stock data to use as a benchmark to compare the performance.

The first metric to calculate is beta. This metric measures the risk of the portfolio in comparison to the overall S&P 500 market. The market always has a beta of 1 so a portfolio that has a beta of 0.15 is theoretically 0.15 times riskier than the market. Our portfolio has a beta of 0.907 which means it is less risky than the overall market throughout the decade.

Another important metric to evaluate is alpha (Jensen Alpha) which calculates the excess return on an investment relative to the return on the S&P 500. It is typically the metric in reference when one says “I beat the market” and is a universally recognized metric of positive portfolio performance. The S&P has an alpha value of 0 and our alpha value was -0.46 which means we did poorly in terms of returns compared to the market.

The last influential metric to assess is the sharpe ratio which describes how much excess return one would receive for the additional risk taken (holding on to a stock). In essence it evaluates how profitable your risks have been and is therefore monitored often when a portfolio contains new or volatile stocks. The S&P 500 always has a sharpe ratio of 0 and our value was -0.145 which means we accrued losses by holding on to these stocks for the decade.

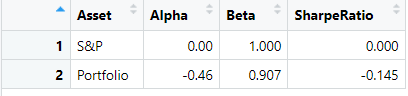


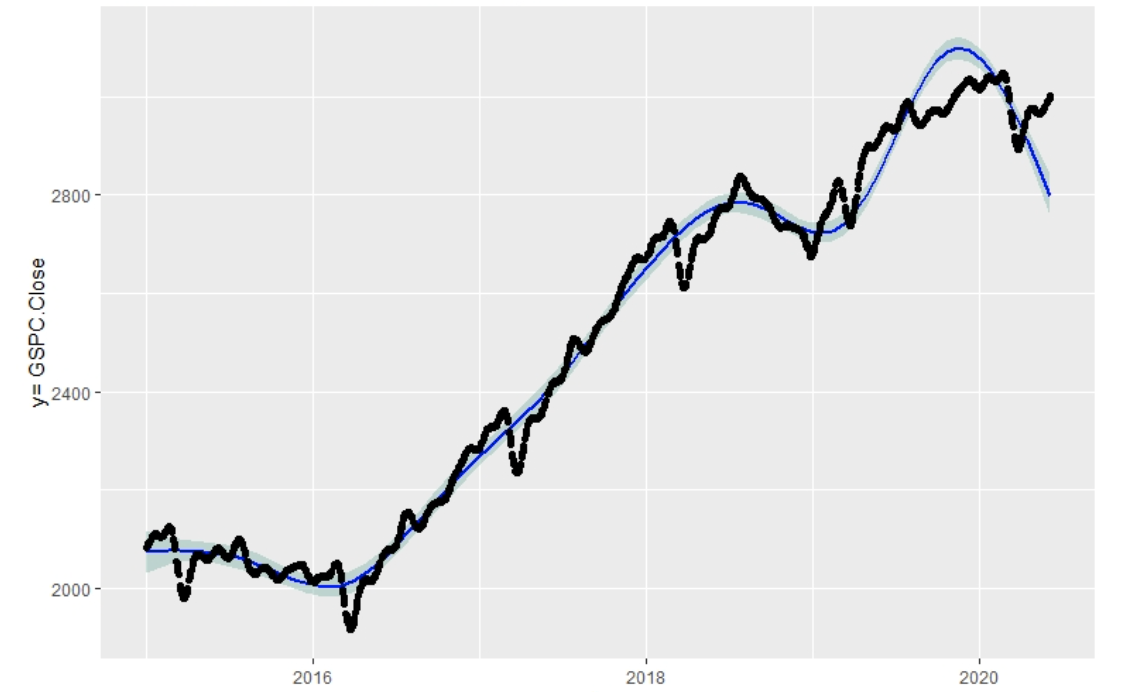
Exhibit 2: Performance analytics on portfolio compared to S&P 500

Time Series Analysis

An ARIMA model which stands for Autoregressive Integrated Moving Average was used to predict future trends in the time series data by looking at the discrepancies between values in the series. The model has three components which is Autoregression (AR) where a changing variable regresses on its old prior values, Integrated (I) represents the differencing of raw observations to allow the time series to be stationary, and Moving Average (MA) where the dependency between an observation and a residual error from a moving average is applied to lagged observations. The time series of stock prices was modeled to divide the historical translation data of the stock in a fixed interval sequence into ARIMA that processed the data to identify the changing pattern of the stock prices over time.

Parameters p,d,q were used. P denoting the number of lag observations used, d as the number of times the actual data are differenced to become stationary and q being the size of the moving average window. When using auto.arima we get non-seasonal ARIMA that doesn’t include the seasonal part of the data while in seasonal ARIMA it is included. Based on the findings we can see that the seasonal part of our data is not significant.

Following observations from the time series data was found: we could observe that there were no huge variations in the opening-closing price and the high-low prices. We can also see a huge dip in the price in 2019 due to the pandemic as well as an overall increase in the stock price from 2017-2018. In 2018 we can see the stock price going progressively down. Closing price was also considered as a useful marker to assess changes in stock price over time.

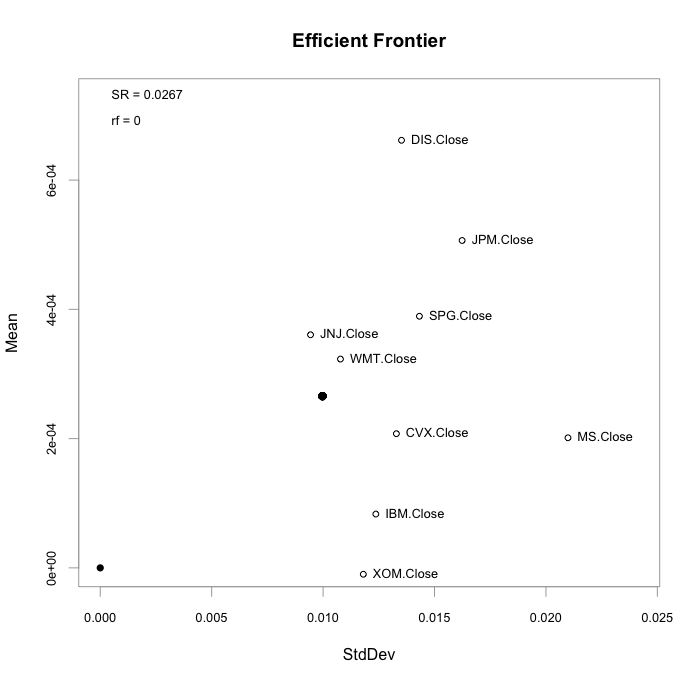


OPTIMIZATION

Introduced by Nobel Laureate Harry Markowitz, the efficient frontier chart is a mainstay in modern portfolio theory, and is a common metric used by many investors to try and optimize a selected portfolio. The efficient frontier is a way to quantify the age-old notion of risk vs. return. Ideally, everyone would want a portfolio that has extremely high returns without any risk. However, this is not realistic. The efficient frontier was created to create an optimal set of portfolios that give you the lowest level of risk for a selected value of expected return. On an Efficient Frontier Graph, the risk is shown on the x-axis, while the expected returns are shown on the y-axis. After seeing from our previous portfolio analysis that the portfolio we selected would not perform ideally, we wanted to try and optimize the portfolio.

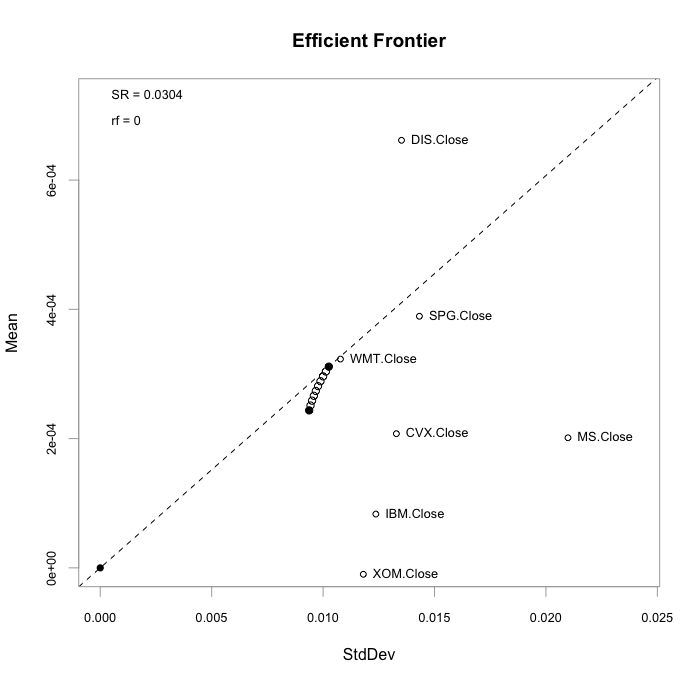
For reference, the closer a stock is to the top left of the graph, the better it is (since it has high reward and low risk. Anything to the top right has a high risk and reward, and the bottom right has high risk for low reward. From looking at the individual stocks on the efficient frontier graph, some conclusions can be drawn. It can be seen that most of the stocks in our portfolio are in the bottom-right quadrant, meaning that they are high risk and low reward. Some particularly bad stocks are Exxon, which has almost no return for a high amount of risk, IBM, which is just slightly better than Exxon, and Morgan Stanley, which has an extremely high risk for a moderate reward. Stocks that stand out as having a better risk-to-reward ratio are Johnson & Johnson, JP Morgan, and in particular, Disney, which has a much higher reward at its level of risk compared to all the other stocks in the graph.

Exhibit 3: The stocks represented on the risk-reward graph



Now, we will try to optimize the portfolio while keeping all the stocks in the portfolio by using the optimization methods in the PortfolioAnalytics package of R.

Exhibit 4: The efficient frontier generated using all stocks in the portfolio.

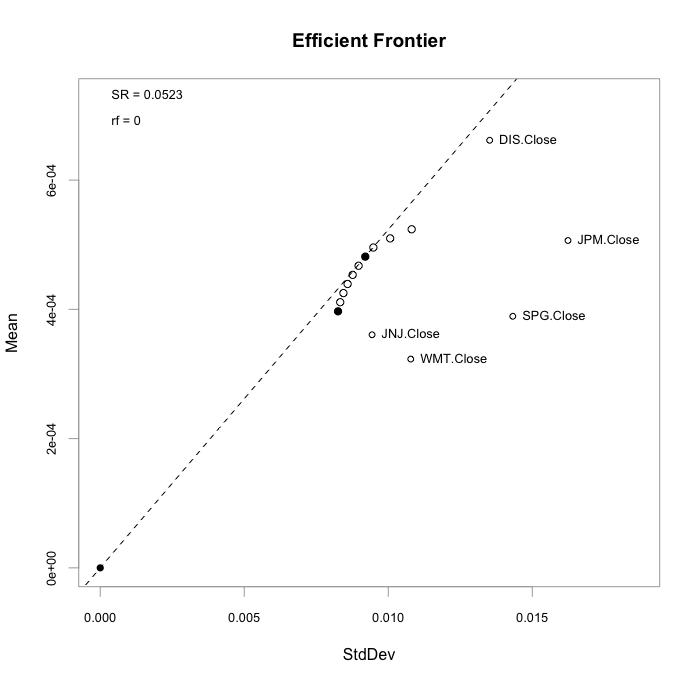


Note: Due to an error in the code that could not be resolved, JNJ and JPM were excluded from the optimization and efficient frontier calculation

The weights chosen by the portfolio optimization method were 0.3 for Disney, and 0.1 for all the other stocks included in the portfolio. This makes sense, as Disney had a much better risk-reward ratio compared to every other stock in the portfolio, so it would be weighted higher when optimizing the portfolio for highest expected returns. For this optimized portfolio, the R output tells us that the mean(the measure used to represent expected return) would be equal to 0.0003113, and the standard deviation (the measure used to represent the risk factor) would be equal to 0.01026. The efficient frontier (the optimal set of portfolios that give you the lowest level of risk for a selected value of expected return) for this optimized portfolio can be seen on the graph as the curved line of points near the middle of the graph. The efficient frontier is near the values of the mean and standard deviation given by the portfolio optimization, which makes sense.

Although this is the best optimized version of the portfolio including all stocks, this portfolio includes many stocks that are in the bottom-right quadrant of the graph, and these stocks have much more risk for less reward. A way to further optimize this portfolio would be to cut out the stocks with low reward and high risk, and build an optimized portfolio from only the other stocks. To do this, a new portfolio was created including only Walmart, Simon Property Group, Disney, Johnson & Johnson, and JP Morgan. After optimizing the portfolio, the weights given for each stock were 0.1 for Walmart and Simon Property Group, 0.1825 for Johnson & Johnson, 0.2175 for JP Morgan, and finally, 0.4 for Disney. These weights are what we would expect: Walmart and Simon Property Group are the lower performing stocks in this portfolio, Johnson & Johnson and JP Morgan perform better and have higher reward, and finally, Disney is the clear best performer here, so it would have the highest weight. For this new portfolio, the expected returns/mean value was given to be 0.0005119, and the risk factor/standard deviation value was 0.01017. This risk value is about the same as the risk value for the previous portfolio, but the expected returns are higher. This is to be expected, as the stocks with low returns and high risk have been cut out. The new efficient frontier graph is shown below.

Exhibit 5: The new efficient frontier using only selected stocks



LIMITATIONS AND IMPROVEMENTS

In terms of the optimization of the portfolio, there were some limitations and improvements that could have been made. The biggest limitation that we faced was that the specific portfolio that we chose had many stocks in the bottom right quadrant of the efficient frontier chart (low reward, high risk), which made it difficult to find a portfolio that would lead to good return on investment for an investor.

In terms of improvements that could have been made in the optimization step, when creating our initial efficient frontier chart including all the stocks, including the Johnson & Johnson and JP Morgan stocks lead to a weird error where the efficient frontier portfolios would not show up on the graph, and so those two stocks had to be excluded in order to get output that could have been analyzed. If this bug could have been figured out, we would have had a more accurate efficient frontier chart, leading to more accurate analysis and next steps to be taken to improve our portfolio further.

RESULTS AND CONCLUSIONS

After analyzing the portfolio using k-means, performance analytics, and time-series, we were able to use techniques such as the efficient frontier and ROI analysis in order to create a new portfolio that would provide us with higher returns for the same amount of risk. The key to this was analyzing the risk-to-reward ratios of different stocks, and narrowing our portfolio to only include the stocks that weren’t dead weight (low reward, high risk). In the end, we were able to almost double the expected returns metric at the same risk factor for our previous optimized portfolio.

Our portfolio performed poorly compared to industry standard benchmarks from 2009 to 2019.

We believe it would be beneficial to look into other stocks to invest in that generate larger steady returns over a long period of time. However, if one desires to continue to hold the current stocks it would be advisable to invest a larger share of investment capital into Walmart, Disney, Johnson & Johnson, JP Morgan and Simon Property Group as they are most likely to succeed in the immediate and distant future