*Demonstrates the Markowitz Optimal Portfolio methodology, uses quadratic programming, Excel’s solver, and for loops in R.*

**Project**

**6**

P6

ALY6050 Intro to Enterprise Analytics

Project 6 – The Optimal Portfolio

**PREPERATION:**

By: John DiSessa

For: Professor Behboudi

On: April 2nd, 2022

Introduction

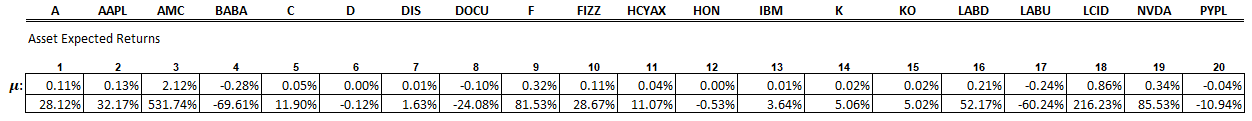
We were tasked with developing different optimal portfolios depending on various return levels. We used the Markowitz Optimal Portfolio methodology in both Excel and R to determine the investment weights of our 20 potential assets. We were given 252 market days of stock prices (December 2020 – December 2021) for each of the 20 assets.

Analysis

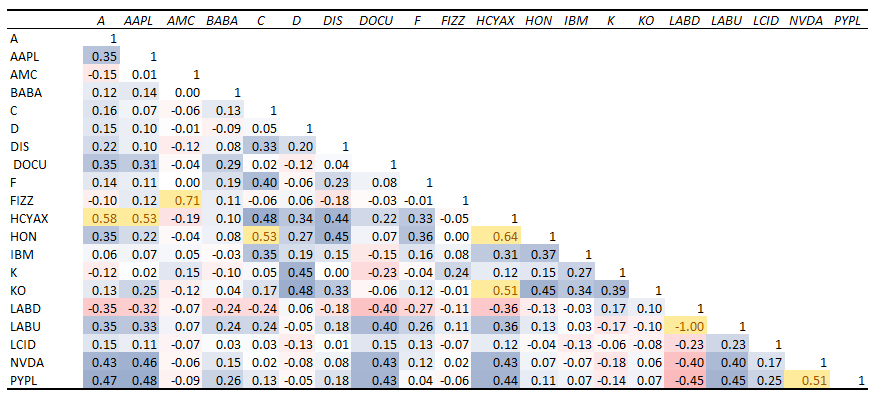
**Part One**

The main goal for each calculation was to minimize portfolio risk as long as the certain return levels were met. If two different solutions led to the same return, we had to be sure our model chose the one with the least amount of risk. Since we had 20 assets to choose from, we were able to ensure diversification in each portfolio. The function we optimized to minimize risk and meet the desired return levels was . Essentially, the formula compares the average daily return for each asset over the year with the variance of each asset’s price over the year. We were also subject to some constraints, such as meeting whichever baseline return we selected, our selected weights for the various assets must total 100%, and selling and shorting were now allowed so all weights must be at least 0%.

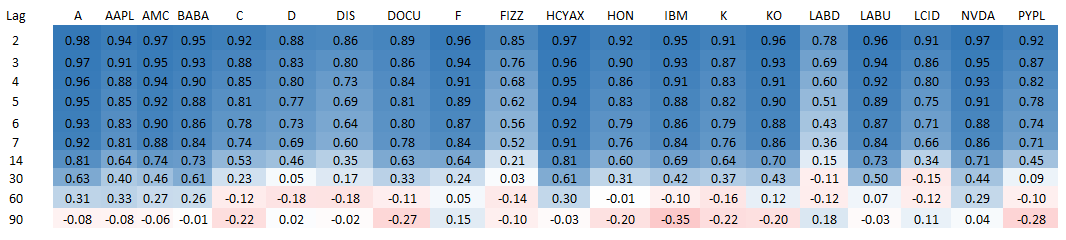
Our first step was to calculate the daily return rates for each asset each day of our time frame. With 252 days we had 251 daily return rates ((asset *j* – asset *j-1*) / asset *j-1*) for each stock. The first row shows the average daily return for each stock in our dataset. The second row shows the yearly return if every day, the stock increased of decreased by its average.



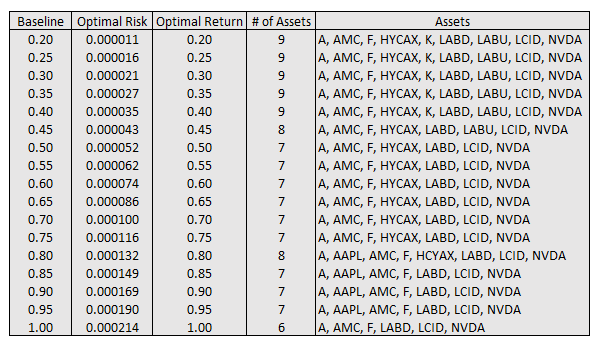
Next we created a covariance matrix to see both the variances of returns in individual assets as well as any significant covariance between any two assets. For the variances in returns of each asset, we decided to check if any asset’s variation was more than two standard deviations above the mean variation for all 20 assets. The only asset with a variance that high was AMC (also more than three standard deviations above the mean). Since it’s yearly return was approximately 531%, we expected to see a high level of variance here. This can be explained by the historic failed short on the stock, causing it to spike in May and June of 2021 due to a large group of Reddit users. We expected this variance to be so far above the mean because the stock market had never seen an event like that happen so we can consider that to be an anomoly. We did notice some high and low covariances so we decided to create a correlation matrix to identify which pairs of stocks were highly correlated, how strong their correlations were, and if they were postively or negatively correlated. Since we have 20 assets and the correlation matrix is congested, we color-coded each correlation. Positive correlations are blue and negative correlations are red. However, we were only concerned with correlations greater than .5 or less than -.5. Those highly correlated pairs are marked in bright yellow. You will first notice that HCYAX was highly correlated with A, AAPL, HON, and K. HCYAX is a fund, not an individual company stock, so it likely contains all four stocks in its collection. In fact, AAPL is indeed its largest holding according to Direxion’s website. Direxion also created the LABD and LABU leveraged ETFs. These funds seek to return -300% or 300% each day. These funds contain the exact same portfolio except one is used to buy and one is used to short. They are almost perfectly negatively correlated. As one increases in price the other decreases by the same amount.



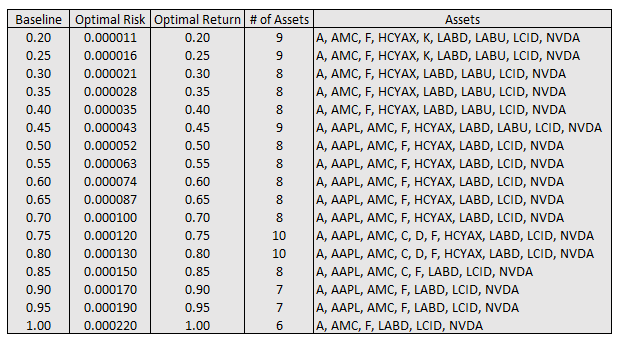
We quickly took a look at the autocorrelation to see if there was anything unusual, however we got results that were to be expected. Strong positive autocorrelation, as indicated in blue, means that if the stock price increased in one day it was very likely to increase the next day (or for any time gap chosen). We looked at the stock prices within a few, 7, 14, 30, 60, and 90 days. Naturally, stock prices are heavily influenced by the previous days price since momentum for stocks exist in the short term as investors flock to the high returning stocks. However, as the timeframe increases, autocorrelation decreases because the short-term momentum goes away and investors are left to analyze the stock price to the company’s value. The increased time frame also allows investors to better understand the influences of company decision making and their impacts.



Now that we identified any potential risks in our optimal portfolio, we used both Excel Solver and R to determine the optimal portfolios in order to achieve baseline annual returns between 20% and 100%. If the solution returned a value for an asset that was less than .5% of the portfolio, we excluded that asset from our solution since the fees and administrative work involved with selecting assets with such a small percentage of the allocation would not be worth it. The first table shows the results from Excel Solver. Since our goal was to minimize risk while meeting the baseline returns, each optimal portfolio did not achieve a return higher than the baseline. With so many assets to choose from, there was an unlimited number of potential outcomes. We were not surprised to see the returns exactly meet the required baselines. Our optimal risks were the average daily variances in returns. In both our Excel Solver and R solutions, the annual risk to achieve 20% returns was about .28% and the annual risk to achieve 100% returns was about 5.4%. When looking at the assets selected in each scenario, you will notice that AMC is chose in every case. However, due to its unprecedented return and volatility, we would not expect that asset to replicate a performance again so we would recommend excluding AMC from all portfolios. Most of the portfolios also chose both LABD and LABU which, as previously discussed, are almost perfectly correlated. Realistically there is no point to holding both assets so we would recommend only selecting LABD to mitigate risk. LABD returned 52% while LABU lost 60% so we would clearly choose LABD. Next, you’ll notice HCYAX is in many portfolios as well. Since this is a fund highly correlated with A, AAPL, HON, and K, we would not recommend holding these assets and HCYAX at the same time. If you are looking for a higher return, you should select A and AAPL. If you are looking for a lower risk, select HCYAX instead. For both aggressive and conservative investors, we would recommend holding A, F, LABD, LCID, and NVDA since these were chosen in all scenarios given their high returns and low variance. Conservative investors should considering adding K to their portfolio since it returned 5% with almost no variance. Conservative investors should aim for about 23% HCYAX and 31% LABD since those are funds that are divsersified by holding many individual stocks. The other 45% of the portfolio should be spread out among the other assets. Aggressive investors should consider adding AAPL to their portfolio since it returned 32% with a moderate variance. Even though many of the names are similar to the conservative portfolios, the allocation percentages are very different. NVDA should be the highest allocated asset (about 30%) since it was the best value asset in the portfolio due to it having the highest return to variance ratio. F should be the next asset chosen with about 26% of the allocation. Both A and and LABD should each be about 15% or 16% too. Even though LCID had the highest return of all assets excluding AMC, it’s variance was so high that we can only justify allocating about 8% of the portfolio to it.



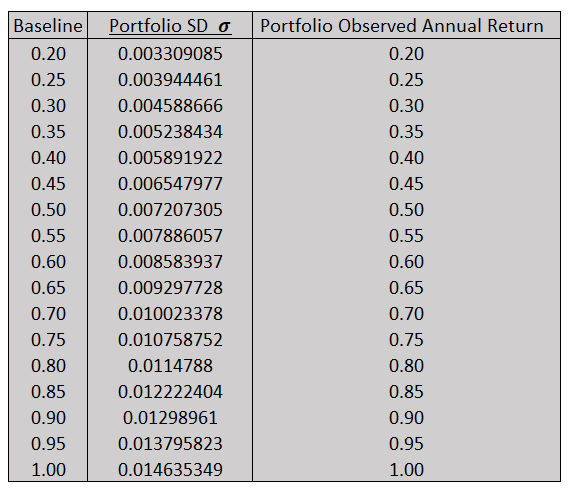
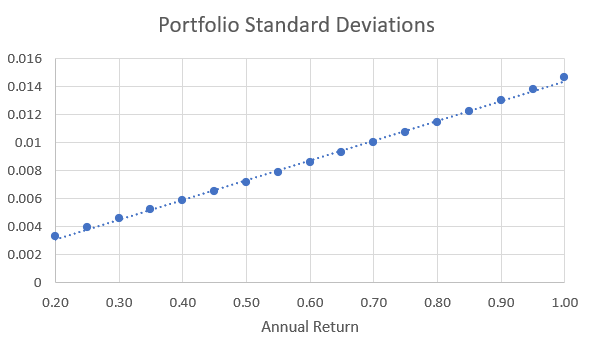
The solution from R yields almost identical results for the top 3 conservative portfolio options and the top 3 aggressive options. However, all of the balanced options contained AAPL which is a great choice since it retuned 32% with moderate variance. We were surprised that AAPL was not selected from the Excel Solver solution but would still recommend this option to moderate investors. The 75% and 80% returns included C and D assets which was strange considering no other portfolio in our research included them. We would not recommend investing in these assets. The variances at each baseline were almost identical to the Excel Solver solutions so there were only minor differences in portfolio allocation percentages.



**Part Two**

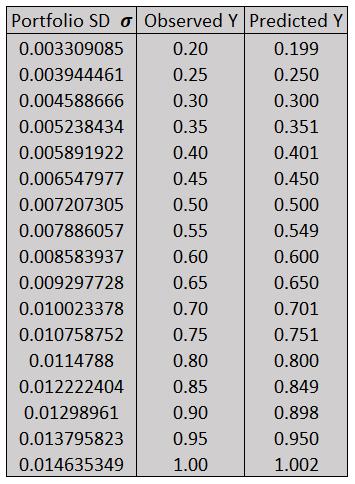
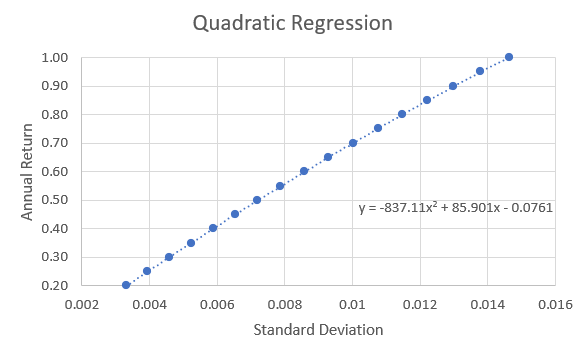
In this part, we looked at the standard deviations and plotted them against our annual returns. As you can see in the chart below, we likely have a linear relationship between annual return and portfolio standard deviation. We calculated the slope of our line of best fit to be .0141 and our y intercept to be .0003 so we can use the following equation to predict out standard deviation if given an expected annual return: . We then calculated our R2 value to see how well the variation in standard deviation can be explained by our annual return. Our R2 value of .9987 means that 99.87% of the variation in standard deviation can be captured by our model by changing the annual return value. We also looked into a similar relationship by treating our standard deviation as the X variable and using it to predict an annual return. The formulation for that relationship was

. That model also had an R2 of .9987.

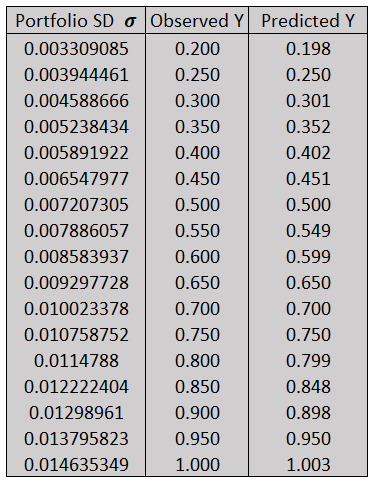
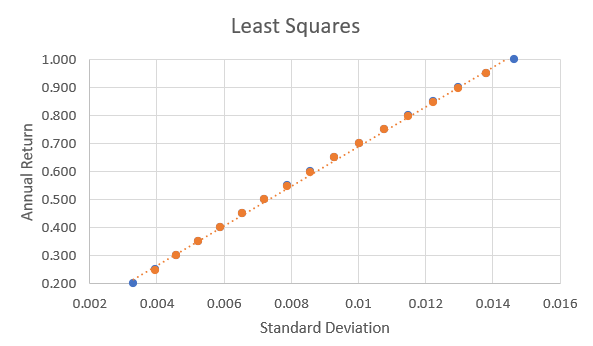
 

We then looked into quadratic regression and the least-squares methods to see if we could get a more accurate fit. For these analyses, we considered standard deviation as the X variable and annual return as the predicted Y variable.

For quadratic regression, we used Excel to calculate the A, B, and C values for our best fit equation: . This gave us an R2 value of .9987 which was identical to our linear model. Our SSE (Sum of Squares Error) was .000014 which we can use to compare this model to our Least Squares model and see which prediction model best fits the actual data. Our quadratic regression equation was so accurate at predicting annual return that the plotted points for both observed and predicted return overlapped.

We then used the Least Squares method to see if it was more or less accurate than the quadratic regression method. Instead of using , we used . This time we calculated R2 ourselves using the formula . After using the formula to get predicted Y values, we used Excel’s Solver to minimize the SSE value. Solver gave us the values of A = -2.468, B = 20.886, and C = -.013. We calculated the R2 value to be .860 which was far less than our .9987 values from other models. Also, our SSE value of .000030 is larger than our SSE value of .000014 from the quadratic regression equation so we can say that the least squares method was not as accurate for us at predicting portfolio standard deviations. Our A, B, and C values are simply mathematical coefficients used for the quadratic formula to complete the square. They have no relevant context to our asset prices, variances, or standard deviations, and cannot be compared between the quadratic regression and least squares models.

Summary

We used Excel and R to utilize the Markowitz Optimal portfolio methodology and calculate various optimal portfolios. Out of the twenty potential assets, we would recommend conservative investors have just over half of their portfolio in HACYX and LABD with the other half containing A, F, LCID, and NVDA. For aggressive investors, we would recommend having about a third of their portfolios holding NVDA, a quarter in F, and the rest made up of A, LABD, and LCID. There were some pairs of assets that were strongly correlated with each other so we would not recommend holding both HCYAX and A, AAPL, HON, or K. LABD and LABU are inverse ETFs of each other with symmetrically opposite goals so it is not wise to hold both of these assets either. If one was interested in predicting annual returns given portfolio standard deviations or vice versa, we would recommend using our linear models due to their almost equal accuracy but would be simpler to use.

Summary

*Bloomberg.com*, Bloomberg, https://www.bloomberg.com/news/articles/2021-05-27/amc-s-four-day-surge-slaps-short-sellers-with-1-3-billion-loss.

“HCYAX HCYIX.” *Direxion*, https://www.direxion.com/product/hilton-tactical-income-fund.

“How to Interpret R-Squared in Regression Analysis?” *How to Interpret R-Squared in Regression Analysis?*, https://www.knowledgehut.com/blog/data-science/interpret-r-squared-and-goodness-fit-regression-analysis.

“LABD LABU.” *Direxion*, https://www.direxion.com/product/daily-sp-biotech-bull-bear-3x-etfs.

Smith, Tim. “What Is Autocorrelation?” *Investopedia*, Investopedia, 19 Mar. 2022, https://www.investopedia.com/terms/a/autocorrelation.asp.