**OPTIMIZATION PROJECT REPORT 2**

**INTEGER PROGRAMMING**

**GROUP 10 – Disha Gandhi, Fernando Chapa, Krish Engineer, Kyle Tobia**

**INTRODUCTION**

An index fund is a type of mutual fund or exchange-traded fund (ETF) with a portfolio constructed to match or track the components of a financial market index, such as the NASDAQ-100, Standard & Poor’s 500 Index (S&P 500) or other similar indexes. Constructing an index fund that tracks a specific broad market index can be done simply by purchasing all the stocks in the index, with the same weights as in the index. However, this approach is impractical (many small positions) and expensive. An index fund with q stocks, where q is substantially lower than the size of the target population (n) is much more ideal. In this project we try to create a smaller portfolio to track the NASDAQ-100 index.

**THE PROJECT**

We will use 2 methods to select stocks from the NASDAQ-100 index which are as follows:

1. Selecting stocks and calculating weights
2. Calculating weights only

**METHOD 1: SELECTING STOCKS AND CALCULATING WEIGHTS**

For this approach, we first determined the correlation between each stock in the index and the fund to increase similarity between the index stock and the fund that represents it. For this the objective function and constraints are:

Text, letter

Description automatically generated

A snapshot of this correlation can be seen below:

Table

Description automatically generated

Our next step is to calculate the weighted return of each chosen stock to match the returns of the NASDAQ-100 for which the objective function and constraints are as follows:

Schematic

Description automatically generated

Text, letter

Description automatically generated

The return of each stock, , is calculated using the pandas pct\_change() function which computes the percentage change from the previous day. We calculated the return for both 2019 and 2020, as shown in the tables below:

Table

Description automatically generated

**2019 Stock Return**

Table

Description automatically generated**2020 Stock Return**

1. **Select 5 Stocks**

We started off by selecting 5 stocks and observed that the difference between the total return of these stocks and the return of index ‘NDX’ was around 0.789 in 2019 whereas the difference was 0.870 in 2020.

1. **Increase the number of stocks in the portfolio**

We increased the selection to 10, 20, 30,...,100 stocks and compared the performance of the various portfolios created. We used the absolute difference between the return of the portfolio and the return of the index as our comparison metric. The smaller the difference between the 2 returns, the better is the tracking of the portfolio with respect to the index.

Generally, the increase in the number of stocks in the portfolio results in a decrease in the difference between the returns on the portfolio with respect to the index, which is correct since we are replicating the index with higher accuracy. However, there are certain exceptions in both years wherein the difference increases. For instance, in 2019 the performance on choosing 60 stocks was slightly worse than choosing 50 stocks, whereas in 2020 the performance reduces after the 30 stock selection and there’s a large spike when we go from selecting 50 stocks to 60 stocks. Hence, we can infer that the optimal number of stocks we should be selecting for this model is 30 to optimally replicate the NASDAQ.

Chart, line chart

Description automatically generated

**Portfolio performance through weight calculation after stock selection**

**METHOD 2: CALCULATING WEIGHTS ONLY**

With this approach, the stock selection integer programming problem was disregarded. Instead, a mixed integer programming problem where the number of non-zero weights is constrained to be an integer is used. To do this, we also introduced a set of binary variables and added some constraints that force . The objective function and constraints for this method are as follows:

Text

Description automatically generated

The smallest value for M here would be 1 since the largest value that can have is 1 given that the sum of the weights must be 1. Since this method had 2T + n + 2 constraints and 2n + T decision variables we limited its runtime to 10 hours (where, T = Time period and n = number of stocks).

**Select 5, 10, 20,…..,100 stocks in the portfolio**

We used a similar approach as the 1st Method wherein we selected 5, 10, 20,……., 100 stocks for the second method and used the same absolute difference metric. Thus, the smaller the value of the metric the close is the tracking of the portfolio to the index. Through the graph below we can observe the expected i.e. as the number of stocks in the portfolio basket increases, the performance also improves. This relation holds true for both 2019 and 2020. Diminishing returns start to settle in after 40 stocks which is larger compared to the method one where diminishing returns settled in after 30 stocks, however the performance stabilizes as well which indicates that the performance of the 40 stock basket is quite similar to the portfolio with 100 stocks.

Chart, line chart

Description automatically generated

**Portfolio performance through calculating weights only**

**INSIGHTS AND RESULTS**

The graph below shows us the daily returns of the NASDAQ-100 for 2019 and 2020. As we can see the daily returns for 2019 are more stable for the year 2019 as compared to the year 2020 which supports are findings through the 2 methods that the portfolio in 2019 outperforms the portfolio in 2020 due to more stability.

Chart, line chart

Description automatically generated

**Index returns for 2019 and 2020**

The 2 heatmaps below show the similarity in the returns of stocks on the NASDAQ-100 index for the years 2019 and 2020. As shown, the returns of stocks for 2019 are highly correlated as compared to the correlation in returns of stocks for 2020. This can be explained by the higher volatility in the NASDAQ-100 index as seen above.

Chart

Description automatically generated

**Heatmap to show correlation/similarity in returns of stocks for 2019**

Chart

Description automatically generated

**Heatmap to show correlation/similarity in returns of stocks for 2020**

**Comparison Between The Two Methods**

Table

Description automatically generated Table

Description automatically generated

**Performance Without Stock Selection Performance With Stock Selection**

Method 1: In 2019, selecting 60 equities performed worse than selecting 50. After selecting 30 companies in 2020, the performance of the replicating portfolio reduces as the difference remains constant and even rises between m = 50 and m = 70. Therefore, 30 stocks would be the ideal number of stocks to select for replicating the NASDAQ-100 index efficiently.

Method 2: Using this method, as expected the performance keeps improving with an increasing number of stocks. In 2020, there is a single exception where the performance of choosing 60 stocks was worse than choosing only 50 stocks. In 2019, there are diminishing returns after choosing 40 stocks as the loss stays stagnant after. As a result, an optimal number of stocks would be to choose 40 stocks to model the NASDAQ-100 index.

**CONCLUSION**

It is clear from the above plot that Method 2 works better on the data giving us a better and more stable performance as the number of stocks in the basket increase. Although the performance is similar across both the methods, there is a lot more variance with Method 1. Since Method 2 is the more stable and better performing method, our recommendation is to choose 40 stocks to replicate the NASDAQ-100 index and model the optimization to directly calculate the weights without selecting the stocks (Method 2). We can improve our model by calculating the cost of adding more stocks to our portfolio and modelling the cost into our optimization constraints.