Optimization 1 | Project 2 | Integer Programming

Group 21 - Shreyansh Agrawal (sa55742), Ankit Muthiyan (am223225),
Parthiv Borgohain (pb25347), Denise Neuman (dn9264)

1. Objective

A market index is composed of a certain number of stocks with different weights. Market indexes are used to evaluate market performance representative of a broad market population. Creating a fund like the market index as your portfolio can be done by purchasing all the stocks in the index fund with their weights the same as the index. Though it may appear easy to create a portfolio like an index fund, it usually is infeasible in the real world.

The market index's performance can also be imitated by creating a portfolio of ‘m’ stocks within a reasonable margin of a performance difference with the index. The goal of this project is the same - to create a portfolio of ‘m’ stocks, an index fund, that can track the NASDAQ-100 index by identifying those ‘m’ stocks and optimizing the difference in portfolio performance.

To create an index fund, we need to decide the number of such stocks, identify those stocks and come up with weighted allocations that can resonate with the NASDAQ-100 index.

There are two questions to answer, how many stocks and which stocks? So initially, we plan to devise an optimization problem to find the 5 best stocks and their weights to create an index that has the lowest error (exact metric defined in the methodology) with the NASDAQ-100. Then we created a portfolio by varying the number of stocks in the index to find a point of diminishing return. The optimization will be done on data for the year 2019 and the performance will be evaluated on 2020’s stock returns data.

2. Methodology

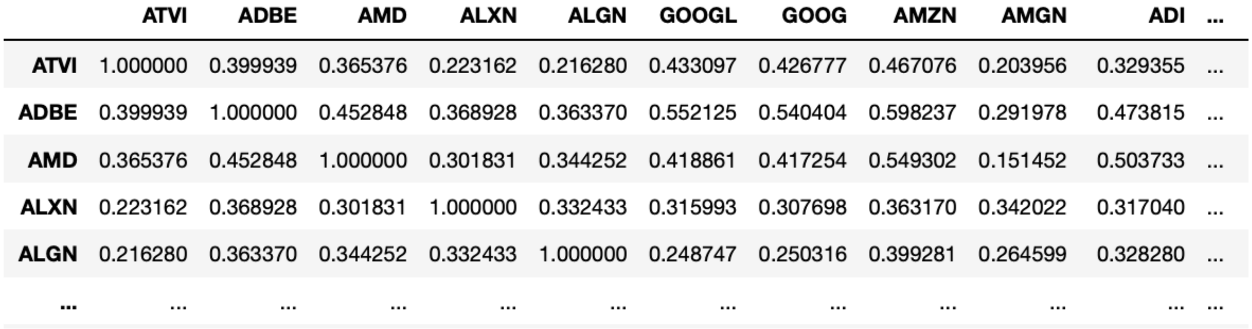
There are two things we need to decide for a given value of ‘m’ (number of stocks in our index)

1. Identify the stocks that best represent the value of all stocks in the index.
2. Calculate the weightage for all ‘m’ stocks.

For each of these steps, an optimization model is required.

2.1 Identifying Stocks

To identify the stocks that best represent the value of all stocks in the index, we first compute the daily returns for each stock and use these returns to generate a correlation matrix between all stocks in the index as shown in Figure 1.



	ATVI	ADBE	AMD	ALXN	ALGN	GOOGL	GOOG	AMZN	AMGN	ADI	...
ATVI	1.000000	0.399939	0.365376	0.223162	0.216280	0.433097	0.426777	0.467076	0.203956	0.329355	...
ADBE	0.399939	1.000000	0.452848	0.368928	0.363370	0.552125	0.540404	0.598237	0.291978	0.473815	...
AMD	0.365376	0.452848	1.000000	0.301831	0.344252	0.418861	0.417254	0.549302	0.151452	0.503733	...
ALXN	0.223162	0.368928	0.301831	1.000000	0.332433	0.315993	0.307698	0.363170	0.342022	0.317040	...
ALGN	0.216280	0.363370	0.344252	0.332433	1.000000	0.248747	0.250316	0.399281	0.264599	0.328280	...
...	...	...	...	...	...	...	...	...	...	...	...

Figure 2.1. Part of correlation matrix for returns table for year 2019

For this optimization problem, the objective is to maximize the correlation that exists between our m-stock portfolio and the whole index. To achieve this, we consider 100 Yj variables and 100*100 = 10,000 Xij variables.

- The Yj binary variables represent whether or not a stock is present in our m-stock fund.
- The Xij binary variables represent whether or not a stock j is the most similar to stock i.

In addition, some constraints must be met:

- The first constraint $\sum_{j=1}^{n} \quad y_j = m$ indicates the number of stocks that must be present in the m-stock fund.
- The second constraint $\sum_{j=1}^{n} \quad x_{ij} = 1 \; for \; i = 1,2,...,n$ states that each stock i can only be “most similar” to one stock j.
- The third constraint $x_{ij} \leq y_j \; for \; i,j = 1,2,...,n$ indicates that the similarity between stock i and j can only occur if stock j is selected to be part of the m-stock fund.
- The last constraint denotes that x_{ij} and y_j can only be either 0 or 1. This constraint is not incorporated into the constraint matrix, instead, the variable type is set as ‘Binary’ in Gurobi.

2.2 Calculating weights

After component stocks that maximize the similarity between the fund and the index have been chosen, the weights of these stocks also need to be determined to complete the fund set up. Therefore, the objective of the weights selection model is to determine the weight for each of the component stocks such that discrepancies between the index return and fund return over time are minimized. In other words, with q representing the index’s return on a given day “t”, “w_i” representing the weight of a selected stock “i”, and “r_{it}“ representing the return of stock “i” on a given day “t”, the goal is to minimize the following:

$$\min_{w} \sum_{t=1}^{T} \left| q_t - \sum_{i=1}^{m} w_i r_{it} \right|$$

While using the absolute discrepancies between index and fund returns prevent the negative and positive discrepancies from canceling each other out, it causes additional issues as there is no direct way to represent absolute values within the objective function. One way to cope with this issue is to break q up into two parts by defining an arbitrary decision variable

$$\left| q_t - \sum_{i=1}^{m} w_i r_{it} \right|$$

up into two parts by defining an arbitrary decision variable yi for each selected stock “i” such that:

$$(1)\ y_i \geq q_t - \sum_{i=1}^{m} w_i r_{it}$$

$$(2)\ y_i \geq -\left(q_t - \sum_{i=1}^{m} w_i r_{it}\right)$$

With the above objectives, a model that optimizes weights for selected stocks can be created with formulations stated below:

Decision Variables:

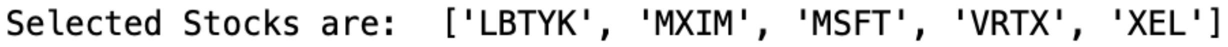
1. w_i : weights of stock “i” selected for portfolio
2. y_i : max value of the possible difference between index return at t and weighted return of selected indexes

Constraints:

1. Sum of weights w_i for m stocks in the fund index equals 1
2. The difference between the index returns and the weighted return of the stocks should be less than y_i

3. Best 5 Stocks

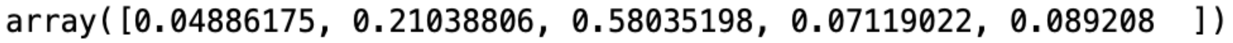
After running the model with m = 5. It was found that the five most significant stocks of the index are Liberty Global PLC (multinational telecommunications company), Maxim Integrated (analog and mixed-signal integrated circuits company), Microsoft (multinational technology corporation), Vertex Pharmaceuticals Incorporated (a biopharmaceutical company), Xcel Energy Inc (utility holding company) as shown in Figure 3.1.



```
Selected Stocks are:  ['LBTYK', 'MXIM', 'MSFT', 'VRTX', 'XEL']
```

Figure 3.1. Selected stocks outputted from the stocks identification optimization model

Figures 3.2 and 3.3 depict the weights obtained for each of these stocks as a result of the portfolio weights optimization model. As observed, Microsoft has the highest weight, followed by Maxim Integrated.



```
array([0.04886175, 0.21038806, 0.58035198, 0.07119022, 0.089208  ])
```

Figure 3.2. Weights obtained for each of the 5 stocks

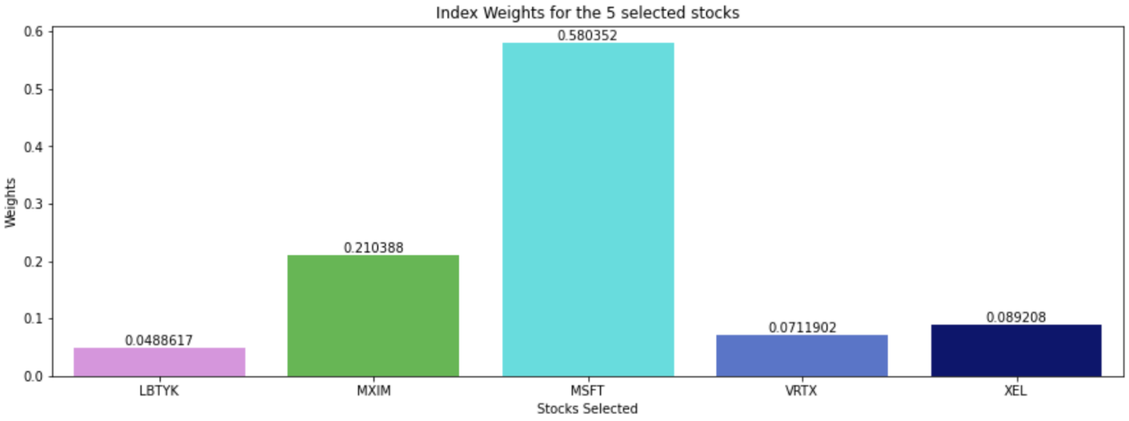


Figure 3.3. Weights obtained for each of the 5 stocks

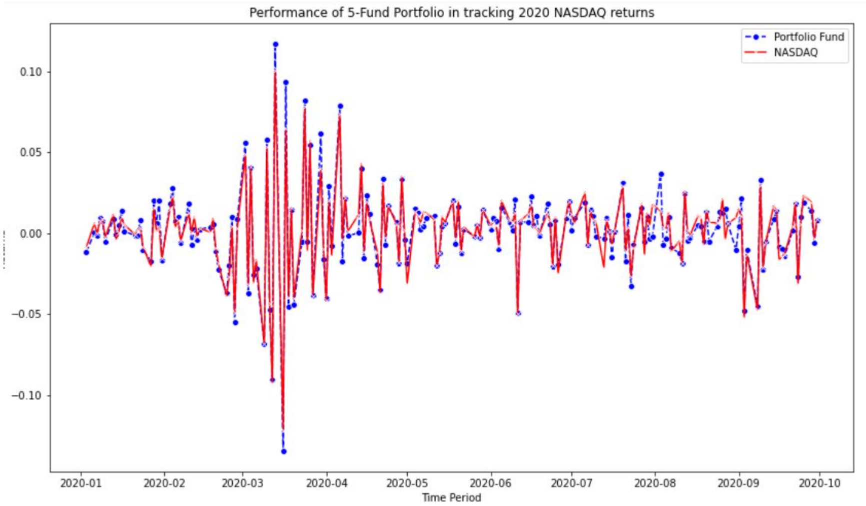
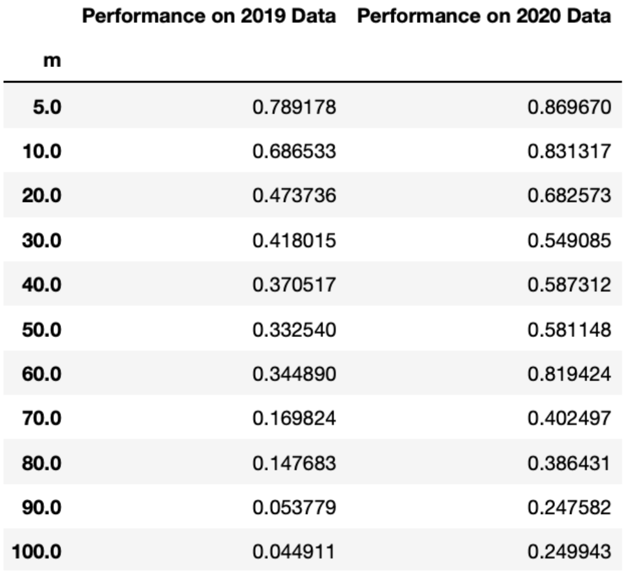


Figure 3.4. Daily Returns of 5 Stock Portfolio vs NASDAQ Index for 2020

As we can see in Fig 3.4, the 5 Stock Portfolio does a fairly reasonable job of tracking the daily returns of the NASDAQ index for the year 2020.

4. Optimal number of stocks

To get the optimal number of stocks, many models were run testing different numbers of stocks to include in the new fund. By looping over many values for m (the number of selected stocks) it was possible to obtain the results of many models. Figure 4.1 shows the absolute deviation comparison for various values of m.



m	Performance on 2019 Data	Performance on 2020 Data
5.0	0.789178	0.869670
10.0	0.686533	0.831317
20.0	0.473736	0.682573
30.0	0.418015	0.549085
40.0	0.370517	0.587312
50.0	0.332540	0.581148
60.0	0.344890	0.819424
70.0	0.169824	0.402497
80.0	0.147683	0.386431
90.0	0.053779	0.247582
100.0	0.044911	0.249943

Figure 4.1. Absolute Deviation of Portfolio Fund with NASDAQ for 2019 and 2020 for various values of m

As observed in Figure 4.2, when looking at the performance on the 2020 data, there is a 49.23% (0.549085 to 0.819424) increase in deviation from 30 to 60 stocks, making it better to use the lower number of selected stocks in this case. Specifically, a 41% (0.581148 to 0.819424) spike in deviation occurs from 50 to 60 stocks. This makes the 30 stock fund a good option. This is also supported by the fact that, when looking at the performance on the 2019 data, the deviation of the 60 stock model is only 17.49% (0.418015 to 0.344890) lower than the 30 stock model.

Another good option is the 70 stock fund, which in both cases has a very similar deviation to the 80 stock option. Furthermore, the smallest deviation for the year 2020 occurs when using 90 stocks.

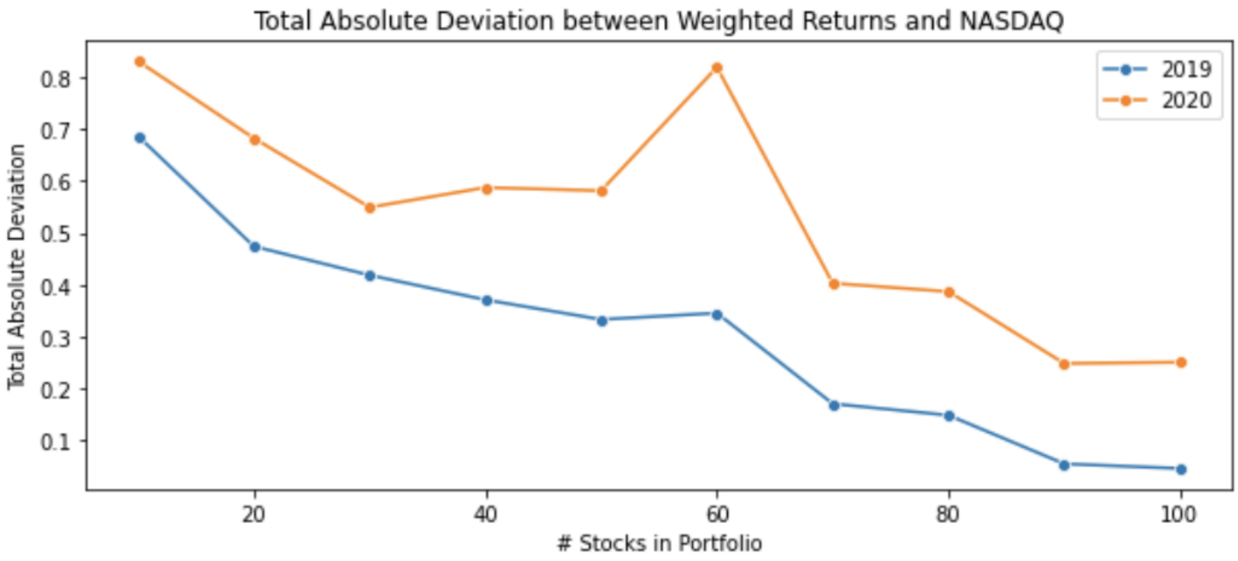
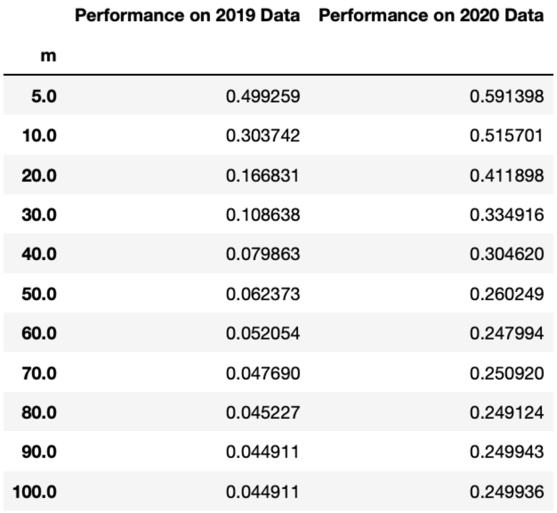


Figure 4.2. Total Absolute Deviation between Weighted Returns and NASDAQ

As a second effort, a MIP model was designed, in which not all the variables were binary. In this case, the 100 Yj variables are still binary, but the 100*100 = 10,000 Xij variables are now constrained by the “Big M” approach to force them to be integers.

The output of this model can be observed in Figures 4.3 and 4.4, where the absolute deviation for each m-stock model is presented for both the years 2019 and 2020.



	Performance on 2019 Data	Performance on 2020 Data
m		
5.0	0.499259	0.591398
10.0	0.303742	0.515701
20.0	0.166831	0.411898
30.0	0.108638	0.334916
40.0	0.079863	0.304620
50.0	0.062373	0.260249
60.0	0.052054	0.247994
70.0	0.047690	0.250920
80.0	0.045227	0.249124
90.0	0.044911	0.249943
100.0	0.044911	0.249936

Figure 4.3. Absolute Deviation of Portfolio Fund with NASDAQ for 2019 and 2020 for various values of m from MIP model

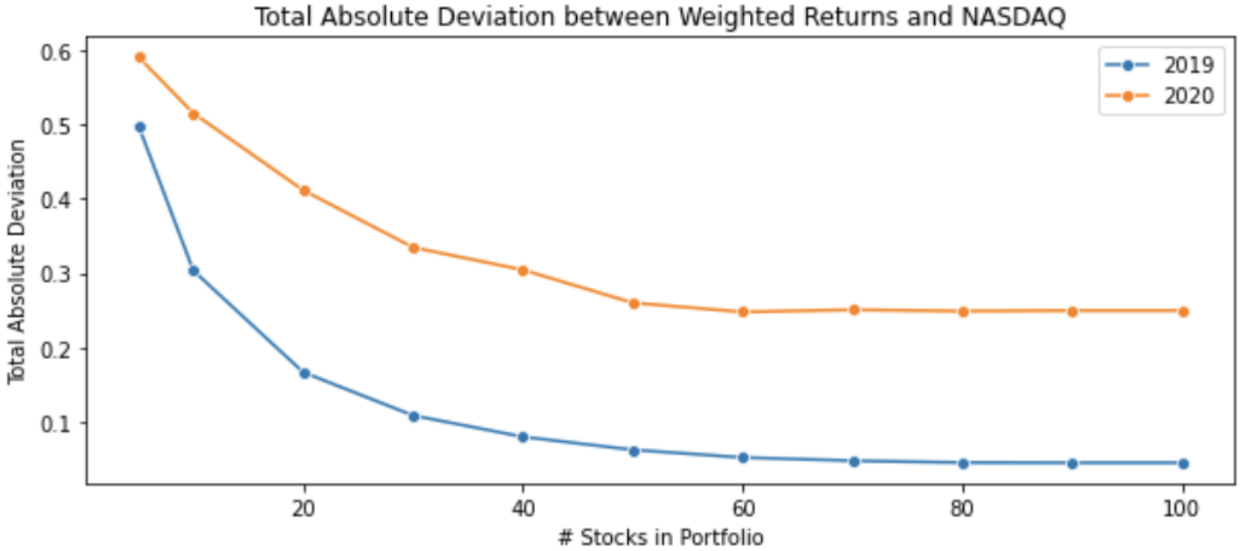


Figure 4.4. Total Absolute Deviation between Weighted Returns and NASDAQ from MIP model

As observed in Figure 4.4, at around m=50 the performance stabilizes, and adding additional stocks does not improve performance by much. We can compare this to the first model, where we saw the performance increase with the addition of more stocks. Also, the magnitude of deviations is smaller in the second method compared to the first. So, the second method yields better and more stable performance. However, it is computationally far more expensive. Around m=50 seems to be the optimal selection, but if we wish to have a lower number of stocks in the fund, then 30 stocks look like the right number, as it can be interpreted as an "elbow" for the 2020 line in the graph.

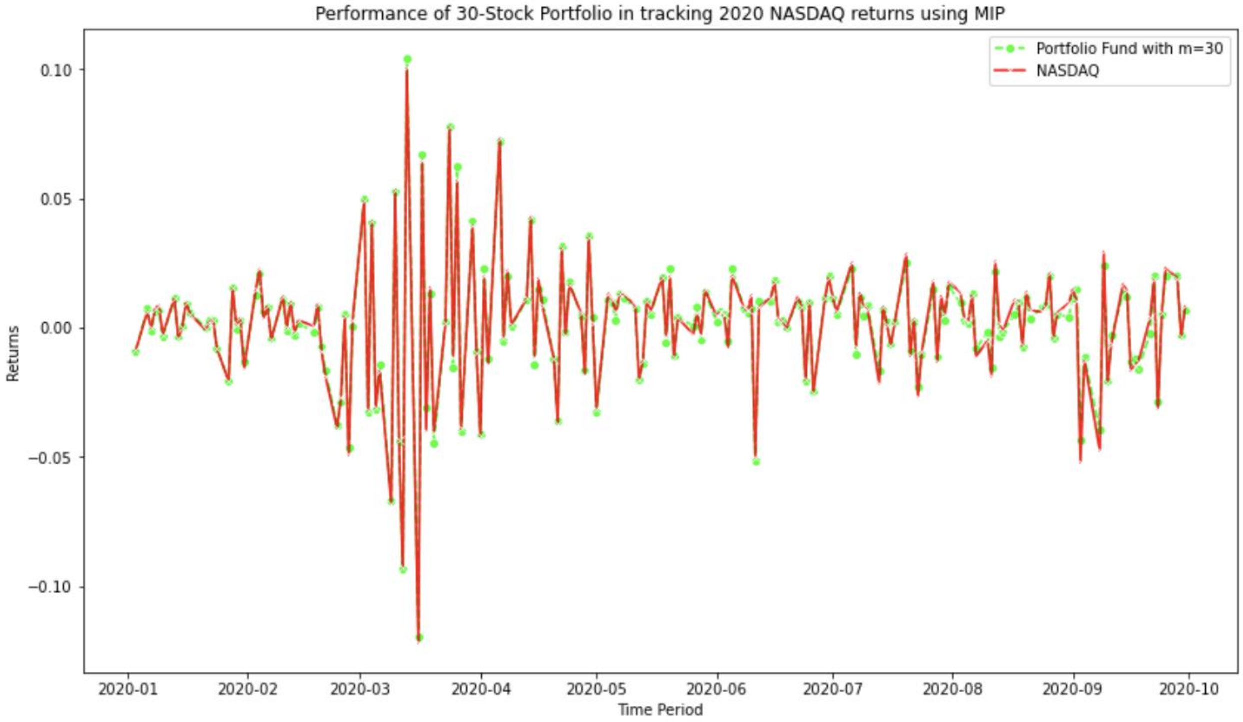


Figure 4.5. Daily Returns of 30 Stock Portfolio vs NASDAQ Index for 2020

Figure 4.5 shows that the 30 Stock Portfolio does a good job of tracking the daily returns of the NASDAQ index for the year 2020.

5. Recommendations

Comparing the two methods, Method 2 clearly performs better at almost all values of ‘m’ with lower error. Its performance effectively peaks at around m = 50, after which it plateaus. Method 1 has a high variance, and its performance continuously improves as ‘m’ increases. Hence, we recommend using Method 2 for modeling our Index fund to mirror the NASDAQ-100.

We recommend using ‘m’ to be around 30, where we see an ‘elbow’ in the graph. This is a small enough number of stocks in our fund that will reasonably mirror. We could increase the value of ‘m’ further if the cost of adding stocks to our portfolio is not too high. The exact value of m can be refined by taking into consideration the cost of adding more stocks to our portfolio (rebalancing costs, price response to trading, etc.) vs. the cost of error while mirroring the NASDAQ-100.

However, one thing to note here is that Method 2 requires considerably higher computing and time. Only if resources are unavailable, should we use Method 1 for modeling our Index fund.

We have currently run the optimization model for ~10 hours, but we could achieve a slightly better performance if we have more computing power and an increase in the runtime.