RM 294 – Optimization I

Project 2 – Stochastic Control & Optimization

Group 24

Andre Han (AH47498), Serena Wu (SW38965), Tsin Kei Derek Tong (TT24754), Vishal Gupta (VG22846)

Objective

A market index is composed of a certain number of stocks with different weights associated with them. 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 simply 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 performance of the market index 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 of portfolio performance.

Approach

To create an index fund, two sets of decision needs to be taken -

1. Which stocks to choose from?
2. What weights to assign to the selected stocks?

Initially, we will try to separate the problem into two parts, first to get the stocks to be kept in the portfolio, and the next part will try to get the weights of the selected stocks. And finally, we will combine the two into a single mixed-integer programming problem to create the best portfolio tracking the index with 'm' stocks.

To recommend the best combination of component stocks, different numbers of stocks 'm' will be tried and the number of stocks (m value) that gives a significantly low difference in stock performance over the next year (next to the year over which optimization is done) will be selected along with the weights that minimize the difference.

Stock Selection

Formulation

Each stock in the index will have to be represented by the component stock in the fund. Hence, the objective of the selection model is to maximize the similarity between all the stocks in the index with their representative component stocks in the portfolio.

Decision Variables

1. y_i : whether stock 'j' is selected in the portfolio or not
2. x_{ij} : is i^{th} stock found to be most similar to j^{th} stock

Constrains

1. Exactly m stocks in the fund
2. Each stock i has exactly one representative stock j in the index
3. Stock j is represented by j only and only if j is in the fund

Total Decision variables - n^2 **+** n

Total Constraints - n^2 **+** n **+ 1**

$$\max_{x,y} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij}$$

$$s.t. \sum_{j=1}^{n} y_j = m.$$

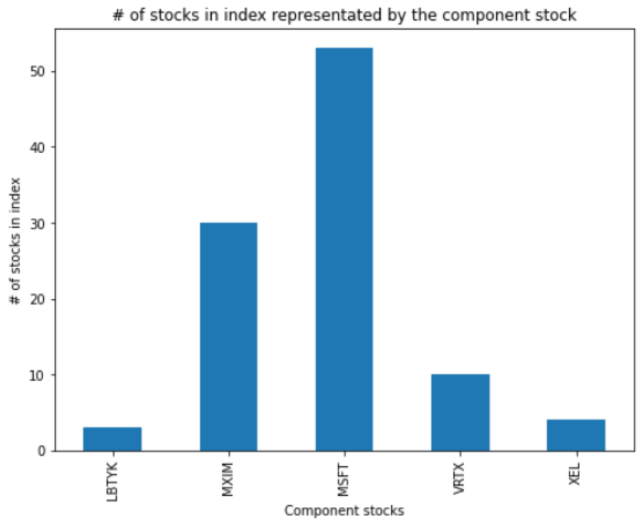
$$\sum_{j=1}^{n} x_{ij} = 1 \quad for\ i = 1,2,\ldots,n$$

$$x_{ij} \leq y_j \qquad for\ i,j = 1,2,\ldots,n$$

$$x_{ij}, y_j \in \{0,1\}$$

Optimal selection

Each component stock will be associated with a certain number of stocks in the index. And as the number of component stocks in the index fund goes up the associated number of stocks will go down. The component stock with the highest number of associated stocks at lower values of ‘m’ will be indicative of the higher relative importance of keeping it in the index fund. For m=5, the stock selected with associated index stock is provided in the figure.



Portfolio Weights Selection

Formulation

After component stocks that maximize the similarity between the fund and the index have been chosen, 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_t representing the index's return on a given day "t", w_i representing the weight of a selected stock "i", andr_{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 $\left| q_t - \sum_{i=1}^{m} w_i r_{it} \right|$ up into two parts by defining an arbitrary decision variable y_i 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 the portfolio
2. y_t : max value of possible difference between index return at *t* and weighted return of selected indexes

Constrains

1. Sum of weights w_i for *m* stocks in the fund equals 1
2. Difference of index returns and weighted return of stocks should be less than y_t

Total Decision variables - $T + m$

Total Constraints - $2 * T + 1$

X vector order for decision variables

X = $[y_1, \dots y_T, w_1, w_2, \dots, w_m]$

Optimal Selection

For each of the "m" component stocks selected, a weight w_i with $0 \le w_i \le 1$ will be assigned such that these weights add up to 1. Each w_i indicates the suggested amount of money or percentage of the portfolio that should be dedicated to stock "i" within the fund. Furthermore, by investing in accordance with the set of optimized w_i on the selected stocks, the discrepancies in return between the index and fund will be minimized. In other words, the fund will track closely the performance of the index.

Out-of-sample vs. In-sample Performance

Since stocks and weights are selected using 2019 stocks returns, the fund performance measured in terms of discrepancies should be better off in 2019 (in-sample) than in 2020 (out-of-sample) data. This is because the 2020 data has not yet happened. Therefore, a gap between the fund performance in 2019 and 2020 is expected with 2020 having a slightly worse performance (i.e. higher discrepancy).

Method 1: Iterations for Different Stock Selections

In the previous section, we selected 5 stocks (LBTYK, MXIM, MSFT, VRTX, XEL) and computed their respective weights. Now that we constructed our portfolio, we will compare its performance by looking at the discrepancies between the returns of the index and fund using 2020 data.

The performance score can help us quantitatively evaluate our portfolio because it sums the absolute difference of returns. Figure 1 shows the returns of our portfolio and the NASDAQ-100 index by trading date. The plots of the returns look very similar. The performance score when calculated using 2020 data is approximately 1.11. If we divide this performance score by the 251 periods we use for evaluation, the average difference in performance per period is 0.0044.

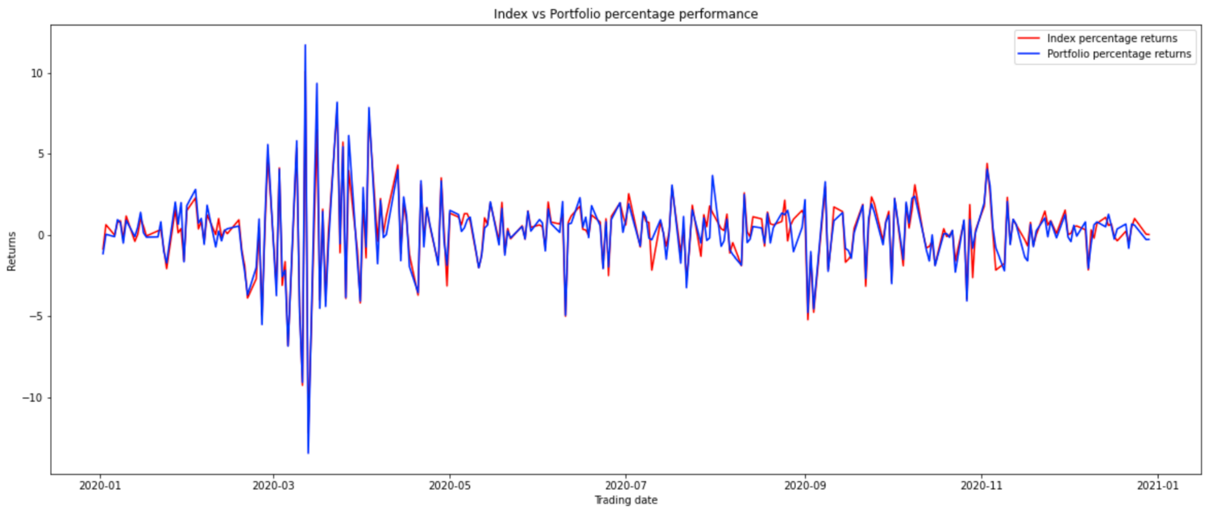


Figure 1: Index vs. Portfolio Percentage Performance

The performance of portfolios may differ based on the number of stocks we decide to include, so we need to identify the optimal amount of stocks to include in our portfolio to match the returns of the index as closely as possible. By using the previously presented stock selection and weight selection method, we repeatedly constructed portfolios with sizes 5, 10, ..., 100.

As shown in Figure 2 below, the performance decreases steadily as the number of stocks included increases when using 2019 data. In general, the performance also decreases as stock amount increases when we are using 2020 data to evaluate it. However, there is a noticeable spike when the number of stocks included in the portfolio reaches 60. This implies that by adding more stocks at this point, the aggregated absolute difference in returns increases.

Figure 2: Performance Score Change w.r.t. Additional Stocks in the Portfolio

Method 2: MIP formulation - Stock selection combined with Weight estimation

We also looked towards an alternative approach where we ignored the stock selection constraint and turned our weight selection process into a MIP. The results of this methodology initially yield a similar trend to our previous portfolio model, with a performance score inversely related to the size of m. However, after the portfolio reaches a size of 60, performance plateaus at 0.05 for the in-sample and 0.37 for the out-of-sample. It's also worth noting that the anomalous spike in performance we saw at 60 stocks is no longer present. Thus, model 2 not only yields better and more consistent performance scores but also achieves it with a smaller portfolio size of 60.

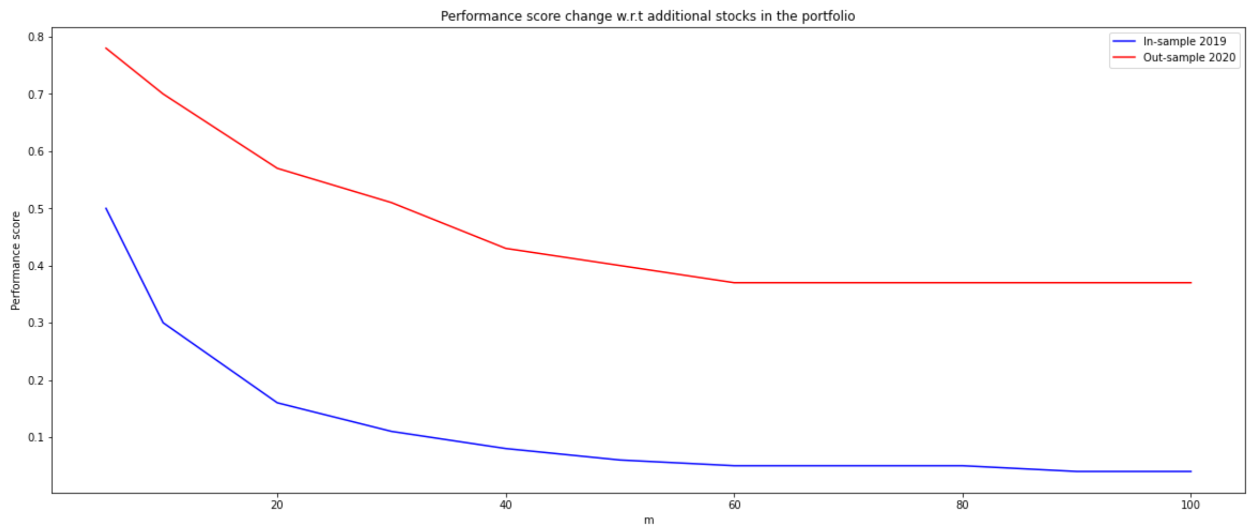


Figure 3: Performance Score Change w.r.t. Additional Stocks in the Portfolio Using MIP Formulation

Recommendations

When we compare the outcomes of our methodologies, Method 2 attains better performance scores at nearly every level of m. Additionally, Method 2 reaches its best performance by 60 stocks, whereas Method 1 sees a continuous decline as the number of stocks increases, reaching its best performance with all 100 stocks. If processing time is a concern, we recommend Method 1 as our choice, as the two-step process of stock selection followed by weight optimization takes considerably less time to obtain a portfolio distribution. In all other cases, we recommend method 2 as our primary model of choice due to it being a better reflection of the NASDAQ with close to half the portfolio size.