MIS 381 N

Stochastic Control and Optimization: Project 2

Index tracking

Money management strategies are largely classified as either 'active' or 'passive'. Active portfolio management tries to achieve superior performance by using technical and fundamental analysis, as well as forecasting techniques. On the other hand, passive portfolio management avoids any forecasting techniques and rather relies on diversification to achieve a desired performance. The most common passive strategy is that of "indexing." Here the goal is to choose a portfolio that mirrors the movements of broad market population or a market index. Such a portfolio is called an index fund. For example, the QQQ Index fund tracks the NASDAQ-100 index. Try plotting the stock price of QQQ along with the NASDAQ-100 index (^NDX) on say finance.yahoo.com.

Constructing an index fund that tracks a specific broad market index could be done simply purchasing all the stocks in the index, with the same exact weights as in the index. However this approach is impractical (many small positions) and expensive (rebalancing costs may be incurred frequently, price response to trading). An index fund with q stocks, where q is substantially smaller than the size of the target population (n), seems desirable.

The project

In this project, we will create an Index fund with 25 stocks to track the NASDAQ-100 index. We will do this in multiple steps. First, we will formulate an integer program that picks q out of n stocks for our portfolio. This integer program will take as input a 'similarity matrix', which we will call ρ. The individual elements of this matrix represent:

$$\rho_{ij} = \text{similarity between stock } i \text{ and stock } j$$

For example, $\rho_{ij} = 1, \rho_{ij} \leq 1$ for $i \neq j$ and ρ_{ij} is larger for more similar stocks. An example of this is the correlation between the returns of stocks i and j. But one could choose other similarity indices ρ_{ij}.

Next, we will construct a portfolio by choosing appropriate weights for the 25 stocks and finally evaluate how well our index fund does as compared to the NASDAQ-100 index.

The Integer Program

We will formulate a problem that clusters the stocks into groups of similar assets and selects one representative asset from each group to be included in the index fund's portfolio. The model is based on a similarity matrix.

$$Z = \max \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij}$$

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

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad for\ i = 1, \dots, n$$

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

$$x_{ij}, y_j = 0\ or\ 1 \qquad for\ i = 1, \dots, n; j = 1, \dots, n$$

The variables y_j describe which stocks j are in the index fund ($y_j = 1$ if j is selected in the fund, 0 otherwise). For each stock $i = 1, .., n$, the variable x_{ij} indicates which stock j in the index fund is most similar to i ($x_{ij} = 1$ if j is the most similar stock in the index fund, 0 otherwise).

The first constraint selects q stocks in the fund. The second constraint imposes that each stock i has exactly one representative stock j in the fund. The third constraint guarantees that stock i can be represented by stock j only if j is in the fund. The objective of the model maximizes the similarity between the n stocks and their representatives in the fund.

Calculating weights

Once the model has been solved and a set of q stocks has been selected for the index fund, a weight w_j is calculated for each stock j in the fund:

$$w_j = \sum_{i=1}^{n} V_i x_{ij}$$

where V_i is the market value of stock i. So w_j is the total market value of the stocks represented by stock j in the fund. The fraction of the index fund to be invested in stock j is proportional to the stock's weight w_j, i.e.,

$$\frac{w_j}{\sum_{f=1}^{n} w_f}$$

- We have provided the following files to you:

Files that have been provided

a. **N100StkPrices.csv** contains daily stock prices for the NASDAQ 100 stocks in 2012. You will use this to calculate the similarity matrix ρ
b. **N100Monthly.csv** is the given monthly stock prices for 2013. You will use them to evaluate how closely your portfolio did in 2013. The monthly NASDAQ 100 in 2013 is 2731.53, 2738.58, 2818.69, 2887.44, 2981.76, 2909.60, 3090.19, 3073.81, 3218.20, 3377.73, 3487.82, and 3592.00. You may also use the index value, 2660.93,in Dec 2012
c. **readData.R** is a script to read the stock prices into R and clean the data (handles missing values, changed stock tickers, etc.).
   For the daily data, it computes a price matrix (rows standing for dates and columns for the 100 stocks), the list of stocks' tickers, the dates and a shares matrix (similar to the price matrix in structure) that contains the total number of shares outstanding in the company (you will need this to compute market capitals of each company).
   For the monthly data, it computes the same things except the total number of shares.

The Specifics

1. Calculate the daily returns for each stock using the 2012 price data.

2. As our initial candidate for the similarity matrix, find the correlation matrix for the returns of the 100 stocks. Note that there will be missing data in the price matrix (NA which stands for Not Available). You need to specify 'use' argument in the 'cor' function in order to handle NAs.

3. Code the integer program above as another function that returns the weights for each of the stock that needs to be in your portfolio

   weights = constructFund(rho, q, priceMat, sharesMat, unique_tickers, unique_dates)

   This will amount to simply formulating the integer program, solving it and then using the market capitalization of each company on the last date to compute weights. The output weights will be a vector of size n with only q non-zero elements denoting the weights.

   (*Note: this will have a very large number of variables and constraints. As you program, you may want to avoid printing out your output.* ***In the final code you submit, please make sure to have no output at all– especially the x values and your coefficient matrix!***)

4. Use your weights to construct an index portfolio at the end of 2012.
   Compare how this index portfolio performs monthly in 2013 as compared to the NASDAQ 100 index using the 2013 stock data provided. Here you may assume that you can directly invest in the Index as if it is a stock. Present your findings using any visualizations or tabulations. You can assume that you will be investing 1 million in your fund. In this case, your shares of each stock you choose to construct the portfolio should be large. As a result, you can leave the shares as non-integers, because the effect that the non-integer parts of shares have should be marginal.

5. Earlier you used correlation as the similarity measure. Now instead create your own similarity measure and put it in a function similarityMat that has the same inputs and outputs

   rho = similarityMat(priceMat, sharesMat, unique_tickers,unique_dates)

   Use this rho in your function call to constructFund and as in step 4, evaluate the performance of this fund as well. Please compare the new fund to the previous fund. Try to explain why the performance of the new fund is better (or worse).

The Deliverables:

1. The following R files:
   a. similarityMat.R
   b. constructFund.R

2. A report that contains
   a. A description of the similarity metric you used.
   b. All your visualizations and tabulations for the two funds you constructed.

   To run your scripts we will simply use this sequence:

```
source('readData.R')

q=25;

source('similarityMat.R')

source('constructFund.R')

rho = similarityMat(priceMat, sharesMat, unique_tickers,unique_dates);

weights = constructFund(rho, q, priceMat, sharesMat, unique_tickers, unique_dates);

<Followed by our evaluation code>
```

Submission Instructions:

Name your report as **project2_gZ.pdf** (Z is your group number). Upload your report, the file similarityMat.R, and the file constructFund.R as 3 separate files to Canvas.

(We want to use speedgrader which requires separate files.)