2021.M2.BigDataAnalysis Group Homework1

Use **Filesync database and MATLAB/Python** to finish the following tasks, note, you don't need to consider any rebalance of CSI300 Index after April 30, 2020.

Problem 1

(1) Select the stocks of the 5 SW-level1 sectors with the largest weights in the CSI300 Index on April 30, 2020;

(2) Construct portfolios for each of the 5 sectors using their weight in CSI300 Index on April 30, 2020 as initial weight.

The result should be:

- You should include a csv file as of the following format with name **problem1.csv**:

StockCode	SectorCode	SectorWeightRank	WeightInIndex	WeightInSectorPortfolio
XXXXX.SH	金融(申万)	1	0.01	0.02
...	...	...	...	...
XXXXX.SH	医药(申万)	5	0.001	0.05

SectorCode column: must use Chinese character to explicitly show the sector name

SectorWeightRank column: 1 means the sector has the largest weight in CSI300 Index as of April 30,2020

WeightInIndex column: weight, for all stocks in CSI300 Index, the total sum should be 1 or 100

WeightInSectorPortfolio column: weight, for all stocks in the same sector, the total sum should be 1 or 100

Problem 2

Plot the total return figure for the 5 portfolios you created in problem 1 between April 30, 2020 and June 30, 2020.

Note: you should use adjusted price for plotting.

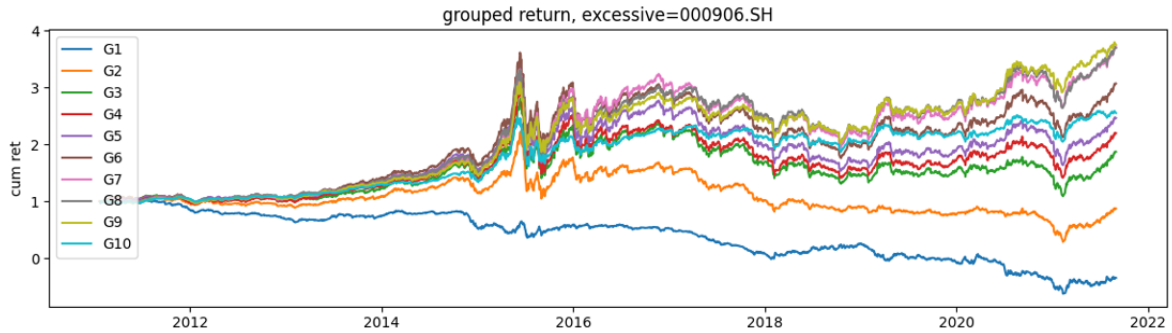
The result should be:

- You should include a csv file as of the following format with name **problem2.csv**:

TradeDate	SectorCode1	SectorCode2	SectorCode3	SectorCode4	SectorCode5
20200430					
...					
20200630					

the **problem2.csv** is a $T \times N$ table, the table is indexed by TradeDate, the columns is the SectorCode, which is the name of the SW-level1 sector, the (i, j) element of the table should be the daily return(not accumulative return) of TradeDate i and SectorCode j.

- You should include a png file whose content is of the following format with name **problem2.png**(other formats of figure is accepted) :



Note: this is only **a sample of figure**, it means, you should include: ***title*, *legend*, *x-axis label*, *y-axis title***

Problem 3

Visualize the 5 sectors' change in the weight of CSI300 Index between April 30, 2020 and June 30, 2020.

The result should be:

- You must include a word file with name **problem3.docx/problem3.doc** to show and explain your result, this result is **not** required to be generated by script automatically.

Submission Guideline

1. You must include a MATLAB/Python script with name **main.m/main.py** which can reproduce all your result, to be more specific, you script should generate **problem1.csv**, **problem2.csv**, **problem2.png**(other formats of figure is accepted) automatically
2. You must include a word file **problem3.docx/problem3.doc**
3. All files in 1. & 2. should be included in a folder with name **2021.M2.BigDataAnalysis.HW1-Group#**
4. You should send your e-mail to both TA and the professor before **December.12 23:59**, e-mail title should be **2021.M2.BigDataAnalysis.HW1-Group#**, e-mail content should explicitly include your group members.