FIN3080 Assignment1 Report

121090885 朱元昊

Question1:

The question asks us firstly to construct quarterly records on *stock return, market capitalization, firm age, total asset, R&D investment, total income, P/E ratio, P/B ratio* for all listed firms in the A-share market.

For *stock return* and *market capitalization*, we can find monthly stock return and total market value through downloading the database of *Stock Trading*. In which we choose the information of stock code, date, closing price, total market value, and return without dividends (Since we really care about the stock's return, it is sensible not to consider about the dividends). We first transform it into quarterly data, using the closing price of the last month of every quarter to be the quarterly closing price, and the product of every (monthly return + 1) and then use the total product - 1 to be the quarterly return, then we use the grouping method to get the mean value of the total market value, totally we can get the quarterly market capitalization, closing price and return.

Then we go to find the information of *firm age, total asset, R&D investment, total income*. We can download the data of the firm's basic information, balance sheet and the income statement. From the firm's balance sheet, we can get the information of firm's total asset, and its current asset with non-current asset, in order to calculate the book value. From the basic information we can find the firm's establish date and its stock's issue date. And from the income statement we can calculate the firm's many kinds of income, and also the total comprehensive income to be our total income and total profit after tax.

To calculate P/E and P/B ratios, we use the formula that P/E = Market capitalization/Earnings which equals to Market capitalization/Total profit after tax, and P/B = Market capitalization/Book value.

Next step is to merge the datasets by time. We group the data with all the quarters available, then pick out the quarters that the datasets have the same, remove the quarters which only a single or several datasets have. Then for every company, we get the quarterly information.

Now we group the data again to have the market type data, actually, we need to groupby against the market type, while removing the individual firm's information, time information

and keep the market type information. Finally, we get 4 kind of market type, and each of them contains the statistics of the data. We here give results that contain all the statistics of the above 8 variables.

Some notes: When handling the missing value, I ignore them when calculating the mean, and add 0 back to the quarterly datasets. For example, the R&D expense, I believe that it is because the company doesn't invest into this field so that it is actually 0, so let 0 replace the missing value space. Also, when handling the other assets term, use 0 to replace as well.

And consider about the A-share market, so for main board, I only keep the stocks that its code start with 600, 601, 603, 605 and 00. For SME board, I keep the stocks that its code start with 002. For GEM board, I keep the stocks that its code start with 300. Since they are just different kind of filtering work, their code are similar.

When calculating the P/E, P/B ratios, it will some time occur Na and Inf, for this case, Inf means profit or book value =0, we shall let them=0 since it has broken down.

Results:

For the Main board, the statistics are

Stock return:

Count	Median	Mean	p25	p75	SD
1.905390e+05	0.000000e+00	3.627794e-02	-1.128257e-01	1.209580e-01	2.819922e-01

Market Capitalization:

Count	Median	Mean	p25	p75	SD
1.905390e+05	3.992520e+09	1.267163e+10	1.986543e+09	8.757561e+09	5.709955e+10

Firm Age:

Count	Median	Mean	p25	p75	SD
190539.000000	15.750000	15.888375	10.500000	21.000000	7.130921

Total Asset:

Count	Median	Mean	p25	p75	SD
1.905390e+05	2.561267e+09	5.201319e+10	1.082255e+09	6.649469e+09	7.565963e+11

R&D Investment:

Count	Median	Mean	p25	p75	SD
190539	0	36211895	0	0	316663227

Total Income:

Count	Median	Mean	p25	p75	SD
190539	18579296	521952362	0	134890735	5947165223

P/E Ratio:

Count	Median	Mean	p25	p75	SD
1.905390e+05	4.380488e+01	-5.417721e+01	1.511684e+01	9.625061e+01	5.728640e+04

P/B Ratio:

Count	Median	Mean	p25	p75	SD
190539.000000	1.434198	3.717689	0.757091	2.513349	266.388874

For GEM board:

Stock return:

Count	Median	Mean	p25	p75	SD
4.803100e+04	0.000000e+00	3.252821e-02	-1.096483e-01	1.141862e-01	2.657269e-01

Market Capitalization:

Count	Median	Mean	p25	p75	SD
48031	4155389653	8237277311	2243881250	8002550227	18882894515

Firm Age:

Count	Median	Mean	p25	p75	SD
48031.000000	14.750000	15.020950	10.250000	19.250000	6.564223

Total Asset:

Count	Median	Mean	p25	p75	SD
4.803100e+04	2.239225e+09	6.598835e+09	1.053135e+09	4.708819e+09	4.065509e+10

R&D Investment:

Count	Median	Mean	p25	p75	SD
48031	0	31939046	0	14476996	159588229

Total Income:

Count	Median	Mean	p25	p75	SD
4.803100e+04	4.470115e+07	1.375263e+08	5.138112e+05	1.294541e+08	6.153395e+08

P/E Ratio:

Count	Median	Mean	p25	p75	SD
4.803100e+04	5.109814e+01	2.039472e+01	2.026194e+01	1.014990e+02	1.126306e+04

P/B Ratio:

Count	Median	Mean	p25	p75	SD
4.803100e+04	1.774139e+00	2.219113e+00	9.728772e-01	2.899342e+00	2.043467e+00

For SME board:

Stock return:

Count	Median	Mean	p25	p75	SD
3.297100e+04	0.000000e+00	3.232517e-02	-1.209784e-01	9.915318e-02	3.060824e-01

Market Capitalization:

Count	Median	Mean	p25	p75	SD
3.297100e+04	3.658613e+09	7.055519e+09	2.017230e+09	6.554217e+09	2.433521e+10

Firm Age:

Count	Median	Mean	p25	p75	SD
32971.000000	15.250000	15.471786	11.500000	19.000000	5.497016

Total Asset:

Count	Median	Mean	p25	p75	SD
32971	1378806538	2537802266	747981023	2643326034	7427370344

R&D Investment:

Count	Median	Mean	p25	p75	SD
32971	0	26741913	0	26047874	113921987

Total Income:

Count	Median	Mean	p25	p75	SD
32971	34051715	73238730	5441226	81054457	408588130

P/E Ratio:

Count	Median	Mean	p25	p75	SD
32971.00000	64.40079	140.28668	17.63177	127.44881	11302.88123

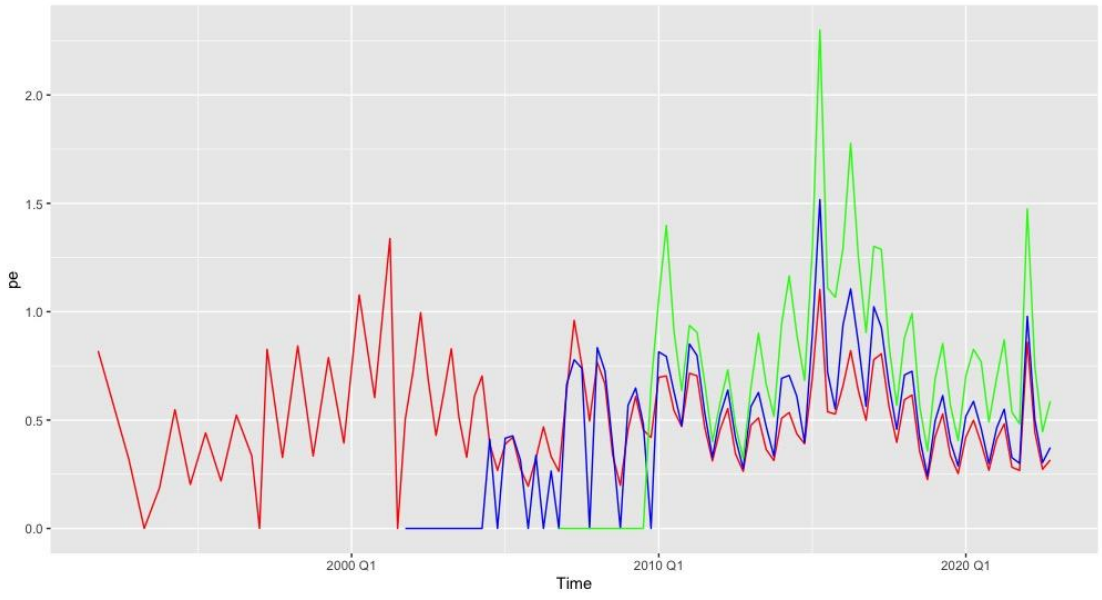
P/B Ratio:

Count	Median	Mean	p25	p75	SD
32971.000000	2.315762	2.910488	1.287061	3.824192	2.816517

By comparing the statistics we can see that, the number of the stocks in the Main board is larger than GEM than SME, the Total Market Capitalization for the Main board is larger than SME then GEM, firms in the main board have higher total asset than GEM and SME board. GEM board has the highest P/E ratio, while the main board has the lowest, even negative.

Question2:

It let us to draw 3 time series line, it is shown below:



Where the red, blue, green line represent the main board, GEM board, and SME board respectively.

From the graph we can see that it may be a good time to enter into the market, because both three lines are likely at the bottom of the mountain, the lower P/E ratio means that the stock is undervalued and it may bring investors high return.

Since now it is more suitable to enter into both three markets, it is more likely for one to long the index ETFs, and when the index or the P/E goes higher, we sell the ETFs or even short the ETFs, because we may see clearly that the ratio curve has some seasonal effect, and we can do sell or short at the top of the mountain.