

FIN 3080 homework 1 report by 122090436

Problem 1:

(a) first find all the data mentioned in problem 1 on dataset CSMAR, clearing it and merging it based on stocks' names and their trading dates. For the quarterly data that cannot match with the monthly counterpart, this merging process is solved by adding an additional column called "lagged variables" to let the numbers of quarterly data in line with the total amount of data counted by month. Anyway, the manually derived data is shown below: (use the **latest** quarterly data for *EPS(TTM)*, *net asset per share (TTM)* to match monthly data)

```
1 # question 1 (a)
2 final_data["monthly_PE_ratio"] = final_data["monthly_closing_price"] / final_data["EPS"]
3 final_data["monthly_PB_ratio"] = final_data["monthly_closing_price"] / final_data["net_asset_per_share"]
4 final_data["quarterly_R&D_expense_by_total_asset"] = final_data["rd_expense"] / final_data["total_assets"]
5 # quarterly firm ages have been derived before the merging of data set with unit (year)
6 final_data["quarterly_firm_ages(by_year)"]
7 # The information including the updated data mentioned in question 1 (a)
8 final_data.info()
```

✓ 0.2s

Python

How manually variables are defined

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 659669 entries, 0 to 659668
Data columns (total 24 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   stock_code                           659669 non-null object
 1   trading_month                        659669 non-null datetime64[ns]
 2   monthly_closing_price                659669 non-null float64
 3   tradable_market_value                659669 non-null float64
 4   total_market_value                  659669 non-null float64
 5   monthly_return                      659669 non-null float64
 6   stock_acronym                       659669 non-null object
 7   date_listed                         659669 non-null datetime64[ns]
 8   establishment_date                  659669 non-null datetime64[ns]
 9   market_type                         659669 non-null int64
10   quarterly_firm_ages(by_year)        659669 non-null float64
11   ending_date                         659669 non-null datetime64[ns]
12   statement_type                      659669 non-null object
13   total_assets                        659669 non-null float64
14   total_liabilities                   659669 non-null float64
15   stock_short_name                    659669 non-null object
16   rd_expense                          659669 non-null float64
17   ROA                                659669 non-null float64
18   ROE                                659669 non-null float64
19   EPS                                659669 non-null float64
20   net_asset_per_share                 659669 non-null float64
21   monthly_PE_ratio                    659669 non-null float64
22   monthly_PB_ratio                    659669 non-null float64
23   quarterly_R&D_expense_by_total_asset 659663 non-null float64
dtypes: datetime64[ns](4), float64(15), int64(1), object(4)
memory usage: 125.8+ MB
```

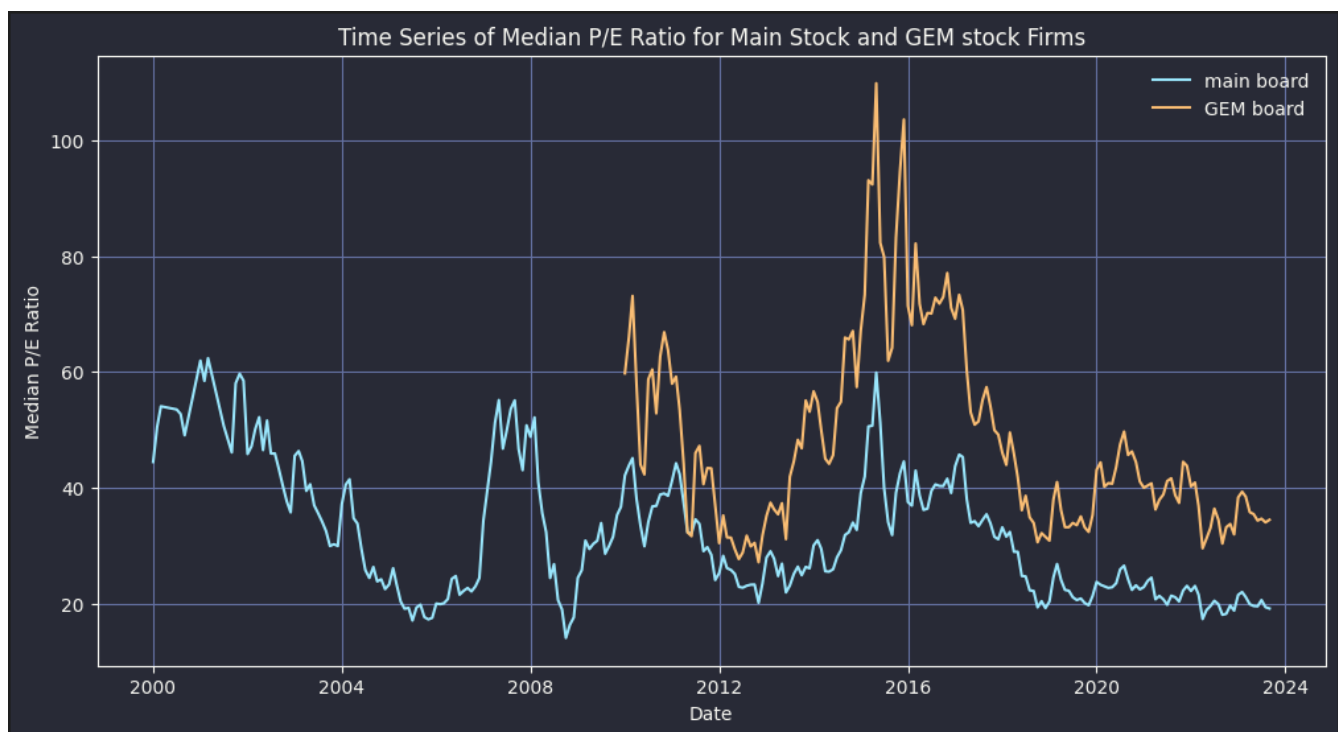
A brief introduction of all variables in the dataset after preprocessing

(b) Given the *market_type*, we can divide stocks into main board (mainly from SSE and SZSE) and GEM board (from ChiNext and STAR). Below image this the comparison of their data such as *monthly stock returns*, *P/E ratios*, *P/Bratios* and *quarterly ROA*, *ROE*, *R&D expense/total asset ratios*, and *firm ages by market type*.

<pre> 1 # summary statistics for main board in target financial metrics 2 main_board[metrics].describe() </pre>							
✓ 0.1s							
	monthly_return	monthly_PE_ratio	monthly_PB_ratio	ROA	ROE	rd_expense	quarterly_firm_ages(by_year)
count	543253.000000	5.432530e+05	5.432530e+05	543253.000000	543253.000000	5.432530e+05	543253.000000
mean	0.012166	4.646797e+01	inf	0.051396	-0.072232	5.253487e+07	26.703933
std	0.157398	6.242943e+03	NaN	0.323017	12.223970	4.889450e+08	5.255688
min	-0.891827	-1.288000e+06	-2.825000e+05	-64.817721	-8782.578891	-1.602876e+07	6.221685
25%	-0.069053	1.257371e+01	1.596606e+00	0.023456	0.022863	0.000000e+00	23.456455
50%	-0.000493	2.823412e+01	2.549538e+00	0.051216	0.070517	0.000000e+00	26.558440
75%	0.076000	5.696421e+01	4.165078e+00	0.087189	0.124388	3.968285e+06	30.659741
max	22.052631	1.279000e+06	inf	29.726154	343.759902	4.975324e+10	66.188830

<pre> 1 # summary statistics for GEM board in target financial metrics 2 GEM_board[metrics].describe() </pre>							
✓ 0.0s							
	monthly_return	monthly_PE_ratio	monthly_PB_ratio	ROA	ROE	rd_expense	quarterly_firm_ages(by_year)
count	112712.000000	1.127120e+05	112712.000000	112712.000000	112712.000000	1.127120e+05	112712.000000
mean	0.011278	1.029826e+02	4.701128	0.056474	0.010731	4.832735e+07	21.422281
std	0.167186	1.042578e+04	18.276651	0.103146	1.971290	2.069396e+08	4.929442
min	-0.836559	-6.928571e+05	-733.511140	-1.887798	-315.152322	-1.130706e+06	6.005394
25%	-0.083956	2.397691e+01	2.378533	0.026690	0.031409	0.000000e+00	18.309296
50%	-0.004545	4.211292e+01	3.510639	0.061452	0.077258	1.247615e+07	21.378427
75%	0.081054	7.367487e+01	5.503460	0.100220	0.123909	4.226910e+07	24.116277
max	6.399758	2.087500e+06	3484.119346	1.292383	7.315695	1.551045e+10	42.985545

Problem 2:

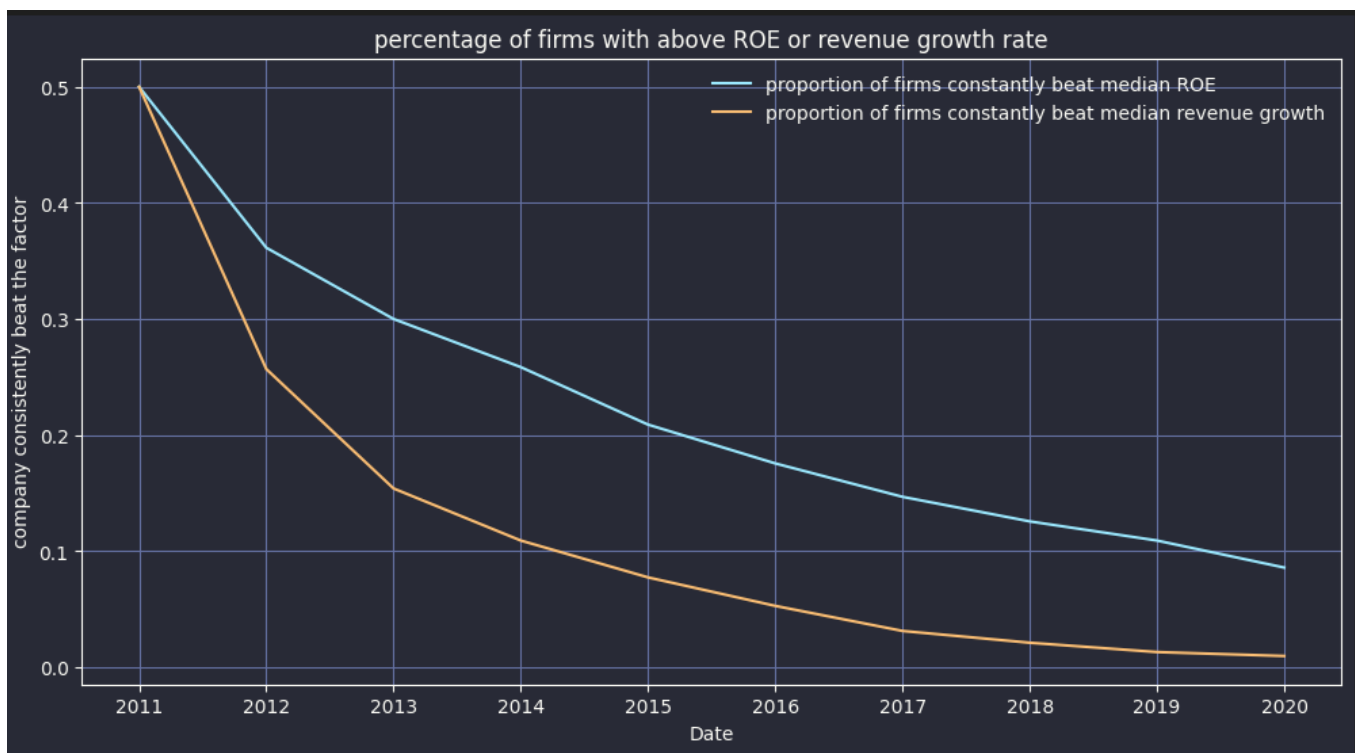


The graph above is the two-time series for annual median P/E ratio by market type (main board & GEM board)

(i) From above graph, it is advisable to consider new investments in both main and GEM markets, since their median P/E ratio is relatively low compared to their corresponding previous data, implying their market value might be **undervalued**. Investment opportunity may be found within this market in Sep. 2023.

(ii) Trend following strategy: If one market is observed to be consistently outperforms the other market, then I can catch the inertia of it and long to the market with rising P/E ratio and short the market with declining P/E ratio based on trading index ETF. This strategy is driven by capturing the market sentiment in a short term and hopes to generate profits through applying practical and efficient algorithms to do the quantitative trading. Moreover, this strategy can identify whether the market is extremely hot or cold. Say, if the P/E ratio of a market is conspicuously higher than the other for no reasons, then the strategy is considered to short the market which is over hot and long the counterpart which is over cold to control risk and be sustainable.

Question 3:



From the above graph, we first find the corresponding stocks that beat the median of ROE and total revenue growth at the year 2011, and then observe them over a decade. As we can see, it is very hard to beat the market constantly over time, even though we set the threshold as "median level". Over ten years, there are less than 10 percent of firms can beat the median factor, indicating hardness to outperform the market consistently.