Problem 1

Data Collection and Clarification

The data are all downloaded from CSMAR as required by the problem.

Some necessary clarifications:

1. Monthly return without dividend reinvested (eliminate the such influence of dividend) used.

2. Return on Assets - B and Return on Equity - B are used. B here means that *average* balance of total assets / equity are used in the denominator. This may help cope with extreme values.

3. Net Assets per Share is used to present book value in P/B ratios.

4. Noted that there are many types of Earning per share. Earning per share - TTM is used to present the statistics as it includes twelve months' data, which is less influenced by extremes and corresponding with real market better. In addition, Earning per share - 1 will also be used to show the robotness of conclusion.

Question (a)

Data Manipulation Process

Data cleaning. Unnecessary data such as "statement type" (parent statements are omitted) and "firm name" are deleted using *drop* function.

Date conversion and merging. Converts the date columns in both dataframes to a period type representing months. In EPS_bookValue, it additionally adjusts the date to the start of the next month to align with the reporting periods. Merges on date columns. Filling missing values by forward filling within each stock_code.

P/E P/B ratios. Calculates the P/E and P/B ratios using the closing_price, earnings_per_share, and book_value_per_share columns. These ratios are fundamental financial metrics used to assess the valuation of stocks.

R&D/Asset ratio. Converts the date columns in the R&D and asset_liability dataframes to a quarterly period type. Merging as usual. Calculates the ratio of R&D expense to total assets for each entry in the merged dataframe.

Firm ages Calculation. Converts the quarter end dates to string, applies a function to transform quarter indicators to the last day of the respective quarter, and calculates the firm age by comparing this date to the establishment date. The age is expressed in years.

Results shown in the file.

Question (b)

Summary Statistics. By using *.describe*() function in *pandas*, we can get the data required easily form merged dataframes. All the numbers are round to two decimals. Firm age in year.

Table 1: Main Board

	monthly return	P/E	P/B	ROA	ROE	RD/asset	firm age
count	323852.0	313535.0	322083.0	186929.0	185781.0	73781.0	193641.0
mean	0.01	42.6	1.19	0.02	0.03	0.02	16.1
std	0.15	5552.91	828.13	0.05	0.65	0.02	6.65
min	-0.88	-1288000.0	-282500.0	-1.87	-174.89	-0.0	-0.6
25%	-0.07	14.61	1.64	0.0	0.01	0.0	11.32
50%	-0.0	28.51	2.54	0.01	0.03	0.01	16.01
75%	0.07	52.71	4.02	0.03	0.06	0.02	20.54
max	12.75	410000.0	9346.11	4.49	14.02	0.87	65.79

Table 2: GEM Board

	monthly return	P/E	P/B	ROA	ROE	RD/asset	firm age
count	113574.0	107601.0	113631.0	64574.0	64416.0	41542.0	66239.0
mean	0.01	95.8	4.75	0.02	-0.0	0.02	16.39
std	0.18	10626.56	18.23	0.05	6.92	0.03	5.44
min	-0.86	-692857.14	-733.51	-1.88	-1756.04	-0.0	0.45
25%	-0.08	24.4	2.38	0.0	0.01	0.01	12.53
50%	-0.0	41.94	3.53	0.02	0.02	0.02	16.18
75%	0.08	72.0	5.55	0.04	0.05	0.03	19.85
max	6.4	2087500.0	3484.12	0.97	1.32	2.7	42.57

Findings. This statistics reveals that:

1. the total number of observations is greater in the main board compared to the GEM board.

2. When looking at monthly returns, both boards average similarly, yet the GEM board displays significantly higher maximum values and volatility, indicating a greater risk-return ratio.

3. In terms of PE ratios, the GEM board generally records higher (and more negative) figures than the main board. Similarly, for PB ratios, the GEM board exhibits higher (and more negative) statistics, although its maximum and standard deviation are lower than those of the main board.

4. Regarding ROA and ROE metrics, aside from outlier values, both boards share comparable statistics. The ratio of R&D expense to total assets is notably higher on the GEM board.

5. In the case of quarterly firm ages, indicating the duration firms have been in operation, the main board surpasses the GEM board across all metrics. Summarizing, this comparison indicates

that firms on the main board are generally more stable, whereas those on the GEM board present higher risks and returns.

Problem 2

Graph. From 2000-01 to 2023-09. Noted that Earnings per share - TTM1 is used in this graph, which is very similar to real figure published by other institutions.

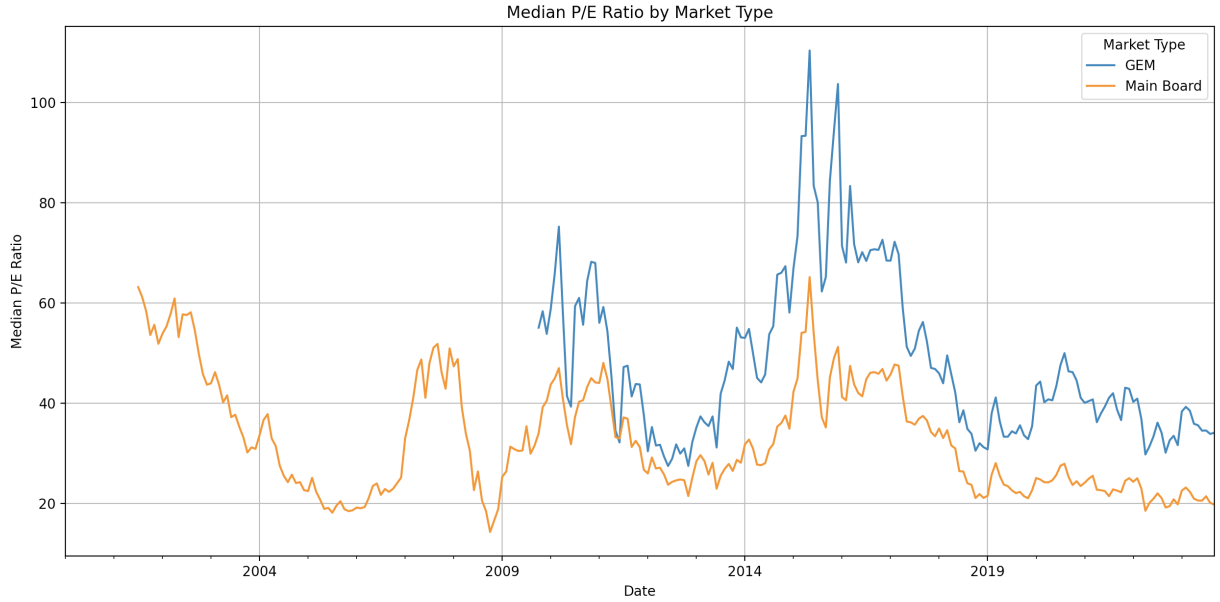

Figure 1: Time-series for Median P/E ratio by Market Type.

Question (i)

Yes. As of September 2023 (the end of the graph), it's recommended to explore new investment opportunities in both markets. The analysis of the trend line indicates that we might be at a low point, P/E ratio is really low (nearly the lowest in the history), suggesting that many companies could be undervalued at present. Historical data shows regular patterns of fluctuation, hinting at a probable future increase in market values. Therefore, entering either market now is considered a profitable move.

Question (ii)

A strategy of buying or going long on index ETFs during periods of low P/E ratios and selling or shorting them during high P/E periods may be a profitable index EFF. The reason comes from the distinct patterns observed in the Price-to-Earnings (P/E) ratios, characterized by lower values on the main board and more significant fluctuations on the GEM board. Investors might also

benefit from adjusting their investment allocations between the main board and GEM board based on their risk tolerance. Those open to speculation might lean towards the GEM, whereas more cautious investors might prefer the stability of the main board.

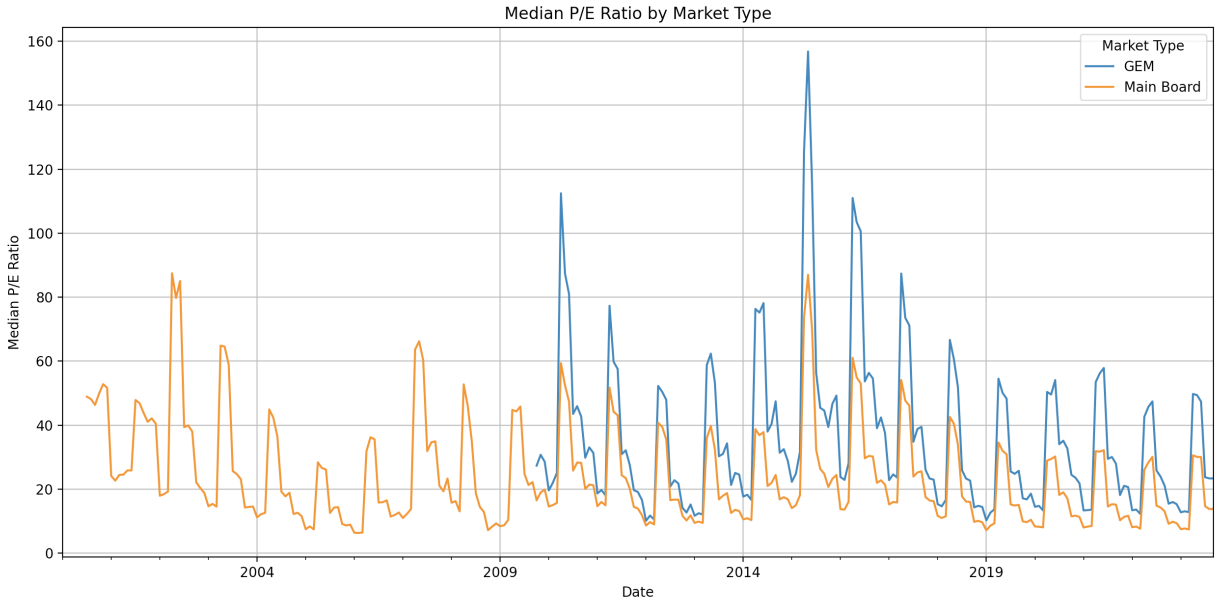

Figure 2: Another Time-series for Median P/E ratio by Market Type, on non TTM EPS data.

Based on the images obtained from non TTM EPS, the P/E ratio fluctuations in the A-share market are more pronounced, which once again strongly supports the above conclusion.

Problem 3

Table. From 2011 to 2020. All numbers are roud to four decimals.

Table 3: The Annual Median for ROE and Total Revenue Growth Rate

year	ROE median	revenue growth median
2011	0.0912	0.1613
2012	0.0731	0.0595
2013	0.0684	0.1057
2014	0.0701	0.0749
2015	0.0689	0.0311
2016	0.0753	0.0946
2017	0.0819	0.1586
2018	0.0701	0.1073
2019	0.0718	0.0648
2020	0.0781	0.0333

Graph. From 2011 to 2020.

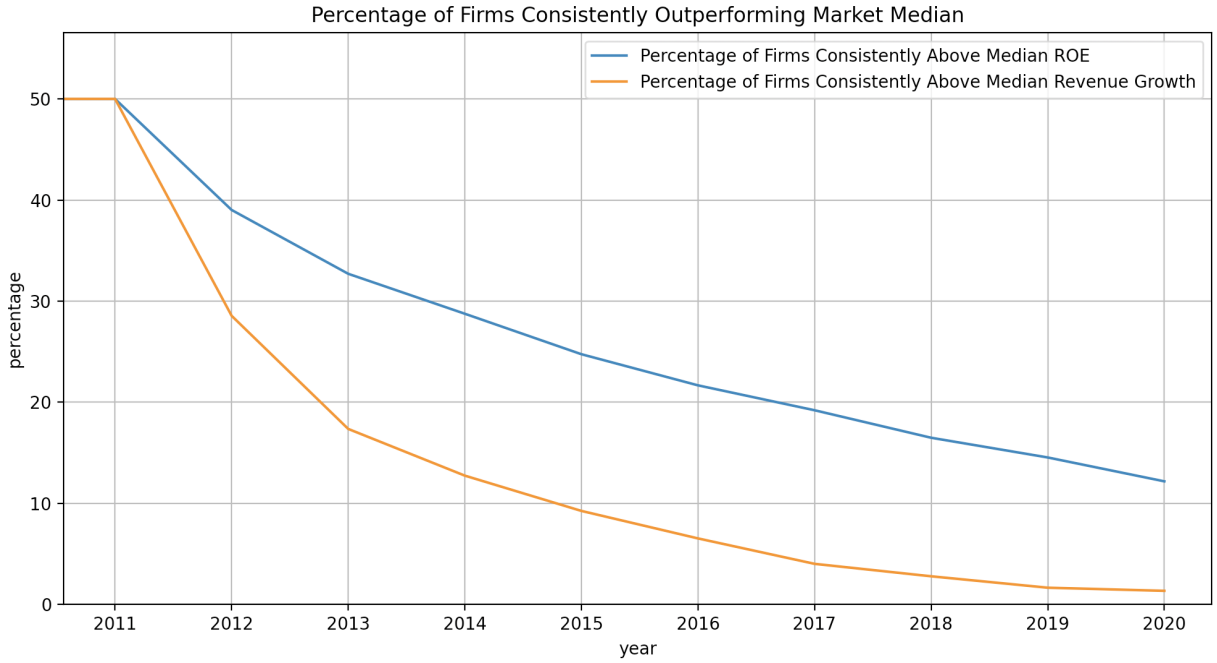

Figure 3: Time-series of the percentages of companies that consistently maintain above-median ROE and total revenue growth rate

Noted that it represents the percentage of firms that outperform consistently than the median ROE and revenue growth rate, repectively.