Problem 1

Notification on Data Collection

All the required data are downloaded from CSMAR. Below are some notifications:

1. Monthly return with dividend reinvested is used. This is to reflect the true aggregate return of stocks, including how well dividends are reinvested.
2. Return on Assets - A and Return on Equity - A are used. On the one hand, ROE here is used to calculate P/B ratio by using P/B = ROE × P/E. Since P/B and P/E all need to be monthly updated, ROE should reflect return in a specific period. In this case, ROE-A satisfies our needs since it embodies the financial situation in a relatively short period. Correspondingly, we choose Assets - A.
3. Earnings per share - TTM is used. It reflects the overall profitability of the company in the last twelve months, which is conducive to long-term trend analysis.

Problem1. a

Basic Financial Metrics Calculation

Note that useless columns like "statement type" and "firm name" are deleted using `drop()` function, with parent statement omitted first. All the results are shown in the .ipynb file.

1. **P/E ratios.** Calculate P/E using closing price and EPS. First transform the trade date and record date to month period using `to_period()` function. Then adjust the record month period to its next month, starting from which the EPS data can be used. Merge two dataframes by aligning stock code and month. Forward fulfill missing EPS data. Calculate the result.
2. **P/B ratios.** Calculate P/B using P/E and ROE. Transform the date column into mergeable form using the trick in 1. Then merge two dataframes by aligning stock code and month. Forward fulfill missing ROE data. Calculate the result.
3. **R&D Expense/Total Asset ratios.** Transform the record dates in balance and rdexpense dataframes to quarter period using `to_period()` function. Merge two dataframes by aligning stock code and date. Calculate the result.
4. **Firm Ages.** Merge balance and company profile dataframes on stock code and firm name. Then drop the NaN data in record date and establish columns. Calculate the result. Note that the firm age is expressed in year.

Problem1. b

Summary Statistics

Use previous merged dataframes to get statistical results using `describe` function. The summary statistics are classified by market types and shown in two tables below.

Table 1: Main Board

	Return	P/E	P/B	ROA	ROE	RD Expense/Asset	Age
count	323852	314519	314159	124654	123901	58296	113508
mean	0.012	34.67	3.57	0.031	0.038	0.017	6259.50
std	0.152	5590.70	719.70	0.109	1.331	0.019	2424.71
min	-0.882	-1288000.0	-133010.59	-18.59	-274.05	-0.002	1
25%	-0.069	14.63	0.589	0.007	0.015	0.004	4588.75
50%	-0.001	28.59	1.39	0.023	0.044	0.011	6253
75%	0.073	52.98	2.67	0.050	0.089	0.023	7835
max	12.752	410000.0	153659.8	12.76	64.06	0.872	24744

Table 2: GEM Board

	Return	P/E	P/B	ROA	ROE	RD Expense/Asset	Age
count	113574	107601	107601	48755	48586	33970	44581
mean	0.012	95.24	4.221	0.039	0.047	0.025	6143.41
std	0.182	10627.14	267.78	0.087	0.727	0.035	2004.16
min	-0.864	-692857.1	-5368.59	-4.600	-96.504	-0.002	164
25%	-0.084	24.40	0.731	0.009	0.014	0.009	4722
50%	-0.005	41.96	1.898	0.029	0.043	0.018	6068
75%	0.081	72.03	3.610	0.063	0.093	0.031	7414
max	6.400	2087500.0	52705.2	0.789	77.13	2.696	15538

Findings

This statistics reveals that:

1. Notice that the number of observations in the main board is greater than that in GEM board. This refelcts that the number of listed companies in the main board is larger, and it has bigger market size.

2. Considering monthly returns, main board and GEM board exhibit very similar average. The GEM board has much higher standard deviation and the main board has higher maximum. Hence we conclude the main board has greater potential gains, and the GEM board appears riskier.

3. The GEM board has generally higher P/E ratio than main board according to the p25, p50, p75 statistics. However, its minimum and maximum appear more extreme. This

reflects that the growth expectations of companies in the GEM board are stronger, but the valuations can be frothy.

4. Considering P/B ratio, the GEM board has higher average, which may imply that companies in it are overvalued.

5. The mean ROA and ROE of the GEM Board are slightly higher. The minimum ROA and ROE values of the Main Board are lower. Hence we conclude the GEM board is slightly more profitable, but the main board has more extreme loss-making companies.

6. On average, the proportion of R&D Expense in the GEM board is higher. Hence we may conclude companies in the GEM board are more focused on research and development investment, which may reflect their greater innovation and growth.

7. On average, the firms in the main board are older. Hence they may appear more mature and stable.

Problem 2

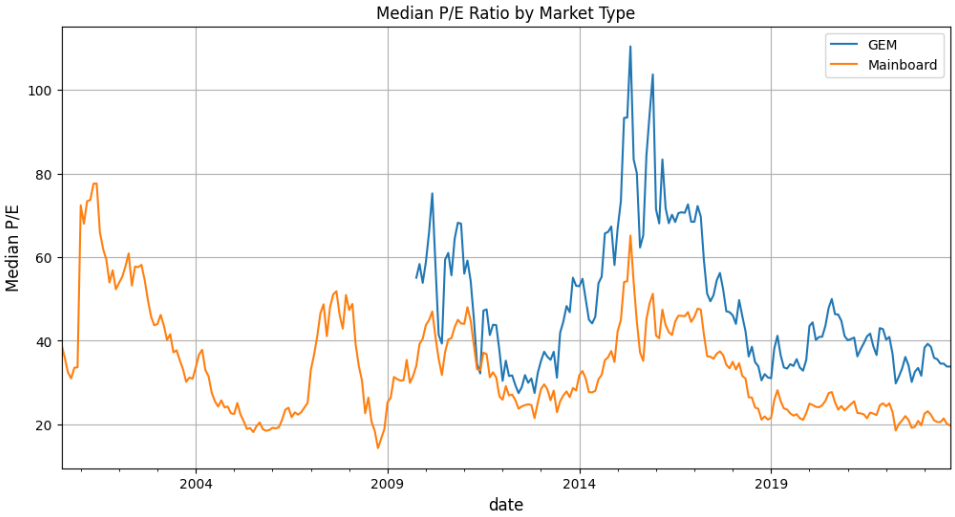

Figure 1: Time-series for Median P/E ratio by Market Type

Problem2. i

Yes. According to the graph, on September 2023, the P/E ratio almost arrived at the historically lowest point for both GEM and Mainboard. This indicates that the firms can be highly undervalued. Since historical data reflects a fluctuating pattern, we can expect that investment at this point can be lucrative.

Probelm2. ii

The strategy takes advantage of cyclical fluctuations in P/E ratios according to the graph. Buy low and sell high index ETFs. Specifically, we can extract an upcycle of a time serie. Utilize the highest and lowest value to approximate every decile of P/E ratio in a cycle. Then we use these deciles as the threshold for making investment decisions. To illustrate, we analyse the GEM time serie in figure 1. Assume that we now have 3000$ for investment, using above strategy, we may have following investment plan. (B, S indicate Buy and Sell respectively.)

Table 3: Investment Plan for GEM

	0-10%	10-20%	20-30%	30-40%	40-50%	70-80%	80-90%
P/E	30-38	38-46	46-54	54-62	62-70	86-94	94-102
Trade	B 1000$	B 800$	B 600$	B 400$	B 200$	S 2000$	S 1000$

Here we omit the last decile segment to avoid loss in extreme market condition. Besides, notice that in figure 1, GEM time serie has apparent greater fluctuation than Mainboard serie. Therefore, we can also adjust our allocation in two markets to achieve higher return under lower risk.

Problem 3

Table 4: The Annual Median

year	ROE median	revenue growth median
2011	0.0912	0.1787
2012	0.0731	0.0675
2013	0.0684	0.1070
2014	0.0701	0.0807
2015	0.0689	0.0408
2016	0.0753	0.1115
2017	0.0819	0.1757
2018	0.0701	0.1099
2019	0.0718	0.0682
2020	0.0781	0.0413

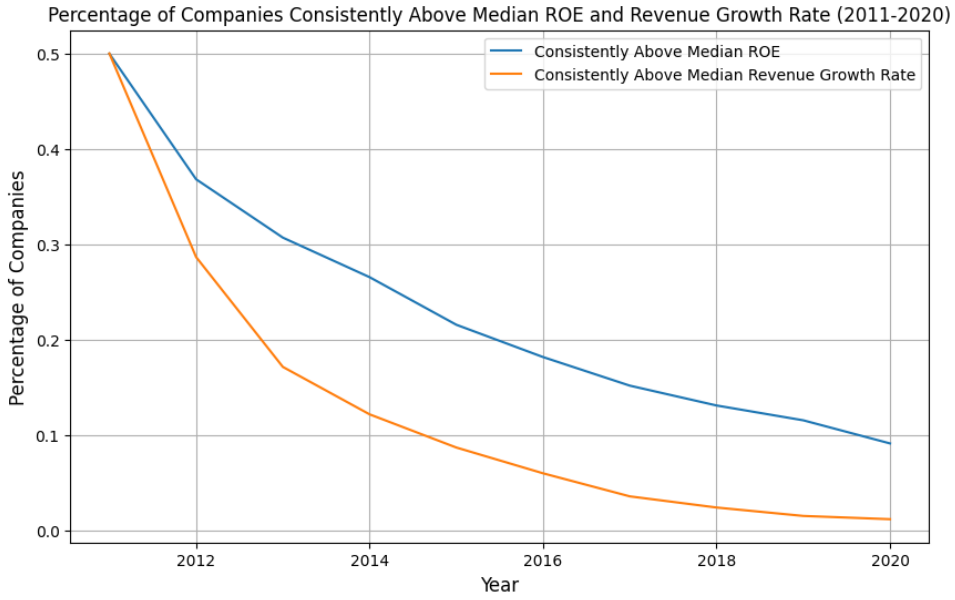

Figure 2: Time-series of percentage of companies that consistently maintain above-median ROE and revenue growth rate