Problem 0

All the required data are downloaded from CSMAR.

Derive P/B Ratio

1. Unnecessary columns like "ShorName_EN" and "Typrep" are deleted using `drop()` function, with parent statement omitted first.

2. 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 NAPS data can be used.

3. Merge two dataframes by aligning stock code and month. Forward fulfill missing NAPS data. Calculate P/B using closing price and NAPS.

4. Extreme P/B data that is less than 5th percentile or greater than 95th percentile are removed as required.

Problem 1

P/B ratio at Dec.2010, ROE-TTM at 2010Q4 and Stock Volatility at 2010/12/31 are extracted and merged into one dataframe. Then we use OLS regression, results shown as below:

OLS Regression Results

Dep. Variable:	pb_ratio	R-squared:	0.133
Model:	OLS	Adj. R-squared:	0.131
Method:	Least Squares	F-statistic:	104.9
Date:	Fri, 14 Mar 2025	Prob (F-statistic):	4.14e-43
Time:	20:12:10	Log-Likelihood:	-2808.0
No. Observations:	1374	AIC:	5622.
Df Residuals:	1371	BIC:	5638.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2057	0.279	0.738	0.461	-0.341	0.753
roe	0.7093	0.327	2.167	0.030	0.067	1.351
volatility	8.6747	0.611	14.204	0.000	7.477	9.873

Omnibus:	107.494	Durbin-Watson:	1.800
Prob(Omnibus):	0.000	Jarque-Bera (JB):	149.290
Skew:	0.637	Prob(JB):	3.82e-33
Kurtosis:	3.992	Cond. No.	14.6

Figure 1: OLS Regression Result

In this regression, P/B is dependent variable, ROE and Stock Volatility are independent variables. The regression is implemented at a single time point. The subscript i indicates different listed companies. And we get the regression result:

$$P/B_i = 0.2057 + 0.7093\,\text{ROE}_i + 8.6747\,\text{Stock Volatility}_i + \epsilon_i$$

Interpretation of Regression Table

1. **F-statistics**. The F-statistic tests the overall significance of the model. Its low p-value (4.14e-43) indicates that the model is statistically significant, i.e, at least one of the predictors has a great correlation with P/B.

2. **Intercept and Coefficients**. The intercept is 0.2057 with p-value = 0.461 > 0.05. This means that when ROE and Stock Volatility are zero, the P/B is not significantly different from zero; The coefficient for ROE is 0.7093 with p-value = 0.03 < 0.05, indicating that ROE is statistically significant at 5% level. This suggests that ROE has a meaningful impact on the P/B ratio; The coefficient for Stock Volatility is 8.6747 with p-value = 0.0 < 0.05. Therefore Stock Volatility has a strong impact on P/B ratio.

3. **R-squared**. R-squared = 0.133. This implies that 13.3% of the fluctuation in P/B can be explain by the model. While the model has some explanatory power, a significant portion of the variance remains unexplained.

4. **Diagnostics Tests**. Omnibus test and Jarque-Bera test check the normality of the residuals. Since both of them have low p-value, we opine that the residuals are not normally distributed. And positive skewness further confirms that the residuals are skewed to the right.

Further Discussion

1. **ROE**. Companies with higher ROE tend to have higher P/B ratios. This aligns with the intuition that more profitable companies, i.e, companies with higher ROE, are valued highly by investors.

2. **Stock Volatility**. Generally, higher volatility is associated with higher risk, leading to lower valuations. Yet in this case, higher volatility is associated with higher P/B ratios. This might indicate that investors are willing to pay a premium for stocks with higher volatility. It could be due to expectations of higher returns.

3. **Model**. The low R-squared suggests that while ROE and Stock Volatility are significant predictors, there are other great factors ignored by the model. The non-normal distribution of residuals indicates there might be ommitted variable bias, non-linear relationships, etc.

Problem 2

Companies are divided into ten groups based on their P/B ratio from the previous month, monthly updated. Thus we formulate ten up-to-date portfolios. Calculate the average return of each porfolio in every month. Then calculate the cross-time average return of each portfolio. The graph illustration is shown as below:

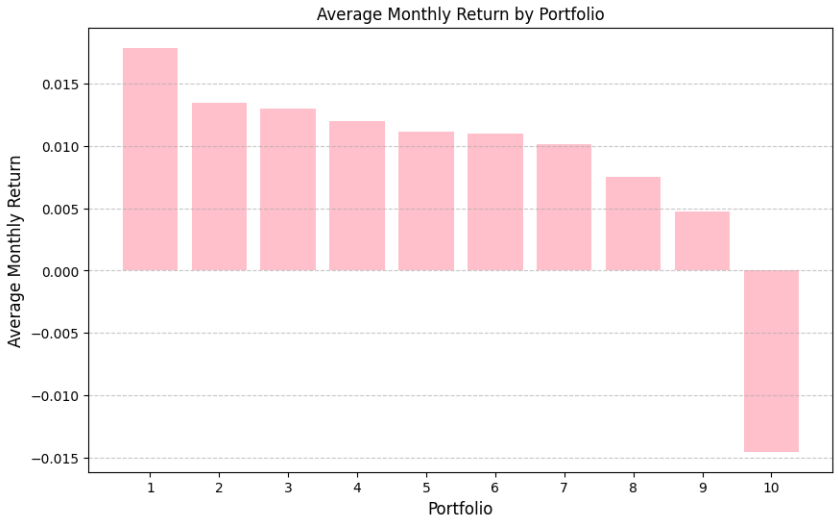

Figure 2: Average Monthly Return by Portfolio

Discussion of Findings

1. **General Trend**. According to the chart, portfolios with lower P/B tend to outperform those with higher P/B over the period from Jan 2010 to Dec 2024. Specifically, Portfolio 1 has the highest average return of 0.017900, while Portfolio 10 has the lowest average return of -0.014539. The result can be interpreted by that undervalued stocks, i.e, those have low P/B provide higher returns compared to overvalued stocks, i.e, those have high P/B.

2. **Negative Return of Portfolio 10**. This further confirm that stocks with very high P/B could underperform significantly. This is because overvaluation drags down the return. Market corrections affect growth stocks more severely.