

Preliminary Data Processing

Data Collection

Because all the files required to be downloaded from CSMAR in the homework document have clearly specified the download methods, form names, and project contents, etc. Therefore, the data used in my code is completely in accordance with the assignment requirements.

The time range for all data is from September 2009 to October 2023, to avoid data gaps caused by factors such as the “last-month” requirement specified in the questions.

Data Manipulation

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

Date conversion. Dates in all involved dataframes are converted to a monthly period format, with additional adjustments for NAPS & ROE data by adding one month, to make it function as the closing data at the end of quarters.

Date merging. The merging process links data on stock codes and the adjusted dates, using a left join approach. Finally, missing value is forward-filled for the ‘ROE’ and ‘NAPS’ columns within each stock code group, ensuring that each stock has continuous data for these metrics.

Extreme data removal. The code calculates the P/B ratio, then filters out records with extreme P/B ratios, specifically those below the 5th percentile or above the 95th percentile as required.

Problem 1

Manipulation

In the multiple linear regression, the dependent variable is P/B ratio, the two independent variables are ROE (return on equity) and stock volatility respectively. i represents observations, i.e. different companies. The time range is 2010Q4.

The regression result is: :

$$P/B_i = 0.1406 + 1.7876 \times ROE_i + 8.6811 \times volatility_i + \epsilon_i$$

Graph. As shown below, on 2010Q4.

OLS Regression Results						
=====						
Dep. Variable:	PB		R-squared:	0.133		
Model:	OLS		Adj. R-squared:	0.132		
Method:	Least Squares		F-statistic:	106.2		
Date:	Fri, 29 Mar 2024		Prob (F-statistic):	1.23e-43		
Time:	22:05:21		Log-Likelihood:	-2864.5		
No. Observations:	1385		AIC:	5735.		
Df Residuals:	1382		BIC:	5751.		
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	0.1406	0.287	0.490	0.624	-0.423	0.704
ROE	1.7876	0.414	4.314	0.000	0.975	2.601
volatility	8.6811	0.623	13.926	0.000	7.458	9.904
=====						
Omnibus:	124.785		Durbin-Watson:	1.786		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	170.268		
Skew:	0.721		Prob(JB):	1.06e-37		
Kurtosis:	3.932		Cond. No.	14.6		

Figure 1: Multiple Linear Regression Results

Regression Results

Intercept Coefficient (β_0): The intercept is 0.1406, but it is not statistically significant ($p > 0.05$), suggesting that when ROE and stock volatility are zero, the P/B ratio is not significantly different from zero.

ROE Coefficient (β_1): The coefficient for ROE is 1.7876, and it is highly statistically significant ($p < 0.001$). This means that for each unit increase in ROE, the P/B ratio is expected to increase by approximately 1.7876, holding stock volatility constant.

Volatility Coefficient (β_2): The coefficient for stock volatility is 8.6811, which is also highly statistically significant ($p < 0.001$). This suggests that for each unit increase in stock volatility, the P/B ratio is expected to increase by approximately 8.6811, holding ROE constant.

R-squared: The R-squared value is 0.133, which means that approximately 13.3% of the variation in the P/B ratio is explained by the model. This is a relatively low value, indicating that ROE and volatility cannot explain a large portion of the variability in the P/B ratio.

Discussion of Findings

The significant positive relationship between ROE and the P/B ratio suggests that more profitable companies (higher ROE) tend to have a higher P/B ratio.

The significant positive relationship between stock volatility and the P/B ratio suggests that companies with higher stock price volatility are associated with higher P/B ratios. This could be considered to be related to risk premium.

The model's low R-squared value suggests that there are many other factors not included in the model that may also be important in explaining the P/B ratio.

The non-normal distribution of the residuals could be a concern and may indicate that the model's predictions are quite biased or inefficient, so there's still much to improve. But after all, while the model has found statistically significant relationships, we can conclude that P/B ratio is positively related and probably influenced by ROE and stock volatility.

Possible Reasons

The P/B ratio can go up when a company's Return on Equity (ROE) increases because a high ROE suggests the company is using its assets efficiently to generate profits. This makes the company more attractive to investors, pushing up its stock price relative to its book value.

As for volatility, if investors expect higher returns for taking on more risk, a volatile stock with the potential for high returns could see its price increase as investors are willing to pay more, again boosting the P/B ratio. So, in a nutshell, good ROE makes the company look like a money-making champ to investors, and some investors dig the thrill of a roller-coaster stock, willing to bet more money on it, both of which can pump the P/B ratio up.

Problem 2

Manipulation

I categorize companies into ten percentile brackets every month from January 2010 to December 2023, using their price-to-book (P/B) ratio from the previous month as the basis for classification. For each month, I calculate the mean return of each percentile group, treating this average as the group's return for that month. I then compute the overall average return for all ten groups. The graph is shown at the end of Discussion.

Discussion of Findings

From the chart, it can be roughly concluded that average monthly return and P/B ratio is negatively related. The analysis is as followed:

1. Why low P/B ratio companies tend to have higher return:

Stocks that are cheap relative to their book value (i.e., have lower P/B ratios) may be the result of market underestimating the potential of these actually valuable firms, leading to higher returns as the market corrects its valuation over time, thus offering a higher average return.

This kind of company include value-oriented, financial sound companies, especially in the tradi-

tional industry. Investors may think they are too mature and old to earn more money, thus causing undervaluation while they are actually still robust and profitable.

2. Why high P/B ratio companies tend to have lower and even negative return:

The market may overvalue the actual worth of high P/B companies. Therefore, as the market corrects this overvaluation, the returns are low and even negative.

These high P/B ratio companies are generally growth-oriented, asset-light companies, mostly in technology industry. It is true that some of these companies may grow very well (like Apple in the past). However, most of these new companies' growth tend to be overvalued by the investors, finally causing low return and loss (dot com bubble is a typical example).

Graph. Group 1 has the lowest P/B, while Group 10 has the highest P/B.

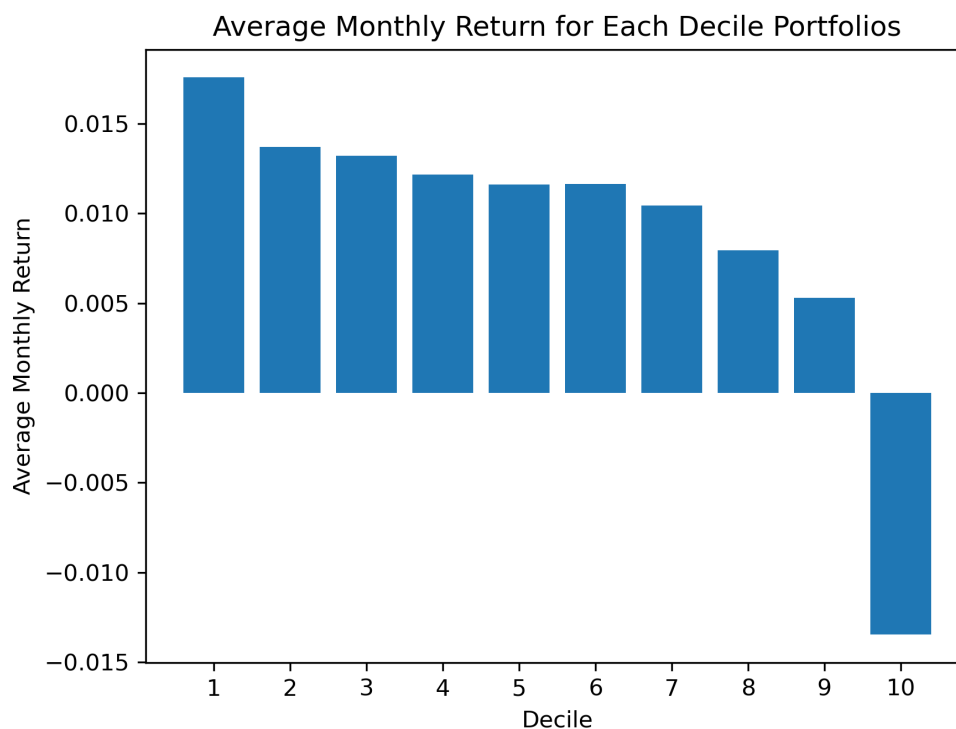


Figure 2: Average Monthly Return for Each Decile Portfolios

– END –