

Problem Set 6

Haocheng Yang

March 19, 2023

Abstract

This document describes how I clean my dataset and why do I visualize it

1 Data Cleaning: Depreciation and Amortization Comparison

In this problem set, I used WRDS database, because I find the database is more efficient for me to collect depreciation information for each company. I used python to clean and visualize my datasets. First, I used a SQL query to extract depreciation and amortization information for all companies in the US from 1986 to 2018. I excluded companies whose sales are lower than 10 million dollars and whose stock prices are lower than 1 dollar. Second, I dropped the column of sale and the column of stock price.

Third, because depreciation expenses for individual companies are not available in WRDS after the year 1998, I calculated depreciation expenses, 'XDP', by subtracting individual amortization expenses, 'am', from the sum of depreciation and amortization which is 'dp' in my dataset.

In addition, I used total assets as a scaler to standardize depreciation expenses and amortization expenses.

What's more, I used the 'groupby. mean' method to show the mean depreciation percentage for each year and also show the mean amortization percentage for each year.

Finally, I utilized matplotlib to draw a trend graph for depreciation and amortization percentages. The graph is presented in PS6a. It's pretty clear that the mean depreciation percentage for all companies is decreasing and the mean amortization percentage is increasing. This could tell us companies are allocating their attention to investment in intangible assets.

2 CAPEX and RD

In order to verify my thought in the previous part. I investigated the over-time change in capital expenditure percentage and RD expense percentage for all US companies from 1987 to 2018.

The graph is shown in PS6b Yang. We can see the mean capital expenditure investment in fixed assets is decreasing and the investment in research and development is inclining. Companies are more aware of the importance of intangible assets and self-generated intellectual assets. However, the data I use is limited to the companies that have both RD and capital expenditure. Some companies might only have CAPEX.

3 Changes in the Number of Companies

In this part, I investigated the change in the number of companies that have either CAPEX or RD. I used gvkey as the unique identifier to count the number of companies. The graph is presented in PS6c. The trend for the change in the number of companies that have CAPEX is similar to the trend of RD companies. A similar trend might be due to the macroeconomic environment. The trend is contradictory to my previous conclusion, I will do further investigation later.