Data Task

Hai Van Le

April 2024

1 Summary

For this data task, we looked at the Herfindahl-Hirschman Index (HHI), which is a measure of market concentration. This index allows us to see how competitive a market is by looking at how many firms are able to compete in a given industry area. The SIC is standard industrial classification code and NAICS is The North American Industry Classification System. As their names suggest, SIC is more globally recognized while NAICS is exclusive to Northern-American-based companies. The HHI index ranges between 0 (perfectly competitive) and 1 (full monopoly) based on how we calculate it. An HHI from 0.15 to 0.25 illustrates a concentrated market, > 0.25 - a highly concentrated market, and < 0.15 representing a competitive marketplace.

1.1 Task 1 and 2

For task 1 and 2, I noticed that calculating HHI using sales data provided a more consistent output as in figure 1. We can see that the HHI score has fallen gradually from early 2000s to 2020. One reason might be because of the way I merged BEA industry output with the individual company sales data using the NAICS-SIC crosswalk. The graph of HHI using BEA data has more extreme outliers as shown in figure 2. If I removed all the extreme outliers as in figure 3, the result was still not as informative as the first figure.

1.2 Task 3 and 4

For task 3 and 4, IHHI is measured using the sum of all squared holdings. If the index is 0, it means less concentration in ownership i.e. all shareholders hold infinitesimally small stakes. If the index is closer to 1, one share-holder owns the entire company. The Investor Herfindahl–Hirschman Index (IHHI) in figure 4 shows that common ownership has gradually increased throughout the years. Removing the outliers help us see the concentration better, as in figure 5.

1.3 Task 6

If I were to write a paper investigating the relationship between profits and common ownership, I would collect company profit data from Compustat (Wharton Data Research Services(WRDS)). I need to be careful about whether the Compustat data is accounting or economic profit and how closely related they are (do they move in line with each other so changes in one measure reflect changes in the other?.) Some financial metrics on profitability I think of using is net income, return on asset, return on equity, gross profit margin, to name a few. The data on common ownership can also be obtained from WRDS Institutional Ownership database - Thomson-Reuters Institutional Holdings (13F). Finally, I'd regress each of the financial metric on the ownership variable.

2 Conclusion

Thank you so much for the opportunity! As a full-time student, I only had 6 hours for this data task. I hope it tells a story and shows my effort.

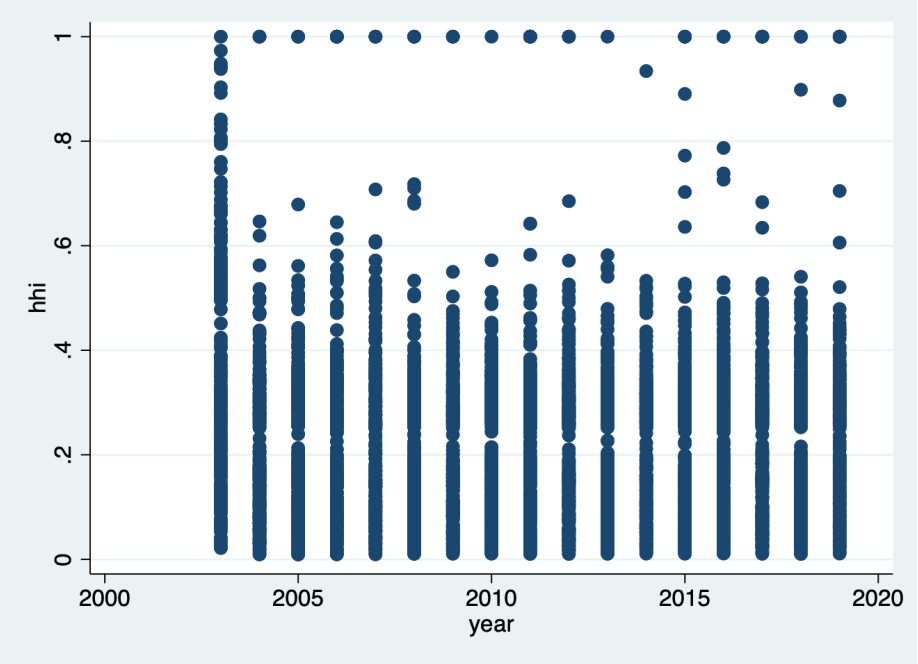

Figure 1: Task 1 - HHI using Sales data

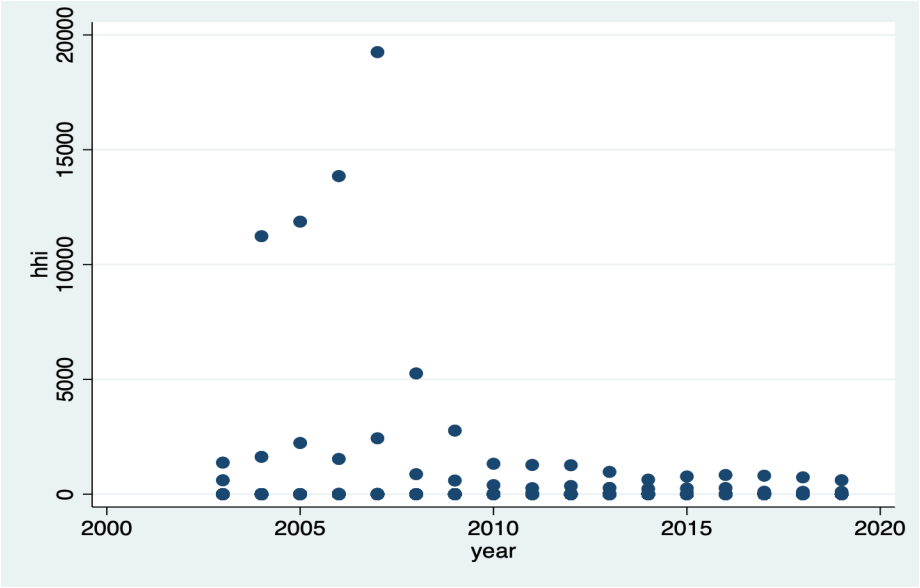

Figure 2: Task 2 - HHI using BEA data

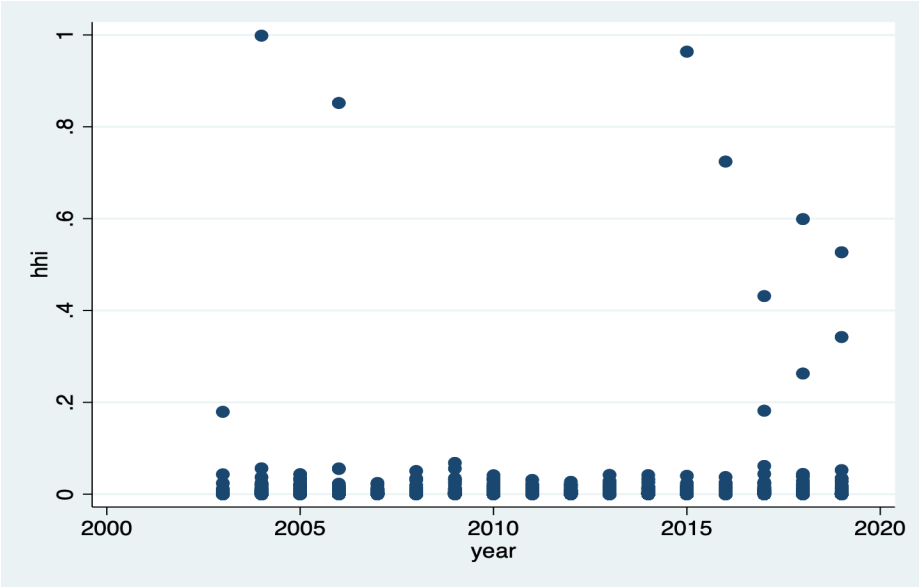

Figure 3: Task 2 - HHI using BEA data without extreme outliers

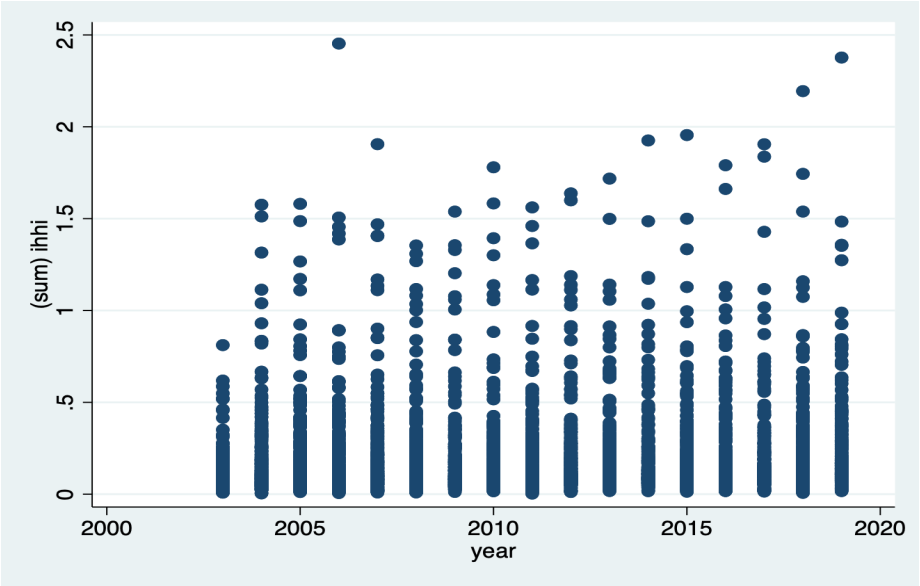

Figure 4: Task 3 - IHHI

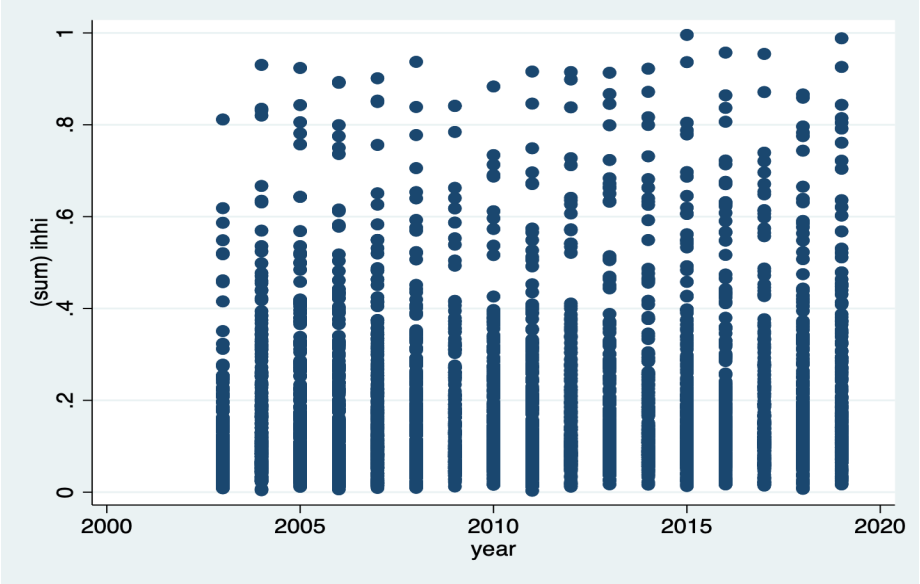

Figure 5: Task 3 - IHHI without outliers