John Schleider

Data Science Reflections on Honors Thesis

Overview

During my senior year, I researched the relationship between the risk of investing in a company and that company's monopoly power. My thesis required applying data science skills at several steps: collecting data, filtering, computing monopoly statistics at a firm level, and finally running regressions to show that risk and monopolism are negatively related across several measures of monopoly power. My economics honors thesis allowed me to grow and develop my skills in so many ways, and the project represents the culmination of all the data science knowledge I've gained at Washington and Lee.

Collect and analyze data in a reproducible and ethically responsible manner

I collected data from the Wharton Research databases, specifically the S&P Compustat that has quarterly firm-level financial data spanning decades. The dataset's reliability and accessibility to researchers makes it an excellent choice for studies on firm microeconomic variables like these.

Obtain data through searching, scraping, mining or experimental methods

The Wharton Research database did not include my key variables of interest—systematic risk and measures of monopoly power. I computed those with the data given. First, I regressed changes in stock prices on changes in the S&P 500 index to get measures of risk for several thousand firms in the sample. Then I regressed firm-level profits on revenues to compute measures of marginal profitability that approximate monopoly power.

Parse, transform and generate wide-ranging data sets for analysis

I filtered the dataset to include firms that had all the necessary minimum data for computing the monopoly statistics and systematic risk. After this, I made several subsets that I used in my core analysis and robustness checks. First, I filtered outliers away by trimming the top and bottom percentiles, as the outliers in the original data were not entirely credible. I used two standards for filtering outliers, which resulted in a weakly filtered set and a strongly filtered set, to show that the results I found are not just due to special filtering. Then, I divided firms by industry (manufacturing vs not manufacturing) and size (above $8 billion in quarterly revenue) to see if there were differing results when analyzing these groups individually.

Statistically analyze data to summarize, draw inferences and make predictions

I regressed systematic risk on measures of monopoly power and found that the relationship depends on the measure of monopoly power used. Importantly—and unlike other literature on the subject—I test several measures of monopoly power. After examining these measures for

theoretical robustness, I conclude that the best measures of power indicate a negative relationship between risk and monopolism.

Identify patterns and relationships in datasets using visualization and algorithms

When testing several measures of monopoly power, I examined whether they agreed with each other. I found that marginal profitability metrics weakly agree with each other, and that concentration measures like market share strongly agree with each other. This analysis required all the data steps above, and I conveyed my results through a correlogram for easier interpretation.

Communicate data methods and conclusions to diverse audiences

Throughout the paper, I was sure to relate my findings to both economists and practitioners, which could include antitrust regulators and company managers. The results indicate that monopolistic firms have lower risk, which furthers those firms' monopoly power by making risk a barrier to entry. The results also suggest monopolies can innovate more cheaply than competitive firms due to their lower cost of capital, which emerges from their lower risk. Though my paper includes a lot of theory about power and firms, the actual statistical methods I use are simple heteroskedastic regressions. I believe my paper enables others to follow the methods and results of the analysis, and should enrich any reader's understanding of risk and monopoly power.