

PS 239T Final Project: API Data for Chinese Companies Listed in New York

Motivation

My efforts here are geared toward my second-year paper and a broader research agenda on how certain Chinese companies use a financial structure known as the Variable Interest Entity (“VIE”) to list on the New York Stock Exchange (NYSE) and the NASDAQ in circumvention of China’s FDI regimes in areas of the economy otherwise characterized by considerable regulatory oversight. I am sorry that I am not sorry insofar as the argument I will be making will ultimately be a descriptive one rather than a causal one. For this part of the research, I am leveraging the fact that the information all publicly listed companies in the US are required by to disclose is accessible through the Security and Exchange Commission’s on-line database, EDGAR. But rather than scraping EDGAR and combing through company reports individually, I can make use of an API, EDGAR Online, maintained by Donnelley Financial LLC (<http://developer.edgar-online.com>). The task here is really just to get a descriptive sense of whether there are any systematic differences in the financial character of Chinese companies that list using the VIE and those who do not.

Data Generation

I began the process by consulting the NYSE and NASDAQ websites, which provide a complete listing by country of the companies currently trading (see <https://www.nyse.com/get-started/international/documents-reports> and <https://www.nasdaq.com/screening/companies-by-region.aspx?region=Asia&country=China>). From these sources I compiled a listing of 179 Chinese companies currently listed on either the NYSE or NASDAQ with the year of their initial listing on those markets. Because US law requires the disclosure of VIE usage, I then did a key-word search for “Variable Interest Entity” in the annual reports published on EDGAR for these companies to create a dummy variable for VIE usage.

With this list I then began digging into the API, which had several limitations. Interestingly, a get request does not always return the same number of variables; rather than having a set number of fields and leaving the string empty or entering a missing value, a get request only returns the available fields for a given company. This created headaches, as it was necessary to pin down a set of key financial metrics that were available for all countries. Secondly, a request for the latest year’s data sometimes can return data from a different fiscal year, though all but one returned data from fiscal year 2016 or 2017. Again, this was a headache for looped get requests. In the interest of keeping the number of observations as large as possible, I decided to use the data from the most recent year available for each company. Finally, the get request for 11 of the companies returned for some reason returned an empty set, leading to errors when running my get request loop. I therefore eliminated these companies from the data, leaving 168 observations.

Materials Posted

1. The R script used to for final data visualization
2. The R script used to make get requests from the API in a loop and produce a data frame
3. Two preliminary R scripts used to explore the structuring of the API and what get requests return
4. The CSV for listing the 168 observations, their initial year of listing, and the dummy variable for VIE usage.

Plots Produced

Note: these plots were all done in black and white in anticipation of using them for my paper submission (which is unlikely to be read in color) and with color blind viewers in mind. These data suggest that financially there really is no major systematic difference between the companies that use and that do not use a VIE mechanism, though the phenomenon seems to be more tightly packed around the middle 2010s.

Breakdown by VIE Usage



