

ExecuComp Low Match Rate Explanation

I checked through the ExecuComp results, and tried string matching, but the gvkey matches worked as intended.

I then tried digging into the dataset to find out the reasons for the low match rate:

- The ExecuComp dataset only has 3230 unique firms/gvkeys, of which 3157 (~98%) of them are headquartered in the US.
- ExecuComp tracks the firms currently in, and firms previously in the S&P 1500, so the US focus made sense.
- But this also means that there will be lots of non-matches to isins.xlsx.
- There are 3424 unique gvkeys in isins.xlsx, of which 1246 have a match in ExecuComp. The match rate is ~36%.
- For some gvkeys in ISIN.xlsx, there is a country tagged to it. If we only filter out gvkeys tagged with 'USA', the match rate is $1180/1750 = \sim 67\%$. This accounts for ~95% of all matches.
- On the other hand, the match rate for all other gvkeys in ISIN.xlsx is $66/1674 = \sim 4\%$. This accounts for ~5% of all matches.
- The low match rate is thus due to (1) ExecuComp having little data on firms outside the US (and firms in general), and (2) isins.xlsx containing many firms outside the US.