"Common Ownership, Competition, and Top Management Incentives"

README For Data Archive

Miguel Antón, IESE Business School, manton@iese.edu
Florian Ederer, Yale School of Management, florian.ederer@yale.edu
Mireia Giné, IESE Business School, mgine@iese.edu
Martin Schmalz, University of Oxford Saïd Business School, martin.schmalz@sbs.ox.ac.uk

1. The Structure of the Data Archive

In the folder, we provide details for obtaining, cleaning, and constructing datasets used for the paper "Common Ownership, Competition, and Top Management Incentives" and the code used for analysis in the paper. Please refer to the paper and a separate Online Appendix for additional details related to our analysis.

The folder includes several key components to which we refer below.

----- jpe_data_archive
|--- README
|--- do_analysis_main.do

Master Code Files

- do_analysis_main.do reproduces all tables and figures in the paper.

The datasets used in this paper are proprietary and therefore data files are not included in the archive. However, they are available for purchase from the sources listed below.

2. Main Data Source and Sample Construction

All tables and figures in the paper are constructed using a dataset which is based on the combination of several datasets from: COMPUSTAT, ExecuCOMP, CRSP, Thomson Reuters, Hoberg-Phillips FIC Industry Classification Codes, and Schwartz-Ziv and Volkova (2021).

2.1 Executive Compensation Data

The main dependent variable of the paper is wealth-performance sensitivity (WPS). As explained in Section 5.1 of the paper, the benchmark measure is WPS from Edmans, Gabaix, and Landier (2009) which we denote by WPS EGL. It is defined as the dollar change in CEO wealth for a 100 percentage point change in firm value divided by annual pay. The second measure is WPS from Jensen and Murphy (1990), WPS JM, defined as the change in CEO wealth for a $1,000 increase in firm value. The third measure is WPS from Hall and Liebman (1998), WPS HL, defined as the dollar change in CEO wealth for a percentage change in firm value. All three measures are studied and computed in Edmans et al. (2009), as BI, BII, and BIII respectively.

The data for the WPS EGL measures are available from Alex Edmans's website.[1] We thank Alex for sharing with us the code that produced the WPS EGL as well as the other measures. We have updated the code in two ways. First, we have updated their sample up to 2019, following each of the steps described in their paper, and using ExecuComp as a source. Secondly, we have computed the WPS measures for all executives present in ExecuComp, and not only for CEOs. This allows us to complete the analysis for a broader set of executives (Table 8).

The outcome panel *master_compensation* contains WPS measures at the *year-firm-executive* level. The WPS measures are winsorized at 1% to reduce the impact of outliers in the sample.

2.2 Ownership Data

We collect ownership data from three sources:

- Thomson Reuters (institutional ownership in 13F).
- SEC 13F filings.
- Schwartz-Ziv and Volkova (2021), which includes blockholdings in 13D and 13G.

Data Processing

We first augment the Thomson Reuters 13F data by scraping SEC 13F filings following Ben-David et al. (2020), which resolves the issues of stale and omitted institutional reports, excluded securities, and missing holdings from 2000 onwards. The code from Ben-David et al. (2020) can be accessed through WRDS.[2] We also follow the standard procedure in the literature, adjusting the holdings for stock splits, and using CRSP cusip-permno merging table to link with the other databases.

We then manually aggregate holdings at the manager level (e.g., aggregating the different funds under Capital Group, INVESCO, Fidelity, etc.), using the name of the fund. We complement this information and check its validity with the sample of aggregation from Azar, Schmalz, and Tecu (2018).

Finally, we complement the institutional holdings with non-institutional holdings (both from directors and from individuals), following Schwartz-Ziv and Volkova (2021), using data provided by the authors. This means that if there are large stakes owned by non-executive directors, they are part of the analysis. Companies have on average 2.8 blocks (ownership above 5%) of institutional investors, and 1.7 blocks of non-institutional investors. Each of those 2.8 blocks of

[1] See https://alexedmans.com/data/ for more information.
[2] An extensive research note on this issues is available here https://wrds-www.wharton.upenn.edu/documents/752/Research_Note_-Thomson_S34_Data_Issues_mldAsdi.pdf

institutional investors have an average of 8.4% ownership, and each of the 1.7 blocks of non-institutional investors hold on average 12.8%. If these directors have large stakes in other firms, these stakes are likewise measured as blocks. We do not make additional assumptions about the ownership of blockholders in other firms. Implicitly, our assumption is that if no ownership is reported in either 13D, 13F, or 13G forms, then no such ownership exists. That is, if a non-13F blockholder in one firm is not a blockholder in a related firm, the ownership stake in the related firm is assumed to be zero. In short, we do not make assumptions that would override the regulatory data.

The outcome panel *master_ownership* contains holdings at the *year-firm-manager* level. Because the majority of the companies have fiscal year end in December, we restrict our analysis to yearly ownership based on December holdings as is common in the literature.

2.3 Common Ownership

Once we have the clean and complete ownership panel data *master_ownership*, we proceed to compute the different measures of common ownership, following the steps described in the paper in Section 5.2. The steps to compute each of the measures are carefully defined in that section of the paper.

The outcome panel *master_common_ownership* contains all common ownership measures equal- and value-weighted by *firm* and *year*: kappas (and its components cosine similarity and IHHI), Top5, AP, HJL, MHHID, and MHHID 1/N.

2.4 Industry Definitions

We use three main industry classifications: SIC from Compustat (sich, historic SIC[3]), SIC from CRSP (siccd, historic SIC [4]), and FIC from Hoberg-Phillips (icode400 from fic_data_extended.txt[5]). All three industry measures are at the four-digit industry level.

To carry out the sensitivity analysis to coarser industry definitions (shown in Appendix Table C8), we compute all measures of common ownership again using industry classifications at the three-digit level (i.e., taking the first three digits from each of the industry classification codes).

---

[3] These definitions are obtained from comp.funda, the Fundamentals Annual database of Compustat.
[4] Similarly, the CRSP industry definitions are available from crsp.msenames.
[5] Finally, the HP industry definitions are available here https://hobergphillips.tuck.dartmouth.edu/industryclass.htm

The outcome panel *master_industry* contains all industry classifications measures (SIC Comp, SIC CRSP and Hoberg-Phillips, both 4- and 3-digit measures) by *firm* and *year.*

2.5 Merging the Main Dataset

We merge *master_compensation* with *master_common_ownership* and *master_industry* by firm (*permno*) and year.

The outcome of the merge is the panel *master_main* at the *year-firm-executive* level.

This dataset *master_main* is used to compute the main empirical tables: Tables 4, 5, 6, and 7, restricted to only CEOs.

Table 8 uses *master_main* without the restriction to CEOs where we keep information on all executives.

2.6 Difference-in-Differences using S&P500 additions: Data and Sample Construction

The data sources to build the diff-in-diff analysis are: Compustat comp.idxcst_his (where we have all index components), and the *master_main* –the combination of the databases described in 2.5 of this document. We restrict the sample to S&P500 components—gvkeyx='000003'. This sample allows us to extract all inclusions in the S&P500. Some of these inclusions are promotions (stocks that were included in the S&P400 or S&P600), with the gvkeyx='024248' and '030824' respectively. We restrict our sample to the competitors of the true inclusions, rather than promotions which we exclude.

Once we have the sample of true inclusions, we merge this information with the industry classification for every year. We identify as *TREATED* firms those that are in the same SIC-4 industry as the added firm and that are already members of the S&P 500. The *CONTROL* firms are those firms that are in the S&P 500 but not in the same SIC-4 industry as the added firm and that do not experience an inclusion in their industry in the same year of the inclusion event. The estimation is run on a sample with five pre- and five post-years of the event treatment year. All information on the detailed analysis of the diff-in-diff can be found in Section 6.3 of the paper.

The outcome panel *master_did* is at the *firm-year-executive* level, and is used to compute Table 9, Table C9, and Figure 4.

In Table C10 we run changes in Common Ownership on the variable *Treat.* To do so we build a new panel *master_did_change* based on *master_did*, where we only take the common ownership

measure as of the year of the addition *y=0* (of all years around addition, between -5 and +5), and subtract the common ownership measure as of the year before the addition *y*=-1.

Finally, to compute Figure C1, we build a dataset *master_did_additions* of the additions themselves, not the competitors. We proceed as indicated above, but instead of keeping the competitors, we keep the additions themselves, with the same structure: *firm-year-executive* level with years around addition (from *y=-5* to *y=+5*).

For simplicity and speed in our computations in the diff-in-diff analysis, we have one dataset for each combination of common ownership measure and industry classification. For example, for kappa with SIC CRSP industry classification, we have *master_did**sich_crsp4vwk_*,** referred to in the code as a macro variable: *master_did `SIC' `CO'.*

3. Summary of Datasets for Analysis

As a summary, the datasets used for the analysis and the corresponding tables and figures are the following:

- **Table 1:** This table corresponds to the Summary of the Model Setup (end of Section 2.4).
- **Table 2:** Panel A comes from S&P Capital IQ (Q2 2016), and Panel B is taken from Azar et al. (2018).
- **Table 3:** This table corresponds to the theoretical predictions and empirical evidence (in sub-Section 3.2.2)
- **Tables 4, 5, 6, 7, C1, C2, C3, C4, C5, C6, C7, C8:** main analysis with panel *master_main* restricted to only CEOs
- **Table 8:** with panel *master_main* with all executives.
- **Table 9, C9:** with panel *master_did.*
- **Table C10:** with panel *master_did_change.*
- **Figure 1:** model timeline (Section 2.1).
- **Figure 2:** model setup (Section 3.2).
- **Figure 3:** distributions of Kappa and Cosine Similarity (coded in SAS) with panel *master_did_change.*
- **Figure 4:** pre- and post-trends of basic diff-in-diff with panel *master_did.*
- **Figure C1:** pre- and post-trends for additions themselves with panel *master_did_additions.*