Replication code for "The Proxy Advisory Industry: Influencing and Being Influenced"
Journal of Financial Economics

Chong Shu

Overview

The code in this replication package consists of five pseudo datasets and two replication files: replication_tablse.do (using stata) for tables and replication_figures.py (using python) for figures.

Data Availability

- ISS Voting Analytics can be accessed via WRDS (subscription required)
- CRSP Mutual Fund can be accessed via WRDS (subscription required)
- Glass Lewis recommendations are obtained via a Public Records Law (PRL) request directed to North Carolina Department of State Treasurer.
- Mutual funds' statutory prospectuses are obtained via the SEC website.
- Mutual funds' N-PX forms for vote disclosure are obtained via the SEC website.

(I certify that the author of the manuscript has legitimate access to and permission to use the data used in this manuscript.)

Pseudo Dataset list:

- *main_data_pseudo.dta:* Mutual fund votes.
- *did_pseudo.dta:* Mutual fund voting systems.
- *robo_pseudo.dta:* Mutual fund robo-voting flag.
- *PA_change_pseudo.dta:* ISS and Glass Lewis recommendations.
- *market_share_pseudo:* Market shares of ISS and Glass Lewis customers.

According to JFE's data disclosure policy, the following variables are randomly generated because they originate from third-party vendors, and I do not have permission to disclose them.

- *contentious*: whether ISS opposes management (source: ISS)
- *contentious_GL*: whether GL opposes management (source: ISS)
- itemonagendaid: item on agenda id (source: ISS)
- *asset_log, num_log, age, index_fund_share, inst_fund_share, esg_share, exp_ratio, mgmt_fee*: control variables (source: CRSP)
- *agreeISS, agreeGL:* whether the fund votes in the same direction as ISS or GL (source: ISS and PRL)

- *iss_for , GL_for:* ISS and GL recommendations (source: ISS and PRL)
- *agreeISS_contentious_management, agreeGL_contentious_management, agreeISS_contentious_shareholder, agreeGL_contentious_shareholder:* Fraction of contentious votes that agree with ISS or GL recommendations (source: ISS)
- *ISS_prospectus*: whether fund prospectus mentions ISS (source: ISS, SEC).
- *GL_prospectus*: whether fund prospectus mentions GL (source: ISS, SEC).
- *iss_change, gl_change*: whether ISS or GL changes recommendations (source: ISS and FOIA)
- disagreeISS_last, disagreeGL_last: % of investors' past votes disagreed with ISS or Glass Lewis (source: ISS)
- *mgmt._cd*: fund management identifier (source: CRSP)
- *tna*: total net assets (source: CRSP)

Computational Requirements

Stata (code was last run with version 15)

- reghdfe

Python (3.0+)

- matplotlib
- pandas
- numpy

Memory and Runtime Requirements

Approximate time needed to reproduce the analyses on a standard (2024) desktop machine: <10 minutes.

The code was last run on an "Intel(R) Core(TM) i7-10810U CPU @ 1.10GHz 1.61 GHz" laptop with Windows and 32.0 GB RAM.

Description of Programs/Code

replication_table.do

replication_figures.py

List of Tables and Programs

	Replication file location	Data set needed
Table 2	*replication_tables.do*	*main_data_pseudo.dta*
Table 3	*replication_tables.do*	*main_data_pseudo.dta*
Table 4	*replication_tables.do*	*did_pseudo.dta*
Table 6	*replication_tables.do*	*robo_pseudo.dta*
Table 7	*replication_tables.do*	*PA_change_pseudo.dta*
Table 8	*replication_tables.do*	*PA_change_pseudo.dta*
Table 9	*replication_tables.do*	*PA_change_pseudo.dta*
Figure 3	*replication_figures.py*	*market_share_pseudo.dta*
Figure 4	*replication_figures.py*	*main_data_pseudo.dta*
Figure 5	*replication_figures.py*	*main_data_pseudo.dta*

Third Party Data

(JFE Data and Code Sharing Policy: In the case of administrative data or third-party data whose owners can be disclosed, the authors must submit a detailed description of how the raw data were obtained or generated, including data sources and the specific date(s) on which the data were downloaded or obtained. Authors also have to provide details to outside researchers on how to obtain the data. When the data provider allows, observation identifiers in the raw data should be disclosed.)

The initial sample consists of the complete mutual fund voting records spanning from 2007 to 2021. Since 2003, mutual funds have been required to disclose their full voting records annually on Form N-PX to the SEC every August. I collect these forms directly from the SEC's EDGAR website. Each N-PX form is subsequently linked to the ISS Voting Analytics database using its accession number, a unique identifier for EDGAR submissions. ISS Voting Analytics compiles a dataset of mutual fund votes from these N-PX forms, providing details on each proposal's final voting outcome and ISS's recommendations. Because accession numbers only appear in the Voting Analytics dataset starting from 2007, my sample is confined to the period between 2007 and 2021. Votes pertaining to procedural matters, such as meeting adjournments or auditor ratifications, and say-on-pay proposal frequencies are omitted. The final sample comprises 140 million votes from 23,264 mutual funds.

While the Voting Analytics database provides recommendations from ISS, those from Glass Lewis are not publicly available. I obtain Glass Lewis's recommendations for the period between 2008 and 2021 through a Public Records Law request directed to a large public pension fund. I request the name of the pension fund's proxy advisor, as well as the recommendations received from this advisor. I then match these recommendations with the main voting dataset using

company names, meeting dates, and item numbers. The Public Records Law request enables me to locate Glass Lewis's recommendations for 3,255 companies, covering approximately 80% of the total assets for companies included in my main dataset.

I obtain mutual fund characteristics from the CRSP Mutual Fund Database. These characteristics are merged with the ISS voting dataset using CIK numbers, which are unique ten-digit identifiers assigned to filers by the SEC. I gather each mutual fund's CIK number from the header files of its N-PX forms. In line with previous studies, I conduct my analyses at the fund-family level. Votes that are not unanimous across funds within a family are omitted. A fund family is defined as a unique fund management entity in the CRSP dataset, as ISS's definition of fund management has inconsistent fund identifiers across years and occasionally misrepresents a fund family. After this procedure, the aggregated dataset contains 17.5 million votes from 809 fund families, covering 455,433 proposals from 8,842 companies spanning 2007 to 2021. To reduce verbosity, I sometimes refer to a fund family simply as a fund in the remainder of the paper.

Each fund's statutory prospectus is collected from EDGAR to determine if it references contractual relationships with proxy advisors. Information about the mutual funds' portfolio holdings is obtained from the Thomson Reuters s12 database. I use the MFLINKS table from WRDS to connect the portfolio holding data with the main dataset. Finally, information on mutual funds' ideology preferences is provided by Bolton (2020).