

The Impact of Financial Reporting Mandates on Labor Unions

Data Description Sheet and ReadMe File

1 Data Description Sheet

1. A description of which author(s) handled the data and conducted the analyses.

All authors handled the data and conducted the analyses.

2. A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

A listing of all of the raw data sources is provided in the accompanying README file, with detailed information on download dates and sources.

3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results). In particular, the authors should indicate if an organization or data provider imposes restrictions on the publication of the results, has not given the authors full control of the relevant data, requires that the results have to be reviewed or approved prior to public release of the paper or publication. This information should be provided to the editors upon submission.

The data used in this draft of the paper is all publicly available or are commercially licensed data.

4. A complete description of the steps necessary to download, obtain or collect as well as process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

See the README file below, the accompanying code, and the online appendix for relevant documentation.

5. After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.

See the README file below, the accompanying code, and the online appendix for relevant documentation. Several of our datasets require manual matching and/or collection, which we disclose in the accompanying README file.

6. The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the manuscript. A brief description that enables other researchers to understand and run the code should be provided. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, and which commands were used in the analysis, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of disclosing the proprietary portion of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.

See the README file below, the accompanying code, and the online appendix for relevant documentation.

7. A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.

See the README file below, the accompanying code, and the online appendix for relevant documentation.

8. **An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.**

The data and programs will be maintained by the corresponding author for at least six years, consistent with the NSF guidelines.

2 Overview and Data Structure of Replication Package

- This file describes the structure of the directory to replicate the results from **The Impact of Financial Reporting Mandates on Labor Unions**
- The replication package contains three main folders:
 - **01 - Code:** Contains all of the code to transform the data from its raw form through to the analysis.
 - **02 - Data:** Contains all of the raw, intermediate, and output data files.¹
 - **03 - Results:** Stores log files outputted from the code and directories where tables/figures would be outputted.

3 Description of Folders in Replication Package

Folder 1: 01 - Code

The “01 - Code” folder is organized such that the code files for different sets of analyses (and datasets) are contained within their respective folders. To reproduce the results from the main text in their entirety, you can use master.do and the notes contained within the master do-file (which requires running a couple of separate programs using Python). Note that any table/figure that is not mentioned below does not require code to create (e.g., Figure 3 is just an illustration of the union process made by the authors). Below is a brief description of the files and subfolders in the “01 - Code” folder, roughly in the order they appear in the master do file:

- **master.do** — This is the master do-file. If attempting to reproduce all the results from the paper, please run this do-file only and it will contain all code necessary to reproduce all Tables/Figures in the main text. To run on your machine, please make sure to change pathnames in the master.do file.
- **IFRMLU01_Financial.do** — This do-file creates the main union financial reporting database (union_fr_combined.dta). This do-file also creates the histogram in Panel A of Figure 5.
- **IFRMLU02_Charges.do** — This do-file creates the main charges database used in the analysis (charged_charges_unionyear.dta). This do-file also creates the descriptives in Table 1.
- **IFRMLU03_OpenSecrets_VariableConstruction.do** — This do-file creates the main political influence datasets used to create Panels B - D in Table 3. Since the datasets require special formatting, they are not directly created from Stata in this

¹We cannot include data in this replication package because most of the data can only be accessed via agreements with the data vendor. Nevertheless, we have structured the directories in the replication to match the directory structure required by the code. Below, we also provide detailed descriptions of the content of each raw, intermediate, and output data file that would be in each of these directories.

do-file. However, this do-file produces the datasets (and imposes all conditions needed) to reproduce the numbers in the panels.²

- **IFRMLU04_coding_CLs.do** — This do-file creates the descriptives in Panel A of Table 3.
- **IFRMLU05_lobbying.do** — This do-file computes the descriptives in Panel B of Table 3.
- **IFRMLU06_first_stage.do** — This do-file runs the “first-stage” analysis, which assess the impact of the threshold on LM-2 filing, and creates Table 4.
- **IFRMLU07_election_outcomes_RD.do** — This do-file runs the RD election analysis and creates Panels A and B of Table 5.
- **IFRMLU08_charges_RD.do** — This do-file runs the charges analysis and creates Panels A and B of Table 6.
- **IFRMLU09_administrative_expenses.do** — This do-file runs the analysis on professional fees and disbursements and creates Table 7.
- **IFRMLU10_labor_consultants.do** — This do-file runs the mediation analysis of labor consultant presence on election outcomes and creates Panels A and C of Table 8.
- **IFRMLU11_labor_consultant_compensation.do** — This do-file runs the mediation analysis of labor consultant compensation on election outcomes and creates Panel B of Table 8.
- **IFRMLU12_election_outcomes_DID.do** — This do-file runs the difference-in-differences analysis on unions’ election outcomes and creates Panels A and B of Table 9.
- **IFRMLU13_charges_DID.do** — This do-file runs the difference-in-differences analysis on unions’ charges and creates Panels A and B of Table 10.
- **IFRMLU14_state_level_analysis.do** — This do-file creates the aggregate treatment measures, runs all aggregate state-level analyses and creates Panels A, B, and C of Table 11.
- **IFRMLU15_geographic_maps.do** — This do-file creates descriptive data needed to produce the state-level and county-level maps in Figure 2.
- **county_map.py** — This .py file generates the actual maps from Figure 2.
- **Bunching** — The only code and data contained separately are the programs used to conduct the “bunching” estimation presented in Figure 6. All of the relevant code is contained in this folder, including a separate README file. Please see the documentation in the folder for more detail.

²In particular, for the code that generates the datasets needed for Panel C of Table 3, please see the section in the do-file labeled 2c(2) (the “Department of Labor” lobbying statistics and 2e(2) (the “All Labor-Related Issues” lobbying statistics. For the dataset needed for Panel D of Table 3, please see section 2g.

Folder 2: “02 - Data”

Below is a brief description of the files and subfolders in the “02 - Data” folder:

- identifiers:
 - This folder contains four identifier files that list the unique f_nums (i.e., the identifier from the financial reporting database) that are included in each of our analysis samples. One file is included for our RD elections sample, RD charges sample, DID elections sample, and DID charges sample. These identifiers represent the set of unique unions in the pre-regression sample for each set of analyses.
- orig:
 - Charges:
 1. **Raw_Data/cases_XXXX.csv**: These files are the raw data files containing charges and petitions obtained from the NLRB’s [Charge and Petitions Database](#), separately for each state (using their state filter). The data for this project was downloaded in May 2022.
 - Elections:
 1. **Recent/recent_election_results_all_dates_20220526171421.csv**: This file contains all of the recent union election results (i.e., we describe these in the paper as maintaining comprehensive coverage post-2008). The data was obtained from the NLRB’s recent election data at [this link](#). These data were downloaded in May 2022.
 2. **CATS/CATS_1999-2011_R_Cases**: This file contains all of the election results housed under the NLRB’s previous case tracking system, CATS, which offered coverage from 1999 - 2011. The data was obtained from the NLRB’s archived case data, which is described [here](#) and housed via data.gov. Included in this folder is code to convert the .xml files (i.e., the format that is available for download) to csv files. The data for this project was downloaded in May 2022.
 - Labor_Consultants:
 1. **Data_Collection_List.xlsx**: This dataset contains all of the records from LM-20s filed with the OLMS, which are available via [this link](#). These reports were all downloaded from this database, and then information was manually extracted from those forms. The information extracted includes employer name, union name (if any), the consulting firm’s name, state, and year, which were all manually collected by a research assistant from these reports. The reports were downloaded and information was extracted in June 2022.
 2. **Data_Collection_List_LM21_Collected.xlsx**: This dataset contains all of the records from LM-21s filed with the OLMS, which are available via [this link](#). These reports were all downloaded from this database, and then information was manually extracted from those forms. The information extracted includes employer name, union name (if any), the consulting firm’s name, payment and disbursement amounts, state, and year were all

manually collected by a research assistant from these reports. The reports were downloaded and information was extracted in June 2022.

– Maps:

1. **County_Business_Pattern/cbp.dta**: This data contains employment data from the County Business Patterns (CBP) database. This data was downloaded in July 2022.
2. **County_Business_Pattern/zip_county**: This folder contains data from the County Business Patterns (CBP) database with the share of residents, businesses, etc. for each ZIP code within a county. This data was downloaded in July 2022.

– Political_Influence

1. **Campaign_Finance/CampaignFinXX**: These data come from OpenSecret's Bulk Campaign Finance database, which can be accessed [here](#). These data were downloaded in February 2024.
2. **Lobbying**: This directory contains the following rep_lob_agency.txt, rep_lob_bills.txt, rep_lob_issue_lmrda.txt, rep_lob_lobbying.txt, rep_lob_lobbyist.txt. These data come from OpenSecret's Bulk Lobbying database, which can be accessed [here](#). These data were downloaded in February 2024.
3. **Comment Letters/comment_letters_data.csv**: This file contains manually coded comment letters (which were originally obtained from [Regulations.gov](#)). The paper contains the Docket ID for the rules that comment letters have been collected for; please search up these Docket IDs to access the rule and all associated comments. This data was collected in February 2024.
4. **candidates_political_affiliation.xlsx**: This dataset contains candidates' political affiliations, which were obtained via Google searches to create this dataset.

– State_Level_Analysis:

1. **compustat_1995_2016.dta**: This data contains the entire Compustat database, as downloaded from WRDS, covering the years 1995 to 2016. This data was downloaded in February 2024.
2. **CPS_ORG_Data/cepr_org_XXXX**: This data contains the entire Current Population Survey ORG database, downloaded from the [CEPR repository](#). This data was downloaded in November 2022.
3. **popest_annual.xls**: This data contains data on state-level population from 2000 onward, downloaded from [the United States Census Bureau](#). This data was downloaded in November 2022.
4. **reviews_external_unc.csv**: This data contains data on Glassdoor reviews, with accompanying state data. This data was obtained in October 2023.
5. **state_GDP.xlsx**: This data contains data on state-level GDP covering 2000 - 2021 from the Bureau of Economic Analysis' [database](#). This data was obtained in November 2022.
6. **union_membership_2000_2020.xlsx**: This data contains data on state-level union membership covering 2000 - 2020 from the [Union Membership and Coverage Database](#). This data was obtained in November 2022.

7. **union_state_wages:** This data contains data on state-level wages by occupation covering 2000 - 2020 from the [BLS' OEWS database](#). This data was obtained in November 2022.
- Union_FR:
 1. **Raw_Data/XXXX:** This dataset contains all of the union financial reporting data from 2000 onward. These data can be obtained from [this link](#) via the OLMS. This data was downloaded in May 2022.
 - intermediate:
 - Charges:
 1. **union_crosswalk:** These files (union_name_combined_2.xlsx and union_names_combined_1_fnum_filled.xls) contain the crosswalk from the charges dataset to the financial reporting database. Unions were manually matched, using the process described in the main text of the paper (which includes the name and local number of the union).
 - Elections:
 1. **CaseNumber_fnum_crosswalk.dta:** This file contains the crosswalk from the union election databases and the financial reporting database. Unions were manually matched, using the process described in the main text of the paper (which includes the name and local number of the union).
 2. **combined_union_election_data.xlsx:** The raw records from the CATS and Recent election databases appended to each other, with an additional variable “dataset” that indicates whether the data stem from the CATS or the recent NLRB database.
 - Political_Influence:
 1. **Campaign Finance/committee_matched_checked.xls:** CID (candidate ids) and candidate names.
 2. **Campaign Finance/candidate_committee_unique.dta:** Committee-level IDs for relevant candidates' committees (receiving money).
 3. **Campaign Finance/candidates_all.csv:** Candidates table that gives us the list of unique candidate names and their personal IDs.
 4. **Campaign Finance/committee_matched_checked_unique.csv:** Person-level IDs for relevant candidates (receiving money).
 5. **Campaign Finance/individual.dta:** donation of individuals with employers as labor unions.
 6. **Campaign Finance/pac_labor_for_matching.dta:** CmteID (pac id) and PACShort (pac name).
 7. **Campaign Finance/pac_labor_unique.dta:** Committee-level IDs for relevant labor union pac (donating money).
 8. **Campaign Finance/pac_to_candidate.dta:** unions' pac to relevant candidates.
 9. **Campaign Finance/pac_to_pac.dta:** unions' pac to relevant candidates' pac.

10. **Lobbying/all_agencies.dta**: Government agencies targeted by each lobbying transaction.
 11. **Lobbying/business_lobby_allagencies.dta**: Agency-year-level lobbying amounts and rank by lobbying from business entities.
 12. **Lobbying/business_lobby_allissues.dta**: Issue-year-level lobbying amounts and rank by lobbying from business entities.
 13. **Lobbying/business_lobby.dta**: Union-year-level lobbying amounts by business entities on labor-related issues.
 14. **Lobbying/labor_lobby_allagencies.dta**: Agency-year-level lobbying amounts and rank by lobbying from labor unions.
 15. **Lobbying/labor_lobby_allissues.dta**: Issue-year-level lobbying amounts and rank by lobbying from labor unions.
 16. **Lobbying/labor_lobby_labor_id.dta**: Transaction IDs for labor unions' lobbying on labor issues.
 17. **Lobbying/labor_lobby.dta**: Union-year-level lobbying amounts by labor unions on labor-related issues.
 18. **Lobbying/lob_issue_all.dta**: Lobbying issues corresponding to each lobbying transaction.
 19. **Lobbying/lob_issue_id.dta**: Specific lobbying issue IDs within labor issues (more granular than issue-level classification and used to merge with lobbying bills).
 20. **Lobbying/lob_issue.dta**: Lobbying transaction IDs of labor-related issues.
- State_Level_Analysis:
 1. **aggregate_firm_state_counts.dta**: This dataset contains the count of different state names/acronyms in firms' 10-K reports using the approach in Garcia & Norli (2012).
 2. **state_acronyms.dta**: Crosswalk between state names and state acronyms.
 - Maps:
 1. **County_Business_Pattern**: This folder contains two files, `county_zip_crosswalk.dta` and `zip_county_crosswalk.dta`, which contain crosswalks between county codes and zip codes.
- output:
 - Charges:
 1. **charged_charges.dta**: This dataset contains, on the charge-level, linked records between the charge record and the union's firm from the financial reporting database.
 2. **charged_charges_unionyear.dta**: The main analysis file for charges filed against unions, with the number of charges filed against each union in each year (conditional on there being at least one charge filed).
 - Maps:
 1. **union_county_emp_map.csv**: This data contains the county-level employment data used to plot Panel C of Figure 2.

2. **union_county_ratio_map.csv**: This data contains the county-level unionization data used to plot Panel D of Figure 2.
 3. **union_state_map_both.csv**: This data contains the state-level unionization data from the financial reporting database and the BLS' data used to plot Panels A and B of Figure 2.
- State_Level_Analysis:
 1. **combined_BLS_OEWS_wages.dta**: The main analysis file containing BLS' OEWS wage data that analysis is performed on.
 2. **glassdoor_average_firm_year.dta**: The main analysis file containing Glassdoor ratings data that analysis is performed on.
 3. **union_CPS_ORG_wages.dta**: The main analysis file containing the Current Population Survey's (CPS) ORG wage data that analysis is performed on.
 - Union_FR:
 1. **union_fr_combined.dta**: The main financial reporting database that analysis is conducted on.
 2. **union_geo_zip.dta**: Zip-code level data on the number of union members and the share of union members belonging to unions above the regulatory reporting threshold.

Folder 3: “03 - Results”

The “03 - Results” folder would contain all output generated by the code files in “01 - Code.” The folder is organized into Figures, Tables, and Log files (which are created for each do-file, but can also be generated all at once using the master.do file).³ Below is a brief description of the subfolders in the “03 - Results” folder:

- **Figures** — The figures from the main text that are outputted by the code base.
- **Tables** — The tables from the main text that are outputted by the code base.⁴
- **Logs** — The log-files produced from each do-file. Full log files of each code file in this replication package is provided in this folder, including:
 1. IFRMLU01_Financial.log
 2. IFRMLU02_Charges.do.log
 3. IFRMLU03_OpenSecrets_VariableConstruction.log
 4. IFRMLU04_coding_CLs.log
 5. IFRMLU05_lobbying.log
 6. IFRMLU06_first_stage.log

³For the purposes of this replication package, only the log files are produced as output, but all Figures/Tables in the main text can be produced by running the code base.

⁴The only tables not directly outputted below are Tables 2 and 3, which require special formatting. The summary statistics shown in Panels A, B, and C of Table 2 are shown at the end of each log file. The summary statistics for Table 3 are produced by multiple do-files, as described in prior sections of this README file.

7. IFRMLU07_election_outcomes_RD.log
8. IFRMLU08_charges_RD.log
9. IFRMLU09_administrative_expenses.log
10. IFRMLU10_labor_consultants.log
11. IFRMLU11_labor_consultant_compensation.log
12. IFRMLU12_election_outcomes_DID.log
13. IFRMLU13_charges_DID.log
14. IFRMLU14_state_level_analysis.log
15. IFRMLU15_geographic_maps.log