

Journal of Accounting Research Data Policy

Data Description Sheet for:

Partisanship, Information, and Punishment for Misconduct at Work

By Vivek Pandey, Joanna S. Wu, and Yuanzhe Zhang

August 2026

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

Yuanzhe Zhang handled the data and conducted the analyses.

Vivek Pandey and Joanna Wu conceptualized the tests and 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 can vouch for the stated source of the raw data.*

We obtained data from following sources:

Political contributions: We obtained political contributions data from DIME database at Stanford in July 2024.

Financial advisers:

- We obtained the 2007 to 2015 financial adviser misconduct data from Mark Egan, Gregor Matvos, and Amit Seru via email in July 2024.
- We obtained “detailed reports” for each adviser from Investment Adviser Public Disclosure in May 2025 (website: <https://adviserinfo.sec.gov/>).
- We obtained race and ethnicity prediction code from Rajashekar Chintalapati, Suriyan Laohaprapanon, and Gaurav Sood (website: <https://pypi.org/project/ethnicolr/>).
- We obtained location information from U.S. Census Gazetteer in July 2024 (website: <https://www.census.gov/geographies/reference-files.html>)

Advisory firm characteristics:

- We obtained firm restatement data from Audit Analytics Database in April 2025.
- We obtained firm internal control weakness data from Audit Analytics and Compustat Database in April 2025.
- We obtained firm management guidance data from IBES Database in April 2025.

- We obtained firm headquarter information from SEC Information About Registered Investment Advisers and Exempt Reporting Advisers in Jan. 2026.
(website: <https://www.sec.gov/data-research/sec-markets-data/information-about-registered-investment-advisers-exempt-reporting-advisers>).
- We obtained firm level voter registration data from Politics at Work in Jan. 2026
(Website: <https://politicsatwork.org/>).
- We obtained data on whether the firm hires independent public accountant to create an internal control report from historical Form ADV Part 1 data files in March 2026
(Website: <https://www.sec.gov/files/adv-filing-data-20001019-20111104.zip>).

Local political ideology: We obtained State and County Presidential Election Returns from the MIT Election Lab in Sept. 2024.

All authors can vouch for the sources of the above datasets.

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, and any restrictions imposed by the organization on the authors). 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 must be reviewed or approved prior to public release of the paper or publication.*

The datasets used in our study were not obtained from any organization on a proprietary basis. All data used in the paper are available from publicly or commercially available sources noted above.

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.*

We describe the detailed procedures used to construct the sample and key variables in Section 2, “Institutional Setting, Data, and Key Variables.” Data used in subsequent analyses are described in detail in the corresponding sections. A complete list of variable definitions and construction procedures is provided in Appendix A. Variables used only in the Online Appendix are described in the relevant sections of the Online Appendix.

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.*

Attached to this datasheet, please find the replication package, which includes (1) Stata and Python code for data cleaning and regression analysis and (2) a README file that explains the execution environment and directory structure. Please note that all files are designed to be run on Windows.

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.*

Please see the replication package README file for detailed instructions. For sample construction, please refer to the "sample cleaning code" folder, which contains the code used to clean and process the relevant datasets. For the analyses, please refer to the "regression table code" folder, which contains the code used to generate all tables in the paper and Online Appendix. The unique identifiers used in the paper are provided in the "Identifiers.csv" file. Please note that differences in software versions and Stata package updates may affect the final results, so slight discrepancies may be expected.

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.

The log files corresponding to each code file are included in the relevant folders. For the sample-cleaning logs, please see the “sample cleaning log” folder. For the regression-analysis logs, please see the “regression table log” folder.

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 authors agree to maintain the data and programs used in this paper for the six-year time period following the suggestion of the National Science Foundation.