Data Description Sheet

"The Anti-SLAPP Statutes and Corporate Environmental Policies"

Kevin C.W. CHEN, Zhihong CHEN, and Chengzhu SUN

September 9, 2026

In compliance with the *JAR* Data Policy, we provide the following information regarding the empirical data used in the paper.

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

Chengzhu SUN and Zhihong CHEN handled the data and conducted the analyses. All authors engaged in the research design.

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.

Zhihong Chen collected (1) the anti-SLAPP statutes for each state from Thomson Reuters' West Law database (via HKUST library) in May 2021; (2) the IFS rating of anti-SLAPP statutes is obtained from https://anti-slapp.org/your-states-free-speech-protection#scorecard in June 2022, (3) the Integrated Database (IDB) of the Federal Judicial Center from https://www.fjc.gov/research/idb in June 2023, and (4) the data on the applicability of anti-SLAPP laws in the federal courts in each state from https://beta.antislapplawyers.com/federal/courts/map-of-applicability-in-federal-court/ in April 2025. He constructed the "SLAPP-like suits" data from the IDB database.

Chengzhu Sun downloaded the data on plant-year level toxic emissions, production ratios, and identifiers from the EPA TRI basic plus database (https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-plus-data-files-calendar-years-1987-present) in March 2025 and basic database (https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-present) in October 2022. She obtained the plant's historical employee number, location, and linkable from the National Establishment Time-Series (NETS) Publicly Listed Database in June 2023 (Data vendor contact information: Don Walls dwalls2@earthlink.net). She obtained the data on county-year level air quality index from the EPA website (https://www.epa.gov/outdoor-air-quality-data) in February 2022. She collected the firm-state level pollution-related news and protest-related news from the Factiva platform (via HKUST library) in October 2022 and in March 2025, respectively. She obtained the data on plant-year level pollution abatement from the EPA TRI basic plus dataset (https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-plus-data-files-calendar-years-1987-present) in March 2025.

She collected the data on county-year level population, income per capita, and employment data from the U.S. Bureau of Economic Analysis (https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas) and U.S. Census Bureau (https://www.census.gov/) in October 2022. She collected the state-year level population, income per capita, and employment data from the U.S. Census Bureau in April 2025. She collected the number of newspapers in each county from the annual list of U.S. local daily newspapers between 1990 and 2014 from the Editor and Publisher Yearbooks in the New York Public Library in September 2019. She obtained the data on the number of NGOs at the county level from the Northeast Regional Center for Rural

Development (NRCRD; https://nercrd.psu.edu/data-resources/county-level-measure-of-social-capital/) in October 2022. She obtained the data on state-level citizens' environmental awareness from the Pew Research Center (https://www.pewforum.org/religious-landscape-study/state/). She collected the State Environment-related Expenditures from the U.S. Census Bureau in April 2025 and the effective provisions in pollution-control regulations of each state from Legal Information Institute (LII) at Cornell Law School, accessible at https://www.law.cornell.edu/regulations, in September 2024. She obtained data on accounting data from Compustat in February 2024, U.S. equity mutual funds data from the CRSP survivor-bias-free Mutual Fund database in May 2022, mutual fund holding data from Thomson Reuters Mutual Fund Holdings - (S12) in May 2022, and environmental performance from the MSCI KLD database in September 2016.

We collected local pollution and protest-related news data using the Factiva Expert Search function (via the HKUST library) in May 2022 and April 2025. To do so, we first mapped each firm's unique identifier (GVKEY) in our sample to its corresponding Factiva company ID (fncode). Using these fncodes, we then retrieved each firm's state-level news coverage, excluding national newspapers (i.e., *The New York Times*, *The Wall Street Journal*, *The Washington Post*, and *USA Today*). In particular, we searched the Factiva News database to identify the two types of news. We restricted the news articles to those in the environment sub-category provided by the Factiva Expert Search function. We rely on the Region sub-category of the search function and require the articles to be about state *s*. For each parent firm, to identify the pollution-related news, we further required the news article to have at least one keyword for pollution ("toxic", "pollution", "pollutant", "hazardous", "air quality", "water quality", and "waste.") To identify the news about protests by environmental activists, we further required the news article to have at least one keyword for protests ("boycott*," "mobiliz*," "campaign*," "fight*," "rally," "rallied," and "protest*") and at least one keyword for environmental activists ("environmentalist*" or "environment*" combined with "activist*," "group*," or "organization*.")

The historical corporate headquarters dataset was generously shared by the research team of Zhihong Chen, Qingyuan Li, and Yongbo Li, who originally compiled it for their paper, 'Freedom of Expression and Stock Price Crash Risk: Evidence from a Natural Experiment,' published in the *Journal of Law & Economics*. They first identified the historical headquarters state of each firm using the 10- K header file, which covers observations between 1993 and 2021, and the Compustat database, which provides the data for the headquarters' state in 2022. The 10-K header file is available at the Software Repository for Accounting and Finance (https://sraf.nd.edu/sec-edgar-data/10-x-header-data/). If these historical headquarters' state data are missing, they filled the data by combining the data compiled by Heider and Ljungqvist (2015) and Bai, Fairhurst, and Serfling (2000) with the company header history file in the legacy Compustat/CRSP merged database (Bena, Ortiz-Molina, and Simintzi 2022).

All authors vouch for the stated sources of the raw data.

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).***

None of the data were obtained from an organization on a proprietary basis.

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 data in Section 4 of the manuscript. The variable definition and data source for each variable are detailed in Appendix B. For further details, see item #2 above and the code files available on the Journal of Accounting Research webpage as discussed below in item 5.

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) codes to be run on SAS and STATA programs, and (2) a Readme file that explains the execution environment and directory structure.

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.

The SAS and STATA programs are available on the Journal of Accounting Research webpage. We enclose these SAS and STATA code files pertaining to the data as follows:

a) TRIclean_share.sas processes the raw emission and abatement data from TRI datasets, along with historical data on plant-level employee numbers, location and link to public firm identifier from Establishment dataset

b) County_data_clean.sas processes the raw data on county-level pollution, income per capital and employment rate

c) Local_SRMF_share.sas processes the raw data from CRSP survivor-bias-free Mutual Fund database and MSCI KLD ESG datasets to identify local socially responsible investors.

d) Sample construction_share.sas merges the data generated from step a, b and c with all other data and sets up our final datasets for regression analyses

e) Regression code_share.do runs the paper's summary statistics and primary regression models in the manuscript.

We also attach a list of all distinct plant identifiers used in the paper's primary sample.

Note that Stata packages change over time which can cause slight differences in results. For example, the REGHDFE package is regularly updated, and occasionally these updates can have material impacts such as: (1) changes to the syntax of a REGHDFE call; (2) changes to the calculations of clustered standard errors. Slight differences are to be expected, depending on vintage.

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.

We enclose the log files for each SAS or STATA code

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.

We commit to retaining all data and code for at least six years, in line with NSF guidelines.