

## Data Description Sheet

for

### **Unintended Regulatory Spillovers: The Impact of Bank Regulatory Oversight on Environmental Enforcement**

John Gallemore (University of North Carolina)  
Vincent Giese (University of Mannheim)

August 2026

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

Vincent Giese handled the raw data, merged all datasets, and conducted all empirical analyses tabulated in the paper. John Gallemore downloaded and analyzed (using the ChatGPT API) the content of the bank enforcement actions used in a supplemental analysis.

#### **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.**

Data were downloaded or obtained in August 2023, September 2023, December 2023, March 2024, July 2024, June 2025, November 2025, and May 2026. Both authors have access to the data and programs and can vouch for the stated sources of the raw data.

**EPA data:** Nonattainment status data were obtained from the U.S. Environmental Protection Agency (EPA) and emissions data from the Toxic Release Inventory of the EPA.

**Small business lending:** Small business lending data were collected from the database maintained under the Community Reinvestment Act and published on the website of the Federal Financial Institutions Examination Council (FFIEC).

**Bank enforcement actions:** Data on bank enforcement actions (including bank identifiers and initiation and termination dates) were collected from Standard & Poor's Capital IQ. We obtained the PDFs of the bank enforcement actions directly from the relevant regulator's (OCC, FDIC, and Federal Reserve) website.

**Bank accounting data:** Bank accounting data are reported via the Call Reports and are published on the Chicago Fed website and the FFIEC Central Data Repository's Public Data Distribution site. We obtained this data via WRDS.

**Bank branch locations:** Bank branch location data were collected from the Federal Deposit Insurance Corporation Summary of Deposit files.

**Regulatory field offices:** Data on banks' distance to the next field office of their primary federal regulator compiled from regulatory websites, using the Wayback Machine for historical versions of the websites. For banks supervised by the OCC, we obtained the historical office locations directly from the authors of Gopalan et al. (2021).

**Bank regulatory leniency measures:** Data on bank regulators' leniency were obtained from Agarwal et al. (2014).

**Bank structure:** Bank structure data are published by the FFIEC National Information Center. We obtained this data via WRDS.

**County characteristics:** Data on demographic and economic county characteristics were obtained from the Bureau of Labor Statistics and the Bureau of Economic Analysis, and population data were obtained from the U.S. Census.

**Credit union and U.S. Small Business Administration data:** Data on lending by credit unions were obtained from the National Credit Union Administration, as well as the U.S. Small Business Administration.

**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, or requires that the results must be reviewed or approved prior to public release of the paper or publication.**

Except for the data shared by the authors of Gopalan et al. (2021), no other proprietary data have been used in the analyses. All other data can be accessed with a subscription (to the databases listed above, under 2).

**4. A complete description of the steps necessary to collect and 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.**

A description of the steps necessary to collect and process the data used in the final analyses is included in the manuscript.

**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.**

We confirm that all manipulations of the data were done via computer programs which are provided via the replication package. For one of the analyses in the online appendix, the raw data was scored using ChatGPT's API, which created a spreadsheet with variables that were not modified outside of a computer program.

**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 replication package provides code that converted raw data into the final data (1\_Data\_Import\_Cleaning, 2\_Data\_Processing), creates the sample used for the analyses (3\_Sample), and conducts the statistical analyses and creates the tables and figures reported in the paper (4\_Tests). We also provide additional code that conducts certain robustness tests (5\_CRA\_aggregate\_Import\_Cleaning, 6\_Robustness\_Tests). We also provide a master do file ("\_Masterfile.do") that executes these individual programs in the correct order, and a 'read me' file (GG\_Replication\_Package\_ReadMe) that walks interested readers through the replication package. Finally, we provide a Python script (7\_ea\_pdf\_classifier\_notebook) that took the raw enforcement action PDFs and analyzed them using a prompt to ChatGPT's API; the resulting data from this step is used in a test for the online appendix as indicated in 6\_Robustness\_Tests.

**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**

The replication package provides the comprehensive log file (ComprehensiveLogFile\_GG) that shows the execution of the entire 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 will maintain the data and programs from the paper for at least six years.