

ESG Rating Competition and Rating Quality

Compliance with Data Policy for the *Journal of Accounting Research*

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

The author Cai Chen handled the data and conducted the analyses. The author Svenja Dube and Shiran Froymovich provide access to the ESG data sources.

2. *A detailed description of how the raw data were obtained or generated, including data sources, the 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.*

- In this paper, we obtain data from the following sources: Compustat, ASSET4, KLD, RepRisk, Sustainalytics, I/B/E/S, Thomson Reuters, Corporate Register, Bloomberg ESG Disclosure score.
- Most of the data are directly downloaded from WRDS via our schools' subscriptions. Svenja Dube obtained the following databases while being an Assistant Professor at Fordham University: Corporate Register, Sustainalytics (legacy), Bloomberg ESG Disclosure score.
- All the ESG related data (i.e., ASSET4, KLD, RepRisk, Trucost) and financial data from Compustat are downloaded by Cai Chen on July 19, 2022.
- I/B/E/S data on analysts following and Thomson Reuters data on institutional ownership were downloaded by Cai Chen in October 2022.
- The Corporate Register data was purchased from Corporate Register in May 2020.
- The Sustainalytics data was purchased by Fordham University prior to Svenja Dube joining the university. The shared sample is dated September 2019.
- The Bloomberg Disclosure score was downloaded from Bloomberg in May 2021.
- The violation data was obtained from Good Jobs First as of 11 Jan 2022.

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., nondisclosure 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.

Not applicable to the paper as all sources of the data have been identified in the paper.

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.

We describe our data in the “Research Design and Data” section of the paper and define all variables in Appendix A. The steps necessary to process the data are described in the computer programs referred to in part 6 of this Data Description Sheet.

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.

All manipulation of the data is done via computer programs provided in part 6. The only manual procedure is the manual matching between ASSET4 and KLD detailed metrics. But all the variables created after the matching was also done by computer programs.

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 “Code Package” contains the required codes to convert different sets of raw data into the final datasets used in the analysis. The “code1_data cleaning” contains procedures of converting raw data into the datasets used in the main difference-in-difference analyses.

“code2_main DiD” contains the steps of the main difference-in-difference analyses. “code3_additional” contains the steps of additional analyses including staggered difference-in-difference, stacked difference-in-difference, and cross-sectional analyses. The excel file “identifiers” provides the firm identifiers (i.e., GVKEY) for the final sample of our main analysis.

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 file recording the execution of the programs from step 6 that covers data conversion and analytical procedures to generate the tables in the paper is provided in “logfile” file.

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 assure to maintain the data and programs used in this paper for the six-year time period suggested by the National Science Foundation.