

Data Description Sheet

“The Determinants of ESG Ratings: Rater Ownership Matters”

Dragon Yongjun Tang, Jiali Yan, Yaqiong Yao

August 14, 2025

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

Jiali Yan and Yaqiong Yao handled the downloaded data and performed the analyses described in the paper.

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.*

The data were retrieved from multiple sources. Section 3 of the paper outlines the sample construction procedure, and the Appendix provides detailed definitions for all variables along with their corresponding data sources. Unless otherwise noted, the initial data collection took place during May and June 2017.

1. Firm-level financial data were downloaded from Compustat (via WRDS), stock return data from CRSP (via WRDS), and combined data from the CRSP/Compustat Merged database (via WRDS).
 2. Institutional ownership data were obtained from the Thomson Reuters Institutional Holdings (13F) database (via WRDS), and analyst coverage data were sourced from the IBES Detail file (via WRDS).
 3. ESG strength and concern data were obtained from the KLD database.
 4. Alternative ESG data were collected from four sources:
 - ESG ratings from Bloomberg and Refinitiv
 - TruValue Labs data accessed via FactSet (September 2021)
 - Sustainalytics data obtained from Morningstar (January 2024)
 5. ESG risk incident data were retrieved from RepRisk, with updates current as of February 2022.
 6. Violation Tracker data were sourced from Good Jobs First in May 2021.
 7. ESG ratings were classified as material or immaterial using the SASB materiality map (as of May 2020).
 8. Segment data were downloaded from Compustat Historical Segments in May 2023.
 9. The 10-K filing size variable, used as a firm complexity measure, was obtained from the Software Repository for Accounting and Finance at the University of Notre Dame in June 2023.
 10. Data on SRI shareholder proposals were obtained from Institutional Shareholder Services (updated as of April 2018).
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).

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 the sample in the “Data and Summary Statistics” section of the paper, and variable definitions are provided at the end of the appendix of the paper.

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 data analyses were conducted using computer programs, and the corresponding code has been uploaded to the JAR website.

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.*

All computer programs used for the data analyses have been uploaded to the JAR website. A description of the code is provided in the “Readme.rtf” file within the replication package, and firm identifiers are included in the “firm_identifiers.rtf” file.

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 comprehensive log file showing the execution of the entire code has been uploaded to the JAR website.

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 confirm that the data and programs will be retained for at least six years.