

Data Description Sheet

Financial Reporting Around Private Firms' Securities Offerings

Yiran Kang

This datasheet is provided upon submission of a manuscript for fulfillment of the Journal of Accounting Research data policy.

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

Yiran Kang handled the data and conducted the 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 is able to vouch for the stated source of the raw data.

This manuscript uses both public and private data.

For public data, two data sources are used. The first part of the data is downloaded from the SEC website (page link: <https://www.sec.gov/dera/data/form-d> and the SEC EDGAR. The raw data are downloaded from the SEC website on 1/23/2021. The second part of the data is provided by the California Department of Financial Protection and Innovation. The data is public available, and the California DFPI provides the batch downloaded version. The data is obtained on 2/12/2022.

The private data is obtained through a survey instrument, and the instrument is provided in the Appendix B of the manuscript. The survey is administered between December 2022 and March 2023.

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).

The private data used in the manuscript is obtained through an anonymous survey instrument with IRB approval from Columbia University (protocol #AAAU0025). To protect the privacy of the respondents, all personal information is kept confidential, and data is reported on the aggregate level without any identifying information.

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.

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.

I describe the data collection process in the data and methodology section of the paper. Since this paper relies on a survey instrument to collect the main dataset for analysis, a copy of the survey instrument is included in Appendix B of the paper.

Upon Acceptance

To be provided *upon acceptance* of the paper and prior to publication:

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 relevant codes to clean and process the data and generate the results. The “R_code” file contains a detailed step-by-step description about how the raw data is cleaned and process, as well as the intermediate steps to

generate input of the regression analyses. The “Stata_master.do” file contains a detailed description on how the cleaned data is used to generate the regression results.

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.*

A log file detailing the steps to clean and process the dataset has been included in the code package. The log file refers to the step-by-step description provided in item 6.

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.*

I will maintain the data and programs for at least six years.