

Data Description Sheet

Paper: Mimicking Regulatory Peers

Author: Minjae Kim, University of Alberta (minjae9@ualberta.ca)

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

I handled all 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 can vouch for the stated *source* of the raw data.

The data used in this paper were obtained from the following sources:

- Reports of Condition and Income (Call Reports), FFIEC: I downloaded the quarterly Call Report SAS data files (.sas7bdat) from Wharton Research Data Services (WRDS) in January 2021, and downloaded additional variables in July and August 2024 for analyses added during the revision process. The files were assembled into a quarterly bank panel using Python.
- FDIC Failed Bank List: public download from fdic.gov, obtained in November 2021. Used to construct bank failure outcomes.
- CRSP–FRB link: Federal Reserve Bank of New York public download. Used to link bank RSSD identifiers to CRSP PERMCOs.
- FDIC Summary of Deposits (number of offices and MSA location): public download, obtained in August 2021.
- SRISK: obtained directly from the NYU Stern Volatility Laboratory (V-Lab) upon request in December 2023; the series is also publicly available at <https://vlab.stern.nyu.edu>.

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

The paper uses no proprietary data.

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.

A detailed description of the data, sample construction, and variable definitions is provided in the paper and the Online Appendix. The code provided with the paper performs every step from the raw data files to the final analysis datasets and all reported results.

- 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.**

After the raw data were obtained, all manipulations of the data were performed via the computer programs provided; no manipulations were made manually or outside the code.

- 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 computer programs used to convert the raw data into the final datasets, to execute the analyses, and to generate the tables are provided, together with a README describing how to run them. Identifiers for the final sample (RSSD IDs) are provided as CSV files.

- 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.

Three comprehensive log files showing the execution of the entire code, from the raw data to every statistical analysis presented in the manuscript's tables, are provided. No portion of the log is masked.

- 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, consistent with National Science Foundation guidelines.