

Journal of Accounting Research Data Policy
Data Description Sheet to
“A Tale of Two Banks: When Credit Loss Models Meet Economic Crisis”

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

Response:

Both authors were involved in the empirical analyses. Difang Huang collected the data and performed the analyses outlined 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.*

Response:

Difang Huang conducted the initial data analysis during five visits to the data provider's offices between October 2021 and May 2025.

We acquired loan-level and bank-level data from the Office of the China Banking and Insurance Regulatory Commission, which collects this information from the monthly credit registry maintained by all commercial banks in the province. We obtained firm-level data from the Bureau of Statistics, which collects financial and operational information from all registered enterprises in the province. We constructed two location-based variables using publicly available data. The "Top 5 Infected Areas" indicator identifies areas among the top five in cumulative COVID-19 infection cases from January 1, 2020, to April 1, 2020, based on data from the Chinese Center for Disease Control and Prevention. The "Top 5 Restricted Areas" indicator identifies areas among the top five in lockdown severity from January 1, 2020, to April 1, 2020, based on population mobility data from Baidu Qianxi. The Baidu Qianxi platform provides real-time and historical migration patterns based on location-based services data, which we used to measure the reduction in traffic flow as a proxy for lockdown severity during the specified period. Both authors can vouch for the stated sources of the raw data.

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). 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.*

Response:

The authors will provide the editors with confidential contact information for a representative of the data provider who can confirm the data access arrangement, as well as copies of the relevant non-disclosure agreements and data sharing terms that prohibits us from sharing raw data we have received. Publication of aggregate summary statistics and test results is permitted. Both authors can vouch for the authenticity and stated sources of the raw data used in this study.

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.*

Response:

Section 3 of the paper provides a comprehensive description of the data collection and processing methodology. This section details the sources of proprietary data, explains the sample selection criteria and any exclusion rules applied, describes the variable construction process, and outlines the steps taken to merge datasets and prepare the final analytical sample. Additional technical details regarding data processing procedures are available in the Online Appendix. In addition, all variables and data are described in the table captions.

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.

Response:

All data manipulations after obtaining the raw data were performed using computer programs. The replication package attached to this datasheet includes: (1) Stata code (.do files) that execute all data processing and statistical analyses; and (2) a Readme file that documents the execution environment, directory structure, package dependencies, and instructions for replication. The master.do file orchestrates the complete workflow from data processing through final output generation. Due to the proprietary nature of the raw data and the requirement to access it on-site at the data provider's offices, we are unable to provide code that downloads the data. The replication package documents all steps taken after obtaining the raw data, ensuring full transparency of our data manipulation and analysis procedures.

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.*

Response:

Please see the replication package for all materials related to data processing, statistical analysis, and output generation. While the underlying data is proprietary, we provide all non-proprietary code used in the analysis. The primary Stata program (master.do) executes the complete workflow to generate the results presented in the paper, including execution of all statistical and econometric analyses, and generation of tables presented in the manuscript.

A readme file included in the replication package provides detailed instructions for running the code and understanding the data processing steps, including how the final sample was formed, how variables were defined and constructed, and which specific commands and options were used in each analysis. The study uses Unified Social Credit Identifiers (USCI) in China as the primary firm identifiers, and these identifiers for the final sample are available for the verification of sample construction and results.

Replicators should note that Stata packages evolve over time, which may cause minor differences in replicated results. Specifically, the reghdfe package receives regular updates that can occasionally affect command syntax requirements and calculations of clustered standard errors. Additionally, the psmatch2 package may produce slight variations across replications due to the random ordering of observations when breaking ties in propensity score matching. When multiple control observations have identical propensity scores, the algorithm selects matches based on the current sort order of the dataset, which can vary across Stata sessions.

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.*

Response:

Please see the log file 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.

Response:

The data providers will maintain the proprietary datasets used in this study. The authors commit to maintaining all analytical programs, code, and documentation for a minimum of six years following publication, in accordance with National Science Foundation guidelines.