

# Reporting Regulation and Private Firms' Bank Credit

## Data Description Sheet

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

Antonio Moreta was solely responsible for handling the data and conducting all the analyses presented in this 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 can vouch for the stated source of the raw data.**

Raw data were obtained from three different confidential databases provided by Banco de España. The first database is called *Central de Balances Integrada* (CBI). It contains firm-level administrative data from Spanish firms legally obliged to deposit annually their financial statements in the Commercial Registry. The Banco de España digitalizes this information and performs comprehensive checks to ensure the accuracy of the database. The raw data corresponds to a download conducted at the beginning of 2023 regarding all the available data contained in the database for fiscal years 2012 to 2020 (both included). In my analysis, I use information on multiple balance sheet and income statement items, the number of employees, the type of financial statements disclosed (i.e., detailed, mixed, or abbreviated), fiscal year, firm and industry identifiers, and zip code.

The second database is the Central Credit Registry of Spain or *Central de Información de Riesgos* (CIR), which is maintained by the Banco de España in its capacity as the primary banking supervisory agency. It contains comprehensive monthly loan-level information on the universe of outstanding loans issued to nonfinancial firms by all banks operating in Spain. The primary raw data consists of monthly extracts provided by employees of the Banco de España in 2021. I use the December extracts from each year from 2013 to 2020 to ensure that the data from CBI and CIR are measured at the same point in time, as over 99% of firms have their fiscal year ending in December. Note that CIR variables are forward-filled in my analysis, so my sample period ends in 2019. I drop loan-year observations from non-depository institutions or with negative or zero outstanding credit amounts. From these extracts, I build a loan-level panel that includes firm and bank identifiers, the outstanding credit amount, and loan types. I complement this panel by downloading and merging information on interest rates, which is only available from 2018, and by manually identifying the province where the headquarters of each bank are located. This panel allows me to construct the majority of the aggregated firm-level variables that I subsequently merge with the firm-level information from CBI (e.g., the outstanding credit amount from “incumbent” and “new” banks, the number of banks in a relationship with the firm, the length of the lending relationship, the weighted average interest rate of the firm, etc...).

The third source of data employed in the paper includes all banks' requests for information from the CIR on potential borrowers. By law, a bank may only request CIR information on potential borrowers when the borrower “seriously approaches the bank to obtain credit” (i.e., applies for a loan). Because banks receive monthly updates on the credit exposures and loan defaults of

their current borrowers, further requests for CIR information are only useful if the firm is not already a borrower. Therefore, I interpret banks' requests for information as *loan applications* from potential borrowers.

The main sample used in the analyses is restricted to fiscal years from 2016 to 2019 (both included), aligning with the implementation of the size-based reporting thresholds described in the paper (see Section 3), and reflecting the last fiscal year for which I have available CIR data (i.e., 2019 since CIR variables are forwarded in the analyses). It includes nonfinancial limited liability Spanish firms within the 10.5% optimal bandwidth from the binding reporting threshold, which is determined following [Calonico, Cattaneo and Titiunik \(2014\)](#). The 10.5% optimal bandwidth excludes from the sample firms that fall below the size-based auditing thresholds. I also drop from the sample public firms or firms belonging to a consolidated group. I further drop firm-years in the top and bottom 5% percentile of the distributions of assets, sales, and employees. This trimming process ensures that "small" and "medium" firms are comparable along all size dimensions (not only the binding dimension), and helps to obtain a more balanced sample, given the correlation of firm size with several firm characteristics. Finally, I require all firm-years in the sample to have non-missing data for any of the variables included in the main specification. For influential studies using the same three databases, I refer to [Jiménez, Ongena, Peydró, and Saurina \(2012, 2014\)](#).

**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 contact information of the representative of the organization who can confirm that I obtained the data:

**Carlos Pérez Montes**

Head of the Department of Financial Stability and Macprudential Policy

Banco de España

[carlos.perezmontes@bde.es](mailto:carlos.perezmontes@bde.es)

In my analyses, I use three proprietary datasets provided by *Banco de España*, which cannot be publicly distributed. *Banco de España* does not impose any restrictions on the publication of the current content of the paper. Furthermore, *Banco de España* generally does not restrict the publication of additional results, including those arising from a revised version following the refereeing process, provided that the findings are aggregated and anonymized. The results presented in the paper have been reviewed and approved by staff of *Banco de España* exclusively for the purpose of ensuring compliance with their disclosure policy. These results are derived from research conducted between 2021-2023 during a research visit at the *Banco de España*, where I was part of a project, and I was granted access to the databases. Any additional analyses were conducted by submitting scripts to *Banco de España* employees, who then generated the results. *Banco de España* retains control of the datasets.

**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.**

The CIR extracts were provided by Banco de España staff in 2021, the CBI database was downloaded from the Banco de España in early 2023, and the loan applications database was provided by Banco de España staff in 2023. All three sources are proprietary and confidential administrative datasets.

The processing steps required to construct the final analysis samples are fully documented and reproducible in the replication package (do-files 10, 11, 12, 13, and 20). At a high level, the processing involved:

- Cleaning the CIR database, aggregating loan-level information to the firm-year level, and preparing identifiers and variables for merging with CBI;
- Cleaning the CBI database, identifying the sample around the reporting thresholds, and merging with CIR variables;
- All cleaning details and filters are fully described in Point 2 of this *Data Description Sheet*, and Section 4 of the manuscript.

All necessary steps, including variable definitions and coding, are detailed in the replication do-files listed above. Appendix A of this *Data Description Sheet* provides a variable mapping between the variables reported in the manuscript and the corresponding variable names used in the replication code. No survey instruments, experiments, or subject instructions are applicable in this setting.

**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 manipulations from the raw data to construct the final datasets and output presented in the paper are performed exclusively through the do-files attached in the replication package.

**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 included in the replication package (1) convert the raw data into the final dataset used in the analysis, (2) execute all statistical and econometric analyses, and (3) generate the tables and figures used in constructing the results presented in the manuscript. The code is in Stata format and organized in a sequential fashion that mirrors the structure of the paper. Do-files 40, 50, and 60 sequentially reproduce the output presented in Sections 6, 7, and 8 of the paper, respectively. All intermediate data construction steps, variable definitions, trimming rules, and sample formation procedures are explicitly coded in earlier do-files (10, 11, 12, 13, and 20).

To facilitate replication, a README file is included in the replication package. This README describes the folder structure, explains the role of each do-file, and provides a roadmap for running the code. In particular, it explains that the file “00 Master Do.do” runs all scripts sequentially and generates the corresponding log file.

Because the databases provided by Banco de España are confidential, it is not possible to provide the firm-level identifiers of the final sample. However, the replication package contains the complete code necessary to transform the raw confidential data into the final samples and reproduce all results reported in the manuscript.

**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 comprehensive log file documenting the execution of the entire replication package is included. This log file records all steps, from the processing of the raw databases into the final analysis samples to the execution of the econometric analyses and the generation of the tables and figures reported in the manuscript. The log file is generated automatically by running the master do-file (“00 Master Do.do”).

**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.**

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

\*References

- Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression- discontinuity designs. *Econometrica* 82, 2295–2326.
- Jiménez, G., Ongena, S., Peydró, J.L., Saurina, J., 2012. Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications. *American Economic Review* 102, 2301–2326.
- Jiménez, G., Ongena, S., Peydró, J.L., Saurina, J., 2014. Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? *Econometrica* 82, 463

## Appendix A:

### Variable Name Mapping

Variable in Manuscript	Variable in Code	Variable in Manuscript	Variable in Code
<i>Mixed</i>	mixed	<i>STcredit (NewLoans)</i>	wmarginal_st_exposure
<i>Medium</i>	MEDIUM	<i>LTcredit (NewLoans)</i>	wmarginal_lt_exposure
<i>Credit</i>	wexposure_cfab_end_assets	<i>log(#App)</i>	log_Napplications
<i>LDT2</i>	LDT2	<i>NewLoan</i>	new_loan
<i>LDT2*Medium</i>	LDT2_MEDIUM	<i>#Banks</i>	Nbanks
<i>I<sup>1</sup></i>	threshold1	<i>NewBank</i>	new_bank_2
<i>LDT1</i>	LDT1	<i>NewBCredit</i>	wnew_credit2_assets
<i>LDT1* I<sup>1</sup></i>	LDT1_threshold1	<i>MainBCredit</i>	wcredit_main_bank_assets
<i>I<sup>3</sup></i>	threshold3	<i>MainBShare</i>	share_credit_main_bank
<i>LDT3</i>	LDT3	<i>RelLenght</i>	mean_rel_length_total
<i>LDT3* I<sup>3</sup></i>	LDT3_threshold3	<i>DiffProvince</i>	atleast_1_diff_province
<i>Liquidity</i>	wLIQUIDITY	<i>HHI(&gt; p50)</i>	high_HHI
<i>Cash</i>	wCASHSTI	<i>RelLength(&gt; p50)</i>	high_rel_lenght
<i>Equity</i>	wEQTOTA	<i>Age(&gt; p50)</i>	high_AGE
<i>Inventories</i>	wINVT	<i>ROA(&gt; p50)</i>	high_ROA
<i>EBITDA</i>	wEBITDA_MARGIN	<i>Intan(&gt; p50)</i>	high_INTAN
<i>O-Score</i>	wOSCORE	<i>SD ROA(&gt; p50)</i>	high_SD_ROA
<i>log(Age)</i>	log_age	<i>PrevDefault</i>	high_previous_default_4y
<i>log(AR Cycle)</i>	log_ARcycle	<i>Equity(&gt; p50)</i>	high_wEQTOTA
<i>log(AP Cycle)</i>	log_APcycle	<i>Equity (Table 8)</i>	wfequity
<i><math>\zeta_{i,t}</math></i>	ind1_year	<i>Liabilities (Table 8)</i>	wfliabilities
<i><math>\alpha_{i,t}</math></i>	ccaa_year	<i><math>\Delta</math>Assets (Table 8)</i>	wfassets_growth
<i>Unscaled Bank Credit</i> <i>(thousands of euros)</i>	wend_of_year_exposure_cfab	<i>ROA (Table 8)</i>	wfni
<i>Natural Logarithm of Unscaled Bank Credit</i>	log_wcredit_cfab	<i>Cash (Table 8)</i>	wfchsti
<i>log(IntRate)</i>	log_interest_rate	<i>IntExp (Table 8)</i>	wfinte
<i>ABL</i>	wend_of_year_exposure_abl	<i>LTdebt(BS) (Table 8)</i>	wfdltl
<i>CFL</i>	wend_of_year_exposure_cfl	<i>STdebt(BS) (Table 8)</i>	wfcrc
<i>CreditLine</i>	wend_of_year_exposure_cl	<i>OtherLiab (Table 8)</i>	wfresto
<i>TermLoan</i>	wend_of_year_term1	<i>TradePayables (Table 8)</i>	wfprv
<i>STcredit</i>	wst_exposure	<i>Credit</i>	credit_hat
<i>LTcredit</i>	wlt_exposure	<i>Post</i>	POST
<i>log(IntRate) (New Loans)</i>	log_marginal_rate	<i>Winsor 2.5% Credit (Table 10)</i>	w2exposure_cfab_end_assets
<i>Credit (New Loans)</i>	wmarginal_exposure_cfab	<i>All Loan Types (Table 10)</i>	wexposure_end_assets
<i>ABL (New Loans)</i>	wmarginal_exposure_abl	<i>LDTB2</i>	LDTB2
<i>CFL (New Loans)</i>	wmarginal_exposure_cf	<i>Large</i>	BIG
<i>CreditLine (New Loans)</i>	wmarginal_exposure_cl	<i>Detailed</i>	detailed
<i>TermLoan (New Loans)</i>	wmarginal_exposure_term1		