Replication package for "Subsidizing Business Entry in Competitive Credit Markets" by Vincenzo Cuciniello, Claudio Michelacci, and Luigi Paciello

The STATA codes in the directory "ReproductionStata" generate the datasets used for the analysis by combining several confidential data with public-use data and run the regressions discussed in the paper. The Matlab codes in the directory "ReproductionMatlab" are used to solve and calibrate the model as well as to run simulations and calculate the optimal business creation subsidy as discussed in the main paper. The analysis was conducted using Stata and MATLAB, with the code last run on Stata 18.0/MP and MATLAB 2024a.

Data Availability

Only authorized employees of the Bank of Italy are permitted to access and merge together the datasets used in the paper.

We combine six data sources for Italy: i) the Credit Register managed by the Bank of Italy; ii) the Business Register by the Chambers of Commerce (InfoCamere); iii) Universo Imprese INPS (UNINPS) from the Italian National Social Security Institute (INPS); iv) the Company Accounts Data Service (CADS) assembled by CERVED Group; v) the Survey of Industrial and Service Firms (INVIND) conducted by the Bank of Italy; vi) the Labor Force Survey and other data from the Italian National Institute of Statistics (ISTAT); and vii) information regarding *I Stay in the South (ISS)* beneficiaries from the Government Department for Cohesion Policies and the National Government Agency for Inward Investment and Economic Development (Invitalia). Sources i-iv and vii are administrative data, while sources v and vi are survey data. Only vi and vii are publicly available, while access to i-v is restricted and subject to strict confidentiality requirements (see below). Firm-level information from different datasets is matched using the company's fiscal code.

Central Credit Registry and TAXIA (highly confidential)

Bank loans to businesses are sourced from the Central Credit Registry (CCR), managed by the Bank of Italy. The CCR provides monthly data on loans from banks and other financial institutions to borrowers, reporting (i) all loans where total exposure to a single intermediary exceeds a threshold (€75,000 before January 2009 and €30,000 afterward), and (ii) all loans of borrowers with non-performing loans (NPLs) classified as bad loans ("Sofferenze"). The threshold is reduced to €250 if the borrower has some bad debt. CCR reports loan amounts, insolvency status, collateral or guarantees, origination dates, and maturity dates. Banks and financial intermediaries are legally required to report this information. Since September 2018, CCR and AnaCredit data have been merged.

Interest rate data charged by financial intermediaries are available through TAXIA, which is part of CCR. TAXIA provides quarterly data on interest rates for clients with total bank exposure exceeding €75,000, excluding NPLs. It covers over 80% of all loans granted by the banking system. Interest rates are the weighted average of the rates charged to customers, including interest, fees, and commissions, using the formula: R = (Interest, fees, and commissions in the year) * 365 / DWB *100 where DWB is the daily weighted balance equal to the product between the daily outstanding amount and the days in which that amount was outstanding.

Direct access to the datasets, including firms' fiscal identifiers and permission to merge them with external data sources, was granted solely because one of the authors is an employee of the Bank of Italy. These options are not available to external users.

References:

-Anacredit https://www.ecb.europa.eu/ecb-and-you/explainers/tell-me-more/html/anacredit.en.html
-Banca d'Italia Segnalazioni sui tassi di interesse analitici. 2019. https://www.bancaditalia.it/statistiche/raccolta-dati/segnalazioni/normativa-segnalazioni/Circ_251_1_agg.pdf
-Banca d'Italia Segnalazioni sui prestiti della Centrale dei rischi. 2022. https://www.bancaditalia.it/compiti/vigilanza/normativa/archivio-norme/circolari/c139/Circolare_139_20_agg.pdf

INFOCAMERE (CONFIDENTIAL)

InfoCamere, managed by the Italian Chambers of Commerce, is built using the digitized Business Register (available since 1993). It provides detailed information on all businesses, including their name, legal form, type of activity, tax code, geographic location, board members and directors, and, if applicable, the reason for business exit. In our analysis, we focus on limited liability companies (corporations) excluding sole proprietorships. Business exits with bankruptcy are identified through insolvency proceedings, which are formally codified by Italian bankruptcy law, involve court and public authority oversight.

Since 2010, InfoCamere also includes data on shareholders and their shareholdings for all businesses. Researchers interested in accessing InfoCamere data can subscribe via the platform at https://accessoallebanchedati.registroimprese.it/abdo/?lang=en.

Reference

-https://www.registroimprese.it/visura-camerale-ordinaria-storica-e-certificato

UNINPS (CONFIDENTIAL)

For welfare-related purposes, INPS systematically collects employment and wage data from all private, non-agricultural employers. Since 1990, UNINPS provides data on the number of employees and wage payments for businesses with at least one employee. Using UNINPS, we determine the annual average number of payroll employees per business, as well as the year of birth of all firm's establishments, their exit, and the age (in years) of each firm. We classify a firm as new if it employs workers for the first time. The firm exits if its workforce drops permanently to zero, using multi-year observations after exit to confirm the exit status.

Firm's age is calculated as the number of years since it first employed some workers, calculated for all firms born since 1990. Firms exit with "bankruptcy" if they exit with some outstanding bad loans (as recorded in the CCR) or if a formal bankruptcy procedure is recorded in InfoCamere's insolvency proceedings. To analyze the dynamics of firms over their life cycle, we track firms over time and calculate annual averages from 2007 to 2015, grouping firms aged 17 years or older into a single category (age 20), with 2007 marking the first year firms have reached age 17.

INPS has introduced the "VisitINPS Scholars" program, granting researchers access to the entire INPS archives. Applicants must present a research proposal aligned with topics identified by the Scientific Committee or related to INPS's "Strategic Research Areas" on labor market and welfare issues. For more information, see
https://www.inps.it/it/it/dati-e-bilanci/attivit--di-ricerca/programma-visitinps- scholars.html

Reference
-Istituto Nazionale della Previdenza Sociale. 2019. Archivi delle denunce mensili. Universo delle imprese italiane. Years: 2006-2019. https://www.inps.it/

CADS (CONFIDENTIAL)
CADS, managed by Cerved Group, is a proprietary database containing the balance sheets and income statements of all Italian limited liability companies since 1993. Firms are legally required to submit their balance sheets to their local Chamber of Commerce, and Cerved assembles this information updating the database annually.

We use CADS to calculate total financial debt and value added. Total financial debt includes bank debt and other financial liabilities, excluding trade payables. Value added is calculated as the sum of production value and provisions, minus expenditures on raw materials and intermediate inputs. The leverage ratio is the ratio of total financial debt to a firm's value added.

Researchers can access CADS contacting CERVED, see https://www.cerved.com/contatti.

Reference
-Cerved Spa. 2019. Base Dati Centrale dei Bilanci/SIEF Database. Years: 1995-2019, Cerved Spa. https://www.cerved.com/.

INVIND (CONFIDENTIAL)
INVIND, available since 1995, is a representative sample of firms with at least 20 employees in industrial sectors (manufacturing, energy, and extractive industries) and non-financial private services, headquartered in Italy. INVIND covers approximately 70% of total sales in Italy. It uses a one-stage stratified sampling method based on 11 sectors, firm size, and region (see *Bank of Italy, 2014* for further details). Recent waves include around 4,000 firms (3,000 industrial firms and 1,000 service firms). Firms in mineral extraction, construction, or involved in extraordinary operations (e.g., mergers, spin-offs) are excluded. INVIND provides data on expected and realized changes in prices and sales, as well as firms' expected demand elasticity, enabling identification of unexpected demand shifts assuming a log-linear iso-elastic demand model as detailed in the online Appendix and further discussed in Cuciniello, Michelacci and Paciello (2024).

External researchers can access INVIND through the REX system, which ensures data confidentiality by allowing analyses without direct access to microdata. After logging into the portal, users submit codes writen in supported programming languages and download the output. Data merges with public datasets owned by researchers can be requested via RemoteExecution@bancaditalia.it, while technical issues can be addressed at autoregistrazione@bancaditalia.it. More details are available at: https://www.bancaditalia.it/statistiche/basi-bati/rdc/rex/index.html?com.dotmarketing.htmlpage.language=1.

Reference

-Bank of Italy (2014). Survey of industrial and service firms. Technical report 34, Bank of Italy.
-Cuciniello, V., C. Michelacci, and L. Paciello (2024). Firm debt and _rm exit dynamics in North and South of Italy. EIEF Working Paper 24/12.

ISTAT

The Italian Labour Force Survey (ILFS), conducted by ISTAT, is a cross-sectional household survey that collects annual labor market data from approximately 250,000 households and 600,000 individuals. As part of the European Labour Force Survey, it serves as Italy's primary source of labor market statistics. The Intercensal Reconstruction of Resident Population (IRRP) for 2002–2018 provides detailed municipal-level population estimates by sex, age, and citizenship, combining census data (2001, 2011, 2018) with demographic flows (for births, deaths, migrations, and citizenship acquisitions). ILFS data is used to compute total employment at the provincial level, IRRP for provincial population estimates, and national accounts data to determine provincial gross value added.

Reference

-Istituto Nazionale di Statistica. 2021. Conti e aggregati economici nazionali annuali. Years: 2006-2019. https://www.istat.it/

OpenCoesione and Invitalia

Information on beneficiaries of the *I Stay in the South* program, including reimbursement timings and amounts, is obtained from the websites managed by the Government Department for Cohesion Policies (https://opencoesione.gov.it) and the National Government Agency for Inward Investment and Economic Development (https://www.invitalia.it/trasparenza/sovvenzioni-contributi-sussidi).

References

-Dipartimento per le politiche di coesione e per il sud: Resto al Sud. https://politichecoesione.governo.it/it/politica-di-coesione/strategie-tematiche-e-territoriali/misure-speciali-per-imprese-e-lavoro-al-sud/resto-al-sud/
-Invitalia: Resto al Sud. https://www.invitalia.it/cosa-facciamo/creiamo-nuove-aziende/resto-al-sud

DATASET LIST

1. va.dta contains data at the province-year level from ISTAT on GDP. [Provided]
2. invind.dta contains data at the firm-year level from INVIND. [Not provided due to confidentiality restrictions]
3. uninps.dta contains data at the firm-year level from UNINPS. [Not provided due to confidentiality restrictions]
4. taxia.dta contains data at the firm-year level from TAXIA. [Not provided due to strict confidentiality restrictions]
5. ccr.dta contains data at the firm-year level from CCR. [Not provided due to strict confidentiality restrictions]
6. cerved.dta contains data at the firm-year level from CADS. [Not provided due to confidentiality restrictions]
7. LFS_codpro.dta contains data at the province-year level from ISTAT on employment and population. [Provided]
8. nfirm_xx.dta contains data from UNINPS foe ach firm age at the province-year level. [Not provided due to confidentiality restrictions]
9. nfirm_ce_xx.dta contains data from UNINPS-CADS-CCR for each firm age at the province-year levels. [Not provided due to strict confidentiality restrictions]
10. nfirm_cebp_xx.dta contains data at the province-year level from UNINPS-CADS-INFOCAMERE-CCR on a balanced panel of firms sorted by firm age. [Not provided due to strict confidentiality restrictions]
11. infoc_panel_rich.dta contains data at the firm level from INFOCAMERE on market exit cause. [Not provided due to confidentiality restrictions]
12. nfirms_ex_ce.dta contains data aggregated at the province-year level from UNINPS-CADS-INFOCAMERE-CCR on firms exiting the market with bankruptcy or bad loans. [Not provided due to strict confidentiality restrictions]
13. soci_SC_SPxx.dta contains data at the firm-year level from INFOCAMERE on shareholders. [Not provided due to confidentiality restrictions]
14. bankdebt.dta contains data at the firm-month level from CCR. [Not provided due to strict confidentiality restrictions]
15. latitude.dta, itprovdb.dta, and provcoord.dta contain data on latitude and longitude (geocode) of each province. [Provided]
16. mass_entry_n.txt and mass_entry_s.txt contain model simulated data for North and South respectively. [Provided]
17. data_n.txt and data_s.txt contain data for province age year for model simulated data in the North and South, respectively. [Provided]
18. OpenCoesione.dta and invitalia.dta contain information on beneficiaries of the I Stay in the South program. [Provided]
19. province_codici.dta contains the ISTAT codes for provinces. [Provided]

COMPUTING ENVIRONMENT

The empirical analysis was conducted on a Shared Server with an Intel® Xeon Gold 6248 processor and 1.2TB of RAM, subject to the following limits:
fsize: 128 GB
data: 512 MB
stack: 64 MB
as (address space): 128 GB

Running all codes to construct data sets takes approximately 8 days. The analysis was performed using Stata 18.0/MP, and the total runtime for the code on the Shared Server is about 3 hours.

The Matlab code was run with Matlab Release 2024a on a Mac Studio M2 Max 64GB. The total time required to run Main_calibration.m is about 1 day, and approximately 2 hours for each of the experiment of Main_opimal_subsidy.m.

DESCRIPTION OF CODES AND INSTRUCTIONS FOR REPLICATION

STATA code

1. MainFinalV1.do: generates all empirical tables and figures in particular Table 1, Table 2, Table 3, Table 5, Table 6 and Figure 7.
2. descriptive_dataset.do: called by MainFinalV1.do creates the working dataset "descriptive.dta."
3. event_study_dataset.do: called by MainFinalV1.do creates the working dataset "event_study_dataset.dta."
4. Createdaset.do: called by MainFinalV1.do, uses "descriptive.dta", "event_study_dataset.dta", and "latitute.dta" to generate the working datasets "calibrationdataset.dta", "EffectSubsidydataset.dta", and "Table3_dataset.dta".
5. calibrationV1.do: called by MainFinalV1.do, uses "calibrationdataset.dta" to produce Table 1.
6. AgeProfileV1.do: called by MainFinalV1.do, uses "calibrationdataset.dta" to produce "Figure 7".
7. EffectSubsidyV1.do: called by MainFinalV1.do, uses "EffectSubsidydataset.dta", "mass_entry_n.txt", "mass_entry_s.txt", "data_n.txt", and "data_s.txt" to produce Tables 5 and 6.
8. EventSubsidyV1.do: called by MainFinalV1.do, uses "Table3_dataset.dta" to produce Table 3.

Main MATLAB codes

There are two main codes **Main_calibration.m** and **Main_opimal_subsidy.m.** The two codes run independently and calls some intermediate codes (discussed below) available in the folder "codes".

1. **Main_calibration.m**: calibrates the model producing the parameter values reported in Table 2 obtained from matching some empirical statistics also reported in Figure 7; **Main_calibration.m** also produces the age profiles on model simulated data also reported in Figure 7. To run the code, the researcher should change the "root" folder path at the beginning of the code in **Main_calibration.m**.
2. **Main_optimal_subsidy.m**: it uses the parameter values of Table 2 to produce Figures 3-15 in the main text of the paper. To run the code, the researcher should change the "root" folder path and assure it is aligned with the path set in **Main_calibration.m**. At the beginning of the code there is global variable incid_expe. By setting "incid_expe" equal to 1, 2, 3, 4, the researcher can select the different experiments discussed in the paper; incid_expe=1 yields Figures 8-11 & 14; incid_expe=2 yields Figure 12; incid_expe=3 yields Figure 15a, data_n.txt, data_s.txt, mass_entry_n.txt, and mass_entry_s.txt; incid_expe=4 yields Figure 13; incid_expe=5 yields Figure 15b.

Additional Matlab codes:

1. fun_buyback.m: it computes the firm repayment policy upon receiving the transfer
2. fun_EntryEq.m: it computes the level of debt at entry and the value added that solves the FE and FC conditions in the paper.
3. fun_FEFC.m: it is an auxiliary file needed for running fun_EntryEq.m and fun_EntryEq_eval.m
4. get_calibration_region.m: it calibrates the parameters of each region (North and South) by matching to statistics reported in Figure 7
5. get_distribution_b.m: it computes the invariant distribution of the debt value ratio b
6. get_distribution_bZ.m: it computes the invariant joint distribution of the debt value ratio b and firm technology Z
7. get_equilibrium_north.m: it loads the calibration for the North at the initial steady state and solves for the steady state equilibrium in the Northern region
8. get_equilibrium_south.m: it loads the calibration for the South at the initial steady state and it solves for the steady state equilibrium in the Southern region for different policy values for government transfers; it calls the file get_transition_dynamics.m which computes transition dynamics between steady states
9. get_transition_dynamics.m: it computes the transition dynamics from an initial steady state to the final steady state.
10. get_figure_distribution.m: it produces Figure O6
11. get_figure_rachet_effect.m: it produces Figure 5 a-b, Figure 6 a-b; it also computes the percent increase in employment after the subsidy
12. get_figures_decomposition.m: it produces Figure 12 and O7
13. get_figures_policy.m: it produces Figures 3, 4, and O2
14. get_figures_ss.m: it produces Figure 8
15. get_figures_transition.m: it produces Figures 9, 10, 11, and Table 4
16. get_firm_policy.m: it computes the firm policy in steady state
17. get_firm_policy_analyticalcase.m: it computes the firm policy in the analytical case discussed in the paper

18. get_firm_policy_loopVA.m: it computes the firm policy as a function of market profitability and produces Figure 03.
19. get_firm_simulations.m: it simulates firms over their lifecycle
20. get_ISS_subsidy.m: it maps the ISS subsidy to parameters in the model
21. get_LCP.m: it solves the Linear Complementary problem (LCP)
22. get_matrices.m: it computes the transition matrices for the Hamilton-Jacobi-Bellman (HJB) equation
23. get_optimal_subsidy_dixit.m : it produces Figure 13
24. get_params.m: it sets the parameters as in Table 2
25. get_shakout_distribution.m: it computes the shift in the distributions immediately after the policy shock
26. get_sim_regression.m: it produces simulations of firms over their lifcycle needed to calculate Table 6
27. get_transition_dynamics.m: it produces the transition dynamics of the economy between steady states
28. read_pars_x.m: it sets the parameters as in Table 2
29. settt.m: auxiliary code needed for formatting and printing
30. vline.m: it creates vertical lines
31. get_sim_exit: It simulates
32. get_simulation_incidence.m: it produces Figure 15b by simulating firm exit rates through get_sim_exit