Instructions on the Data and Code for:
"Small and Vulnerable: SME Productivity in the Great Productivity Slowdown"*

Sophia Chen

International Monetary Fund

Do Lee

New York University

September 11, 2022

Contents

---

1 Instructions for Replication

The code in this replication package produces all the figures and tables in the main paper. The code was written for Stata MP version 14. To replicate the results of the paper:

1. Obtain the restricted data and place them in `./data/raw/` (Please see Section 3.2 for details)

2. Run the master code `./code/main.do`

2 Output

The master code calls other supporting `.do` files in the folder and produces the paper's main results:

1. Figures (`.pdf` files) saved in `./figures/`

   (a) `./figures/figure_1.pdf`: Figure 1 produced by `./code/figure_1.do`

      i. The figure uses data from `./data/cleaned/sample_tfp_2008.dta`, which is a annual panel of firms in the Orbis Historical database and their estimated TFP. Please see Section 3.3 for details on how to reproduce this data.

2. Tables (`.tex` files) saved in `./tables/`

   (a) `./tables/table_#.tex`: Table # produced by `./code/table_#.do` for $\# = 1, \ldots, 5$.

      i. The tables use data from `./data/cleaned/sample_did.dta`, which is a firm-level dataset containing all the variables used in the Difference in Differences regressions in this paper. Please see Section 3.3 for details on how to reproduce this data.

3 Data

This section describes the raw data required to replicate the results in this paper. Section 3.1 describes the publicly available data sources and their location within this replication package. Section 3.2 describes the restricted data sources.

Note: We do not have permission to redistribute the restricted data sources in this replication package. The data files from the restricted data sources have been replaced with pseudo data. The pseudo data clarifies the format of the data and demonstrates that the code runs without error. But otherwise, the pseudo data does not include any information that is protected by copyright or confidentiality agreements. Running the program with the pseudo data will produce outputs in the same format as in the paper with different numbers. Researchers interested in replicating this paper will need to obtain the restricted data from the original sources listed in the section.

3.1 Publicly available data

1. BIS US dollar exchange rates (End of period):
   `./data/raw/exchange_rates/WEBSTATS_XRU_CURRENT_DATAFLOW_csv_col.csv`

   (a) We use annual end of period US dollar exchange rates available from the BIS at `https://www.bis.org/statistics/full_xru_csv.zip`, downloaded on 06/25/2021.

   (b) `./code/clean_exchange_rates.do` converts the raw `.csv` file into Stata `.dta` format. These exchange rates are later used in `./code/clean_orbis.do` to convert nominal variables recorded in U.S. dollars from the Orbis Historical database to nominal local currency units.

2. Local currency deflators: `./data/raw/stan/STANI4_2020.csv`

   (a) The underlying data comes from the SNA08, ISIC REV.4 version of the OECD Structural Analysis Database (STAN), 2015 reference year. The data was downloaded as `.csv` files through the OECD website on 06/19/2021, available at `https://www.oecd.org/sti/ind/stanstructuralanalysisdatabase.htm`.

   (b) `./code/clean_stan.do` takes the raw `.csv` file as input and produces local currency deflators at the two-digit NACE Rev. 2 level in `./data/cleaned/STANI4_2020_NACE2.dta`. These deflators are later used in `./code/clean_orbis.do` to convert nominal local currency units from the Orbis Historical database to 2005 local currency units.

3. Purchasing Power Parity (PPP) exchange rates: `./data/raw/ppp/benchmark_2005.xlsx`

   (a) We use the "GDP and major sector value added relative prices for 2005, USA GDP=1" from Inklaar et al. (2005). The data file was downloaded on 06/17/2021 from the "Supporting Information" to the original paper: `http://onlinelibrary.wiley.com/doi/10.1111/roiw.12012/abstract`.

   (b) `./code/clean_ppp.do` converts the raw excel file into Stata `.dta` format, which is later used in `./code/clean_orbis.do` to convert 2005 local currency units from the Orbis Historical database to 2005 U.S. dollars.

3.2 Restricted data

1. Orbis Historical Database from Bureau van Dijk: `.dta` files in `./data/raw/orbis/`

   (a) Orbis is our main firm-level data source. The raw data has been downloaded on 04/30/2018 by connecting to a local SQL database at the IMF library.

(b) The files are saved as separate `.dta` files for each country where the suffix in each file name indexes the country's ISO code. There are three `.dta` files for each country. Files beginning with `financial` contains the historical financial dataset, which contains information on each firm's financial accounts and productive activities (e.g., output, capital, and employment). Files beginning with `industry_classifications` and `legal_info` contain the industry classification (NACE Rev. 2, SIC, and NAICS codes) and legal information (date of incorporation, listed/delisted status) for each firm in the Orbis historical database, respectively.

2. Firm-bank relationships from the AMADEUS database: `./data/raw/orbis/Banker.dta`

(a) This data file contains the BANKER variable in the AMADEUS database, which lists the name of up to five of the most important creditor banks for each firm. The bank names are used as an input to `./code/clean_fitch.do` and `./code/clean_markit.do` to bring in additional information about these banks.

(b) The original source of the information is Kompass, which provides directories for companies in more than 70 countries. Kompass collects data from firm registries, Chambers of Commerce, and phone interviews with firm representatives. Firms can also voluntarily register with the Kompass directory.

3. Fitch Connect: `./data/raw/fitch/bnk_hist_usd.dta`

(a) Fitch Connect contains financial data on banks. We use information about the regulatory Tier 1 capital (in percentage of Risk-Weighted Assets) for each bank in this database.

(b) The raw data has been downloaded from the IMF library on 08/24/2018 as flat files (complete database files) in Stata `.dta` format. `./code/clean_fitch.do` processes this raw data and matches the banks in Fitch Connect with banks in the AMADEUS database using a probabilistic record linkage algorithm. The Tier 1 capital from Fitch Connect are later used in `./code/table_4.do`.

4. Markit Credit Default Swap database: `./data/raw/markit/markit.dta`

(a) The Markit database contains bank CDS spreads. We use the end-of-day 5-year CDS spreads from banks that are incorporated in the countries covered in our sample.

(b) The raw data has been downloaded from Wharton Research Data Services (WRDS) on 02/17/2020 in Stata `.dta` format. `./code/clean_markit.do` processes this raw data and matches the banks in Markit with banks in the AMADEUS database using a probabilistic record linkage algorithm. These CDS spreads are later used in `./code/table_4.do`.

5. Relationship bank share: `.dta` files in `./data/raw/bepsii/`

   (a) We use data from the Banking Environment and Performance Survey (BEPS) collected in Beck et al. (2018), which classifies banks as relationship or transaction banks based on face-to-face interviews with bank chief executive officers. We obtained this survey data by contacting the authors of Beck et al. (2018) and signing a confidentiality agreement with the European Bank for Reconstruction and Development (EBRD).

   (b) `./code/clean_relationship_bank_share.do` takes the raw data files as input to calculate the share of relationship banks relative to total banks in each country. This relationship bank share data is later used in `./code/table_4.do`.

6. Worldwide Patent Statistical Database (PATSTAT): `./data/raw/patstat/patent_app.dta`

   (a) PATSTAT is maintained by the European Patent Office and is the most comprehensive cross-country patent database. It contains bibliographic data on patents from 90 patent-issuing authorities, covering close to the population of all patents worldwide.

   (b) `./code/clean_patent_applications.do` takes this data as input to calculate each firm's annual patent applications. We match applicant firms from PATSTAT to firms in the Orbis dataset using firm-patent matches provided by Bureau van Dijk. This patent data is later used in `./code/table_5.do`.

3.3 Data cleaning procedures

The following code combines the raw data into intermediate data files required to replicate the output:

1. `./code/clean_orbis.do` cleans the raw Orbis data in `./data/raw/orbis/.*_##.dta` and saves them in `./data/cleaned/cleaned_orbis_##.dta` where `##` indexes the ISO code for each country. The cleaned files contain panel data at the firm and year level.

2. `./code/estimate_tfp.do` takes the cleaned Orbis data, estimates firm-level productivity, and saves the estimates in `./data/cleaned/sample_tfp_2008.dta`. The productivity estimates are at the firm and year level.

3. `./code/make_did_sample.do` and `./code/merge_did_sample.do` produces `./data/cleaned/sample.dta`, the final sample for the Difference in Differences regressions in the paper. First, the code combines the cleaned Orbis data and the productivity estimates. Then the code transforms this combined firm-year level panel into firm-level averages over 4 year windows before (2004–2007) and after (2008–2011) the global financial crisis.

References

Beck, Thorsten, Hans Degryse, Ralph De Haas, and Neeltje van Horen, "When arm's length is too far: Relationship banking over the credit cycle," *Journal of Financial Economics*, 2018, *127* (1), 174–196.

Inklaar, Robert, Mary O'Mahony, and Marcel Timmer, "ICT and Europe's productivity performance: Industry-level growth account comparisons with the United States," *Review of Income and Wealth*, December 2005, *51* (4), 505–536.