"Capital Accumulation and Structural Transformation"
by Paula Bustos, Gabriel Garber, Jacopo Ponticelli
Readme file for Replication

Folders

- "Figures": by default, the code *CAST_Replication_Figures.do* saves figures into this folder.
- "Tables": by default, the code *CAST_Replication_Tables.do* saves tables into this folder.
- "Maps": contains the shapefiles and auxiliary files needed to plot maps in Stata.

Code

- *CAST_Replication_Tables.do* is a Stata do-file that replicates the tables in the paper and in the Online Appendix.
- *CAST_Replication_Figures.do* is a Stata do-file that replicates the figures in the paper and in the Online Appendix.
- The log-files *CAST_Replication_Tables.log* and *CAST_Replication_Figures.log* are created by the respective do-files.
- The code uses the databases inside the *ReplicationFiles* folder. If the do-files are opened directly inside this folder, they should run without the need to re-direct the path.

Databases

- *CAST_data1.dta:* database with agricultural census outcomes at municipality level, unit of observation is AMC/year (2006)
- *CAST_data2.dta:* database with banking sector outcomes at municipality level, unit of observation is AMC/year (1996 to 2010)
- *CAST_data3.dta*: database with bank-level outcomes, unit of observation is bank/year (1996 to 2010)
- *CAST_AMC_codes.dta* database with the correspondence between municipalities and AMC.
- We also include a set of smaller auxiliary databases to replicate Figures 1, 2, 3, 4, 6, C8. These datasets are named*: fig1.dta, fig2.dta, fig3a.dta, fig3b.dta, fig4.dta, fig6.dta* and *figC8.dta*.

Data sources and definition of variables

FAO-GAEZ data on soy potential yields under different technological assumptions can be downloaded from the FAO website: http://gaez.fao.org/Main.html.

Data from the Agricultural Census at municipality level is available online from the website of the Brazilian statistical institute IBGE at: http://www.sidra.ibge.gov.br/.

Data on deposits and loans at municipality level from the ESTBAN dataset is available from the Central Bank of Brazil website at the following link: https://www.bcb.gov.br/

Loan-level data is sourced from the Credit Registry (SCR) of the Central Bank of Brazil. Firm-level variables are constructed using the employer-employee dataset RAIS, which is also available at the Central Bank of Brazil. These two datasets can be merged using firm unique identifiers in the computers of the Central Bank of Brazil. These datasets are not publicly available, but interested researchers can obtain access to the data by contacting the Research Department of the Central Bank of Brazil (DEPEP). Information on external collaborations with the Central Bank of Brazil can be found at this link: https://www.bcb.gov.br/en/publications/externalcollaborationresearch.

Interested researchers can contact the bank to get access to the data at this link: https://www3.bcb.gov.br/faleconosco/#/contact-us

We are happy to assist researchers that want to re-create our final datasets starting from raw data.

A note on Brazilian geographical identifiers

Borders of Brazilian municipalities often change. To make them comparable overtime, IBGE has defined Áreas Mínimas Comparáveis (AMCs), or smallest comparable areas. This is our unit of observation for most of the analysis. The file *CAST_AMC_codes.dta* contains the correspondence between AMC and municipality codes. The original correspondence between municipality codes and the AMC codes is maintained by the government agency Instituto de Pesquisa Econômica Aplicada (IPEA). The relevant correspondence for this paper can be downloaded at: http://www.ipeadata.gov.br/doc/Municipios1997_AMCs.zip. Currently (December 2019), this correspondence is updated to the year 2000. Between 2000 and 2010, the Brazilian government allowed the formation of new municipalities. These new municipalities are created by dividing the area that in 2000 belonged to a single municipality into 2 or more smaller units assigned to different municipalities. In order to correctly match geographical units after the year 2000, we updated the AMC correspondence to account for the change in borders.

The file *CAST_AMC_codes.dta* contains the correspondence between municipalities and AMCs that we use in the paper. This is the same correspondence file used in Bustos et al. (2019). The variables *cod_uf*, *cod_munic*, *cod_amc* are official IBGE codes for Federal Units, municipalities and AMCs. The variable *AMC* is a numeric variable that identifies AMC in the datasets used to replicate our results. The variable *flag* is a dummy that is equal to 1 whenever we updated the AMC to take into account changes in borders that happened between 2000 and 2010.

References

Bustos, P., B. Caprettini, and J. Ponticelli (2016). “Agricultural Productivity and Structural Transformation: Evidence from Brazil". American Economic Review 106 (6), 1320-1365