Replication files of:

Agricultural Productivity and Structural Transformation Evidence from Brazil

Paula Bustos, Bruno Caprettini, Jacopo Ponticelli

These folders contain code and files sufficient to replicate tables and figures in the paper "Agricultural Productivity and Structural Tranformation. Evidence from Brazil".

Folders

- "Figures": by default, the code *APST_replication figures.do* saves into this folder all the figures reported in the paper.
- "Tables": by default, the code *APST_replication tables.do* saves into this folder all the tables reported in the paper.
- "Maps": stores shapefiles and auxiliary files needed to plot maps in Stata. Inside this folder there are 2 maps: one at the AMC (municipality) level: *Brazil_AMC.XXX*; and one at micro-region level: *Brazil_microregion.XXX* (where XXX is one of: "cpg," "dbf," ".prj," ".sbn," ".sbx," ".shp," ".shp.xml," ".shx," and ".xml"). *All* of these files are needed to plot the maps.

Code

- *APST_replication tables.do* is a Stata do-file that replicates tables in the main body of the paper as well as tables in the appendix.
- *APST_replication figures.do* is a Stata do-file that replicates figures in the main body of the paper as well as figures in the appendix.
  Note: this do-file automatically downloads 2 user-written programs: *shp2dta.ado* and *spmap.ado*. They are needed to produce the maps in figures 2, 3 in the paper and figure A7 in the appendix. If these two programs are already installed, the code does not overwrite them.

- the log-files *APST_replication tables.log* and *APST_replication figures.log* are created by the respective do-files.
- *ols_spatial_HAC.ado* is called inside the file *APST_replication tables.do* and it is needed to estimate standard errors corrected for spatial dependence as suggested by Conley (1999) for tables A10, A11 and A12.

The code uses the databases inside the folder. If the do-files are opened directly inside this folder, they should run without the need to re-direct the path.

Databases

- *APST_AMC.dta* database used to produce all the municipality-level estimates.
- *APST_micro.dta* database used to run all the micro-region-level estimates.
- *APST_figures.dta* database used to create all the figures except the maps and the 2 figures with the evolution of the price of soy and maize (figures A8 and A9 in the appendix).
- *APST_prices.dta* database used to create 2 figures with the evolution of the price of soy and maize (figures A8 and A9 in the appendix).
- *APST_AMC_codes.dta* database with the correspondence between municipalities and AMC.

Data sources and definition of variables

See appendix, section A4 for a detailed description of the way we construct every variable in the final datasets.

FAO-GAEZ data on soil suitability of soy and maize under different assumptions of technological level can be downloaded from the FAO website: http://gaez.fao.org/Main.html.

Data from the Brazilian agricultural Census at the municipality level is freely available online from the IBGE Sidra repository at: http://www.sidra.ibge.gov.br/.

Data from the Brazilian Population Census for the years 1980-2010 is proprietary and can be purchased on the website of the Brazilian Statistical office at: http://loja.ibge.gov.br/.

Firm-level data we use for the robustness check shown on Table A9 in the appendix come from the Pesquisa Industrial Anual (PIA), the Yearly Industrial Survey carried out by the Instituto Brasileiro de Geografia e Estatística (IBGE), the Brazilian National Statistical Office. This firm-level dataset is not publicly available, but interested researchers can obtain access to the data in the IBGE office of Rio de Janeiro by contacting the Centro de Documentaçao e Disseminaçao de Informaçoes - CDDI of the IBGE in Rio de Janeiro.

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 Área Mínima Comparável (AMC), smallest comparable areas. This is our unit of observation for most of the analysis. It is also the level of aggregation of the data stored in the file *APST_AMC.dta* and in the map *Brazil_AMC.shp*.

In this folder we provide the correspondence between AMC and municipalities that we use in our analysis. The original correspondence between municipality codes and the AMC codes is maintained by the government agency Instituto de Pesquisa Econômica Aplicada (IPEA). The most recent correspondence can be downloaded at: http://www.ipeadata.gov.br/doc/Municipios1997_AMCs.zip. Currently (December 2015), this correspondence is updated to 2000, when there were 5'507 municipalities. Between 2000 and 2010, the Brazilian government allowed the formation of 57 new municipalities: all of these new municipalities appear in the Population Census of 2010, and some of them also in the Agricultural Census of 2006. Moreover, all of these new municipalities are created by dividing the area that 2000 belongs to a single municipality into 2 or more smaller units assigned to different municipalities. In order to match correctly geographical units after the year 2000, we updated the AMC correspondence to account for the change of borders.

The file *APST_AMC_codes.dta* contains the correspondence that we use. The variables *cod_uf*, *cod_munic*, *cod_amc* are official IBGE codes for Federal Units, municipalities and AMC. The variable *AMC* is a numeric variable that identifies AMC in the database *APST_AMC.dta*. 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.