January 05, 2024

Replication materials for

"The Unintended Consequences of Infrastructure Development" Review of Economics and Statistics

By Antonella Bancalari

1. Software used:

Stata/MP 16.1 for Mac (Apple Silicon), Revision 11 Oct 2022, MP-Parallel Edition.

Master do-file (Master.do) installs a series of user-written commands. Internet connection is required to download and install these commands.

2. Overview:

The data and Stata do-files provided replicate the figures and tables of the paper "The Unintended Consequences of Infrastructure Development".

The user should create a folder with the following structure:

./Code/	- Folder with all the do-files provided (8 in total)
./Data/	- Data folder
./Data/APPENDIX	- Subfolder with additional data required to replicate the Appendix (UN_*.dta and MINSA_births.dta)
./Data/GIS	- Subfolder with spatial data
./Graphs/	- Folder where graphs are stored
./Tables/	- Folder where Tables are stored
./Graphs/APPENDIX	- Subfolder where graphs for Appendix are stored
./Tables/ APPENDIX	- Subfolder where Tables for Appendix are stored

To run the code, the user must start the file "Code/Master.do", having first created the path containing the folders listed above.

"Code/Master.do" first installs user-written programs which are not part of the standard package of Stata. Internet connection is required to download and install the corresponding files. To install these additional commands, the user must not comment out "setup_proj" in line 52 of the master do-file.

The output shown in the paper is based on the following version of the commands:

estout - version 3.30 25 Mar 2022

coefplot - version 1.8.5 23 Nov 2021

ivreg - version 2 4.1.11 22 Nov2019

spmap - version 1.3.2 - 09 Jan 2018

reghdfe - version 5.7.3 13 Nov 2019

event_plot – version 01 June 2021

did_multiplegt – version 2019

drdid - version 1.71 bug with weights

csdid – version v1.72

eventstudyinteract - version 0.1 24 Jan 2022

heatplot - version 1.0.6 13 Oct 2020

bacondecomp – version 1.0.2 15 Sep 2019

"Code/Master.do" next sets *globals* that are necessary to replicate the exhibits in the paper. The user must make sure that these *globals* are set before running the do-files.

Execution of the master do-file calls the other dofiles, which must run in the following order:

```
./Code/0_programs.do        //author-written programs
./Code/1_vitalstats.do      //set-up vital statistics
./Code/2_district.do        //set up district-level data
./Code/3_projects.do        // set up project data
./Code/4_setup.do           //set up final dataset
./Code/5_analysis.do        //provides results presented in the paper*
./Code/6_appendix.do        //creates exhibits for Appendix*
```

* Both codes clearly indicate which programs correspond to what results in the paper and Appendix.

3. Data

The folder "./Data/" contains all the input data required to construct the main dataset and the clean dataset created to run the analyses.

3.1. Input data:

- **minsa_deaths_append.dta:** this file contains mortality data by ICD-10 category at the district-level between 2003 and 2015. This information was provided by the Ministry of Health of Peru after filing a formal request for information.

- **inei_poblacion_5y.dta:** this file contains data on population forecasts for children under the age of five at the district-level between 2005 and 2015. The information was provided by the Institute of Statistics and Informatics (INEI).
- **minsa_births.dta:** this file contains births data at the district-level between 2003 and 2015. This information was provided by the Ministry of Health of Peru after filing a formal request for information.
- **mvcs_devengado_analysis.dta:** this file contains data on all sewerage infrastructure projects started between 2003 and 2016. It is a panel spanning 2003-2016 at the project-level. The information was provided by the Ministry of Housing, Construction and Sanitation after filing a formal request for information and meeting the data officers in person.
- **geo.dta:** this file contains cross-sectional geographical data at the district-level. The spatial files carrying this information were provided by the Ministry of Environment of Peru after filing a formal request for information. The spatial data is stored in the sub-folder **"GIS".** The district-level information was extracted using QGIS version 3.16.4 – Hannover for macOS 10.16.

3.2. Input data used for Appendix:

- **ineipop.dta:** this file contains data on population forecasts for all ages at the district-level between 2005 and 2015. The information was provided by the Institute of Statistics and Informatics (INEI).
- **censo05_popu1y5y.dta:** this file contains data on infant and under-five population from the 2005 Peruvian Census aggregated at the district-level from the individual-level data provided by the Institute of Statistics and Informatics (INEI). Similar data can be obtained from REDATAM in the INEI portal: http://censos1.inei.gob.pe/Censos2005/redatam
- **censo05.dta:** this file contains data on selected variables from the 2005 Peruvian census aggregated at the district-level from the household-level data provided by the Institute of Statistics and Informatics (INEI). Similar data can be obtained from REDATAM in the INEI portal: http://censos1.inei.gob.pe/Censos2005/redatam
- **censo.dta:** this file contains data on selected variables from the 2005, 2007 and 2017 Peruvian census at the district-level obtained from REDATAM in the INEI portal at:
  http://censos1.inei.gob.pe/Censos2005/redatam
  http://censos1.inei.gob.pe/Censos2007/redatam/
  https://censos2017.inei.gob.pe/redatam/
- **renamu.dta:** this file contains data on selected variables from the National Registry of Municipalities (RENAMU) at the district-level spanning 2004-2016. The data is publicly available from the INEI Microdatos portal: https://proyectos.inei.gob.pe/microdatos/
- **siaf_expenditure.dta:** this file contains data on municipal expenditure at the district-level spanning 2007-2015 from the National System of Financial Administration (SIAF) of the Ministry of Economy and Finance. The data is publicly available from the MEF portal "Consulta Amigable – Consulta de Ejecucion del Gasto": https://apps5.mineco.gob.pe/transparencia/Navegador/default.aspx

- **jne_indic.dta:** this file contains political data at the district-level for the local elections since 1993. The information was provided by the National Jury of Elections (JNE) after filing a formal request for information.
- **APPENDIX:** this folder contains data downloaded in 2018 from the UN Inter-Agency Group for Child Mortality Estimation https://childmortality.org/data/Peru
  Additionally, it contains the file "**MINSA_births.dta**" with births at the national level computed by collapsing for each year the **minsa_births.dta** described above.

3.3. Clean data used for analysis:

- **analysis_district_clean_all.dta**: final dataset used for the main analysis of the paper with the sample of analysis.

3.4. Clean data used for Appendix:

- **projects.dta**: clean dataset at the project-level used in the Appendix to compute project-level descriptive statistics.
- **analysis_district_clean.dta:** clean dataset used in the Appendix, before restricting to the final sample of analysis.