

README – Replication Package

Paper: Higher Education Expansion and Firm Organization

Authors: Xavier Cirera, Marcio Cruz, and Antonio Martins-Neto

Overview

This replication package provides the Stata codes used to reproduce the results of the paper Higher Education Expansion and Firm Organization.

The package does not include the underlying microdata (RAIS, Higher Education Census, BADEPI, and SECEX/MDIC firm-trade database), which are either restricted-access or publicly available through external portals. Instructions for obtaining these datasets are provided below.

Folder Structure

The main folders included in the package must be located in the same directory for the replication to run correctly:

- Program/
- Inputs/
- Outputs/
- Censo Ensino Superior/

Within Outputs/ and Censo Ensino Superior/ there are additional required subfolders created automatically by the codes.

Dependencies

The analysis was conducted using Stata 17. The following Stata packages must be installed prior to running the codes:

- `ssc install reghdfe`
- `ssc install ftools`
- `ssc install outreg`
- `ssc install drdid`
- `ssc install csdid`
- `ssc install cem`
- `ssc install event_plot`
- `ssc install eventstudyinteract`
- `ssc install avar`
- `ssc install did_imputation`
- `ssc install estout`

How to Run

1. Update all directory paths in the master file O.Main dofile.do.
2. Run the do-files in the sequence described below.
3. Figures and tables will be generated in the Outputs/ folder.

Code Description

Stata

0. Clean_Census.do – Cleans the Higher Education Census, defines treated microregions, and prepares Figures 2 and 3A.
1. Establishment level information.do – Cleans RAIS and collapses data to the establishment level.
- 1s. Matching 1-1 and estimates.do – Performs establishment-level matching and estimates results for Figure A1, Figures 5–9, and Figures A2–A4.
2. Matching 1-1 less restrict.do – Performs coarsened exact matching with less restrictive criteria; generates Appendix B figures (B1–B2).
3. Descriptive matched.do – Produces descriptive statistics, Table 2, Table A1, and Table B1.
4. Productivity.do – Estimates productivity regressions (Table 3).
5. Figures export and patent.do – Prepares Figure 4.

Python

Maps.py is a Python script that generates the map in Figure 1. The code uses inputs from the 0. Clean_Census.do, so please run the Stata codes.

Data Sources

This package relies on multiple administrative and statistical datasets. The data are not included in this package but can be obtained from the following sources:

RAIS – Relação Anual de Informações Sociais: Annual census of formal workers administered by Brazil's Ministry of Labor. Contains employer and employee identifiers, municipality and industry codes, wages, and demographics. Access is restricted and was obtained through a formal agreement between the Ministry of Labor and the World Bank. Not included in replication package.

Accessed: October 2019

Higher Education Census – Censo da Educação Superior: Annual census of higher education conducted by Brazil’s Ministry of Education. Covers students, faculty, institutions, enrollment, graduation, and fields of study.

Available at: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>

Accessed: May 2024

BADEPI – Base de Dados Estatísticos de Propriedade Industrial: Statistical database of intellectual property maintained by INPI. Includes patents, trademarks, industrial designs, software registrations, and geographical indications. Raw data downloaded from the data URL was cleaned by consolidating multiple raw files into a harmonized dataset of patent applications, applicants, inventors, and classifications, creating variables for the year of application filing, year of granting, and year of rejection, and generating applicant-level counts of total and granted patents. Records without firm identifiers or foreign entities are dropped. The file "BADEPI.dta" contains one observation per patent application with applicant, inventor, and classification information.

Available at: <https://www.gov.br/inpi/pt-br/inpi-data/dados-e-series-temporais/badepi>

Accessed: June 2023

SECEX/MDIC – Importing and Exporting Firms Database: Establishment-level data identifying Brazilian firms that engaged in imports and exports between 1997 and 2021. Variables include CNPJ, firm name, legal nature, CNAE code, municipality, state, and trade activity (export, import, or both). Raw data was imported into Stata and saved into the file "Exporters_Importers.dta".

Available at: <https://basedosdados.org/dataset/5ebd8129-ae10-4ec6-bf97-8a43d0166fdb?table=253313e1-c0e9-402b-8649-986a4b94f654>

Accessed: June 2022

Citation

If you use these codes, please cite:

Cirera, Xavier; Cruz, Marcio; and Martins-Neto, Antonio. Higher Education Expansion and Firm Organization. [Year].