Overview

The code in this replication package runs the codes using Stata to generate the 11 main figures, 6 main tables and every figure and table in the appendices in the paper using several data sources. The replicator should expect the code to run for about one day and a half.

Data Availability and Provenance Statements

The data for this project are confidential, and it was obtained with Data Use Agreements with the Department of National Planning (DNP) of Colombia. DNP linked multiple sources of information into a single database, including information from the SABER 11 exams, the third version of Colombia's proxy-means testing instrument (SISBEN III), the census of postsecondary attendance (SNIES), the Saber Pro and Saber TyT college graduation exams, and social security records (PILA). DNP does not have a specific protocol to grant access to data to individual researchers. Researchers interested in accessing these datasets can submit their request to Patricia Moreno (pmoreno@dnp.gov.co) or contact these institutions directly. It can take several months to negotiate data use agreements and access the data. The author will assist with any reasonable replication attempts for a year following publication.

Summary of Availability

- ☐ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ✓ **No data can be made** publicly available.
- ☐ Confidential data used in this paper and not provided as part of the public replication package will be preserved for __ years after publication, in accordance with journal policies.

Details on each Data Source

Main datasets are described below:

Data file	Source	Notes	Provided
`PILA.dta`	DNP, Minsalud	Confidential. Contains labor market outcomes of the study population	
`Matricula_graduados_2013_2020_red.dta`	DNP, Mineducacion	Confidential. Contains information about who is in the higher education system (associate and bachelor's degree program)	

Matricula_posgrado_2013_2020_red	DNP, Mineducation	Confidential. Contains information about who is in the higher education system (graduate degree program)
S11_20101_20202.dta	DNP, Icfes	Confidential. Contains information about test takers of Saber 11 exam
saber_pro_tyt_Spro_2012_2021.dta	DNP, Icfes	Confidential. Contains information about test takers of Saber TyT and Saber Pro exams
arreglo_GE.dta	DNP, ICETEX	Confidential. Contains information about Generación E beneficiaries
arreglo_pilo.dta	DNP, ICETEX	Confidential. Contains information about Ser Pilo Paga beneficiaries
cruce_saber11_sisben_iii.dta	DNP	Confidential. Contains the merge between S11_20101_20202 and SISBEN III databases
base_maestra.dta	DNP	Confidential. Contains the linked data
data_RD.dta	DNP	Confidential. Contains the filtered dataset used for all analysis.

Computational requirements

Software Requirements

- ✓ The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.
- Stata (code was last run with version 18)
  - egenmore
  - reghdfe
  - ereplace
  - gtools
  - ftools
  - rdrobust
  - winsor2

The program "`Code/Master.do`" will install all dependencies locally and should be run once.

Memory, Runtime, Storage Requirements

Summary

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-2 hours
- ☐ 2-8 hours
- ☐ 8-24 hours
- ✓ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days

Approximate storage space needed:

- ☐ < 25 MBytes
- ☐ 25 MB - 250 MB
- ☐ 250 MB - 2 GB
- ☐ 2 GB - 25 GB
- ✓ 25 GB - 250 GB
- ☐ > 250 GB
- ☐ Not feasible to run on a desktop machine, as described below.

Details

The code was last run on an Intel(R) Xeon(R) w5-2455X 3.19GHz workstation with 500GB of free space

Description of programs/code

- Programs in `Code/Output` generate all tables and figures in the main body of the article. The file `Code/Master.do` will run them all. Each program called from `Master.do` identifies the table or figure it creates.

- Several of the tables are exported in dta or xlsx format. These files are used to create the tables in the document.

Instructions to Replicators

- Edit `Code/Master.do` to adjust the default path in line 28.
- Run `Code/Master.do` to run all steps in sequence.

List of tables and programs

The provided code reproduces:

- ☐ All numbers provided in text in the paper
- ✓ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.