Replication files for the paper

Egger, Eva-Maria (2021), "Internal Migration and Crime in Brazil"

All analysis was executed in STATA 16.

Original data sources by variables are:

- Homicide rates: Sistema de Informacao sobre Mortalidade (SIM-DATASUS) freely accessible via Instituto de Pesquisa Econômica Aplicada (IPEA) https://www.ipeadata.gov.br
- Migration: Censo Demografico - available freely at the website of the Brazilian National Institute of Geography and Statistics

  https://www.ibge.gov.br/estatisticas/sociais/trabalho/9662-censo-demografico-2010.html?=&t=microdados (Data has to be downloaded for each state.)
- Local labor market outcomes: Pesquisa Nacional de Amostra de Domicilios (PNAD) - available freely at the website of the Brazilian National Institute of Geography and Statistics

  https://www.ibge.gov.br/estatisticas/sociais/trabalho/19897-sintese-de-indicadores-pnad2.html?=&t=microdados
- Manufacturing sector employment and wages: Relação Anual de Informações Sociais (RAIS) - available freely through https://basedosdados.org/dataset/br-me-rais ("Versão Pública")
- Other variables (e.g. population, GDP) downloaded from Instituto de Pesquisa Econômica Aplicada (IPEA) at Microregiao level via https://www.ipeadata.gov.br

The following files are available for replication of the study:

- Code in order of required execution:
  1. "dataprep_individuals.do": Do-file documenting the cleaning of the Census data and creation of migration variables at individual level. Code for transparency, datasets have to be downloaded.
  2. "dataprep_RAIS.do": Do-file to import RAIS data and construct local labor demand shocks.
  3. "dataprep_PNAD.do": Dofile combining data from all PNAD rounds 2001 to 2009 and then aggregated to microregiao level. Code for transparency, datasets have to be downloaded.
  4. "dataprep_MRlevel.do": Do-file combining all data from different sources to construct panel of microregioes with main variables. Code for transparency, datasets have to be downloaded.
  5. "dataprep_MRlevel_distance": Do-file creating instrumental variable and migration rates with different distance cut-offs. Code for transparency, datasets from preceeding files.
  6. "dataprep_combineall": Do-file combining all data to analysis files. Code for transparency, datasets resulting from preceeding files.

7. "Main analysis.do": The do-file contains all code to produce tables and figures of the paper. It uses the shared datasets (see below).

- Data files for replication of analysis (do-file "Main analysis.do"):
  - "MR_analysis": Main data set at microregiao level to conduct main analysis.
  - "Individuals_analysis": Data at individual level to conduct analysis of channels.
  - "MR_GPSS": Data at microregiao level in format to compute Rothemberg weights as in Goldsmith-Pinkham et al. (2020)
  - "MR_GPSStests": Data at microregiao level in format to run misspecification and overidentification tests.
  - "MR_coord": Longitude and latitude coordinates to create maps.

Contact: egger@wider.unu.edu