

Territorial Inequalities: A Note on State Discontinuity

Lidia Ceriani, Luis F. Lopez-Calva, and Samuel D. Restrepo-Oyola

March 2025

1 Overview

This note introduces the package that reproduces all results presented in the working paper “*Territorial Inequalities: A Note on State Discontinuity*” by Lidia Ceriani, Luis F. Lopez-Calva, and Samuel D. Restrepo-Oyola. We explain how to download raw data from the Integrated Public Use Microdata Series (IPUMS) International, the Armed Conflict Location & Event Data (ACLED), and other complementary sources. We then explain how replicators can execute the code to reproduce the results. The figures, tables, and numbers in the paper were generated using Stata, while all maps were created in R. We encourage you to first download the reproducibility folder to your computer.

2 Data Availability

All data are publicly available, but some cannot be redistributed or published.

2.1 IPUMS International

IPUMS International is the world’s largest collection of publicly available individual-level census data. However, a clause prohibits the redistribution of any microdata extracted from the IPUMS International database, as it is intended solely for use by the researcher granted the license to access and download the data. The dataset used to replicate all our results was obtained in March 2024. To download the data, please follow these steps:

2.1.1 Step 1: Apply for Access

To access IPUMS International data, we first need to apply for restricted microdata access. You can submit your application using the following link: [Apply for access](#)

2.1.2 Step 2: Log In

Once you register and gain access to the restricted microdata, you will be able to [Log In](#)

2.1.3 Select Samples and Data

The next step is to navigate to the SELECT DATA window. There, you’ll find options to select harmonized variables and samples.

The samples to be selected for replicating the main analysis are as follows:
Armenia 2001; Benin 2013; Bolivia 2012; Brazil 2010; Cambodia 2013; Colombia 2005; Costa Rica 2011; Dominican Republic 2010; Ecuador 2010; El Salvador 2007; Ghana 2010; Guatemala 2002; Haiti 2003; Honduras 2001; Iran 2006; Laos 2005; Lesotho 2006; Liberia 2008; Mali 2009; Mexico 2015; Mozambique 2007; Nepal 2011; Nicaragua 2005; Peru 2007; Rwanda 2002; Senegal 2002; Sierra

Leone 2015; Sudan 2008; Tanzania 2012; Togo 2010; Venezuela 2001; Zambia 2000.

We recommend selecting the following set of harmonized variables for replicating the main findings: COUNTRY; YEAR; SAMPLE; SERIAL; HHWT; GEO1_AM2001; GEO1_BJ2013; GEO1_BO2012; GEO1_BR2010; GEO1_KH2013; GEO1_CO2005; GEO1_CR2011; GEO1_DO2010; GEO1_EC2010; GEO1_SV2007; GEO1_GH2010; GEO1_GT2002; GEO1_HT2003; GEO1_HN2001; GEO1_IR2006; GEO1_LA2005; GEO1_LS2006; GEO1_LR2008; GEO1_ML2009; GEO1_MX2015; GEO1_MZ2007; GEO1_NP2011; GEO1_NI2005; GEO1_PE2007; GEO1_RW2002; GEO1_SN2002; GEO1_SL2015; GEO1_SD2008; GEO1_TZ2012; GEO1_TG2010; GEO1_VE2001; GEO1_ZM2000; ELEC-TRIC; WATSUP; PERNUM; PERWT; CHBORN; CHSURV; LIT.

Please note that the variables COUNTRY, YEAR, SAMPLE, SERIAL, HHWT, PERNUM, and PERWT will be automatically added to your cart once you select your samples and variables of interest. **Importantly**, please remember that IPUMS International may update its datasets, and re-running the code for replication could yield slightly different results.

2.1.4 Step 4: Create a Data Extract

Once you have customized your selection of variables, you will need to CREATE A DATA EXTRACT and submit it. IPUMS International will then create a dataset based on your specifications and will notify you when it is ready for download.

2.1.5 Step 5: Download the Data and Command Files

On the My Data page (Download or Revise Extracts), you'll find your extract requests along with their creation dates, optional descriptions, and various links on the right side, including data, codebook, SPSS, SAS, STATA, and R. To download the data file and the command file, click on the "DOWNLOAD .DAT". To download the command file, right-click on the links labeled "SPSS," "SAS," "Stata," or "R," and select "Save Link As."

2.1.6 Step 6: Decompress the Data File

The downloaded data file should have the suffix ".dat.gz" (e.g., "ipumsi_00001.dat.gz"). You need to decompress this file. After decompression, you will see a file with the suffix ".dat" (the ".gz" part will be removed). Make a note of the path to the ".dat" file on your computer.

2.1.7 Step 7: Modify the Command File and Save the Data

Let us assume we are working on a machine running Windows. The next step is to update your command file to reflect the location of the ".dat" file on your computer. Instructions may vary slightly depending on the statistical package you're using.

The example below assumes you have a data file named "ipumsi_00001.dat" located in a folder on your "C:" drive called "Reproducibility_Package\Data\Raw\ipumsi_00001.dat," resulting in the full path being:

```
"C:\Reproducibility_Package\Data\Raw\ipumsi_00001.dat\ipumsi_00001.dat"
```

Using Stata

Open Stata. Then change the directory to where your ".dat" and ".do" files are stored. It's recommended to keep both the do file and the .dat file in the same folder. If your files are saved in the "Reproducibility_Package\Data\Raw\ipumsi_00001.dat" folder on your C drive, type:

```
cd "C:\Reproducibility_Package\Data\Raw\ipumsi_00001.dat"
do ipumsi_00001.do
```

For additional information, please refer to the [webpage](#), especially if you are using a different operating system or statistical software such as SAS, SPSS, or R. The code will run for about 15 minutes on a laptop with 32 gigabytes of RAM. Once you run the code, you'll generate a dataset that can be manipulated using your chosen software. This dataset will contain the information you requested. We recommend saving the dataset once the code execution is complete. We have saved this dataset as "ipums_countries_final.dta" in the Data\Raw folder. The size of this dataset is approximately 12 gigabytes. We accessed and created this dataset in March 2024. After saving the dataset, you can delete the file ipumsi_00001.dat.gz, along with the folder ipumsi_00001.dat, which contains the do file and the .dat file.

2.1.8 Additional Dataset for the Econometric Model in Section 5

From here on, we assume you already have an account on IPUMS International. To download the dataset used for our econometric model in Section 5, you should select the following samples: Bolivia 2012; Brazil 2010; Colombia 2005; Costa Rica 2011; Ecuador 2010; El Salvador 2007; Guatemala 2002; Mexico 2015; Nicaragua 2005.

We recommend also downloading the following harmonized variables: COUNTRY; YEAR; SAMPLE; SERIAL; HHWT; URBAN; POPDENSCEO1; AREAMOLLWCEO1; GEO1_BO2012; GEO1_BR2010; GEO1_CO2005; GEO1_CR2011; GEO1_EC2010; GEO1_SV2007; GEO1_GT2002; GEO1_MX2015; GEO1_NI2005; ELECTRIC; WATSUP; PERNUM; PERWT; CHBORN; CHSURV; NATIVITY; INDIG; LIT.

After tailoring your dataset, please follow steps 4 through 7 as previously mentioned. In other words: create a data extract, download the data and command files, decompress the data file, modify the command file, and save the data. We used Stata and saved the dataset as:

"ipums_countries_controls_final.dta" in the Data\Raw folder. The size of this dataset is approximately 3 gigabytes. We accessed and created this dataset in March 2024. After saving the dataset, you can delete the file with the "dat.gz" suffix, along with the folder containing the do file and the .dat file.

2.2 ACLED

Data from ACLED are publicly available, but users cannot provide, permit, or allow direct access to any of ACLED's original/raw data to any other user. To download the data, use the data-export tool. First, create a free account through the ACLED Access Portal and generate a unique access key. You can find all relevant information at the link below.

- **File name:** ACLED_2023-01-01-2023-11-07.csv
- **URL:** <https://acleddata.com/data-export-tool/>
- **Access year:** March 2024
- **Variable names:** When using the data-export tool you should keep the "ALL" option activated for the following variables: Event Type; Sub Event Type; Actor Type; Actor; Region; Country; and Location. The time frame considered (yy/mm/dd) goes From: 2023-01-01; To: 2023-11-07

Importantly, please remember that ACLED may update its datasets, and re-running the code for replication could yield slightly different results.

2.3 Complementary data

We also use ISO code information from the World Development Indicators (WDI). This dataset is provided as `CLASS.xlsx` in the `Data\Raw` folder. The information is publicly available.

- **File name:** Current classification by income
- **URL:** <https://datacatalog.worldbank.org/search/dataset/0037712>
- **Access year:** December 2024

To create the maps in the Annex, we also downloaded first-level geography shapefiles from IPUMS International. These files have been placed in the `Data\Raw\Shapefiles` folder and are publicly available. If you wish to download the files again, please ensure the following files are placed in the `Data\Raw\Shapefiles` folder:

- **File names:** Armenia 2001 (`geo1_am2001.zip`); Benin 2013 (`geo1_bj2013.zip`); Brazil 2010 (`geo1_br2010.zip`); Costa Rica 2011 (`geo1_cr2011.zip`); Honduras 2001 (`geo1_hn2001.zip`); Haiti 2003 (`geo1_ht2003.zip`); Laos 2005 (`geo1_la2005.zip`); Mali 2009 (`geo1_ml2009.zip`); Mozambique 2007 (`geo1_mz2007.zip`)
- **URL:** [IPUMS GIS Data](#)
- **Access year:** December 2024

3 Instructions for Replicators

Replicators should follow these steps to successfully run the package:

- Please download the reproducibility package to your computer.
- Download data as outlined in Section 2. After completing those steps, you should have three additional datasets in the `Data\Raw` folder, along with the one provided in the package. Therefore, the `Data\Raw` folder in the replication package should contain the following files and folder:

- ☐ Shapefiles
- ☐ `ACLED_2023-01-01-2023-11-07.csv`
- ☐ `CLASS.xlsx`
- ☐ `ipums_countries_controls_final.dta`
- ☐ `ipums_countries_final.dta`

- Update the following files with your directory path in the `Code` folder:
 - `Main_dofile.do` (line 19)
 - `Annex_Maps.R` (line 19)
- The `Main_dofile.do` will install all necessary dependencies, but only if they are not already installed.
- The `Main_dofile.do` will also run all subscripts and produce all the tables and figures from the main text, as well as the numbers mentioned in Section 4.2.1 (pages 12-13) and all the tables from the Annex.
- Run the `Main_dofile.do` script.
- Next, run `Annex_Maps.R` to reproduce the maps in the Annex. This script will install any required dependencies as needed but only if they are not already installed.

4 List of Exhibits

Exhibit name	Output filename	Script	Note
Table 1	-	-	Manually created
Table 2	-	-	Manually created
Table 3	tab3_countries.csv	05_Analysis.do	Found in Outputs\Main\Tables
Table 4	tab4_discontinuity.tex	05_Analysis.do	Found in Outputs\Main\Tables
Table 5	tab5_reg_dens_fe_pooled.tex	06_Regression_Analysis.do	Found in Outputs\Main\Tables
Table 6	tab6_reg_effect_fe_pooled.tex	06_Regression_Analysis.do	Found in Outputs\Main\Tables
Figure 1(a)	fig1_sankey_EAP.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 1(b)	fig1_sankey_ECA.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 1(c)	fig1_sankey_LAC.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 1(d)	fig1_sankey_MNA.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 1(e)	fig1_sankey_SAS.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 1(f)	fig1_sankey_SSA.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 2(a)	fig2_hbox_Density_r_5.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 2(b)	fig2_hbox_Density_r_10.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 2(c)	fig2_hbox_Density_r_20.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 3	fig3.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 4	fig4_Bumpgraph_Discontinuity.png	05_Analysis.do	Found in Outputs\Main\Figures
Figure 5(a)	fig5_bumpgraph_robust_05.png	08_Robustness.do	Found in Outputs\Main\Figures
Figure 5(b)	fig5_bumpgraph_robust_1.png	08_Robustness.do	Found in Outputs\Main\Figures
Figure 5(c)	fig5_bumpgraph_robust_2.png	08_Robustness.do	Found in Outputs\Main\Figures
Table A.1	-	-	Manually created
Table A.2	-	-	Manually created
Table A.3	tabA3_discontinuity_robust_check.csv	08_Robustness.do	Found in Outputs\Annex\Tables
Table A.4	tabA4_pairw_correl.doc	06_Regression_Analysis.do	Found in Outputs\Annex\Tables
Figure A.1	A1_cr.png	Annex_Maps.R	Found in Outputs\Annex\Maps
Figure A.2	A2_ml.png	Annex_Maps.R	Found in Outputs\Annex\Maps
Figure A.3	A3_mz.png	Annex_Maps.R	Found in Outputs\Annex\Maps
Figure A.4	A4_am.png	Annex_Maps.R	Found in Outputs\Annex\Maps
Figure A.5	A5_br.png	Annex_Maps.R	Found in Outputs\Annex\Maps
Figure A.6	A6_density.png	Annex_Maps.R	Found in Outputs\Annex\Maps
-	text_numbers.tex	05_Analysis.do	Numbers given on pages 12-13, Found in Outputs\Main\Tables

Table 1: List of Exhibits

5 Computational Requirements

5.1 Software and Hardware Requirements

All figures, tables, and numbers mentioned in the paper were created using Stata, and the same goes for the tables in the Annex. The maps in the Annex were created exclusively with R. The following is a list of required versions, dependencies, libraries, and packages:

- Stata/MP 18.0 for Windows (64-bit x86-64)
 - alluvial version ‘1.5’
 - asdoc version ‘2.3.9.5’
 - bumpline version ‘1.4’
 - colrspace version ‘1.2.1’
 - estout version ‘3.31’
 - palettes version ‘1.2.7’
 - sankey version ‘1.81’
 - tabout version ‘2.0.8’
 - wbopendata version ‘16.3’
- R version 4.3.2 for Windows (x86-64)
 - dplyr version ‘1.1.4’
 - ggplot2 version ‘3.4.4’
 - ggpubr version ‘0.6.0’
 - readxl version ‘1.4.3’
 - sf version ‘1.0.15’

5.2 Memory and Runtime and Storage Requirements

One of the authors used Microsoft Windows 11, and the results presented in the paper and the Appendix were generated on a machine with the following specifications:

- Processor 11th Gen Intel(R) Core(TM) i7-1185G7 [at] 3.00GHz, 2995 Mhz, 4 Core(s), 8 Logical Processor(s)
- Installed RAM: 32.0 GB

The final size of the reproducibility folder is around 15 GB. Replicators can expect to run all the code in approximately 45 minutes under these conditions.

6 Folder Structure

- Code
- Data
 - Cleaned
 - New
 - Raw
 - Shapefiles
- Outputs
 - Annex
 - Manuscript
 - Maps
 - Tables
 - Main
 - Figures
 - Tables