_READ-ME_State Terror and Long-run Development: The Persistence of the Khmer Rouge

Don Grasse

2023-03-19

Replication Instructions

This is the read me file for replication of the manuscript "State Terror and Long-run Development: The Persistence of the Khmer Rouge.' '

Data in Data Folder

The replication package includes the following datasets and scripts to produce results.

Dataframes for analysis:

- Dataframe of 2008 villages: 'main_0310.rds'
- Dataframe of 1998 villages: 'census_1998_0310.rds'
- Dataframe of spatial outcomes for balance: 'grid1x1_2022.rds'
- Dataframe of villages for placebo analysis: 'NR3df.rds'
- Dataframe of commune polygons: 'gadm36_KHM_3_sf.rds'
- Dataframe of province polygons: 'gadm36_KHM_1_sf.rds'
- Dataframe of Violence Against Women Survey: 'VAW Cambodia 2015 original full STATA9.dta'
- Dataframe for Grid Cell Elevation used for Figure 2: 'grid_cell_df_map.rds'
- Dataframe of Commune Election Results: eledf.rds
- Dataframe of Commune Migration: 'im_df.rds'
- Shapefile of schools: 'school_of_cambodia.shp'
- Survey of EWR exposure: 'baselinesurveymineerw'
- Causalities from landmines and EWR: 'mine-and-erw-causality-2005-2013'
- Shapefile of prisons, mass graves, memorials from DK 'prisons_burials_memorials'
- Shapefile of roads: 'Major_Roads'
- Shapefile of rivers: 'Major_Rivers'
- DK era zones: 'Democratic_Kampuchea_Zones'

Individual level microdata used to study the impacts of the regime on individual health, labor market outcomes, and wealth, are provided by the National Institute of Statistics Cambodia (NIS) and the Demographic Health Survey (DHS). These data cannot be provided for replication. Instead, an interested scholar must

contact these organizations to receive the data for themselves. Further information is provided in the data-access-developmental-legacies-of-dictatorship.rtf file.

Dataframes that cannot be included in the replication files but are required to reproduce the main results include

- Dataframe of education by cohorts: 'educDD.rds'
- Dataframe of children health: 'dhs_kr.rds'
- Dataframe of individual wealth data from DHS 'dhs_df.rds'
- Dataframe of individual income from LFS 'lfs_income.rds'
- Folder of DHS data including survey clusters 'DHS'
- DHS clusters from 2000-2014 'dhs_shape_stack.rds'

Analysis

The file _main.R will load all necessary packages, extra functions, and data to produce all figures and tables that appear in the main text and appendix. A scholar interested in reproducing just one figure or table can do so using the argument source(.) where . refers to the name of the script that the scholar seeks to recreate. Note that tables are edited in .tex for aesthetics, and thus the output does not exactly match what appears in the final .PDFs, where additional labels and notes are added manually to improve transparency and readability.

A researcher wishing to reproduce an individual table or figure can do so by looking inside of the tabs-in or figs-in folder. Each table and figure is named as it appears in the manuscript - for instance, code to produce Table 2 in the main text is in tab2-baseline-in.R and Figure A1 is in figA1-balance-in.R.

PLEASE NOTE: Since the data required to conduct the entire analysis cannot be shared by the author, the script _main.R will not run cleanly, because it will seek to use data that does not exist in the data folder I was able to make available. As such, one will need to obtain the necessary data to run the entire script themselves.

The file packages.R includes all packages necessary for the analysis. Again, if a researcher wishes to produce just an individual table or figure, they will want to consult the packages.R script to load everything that is necessary.

The file grd-helper-functions.R includes functions created by the author to conduct the analysis.

As always, take care to set the working directory! Each individual script assumes you have loaded the packages and set the working directory as set in the _main.R script. So make sure you check that first even if you just want to look at one particular table/figure.

Technical Information

R version 4.2.1 (2022-06-23) \\ Platform: x86_64-apple-darwin17.0 (64-bit) \\ Running under: macOS Monterey 12.3.1

Please note that the robust standard error estimation is computationally expensive because the range of the spatial autocorrelation is being estimated with MLE to inform how much spatial dependence should be tolerated when constructing uncertainty estimates. As such, many of the models take a long time. . . if there seems to be a delay, nothing is wrong with your machine, just trust the process.

Questions?

Questions? Please contact Donald Grasse <donald dot grasse dot 93 at gmail dot com>.