

README: Reproducibility Package for “Population Thresholds and Classification Stability: Evidence from the World Bank’s Small States Classification Rule”

Last updated: June 10, 2026

Corresponding author: Baran Han bhan3@worldbank.org

1. Purpose of this package

This reproducibility package contains the Stata code and input data needed to reproduce the tables and figures in the paper. The analysis uses World Development Indicators (WDI) data, archived WDI population vintages, and an author-compiled country classification file.

Run the package from the master file:
codes/0.master.do

2. Computational requirements

The package is written for **Stata 18 for Windows 11**

The following Stata packages are required:

```
ssc install wbopendata, replace  
ssc install estout, replace
```

An internet connection is needed to download WDI data through wbopendata, unless the WDI datasets created by the code are already included in the package.

3. Folder structure

The package should be organized as follows:

```
reproduction package/  
  README.docx  
  codes/  
    0.master.do  
    1.setup.do  
    2.xs_generate_baseline_regression.do  
    3.graphs.do  
    4.population_data_extraction.do  
    5.simulationcodes.do  
  data/  
    country_categorization_distance.xlsx  
    population/  
      WDI_excel_2019_06_28/WDIEXCEL.xlsx  
      WDI_excel_2020_07_01/WDIEXCEL.xlsx  
      WDI_excel_2021_07_30/WDIEXCEL.xlsx
```

WDI_excel_2022_06_30/WDIEXCEL.xlsx
WDI_excel_2023_06_29/WDIEXCEL.xlsx
WDI_excel_2024_07_01/WDIEXCEL.xlsx
WDI_excel_2025_07_02/WDIEXCEL.xlsx
output/
log/

The output/ and log/ folders store files created by the code.

4. Instructions for Replicators

1. Unzip the reproduction package.
2. Open codes/0.master.do.
3. Set the top-level directory in the first line:
4. Confirm that the required input files are in the data/ folder.
5. Run codes/0.master.do.

global dir "YOUR/PATH/reproduction package – to send"

The master file runs the code in this order:

- 1.setup.do
- 2.xs_generate_baseline_regression.do
- 3.graphs.do
4. population_data_extraction.do
- 5.simulationcodes.do

5. A List of Exhibits

Paper item	Code file	Output
Table 1	2.xs_generate_baseline_regression.do	output/table_xs_baseline.csv
Figure 1	3.graphs.do	output/fig_lpoly_outcomes_lnpop.png
Table 2	5.simulationcodes.do	Printed in Stata and saved in log/log.smcl
Table 3	Not code-generated	Descriptive table in the paper
Table 4	5.simulationcodes.do	Printed in Stata and saved in log/log.smcl, see 5.1. for details
Appendix Figure A1	3.graphs.do	output/figA1_partial_residuals.png

The simulation tables are currently printed to the Stata Results window and captured in the log file. If separate output files are required, export these tables from 5.simulationcodes.do to CSV, Excel, or text format.

5.1. Reproducing Table 4

Table 4 is based on the simulation output produced by codes/5.simulationcodes.do. The underlying results are printed to the Stata Results window and saved in log/log.smcl. The key output is:

```
list iso3 country_name switch_r1-switch_r4, noobs sep(0)
```

where switch_r1–switch_r4 report the number of Small State classification changes between 2019 and 2025 under each implementation rule.

Interpretation of the output:

- 0 indicates no change in classification during the sample period.
- 1 indicates one classification change (for example, Small State → non-Small State).
- 2 indicates two classification changes (Small State → non-Small State → Small State).

For the countries discussed in Table 4:

- **Equatorial Guinea (GNQ)** records one switch under all four rules, corresponding to declassification.
- **Trinidad and Tobago (TTO)** records two switches under Rules 1 and 4, corresponding to declassification followed by reclassification.
- **Kosovo and North Macedonia**, have always had populations above 1.5 million and have never entered the Small State classification; accordingly, they record zero switches under every rule.
- **Eswatini, Mauritius, and Timor-Leste** have always had populations below 1.5 million and have remained Small States throughout the sample period; accordingly, they record zero switches under every rule.

To identify the year of each classification change reported in Table 4, run:

```
list iso3 vintage population r1 r2 r3 r4, sepby(iso3) noobs
```

The log file shows the specific years in which switches occurred.

Table 4 in the manuscript is a narrative summary of these simulation results. It is not generated as a formatted table by the code; rather, it translates the logged simulation output into the descriptive classification outcomes reported in the paper.

6. Data Availability Statement

This package uses publicly available World Bank data and one author-compiled country classification file. All data were accessed or downloaded on June 10, 2026.

Data source 1: World Development Indicators, current WDI data

File name:

data/wdi_panel.dta [created by code]

data/wdi_panel_clean.dta [created by code]

Brief description:

The main macroeconomic and demographic variables are downloaded from the World Bank World Development Indicators for 2000-2024 using the Stata command `wbopendata`.

Indicators used:

SP.POP.TOTL: Population

AG.LND.TOTL.K2: Land area

TX.VAL.FUEL.ZS.UN: Fuel exports

TT.PRI.MRCH.XD.WD: Net barter terms of trade index

NE.TRD.GNFS.ZS: Trade (% of GDP)

NY.GDP.MKTP.PP.KD: GDP, PPP, constant 2021 international dollars

NY.GDP.PCAP.PP.KD: GDP per capita, PPP, constant 2021 international dollars

Citation:

World Bank. World Development Indicators. Washington, DC: World Bank.

URL or access instructions:

<https://databank.worldbank.org/source/world-development-indicators>

Access or download date:

June 10, 2026.

Download steps:

The code downloads these indicators directly using `wbopendata` in `codes/1.setup.do`. No manual download is required if the user has an internet connection and the `wbopendata` Stata package installed.

License or access restrictions:

The WDI data are publicly available from the World Bank and are subject to the World Bank data terms of use. Because current WDI data may be revised over time, exact numerical reproduction of the baseline regressions requires using the same WDI vintage as the submitted package.

Data source 2: World Development Indicators Archives, population vintages

File name:

data/population/WDI_excel_2019_06_28/WDIEXCEL.xlsx

data/population/WDI_excel_2020_07_01/WDIEXCEL.xlsx

data/population/WDI_excel_2021_07_30/WDIEXCEL.xlsx

data/population/WDI_excel_2022_06_30/WDIEXCEL.xlsx

data/population/WDI_excel_2023_06_29/WDIEXCEL.xlsx

data/population/WDI_excel_2024_07_01/WDIEXCEL.xlsx

data/population/WDI_excel_2025_07_02/WDIEXCEL.xlsx

Brief description:

These archived WDI Excel files are used to construct the population-vintage panel for the classification-stability analysis. The code extracts total population, SP.POP.TOTL, from each vintage.

Citation:

World Bank. World Development Indicators Archives. Washington, DC: World Bank.

URL or access instructions:

<https://datatopics.worldbank.org/world-development-indicators/wdi-archives.html>

Access or download date:

June 10, 2026.

Download steps:

Go to the World Bank WDI Archives page. Download the WDI Excel archive for each listed vintage.

Unzip each archive and place the WDIEXCEL.xlsx file in the corresponding folder under data/population/, using the folder names shown above. The script codes/4.

population_data_extraction.do then imports the files and creates:

data/population/vintage_population_panel.dta

data/population/vintage_population_panel.csv

License or access restrictions:

The WDI archive files are publicly available from the World Bank and are subject to the World Bank data terms of use. Exact reproduction of the population-vintage analysis requires the archived WDI Excel files listed above.

Data source 3: Author-compiled country classification and distance file

File name:

data/country_categorization_distance.xlsx

Brief description:

This file contains country classification variables, sample restrictions, lending and income group variables, regional identifiers, and distance-from-markets variables used in the analysis. The distance-from-markets variable is extracted from Table 6 of the UNCTAD report cited below.

Citation:

UNCTAD. 2021. "The Least Developed Countries Report 2021: The Least Developed Countries in the Post-COVID World." Table 6. Geneva: United Nations Conference on Trade and Development.

URL or access instructions:

https://unctad.org/system/files/official-document/ser-rp-2021d10_en.pdf

Access or download date:

June 10, 2026.

Download steps:

The file `data/country_categorization_distance.xlsx` is included in the reproduction package. No separate download is required. The distance-from-markets variable can be checked against Table 6 of the UNCTAD report linked above.

License or access restrictions:

The file is author-compiled and is included for reproducibility. The underlying distance-from-markets information is from a publicly available UNCTAD report. Any third-party source materials remain subject to their original terms of use.

Datasets created by the code

The following files are created by the code and do not need to be supplied separately:

`data/wdi_panel.dta`

`data/wdi_panel_clean.dta`

`data/distance.dta`

`data/small states definition vulnerability work_merged.dta`

`data/small states definition vulnerability work_merged2.dta`

`data/small_states_xs_2000_2024.dta`

`data/population/vintage_population_panel.dta`

`data/population/vintage_population_panel.csv`