

README File on Data and Codes for "Unreliable Electricity in Developing Countries: The Role of Weak Institutions"

Andu Berha

Stuti Khemani

December 9, 2025

Initial Instructions

At the beginning of each do file in Stata, please ensure you set the path to the corresponding directory on your computer.

Data Availability Statement (DAS)

- **WDI Income and Population Data** (`wdi_income_data.csv`). This dataset provides the GDP per capita and population density variables. The data are derived from the World Development Indicators (WDI), which are publicly accessible through the World Bank's DataBank portal at [World Development Indicators](#). Accessed: July 2025.
- **UPBEAT Utility Performance Data** (`UpbeatRawUtilityData/*.csv`). These raw CSV files (for each country) contain the utility-level performance indicators. The UPBEAT database is a public World Bank/ESMAP product available through the official download portal at [utilityperformance.energydata.info](#). The dataset is licensed under [CC BY 4.0](#), which permits reuse with attribution and disclosure of modifications. Accessed: July 2025.
- **Worldwide Governance Indicators** (`wgidataset.dta`). This file provides the Worldwide Governance Indicators used to supplement institutional measures in the empirical analysis. The WGI dataset is publicly available from the World Bank at [Worldwide Governance Indicators](#). Accessed: June 2025.
- **ICRG Governance Data** (`ICRG_data.xlsx`). This proprietary dataset contains the International Country Risk Guide (ICRG) governance components, including corruption. Because the data are restricted by the PRS Group, the public replication package does *not* redistribute the raw file. World Bank Group users may access ICRG through the WBG Library by searching

“political risk” and navigating to the ICRG *Country Data* interface.¹ External users must obtain licensed access directly from the PRS Group. Accessed: June 2025.

- **World Bank Region Classification** (`WB_list_of_economies.xlsx`). This file provides standardized geographic region identifiers. It is publicly available from the World Bank’s classification resources at [World Bank](#). Accessed: June 2025.
- **Income Group Classification** (`income_group_classification.xlsx`). This dataset contains the World Bank income group categories over time. It is publicly accessible from the World Bank’s Country and Lending Groups repository at [World Bank](#). Accessed: June 2025.

Rights and permissions: The authors accessed all datasets under the applicable terms of use. Redistribution of proprietary ICRG data is not authorized under the public version of this package.

Software

- Stata v17 or later.

Required Packages

The replication scripts rely on the following user-written packages:

- `reghdfe` and `ftools`
- `estout`
- `winsor2`
- `palettes` and `colrspace` (for consistent palette utilities)

Directory Structure

```
ReplicationPackage/  
  Code/  
    cleaning/  
      country_cleaning.do  
      utility_cleaning.do
```

¹The variables included in the input file are Corruption, Law & Order, Bureaucracy Quality, Civil Disorder, Democratic Accountability, Government Cohesion, Government Stability, and War.

```

    name_cleaning.do
    master_merge.do
analysis/
    do_tables.do
    do_figures.do

Data/
    UpbeatRawUtilityData/          (raw UPBEAT CSV files)
    WB_list_of_economies.xlsx
    income_group_classification.xlsx
    wdi_income_data.csv
    wgidataset.dta
    ICRG_data.xlsx                 (proprietary; see DAS)

    country_data.dta               (generated)
    utility_data.dta               (generated)
    final_analysis_data.dta        (generated)

Output/
    tables/
    figures/

```

The `country_data.dta`, `utility_data.dta`, and `final_analysis_data.dta` files are produced by the scripts listed below. They may be included in the package to facilitate reproducibility verification, but the code is written to regenerate/overwrite them from the documented inputs.

Core Intermediate and Analysis Files

- `country_data.dta`: Harmonized country-year panel built from WDI, WGI, ICRG, and World Bank classification files.
- `utility_data.dta`: Harmonized utility-year panel constructed from UPBEAT raw CSV inputs.
- `final_analysis_data.dta`: Unified utility-country-year dataset used to generate all regression tables and figures.

Shared Dependency: name_cleaning.do

The cleaning scripts call name_cleaning.do to standardize country names before reshaping and merging. This ensures stable joins across WDI, ICRG, WGI, World Bank classification files, and UPBEAT-derived country identifiers.

Script Descriptions

Code/cleaning/country_cleaning.do

Purpose: Creates a harmonized country-year dataset by integrating:

Inputs (expected in Data/): WB_list_of_economies.xlsx, income_group_classification.xlsx, wdi_income_data.csv, ICRG_data.xlsx, wgidataset.dta.

Output: Data/country_data.dta.

Code/cleaning/utility_cleaning.do

Purpose: Constructs a utility-year panel from UPBEAT raw CSV files by stacking, reshaping, and mapping indicator labels into standardized variable names.

Inputs: Data/UpbeatRawUtilityData/*.csv and Code/cleaning/name_cleaning.do.

Output: Data/utility_data.dta.

Code/cleaning/master_merge.do

Purpose: Merges the cleaned country-year and utility-year datasets and applies final harmonization rules to create the unified analysis file.

Inputs: Data/country_data.dta, Data/utility_data.dta.

Output: Data/final_analysis_data.dta.

Code/analysis/do_tables.do

Purpose: Reproduces all main-text and appendix tables using final_analysis_data.dta.

Outputs: Table files saved to Output/tables/.

Code/analysis/do_figures.do

Purpose: Reproduces all figures using `final_analysis_data.dta`.

Outputs: Figure files saved to `Output/figures/`.

Recommended Run Order

The user can regenerate the cleaned datasets and reproduce all results with minimal edits. In most cases, the only required change is updating the working directory at the top of each `.do` file. From the package root, execute the workflow in the following sequence:

```
* 1. Country-level cleaning
do "Code/cleaning/country_cleaning.do"

* 2. Utility-level cleaning
do "Code/cleaning/utility_cleaning.do"

* 3. Build unified analysis data
do "Code/cleaning/master_merge.do"

* 4. Reproduce tables and figures
do "Code/analysis/do_tables.do"
do "Code/analysis/do_figures.do"
```

One-Button Execution

For convenience, the full pipeline can also be executed using the one-button driver script `main_script.do`. After updating the working directory in each code file, run:

```
do "main_script.do"
```

Reproducibility Notes

Consistent with the Data Availability Statement, some inputs and generated data files cannot be redistributed publicly, but they may be included in the package for reproducibility verification only. In particular:

Data/

ICRG_data.xlsx (proprietary; see DAS)

country_data.dta (generated intermediate)

final_analysis_data.dta (generated analysis file)