README

Replication package for:

"Financial Inclusion and Electricity Uptake: Global Evidence and District-Level Mechanisms"

Megan Lang - mlang@worldbank.org
Alpha Ly - alpha.ly@ensai.fr

World Bank Policy Research Working Paper Series
Last updated: May 2026

1. Overview

This README describes the steps required to replicate all tables and figures in "Financial Inclusion and Electricity Uptake: Global Evidence and District-Level Mechanisms" by Megan Lang and Alpha Ly.

The replication package is organised into two analytical parts:

- **Global level:** cross-country panel analysis.
- **District level:** DHS-based district-level analysis for Sub-Saharan Africa.

To replicate all results, run the master file

`Replication.do`

which sequentially calls several data process sub-programs and the main estimation scripts.

2. Data Availability

The table below summarizes data access status for each source.

Dataset	Publicly available?	Notes
GSMA Mobile Money Metrics	Yes (open data)	Registration is required
World Development Indicators	Yes (open data)	
U.S. EIA International Energy Statistics	Yes (public domain)	
World Inequality Database (WID)	Yes (open access)	Free for research use
DHS Surveys (HR, IR, MR, PR recodes)	Restricted	Registration required at dhsprogram.com
Collins Mobile Coverage Explorer	Restricted	Commercial licence required
SEDAC Gridded Population of the World (GPW)	Restricted	NASA EarthData registration needed

2.1 Data Sources

Below are the full details for each dataset used. Most are already included in the replication package, but a replicator will need to download and save three of them as explained below (DHS, mobile coverage, and population density).

1. GSMA Mobile Money Metrics

Source	GSMA
URL	Go to https://www.gsma.com/mobile-money-metrics/ and download the three datasets: the Mobile Money Deployment Tracker, the Mobile Money Regulatory Index, and the Mobile Money Prevalence Index (MMPI). Then manually assemble them into a single Excel workbook with one sheet per dataset, following the structure of the attached Mobi_money.xlsx file (sheets named Deployment, Regulation, and MMPI).
Citation	GSMA. Mobile Money Metrics and Adoption Data.
License	Open data
Access year	2023
Filenames	Mobi_money.xlsx
Directory	...\Replication\Global level\Bases_Used and ...\Replication\District level\4_Country_Bases_Used

2. World Development Indicators

Source	World Bank
URL	https://databank.worldbank.org/source/world-development-indicators
Citation	World Bank. World Development Indicators. Washington, DC.
License	Open data
Access year	2023
Filenames	P_Data_Extract_From_WDI_2.xlsx | P_Data_Extract_From_WDI_2_female.xlsx
Directory	...\Replication\Global level\Bases_Used and ...\Replication\District level\4_Country_Bases_Used

3. U.S. Energy Information Administration

Source	U.S. Energy Information Administration (EIA)
URL	https://www.eia.gov/international/data/world Note: To download the raw data, select “Electricity”, then choose “Electricity capacity”, sort by Country/Region, then select “Data Options” to add “Electricity generation”, keep the period 1990-2021, select “Download Options”, and then “Export as CSV (Table)”.
Citation	U.S. EIA. International Energy Statistics.
License	Public domain (U.S. government)
Access year	2023

Filename	us_eia_final.xlsx
Directory	...\Replication\Global level\Bases_Used and ...\Replication\District level\4_Country_Bases_Used

4. World Inequality Database (WID)

Source	World Inequality Lab
URL	https://wid.world/ Note: Click on "Data tables", then under "Indicators" choose "Wealth inequality, Top 10% share", add "All countries", select the period 1990–2021, retrieve the data, choose the structure "country-indicator-percentile-year-1234", and click "Download".
Citation	World Inequality Lab. World Inequality Database.
License	Open access (free for research)
Access year	2023
Filename	wealth_top10.dta
Directory	...\Replication\Global level\Bases_Used

5. DHS Surveys

Source	Demographic and Health Surveys (DHS) Program
URL	https://dhsprogram.com/data/dataset_admin/index.cfm Note: Select "Register", then "Create New Project", choosing the SSA region and the period 2003-2021. Choose "Download Manager - Download Multiple Surveys", then select "HR - Household Recode", "PR - Household Member (Person) Recode", "IR - Individual Recode (Women)", and "MR - Men's Recode". Check all countries, select the Stata system file, build the URL file list, and Copy and Paste the list into your download manager (ours is Chrono) to download the datasets.
Citation	Demographic and Health Surveys.
License	Restricted (registration required; free for research)
Access year	2023
Download and save as	Download HR, IR, MR, and PR recode files from the URL above. Do not rename the files, simply save them in the folder indicated below.
Save in the directory	...\Replication\District level\1_DHS_Raw

6. Collins Mobile Coverage Explorer

Source	Collins Bartholomew
URL	https://www.collinsbartholomew.com/mobile-coverage-explorer/
Citation	Collins Bartholomew. Mobile Coverage Explorer Dataset.
License	Restricted (commercial licence required)
Access year	2023

Download and save as	Download the GSM (1998–2020), 3G (2006–2020), and 4G (2013–2020) datasets from the URL above. Extract the .dbf databases using QGIS and save them following this naming convention: GID55_Global_3G_YYYY, ADM1_Global_3G_YYYY, ADM2_Global_3G_YYYY, where YYYY is the corresponding year. Apply the same procedure to GSM and 4G.
Save in the directory	...\Replication\District level\5_Network_data

7. SEDAC – Gridded Population of the World (GPW)

Source	NASA Socioeconomic Data and Applications Center (SEDAC) / CIESIN
URL	https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-gpwv4-popdens-r11-4.11
Citation	CIESIN (Columbia University). Gridded Population of the World (GPW), v4.
License	Restricted (NASA EarthData registration needed)
Access year	2023
Download and save as	Download the population density raster files from the URL above (2000–2020). Extract them at the GID55 level (55x55 grid) using QGIS, then save each file following this naming convention: CEDAC_POPYYYY, where YYYY is the corresponding year. Save a copy of the network data processed above at the GID55 level (55x55 grid) only, using the following naming convention: Global_GSM_YYYY, Global_3G_YYYY, Global_4G_YYYY, where YYYY is the corresponding year.
Save in the directory	...\Replication\District level\5_Network_POP_data

8. World Bank Country and Lending Groups

Source	World Bank Group
URL	https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups Note: You can download the classification here: https://ddh-openapi.worldbank.org/resources/DR0095333/download
Citation	World Bank. World Bank Country and Lending Groups Classification.
License	World Bank Open Data
Access year	2023
Filename	Region_IncomeGp.dta
Directory	...\Replication\Global level\Bases_Used

9. Country classifications (Monde.xls)

Source	Compiled by the authors
Description	Monde.xls consolidates four country-level dummy classifications: **Afrique** – African countries; see African Union Member States, https://au.int/en/member_states/countryprofiles2. **OCDE** – OECD member states, reflecting membership as of 2016–early 2018; see OECD "Members and Partners", https://www.oecd.org/en/about/members-partners.html. **PED** (Pays en développement) – developing economies, operationally defined as the complement of OECD membership. **Heavens** – tax havens from the OECD list of 2 April 2009. See Table 1 in Jackson, J.K. (2010), The OECD Initiative on Tax Havens, CRS Report R40114, https://www.everycrsreport.com/reports/R40114.html.
Citation	See URL field for the four classifications.
License	Open data (compiled from public sources)
Access year	2023
Filename	Monde.xls
Directory	...\Replication\Global level\Bases_Used

10. SSA_DHSClusters_Grids_ADM_SHPiso and Figure 5 (derived shapefiles)

Description	A merged shapefile of DHS Sub-Saharan Africa survey clusters (2003–2021), joined with PRIO-GRID cells and GADM administrative boundaries. Used to produce Figure 5 and to merge spatial identifiers into the district-level analysis dataset.
Inputs	DHS GPS cluster files for all Sub-Saharan African surveys, 2003–2021. GADM world boundaries filtered to Africa. PRIO-GRID cell shapefile (priogrid_cellshp).
Software	QGIS
Procedure	**Step 1.** Go to https://dhsprogram.com/data/dataset_admin/index.cfm and select "Register", then "Create New Project", choosing the SSA region and the period 2003–2021. Choose "Download Manager – Download Multiple Surveys", then select "GPS – Geographic Data". Check all countries, select the Flat File format, build the URL file list, and copy and paste the list into a download manager (e.g. Chrono) to download the datasets. Merge all surveys into a single shapefile (SSA_DHSClusters_merged). For Figure 5, also export the cluster coordinates as clusters_csv. **Step 2.** Go to https://gadm.org/download_world.html and download the file as six separate layers (one per level of subdivision), structured as a geopackage database. The package covers all countries at ADM0 (country), ADM1 (region), ADM2 (district), and lower-level subdivisions. Filter to obtain Africa_shp. For Figure 5, also produce pays_afrique, AfricaADM1, and AfricaADM2. **Step 3.** Go to https://grid.prio.org/#/extensions and download priogrid_cellshp. Clip priogrid_cellshp using Africa_shp to obtain priogrid_cellshp_Africa. **Step 4.** Take SSA_DHSClusters_merged and join priogrid_cellshp_Africa attributes by location (first-match, one-to-one), then save as SSA_DHSClusters_Grids.

	Step 5. Take SSA_DHSClusters_Grids and join Africa_shp attributes by location (first-match, one-to-one), then save as SSA_DHSClusters_Grids_ADM. **Step 6.** Take SSA_DHSClusters_Grids_ADM and join Africa_shp attributes by location (first-match, one-to-one), then save as SSA_DHSClusters_Grids_ADM_SHPiso.
Output file	SSA_DHSClusters_Grids_ADM_SHPiso.dbf (and accompanying .shp / .shx / .prj) and Shapefiles for Figure 5.
Directory	...\Replication\District level\4_Country_Bases_Used and ...\Replication\District level\7_MoMo_Repository\Figures_Tables

11. World shapefile (WB)

Source	World Bank Official Boundaries (ADM0, 10m generalization)
URL	Download from https://doi.org/10.1371/journal.pone.0315166.s003 (Supporting Information of the PLOS ONE article).
Citation	World Bank. World Bank Official Boundaries. Washington, DC. Redistributed in: Chen S., Piao X., Xie J., Managi S. (2023).
License	World Bank Open Data; PLOS ONE Supporting Information distributed.
Access year	2023
Filename	wb_countries_admin0_10m
Directory	...\Replication\Global level

2.2 Statement about Rights

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in the manuscript.

3. Instructions for Replicators

Follow the steps below to replicate all results from start to finish.

1. **Install required software.** Stata 18 or 19 is required. On first run, the master script automatically installs four community packages: `ebalance`, `ivreg2`, `ranktest`, and `xtivreg2`.
2. **Set the working directory.** Open `Replication.do` and verify that the global `working_dir` resolves correctly to the root of the replication folder. The script uses `c(pwd)`, so running it from the root directory is sufficient.
3. **Run the master do-file.** Execute **`Replication.do`**. This calls all sub-scripts in order and writes a timestamped log file to `LogFile\`.

Note on Stata versions: Results were estimated in both Stata 18 and Stata 19. Minor numerical differences may appear depending on the version used; see Stata's version FAQ for details.

4. List of Exhibits

The provided code reproduces all tables and figures in the paper and appendices. Exhibits marked “No code required” are conceptual diagrams or manually compiled tables.

Main Results – Figures

Exhibit	Output filename	Script	Note
Figure 1	—	—	No code required
Figure 2	MM_diffusion.pdf	Estimates_global.do	Global level
Figure 3	stat1.pdf	Estimates_global.do	Global level
Figure 4	estimates.pdf	Estimates_global.do	Global level
Figure 5	Layout 1.pdf	Project v3.qgz	The QGIS Project is in the following directory: “...\Replication\District level\7_MoMo_Repository\Figures_Tables\Figure 5”. All shapefiles are included in the folder. The electrification data used are from the DHS data referenced in the paper.
Figure 6	BoxPlot.pdf	9_Estimates.do	District level

Main Results – Tables

Exhibit	Output filename	Script	Note
Table 1	Table1_.tex	Estimates_global.do	Global level
Table 2	Table3.tex	Estimates_global.do	Global level
Table 3	District_baseline.tex	9_Estimates.do	District level

Appendix A – Figures

Exhibit	Output filename	Script	Note
Figure A1	robustness.pdf	Estimates_global.do	
Figure A2	estimates_.pdf	Estimates_global.do	
Figure A3	estimates_timetrend.pdf	Estimates_global.do	
Figure A4	estimates_idt.pdf	Estimates_global.do	
Figure A5	High_elec.pdf	Estimates_global.do	
Figure A6	honestdid.pdf	Estimates_global.do	
Figure A7	Mechanisms.pdf	Estimates_global.do	
Figure A8	heterogen_opera.pdf	Estimates_global.do	

Figure A9	supply0.pdf	Estimates_global.do	
Figure A10	supply12.pdf + supply34.pdf	Estimates_global.do	Two files
Figure A11	fiscal_monetary.pdf	Estimates_global.do	
Figure A12	FDI_GDP.pdf	Estimates_global.do	

Appendix A – Tables

Exhibit	Output filename	Script	Note
Table A1	Table22.tex	Estimates_global.do	
Table A2	Table23.tex	Estimates_global.do	
Table A3–A5	—	—	No code required
Table A6	Table3_.tex	Estimates_global.do	
Table A7	Table3_timetrend.tex	Estimates_global.do	
Table A8	Table3_idt.tex	Estimates_global.do	
Table A9	Table5.tex	Estimates_global.do	
Table A10	Table7.tex	Estimates_global.do	
Table A11	Table9.tex	Estimates_global.do	
Table A12	Table11.tex	Estimates_global.do	
Table A13	Table13.tex	Estimates_global.do	
Table A14	Table15.tex	Estimates_global.do	
Table A15	Table17.tex	Estimates_global.do	
Table A16	Table19.tex	Estimates_global.do	
Table A17	Table21.tex	Estimates_global.do	
Table A18	Table_consis2.tex	Estimates_global.do	
Table A19	Table2.tex	Estimates_global.do	
Table A20	Table2_.tex	Estimates_global.do	
Table A21	Table2_timetrend.tex	Estimates_global.do	
Table A22	Table2_idt.tex	Estimates_global.do	
Table A23	Table4.tex	Estimates_global.do	
Table A24	Table6.tex	Estimates_global.do	
Table A25	Table8.tex	Estimates_global.do	
Table A26	Table10.tex	Estimates_global.do	

Table A27	Table12.tex	Estimates_global.do	
Table A28	Table14.tex	Estimates_global.do	
Table A29	Table16.tex	Estimates_global.do	
Table A30	Table18.tex	Estimates_global.do	
Table A31	Table20.tex	Estimates_global.do	
Table A32	Table_consis.tex	Estimates_global.do	

Appendix B – Tables

Exhibit	Output filename	Script	Note
Table B1	StatMomo.tex + StatUptake.tex + StatAll.tex	9_Estimates.do	Three files
Table B2–B4	—	—	No code required
Table B5	ttest.tex	9_Estimates.do	
Table B6	balance.html	9_Estimates.do	
Table B7	District_cluster.tex	9_Estimates.do	
Table B8	District_IVOLS.tex	9_Estimates.do	
Table B9	District_teffect.tex	9_Estimates.do	
Table B10	District_IVFE.tex	9_Estimates.do	

5. Requirements

5.1 Software

Software	Details
Stata	Version 18 or 19 (results generated in Stata 18 and 19)
ebalance	Entropy balancing — installed automatically by Replication.do
ivreg2	IV regression — installed automatically
ranktest	Rank tests for IV identification — installed automatically
xtivreg2	Panel IV regression — installed automatically

5.2 Computational requirements

Specification	Details
Processor	Intel Core i7, 2.6 GHz (used for benchmarking)
RAM	8 GB
Operating system	Windows 10
Approximate runtime	~3 hours on the above configuration

Actual runtime will vary with machine specifications.

6. Code Description

Master script

`Replication.do` — installs packages, sets the working directory, and calls all sub-scripts in sequence. Run this file to replicate everything.

Global level

Script	Description
Data_global.do	Cleans the raw data and builds the final global dataset
Estimates_global.do	Estimates the results for the global-level analysis

District level

Script	Description
1_DHS_Clean.do	Cleans the raw DHS dataset
2_DHS_Append - HR/IR/MR/PR.do	Appends cleaned DHS files across survey rounds by recode type (household, individual, male, PR)
3_!PHmain.do	Calculates population & pousing indicators (partly adapted from Shireen Assaf and Courtney Allen)
4_!RCmain.do	Calculates respondents characteristics indicators (partly adapted from Shireen Assaf and Courtney Allen)
5_!WEmain.do	Calculates women's empowerment indicators (partly adapted from Shireen Assaf and Courtney Allen)
6_SSA_Merge.do	Merges DHS datasets
7_Country_data.do	Processes country-level raw datasets
7_Mobile_Network.do	Processes mobile network coverage raw data
7_Mobile_Network_POP.do	Computes mobile network coverage rates using population density

8_0_HR_GPS_Grid_Merge.do	Spatially merges DHS household recode to grid cells using GPS coordinates
8_1_MR_IR_Append.do	Appends MR and IR recode files
8_2_MergePR_toMR_IR.do	Merges PR recode into the MR/IR dataset
8_3_MergeHR_toMR_IR_PR.do	Merges the household recode into the combined MR/IR/PR dataset
8_4_DHS_CountryLevelData.do	Merges country-level covariates into the final DHS dataset
9_Estimates.do	Estimates the results for the district-level analysis

7. Folder Structure

The replication package is organised as follows:

```
Replication/
├── Replication.do                  ← Master script (run this)
├── LogFile/                         ← Timestamped log files
├── Global level/
│   ├── Bases_Used/                    Raw data files
│   ├── Bases_Created/                 Final global dataset
│   ├── MoMo_Repository/               Figures and estimation tables
│   ├── wb_countries_admin0_10m/       World shapefile
│   ├── Data_global.do
│   └── Estimates_global.do
└── District level/
    ├── 1_DHS_Raw/                     DHS raw + cleaned data
    ├── 2_DHS_Append/                  Appended DHS files
    ├── 3_DHS_indicators/              Intermediate and final datasets
    ├── 4_Country_Bases_Used/          Country-level raw dataset
    ├── 4_Country_Bases_Created/       Country-level final dataset
    ├── 5_Network_data/                Mobile network raw data (Collins)
    ├── 5_Network_final/               Processed mobile network dataset
    ├── 5_Network_POP_data/            Population density raw data (SEDAC)
    ├── 5_Network_POP_final/           Mobile network coverage rates
    ├── 6_SSA_DHSClusters_Grids_ADM_SHPiso/  DHS clusters + spatial identifiers
    └── 7_MoMo_Repository/             Figures and estimation tables
```