

Replication Package README

Overview

This replication package contains the data and Stata code used to reproduce the tables, figures, descriptive statistics, and regression results reported in the associated World Bank paper on the implementation and completion of foreign direct investment (FDI) projects in low-income countries.

The package is designed as a one-button replication workflow. The replicator should open `main.do`, change the top-level project directory to the location of the replication package on their computer, and run the file. `main.do` creates the required output folders, checks for the required user-written Stata command, and executes the analytical scripts in the required order.

The project-level analytical source dataset is `Data/master_data.dta`. The econometric dataset `Data/econometrics_data.dta` is generated by the replication code and is not a source input.

Publication: From Announcement to Completion: Tracking Big FDI Projects in LICs

Data Availability

Some data cannot be made publicly available.

The analysis combines a World Bank-created project-level dataset, publicly available macroeconomic and conflict data, public IMF balance-of-payments data, and licensed fDi Markets data. The source and access conditions for each input dataset are described below.

Statement about Rights

The authors had legitimate access to all data used in the study. Publicly available data are cited below with their relevant access information. The World Bank-created project-level dataset is being deposited in the World Bank Development Data Hub (DDH). fDi Markets data were accessed through the World Bank Group's institutional subscription and are subject to the provider's licensing and redistribution restrictions. The underlying project-level fDi Markets database is not publicly redistributed as part of this package. Aggregated statistics derived from fDi Markets are used for Figure 1, subject to the applicable subscription terms.

Third-party reproducibility: The full replication package cannot be run by an external third party because the project-level source dataset contains data derived from licensed fDi Markets records and is classified as Confidential. The World Bank Reproducibility Team may run the package using the archived confidential source dataset under the applicable internal access conditions. No code modifications are required beyond changing the top-level project directory in `main.do`. Dataset must be deleted after the process.

Data Sources

1. Project-level FDI project status dataset

- **File:** `Data/master_data.dta`
- **Dataset title:** *FDI Project Status and Outcomes in Low-Income Countries, 2011–2020*
- **Provider/creator:** World Bank
- **Description:** Project-level dataset containing FDI project characteristics, implementation status and outcomes, and supporting evidence used to assess project implementation.
- **Access:** World Bank Development Data Hub (DDH).
- **DDH record:** [TO ADD: DDH URL / DOI or catalogue identifier]
- **Access date:** [TO ADD]
- **Redistribution:** Subject to the access and licensing terms stated in the DDH record.
- **Notes:** This archived dataset is the analytical starting point for the replication package.

2. GDP growth data

- **File:** `Data/GDP_Growth_Master.xlsx`
- **Underlying sources:** International Monetary Fund (IMF) World Economic Outlook (WEO), April 2026, and World Bank Macro Poverty Outlook (MPO), April 2026.
- **Description:** Country-year real GDP growth data used to construct the growth-shift variable in the econometric analysis. IMF World Economic Outlook data are used for all countries except Syria and Yemen, for which World Bank Macro Poverty Outlook data are used where the corresponding IMF WEO observations are unavailable.
- **IMF WEO source:** April 2026 World Economic Outlook Database
<https://data.imf.org/Datasets/WEO>
- **World Bank MPO source:** Macro Poverty Outlook, April 2026
<https://www.worldbank.org/en/publication/macro-poverty-outlook>
- **Citation:** International Monetary Fund. 2026. *World Economic Outlook Database, April 2026*. Washington, DC: International Monetary Fund.
- **Citation:** World Bank. 2026. *Macro Poverty Outlook, April 2026: Country-by-Country Analysis and Projections for the Developing World*. Washington, DC: World Bank.
- **Redistribution:** Both underlying sources are publicly available, subject to the respective IMF and World Bank terms of use.
- **Notes:** `prepare_econometrics.do` uses this file to construct the project-level growth-shift measure.

3. UCDP/PRIO conflict data

- **File:** `Data/UcdpPrioConflict_v25_1.dta`
- **Provider:** Uppsala Conflict Data Program (UCDP) / Peace Research Institute Oslo (PRIO)
- **Dataset/version:** UCDP/PRIO Armed Conflict Dataset, version 25.1
- **Description:** Country-year conflict data used to identify war-intensity conflict occurring in the project announcement year or the following two years.
- **Source URL:** <https://ucdp.uu.se/downloads/olddw.html>
- **Access date:** June 2026
- **Citation:** Davies, Shawn, Therése Pettersson, Margareta Sollenberg, and Magnus Öberg (2025), "Organized violence 1989–2024, and the challenges of identifying civilian victims," *Journal of Peace Research* 62(4).
- **Redistribution:** Publicly available under the UCDP terms of use and applicable citation requirements.

4. fDi Markets aggregate series used in Figure 1

- **File:** `Data/figure1_fdim_totals.dta`
- **Underlying source:** fDi Markets, Financial Times
- **Description:** Author-prepared annual aggregate series of announced greenfield FDI used in Figure 1. The file contains aggregate annual values rather than the underlying project-level fDi Markets records.
- **Access:** Underlying fDi Markets data were accessed through the World Bank Group's institutional subscription.
- **Access date:** August 2025
- **Redistribution:** The underlying project-level fDi Markets database is proprietary and is not publicly redistributed. Relevant extracts may be shared internally with the World Bank Reproducibility Team for verification subject to institutional subscription conditions and deletion after verification. Publication of derived fDi Markets data is limited to aggregated information and remains subject to applicable provider terms.
- **Notes:** The package uses this prepared annual aggregate support file solely to reproduce Figure 1.

5. Balance-of-payments FDI series used in Figure 1

- **File:** `Data/figure1_imf_bop.dta`
- **Underlying sources:** IMF Balance of Payments (BOP), IMF International Financial Statistics (IFS), IMF World Economic Outlook (WEO), World Bank World Development Indicators (WDI), national authorities, UNCTAD World Investment Report,

Bank of Mauritius, Magyar Nemzeti Bank (MNB) Statistics, and DECPM annual estimates based on year-to-date data. IMF BOP, IMF IFS, IMF WEO, World Bank WDI, and selected national-authority series were accessed through Haver Analytics.

- **Description:** Author-prepared annual aggregate FDI series used in Figure 1. The series was compiled from multiple balance-of-payments and related statistical sources, with DECPM annual estimates used in some cases based on available year-to-date data.
- **Source information:**
 - UNCTAD World Investment Report: <https://unctad.org/topic/investment/world-investment-report>
 - Bank of Mauritius, Monthly Statistical Bulletin: <https://www.bom.mu/publications-statistics/monthly-statistical-bulletin>
 - Magyar Nemzeti Bank (MNB), Foreign Direct Investment Flows: <https://statisztika.mnb.hu/statistical-topics/balance-of-payments-and-related-statistics/foreign-direct-investments/fdi-flows>
 - Haver Analytics was used to retrieve IMF BOP, IMF IFS, IMF WEO, World Bank WDI, and selected national-authority series.
- **Access date:** August 2025
- **Redistribution:** The underlying data include both publicly available sources and series accessed through the World Bank Group's institutional access to Haver Analytics. Publicly available data are subject to the terms of their respective providers; underlying Haver-sourced data are subject to applicable subscription and redistribution terms. The replication package contains only the prepared annual aggregate series used to reproduce Figure 1.
- **Notes:** The package uses this prepared annual aggregate support file solely to reproduce Figure 1.

Generated Data

`Data/econometrics_data.dta` is generated automatically by `Code/prepare_econometrics.do`. It is an intermediate analytical dataset and does not need to be obtained separately.

Instructions for Replicators

These instructions apply to authorized users with access to the confidential source dataset.

1. Download or copy the full replication package while preserving the folder structure.
2. Open `main.do` in Stata.
3. Change only the following line so that it points to the local copy of the replication package:

```
global project "C:\path\to\Replication Package"
```

4. Run `main.do`.

The main script will:

- define all project subdirectories;
- create the output subfolders if they do not already exist;
- create a local `Code/ado` directory;
- install the `estout` package from SSC if it is not already available;
- reproduce the main descriptive tables;
- reproduce the annex descriptive tables;
- reproduce the main figures;
- construct `Data/econometrics_data.dta`; and
- reproduce the main regression table and Annex 5 econometric tables.

No manual changes to the individual analysis do-files should be required.

List of Exhibits

Paper exhibit	Replication script	Output
Table 1	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_1.csv</code>
Table 2	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_2.csv</code>
Table 3	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_3.csv</code>
Table 4	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_4.csv</code>
Table 5	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_5.csv</code>
Table 6	<code>Code/main_tables.do</code>	<code>Output/Main_Tables/Table_6.csv</code>
Table 7: Determinants of FDI Project Completion	<code>Code/prepare_econometrics.do;</code> <code>Code/regression_tables.do</code>	<code>Output/Regression_Tables/Table7_Probit_AMEs.rtf</code>
Figure 1	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_1.png</code>
Figure 2	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_2.png</code>
Figure 3	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_3.png</code>
Figure 4	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_4.png</code>
Figure 5	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_5.png</code>
Figure 6	<code>Code/main_figures.do</code>	<code>Output/Main_Figures/Figure_6.png</code>
Annex 4 Table 1	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_1.csv</code>
Annex 4 Table 2	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_2.csv</code>
Annex 4 Table 3	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_3.csv</code>
Annex 4 Table 4	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_4.csv</code>
Annex 4 Table 5	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_5.csv</code>
Annex 4 Table 6	<code>Code/annex_tables.do</code>	<code>Output/Annex_Tables/Annex_4_Table_6.csv</code>
Annex 5 Table 1: Descriptive Statistics	<code>Code/prepare_econometrics.do;</code> <code>Code/regression_tables.do</code>	<code>Output/Annex_Tables/Annex5_Table1_Descriptive_Statistics.rtf</code>
Annex 5 Table 2: Correlation Matrix	<code>Code/prepare_econometrics.do;</code> <code>Code/regression_tables.do</code>	<code>Output/Annex_Tables/Annex5_Table2_Correlation_Matrix.rtf</code>

Rounding note: In Annex 4 Table 4, the median completion time for Renewable Energy is reported as 58.4 months in the paper and 58.3 months in the Stata replication output. The difference is due solely to floating-point precision: Stata stores

the underlying median as 58.349999999999994, which rounds to 58.3 at one decimal place, while the corresponding Excel calculation rounds to 58.4.

Requirements

- **Statistical software:** Stata 18
- **User-written Stata dependency:** `estout` (installed automatically from SSC by `main.do` if unavailable)
- **Operating system:** Developed and tested on Windows
- **Internet connection:** Required on the first run only if `estout` is not already installed
- **Estimated runtime:** A clean full run, including recreation of generated data and installation of required user-written commands, completed in approximately 40 seconds on the development machine.
- **Memory/storage requirements:** No unusual hardware requirements are expected.

Code Description

`main.do`

Master replication script. Defines the project directories, creates output directories, configures the local ado directory, checks/install dependencies, and runs all replication scripts in order.

`Code/main_tables.do`

Produces Main Tables 1–6 from `Data/master_data.dta`.

`Code/annex_tables.do`

Produces Annex 4 Tables 1–6 from `Data/master_data.dta`.

`Code/main_figures.do`

Produces Main Figures 1–6. Figure 1 additionally uses `figure1_fdim_totals.dta` and `figure1_imf_bop.dta`.

`Code/prepare_econometrics.do`

Constructs the econometric analysis dataset from `master_data.dta`, GDP growth data, and UCDP/PRIO conflict data. The resulting `Data/econometrics_data.dta` is generated automatically.

`Code/regression_tables.do`

Produces Table 7 and Annex 5 Tables 1–2 from the generated econometric dataset.

Folder Structure

```
Replication Package/
├── main.do
├── README.md
├── [final manuscript]
├── Data/
│   ├── master_data.dta
│   ├── GDP_Growth_Master.xlsx
│   ├── UcdpPrioConflict_v25_1.dta
│   ├── figure1_fdim_totals.dta
│   ├── figure1_imf_bop.dta
│   └── econometrics_data.dta      [generated by code]
├── Code/
│   ├── main_tables.do
│   └── annex_tables.do
```

```
| | | main_figures.do
| | | prepare_econometrics.do
| | | regression_tables.do
| | | ado/ [created automatically]
| | |
| | | Output/
| | | | Main_Tables/
| | | | Annex_Tables/
| | | | Main_Figures/
| | | | Regression_Tables/
| | |
| | | ai_documentation/
| | | | ai_log.md
```

AI Use Statement

Artificial intelligence tools were used in two limited ways in this project.

First, AI was used to assist with the creation of the project-level dataset, including reviewing publicly available information and supporting the classification of FDI project implementation status and outcomes. Because these AI-assisted outputs contributed directly to the analytical dataset used in the paper, detailed documentation of this workflow—including tool and model information, prompts, source materials, raw outputs, classification procedures, and human validation—is provided in:

[ai_documentation/ai_log.md](#)

Second, AI was used as a coding assistant to help debug selected lines of Stata code used in the replication package. The final Stata scripts included in this package are the auditable analytical outputs.