

README for the Reproducibility Package for Hooked on Subsidies: The Case for Reform

Overview

This repository contains the complete reproducibility package for the flagship report "Hooked on Subsidies: The Case for Reform."

The package includes:

- All files (code/Excel) required to reproduce each exhibit in the report
- Fully documented datasets
- A chapter-by-chapter reproducibility structure
- Replicators' outputs for verification
- Instructions for running R, Stata, Python, Excel workflows.

Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Instructions for Replicators

To reproduce all results in this report, please follow the steps below.

The report is organized into **8 standalone chapters** (the ones that have analytical outputs), each functioning as an individual reproducibility package with its own README, data folder, and code folder.

To reproduce a chapter:

1. Open the chapter folder
2. Read the chapter-specific README
3. Follow the instructions to reproduce the results.

Find below a summary of instructions by chapter and software:

Chapter	Software Used	Instructions
Chapter 1	- Stata - Excel	1. Obtain access to the data not included in the package (see Dataset sections for more details). Place it in the appropriate folder. 2. Update the global in do-file “_Master” in line 2 to your folder’s location, and run the do-file. 3. Follow the instructions in the chapter-specific README to reproduce the figures in the “Excel graph template.xlsx”

Chapter 4.1	- Stata	1. Obtain access to the data not included in the package (see Dataset sections for more details). Place it in the appropriate folder. 2. Update the global in do-file “Main” in line 19 to your folder’s location, and run the do-file.
Chapter 4.2	- Stata	Update the global in the do-file “Create figure for R&D section.do” in line 40 to your folder’s location, and run the dofile.
Chapter 4.3	- Python	Recreate the environment using the environemnt.txt file and run “MVPF_calculation.py”
Chapter 5.1	- Stata	1. Obtain access to the data not included in the package (see Dataset sections for more details). Place it in the appropriate folder. 2. Update the global in do-file “Section 5.1/Codes/0_master_do_file.do” in line 12 to your folder’s location, and run the do-file. 3. Update the global in do-file “Section 5.1b/Codes/0_master_do_file.do” in line 12 to your folder’s location, and run the do-file.
Chapter 5.2	- R - Excel	1. Obtain access to the data not included in the package (see Dataset sections for more details). Place it in the appropriate folder. 2. Open the R Project Chapter 5.2.Rproj. Restore the R environment using renv::restore() and follow the prompts. Alternatively, install the required packages manually using the information provided in the renv.lock file. 3. Run the files “01_Data_Merging_20241025_B.Rmd” and 02_Data_Merging_NewIBNET_20241030.Rmd 4. Follow the instructions in the chapter-specific README to reproduce the results in Excel.
Chapter 5.3	- Excel	The figure is prepared in Excel manually using data documented under data sections.
Chapter 5.4	- Excel	1. Obtain access to the data not included in the package (see Dataset sections for more details). Place it in the appropriate folder. 2. Follow instructions in the chapter-specific README to reproduce the results in Excel.

Software Requirements

The exhibits in the report were verified by the replicator using the following software: Stata 19.0 MP; R version 4.5.1 ; Python 3.10.19; Excel version 2512

Data Availability Statement

The report uses open data (included in the package), limited-access data (available to World Bank staff via the Development Data Hub), and forthcoming data (soon on DDH).

- Only the open data is included.
- Other datasets are on, or will be added to, the WB DDH; users must obtain these separately to replicate the package.

Detailed Data Availability Statement can be found in the chapter-specific README. Find below the list of the datasets used:

Dataset	Provided	Files (file names + location)	URL
World Economic Outlook Database (IMF)	Yes	- Chapter 1/Raw data folder/WEO_country_bridge_table.dta; - WEOApr2025all.xlsx. - Chapter 5.3/Agriculture.xlsx. (sheet: WEO GDP deflator data) - Chapter5.2/Data_Cleaning/00_CountryDataWithMasterKey.xlsx. (sheet: IMF_Infla)	Link
Government Finance Statistics (IMF)	Yes	- Chapter 1/Raw data folder/Preferred level of government list.xlsx - Chapter 1/Raw data folder/GFS_country_bridge_table.dta - Chapter 1/Raw data folder/GFS_SOO.csv.	Link
World Bank Country and Lending Groups	Yes	- Chapter 1/Raw data folder/OGHIST_2025_07_01.xlsx - Chapter 4.1/data/raw_data/OGHIST_2025_10_07.xlsx - Chapter 5.3/Agriculture.xlsx (sheet: Income groups) - Chapter 5.2/Data_Cleaning/CLASS.xlsx - Chapter 5.1/Section 5.1/Data/CLASS_2025_10_07.xlsx	Link
Global Tax Expenditures Database (GTED)	Yes	- Chapter 1/Raw data folder/GTED_FullDatabase_20250505.xlsx - Chapter 4.1/data/raw_data/GTED Data Download GTED (202411).xlsx - Chapter 5.1/Section 5.1/Data/GTED_FullDatabase_20240125.xlsx	Link
IMF Fossil Fuel Subsidies	Yes	- Chapter 1/Raw data folder/Fossil_Fuel_Subsidies.csv	Link
BOOST Aggregator (Sept 2025)	No	- Chapter 1/Raw data folder/aggregator_basic.xlsm - Chapter 1/Raw data folder/BOOST_selection.xlsx - Chapter 1/Raw data folder/BOOST_country_bridge_table.dta	Link
Fiscal Survey (July 2025)	No	- Chapter 1/Raw data folder/WBFD core variables in columns.dta	Link

Corporate Income Tax Incentives Database (World Bank)	No	- Chapter 4.1/data/raw_data/ World Bank CIT Incentives Database - Streamlined data by incentive Jan 2025.xlsx	Forthcoming
R&D Tax Incentives (OECD)	Yes	- Chapter 4.2/data/Updated /RDTAX_02072024152147036	Link
GDP per capita (OECD)	Yes	- Chapter 4.2/data/Updated /PDB_LV_13062024044944980.	Link
VAT Threshold Dataset (IMF TPAF)	Yes	- Chapter 5.1/Section 5.1/Data/VAT_threshold.xlsx	Link
World Development Indicators (WDI)	Yes	- Chapter 1/Raw data folder/WDICountry.csv - Chapter 5.1/Section 5.1/Data/P_Data_Extract_From_WDI.xlsx - Chapter 5.3/Agriculture.xlsx (sheet: GDP) - Chapter 5.2/Data_Cleaning00_CountryDataWithMasterKey.xlsx (sheet: WDI_GDP; WDI_Pop; WDI_Exch_Rate)	Link
CEQ Country-Year Decile Dataset (DDH)	No	- Chapter 5.1/Section 5.1/Data/CEQPartB_deciledata_2024_march.dta	Link
Simulation Results Dataset (DDH)	NO	- Chapter 5.1/Section 5.1b/Data/simulation_results.xlsx	Link
Agriculture Support Data (compiled)	Yes	- Chapter 5.3/Agriculture.xlsx	
Climate Policy Assessment Tool (CPAT) Dataset (DDH)	No	- Chapter 5.4/CPAT 1.0pre_358_SFR.xlsb - Chapter 5.4/Multiscenario Tool 1.332_SFR.xlsb - Chapter 5.4/FIGURE_6.11_REPLICATION	Link
WHO/UNICEF Joint Monitoring Programme (JMP)	Yes	- Chapter 5.2/Data_Cleaning/00_CountryDataWithMasterKey.xlsx (sheets: JMP_Water; JMP_WasteWater; JMP_Code)	Link

Country and Equity Risk Premiums (Damodaran)	Yes	- Chapter 5.2/Data_Cleaning/00_CountryDataWithMasterKey.xlsx (sheet: Damo_2024)	Link
Marginal Value of Public Funds (MVPF) of Green Subsidies	Yes	- Chapter 4.3 /FTT Power - Results S1-2.xlsx - Chapter 4.3/FTT Power - Results S3.xlsx - Chapter 4.3/FTT Transport - Results S1.xlsx - Chapter 4.3/FTT_countries.csv; MVPF_aggregation.xlsx	Link
IBNET Water Utility Datasets	No	- Chapter 5.2/Data_Cleaning/IBNET_2023.xlsx - Chapter 5.2/Data_Cleaning/New_IBNET_2024.xlsx - Chapter 5.2/ Data_Cleaning/New_IBNET_KPI_2024 - Chapter 5.2/Final_Calculations/IBNET_WaterTariff_Calculations_20241031.xlsx - Chapter 5.2/Tariffs data for Fig 6-6.xlsx - Chapter 5.2/Data_Cleaning /IBNET_Utility_MasterFile.xlsx	Forthcoming

Contacts

- Eric Lacey (elacey@worldbank.org)
- Fernando Andres Blanco Cossio (fblanco@worldbank.org)