

That README for the Reproducibility Package for TAXES, SPENDING AND EQUITY: International Patterns and Lessons for Developing Countries

Overview

This package contains the Stata files and excel files necessary to produce the figures and tables in the paper. The directory is divided into:

- 01_data: data inputs for calculations in MS Excel or Stata
- 02_do_files: Stata calculation files
- 03_output_data: output tables containing the data for in figures
- The excel file WSb_figures_2025.xlsx containing the figures in the report.

Data Availability

Source	Dataset	Link	Year	Access
Fiscal incidence data from the 2022 Poverty and Shared Prosperity Report	PSPR_CEQPartA&OECD.dta & PSPR_CEQPartB&EUROMOD.dta	For access, email: Matt Wai Poi mwaipoi@worldbank.org or, Mariano Sosa msosa@worldbank.org	2023	Limited
Government Revenue Dataset	Merged_UNUWIDER.dta	https://www.wider.unu.edu/project/grd-government-revenue-dataset	2023	Public
IMF Government Finance Statistics	IMF_GFS_Expenditure_by_Functions.xlsx	https://data.imf.org/en/datasets/IMF.STA:GFS_COFOG	2023	Public
World Bank Boost database	aggregator_reduced.xlsm	www.worldbank.org/en/programs/boost-portal . For data access reach out to mmastruzzi@worldbank.org .	2021	Limited
World Development Indicators	gdp_curr_lcu.dta & population_wbopen.dta; API_SP.POP.TOTL_DS2_en_csv_v2_3852487.csv	Data access through Stata <code>wbopendata</code> commands	2023	Public
International Institute for Sustainable Development	iisd_FossilFuelSubsidiesTracker_CountryData	https://www.iisd.org/	2022	Included in the package
Author's compilation using data from WB and IMF.	WB_IMF_etry_classification.dta	See Instructions for replicators below. For access email: msosa@worldbank.org	2023	Public

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

Download reproducibility package entirely

Open master do file: 00_master.do

Adjust global "path"

Run 00_master.do

Outputs tables from Stata are stored in the file WSb_stataoutputs.xlsx. The file WSb_figures_2025.xlsx creates the figures in the report using the table results stored in WSb_stataoutputs.xlsx

Instructions to replicate figures using data from the World Bank's 2022 Poverty and Shared Prosperity report

Data files:

This paper used data assembled and prepared for World Bank. Poverty and Shared Prosperity 2022: Correcting Course. Washington, DC: World Bank (2022 PSPR). The 2022 PSPR report is accessible in the link below but not the data:

<https://www.worldbank.org/en/publication/poverty-and-shared-prosperity>

The authors of this paper used two data packages from the 2022 PSPR. One for fiscal incidence analysis data and one for government revenue data.

The fiscal incidence analysis package contains two data files.

1. “PSPR_CEQPartA&OECD.dta”
2. “PSPR_CEQPartB&EUROMOD.dta”

These data files were collected in a non-codable manner drawing from three sources. Data can be traced back to these sources. First, the [CEQ cross-country database](#), a joint work of the World Bank’s Global Solutions Group 3 (GSG3) of the Poverty Department and Tulane University’s CEQ Institute. The World Bank’s GSG3 hosts and updates the CEQ cross-country database. Access is restricted to World Bank staff. Data is accessible through the GSG3 site, Data and Resource section. [GSG3: Fiscal and Social Policies..](#) Secondly, the OECD Income Distribution Database (IDD) accessible at <https://data-explorer.oecd.org/>. Finally, EUROMOD - Tax-Benefit microsimulation model for the European Union accessible at <https://euromod-web.jrc.ec.europa.eu/download-euromod#section-3390>.

The government revenue data package (01_data/Revenue_composition) contains two data files:

1. API_SP.POP.TOTL_DS2_en_csv_v2_3852487
2. Merged_UNUWIDER

The data was collected by the authors for the 2022 PSPR drawing from Government Revenue data from ICTD accessible at <https://www.ictd.ac/dataset/government-revenue-dataset/>

Do files

01_ceq_calculations.do

This file handles the data contained in the fiscal incidence analysis package

03_graphs_report.do

This file handles the data contained in the government revenue data package

Instructions to replicate Figure 4.1a and Figure 4.1b

Handling of World Bank Boost database: 02a_boost_prep.do

File “aggregator_reduced.xlsm” was provided by Massimo Mastruzzi mmastruzzi@worldbank.org containing public spending categories. The file “aggregator_reduced.xlsm” contains an ad-hoc classification and categorizes spending items in to large spending categories. The file 02a_boost_prep.do” imports data from “aggregator_reduced.xlsm”, does necessary data handling tasks, cleans the dataset, sets its format and calculates spending levels as share of GDP. This file merges with two datasets generated in the code and then greened out. “gdp_curr_lcu.dta” and “population_wbopen.dta”. These files contain GDP and population estimates as obtained in the WDI database access through the command WBOPENDATA in August 2023. Activating these commands will allow for a recalculation of the files but the GDP and population estimates may have changed since the time of access.

Handling of IMF GFS data: 02b_img_gfs_prep.do

Database of the GFS Government expenditures by Function were downloaded from https://data.imf.org/en/datasets/IMF.STA:GFS_COFOG. Additional classification variables for the countries are added by merging with the file “WB_IMF_etry_classification.dta”. The file WB_IMF_etry_classification was manually put together by authors using income and regional classification from IMF-WEO and World Development Indicators.

Preparing the data for the figure: 02c_spending_BOOST-GFS.do

This file merges the two databases for the two sources described above BOOST and IMF GFS. It further handles, cleans and selects relevant data. It also classifies countries based on OECD membership using the file “OECD_etry_list.dta”, which was manually constructed by authors based on the country list in this page <https://www.oecd.org/en/about/members-partners.html>

Instructions to replicate Figure 4.2a and Figure 4.2b

The file iisd_FossilFuelSubsidiesTracker_CountryData.xlsx contains the raw data from <https://www.iisd.org/> and auxiliary calculations necessary for the analysis included in the file WSB_figures.xlsx and in the paper.

Note on analysis presented in this manuscript from other publications

This document presents analyses originally published in prior works, which the authors of this manuscript are authorized to use. Appropriate authorship and attribution have been acknowledged. The referenced figures are:

- Figure 3.3 Panel A shows unpublished results from World Bank. 2023c. Indonesia Poverty Assessment: Pathways towards Economic Security.
- Figure 3.3 Panel B shows unpublished results from Amjad, Beenish; Carassco, Haydeeliz; Meyer, Moritz. 2025. The Effects of Taxes and Transfers on Inequality and Poverty in Pakistan. World Bank

- Figure 3.4 Panel A, Figure 3.4 Panel B and, Figure 3.4 Panel C from Bachas, Gadenne and Jensen (2023).
- Figure 3.5 Panel A, and Figure 3.5 Panel B are unpublished results from 2023a. Thailand Public Spending and Revenue Assessment: Promoting an inclusive and sustainable future.
- Figure 3.7 coming from Thomas (2022)
- Figure 4.2 Panel A, Figure 4.2 Panel B, Figure 4.2 Panel C, and Figure 4.2 Panel D coming from the 2022 PSPR
- Figure 4.5 Panel A, and Figure 4.5 Panel B shows unpublished results from 2023a. Thailand Public Spending and Revenue Assessment: Promoting an inclusive and sustainable future

Software Requirements

Stata/MP 18

Code Description

00_master.do: defines global paths and runs all calculations combined.

01_ceq_calculations.do: runs calculations using fiscal incidence data.

02a_boost_prep.do prepares data from Boost for combination with IMF-GFS data.

02b_img_gfs_prep.do prepares data from IMF-GFS for combination with Boost data.

02c_spending_BOOST-GFS.do: combines IMF and Boost data and prepares data for figures.

Folder Structure

WSB_Reproducibility contains all files necessary to reproduce the figures in the paper. Outputs tables from Stata are stored in the file “WSB_stataoutputs.xlsx”. The file “WSb_figures_2025.xlsx” creates the figures in the report using the table results stored in WSB_stataoutputs.xlsx

01_data: contains all data used needed the paper

02_do_files: contains all do files needed for the paper