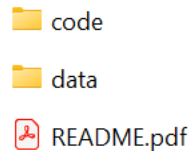


README

Data and Code for International Poverty Rates 2023-24

1. Overview

This document provides information needed to replicate the poverty rates for FY 2023-24. The folders and files included in this replication package are the following:



The code and data folders should be in the same folder (e.g., the folder they were archived in) to run the project do files.

2. Computational requirements

Software Requirements: MS Excel, Stata, R, and Nesstar Explorer are necessary for the full replication of the results in the document.

2.1. Stata

Stata version 14 or higher is required. The team produced this workflow using Stata 18. The guidelines below, sourced from <https://www.stata.com/products/compatible-operatingsystems/>, outline the following computational requirements:

Platforms

Stata for Windows®

- Windows 11*
- Windows 10 *
- Windows Server 2022, 2019, 2016, 2012R2 *

* Stata requires 64-bit Windows for x86-64 processors made by Intel® or AMD (Core i3 equivalent or better)
[Find out if your OS is 64-bit compliant.](#)

Stata for Mac®

- Mac with Apple Silicon or Intel processors
- macOS 11.0 (Big Sur) or newer for Macs with Apple Silicon and macOS 10.13 (High Sierra) or newer for Macs with Intel processors

Stata for Linux

- Any 64-bit (Core i3 equivalent or better) running Linux
- Minimum requirements include the GNU C library (glibc) 2.17 or better and libcurl4
 - Check the output of `ldd -v` within a terminal
- For xstata, you need to have GTK 2.24 installed

Hardware requirements

Package	Memory	Disk space
Stata/MP	4 GB	2 GB
Stata/SE	2 GB	2 GB
Stata/BE	1 GB	2 GB

Stata for Linux requires a video card that can display thousands of colors or more (16-bit or 24-bit color)

The packages needed to run the code are: fastreshape, ainequal, distinct, mdesc, egenmore, glcurve, unique, _gwmean, winsor2, fre, estout and povdeco. The installation of these commands is written in the master.do file. There's an ado folder in the code folder because the original package code had to be slightly modified to avoid a file path issue. The adopath is specified in the master.do file.

3. Memory and runtime requirements

The analysis can be run on a standard desktop computer. The replicator should expect the code to run up to ~1 hour. The initial package weighs around ~8GB, and the final set of folders and files after execution is considerably heavier because the workflow will create many intermediate datasets. If working in a cloud environment, high Wi-Fi speed is of utmost importance. Many datasets are considerably heavy (~1GB individually), and this could slow synchronicity in the cloud. The recommended approach is to pause the synchronization to the cloud during the whole execution of this package or work in a local drive. At the beginning of 07_pov_ineq.do file, the code will delete heavy intermediate datasets. The user can keep these lines by commenting the code out, and the final folder will weigh around ~35 GB.

4. Folder and contents

The folder **code** contains the do-files needed to reproduce the poverty rates. The do files are sequential and numbered from 00 to 08 with sub-sequences along the way.

The folder called **data** contains some of the raw data required for this package to run correctly. There are two subfolders in it: **ces** and **other**. To get all the data necessary for execution, the user needs to download the microdata from the Consumption Expenditure Survey (CES) published by India's government, as explained next in Section 5, and save it in **data → raw → ces**. The folder **data → raw → other** contains other raw data, such as price indices, population figures, item equivalence tables, etc., which the team harmonized for the user's convenience and comes with the package. It is a set of four Excel files:

 cpi_R_U_Jan2022_May2025.xlsx

 item_list.xlsx

 Population_Projections_Health_Ministry.xlsx

 Population_Projections_WDI.xlsx

- (i) Consumer Price Index downloaded from MoSPI.
- (ii) The item list and equivalences using the item information from the 2023-24 CES questionnaires (also 2011-12, 2022-23 for comparison)
- (iii) Population projection numbers publicly shared by the Ministry of Health and Family Welfare (MoHFW).
- (iv) Population projections sourced from World Development Indicators (WDI), while the urban population and extrapolations are based on the team's calculations.

The user can download the underlying data, as explained in Section 5. In addition to the raw and other data included in this package, the workflow will create a folder called “processed,” which houses all the intermediate data used during the computations. This folder also houses a special subfolder for sensitivity checks.

The execution of the **master.do** file from beginning to end will create the final consumption vector (**output → welfare_pov_final.dta**), and all the tables. The code will create the necessary folders and

subfolders to ensure the workflow runs smoothly. Each empirical table will be a worksheet of the spreadsheet “**tables.xlsx**” which will be exported in **output → excel**.

5. Data availability

The raw data of CES can be downloaded from the Ministry of Statistics and Programme Implementation (MoSPI) microdata library, a web-based cataloguing portal powered by the National Data Archive (NADA). The user needs to create an account and log in to <https://microdata.gov.in/NADA/index.php/home> before clicking the links below. The raw files will be downloaded in a zip/rar format. Then the user needs to extract fifteen CSV files and save them in the **raw→ces** folder. The contents of this package reflect the data as downloaded in February 2026.

The file that the user needs to download and its link is the following:



Filename: **HCES_Data_2023-24_Csv.zip**

URL: <https://microdata.gov.in/NADA/index.php/catalog/237/get-microdata>

Survey year: 2023-24

The user will find a list of 15 CSV files that look like the following:

-  LEVEL - 01 (Section 1 and 1_1).csv
-  LEVEL - 02 (Section 3).csv
-  LEVEL - 03.csv
-  LEVEL - 04 (Section 4_1).csv
-  LEVEL - 05 (Sec 5 6).csv
-  LEVEL - 06 (Section 7).csv
-  LEVEL - 07 (Section 4_2).csv
-  LEVEL - 08 (Section 8_1).csv
-  LEVEL - 09 (Section 9 10 11).csv
-  LEVEL - 10 (Section 12).csv
-  LEVEL - 11 (Section 4_3).csv
-  LEVEL - 12 (Section 13).csv
-  Level - 13 (Section 14).csv
-  LEVEL - 14 (Section A1,B1 C1).csv
-  LEVEL - 15 (Section 1_1, A2,B2 C2).csv

The user needs to save these CSV files in “**data\raw\ces**”. In this folder, there are five pre-existing Stata datasets produced by the reproducibility package for the 2011-12 and 2022-23 poverty estimates, which are considered raw inputs for this pipeline. The reproducibility package can be found at: <https://reproducibility.worldbank.org/index.php/catalog/280> .

 master_consumption_2022.dta
 master_hh_2011.dta
 master_hh_2022.dta
 master_item_list_2022.dta
 welfare_agg_2011_2022_all.dta

In that package, the file 02_master_datasets_create.do produces the files “master_consumption_2022.dta”, “master_hh_2011.dta”, “master_hh_2022.dta”, “master_item_list_2022.dta” and do file 06_welfare_agg.do creates the file “welfare_agg_2011_2022.dta” (in that pipeline, the intermediate file created is called “welfare_agg_21.dta” in reference to 2021 PPP, but it was renamed here to avoid confusion with survey year).

The underlying data of the harmonized information in the Excel files in **data → raw → other** can be downloaded and verified from MoSPI, the Ministry of Health and Family Welfare (MoHFW), and the The World Bank website

- **CPI Index Current Series (Base 2012) - Jan'2013 onwards**

- Filename: cpi_R_U_Jan2022_May2025.csv
- Date range: January 2022 – May 2025
- Source: Consumer Price Indices Warehouse, MoSPI
- Access year: 2025
- URL : <https://cpi.mospi.gov.in/> (Select Time Series > Base 2012 > Current Series . State: Click *Check All*. Group: Select *All Groups and Subgroups*. Order by: Group. Click View Indices, then download the data in .csv format. Once downloaded, the file was reshaped to long format so that each row represents a unique combination of state × month × year × subgroup, with separate entries for the rural, urban, and combined index values)

- **Population Projection for India and States, 2011-2036**

- Filename: Population_Projections_Health_Ministry.xlsx. It contains two sheets – “MoH_pop_projections_original” is an Excel conversion of the output tables originally published in PDF format by the Ministry of Health and Family Welfare (MoHFW). “stata_sheet” draws from Tables 8 and 9 of MoH_pop_projections_original, covering state-level population projections at the urban, rural and combined levels from 2011 to 2036, reformatted for use in Stata.
- Source: National Health Ministry
- URL: https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf
- Access year: 2025

- **Population estimates and projections, World Development Indicators**
 - Filename: DataBank | World Development Indicators
 - Source: DataBank World Bank
 - URL: <https://databank.worldbank.org/source/population-estimates-and-projections>
 - Access year: 2025
 - Variable name: Rural population, Urban population

6. Instructions for correct execution

All the workflow can be executed from the **master.do** file.

The workflow is organized sequentially, and even though some of the sequences could be added in a loop, this was purposely avoided for clarity. It's worth noting that some sections depend on the value of a global variable that defines a particular scenario. If the user wants to reproduce only one part of the code, they must consider where the global is initialized and dropped, and execute that chunk of code with the precaution of not keeping the global in memory while running other code.

Below is a step-by-step execution instruction:

1. **Set the user path in the 00_master.do file:** Copy the folder path where the output of the execution of this package is to be saved and put it in the user global variable using "" as a string. For instance, if the package was saved in C:/Users/username/reproducibility_package, then the path of the global variable user should be written like this:

```
global user "C:/Users/username/reproducibility_package"
```
2. **Execute the master.do file:** Select all the code in the **00_master.do** file and execute it. The execution could take up to 1 hour. As explained before, check the output folder for tables.

7. Do files

This workflow will be executed for several scenarios as explained in the technical note. Do-files 01, 01a, 01b, 02, 02a, 02b, 02c, need to run only once because they set the ground and do not define any scenario. Then do files 03-06 are run for every scenario except 05, which is the spatial deflators and needs to be run only in the baseline scenario. The workflow is structured around 17 main do-files.

00_master. This file defines the global variable that sets the working directory path. It also creates a global variable $\{s\}$, which pivots across scenarios to estimate the marginal effects of these adjustments on the welfare aggregate. The file contains an ordered sequence of executions to replicate all the numbers and figures.

01_ces_clean.do: imports the raw microdata for 2023-24 and does some preliminary cleaning

01a_cpi_clean.do: structures the CPI information with equivalences to be applied during the workflow

01b_population_MoH: introduces the population figures in the workflow

02_master_datasets_create: creates a set of master datasets to centralize all the information at the HH-level characteristics, HH consumption, item level, etc. This helps to facilitate the coding and circumvent the need to go to the raw files in each do file.

02a_match_survey_cpi_items.do: creates a match between items from the consumption survey to items in CPI tables.

02b_PDS_imputation_20231: creates a vector of imputed market-equivalent values for PDS subsidized items for 2023-24

02c_nonfood_school_imputation_2023.do: estimates the imputed value of free items received when attending government educational institutes

03_consumption_adjustments.do: reads the value of the global `{s}` in the master do file and does the corresponding adjustments depending on the scenario in which the marginal effects modify the welfare aggregate. It includes six distinct scenarios that modify the nominal expenditure aggregate (before within-year and spatial deflation):

- **Baseline scenario (`{s} = ""`)** – This scenario is the one reported by MoSPI, just nominal expenditure without any adjustments.
- **Food and non-food imputations (`{s} = "_pds"`)** – Building on the nominal expenditure aggregate, this scenario imputes the prices of subsidized and free food and non-food items. The output is a nominal expenditure aggregate plus the imputations.
- **Removal of lumpy expenditure (`{s} = "_lumpexp"`)** – This scenario excludes lumpy expenditures, such as durables, furniture, and jewelry, from the nominal expenditure aggregate.
- **Removal of rents (`{s} = "_rents"`)** – This scenario removes rental expenses from the nominal expenditure aggregate.
- **Removal of hospitalization expenses (`{s} = "_hosp"`)** – This scenario removes hospitalization expenses from the nominal expenditure aggregate.
- **Final scenario (`{s} = "_all"`)** – The final scenario combines all previous adjustments.

04_deflators_wy: carries out the within-year deflation process aimed at adjusting expenditures, so they are all expressed in survey-year average prices. This deflation is executed for each of the six scenarios, and the output is a within-year deflated expenditure aggregate

05_deflators_spatial: performs the spatial deflation process to adjust expenditures to all be expressed in India's average prices. It starts from the within-year deflated expenditure aggregate. The spatial deflators are calculated for the baseline scenario only and are used for all other scenarios and sensitivity analyses.

06_welfare_agg.do: This file creates the Mean Per Capita Expenditure (MPCE) after the expenditures have been adjusted within-year and spatially. For the baseline scenario, the output is a **real expenditure aggregate (EA)** (no adjustments to aggregate, just deflation), for the `_all` scenario, the output is a **real welfare aggregate (WA)** (all adjustments to WA + within-year + spatial deflation).

07_pov_ineq.do: Computes the international poverty rates and Gini Index based on the real welfare aggregate for the scenario with all adjustments taking place (`_all`) only. This do file contains

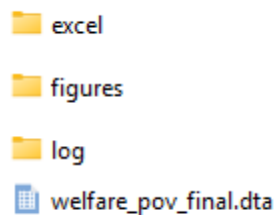
parameters such as the Purchasing Power Parity (PPP) factor and the International Poverty Lines that can be easily changed to show alternative scenarios.

08_tables.do: This do file tabulates relevant intermediate output and presents it in a table format clearly to match the tables in the report.

09_figures.do: This do file creates the figures that match the report.

8. Output

The main output folder created by the workflow contains the consumption vector (**welfare_pov_final.dta**), which includes the **real welfare aggregate** and the poverty dummies. The folder would look like this:



All the tables will be located in **output/excel/tables.xlsx** and all the figures will be located in **output/figures**.