





README

Data and Code for “India: Trends in Poverty, 2011-12 to 2022-23. Methodology note”

1. Overview

This document provides information needed to replicate the results in “**India: Trends in poverty, 2011-12 to 2022-23. Methodology note**”. The folders and files included in this replication package are the following:

-  code
-  data
-  README.pdf
-  Technical Note_India Poverty Measurement.pdf

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. The installation of these commands is written in the master.do file.

2.2. R and R Studio

R is maintained by an international team of developers who make the language available through the web page of [The Comprehensive R Archive Network](#). The top of the web page provides three links for downloading R. Follow the instructions based on your operating system.

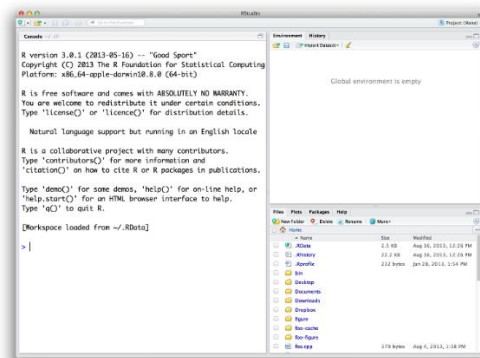
2.2.1. Using R

R isn't a program that the user can open and start using, like Microsoft Word or Internet Explorer. Instead, R is a computer language, like C, C++, or UNIX. The user can use R by writing commands in the R language and asking their computer to interpret them. In the old days, people ran R code in a UNIX terminal window. Now, almost everyone uses R with an application called RStudio.

2.2.2. RStudio

RStudio is an application like Microsoft Word—except that instead of helping you write in English, RStudio helps you write in R. The RStudio interface looks the same for Windows, Mac OS, and Linux.

The user can [download RStudio](#) for free. When the user opens RStudio, a window appears with three panes, as shown in the following figure. The largest pane is a console window. This is where the user will run their R code and see results. The console window is exactly what the user would see if they ran R from a UNIX console or the Windows or Mac GUIs. Everything else the user sees is unique to RStudio. Hidden in the other panes are a text editor, a graphics window, a debugger, a file manager, among other things.



Even if the user uses RStudio, they still need to download R to their computer. RStudio helps use the version of R that lives on the user's computer, but it doesn't come with a version of R on its own. Once the user has both R and RStudio on their computer, they can begin using R by opening the RStudio program.

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 ~4.5 hours. 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 ~61 GB as compared to ~85GB without this deletion.


4. Folder and contents


The folder **code** contains the do-files needed to reproduce the output presented in sections 4 and 5 of the technical note. The do files are sequential and numbered from 00 to 08 with sub-sequences along the way. It also houses a sub-folder called **sensitivity_checks** to produce estimates outlined in section 6.

The folder called **data** contains some of the raw data necessary for the correct execution of this package. 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_subgroups_Jan11_May24.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 2011-12 and 2022-23 CES questionnaires.
- (iii) Population projection numbers publicly shared by the Ministry of Health and Family Welfare (MoHFW)¹.
- (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 and figures except the housing sensitivity scenario, which needs to be run separately. The code will create the necessary folders and

¹ The data originally shared by the Ministry was in PDF. The team converted Table 8 and 9 from the report to excel.

subfolders for the smooth execution of the workflow. Each empirical table will be a worksheet of the spreadsheet “**tables.xlsx**” which will be exported in **output → excel** and figures will be exported to the folder **output → figures**. These Excel and figures subfolders contain all the elements in the document, but the workflow creates other intermediate files as well that the user can use for deep dives or ad-hoc analysis.

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 and open some files using Nesstar Explorer and then save them in the **raw→ces** folder. There are two subfolders: 2011-12 and 2022-23. At the same time, there are two folders inside 2011-12: URP for schedule type 1 and MMRP for schedule type 2. The contents of this package reflect the data as downloaded in April 2025.

The three files that the user needs to download and their links are the following:

Household Consumption Expenditure Survey: 2022-23

India,

NSSO

Collection: Household Consumption Expenditure

ID: DDI-IND-MOSPI-NSSO-HCES22-23 Last modified: Dec 20, 2024 Views: 65348

 Public use data files

Filename: **Unit level data of HCES 2022-23 round.zip**

URL: <https://microdata.gov.in/NADA/index.php/catalog/224/download/2548>

Survey year: 2022-23

Access year (public release): 2024

Household Consumer Expenditure, NSS 68th Round Sch1.0 Type 2 : July 2011 - June 2012

Type - 2

India, 2011-2012

National Sample Survey Office, NSSO

Collection: Household Consumption Expenditure

ID: DDI-IND-MOSPI-NSSO-68Rnd-Sch2.0-July2011-June2012 Last modified: Mar 25, 2019 Views: 480723

 Public use data files

Filename: **Nss68_1.0_Type2_new format.rar**

URL: <https://microdata.gov.in/NADA/index.php/catalog/126/download/1595>

Survey year: 2011-12

Access year (public release): 2013

India, 2011-2012

National Sample Survey Office, NSSO

Collection: Household Consumption Expenditure

ID: DDI-IND-MOSPI-NSSO-68Rnd-Sch1.0-July2011-June2012 Last modified: Mar 25, 2019 Views: 332257

[Public use data files](#)

Filename: **Nss68_1.0_Type1_new format.rar**

Source: Online microdata library, MoSPI

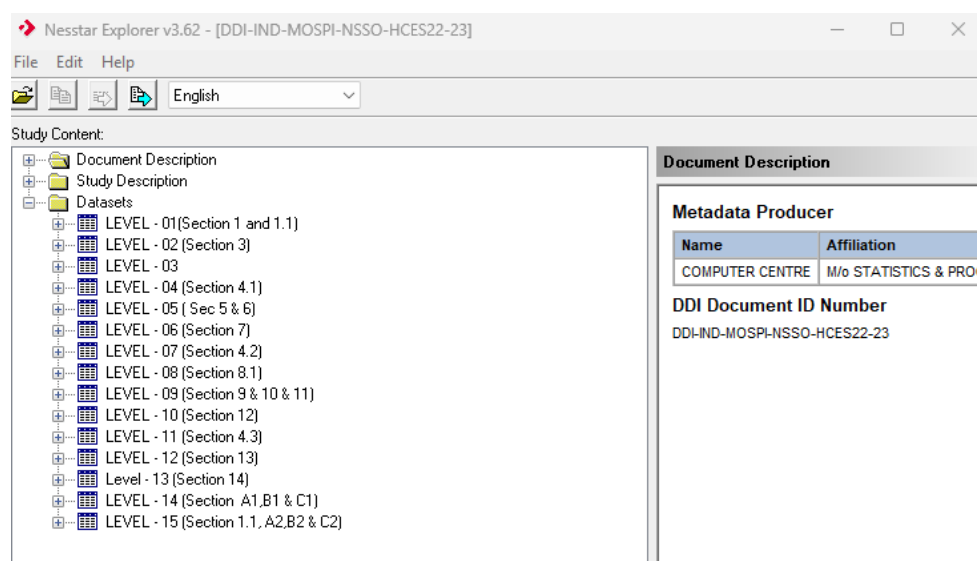
URL: <https://microdata.gov.in/NADA/index.php/catalog/1/download/1596>

Survey year: 2011-12

Access year (public release): 2013

Once these three files are downloaded, the user needs to install Nesstar Explorer (executing the file **NesstarExplorerInstall.exe** in a folder with the 2022-23 data).

To download the data for 2022-23, the user needs to open the file **DDI-IND-MOSPI-NSSO-HCES22-23.Nesstar** is contained in the unit-level data of **HCES 2022-23 round.zip**. The Nesstar screen looks like the one shown in the screenshot below.



The user needs to click on a dataset and select **File > Export Dataset**. The export format should be STATA 8 (*.dta) and the export folder file path should be "...\\data\\raw\\ces\\2022-23" .

Since the file names are not compact and have special characters, it leads to some issues with downloading due to their long names (especially in 2011). The team has used a compact naming convention to export the files in the raw data folders. It is important to follow this convention for the correct execution of the workflow. The following was the naming convention for 2022-23:

Dataset name in Nesstar	Saved in data/raw/ces/2022-23
LEVEL - 01(Section 1 and 1.1)	level01_2022.dta
LEVEL - 02 (Section 3)	level02_2022.dta
LEVEL - 03	level03_2022.dta
LEVEL - 04 (Section 4.1)	level04_2022.dta
LEVEL - 05 (Sec 5 & 6)	level05_2022.dta

LEVEL - 06 (Section 7)	level06_2022.dta
LEVEL - 07 (Section 4.2)	level07_2022.dta
LEVEL - 08 (Section 8.1)	level08_2022.dta
LEVEL - 09 (Section 9 & 10 & 11)	level09_2022.dta
LEVEL - 10 (Section 12)	level10_2022.dta
LEVEL - 11 (Section 4.3)	level11_2022.dta
LEVEL - 12 (Section 13)	level12_2022.dta
Level - 13 (Section 14)	level13_2022.dta
LEVEL - 14 (Section A1,B1 & C1)	level14_2022.dta
LEVEL - 15 (Section 1.1, A2,B2 & C2)	level15_2022.dta

For CES 2011-12 Type1 (Uniform Recall Period), the user needs to open the Nesstar file **nss68_consumer_expenditure_type1.Nesstar** inside the file **Nss68_1.0_Type1_new format.rar** in the folder **survey0 → data** , then follow the same steps as above. The export format should be STATA 8 (*.dta) and the export folder file path should be “...\data\raw\ces\2011-22\URP” . The naming convention to download these files was

Dataset name in Nesstar	Saved in data/raw/ces/2011-12/URP as:
Identification of Sample Household - Block 1 and 2 - Level 1 - 68	block0102_2011_urp.dta
Household Characteristics - Block 3 - Level 2 - 68	block03a_2011_urp.dta
Household characteristics - Block 3 - Level 3	block03b_2011_urp.dta
Demographic and other particulars of household members - Block 4 - Level 4 - 68	block04_2011_urp.dta
Consumption of cereals-pulses- milk and milk products during the last 30 days - Block 5.1- 5.2- 6 - Level 5 - 68	block0506_2011_urp.dta
Consumption of clothing, bedding and footwear during last 30 and 365 days - Block 7 and 8 - Level 6 - 68	blocks0708_2011_urp.dta
Expenditure on Education and Medical (institutional) goods and services - Block 9 - Level 7 - 68	block09_2011_urp.dta
Expenditure on miscellaneous goods and services including medical(non-institutional), rents and taxes during the last 30 days. Block 10 - Level 8 -68	block10_2011_urp.dta
Expenditure for purchase and construction (including repair and maintenance) of durable goods for domestic use- Block 11 - Level 9 - 68	block11_2011_urp.dta
Information on Ayurveda, Yoga, Naturopathy, Unani, Siddha, Homeopathy(ASYUSH) - Block 13 - Level 10 - 68	block13_2011_urp.dta
Summary of Consumer Expenditure - Block 12 - Level 11 - 68	block12_2011_urp.dta

For CES 2011-12 Type2 (Mixed Modified Recall Period), the user needs to open the Nesstar file **nss68_consumer_expenditure_type2.Nesstar** in the file **Nss68_1.0_Type2_new format.rar** in the folder **survey0 → data** , then follow the same steps as above. The export format should be STATA 8 (*.dta) and the export folder file path should be “...\data\raw\ces\2011-22\MMRP” . The naming convention to download these files was

File name in Nesstar	Saved in data/raw/ces/2011-12/MMRP as:
-----------------------------	---

Identification of Sample Household - Block 1 and 2 - Level1-type2-68 Round	block0102_2011_mmrp.dta
Household characteristics - Block 3 - Level2 - type2 - 68	block03a_2011_mmrp.dta
Household characteristics - Block 3 - level3 - type2 - 68	block03b_2011_mmrp.dta
Demographic and other particulars of household members - Block 4 - Level 4 - Type 2 - 68	block04_2011_mmrp.dta
Consumption of cereals-pulses- milk and milk products during the last 30 days - Block 5.1- 5.2- 6 - Level 5 - type 2 - 68	block0506_2011_mmrp.dta
Consumption of clothing, bedding and footwear during last 365 days - Block 7 and 8 - Level 6 - Type 2 - 68	blocks0708_2011_mmrp.dta
Expenditure on Education and Medical (institutional) goods and services during the last 365 days - Block 9 - Level 7 - type 2 - 68	block09_2011_mmrp.dta
Expenditure on miscellaneous goods and services including medical(non-institutional), rents and taxes during the last 30 days. Block 10 - Level 8 - type 2 - 68	block10_2011_mmrp.dta
Expenditure for purchase and construction (including repair and maintenance) of durable goods for domestic use- Block 11 - Level 9 - Type 2 - 68	block11_2011_mmrp.dta
Summary of Consumer Expenditure - Block 12 - Level 11 - type 2 - 68	block12_2011_mmrp.dta
Information on Ayurveda, Yoga, Naturopathy, Unani, Siddha, Homeopathy(ASYUSH) - Block 13 - Level 10 - type 2 - 68	block13_2011_mmrp.dta

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 Back Series (Base 2012) - Jan'2011 to May'2013**
 - Filename: CPIIndex_Jul11-To-Jun12.csv
 - Source: Consumer Price Indices Warehouse, MoSPI
 - URL: https://cpi.mospi.gov.in/TimeSeries_BackSeries_2012.aspx
 - Access year: 2012
- **CPI Index Current Series (Base 2012) - Jan'2013 onwards**
 - Filename: CPIIndex_Aug22-To-Jul23.csv
 - Source: Consumer Price Indices Warehouse, MoSPI
 - URL: https://cpi.mospi.gov.in/TimeSeries_2012.aspx
 - Access year: 2022
- **Population Projection for India and States, 2011-2036**
 - Filename: Report_Population_Projection_2019.pdf
 - Source: National Health Mission
 - URL: https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf
 - Access year: 2019
- **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

Most of the workflow can be executed from the **master.do** file. The only exception is the housing sensitivity scenario, which needs an input generated in R. For this, the user will need to run a script in R and then execute more Stata code to get the tables and figures related to this topic

The workflow is organized sequentially, and even though some of the sequences could be added in a loop, it was purposely avoided for clarity of the sequence. It's worth noting that some sections depend on the value of a global variable that defines a particular scenario. In case the user wants to reproduce only one part of the code, they must consider where the global is initiated and dropped, and execute that chunk of code with the precaution of not keeping the global in memory and 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 several hours (~4.5). As explained before, check the output folder for tables and figures. In line 202, there is an exit command so the user can then do the more manual housing section in R and Stata. This is the last chunk of code in the master.do file and looks like this:

```
202     exit
203
204     *Housing sensitivity_scenario
205     do "${user}/code/sensitivity_checks/add_housing/02a_housing_2022_prematch.do"
206     *---> (((Run R script 02b_matching_rents_2022.R for matching exercise))) <-----
207     do "${user}/code/sensitivity_checks/add_housing/02c_housing_2022_postmatch.do"
208     do "${user}/code/sensitivity_checks/add_housing/03_consumption_adjustments.do"
209     do "${user}/code/sensitivity_checks/add_housing/04_deflators_wy.do"
210     do "${user}/code/sensitivity_checks/add_housing/06_welfare_agg.do"
211     do "${user}/code/sensitivity_checks/add_housing/07_pov_ineq.do"
212     do "${user}/code/sensitivity_checks/add_housing/09_housing_tables_figures.do"
213     /* -----
214     THE END
215     ----- */
```

The following are the step-by-step instructions to run this part:

1. Execute the "**02a_housing_2022_prematch.do**" file from the first line of the last chunk of code in the master.do file

2. With R Studio, open the R file **02d_matching_rents_2022.R** located in **code/sensitivity_checks/add_housing** and copy the user's folder path of subfolder **data/processed/ces/sensitivity_checks/add_housing** in the `setwd("")` command between the `"`. For example:

```
setwd("C:/Users/username/reproducibility_package/data/processed/ces/sensitivity_checks/add_housing")
```

Please note R uses forward slashes / to separate paths (instead of backward slashes \)

3. Select the R script and execute it
4. Wait for the end of the execution, which could take up to 20 minutes, and the output will create a dataset called **train_R.dta**
5. Check that the **train_R.dta** was created in the folder `"data/processed/ces/sensitivity_checks/add_housing/train_R.dta"`
6. Execute the last 6 lines of code that run: **02c_housing_2022_postmatch.do**, **03_consumption.do_adjustments.do**, **04_deflators_wy.do**, **06_welfare_agg.do**, **07_pov_ineq.do** and **09_housing_tables_figures.do**

The execution of these lines will add the tables and figures related to housing to the output. Specifically, Table 8 and Table A.6.1. and Figures 4, A.6.1, A.6.2, A.6.3, and A.6.4

7. Do files and R script

This workflow will be executed for several scenarios as explained in the technical note. Do-files 01, 01a, 01b, 02, 02a, 02b, 02c, 02d, 02e 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 18 main do-files.

00_master.do: This file creates the global variables to set the path of the working directory. It also creates a global variable `$(s)` which will work to pivot among different scenarios to understand 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 2022-23, 2011-12 (MMRP), 2011-12 (URP) 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_2011.do: creates a match between items from the consumption survey to items in CPI tables for 2011

02b_match_survey_cpi_items_2022.do: creates a match between items from the consumption survey to items in the CPI tables for 2022

02c_PDS_imputation_2011: creates a vector of imputed market-equivalent values for PDS subsidized items for 2011

02d_PDS_imputation_2022.do: creates a vector of imputed market-equivalent values for PDS subsidized and free items for 2022.

02e_nonfood_school_imputation_2022.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: carries out the spatial deflation process aimed at adjusting expenditures so they are all 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 used for every other scenario or sensitivity analysis.

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 tabulates relevant intermediate output and constructs a figure to match the figures in the report.

Code for sensitivity checks

Inside the code folder, there is a subfolder for sensitivity checks. The sensitivity checks try to mimic the main workflow, but there are some alterations in some branches of the workflow.

Outliers (subfolder: outliers)

- This exercise mostly mimics the main workflow, with the exception that do file **04_outliers_flag_deflators_wy.do** creates an identifier variable for outliers of unit values. The file runs a process of unit value standardization to define the range outside which an observation can be classified as an outlier (3 or 4 SD). This do file also carries out the within-year deflation. In addition, there are some tabulations that are exported in Excel that are turned off in the code, and the user can activate them. These tabulations show the outliers by item and could be useful for exploration for the interested user.
- **06_outliers_welfare_agg.do:** Creates the welfare aggregate for this sensitivity scenario
- **07_outliers_pov_ineq.do:** Calculates poverty and inequality for this sensitivity scenario

Intertemporal (subfolder: inter_temporal)

- **05_intertemporal_uv_v1.do:** Creates temporal deflators based on unit values using all the items included in 2011-12 and 2022-23 survey. In addition, there are some tabulations that are exported in Excel that are turned off in the code, and the user can activate them. These tabulations show the inflation by item and items not common in both years and could be useful for exploration for the interested user.
- **05_intertemporal_uv_v2.do:** Creates temporal deflators based on unit values using only the common items between the 2011-12 and 2022-23 survey. In addition, there are some tabulations that are exported in Excel that are turned off in the code, and the user can activate them. These tabulations show the inflation by item and items not common in both years and could be useful for exploration for the interested user.
- **06_intertemporal_welfare_agg.do:** Creates the welfare aggregate for this sensitivity scenario
- **07_intertemporal_pov_ineq.do:** Calculates poverty and inequality for this sensitivity scenario

Household Size (subfolder: hhsz)

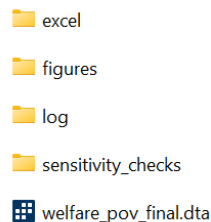
- **06_welfare_agg_h3.do:** Creates the WA with household size from CSQ schedule
- **06_welfare_agg_h4.do:** Creates the WA with household size from DGQ schedule
- **07_pov_ineq_h3.do:** Calculates poverty and inequality with household size from CSQ schedule
- **07_pov_ineq_h4.do:** Calculates poverty and inequality aggregate with household size from DGQ schedule

Housing (subfolder: add_housing)

- **02a_housing_2022_prematch.do:** Cleans the data and creates a dataset that will work as an input for the next R script which will carry the matching exercise.
- **02b_matching_rents_2022.R:** Carries out a matching exercise to select owners that have similar characteristics to renters and can be used to expand the sample of renters in areas where the rental markets are thin (e.g., rural areas).
- **02c_housing_2022_postmatch.do:** Creates a vector of estimated hedonic rents
- **03_consumption_adjustments.do:** Adds the estimated vector of hedonic rents to the nominal expenditure aggregate
- **04_deflators_wy.do:** Carries out within-year deflation for this sensitivity scenario
- **06_welfare_agg.do:** Creates the welfare aggregate for this sensitivity scenario
- **07_pov_ineq.do:** Calculates poverty and inequality for this sensitivity scenario

8. Output

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



In analogy to *code* and *data processed*, this folder houses a special subfolder for sensitivity checks. All the empirical tables from the report will be located in **output/tables.xlsx** except the non-empirical tables (Tables 1,2,3,4,5,7,9, A.1.1 and A.1.3). All Figures will be located in **output/figures** except the non-empirical figures (Figure 1 and A.5.1.)