# Guide to data files for replication of "Does Inequality Affect the Needs of the Poor?"

## Overview

The code in this replication package constructs the analysis file from two data sources (NSS consumer surveys, NSS employment unemployment surveys) along with secondary data source and cross-files using Stata 17.0. One main file runs all of the code to generate the data for the 18 figures and 24 tables in the paper and appendix. The replicator should expect the code to run for about 48 to 76 hours.

## 1 Data Availability and Provenance

This paper draws on several data sets that are available upon registration, and secondary data such as cross-files that are not publicly available but described in the manuscript and replication files. Interested researchers can download the publicly available data and reproduce the results of the paper using their own cross-files.

### 1.1 Statement about Rights

- ✓ I certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ✓ I certify that the authors of the manuscript have documented in the LICENSE.txt that this is **confidential data not for publication**.

### 1.2 Summary of Availability

- ✓ **No data can be made** publicly available.

### 1.3 Details on each Data Source

**Available upon registration**

1. Household Consumer Expenditure data from the National Sample Surveys of India (38th, 43rd, 50th, 55th, 61st and 66th rounds).

Citation: National Sample Survey Office (NSSO). Household Consumption of Various Goods and Services in India. Ministry of Statistics and Programme Implementation, Government of India.

Data can be downloaded from the Microdata Indian Government website: `https://microdata.gov.in/nada43/index.php/catalog/central/about`

2. Employment and Unemployment data from the National Sample Surveys of India (50th, 55th, 61st and 66th rounds).

   Citation: National Sample Survey Office (NSSO). Employment and Unemployment Situation in India. Ministry of Statistics and Programme Implementation, Government of India.

   Data can be downloaded from the Microdata Indian Government website: `https://microdata.gov.in/nada43/index.php/catalog/central/about`

**Freely available**

1. Poverty line data from a report published by the Planning Commission of the Government of India.

   Citation: Planning Commission (2014). Report of the Expert Group to Review the Methodology for Measurement of Poverty. Government of India.

2. Inequality data from the World Inequality Database - India

   Citation: World Inequality Lab. World Inequality Database - India. Retrieved [March 13, 2021]

   Data can be downloaded from the World Inequality Database website: from `https://wid.world/country/india/`

3. Shape files for India from the GADM database

   Citation: GADM. (2011). Database of Global Administrative Areas (version 2.0) [India].

   Data can be downloaded from the GADM Database website: from `https://gadm.org/download_country.html`

Note: All websites accessed and correct as of September 23, 2024.

**Not available: cross-files and other secondary data**

- FItemXX_nutrients.dta: XX stands for each NSS survey round. This file contains the quantity and value of consumption for each good in the NSS surveys and for each household. It also contains calories, proteins and fats for each item calculated using nutrient conversion adjustment factors based on the charts presented in the relevant NSS report for that survey round. The raw NSS files were graciously provided by David Atkin for the

authors' own use. The various variable names are self-explanatory and correspond to the NSS schedule 1 questionnaires.

- clean_nssXX.dta, indXX.dta, medpriceXX_fsu_cal.dta: The cleaned household and individual data from NSS schedule 1, and the village-level median prices by item. The files were graciously provided by David Atkin for the authors' own use, and appropriately modified by Eve Colson-Sihra in a common work. They include geographical markers harmonized across rounds.

- Other files: additional files contain the various concordances (geographical, item) required for the replication.

**Data Organization**

The data is confidential and not for publication. The source data files are organized as follows:

- Folders Data/Raw/nssXX for XX=50,55,61,66: contain household consumer expenditure survey for NSS round XX, FItemXX_nutrients.dta, clean_nssXX.dta, indXX.dta, medpriceXX_fsu_cal.

- Folders Data/Raw/nssXX for XX=38,43: contain clean_nssXX.dta, indXX.dta.

- Folders Data/Raw/employment/nssXX for XX=50,55,61,66: contain employment-unemployment survey for NSS round XX.

- Folder Data/Raw/cross contains concordance for occupations (changingcodes7087_full.dta, concordance_nco_6804.dta), items (itemcodeall.dta, meter_to_unit.dta, nss66_concordance.dta, q_to_cal.dta), districts (stratum_to_district_urban_nss50.dta, dist_od_XX for XX=50,55,61,66) and the Indian government poverty line across rounds (poverty_line.dta).

- Folder Data/Raw/other contains inequality data WID_Data_13032021-123538.csv and the folder "map" with shape files for districts in India.

# 2 Computational requirements

## 2.1 Software requirements

✓ The replication package contains one program to install all dependencies and set up the necessary directory structure.

- Stata (code was last run with version 17.0): egenmore, reghdfe, estout, ineqdeco, shp2dta, spmap, addplot

- The program "setup_ineqneed.do" will install all dependencies locally, and should be run once.

Approximate time needed to reproduce the analyses on a standard 2024 desktop machine is 2-3 days. Approximate storage space needed is 15 GB.

# 3 Description of code

### Master and Set-up Files:

- setup_ineqneed.do provides the required Stata packages for replication.

- master_ineqneed.do runs all do files from the raw data until the production of tables and figures. Each file is documented via comments within this master file. The programs were last run top to bottom on September 29, 2024.

### Intermediate Files:

Each file in Intermediate-Files extracts and reformats datasets from the raw data. *Note: These should be run in order, such as in master_ineqneed.do.*

- make_dataset_hh.do

- clean_raw_cons50.do

- clean_raw_cons55.do

- clean_raw_cons61.do

- clean_raw_cons66.do

- make_food_in_calories.do

- make_harmonized_quantity.do

- make_data_consumption_prices.do

- make_tempotherexp.do

- make_dataset_malnutrition.do
- make_rd.do
- make_iv_mpce.do
- make_iv_rd.do
- make_dataset_analysis.do
- make_dataset_statcons.do
- make_cal_per_rs.do
- make_ineq_corr_d.do
- make_rs.do
- make_iv_rs.do
- make_iv_mpce_nco2.do
- make_iv_rd_nco2.do
- make_rd_withmiddle.do
- make_rd_withupper.do
- make_iv_rd_middle.do
- make_iv_rd_upper.do
- make_distance_shares.do
- make_iv_price.do
- make_stone_price_index.do
- make_pricequality.do
- make_dataset_gmm_estimation.do
- make_dataset_gmm_estimation_all.do
- make_dataset_reducedform.do
- make_est_gmm_base.do
- make_est_gmm_lps.do
- make_est_gmm_all.do
- make_est_gmm_nonpoor.do

- make_dataset_gmm_results.do
- clean_raw_cons61_charity.do
- clean_raw_cons66_charity.do
- make_food_in_calories_charity.do
- make_harmonized_quantity_charity.do
- make_data_consumption_prices_charity.do
- make_data_debt.do
- make_rd_debt.do
- make_regdebt.do

**Analysis Files:**

- Each file in Output-Files is named after the table or figure it replicates (e.g., table1.do replicates Table 1 of the manuscript)
- These files only create tables and figures, no dataset.

## Step-by-Step Instructions to Run Replication Package Code

1. Gather the Raw Data Outlined Above in a Single Folder Called "*Raw*":
   - All Household Consumer Expenditure data downloaded from the National Sample Surveys of India (38th, 43rd, 50th, 55th, 61st and 66th rounds). These datasets should be placed in their respective sub-folders called "*nss38*", "*nss43*", "*nss50*", "*nss55*", "*nss61*" and "*nss66*"
   - All Employment and Unemployment data from the National Sample Surveys of India (50th, 55th, 61st and 66th rounds). These datasets should be placed in a sub-folder called "*employment*" and sub-sub-folder for each nss round.
   - All cross-files and other secondary datasets described above. These datasets should be placed in their respective sub-folders "*cross*" and "*other*".
2. Create a Project Folder:
   - Place the setup_ineqneed.do and master_ineqneed.do files in the Project Folder
   - Include three subfolders: "*Data*", "*Intermediate-Files*", and "*Output-Files*".
     - Place the "Raw" folder in the "*Data*" folder.
     - Place all intermediate do files in the "*Intermediate-Files*" folder.
     - Place all analysis do files in the "*Output-Files*" folder.

3. Set Up the Working Directory:
    - Input your working directory into the master_ineqneed.do file where prompted.
4. Run setup_ineqneed.do:
    - This will install all dependencies locally, and should be run once.
5. Run master_ineqneed.do:
    - This will create four subfolders: "Temp" and "Analysis" in "Data", "Charity" in "Temp", and "Output" in the Project Folder.
    - This will call each of the intermediate files in turn to create the temporary data, which will appear in a "*Temp*" subfolder, and the analysis data, which will appear in an "*Analysis*" subfolder
    - This will create all the tables and figures of the paper, which will appear in the "*Output*" folder.

## Further Details

Any further questions on how to replicate the paper should be directed to the authors: bellet@ese.eur.nl, eve.cs@mail.huji.ac.il