Datasets and Code for "Why Guarantee Employment? Evidence from a Large Indian Public-Works Program" by Laura V. Zimmermann.

Datasets relevant to the analysis (files were created on Windows 11 operating system):

1. National Sample Survey (NSS) datasets
   The labor market data used in the analysis comes from the NSS, which is a large representative household survey collected by the Indian government. The paper uses data from the employment survey of the 61st round (2004-2005) and the 64th round (2007-2008). The version of the datasets used in the paper is proprietary and not publicly available. Validated unit level data on CD-ROMS was bought from the Indian Ministry of Statistics and Programme Implementation (MOSPI), more specifically the Deputy Director General, Computer Centre, M/O Statistics and PI, East Block No. 10 R.K. Puram, New Delhi-110066 by paying the price along with packaging and postal charges as well as giving an undertaking duly signed in a specified format. For more information on current data access options, please see http://microdata.gov.in/nada43/index.php/catalog/central.

2. Crop statistics data (Stata 17 format)
   The Stata dataset crop_database_raw.dta uses publicly available district-level crop statistics by the Indian Directorate of Economics and Statistics on production and yields (output per hectare) for the same years as the NSS data.

3. Dist_phase_nregs.dta (Stata 17 format)
   This dataset contains the NREGS information for each phase and is merged to the labor market and crop statistics datasets for the main analysis. Details on how the NREGS algorithm was constructed can be found in the paper and in the Online Appendix section "Additional Details on Algorithm Prediction Success".

This folder contains the following do-files (all created using Stata 17 format)

1. Setup do-files for labor market dataset
   To create Stata datasets from the raw proprietary NSS files, carry out the steps below. A series of .do files will read in the raw datasets using dictionary files provided in the dict sub-folders.
   a. The folder NSS dataset creation contains two folders, one for round 61_10 and one for round 64_10. Put the raw data files from MOSPI into the raw_data folders for 61_10 and 64_10. Unzip any files if necessary.
   b. Adjust the file path for "topdir" to the correct directory in the "level" do-files in the round 61 and round 64 do_files folders.
   c. Run extract_all_levels.do. This reads in the datasets and saves them as Stata files in the data_output folder.
   d. Merge NSS files by running merge_61_64_files.do. This creates the dataset that is used for the labor market sample creation and analysis.

2. Analysis sample creation do-files

Numbered 2a through 2e, the analysis_sample_creation.do files create the analysis samples used for the tables in the paper.

3. Crop statistics sample creation
   The do file 3a.crop_sample_creation.do cleans the raw crop statistics data and creates the sample used for the analysis in the paper.

4. Analysis
   The do-file Analysis.do replicates all the tables and figures in the paper.