Women at Work: Evidence from a Randomized Experiment in Urban Djibouti

By Florencia Devoto, Emanuela Galasso, Stefanie Brodmann, and Kathleen Beegle

Overview

The code in this replication package produces analytical output, drawing on from the public version of the Djibouti Safety Net Interlinking Nutrition and Workfare Impact Evaluation Surveys 2014-2016, which has data from baseline, midline and endline surveys of households. In addition, the package includes three auxiliary data files.

Also, the package includes the dofile code applied to the national multi-topic household survey (Enquête Djiboutienne Auprès des Ménages pour les Indicateurs Sociaux, EDAM) 2017 for some comparison labor statistics cited in the paper.

The replicator should take approximately 17 hours to finish.

This package consists of two parts:

(A) Stata analysis for almost all tables and statistics in the text (10 minutes)

(B) R files for the analysis in Tables 10 & 11. (17 hours if you set the run_ml = TRUE and 2 hours if you set run_ml = FALSE)

Data Availability and Provenance Statements

(1) The Djibouti Safety Net Interlinking Nutrition and Workfare Impact Evaluation Surveys 2014-2016 are in the public domain via the World Bank Microdata catalogue:

- https://microdata.worldbank.org/index.php/catalog/2227
- https://microdata.worldbank.org/index.php/catalog/5791
- https://microdata.worldbank.org/index.php/catalog/5790

(2) The two auxiliary data sets included pertain to

i. designdata.dta: This tracks information on sample household's stratification assignment, based on nutritional session assignment, public works sites and treatment/control status, as further described in the paper.

ii. pmt.dta: includes the computed proxy means test which is the score assigned to households on the basis of several traits of the household (related to wealth indicators). The proxy means test and formula were proprietary (in regards to weights assigned different attributes) to the social protection program of the government of Djibouti. The proxy means score indicates household wealth status based on the formula. These can be shared publicly in this reproducibility package; they are not in the World Bank Microdata catalogue.

(3) The EDAM 2017 data are not in the public domain. See access information here: https://microdata.worldbank.org/index.php/catalog/3463

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.

Summary of Availability
• Some data cannot be made publicly available.

(A) Stata Analysis

Dataset list
All datafiles are in Stata (.dta) format and open with Stata version 15 and above. These files are all located in the folder "data_original"

Datafile	*Note*	*Provided*
dji_baseline	Baseline Survey	Yes (link for
dji_midline	Midline Survey	direct
dji_endline	Endline Survey	download)
designdata.dta	Household stratification	Yes
pmt.dta	Proxy means test for sample households	Yes
EDAM2017_Section_01-04_PUF.dta	EDAM 2017 household survey data	No (access can be requested through the link)

Computational requirements
Stata 15 (and above) can be used to run the codes.

Description of programs/code: STATA
• master.do runs the full set of Stata dofiles as follows

• midline_analysis0.do: prepares the data for analysis of midline results; creates data_analysis\midline_analysis.dta

• midline_analysis1.do: produces midline analysis reported in Tables 1-9, A1-A5, and B1-B2; saved in "output\" folder

• endline_analysis0.do: prepares the data for analysis of endline results; creates data_analysis\endline_analysis.dta

• endline_analysis1.do: produces endline analysis reported in Tables 1-9, A1-A5, and B1-B3; saved in "output\" folder

• stats_misc.do: produces various statistics noted in the paper but not in tables, saved in "output\stats_misc.txt"

• figures.do: produces Figures 2-5 in the paper, saved in "output\" folder

• edam2017_labor.do: computes a variety of labor statistics from the EDAM 2017 data which are cited in the paper, saved "output\ edam2017_labor.dta". The original data file are not publicly available.

Instructions to Replicators

• Copy "Reproducibility Package" folder which has the following subfolders (and which comes with files in the first two folders):
- Dofiles [8 dofiles]
- data_original [6 datafiles]
- data_analysis
- output

• In the Dofile folder, open "master.do" and change rows 7-9 as needed to reflect the directory where you saved the "Reproducibility Package"
• Run master.do

(B) R files

This section describes the R code and data to compute the analysis in Tables 10 and 11 of the paper.

Folder's structure and description of files:

- **Codes**

  o **Main scripts :**
    o **main.R**: You can compute all results using only this file. The code will clean the data, run ML algorithms and produce BLP, GATES and CLAN as well as a word summary document.
  o **Utils :**
    o **0_install_packages.R**: sourcing this file will install required packages
    o **0_helper_functions.R**: functions to help the estimation procedure and files storage
    o **1_clean_djibouti_x.R**: functions to clean raw data
    o **2_gml_function.R** : this function runs the whole Generic Machine Learning procedure (ML, regression, final table) and is used only in the main script
    o **2b_methods_functions.R**: functions used in the estimation process
    o **3_visualization_functions.R**: functions used in the (Rmd) code that produces the final output in word document
- **Results**: *it will be automatically generated when executing the codes*
  o DjiboutiMidline (or DijboutiEndline)
    ▪ [Outcome_name1]
      • [treatmentvstreatment]
        o B_and_S: These are the B and S estimations obtained from each ML method, for each replication

- o bestTune: These are the hyperparameters of the best model for each ML method, for each replication
- o sampled: We save here the ids of observations in the auxiliary and main samples
- o settings: We also keep here the tuning parameters of the ML algorithms
- o output_table: This folder contains BLP, GATES and CLAN results that are gathered in one unique excel file
    - ▪ results_[outcome_name1] _ treatment_XvsX.xlsx
  - • [treatment_XvsY]
  - • …

  - ▪ [Outcome_name2]
  - ▪ …
- o Results/midline and endline/ results_djibouti.docx: final words output

Instructions to Replicators

1. Copy the following data generated by stata:

   - o data_analysis\midline_analysis.dta
   - o data_analysis\endline_analysis.dta

   to the \data folder

0. Open the Rproject.

1. Recover the environment by doing renv ::restore() and follow the prompts. The code 0_install_packages.R has all the needed packages, but the renv.lock contains the version the replicators used to reproduce the results.

2. Edit line #100 of main.R code to specify whether you are running:
   - o the midline heterogeneity (name_of_experiment = 'DjiboutiMidline')
   - o or the endline heterogeneity (name_of_experiment = 'DjiboutiEndline')

3. Run main.R code. By running the program, you will:
   - o Obtain the cleaned datasets
   - o Estimate and store all GML estimation results.
   - o Produce the output documents that aggregate all results in Word format in the \results folder
4. You have the option to run the re-run all ML estimations set the parameter in line 171 to TRUE, but this would take around 2 days to run. If set to "FALSE" it takes round 2 hours.