README for Replication files for Long-term and lasting impacts of personal initiative training on entrepreneurial success

2025-01-29

Francisco Campos, Michael Frese, Leonardo Iacovone, Hillary C. Johnson, David McKenzie, Mona Mensmann

AER Insights

Overview

The code in this replication package merges together different survey rounds and then constructs all the Tables and Figures using Stata. Four do files are used for cleaning the data and constructing the key variables to generate a data set for analysis. One main do file then constructs the 2 figures and 3 tables in the paper, along with all tables and the figure in the supplementary appendix. The replicator should expect the code to run for less than 5 minutes.

Data Availability and Provenance Statements

The data used in this paper come from a baseline dataset, four rounds of short-run data collection, and an in-person and phone survey collected after 7 years. These datasets were collected by the research team for the purpose of this impact evaluation, and anonymized versions of all of these datasets are contained in this replication package.

The "questionnaires" folder includes:

- In-person survey questionnaire: LongTermInPersonFollowup_English.pdf and LongTermInPersonFollowup_French.pdf

- Phone survey questionnaire: LongTermPhoneSurvey_English.pdf and LongTermPhoneSurvey_French.pdf

- Short-term questionnaires are in the World Bank data library: here DOI here

To adjust for inflation, we converted all nominal values of financial variables to real September 2021 CFA using the consumer price index published by the Institut National de la Statistique et des Etudes Economiques et Démographiques (INSEED-TOGO), a public establishment attached to the Togolese Ministry in charge of statistics (INSEED-TOGO, 2024).

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 permission to redistribute/publish the data contained within this replication package.

Summary of Data Availability

All data used in this paper are publicly available. The data is included here: World Bank Data Library

The replication package has been deposited in the World Bank's Reproducible Research Repository and is available under a Modified BSD3 license (Campos et al, 2025): here

Datasets Contained in this Replication Package

RAW DATA (contained in the subfolder rawdata)

Short-run raw data files

- BL_raw_an_old.dta - Baseline dataset
- FU1_raw_an_old.dta - Follow-up 1 dataset
- FU2_raw_an_old.dta - Follow-up 2 dataset
- FU3_raw_an_old.dta - Follow-up 3 dataset
- FU4_raw_an_old.dta - Follow-up 4 dataset

Long-run raw data files

- final_dataset_ipv2.dta - Long-run in-person dataset
- final_dataset_psv2.dta - Long-run phone survey dataset
- nonresponsedata.dta - Data collected about non-responders to in-person survey

FINAL DATA (contained in the subfolder finaldata) Togoanalysisdata.dta - This is the constructed master dataset formed from the raw data files above. If you just wish to reproduce the tables and figures, you can use this file without needing to use the raw data. To do this, set the line `Data_Analysis` to 1 in the MasterLongRun.do file (line 80), and set the other sections to 0.

Computational Requirements

Analysis was conducted in Stata 18.0 on a standard Dell Latitude 7440 laptop running Windows 11 with 32GB of installed RAM. Total run time was approximately 2 minutes. Less than 1GB of storage is required for the data and output.

The file `MasterLongRun.do` is used to set file paths and manage dependencies. It uses the package `ietoolkit` to set file paths and manage Stata add-on package versions. It prompts the user to install this package if not present. It also manages access to Stata .ado add-ons, which are contained as follows:

STATA ADO FILES

- These are provided in the `packages` folder and are copies of downloadable Stata add-on packages. These include:
  - ststo - package for collecting results and making tables
  - mat2txt - package for outputting matrices as tables
  - pdslasso - package used for post double-selection lasso estimation (note this in turn relies on `rlasso` and other included components)
  - grc1leg - for combining multiple figures and giving them a common legend
  - winsor2 - for winsorizing data

- cibar - for graphing means by treatment with confidence intervals
- qreg2 - for quantile regressions
- distinct - checks for distinct cases

Controlled Randomness

The upper and lower bounds in Table A5b use a random number generator to break ties when assigning ranks. The file `VariableConstructionTogo1.do` sets the seed for sorting on lines 467 and 468.

Description of Programs and Code

The following Stata do files are used:

- MasterLongRun.do - the master do file that can be used to run all of the other do files
- ConstructShortrun.do - merges together the baseline and four short-run follow-ups
- CleanShortrundata1.do - gets the variables from the short-run dataset needed for this paper
- CleanLongrundata.do - combines the in-person and phone long-run surveys with the earlier rounds of data
- VariableConstructionTogo1.do - constructs the outcomes used in the analysis
- AnalysisTogo1.do - replicates all Figures and Tables in the paper

Instructions for Replicators

1. Create the following file structure:
   - A subfolder called `Data`, with subfolders `Raw`, `Intermediate`, and `Final`. Raw data should be copied into the raw data subfolder, and the file `Togoanalysisdata.dta` should either be copied into the `Final` subfolder or will be created by running all sections of the master do file.
   - A subfolder called `Output`, with a subfolder called `figures`.
   - All do files should be copied into a do-file folder.
2. Edit `MasterLongRun.do` to adjust the default path on lines 30-31.
3. The replicator should decide whether they wish to start with the constructed dataset (`Togoanalysisdata.dta`) or start with the raw data files.
4. Run `MasterLongRun.do`.

References

Campos Francisco, Michael Frese, Leonardo Iacovone, Hilary C Johnson, David McKenzie, and Mona Mensmann (2025). "Reproducibility package for Long-term and lasting impacts of personal initiative training on entrepreneurial success." The World Bank Reproducible Research Repository, DOI.

Institut National de la Statistique et des Etudes Economiques et Démographiques (INSEED-TOGO). "Indice National Harmonisé des Prix à la Consommation," INSEED-TOGO [Accessed July 1, 2024].