```
**************************************************************************
**** README for Replication files for Long-term and lasting impacts of
personal initiative training on entrepreneurial success          ***
*** Francisco Campos, Michael Frese, Leonardo Iacovone, Hillary C.
Johnson, David McKenzie, Mona Mensmann                           ***
*** AER Insights                                                 ***
**************************************************************************
********************************************************
```

March 14, 2025

```
********************************************
*** 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 include:
* 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:
https://microdata.worldbank.org/index.php/catalog/2860/related-materials

To adjust for inflation we converted all nominal values of financial
variables to real September 2021 CFA using the consumer price index
published at 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 raw data are available in the World Bank's Open Data Library (McKenzie, 2025). This library initially contained the short-run data and replication files for Campos et al. (2017), and was updated to add the long-term surveys. Access through the microdata library requires confirming a confidentiality agreement and limits redistribution.

These data, and the supporting do files necessary for replication of the AER Insights paper, have also been deposited in the World Bank's Reproducible Research Repository, and are available under a Modified BSD3 license (Campos et al, 2025): https://doi.org/10.60572/j9n1-g450

*** The following datasets are contained in this replication package, and are stored in Stata version 18.0:

*** 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 Lattitude 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:
    * eststo - package for collecting results and making tables
    * mat2txt - package for outputing 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
* 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, 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
* Edit MasterLongRun.do to adjust the default path on lines 30-31 (replacing the current text "your name" & "Your filepath to the replication folder here") and download ietoolkit if not installed.
* The replicator should decide whether they wish to start with the constructed dataset (Togoanalysisdata.dta), and simply run the Tables and Figures from there, or start with the raw data files and first construct this final dataset. This is controlled in lines 76-80 of the

MasterLongRun.do file. Switch the section to 1 if you wish to run it, or 0 otherwise.
* Run MasterLongRun.do

```
*********************************************************
**** Code Locations for Figures and Tables      *******
*********************************************************
```

*** All Tables and Figures are produced by the file AnalysisTogo1.do. Tables and Figures are clearly indicated, and appear at
*** the following lines, and are stored in the listed output files:
** Figure 1: line 18, output file: Fig1_combined.pdf
** Figure 2: line 106, output file: Figure2Paper.pdf
** Table 1: line 238, output file: Table1Paper.csv
** Table 2: line 408, output file: Table2Paper.csv
** Table 3: line 733, output file: Table3Paper.csv
** Figure A1: line 853, output file: FigureA1Paper.pdf
** Table A1: line 1014, output file: TableA1.xls
** Tables A2a and A2b: line 1094, output file: TableA2a.xls and TableA2b.xls
** Table A3: line 1221, output file: TableA3.xls
** Tables A4a and A4b: line 1263, output file: TableA4a.xls and TableA4b.xls
** Tables A5a and A5b: line 1480, output file: TableA5a.csv and TableA5b.csv
** Table A6: line 1561, output file: TableA6.csv
** Table A7: line 408, output file: TableA7.csv
** Table A8: line 1611, output file: TableA8.csv
** Table A9: line 1770, output file: TableA9.csv
** Tables A10a and A10b: line 1824, output file: TableA10a.csv and TableA10b.csv
** Table A11: line 2034, output file: TableA11.csv
** Table E1: line 2060, output file: TableE1.xls

```
****************************************************
*** 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, https://doi.org/10.60572/j9n1-g450

Campos Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hilary C Johnson, David McKenzie and Mona Mensmann (2017), "Teaching Personal Initiative beats Traditional Training in Boosting Small Business in West Africa", Science, 357(6357): 1287-1290.

Institut National de la Statistique et des Etudes Economiques et Démographiques (INSEED-TOGO) Indice National Harmonisé des Prix à la Consommation,
https://inseed.tg/inflation-prix/ [accessed July 1, 2024]

McKenzie, David (2025) "Firm Surveys for Comparing Personal Initiative Training to Traditional Business Training 2013-2021", World Bank MicroData Library, https://microdata.worldbank.org/index.php/catalog/2860