

Replication for “Training Microentrepreneurs over Zoom: Experimental evidence from Mexico” by Elwyn Davies, Peter Deffebach, Leonardo Iacovone, and David McKenzie

Overview

The code in this folder generates all tables and figures in the paper Training Microentrepreneurs over Zoom: Experimental evidence from Mexico” by Elwyn Davies, Peter Deffebach, Leonardo Iacovone, and David McKenzie. The replicator should expect the code to run for 20 minutes. Downloading the necessary R packages may take additional time.

Data Accessibility Statement

Data sources

There are three data sources used in this paper

1. Original newly collected data from our experiment. Data file: `data/merged_all_surveys.dta`
2. Data from the baseline survey of a 2018 experiment with CREA. See [here](#) for a link to the published article an the AER P&P. Replication data available [here](#). We certify that we have permission to distribute the 2018 CREA data. However for simplicity we only provide variables used in the analysis, Table 1 of the main paper, which shows summary statistics of previous CREA trainees to better compare to our sample. The `.do` files which clean the raw data are provided, but cannot be run. Data file: `data/crea_2018_cleaned.dta`
3. Data from the 1st quarter 2023 National Survey of Occupation and Employment in Mexico (ENOE). Download the data [here](#). We do not have express permission to distribute the data ourselves. For simplicity we only provide variables used in the analysis, Table 1 of the main paper, which shows summary statistics of women entrepreneurs in Mexico to better compare to our sample. The `.do` files which clean the raw data are provided, but cannot be run. Data file: `data/enoe_cleaned.dta`

License for data

The code is licensed under a Creative Commons/CC-BY-NC/CC0 license. See [license.txt](#) for details. The original survey data (`data/merged_all_surveys.dta`) is also licensed uncer a a Creative Commons/CC-BY-NC/CC0 license. For other data used in this project, consult the original sources.

Summary of data

Source	Source	Notes	Provided
<code>data/merged_all_surveys.dta</code>	Original		Yes
<code>data/crea_2018_cleaned.dta</code>	2018 CREA Experiment	Cleaning <code>.do</code> provided	Cleaned, selected variable only
<code>data/enoe_cleaned.dta</code>	INEGI	Cleaning <code>.do</code> provided	Cleaned, selected variable only

Description of folder contents

This folder structure of this document is as follows:

- `ado`: Holds the Stata files
- `data`: Holds the cleaned dataset on women entrepreneurs in Mexico and Guatemala. Also holds the 2018 CREA data and the ENOE data.
- `out`: Will store all the tables and figures produced in the analysis.
- `out_example`: Stores the tables and figures produced by the code when I ran it. These match the tables and figures in the paper. It will be useful for checking results match.
- `src`: Holds all code used in the production of the tables and figures. See instructions below for more details.
- `tex`: Contains the LaTeX documents for compiling the results in the main paper and the pre-analysis plan.
- `training_session_data`: Contains additional data on the content of the training sessions.

Computational requirements

Software requirements

Stata and R are used in the analysis of this project. This project was written using Stata 18 and has not been tested on older versions. I recommend using the latest version of R, 4.3.1, to run the R code. As described below, we use `renv` to keep track of dependencies, and only the exact versions of dependencies are listed. As a consequence, if a package listed does not support older versions of R, there may be issues instantiating the environment.

Package dependencies

The `ado` folder lists all dependencies for the Stata portion of the code. The `renv.lock` file lists all dependencies for the R portion of the code.

Deterministic randomness

Seeds are set at

- Line 9 of `src/heterogeneous_effects/run_generic_m1.R`
- Line 105 of `src/heterogeneous_effects/define_get_churn.R`
- Line 9 of `src/heterogeneous_effects/zaminfluence_mexico.R`

Instructions for replication

Note: I do some post-processing in Julia to clean up significant figures for the tables. For example, I round sales outcomes, which are in the thousands, to the nearest integer. Any differences between the tables produced in this folder and what appears in the working paper should be due to this rounding.

Instructions for replicating tables and figures made in Stata:

1. Open Stata.
2. Set the working directory to the one *this* file lives in. There is no other path to set. We will never change directories during this code.
3. We record all dependencies in Stata using the *ado* folder and by adding this folder to the front of the list of *ado* paths in Stata. This means you should not need to use *ssc* at all. If something pops up that you need to download, please download it. *Note*: you should not need to download anything to use the LaTeX table maker *frmtable*. But if you encounter an error, do not download *frmtable* via *ssc*. The *frmtable* command is exported by *outreg*, so you should download *outreg*.
4. Run *do main.do* and all code that uses Stata (to create the vast majority of tables and figures) will run. This is the only *do*-file you should need to run.

Instructions for replicating tables and figures made in R

Replicating all tables and figures except for Table B.5

The following are instructions to make a tables and figures produced in the R scripts *except* the ZAM Influence table (Table B.5). As discussed below, we use *renv* to keep track of dependencies. But since the *zaminfluence* package is not registered on CRAN, we couldn't automate the downloading process.

1. Open R.
2. Again, set the working directory to the one *this* file lives in. If you open R in this directly, *renv* might bootstrap itself and install and activate.
3. We record all the dependencies in R using the R package *renv*. Please download *renv* using `install.packages("renv")` and then run `renv::activate()` and `renv::restore()`. R should now download all the necessary dependencies.
4. Run `source("main.R")`. This will take a long time to run, as the GenericML machine learning process is computationally heavy.

Replicating Table B.5

To replicate Table B.5, we are going to install the *zaminfluence* package without using *renv*. There is a separate step here because it very hard to install this package.

To install the the *zaminfluence* package, run

```
renv::deactivate()

install.packages("devtools") # Skip if already installed
library(devtools)
devtools::install_github("https://github.com/rngiordan/zaminfluence/",
                        ref="master",
                        subdir="zaminfluence",
                        force=TRUE)

install.packages("haven") # Skip if already installed
install.packages("tidyverse") # Skip if already installed
install.packages("fixest") # Skip if already installed
install.packages("gridExtra") # Skip if already installed
```

To create Table B.5, run

```
source("src/heterogeneous_effects/zaminfluence_mexico.R")
```

Remember to run `renv::activate()` after finishing to re-load the package environment.

List of Tables and Figures

Table or Figure	File name	File created	Table definit
Table 1	out/tables/controls_balance.tex	src/replication/main_paper.do	src/table_programs/define_balanc
Table 2	out/tables/primary_outcomes_lasso_twowave_long.tex	src/replication/main_paper.do	src/table_programs/define_lasso_
Table 3	out/tables/adoption_table_itt.tex	src/replication/main_paper.do	src/table_programs/define_adopti
Table A.1	out/tables/compare_combined_treatment_primary.tex	src/replication/main_paper.do	src/table_programs/define_compe
Table A.2	out/tables/compare_combined_treatment_primary_6m.tex	src/replication/main_paper.do	src/table_programs/define_compe
Table B.1	out/tables/attrition_treatment_effect.tex	src/replication/main_paper.do	src/table_programs/define_attriti
Table B.2	out/tables/attrition_balance_all_periods.tex	src/replication/main_paper.do	src/table_programs/define_attriti
Table B.3	out/tables/primary_outcomes_lasso_twowave_long_matched_sample.tex	src/replication/main_paper.do	src/table_programs/define_lasso_
Table B.4	out/tables/primary_outcomes_bounding.tex	src/replication/main_paper.do	src/table_programs/define_ancov:
Table B.5	out/tables/zamtable.tex	src/heterogeneous_effects/zaminfluence.R	

Table or Figure	File name	File created	Table definition
Table C.1	out/tables/secondary_outcomes_lasso_twowave_long.tex	src/replication/main_paper.do	src/table_programs/define_lasso_
Table C.2	out/tables/adoption_table_marketing_itt.tex	src/replication/main_paper.do	src/table_programs/define_adopti
Table D.1	out/tables/hetero_new.tex	src/heterogeneous_effects/run_generic_ml.R	
Table D.2	out/tables/primary_outcomes_hetero_baseline_end.tex	src/replication/main_paper.do	src/table_programs/define_lasso_
Table D.3	out/tables/primary_outcomes_hetero_baseline_6m.tex	src/replication/main_paper.do	src/table_programs/define_lasso_
Figure 1	out/figures/quantile_reg_combined.png	src/replication/main_paper.do	src/table_programs/define_make_
Figure 2	out/figures/diff_plots_combined.png	src/replication/main_paper.R	src/table_programs/define_plottin
Figure A.1	out/figures/sessions_topic_dist.png	src/training_sessions/analyze_training_sessions.do	
Figure A.2	out/figures/num_sessions_attended.png	src/main_paper.R	
PAP Table A.1	out/tables/primary_pap_end.tex	src/replication/pap.do	src/table_programs/define_compe
PAP Table A.2	out/tables/primary_pap_6m.tex	src/replication/pap.do	src/table_programs/define_compe
PAP Table A.3	out/tables/primary_pap_pooled.tex	src/replication/pap.do	src/table_programs/define_lasso_
PAP Table B.1	out/tables/primary_outcomes_pap_lasso_twowave_long.tex	src/replication/pap.do	src/table_programs/define_lasso_

Instructions for compiling the pdf

Using Latex, compile the documents `paper.tex` and `pap.tex`. I have included `paper_example.tex` and `pap_example.tex` which look in the folder `out_example` for tables and figures. Use these to debug any LaTeX compilation errors.