

READ ME

Overview

This file describes the procedures to replicate the results for the paper “Sexual Harassment and the Survival Rates of Women-owned Firms” by S. Delecourt, S. Papineni, N. Buehren, and A. Alibhai. The code in this replication package constructs the analysis files using Stata.

Replication folder directory

This project is organized into several key folders. “raw” holds the unprocessed, original datasets. The “temporary” folder stores the cleaned and pre-processed data, and “clean” contains the clean datasets ready for analysis. “output” includes all the results for tables and graphs generated (in both the main body of the text and the appendix). The “code” folder contains the do-files used for data processing, analysis, and generating output. The replicator can run all cleaning and analysis files from start to end without interruption using the master do-file `run_project.do`. The replicator needs to define a personal directory path in the master do-file before running the program.

Do-files

run_project.do runs code to set the directory, generates all relevant files for analysis, and runs each of the do-files for cleaning and analysis (further described below).

cleaning.do runs code to clean data. The corrections were provided by the field team in Somalia who collected the data and were applied to the raw datasets.

analysis.do runs the code to construct all the tables in the paper using survey data.

Data

To replicate the results, the do files draw on multiple datasets – and should run from start to end without interruption. To perform the analysis, we rely on the following datasets:

1. **wedp_data_clean.dta** dataset includes all variables created for analysis across the two survey rounds with data collected from the respondent.

All results (main and appendix) will be stored in the folder “output”.

Data Availability Statements

Statement about Rights

The data contained in this package was collected by the authors and the authors have legitimate access to and permission to use the data in this manuscript.

Summary of Availability

The authors have permission to redistribute/ publish the data contained within this replication package. All data has been de-identified of personal information as per IRB requirements.

Computational requirements

Software Requirements

Stata (code was last run with version 19.5).

- winsor2
- ietoolkit
- balancetable
- gr0070 (scheme_plotplainblind)
- sg97_5 (frmtable)

The replicator will need to install all dependencies locally if not already installed.

Memory, Runtime, Storage Requirements

Approximate time needed to reproduce the analyses on a standard laptop: **5 minutes**. The replicator should expect the code to reproduce the main and appendix tables in less than 5 minutes.

List of exhibits

The provided code reproduces all tables and figures in the paper. See following table for the list of exhibits and corresponding do-file with script that runs it.

Exhibit name	Output filename	Script	Output folder
Table 1	prevalence.doc and prevalence.tex	analysis.do	output/tables
Table 2	reg_sh.tex	analysis.do	output/tables
Table 3	raw.doc and raw.tex	analysis.do	output/tables
Table 4	bizclosed.tex	analysis.do	output/tables
Table 5	endline_earnings_reg.tex	analysis.do	output/tables
Figure 1	profits2019.png	analysis.do	output/graphs
Figure 2	profits2014.png	analysis.do	output/graphs
Table A1	descriptives.tex	analysis.do	output/tables
Table A2	reg_attrition.tex	analysis.do	output/tables
Table A3	attrition.doc and attrition.tex	analysis.do	output/tables
Table A4	sector_harassed.tex	analysis.do	output/tables
Table A5	reg_profits_sh.tex	analysis.do	output/tables
Table A6	regprofits.tex	analysis.do	output/tables
Table A7	bizclosed_marriedchildren.tex	analysis.do	output/tables
Table A8	attrited_as_closed.tex	analysis.do	output/tables
Table A9	matching.tex	analysis.do	output/tables
Table A10	doublerobust.tex	analysis.do	output/tables

Notes: The table above provides a one-to-one mapping between manuscript exhibits, output files, and the scripts used to generate them. Tables 1, 3 and A3 are manually created first as doc files and then converted to tex files. The rest of the exhibits are automatically created using Stata.