

README for “Wage Subsidies to Promote Female Hiring: Evidence from Pakistan”

Overview

The code in this replication package cleans data sources used in the paper “Wage Subsidies to Promote Female Hiring: Evidence from Pakistan” using Stata and R to generate all results in the paper and appendices. One master dofile runs all of the Stata code to generate the data and produce the results.

Data Availability

This paper uses firm, and employee data collected from firms in Pakistan by the authors in collaboration with Rozee.pk. In addition, the paper uses admin data from Rozee.pk job portal.

The data for this paper were collected over four rounds. The first round was a baseline of firms in the study. The second round is the intervention with only firms in the treatment group. The third round is the hiring follow-up survey. The fourth round is the endline survey.

Statement About Rights

We certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript

Dataset List

Input datasets are provided in the directory `dta -> raw`:

1. `rozee__survey__raw`
2. `portal__outofsample__raw`
3. industry codes
4. edu codes
5. `jobs__ads`

Computational Requirements

Software Requirements

- Stata (code was last run with version 19 with backward compatibility to version 17) with the following add-on files. The program `master.do` will install all dependencies locally when the switch/global macro `firsttime` is set to 1.
- R version 4.5.1 was used to run the R script.

Memory and Runtime Requirements

The full sequence of data preparation and analysis will take approximately under 10 minutes to run. The program installation takes also takes under 10 minutes.

Description of Programs/Code

Program	Description
do/master.do	Runs all do files listed below and installs all necessary Stata add-on files.
do/1. data prep.do	Cleans anonymized survey data and constructs all relevant variables used for analysis. The following dataset for analysis is produced: rozee_survey.dta
do/2. portal data prep.do	Cleans anonymized portal data and constructs all relevant variables used for long term job posting behavior. The following dataset for analysis is produced: portal.dta
do/4. outofsample portal data prep.do	Cleans anonymized portal data and constructs all relevant variables used for out of sample data comparison. The following dataset for analysis is produced: portal_outofsample.dta
do/Table1_sumstats.do	Summary statistics of baseline survey data
do/Table2_ITThiredwomen.do	Treatment effects on main outcome of interest
do/Table3_ITThiringprocess.do	Treatment effects on other outcomes
do/Table4_ITTportal.do	Long-term treatment effects on job posting behavior
do/TableOA1_balance.do	Balance tables comparing T and C groups
do/TableOA2_samplecomparison.do	Comparison of survey sample with other no-gender preference postings
do/TableOA3_comparison_acrossgenderpref.do	Comparison of job postings across gender preferences
do/TableOA4_attritionHFS.do	Correlates of attrition during Hiring Follow Up Survey
do/TableOA5_ToT.do	ToT estimates on main outcome
do/TableOA6_attritionEL.do	Correlates of attrition during Endline
do/TableOA7_ITTfirm_el.do	Treatment effects on firm outcomes at Endline
do/TableOA8_ITThm_el.do	Treatment effects on manager outcomes at Endline
do/TableOA9_portalsumstats.do	Summary statistics of job postings on portal
do/TableOA10_ToTportal.do	Long-term ToT effects on job posting behavior
do/Figure 1_kernelIV.do	Distribution of treatment effects using Kernel IV estimates
do/Figure 2_HTE.do	Heterogeneous treatment effects by job and firm characteristics
do/Figure OA2_subsidybid_kdens.do	Distribution of incentivized and un-incentivized subsidy bids
do/Figure OA3_HTEmanager.do	Heterogeneous treatment effects by other manager characteristics
do/Figure OA6_dynamic.do	Dynamic treatment effects on long term job posting behavior
do/Figure OA7_postings.do	Distribution of job postings on Rozee

Note: The only R script is `do/3.htes_fig0A4-5.R` to implement the Generalized Random Forest (GRF) algorithm to generate the distribution of firm average treatment effects, and plot the distribution over selected covariates.

Instructions to Replicators

New users should follow these steps to run the package successfully:

Update the following files with your directory paths: master.do

Ensure all required software and dependencies are installed as listed in the Requirements section by setting the swithc/global firsttime to 1 (line 46 of master.dofile).

Run the master.do file and the single R script separately.

List of Tables and Figures

The published replication package allows users to replicate all tables and figures, as detailed below.

Figure/Table	Do File	Output File
Table 1	do/Table1_sumstats	output/tex/sumstats.tex
Table 2	do/Table2_ITThiredwomen.do	output/tex/itt_hiringwomen.tex
Table 3	do/Table3_ITThiringprocess	output/tex/itt_hiringwomenother.tex
Table 4	do/Table4_ITTportal.do	output/tex/admin_itt.tex
Table A1	do/TableOA1_balance.do	output/tex/Balance.tex
Table A2	do/TableOA2_samplecomparison.do	output/tex/sample_comparison_no pref.tex
Table A3	do/TableOA3_comparison_acrossgenderpref.do	output/tex/across gender pref.tex
Table A4	do/TableOA4_attritionHFS.do	output/tex/attrition_HFS.tex
Table A5	do/TableOA5_ToT.do	output/tex/tot.tex
Table A6	do/TableOA6_attritionEL.do	output/tex/attrition_el.tex
Table A7	do/TableOA7_ITTfirm_el.do	output/tex/itt_main_el.tex
Table A8	do/TableOA8_ITThm_el.do	output/tex/itt_hmoutcomes_el.tex
Table B1	do/TableOA9_portalsumstats.do	output/tex/portal_sumstats.tex
Table B2	do/TableSA10_ToTportal.do	output/tex/admin_tot.tex
Figure 1	do/Figure 1_kernelIV.do	output/figures/HTE_KernelIV.pdf
Figure 2	do/Figure 2_HTE.do	output/figures/het_firmjob.png
Figure A2	do/Figure OA2_subsidybid_kdens.do	output/figures/subsidy_density.png
Figure A3	do/Figure OA3_HTEmanager.do	output/figures/het_manager.png
Figure A6	do/Figure OA6_dynamic.do	output/figures/dynamiceffects_binned.png
Figure A7	do/Figure OA7_postings.do	output/figures/postings.png

Note: The only R script is `do/3.htes_fig0A4-5.R` to generate Figures A4 and A5.