

README for “Job Search and Hiring with Limited Information about Workseekers’ Skills”

Overview

The code in this replication package cleans three data sources used in the paper (Carranza et al., 2022; Statistics South Africa, 2016, 2017) using Stata and then generates all results in the paper and appendices. One master file runs all of the code to generate the data and produce the results, other than 2 tables that are manually constructed using external data. The replicator should expect the full code to run in less than 15 minutes.

Data Availability and Provenance Statements

This paper uses data on the labor market in South Africa, collected from five sources. The first source is a workseeker-level survey dataset collected by the authors, which records participants’ employment, job search, and beliefs relevant to their employment, job search, and skills. The second source is an administrative dataset obtained from the Harambee Youth Employment Accelerator. This records participants’ scores on skill assessments. We merge these sources using participants’ national identity numbers and names. The third source is a correspondence study dataset collected by the authors, which records job application characteristics and outcomes. The fourth source is a firm-level dataset containing survey and choice experiment data. These data are provided as part of the replication package (Carranza et al., 2022).

One table in Appendix D uses publicly-available data from Statistics South Africa (Statistics South Africa, 2016, 2017). These data can be accessed from the URLs included in the bibliographic entries and are not included in the replication package.

Statement about Rights

The authors of the manuscript have legitimate access to and permission to use the data used in this manuscript. See the accompanying letter granting permission from the external data provider, the Harambee Youth Employment Accelerator. Users interested in accessing the original data that the Harambee Youth Employment Accelerator shared with the authors can contact the organization at <https://harambee.co.za/>, although the organization has not indicated that they will respond to any such requests.

License for Data

Data provided by Harambee remain the property of the provider and can be used only to replicate results reported in the paper and appendices. Survey data generated by the authors are licensed under a Creative Commons/CC-BY-NC/CC0 license.

Summary of Availability

Some data **cannot be made** publicly available. The personally identifiable information used to merge different datasets are not included in this replication package.

Dataset list

File `audit/data/audit_data.dta` contains the job application-level data from the correspondence study conducted by the authors. This dataset is provided in the replication package.

File `firm/data/raw/firm_data_trimmed.dta` contains the firm-level data collected by the authors. This dataset is provided in the replication package.

File `workseeker/data/raw/wsdata_signalling_interim.dta` contains the merged data workseeker-level data from Harambee and the authors’ own surveys. This dataset is provided in the replication package. The merging process uses personally identifiable information so that stage of the data creation process is excluded from replication package.

The Quarterly Labour Force Survey data are in `workseeker/data_qlfs/lmdsa-2016-1.0-stata11.dta` and `workseeker/data_qlfs/lmdsa-2017-1.0-stata11.dta`.

Computational requirements

Software Requirements

- Stata (code was last run with version 17) with add-on files:
 - `egenmore` (as of 2022/09/21)
 - `estout` (as of 2022/09/21)
 - `esttab` (as of 2022/09/21)
 - `ftools` (as of 2022/09/21)
 - `reghdfe` (as of 2022/09/21)
 - `qreg2` (as of 2022/09/21)
 - `winsor` (as of 2022/09/21)
 - `winsor2` (as of 2022/09/21)
- The program `0_master.do` will install all dependencies locally when the switch `/ global macro installado is set to 1.`

Memory and Runtime Requirements

The full sequence of analysis will take approximately 15 minutes to run on a standard 2022 desktop machine. The code was last run on a **4-core Intel-based desktop with Linux Ubuntu version 18.04**.

Description of programs/code

- Program `workseeker/scripts/0_master.do` runs all programs listed below and installs all necessary Stata add-on files.
- Program `workseeker/scripts/1_decisions.do` sets the values of parameters used by subsequent programs.
- Program `workseeker/scripts/2a_create_anonymous_id.do` creates anonymous identifiers used throughout the rest of the analysis.
- Program `workseeker/scripts/2b_deidentify_datasets.do` removes identifiable information and merges datasets from different sources together.
- Program `workseeker/scripts/1_decisions.do` sets the values of parameters used by subsequent programs.
- Program `workseeker/scripts/3a_clean_baseline.do` prepares data from the workseeker baseline survey and administrative data from Harambee.
- Program `workseeker/scripts/3b_clean_sms.do` prepares data from the first workseeker-level follow-up survey.
- Program `workseeker/scripts/3c_clean_endline.do` prepares data from the second workseeker-level follow-up survey.
- Program `workseeker/scripts/3d_create_indices.do` constructs indices using data from multiple waves of workseeker-level survey data.
- Program `workseeker/scripts/3e_trim_dataset.do` drops variables used in data construction but not needed for paper analysis.
- Program `workseeker/scripts/3f_missingvalues.do` addresses missing data.
- Program `workseeker/scripts/4a_paper.do` generates most tables and figures in the paper.
- Program `workseeker/scripts/4b_appendices.do` generates most tables and figures in the appendix.
- Program `workseeker/scripts/4c_external_comparison.do` combines the authors' workseeker-level survey data with external data from the Quarterly Labour Force Survey to generate one table in the appendix.
- Programs in `workseeker/scripts/aux/` are sub-routines called by another program from the list above.
- Program `audit/scripts/audit_analysis.do` prepares the job application-level data and runs all analysis on these data.

- Program `firm/scripts/firms_clean_data.do` prepares the firm-level data for analysis.
- Program `firm/scripts/firms_analysis.do` runs all analysis on the firm-level data.

License for Code

The code is licensed under a MIT/BSD/GPL/Creative Commons license.

Instructions to Replicators

Run `0_master.do`, which calls all other necessary programs except `2a_create_anonymous_id.do` and `2b_deidentify_datasets_id.do`, which required datasets containing personally identifiable information that are not included in this replication package.

List of tables and programs

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

Figure/Table	Program	Output file	Note
Table 1	<code>4a_paper.do</code>	<code>workseeker/output/employ.tex</code>	
Table 2	<code>4a_paper.do</code>	<code>workseeker/output/employ_decomposition.tex</code>	
Table 3	<code>4a_paper.do</code>	<code>workseeker/output/mechanisms.tex</code>	(1)
Table 4	<code>audit_analysis.do</code>	<code>audit/output/audit_effects.tex</code>	
Figure 1	NA	NA	(2)
Figure 2	<code>4a_paper.do</code>	<code>workseeker/output/earnings_cdfs.pdf</code> <code>workseeker/output/earnings_qtes.pdf</code>	
Figure 3	NA	NA	(2)
Table A1	NA	NA	(3)
Table A2	<code>4b_appendices.do</code>	<code>workseeker/output/skillcor.tex</code>	
Table A3	<code>4b_appendices.do</code>	<code>workseeker/output/skillterciles.tex</code>	
Figure C1	<code>4a_paper.do</code>	<code>workseeker/output/earnings_pdfs.pdf</code>	
Table D1	<code>4b_appendices.do</code>	<code>workseeker/output/summstats_bl.tex</code>	
Table D2	<code>4b_appendices.do</code>	<code>workseeker/output/summstats_el.tex</code>	
Table D3	<code>4b_appendices.do</code>	<code>workseeker/output/baltests.tex</code>	
Table D4	<code>4b_appendices.do</code>	<code>workseeker/output/summstats_compare.tex</code>	
Table D5	NA	NA	(4)
Table D6	<code>4b_appendices.do</code>	<code>workseeker/output/attrit_byt_withplc.tex</code>	
Table D7	<code>4b_appendices.do</code>	<code>workseeker/output/attrit_byx.tex</code>	
Table D8	<code>4b_appendices.do</code>	<code>workseeker/output/robustness_nocovs.tex</code>	
Table D9	<code>4b_appendices.do</code>	<code>workseeker/output/robustness_extracovs.tex</code>	
Table D10	<code>4b_appendices.do</code>	<code>workseeker/output/beliefs.tex</code>	
Table D11	<code>4b_appendices.do</code>	<code>workseeker/output/employ_het.tex</code>	(5)
Table D12	<code>4b_appendices.do</code>	<code>workseeker/output/prespec.tex</code>	(6)
Table D13	<code>4b_appendices.do</code>	<code>workseeker/output/employ_decomposition_pubvspvt.tex</code>	
Table D14	<code>4b_appendices.do</code>	<code>workseeker/output/employ_selection.tex</code>	
Figure D1	<code>4b_appendices.do</code>	<code>workseeker/output/selfesteem_cdfs.pdf</code>	
Figure D2	<code>4b_appendices.do</code>	<code>workseeker/output/earnings_allgroups_qtes.pdf</code>	
Table E1	<code>audit_analysis.do</code>	<code>audit/output/audit_workseekers.tex</code>	
Table E2	<code>audit_analysis.do</code>	<code>audit/output/audit_applications.tex</code>	
Table F1	<code>4b_appendices.do</code>	<code>workseeker/output/employ_plc.tex</code>	
Figure F1	NA	NA	(2)
Table G1	<code>firms_analysis.do</code>	<code>firm/output/firm_summstats.tex</code>	
Table G2	<code>firms_analysis.do</code>	<code>firm/output/firm_skillrank.tex</code>	
Figure G1	NA	NA	(2)
Figure G1	NA	NA	(2)

Table notes:

1. Includes the intermediate inputs mechanismsZ.tex for $Z = 1, 2, 3$.
2. Shows intervention material, not analytical outputs.
3. Shows data obtained directly from Harambee staff without any subsequent analysis that could be replicated.
4. Shows results from other papers, cited in the cited.
5. Includes the intermediate results employ_het_byskillindex.tex and employ_het_bysignal.tex.
6. Includes the intermediate results prespecZ.tex for $Z = 1, 2, 3, 4, 5$. Requires some manual reformatting to match the version displayed in the paper.

References

- Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin**, “Data and Replication Code for: Job Search and Hiring with Limited Information about Workseekers’ Skills,” 2022. *American Economic Association* [publisher], Inter-university Consortium for Political and Social Research [distributor]. <http://doi.org/10.3886/E172902V1>.
- Statistics South Africa**, “Labour Market Dynamics in South Africa 2018 [dataset]. Version 1,” 2016. Pretoria: Statistics South Africa [producer], 2017. Cape Town: DataFirst [distributor], 2018. DOI: <https://doi.org/10.25828/kkzz-mr59>.
- , “Labour Market Dynamics in South Africa 2018 [dataset]. Version 1,” 2017. Pretoria: Statistics South Africa [producer], 2018. Cape Town: DataFirst [distributor], 2019. DOI: <https://doi.org/10.25828/wyv8-6k31>.