README

Manuscript title: "Why Don't Jobseekers Search More? Barriers and Returns to Search on a Job Matching Platform"
Authors: Kate Vyborny ⓡ Robert Garlick ⓡ Nivedhitha Subramanian ⓡ Erica Field
Date: 15 March 2025

Overview:

This replication package contains the code and data used to generate 29 tables and 13 figures in the paper, including those in the online appendix. The code, written in Stata, constructs the analysis data files by integrating multiple data sources. These sources include platform data, and survey data collected from households, jobseekers, and firms.

The file ***01_masterdofile.do*** provides the structure for organizing the replication process. The code is divided into three main subsidiary do-files which perform the following tasks:

1. Setup
2. Code to generate datasets
3. Code to replicate all tables and figures

Code files are organized into two main sub folders: ***cleaning*** and ***analysis***. The ***cleaning*** folder contains 27 do-files to prepare datasets. The ***analysis*** folder contains 44 do-files to generate tables and figures in the paper, including those in the online appendix.

The replicator should expect the code to run for approximately 10-15 hours.

Data availability and provenance statements:

The data used in this analysis originates from a combination of administrative sources and surveys collected directly from households, firms, and jobseekers connected to a job search platform based in Lahore, Pakistan. All collected data has been de-identified to ensure confidentiality. This research received ethics approval from Duke University (#2019-0067).

Additionally, external data from Pakistan's Labor Force Survey (LFS), which is publicly available and maintained by government statistical authorities, was incorporated into the analysis.

The details of the data sources used as part of analysis are mentioned below:

- Jobseeker data sources:
    - Household data:
        - Household data comes from the initial household sign-up, which includes a basic roster, elicitation of interest in the Job Talash platform, and basic neighborhood characteristics. Only those who expressed interest in the platform in that survey were approached for enrollment in the study. Please refer to the *"Household Survey"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
        - Both semi-cleaned and cleaned data have been provided.
    - Jobseeker baseline data:
        - Baseline data comes from the signup instrument targeting jobseekers who expressed an interest in signing up with the platform as part of the household survey. Please refer to the *"Subscriber Signup Surveys"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
        - Both semi-cleaned and cleaned data have been provided.
    - Jobseeker post signup (or updates) data:
        - Jobseekers as part of the platform could update their CV information post signup. This data was collected through various phone surveys where the jobseeker provided their most up to date CV information. Please refer to the *"Jobseeker Updates Surveys"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
        - Both semi-cleaned and cleaned data have been provided.
    - Jobseeker application data:
        - The application data come from the screener survey, where job seekers indicate which jobs they want to apply for after hearing details about the list of jobs they were matched with in a given round. Please refer to the *"Jobseeker Screener/Application Surveys"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.

        - Both semi-cleaned and cleaned data have been provided.
    - Jobseeker survey response data on both on-platform and off-platform search activity:
        - These data come from the jobseeker-facing survey that collects information on and off platform job search, applications and interviews. Also includes survey questions capturing beliefs about the quality of job matches sent. Please refer to the "*Jobseeker Survey about On and Off-Platform Search"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
        - Both semi-cleaned and cleaned data have been provided.
- Firm data:
    - Firm baseline data & detailed data on the vacancies listed by firms on the platform:
        - Baseline firm data come from the signup survey, where firms were visited in person, and their decision to sign up was recorded. The survey also collected detailed information on each firm's hiring process. Additionally, information on the vacancies firms wanted to list on the platform was recorded either at the time of signup or afterward through the same survey. Please refer to the *"Firm Baseline and Ad-listing Survey"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
        - Both semi-cleaned and cleaned data have been provided.
    - Firms' feedback on the CVs received through the platform:
        - After jobseekers inform the platform of their application decisions, the platform forwards their applications to firms. Data on shortlisted CVs, candidates invited for interviews, and off-platform applications for these jobs are collected through the firm feedback survey. Please refer to the *"Firm Feedback Survey"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
            - Both semi-cleaned and cleaned data have been provided.
        - Additionally, as part of the firm midline survey, we also collected data on these jobs from firms. . Please refer to the *"Firm Midline Survey"* questionnaire which can be found in the *"\Additional Documentation\Questionnaire"* folder.
- Administrative data:
    - Jobseeker-vacancy pair (or "match") data:

        - After signing up, jobseekers are matched to each listed vacancy using a simple algorithm: the jobseeker must have at least the required years of education and experience, match any gender requirement, and have indicated interest in this occupation. We refer to each jobseeker-vacancy pair, for which the respondent qualifies and has indicated interest in the occupation, as a match. This data also contains the main treatment variable of interest, the "phone call treatment."
        - Both semi-cleaned and cleaned data have been provided.
    - Hypothetical jobseeker-vacancy pair (or "match") data:
        - We also construct matches that the jobseeker would have received if they had retained their original job preferences, education and years of experience, rather than updating this information at some time after their initial registration.
        - Cleaned data has been provided.
- Analysis datasets:
    - We combine data from multiple sources, as mentioned above, and generate a dataset at the match level which is mostly used as part of analysis. We further collapse this dataset at alternative units of observation, such as the job level, jobseeker level, and jobseeker*round level.
        - All datasets have been provided.
- Miscellaneous datasets:
    - These are one-off datasets that are either used in analysis do-files or as part of the cleaning do-files to obtain variables such as dataset containing predicted probabilities of getting interviews, the calling order in which an enumerator might attempt to contact jobseekers for screener surveys, the scoring of CVs done for a few jobs to which jobseekers applied, etc.
        - All cleaned datasets have been provided.

Summary of availability

Data Source/Survey name	**Semi Cleaned/Raw data**		**Cleaned/output data**	
	Name	**Provided**	**Name**	**Provided**
Jobseeker side data				
Household survey	HHlisting_long_anony.dta	Yes	hhdata_prep_se.dta	Yes

Subscriber signup survey	JT_JA_signup_harmonized.dta Jaebs_prep.dta	Yes	jt_ja_harmonized_prep_se.dta	Yes
Jobseeker update surveys	jt_ja_harmonized_prep_se.dta; JTupdates_JA_harmonized.dta	Yes	jt_ja_harmonized_updates_prep_se .dta	Yes
	CV_updates_Highed_attempts.dta; CV_updates_Lowed_attempts.dta; CV_updates_Lowed_FS_attempts; CV_updates_uniq_ID_list_he_2022_0 8_11_b5.csv cv_updates_highed_data_25 Jun 2021.dta; cv_updates_highed_data_2020_11_30 .dta		CV_updates_highed_data.dta; jsfollowup_prep_se.dta; Updates_survey_var_JT.dta	Yes
Jobseeker surveys about on and off platform search	20percent_prep.dta; 20percent_bonuscalls_prep.dta 20percent_prep_attempt.dta JT_follow_up_survey_rand.dta; JT_rand_preload_10percent_hybridm od.dta; JT_rand_preload_10percent_hybridm od.dta 20percent_prep_01 Nov 2022.dta	Yes Yes	jt_ja_harmonized_JSF_CV_update s.dta	Yes
Jobseeker application/screener surveys	JT_JA_screener_appended.dta	Yes	pickup_prep_se.dta; screener_prep_se.dta; cv_tracking_batches.dta	Yes
Firms data				
Firm baseline survey & ad-listing survey	FirmSurvey_nodup_anon_se.dta	Yes	firms_prep_se.dta	Yes
	FS_adlisting_anon_se.dta	Yes	adlisting_prep_se.dta	Yes
Firm feedback survey	FS_feedback_cleaned_anon_se.dta; FS_MidlineSurvey_feedback_cleaned. dta FS_del_cleaned_unique_jobid.dta FS_del_cleaned_unique_memidjobid_ anon.dta FS_interview_med.dta FS_non_compliance_int_med.dta	Yes	feedback_joblevel_prep_allvars_se .dta; feedback_joblevel_prep_se.dta; feedback matchlevel_prep_allvars se.dta; Feedback_matchlevel_prep_se.dta intmed_prep.dta	Yes
Administrative data				
Matches data	matches_JT_JA_consolidated.dta	Yes	matches_prep_se.dta match_level_oct2022.dta	Yes

Hypothetical matches data			matches_data_sensitivity_cons.dta	Yes
Analysis data				
Intermediary datasets	hhdata_prep_se.dta; jt_ja_harmonized_prep.dta; jt_ja_harmonized_prep_se.dta; matches_prep_se.dta; pickup_prep_se.dta; screener_prep_se.dta; FS_del_cleaned_unique_memidjobid_anon.dta; adlisting_prep_se.dta; firms_prep_se.dta; feedback_matchlevel_prep_se.dta; feedback_joblevel_prep_se.dta; JT_noid_duprem.dta; cv_tracking_batches.dta	Yes	Match_level_analysis.dta	Yes
Jobseeker x vacancy (or matches) level	match_level_analysis.dta	Yes	individual_round_level_analysis.dta; individual_level_analysis.dta; job_level_analysis_se.dta	Yes
Miscellaneous data				
Data with predicted probabilities of interviews	match_level_analysis.dta firms_prep_se.dta individual_level_analysis.dta	Yes	se_matchlevel_latent_interview_probability_variation_v1.dta se_matchlevel_latent_interview_probability_prep_01_replication_variaton_v1.dta	Yes
Scoring of CVs			FR_jobs_cv_scoring.dta; CC_jobs_cv_scoring.dta	Yes
Call order in which the enumerator call JS for screener surveys			post_sep2020_call_order.dta; pre_sep2020_call_order_recovered	Yes
External data				
LFS data			LFS 2018-19.dta	Yes

Details on each data source

All relevant datasets used in the analysis presented in this paper are available in the directory: *GEA_Replication_Package/Data*. Additionally, CSV versions of these datasets can be accessed in the subdirectory: *GEA_Replication_Package/Data/Raw CSV*.

Public use data sourced

Data on the Pakistan Labor Force Survey 2018-2019 were downloaded from the Pakistan Bureau of Statistics (2018). Data can be downloaded from https://www.pbs.gov.pk/content/lfs-2018-2019-microdata. The data are in the public domain.

Software requirements

- The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.
- The code is based on Stata (version 17 and 18). All necessary packages can be installed from the preparation do-file.
    - The script "*se_sub_setup.do"* within the *01_masterdofile.do* will install all dependencies locally, and should be run once.
- The output either consists of tables and figures. All tables are saved as .tex files, therefore a program that can read .tex files is required.
- The code uses backslashes in file paths, which will run on Windows machines but not on Linux or maxOS machines.

Memory, runtime, storage requirements

- Approximate time needed to reproduce the analysis on a standard desktop machine: 10-15 hours.
- Approximate storage space needed: 16 GB

Description of programs/code

Program in *01_masterdofile.do* is designed to reproduce all tables and graphs included in the paper following the given sequence:

- **Step 1**: Install the relevant packages and save relevant macros/globals using *".\Code\Cleaning se_sub_setup.do"*

- This do-file loads the user-written packages, and includes globals for variables and samples, used in the cleaning and analysis do-files.

- **Step 2**: Prep data using
  ".\Code\Cleaining/summary_se/WMP_JOBMATCHING_ANALYSIS_PREP.do" :
  - This do-file consists of 3 subsidiary do-files:
    - *."\Code\Cleaning/summary_se/ se_sub_cleaning_main_parta.do"*
      - This do-file cleans individual datasets
    - *".\Code\Cleaning/summary_se/se_sub_cleaning_main_partb.do"*
      - This do-file generates analysis datasets (from merging individual datasets)
    - *".\Code\Cleaning/summary_se/se_sub_cleaning_main_partc.do"*
      - This do-file generates predictive probabilities of interview
    - Please refer to the table below to understand the sequence of the do-file in detail

Flow of *WMP_JOBMATCHING_ANALYSIS_PREP.do*			
Part A: Cleaning of individual datasets (*se_sub_cleaning_main_parta.do*)			
No	**Dofile name**	**Purpose**	**Output data name**
1.	se_sub_cleaning_01_hhdata_prep.do	Clean household listing dataset	hhdata_prep_se.dta
2.	se_sub_cleaning_02_js_signup_prep.do	Clean jobseeker baseline data	jt_ja_harmonized_prep_se.dta
3.	se_sub_cleaning_03_js_updates_prep.do	Clean jobseeker updates data	jt_ja_harmonized_updates_prep_se.dta
4.	se_sub_cleaning_04_js_surveyvars_prep.do	Clean jobseeker survey data regarding on and off-platform search and work	jt_ja_harmonized_JSF_CV_updates.dta
5.	se_sub_cleaning_05_firmbl_adlisting_prep.do	Clean firm baseline and adlisting data	firms_prep_se.dta adlisting_prep_se.dta
6.	se_sub_cleaning_06_firmfeedback_prep.do	Clean firm feedback data on CVs received	feedback matchlevel_prep_allvars se.dta feedback_matchlevel_prep_se.dta feedback_joblevel_prep_allvars

			se.dta feedback_joblevel_prep_se.dta
7.	se_sub_cleaning_07_matches_prep.do	Clean match data	matches_prep_se.dta
8.	se_sub_cleaning_08_screenersurvey_prep.do	Clean screener survey data	screener_prep_se.dta pickup_prep_se.dta

Part B: Generation of analysis datasets (generated from merging individual datasets created in Part A) (*se_sub_cleaning_main_partb.do*)

No	Dofile name	Purpose	Output data name
1.	se_sub_match_level_analysis.do	Merges intermediary datasets to generate the match level dataset	match_level_analysis.dta
2.	se_sub_individual_level_analysis.do	Collapses the match level dataset to generate dataset at individual level	individual_level_analysis.dta
3.	se_sub_individual_round_level_analysis.do	Collapses the match level dataset to generate dataset at individual x round level	individual_round_level_analysis.dta
4.	se_sub_job_level_analysis.do	Collapses the match level dataset to generate dataset at job level	job_level_analysis_se.dta

Part C: Generation of predicted interview probability (using analysis datasets generated in part B) (*se_sub_cleaning_main_partc.do*)

No	Dofile name	Purpose	Output data name
1	se_sub_cleaning_main_partc.do	Use match level dataset to predict latent interview probabilities using lasso	se_matchlevel_latent_interview_probability_variation_v1.dta

- **Step 3:** Replicate tables and figures in the paper using ".\Code\Analysis\Main*main_se_analysis_paper.do"*
    - Scripts in ".*Code\Analysis\Subsidiary"* generate all the tables and figures in the paper. Each program called from this location identifies the table or figure it creates (e.g., table_3_se_sub_matchlevel_main_result.do). Code

for all tables in the subsidiary folder are referenced as part of the *main_se_analysis_paper.do*

- Please refer to the below table for the flow of the *main_se_analysis_paper.do*

List of tables and programs

The provided code reproduces:

1) All numbers provided in text in the paper
2) All tables and figures in the paper
3) Selected tables and figures in the paper, as explained and justified below

Flow of *main_se_analysis_paper.do*		
Figure/Table #	**Program**	**Output file**
Table 1	table_1_2_se_sub_summary_stats_js_jobs.do	table1_summstats_respond_v2.tex
Table 2	table_1_2_se_sub_summary_stats_js_jobs.do	table1_summstats_vacancy.tex
Table 3	table_3_se_sub_matchlevel_main_result.do	se_matchlevel_paper_table2_unweighted.tex
Figure 1	figure_1_se_sub_app_decision_by_treat.do	combined_graph_simple_v1.png
Appendix A		
Table A.1	table_a1_panel_1_se_sub_lfs_person_table.do table_a1_panel_2_3_se_sub_lfs_person_table.do	summstats_compare.tex summstats_compare_female.tex summstats_compare_male.tex
Figure A.1	figure_a1_se_sub_lfs_jtvacancies_salaries_overlap.do	se_sub_lfsjt_salary_lhr.png
Appendix B		
Figure B.1	figure_b2_late_bounds_with_defiers.do	se_person_level_itt_ntimes_applied_v1.png se_person_level_itt_ntimes_inter_v1.png
Table B.1	table_b1_se_sub_hte_matchlevel_3panels.do	se_matchlevel_hte_iv_panel_Updated.tex
Table B.2	table_b2_se_sub_match_main_2panel_interacted_proxies_V.do	se_matchlevel_panela_interacted_proxies_V.tex
Table B.3	table_b3_se_sub_match_sensitiveanalysis_9columns.do	se_matchlevel_sensative_hh_9columns_apply.tex

		se_matchlevel_sensative_hh_9columns_interview.tex paper/se_matchlevel_sensative_hh_9columns_iv.tex
Figure B.2	figure_b2_late_bounds_with_defiers	se_match_ivbounds.pdf
Table B.4	table_b4_se_sub_person_non_app_measures.do	se_personlevel_iv_quality_fullsample_wave.tex
Table B.5	table_b5_se_sub_match_changes_js_profile_pref.do	se_matchlevel_sensative_hypo.tex se_matchlevel_int_sensative_hypo.tex se_matchlevel_int_iv_sensative_hypo.tex
Table B.6	table_b6_se_sub_alt_test_constant_returns.do	se_matchlevel_paper_decomposing_v1
Table B.7	table_b7_se_sub_person_search_work_te.do	se_person_instruments_non_res_paper_search_work.tex
Table B.8	table_b8_se_sub_person_intensive_search_te.do	se_personlevel_intensive_search_un_heck.tex
Table B.9	table_b9_se_sub_person_survey_engage.do	se_person_survey_engage_paper.tex
Table B.10	table_b10_se_sub_person_rand_insturments_answer_survey_modules.do	se_person_instruments_non_res_paper_firststage.tex
Appendix C		
Table C.1	table_c1_se_sub_matchlevel_main_result_other_treatments.do	se_matchlevel_paper_table2_unweighted_cross_1
Figure C.1	figure_c1_se_sub_app_decision_by_treat.do	model_cost_axis_labels.png
Figure C.2	figure_c2_3_se_sub_te_app_quint_match_values.do	match_salary_min_cons_ad_m_q.eps match_index_all_vars_v_m_q.eps
Figure C.3	figure_c2_3_se_sub_te_app_quint_match_values.do	match_salary_min_cons_ad_p_q.eps match_index_all_vars_v_p_q.eps
Figure C.4	figure_c4_se_sub_person_round_napplications.do	nmatches_byround_1vs2_fe.eps nmatches_byround_1vs2_fe_scale.eps
Figure C.5	figure_c5_6_se_sub_te_app_quint_mean_max_batch_vals.do	personround_between_meansalaryonround_q.eps personround_between_meanvindonround_q.eps
Figure C.6	figure_c5_6_se_sub_te_app_quint_mean_max_bat	personround_between_maxsalaryonround_q.

	ch_vals.do	eps personround_between_maxvindonround_q.eps
Table C.5	table_c5_11_se_sub_match_js_charact_marg_infra_apps.do	se_matchlevel_compliermeans_02
Table C.6	Currently, we are solving some issues with the table, hence not including its code.	
Table C.7	table_c7_se_sub_matchlevel_pecuniary_time_costs.do	se_matchlevel_smsfreecallback_v2.tex
Table C.8	table_c8_se_sub_reminder_text_messages.do	se_sub_matchlevel_procrastinationb_remindersms.tex
Table C.9	table_c9_se_sub_length_app_timing_call.do	se_sub_matchlevel_procrastinationa_test.tex
Figure C.7	figure_c7_se_sub_figure_joborder_applications.do	se_joborder_applications.eps
Table C.10	table_c10_se_sub_response_btw_periods.do	se_sub_personroundlevel_4_5_combined_wcontrols.tex
Table C.11	table_c5_11_se_sub_match_js_charact_marg_infra_apps.do	se_matchlevel_compliermeans_04
Table C.12	table_c12_se_sub_matchlevel_main_result_js_fe.do	se_matchlevel_paper_table2_unweighted_crossover_jt_fe.tex
Table C.13	table_c13_se_sub_matchlevel_main_result_crossover_matches.do	se_matchlevel_paper_table2_unweighted_cross
Table C.14	table_c14_se_sub_matchlevel_main_result_excl_rounds_JCs_job_info.do	se_matchlevel_paper_table2_unweighted_jobdetails
Table C.15	table_c15_se_sub_recall_receiving_matches.do	se_personcycle_gotcallsms_paper_fullsample_v14
Table C.16	table_c16_se_sub_beliefs_returns_to_search.do	se_person_instruments_non_res_paper_beliefs
Table C.17	table_c17_se_sub_treatment_effects_applying_random_chosen_vacancies.do	se_joborder_outcomes_d_separate_6.tex
Appendix D		
Figure D.1	tab_fig_d1_d2_se_sub_matchlevel_spillovers.do	currentse_spillovers_variation_density.pdf currentse_spillovers_variation_density_bysiz

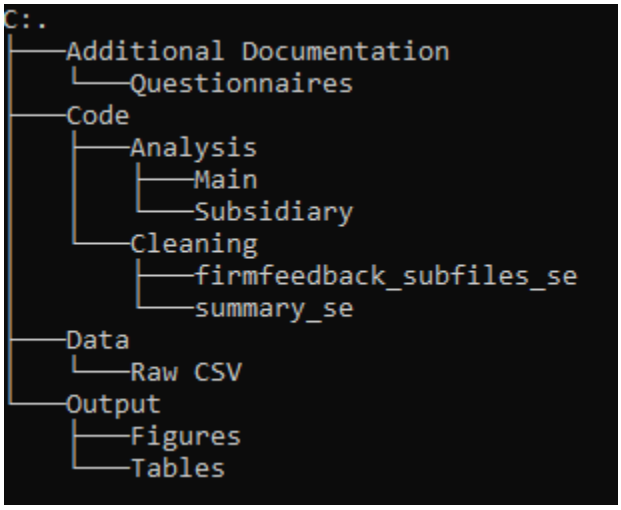
		e.pdf
Table D.1	tab_fig_d1_d2_se_sub_matchlevel_spillovers.do	se_spillovers_main.tex
Figure D.2	tab_fig_d1_d2_se_sub_matchlevel_spillovers.do	currentse_spillovers_treateffect.pdf
Table D.2	tab_fig_d1_d2_se_sub_matchlevel_spillovers.do	se_spillovers_vacancylevel.tex

Questionnaires

- The printables for all questionnaires can be found at the following location: *"\Additional Documentation\Questionnaire"*.
- Please note that the 'Subscriber Signup Surveys' contain three instruments: one for the Job Talash pool (jobseekers initially enrolled through the household survey, which serves as the sample for this paper) and two for other jobseeker samples. Additionally, the 'Jobseeker Updates Surveys' also contain multiple instruments corresponding to different enrollment samples.

Instructions to replicators

1. Download all files in this dataset to one directory.
2. Unzip all zipped files.
3. At this stage, the directory should have the following structure of subdirectories and subsubdirectories:



4. Run *01_masterdofile.do* to run the full sequence of data preparation and analysis, which will generate all tables and figures listed above

Acknowledgements

Aiman Farrukh, Harmalah Khan, and Sahil Nisar did exceptional work preparing this replication package.