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: 09 July 2026

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 78 do-files to prepare datasets. The ***analysis*** folder contains 47 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 Job Talash, a researcher-operated job search platform based in Lahore, Pakistan. Data collected through the platform is jointly owned by the World Bank, Center for Economic Research of Pakistan (CERP), and Duke University. 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.

Nature of the data included in this package: The datasets reflect operational data generated by the platform across the full study period, drawing on multiple sources collected at different points in time through phone surveys, in-person visits, and administrative records. These include baseline surveys of jobseekers and firms enrolled throughout the study, CV and profile updates made by jobseekers over time, job advertisement details listed by firms, administrative records such as job matches generated using the most up-to-date jobseeker and firm information at the time of matching, application decisions recorded through jobseeker surveys, CVs shared with firms, and feedback data collected from firms on those CVs.

Given the complexity and multiplicity of these data streams, which the team progressively consolidated over the study period, there is no single raw file or unified cleaning script that captures this process in a shareable, tractable form. Reconstructing and documenting the full data pipeline back to its original sources would require a substantial investment of time and resources.

To support replication as fully as possible, the team has therefore included the semi-cleaned data as the starting point. This represents the most complete and responsible version of the data we are able to share, and the 'Processed Intermediary Dataset' label in this README document reflects this.

Data Sources: 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: All datasets listed in the table below are provided in the replication package.

**The semi-cleaning conducted to produce "Processed Intermediary Datasets" includes appending and merging multiple administrative and survey rounds datasets, harmonizing variable names, cleaning inconsistent or invalid observations, recoding variables, generating identifiers and intermediate variables, applying variable labels, and conducting basic quality checks.*
***Access date is the duration of data collection for this study.*

Data Source/Survey name	**Processed Intermediary Dataset***		**Cleaned/output data**	
	Name & Access Date**	**Description**	**Name**	**Description**
Jobseeker side data				
Household survey	HHlisting_long_anony.dta Access date: October 2016-September 2017	Semi-cleaned household listing survey data containing information of households that showed interest in the job-matching services as well as data on other variables such as demographic and socioeconomic information on household members, and household characteristics. The dataset serves as the primary source for preparing household-level data subsequently used for analysis.	hhdata_prep_se.dta	Cleaned household listing survey data, with corrections for data entry errors and non-response observations, and preparation of key demographic and socioeconomic variables subsequently used in the construction of the match-level dataset for analysis.
Subscriber signup survey	JT_JA_signup_harmonized_se.dta Access date: December 2016 - October 2020	Consolidation of jobseeker-level baseline survey data sets into a harmonized dataset.	jt_ja_harmonized_prep_se.dta	Cleaned and consolidated baseline dataset to be used in the preparation of (i) other intermediary datasets and (ii) the final analysis dataset.
	jaebs_prep.dta Access date: July 2018 - October 2020	Data set containing enumeration block identifiers for jobseekers.		
Subscriber CV Information Updates survey	jt_ja_harmonized_prep_se.dta Access date: August 2017 - September 2022	Cleaned and consolidated baseline dataset to be used in the preparation of (i) other intermediary datasets and (ii) the final analysis dataset.	jt_ja_harmonized_updates_prep_se.dta	Cleaned and consolidated CV information updates dataset to be used in the preparation of (i) other intermediary datasets and (ii) the

	JTupdates_JA_harmonized.dta Access date: August 2017 - September 2022	Consolidated jobseeker-level update datasets containing changes to CV information over time into a single harmonized dataset.		final analysis dataset.
Jobseeker application/screener surveys	JT_JA_screener_appended.dta Access date: August 2017 - September 2022	Semi-cleaned application data 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.	pickup_prep_se.dta;	Jobseeker-round level dataset derived from the application (screener) data, containing variables that indicate the survey completion status of each jobseeker.
			screener_prep_se.dta;	Application data collapsed to the jobseeker-vacancy level, containing information on applications submitted by jobseekers to matched vacancies.
Jobseeker surveys about on and off platform search	CV_updates_highed_data Access date: July 2020	Backend dataset for the follow-up phone survey used to update jobseekers' CV information, with jobseekers assigned to batches that determines the scheduling of survey calls.	jt_ja_harmonized_JSF_CV_updates.dta	Primary purpose is to consolidate responses on job search and work and related questions from different survey instruments into one dataset.
	CV_updates_Highed_attempts.dta; Access date: August 2020 till September 2022	Semi cleaned phone call attempts level data data based on changes to CV information over time obtained from high-education jobseekers i.e. matriculation and above.		
	CV_updates_Lowed_attempts.dta; Access date:	Semi cleaned phone call attempts level data data based on changes to CV information over time obtained from low-education jobseekers i.e. below		

	June 2020 - February 2021	matriculation.		
	CV_updates_Lowed_FS_attempts Access date: April 2021 - September 2022	Semi cleaned phone call attempts level data data based on changes to CV information over time obtained from low-education jobseekers i.e. below matriculation. This data was collected through a short followup survey focusing on updates to the core job matching information.		
	cv_updates_uniq_ID_list_he_2022_08_11_b5.csv Access date: August 2022	Data set containing information on the call order random assignment based on which jobseekers are attempted for a phone survey within a batch.		
	20percent_prep.dta; Access date: June 2021 - September 2022	Jobseeker-round level survey response data on both on-platform and off-platform search activity.		
	20percent_bonuscalls_prep.dta Access date: July 2021 - July 2022	Jobseeker-round level survey response data on both on-platform and off-platform search activity. This survey was an add-on to existing activity targeting jobseekers who have been non-responsive during earlier rounds.		
	20percent_prep_attempt.dta Access date: June 2021 - September 2022	Phone call attempt level survey response data on both on-platform and off-platform search activity.		
	JT_followup_survey_rand.dta	Dataset on the random assignment of the recall period duration for any job		

	Access date: September 2019	search related activity for employment and search modules.		
	JT_rand_preload_10percent_hybrid mod.dta; Access date: May 2021	Dataset containing randomization assignments for secondary variations, and survey modules and cycles during which jobseekers were contacted to collect data on on-platform and off-platform search activities.		
	all_optout.dta Access date: December 2016 - September 2020	Jobseekers who have opted out of the job matching service.		
	JT_noid_updated.dta; JA_updated.dta Access date: August 2017 - September 2022	Jobseeker-level dataset containing updated CV information obtained from different surveys.		
	search_work_resp_ac.dta Access date: January 2021	Responses on job search and work-related questions recovered from qualitative information recorded by enumerators in the additional comments section of different surveys.		
	jsfollowup_prep.dta Access date: August 2019 - January 2020	Pilot data obtained from followup-survey for CV updating activity.		
	JT_JA_screener_appended.dta Access date:	Application data from the screener survey, where job seekers indicate which jobs they want to apply for after hearing details about the list of jobs		

	August 2017 - September 2022	they were matched with in a given round.		
	pickup_prep_se.dta <u>Access date:</u> August 2017 - September 2022	Jobseeker-round level dataset derived from the application (screener) data, containing variables that indicate the survey completion status of each jobseeker.		
	individual_level_subscriber_analysis.dta <u>Access date:</u> August 2017 - September 2022	Main matches dataset for analysis collapsed at jobseeker-level.		
Firms data				
Firm baseline survey & ad-listing survey	FirmSurvey_nodup_anon_se.dta <u>Access date:</u> May 2017 - September 2022	Firm baseline survey dataset at the firm level. Contains information collected during firm enrollment on the platform, including firm characteristics, organizational composition, recruitment practices, and other baseline firm attributes.	firms_prep_se.dta	Prepared firm-level baseline dataset containing all baseline variables used in the analysis. Includes information on firm characteristics, composition, and hiring practices.
	FS_adlisting_anon_se.dta <u>Access date:</u> July 2017 - September 2022	Adlisting dataset at the vacancy level. Contains information on vacancy characteristics, including vacancy descriptions, required qualifications, gender preferences, education and experience requirements, and other vacancy-related details collected at the time of firm adlisting on the platform.	adlisting_prep_se.dta	Vacancy-level dataset containing all vacancy advertisement variables used in the analysis. Includes information on vacancy characteristics, requirements, and other details.
Firm feedback	FS_feedback_cleaned_anon_se.dta	Firm feedback dataset at the	feedback_joblevel_prep_allvars_	Feedback dataset at the vacancy

survey *(Please note that the intermediary datasets in this section are used in the construction of all output datasets.)*	 <u>Access date:</u> September 2017 - September 2022	survey-interaction level. Contains information collected from firms on vacancy outcomes, CV delivery, applications, interviews, and hiring outcomes for vacancies listed through the platform.	se.dta	level containing all firm feedback variables. Includes information on CV delivery, and on-platform and off-platform interview outcomes, hiring outcomes, and vacancy status.
	FS_MidlineSurvey_feedback_cleaned.dta <u>Access date:</u> October 2020 - September 2022	Midline firm feedback survey dataset collected from firms that previously listed vacancies through Job Talash. The unit of observation is the jobseeker–vacancy match and includes jobseeker-level feedback and vacancy outcomes.	feedback_joblevel_prep_se.dta	Feedback dataset at the vacancy level containing key vacancy outcome variables. Includes measures of vacancy status, applications, interviews, and hiring outcomes.
	FS_del_cleaned_unique_jobid.dta <u>Access date:</u> June 2018 - September 2022	CV delivery dataset at the vacancy level. Contains information on CV delivery status and related vacancy-level outcomes.	feedback_matchlevel_prep_allvars_se.dta	Prepared feedback dataset at the JT candidate and vacancy match level containing all feedback variables. Includes match level information on CV delivery, and on-platform and off-platform interview outcomes, hiring outcomes, and vacancy status.
	FS_non_compliance_int_med.dta: <u>Access date:</u> September 2017 - September 2022	Vacancy level dataset of firms that did not utilize the service for that vacancy.	feedback_matchlevel_prep_se.dta	Feedback dataset at the jobseeker–vacancy match level containing key outcome variables. Includes jobseeker-level information on CV delivery, interview, hiring outcomes, and vacancy status.
	FS_interview_med.dta: <u>Access date:</u>	Additional dataset that includes information on jobseeker-vacancy level interview outcomes.		

	September 2017 - September 2022			
	cv_tracking_batches.dta <u>Access date:</u> August 2017 - September 2022	Dataset containing records of CVs shared with firms against jobseeker applications, based on the completeness of the jobseekers' CVs.		
Spillover Effects	match_level_analysis.dta <u>Access date:</u> August 2017 - September 2022	We combine data from multiple sources and generate a dataset at the match level which is mostly used as part of analysis.	triminput	Data prepared from main jobseeker-vacancy level dataset for vacancy-level spillover effects.
Administrative data				
Matches data	matches_JT_JA_consolidated.dta <u>Access date:</u> August 2017 - September 2022	Jobseeker-vacancy level administrative dataset containing matches between jobseekers and vacancies based on jobseekers' profiles, and preferences, as well as firms' requirements. This data also contains the main treatment variable of interest, the "phone call treatment."	matches_prep_se.dta	Jobseeker-vacancy level dataset prepared by generating variables required for analysis, and retaining the ones that are required.
Analysis data				
Main matches dataset for analysis	hhdata_prep_se.dta; jt_ja_harmonized_prep_se.dta; jt_ja_harmonized_updates_prep_se matches_prep_se.dta; pickup_prep_se.dta; screener_prep_se.dta; adlisting_prep_se.dta; firms_prep_se.dta; feedback_matchlevel_prep_se.dta; feedback_joblevel_prep_se.dta; cv_tracking_batches.dta	Description on these datasets is provided above.	match_level_analysis.dta	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.

	JT_noid_duprem.dta Access date: December 2016 - October 2020	Jobseeker-level baseline data.		
	intmed_prep.dta Access date: September 2017 - September 2022	Additional dataset that includes information on jobseeker-vacancy level interview outcomes.		
	FS_del_cleaned_unique_memidjobid_anon.dta Access date: June 2018 - September 2022	CV delivery dataset at the jobseeker-vacancy level. Contains jobseeker-level information on CV delivery for applications submitted through the platform.		
Jobseeker Jobseeker-round-level, and Job-level analysis data sets	match_level_analysis.dta Access date: August 2017 - September 2022	We combine data from multiple sources and generate a dataset at the match level which is mostly used as part of analysis.	individual_round_level_analysis.dta	Main matches dataset for analysis collapsed to jobseeker-round level
			individual_level_analysis.dta	Matches dataset for analysis collapsed at jobseeker-level and merged with data on jobseekers who never subscribed for the job-matching service.
			job_level_analysis_se.dta	Main matches dataset for analysis collapsed to the vacancy level, containing variables based on jobseeker applications and feedback received from firms on the CVs shared with them.
Data with predicted probabilities of	match_level_analysis.dta individual_level_analysis.dta job_level_analysis_se.dta	Description on these datasets is provided above.	se_matchlevel_latent_interview_probability_variation_v1.dta	Data on the predicted probability of application for each jobseeker–vacancy match.

interviews	firms_prep_se.dta.dta 20percent_prep.dta		Access Date: August 2017 - September 2022	
External data				
LFS data			LFS 2018-19.dta Access date: 2018 - 2019	Labour Force Survey dataset obtained from the Pakistan Bureau of Statistics (PBS) website. The original data was provided in SPSS format and subsequently converted to Stata format.

Miscellaneous data: Auxiliary datasets for which the input files are extensive; therefore, the data preparation steps are not included in the replication package.

Data Source/Survey Name	**Dataset Name**	**Description**
Hypothetical matches data	match_level_analysis_sensitivity.dta Access date: August 2017 - May 2022	Hypothetical matches dataset created through holding fixed jobseekers' initial occupational preferences, education, and experience. This dataset is generated after each matching round by rerunning the matching algorithm on the latest jobseeker–vacancy data to produce a hypothetical dataset, which is then appended to the existing data from previous rounds.
Scoring of CVs	FR_jobs_cv_scoring.dta; CC_jobs_cv_scoring.dta Access date: March 2022	Data on CV quality scores from the hiring managers for a subset of jobs on the platform for the CVs of jobseekers matched to those jobs. Each CV was scored by the field managers

		hired on-platform.
Call order in which the enumerator call JS for screener surveys	post_sep2020_call_order.dta Access date: September 2020 - February 2021	Call order data for survey rounds from September 2020 till February 2021. This dataset was generated by appending the randomized survey call-order across different matching rounds for jobseekers who were contacted for the screener survey.
	pre_sep2020_call_order_recovered.dta Access date: March 2018 - February 2020	Call order data recovered for survey rounds prior to September 2020, when this data was not being consolidated. This dataset was generated by appending the randomized survey call-order across different matching rounds for jobseekers who were contacted for the screener survey.
CV Information Updates Data	cv_updates_highed_data_25 Jun 2021.dta Access date: June 2021	Survey backend dataset on the eligibility of jobseekers for the follow-up phone survey used to update jobseekers' CV information.
	cv_updates_highed_data_2020_11_30.dta Access date: November 2020	Dataset identifying jobseekers who were part of either the pilot or the scale-up of the follow-up phone survey used to update jobseekers' CV information.
	Updates_survey_var_JT.dta Access date: April 2021 - September 2022	Consolidated dataset combining information from multiple survey instruments on the dates when jobseekers were contacted for a survey.
Jobseeker-vacancy level data for two additional rounds	match_level_oct2022.dta Access data:	Additional two rounds of jobseeker-vacancy level matches data beyond the analysis cutoff, included to capture the "crossover"

	October 2022-November 2022	experiment implemented after the cutoff period.

Details on each data source

All relevant datasets used in the analysis presented in this paper are available in the directory:" */Data"*. Additionally, CSV versions of these datasets can be accessed in the subdirectory ":*/Data/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/wp-content/uploads/2020/07/LFS-2018-19.zip. The data are in the public domain. The questionnaire related to the survey can be found here: "\Additional Documentation\Questionnaires\9. LFS Questionnaire-2018-19-Final.pdf"

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_08_screenersurvey_prep.do	Clean screener survey data	screener_prep_se.dta pickup_prep_se.dta
5.	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
6.	se_sub_cleaning_05_firmbl_adlisting_prep.do	Clean firm baseline and adlisting data	firms_prep_se.dta adlisting_prep_se.dta
7.	se_sub_cleaning_06_firmfeedback_prep.do	Clean firm feedback data on CVs received	feedback matchlevel_prep_allvarsse.dta feedback_matchlevel_prep_se.dta feedback_joblevel_prep_allvarsse.dta feedback_joblevel_prep_se.dta
8.	se_sub_cleaning_07_matches_prep.do	Clean match data	matches_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
    a) *Please note that for some .eps figures included in the paper, we also generate .png versions of the figure. Since these don't go in the paper, we don't mention them here, but they do appear in the code and output folders.*
    b) Please note that we have not included code for replicating Tables A.2 and A.3 because these largely contain descriptive/text-based content. Likewise, we have not included code for Figures A.2 and A.3 because these are flowcharts and images.
        i) For Table A.3, the mean applications across platforms were extracted by the authors from various published and working papers and converted into monthly rates. Therefore, the code for this table is not available.
3) Selected tables and figures in the paper, as explained and justified below

Flow of *main_se_analysis_paper.do*		
Figure/Table #/Stats	**Program**	**Output file**
One off statistics	se_sub_oneoff_stats.do	se_stats_oneoff.tex
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	figA1.pdf
Table A.2		Table_A.2.tex
Table A.3		Table_A.3.tex
Figure A.2		SMS_template_latest.pdf
Figure A.3		SE_T_vs_C_flowchart.pdf
Appendix B		
Figure B.1	figure_b1_se_sub_person_application_interviews_ntimes.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_planinterview_panel_Updated.tex; se_matchlevel_hte_selected_int_conso_bound_panel_Updated.tex; 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; se_matchlevel_panelb_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; 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.tex
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.tex
Table C.2	table_c2_se_stats.do	Table_C.2.tex table_c2_se_stats_log.smcl
Table C.3	table_c3_se_stats.do	Table_C.3.tex table_c3_se_stats_log.smcl

Table C.4	table_c4_se_stats.do	Table_C.4.tex table_c4_se_stats_log.smcl
Figure C.1	figure_c1_se_sub_app_decision_by_treat.do	combined_graph_simple_v2.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_batch_vals.do	personround_between_maxsalaryonround_q.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.tex
Table C.6	table_c6_se_sub_dispersioneffects_value_vacancy.do	effects_on_dispersion_2.tex
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

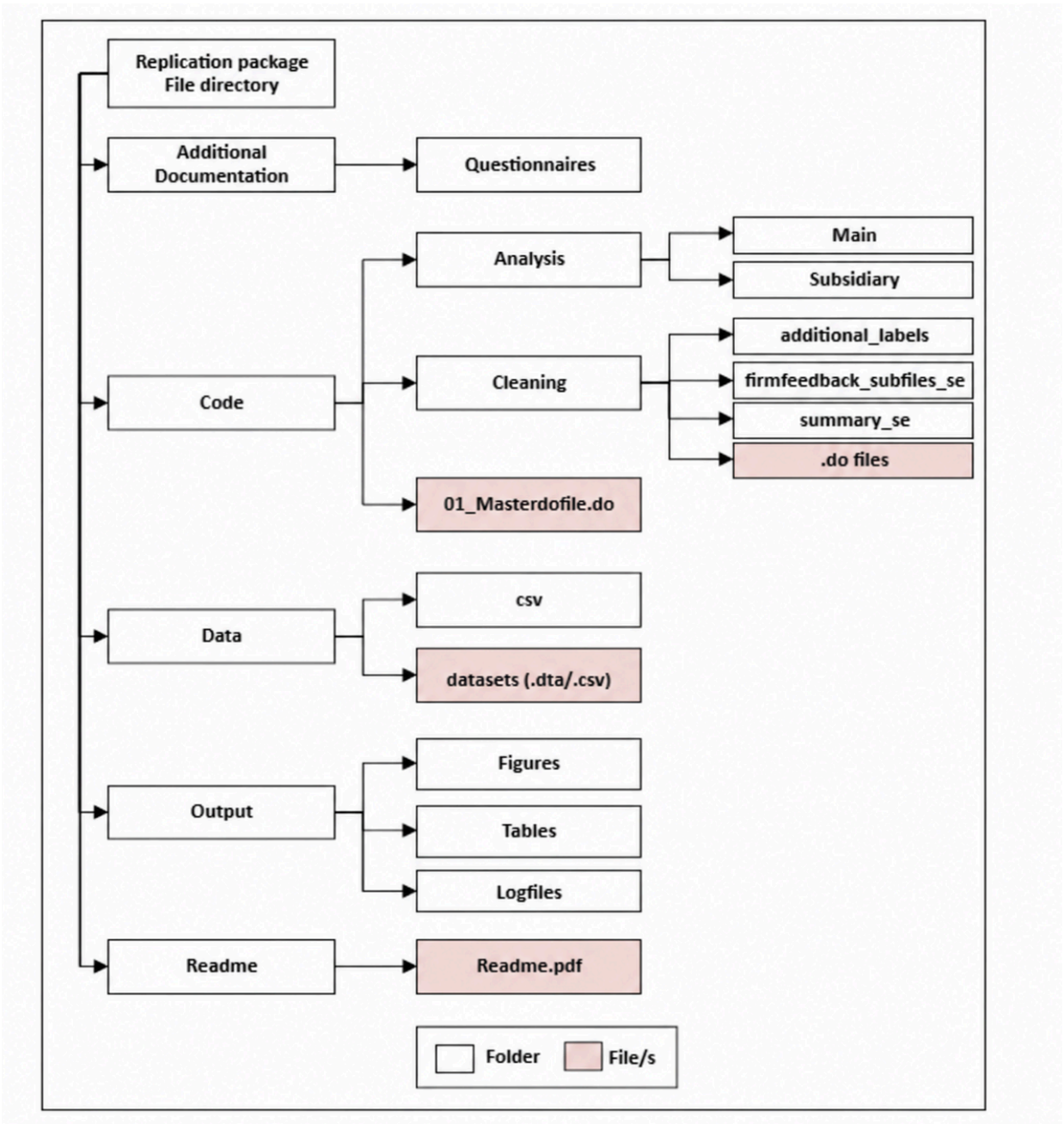
Table C.10	table_c10_se_sub_response_btw_periods.do	se_sub_personroundlevel_4_5_combined_wcontrols.tex
Table C.11	table_c5_11_se_sub_match_js_charact_marg_infra_apps.do	se_matchlevel_compliermeans_04.tex
Table C.12	table_c12_se_sub_matchlevel_main_result_js_fe.do	se_matchlevel_paper_table2_unweighted_crossover_jt_fe.tex
Table C.13	table_c13_se_sub_matchlevel_main_result_crossover_matches.do	se_matchlevel_paper_table2_unweighted_cross.tex
Table C.14	table_c14_se_sub_matchlevel_main_result_excl_rounds_JCs_job_info.do	se_matchlevel_paper_table2_unweighted_jobdetails.tex
Table C.15	table_c15_se_sub_recall_receiving_matches.do	se_personcycle_gotcallsms_paper_fullsample_v14.tex
Table C.16	table_c16_se_sub_beliefs_returns_to_search.do	se_person_instruments_non_res_paper_beliefs.tex
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	se_spillovers_variation_density se_spillovers_variation_density_bysize
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	se_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:


Replication package
File directory
Additional
Documentation
Questionnaires
Code
Analysis
Main
Subsidiary
Cleaning
additional_labels
firmfeedback_subfiles_se
summary_se
.do files
01_Masterdofile.do
Data
csv
datasets (.dta/.csv)
Output
Figures
Tables
Logfiles
Readme
Readme.pdf
Folder
File/s

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.