READ ME:

CREATION OF STUDY DATASETS

Source

All the datasets we use come from the French public employment agency, Pôle Emploi. They can be exploited by researchers after the research project is approved, and under the rules defined in the "data access convention". For data access, researchers should contact the Direction des Statistiques, des Etudes et de l'Evaluation of the Département Etudes et Recherche of Pôle Emploi. In particular, Anita Bonnet, was the person in charge of the data access for our project (anita.BONNET@pole-emploi.fr).

First, we created a list of unique identifiers from the online application dataset ("Telecandidature" in SISP). We implemented a linkage *at the individual level* to extract all the information for these individuals from the following administrative datasets:

1. unemployment episodes from unemployment registers ("DE" in FH),
2. employment episodes from employers' mandatory declarations ("DPAE")
3. unemployment benefits claim ("ODD" in FNA)
4. unemployment benefits collection episodes ("PJC" in FNA)

We also link *at the vacancy level* with administrative datasets for vacancies ("OE" in STMT).

The data construction proceeds in several steps.

Individual spells datasets

We identify all workers with at least one observation in "Telecandidature". For these individuals, we extract information on all unemployment episodes in "DE". We restrict to unemployment spells when the individual is available to work and required to search for a job (Pole Emploi category A, B, or C). We clean the unemployment spells to make sure that there is no overlap between several spells, and we merge spells that are less than 30 days apart. We keep from this dataset information on the dates of start and end of unemployment spells (based on "datins" and "datann") and on individual demographic variables (education, etc.) We use the information from all unemployment spells to build variables on the cumulated unemployment duration prior to each unemployment episode. We then restrict our sample to unemployment spells that start after January 1st 2013. We hence have an unemployment spells dataset ("Data_Spell").

We merge at the individual level the dataset "ODD" with "Data_Spell". We identify the unemployment spells for which we observe unemployment benefits claims in "ODD" during the spell. We create two subsamples of "Data_spell": one with those who have *no benefits claim* during their spell ("Data_Spell_noUI"), and one with those who have a benefits claim corresponding to the standard UI system ("Data_Spell_UI"). We restrict the sample in "Data_Spell_UI" to spells with the benefits claim that correspond to the standard UI system (based on "kcrd" and "kcaaj"). Importantly, we get the variable for the potential benefits duration ("kpjdxp").

We merge at the individual claim level the dataset "Data_Spell_UI" with the dataset "PJC". Each individual benefit claim can be associated with several different benefits spells in "PJC". We ensure that individuals have a benefits spell shortly following the UI claim. We identify the periods of actual benefits collection, based on the variable "kcpjc". We count the days with an interruption in benefits collection before the end of the potential benefits duration. We exclude these periods such that the end of benefits collection corresponds with the potential benefits duration (note that we show in robustness check that this choice does not affect our results).

Employment and job search outcomes

We create a data set with all the job applications sent by individuals in "Data_Spell_UI" and "Data_Spell_noUI", using the online applications dataset "Telecandidature". We link this dataset with the vacancy dataset "OE", at the vacancy level. This linkage allows us to collect information on the characteristics of the job applied to: the posted hourly wage, the weekly hours, the experience level required, the term of the contract, the firm identifier. We append the data on online

applications from “Data_Spell_UI” and “Data_Spell_noUI”, into a unique dataset for workers eligible for UI and those ineligible: the “All_applications” dataset.

We link the individual spell datasets “Data_Spell_UI” and “Data_Spell_noUI” with all employment episodes tracked in DPAE. For each unemployment spell, we identify the first contract start date that is posterior to the date of start of the spell—-when there is one. We de-duplicates unemployment spells and create a dataset with the re-employment date for each unemployment spell for workers eligible for UI and those ineligible: “All_Remployment_date”.

Study datasets

We merge “All_applications” with “Remployment_date”. We convert the dates of employment episodes and online application in spell time, by subtracting their date from the date of the start of the unemployment spell. For spells in “Data_Spell_UI”, we also subtract the number of days with benefits interruption prior to this date, to reflect the time of benefits collected (note that we show in robustness check that this choice does not affect our results). We only keep the applications that are sent before an exit from the unemployment register, or a new employment contract, and for up to 3 years. For each month of the unemployment spell, we build the following search outcomes:

- We count the number of application sent each month of the unemployment spell;
- We compute the average predicted wage, based on the posted wage and the characteristics of the vacancies applied to;
- We compute other measures of job quality: we take the average characteristics among jobs applied to, for job characteristics that are continuous variables (such as the number of hours per week); and the proportion of jobs with desirable characteristics, for job characteristics that are discrete variables (such as open-ended contracts)
- We compare individual characteristics and vacancies characteristics to detect when applicants are applying to jobs below their education (“educ”) or skill level (“qualif”).
- We build a measure of the distance between the city where the applicant lives, and the city where the job is located. We collect information on the latitude and longitude corresponding to the center of each city and compute the geographic distance between all pairs of cities. We add the distance between city centers, and the radius of each city. For job seekers applying to a plant located in their own city, the measure of distance hence corresponds to the city diameter.

We additionally identify the month of the first contract following the start of unemployment spells.

Based on this database, we create 2 type of datasets for the rest of the empirical analysis. First, we build a cross-sectional database with the individual characteristics of the job seeker at each unemployment spell, which can be merged with the other data bases at the spell level. Second, we create a monthly panel at the unemployment spell and month level. Our main dataset of analysis is a quarterly panel: we build it by averaging the variables of the monthly panel.

PRODUCTION OF MAIN RESULTS FROM STUDY DATASETS

Datasets

The mail empirical analysis is based on the 2 datasets constructed as described above:

1. Cross_section: The cross-sectional data base, containing the individual characteristics of the job seeker.
2. Panel_quarterly: The quarterly panel, at the spell and quarter level, containing the search and employment outcomes.

Samples

We also define several samples that are key for the rest of the analysis:

- The roundPBD_sample contains the spells of unemployed workers ineligible for UI, or eligible for 6, 12, 24 or 36 months of benefits. We use this sample for the descriptive statistics. We restrict the sample to individuals who have at least one application during their unemployment spell.
- The analysis_sample contains the spells of unemployed workers ineligible for UI, or eligible for 12 to 24 months of benefits. We restrict the sample to individuals who have at least one

application during their unemployment spell, and who stay unemployed more than 1 month, such that we can identify within spell variation in search effort. We use this sample to produce all the results.
- The no_missing_wage sample is a subsample from the analysis_sample, with at least two quarters when the target wage is not missing, such that we can identify within spell variation in target wage. We use this additional sample restriction in all the analyses of target wage.
- The balanced_sample is a subsample from the analysis_sample, made of longer spells as explained in the paper. We use this additional sample restriction each time we want to neutralize changes in the sample composition.

Main variables

- spell_identifier: Uniquely identifies each unemployment spell, i.e. each individual and registration date
- ind_identifier: Uniquely identifies each individual
- PBD: Potential benefits duration
- quarter_*: Quarter of the unemployment spells dummies
- quarter_to_exhaustion: Potential benefits duration - spell time, for UI recipients. It is positive before benefits exhaustion and negative after.
- quarter_to_exhaustion_*: Dummies, based on quarter_to_exhaustion. They are always equal to 0 for individuals not eligible for UI.
- Count_Application: Measure of search effort
- explained_monthly_w: Measure of target monthly wage
- Quarter_FirstJob: Quarter when the first contract posterior to the registration date is observed
- last_explained_monthly_w2: Proxy for re-employment wage

Dofiles

To produce the main results in the paper, you can run the Master dofile, which calls the dofiles producing the main outputs: each dofile has the name of the resulting table or figure.