**README**

**Softwares**

The softwares used to extract, clean and analyze the data are Stata 15, SAS and R.

**Datasets used**

We use French administrative data which cannot be made publicly available but which can be accessed through a secure server (CASD) after a successful application process. All the relevant information to apply for the data can be found here: <https://cdap.casd.eu>

Here are the datasets we asked for

* **FH-DADS :** <https://www.casd.eu/en/source/matching-of-the-job-seekers-history-file-and-the-annual-declaration-of-social-data>

This dataset comes out in different annual files:

* dads (information on employment spells)
* de (information collected at unemployment registration)
* pjc and di (information on unemployment insurance claims)
* m0 (job applications and hiring outcomes)
* **Panel DADS :** <https://www.casd.eu/source/panel-des-declarations-annuelles-de-donnees-sociales>
* **Panel DADS EDP :** <https://www.casd.eu/source/panel-dads-apparie-avec-ledp>

On the CASD portal information on GPS coordinates of municipalities’ centroids are available to all.

We use data on vacancies posted on the French public employment service (Pole employ). These data can be accessed through an agreement with Pole emploi (contact the head of studies and statistics, Cyril Nouveau in 2020 at [cyril.nouveau@pole-emploi.fr](mailto:cyril.nouveau@pole-emploi.fr)). Files are named Vacancies2010\_clean, Vacancies2011\_clean, and Vacancies2012\_clean.

We use the public-use files of the survey of Unemployed Workers in New Jersey by Krueger and Mueller (2016) available at <https://www.aeaweb.org/articles?id=10.1257/pol.20140211>

**Dofiles**

To reproduce all the paper: run the dofile overall\_dofile which will call the various dofiles described below in the relevant order

All datasets above mentioned are to be put in the source folder

All dataset created by the programs below are stored in the data folder

All tables and figures created by the programs below are stored in the output folder

1. **Cleaning the data**

* 0\_prepare\_de.do

Cleans the “de” file from FH-DADS (see above)

* 0\_prepare\_dads.do

Cleans the “dads” file from FH-DADS

* 0\_prepare\_edp.do :

Cleans the Panel DADS EDP data used in panel b and d of Figures 3, 4 and C

* Convert\_stata\_paneldads2015.sas

Creates Paneldads\_pre2009 and Paneldads\_post2009 used for Figure 1

1. **Preparing the main datasets**

* 1\_construction\_main\_dataset:

It creates the dataset basedetravail\_final\_QJE used in most of the paper using the raw di dataset and the prepared de and dads files from 0\_prepare\_de.do and 0\_prepare\_dads.do

* 1\_construction\_application\_dataset:

Creates the dataset m0\_vac\_de\_travail used for Table 7, 8, Figure 7 and Appendix B

* 1\_construction\_job2job\_dataset:

Creates the dataset job2job\_2004\_2012 used for Appendix Table D11

* 1\_create\_joint\_density:

Creates the dataset used for Figure 5

1. **Reproducing the Tables and Figures of the paper and its online appendix**

**Figure 1**

* 2\_Figure 1.do

It creates Figure 1. The input data is Paneldads\_pre2009 and Paneldads\_post2009 from Convert\_stata\_paneldads2015.sas

**Descriptive evidence (section 2 and 3)**

* 2\_Tables\_1\_to\_3\_figures\_2\_to4.do

It creates the main Tables and Figures of sections 2 and 3. The input data are basedetravail\_final\_qje created in the dofile 1\_construction\_main\_dataset and edp\_data created in 0\_prepare\_edp

* 2\_KM.do

Creates the results of Table IV using data from Krueger and Mueller (2016). This can be run outside of the CASD environment.

* 9\_Appendix\_Tables\_D1\_toD11\_figures\_C3\_C4\_C6.do

It creates the Appendix Tables and Figures referred to in sections 2 and 3. The input data are basedetravail\_final\_qje created in the dofile 1\_construction\_main\_dataset. Do, edp\_data created in 0\_prepare\_edp and job2job\_2004\_2012 created in 1\_construction\_job2job\_dataset

* 9\_Appendix\_Figure\_C5

It creates Figure C5 using as input the raw data paneldadsedp2010

**Estimation of the WTP for a shorter commute parameter (section 4)**

* 3\_Figure5.R

Creates Figure 5. It uses the dataset produced by 1\_create\_joint\_density

* 4\_Tables\_5\_D12\_D13\_D14.do

This dofile creates the dataset indiff\_curve\_dataset used to estimate the WTP parameter. It then runs the main estimation of alpha as well as robustness checks. To get the estimates of the various columns, filter have to be changed manually (line 160 to 195). The standard errors are obtained when running the bootstrapped programs.

* 4\_Tables\_A1\_A2.do

It creates the estimates of alpha under Interpretation 2. The input data is indiff\_curve\_dataset created at the beginning of the dofile Tables\_5\_D12\_D13\_D14.do

**Calibration (section 5)**

* 5\_Table\_D15\_D16\_calibration

It produces 3 datasets with relevant input for the calibration tables: calibration\_sum\_stat, calibration\_wage\_distribution and calibration\_distance\_distribution.

* The next step is to run the R scripts **5\_decomposition\_interp1.R** and **5\_decomposition\_interp2.R**, which generate the material for Tables VI, D15, D16 (5\_decomposition\_interp1.R) and Table A3 (5\_decomposition\_interp2.R). The two R scripts rely on auxiliary functions defined in **decomposition\_funs.R** (called in both scripts).
* 9\_Appendix\_Figure\_C8

It creates Figure C8. The input data are basedetravail\_final\_qje created in the dofile 1\_construction\_main\_dataset and vac\_com\_quarter\_2005\_2012 (quarterly data of vacancy postings at the municipality level)

**Evidence from applications (section 6)**

* 6\_Tables\_7\_B1\_B2\_Figure\_7.do

It creates the Tables of section 6 and Appendix B, except table 8 which is produced in a separate dofile. The input dataset is m0\_vac\_de\_travail created in 1\_construction\_application\_dataset

* 6\_Table\_8\_choice\_model.do

It creates Table 8. The input dataset is m0\_vac\_de\_travail created in 1\_construction\_application\_dataset