Read me file, Replication package for:

Trading Non-Tradables: The Implications of Europe's Job Policy

Mathilde Munoz

1- General description

This folder contains the necessary information to replicate the results presented in the main text and supplementary online appendix of the paper. The analysis builds mostly on **confidential data** collected for six countries (France, Belgium, Germany, Austria, Luxembourg, Portugal) as well as an administrative dataset on posting flows covering all EU member states. Nevertheless, those datasets can be used by other researchers after **following the steps for data requests and access detailed in this document.** The programs provided in this replication package allow to reproduce the results based on the confidential administrative datasets once data access has been granted.

Because the analysis builds on several countries, and each country has a specific dataset structure and procedure to access and use the datasets, the replication package provides one main program per country/dataset analyzed in the paper. This read me file describes each program, the figures it reproduces in the paper, and the datasets used, as well as the necessary steps needed for other researchers to replicate the results.

When getting started on the process to request and access datasets on postings, researchers and students should feel free to contact me at mathildemunoz@berkeley.edu to ask for help and advice in how to proceed. If you have any question about data availability for your research projects (period coverage, variable availability), you should also feel free to contact me.

2- Tools and specifics to run the programs

All programs are coded using Stata. The versions of Stata vary because not all servers (recall that those administrative sources are often kept in different countries and servers) have the same version installed. I will detail those specifics when describing each program.

Because the analysis must be carried in different countries, and in different servers (because each country has different confidentiality and access constraints), it is not possible to use a "main" code that calls all programs. Rather the researcher will have to perform the analysis in each country (often by travelling there) and on each dataset to replicate all the results presented in the main paper. For this reason, the replication package provides one program per main figure (and country analyzed). Each of those programs will have different data and output paths that will correspond to each server used in each of those countries.

3- Programs and data description

Fig1.do: This program provides the necessary steps to replicate Figure 1 of the main paper. It is produced using two main datasets.

- EULFS microdata: those micro-files are needed to build descriptive statistics on the stock and flows of international EU migrants. The micro files are confidential and researchers must apply for access following the guidelines described here: https://ec.europa.eu/eurostat/documents/203647/771732/How_to_apply_for_microdata_access.pdf . The process of application takes approximately 6 months and upon approval, the research will receive micro-files in CDROMs formats. The files can be accessed in the yearly or quarterly format; this paper uses the yearly format. It is important to always use the weighting variable provided with the survey (COEFF in the yearly files) when carrying analysis using the EU-LFS micro files.
- E101/A1 data: You need to fill a special request to the European Commission (DG employment) to get access to the social security posting forms A1. You should start by filling the regular contact form https://ec.europa.eu/social/main.jsp?catId=2#:~:text=Call%20the%20free%20phone%20number,from%20anywhere%20in%20the%20EU with a description of the request and purpose of the research project. After filling this initial request, the EC should initiate a data access request in their system. If the EC grants approval, the files should be communicated by excel files and should report number of issued posting forms by member state, destination countries (for article 12 posting), and duration of posting contracts. Files for latest years (2012-2020) also have better information on sectors.

The program Fig1.do starts from the raw EU-LFS datasets to build stock of EU migrants and cross-border migration flows. Those series are then combined with the serie on the number of posted workers measured through the A1 forms provided by the EC. The resulting aggregate statistics for year 2016 are provided in .xlsx format in the program folder. This program can be used in your personal computer, as the dataset will be directly provided to you upon approval.

Fig2.do: this program provides the program that reproduce the Figure 2 of the main text, as well as supplementary figures in Appendix B1. The main dataset used to reproduce Figure 2 is based on administrative posting registries on incoming posted workers collected for 4 countries: France, Austria, Germany and Belgium.

- Data for Germany: the data for Germany comes from SOKA-BAU, the national union for the construction sector that collects information on incoming posting since 2000. The researcher should send an email to the SOKA-BAU contact (arbeitnehmer@soka-bau.de) with a brief description of the research project and the intent to use the data. If approval is granted, the data is provided by excel files.
- Data for Austria: the data for Austria is collected from the BUAK, the national union for construction workers in Austria, that collects information on incoming posted workers since 2006. The researcher should send an email to the BUAK contact (kundendienst@buak.at) with a brief description of the research project and the intent to use the data. If approval is granted, the dataset is provided by .pdf files and needs to be converted to excel file (or another format) by the researcher manually.
- Data for Belgium: the data for Belgium come from LIMOSA and more specifically the National Social Security Office. The aggregate data at the origin-year level can be requested

by describing the research project and the intent to use the data at this link: https://www.onss.be/contactez-nous. Upon approval, the data will be sent by excel files. It is also possible to access the underlying micro-data, as I will explain in "Fig5.do".

- France: The data on origin-year posting inflows has been collected through the "DPD collectes" since 2000. The data can be accessed by sending an email to (dgt.dir@travail.gouv.fr) with a brief description of the research project and the intent to use the data. I will provide more information in the description of "Fig5.do".

The main program for the reproduction of Figure 2 uses a dataset appending all those 4 data sources together. The program starts by performing the difference-in-differences analysis around the posting liberalization reforms described in Table B5 of the paper, using different sets of fixed effects. The data program then runs several robustness checks presented in appendix B1, and uses the same data replacing posting reforms with migration reforms (described in Table B5) to run the same gravity model.

To run this program, it is necessary to have reghdfe, ftools, multipledid_gt and ppmlhdfe installed. The program has been run using Stata version 17. Note: the command multipledid_gt may or may not produce a graph after you run it depending on the version of stata you use, and depending on when this adds-on command was installed.

Fig3.do: This program reproduces Figure 3 of the main text. It is based on the granular posting forms "SIPSI" collected by the French administration since mid-2016. The dataset has been created for the paper "TNT" and we have worked together with the DARES to make it accessible to all researchers. While I had access to this dataset before its availability in the CASD, the datasets and all variables are now available and described here https://www.casd.eu/en/source/online-information-system-for-declaring-employee-secondment/. To get access to the dataset, the researcher should create a project in the CDAP portal where it will be possible to request this data source. Every 6 months, the CDAP meets and decide whether the researchers should get approval for the data request. If you request is approved, the data will be accessed through a secured databox, It takes approximately 6/7 months to get access to the dataset, and the CDAP may request that you access the data only in France. You may also have to present you research project in person if the CDAP requires it. The data provided through CASD is in csv forms and has three main files (see link above for a description).

This program will have to be run in the external CASD server, and will require that you travel to France (or in a EU country) in order to access a secured data access.

Fig 4.do: This program reproduces Figure 4 of the main text in the paper. It uses micro administrative data provided by the Banco of Portugal. The analysis must be performed in the external server provided by the Banco of Portugal. The process to get access to those datasets is well described here: https://bplim.bportugal.pt/content/access-0. The researchers should send their application to the email address provided in the webpage linked above. Upon approval, the researchers will get access to the external server where raw data will be posted. The programs always start from the raw datafiles provided by BPLIM. For the paper TNT, the CBHP dataset

were provided for years 2006-2017 only—the researchers should make sure to obtain the same dataset archive in order to replicate the results using the same sample.

The program will have to be run in the external BPLIM sector—the outputs and data paths correspond to the external server.

Fig5_Tab1.do: This program allows to replicate the main results presented in Figure 5 and Table 1, as well as most of the supplementary analysis on the French labor market presented in Appendix D1. To produce this analysis, the researcher needs to get access to the DPD dataset that measures posting forms at the province-sector-year level. To get access to the dataset, the researcher should send a description of the research project to the contact listed in the description of Fig3.do program. In addition to the dataset on posting, the program uses publicly available datasets on French provinces:

- Sectoral employment statistics for years 1989-2015 is "T202 : Emploi salarié en fin d'année par département et région de France (hors Mayotte), selon le secteur d'activité (A38) et le sexe". You can download this file here: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwigrqio2KD_AhUBIX0KHVLGD5wQFnoECAgQAQ&url=https%3A%2F%2Fwww.insee.fr%2Ffr%2Fstatistiques%2Ffichier%2F3524476%2FT202.xls&usg=AOvVaw35OJwoW4bFp8kAwTA3gyrN . Note that the sectoral classification used by the INSEE is not exactly the same than in the DPD dataset (this is because the DPD dataset doesn't have a statistical purpose, it is a dataset that is used by labor inspectorates to perform labor inspections. That is way the classifications in terms of sectors do not always map with definitions used by other institutes in employment datasets). You will need to extract sectors TGUMNZ, TGURURZ and TGUIZ from TGU to get the tertiary sector in the sense of the DPD (what I call TGUO).
- Population statistics: "Source : Insee - Estimations de population". I use working age population e.g 20-39 and 40-59 categories. You can download the dataset here: https://www.insee.fr/fr/statistiques/1893198
- Share of migrants and employment by citizenship: micro census files can be downloaded here https://www.insee.fr/fr/statistiques/4248862?sommaire=2414232#consulter . As detailed in the appendix, the "years" in the census file are not true years because of the rotating dimension of the census after 2004. Only 1/5 of the census is updated each year, meaning that the true variation occurs only every five years after 2004. Before that, the censuses were decadal.
- Unemployment rates at the province-year level can be found here: https://www.insee.fr/fr/statistiques/series/102760732

The program provides the necessary code to reproduce Figure 5 and Table 1, as well as the main robustness checks and supplementary analysis presented in Appendix D1 and D3. The program can be run on the researcher personal computer, as the dataset should be provided directly to the researcher upon approval.

To run this program, it is necessary to have reghdfe, ftools, coefplot and egenmore installed. The program has been run using Stata version 17.

Fig6.do: This program replicates Figure 6 of the main text. It uses administrative, confidential data from Belgium. The LIMOSA dataset on incoming posted workers, as well as the Belgian matched employer-employee dataset are described here: https://www.ksz-bcss.fgov.be/fr/dwh/source/variables/bysource

To access the data you need to follow the steps described here: https://dwh-live.bcss.fgov.be/fr/dwh/dwh_page/content/websites/datawarehouse/menu/description-du-cadre-general.html . This involves sending a detailed description of the research project and the intent use of the dataset. It is important to note that only researchers **affiliated to an EU institution** are allowed to request the dataset. The dataset cannot be used outside the CBSS. **You will need to travel to Brussels and use the data "on-site" in the secured data center.** You will run your programs in the internal server of the CBSS – the data and outputs paths in the programs correspond to this server.

The program replicates Figure 6, starting from raw datafiles on received posting forms in Belgium combined with matched employer-employee data on Belgian workers. The datasets are merged using a common firm identifier for the Belgian client firms. The program estimates the event-study coefficient plotted in Figure 6, and performs additional tests presented in Appendix D4.

To run the program, you will need reghdfe, ftools,egenmore, did_imputation and coefplot. The program has been run using Stata 16.

Fig7_Tab2.do This program replicates Figure 7 of the main text, results presented in Table 2, as well as most of the supplementary analysis presented in Appendix D4 for France.

The analysis is performed using confidential French administrative data on posting, French firms, and French employment. The datasets should be accessed through the CASD, after filling an access request at the CDAP (see link provided in description of Fig3.do). The datasets needed are:

- SIPSI: Posting forms recorded in France from 2017 to 2020. The dataset is in .csv format and has three datasets which are described here: https://www.casd.eu/en/source/online-information-system-for-declaring-employee-secondment/.
- FICUS: balance-sheets data for firms in France. This dataset is particularly useful because it has names and ID of companies in order to maximize the matching with the DPD data. Note that at the time of TNT, 2018 was the latest year available. The dataset is provided in .sas format. It can be converted to .dta using "import sas" but if the configuration of the CASD server is too small (my case), you need to use SAS and convert each file manually or with a short SAS loop.
- DADS postes: The matched employer-employee dataset covering all job spells in France each year. The dataset is described here: https://www.casd.eu/source/base-tous-salaries-fichier-postes/?tab=1 . Note that each year, the dataset provides information on current year and past year job spell, so can actually be converted into a 2-year panel. Because of the size of the dataset, it is provided in 4 separate .SAS files.

The first part of the program cleans the raw data and proceeds to the matching of the posting dataset with the French datasets on domestic employment and firms' balance-sheets data, using the firm identifiers (SIREN) and names. The second part of the programs proceeds to produce descriptives on using firms (compared to non-using) and compares the wages of posted workers and domestic workers at the same workplace.

To use those datasets, **you will need to travel to France or the EU and work on a secured data box.** All programs will be run on this server, and you will need to request approval for outputs extraction. All data and output paths correspond to the CASD server.

Fig8-9.do: This dataset used confidential firm-level balance sheets data from BPLIM—see description of Fig4.do to obtain information on the data access process. The program must be run in the external server provided by the central bank of Portugal. To run the program, you need coefplot, reghdfe, ftools, multipledid_gt, egenmore.

The program starts from the raw panel dataset (CBHP) provided by BPLIM for the period 20006-2017: the researcher should make sure to access the same dataset to replicate exactly the results in TNT. It reproduces Figure 8-9 as well as most of the supplementary analysis presented in Appendix D1.

Fig10.do: Same datasets and remarks than before apply. The program starts from the same raw datasets to produce Figure 10.

Fig11.do : This program replicates the series for Luxembourg plotted in Figure 11 (series for Portugal are produced in Fig8-9.do in the Portuguese servers). The dataset is the matched employer-employees dataset in Luxembourg. The dataset and the data access process is described here: https://igss.gouvernement.lu/fr/microdata-platform.html

Only researchers affiliated to a Luxembourg-based organization can request the dataset. If the request is approved, the researcher must obtain a "luxtrust" which requires to get an approved ID by a notary or lawyer in your country. You will then receive a connection device and a card to connect to the external server where you will be able to perform this analysis. All data and outputs paths relate to this server.

Folder "more_appendices": while most supplementary figures and tables are reproduced in the programs described above, this folder contains the program to replicate the remaining appendix tables and figures. Those programs build on the datasets built in the main code.