Readme file for "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement"

Lorenzo Caliendo (Yale University)
Luca David Opromolla (Banco de Portugal)
Fernando Parro (Pennsylvania State University)
Alessandro Sforza (University of Bologna)

Data Sources

Our data set is built from different sources. Our main data source is the EU-LFS which we use to construct the gross migration flows. We use the information contained in the WIOD database to construct bilateral trade flows, real wages and share of labor compensation in value added, EUKLEMS to complement the information in WIOD for the years 2009-2013, WITS database to construct bilateral tariffs and the Portuguese matched employer-employee panel data set (Quadros de Pessoal or QP) to perform one of the estimates for the elasticity of substitution.
While WITS, WIOD and EUKLEMS are publicly available, EU-LFS is only available upon request to the European Labour Force Survey office and Quadros de Pessoal is available from Statistics Portugal (*INE*) through its protocol with the *Direção-Geral de Estatísticas da Educação e Ciência*.[1]

Codes:

Step 1.
Prepare gross migration flows and migration shares. COPS_grossflows.do

This dofile prepares the dataset of gross migration flows (See Online Appendix B for a detailed description of the data and each step of the procedure). This is the core do file, which takes the raw migration data and creates the gross migration flows and the migration share. The structure of this code is detailed at the beginning of the do file.

[1] https://www.dgeec.mec.pt/np4/46/

We import the raw micro data into Stata using the Setup_EULFS_1983-2016_y.do and Labels_EULFS_1983-2016.do Stata dofiles made available by the German Microdata Lab of GESIS - Leibniz Institute for the Social Sciences (http://www.gesis.org/en/gml/). We restrict the imported dataset to the following set of variables:

country - Current country of residence (ISO country classification)
countr1y - Country of residence one year before survey (ISO country classification)
refyear - Year of survey
coeff - Yearly weighting factor (Numbers in thousands)
refweek - Reference week
age - Age of interviewed person (in 5-year age bands, 0-4, 5-9,...)
national - Nationality
hat97lev - Highest educational attainment level (ISCED 1997 classification, till 2013)
hat11lev - Highest educational attainment level (ISCED 2011 classification, from 2014)
ilostat - ILO Work status
wstat1y - Situation with regard to activity one year before survey

Step 2

Prepare the tariff data for the model. COPS_tariffs.do

This dofile prepares the tariff data to be used in the model. We use the information contained in the WITS database and use the effectively applied tariff rates from TRAINS dataset and complement missing or incomplete information using the WTO dataset. The dofile constructs tariff matrices using TRAINS data, WTO data, and a combination of the two, which we then use to produce the bilateral tariff matrices for the model. The structure of the code is detailed at the beginning of the dofile.

Step 3

Prepare real wages and share of labor compensation in value added. COPS_wiod.do

This do file uses the socio-economic accounts from WIOD to prepare the files to build the real wages and the shares of labor compensation in value added. We complement the information contained in WIOD with two additional

datasets. First, to construct a complete series of employment level by skill, we combine the information from WIOD for the years 2000 to 2008 with the information from EUKLEMS for the years 2009 to 2014 (COPS_euklems.do). Second, to construct the series of real wages we use the information on labor compensation from the Socio Economic Accounts in the WIOD database and the information on the price levels of the countries in our sample from the Penn World Tables version 9.0 (www.ggdc.net/pwt). See appendix B.4 of the Online Appendix for additional details.

Step 4

Prepare trade data. COPS_wiot.do

This do file prepares the trade data starting from the WIOT input output tables. We keep the set of countries consistent with the migration data and we pool all the remaining countries in the rest of the world. Values are in US dollars at current prices. See appendix B.3 of the Online Appendix for additional details.

Step 5

Run regressions to estimate migration costs.
COPS_costs.do

This dofile contains the regressions to estimate the changes in migration costs. A detailed description of the different estimations can be found at the beginning of the dofile.

Step 5

Run regressions to estimate migration elasticity.
COPS_migration_elasticity.do

This dofile contains the regressions to estimate the migration elasticity. A detailed description of the different estimations can be found at the beginning of the dofile.

Step 6

Run regressions to estimate elasticity of substitution.
COPS_elasticity_of_substitution.do

This dofile contains the regressions to estimate the elasticity of substitution. A detailed description of the different estimations can be found at the beginning of the dofile

Step 7

Reduced form evidence. COPS_reducedformevidence.do

Runs the regression of the reduced form evidence presented in table 1 of the paper.

Additional codes

- COPS_datamatlab.do

  Prepares the Matlab matrices for the model estimation.

- COPS_figure2.do

  Creates figure 2 of the paper.

- COPS_figure3.do

  Creates figure 3 of the paper.

Online Appendix

1. COPS_qp_Jul_20_2020D.do

Run regressions to estimate elasticity of substitution using Portuguese matched employer-employee data.

This dofile produces the estimates of the elasticity of substitution using the Portuguese matched employer-employee data described in Online Appendix F of the paper.

2. COPS_migrationdatachecks.do

This dofile compares our migration data to other sources, as described in section B.2.6 of the Online Appendix

3. COPS_figuresD.1.2.3_appendix.do

This dofile creates the figures of section D of the Online Appendix.

4. COPS_transition_prob.do

This dofile creates produces the transition probabilities described in section C of the Online Appendix.