Data and Code for "Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States," by Pol Antras, Evgenii Fadeev, Teresa C. Fort, and Felix Tintelnot. The Review of Economics and Statistics

Overview

This Readme documents the necessary steps to replicate all results in the paper and appendix of: **"Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States."**

The paper uses confidential micro-data from the US Census Bureau and the Bureau of Economic Analysis. We provide the code necessary to construct the micro-datasets within the secure FSRDC network (https://www.census.gov/about/adrm/fsrdc.html), the disclosed output, and the code to reproduce the tables and figures that appear in the paper using the disclosed output.

Data Availability and Provenance Statements

The microdata are available within the FSRDCs and can be accessed by researchers with an approved project. Directions on writing a project proposal and the requisite datasets to request are in the detailed instructions below.

We also use several publicly available datasets. These datasets must be uploaded to the FSRDC system to reproduce our results. We provide these datasets in rdc_replication/input. Details on the public datasets and their provenance follow.

We use 2 publicly available datasets from CEPII:

1) Countries_V202211.dta – a crosswalk containing mappings between various country codings, including ISO2, ISO3, ISO3 numbers, and country (name) strings. Also contains information on the age of the country, as well as previous names for the country. This dataset is used to map the CEPII data to the Census microdata.

2) Gravity_V202211_abb.dta – a dataset with bilateral measures such as distance.

3) The CEPII data were downloaded from:
http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8

We use publicly available country-level data from Antràs, Fort, and Tintelnot (2017)

1) country_data.dta: a dataset with various country-level variables
2) These data are available here: https://faculty.tuck.dartmouth.edu/teresa-fort/data/

We use publicly available data on UN geographic regions

1) UN_country_regions.xlsx
2) These data were downloaded from:
https://unstats.un.org/unsd/methodology/m49/overview/

We use a concordance between HS codes and BEA IO codes constructed using public data from Pierce and Schott (JESM 2012), BEA industry codes, and NAICS codes to concord between the material trailer files and HS codes (to make Appendix Table A8).

1) hs_imp_naics_beax.dta
2) This concordance was constructed from public data on the BEA, Census, and Schott websites.

Statement about Rights

I certify that Teresa Fort (one of the authors of the manuscript) has legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

Some data **cannot be made** publicly available.

Details on each Data Source and Output in Paper

To reproduce the results in the paper using disclosed data, follow the directions in programs/fdi_trade_script.do:

1. Set up the appropriate directories.

2. Include the disclosed data in the input directory.

3. Run the script from the programs directory and the results will appear in output.
    a. Note that some formatting from the Excel tables to the final latex files is necessary.
    b. The script specifies the output in the paper produced by each of the programs.

4. The results in gvc_output.dta are provided to facilitate future research.
    a. Please cite this paper if you use any of the results in that file.
    b. You must cite CES Project 1530 (release on 7/15/2019), and CES Project 6907751 (BEA-FY23-P6907751-R2) in any output that uses the results in these files.

5. The code is all in Stata (version 15).

6. These programs reproduce Appendix Tables A.1, A.2, A.3, A.4, A.5, and A.8.

To reproduce the disclosed results, you will need to access Census and BEA microdata:

1. To gain access to the Census and BEA microdata, follow the directions here on how to write a proposal for access to the data via a Federal Statistical Research Data Center: https://www.census.gov/about/adrm/ced/apply-for-access.html and contact BEA at specialswornresearch@bea.gov

2. You must request the following datasets in your proposal to reproduce all our code and results. Note that we have only checked the underlying datasets used to create the results in our paper. Although the code creates data for other years, those have not been checked and may not be reliable. Note that only US citizens may access the BEA data.

    a. Longitudinal Business Database (LBD), 2007
    b. Longitudinal Foreign Trade Transactions Database (LFTTD), 2007
    c. Foreign Trade Data – Imports (IMP) 2007
    d. Census of Construction (CCN), 2007
    e. Census of Finance, Insurance, and Real Estate (CFI), 2007
    f. Census of Manufactures (CMF), 2007
    g. Census of Mining (CMI), 2007

h. Census of Retail Trade (CRT) 2007
i. Census of Services (CSR), 2007
j. Census of Transportation, Communications, and Utilities (CUT), 2007
k. Census of Wholesale (CWH), 2007
l. Standard Statistical Establishment List (SSEL), 2007
m. Company Organization Survey (COS), 2007, including cos2007wfirmid.dta which contains mappings between the COS and firmid
n. BE-11 files: (note that these files are archived for this replication package - from transfer20160318)
    i. be11_2007r_fa_faentity.sas7bdat (received from BEA and archived with this paper's replication package)
    ii. be11_2007rfa_faform.sas7bdat (received from BEA and archived with this paper's replication package)
    iii. be11_2007r_usr_usrentity.sas7bdat (received from BEA and archived with this paper's replication package)
    iv. be11_2007r_usr_usrform.sas7bdat (received from BEA and archived with this paper's replication package)
    v. We recommend also requesting the corresponding BEA files that are now available on the FSRDC servers.

o. BE-12 files (note that these files are archived for this replication package, we converted these data from CSV to SAS and have archived the raw SAS files)
    i. gvc2007_be15_entity.sas7bdat (received from BEA and archived with this paper's replication package)
    ii. gvc2007_be15_form.sas7bdat (received from BEA and archived with this paper's replication package)
    iii. We note that the filenames say 15, but the data are from the BE-12 2007 Benchmark Survey of Foreign Direct Investment in the United States.
    iv. We recommend also requesting the corresponding Benchmark Survey of Foreign Direct Investment in the United States (Form BE-12) data that are now available on the FSRDC servers. See https://www.researchdatagov.org/product/1963 for more information.

p. Crosswalks (Note that these were early crosswalks developed in conjunction with Fariha Kamal and co-authors. They are now superseded by similar files on the FSRDC servers.)
    i. affiliate_crosswalk_2007.sas7bdat (archived with paper replication package)
    ii. parent_crosswalk_2007.sas7bdat (archived with paper replication package)
    iii. We recommend also requesting the corresponding Crosswalk between BEA Annual Survey of U.S. Direct Investment Abroad (Form BE-11) to Census Business Register and Crosswalk between BEA Benchmark Survey of Foreign Direct Investment in the United States (Form BE-12) to Census Business Register that are now available on the FSRDC servers. See https://www.researchdatagov.org/product/785 and https://www.researchdatagov.org/product/860 for more information.

3. You should also reference "**Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States**" CES Project 6907751 in the proposal and request archived code and data from this project.

   Note that because this work was done on an internal project that contributed to the new BEA-Census data concordance construction, the files and concordances we used were not available on the FSRDC servers at the time of our analysis. We have asked BEA and CES to archive the unredacted programs required to reproduce the results and those BEA datasets that were not available on the FSRDC servers at the time of our analysis (specifically items listed under 2n, 2o, and 2p). We strongly recommend that future researchers also request the corresponding BEA datasets now available on the FSRDC servers so that their results are easily replicable and because those files may contain updates and improvements relative to the earlier files used in our work.

4. You should create a main directory with the following subdirectories:
   - code (place all programs here)
   - input (place all input data here)
   - the other directories are produced by the script

5. The unredacted, internal code is set up to so that each program needs to cd to the main directory at the top, but the remaining directories do not need to be specified.

6. You can now recreate all the tables and figures in the paper by running the afft_script.bash program. To do so, cd to the program directory and enter the command: "qsub afft_script.bash". All the output will be created in the output subdirectory. The script also contains notes about what each program does, including the input datasets used and the output created.

   Note that Appendix Table A.8 (imports by good and firm types) is made in a separate script: make_import_type_03.bash that must be run after the main aft_script.bash file. We kept this code separate since it is also useful for researchers without the MNE data who wish to analyze the types of goods manufacturing firms import and how they relate to their domestic production and input purchases.

7. The programs to construct the basic datasets are written in SAS (this is the native environment for Census data, version 9.4), while the analysis programs are in Stata (version 16.1)

8. These programs produce the disclosed data used by the public programs, and also Tables 1, 2, 3, 4, A.6, and A.9 and Figures 1 and 2.

To reproduce Figure 3:

1. Run the program code_figure3.m using Matlab (version R2022a).

Computational requirements

The public replication files can be run on a 2021 laptop with 16GB of RAM and a 64-bit operating system. The confidential code must all be run on the IRE in the FSRDC network.

Memory and Runtime Requirements

The script for replicating the Census results specifies the memory requirements needed to run the jobs.

The public replication files take less than 10 minutes to run.

The FSRDC code takes under 2 hours to run.

Description of programs/code

The public and the FSRDC programs each have a master script. The public programs script is called fdi_trade_script.do. It lists each program and the specific output (i.e., table and figure numbers) that each program creates. The rdc programs script is called afft_script.bash. It lists each program in the order it needs to run, and it describes what each program does. To replicate the results, researchers need to copy the program and input directories as provided and then simply run each script.

References

Conte, M., P. Cotterlaz and T. Mayer (2022), "The CEPII Gravity database". CEPII Working Paper N°2022-05, July 2022.

Pierce, Justin and Peter Schott. "A Concordance Between Ten-Digit US Harmonized System Codes and SIC/NAICS Product Classes and Industries," *Journal of Economics and Social Measurement*. 37(1-2): 61-96.

U.S. Bureau of Economic Analysis, Annual Survey of U.S. Direct Investment Abroad, 2007.

U.S. Bureau of Economic Analysis, Benchmark Survey of Foreign Direct Investment in the United States, 2007.

US Census Bureau. 1977-2017. *Longitudinal Business Database [database]*. Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 1992-2021. *Longitudinal Foreign Trade Transactions Database [database]*. Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Construction, 1977-2017 [database]*. Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Finance, Insurance, and Real Estate, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Manufactures, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Mining, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Retail Trade, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Services, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Transportation, Communications, and Utilities, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2017. *Census of Wholesale, 1977-2017 [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 1977-2021. *Standard Statistical Establishment List [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 1986-2021. *County Business Patterns Business Register [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2002-2021. *Company Organization Survey [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.

US Census Bureau. 2002-2021. *Company Organization Survey firmid bridge [database].* Federal Research Data Centers [distributor], last accessed 2023-07-06.