

README

This README provides directions on how to replicate the results in:
“Factoryless Goods Producing Firms” by Andrew Bernard and Teresa Fort.

Researchers wishing to replicate the analyses in this paper can do so at any of the U.S. Census Bureau’s Research Data Centers (RDCs). To gain access to an RDC, researchers must write a research proposal and submit it to an RDC administrator. This proposal must list the specific datasets that the researcher plans to use. Additional information on the RDC proposal process is available here: <https://www.census.gov/ces/rdcresearch/>.

The main analyses in this paper use information from the following RDC datasets:

- 2007 Census of Wholesale Trade
- 2007 Census of Manufactures
- 1992-2007 Longitudinal Business Database
- 1992-2007 Foreign Trade Data-Imports

The algorithms in the paper used to match the import data to the Economic Censuses and the LBD directly also use the following datasets:

- 2007 Census of Mining
- 2007 Census of Finance, Insurance, and Real Estate
- 2007 Census of Construction
- 2007 Census of Retail
- 2007 Census of Services
- 2007 Census of Transportation, Communications, and Utilities
- 2007 Business Register

Researchers could use the Longitudinal Firm Trade Transactions Database in lieu of directly matching the import, though the sample and statistics are likely to differ somewhat under this approach.

The following programs are used to construct the datasets for the analyses:

`script_aer.bash`

This program lists all the programs used in the analyses. It can be run from the command line to replicate all results.

The individual programs used in the analyses, listed in the order in which they should be run, are:

SAS_01_Make_Firm_Panel.sas

This program makes a firm panel from 1992-2007 with employment and imports.

SAS_02_Make_2007_Imports.sas

This program makes the 2007 Matched Import data.

SAS_03_Make_Census_Estabs.sas

This program pulls in the 2007 data from the CMF and CWH

SAS_04_Make_FGP_Firm_Panel.sas

This program makes the FGP panel dataset.

STATA_01_FGP_Analysis.do

This program calculates the 2007 summary statistics for FGPs.

STATA_02_FGP_Panel_Analysis.do

This program calculates the panel statistics for FGPs.

Note that, because pathnames cannot be disclosed, researchers must enter the correct path for each dataset used and/or produced in the analysis. Directions for specifying the pathnames are in the programs. In addition, the SAS programs require either (a) an autoexec file that specifies the libname for the requisite input datasets (for example, the libname cmf points to the RDC directory for the cmf data), or (b) specifying a libname in the beginning of each program.