Replication Files for Endogenous Production Networks

Daron Acemoglu and Pablo D. Azar

August 23, 2019

This document describes how to use the replication files for the paper *Endogenous Production Networks* by Acemoglu and Azar (2019).

1 Organization of the Replication File Directory

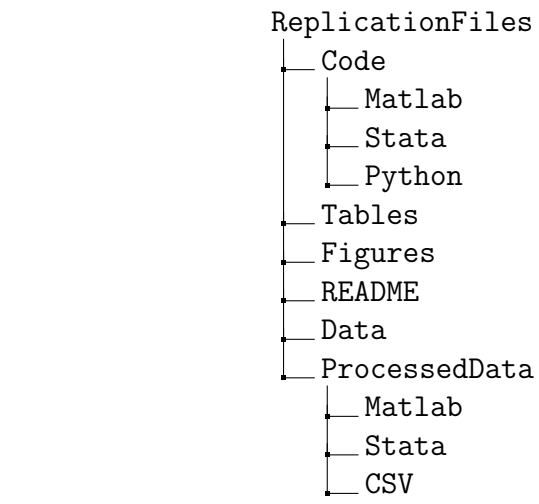
The replication file folder contains 6 subfolders:

1. The *Code* folder contains three subfolders with code in Matlab, Stata and Python, respectively.
2. The *Tables* folder contains intermediate and final tables presented in the paper.
3. The *Figures* folder contains figures displayed in the paper.
4. The *README* folder contains the source for this README file.
5. The *Data* folder contains raw data, including
    - *Full_table.dta*, the harmonized input-output tables from Acemoglu, Autor and Patterson (2017)
    - *CxI_DR_2007_2012_PRO_Det.xlsx*, which contains the 2007 input-output table from the BEA

- *Use_SUT_Framework_2007_2012_DET.xlsx*, which contains value added by industry as well as personal consumption expenditures by commodity for 2007 (from the BEA)
- *CCAllYears.mat* A file with aggregated input-output matrices used to produce Figure 1
- *sic5811.dta*, a file of manufacturing industry data from the NBER-CES survey
- *BEA-BLS-industry-level-production-account-1987-1998.xlsx* a file of non-manufacturing industry data from the BEA and BLS industry level production accounts

6. The *ProcessedData* folder contains intermediate datasets used by the Stata,Matlab and Python programs. The processed data is in three subfolders, called *CSV*, *Stata*, and *Matlab*.

The directory tree of the ReplicationFiles folder is given in the following figure.



2 Replicating Figure 1

To replicate Figure 1 in the text, run the program *./Code/Matlab/Figure1.m*.

3 Replicating Table 1 and Tables C1-C4

To replicate Table 1 and Tables C1-C4, run the following programs in sequence

1. *./Code/Stata/process_data.do* takes as input the harmonized input-output tables from Acemoglu, Autor and Patterson (2017) and outputs the Jaccard distances $d(S_{i,t}, S_{i,t-1})$ for each $(industry, year)$ pair as well as the indegrees $(\mathcal{I}_{i,t})$ for each industry, year pair.

2. *./Code/Stata/merge_data.do* merges the computed network statistics (Jaccard distances, indegrees) with industry data from the BEA, BLS and NBER-CES datasets. It produces an output file with the data necessary to run regressions: the industry change in productivity $\Delta \log TFP_{i,t}$ and the Jaccard Dummies $J_{i,t,10}, J_{i,t,20}, J_{i,t,30}$.

3. *./Code/Stata/run_regressions.do* runs the regressions reported in Tables 1 and Tables C1-C4. These regressions are of the form

$$\Delta \log \hat{TFP}_{i,t} = bJ_{i,t} + c_i + d_t + f\Delta \log TFP_{i,t-1} + e_{i,t}.$$

   It also outputs, for every industry and year pair, the predicted change in log TFP attributable to changes in supplier sets. This predicted change is $\Delta \log \hat{TFP}_{i,t} = \hat{b}J_{i,t}$, where $\hat{b}$ is the estimated coefficient in the regression. When there is no lagged TFP term, the predicted change in TFP for industry i over the entire sample period is

$$\Delta \widehat{\log TFP}_i = \sum_{t=0}^{T} \hat{b}J_{i,t}.$$

   With a lagged TFP term, we need to take into account that the effect of changing supplier sets acts on this period's TFP through two channels. Every period, there is a direct effect of size $\hat{b}J_{i,t}$, and an indirect effect through lagged TFP of size $\sum_{k=1}^{\infty} f^k \hat{b}J_{i,t-k}$. Thus, every period t, the effect of changing sets on productivity (including the indirect effect through lagged TFP) is $\sum_{k=0}^{t} \hat{f}^k \hat{b}J_{i,t-k}$. Adding this up over all periods $t \in \{0, ..., T\}$, we obtain that

$$\Delta \widehat{\log TFP}_i = \sum_{t=0}^{T} \sum_{k=0}^{t} \hat{f}^k \hat{b}J_{i,t-k}.$$

4. *./Code/Matlab/ComputeCounterfactuals.m* computes the predicted change in aggregate log GDP attributable to changes in supplier sets. This predicted change is

$$\Delta \widehat{\log GDP} = \beta' \mathcal{L} \Delta \widehat{\log TFP}$$

where β is the vector of consumption shares by industry, $\mathcal{L}$ is the Leontief input-output matrix, and $\Delta \log \hat{TFP}$ is the vector of predicted changes in industry TFP over the entire sample period, computed by *./Code/Stata/run_regressions.do*.

5. *./Code/Python/writeTable.py* collates the results into the tables presented in the paper. The tables produced by this file are

   - *./Tables/big_table_tfp.tex*, which corresponds to Table 1.
   - *./Tables/big_table_tfp_10.tex* which corresponds to Table C1.
   - *./Tables/big_table_tfp_30.tex* which corresponds to Table C2.
   - *./Tables/big_table_tfp_weighted.tex* which corresponds to Table C3.
   - *./Tables/big_table_tfp_1997.tex* which corresponds to Table C4.

4 Replicating Appendix D

To replicate the results in Appendix D, including figures 3 and 4, run the program *./Code/Matlab/ComputeSimulations.m*