

README

The Effect of High-Tech Clusters on the Productivity of Top Inventors

Enrico Moretti

Overview

This file describes how to replicate the estimates in the paper The Effect of High-Tech Clusters on the Productivity of Top Inventors

To replicate the paper, please run the stata file MAIN.do. This file calls all the files needed to produce all the tables and figures in the paper

Software: Stata Version 13. I wrote and run the codes on Unix. The replicator should expect the code to run for about 15 hours.

Data Availability and Provenance Statements

- ☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

If box above is checked and if no simulated/synthetic data files are provided by the authors, please skip directly to the section on [Computational Requirements](#). Otherwise, continue.

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

- ☒ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on Each Data Source

The paper uses several publicly available data sources. The main dataset is COMETS. All the regressions are based on COMETS data.

A) COMETS

The Connecting Outcome Measures in Entrepreneurship, Technology, and Science (COMETS) database (Zucker and Darby, 2014) was downloaded from

<https://www.kauffman.org/entrepreneurship/research/comets/data-files/>

The Codebook can be found here

<https://www.kauffman.org/entrepreneurship/research/comets/codebook/>

Datafiles in Stata:

patent_inventors.dta

patent_zd_cats.dta

patent_assignees.dta

patent_cite_counts.dta

patent_citations.dta

B) NBER patent database

The NBER patent data base (Hall, Jaffe and Trajtenberg, 2001) was downloaded from

<https://www.nber.org/research/data/us-patents>

Datafile in text format:

apat63_99.txt

C) KODAK STOCK PRICES

The data is from a table published by 1stock1.com (2017). The table was downloaded in 2017 from

http://www.1stock1.com/1stock1_173.htm

The data were hand-entered into excel by me and saved as a csv file.

Datafile in csv format

kodak_stock.csv

D) INPUT OUTPUT TABLES

The BEA Table “Use Table 2007, 2012: [405 Industries](#)” (BEA 2012) was downloaded from <https://www.bea.gov/industry/input-output-accounts-data>

Datafile in csv format

Use_SUT_Framework_2007_2012_DET.csv

E) CROSSWALK FROM NAICS TO PATENT CLASSES

This file is from Nikolas Zolas -- United States Patent Classification (USPC) Crosswalks - Version 1610. It was downloaded from

<https://sites.google.com/site/nikolaszolas/PatentCrosswalk>

Datafile in text format

uspc_Class_to_naics07_6.txt

F) BEA Economic Area names

This data is from Table 1 in Johnson (2004) downloaded from

https://www.csus.edu/indiv/j/jensena/sfp/ea_desc.htm

I hand entered the table and created a data file

Datafile in stata format

bea_codes.dta

G) COUNTY CHARACTERISTICS

This file has county characteristics and it is from Moretti (2021).

Datafile in stata format

tva1.dta

H) CROSSWALK from FIPS TO Economic Areas

The crosswalk that links county fips codes to bea economic areas codes is from

moretti, enrico. county - economic area crosswalk. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-15.

<https://doi.org/10.3886/E140581V1>. It can be downloaded here:

<https://www.openicpsr.org/openicpsr/project/140581/version/V1/download/terms?path=/openicpsr/140581/fcr:versions/V1&type=project>

Datafile in text format

bea_county_crosswalk.txt

Computational requirements

Stata (code was last run with version 13).

- estout (as of 2018-06-12)
- reghdfe (as of 2019-07-12)
- ftools
- ivreghdfe
- ivreg2
- ranktest
- The program does not install dependencies since installation is not allowed on my unix system. The replicator needs to install them.

I wrote and run the codes on a Unix system. The replicator should expect the code to run for about 15 hours. The code was run on a node with two 14-core CPUs. The node has 132 GB dedicated RAM.

Description of programs/code

- To replicate the paper, please run the stata file MAIN.do. This file calls all the files needed to produce all the tables and figures in the paper . Each do file is called using the stata command "include". MAIN.do contains a brief explanation of what each .do file does
- Programs in data/ will extract and reformat all datasets referenced above, and they will save new data files
- Programs in data2/ will create and save new data files needed for the regressions
- Programs in the root directory generate all tables and figures in the main body of the article.
- Output files are in the directories tables/ to tables36/

The directories are organized as follows:

- a) The root directory contains the file MAIN.do and the files that create the tables and figures
- b) The directory data/ contains the original datasets used, the do files that read them and the stata data files created after reading them, cleaning them and reformatting them. The subdirectory data/documentation contains the documentation
- c) The directory data2/ contains all other data files created in the process of running the regressions and some do files that create intermediate data files
- d) The directories tables/ to tables36/ contain all the tables and figures

All the tables come in 2 formats: .tex (latex) and .txt (text)

Instructions to Replicators

- Edit MAIN.do to adjust the default path
- The data files referenced above should be in the directory data/ Each file should be unzipped

- Create directories named data2/ tables/ tables3/ tables22/ tables36/ tables23/ tables4/ tables11/ tables30/ tables2/ tables31/ tables20/ tables34/
- Run MAIN.do to run all steps in sequence

Details: List of al the .do files

This file MAIN.do calls all the files needed to produce all the tables and figures in the paper.

The following is a list of .do files called by MAIN.do, their location and a brief description of what each file does. Note that \$main/ is the root directory

- \$main/data/create_COMETS_Patent_ExtractForEnrico.do

This file reads and reformats COMETS data

- \$main/data/read_apat

This file reads and reformats APAT data

- \$main/data/read1

This file reads and reformats Input-Output Tables

- \$main/data_3.do

This file creates the main working dataset

- \$main/kodak_stock.do

This files creates Figure 2 -- Top panel

- \$main/reg3.do and \$main/reg27.do

These 2 files generate Table 2 (col 1 to 4) Table 4 (col 1 to 4); Fig 2 (bottom panel) and Fig 3

- \$main/data/read

This file reads tva1.dta and bea_county_crosswalk.txt and creates read.dta

- \$main/reg22.do

This file generates Table 2 (col 5) Table 4 (col 5)

- \$main/reg.do

This file generates Table 3, Table 6 (rows 1-8), 8 (top panel); Fig 5, Fig A1

- \$main/reg23.do

This file generates Fig 6

- \$main/reg2.do

This file generates Table 1, 10, 11, 12 (column1); Appendix Table A1; Fig 1 and 4

- \$main/reg29.do

This file generates Table 12 (column2)

- \$main/iv_new.do

This file creates the instrumental variables

- \$main/reg11.do and \$main/reg13.do

These 2 files create Table 5

- \$main/data2/citations_received.do

\$main/data2/citations_made.do

\$main/data2/density_team.do

`$main/data2/density14_W.do`

`$main/data2/density_14_3_resaped.do`

`$main/reg20.do`

These files create Tab 6 (col 9,10); tab 7; tab8 (panel b); appendix tab A3, A5, A8

- `$main/reg5.do`

This file creates Appendix Table A2

- `$main/reg36.do`

This file creates Appendix Table A4

- `$main/reg31.do` and `$main/reg30.do`

These two files create Appendix Table A6

- `$main/data2/data_4.do` and `$main/reg4.do` and `$main/reg34.do`

These files create Appendix Table A7

Where to Find Tables and Figures

All the tables and figures are in the directories called tables/ tables2/ to tables36/

The names of the directories do NOT correspond to the table and figure numbers. Here I explain exactly where to find each table and figure included in the paper:

- Table 1

Top panel is in tables2/tables2d.txt

Middle panel is in tables2/tables2c.txt

Bottom panel is in tables2/tables2b.txt

- Table 2

Top panel column 1 to 4 is in tables3/tables2.txt

Bottom panel column 1 to 4 is in tables3/tables3.txt

Top panel - column 5 is in tables22/tables2.txt

Bottom panel - column 5 is in tables22/tables3.txt

- Table 3 is in tables/table1.txt

- Table 4

First stage -- column 1 to 4 is in tables3/tables6.txt

First stage --column 5 is in tables22/tables6.txt

Second stage -- column 1 to 4 is in tables3/tables5.txt

Second stage --column 5 is in tables22/tables5.txt

- Table 5

OLS is in tables11/table20.txt

2SLS is in tables11/table22.txt

First stage is in tables11/table21.txt

F-stat are in Main.log (search for "XXX FTEST")

- Table 6

Top panel -- rows 1 to 8 is in tables/table4.txt

Top panel -- rows 9 and 10 is in tables20/table6.txt

Bottom panel -- rows 1 to 8 is in tables/table5.txt

Bottom panel -- rows 9 and 10 is in tables20/table7.txt

- Table 7 is in tables20/table72.txt

- Table 8
Top panel is in tables/table90.txt
Bottom panel is in tables20/table2.txt
- Table 9 is in tables4/table2.txt
- Table 10 is in tables2/tables3d.txt
- Table 11 is in MAIN.log marked by "XXX TABLE 11"
- Table 12
Column 1 is in Main.log marked by "XXX TABLE 12 col 1"
Column 2 is in Main.log marked by "XXX TABLE 12 col 2"
- Appendix Table A1 is in tables2/table4.txt
- Appendix Table A2 the examples are in Main.log. Search for "XXX SPILLOVERS"
- Appendix Table A3 is in tables20/table8.txt
- Appendix Table A4
First stage is in tables36/table21.txt
2SLS is in tables36/table22.txt
- Appendix Table A5 is in tables20/table3.txt
- Appendix Table A6
Top panel is in tables31/table1.txt

Bottom panel is in tables30/table1.txt

- Appendix Table A7

Top panel is in tables4/table9.txt

Bottom panel is in tables34/table9.txt

- Appendix Table A8 is in tables20/table87.txt

- Figure 1 is in tables2/fig7.gph

tables2/fig8.gph

tables2/fig9.gph

- Figure 2

Top panel is in tables/kodak_stock.gph

Bottom panel is in tables3/fig1.gph

- Figure 3 in in tables3/fig11.gph

- Figure 4 is in tables2/fig4.gph

- Figure 5 top panel is in tables/fig1.gph

The bottom panel is in tables/fig1b.gph

- Figure 6 is in tables23/fig1_m.gph and fig1b_m.gph

- Appendix Figure A1 is in tables/fig2.gph and fig2b.gph

List of tables and programs

The provided code reproduces:

- ☐ X All numbers provided in text in the paper
- ☐ X All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

References

- BEA, U.S. Bureau of Economic Analysis. 2012. Table “2007, 2012: 405 Industries.” Available from <https://www.bea.gov/industry/input-output-accounts-data> (accessed Sept 1, 2019).
- Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). "The NBER Patent Citation Data File" NBER Working Paper 8498. Available at <https://www.nber.org/research/data/us-patents> (accessed on Sept 20, 2019)
- Johnson Kenneth P. “Redefinition of the BEA Economic Areas” (2004) downloaded from https://www.csus.edu/indiv/j/jensena/sfp/ea_desc.htm (accessed on April 3, 2019.)
- Moretti, Enrico, 2021, "Replication Data for ""Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority""", <https://doi.org/10.7910/DVN/G1UYL0>, Harvard Dataverse, V1.
- moretti, enrico. county - economic area crosswalk. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-15. <https://doi.org/10.3886/E140581V1>
- Zolas, Nikolas “United States Patent Classification (USPC) Crosswalks - Version 1610” (2016). <https://sites.google.com/site/nikolaszolas/PatentCrosswalk>

- Zucker, Lynne G. and Michael R. Darby, Connecting Outcome Measures in Entrepreneurship Technology and Science (COMETS) database, public-use version COMETS 2.0 at <https://kauffman.org/entrepreneurship/research/comets/>, published to the web February 26, 2014.
- 1stock1.com, Table with Kodak Prices (2017), http://www.1stock1.com/1stock1_173.htm (Accessed on Sept 30, 2017.)