Replication repository for Bartik Instruments: What, When, Why and How

Summary

Software requirements: Stata and SAS (for constructing the Card immigration analysis dataset).

This code replicates the figures and tables from Goldsmith-Pinkham, Sorkin and Swift (2020). The main file for rerunning the code can be run using master.do. The individual do-files are outlined below. The do-files use finalized datasets, which are constructed from various data sources, outlined below.

- The canonical Bartik analysis (BAR) is replicated using Census and ACS data from IPUMS (Ruggles et al., 2015) and uses cross-walks generously provided by David Dorn on his website.
- The China shock analysis (ADH) is replicated using a combination of data sources:
  - the replication file (Autor, Dorn and Hanson, 2013*b*) from (Autor, Dorn and Hanson, 2013*a*). Specifically, the BHJ files below draw (in part) from these replication files
  - the replication data (Borusyak, Hull and Jaravel, 2019*b*) from (Borusyak, Hull and Jaravel, 2019*a*). Specifically, `data/Lshares.dta` and `data/shocks.dta` come from these files (under `shift-share/ADH/Data/`)
  - and unpublished data (Adao, Kolesar and Morales, 2019*a*) from (Adao, Kolesar and Morales, 2019*b*) that was generously provided through personal correspondence with Michal Kolesár. Specifically, `data/ADHdata_AKM.csv` file comes from Michal.
- The Card immigration analysis (CARD) is replicated using replication code provided by David Card from (Card, 2009) and data from the 1990 Census and 2000 Census) provided through the ICPSR.
  - The construction of this dataset requires the use of SAS, but the final analysis dataset is in Stata.

Code process

The `master.do` file executes the following code:

1. `do make_BAR_table.do` constructs Table 3 from the paper and uses `input_BAR2.dta`, the finalized Bartik analysis file. [NOTE: This code is slow due to bootstrapping.]
2. `make_rotemberg_summary_BAR.do` constructs Table 1, Figure 1, and Appendix Figure A5. It uses `input_BAR2.dta`, the finalized Bartik analysis file.
3. `make_rotemberg_summary_BAR_appendix.do.do` constructs Appendix Table A4. It uses `input_BAR2.dta`, the finalized Bartik analysis file.
4. `make_char_table_BAR.do` constructs Table 2. It uses `input_BAR2.dta`, the finalized Bartik analysis file.
5. `do make_ADH_table.do` constructs Table A3 from the paper and uses `ADHdata_AKM.csv`, `Lshares.dta` and `shocks.dta`. [NOTE: This code is slow due to bootstrapping.]
6. `make_rotemberg_summary_ADH.do` constructs Table A1, Appendix Figure A2 and Appendix Figure A3. It uses uses `ADHdata_AKM.csv`, `Lshares.dta` and `shocks.dta`.
7. `make_pretrends_ADH.do` makes Appendix Figure A1 and Appendix Figure A4. It uses `workfile_china_preperiod.dta`, `ADHdata_AKM.csv`, `Lshares.dta` and `shocks.dta`.
8. `make_char_table_ADH.do` constructs Table A2. It uses uses `ADHdata_AKM.csv`, `Lshares.dta` and `shocks.dta`.
9. `make_CARD_table_hs.do` and `make_CARD_table_college.do` make Table 6. They use `input_card.dta`.
10. `make_rotemberg_summary_CARD_hs.do` and `make_rotemberg_summary_CARD_college.do` make Table 4, Figure 4 and Appendix Figure A6. They use `input_card.dta`.
11. `make_char_table_CARD.do` makes Table 5. It uses `input_card.dta`.
12. `make_pretrends_CARD.do` makes Figures 2 and 3. It uses `input_card.dta`.

See `file-name_to_exhibit_map.txt` for a mapping between file names and the exhibit names.

Data construction for canonical Bartik

The following steps below allow researchers to recreate `input_BAR2.dta` themselves.

The file is created using two do-files:

1. `create_bartik_data.do`, which creates `Characteristics_CZone.dta` and `shares_long_ind3_czone.dta`, and takes nine inputs:
    1. `IPUMS_data.dta`

2. IPUMS_ind1990.dta
3. IPUMS_geo.dta
4. IPUMS_bpl.dta
5. cw_ctygrp1980_czone_corr.dta
6. cw_puma1990_czone.dta
7. cw_puma2000_czone.dta
8. czone_list.dta

2. make_input_bar.do, which creates input_BAR2.dta and takes two inputs:
    1. Characteristics_CZone.dta
    2. shares_long_ind3_czone.dta

These files are described in further detail below:

IPUMS_data.dta

Our large base dataset downloaded from IPUMS here: https://usa.ipums.org/usa/data.shtml
Note that of the 2009-2011 ACS samples were pooled to form the 2010 sample.

Samples:

1. 1980 5% state;
2. 1990 5%;
3. 2000 5%;
4. 2009 ACS; 2010 ACS; 2011 ACS

Variables:

```
year; datanum; serial; hhwt; statefip; conspuma;  cpuma0010;
gq; ownershp; ownershpd; mortgage; mortgag2; rent; rentgrs;
hhincome; foodstmp; valueh; nfams; nsubfam; ncouples; nmothers;
nfathers; multgen; multgend; pernum; perwt; famsize; nchild;
nchlt5; famunit; eldch; relate; related; sex; age; marst; birthyr;
race; raced; hispan; hispand; ancestr1; ancestr1d; ancestr2;
ancestr2d; citizen; yrsusa2; speakeng; racesing; racesingd;
school; educ; educd; gradeatt; gradeattd; schltype; empstat;
empstatd; labforce; occ; ind; classwkr ; classwkrd; wkswork2;
uhrswork; wrklstwk; absent; looking; availble; wrkrecal; workedyr;
inctot; ftotinc: incwage; incbus00; incss; incwelfr; incinvst;
incretir; incsupp; incother; incearn; poverty; occscore; sei;
hwsei; presgl; prent; erscor90; edscor90; npboss90; migrate5;
migrate5d; migrate1; migrate1d; migplac5; migplac1; movedin;
vetstat; vetstatd; pwstate2; trantime
```

`IPUMS_ind1990.dta`

An additional dataset of 1990 standardized industries to merge onto the main dataset, again downloaded here: https://usa.ipums.org/usa/data.shtml Note that in the ACS samples, 2009-2011 were pooled to form the 2010 sample. Merging with the main dataset occurred by matching year-serial-pernum.

Samples:

1. 1980 5% state;
2. 1990 5%;
3. 2000 5%;
4. 2009 ACS; 2010 ACS; 2011 ACS

Variables:

`year; datanum; serial; hhwt; gq; pernum; perwt; ind1990`

`IPUMS_geo.dta`

An additional dataset of geographies to merge onto the main dataset, again downloaded here: https://usa.ipums.org/usa/data.shtml

Samples:

1. 1980 5% state;
2. 1990 5%;
3. 2000 5%;
4. 2009 ACS; 2010 ACS; 2011 ACS

Variables:

`year; datanum; serial; hhwt; gq; pernum; perwt; county; countyfips; cntygp98; puma`

`IPUMS_bpl.dta`

An additional dataset of birthplace to merge onto the main dataset, again downloaded here: https://usa.ipums.org/usa/data.shtml

Samples:

1. 1980 5% state;
2. 1990 5%;
3. 2000 5%;
4. 2009 ACS; 2010 ACS; 2011 ACS

Variables:

`year; datanum; serial; hhwt; gq; pernum; perwt; bpl`

Data construction for Card (2009)

1980

1. `read80.do` - reads the state-specific files of the 1980 5% extracts (available from ICPSR), does minimal data cleaning, merges all state-specific files. The output is `all80.dta`. Takes as input:
    i. Census of Population and Housing, 1980 [United States]: Public Use Microdata Sample (A Sample): 5-Percent Sample (ICPSR 8101). Download it here: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8101/summary.
2. `read_all80.sas` - creates `all80.sas7bdat`. Takes as input `all80.dta`.
3. Run the scripts provided by Card.
    i. `np2.sas` - creates a working data set of wage-earners age 18+, with recodes, etc. This is `np80.sas7bdat`. These data are used to build wage outcomes. Takes as input `all80.sas7bdat`. *reads the code in `smsarecode80.sas` to re-code msa's.
    ii. `allnp2.sas` - creates a working data set of EVERYONE age 18+, with recodes, etc. This is `supp80.sas7bdat`. These data are used to build supply variables. Takes as input `all80.sas7bdat`. *reads the code in `smsarecode80.sas` to re-code msa's.
    iii. `cell1.sas` - creates a big summary of data by cell ==> `bigcells.sas7bdat`. Takes as input `np80.sas7bdat`.
    iv. `t1.sas`- creates a big summary of data by cell ==> `allcells.sas7bdat`. Takes as input `supp80.sas7bdat`.
    v. `supply1.sas` - gets supply measures ==> `cellsupply.sas7bdat`. Takes as input `np80.sas7bdat`.

vi. `imm1.sas` - gets counts of immigrants by sending country in each city ==>`ic_city.sas7bdat` (IC is Card's classification of sending countries). Takes as input 'supp80.sas7bdat.

vii. `indist.sas` - gets fraction of workers in manufacturing by city. Takes as input `np80.sas7bdat`.

4. Export some datasets to Stata:

   i. `cell1_to_stata.sas` - creates datasets on wages of immigrants and natives by education class. Exports them to Stata (`1980_bigcells_new1.dta`, `1980_bigcells_new2.dta`, `nw80.dta`, `iw80.dta`, `nw801.dta`, `nw802.dta`, `nw803.dta`, `nw804.dta`, `iw801.dta`, `iw802.dta`, `iw803.dta`, `iw804.dta`). Takes as input `bigcells.sas7bdat`.

   ii. `t1_to_stata.sas` - creates `1980_allcells_new2.dta`. Takes as input `allcells.sas7bdat`

   iii. `indist_to_stata.sas` - creates `1980_mfg.dta`. Takes as input `mfg.sas7bdat`

1990

1. `read90.do` - reads the state-specific files of the 1990 5% extracts (available from ICPSR), does minimal data cleaning, merges all state-specific files. The output is `all90.dta`. Takes as input:

   i. Census of Population and Housing, 1990 [United States]: Public Use Microdata Sample: 5-Percent Sample (ICPSR 9952). Download it here: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/9952.

2. `read_all90.sas` - creates `all90.sas7bdat`. Takes as input `all90.dta`.

3. Run the scripts provided by Card.

   i. `np2.sas` - creates a working data set of wage-earners age 18+, with recodes, etc. This is `np90.sas7bdat`. These data are used to build wage outcomes. Takes as input `all90.sas7bdat`. *reads the code in `smsarecode90.sas` to re-code msa's.

   ii. `allnp2.sas`- creates a working data set of EVERYONE age 18+, with recodes, etc. This is `supp90.sas7bdat`. These data are used to build supply variables. Takes as input `all90.sas7bdat`. *reads the code in `smsarecode90.sas` to re-code msa's.

   iii. `cell1.sas` - creates a big summary of data by cell ==> `bigcells.sas7bdat`. Takes as input `np90.sas7bdat`.

   iv. `t1.sas`- creates a big summary of data by cell ==> `allcells.sas7bdat`. Takes as input `supp90.sas7bdat`.

v. `supply1.sas` - gets supply measures ==> `cellsupply.sas7bdat`. Takes as input `np90.sas7bdat`.

vi. `imm1.sas` - gets counts of immigrants by sending country in each city ==>`ic_city.sas7bdat` (IC is Card's classification of sending countries). Takes as input 'supp90.sas7bdat.

vii. `indist.sas` - gets fraction of workers in manufacturing by city. Takes as input `np90.sas7bdat`.

4. Export some datasets to Stata:

   i. `cell1_to_stata.sas` - creates datasets on wages of immigrants and natives by education class. Exports them to Stata (`1990_bigcells_new1.dta`, `1990_bigcells_new2.dta`, `nw90.dta`, `iw90.dta`, `nw901.dta`, `nw902.dta`, `nw903.dta`, `nw904.dta`, `iw901.dta`, `iw902.dta`, `iw903.dta`, `iw904.dta`). Takes as input `bigcells.sas7bdat`.

   ii. `t1_to_stata.sas` - creates `1990_allcells_new2.dta`. Takes as input `allcells.sas7bdat`

   iii. `indist_to_stata.sas` - creates `1990_mfg.dta`. Takes as input `mfg.sas7bdat`

2000

1. `read2000.do` - reads the state-specific files of the 2000 5% extracts (available from ICPSR), does minimal data cleaning, merges all state-specific files. The output is `all2000.dta`. Takes as input:

   i. Census of Population and Housing, 2000 [United States]: Public Use Microdata Sample: 5-Percent Sample (ICPSR 13568). Download it here: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/13568.

2. `read_all2000.sas` - creates `all2000.sas7bdat`. Takes as input `all2000.dta`.

3. Run the scripts provided by Card.

   i. `np2.sas` - creates a working data set of wage-earners age 18+, with recodes, etc. This is `np2000.sas7bdat`. These data are used to build wage outcomes. Takes as input `all2000.sas7bdat`.

   ii. `allnp2.sas`- creates a working data set of EVERYONE age 18+, with recodes, etc. This is `supp2000.sas7bdat`. These data are used to build supply variables. Takes as input `all2000.sas7bdat`.

   iii. `cell1.sas` - creates a big summary of data by cell ==> `bigcells.sas7bdat`. Takes as input `np2000.sas7bdat`.

iv. `t1.sas`- creates a big summary of data by cell ==> `allcells.sas7bdat`. Takes as input `supp2000.sas7bdat`.

v. `supply1.sas` - gets supply measures ==> `cellsupply.sas7bdat`. Takes as input `np2000.sas7bdat`.

vi. `imm3.sas` - gets counts of immigrants by sending country in each city ==> `ic_citynew.sas7bdat` (IC is Card's classification of sending countries). Takes as input `supp2000.sas7bdat`.

vii. `imm2.sas` - gets a count of immigrants present in 2000 by IC - this is used to construct the instrumental variable ==> `byicnew.sas7bdat`. Takes as input `supp2000.sas7bdat`.

viii. `inflow3.sas` - constructs the supply push instrument by "education and experience cell" and city. This is `newflows.sas7bdat`. Takes as input `ic_city.sas7bdat` (output of `imm1.sas' in 1980) and`byicnew.sas7bdat`(output of`imm2.sas' in 2000).

4. Export some datasets to Stata:

   i. `cell1_to_stata` - creates datasets on wages of immigrants and natives by education class. Exports them to Stata (`2000_bigcells_new1.dta`, `2000_bigcells_new2.dta`, `nw.dta`, `iw.dta`, `nw.dta`, `nw.dta`, `nw.dta`, `nw.dta`, `iw.dta`, `iw.dta`, `iw.dta`, `iw.dta`). Takes as input `bigcells.sas7bdat`.

   ii. `t1_to_stata` - creates `2000_allcells_new1.dta` and `2000_allcells_new2.dta`. Takes as input `allcells.sas7bdat`.

   iii. `inflow3_to_stata` - exports 'newflows.sas7bdat' to dta.

Replicate Table 6 of Card (2009) and construct input dataset for Bartik analysis

1. `table6.do` - replicates Table 6 of Card (2009) and constructs the dataset `input_card.dta`. Takes as input the Stata datasets exported from SAS (cited above) for 1980, 1990, and 2000.

References

Adao, Rodrigo, Michal Kolesar, and Eduardo Morales. 2019*a*. "Replication data for: Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics [publisher]*. Unpublished data provided in personal communication by Michal Kolesar [distributor].

Adao, Rodrigo, Michal Kolesar, and Eduardo Morales. 2019*b*. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics*, 134(4): 1949–2010.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013*a*. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–2168.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013*b*. "Replication data for: The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Association [publisher]*, https://doi.org/https://doi.org/10.3886/E112670V1. Inter-university Consortium for Political and Social Research [distributor].

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019*a*. "Quasi-experimental Shift-share Research Designs." Working Paper.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019*b*. "Replication data for: Quasi-experimental Shift-share Research Designs." *Github repository, last accessed February 22, 2020*. https://github.com/borusyak/shift-share.

Card, David. 2009. "Immigration and Inequality." *American Economic Review: Papers and Proceedings*, 99(2): 1–21.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]*. Minneapolis: University of Minnesota.