

The code in this replication package constructs the main analysis files for Borusyak, Hull, and Jaravel (2021) from several primary sources using Stata. Analyses are then conducted in Stata and R. We separate the package into three subpackages:

- The primary Autor et al. (2013, **ADH**) application
- The supplementary **Bartik (1991)** application in the paper's appendix,
- The **Monte Carlo** simulations in the paper's appendix.

The replicator should expect the ADH code to run for less than 15 minutes; it will produce two figures and eight table files. The replicator should expect the Bartik code to run for around two days and produce one table file. The replicator should expect the Monte Carlo code to run for around one day and produce three table files.

Data Availability and Provenance Statement

ADH

Data are combined from several primary sources and are provided as part of this archive (in *ADH/Raw*) as part of the public domain. The sources are as follows:

- Data and code provided by David Dorn (with permission to include in our replication archive):
 - *czone_industry_1980.dta, czone_industry_1990.dta, czone_industry_2000.dta*: industry-by-commuting zone employment data based on CBP
- Data provided by Gordon Hanson (with permission to include in our replication archive):
 - *year1991.dta, year2000.dta, year2007.dta*: ADH trade data
 - *hs6_sic_gen_all_95_05_new.dta*: concordance from HS6 to SIC codes
- Data from the ADH replication archive (package P3 at <https://www.ddorn.net/data.htm>)
 - *workfile_china.dta, workfile_china_preperiod.dta*: commuting zone level files
- Data from the Acemoglu et al. (2016) replication archive (package P6 at <https://www.ddorn.net/data.htm>)
 - *dta/sample/main.dta* renamed into *aadhp_main.dta*: main analysis file
- NBER-CES Manufacturing Industry Database (1987): can be downloaded at <https://data.nber.org/nberces/nberces5811/sic5811.dta>: *sic5811.dta*

Data files in the *ADH/Data/*, *ADH/Temp/*, and *ADH/Results/* folders are produced by the code described below. We include these files only to facilitate replication.

Bartik (1991)

Data come from the Goldsmith-Pinkham et al. (2020, GPSS) replication archive (https://github.com/paulgp/gpss_replication/tree/master/data/raw) and from IPUMS USA (Ruggles et al., 2020).

We include copies of the GPSS data, which are in the public domain, in this archive (in *Bartik 1991/Raw*). These files are as follows:

- *czone_list.dta*: List of commuting zone IDs
- *cw_puma1990_czone.dta*, *cw_puma2000_czone.dta*: commuting zone to public use microdata area crosswalks, 1990 and 2000
- *cw_ctygrp1980_czone_corr.dta*: commuting zone to city group crosswalk, 1980

The following data extracts from IPUMS (<https://usa.ipums.org/usa/>) will need to be added to the *Bartik 1991/Raw* folder, as described in the GPSS replication archive: *IPUMS_data.dta*, *IPUMS_ind1990.dta*, *IPUMS_geo.dta*, *IPUMS_bpl.dta*. Each extract should be included from the following samples: 1980 5% state; 1990 5%; 2000 5%; 2009 ACS; 2010 ACS; 2011 ACS. The variables that should be included in each extract are as follows:

- *IPUMS_data.dta*: 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; hispan; 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; wrkreca; 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*: year; datanum; serial; hhwt; gq; pernum; perwt; ind1990
- *IPUMS_geo.dta*: year; datanum; serial; hhwt; gq; pernum; perwt; county; countyfips; cntygp98; puma
- *IPUMS_bpl.dta*: year; datanum; serial; hhwt; gq; pernum; perwt; bpl

Data files in the *Bartik 1991/Data/* and *Bartik 1991/Results/* folders are produced by our code described below. We include these files only to facilitate replication.

Monte Carlo

This subpackage uses data from the *ADH/Data* folder, which is produced from the programs described below, plus an additional raw dataset *cbp_czone_merged.dta* from the Acemoglu et al. (2016) replication archive, which is in the public domain and can be downloaded from package P6 at <https://www.ddorn.net/data.htm>. We provide it as part of this archive (in *MonteCarlo/Raw*). It contains industry-by-commuting zone employment.

Data files in the *MonteCarlo/Data/* and *MonteCarlo/Results/* folders are produced by the code described below. We include these files only to facilitate replication.

Statement about Rights

We certify that we, the authors of the manuscript, have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

All data in this replication archive are publicly available.

Computational requirements

Each of the three subpackages contains a “master” Stata file which executes the primary data cleaning and analysis. Supplemental analyses are then conducted in R. We give a description of programs in the Instructions to Replicators, below.

Software Requirements

- Stata (code was last run with version 16.1), with packages:
 - `distinct` (as of 2018-05-12)
 - `estout` (as of 2020-02-04)
 - `avar` (as of 2018-05-12)
- R 4.0.3, with packages:
 - VGAM 1.1-5
 - AER 1.2-9

The *ADH/Code* folder contains additional user-generated packages used in the **ADH** and **Monte Carlo** procedures:

- *Code/iv.R*: shift-share IV regression code from the replication package of Adão et al. (2019). Available at <https://github.com/kolesarm/ShiftShareSE>
- *Raw/subfile_sic87dd.do*: code to aggregate some SIC4 industries as in ADH. Provided to us by David Dorn with permission to include in this replication archive
- *Code/ssaggregate.ado* (with help file *ssaggregate.sthlp*): code written by us to aggregate data from the observation level to the shock level for running equivalent shock-level IV regressions. Written by the authors
- *Code/ssiv_null_imposed.ado*: code written by us to compute standard errors with the null imposed for equivalent shock-level regressions. Written by the authors
- *Code/binscatter2.ado*: our modification of Michael Stepner’s *binscatter.ado* (available at <https://michaelstepner.com/binscatter/>), which adds extra functionality
- *Code/weakssivtest.ado*: our modification of *weakivtest* by Pflueger and Wang (2015) for shift-share IV regressions

Memory and Runtime Requirements

The code was last run on a 4-core Intel-based laptop with Windows 10 version 1909. Approximate time needed to reproduce the analyses on a standard (2021) desktop machine is as follows:

- **ADH**: 15 minutes. **Bartik (1991)**: two days. **Monte Carlo**: one day

Instructions to Replicators

We provide instructions separately for each subpackage.

ADH

1. Edit both the *adh* global in line 1 of *Code/master_adh.do* and the *setwd()* command in line 1 of *Code/TableC2_akm_se* to the local directory.
2. Run *Code/master_adh.do* in Stata. This will run the following programs:
 - *Code/build_data.do* will reconstruct the ADH analysis file.
 - *Code/Transformation.do* will convert the ADH analysis file into an industry-level file, following the numerical equivalence in the paper.
 - *Code/Table1.do*, *Code/Table2.do*, *Code/Table3.do*, *Code/Table4_C1_C2_C4.do*, *Code/TableC3.do*, *Code/TableC51.do*, *Code/FigureC1.do* will produce the corresponding tables and figures (except for the Adao et al. (2019) confidence intervals for Table C2).
 - *Code/export_to_r.do* will prepare data for the R code.
3. Run *Code/TableC2_akm_se.R* in R: This will produce the Adao et al. (2019) confidence intervals for Table C2.

This subpackage was last run top-to-bottom on February 4, 2021.

Bartik (1991)

1. Edit the *main* global in line 1 of *Code/master_barik1991.do* to the local directory
2. Run *Code/master_barik1991.do* in Stata. This will run the following programs:
 - *Code/create_bartik_data.do*, *Code/make_input_bar.do*, and *Code/prepare_gpss_data.do* will create three intermediate data files and one final analysis file.
 - *Code/Table_C6.do* will create Table C6.

This subpackage was last run top-to-bottom on February 4-5, 2021.

Monte Carlo

1. Edit the *main* global in line 1 of *Code/master_montecarlo.do* and the *setwd()* commands in line 2 of *Code/TableC8_gen_data.R* and *Code/TableC7.R* to the local directory. Make sure that the ADH replication folder is available in *../ADH*.
2. Run *Code/master_montecarlo.do* in Stata up to the “stop” command. This will run the following programs:
 - *Code/import_aadhp.do* will create an intermediate analysis file based on Acemoglu et al. (2016).
 - *Code/adh_aggregate_sic3.do* and *Code/adh_aggregate_sic2.do* will create intermediate analysis files based on ADH.
 - *Code/export_data.do* will produce the final analysis file.
3. Run *Code/TableC7.R* in R. This will create Table C7.
4. Run *Code/TableC8_gen_data.R* in R. This will create the data for Table C8.
5. Run the remaining part of *Code/master_montecarlo.do* after the “stop” command. This will run *Code/TableC8.do*, which creates Table C8.

This subpackage was last run top-to-bottom on February 5-6, 2021.

List of Tables and Programs

Figure/Table #	Program	Line #	Output file
Table 1	<i>ADH/Code/Table1.do</i>	115	<i>ADH/Results/Table1.dta</i>
Table 2	<i>ADH/Code/Table2.do</i>	15	<i>ADH/Results/Table2.smcl</i>
Table 3	<i>ADH/Code/Table3.do</i>	15-16, 46-47	<i>ADH/Results/Table3A.csv</i> , <i>ADH/Results/Table3B.csv</i>
Table 4	<i>ADH/Code/Table4_C1_C2_C4.do</i>	82-86	<i>ADH/Results/Table4_C1_C2_C4.csv</i>
Figure C1	<i>ADH/Code/FigureC1.do</i>	4, 7	<i>ADH/Results/FigureC1 FS.pdf</i> , <i>ADH/Results/FigureC1 RF.pdf</i>
Table C1	<i>ADH/Code/Table4_C1_C2_C4.do</i>	82-86	<i>ADH/Results/Table4_C1_C2_C4.csv</i>
Table C2	<i>ADH/Code/Table4_C1_C2_C4.do</i> <i>ADH/Code/Table2_akm_se.R</i>	82-86	<i>ADH/Results/Table4_C1_C2_C4.csv</i> , <i>ADH/Results/TableC2_akm_se.csv</i>
Table C3	<i>ADH/Code/TableC3.do</i>	11	<i>ADH/Results/TableC3.do</i>
Table C4	<i>ADH/Code/Table4_C1_C2_C4.do</i>	82-86	<i>ADH/Results/Table4_C1_C2_C4.csv</i>
Table C5	<i>ADH/Code/TableC5.do</i>	28	<i>ADH/Results/TableC5.csv</i>
Table C6	<i>Bartik 1991/Code/TableC6.do</i>	87	<i>Bartik 1991/Results/TableC6.csv</i>
Table C7	<i>MonteCarlo/Code/TableC7.R</i>	36, 90	<i>MonteCarlo/Results/AADHP.csv</i> , <i>MonteCarlo/Results/ADH.csv</i>
Table C8	<i>MonteCarlo/Code/TableC8.do</i>	90-103	<i>MonteCarlo/Results/TableC8.smcl</i>

Acknowledgements

We thank David Dorn and Gordon Hanson for providing some of the code and data in the Autor et al. (2013) package. We thank Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift for providing some of the code and data in the Bartik (1991) package.

References

- Acemoglu, Daron, David H. Autor, David Dorn, Gordon H. Hanson, and Brendard Price. 2016. "Import Competition and the Great U.S. Employment Sag of the 2000s." *Journal of Labor Economics* 34:S141–S198
- Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *The Quarterly Journal of Economics* 134:1949–2010
- Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103:2121–2168.

Bartik, Timothy J. 1991 *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110:2586-2624

Pflueger, Carolin E. and Su Wang. 2015. "A Robust Test for Weak Instrument in Stata." *Stata Journal* 15:216-225

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Steven M. Manson. 2020. "IPUMS USA: Version 10.0 [dataset]." Minneapolis, MN: *Minnesota Population Center, IPUMS*. <https://doi.org/10.18128/D010.V10.0>