U.S. Job Flows and the China Shock

Data and Programs

Brian Asquith, Sanjana Goswami, David Neumark, and Antonio Rodriguez-Lopez

August 2018

This document describes the Stata programs and datasets used to create the results reported in the paper.

Basic overview of the folders:

- The folder "do/build/national" contains the programs used to construct the industry-level sample used in our national analysis.
- The folder "do/build/regional" contains the programs used to construct the commuting zone-level sample used in our analysis of local labor markets.
- The folder "do/estimate" contains scripts that perform all of the analyses reported in the paper (tables, figures, and appendix).
- The folder "data" contains some raw and cleaned data files prepared by the scripts in do/build. It also contains certain data files taken from Acemoglu, Autor, Dorn, Hanson, and Price (2016) and Pierce and Schott (2016).
- The folder "out" contains all analytical output.
- The folder "xwalks" contains the geographic and industry crosswalks used in the paper. These have been taken from either Acemoglu, Autor, Dorn, Hanson, and Price (2016) or Pierce and Schott (2016) and have been indicated as such.

Samples:

Industry-level sample: Our industry-level analyses are conducted at the level of 392 manufacturing and 87 non-manufacturing industries, identified using 4-digit SIC87DD codes. These are the same industries as in Acemoglu, Autor, Dorn, Hanson, and Price (2016).

Commuting zone-level sample: Our regional analysis encompasses 722 commuting zones comprising the whole of the mainland United States (excluding Alaska and Hawaii). Commuting zones are aggregations of counties (Tolbert and Sizer (1996); David, Dorn, and Hanson (2013)).

Data files:

Data sets are available in Stata format in several subfolders in "/data". Wherever data files are pulled directly from Acemoglu, Autor, Dorn, Hanson, and Price (2016), we retain their original file name followed by _AADHP. Similarly, wherever data files are pulled directly from Pierce and Schott (2016), we retain their original file name followed by _PS. Following is a brief description of each data file:

/data/cbp/emp_counts:

- *cbp_czone_emp_total.dta* provides the net employment levels at the commuting zone and industry level from CBP data for the years 1988-2011. The original data file of AADHP is: "/AADHP-JOLE-Sag-Replication-2015-05-13/dta/cbp/county/clean/emp_counts/cbp_czone_merged.dta" and it contains data from 1988, 1991, 1999, 2007 and 2011. We run their "/AADHP-JOLE-Sag-Replication-2015-05-13/do/build/regional/cbp_co_reader.do" to retrieve the remaining years.

/data/cbp/classification:

- *industry_classification_AADHP.dta* classified which industries are tradable, exposed-tradable and non-exposed non-tradable. This data file is taken from "/AADHP-JOLE-Sag-Replication-2015-05-13/dta/cbp/county/clean/classification/industry_classification.dta"

/data/trade/clean:

- *trade_variables.dta* consists of measures of Chinese import exposure at the industry-year level as well as some controls used by AADHP.

/data/trade/raw:

- *main_AADHP.dta* is taken from "/AADHP-JOLE-Sag-Replication-2015-05-13/dta/sample/main.dta". This is the industry-level file they use in their regressions.
- *trade_czone_variables_cbp.dta* consists of all the measures of Chinese import exposure at the commuting zone level weighted by CBP employment.

/data/pntr/clean:

- *spread.dta* consists of the PNTR measure at the industry level.
- *spread_czone.dta* consists of the PNTR measure (7, 8 and 12 year differences) at the commuting zone level.

/data/pntr/raw:

- *gaps_by_naics6_20150722_fam50_PS.dta* is taken from Pierce and Schott (2016): "data_files_aer_2013-1578/gaps_by_naics6_20150722_fam50.dta". This file contains their PNTR measure at the family level.

/data/bartik:

- *bartik_variables.dta* consists of the Bartik shock at the commuting zone level for 7, 8 and 12 year differences weighted by CBP as well as NETS employment.

/data/io/clean/exposure:

- *first_order_exposure.dta* consists of the upstream and downstream measures of first-order import exposure.
- *leontief_exposure.dta* consists of the upstream and downstream measures of full higher-order import exposure.

/data/io/raw/shares:

- *gross_shares_AADHP.dta* consists of share variables needed to construct first-order upstream and downstream measures of import exposure. It is taken from "/AADHP-JOLE-Sag-Replication-2015-05-13/dta/io/clean/shares/gross_shares.dta"
- *leontief_shares_AADHP.dta* consists of share variables needed to construct higher-order upstream and downstream measures of import exposure. It is taken from "/AADHP-JOLE-Sag-Replication-2015-05-13/dta/io/clean/shares/leontief_shares.dta"

/data/controls:

- *mfgsh.dta* consists of the share of manufacturing employment in total employment for both CBP and NETS data

/data/popest:

- *czone_pop_1990_2012_AADHP.dta* is a commuting zone level data set with information on total population and working-age population and is taken from /AADHP-JOLE-Sag-Replication-2015-05-13/dta/popest/clean/czone_pop_1990_2012.dta

Final datasets:

/data/sample:

- *main_industry.dta* is the industry-level sample with 392 manufacturing industries and 87 non-manufacturing industries. It contains CBP data, import exposure and PNTR measures.
- *main_czone.dta* is the commuting zone level sample where all variables are classified into three sectors: exposed, non-exposed tradable and non-exposed non-tradable. It contains CBP data and measures of the China shock.

NETS data required to run the "do/build" and "do/estimate" files:

We cannot provide NETS data since it is proprietary information. In order to run all the do files smoothly, we recommend constructing the following data files:

/data/nets/emp_counts:

- *nets_czone_emp_total.dta* should contain the net employment levels at the commuting zone and industry level from NETS data for the years 1992-2011.

Variable Name	Description
year	1992-2012
sic87dd	479 industries in 4 digit SIC codes
czone	722 commuting zone codes
nets_emp	Total employment level for the corresponding year

/data/nets/national:

- *nets_industry.dta* should contain total employment, job flows data (births, deaths, expansions and contractions) and their decomposition into the three productivity terciles at the 4-digit "SIC87DD' industry-level. These variables should be the 7, 8 and 12-year differences derived from NETS establishment-level data as follows:

Variable Name	**Description**
year	1992-2012
sic87dd	479 industries in 4 digit SIC codes
nets_emp_total	Total employment level for the corresponding year
nets_birth_empchange_7	The 7 year change in employment in New establishments in 1999
nets_death_empchange_7	The 7 year change in employment in Dead establishments in 1999
nets_exp_empchange_7	The 7 year change in employment in establishments with Employment Expansions in 1999
nets_con_empchange_7	The 7 year change in employment in establishments with Employment Contractions in 1999
nets_birth_empchange_7_t1	The 7 year change in employment in the first productivity tercile of New establishments in 1999
nets_birth_empchange_7_t2	The 7 year change in employment in the first productivity tercile of New establishments in 1999
nets_birth_empchange_7_t3	The 7 year change in employment in the third productivity tercile of New establishments in 1999

  Similarly, *nets_birth_empchange_8* would be the 8 year change in employment in New establishments in 2007 and *nets_birth_empchange_12* would be the 12 year change in employment in New establishments in 2011. The other job flows should be constructed in a similar fashion and the same approach would follow for the terciles variables.

- *nets_czone_industry.dta* should contain total employment and job flows data (7, 8 and 12-year differences) at the commuting zone-industry level. There are 722 commuting zones and 479 industries (392 manufacturing and 87 non-manufacturing industries). The variables should be named exactly as described above.

/data/nets/county:

- *nets_czone.dta* should have total employment and job flows data (7, 8 and 12-year differences) at the commuting zone level. The variables should be named exactly as described above.

- *nets_czone_terciles.dta* should have total employment and job flows data decomposed into three productivity-based terciles at the commuting zone-sectoral level.

/data/trade/raw:

- *trade_czone_variables_nets.dta* should consist of all the measures of Chinese import exposure at the commuting zone level weighted by NETS employment. This file can be constructed by simple running "do/build/regional/nets_trade_czone_builder.do".

References

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): "Import competition and the great US employment sag of the 2000s," *Journal of Labor Economics*, 34(S1), S141–S198.

David, H., D. Dorn, and G. H. Hanson (2013): "The China syndrome: Local labor market effects of import competition in the United States," *The American Economic Review*, 103(6), 2121–2168.

Pierce, J. R., and P. K. Schott (2016): "The surprisingly swift decline of US manufacturing employment," *The American Economic Review*, 106(7), 1632–1662.

Tolbert, C. M., and M. Sizer (1996): "US commuting zones and labor market areas: A 1990 update," .