

Data for “The China Syndrome: Local Labor Market Effects of Import Competition in the United States”

David H. Autor
MIT and NBER

David Dorn
CEMFI and IZA

Gordon H. Hanson
UCSD and NBER

January 9, 2013

Data files (folder */dta*)

The online data archive contains the following data sets in Stata format:

- *workfile_china.dta* provides data for 722 Commuting Zones times 2 time periods (1990-2000, 2000-2007). All variables with the prefix “*l_*” refer to start-of-period values while variables with the prefix “*d_*” correspond to 10-year equivalent changes. This file is the basis for all Tables in the paper except Tables 1 and 2.
- *workfile_china_preperiod.dta* provides a reduced set of variables for 722 Commuting Zones times 4 time periods (1970-1980, 1980-1990, 1990-2000, 2000-2007). This data is used for Table 2.
- *workfile_china_long.dta* provides a reduced set of variables for 722 Commuting Zones over the long period 1990-2007 (used to create Figure 2).
- *figure1_data.dta* contains time-series data for Chinese import penetration and U.S. manufacturing employment, 1987-2007 (used to create Figure 1).
- *import_levels.dta* and *export_levels.dta* contain data for U.S. aggregate imports and exports in 1991/92, 2000, and 2007, as well as corresponding values for a group of eight other wealthy countries (used to create Table 1).
- *sic87dd_trade_data.dta* contains the basic trade flow data that is used in the paper. Trade flows are differentiated by exporter, importer, *sic87dd* industry, and year (1991-2007).

Computation of results (folder */do*)

The following Stata do-files compute the results of the paper:

- *czone_analysis_ipw_final.do* computes results for Tables 3 to 10, and Appendix Tables 1 to 5.
- *czone_analysis_preperiod_final.do* computes results for Table 2.
- *import_stats_final.do* computes descriptive statistics for Table 1.
- *figure1.do* plots Figure 1.
- *czone_plot_import_long_final.do* plots Figure 2.

Raw results (folders */log* and */gph*)

The folders */log* and */gph* contain all regression log files and graphs that are produced by the files in the folder */do*.

Tables and figures (folder */tab-fig*)

The Excel files *ADH-ChinaSyndrome-AER-2011-1050-Tables.xls* and *ADH-ChinaSyndrome-AER-2011-1050-Figures.xls* provide the full set of tables and figures that are presented in the paper.

Additional documentation (folder */other*)

Construction and documentation of *sic87dd* and *ind1990dd* industry codes:

- *cw_n97_s87.dta*: Weighted crosswalk from *naics97* to *sic87* codes.
- *subfile_sic87dd.do*: Definition of *sic87* codes (based on *sic87*).
- *subfile_ind1990dd.do*: Definition of *ind1990dd* codes (based on *ind1990*).
- *Manufacturing_Industry_Codes.pdf*: Correspondence table for *sic87*, *sic87dd*, and *ind1990dd* codes.

Data cleaners for County Business Patterns data:

- *cbpYYYY_imputations.do*: Data cleaner and imputation file for CBP waves YYYY=1980, 1990, 2000.
- *cbpYYYY_to_cz.do*: Aggregation of CBP data to commuting zone level, years YYYY=1980, 1990, 2000.

See David Dorn's web page for further description of these files.