

Data for “When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men”

David H. Autor
MIT and NBER

David Dorn
University of Zurich and CEPR

Gordon H. Hanson
UCSD and NBER

April 3, 2018

Data files (folder */dta*)

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

- *workfile9014wwd.dta* provides data for 722 Commuting Zones (CZs) times 2 time periods (1990-2000, 2000-2014). This file is the basis for all tabulated regression results. Each CZ is identified by the variable *czone*. The variable *statefip* indicates the state that comprises most of the CZ's population, while the dummy variables *reg_** indicate the Census division to which that state belongs (omitted category New England). Time periods are identified by their start year (*yr*) and an indicator for the second period (*t2*). The weight variable *timepwt24* indicates each CZ's share in the start-of-period population of the 48 U.S. mainland states, multiplied by 1/10 of the period length (10 or 14 years). The variable *d_impusch_p9* captures a CZ's exposure to growing import penetration from China, while *d_impuscm_p9cen* and *d_impuschf_p9cen* and the male and female components of that import shock. The variables *d_impotch*_p9cen_lag* denote the corresponding instruments using other developed countries' imports from China. All other variables with prefix "*d_*" are outcome variables of the regression analysis and are measured as 10-year equivalent changes. All variables with prefix "*l_*" are start-of-period values that are used either for descriptive statistics or as control variables. Finally, variables starting with "*cum_mort*" measure cumulate per-decade mortality.
- *czone9014_percentiles.dta* provides additional data for 722 Commuting Zones (CZs) times 2 time periods (1990-2000, 2000-2014). It contains the initial levels and subsequent changes of all percentiles of the gender-specific distributions of annual wage and salary earnings in a CZ. The underlying earnings distributions include individuals with zero earnings, and positive earnings are inflated to \$2015 using the personal consumption expenditure index.
- *workfile7090wwd.dta* provides data for 722 Commuting Zones (CZs) for two lagged time periods (1970-1980, 1980-1990). This file is used to analyze pre-exposure changes in manufacturing employment (Appendix Table 1, panel III).
- *graph_bygender.dta* and *graph_gap.dta* contain the regression estimates that are plotted in Figure 1.

Computation of results (folder */do*)

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

- *czone_analysis_wwd.do* computes all tabulated regression results and descriptive statistics.
- *czone_analysis_wwd_earnptile.do* computes the regression results in the files *graph_bygender.dta* and *graph_gap.dta*. It uses the auxiliary file *percentiles_help.do*.
- *figure1.do* creates the two panels of Figure 1.

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

The folders */log* and */gph* contain the regression log file created by *czone_analysis_wwd.do* and the graphs created by *figure1.do*.

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

The Excel file *ADH-WWD-AERi-Final-TabsFigs.xlsx* provide the full set of tables and figures that are presented in the paper.

Additional documentation (folder */other*)

We construct a new population-weighted crosswalk to map microdata from the 2013-2015 American Community Survey into Commuting Zones (CZs). This crosswalk relates the Public Use Micro Areas (PUMAs) defined by the 2010 Census to the 1990 CZs:

- *cw_puma2010_czone.dta*

See David Dorn's web page for similar crosswalks based on earlier PUMA definitions. We also construct a crosswalk of industry codes from the 1980, 1990 and 2000 Census to 1987 SIC codes. The first three files below convert the original Census codes to a balanced industry panel whose *ind1990ddx* codes are a slight aggregation of the 1980 and 1990 Census codes, while the 2000 Census codes are matched with a weighted crosswalk. The final file then maps each *ind1990ddx* code to a 4-digit 1987 SIC code:

- *cw_ind1980_ind1990ddx.dta*
- *cw_ind1990_ind1990ddx.dta*
- *cw_ind2000_ind1990ddx.dta*
- *cw_ind1990ddx_sic87.dta*