

Data for “The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market”

David H. Autor
MIT and NBER

David Dorn
CEMFI and IZA

December 2012

Data files (folder */dta*)

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

- *workfile2012.dta* provides data for 722 Commuting Zones times 5 time periods (1950-1970, 1970-1980, 1980-1990, 1990-2000, 2000-2005). All variables with the prefix “*l_*” refer to start-of-period values while variables with the prefix “*d_*” correspond to 10-year equivalent changes.
- *occ1990dd_data2012.dta* provides data for the 330 occupations of the *occ1990dd* occupation panel. Variables are defined for the years 1950, 1970, 1980, 1990, 2000, and 2005.
- *emp-bypctile-1980-2005-rewt-0-*.dta*, *emp-bypctile-1980-2005-rewt-1980-*.dta*, *dwg-byperc-1980-2005-*.dta*, and *occ-means-by-decade-1980-2005-*.dta* provide occupation-level data for polarization graphs, drawing either on the full set of commuting zones (suffix *czall*) or only on commuting zones with high or low routine employment share (suffix *czhigh-abovemean-RSHA* or *czlow-abovemean-RSHA*).
- *plotpoints-svc-by-quintile-1970-2005.dta* and *RTI-bypctile.dta* provide data for employment shares in the bottom quintile of the occupational skill distribution, and for routine intensity by skill percentile.

These data files are the basis for all reported empirical results except Panel B of Table 7 which draws on individual-level Census microdata.

Crosswalks (folder */crosswalks*)

Mapping of Census occupation codes to *occ1990dd* occupation panel:

- *occ1950_occ1990dd.dta*: 1950 Census occupation codes to *occ1990dd* codes.
- *occ1970_occ1990dd.dta*: 1970 Census occupation codes to *occ1990dd* codes.
- *occ1980_occ1990dd.dta*: 1980 Census occupation codes to *occ1990dd* codes.
- *occ1990_occ1990dd.dta*: 1990 Census occupation codes to *occ1990dd* codes.
- *occ2000_occ1990dd.dta*: 2000 Census 5%-sample occupation codes to *occ1990dd* codes.
- *occ2005_occ1990dd.dta*: ACS occupation codes to *occ1990dd* codes.

Other resources relating to *occ1990dd* codes:

- *subfile_occ1990dd_occgroups.do*: aggregation of *occ1990dd* occupations to occupation groups.
- *Dorn_Thesis_Appendix.pdf*, section A1: additional documentation regarding the construction of *occ1990dd* codes.

Mapping of geographic units in Census IPUMS samples to Commuting Zones:

- *cw_sea1950_czone.dta*: 1950 Census State Economic Areas to 1990 Commuting Zones.
- *cw_ctygrp1970_czone.dta*: 1970 Census County Groups to 1990 Commuting Zones.
- *cw_ctygrp1980_czone.dta*: 1980 Census County Groups to 1990 Commuting Zones.
- *cw_puma1990_czone.dta*: 1990 Census Public Use Micro Areas to 1990 Commuting Zones.
- *cw_puma2000_czone.dta*: 2000 Census Public Use Micro Areas (also used in 2005 ACS) to 1990 Commuting Zones.

Note: In each file, the variable *afactor* indicates which fraction of a SEA/County Group/PUMA maps to a given CZ. To allocate observations from Census microdata to Commuting Zones, one has to merge the geographic unit of the Census file to the corresponding Commuting Zone crosswalk using a many-to-many merge (command *joinby* in older versions of Stata). The labor supply weight or person weight from the Census then needs to be multiplied with *afactor*.

Other resources relating to Commuting Zones:

- *cw_cty_czone.dta*: mapping of 1990 FIPS county codes to 1990 Commuting Zones.
- *cw_czone_state.dta*: mapping of 1990 Commuting Zones to FIPS state codes (indicating the state that contains the largest fraction of a CZ's population).
- *cw_czone_division.dta*: mapping of 1990 Commuting Zones to Census Division dummies.
- *Dorn_Thesis_Appendix.pdf*, section A2: additional documentation regarding Commuting Zones.

Computation of results (folder */do*)

The following Stata do-files produce our main results:

- *czone_analysis_final.do* computes the results for Table 3, Table 4, Table 5, Table 6, Table 7 (Panel A), Appendix Table 3, Appendix Table 4 (Panel A), and the two Online Appendix Tables.
- *occ1990dd_stats.do* computes the results for Table 1, Table 2, Appendix Table 1, Appendix Table 2, Appendix Table 4 (Panel B), and the data input for Figure 3.
- *microwage1980_2005_regression.do* computes the results for Table 7 (Panel B). This file draws on individual-level microdata from the 1980 Census and the 2005 ACS which is not contained in this data distribution due to very large file size.
- *plot-demp-bypctile-1980-2005-rewt.do* computes the employment polarization graphs for Figure 1 (Panel A), Figure 2 (Panel A), and Figure 5 (Panel A).

- *calc-plot-dwage-by-occ-pctile-1980-2005.do* computes the wage polarization graphs for Figure 1 (Panel B), Figure 2 (Panel B), and Figure 5 (Panel B).
- *plot-svc-growth-barchart-p20.do* computes the bar chart of Figure 3.
- *calc-plot-rti-by-wage-pctile.do* computes the graph of Figure 4.
- *czone_plot_final.do* computes the scatterplots for Figure 6.

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 *Autor-Dorn-AER-Final-Tables.xls* and *Autor-Dorn-AER-Final-Figures.xls* provide the full set of tables and figures that are presented in the paper and in the online appendix.