DESCRIPTION OF DATA AND COMPUTER CODE

Structural Change and the Fertility Transition

By Philipp Ager and Benedikt Herz

Review of Economics and Statistics (MS 21723)

April 2019

The computer code replicates the results in "Structural Change and the Fertility Transition" by Ager and Herz, published in the Review of Economics and Statistics.

STATA version 15.1 was used to execute the code. Running the code requires a computer with a large amount of internal memory. We used a machine with 128GB of RAM. Please note that the user-written commands spmap, outreg2, and reghdfe (version 5.2.9 06aug2018) need to be installed. The file "*make_all.do*" produces all data files from original data as well as all tables and figures in the main text and the Online Appendix. It will take several hours to run this file. The output of a complete run of "*make_all.do*" can be found in the log-file ./output/log_RESTAT_MS21723.smcl. To only produce the tables and figures in the main text the user can execute the STATA do-file replicationfile_RESTAT_MS21723.do.

Main STATA do-files

Name	Description
replicationfile_RESTAT_MS21723.do	Creates Tables 1, 2, 3, 4, and 5 as well as Figures 1, 3, 4, and 5. Also creates Online Appendix Tables 4 and 7. All tables and figures are saved in the folder "Ager Herz (2019) RESTAT MS21723/output".
replicationfile_onlineappendix_RESTAT_MS21723.do	Creation of the tables and figures in the Online Appendix (except Tables 4 and 7).
replicate_numbers_cited_in_paper.do	Replicates the number and figures quoted in the paper.
make_all.do	Creates the final data sets from raw data by calling secondary STATA do-files. Also calls the three do-files listed above to produce results.

Main STATA data-files

Name	Description	Main variables
fertility.dta	13.5 million married women (16 to 49-year-old) with spouse present	• nchlt5: number of children under age 5 • lhh_income: household income • bw_intensity: Boll Weevil Intensity measure, defined as the product of boll weevil incidence and cotton dependency.
mother_panel.dta	Data set containing mothers' fertility histories	• births: indicator variable that equals 1 if mother gave birth in given year • bw_intensity: see above
structural_change.dta	61 million individuals of working age (10 to 65-year-old)	• manufacturing: indicator variable that equals 1 if individual works in manufacturing • farming: indicator variable that equals 1 if individual works in farming • bw_intensity: see above
schooling.dta	7.5 million children (10 to 15-year-old) who are listed together with their mothers in the Census	• childlabor: indicator variable that equals 1 if child works • regular_school: indicator variable that equals 1 if child regularly attends school • idle: indicator variable that equals 1 if child idle • bw_intensity: see above
linked_sample_males_only_heads.dta	Linked records from the 1880 complete-count database to the one percent samples of the 1900, 1910, and 1920 Censuses	• out_of_farm: indicator variable that equals 1 if individual was on farm in 1880 but not in 1900 (or 1910/1920) • bw_intensity: see above
twoinstruments.dta	County-level data used for Table 4	• nchlt5_2039: number of children under age 5 of 20-39 year olds • nchlt5_2039_ag_hh: number of children under age 5 of 20-39 year olds (agricultural households only) • l_agincome: agricultural income (in logs) • lmanufacturing_cty: share of individuals in manufacturing (in logs)
rosenwald_schools_1920_1930_final.dta	County-level information on Rosenwald schools for 1920 and 1930	• schools_pop: schools per 1,000 inhabitants • teachers_opp: teachers per 1,000 inhabitants • bw_intensity: see above • bw_childlabor: interaction of bw_intensity and indicator that equals 1 if county c's childlabor share in 1910 was above the median

event_study.dta	County-level data set used for the event study shown in Figure 3	• births: number of births by 15-44 year old women in county • medium_relative_t_X: lead/lag of boll weevil arrival in counties with cotton-dependency in 2nd or 3rd quartile • high_relative_t_X: lead/lag of boll weevil arrival in counties with cotton-dependency in 4th quartile
LA_casestudy.dta	County-level data set used for the Louisiana case study shown in Figure 4	• births: number of births by 15-44 year old women in county • high_relative_t_X: lead/lag of boll weevil arrival in counties with cotton-dependency in 4th quartile

Original data-files; make_all.do uses these to construct the .dta-files listed above:

Name	Description
IPUMS_full_count.dta	IPUMS USA full-count data for years 1880, 1900, 1910, 1920, 1930
super_fips1880_1930.dta	Data set that defines consistent "super counties" from 1800 to 1930
BW by county.xls	Excel data set; Information on arrival of boll weevil by county (provided by Fabian Lange, Alan Olmstead, and Paul Rhode)
1900_schlmnth.dta	Supplementary data set with variable schlmnth for year 1900 from IPUMS USA
ag_beforesupercty.dta	Agricultural data from Haines, Fishback and Rhode (2015).
brannenplantcounties_1910.dta	Indication of whether a county is a "plantation county" by Brannen (1924).
./rosenwald/icpsrcnt.xls	County-level information on Rosenwald schools; provided by Aaronson and Mazumder (2011).
fishback_mtco2129.xls	Data on child mortality and stillbirths provided by Fishback, Haines and Kantor (2007).
linked_1880_1900_males.dta, linked_1880_1910_males.dta, linked_1880_1920_males.dta	Linked Data Samples from IPUMS USA; individuals linked from the 1880 100% Census to the 1% Census of 1900, 1910, and 1920 from
folders "DS006" to "DS0032"	Agricultural data from Haines, Fishback and Rhode (2015).
Folder "wage data excel files"	See data appendix for further details
pop_beforesupercty.dta	Population data (see population data replication file Ager, Brueckner, and Herz (2017))