

Replication files for

“The Not-so-hot Melting Pot: The persistence of outcomes for descendants of the Age of Mass Migration”

This file explains the data creation and analysis for “The Not-so-hot Melting Pot: The persistence of outcomes for descendant of the Age of Mass Migration”

Note, the main data file cannot be included since it is restricted-access from IPUMS. Individuals can apply for access to the historical data from IPUMS. See https://usa.ipums.org/usa/complete_count.shtml for more information.

Creation of linking files:

- 1) Build set of potential links
 - a. build18801910links.do – builds set of potential links between 1880 and 1910
 - b. build19101940 links.do – builds set of potential links between 1910 and 1940
 - c. build19101920 links.do – builds set of potential links between 1910 and 1920
- 2) Draw random sample of 2000 from set of potential links and handlink them
 - a. 18801910handlinked.dta
 - b. 19101940handlinked.dta
 - c. 19101920handlinked.dta
- 3) Model the best link with probits
 - a. probitmodelling_18801910.do
 - b. probitmodelling_19101940.do
 - c. probitmodelling_19101920.do
 - d. Then you apply the probit results to the set of potential links and keep only those who meet the cutoffs described in Appendix B.
- 4) Merge the 1880-1910 and 1910-1940 datasets
 - a. build188019101940data.do – this creates the main dataset used called “188019101940_G3_G2_G1.dta”
- 5) Clean the linked data
 - a. clean_step1.do – sets the sample frame and adds in occupational income variables.
 - b. clean_step2.do – weights the data to be representative of the underlying population
 - c. clean_step3.do – collapses the data to the first-generation source country level and takes the averages.

6) Analysis

- a. mainanalysis.do – creates all major tables
- b. robustto1920.do – analysis when accounting for measurement error by using a second father observation in the 1920 census.

General comments:

- 1) The runproject.do creates all the tables after creating linked census data from steps (1)-(4)
- 2) The first step of building the potential set of links takes a long time and can use a lot of memory.
- 3) I note the generation by adding postscripts to variable names. For example, I add _G11880 for grandfathers in 1880, _G21910 for fathers in 1910, and _G31940 for grandsons in 1940.