1. **Software**

Stata/SE 15.1
- User installed packages: clus_nway, version 3.0, Oct. 10, 2017 (http://fmwww.bc.edu/RePEc/bocode/c/)

R version 4.3.2 (2023-10-31) -- "Eye Holes"
- Platform: x86_64-redhat-linux-gnu (64-bit)
- User installed packages: foreign

2. **Raw data sources**

In addition to the data sources described in the read me file for the main text, Appendix A uses IPUMS Linked Representative Census Samples, 1850-1930. Complete documentation can be found at the following link: https://usa.ipums.org/usa/linked_data_samples.shtml.

Appendix A makes extensive use of the 1940 census with first names, which is not publicly available. See detailed instructions for obtaining access to these raw data files from the NBER in the README_1940_data file in the "1940_data_details" folder of this replication package. For convenience, detailed replication instructions below clearly indicate steps that must be done on the NBER server.

3. **Guide to folders**

Folder Name	Description
raw_data	All raw data files or data files that cannot be produced from raw data provided as part of this replication package can be found here.
generated_data	All data files produced from raw data files using code in this replication package are stored here.
stata_code	All .do files used to produce tables for this appendix can be found here.
batched_R_code	All .R files used to run simulations for this appendix are stored here.
simulation_log_files	All .Rout files produced by simulations, for users reference.
output	Contains text files with figures that comprise all tables in this appendix.

4. **Detailed instructions**

4.1. Appendix A.1.1

FILE TYPE	FILE NAME	NOTES
Required data file(s)	linked_1880-1910_males.dta; linked_1850-1880_males.dta; 1880_fathers.dta; 1850_fathers.dta; panel_trends_30yr_income.dta	The first two data files are raw data downloaded from IPUMS.org. The last three are data files generated by the replication files for the main paper.
Do file(s)	00_TableA1.do	
Data file(s) created	none	
Table output file(s) created	tableA-1.txt	This do file produces columns (1) and (2) of this table. The remaining columns are calculated from these values according to the formulas given in the column headings.

4.2 Appendix A.1.2

NOTE: All steps in section 4.2 must be run on NBER server with special permission

4.2.1 Generate two separate 1% samples from the full count 1940 census.

FILE TYPE	FILE NAME	NOTES
Required data file(s)	northeast.dta; midwest.dta; south.dta; west.dta	These are cleaned full count 1940 census data files. Codes and procedures for generating these files are included in the replication package for the main paper.
Do file(s)	generate_1940_extracts.do	Change paths to own working directory.
Data file(s) created	All_1pct.dta	
Table output file(s) created	None	

4.2.2 Generate smaller files stripped of identifying information, which are eligible for public use.

FILE TYPE	FILE NAME	NOTES
Required data file(s)	All_1pct.dta	Generated in step 4.2.1

Do file(s)	simulations_prelim_1.do; simulations_prelim_2.do	Can be run in any order. Change paths to own working directory.
Data file(s) created	simulation_dataset_1.dta; simulation_dataset_2_child.dta; simulation_dataset_2_adult.dta	These contain extracts with first names from restricted use 1940 data, so they are only available on NBER server.
Table output file(s) created	None	

4.2.3 Generate tables A-2, A-3, A-4, and A-9. This step generates text files which contain all of the information in these tables. We filled in the tables manually using numbers from these text files. This step also creates data files used in Appendix A.2.

FILE TYPE	FILE NAME	NOTES
Required data file(s)	simulation_dataset_1.dta; simulation_dataset_2_child.dta; simulation_dataset_2_adult.dta	These are included as raw data. Codes and procedures for generating them are given above (4.2.2).
Do file(s)	01_TableA2_TableA4.do; 02_TableA3_TableA9.do	Can be run in any order.
Data file(s) created	data_for_R_1.dta; dataset_adult_for_R_2.dta; dataset_child_for_R_2.dta	To be used to generate tables for Appendix A.2. These datasets are included in the replication package in the folder "generated_data"
Table output file(s) created	tableA-2.txt, tableA-3.txt, tableA-4.txt, table A-9.txt	Tables were manually formatted based on figures in these text file.

4.3. Appendix A.2

NOTE:

4.3.1 Run the first set of simulations that generate tables A-5, A-6, A-7, A-8. This is done in R, which reads Stata data files generate above (Appendix A.1.2 step 3). We did this in batches on an external server to save time. **NOTE: Aggregate run time for this step is roughly 120 hours. We are including raw output and log files for users who are not able to reproduce this step themselves.**

FILE TYPE	FILE NAME	NOTES

Required data file(s)	data_for_R_1.dta	Generated above (4.2.3).
R code file(s)	simul_intgmob_1_f1.R; simul_intgmob_1_f2.R; simul_intgmob_1_f3.R; simul_intgmob_1_f4.R; simul_intgmob_1_f5.R; simul_intgmob_1_f6.R; simul_intgmob_1_f7.R; simul_intgmob_1_f8.R; simul_intgmob_1_f9.R; simul_intgmob_1_f10.R	Run in parallel on an external server. Each file runs 5 iterations of same program (with different seed set at start).
Data file(s) created	res1.dta; res2.dta; res3.dta; res4.dta; res5.dta; res6.dta; res7.dta; res8.dta; res9.dta; res10.dta	
Table file(s) created	none	

4.3.2 Run the second set of simulations that generate tables A-10, A-11, A-12. This is done in R, which reads Stata data files generate above (Appendix A.1.2 step 3). We did this in batches on an external server to save time. **NOTE: Aggregate run time for this step is roughly 240 hours. We are including raw output files for users who are not able to reproduce this step themselves.**

FILE TYPE	FILE NAME	NOTES
Required data file(s)	dataset_adult_for_R_2.dta; dataset_chid_for_R_2.dta	Generated above (4.2.3).
R code file(s)	simul_intgmob_2_f1.R; simul_intgmob_2_f2.R; simul_intgmob_2_f3.R; simul_intgmob_2_f4.R; simul_intgmob_2_f5.R; simul_intgmob_2_f6.R; simul_intgmob_2_f7.R; simul_intgmob_2_f8.R; simul_intgmob_2_f9.R; simul_intgmob_2_f10.R	Run in parallel on an external server. Each file runs 5 iterations of same program (with different seed set at start).
Data file(s) created	v2res1.dta; v2res2.dta; v2res3.dta; v2res4.dta; v2res5.dta; v2res6.dta; v2res7.dta; v2res8.dta; v2res9.dta; v2res10.dta	
Table file(s) created	none	

4.3.3 Process output from simulations into appendix tables A-5, A-6, A-7, A-8, A-10, A-11, A-12.

FILE TYPE	FILE NAME	NOTES

Required data file(s)	res1.dta; res2.dta; res3.dta; res4.dta; res5.dta; res6.dta; res7.dta; res8.dta; res9.dta; res10.dta; v2res1.dta; v2res2.dta; v2res3.dta; v2res4.dta; v2res5.dta; v2res6.dta; v2res7.dta; v2res8.dta; v2res9.dta; v2res10.dta	Generated above (4.3.1 & 4.3.2).
Do file(s)	03_TableA5_A6_A7_A8.do; 04_TableA10_A11_A12.do	Can be run in any order
Data file(s) created	none	
Table output file(s) created	tableA-5.txt; tableA-6.txt; tableA-7.txt; tableA-8.txt; tableA-10.txt; tableA-11.txt; tableA-12.txt	