Replication Package for "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century"

Authors: Elisa Jácome, Ilyana Kuziemko, and Suresh Naidu
May 2, 2024

1 Structure of Replication Package

The replication package consists of do-files placed into two main folders: "build" and "analysis."

- **Build**: The files in this folder bring in the raw data as inputs and compile all of the datasets that are used in the analysis. This folder contains three sub-folders with code for cleaning and saving all of the data.
  - 1_DataSources: this folder contains 21 folders. These files are described in Sections 3–5 of this README file.
  - 2_Append: this folder contains all of the do-files used to combine all of the survey datasets and measures of predicted parental income. These files are described in Section 6 of this README file.
  - 3_Output: The output from 2_Append is saved in this folder.
- **Analysis**: The do-files in this folder construct all the figures and tables in the paper using the outputs from the "build" folder.
  - Folders "_main_figures_tables," "appendix_a," "appendix_b," "appendix_c," "appendix_d," "appendix_e": within each of the sub-folders, researchers will find do-files for generating each table or figure. The title of the do-file corresponds to the figure or table number in the paper.
  - Folder "output": output is saved in this folder.
  - Folder "supplementary_data": any additional data used to create figures or tables that was not cleaned by the authors of this project is contained here. These files are described in Section 7 of this README file.

2 Setting up your paths and running files

* To set up directories, run "_paths.do" first. Make sure to add your paths to the do-file.
* You can run "_run_all_build" to run all of the do-files in the "build" folder. You can run "_run_all_analysis" to run all of the do-files in the "analysis" folder.

3 Predicted Parental Income: Setting up Census, 1900, and 1936 data

This set of do-files bring in Census samples and other data sources in order to construct predicted parental incomes at the occupation x race x south level as well as any other variations used throughout the paper. All of these files can be found in the "1_DataSources" folder of "build."

3.1 1940 Census on NBER server

We use the full-count 1940 Census available via the National Bureau of Economic Research server. All of these do-files were run on the NBER server and can be found in the "Census-Data/code/NBERserver" folder of the replication package. We do not include the relevant data in this package, as any distribution fo this data would be a violation of data use agreements.

Researchers with access to the NBER server can access the data in "/homes/data/cens1940/1940_-2-0/100files/" on the server. Researchers without access to the NBER server who wish to access the restricted 1940 full-count Censuses should email ipums@umn.edu (`https://usa.ipums.org/usa/complete_count.shtml`).

1. 0a_Income_Scores_1940clean.do

    - *Description:* Using the 1940 full Census, this file locates (1) men aged 30–50 who are fathers of a child younger than 18 in the same household and (2) women aged 30–50 who are mothers of a child younger than 18 in the same household. Income measures for both parents are cleaned.

2. 0b_Income_Scores_1940_setup.do

    - *Description:* Using the sample of 1940 Census fathers aged 30–50 (created in 0a), this file further restricts to Black and white fathers. Other variables necessary to create average predicted income at various levels (files 1a-1f) are also cleaned.

3. Do-files 1a–1f

    - *Description:* These files take the sample of 1940 Census fathers aged 30–50 from 0b and construct predicted parental income (i.e., "income scores") at several levels of variation.

4. 2_Income_Scores_1940_byracebysouth_mothers.do

- *Description:* Using the sample of 1940 Census mothers aged 30–50 (created in 0a), this file further restricts to Black and white mothers. Average predicted income for mothers is calculated at various levels.

5. 3_IncomeRatios_1940_byracebysouth.do

    - *Description:* Using the 1940 sample of Census fathers aged 30–50, this file calculates income ratios at various levels. This file is utilized to construct the 1900-based measures of parental income.

6. 4_Fathers_byOcc.do

    - *Description:* Calculates share of 1940 Census fathers that report each of the coarsened (ANES–based) occupations.
    - *Notes:* This file is brought into the do-files that create Tables C.2 & C.3.

3.2 Census and American Community Surveys from IPUMS, 1950–2019

Throughout the paper, we use sub-samples of the decennial Census as well as the American Community Survey available through IPUMS (`https://usa.ipums.org/usa/`) to construct predicted parental income. Downloading the data requires creating an account with IPUMS. IPUMS does not currently provide the ability to store or reference custom extracts, but it allows for redistribution for the purpose of replication.

1. Cleaner do-files used to extract data in the "CensusData/input/RawData" folder.

    - *Description:* These seven files are the do-files made available by IPUMS after the data download and they include the list of variables in each file. The output from these files is then used in all subsequent files.

2. Do-files 2a–2d,3a–3d,4a–4d,5a–5d,6a–6d, and 7a–7c

    - *Description:* These files clean the Census and ACS data and construct average predicted parental income (i.e., "income scores") by decade at several levels of variation. The name of each do-file includes the Census decade utilized as well as whether the file pertains to Census fathers or Census mothers.
    - *Note:* For 1950–1990, the "a" & "c" files clean variables used to make income scores. The "b" & "d" files construct income scores.

3. 8_IncomeScores_CWfixes.do

    - *Description:* This file adjusts 1940-based income scores for self-employed and farmer income at several levels of variation using the 1960 Census.

- *Note:* All files merged in to adjust income are constructed in the Census subfolder "1940_Adjustments" described below.

4. 9a_IncomeScores_othervariations.do &
   9b_IncomeScores_othervariations_mothers.do

   - *Description:* Using the 1960–1990 Censuses, the 9a file constructs father income scores at the race-only, south-only, occ x south, and race x south levels. The 9b file constructs income scores for mothers at these levels.

5. 10_IncomeScores_Append.do

   - *Description:* Merges and saves 1940–1990 income scores by level of variation, first for father income scores and then for mother income scores.

In "1940_Adjustments" Subfolder: All of these do-files make additional adjustments to predicted parental income and can be found in the "CensusData/code/1940_Adjustments" folder of the replication package.

5. 1a_Census1960_fathers30to50.do *&*
   1b_Census1960_mothers30to50.do

   - *Description:* Using the 1960 Census (5% sample), the 1a file locates Black and white men aged 30–50 who are fathers of a child younger than 18 in the same household. Other variables necessary to create average predicted father income are cleaned. 1b does the same for mothers in the Census.

6. 1c_Census1960_farmer_ratios.do *&*
   1g_Census1960_farmer_ratios_mothers.do

   - *Description:* Using the output from 1a, file 1c calculates the ratio of farmer to farm laborer average income at various levels of variation for fathers. The 1g file uses the output from 1b to do the same for mothers.

7. 1d_Census1960_self_emp_ratios.do *&*
   1h_Census1960_self_emp_ratios_mothers.do

   - *Description:* Using the output from 1a, file 1d calculates the ratio of self-employed to non-self-employed average income at various levels of variation for fathers. File 1h uses the output from 1b to do the same for mothers.

8. 1e_Census1960_NILF_ratios.do

- *Description:* Using the output from 1a, this file calculates the ratio of income between those not in the labor force and those who are employed at the race x south level only.

9. 1f_Census_1960_income_to_earnings_ratios.do

    - *Description:* Using the output from 1a, this file calculates the ratio of total income to income from earnings, first for personal income and then for family income and household income. All ratios are at the occupation x race x south level.

Citation: Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024. https://doi.org/10.18128/D010.V15.0

3.3 1900 Income Scores

This set of do-files brings in two other data sources to construct predicted parental income for the earliest cohorts. All of these do files can be found in the "1900_IncomeScores" folder of the replication package.

1. 1a_1901CostLiving_construct.do

    - *Description:* Creates 1900-based income scores at various levels using data from Preston et al. (2003).

2. 1b_CensusAg_farmerincome_construct.do

    - *Description:* Imports data and calculates average net earnings for farmers at various levels. Data come from the United States Department of Agriculture Census of Agriculture Historical Archive, Tables 13 and 14.

3. 2_1900incomescores_combine.do

    - *Description:* Combines income scores from 1a with average net earnings for farmers from 1b. The output is the "1900" income scores at various levels.

3.4 1936 Consumption Survey

This set of do-files brings in the 1936 Expenditure Survey to construct predicted parental income for the earliest cohorts of farmer fathers. All of these do files can be found in the "ConsumptionSurvey_1936" folder of the replication package.

1. 1a_clean_1936.do

- *Description:* This file takes the two raw data sets (included in "RawData" folder, available through ICPSR) and cleans the necessary variables to generate measures of predicted parental income.
- *Notes:* This file calls in 1b_occ_coding_1936.do, which assigns individuals to the consistent occupational categories used in this paper.

2. 2a_IncomeScores_1936.do
    - *Description:* Using the output from 1_clean_1936.do, this file creates income scores at the occupation-only, occupation x race, and occupation x race x south levels of variation.
3. 2b_IncomeScores_othervariations_1936.do
    - *Description:* Creates income scores at the race-only, south-only, race x south, and occupation x south levels of variation.
4. 2c_IncomeScores_1936_merge.do
    - *Description:* Merges the output of 2a and 2b to create one dta file with all 1936 income scores.

Citation: United States Department of Labor. Bureau of Labor Statistics. Cost of Living Division, United States Department of Agriculture. Bureau of Home Economics. Economics Division, United States National Resources Committee. Consumption Research Staff. Industrial Section, United States Central Statistical Board, and United States Works Progress Administration. Study of Consumer Purchases in the United States, 1935-1936. Inter-university Consortium for Political and Social Research [distributor], 2009-06-29. https://doi.org/10.3886/ICPSR08908.v3

4 Cleaning survey datasets

This group of do-files cleans each individual survey separately and harmonizes the occupations to match the coarsened occupations used throughout the paper. The raw survey data is publicly available. All of these do-files can be found in the "build/1_DataSources" folder of the replication package.

4.1 Americans View Their Mental Health, 1957

1. 0_AVTMH1957_import_clean.do
    - *Description:* Imports raw data.
2. 1_AVTMH1957_cleaner.do
    - *Description:* Cleans dataset.

- *Note:* Runs "1a_AVTMH1957_R_hhsize.do," which codes variables related to household size of respondent.

Citation: Gurin, Gerald, Veroff, Joseph, and Feld, Sheila. Americans View Their Mental Health, 1957. Inter-university Consortium for Political and Social Research [distributor], 1992-02-16. https://doi.org/10.3886/ICPSR03503.v1

4.2 Americans View Their Mental Health, 1976

1. AVTMH1976_cleaner.do

    - *Description:* Imports raw data and cleans original dataset.
    - *Note:* Runs "07948-0001-Supplemental_syntax.do," ICPSR-provided code to clean raw data. Also runs "1a_AVTMH1976_R_hhsize.do," which codes variables related to household size of respondent.

Citation: Veroff, Joseph, Douvan, Elizabeth, and Kulka, Richard. Americans View Their Mental Health, 1976. Inter-university Consortium for Political and Social Research [distributor], 2005-11-04. https://doi.org/10.3886/ICPSR07948.v1

4.3 American National Election Studies, 1956–1970

1. 0_xwalk_xsectionid_to_panelid.do

    - *Description:* Tags panel respondents that should be kept in 1958 and 1960 (because they only meet all sample inclusion restrictions in one year). Output is then merged onto 1_ANES_cleaner.do in 1958 and 1960.

2. 1_ANES_cleaner.do

    - *Description:* Imports raw data for 1956, 1958, 1960. 1964, 1966, 1968, and 1970 cross sections. Cleans each dataset and appends them into one dataset.

3. 2_ANES_InterGenMobility.do

    - *Description:* Brings in the dataset created in previous step and creates more variables necessary for the main analysis.

Note: Each individual dataset can be downloaded from ICPSR via their American National Election Study (ANES) Series (`https://www.icpsr.umich.edu/web/ICPSR/series/3`).

4.4 Occupational Changes in a Generation

Citation: Blau, Peter M., Duncan, Otis Dudley, Featherman, David L., and Hauser, Robert M. Occupational Changes in a Generation, 1962 and 1973. [distributor], 1994-05-20. https://doi.org/10.3886/ICPSR06162.v1

1962 Wave

1. 0_ocg1962_data_input.do
    - *Description:* Imports raw data.
2. 1_ocg1962_clean.do
    - *Description:* Cleans dataset.

1973 Wave

1. 0_ocg1973_data_input.do
    - *Description:* Imports raw data.
2. 1_ocg1973_clean.do
    - *Description:* Cleans dataset.

4.5 National Fertility Study, 1970

1. NFS70_clean.do
    - *Description:* Imports raw data and cleans dataset.
    - *Note:* Runs "20003-0001-Supplemental_syntax.do," ICPSR-provided code to clean raw data.

Citation: Westoff, Charles F., and Ryder, Norman B. National Fertility Survey, 1970. Inter-university Consortium for Political and Social Research [distributor], 2008-08-08. https://doi.org/10.3886/ICPSR20003.v1

4.6 General Social Surveys, 1977–2018

Occupation and family income variables are cleaned over the course of 3 do-files (i.e., 1a,1b, and 1c) because the GSS-provided occupation codes change over time.

1. 1a_GSS_77to90_clean.do
    - *Description:* Imports full GSS-provided dataset (1972–2018) and restricts sample to 1977-1990 years. Cleans occupation and family income variables. Crosswalks occupation codes to 1950 ANES coarsened occupations.
2. 1b_GSS_88to10_clean.do
    - *Description:* Imports full GSS-provided dataset (1972–2018) and restricts sample to 1988-2010 years. Cleans occupation and family income variables. Crosswalks occupation codes to 1950 ANES coarsened occupations.

3. 1c_GSS_10to18_clean.do

    - *Description:* Imports full GSS-provided dataset (1972–2018) and restricts sample to 2012-2018 years. Cleans occupation and family income variables. Crosswalks occupation codes to 1950 ANES coarsened occupations.

4. 2_GSS_join_77to18_clean.do

    - *Description:* Appends 1a, 1b, and 1c files. Imports full GSS-provided dataset (1972–2018) and merges on (appended) file. Cleans merged dataset.

Citation: Davern, Michael; Bautista, Rene; Freese, Jeremy; Herd, Pamela; and Morgan, Stephen L.; General Social Survey 1972-2022. [Machine-readable data file]. Principal Investigator, Michael Davern; Co-Principal Investigators, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. NORC ed. Chicago, 2024. 1 datafile (2022 Release 2a)

4.7 National Survey of Families and Households, 1987-1988

1. 1_NSFHH_cleaner.do

    - *Description:* Imports raw data and cleans dataset.
    - *Note:* Runs 06041-0001-Supplemental_syntax, ICPSR-provided code to clean raw data.

Citation: Bumpass, Larry L., Sweet, James A., and Call, Vaughn R.A. National Survey of Families and Households, Wave 1: 1987-1988, [United States]. Inter-university Consortium for Political and Social Research [distributor], 2017-08-31. https://doi.org/10.3886/ICPSR06041.v2

4.8 National Survey of Black Americans, 1980

1. nsba_clean.do

    - *Description:* Imports raw data and cleans dataset.
    - *Note:* Runs 08512-0001-Supplemental_syntax.do, the provided code to clean raw data.

Citation: Jackson, James S. (James Sidney), and Gurin, Gerald. National Survey of Black Americans, 1979-1980. Inter-university Consortium for Political and Social Research [distributor], 2005-11-04. https://doi.org/10.3886/ICPSR08512.v1

4.9 National Longitudinal Survey of Mature Women, 1967

1. 1_NLSMW_clean.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*

- Runs default-value-labels.do, husbandocc-value-labels.do, and father_edu-value-labels.do. These NLS-provided do-files carry out a preliminary cleaning of the raw data.
- Merges in husbandocc, parents_edu, siblings_fertility_hhsize, motherocc_-Rage15, and who_Rlivew_age15 dta files, which add other relevant variables to the raw data.

Citation: Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Mature Women, 1967-2003 (rounds 1-22). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2004.

4.10 National Longitudinal Survey of Older Men, 1966

1. 1_NLSOM_cleaner.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs nls66_for_mobility-value-labels_OLD.do, NLS-provided code to clean main raw dataset.
        - Runs numkids_inhh-value-labels.do and Rnat_classwrker_formobility-value-labels.do, which add other revelant variables to the raw data. The NLS-written code in these do-files lightly cleans the added variables.

Citation: Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Older Men, 1967-1990 (rounds 1-13). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 1999.

4.11 National Longitudinal Survey of Youth, 2002

1. cleanup_nlsy79_xsec.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs NLSY79-ISM-value-labels.do, NLS-provided code to clean main raw dataset.
        - Merges in employment_status_byintdate, marstat_expanded, and parentsocc_-Rage14 dta files, which add other relevant variables to the raw data.

Citation: Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1979 cohort, 1979-2016 (rounds 1-27). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2019.

4.12 National Longitudinal Survey of Young Men, 1981

1. 1_cleanup_nlsym66_newxsec.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs NLSYM66-ISM-value-labels.do, NLS-provided code to clean main raw dataset.
        - Merges in hhsize, Rwho_livewith_age14, classofwrker_nlsym, and sibs dta files, which add other relevant variables to the raw data.

Citation: Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Young Men, 1966-1981 (rounds 1-12). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 1992.

4.13 National Longitudinal Survey of Young Women, 1988

1. 1_cleanup_nlsyw68_newxsec.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs NLSYW68-ISM-value-labels.do, NLS-provided code to clean main raw dataset.
        - Merges in hhsize, altmotherocc_Rage14, momwork_Rage14, classofwrker, and sibs dta files, which add other relevant variables to the raw data.

Citation: Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Young Women, 1968-2003 (rounds 1-23). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2004.

4.14 Panel Study of Income Dynamics

Citation: Panel Study of Income Dynamics, public use dataset [restricted use data, if appropriate]. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

1997 Wave

Note: The 1997 folder must be run before the 2017 folder.

1. 0_fims_sibmap_wide.do

    - *Description:* Imports sibling relationships from PSID FIMS database and saves a wide format dta file.

2. psid1997_clean.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs J320143.do, PSID-provided code that imports the raw data.
        - Runs J320143_formats.do, PSID-provided code that lightly cleans main raw dataset.

2017 Wave

1. psid2017_clean.do

    - *Description:* Imports raw data and cleans dataset.
    - *Notes:*
        - Runs J320108.do, PSID-provided code that imports the raw data.
        - Runs J320108_formats.do, PSID-provided code that lightly cleans main raw dataset.

Combined survey waves

1. 0_PSID_prelimfiles.do

    - *Description:* This file imports raw psid data extracts and creates dta files that are called into subsequent do-files in this folder.

2. 1_PSID_IGEclean.do

    - *Description:* Cleans PSID data for analysis. Calculates actual income (using 1, 5, and 10 years of income) and predicted income for linked PSID fathers and calculates 1-year actual income for linked adult child respondents aged 30–50.
    - *Notes:* Output of this file is used to create Table A.2.

3. 2a_PSID_retrospective_children.do

    - *Description:* Cleans demographic variables, finds actual income (1-year) and binned family income, and finds coarsened modal retrospective father occupation for linked adult children in the PSID who are heads of their household.

4. 2b_PSID_retrospective_fathers.do

    - *Description:* Calculates 1-,5-, and 10-year average actual income around age 40 for fathers of respondents in 2a. 1- and 3-year actual income in 1970 is further calculated, as well as predicted father income at the occupation x race x south level.

5. 2c_PSID_retrospective_verify.do

- *Description:* Merges the output of 2a and 2b, tags discrepancies in father occupation recall between fathers and adult children, creates logged measures of income for fathers and adult children, and creates ranked measures in both generations.

6. 2d_PSID_recall_matrix.do

   - *Description:* This file creates a "recall" matrix (wide format) with adult child's recalled father occupation (when child is between ages 30 and 50) as the rows and father's modal actual reported occupation between 30 and 50 as the columns. A matrix in long-form is also saved.

In "Actual_vs_predictedincome" Subfolder:

7. 1_PSID_bothparents_clean.do

   - *Description:* Cleans demographic variables for adult child respondents linked to at least one parent. Creates 3- and 5-year averages of actual income and parental actual income for respondents.

8. 2_PSID_incscores_bothparents.do

   - *Description:* Creates predicted parental income (retrospective, both parents) for adult child respondents.

5 Additional files used to clean surveys and construct analysis dataset

This additional group of do-files are used throughout the data-cleaning process in Sections 2 and 3 or in Section 5 below.

5.1 Additional files utilizing Census data

1. Number of children by cohort: all of these do-files use Census data from IPUMS and are called in in Section 5. These files can be found in "build/1_DataSources/CensusData/code/NumberChildren_byCohort."

   (a) 1_NumberChildren_1910to1990_byCensus.do

   - *Description:* Using the 1% 1910–1990 Censuses, this file calculates median number of children in the household and median number of family members at the occupation x race x south cell.

(b) 2_NumberChildren_bycohort.do

- *Description:* Using output from 1_NumberChildren, this file calculates weighted average median family size at the occupation x race x south x year level. Year ranges from 1920–1989.

2. Census-based shares for constructing weights: all of these do-files use Census data from IPUMS and are called in in Section 5. These files can be found in "build/1_DataSources/CensusData/code/Shares_for_weights."

   (a) Shares_bysex_byrace.do

   - *Description:* Using the 1950–2010 Censuses, this file calculates sex-by-race shares by Census decade:

   (b) Shares_byage_byedu_bysex_byrace.do

   - *Description:* Using the 1950–2010 Censuses, this file calculates demographic shares by decade. Each share corresponds to a sex-race-education-age group.

3. Other raw data files: these four do-files use Census data from IPUMS and can be found in"build/1_DataSources/CensusData/code/OtherCensus_RawData."

   (a) The first three do-files (with prefixes "1900," "1930_1940" and "1950to2000") are Census-provided cleaner files that we modified and which save output used to make figures or tables. Each file lists the IPUMS USA dataset that corresponds to the do-file and we include the data extract in the replication package in order to run these files (in "CensusData/input/RawData").

   (b) The final file "motheroccej_template_occs.dta" is also included, which is utilized to crosswalk Census occupations of mothers to coarsened occupations.

5.2 Additional files used to clean survey data

1. Do-files in "'build/1_DataSources/Crosswalks" folder

   - *Description:* All of these do-files were hand-coded to translate the occupational codes in a given survey to the set of occupational codes used in this paper. These crosswalks are called in throughout the data-cleaning process to crosswalk the occupational codes in the Census or in a given survey to the 28 coarsened occupations described in Appendix E.

2. CPI_deflator.do in "'build/1_DataSources/CPI" folder

- *Description:* Brings in 1947–2017 CPI data and creates a "deflator" variable that will be used to convert income to 1950$ when cleaning some surveys.

6 Appending predicted parental income to survey data

This set of do-files combines all of the output files from Sections 2 and 3. All of these files can be found in the "2_Append" folder of "build."

1. 1_Append_Datasets.do
    - *Description:* Imports all cleaned surveys, keeps relevant variables, and appends all surveys together.
    - *Notes:* Census data is not imported nor appended.
2. 2_Clean_PooledData.do
    - *Description:* This file takes the appended surveys from 1_Append_datasets.do and merges in many variations of predicted parental income (i.e, "income scores"). Census-adjusted weights are then made and are used to create ranked measures of predicted parental income.
    - *Notes:* Additional variables necessary for the merging of income scores are also constructed.
3. 3_Append_Census_Surveys.do
    - *Description:* Appends cleaned survey data from 2_Clean_PooledData, 1936 data, and 1940–1980 Census data. The output is used to construct appendix figures.
4. 4_TSIV_twosample_build.do
    - *Description:* Appends cleaned survey data from 2_Clean_PooledData, cleaned 1940 Census data, and cleaned 1936 data. The output is used to construct appendix figures.

7 Analysis files

Each analysis file is given the name of the figure or table to which it corresponds in the paper. If the code to create a figure or table is spread over multiple do-files, then the do-files will have the suffix "part1", "part2", and so on. The do-files are cataloged based on where they are in the paper (i.e., in the main text or in Appendices A–E).

There is also a folder titled "supplementary_data" that includes datasets of already-compiled information that we use to make certain figures. See the corresponding figure note for more details.

- "national_percentile_outcomes" is publicly available via Opportunity Insights and is used to create Figure 9.
- "Margo_Census" and "wid_summarystats_1913to2021" are used to create Figure A.1.
- "Occ1950ej_definitions" is a file containing the full name of each occupation, used to create Tables C.2 and C.3.
- To make Figure D.2, we utilize a sample of parents and children from the National Longitudinal Surveys provided to the authors by Jonathan Davis and Bhash Mazumder. This sample corresponds to Figure 3 in their paper "The Decline in Intergenerational Mobility After 1980." For replication purposes, we refer the reader to their replication package available at `https://direct.mit.edu/rest/article-abstract/doi/10.1162/rest_a_01413/119490/The-Decline-in-Intergenerational-Mobility-after`.