Readme File for Replication Package of
"Economic Consequences of Kinship: Evidence From US bans on Cousin Marriage"

Arkadev Ghosh, "Sam" Il Myoung Hwang, and Munir Squires

23 March 2023

03 October 2024 amendment: The authors found a minor coding error that only affected Appendix Figure A.13. This has been fixed in the updated code. There is no substantive change in the corresponding result, and so the paper itself remains unchanged.

OVERVIEW

This README file describes the contents of the replication folder and instructs users how to replicate the results of the paper and appendix. Separate files to reproduce the tables and figures are stored in the `/Code/` folder.

To perform the replications, run "`master.do`" in the code folder. Before doing so, change the global macro "`dir`" (first line of code in `master.do`) to the location of the `CM_Replication` folder on your computer.

We are not able to publicly post some of the data sources used in our analysis. Collapsed versions of these data, sufficient for reproduction of our tables and figures, are included in the replication folder.

We include in the replication files all code needed to generate the results in the paper and appendix.

DATA AVAILABILITY AND PROVENANCE

IPUMS US Full-Count Restricted Census: We obtained individual-level, complete-count Population Census data from 1850-1940 (Ruggles at al., 2015) via an agreement with the Minnesota Population Center. These data can be accessed by other researchers via their own agreements.

Familinx: We use anonymised, publicly available family trees from the Familinx database (Kaplanis et al., 2018), downloadable at familinx.org. This dataset provides exact ancestral links of individuals derived from family trees that have been created and managed by users of the geni.com website. Kaplanis et al. (2018) have cleaned this dataset and removed entries with problematic errors (such as having three parents).

State characteristics:

- *Statehood* – We obtained year of statehood from the United States Census Bureau Website (https://www.census.gov/dataviz/visualizations/048/508.php)
- *Compulsory schooling law* – We use state-level data describing compulsory attendance and child labor laws in the US. These come from Adriana Lleras-Muney's website (https://adriana-llerasmuney.squarespace.com/data).

- *Minimum age at marriage* – We use state-level data on minimum age at marriage in the US. Data for this variable come from three sources: Blank et al. (2007), Syrett (2016), and email correspondence with James Sallee (a co-author of Blank et al. (2007)).
- *Total frontier experience* – We use data describing time spent on the US frontier, which is defined as time spent on "the boundary at which population density falls below two people per square mile" from Bazzi et al. (2020).
- *Railroad coverage* – We use data on the railway networks in the US from Bazzi et al. (2020), which uses data from Atack et al. (2010).
- *Sterilization* – We use state-level data describing the adoption and composition of compulsory sterilization laws in the US. These data come from Lutz Kaelber's website (https://www.uvm.edu/~lkaelber/eugenics/).

Familysearch Marriage Records: The marriage records in our dataset come from original documents which have been scanned, transcribed and made publicly available by Family Search (familysearch.org). We retrieved all available records for marriages in any US state between 1750 and 1940.

Other:
- Lasso Industry Demographic and Occupation (LIDO) Score from Saavedra and Twinam (2020).
- Occupation based income-percentile ranks from Song et. al. (2020).
- US state-decade level population data compiled from the US Census Bureau.

DATA

The following data files produce all figures and tables in the paper and appendix. We flag restricted access data with a * below.

Folder Data/Original

1. `income_pcrank.dta` -- Occupation based income-percentile ranks (Song et. al. 2020)
2. `LIDO_occscore_1950.dta` -- Lasso Industry Demographic and Occupation (LIDO) Score (Saavedra and Twinam, 2020). This measure predicts 1950 earnings using occupation, industry, and demographics.
3. `TotalMR_bystateyear.dta` -- Total number of marriage records at the state-year level
4. `US_state_codes.dta` -- Two-digit U.S. state codes
5. `US_States_Pop.dta` -- US state-decade level population data (compiled from US Census Bureau)
6. `MR_cleaned_1750to1950.dta*` -- Marriage records data from 1750 to 1950
7. `MR_Census_linked.dta* --` This data file includes couple-level census data merged to our marriage records (before cleaning).
8. `MR_Census_Linked_histids.dta*` -- This file includes histids of married individuals in the 1900, 1910 and 1930 census whom we could link directly to our marriage records.

9. `Census_married_1900_30.dta*` -- This file includes census-data (variables used in Table A.2.) for all married individuals in the 1900, 1910 and 1930 censuses.
10. `robustness_interactions.dta` -- This file includes state-decade level data on railroad coverage, min. age of marriage, sterilizations, compulsory schooling, statehood, share of foreign borns and sex-ratios.
11. `year_of_ban.dta` -- This file includes information on when U.S. states banned cousin marriage
12. `state_characteristics.dta` -- Data on several state characteristics for different time periods (used to compare states with early versus late bans).
13. `census_data_uncollapsed.dta*` -- This file includes stacked censuses from 1850 to 1940. The sample is all white males aged 18 to 50, having surnames that also appear in our marriage records. The dataset includes all the variables that are defined in Appendix B.6 of the paper.
14. `FCM_Familinx.dta` -- Marriage-level data indicating first cousin marriages from Familinx (those with a full set of great-grandparents).

Folder Data/Final

1. `MR_Census_linked_clean*.dta` -- This file includes couple-level census data linked to our marriage records, cleaned and ready for OLS analysis.
2. `Isonymy_fs.dta` -- This file includes isonymy data at the surname-state level for the entire period (1750-1940).
3. `isonymy_sy.dta` -- This file includes rates of isonymy at the state-year level for the entire period (1750-1940).
4. `Isonymy_fsy.dta*` -- This file includes isonymy data at the surname-state-year level.
5. `final_collapsed_clean.dta`, `final_collapsed_flhocc.dta`, `final_collapsed_ageb_ms.dta`, `final_collapsed_fbpl.dta`, `final_collapsed_nysiis.dta`, `final_collapsed_scommonq.dta`, `final_collapsed_surnamestate_sz.dta*`, `final_collapsed_surnamestatecm.dta`, `final_collapsed_thresh8.dta`, `final_collapsed_thresh12.dta`, `final_collapsed_obsiso.dta`, `final_collapsed_SFE.dta*` -- All these files are census data merged with marriage records and collapsed and ready for analysis.
6. `geni_isonymy.dta` -- State-year-level data on consanguineous marriages rates (from Familinx) and isonymous marriages rates (from marriage records)
7. `geni_fs.dta` -- State-year-level data on consanguineous marriages rates (from Familinx)

Computational requirements

The code was last run using Stata Version SE 17 on Windows 10 with packages updated on March 23rd, 2023. The code was also run successfully using Stata SE 17 on Mac OS 12.6.

It should run on relatively recent versions of Stata in Windows and Mac OS, but we cannot guarantee that it will. All required Stata packages are loaded in the `master.do` file. No other software is required.

The code was last run on a machine with the following specifications:

- Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz
- 32.0 GB 2400 MHz DDR4 RAM
- Micron 1100 SATA 512 GB SSD
- Windows 10 pro

The code takes approximately 50 minutes to run from start to finish on the above machine.

CODE

Programs in `/Code/01 Prep` merge, collapse, and otherwise prepare datasets for analysis.

Programs in `/Code/02 Analysis/Descriptives` generate summary statistic tables and figures.

Programs in `/Code/02 Analysis/Event-Study` generate tables and figures summarising our findings.

Programs in `/Code/02 Analysis/Familinx_MR` generate tables and figures summarising cousin marriage rates and types using Familinx genealogy data

The program in `/Code/02 Analysis/OLS` generates tables summarising our findings using individual-level data.

Instructions:

First edit the default path then run `master.do` in `CM_Replication/Code` to replicate all the Figures and Tables generated using the data we are able to share publicly. We flag code run on restricted-access data with a * below. These files are commented out and will not run.

The sub-files and the output they each produce are all listed below.

Folder: 01 Prep

1. `1_Create_CM.do`* -- Prepares cousin marriage data from marriage records which are merged to census data for analysis
2. `2_OLS_PrepareAnalysis.do`* -- Takes restricted access couple-level census data merged to marriage records, and cleans and prepares it for analysis.
3. `3_EventStudy_PrepareAnalysis.do`* -- Takes restricted-access individual-level census data, merges with data on cousin marriage rates (by surname) and cousin marriage bans (by state and year of ban), and finally prepares collapsed data ready for analysis.

Folder: 02 Analysis/Descriptives

1. `1_prepostCM_persistence.do` -- Figure 1
2. `2_CM_Rates_Maps.do` -- Figure 2, Figure A.11, Figure A.12,
3. `3_MarriageRecords_Coverage.do` -- Figure A.10
4. `4_Consanguinity_income.do` Figure A.14
5. `5_FCM_Isonymy_SY.do` - Figure B.18, Figure B.19
6. `6_IsoCM.do` -- Table 2
7. `7_BalanceTable_MRCensus_linkedvsnot.do*` -- Table A.2
8. `8_earlyvslate_ban.do` -- Table B.30

Folder: 02 Analysis/OLS

1. `1_LinkedCensusMR_OLS.do*` -- Table 3, Table A.3

Folder: 02 Analysis/Event-Study

1. `1_CM_FiguresTables_FirstStage.do` -- Figure 3, Table A.4
2. `2_CM_FiguresTables_MainResults.do` -- Figure 4, Figure 5, Figure A.1, Figure A.2, Figure A.3, Figure A.7, Figure A.8 , Figure A.13, Table A.5, Table A.6, Table A.7, Table A.8, Table A.9, Table A.13 , Table A.14, Table A.18.
3. `3_CM_FiguresTables_AgeMSHet.do` -- Figure A.4, Figure A.5, Table A.10, Table A.11
4. `4_CM_FiguresTables_WealthHet.do` -- Figure A.6, Table A.12
5. `5_CM_TablesRobustness_Confounders.do` -- Table A.15
6. `6_CM_TablesRobustness_FBPLTreatment.do` -- Table A.16
7. `7_CM_TablesRobustness_NYSIIS_Surnames.do` -- Table A.17
8. `8_CM_TablesRobustness_StatesurnameCM.do` -- Table A.19
9. `9_CM_TablesRobustness_Surnamestatesize_controls.do*` -- Table A.20
10. `10_CM_TablesRobustness_8pthresh.do` -- Table A.21
11. `11_CM_TablesRobustness_12pthresh.do` -- Table A.22
12. `12_CM_TablesRobustness_ObsIsonymy.do` -- Table A.23
13. `13_CM_TablesRobustness_SFE.do*` -- Table A.24
14. `14_CM_TablesRobustness_LastTreatedControl.do` -- Table A.25
15. `15_CM_TablesRobustness_CommUncomm_Surnames.do` -- Table A.26

Folder: 02 Analysis/Familinx_MR

1. `1_geni_isonymy.do` -- Table B.27
2. `2_geni_fs.do` -- Table B.28
3. `3_fcm_types.do` -- Table B.29

REFERENCES

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A Margo (2010) “Did railroads induce or follow economic growth?: Urbanization and population growth in the American Midwest, 1850–1860,” Social Science History, 34 (2), 171–197.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse (2020) “Frontier culture: The roots and persistence of “rugged individualism” in the United States,” Econometrica, 88 (6), 2329–2368.

Blank, Rebecca M, Kerwin Kofi Charles, and James M Sallee (2007) “Do state laws affect the age of marriage? A cautionary tale about avoidance behavior,”Technical report, National Bureau of Economic Research.

FamilySearch. Retrieved 2021-2022, from https://www.familysearch.org/

Kaelber, Lutz (2011). *Eugenics: Compulsory Sterilization in 50 American States*. Retrieved 2022, from https://www.uvm.edu/~lkaelber/eugenics/

Kaplanis, Joanna, Assaf Gordon, Tal Shor et al. (2018) “Quantitative analysis of population-scale family trees with millions of relatives,” Science, 360 (6385), 171–175.

Lleras-Muney, Adriana. (n.d.). *Data*. Retrieved 2022, from https://adriana-llerasmuney.squarespace.com/data

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2015) “Integrated Public Use Microdata Series: Version 6.0 [dataset],” Minneapolis: University of Minnesota.

Saavedra, Martin and Tate Twinam (2020) “A machine learning approach to improving occupational income scores,” Explorations in Economic History, 75, 101304.

Song, Xi, Catherine G Massey, Karen A Rolf, Joseph P Ferrie, Jonathan L Rothbaum, and Yu Xie (2020) “Long-term decline in intergenerational mobility in the United States since the 1850s,” Proceedings of the National Academy of Sciences, 117 (1), 251–258.

Syrett, Nicholas L (2016) American child bride: a history of minors and marriage in the United States: UNC Press Books.

United States Census Bureau (2012). *U.S. Territory and Statehood Status by Decade, 1790-1960*. Retrieved 2022, from https://www.census.gov/dataviz/visualizations/048/508.php