

In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940.

Claudia Olivetti and M. Daniele Paserman

Online Appendix

A Name coding

We assess here the robustness of our results to different name coding schemes. In the benchmark specification, for the purpose of aggregating results by name, we treated each individual occurrence of a first name as a separate name, ignoring middle names or middle initials and common abbreviations. In Appendix Table A1 we relax this restriction.

The first row of each panel in the table reproduces the benchmark estimates from Table 3. The second row presents the intergenerational elasticities when we take into account middle initials as well as the first name (so, for example, “William,” “William J.” and “William H.” are treated as three separate names). The third row presents results when we group together the main root of a name with its most common nicknames (so, for example, “William,” “Bill,” “Billy” and “Willie” are all treated as separate instances of the same name).

In the next row, we deal with the potential misspelling of names by using the Soundex algorithm.¹ Coding names this way results in a substantial reduction in the number of names, and an increase in the average number of occurrences per name. This can have two opposite effects on the estimated elasticity. On one hand, it reduces the occurrence of uncommon names and therefore average income is likely a more accurate measure of actual father’s income. On the other hand, the Soundex algorithm groups together names that may belong to very different socioeconomic groups (for example, Michael and Miguel) which may exacerbate measurement error.

The trends in father-son-elasticities are not much affected by the different name coding schemes. The point estimates in row 2 and 3 are quite similar to the benchmark, while based on the Soundex algorithm the father-son elasticity are higher by 6 to 13 percentage points. As for the father-son-in-law elasticities in all specifications there is less evidence of an upward trend. This is driven by fluctuations in the point estimates in the first part of the sample while all specifications exhibit a sharp decline in the elasticity after 1920.

¹The Soundex is a phonetic algorithm that indexes names by sound, and is specifically designed to assign the same numeric code to similar sounding names. (The U.S. National Archives and Records Administration, <http://www.archives.gov/research/census/soundex.html>)

Appendix Table A1. Sensitivity to Different Name Coding Schemes.

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
A: Fathers-Sons						
Name concept:						
All	0.3500 (0.0240)	0.3133 (0.0201)	0.3440 (0.0166)	0.4953 (0.0153)	0.4756 (0.0119)	0.4340 (0.0119)
Middle initials	0.3400 (0.0231)	0.3112 (0.0191)	0.3291 (0.0156)	0.4189 (0.0136)	0.4391 (0.0111)	0.3963 (0.0112)
Nicknames	0.3673 (0.0246)	0.3310 (0.0207)	0.3412 (0.0176)	0.4489 (0.0160)	0.4264 (0.0124)	0.3951 (0.0125)
Soundex codes	0.4212 (0.0305)	0.4041 (0.0251)	0.4771 (0.0224)	0.5571 (0.0184)	0.5527 (0.0155)	0.5220 (0.0155)
N, no. names (All)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]	[119406, 3866]
N, no. names (M.I.)	[36685, 1419]	[50243, 1789]	[79227, 2676]	[107721, 3910]	[120708, 4607]	[118003, 4590]
N, no. names (Nicknames)	[37172, 1138]	[50947, 1415]	[80315, 2107]	[109098, 3111]	[122502, 3582]	[119421, 3736]
N, no. names (Soundex)	[39262, 887]	[54941, 995]	[84686, 1248]	[116154, 1596]	[130274, 1623]	[127893, 1466]
B: Fathers-Sons in Law						
All	0.3402 (0.0214)	0.4009 (0.0194)	0.3992 (0.0185)	0.4932 (0.0134)	0.4143 (0.0102)	0.3725 (0.0101)
Middle initials	0.3441 (0.0210)	0.3619 (0.0181)	0.3771 (0.0172)	0.4249 (0.0124)	0.3838 (0.0098)	0.3491 (0.0098)
Nicknames	0.4360 (0.0262)	0.4152 (0.0206)	0.4135 (0.0191)	0.4551 (0.0142)	0.3889 (0.0109)	0.3543 (0.0107)
Soundex codes	0.5907 (0.0314)	0.5543 (0.0262)	0.5570 (0.0260)	0.6122 (0.0180)	0.4933 (0.0136)	0.4935 (0.0144)
N, no. names (All)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]	[77001, 3320]
N, no. names (M.I.)	[22954, 1142]	[29682, 1644]	[45239, 2459]	[67637, 3496]	[77968, 4086]	[76069, 3847]
N, no. names (Nicknames)	[23627, 945]	[30152, 1309]	[45814, 1958]	[68445, 2787]	[79327, 3229]	[77006, 3238]
N, no. names (Soundex)	[25482, 566]	[32626, 705]	[48695, 855]	[72906, 1114]	[84544, 1199]	[83099, 8703]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. The "middle initials" row groups separately names with and without middle initials: "John" is a separate category from "John M.," but "John M." and "John H." belong to the same category. The "nicknames" row groups together common nicknames associated to a given first name: "Johnny" is the same as "John". The "Soundex codes" row groups together all names that have the same Soundex code. At the bottom of each panel, the number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income.

B Intergenerational Elasticities at 30-year Intervals

Appendix Table B1: 30-year elasticities

	(1)	(2)	(3)	(4)	(5)
	1850-1880	1870-1900	1880-1910	1900-1930	1910-1940
Sample:					
Sons: baseline	0.2311 (0.0185)	0.3108 (0.0165)	0.3189 (0.0156)	0.3871 (0.0123)	0.3794 (0.0115)
N, no. names	[37778, 1240]	[64972, 1645]	[83447, 2240]	[115713, 3313]	[107971, 3354]
Sons in law: baseline	0.2913 (0.0189)	0.3315 (0.0167)	0.3726 (0.0174)	0.4144 (0.0108)	0.3774 (0.0103)
N, no. names	[26311, 1093]	[43954, 1655]	[56494, 2105]	[87271, 3152]	[81667, 2957]
Sons: Individually linked data					
OLS	0.3849 (0.0214)		0.4186 (0.0129)		
First name fixed effects	0.3606 (0.0257)		0.4078 (0.0145)		
N	[3947]		[8847]		

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

C Imputation of Farmers' Income

The 1900 occupational-earnings distribution obtained from the tabulations in Preston and Haines (1991) does not include an estimate of average income for generic farm owners and farm tenants. We therefore experiment with two different methods for imputing farmers' income.

Our first imputation method assigns to all individuals coded as “farmer” in the 1950 coding scheme a weighted average of income for all farming occupations present in the Preston-Haines classification. Specifically, we record all the occupations in the 1910 Census that were coded as farmers in the 1950 occupational classification. We then calculate the average income (weighted by the sample frequencies in 1910) for the occupations with nonmissing income data based on the Preston-Haines tabulations, and assign this value (\$335.04) to all farmers (see Appendix Table C1 for the details of the calculation). As an alternative, we impute farmers' income based on the methodology originally described by Mitchell et al. (1922) and recently used by Abramitzky et al. (2012). For owner-occupier farmers, we calculate income as the difference between the value of farm products (augmented by the value of rent and food consumed by the family) and the total expenditures on labor, fertilizer, feed, seeds, threshing, taxes and maintenance (this results in an imputed annual income of \$576). For farm tenants, we imputed an income of \$334, which is the income for specialized farm workers (stock raisers, fruit growers, etc.) in the Preston-Haines tabulations. Under this second imputation rule we assign an income of \$475.93 to all “farmers” in the 1950 coding scheme.

References

- [1] Abramitzky, Ran; Leah Platt Boustan and Katherine Eriksson (2012). “Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration,” *American Economic Review*, 102(5), pp. 1832-1856.
- [2] Mitchell, Wesley; Willford I. King, Frederick R. Macaulay and Oswald W. Knauth (1922). *Income in the United States, Its Amount and Distribution, 1909-1919*, New York: National Bureau of Economic Research.
- [3] Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*, Princeton, NJ: Princeton University Press.

Appendix Table C1. Calculation of 1900 Wage Assigned to Farmers

Name of 1910 occupation assigned 1950 occupation of "farmer" at least once	Percent of "farmers" with this occupation in 1910	Default weight assigned to this occupation's wage (out of 1)	1901 wage assigned to this 1910 occupation
Farmers, general farms (owners)	57.55%		576
Farmers, general farms (tenants)	35.97%		334
Farm laborers, home farm	2.07%	0.384	255
Gardeners	1.37%	0.253	413
Dairy farmers	1.05%		576
Stock raisers	0.75%	0.139	334
Fruit growers	0.73%	0.135	334
Poultry raisers	0.17%	0.032	334
Florists	0.15%	0.028	593
Farm laborers, working out	0.04%	0.008	255
Nurserymen	0.04%	0.007	593
Dairy foremen, general farms	0.03%	0.006	750
Sugar cane farmer	0.02%		576
Apiarists	0.02%	0.003	334
Livery stable keepers and managers	0.01%	0.002	502
Coffee farmers	0.01%		576
Stock herders, drovers, and feeders	0.01%	0.001	334
Other and not specified pursuits	0.01%	0.001	334
Garden laborers	0.00%	0.001	255
Orchard and nursery laborers	0.00%	0.001	255
Corn shellers, hay balers, grain threshers, etc.	0.00%	0.000	255
Policemen	0.00%	0.000	887
Default wage to farmers:			335.04
Wage to farmers with income imputation:			475.93

Notes: The table records all the occupations in the 1910 Census that were coded as farmers in the 1950 occupational classification scheme, and the fraction of each occupation out of the total. Column (2) shows the weights used to calculate the default average wage for farmers. This is the average wage assigned to all farmers in Table 4, second row of each panel. Column (3) records the income data for specialized farming occupations from Preston and Haines (1991). See text for details on the calculation of income for owner-occupier farmers and farm tenants.

D Income Imputation Using 1850-1870 Wealth

In the third row of Table 5, we impute occupational income using a measure based on real estate property reported in 1850 to 1870 and personal property from 1860 and 1870. This is advantageous because it corresponds to the earlier periods in our analysis. We calculate mean value of real estate and personal property of household heads by occupation, pooling data from 1850 to 1870 and adjusting for price differences between these three decades. We adjust farmers' personal property downward by the average value of farm equipment and livestock in 1860 and 1870. Similarly, we adjust real estate property by subtracting the average cash value of farms between 1850 and 1870. In both cases we use the national averages calculated from the Census of Agriculture (Haines and ICPSR 2010).

References

- [1] Haines, Michael R., and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. <http://doi.org/10.3886/ICPSR02896.v3>