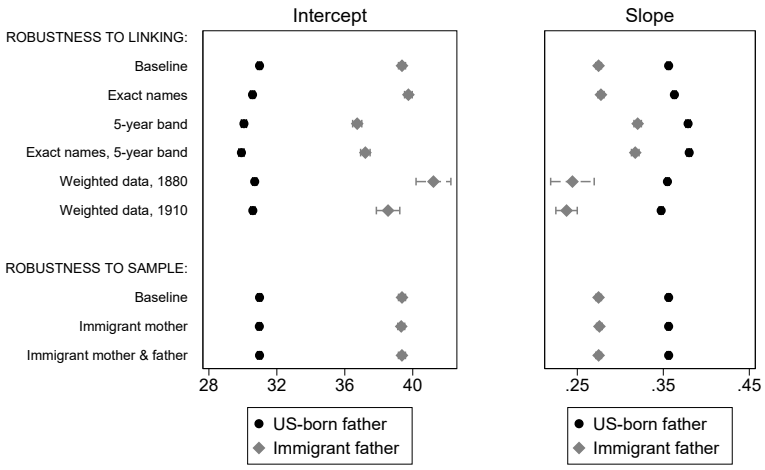
Online Appendix

Intergenerational Mobility of Immigrants in the US over Two Centuries

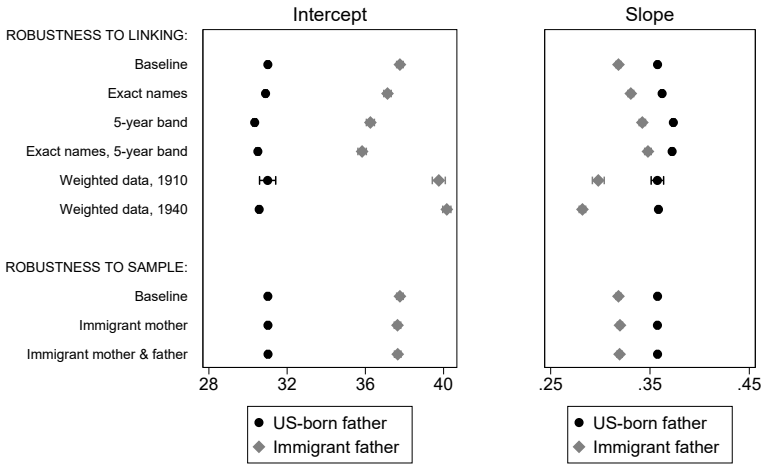
By RAN ABRAMITZKY, LEAH BOUSTAN, ELISA JÁCOME, AND SANTIAGO PÉREZ

Figure A1: Robustness to Weighting & Linking Algorithm and to Sample Construction

(a) 1880–1910 Cohort



(b) 1910–1940 Cohort



NOTE: These figures report the slope and intercept from regressions of son's rank on father's rank, using alternative weighting schemes and linking algorithms as well as alternative approaches for constructing the sample. After showing the baseline estimates, the next four estimates in each panel show the results from the rank-rank specifications that use alternative strategies for linking fathers and sons across the earlier and later Censuses, as described in Appendix A. "Weighted data" refers to weighing observations by the inverse probability of being a linked individual. Each cohort has two possible weights: one that uses childhood characteristics and the earlier cross-sectional Census, and one that uses adult characteristics and the later cross-sectional Census. "Immigrant mother" refers to a sample that uses the mother's birthplace to classify men as sons of immigrants. "Immigrant mother & father" refers to a sample that classifies men as immigrants only if both their mother and father were born outside of the United States.

Table A1: Sample Size, by Cohort and Father's Country of Origin

	1880–1910 cohort		1910–1940 cohort	
	US-born father	Immigrant father	US-born father	Immigrant father
White men, age 0–16	6,184,219	2,814,910	10,492,593	4,079,624
Linked men	1,432,861	602,726	3,224,578	1,068,064
living with father	1,287,253	544,197	2,931,570	965,839
white father, age 30–50	931,754	414,933	2,175,175	740,814
from largest sending countries	931,754	397,314	2,175,175	696,007
non-missing labor market outcomes	893,046	380,590	2,056,119	657,827
Share of white men age 0–16 in final sample	14.44	13.52	19.60	16.12

NOTE: The first row refers to the number of white men ages 0–16 in the earlier Census. Rows 2 through 6 show the sample size as we restrict the sample based on the fathers' and sons' characteristics. "Immigrant father" refers to sons whose fathers were born outside of the United States, and "largest sending countries" refers to the 17 countries considered in the historical cohorts.

Table A2: Comparison of Cross-Sectional and Linked Sample of Sons

(a) Sons of US-born Fathers

	1880-1910 cohort		1910-1940 cohort	
	Non-Linked	Linked	Non-Linked	Linked
Age	37.11	37.27	36.28	37.44
Farmer	0.24	0.25	0.30	0.33
White-collar	0.32	0.33	0.14	0.14
Skilled	0.25	0.25	0.32	0.34
Unskilled	0.19	0.17	0.24	0.19
Income	1,046.49	1,078.48	1,064.45	1,140.82
South	0.35	0.29	0.33	0.31

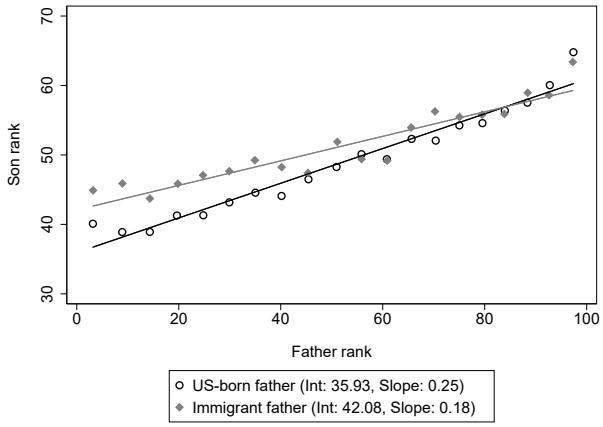
(b) Sons of Immigrant Fathers

	1880-1910 cohort		1910-1940 cohort	
	Non-Linked	Linked	Non-Linked	Linked
Immigrant (full sample)	0.26	0.29	0.22	0.24
Age	37.28	37.70	36.10	38.09
Farmer	0.27	0.27	0.36	0.37
White-collar	0.19	0.22	0.08	0.10
Skilled	0.34	0.33	0.35	0.36
Unskilled	0.20	0.18	0.21	0.17
Income	1,161.67	1,165.21	1,199.60	1,255.94
South	0.08	0.08	0.05	0.06

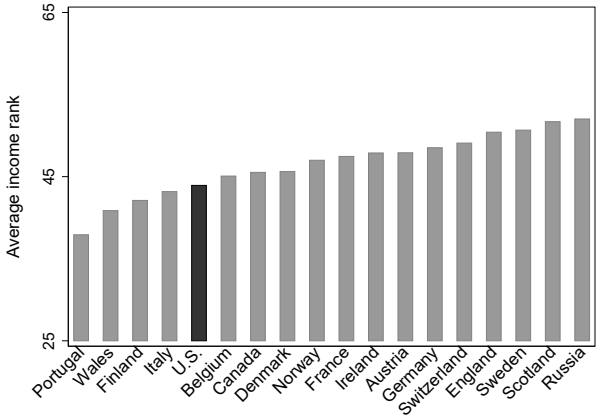
NOTE: In this table, the linked sample has not yet been restricted based on the fathers' characteristics (i.e., whether the father is white, aged 30–50, living with the child in the early Census, and without missing labor market outcomes).

Figure B1: Robustness to Using Actual Income for Fathers and Sons, 1910–1940 Cohort

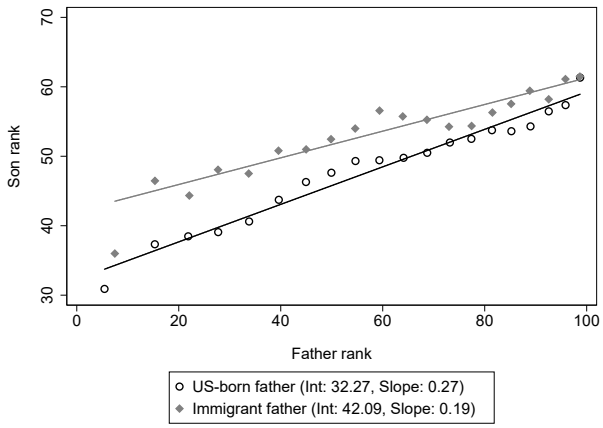
(a) Rank-rank, Actual Income



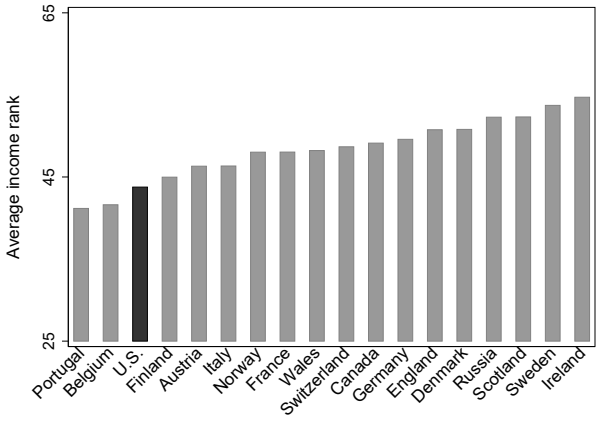
(b) Avg. Rank at 25th Percentile., Actual Income



(c) Rank-rank, Income Score



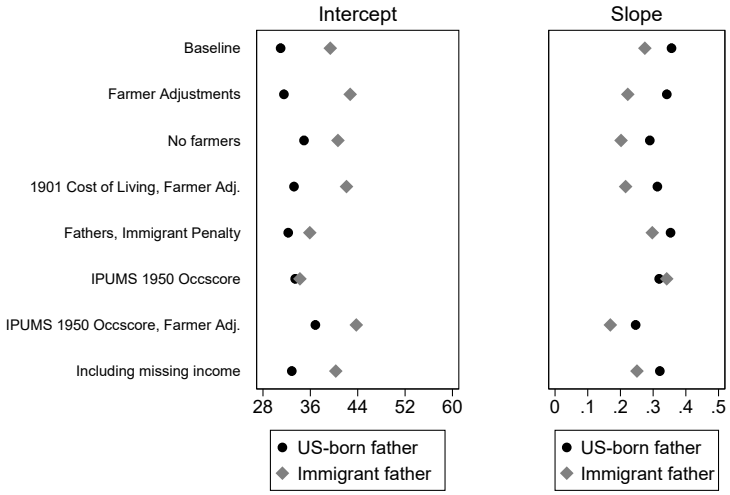
(d) Avg. Rank at 25th Percentile., Income Score



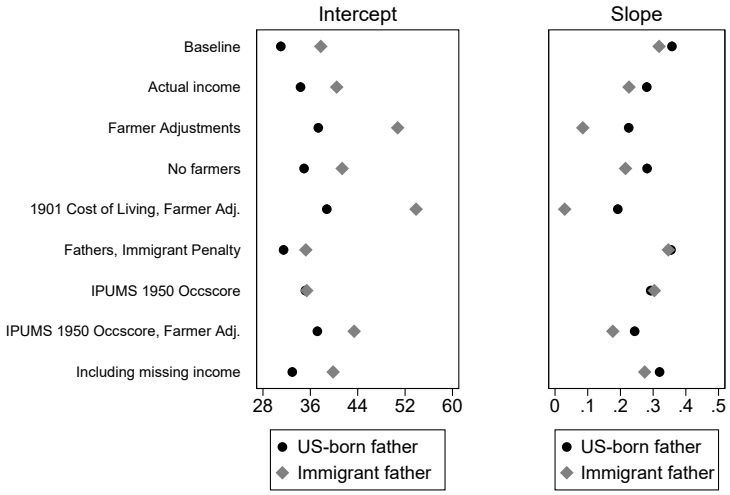
NOTE: These figures use a sample of father-son pairs for which we are able to link both fathers and sons between 1910 and 1940. The sample is restricted to father-son pairs in which both the father and the son received their income from wages. Panels (a) and (b) are based on individual-level income from wages from the 1940 Census, whereas panels (c) and (d) use the predicted income scores used throughout the paper.

Figure B2: Robustness to Using Alternative Income Measures

(a) 1880–1910 Cohort



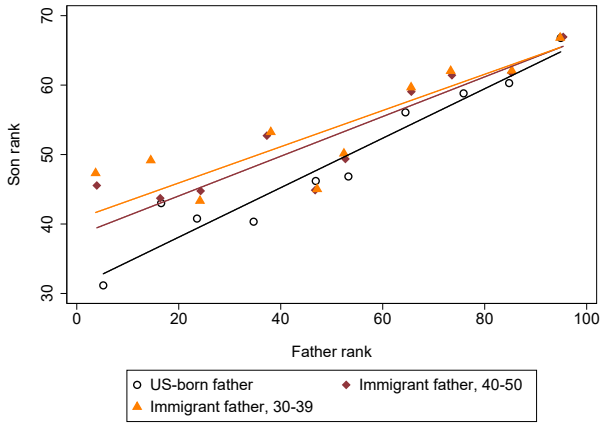
(b) 1910–1940 Cohort



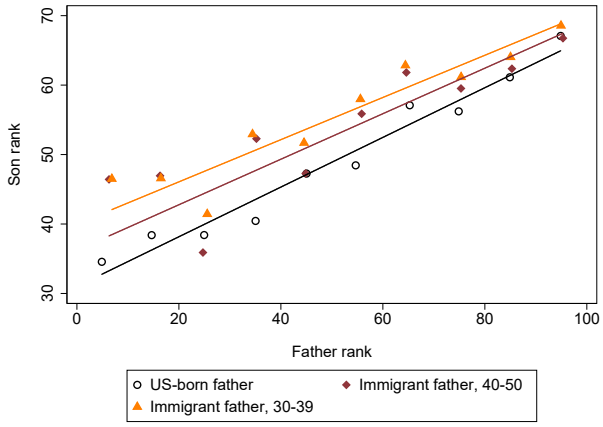
NOTE: These figures report the slope and intercept from regressions of son's rank on father's rank, using alternative approaches for measuring fathers' and sons' incomes. "Actual income" refers to specifications that use actual income for the 1940 sons, adjusted for self-employment and farm income (note we can only run this exercise for the 1910–1940 cohort). "Farmer Adjustments" keeps predicted income for non-farming occupations, but adjusts 1880 and 1910 farmers' incomes using county-level measures derived from the 1900 Census of Agriculture; 1940 sons are adjusted using the 1950 IPUMS "occscore" variable. "No farmers" drops fathers and sons who are farmers from the sample and re-runs the specification using ranks from the non-farming income distribution. "1901 Cost of Living, Farmer Adj." uses the farmer adjustments from above, and also adjusts non-farmer income for all fathers as well as 1910 sons using the average earnings in a person's occupation from the 1901 Cost of Living Survey. "Fathers, Immigrant Penalty" adjusts predicted income so that a father's birthplace is included in the income prediction (or immigrant status for farmers). "IPUMS 1950 Occscore" refers to using the IPUMS 1950 "occscore" variable to measure income. "IPUMS 1950 Occscore, Farmer Adj." refers to using the 1950 "occscore" variable for non-farmers, but scaling up 1880 and 1910 farmers' income so that farmers are ranked roughly 10 percentiles higher on the income distribution (at the same level as in our baseline approach). "Including missing income" refers to specifications in which fathers and sons with missing occupation or income are assigned an income of zero.

Figure C1: Rank-rank Correlations, by Age of Immigrant Fathers

(a) 1880–1910 Cohort: Rank-rank



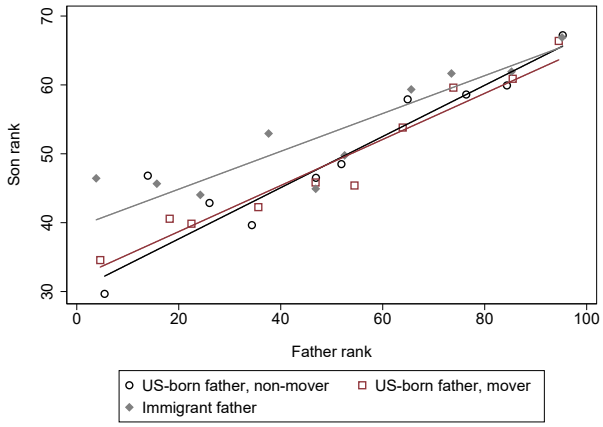
(b) 1910–1940 Cohort: Rank-rank



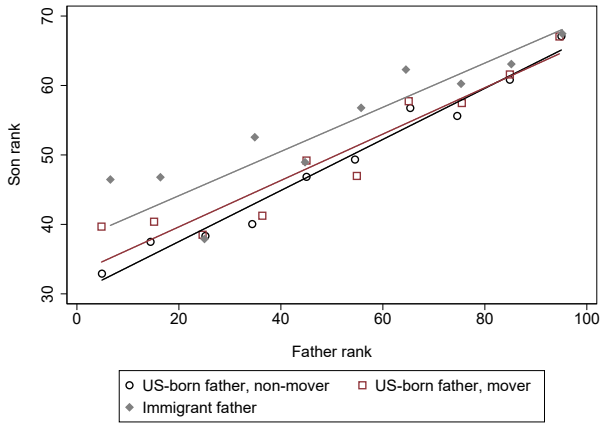
NOTE: The figures plot the mean income rank of children by father's income rank as well as the corresponding regression lines for each group. This figure divides sons with immigrant fathers into two groups based on the father's age: those whose fathers were aged 30–39 and those whose fathers were aged 40–50 in the earlier Census.

Figure C2: Rank-rank Correlations, by US-born Father's Internal Migration Status

(a) 1880–1910 Cohort: Rank-rank



(b) 1910–1940 Cohort: Rank-rank



NOTE: The figures plot the mean income rank of children by father's income rank as well as the corresponding regression lines for each group. This figure divides sons with US-born fathers into two groups based on the father's migrant status: those whose fathers moved within the US (i.e., their state of residence differs from their state of birth) and those whose fathers did not move.

Sensitivity of Results to using the Abramitzky, Mill and Pérez (2020) Linking Algorithm

Our baseline results use the linking algorithm developed in Abramitzky, Boustan, and Eriksson (2012, 2014), which is explained and evaluated in Abramitzky et al. (2019). In this section, we replicate our main results using the algorithm introduced in Abramitzky, Mill and Pérez (2020). This method uses the Expectation Maximization (EM) algorithm to combine age and name distances into a single score reflecting the probability that each potential pair of records is a true match. An advantage of this approach is that it enables researchers to create samples with very low false positive rates, although at the expense of generating smaller sample sizes (see Abramitzky et al. (2019) for an evaluation of this method). Indeed, Online Appendix Table D.1 shows that the samples that we generate in this case are between one fourth and one half the size of our baseline samples.

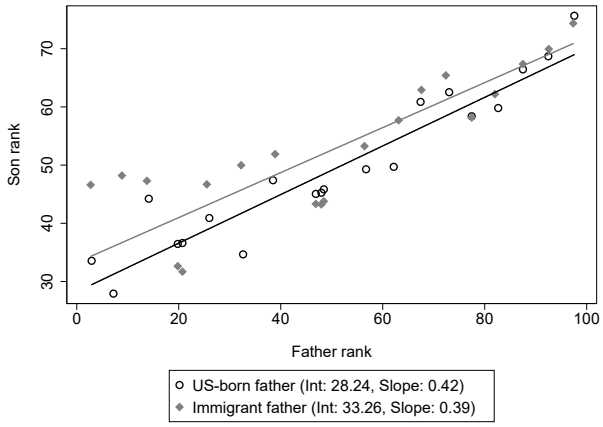
Online Appendix Figures D.1 to D.4 show that our main results hold when using this alternative linking strategy. Specifically, we continue to find that: (1) children of immigrants are more upwardly mobile (Figure D.1), (2) this is true for immigrants from most sending countries (Figures D.2 and D.3), and (3) differences in rates of upward mobility disappear once we compare children of the immigrants and children of the US born who grew up in similar areas (Figure D.4).

The main difference between the results that use this sample and our baseline results is that the rank-rank slopes are higher and the intercepts are lower, both for immigrants and for the US born. This is likely driven by the combination of two facts. First, this method likely generates lower false positive rates, which would tend to increase rank-rank associations. Second, because the method is more conservative with respect to which observations it deems as a match, it is more likely to select individuals with stable characteristics and who report their identifying information with accuracy.

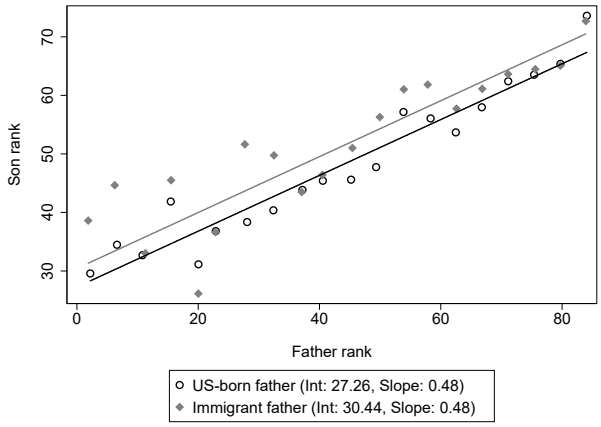
Finally, while the ranking of countries of origin in terms of mobility that we generate using this method is highly correlated with the ranking in our baseline approach, the exact ordering of countries is slightly different. Specifically, the correlation between the country ranks using the baseline samples and these alternative samples is 0.85 for the 1880-1910 cohort and 0.97 for the 1910-1940 cohort.

Figure D1: Intergenerational Mobility of Immigrants and the US-born, Rank-rank Correlations, EM Samples

(a) 1880–1910 Cohort



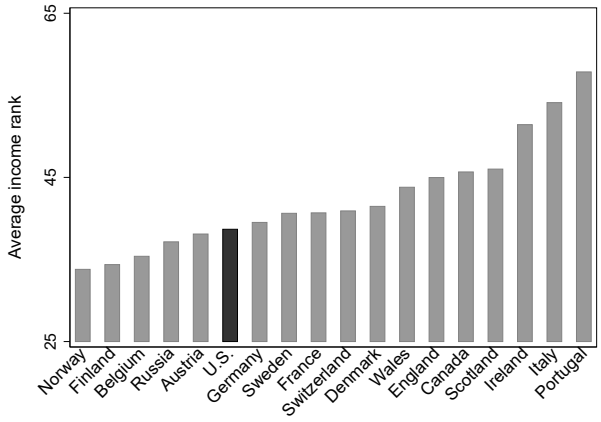
(b) 1910–to 1940 Cohort



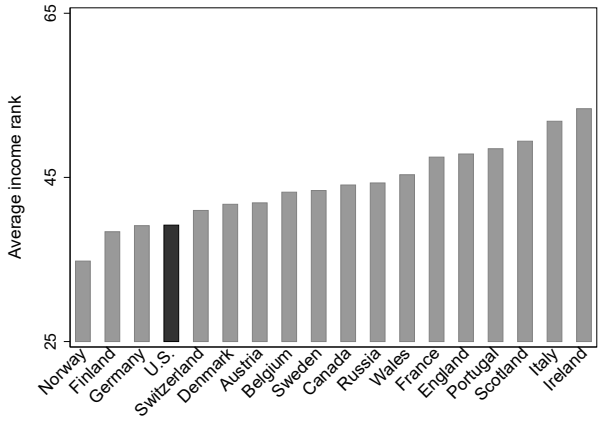
NOTE: This figure replicates Figure 2 in the main text using the linking algorithm in Abramitzky, Mill and Pérez (2020).

Figure D2: Average Income Rank for Children Born to 25th Percentile, by Father's Birthplace, EM Samples

(a) 1880–1910 Cohort



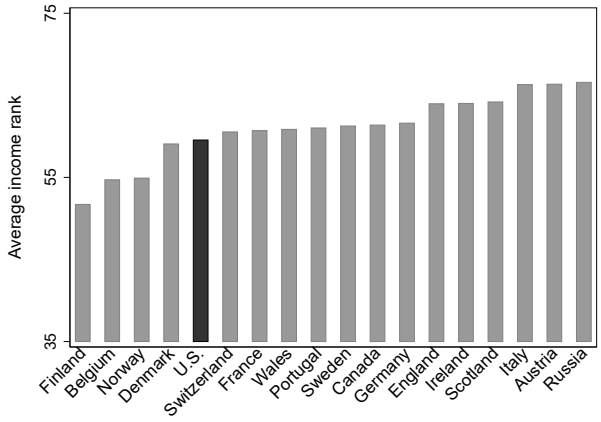
(b) 1910–1940 Cohort



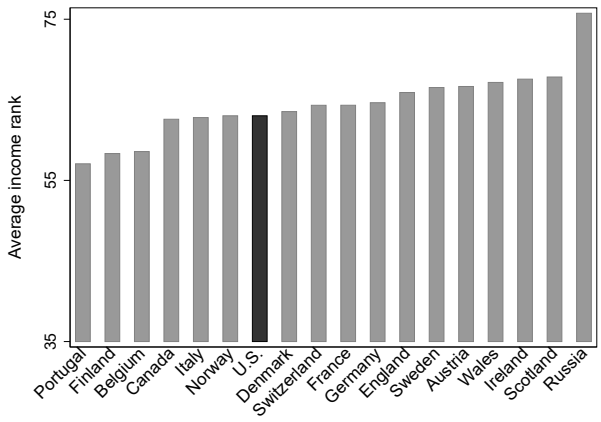
NOTE: This figure replicates Figure 3 in the main text using the linking algorithm in Abramitzky, Mill and Pérez (2020).

Figure D3: Average Income Rank for Children Born to 75th Percentile, by Father's Birthplace, EM Samples

(a) 1880–1910 Cohort



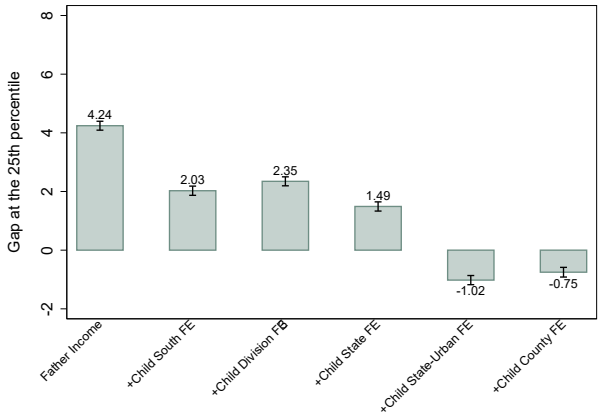
(b) 1910–1940 Cohort



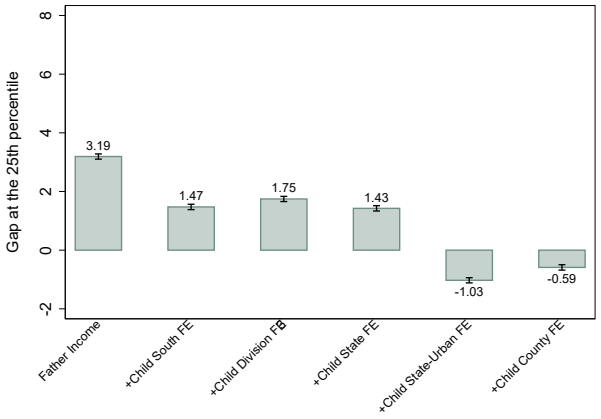
NOTE: This figure replicates Figure 4 in the main text using the linking algorithm in Abramitzky, Mill and Pérez (2020).

Figure D4: Intergenerational Gap at the 25th percentile, Comparing Children in Similar Childhood Locations, EM Samples

(a) 1880–1910 Cohort



(b) 1910–1940 Cohort



NOTE: This figure replicates Figure 7 in the main text using the linking algorithm in Abramitzky, Mill and Pérez (2020).

Table D1: Sample Size, by Cohort and Father's Country of Origin, EM samples

	1880–1910 cohort		1910–1940 cohort	
	US-born father	Immigrant father	US-born father	Immigrant father
White men, age 0–16	6,184,219	2,814,910	10,492,593	4,079,624
Linked men	504,963	183,660	1,517,949	457,833
living with father	461,021	168,543	1,395,790	419,681
white father, age 30–50	332,508	127,964	1,037,349	319,322
from largest sending countries	332,508	122,675	1,037,349	302,560
non-missing labor market outcomes	272,430	97,560	870,967	245,046
Share of white men age 0–16 in final sample	4.41	3.47	8.30	6.01

NOTE: This table shows sample size by cohort and father's country of origin when using the EM samples.