Online Appendix for

"Black Economic Progress in the Jim Crow South: Evidence from Rosenwald Schools"

A. R. Shariq Mohammed and Paul Mohnen

Online Appendix A. Online Appendix Tables and Figures

Table A1: Linked Samples, Match Rates and Sample Sizes

	Black men born in the South (1910–1919)	Black women born in the South (1910–1919)
Panel A. SS-5 linked dataset		
Individuals in SS-5 records	275,887	325,099
Links to 1940 Census at 95% precision	63,003	72,430
Match rate to 1940 Census at 95% precision	0.228	0.223
Links to 1940 and 1920 Censuses at 95% precision	28,521	31,155
Match rate to 1940 and 1920 Censuses at 95% precision	0.103	0.096
Panel B. IPUMS Multigenerational Longitudinal Panel (MLP)		
Individuals in 1920 Census	1,097,206	1,100,288
Links to 1930 Census	483,197	335,845
Match rate to 1930 Census	0.44	0.305
Links to 1930 and 1940 Censuses	127,955	47,603
Match rate to 1930 and 1940 Censuses	0.117	0.043
Panel C. Census Linking Project (CLP), standard ABE method		
Individuals in 1920 Census	1,097,206	—
Links to 1940 Census	134,865	—
Match rate to 1940 Census	0.123	—
Panel D. Census Linking Project (CLP), conservative ABE method		
Individuals in 1920 Census	1,097,206	—
Links to 1940 Census	73,465	—
Match rate to 1940 Census	0.067	—

Notes: The targeted precision rate in Panel A is relative to hand links in the training data (see Online Appendix B for details).

Sources: See text for details.

Table A2: Representativeness of Linked Sample, Southern Black Men (1910–1919)

	Mean			Difference (two-sided *t*-test)	
	Population (10%)	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Born in 1910–1913	0.388	0.2	0.391	-0.188	0.004
				[0]	[0.275]
Born in 1914–1916	0.303	0.356	0.301	0.053	-0.003
				[0]	[0.38]
Born in 1917–1919	0.309	0.444	0.308	0.135	-0.001
				[0]	[0.78]
Born in AL	0.105	0.096	0.105	-0.01	0
				[0]	[0.946]
Born in AR	0.045	0.052	0.044	0.006	-0.002
				[0]	[0.236]
Born in FL	0.033	0.041	0.032	0.008	0
				[0]	[0.718]
Born in GA	0.139	0.116	0.138	-0.023	-0.002
				[0]	[0.518]
Born in KY	0.02	0.025	0.019	0.005	0
				[0]	[0.654]
Born in LA	0.081	0.091	0.08	0.01	-0.001
				[0]	[0.541]
Born in MD	0.022	0.032	0.022	0.01	0
				[0]	[0.909]
Born in MS	0.116	0.098	0.118	-0.019	0.002
				[0]	[0.345]
Born in NC	0.107	0.105	0.108	-0.003	0
				[0.218]	[0.961]
Born in OK	0.014	0.022	0.014	0.008	0
				[0]	[0.995]
Born in SC	0.111	0.1	0.114	-0.011	0.003
				[0]	[0.138]
Born in TN	0.048	0.05	0.047	0.002	-0.001
				[0.144]	[0.333]
Born in TX	0.082	0.085	0.083	0.003	0.001
				[0.132]	[0.754]
Born in VA	0.076	0.089	0.076	0.013	0
				[0]	[0.856]
Live in Northeast	0.059	0.052	0.06	-0.007	0
				[0]	[0.794]
Live in Midwest	0.07	0.069	0.069	-0.001	-0.001
				[0.689]	[0.634]

(table continues on next page)

Table A2 (cont.): Representativeness of Linked Sample, Southern Black Men (1910–1919)

	Mean			Difference (two-sided *t*-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(10%)				
	(1)	(2)	(3)	(4)	(5)
Live in South	0.86	0.866	0.859	0.005	-0.002
				[0.019]	[0.51]
Live in West	0.011	0.013	0.012	0.002	0.002
				[0.003]	[0.008]
Live in state of birth	0.733	0.767	0.742	0.034	0.009
				[0]	[0.003]
Live in urban area	0.439	0.406	0.436	-0.033	-0.003
				[0]	[0.36]
Live on farm	0.375	0.416	0.376	0.041	0.001
				[0]	[0.77]
Homeowner	0.192	0.265	0.19	0.074	-0.001
				[0]	[0.682]
Head of household (or spouse)	0.43	0.369	0.447	-0.061	0.018
				[0]	[0]
Married	0.554	0.481	0.556	-0.074	0.002
				[0]	[0.558]
Any kids	0.282	0.262	0.282	-0.02	-0.001
				[0]	[0.859]
Years of education	5.976	6.932	5.932	0.956	-0.045
				[0]	[0.057]
In labor force	0.906	0.908	0.916	0.002	0.011
				[0.353]	[0]
Self-employed	0.141	0.142	0.148	0.002	0.008
				[0.505]	[0.001]
Occ.: skilled	0.033	0.044	0.037	0.012	0.004
				[0]	[0.002]
Occ.: farmer (homeowner)	0.015	0.022	0.017	0.007	0.001
				[0]	[0.102]
Occ.: farmer (renter)	0.113	0.106	0.117	-0.007	0.004
				[0.001]	[0.099]
Occ.: blue collar	0.143	0.146	0.154	0.003	0.011
				[0.238]	[0]
Occ.: unskilled	0.598	0.584	0.589	-0.014	-0.009
				[0]	[0.006]
Annual wage income	306.183	295.863	321.347	-10.32	15.164
				[0]	[0]
N	95,540	28,521			

Notes: Means in columns 3 and 5 are weighted using inverse propensity score weights. *p*-value for two-sided *t*-test of equality of means in brackets (columns 4–5).
Sources: See text for details.

Table A3: Representativeness of Linked Sample, Southern Black Women (1910–1919)

	Mean			Difference (two-sided *t*-test)	
	Population (10%)	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Born in 1910–1913	0.383	0.222	0.383	-0.161	0
				[0]	[0.99]
Born in 1914–1916	0.31	0.346	0.307	0.035	-0.003
				[0]	[0.29]
Born in 1917–1919	0.307	0.432	0.31	0.125	0.003
				[0]	[0.296]
Born in AL	0.11	0.097	0.111	-0.013	0.002
				[0]	[0.37]
Born in AR	0.045	0.046	0.045	0.001	-0.001
				[0.688]	[0.585]
Born in FL	0.036	0.039	0.036	0.003	0
				[0.011]	[0.889]
Born in GA	0.143	0.122	0.143	-0.021	0
				[0]	[0.932]
Born in KY	0.019	0.03	0.018	0.01	-0.001
				[0]	[0.417]
Born in LA	0.083	0.079	0.083	-0.003	0
				[0.049]	[0.967]
Born in MD	0.02	0.036	0.02	0.016	0
				[0]	[0.818]
Born in MS	0.116	0.085	0.113	-0.031	-0.002
				[0]	[0.251]
Born in NC	0.102	0.117	0.104	0.015	0.002
				[0]	[0.423]
Born in OK	0.014	0.02	0.014	0.006	0
				[0]	[0.819]
Born in SC	0.111	0.096	0.114	-0.016	0.002
				[0]	[0.317]
Born in TN	0.045	0.053	0.043	0.008	-0.002
				[0]	[0.09]
Born in TX	0.082	0.08	0.082	-0.002	0
				[0.385]	[0.778]
Born in VA	0.073	0.099	0.073	0.026	0
				[0]	[0.855]
Live in Northeast	0.074	0.076	0.075	0.002	0.001
				[0.312]	[0.757]
Live in Midwest	0.073	0.079	0.081	0.006	0.008
				[0]	[0]

(table continues on next page)

Table A3 (cont.): Representativeness of Linked Sample, Southern Black Women (1910–1919)

	Mean			Difference (two-sided *t*-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(10%)				
	(1)	(2)	(3)	(4)	(5)
Live in South	0.844	0.834	0.835	-0.009	-0.009
				[0]	[0]
Live in West	0.009	0.011	0.01	0.002	0
				[0.018]	[0.566]
Live in state of birth	0.716	0.728	0.716	0.012	0.001
				[0]	[0.863]
Live in urban area	0.51	0.495	0.507	-0.014	-0.003
				[0]	[0.322]
Live on farm	0.321	0.328	0.322	0.006	0.001
				[0.037]	[0.768]
Homeowner	0.179	0.253	0.18	0.074	0.001
				[0]	[0.64]
Head of household (or spouse)	0.549	0.458	0.564	-0.091	0.015
				[0]	[0]
Married	0.684	0.564	0.675	-0.12	-0.009
				[0]	[0.004]
Any kids	0.44	0.409	0.446	-0.031	0.006
				[0]	[0.053]
Years of education	6.98	8.142	6.995	1.162	0.014
				[0]	[0.511]
In labor force	0.447	0.434	0.409	-0.013	-0.038
				[0]	[0]
Self-employed	0.027	0.023	0.025	-0.004	-0.002
				[0]	[0.058]
Occ.: skilled	0.033	0.061	0.038	0.028	0.005
				[0]	[0]
Occ.: farmer (homeowner)	0.001	0.001	0.001	0	0
				[0.109]	[0.593]
Occ.: farmer (renter)	0.006	0.004	0.005	-0.001	0
				[0.002]	[0.662]
Occ.: blue collar	0.035	0.037	0.038	0.002	0.003
				[0.07]	[0.023]
Occ.: unskilled	0.386	0.34	0.341	-0.045	-0.045
				[0]	[0]
Annual wage income	110.132	123.276	110.223	13.144	0.091
				[0]	[0.955]
N	111,084	31,155			

Notes: Means in columns 3 and 5 are weighted using inverse propensity score weights. *p*-value for two-sided *t*-test of equality of means in brackets (columns 4–5).

Sources: See text for details.

Table A4: The Impact of Rosenwald Schools on Educational Attainment in 1940: Sensitivity to Alternative Controls and Sample Restrictions

	Dependent variable: Years of education in 1940		
	(1)	(2)	(3)
Panel A. Black men (1910–1919)			
Rosenwald exposure	0.438***	0.505***	0.946***
	(0.144)	(0.161)	(0.275)
R^2	0.134	0.081	0.096
N	23,122	23,122	8,890
Panel B. Black women (1910–1919)			
Rosenwald exposure	0.456***	0.473***	0.820***
	(0.143)	(0.153)	(0.245)
R^2	0.117	0.077	0.082
N	24,380	24,380	9,455
Drop 1920 household controls		x	x
Restrict to rural-only counties			x

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate), unless otherwise specified. Column 3 restricts the sample to individuals whose 1920 county of residence had a rural share of 1 in the 1910, 1920 and 1930 Censuses. Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.
Sources: See text for details.

Table A5: The Impact of Rosenwald Schools on Educational Attainment in 1940: Placebo Cohorts

	Dependent variable: Outcome in 1940			
	Years of education	Some high school or more (9+ years of education)	High school or more (12+ years of education)	College or more (16+ years of education)
	(1)	(2)	(3)	(4)
Panel A. Black men (1890–1897)				
Future Rosenwald exp.	-0.182	0.019	0.0002	-0.004
	(0.280)	(0.025)	(0.015)	(0.007)
R^2	0.107	0.062	0.051	0.086
N	5,842	5,842	5,842	5,842
Panel B. Black women (1890–1897)				
Future Rosenwald exp.	0.239	0.021	0.013	0.002
	(0.345)	(0.029)	(0.018)	(0.008)
R^2	0.105	0.060	0.052	0.086
N	4,021	4,021	4,021	4,021

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1900 state of residence fixed effects, 1900 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Future Rosenwald exposure based on shifting year of birth 20 years later. Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A6: Controlling for County-Specific Unobservable Factors

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rosenwald exposure	0.545	0.038	0.039	0.055	-0.031	-0.176**	0.211***	-0.012
× Black men (1910–1914)	(0.478)	(0.058)	(0.052)	(0.076)	(0.058)	(0.089)	(0.074)	(0.032)
Rosenwald exposure	0.126	0.016	0.029	0.021	0.011	-0.108*	0.127***	-0.025
× Black men (1915–1919)	(0.295)	(0.037)	(0.037)	(0.056)	(0.040)	(0.058)	(0.045)	(0.020)
Rosenwald exposure	0.510	0.075	0.168***	0.151*	0.035	-0.238***	0.175***	0.020
× Black women (1910–1914)	(0.461)	(0.060)	(0.062)	(0.090)	(0.057)	(0.088)	(0.068)	(0.044)
Rosenwald exposure	-0.041	0.032	0.076*	-0.039	0.043	-0.104*	0.082*	-0.030
× Black women (1915–1919)	(0.299)	(0.039)	(0.039)	(0.060)	(0.043)	(0.057)	(0.045)	(0.022)
R^2	0.185	0.110	0.341	0.190	0.174	0.115	0.087	0.070
N	47,502	47,502	48,236	30,883	31,346	31,346	31,346	31,346

Notes: All regressions include 1920 county of residence fixed effect, sex-by-cohort fixed effects, sex-by-state of birth fixed effects, cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effect, and 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A7: Alternative Linked Datasets, Black Men and Women (1910–1914)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Black men (1910–1914), Multigenerational Longitudinal Panel (MLP)								
Rosenwald exposure	0.495**	0.038*	0.012	0.012	-0.029	0.035	-0.019	0.008
	(0.218)	(0.019)	(0.012)	(0.024)	(0.031)	(0.034)	(0.025)	(0.008)
R^2	0.154	0.067	0.007	0.134	0.096	0.037	0.039	0.021
N	37,712	37,712	38,276	35,867	35,979	35,979	35,979	35,979
Panel B. Black men (1910–1914), Census Linking Project (CLP), standard ABE method								
Rosenwald exposure	0.442***	0.038***	-0.006	0.008	0.004	-0.022	0.008	0.008
	(0.139)	(0.013)	(0.009)	(0.017)	(0.020)	(0.022)	(0.017)	(0.006)
R^2	0.078	0.040	0.005	0.086	0.070	0.028	0.017	0.008
N	53,190	53,190	54,189	50,211	50,323	50,323	50,323	50,323
Panel C. Black men (1910–1914), Census Linking Project (CLP), conservative ABE method								
Rosenwald exposure	0.607***	0.047***	-0.011	0.002	0.0001	0.00003	-0.007	0.005
	(0.180)	(0.016)	(0.012)	(0.022)	(0.024)	(0.028)	(0.022)	(0.008)
R^2	0.103	0.052	0.009	0.099	0.074	0.032	0.024	0.011
N	28,789	28,789	29,315	27,313	27,372	27,372	27,372	27,372
Panel D. Black women (1910–1914), Multigenerational Longitudinal Panel (MLP)								
Rosenwald exposure	0.169	0.074	0.093	0.087	0.019	-0.038	0.004	0.024
	(0.410)	(0.048)	(0.072)	(0.101)	(0.015)	(0.073)	(0.063)	(0.037)
R^2	0.160	0.118	0.050	0.190	0.034	0.103	0.093	0.089
N	10,476	10,476	10,640	5,360	5,468	5,468	5,468	5,468

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A8: Alternative Linked Datasets, Black Men and Women (1915–1919)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Black men (1915–1919), Multigenerational Longitudinal Panel (MLP)								
Rosenwald exposure	0.321**	0.032**	-0.011	-0.007	0.005	-0.009	0.007	-0.003
	(0.149)	(0.015)	(0.010)	(0.016)	(0.015)	(0.019)	(0.012)	(0.003)
R^2	0.140	0.073	0.015	0.139	0.073	0.028	0.028	0.011
N	65,536	65,536	66,458	58,642	59,523	59,523	59,523	59,523
Panel B. Black men (1915–1919), Census Linking Project (CLP), standard ABE method								
Rosenwald exposure	0.118	0.016	-0.003	-0.011	-0.003	0.007	0.004	-0.007**
	(0.093)	(0.010)	(0.007)	(0.012)	(0.010)	(0.013)	(0.009)	(0.003)
R^2	0.085	0.047	0.012	0.104	0.071	0.026	0.016	0.006
N	54,568	54,568	55,536	49,044	49,800	49,800	49,800	49,800
Panel C. Black men (1915–1919), Census Linking Project (CLP), conservative ABE method								
Rosenwald exposure	0.248**	0.030**	0.001	-0.008	-0.003	0.002	0.006	-0.004
	(0.123)	(0.013)	(0.009)	(0.017)	(0.014)	(0.017)	(0.012)	(0.004)
R^2	0.106	0.058	0.017	0.121	0.071	0.029	0.021	0.011
N	30,240	30,240	30,729	27,246	27,638	27,638	27,638	27,638
Panel D. Black women (1915–1919), Multigenerational Longitudinal Panel (MLP)								
Rosenwald exposure	0.297	0.049*	-0.005	0.028	-0.006	0.034	-0.015	0.001
	(0.218)	(0.029)	(0.031)	(0.031)	(0.005)	(0.028)	(0.019)	(0.014)
R^2	0.137	0.089	0.026	0.150	0.020	0.053	0.043	0.061
N	25,672	25,672	26,087	12,016	12,460	12,460	12,460	12,460

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A9: Alternative Results Based on Non-Migrants in the 1940 Census Cross-Section

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Black men (1910–1914), living in a rural area in 1940, in their state of birth, and in the same house as in 1935								
Rosenwald exposure	1.040***	0.085***	0.019	0.005	0.002	0.019	-0.028	0.005
	(0.215)	(0.020)	(0.015)	(0.041)	(0.047)	(0.039)	(0.037)	(0.008)
R^2	0.091	0.032	0.005	0.202	0.103	0.058	0.028	0.002
N	55,055	55,055	55,947	52,247	52,367	52,367	52,367	52,367
Panel B. Black men (1915–1919), living in a rural area in 1940, in their state of birth, and in the same house as in 1935								
Rosenwald exposure	0.692***	0.075***	-0.005	-0.021	0.012	-0.003	-0.013	-0.001
	(0.165)	(0.017)	(0.011)	(0.022)	(0.018)	(0.022)	(0.016)	(0.002)
R^2	0.095	0.046	0.013	0.226	0.059	0.031	0.023	0.003
N	81,203	81,203	82,375	73,034	73,696	73,696	73,696	73,696
Panel C. Black women (1910–1914), living in a rural area in 1940, in their state of birth, and in the same house as in 1935								
Rosenwald exposure	1.099***	0.102***	0.077**	0.052*	-0.008	-0.020	-0.025*	0.054***
	(0.257)	(0.027)	(0.039)	(0.027)	(0.011)	(0.027)	(0.014)	(0.019)
R^2	0.087	0.041	0.015	0.194	0.028	0.013	0.022	0.014
N	61,263	61,263	62,161	19,846	20,054	20,054	20,054	20,054
Panel D. Black women (1915–1919), living in a rural area in 1940, in their state of birth, and in the same house as in 1935								
Rosenwald exposure	0.743***	0.097***	0.027	0.009	-0.003	-0.001	-0.014	0.020*
	(0.191)	(0.022)	(0.022)	(0.018)	(0.005)	(0.018)	(0.010)	(0.011)
R^2	0.090	0.058	0.014	0.227	0.022	0.019	0.013	0.012
N	73,366	73,366	74,499	26,263	26,863	26,863	26,863	26,863

Notes: Sample restricted to Black men and women living in a rural area in 1940, in their state of birth, and reporting living in the same house in 1940 as in 1935. All regressions include cohort-by-state of birth fixed effects, and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A10: Robustness Checks, Black Men (1910–1914)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Rosenwald exposure based on ages 7–17								
Rosenwald exposure	0.718***	0.054*	-0.056**	0.027	-0.050	0.006	0.019	0.017
	(0.269)	(0.029)	(0.026)	(0.038)	(0.041)	(0.049)	(0.038)	(0.015)
R^2	0.134	0.069	0.031	0.121	0.090	0.047	0.033	0.039
N	6,883	6,883	6,989	6,575	6,588	6,588	6,588	6,588
Panel B. Restrict to individuals living in county of birth in 1920								
Rosenwald exposure	1.249***	0.097**	-0.049	0.024	-0.025	-0.069	0.047	0.039*
	(0.367)	(0.041)	(0.036)	(0.045)	(0.052)	(0.064)	(0.056)	(0.020)
R^2	0.143	0.069	0.029	0.106	0.094	0.043	0.026	0.026
N	4,536	4,536	4,602	4,325	4,331	4,331	4,331	4,331
Panel C. Alternative linked sample based on 90% precision linkages								
Rosenwald exposure	0.653**	0.062**	-0.033	0.030	-0.020	-0.004	0.001	0.016
	(0.281)	(0.030)	(0.024)	(0.034)	(0.038)	(0.048)	(0.038)	(0.015)
R^2	0.119	0.059	0.020	0.117	0.081	0.040	0.032	0.023
N	9,192	9,192	9,342	8,756	8,775	8,775	8,775	8,775
Panel D. Inverse probability weights								
Rosenwald exposure	0.743**	0.072**	-0.026	0.064*	-0.026	-0.076	0.069*	0.024
	(0.296)	(0.036)	(0.022)	(0.034)	(0.039)	(0.047)	(0.036)	(0.019)
R^2	0.147	0.090	0.015	0.148	0.083	0.039	0.036	0.037
N	6,883	6,883	6,989	6,575	6,588	6,588	6,588	6,588
Panel E. Unweighted								
Rosenwald exposure	0.879***	0.085**	-0.031	0.067*	-0.039	-0.063	0.066*	0.027
	(0.289)	(0.036)	(0.022)	(0.034)	(0.039)	(0.048)	(0.037)	(0.018)
R^2	0.142	0.087	0.019	0.132	0.089	0.044	0.034	0.039
N	6,883	6,883	6,989	6,575	6,588	6,588	6,588	6,588

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights in Panels A–C and inverse probability weights in Panel D (see text for details). Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

Table A11: Robustness Checks, Black Women (1910–1914)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Rosenwald exposure based on ages 7–17								
Rosenwald exposure	0.795***	0.078**	0.052	0.121**	0.011	-0.073	-0.007	0.064**
	(0.249)	(0.036)	(0.041)	(0.059)	(0.019)	(0.050)	(0.036)	(0.031)
R^2	0.126	0.080	0.032	0.162	0.066	0.066	0.076	0.047
N	7,711	7,711	7,821	3,201	3,250	3,250	3,250	3,250
Panel B. Restrict to individuals living in county of birth in 1920								
Rosenwald exposure	0.782**	0.115**	0.124**	0.118	0.024	-0.097	0.008	0.043
	(0.335)	(0.048)	(0.060)	(0.076)	(0.023)	(0.064)	(0.049)	(0.042)
R^2	0.137	0.084	0.026	0.158	0.068	0.060	0.058	0.063
N	5,100	5,100	5,176	2,104	2,137	2,137	2,137	2,137
Panel C. Alternative linked sample based on 90% precision linkages								
Rosenwald exposure	0.542**	0.056*	0.081**	0.094*	0.033	-0.109**	0.032	0.037
	(0.228)	(0.034)	(0.038)	(0.054)	(0.023)	(0.047)	(0.033)	(0.029)
R^2	0.103	0.062	0.027	0.144	0.044	0.051	0.055	0.049
N	11,140	11,140	11,298	4,703	4,766	4,766	4,766	4,766
Panel D. Inverse probability weights								
Rosenwald exposure	1.084***	0.115***	0.102**	0.095*	0.011	-0.052	-0.015	0.034
	(0.287)	(0.044)	(0.040)	(0.053)	(0.012)	(0.052)	(0.028)	(0.046)
R^2	0.126	0.090	0.032	0.184	0.049	0.059	0.054	0.064
N	7,711	7,711	7,821	3,201	3,250	3,250	3,250	3,250
Panel E. Unweighted								
Rosenwald exposure	1.063***	0.114**	0.098**	0.123**	0.011	-0.087	-0.009	0.063
	(0.289)	(0.045)	(0.039)	(0.057)	(0.013)	(0.053)	(0.030)	(0.046)
R^2	0.127	0.090	0.031	0.161	0.058	0.063	0.058	0.066
N	7,711	7,711	7,821	3,201	3,250	3,250	3,250	3,250

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights in Panels A–C and inverse probability weights in Panel D (see text for details). Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.
Sources: See text for details.

Table A12: Robustness Checks, Black Men (1915–1919)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Rosenwald exposure based on ages 7–17								
Rosenwald exposure	0.244*	0.035**	-0.028*	-0.005	0.002	0.006	0.003	-0.008
	(0.136)	(0.015)	(0.016)	(0.018)	(0.018)	(0.021)	(0.015)	(0.006)
R^2	0.128	0.069	0.019	0.128	0.071	0.031	0.032	0.020
N	16,239	16,239	16,484	14,505	14,703	14,703	14,703	14,703
Panel B. Restrict to individuals living in county of birth in 1920								
Rosenwald exposure	0.249	0.020	-0.034*	-0.008	-0.007	0.043	-0.016	-0.014*
	(0.159)	(0.018)	(0.019)	(0.021)	(0.021)	(0.027)	(0.018)	(0.007)
R^2	0.137	0.070	0.019	0.107	0.066	0.030	0.032	0.010
N	11,247	11,247	11,405	10,034	10,171	10,171	10,171	10,171
Panel C. Alternative linked sample based on 90% precision linkages								
Rosenwald exposure	0.253**	0.027**	-0.015	-0.009	-0.001	0.020	-0.004	-0.011*
	(0.121)	(0.014)	(0.014)	(0.017)	(0.017)	(0.020)	(0.014)	(0.006)
R^2	0.112	0.059	0.018	0.121	0.069	0.027	0.026	0.016
N	21,412	21,412	21,734	19,027	19,294	19,294	19,294	19,294
Panel D. Inverse probability weights								
Rosenwald exposure	0.308**	0.047**	-0.033**	0.005	-0.007	0.013	0.007	-0.009
	(0.138)	(0.019)	(0.015)	(0.017)	(0.018)	(0.022)	(0.014)	(0.007)
R^2	0.146	0.086	0.025	0.160	0.071	0.033	0.035	0.021
N	16,239	16,239	16,484	14,505	14,703	14,703	14,703	14,703
Panel E. Unweighted								
Rosenwald exposure	0.301**	0.044**	-0.031**	-0.003	-0.0003	0.010	0.004	-0.008
	(0.134)	(0.018)	(0.015)	(0.017)	(0.018)	(0.022)	(0.015)	(0.007)
R^2	0.138	0.081	0.023	0.142	0.072	0.033	0.033	0.021
N	16,239	16,239	16,484	14,505	14,703	14,703	14,703	14,703

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights in Panels A–C and inverse probability weights in Panel D (see text for details). Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.
Sources: See text for details.

Table A13: Robustness Checks, Black Women (1915–1919)

	Dependent variable: Outcome in 1940							
	Years of education	9+ years of education	In labor force	Occupational income score (log)	Farmer	Unskilled job	Blue-collar job	White-collar job
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Rosenwald exposure based on ages 7–17								
Rosenwald exposure	0.313**	0.058***	0.001	-0.020	-0.001	0.011	-0.010	-0.006
	(0.143)	(0.018)	(0.020)	(0.024)	(0.008)	(0.025)	(0.016)	(0.013)
R^2	0.107	0.065	0.024	0.146	0.039	0.029	0.037	0.035
N	16,669	16,669	16,942	6,602	6,805	6,805	6,805	6,805
Panel B. Restrict to individuals living in county of birth in 1920								
Rosenwald exposure	0.326*	0.064***	-0.010	-0.047*	-0.012	0.021	-0.001	-0.003
	(0.170)	(0.022)	(0.024)	(0.026)	(0.009)	(0.029)	(0.019)	(0.015)
R^2	0.121	0.071	0.022	0.114	0.036	0.022	0.025	0.030
N	11,469	11,469	11,661	4,549	4,681	4,681	4,681	4,681
Panel C. Alternative linked sample based on 90% precision linkages								
Rosenwald exposure	0.288**	0.042***	0.017	-0.015	-0.002	0.004	-0.019	0.005
	(0.124)	(0.016)	(0.017)	(0.021)	(0.008)	(0.024)	(0.015)	(0.012)
R^2	0.094	0.057	0.022	0.143	0.039	0.025	0.030	0.025
N	23,036	23,036	23,419	9,168	9,457	9,457	9,457	9,457
Panel D. Inverse probability weights								
Rosenwald exposure	0.406***	0.080***	0.005	-0.021	-0.006	0.019	-0.018	-0.005
	(0.136)	(0.020)	(0.020)	(0.025)	(0.005)	(0.026)	(0.016)	(0.017)
R^2	0.117	0.077	0.027	0.188	0.041	0.032	0.032	0.042
N	16,669	16,669	16,942	6,602	6,805	6,805	6,805	6,805
Panel E. Unweighted								
Rosenwald exposure	0.398***	0.078***	0.009	-0.023	-0.004	0.020	-0.016	-0.006
	(0.135)	(0.020)	(0.020)	(0.024)	(0.006)	(0.025)	(0.016)	(0.017)
R^2	0.111	0.073	0.025	0.153	0.039	0.033	0.032	0.045
N	16,669	16,669	16,942	6,602	6,805	6,805	6,805	6,805

Notes: All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights in Panels A–C and inverse probability weights in Panel D (see text for details). Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.
Sources: See text for details.

Table A14: Decomposing the Impact of Rosenwald Schools on the Probability of Being a Teacher in 1940 by Migration Status, Black Women (1910–1914)

	Dep. var.: Joint outcome of being a teacher in 1940 and residing in:	
	Same county as 1920, rural area	Same county as 1920, urban area or different county than 1920
	(1)	(2)
Rosenwald exposure	0.024**	0.018
	(0.012)	(0.018)
R^2	0.044	0.040
N	3,250	3,250

Notes: Sample restricted to Black women in the labor force in 1940. All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

Sources: See text for details.

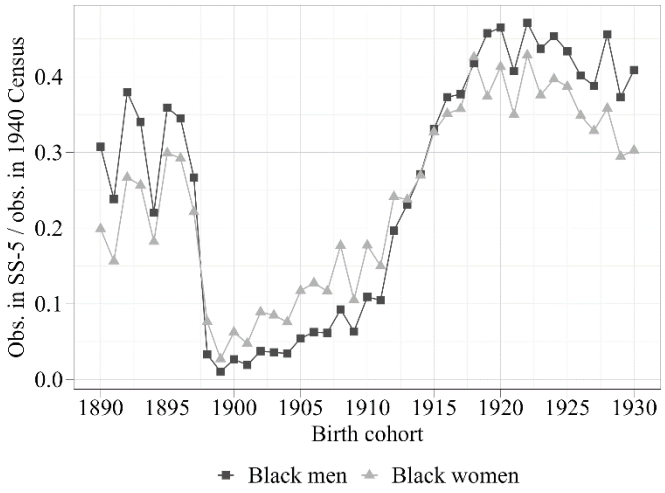
Table A15: The Impact of Rosenwald Schools on Wage Income and Imputed Income in 1940

	Dep. variable: Log annual income in 1940		
	(1)	(2)	(3)
Panel A. Black men (1910–1914)			
Rosenwald exposure	0.120	0.093	0.077
	(0.094)	(0.067)	(0.061)
R^2	0.072	0.071	0.262
N	4,486	6,377	6,377
Panel B. Black men (1915–1919)			
Rosenwald exposure	0.050	0.012	0.021
	(0.047)	(0.028)	(0.025)
R^2	0.067	0.053	0.158
N	9,337	13,925	13,925
Panel C. Black women (1910–1914)			
Rosenwald exposure	0.203	0.061	0.087
	(0.134)	(0.110)	(0.097)
R^2	0.123	0.109	0.308
N	2,571	3,011	3,011
Panel D. Black women (1915–1919)			
Rosenwald exposure	-0.055	-0.064	-0.012
	(0.070)	(0.055)	(0.052)
R^2	0.076	0.059	0.211
N	4,957	6,101	6,101
Impute income for farmers/self-employed		x	x
Occupation fixed effects (3-digit)			x
Sample	Wage workers	All workers	All workers

Notes: Income for farmers and self-employed workers in columns 2–3 imputed using the occupational income scores from Collins and Wanamaker (2022). All regressions include cohort-by-state of birth fixed effects, cohort-by-1920 state of residence fixed effects, 1920 household controls (father occupational income score, indicators for homeownership, father literacy, and mother literacy), and 1910 county controls (share of rural Blacks, rural Black/white literacy rate, rural Black/white school attendance rate). Observations weighted using inverse propensity score weights. Robust standard errors in parentheses, clustered at the county level. *** 1%, ** 5%, * 10% significance.

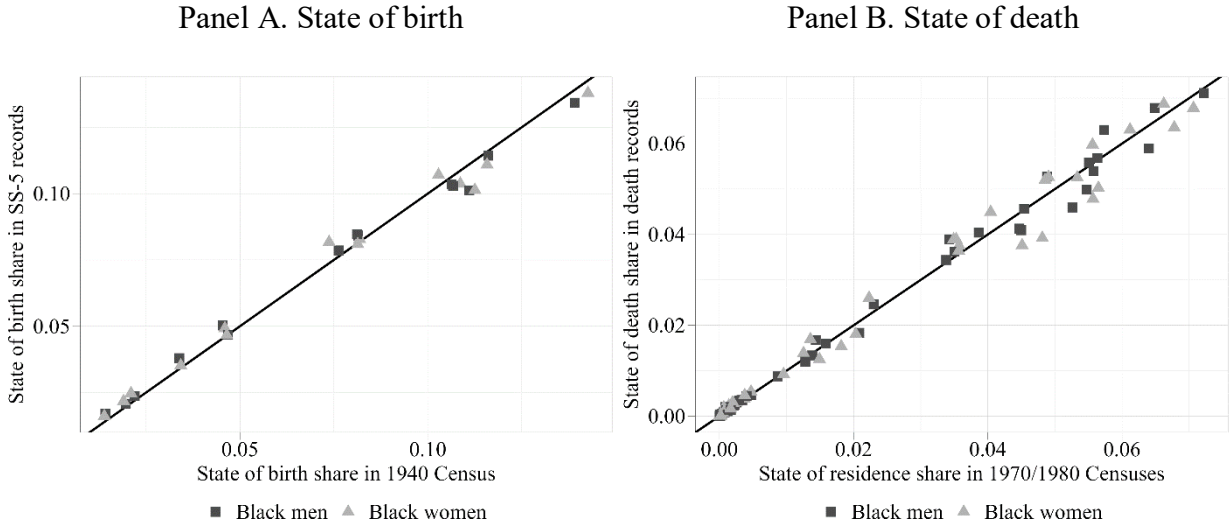
Sources: See text for details.

Figure A1: Cohort Coverage in SS-5 Records, Southern Blacks (1890–1930)



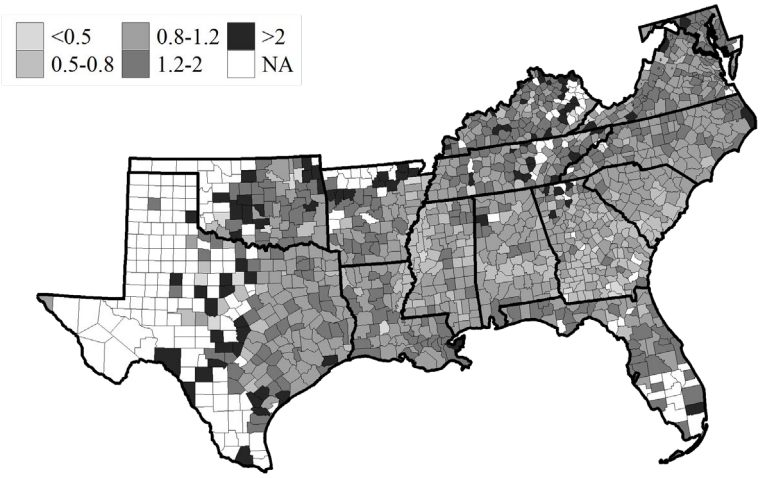
Notes: This figure plots the coverage rate of SS-5 records for Blacks born in the South, defined as the number of individuals in SS-5 records divided by the corresponding number of individuals in the 1940 Census, separately by cohort (*x*-axis) and gender (legend).
Sources: Numident SS-5 records, 1940 Census.

Figure A2: State of Birth and State of Death Coverage in SS-5 Records, Southern Blacks (1910–1919)



Notes: Panel A plots state of birth shares in SS-5 records against the corresponding shares in the 1940 Census among Blacks born in the South between 1910 and 1919, separately gender (legend). For the same cohorts, Panel B plots state of death shares in SS-5 records (among the subset of SSNs that can be found in Numident death records) against state of residence shares in the 1970 and 1980 Censuses, separately gender (legend).
Sources: Numident SS-5 records, Numident death records, 1940, 1970, and 1980 Censuses.

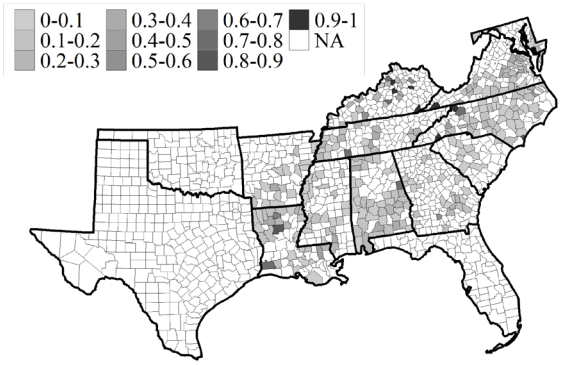
Figure A3: Ratio of 1920 County of Residence Shares in Linked Sample vs. 1920 Census



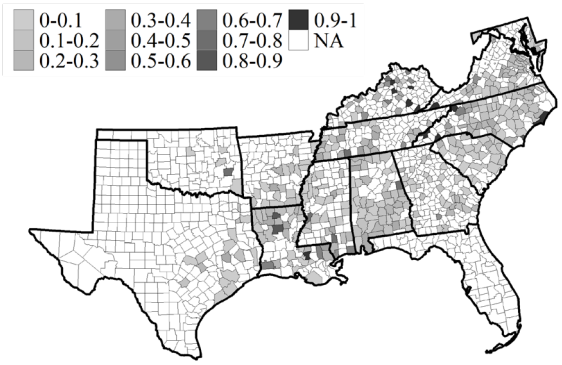
Notes: This figure plots the ratio of 1920 county of residence shares in the linked sample and 1920 county of residence shares in the population for Blacks born in the South between 1910 and 1919.
Sources: See text for details.

Figure A4: Rosenwald Exposure by County

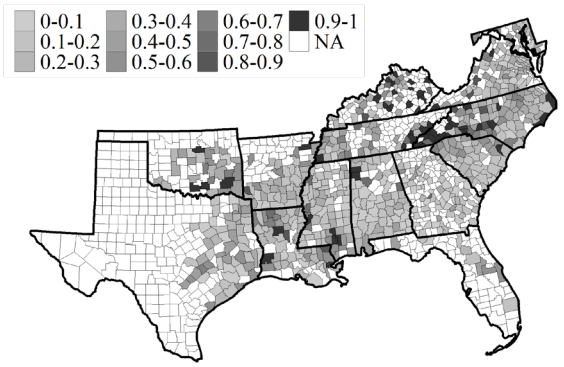
Panel A. 1919



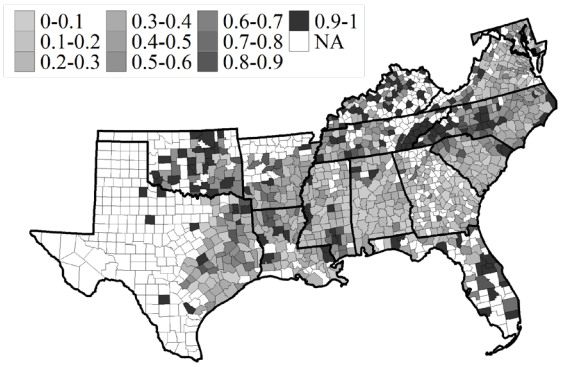
Panel B. 1920



Panel C. 1925

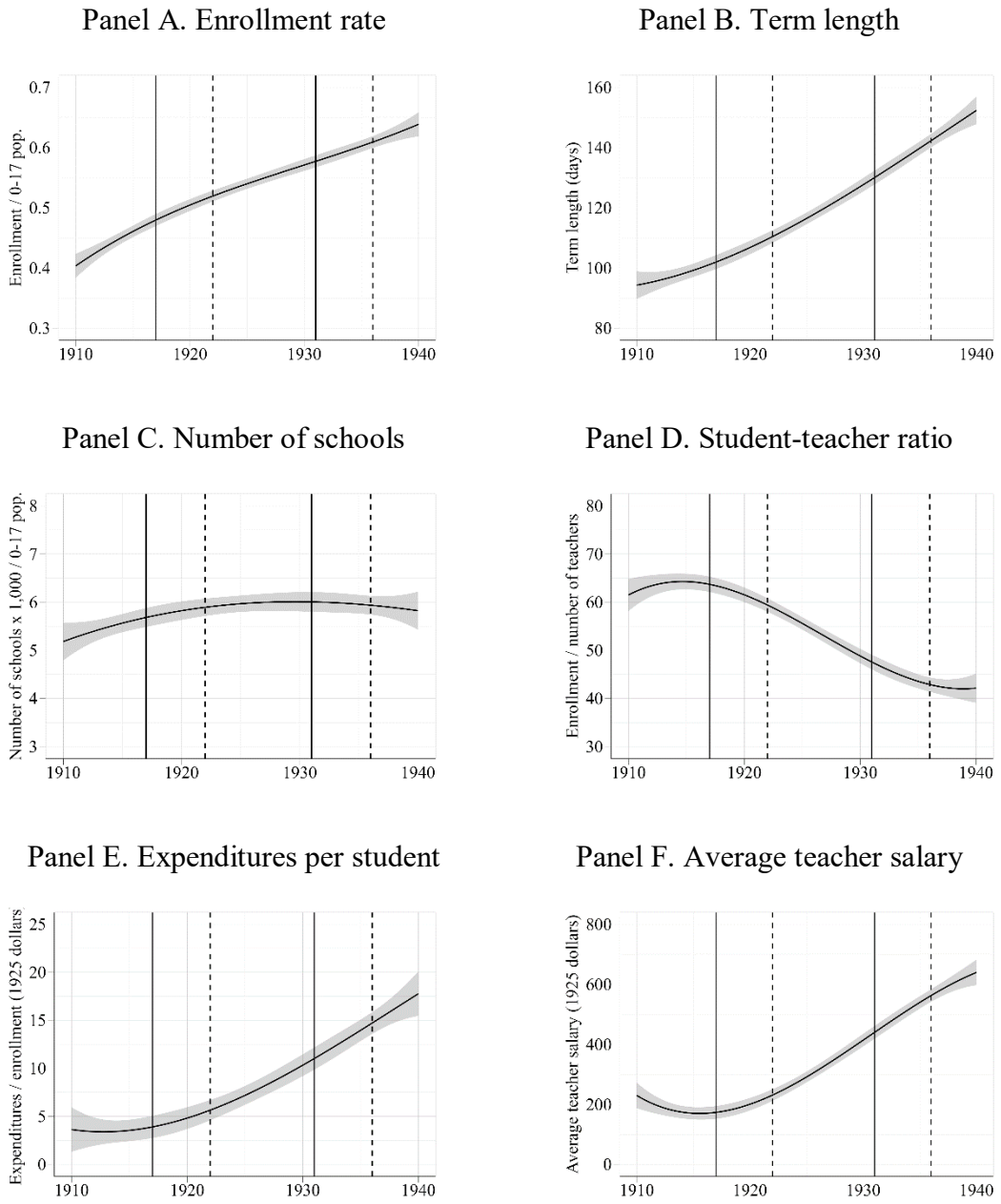


Panel D. 1931



Notes: This figure plots our Rosenwald exposure measure by county, separately by year.
Sources: Aaronson et al. (2014).

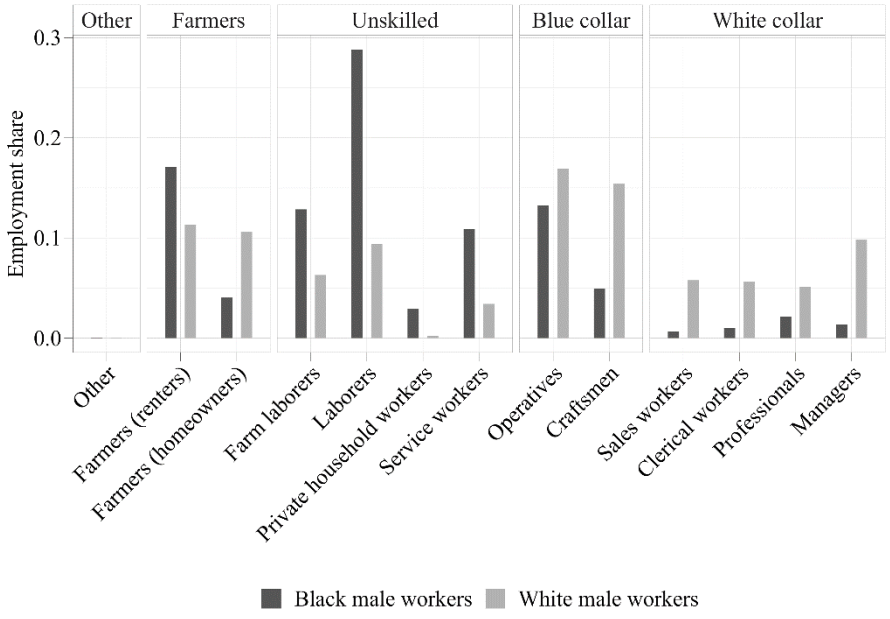
Figure A5: Educational Statistics for Blacks in 10 Southern States, 1910–1940



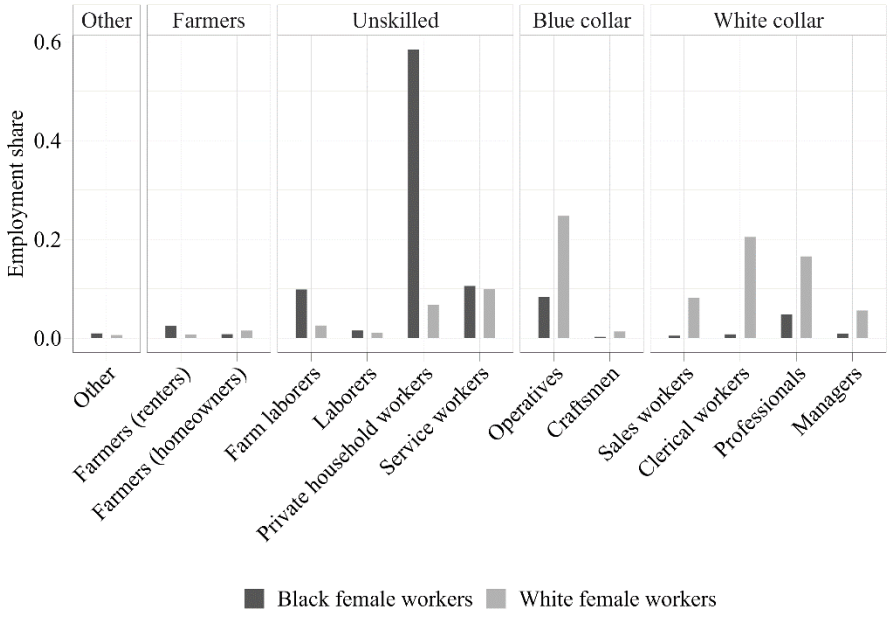
Notes: This figure plots educational statistics for Blacks between 1910 and 1940 in 10 Southern states (AL, AR, GA, KY, LA, MS, NC, SC, TN, TX), smoothed using a cubic function. County-level educational statistics for Black schools by year come from Carruthers and Wanamaker (2019). County-level population counts of Black children aged 0–17 come from the full-count 1910–1940 Censuses (population counts are linearly interpolated across years within counties). County-level statistics are averaged across counties within each year, weighting counties by the 0–17 Black population. The solid (dashed) vertical lines indicate the range of years in which the 1910–1914 (1915–1919) cohorts are aged 7–17 (1917–1931 vs. 1922–1936). Extreme values for 1929 in Panel C and for 1912 and 1914 in Panel E are excluded before smoothing (likely outliers).
Sources: Carruthers and Wanamaker (2019), 1910–1940 Censuses.

Figure A6: Employment Shares by Occupation Group in 1940

Panel A. Black vs. white male workers



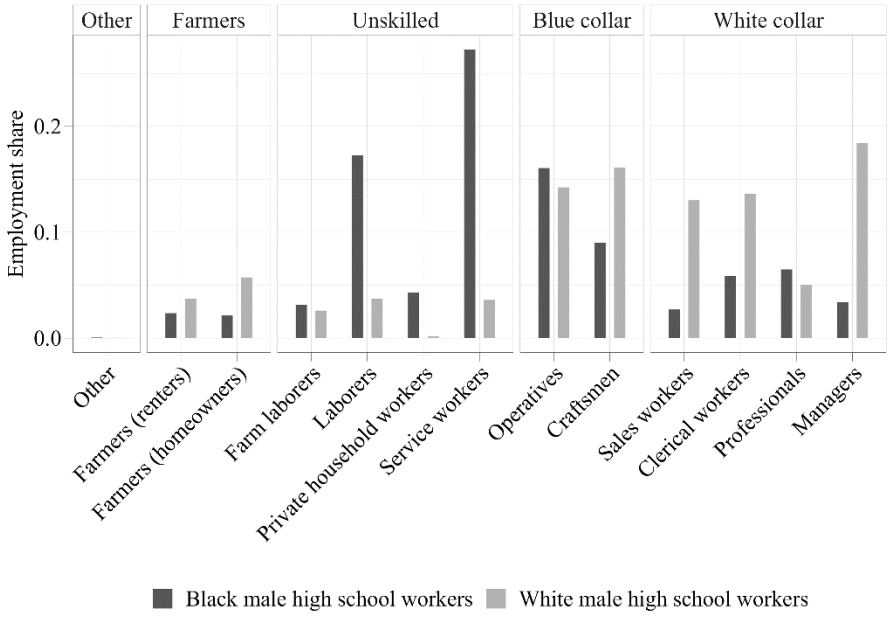
Panel B. Black vs. white female workers



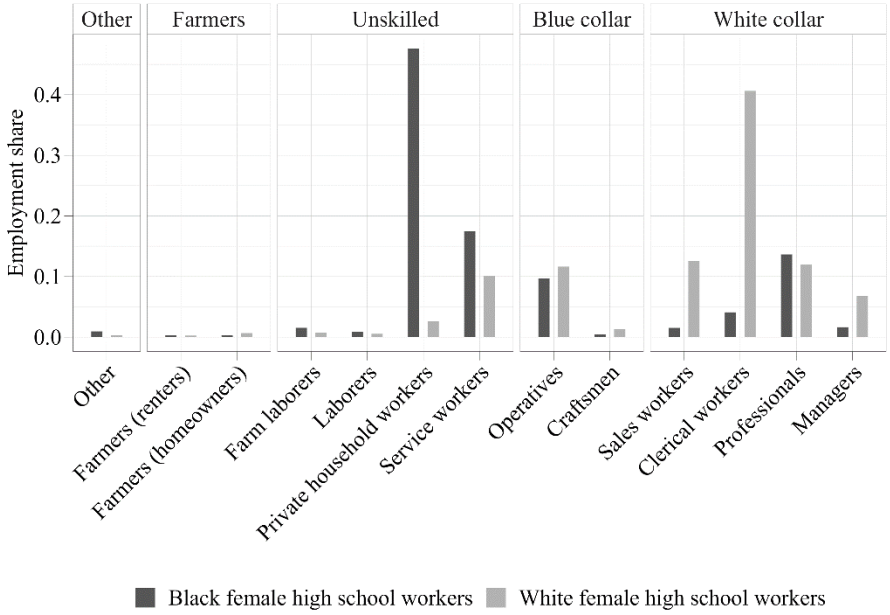
Notes: These figures plot employment shares in different occupation groups (*x*-axis) among Southern-born workers aged 25–54 in 1940, separately by race (legend).
Sources: See text for details.

Figure A7: Employment Shares by Occupation Group in 1940, Workers with 12 Years of Education

Panel A. Black vs. white male workers

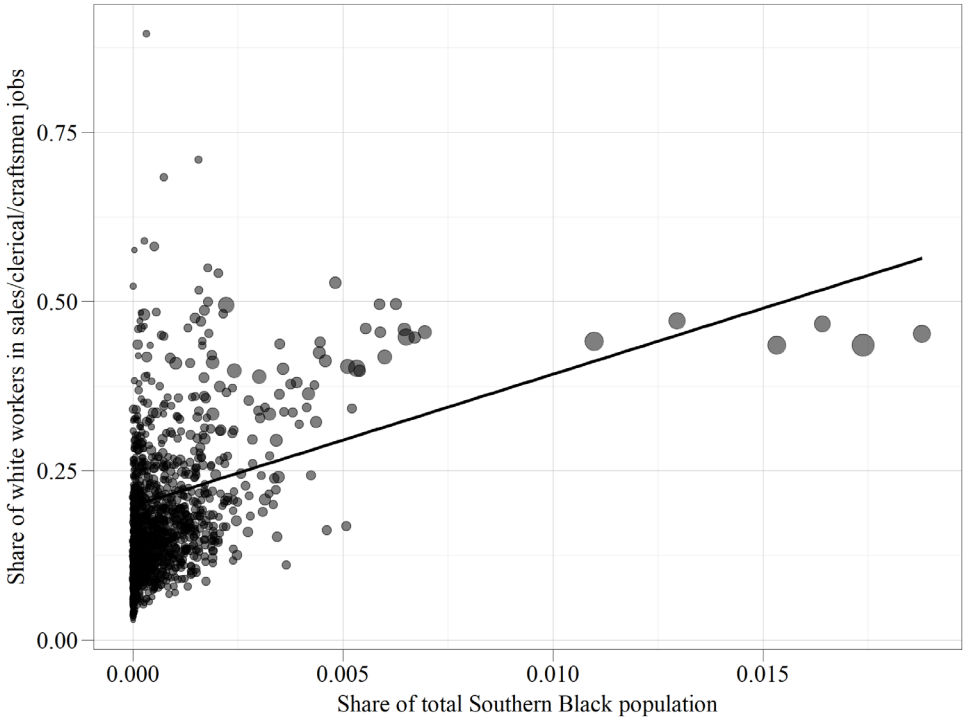


Panel B. Black vs. white female workers



Notes: These figures plot employment shares in different occupation groups (*x*-axis) among Southern-born workers aged 25–54 in 1940 with exactly 12 years of education, separately by race (legend).
Sources: See text for details.

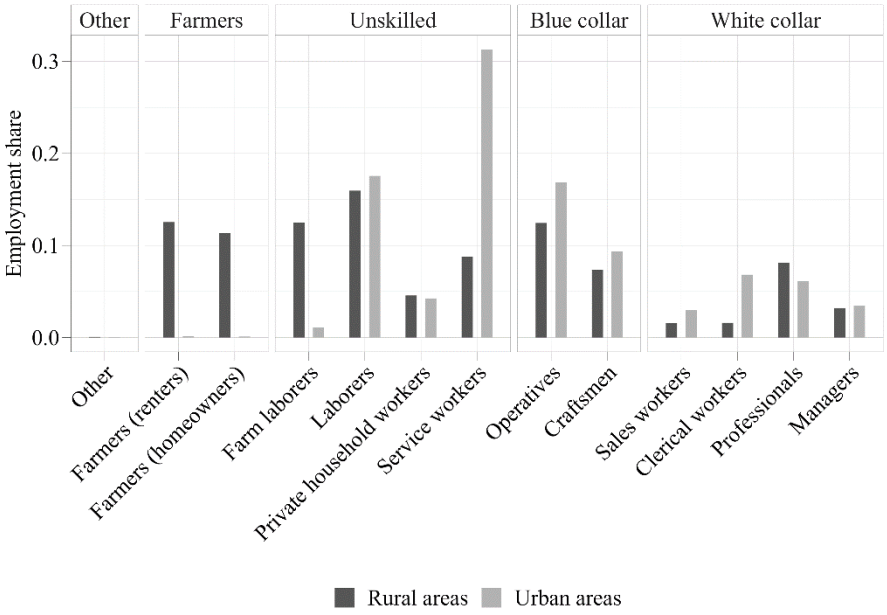
Figure A8: Share of White Workers in Sales/Clerical/Craftsman Jobs vs. Share of Total Southern Black Population in 1940, Southern Counties



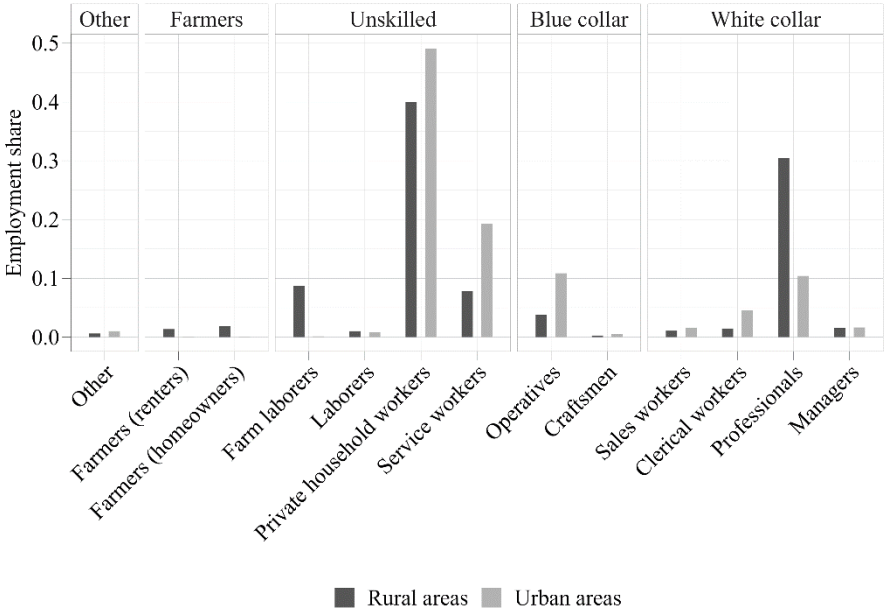
Notes: This figure plots the county-specific share of white workers in sales, clerical, and craftsman jobs against the county's share of the total Southern Black population in 1940. The sample is restricted to 14 Southern states (AL, AK, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA). The line represents the predicted values from a linear regression, where counties are weighted by total population (indicated by the size of the points).
Sources: See text for details.

Figure A9: Employment Shares by Occupation Group and Urban Status in 1940, Black Workers with 12 Years of Education

Panel A. Black male workers



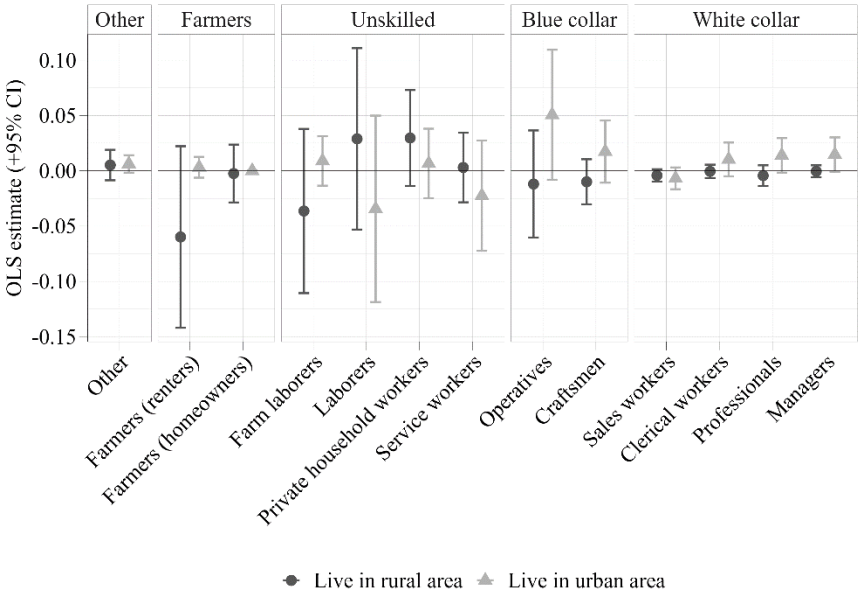
Panel B. Black female workers



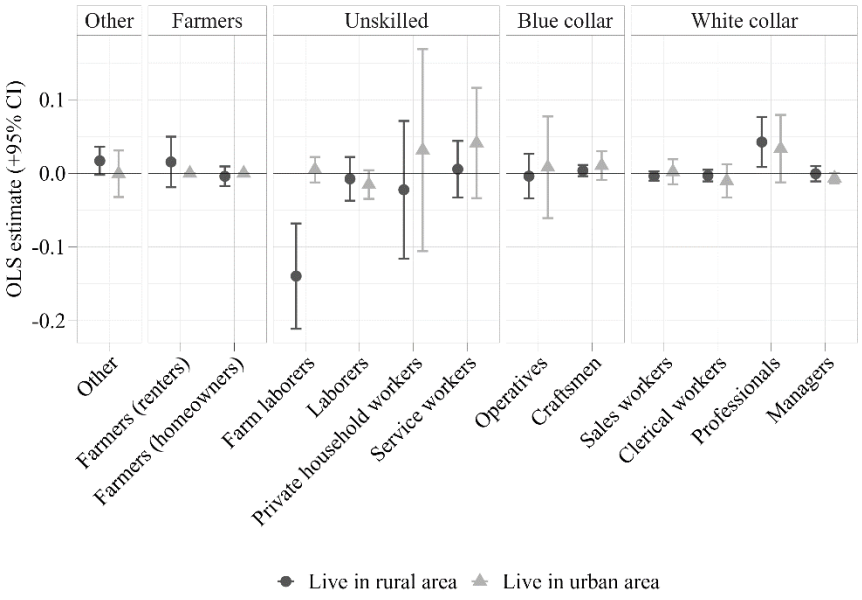
Notes: These figures plot employment shares in different occupation groups (*x*-axis) among Southern-born Black workers aged 25–54 in 1940 with exactly 12 years of education, separately by urban status (legend).
Sources: See text for details.

Figure A10: The Joint Impact of Rosenwald Schools on Occupation Groups and Urban Status in 1940

Panel A. Black men (1910–1914)



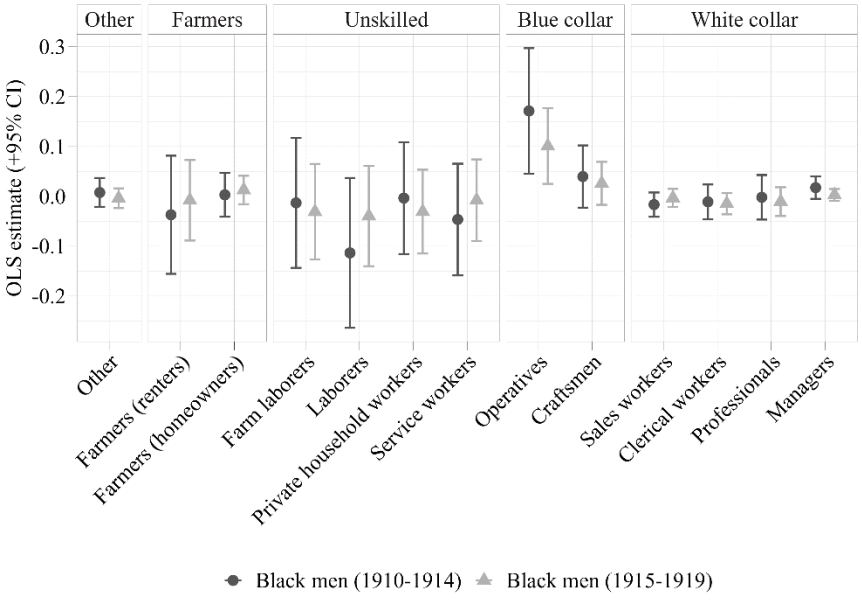
Panel B. Black women (1910–1914)



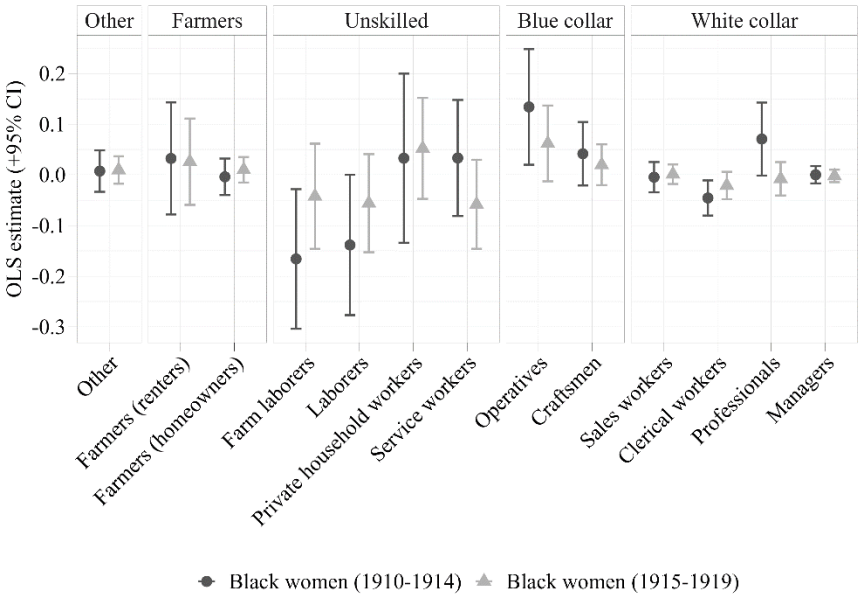
Notes: Each point corresponds to the OLS estimate from equation (2), where the dependent variable is an indicator for the joint outcome of being in a particular occupation group (*x*-axis) and living in a rural or urban area in 1940 (legend). The error bars represent the corresponding 95% confidence intervals.
Sources: See text for details.

Figure A11: The Impact of Rosenwald Schools on Occupation Groups in 1940, Alternative Specification with County Fixed Effects

Panel A. Black men



Panel B. Black women



Notes: Each point corresponds to the OLS estimate from equation (A.1), where the dependent variable is an indicator for being in a particular occupation group in 1940 (*x*-axis). The error bars represent the corresponding 95% confidence intervals.
Sources: See text for details.

Online Appendix B. Details on Linking Procedure

B.1 Linking SS-5 Records to the 1940 Census

The first step in constructing our linked sample involves linking SS-5 records to the 1940 Census. We start by describing how we link men. For each primary SS-5 record, we search for potential candidates in the 1940 Census that match on sex, first or last name initial, state/continent of birth, and a +/-3 year window around the primary record's age in 1940.[1] We then compute Jaro-Winkler name similarity scores between the primary record and *all* potential candidates that satisfy those conditions, and keep the top 25 candidates.

To generate the training data, we randomly draw 1,500 records and make manual linking decisions as follows. For each record and the corresponding 25 potential candidates from the previous step, we display full names (including middles names, if available), age in 1940, country of birth (if born outside the U.S.), and race.[2] The trainer either chooses a single best candidate from the set of candidates or declines to make a link. Failing to make a link occurs either when there is no promising candidate or when there are multiple plausible candidates. Following the LIFE-M training process (Bailey et al., 2023), each case is independently reviewed by two trainers. Cases on which they disagreed are sent to three additional trainers for independent review, mixed in with a random sample of cases on which they agreed. Only links chosen by both initial trainers or by

[1] Blocking on first or last name initials implies that potentials candidates only need to satisfy one of those two conditions.

[2] We display race and country of birth as optional pieces of information to take into account, knowing that they are not always accurately reported.

four out of five trainers are considered links for the purpose of training the model. We are able to manually link 43 percent of male SS-5 records to the 1940 Census.

To train the two-stage model from Bailey et al. (2023) to mimic trainers' linking decisions, we first split the training data into a "training set" and a "test set." The training set is used as input data to train the algorithm, and the test set is be used to assess the out-of-sample performance of the model (which typically lines up well with the in-sample performance). Probabilities are generated using 10-fold cross validation.[3] The male model is able to achieve a "recall" rate of 75 percent at a 95 percent "precision" rate. In other words, fixing the error rate of model-based links relative to human trainers at 5 percent, the model is able to reproduce 75 percent of the links made by human trainers.

The process of linking female SS-5 records is analogous, except for a few key differences. First, we generate two versions of each female record, one with the married name as the last name and one with the birth name as the last name (when both are available). Birth and married names are determined by comparing the last name field with the father last name field. Second, we additionally block on marital status when generating potential candidates. Women linked using their birth name are linked to single women in the census, and women linked using their married name are linked to non-single women in the census (in other words, married, divorced, widowed).

To generate the training data for women, we manually link a random sample of 500 SSNs, which correspond to around 1,500 primary SS-5 records. In this case, a primary record is a particular name variant of a female SSN. Trainers make decisions for each primary record in isolation (in other words, they are instructed *not* to consider whether they previously made a link for another primary record belonging to the same SSN). For individuals that have conflicting links

[3] The full set of model features used to train this model (and all other models) is available upon request.

(in other words, a link using their married name and another link using their birth name), we manually resolve cases by either choosing one link over the other or deciding there is too much ambiguity and discarding both links. Ultimately, we are able to manually link 52 percent of female SSNs to the 1940 Census. The female model is estimated at the primary record level to mirror the training process. Links are then resolved at the SSN level using the following rule: When there are two conflicting link within the same SSN, we simply discard both links. The SSN-level recall rate of the female model is 74 percent at 95 percent precision.

Using the trained male and female models and the corresponding cross-validated probability thresholds, we generate links for the full sample of SS-5 records. Note that while the model links each SSN to a unique census record, it is possible for the same census record to be linked to two different SSNs due to common names. To resolve those cases, we prioritize links that have a higher precision rate, and discard the remaining conflicts.

B.2 Linking Siblings in SS-5 Records

The second step in constructing our linked sample involves linking siblings together in the SS-5 data. The purpose of these linkages is to reconstruct families (children + parents) for the final linking step, in which we link families to households in the 1920 Census. We start with all unique SSNs in the SS-5 data. For each SSN, we identify potential siblings from the remaining SSNs by blocking on father first two last name initials and a +/-10 year window around the primary child's year of birth.[4] For all potential siblings within the block, we compute a Jaro-Winkler score between parent names fields, keep the top 21 potentials, and drop redundant pairs between identical records.

[4] This blocking scheme is necessary for computational reasons, as we do not block on place of birth (siblings can be born in different states/countries).

To generate the training data, we randomly draw 500 records and make manual linking decisions as follows. For each record and the corresponding 20 potential siblings from the previous step, we display the year of birth, state/country of birth, race, and parent full names.[5] Here, trainers can choose multiple links as individuals can have multiple siblings. Note that SS-5 records list the mothers' birth names, which greatly reduces ambiguity. In our training data, we are able to find at least one sibling for 40 percent of individuals.

Similar to SS-5-1940 Census linking, we split the training data into a training set and test set.[6] We then train a random forest model on the training set and use 10-fold cross-validation to generate match probabilities and determine the threshold associated with a 97 percent precision rate. The model is able to achieve a recall rate of 90 percent at 97 percent precision. Using the trained random forest model and the cross-validated probability threshold, we generate sibling links in the full sample of SS-5 records. We then use these sibling links to create a crosswalk between SSNs and family identifiers. The simplest way to think about this problem is that SS-5 records and sibling links are nodes and edges in a network. Family identifiers are then clusters of nodes formed by the edges between them. We use the concept of connected components to define family identifiers: Families are such that any two nodes within a family can be connected via an arbitrary chain of edges, but two nodes in different families cannot. In particular, this means that siblings that are born more than 10 years apart can still belong to the same family if they are both linked to a

[5] In practice, the main consideration is parent name similarity, but year of birth differences and whether the siblings are born in the same state or country is also informative.

[6] In contrast to one-to-one linking where each set constitutes a distinct linking decision, in one-to-many linking we consider each primary record-potential pair as a separate linking decision, even though these decisions were made simultaneously. As a result, the effective size of the training data is roughly $N = 10{,}000$ (500×20).

common sibling born within 10 years of both. In the end, we are able to reconstitute around 5.9 million non-singleton families in the SS-5 data (in other words, families with at least 2 children), such that 35 percent of SSNs are linked to at least one other SSN.

B.3 Linking SS-5 Records to the 1920 Census

The final step in constructing our linked sample involves linking reconstituted families in SS-5 records to households in the 1920 Census, when children are mostly still living with their parents as dependents. We generate candidate households for each family by blocking on father last name initials and the union of children's state/continent of birth. Blocking on the union of children's place of birth implies that the only requirement we impose is that at least one of the potential children's place of birth matches one of the primary children's place of birth. Within each block, we select the top 20 candidate households based on a Jaro-Winkler score between parent names.

We generate training data for a random sample of 500 primary families. We display the following information to trainers: father full name, mother first and middle name, and children's first and middle name, age in 1920, and state/country of birth. We also display the modal race among children. We only display primary siblings that would have been alive in 1920, and mask potential children in the census if there is little chance that they match any of the primary children (based on their age and place of birth). We are able to manually link 48 percent of families to a household in the 1920 Census in the training data. We train the two-stage model from Bailey et al. (2023) using the same steps as in SS-5-1940 linking. The model is able to reproduce 88 percent of links made by trainers at a 95 percent precision rate.

In contrast to SS-5-1940 linking, it is not uncommon for the same census household to be linked to two (or more) different families in the SS-5 data after scaling up the model. The reason is that we were not able to make all possible sibling linkages in the SS-5 data at 97 percent

precision, mainly because parent names are not always consistently reported across siblings' SS-5 records. Although we might not be able to link two SS-5 records solely based on parent name information, it is possible for both of them to be linked to the same census household since the latter additionally exploit information on children. As a result, we end up with apparent conflicts in the data that are not true conflicts but simply reflect the incompleteness of our family identifiers. Rather than throwing away all conflicts, we manually examined a random sample of 200 of them and either chose one link over the others, concluded all links were correct, or chose none if there was too much ambiguity. We then used this data to assess the performance of different conflict resolution rules in terms of their ability to keep all correct links and drop all incorrect links.[7] The rule we ultimately settled on is able to retain 80 percent of correct links and drop 75 percent of incorrect links based on our training data.

Online Appendix C. Reconciling Education Results for Black Men with Aaronson and Mazumder (2011)

In this section, we show that one can reconcile the magnitude of our educational attainment estimates for Black men with existing estimates in Aaronson and Mazumder (2011). Focusing on World War II enlistees born between 1910 and 1928, they report a triple difference estimate of 1.377, and a "simple difference" estimate of 1.193 that does not exploit urban Blacks/whites as control groups, which is more directly comparable to our empirical strategy. When we pool Black

[7] The two extreme rules are to keep all conflicts or to drop all conflicts, which exactly achieve one objective but not the other. The optimal rule lies somewhere in between.

men born in 1910–1919 to maximize comparability, we obtain an estimate of 0.438.[8] What accounts for the difference in magnitudes? Note that, following their analysis of school attendance and literacy based on pooled census data, we include controls for household characteristics in childhood (in other words, father occupational income score, parents' literacy status, homeownership status) and assign Rosenwald exposure based on observed rural status in childhood. However, because World War II enlistee records lack information on household characteristics and rural status in childhood, their analysis of educational attainment omits these household controls and assigns rural status by categorizing enlistees' county at enlistment as entirely rural, predominantly urban, or neither (based on the share of rural residents by county across the 1910 to 1930 Censuses). In particular, counties that are neither entirely rural nor predominantly urban—roughly one out of two counties in the South, in which around 60 percent of all Rosenwald schools were built—are excluded from their analysis.

Online Appendix Table A4 demonstrates that imposing these two restrictions helps reconcile most of the gap between our estimates. The second column in Panel A shows that dropping 1920 household characteristics from the set of controls increases our estimate from 0.438 to 0.505. The third column in Panel A shows that further restricting the sample to rural-only counties yields an estimate of 0.946, which lies within the 95 percent confidence interval of the simple difference estimate cited above.[9] The estimates in Panel B reveal similar patterns for Black women. Although

[8] We opted not to include later cohorts as many of those individuals will not have completed their schooling by 1940.

[9] Aaronson and Mazumder (2011) note that the impact of Rosenwald schools on Black men's educational attainment is around half as large in the subset of excluded counties. This could reflect the fact that these counties contain a mix of treated and untreated individuals (which is why they exclude them to begin with), but it could also

there are no estimates in the literature to compare them to, our findings for Black women are in line the positive effects on school attendance and literacy documented in Aaronson and Mazumder (2011), which cover both Black boys and girls.

Online Appendix D. Robustness Checks

D.1 Controlling for County-Specific Unobservable Factors

As discussed in the empirical strategy section of the paper, the primary threat to identification is the possibility that Rosenwald schools were established in counties with certain characteristics that independently affected later-life outcomes of Black children who grew up in those counties. Our baseline specification mitigates this concern by controlling for key county characteristics that have been found to be correlated with the placement of Rosenwald schools (for example, Black literacy rates, the population share of rural Blacks).

Here, we take a more direct approach and present results from an alternative specification that includes county fixed effects, which will capture any time-invariant county characteristic that may be correlated with the placement of Rosenwald schools. Since in this case identification relies on within-county cross-cohort variation in Rosenwald exposure, we pool all cohorts of Black men and women in our sample and include sex-by-cohort and sex-by-state of birth fixed effects. For comparability with the baseline results, we still allow the impact of Rosenwald exposure to vary

reflect genuine heterogeneity in the effect of Rosenwald schools across counties. Since we directly observe individuals' rural status in 1920, we can include all counties in our analysis without compromising on the accuracy of our Rosenwald exposure measure.

across sex-cohort groups by interacting our exposure measure with relevant indicator variables. Formally, the alternative specification is given by:

$$y_{ibc} = \alpha_c + \delta_{bs(c)} + \sum_g \beta_g \cdot \text{ROSE}_{bc} + \Gamma \cdot \text{X}_i + \varepsilon_{ibc} \quad \text{(A.1)}$$

where g denote demographic groups (men or women born in 1910–1914 or 1915–1919), α_c are 1920 county of residence fixed effects, $\delta_{bs(c)}$ are cohort-by-1920 state of residence fixed effects, and X_i now includes sex-by-cohort and sex-by-state of birth fixed effects (in addition to cohort-by-state of birth fixed effects and 1920 household controls).

The resulting estimates are shown in Online Appendix Table A6. The estimates for years of education and the probability of having 9+ years of education are smaller in magnitude than the baseline estimates (except for the coefficient on years of education for Black women born in 1915–1919 which is negative but close to zero), but not statistically significant. As in our baseline results, exposure to Rosenwald schools has a positive effect on Black women's labor force participation, though not just for older cohorts but also younger cohorts of women. The occupational income score estimates are very similar to the baseline estimates (0.055 and 0.151 for Black men and women born in 1910–1914, relative to baseline estimates of 0.049 and 0.134 respectively).

The occupational reallocation patterns differ somewhat from the baseline results, with some evidence that Black men and women exposed to Rosenwald schools were less likely to hold unskilled jobs and more likely to hold blue-collar jobs in 1940. Nevertheless, consistent with our baseline results, Online Appendix Figure A11 shows that Black men and women made no significant gains in craftsman, sales, and clerical positions. Furthermore, over half of the gains in blue-collar jobs are concentrated in two of the most common blue-collar jobs among Southern Black workers, "Operative and kindred workers (n.e.c.)" and "Truck and tractor drivers." As in

the baseline results, there is a positive and significant effect on the probability of being a storekeeper in 1940 among Black men born in 1910–1914 (estimate of 0.023, relative to a baseline estimate of 0.015). The point estimate for the probability of being a teacher in 1940 among Black women born in 1910–1914 is positive and similar in magnitude to the baseline estimate (0.038 vs. 0.043), but not statistically significant. All in all, taking into account statistical imprecision, we view these alternative results as painting a similar picture as our baseline results.

D.2 Alternative Linked Datasets

Although SS-5 records enable us to link women across censuses, one potential concern is that individuals who appear in SS-5 records are a selected subset of the population. Here, we show that our main findings are robust to exploiting alternative linked datasets. We consider two datasets directly linking individuals across the 1920 and 1940 Censuses: (1) the IPUMS Multigenerational Longitudinal Panel (Helgertz et al., 2020), and (2) the Census Linking Project (Abramitzky et al., 2020), which we respectively refer to as MLP and CLP for simplicity. Like our method, MLP uses a supervised ML approach to generate links. There are two steps in the linking process. The first step links men across adjacent censuses.[10] The second step links members living in the same household as men linked in the first step. Among others, this makes it possible to link women, essentially women living with their husbands or fathers. CLP links men across pairs of censuses based on name and demographic information using the "ABE" automated linking algorithm first

[10] In addition to name and demographic information, MLP exploits rich contextual information as part of the linking process (for example, location, neighbors, household composition). The upside is that it reduces the ambiguity in the linking decisions, producing higher match rates without compromising on accuracy. The downside is that the linked sample will naturally be skewed towards men who do not move between censuses and have stable household structures. The consequences of this latent selection depend on the question of interest.

proposed in Abramitzky et al. (2012). CLP does not attempt to link women. We present results using the "standard" variant of their algorithm, which matches on exact names and allows for age differences of up to two years, as well as the "conservative" variant of their algorithm, which additionally requires names to be unique within a 5-year age band. The main advantage of MLP and CLP is that they link the entire population in the 1920 Census, which eliminates concerns related to the selection of individuals into SS-5 records and naturally results in larger sample sizes (see Panels B–D in Online Appendix Table A1 for match rates and sample sizes). For comparability, we re-weight these linked samples using the same procedure as in our baseline linked sample.

The top three panels in Online Appendix Table A7 show results for Black men born in 1910–1914 based on MLP and CLP. Although the point estimates are smaller in magnitude than our baseline estimates, the patterns are similar: Exposure to Rosenwald schools has a positive and significant effect on educational attainment, and no significant effect on labor force participation or occupational standing. The MLP-based estimates for Black women born in 1910–1914 in Panel D are also somewhat smaller in magnitude than the baseline estimates, though imprecisely estimated.[11] Lastly, while one might worry that the more muted effects among younger cohorts of Black men and women is just an artifact of differential cohort coverage in SS-5 records (see Online Appendix Figure A1), Online Appendix Table A8 shows that the same pattern arises in MLP and CLP. Overall, the fact that the findings are broadly similar across different linked datasets is reassuring.

[11] For women we strongly prefer our SS-5 based estimates as MLP linkages are heavily skewed towards unmarried women. The unweighted share of Black women that are married in 1940 is 56 percent in our linked sample, while the corresponding share in MLP is only 21 percent (relative to a true population mean of 68 percent).

D.3 Results for Non-Migrants Based on 1940 Census Cross-Section

The advantage of linking individuals across the 1920 and 1940 Censuses is that it allows us to accurately determine their childhood exposure to Rosenwald schools. However, one drawback is that individuals with lower socio-economic standing tend to be underrepresented in name-based linked samples since they have more common names for example. Given that the gains from exposure to Rosenwald schools may have been larger for this subpopulation, our results may be underestimating the true gains in the population, despite our re-weighting procedure. Here, we prioritize sample representativeness over accuracy of treatment assignment by focusing on Black men and women in the 1940 Census who likely still live where they grew up, and for which we can therefore approximate exposure to Rosenwald schools based on their current location. More specifically, we focus on individuals living in a rural area in their state of birth who also report living in the same house as in 1935.[12]

Online Appendix Table A9 presents the results using this alternative approach.[13] The patterns are remarkably similar to our baseline estimates. Rosenwald exposure has a positive impact on Black men and women's educational attainment, and the gains are larger among older cohorts than younger cohorts. We also find a positive labor force participation effect among older cohorts of

[12] Eriksson (2020) conducts a similar exercise, though one minor difference is that she assigns exposure based on 1935 county of residence and 1935 rural status (the latter is approximated using the response to the 1935 city of residence question). We instead chose to focus on "non-migrants" for which we can observe actual rural status in 1940.

[13] Note that this specification naturally excludes 1920 household controls. The cohort-by-1920 state of residence fixed effects should mostly coincide with the cohort-by-state of birth fixed effects since we focus on individuals living in their state of birth.

Black women, but not younger cohorts. Importantly, there are no detectable gains in occupational standing, except among older cohorts of Black women. For this subgroup, there are positive (and significant) gains in white-collar employment, though one minor difference is that those gains are offset by losses in both unskilled and blue-collar jobs (rather than primarily unskilled jobs). Moreover, one can show that the gains in white-collar are concentrated in teaching jobs (estimate of 0.041, significant at the 5 percent level), as in our baseline results. Overall, these findings suggest that our results are not driven by sample selection due to linking.

D.4 Additional Robustness Checks

Online Appendix Tables A10 and A11 show that our main findings for Black men and women born in 1910–1914 are not sensitive to several alternative empirical choices.[14] Our default measure of Rosenwald exposure is based on the number of Rosenwald teachers normalized by the rural Black school-age population, averaged over the years an individual was aged 7–13. Panel A instead defines Rosenwald exposure based on the years an individual was aged 7–17 to account for the possibility that some Rosenwald schools may have provided secondary education. The resulting estimates are fairly similar to the baseline estimates.

In our baseline specification, we assigned exposure to Rosenwald schools based on individuals' county of residence in 1920. The advantage is that location in 1920 likely provides a good approximation of actual exposure to Rosenwald schools. However, one worry is that some parents may have endogenously chosen to move to Rosenwald counties in the hope of sending their children to these schools, which may bias the estimates. To the extent that the later-life outcomes of these children would be have been better anyway (for example, due to greater family

[14] Analogous results for Blacks born in 1915–1919 are shown in Online Appendix Tables A12 and A13.

resources), we might be overstating the impact of Rosenwald schools. To address this, Panel B restricts the sample to individuals who resided in their county of birth in 1920, which is the case for around 75 percent of individuals in our sample for which we can determine county of birth.[15] If anything, the point estimates for this subpopulation are slightly larger in magnitude than our baseline estimates, so that selective migration is unlikely to be an important confounding factor.

Panel C shows that point estimates tend to be slightly smaller in magnitude when we target a precision rate of 90 percent instead of 95 percent in the construction of our linked sample, consistent with the notion that classical measurement error induced by false links can lead to attenuation bias (Bailey et al., 2020). Lastly, Panels D and E show that our main findings are also not sensitive to alternative weighting schemes. In Panel D, instead of the inverse propensity score weights described in the data section of the paper, we use simple inverse probability weights that re-weight each sex-by-race-by-cohort-by-state of birth cell by the inverse of the probability of being in our linked sample in 1940. Note that these weights deliberately do not adjust for educational attainment and other endogenous outcomes in 1940. Panel E presents unweighted results. The estimates are similar regardless of the weighting strategy.

References

Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder. "Fertility Transitions Along the Extensive and Intensive Margins." *American Economic Review* 104, no. 11 (2014): 3701–3724.

[15] County of birth information is available for around 90 percent of individuals in SS-5 records via a 12-character place of birth string. We thank Evan Taylor for kindly sharing his crosswalk.

Abramitzky, Ran, Leah Boustan, and Myera Rashid. *Census Linking Project: Version 1.0 [dataset]* 2020.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, no. 5 (2012): 1832–1856.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform? Lessons From U.S. Historical Data." *Journal of Economic Literature* 58, no. 4 (2020): 997–1044.

Bailey, Martha J., Peter Z. Lin, A. R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman. "The creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database project." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 56, no. 3 (2023): 138–159.

Carruthers, Celeste, and Marianne Wanamaker. *County-Level School Enrollment and Resources in Ten Segregated Southern States, 1910–1940*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019.

Collins, William J., and Marianne H. Wanamaker. "African American Intergenerational Economic Mobility Since 1880." *American Economic Journal: Applied Economics* 14, no. 3 (2022), 84–117.

Eriksson, Katherine. "Education and Incarceration in the Jim Crow South: Evidence from Rosenwald Schools." *Journal of Human Resources* 55, no. 1 (2020): 43–75.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. *IPUMS Multigenerational Longitudinal Panel: Version 1.0 [dataset]*, Minneapolis, MN: IPUMS, 2020.