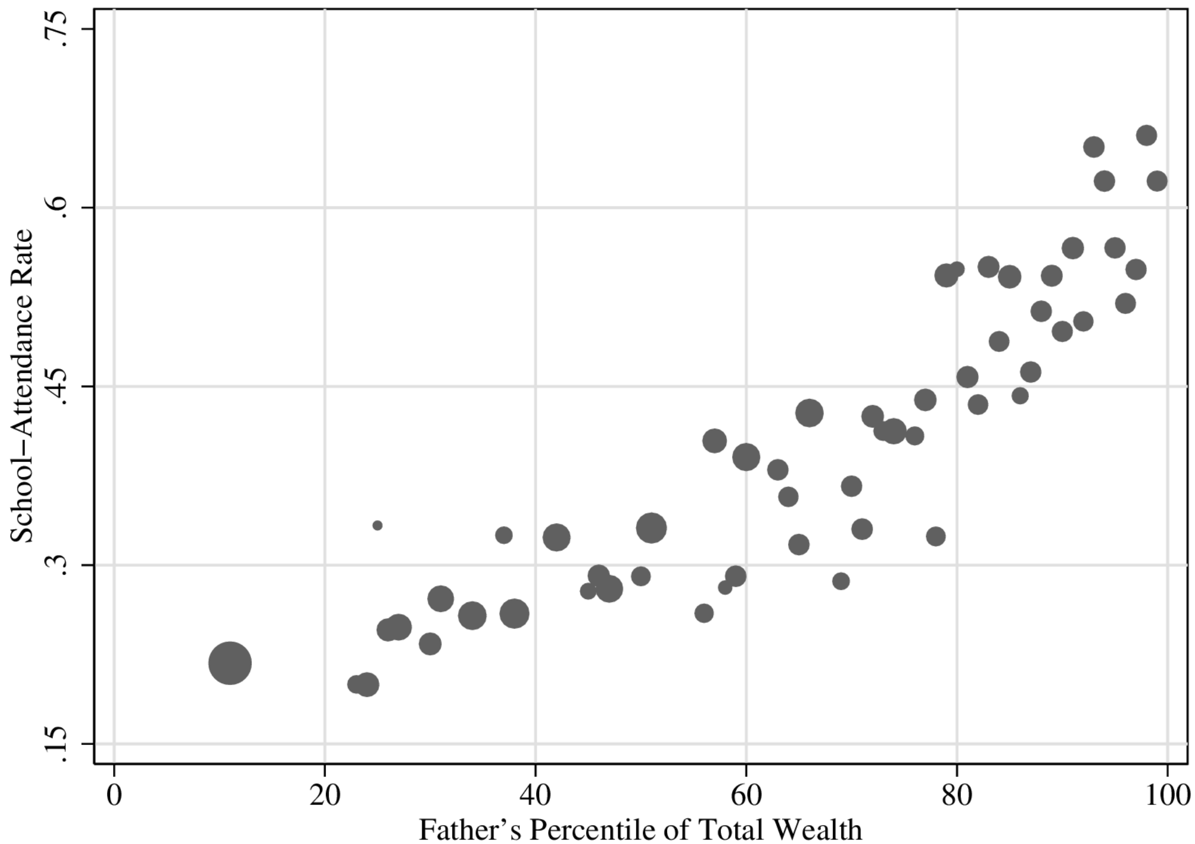
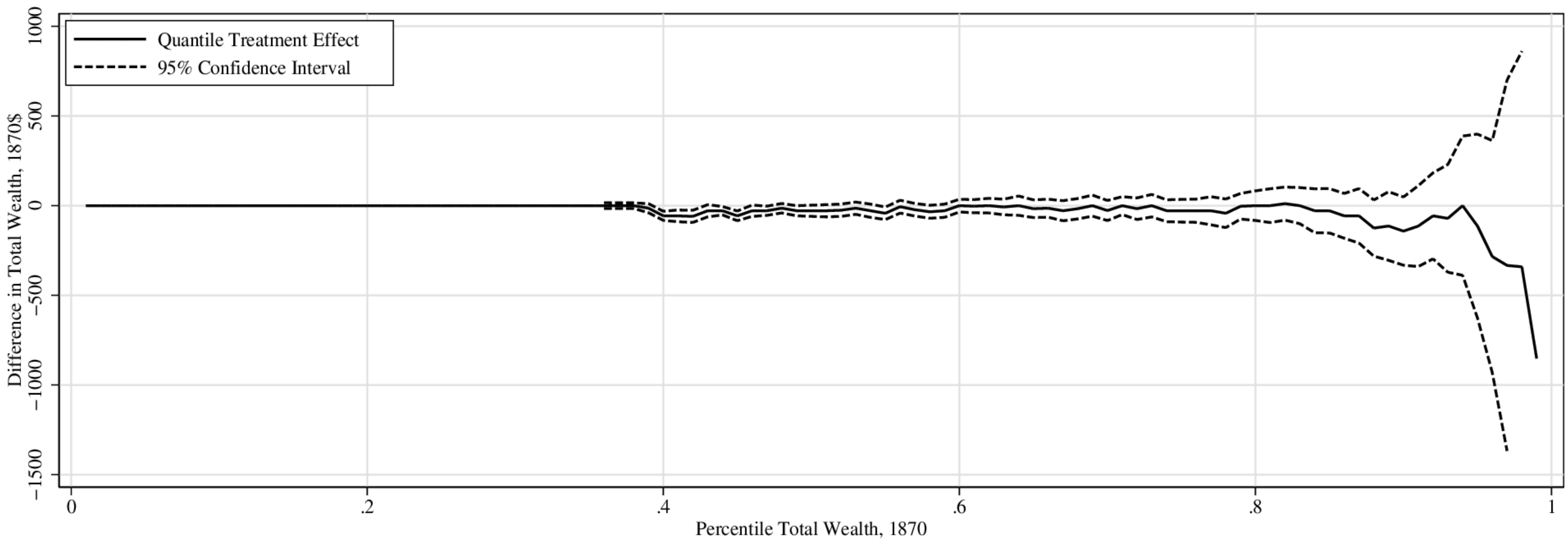
[Appendices for Online Publication]

Appendix Figure 1: The Gradient in School Attendance by Father's Percentile of Total Wealth, 1850



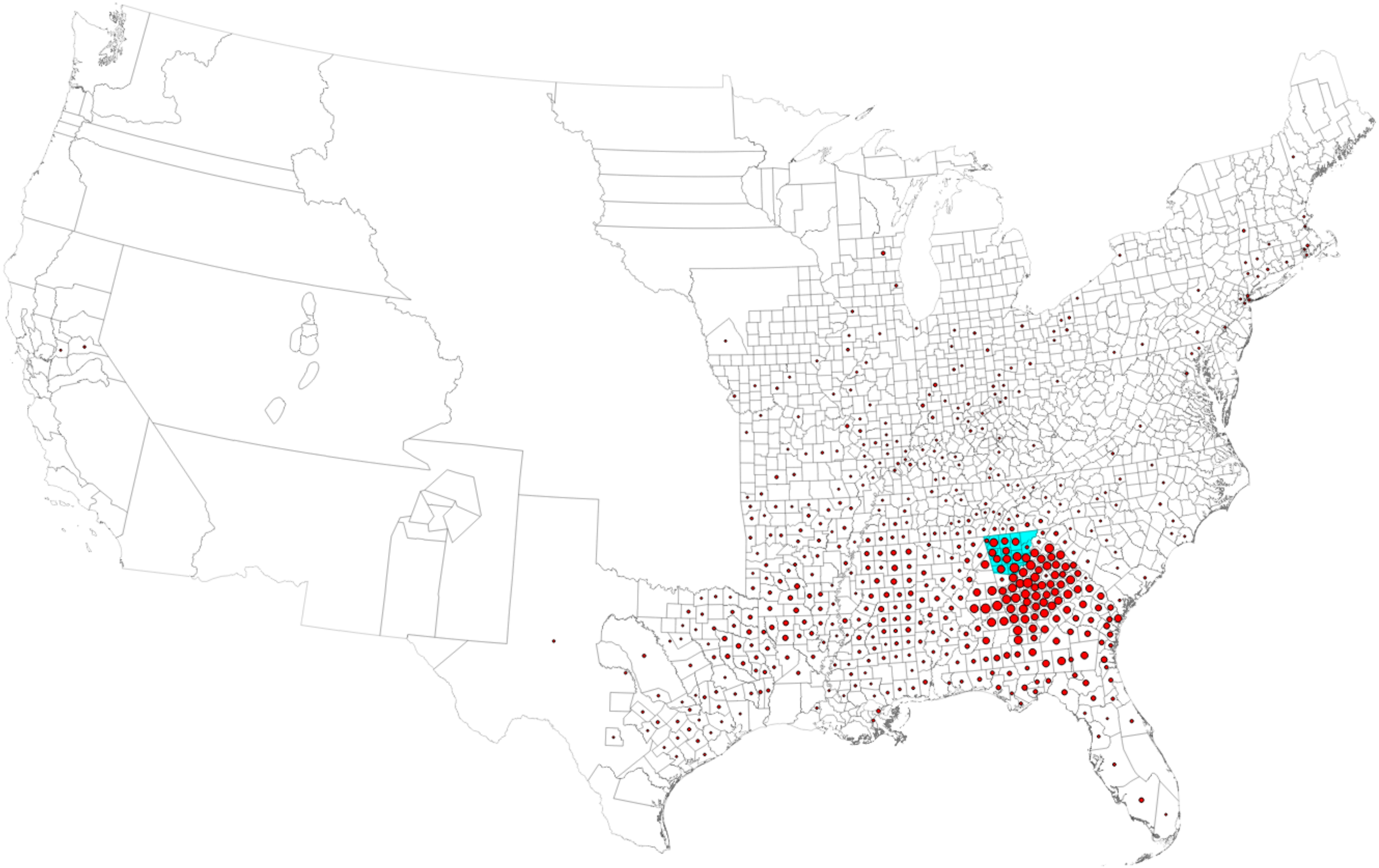
Notes: This figure displays the fraction of children attending school for each percentile of paternal total wealth (real estate plus slaves) in 1850. The data are the main sample of households with lottery-eligible men, as used in Table 2, for example. Approximately 22% of fathers have zero wealth, and such cases are coded to a percentile of 11. Lumpiness in the data yield uneven cell sizes. The dots are sized in proportion with the cell sizes.

Appendix Figure 2: Quantile Regression Estimates of Treatment Effects on Childrens' 1870 Wealth



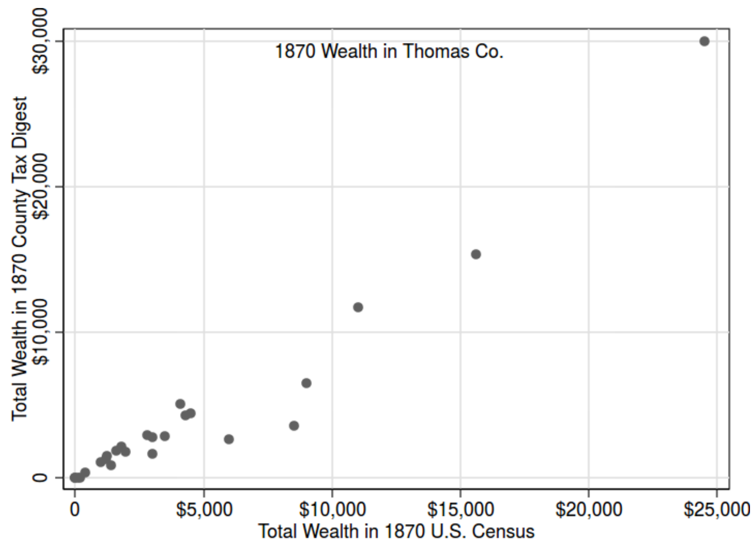
Notes: This figure displays quantile-regression estimates of equation (1) in the text. The coefficient on winning the lottery is reported for various quantiles. Data sources and additional variable and sample definitions are found in the text.

Appendix Figure 3: Old Cherokee County and the 1850 Locations of the Sample



Notes: This figure displays a map of the United States with information on the location (by county) in 1850 of the lottery-eligible households in our main sample. (A cropped version of this map appears in Figure 1.) Black lines indicate the 1850 county boundaries, drawn from the NHGIS database. The area shaded in blue in northwest Georgia denotes old Cherokee County, which was allocated by the Cherokee Lottery of 1832. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. If households in our sample are resident in a county in 1850, we place a red dot at the county centroid. The area of a dot is proportional to the number of sample households resident in that county. Data sources and additional variable and sample definitions are found in the text.

Appendix Figure 4: Wealth in Thomas Co. (Georgia) 1870 Tax Digests and Census Manuscripts.



Notes: This figure is a scatterplot of two measures of wealth in 1870. See Appendix F for details.

Appendix Table 1: Falsification test using South Carolina instead of Georgia to construct sample

Dependent variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lottery-status variables:	Born in Georgia	Born in South Carolina	Number Ga.-born children, pre-lottery	Number SC-born children, pre-lottery	Resides in Old Cherokee County	Number children born post 1832	Real-estate Wealth ($)	Real-estate Wealth >$100	In school, children ages [5,17]
Panel A: South Carolina, basic specification									
Dummy for unique match to Smith (1838) list	0.001	-0.017		-0.019	0.005	0.019	-41.1	0.001	0.025
	(0.004)	(0.013)		(0.019)	(0.007)	(0.077)	(236.2)	(0.016)	(0.021)
Dummy for match to Smith (1838), deflated to 1/n in case of ties	0.004	-0.016		-0.002	0.010	0.001	-15.6	0.001	0.025
	(0.004)	(0.012)		(0.018)	(0.007)	(0.074)	(232.0)	(0.015)	(0.020)
Panel B: South Carolina, including surname fixed effects									
Dummy for unique match to Smith (1838) list	0.000	-0.003		-0.016	0.006	0.016	-93.3	0.005	0.053
	(0.004)	(0.014)		(0.021)	(0.008)	(0.096)	(229.4)	(0.016)	(0.023) **
Dummy for match to Smith (1838), deflated to 1/n in case of ties	0.003	-0.004		-0.004	0.009	-0.029	-72.6	0.015	0.055
	(0.004)	(0.014)		(0.020)	(0.008)	(0.093)	-(72.6)	(0.015)	(0.022) **
Panel C: Analogous results for Georgia, dummy for unique match to Smith list									
Basic specification	-0.004	0.014	0.002		0.022	0.134	295.2	0.002	-0.001
	(0.012)	(0.011)	(0.014)		(0.008) ***	(0.058) **	(154.4) *	(0.011)	(0.011)
Control for surname fixed effects	0.001	0.012	0.009		0.023	0.193	315.8	0.002	-0.003
	(0.014)	(0.012)	(0.016)		(0.008) ***	(0.073) ***	(146.8) ***	(0.011)	(0.011)

Notes: This table displays estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on "winning the lottery" is reported. The sample for Panels A and B consists of all households in the 1850 census with children born in South Carolina during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. The sample for Panel C, which repeats some results from earlier tables, uses households with Georgia-born children in this same window. We use two measures of whether the person won land in the drawing for the Cherokee Land Lottery of 1832. The first measure is coded to 1 if that person is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure takes individuals that “tie” for a match to the Smith list with (n-1) other observations and recodes them to 1/n. Note that these are spurious measures for the South-Carolina samples because the birthplace of their children implies that they lived outside of Georgia at some point during the three years prior to the lottery, and were therefore ineligible. The basic specification also includes dummies for age. The other specification used includes fixed effects for surname (soundex). The dependent variables are indicated in the column headings. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors (shown in parentheses) are heteroskedasticity robust and clustered on the lottery-eligible man if there are multiple observations per household. Data sources and additional variable and sample definitions are found in the text.

Appendix Table 2: Replicate 1850 fertility and schooling results with 1830-50 linked sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Full sample		Children born 1825-1830 and resident in 1830 household?				Children born in Georgia in pre-lottery window & resident in 1850 household?			
			Yes		No		Yes		No	
Outcome variable (in 1850):										
Number of children aged 17 and under in the household	0.085	0.080	-0.028	0.186	0.190	0.172	0.195	-0.037	-0.012	-0.091
	(0.130)	(0.176)	(0.178)	(0.315)	(0.190)	(0.307)	(0.238)	(0.527)	(0.152)	(0.219)
	[4328]	[4328]	[2065]	[2065]	[2263]	[2263]	[1084]	[1084]	[3244]	[3244]
Number of children aged 5 and under in the household	0.051	0.024	0.014	0.081	0.086	0.014	0.053	-0.071	0.052	-0.021
	(0.050)	(0.069)	(0.069)	(0.128)	(0.071)	(0.120)	(0.089)	(0.199)	(0.060)	(0.093)
	[4328]	[4328]	[2065]	[2065]	[2263]	[2263]	[1084]	[1084]	[3244]	[3244]
School attendance, children aged [5,17]	-0.004	-0.001	0.020	0.011	-0.027	0.009	-0.039	0.023	0.023	0.012
	(0.022)	(0.027)	(0.030)	(0.042)	(0.033)	(0.048)	(0.037)	(0.060)	(0.028)	(0.035)
	[11630]	[11630]	[5722]	[5722]	[5908]	[5908]	[3721]	[3721]	[7909]	[7909]
Surname fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes:
This table presents regressions of equation (1) from the text. The unique match to the Smith (1838) list is used as the main RHS variable.
Robust standard errors (clustered on the lottery-eligible man) in parentheses. Sample size in brackets.
None of the coefficient estimates is statistically significant at conventional confidence levels.
Specification includes dummies for age and gender (for children).

Appendix Table 3: Representativeness of main sample

	Linear regression:		Quantile regression:			Summary statistics:	
	Simple	Surname fixed effects	Median	25th %tile	75th %tile	Mean	Median
			Panel A: Outcomes Observed in 1850				
Number of Georgia-native children born in the three years prior to the lottery	1.377	1.379	1.000	1.000	2.000	0.354	0
	(0.017) ***	(0.021) ***	(0.000)	(0.000)	(0.000)		
Age in 1850	-0.073	-0.357	1.000	2.000	-3.000	52.236	50
	(0.320)	(0.476)	(0.489) **	(0.510) ***	(0.786) ***		
Number of children aged 17 and under in the household	0.975	0.980	1.000	2.000	1.000	3.230	3
	(0.091) ***	(0.126) ***	(0.000)	(0.000)	(0.000)		
Number of children aged 5 and under in the household	-0.047	-0.041	0.000	0.000	-1.000	0.771	0
	(0.035)	(0.048)	(0.224)	(0.000)	(0.410) **		
Cannot read or write	0.006	0.015	0.000	0.000	0.000	0.147	0
	(0.013)	(0.017)	(0.000)	(0.000)	(0.000)		
Real-estate wealth	375.7	109.7	200.0	100.0	600.0	2586.3	1000.0
	(250.3)	(325.0)	(90.8) **	(0.0)	(289.7) **		
Frequency of Surname in Southeastern US	-54.0		-28.0	-20.0	-8.0	831.7	366.0
	(41.1)		(18.8)	(9.8) **	(78.0)		
Mean wealth of families in the South with same surname	-19.1		-12.8	-24.4	-15.7	1198.9	1153.0
	(18.2)		(9.4)	(22.0)	(17.2)		
Median wealth of families in the South with same surname	8.5		-4.5	-10.0	6.0	190.8	187.5
	(6.3)		(9.9)	(12.1)	(18.8)		
Mean illiteracy of adults in the South with same surname	0.001		0.000	0.000	0.000	0.174	0.174
	(0.002)		(0.001)	(0.002)	(0.002)		
Mean school attendance of children in the South with same surname	0.000		0.001	-0.002	0.002	0.324	0.323
	(0.002)		(0.001)	(0.002)	(0.003)		
			Panel B: Outcomes Observed in 1830				
Number of free persons in the household	0.149	0.158	0.000	0.000	0.000	5.902	6
	(0.099)	(0.141)	(0.686)	(0.224)	(0.000)		
Number of white children aged 5 and under in the household	0.573	0.553	1.000	1.000	1.000	1.353	1
	(0.038) ***	(0.054) ***			(0.503) **		
Number of white children aged 20 and under in the household	0.322	0.289	0.000	1.000	0.000	3.546	3
	(0.085) ***	(0.121) **	(0.444)		(0.308)		
Number of slaves in the household	-1.259	-1.430	0.000	0.000	-2.000	3.474	0
	(0.264) ***	(0.473) ***			(0.587) ***		

Notes: This table presents regressions of indicated outcomes on a dummy for inclusion in our main sample. The data used are drawn from a sample of Georgia household heads in 1830 that we linked forward to 1850.

Appendix Table 4: Simulated effects of wealth and comparison with estimates from the lottery

		Outcome measure:	Fraction w/ zero	Results from simulations or estimates; Mean and [95% Confidence Interval] EV $700	EV $500	EV $300	EV $900	Estimates from above	
(A)	1850	Number children born post 1832	0	.029 [-.008, .065]	.038 [.003, .073]	.045 [.010, .077]	.018 [-.022, .056]	.134 [.018, .250]	3A1
			.25	.012 [-.023, .041]	.023 [-.006, .047]	.031 [.005, .054]	.001 [-.034, .033]		
			.5	-.004 [-.027, .017] †	.006 [-.013, .028]	.017 [-.003, .036]	-.014 [-.039, .011] †		
(B)	1850	Natural log of total children	0	.005 [-.005, .013]	.007 [-.001, .015]	.009 [.001, .017]	.003 [-.008, .012]	.032 [.003, .061]	3A2
			.25	.002 [-.007, .009]	.004 [-.003, .010]	.006 [.000, .012]	-.001 [-.009, .007]		
			.5	-.002 [-.007, .004]	.001 [-.005, .006]	.003 [-.002, .007]	-.004 [-.011, .003]		
(C)	1850	Attended school (children)	0	.054 [.050, .057] †	.042 [.038, .045] †	.028 [.025, .031] †	.064 [.060, .068] †	-.001 [-.023, .021]	3A3
			.25	.049 [.046, .053] †	.039 [.036, .042] †	.026 [.024, .029] †	.058 [.055, .062] †		
			.5	.043 [.040, .045] †	.034 [.032, .037] †	.024 [.022, .026] †	.049 [.047, .053] †		
(D)	1880 (sons and grandchildren)	Number grandchildren (per son) under 18	0	-.045 [-.073, -.012]	-.034 [-.057, -.003]	-.021 [-.043, .009]	-.055 [-.082, -.018]	-.092 [-.202, .018]	6A2
			.25	-.042 [-.065, -.015]	-.033 [-.051, -.005]	-.021 [-.039, .001]	-.049 [-.075, -.026]		
			.5	-.035 [-.061, -.020]	-.029 [-.046, -.011]	-.020 [-.032, -.002]	-.040 [-.070, -.030]		
(E)	1880 (sons and grandchildren)	Occupational score (sons)	0	.347 [.207, .482]	.229 [.086, .351]	.112 [-.013, .233]	.458 [.319, .598]	-.124 [-.683, .435]	6A4
			.25	.354 [.243, .477]	.247 [.134, .338]	.126 [.025, .216]	.459 [.331, .609]		
			.5	.346 [.260, .442]	.252 [.174, .322]	.142 [.076, .209]	.435 [.311, .550]		
(F)	1880 (sons and grandchildren)	Unable to read and write (sons)	0	-.021 [-.025, -.015] †	-.018 [-.022, -.013] †	-.015 [-.019, -.010]	-.023 [-.028, -.018] †	.004 [-.012, .020]	6A3
			.25	-.018 [-.022, -.014] †	-.016 [-.019, -.011]	-.012 [-.016, -.009]	-.020 [-.024, -.015] †		
			.5	-.015 [-.017, -.011]	-.012 [-.015, -.009]	-.010 [-.012, -.007]	-.016 [-.019, -.013] †		
(G)	1880 (sons and grandchildren)	Unable to read and write (grandchildren)	0	-.030 [-.035, -.025]	-.026 [-.031, -.022]	-.020 [-.025, -.016]	-.033 [-.039, -.028] †	-.002 [-.027, .023]	6A5
			.25	-.025 [-.030, -.022]	-.022 [-.026, -.019]	-.018 [-.021, -.014]	-.028 [-.032, -.024]		
			.5	-.020 [-.023, -.017]	-.018 [-.020, -.015]	-.014 [-.016, -.012]	-.023 [-.026, -.019]		
(H)	1880 (sons and grandchildren)	Attended school (grandchildren)	0	.005 [.001, .009]	.001 [-.003, .004]	-.002 [-.005, .001]	.008 [.004, .013] †	-.020 [-.044, .004]	6A6
			.25	.007 [.003, .010]	.003 [.000, .006]	-.001 [-.003, .002]	.010 [.006, .014] †		
			.5	.008 [.006, .011] †	.005 [.003, .008]	.001 [-.001, .003]	.011 [.008, .013] †		
(I)	1870 (sons only)	Total Wealth ($)	0	70 [-199, 227]	20 [-252, 177]	-27 [-261, 115]	116 [-170, 281]	58 [-113, 228]	6A8
			.25	93 [-123, 218]	46 [-169, 163]	-5 [-213, 110]	135 [-82, 270]		
			.5	111 [-43, 208]	68 [-78, 151]	22 [-131, 99]	146 [-22, 262]		
(J)	1870 (sons only)	Wealth is positive	0	.035 [.028, .043] †	.032 [.024, .039] †	.026 [.018, .033] †	.038 [.029, .048] †	-.031 [-.062, .000]	6A9
			.25	.029 [.023, .036] †	.026 [.020, .033] †	.022 [.016, .027] †	.031 [.025, .040] †		
			.5	.022 [.017, .027] †	.020 [.015, .025] †	.017 [.013, .021] †	.023 [.019, .030] †		
(K)	1870 (sons only)	Natural log of total wealth	0	.320 [.298, .339] †	.239 [.221, .259] †	.129 [.111, .148] †	.382 [.359, .403] †	-.038 [-.132, .056]	6A10
			.25	.292 [.277, .310] †	.231 [.217, .247] †	.144 [.131, .157] †	.345 [.323, .362] †		
			.5	.249 [.238, .266] †	.208 [.195, .219] †	.142 [.133, .152] †	.290 [.274, .303] †		

Notes: This table provides a shift-share analysis with the differences in probability generated by various perturbations of the wealth distribution and the relationship between each outcome and 1850 wealth in the control group. The outcome measures and the year in which they are measured are displayed on the leftmost columns of the table. We use a discretized distribution of 1850 wealth using 100 grid points evenly spaced across 1850 log wealth. For each simulation, we specify the expected value of winning (in 1850$), as denoted in the "EV" column-group headings. For each outcome and expected value of winning, we conduct three simulations with varying degrees of heterogeneity in the value of land winnings. These are denoted in the column "Fraction w/ zero", and indicate the fraction of the simulated winners that receive zero change in wealth. The rightmost columns display estimates from the treatment/control comparisons above. The final column on the right (a number-letter-number sequence) denotes the Table, Panel, and Column from which the estimate is drawn. Each row and column group displays the mean and, in square brackets, the 95% confidence interval from a different simulation. A dagger denotes that the confidence interval for that simulation does not overlap with the confidence interval estimated from the lottery treatment. The data for the simulation are the lottery losers, defined as those with no match to the Smith (1838) list. The statistics for each simulation come from 500 bootstrapped samples of the control group, with the lottery-eligible man being the block for the bootstrap when the outcomes are for their descendants.

Appendix Table 5: Differences by Lottery Status in Characteristics of 1850 Residence

(1) Resides in Old Cherokee County	(2) Resides in Georgia	(3) Miles East	(4) Miles North	(5) School Enroll. Rate	(6) Total Fertility Rate (TFR5)	(7) Total Fertility Rate (TFR19)	(8) Log of Farm Value per Acre	(9) Log of Average Farm Size	(10) Log of Improved Land Ratio	(11) Log Slaves per Area	(12) Log Pop. Density in 1850	(13) Log Pop. Density in 1830	(14) Log Fraction Urban	(15) Resides in town or city	(16) Access to Water Transport	(17) Access to Railroads
Panel A: Basic Specification																
0.022	0.006	4.319	-4.301	-0.003	0.007	0.013	-0.010	-0.014	-0.019	-0.005	-0.045	-0.111	0.073	-0.007	-0.006	0.017
(0.008) ***	(0.011)	(3.652)	(2.227) *	(0.003)	(0.004) *	(0.011)	(0.021)	(0.017)	(0.018)	(0.026)	(0.029)	(0.052) **	(0.065)	(0.003) **	(0.010)	(0.016)
Panel B: Control for Surname Fixed Effects																
0.022	0.005	4.393	-4.902	-0.004	0.006	0.012	-0.015	-0.005	-0.026	-0.004	-0.057	-0.117	0.055	-0.008	-0.001	0.013
(0.008) ***	(0.013)	(3.999)	(2.321) **	(0.003)	(0.004) *	(0.011)	(0.022)	(0.017)	(0.019)	(0.026)	(0.030) *	(0.054) **	(0.067)	(0.004) **	(0.010)	(0.016)
Panel C: Basic Specification, Control for Residence in Old Cherokee County																
---	---	4.655	-6.196	-0.004	0.006	0.009	-0.016	-0.004	-0.017	0.005	-0.048	-0.101	0.073	-0.007	-0.002	0.016
		(4.331)	(2.787) **	(0.003)	(0.004)	(0.011)	(0.022)	(0.016)	(0.018)	(0.025)	(0.029)	(0.052) *	(0.065)	(0.003) **	(0.010)	(0.015)
Panel D: Control for Surname Fixed Effects, Control for Residence in Old Cherokee County																
---	---	4.697	-6.807	-0.005	0.005	0.008	-0.020	0.004	-0.024	0.005	-0.058	-0.108	0.055	-0.008	0.003	0.012
		(4.744)	(2.777) **	(0.003)	(0.004)	(0.011)	(0.022)	(0.016)	(0.019)	(0.025)	(0.030) *	(0.054) **	(0.067)	(0.004) **	(0.010)	(0.015)

Notes: This table displays OLS estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on winning the lottery is reported. The basic specification (shown in Panel A) includes dummies for age. The specification used in Panel B also includes fixed effects for surname (soundex). Panels C and D repeat specifications from Panels A and B, respectively, but also include a dummy variable for residence in Old Cherokee County. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. The dependent variables are the locational county-specific characteristics denoted in the column headings. Location data used in Columns 3 and 4 are county centroids computed from NHGIS data, and are converted into miles east or north of the NAD83 reference point in central Oklahoma. County data used in Columns 5-14 are drawn from ICPSR study #2896. The number of observations for Columns 1-4 is 14375 and for Columns 5-17 is 14237 because of missing data for some (mostly unorganized) counties. A household is coded as a lottery winner if the head is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors are heteroskedasticity robust and, in Columns 3-13 and 15-17, clustered at the (state x county) level to account for multiple observations per county. Data sources and additional variable and sample definitions are found in the text.

Appendix Table 6: Statistics for Predetermined Outcomes, by Lottery Status and by Linkage to 1880

	(1)	(2)	(3)	(4)
	Whole Sample	Lottery "Losers"	Lottery "Winners"	p-value, mean difference [N]
Panel A: Lottery-Eligible Men in 1850 with No Sons Linked to 1880				
Age, in years	52.7	52.6	53.2	0.143
	(9.2)	(9.2)	(9.5)	[5895]
Born in Georgia	0.470	0.473	0.453	0.351
	(0.499)	(0.499)	(0.498)	[5911]
Born in South Carolina	0.199	0.199	0.201	0.912
	(0.399)	(0.399)	(0.401)	[5911]
Born in North Carolina	0.194	0.194	0.198	0.804
	(0.395)	(0.395)	(0.399)	[5911]
Number of Georgia-born children in the three years prior to the lottery	1.274	1.274	1.274	0.996
	(0.497)	(0.496)	(0.507)	[5911]
Cannot read and write	0.151	0.151	0.151	0.884
	(0.358)	(0.358)	(0.356)	[5821]
Number of letters in surname	6.19	6.19	6.16	0.633
	(1.64)	(1.66)	(1.48)	[5911]
Frequency with which surname appears in sample	17.0	17.1	15.9	0.183
	(23.9)	(24.3)	(20.1)	[5911]
Surname begins with "M" or "O"	0.106	0.106	0.104	0.835
	(0.308)	(0.308)	(0.305)	[5911]
Mean wealth of families in the South with same surname	1219.2	1219.1	1220.2	0.947
	(491.4)	(502.9)	(384.5)	[5747]
Median wealth of families in the South with same surname	184.0	184.2	182.6	0.776
	(169.6)	(174.1)	(126.0)	[5747]
Mean illiteracy of adults in the South with same surname	0.173	0.173	0.174	0.625
	(0.043)	(0.043)	(0.037)	[5747]
Mean school attendance of children in the South with same surname	0.323	0.323	0.324	0.496
	(0.051)	(0.051)	(0.048)	[5690]

Notes: Table continues on next page.

Appendix Table 6 (continued): Summary Statistics for Predetermined Outcomes, by Linkage Status to 1880

	(1) Whole Sample	(2) Lottery "Losers"	(3) Lottery "Winners"	(4) p-value, mean difference [N]
Panel B: Lottery-Eligible Men in 1850 with At Least One Son Linked to 1880				
Age, in years	50.2 (7.9)	50.3 (7.9)	49.7 (7.7)	0.015 [8454]
Born in Georgia	0.515 (0.500)	0.514 (0.500)	0.523 (0.500)	0.562 [8464]
Born in South Carolina	0.220 (0.414)	0.218 (0.413)	0.233 (0.423)	0.256 [8464]
Born in North Carolina	0.170 (0.376)	0.171 (0.376)	0.167 (0.373)	0.759 [8464]
Number of Georgia-born children in the three years prior to the lottery	1.374 (0.567)	1.376 (0.568)	1.363 (0.558)	0.468 [8464]
Cannot read and write	0.144 (0.351)	0.145 (0.352)	0.140 (0.347)	0.665 [8450]
Number of letters in surname	6.19 (1.59)	6.21 (1.60)	6.11 (1.53)	0.057 [8464]
Frequency with which surname appears in sample	20.6 (23.7)	20.6 (23.8)	20.5 (22.5)	0.987 [8464]
Surname begins with "M" or "O"	0.098 (0.297)	0.097 (0.296)	0.104 (0.305)	0.508 [8464]
Mean wealth of families in the South with same surname	1192.4 (410.4)	1194.0 (417.5)	1182.4 (362.1)	0.326 [8346]
Median wealth of families in the South with same surname	184.2 (157.0)	184.9 (162.2)	180.2 (119.5)	0.249 [8346]
Mean illiteracy of adults in the South with same surname	0.176 (0.043)	0.175 (0.044)	0.177 (0.040)	0.160 [8346]
Mean school attendance of children in the South with same surname	0.322 (0.052)	0.323 (0.053)	0.322 (0.049)	0.639 [8285]

Notes: This table displays summary statistics for the main data used in the present study. Subsamples are indicated in the panel headings. (Results for the full sample are seen in Table 1.) The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. Column (1) presents means and standard deviations (in parentheses) of variables for this entire sample. Columns (2) and (3) present means and standard deviations of variables for the subsamples of, respectively, lottery losers and winners. Column (4) presents the p-value on the test of zero difference in means between the subsamples of losers and winners. In square brackets, we report the sample size used for this test, although the test involving children or surnames adjust for the clustering of errors. With the exception of the measure of surname length, we use the Soundex version of each name to account for minor spelling differences. For the variables that are means by surname, we use a preliminary copy of the 1850 100% census file to construct average fertility, school attendance, and real-estate wealth among households in Alabama, Florida, Georgia Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia, for each (soundex) surname. (Those individuals that appear in our lottery-eligible sample are excluded from the construction of these indices.) Data sources and additional variable and sample definitions are found in the text.

Appendix Table 7: Outcomes of Next Generation(s) in 1860-80

(0)	(1)	(2)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Sons of Lottery-Eligible Men											Grandchildren			
				1880		1870			1860			1880			
Match to list of winners:	Linked to 1880 census	Linked to 1870 census	Linked to 1860 census	Unable to read and write	Occup. Score	Total Wealth ($)	Wealth Positive	Natural log of Wealth	Total Wealth ($)	Wealth Positive	Natural log of Wealth	Unable to read and write	Enrolled in school	Number children under 10	Number children under 18
1. Estimates of the effect of father or grandfather winning the lottery															
Panel A: Basic Specification															
Binary	0.040	0.036	0.017	0.003	-0.230	-13.2	-0.032	-0.036	-53.0	-0.010	0.030	-0.004	-0.021	-0.055	-0.097
	(0.008) ***	(0.008) ***	(0.004) ***	(0.009)	(0.304)	(70.9)	(0.013) **	(0.032)	(159.4)	(0.023)	(0.105)	(0.014)	(0.012) *	(0.041)	(0.060)
1/n	0.038	0.029	0.012	0.006	-0.226	-47.3	-0.027	-0.030	-102.0	-0.009	0.026	0.003	-0.012	-0.059	-0.089
	(0.008) ***	(0.008) ***	(0.004) ***	(0.009)	(0.304)	(72.3)	(0.013) **	(0.032)	(158.7)	(0.023)	(0.105)	(0.013)	(0.012)	(0.041)	(0.059)
Panel B: Control for Surname Fixed Effects															
Binary	0.030	0.029	0.013	0.006	-0.217	53.4	-0.021	-0.004	-76.3	0.009	-0.063	-0.006	-0.026	-0.044	-0.086
	(0.008) ***	(0.008) ***	(0.004) ***	(0.010)	(0.365)	(68.8)	(0.015)	(0.037)	(173.9)	(0.033)	(0.175)	(0.014)	(0.013) **	(0.046)	(0.066)
1/n	0.029	0.025	0.011	0.010	-0.184	24.2	-0.021	-0.007	-90.6	0.009	-0.001	0.001	-0.020	-0.051	-0.076
	(0.008) ***	(0.008) ***	(0.004) ***	(0.010)	(0.362)	(68.3)	(0.014)	(0.037)	(173.1)	(0.032)	(0.173)	(0.014)	(0.013)	(0.046)	(0.066)
Panel C: Control for Surname Effects and Length of Given Name															
Binary	0.019	0.014	0.001	0.014	0.190	112.3	-0.006	0.050	46.3	-0.014	0.116	-0.005	-0.032	-0.049	-0.120
	(0.009) **	(0.010)	(0.004)	(0.012)	(0.440)	(82.6)	(0.019)	(0.048)	(287.1)	(0.050)	(0.316)	(0.016)	(0.014) **	(0.056)	(0.079)
1/n	0.015	0.008	0.000	0.019	0.284	73.7	-0.005	0.048	48.0	-0.016	0.202	0.007	-0.024	-0.055	-0.099
	(0.009)	(0.009)	(0.004)	(0.011)	(0.435)	(82.6)	(0.018)	(0.048)	(282.4)	(0.050)	(0.304)	(0.016)	(0.014) *	(0.055)	(0.079)
Panel D: Control for Surname Effects; Length of Given Name; Order in and Size of 1850 Household															
Binary	0.017	0.013	0.001	0.013	0.218	110.6	-0.007	0.047	50.5	-0.006	0.102	-0.005	-0.030	-0.053	-0.123
	(0.009) *	(0.010)	(0.004)	(0.011)	(0.440)	(82.5)	(0.019)	(0.048)	(294.9)	(0.050)	(0.323)	(0.016)	(0.014) **	(0.056)	(0.080)
1/n	0.014	0.008	0.000	0.018	0.298	71.1	-0.006	0.043	31.4	-0.010	0.138	0.007	-0.023	-0.059	-0.102
	(0.009)	(0.009)	(0.004)	(0.011)	(0.436)	(82.1)	(0.018)	(0.047)	(284.3)	(0.050)	(0.307)	(0.016)	(0.014)	(0.055)	(0.079)
Panel E: Control for Surname Effects and Length of Given Name; Only Eldest Sons in 1850															
Binary	0.021	0.027	-0.003	-0.032	-0.267	93.1	-0.034	0.035	-1001.4	-0.070	0.526	-0.045	-0.014	-0.025	-0.101
	(0.018)	(0.018)	(0.009)	(0.024)	(0.974)	(160.9)	(0.035)	(0.082)	(1076.9)	(0.212)	(1.784)	(0.029)	(0.027)	(0.126)	(0.188)
1/n	0.017	0.024	-0.003	-0.035	0.211	93.6	-0.040	0.040	-885.9	-0.045	0.699	-0.055	0.007	-0.059	-0.072
	(0.017)	(0.018)	(0.009)	(0.024)	(0.981)	(161.6)	(0.034)	(0.082)	(1020.0)	(0.201)	(1.570)	(0.028) *	(0.027)	(0.124)	(0.187)
2. Estimation sample															
	Children in 1850	Children in 1850, if adult in 1870	Children in 1850, if adult in 1860	1850 children as adults in 1880	1850 children as adults in 1880	1850 children as adults in 1870	1850 children as adults in 1870	1850 children as adults in 1870	1850 children as adults in 1860	1850 children as adults in 1860	1850 children as adults in 1860	Children in 1880, 10-19 years old	Children in 1880, ages 5-19	1850 children as adults in 1880	1850 children as adults in 1880
	[N=40024]	[N=38529]	[N=24510]	[N=14963]	[N=14956]	[N=12235]	[N=12235]	[N=12235]	[N=3306]	[N=3306]	[N=3306]	[N=23544]	[N=40658]	[N=14963]	[N=14963]

Notes: This table displays OLS estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on winning the lottery is reported. The basic specification (shown in Panel A) includes dummies for the 1850 ages of the lottery-eligible man and his children. Specifications with grandchildren also include dummies for their 1880 age x gender. The specification used in Panel B also includes fixed effects for surname (soundex), and the specification in Panel C adds to this dummies for the length (number of letters) of the given name. Panel D reports the results with controls from Panel C and also dummies for the order in which the son appears in the 1850 household and for the number of children in the 1850 household. Panel E repeats the specification from Panel C, but only for the subsample of eldest sons (as measured in the 1850 household). The base sample of children in 1850 is as described in prior tables, and this sample is used in Columns 1, 2 and 3 to estimate the differential probability of linkage to 1860, 1870, and 1880 censuses. The samples in the remaining columns are drawn from the 1860, 1870, or 1880 households of those male children linked from 1850. The dependent variables are indicated in the column headings. For the binary lottery-status variable, a household is coded as a lottery winner if the head is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure takes individuals that "tie" for a match to the Smith list with (n-1) other observations and recodes them to 1/n. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors are heteroskedasticity robust and clustered on the lottery-eligible man if there are multiple observations per household. Data sources and additional variable and sample definitions are found in the text.

Appendix Table 8: Replication of Table 2, by Age of Father

		(1)	(2)	(3)
		Sample restriction on the father's age		
Specification:	Match to list of winners:	None	50 and below	Over 50
Panel A: Post-1832 fertility of lottery-eligible men				
Basic	Binary	0.132	0.153	0.106
		(0.058) **	(0.082) *	(0.080)
	1/n	0.137	0.171	0.095
		(0.056) **	(0.080) **	(0.078)
Surname	Binary	0.184	0.120	0.114
		(0.073) **	(0.099)	(0.100)
	1/n	0.175	0.115	0.104
		(0.072) **	(0.096)	(0.098)
Panel B: School attendance of children aged 5-17				
Basic	Binary	-0.005	-0.006	-0.002
		(0.011)	(0.014)	(0.018)
	1/n	-0.004	-0.003	-0.004
		(0.011)	(0.014)	(0.018)
Surname	Binary	-0.005	-0.008	-0.002
		(0.011)	(0.018)	(0.020)
	1/n	-0.006	-0.010	-0.004
		(0.011)	(0.015)	(0.019)

Notes: See footnote for Table 2. Additional sample restrictions are indicated in the column headings.

Appendix Table 9: Intergenerational Wealth Elasticities (Cross Sectional)

	(1)	(2)	(3)	(4)
	1860		1870	
	Log wealth, exclude zeros	Log wealth, adjusted for truncation	Log wealth, exclude zeros	Log wealth, adjusted for truncation
1. Estimates of the effect of father's wealth in 1850				
Panel A: Basic Specification				
	0.451	0.213	0.277	0.137
	(0.029) ***	(0.014) ***	(0.013) ***	(0.007) ***
Panel B: Control for Surname Fixed Effects				
	0.341	0.181	0.266	0.128
	(0.051) ***	(0.019) ***	(0.015) ***	(0.007) ***
Panel C: Control for Surname Effects and Length of Given Name				
	0.217	0.176	0.254	0.130
	(0.098) **	(0.027) ***	(0.021) ***	(0.009) ***
2. Estimation sample				
	1850 children as adults in 1860	1850 children as adults in 1860	1850 children as adults in 1870	1850 children as adults in 1870
	[N=1676]	[N=3306]	[N=6201]	[N=12235]

Notes: This table displays OLS estimates of the intergenerational wealth elasticity in the sample of lottery-eligibles. Each cell presents results from a separate regression, and only the coefficient on father's wealth in 1850 is reported. The basic specification (shown in Panel A) includes dummies for the 1850 ages of the lottery-eligible man and his children. The specification used in Panel B also includes fixed effects for surname (soundex), and the specification in Panel C adds to this dummies for the length (number of letters) of the given name. The samples consist of the 1860 or 1870 households of those male children linked from 1850. The dependent variables are two measures of log total wealth. The first measure excludes zero wealth from the data. The second measure imputes value of wealth using an estimate of the truncated normal distribution. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors are heteroskedasticity robust and clustered on the lottery-eligible man if there are multiple observations per household. Data sources and additional variable and sample definitions are found in the text.

Appendix Table 10: Correlation of Descendants' Outcomes with 1850 Log Wealth of Lottery-Eligible Man, Control Group Only

	Correlation	P-value
1. 1850, outcomes for children		
School enrollment, ages 5-17	25.07%	.0000
Post-1832 fertility (numbers)	-2.21%	.0247
2. 1870, outcomes for linked sons		
Wealth positive	12.31%	.0000
Wealth in logs	56.75%	.0000
3. 1880, outcomes for linked sons		
Literacy	11.24%	.0000
Occupational income score	10.90%	.0000
4. 1880, outcomes for linked grandchildren		
Number of grandchildren under 18	-6.13%	.0000
School enrollment, ages 5-17	4.28%	.0000
Literacy	13.49%	.0000

Notes: This table displays estimates of the correlation coefficient between the 1850 wealth of the lottery-eligible man and selected outcomes of his descendants. Only the control group (no match to Smith list) is used to estimate these correlations. P-values on the null hypothesis of zero correlation are displayed. The measures of log wealth include an imputed value of missing wealth using an estimate of the truncated normal distribution. Data sources and additional variable and sample definitions are found in the text.

Appendix A. Lottery Participation Rate Calculation

Knowing what fraction of the eligible population actually participated in the 1832 Georgia land lottery requires two pieces of information: the number of registered participants and the number of eligibles. A report in *Niles Weekly Register* (Oct. 27, 1832, p. 131), referenced in Cadle (1991, p. 278), tells us there were 85,000 "names" or "blanks" entered in the drawing. The words "names," "blanks," and "tickets" are used interchangeably throughout this contemporaneous account. The last term – "tickets" – corresponds to the common definition of individual entries in a lottery. "Names" therefore likely refers here to slips of paper with names on them, rather than to a set of distinct individuals, as the lottery rules allowed some individuals (*e.g.* male household heads) to take two draws. We first need to determine how many males age 18+ in 1832 were represented among the 85,000 slips of paper.

Transcribed lists of individual registrants for 1832 in several districts of Columbia, Oglethorpe, and Wilkes Counties are available online. In these districts, there were 882 registered males age 18+ (i.e. neither widows nor orphans under age 18) and 1,568 draws, so the ratio of male names to total draws was 0.5625. Extrapolating from this to the entire lottery, we estimate that there were 85,000*0.5625=47,813 distinct males age 18+ in 85,000 draws.

We now need to calculate the denominator of the participation rate: the state's population of free white males age 18+ resident in Georgia 3+ years as of January 1, 1832 who survived to the lottery start on October 22, 1832. The U.S. Census reports that on June 1, 1830, there were 63,000 free white males age 20+ and 15,000 age 15-19, roughly 3,000 in each of the latter 5 birth cohorts. Of the 15,000 in 1830, approximately 10,750 would have been age 18+ by January 1, 1832 (none of the 15 year olds, 7/12 of the 16 year olds, and all of the 17-19 year olds). So the total white male population age 18+ on January 1, 1832 would have been 73, 750 before accounting for mortality, out-migration before the lottery, and in-migrants present in 1830 who were born before December 31, 1813 but entered Georgia after December 31, 1828 (and thus would not have meet the 3 year residence rule).

We can roughly determine the values we need for mortality, in-migration, and out-migration by examining the components of the change in the state's population between 1830 and 1840 for males age 10+ in 1830. The total population change between 1830 & 1840 for individuals age 10+ in 1830 is

$$\underset{(1)}{\Delta \text{Population}_{1830\text{-}40}} = \underset{(2)}{(\text{Age } 20+)_{1840}} - \underset{(3)}{(\text{Age } 10+)_{1830}} = \underset{(4)}{-\,(\text{Deaths} + \text{Out-Migrants})} \underset{(5)}{+ \text{In-Migrants}}$$

For this cohort, (1)=84,843 and (2)=95,552. Hacker (2010, p. 70) gives 10-year survival rates for this decade's white male U.S. population so (3)=12,633 (or 3,026 from June 1, 1830 to October 22, 1832). To determine (4), note that we find 72.3% of eligibles in our sample still in Georgia June 1, 1850, so their out-migration rate is 27.7% over 17 years, 6 months, and 1 week (from October 22, 1832 to June 1, 1850), or 0.13% per month and 15.6% 1830-40. This is too high for this period, however, as it is calculated based in part on 1840-50 when the set of destinations for potential out-migrants was expanded to include Alabama, Mississippi, and Texas, and it comes after all of Georgia's land was allocated. If we reduce the 1830-40 out-migration rate by ½ to 7.8%, then (4)=7,453 (or 1,786 from June 1, 1830 to October 22, 1832). As a result (5)=9,377 for the decade (or 2,246 from June 1, 1830 to October 22, 1832). For in-migrants, what we actually need, however, is the number entering Georgia from December 31, 1828 through June 1, 1830 (as these would not have been eligible based on the 3 year rule). Assuming the rate of in-migration (relative to end-of-decade population) was the same in these 17 months as it had been in the 1830s (1.6% for 17 months), the in-migrants we actually need to subtract are 1.6% of the 1840 population, or 1,357.

So, eligibles surviving to the lottery are approximately (73,750 – 3,026 – 3,571 – 1,357) = 65,796

We then need to add in under-enumeration of white males age 15+ in 1830. Hacker (2013, p. 93) estimates a 5% undercount in 1850 for Southern-born white males age 10+ in 1850. Assuming and undercount twice as high in 1830 as in 1850 yields an estimate of the white male eligibles on October 22, 1832 of 72,375.

Winners from prior lotteries were ineligible, however, and need to be subtracted before we are finished. In the 1827 lottery, 22,687 plots were awarded; in 1821, 17,388; and in 1820, 28,607. Of these, 76% or 52,198 would have been white males age 18+ in 1820, 1821, or 1827(assuming the same proportions of widows, orphans as in 1832, and males 18+). Approximately 45% of these prior winners would have survived to 1832, so 23,489 need to be subtracted, yielding 48,885 eligibles and a participation rate of 97.8%.

Winners of previous lotteries will be absent from the list of 1832 winners (since they were ineligible to win in 1832), but we have not attempted to purge them from our list of "eligibles" where eligibility was imputed either on the basis of 1850 family structure or presence in the 1830 Georgia census manuscripts. These prior winner can thus end up in our control group. Of those who had been winners in the 1820, 1821, and 1827 lotteries, no more than 20% or 10,439 will have survived to 1850, less than 10% of the age 20+ Georgia white male population in 1850. Like the treated group, they will have received a substantial windfall, so this will bias the case against finding any effect of winning in 1832 on either wealth in 1850 or outcomes for children. The fact that we actually observe an effect of winning in 1832 on 1850 wealth that is statistically significant and roughly equal in magnitude to the value of a parcel of land in the lottery region suggests that this bias is insufficient to dilute the wealth effect. For this reason, we expect that the bias is unlikely to be the source of the lack of an effect on schooling or sons' 1860 or 1870 wealth or 1880 literacy or occupation.

Nonetheless, to account for this possibility, we have partitioned our sample into two groups based on age in 1850: age 50 and below and over age 50. The former will have been considerably less exposed to pre-1832 lotteries than the latter. The results are in online Appendix Table 8. The small positive effect on post-1832 fertility seen in the full sample is also seen in the 50 and below group (though post-lottery fertility for winners is a bit stronger for the younger group, this is no doubt an age effect – they are young enough that their family size was more likely to be increasing after 1832 than those over age 50 in 1850); the lack of any effect on school attendance for the children of lottery winners in the full sample is also seen in this younger group. We view this as evidence that the size of the bias induced by the presence of prior lottery winners in our control group is not great enough to be the source of our finding that lottery winning does not lead to greater investment in children.

Appendix A References:

Cadle, Farris W. 1991. *Georgia land surveying history and law.* University of Georgia Press.

Hacker, J. David. 2010. Decennial life tables for the white population of the United States, 1790-1900. *Historical methods* 43, no. 2: 45-79.

Hacker, J. David. 2013. New estimates of census coverage in the United States, 1850?1930. *Social Science History* 37, no. 1 (2013): 71-101.

Appendix B. Falsification exercise using a placebo sample from South Carolina

We perform a falsification exercise using South Carolina rather than Georgia and do not find statistically significant results. One of the challenges in identifying the treatment effect associated with winning the 1832 Lottery is that our method of imputing lottery status via name matching may introduce biases through sample selection. To check for this possibility, we construct a placebo sample using households with children born only in South Carolina (rather than Georgia) during the same pre-lottery window (the three years prior to the Cherokee Land Lottery of 1832).[1] We use the names among this South Carolina sample to impute a pseudo-lottery-status by linking to the Smith (1838) list. As above, we use both a dummy for a unique match to the Smith list and a variable that allows for probabilistic matches, deflated to $1/n$ case of ties. By the eligibility rules of the Cherokee Land Lottery, any matches from the South Carolina sample to this list must be spurious. It is then reassuring that the fraction of unique matches in the placebo sample derived from South Carolina is only one quarter of the fraction in the Georgia sample.

In Appendix Table 1, we estimate equation (1) using this placebo sample, for the different variables indicating lottery status, and using both the basic specification and the one that includes surname/Soundex fixed effects. These results are found in Panels A and B, with analogous results from the Georgia sample provided for reference in Panel C. The first four columns of Appendix Table 1 show outcomes that were determined prior to the 1832 Lottery, and there are no statistically significant results. (Note that a series of falsifications checks using pre-lottery variables was also performed for the Georgia sample, as shown in Table 1, Panel B.) The remaining columns show post-lottery outcomes such as residing in Old Cherokee County and fertility by 1850. There is no statistically significant pseudo-treatment effect for the South Carolina sample, in contrast to what we find for Georgia.[2] Nor is there a statistically significant effect of treatment for 1850 real-estate wealth, along either intensive or extensive margins (Columns 7 and 8, respectively). In Column 9, there is some evidence of a positive relationship between the pseudo-treatment and school attendance. If we choose to subtract this estimate from the Georgia estimates, it would make the above estimates even less supportive of the idea that wealth allowed families to buy their way around credit constraints to invest more in their children's schooling. Thus, of the 20 estimated effects on the "placebo" sample in Appendix Table 1, only 2 are statistically significant, which is what we would expect at the 5% significance level.

[1] This exercise required the transcription of an additional 55,739 observations.

[2] We generally limited this falsification test to variables that were already available in the 1850 census index. We also transcribed wealth and school attendance in these households from the 1850 Census manuscripts. Our efforts to link to the slave schedule were considerably more skilled-labor intensive, so we did not duplicate these efforts for the placebo sample.

Appendix C. Alternate sample using linked 1830-50 data

In this Appendix, we present an alternative sampling strategy for comparing winners and losers of the 1832 Cherokee Land Lottery.

Quite late into the cycle of this project, we obtained a 100% index of the 1830 Census, which enumerated households more than two years prior to the lottery. Unlike the 1850 Census, the 1830 enumeration did not record names except for that of the household head. It does indicate place of residence, however, and the timing relative to the 1832 lottery allows us to pin down Georgia residence (for those named) almost at the beginning of the three-year window prior to the lottery drawing.

We use these data to construct an alternative sample of lottery eligibles. This sample starts with all households in Georgia in the 1830 census. It therefore ensures that the household heads met the Lottery's Georgia-residency requirement, at least for the early part of the three-year window. We create a dummy variable for a unique (name) match to the Smith list.

We then linked 33,506 (53% of white males age 20+ present in Georgia in 1830) of these men forward to the 1850 Census, whether they were matched to the Smith list as "winners" or not as "losers," to measure post-lottery outcomes. This was done solely on the basis of surname, given name, and residence in the South in 1850 (96 percent of Georgia-born males age 36+ in 1850 – the birth cohorts that would have been age 18+ in 1832 – resided in the South). Only males who were uniquely linked using these characteristics were taken as matches.

An advantage of this approach is that it uses a pre-lottery sample, which might avoid any bias that results from conditioning on the presence of older children in the household in 1850. (Recall that the main strategy of the paper draws a sample of eligibles by virtue of having in the household Georgia-born children with years of birth corresponding to the three-year residency required for the 1832 lottery.) Nevertheless, we would not say that this strategy is more or less problematic because errors might be introduced through the process of linkage. Furthermore, some people observed in 1830 may have died or emigrated from Georgia prior to the 1832 lottery. (Death between 1832 and 1850 would contaminate either strategy, although we argue above that such bias is not great. Emigration from Georgia should not contaminate either sample, as long as such emigration is to elsewhere in the U.S. South in the case of this sample or anywhere in the U.S. in the case of the main sample.)

C.1. Replicating 1850 fertility and school results

When comparing the main and linked-sample results, we find essentially similar estimates for schooling and attenuated estimates for fertility. These estimates are found in Appendix Table 2, which contains results from a regression of various outcomes on the Smith dummy. Specifications with and without surname/soundex fixed effects are shown. The first two columns contain the main result using the full linked sample. The difference in 1850 household fertility and school attendance between lottery winners and losers are statistically similar to results shown above (in Table 2, *e.g.*). (The sample sizes here are somewhat smaller because the filter of unique name matching for census linkage is more stringent than the Georgia-born-kids-in-the-pre-lottery-window filter that we used to construct our sample above.)

We also consider whether the responses look different based on the presence of children in the household. Ideally, we would have a measure from 1832 of the presence of children in the pre-lottery window to match pre-lottery data with what we did above using 1850 data. The best proxy that we can find in the 1830 census is the presence of children born in the previous five years and still resident in the household. Disaggregating the regression results by this proxy yields results that are quite similar to the main results of the paper and to the linked-sample baseline presents in the same table. Next, we disaggregate the results using the presence or absence of children in 1850 born in Georgia during the three-year, pre-lottery window. Again, these results are substantially similar to what we had already found.

We conclude that our method of constructing the sample using children in 1850 to proxy the household's pre-lottery Georgia residence does not seem to be warping the results.

C.2. How representative is our sample of lottery eligible men?

Finally, the new linked 1830-50 sample allows us to ask how our sample used above differs from a representative cross-section of people eligible for the lottery. We take the 1830-50 linked sample and create a dummy variable if an observation would appear in our sample above based on the presence of Georgia-native children born in the pre-lottery window. In Appendix Table 3, we estimate OLS and quantile regressions of selected outcomes on the in-our-sample dummy. (Note that these coefficients tell us the relationship with being in our sample, not the relationship with winning the lottery.) Outcomes observed in 1850 are considered in Panel A, while outcomes from 1830 are shown in Panel B. The most obvious difference is an artifact of the construction of the dummy: in the first row of the table, we see that households that would be in our sample above tend to have more Georgia-native children born in the pre-lottery window. In the next row, we see that the mean age is the same across the samples. Nevertheless, the in-our-sample dummy predicts differences in certain quintiles of the age distribution. These are consistent with our sample above having fewer young men and fewer old men. This accords with the life-cycle pattern of fertility; younger men may not have started a family by 1832 while older men may have finished having children by (three years before) then. The last five rows of Panel A show outcomes related to the surname. There is no clear pattern for these results and, of the 20 coefficients displayed, only one is statistically significant at the 5% level. The next two rows show the number of children in age ranges that are born after the lottery. As can be seen, being in our sample is correlated with higher fertility. This makes sense: households with higher fertility in general would be more likely to have higher fertility any specific window that we chose for designing the sample. In the row that follows, we see that households making it into our sample tend to have more land wealth in 1850, at least at the higher quantiles. We also see in the final row of Panel B that such households have less slave wealth in 1830. We also see in Panel B that, consistent with the fertility results in Panel A, households in our sample had more children in 1830.

In sum, the restriction of having Georgia-native children born in the pre-lottery window yields us a sample of men with somewhat more children, somewhat compressed age distribution, and with the wealth portfolio tilted towards land and away from slaves.

Appendix D. Simulated results using cross-sectional relationships in the control group

Based on the cross-sectional relationship between paternal wealth and sons' outcomes, we would have expected much larger effects of winning the lottery on sons' human capital but not on fertility. We come to this conclusion by conducting a simple shift/share analysis using the expected change in the wealth distribution interacted with the relationship between wealth and various outcomes in the control group. We use the control group to conduct this calculation because we wish to compare the results from the randomized wealth with those for wealth in a sample that did not receive a random wealth disbursement. Some readers might ask why we did not instead set this up as a two-stage-least-squares (2SLS) problem with 1850 wealth as the endogenous regressor. This is inappropriate in that lottery winners may have spent some of their wealth precisely on the human-capital formation of children. This would violate the 2SLS exclusion restriction in that lottery treatment has an effect on child outcomes via a channel *other* than measured 1850 wealth. Such transitional dynamics of wealth would not be present in the control group, which did not receive the extra wealth. However, we would not argue that the relationship between child outcomes and wealth in the control group is necessarily causal, but rather is a useful benchmark. One additional complication that motivates our use of the shift/share analysis (versus a more common comparison of 2SLS and OLS estimates) is that the relationship between wealth and various outcomes might not be linear.

The specifics of the shift/share calculation are as follows. We use 100 grid points, evenly spaced across the distribution of log 1850 total wealth, to discretize the 1850 wealth distribution. Within each cell j there is an estimated average x_j of some outcome. Let the vector of these averages be $\mathbf{x}$ and the probability of being in each cell summarized by the vector $\mathbf{p}=\{p_j\}$. The expected value of this outcome variable across the whole sample is therefore the dot product of $\mathbf{p}$ and $\mathbf{x}$. Suppose the distribution of wealth is perturbed to be $\mathbf{q}$. The change in the expected value of the outcome variable would be $\Delta = (\mathbf{q}-\mathbf{p})\cdot\mathbf{x}$. For a given perturbation of the wealth distribution, we compute the distribution of Δ with 500 bootstraps from the control sample. In the case of child or grandchild outcomes, we use a block bootstrap grouped by the lottery-eligible father.

Results from this exercise are shown in Appendix Table 4. The outcome measures and the year in which they are measured are displayed on the leftmost columns of the table. Each row and column group displays the mean and, in square brackets, the 95% confidence interval from a different simulation. The rightmost columns display estimates from the lottery-based design above. A dagger denotes that the confidence interval for that simulation does not overlap with the confidence interval estimated from the lottery treatment. For each simulation, we specify an expected value of winning, discounted to 1850 and denoted in the "EV" column-group headings. The first expected value we use for the simulation is $700, corresponding to our estimate in Section 2 of the value of land won. We also consider expected values $200 above and $200 and $400 below $700. We focus mostly on the $700 case, but discuss the robustness to alternate assumptions.

We also allow for heterogeneity in the value of land winnings by using a simple, two-point distribution including zero as a possible "prize." For each outcome and expected value of winning, we conduct three simulations with varying degrees of heterogeneity. These are denoted in the column "Fraction with zero," and indicate the fraction t of the simulated winners that receive zero change in wealth. In other words, we use the control sample to construct $\mathbf{p}$ using 1850 total wealth. We then define a perturbed-wealth variable equal to measured wealth for t of the sample and equal to measured wealth plus EV/(1-

t) for the remaining (1- *t*). (Receiving zero wealth is randomly assigned separately for each bootstrapped sample.) In the end, these alternate assumptions do not make much difference. For a given outcome and expected value of the lottery prize, the simulated and estimated confidence intervals tend to overlap either in all three cases or in none at all. (Of the 44 blocks of cells in Appendix Table 4, 38 have either zero or three daggers.) In practice, this relative insensitivity arises from the approximate linearity of the relationship between most outcomes and 1850 wealth, at least across the densest part of the wealth distribution.

The simulations are generally consistent with our estimates of lottery treatment on fertility above. These are seen in row-groups A, B, and D, in Appendix Table 4. The fertility/wealth relationship in the control sample (not shown) is approximately flat in terms of economic significance. (A smoothed plot of 1850 fertility versus 1850 wealth displays an inverse ∪-shape. However, the range on the *y*-axis is quite small and only a minor change in fertility is associated with large changes in wealth.) In the simulations, wealth shocks of various sizes change the fertility rate by only a few children per hundred. The 95% confidence intervals for the simulation typically do not contain the point estimates from above, but they do overlap with the estimated confidence interval.

With a few exceptions noted below, the simulations are generally *not* consistent with our estimates of lottery treatment on human-capital variables, particularly at the low end. These results are found in Appendix Table 4, row-groups C and D-J. First, consider 1850 school attendance in row-group C. By this simulation, a homogeneous $700 wealth shock would increase school attendance by approximately 5.4%. This is different in both statistical and economic terms from the lottery-based estimate of -0.001. The simulation delivers larger (smaller) effects for larger (smaller) wealth shocks. In contrast, while we find positive rather than negative simulated effects for the occupational income score of the sons in 1880, there is generally a substantial overlap between the confidence intervals of the simulation and the estimate. (This outcome is complicated by the fact that essentially the entire sample was involved in farming, thus narrowing the occupational range.) The results for the sons' literacy, however, show economically and statistically significant differences between the simulation and the estimates. In the simulation, a positive wealth shock should have reduced the rate of illiteracy. However, the relationship between 1850 wealth and human capital of the descendants is weaker for the grandchildren than for the sons. Accordingly, there is substantial overlap in the confidence intervals for grandchildren's human capital, except for the $900 wealth shock.

Finally, we examine the 1870 wealth of sons, which the simulations suggest would have been markedly different at the low end (row-groups I-K). By the simulation, we would have expected an increase in the proportion of the sons with positive wealth in 1870, rather than a decrease as was estimated above. Relatedly, the simulations imply a large increase in the natural log of the sons' 1870 wealth, while we observed essentially no change using the lottery-based estimates. While these latter two outcomes are weighted towards changes in the lower tail of the sons' wealth distribution, we also examine the level of wealth in row-group H. In each simulation for 1870 wealth levels, the estimated and simulated confidence intervals have substantial overlap. These results taken together indicate the strongest effect of winning the lottery on the low end of the sons' wealth distribution. We might expect this pattern of results on *a priori* grounds as well in that high-wealth families were presumably less likely to be liquidity constrained.

Appendix E: Detailed Discussion of Lottery Effects in Model of Schooling

Here we present a standard (convex) model of schooling, augmented with a simple constraint, to illustrate our interpretation of the effects of winning the lottery. This model includes both the benefits of school and the opportunity cost of child labor, either on or off the family farm.

Our preferred interpretation is that the wealth shock relaxes the constraint on school investment, if indeed such a constraint had bound in the first place. Below, we also consider what happens if lottery winning directly affects the costs and benefits of schooling. To preview our findings, note that such effects are only operative if time in school is a decision made on the margin (at the first-order condition, in other words). In contrast, the point of the paper is to test whether a borrowing constraint (because of lack of current wealth) rations the school decision to below the level consistent with the FOC. Finally, we present cases from a non-convex model of schooling that exhibits a poverty trap. We also pay particular attention to Moav's 2005 model in which parents' human capital affects the quantity/quality tradeoff.

We start with graphical analyses of a standard schooling model in Appendix Figure E-1, which features the optimal choice of schooling (e*) implied by a model with the standard curvature in production and cost functions. This model generates a declining marginal benefit (MB) of time in school. It also implies a rising marginal cost (MC). The MC curve consists of two components: the direct costs of school, including school tuition and transport, and the opportunity costs of school, such as from foregone earnings from child labor or from the implicit value of the child's contribution to home production. Opportunity costs rise with time in school both because the students are learning something of value and because of the direct maturational benefit of aging.

If the initial marginal benefits are large enough (MB>MC at e=0), there is an interior solution denoted as e*. In our sample, the average school-enrollment rate for children aged 5-17 is more than a third, and this rate peaks around half among children 11 years of age. The often noncontinuous path through school in that era means that this number is a lower bound on the fraction of children attending at least some school. Relatedly, Bleakley and Hong (2013) report that white children in the Antebellum South achieved literacy rates of approximately 85%. Therefore, the assumption of an interior solution seems appropriate for the majority, likely the vast majority, of cases.

Now imagine a borrowing constraint that low-wealth households might confront. Two constraints on the choice of education are seen in online Appendix Figure E-1 (and repeated in other figures in this appendix). Investment in schooling seems like a clear case in which family wealth might matter. The costs (both direct and opportunity) of attending school are immediate while the benefits pay off over the course of that child's working lifetime. If credit markets are dysfunctional (as they most assuredly were in antebellum Georgia), families would have to depend largely on their own resources to finance investments in their children. For simplicity, we represent the borrowing constraint in this graph as simple upper bound on time in school. In a more formal model, it would be an endogenous function of not just parental wealth, but also the intertemporal elasticity of substitution, the gap between MB and MC, and the degree of intergenerational altruism, the nature of permissible intergenerational contracts, inter alia.

The borrowing constraint rations the choice of education below e*. This results in a deadweight loss equivalent to the triangle between MB and MC to the right of the constraint. A positive wealth shock can relax this constraint (moving from A to B, *e.g.*) and reduce the size of the triangle. This is the case that we have in mind when testing the effect of parental wealth on child investment and child outcomes. (Recall that we show a strong and smooth upward-sloping gradient for school attendance with respect to parental wealth in online Appendix Figure 1.)

Next, in online Appendix Figure E-2, we take this model and examine its implications under a series of assumptions about how prices and constraints respond to wealth shocks. Case 2A displays the baseline model. For the remainder of the figure, let us consider how the unconstrained choice might respond to the increase in wealth. For example, if the family owns a larger farm as a result of winning the lottery, this might increase the demand for labor by the children (perhaps in a more pronounced way for the sons). Of course, even if they had a smaller farm or no farm at all, the income from the labor of their children would have expanded the household budget set. Nevertheless, a larger farm most likely was a more complicated enterprise requiring greater coordination, which might make the labor input of one's children more strongly complementary with the labor input of the parents. This would shift up the marginal product of labor of the children, and therefore the marginal cost curve. Case B displays a shift that is neutral across skill: the upward-sloping red curve MC', which is MC plus a constant.

The assumption, however, that this complementary-labor effect is neutral across the skill distribution seems restrictive. Why would this be the case? Indeed, if the need for parents and children to work together is driven by administrative complication and agency issues associated with managing a larger enterprise, then the marginal benefit of skill might actually be higher. Bleakley and Hong (2013) make this observation in regard to the post-Bellum decline in white school attendance in the South. They argue that more management skills are required in a largely self-sufficient plantation than in a similarly-sized farm that is sharecropped (the former being more prevalent antebellum and the latter being more prevalent after the war). Olmsted and Rhode (2008) detail the adoption of new crops and the structural change in agriculture that occurs at various points in the region in response to technological change. While it is tempting to highlight the role a "vital few" had in the technological change that transformed the region's agriculture, ultimately the adoption of new methods depended on the ability of farmers to understand and implement them. (Schultz (1975) makes essentially the same point about the role of skill in, for example, The Green Revolution.) If indeed this does raise the return to school, we would expect the marginal benefits of time in school to shift up, as seen in the downward-sloping red curve (MB'). Therefore, whether having a larger farm causes the optimal choice of education to rise or fall is ultimately ambiguous.

We add back into this graph a borrowing constraint that low-wealth households might confront. If the constraint causes a large loss (the scenario of interest for our paper), then the choice of education is rationed so far away from the FOC in either the baseline or the prime cases that changes in the unconstrained optimum are irrelevant. This is seen in in Case 2C. In Case 2D, we see that a positive wealth shock can relax this constraint and reduce the size of the triangle. This is the case that we have in mind when testing the effect of parental wealth on child investment and child outcomes. Note that, unless the shock is large enough to completely eliminate the triangle, the optimal choice of schooling is still not governed by the first-order condition.

The remaining two cases in online Appendix Figure E-2 are ones in which the shock of winning the lottery is so large that the subsequent education choice satisfies the FOC.

Case 2E represents a shock to the opportunity cost of the child's time that is very large. In this case, the marginal cost shifts to MC" which intersects either the marginal-benefits curve to the left of either the old or new borrowing constraints. In other words, if the marginal-cost curve rises enough because of the wealth shock, it can obscure the presence of a borrowing constraint by depressing the optimum down to what had been the constrained choice absent the wealth shock. A small triangle could thus be masked by a small decline in e*, or a large triangle could be masked by a large decline in e*. Recall that our estimates for the effect on school attendance were essentially zero, and thus we would require precisely this sort of coincidence for Case 2E to describe the results. Indeed, we would need a quadruple coincidence because we estimate the zero effects for boys or girls as well as for younger or older children. Further, we estimate similar effects for literate versus illiterate fathers, younger versus older fathers, and for families with different propensities (measured by the average behavior of others in the 1850 census with the same surname) for fertility, school, literacy, and wealth accumulation. Thus, this coincidence would have to be repeat across ten cuts of the data, almost all of which are themselves associated with different baseline levels of school enrollment. While we cannot rule out such a harmonic convergence using only the data on school attendance, we find it implausible that the relaxation of constraints and changes in e* would exactly offset for all of the subgroups.

In the final case (2F), the wealth shock is large enough to push the borrowing constraint above the unconstrained optimum. This eliminates the triangle of loss and leaves the education decision as determined on the margin by the FOCs. Opportunity-cost effects become relevant, but only after relaxing the borrowing constraint. It is clear that, in this case, low wealth is an impediment to educational investment, and time in school would rise even if the unconstrained optimum falls.

In this model and in similar (strongly) convex models, the analysis above holds for those households rationed below the optimal investment by the constraint. However, this might not be the case in certain nonconvex models. Examples of such models are presented by Galor & Zeira (1993) and Moav (2005). These models feature a poverty trap at zero investment in education, although the mechanisms introduced in each paper have some differences.

To fix ideas, we give a simplified representation of a poverty trap inducing a corner solution at zero in Appendix Figure E-3, Case 3A, which is a simple modification of the original, convex model. By some mechanism, the marginal returns to education are depressed at low levels of investment. (Often this mechanism is a minimum-input requirement or fixed cost, although that is not represented here.) There may be an unconstrained optimum e* as before. The borrowing constraints could ration the household away from this solution, as before. Unlike Appendix Figure E-1, however, this fails the "initial marginal benefits are large enough" condition referenced above, and now a corner solution is possible at an e** of zero. Shifts in the borrowing constraint do not affect the location of this corner solution, although, as the borrowing constraint is relaxed, the non-zero constrained choice is now more likely to be the global optimum.

One difficulty with this analysis is that antebellum Census data on school attendance and literacy indicate that the vast majority of households were not at a corner solution of zero. Thus, if the model in this case were correct, our interpretation of the estimates would not apply to the minority at zero, but our interpretation would apply to the majority off this corner, if a wealth constraint is present.

However, it might be argued that this implication of a corner solution is taking the model too literally. There may be some minimum level of education that is obtained at relatively low cost, but higher levels are more difficult to achieve. Thus, we present Case 3B in which the MB curve has a fillip near zero such that there are two interior, local maxima. This being the case, there could be a large mass at e** that would neither be (locally) influenced by the changing constraint nor at the corner solution of zero. (But they might still be influenced by relaxing the constraint, if the constrained value becomes the global optimum.)

Could the lottery winnings be enough to cause a jump from the low to the high equilibrium in case 3B? Based on the magnitude of the winnings and the costs of school attendance, we argue that this is quantitatively possible, at least for some children. This issue arises if (a) the low skill investment is a local maximum, (b) there is also a higher skill investment that is a global maximum, and (c) the global maximum is unreachable because of a credit constraint. This could be represented in something like Case 3B, if the skill investment at the constraint yields a lower return than the value at the low-skill intersection of MB and MC. Certainly this is possible for a marginal relaxation of the budget constraint, but it is probably not applicable in our case because the shock to wealth was not marginal. In the paper, we reference $18/month as an approximate wage for an unskilled adult male. In an economy in which brawn is important, this figure is probably an upper bound on the value of child labor. So, financing the foregone earnings for a few years of school that met for less than a four-month term (much of which is outside the harvest time) should cost less than $200, which is still smaller than our estimated value of winning the lottery.

The direct costs of schooling were only a fraction of the indirect cost of foregone child labor. The antebellum Georgia educational system consisted of a mix of (1) academies, generally used by the upper class, where fees were $3-5/quarter; and (2) local or county public schools with fees of no more than $3/quarter (some, like Bryan County's, entirely free), though these were only present in some counties. The state also paid tuition (for either academies or local or county public schools) from its "Poor School Fund," for students from any families paying less than the state's poll tax plus $0.50 (Kilpatrick 1921; Biehle, 1968). The poll tax was quite low in this period (generally under $0.40, and zero for a time in the mid-1830s; Wallenstein 1985), so the number of students who benefitted from these subsidies was quite small as well.

On a somewhat different track, the Moav (2005) model merits additional attention as a possible explanation for our results, and we will evaluate this now. In that model, the mechanism that generates a poverty trap at low levels of education is twofold. First, there is a quantity/quality trade-off. Second, high skilled parents have a comparative advantage in rearing high-skilled children. Wealth-constrained parents stuck in the low-skill trap might be so far away from education being a worthwhile investment that any marginal relaxation of their budget constraint goes to fertility rather than education (as in our results).

The relevance of this mechanism for our context is, we believe, testable. In the Moav model, wealth-constrained parents are more likely to be at the corner solution of zero education if the parents themselves had low education/skill. Parents with greater skill would choose to take some of this lottery wealth and put some of it in fertility and some other part in education. (The evidence discussed in our paper and in this appendix rejects the view of all parents being at the only-fertility/no-skill corner solution.)

This test starts with two ingredients. First, we need a measure that predicts parents' skill or otherwise predicts their skill or fertility investment in their children. The most obvious one in the 1850 census is whether the father can read or write. Approximately 15% of fathers are illiterate in our sample. We also construct a measure of the propensity to invest in quantity or quality by taking averages of other people in the South in 1850 Census with the same surname. (Anyone in our sample is excluded from the construction of these indices.) This tells us about common investment tendencies among people who are most likely cousins connected along some patriline. (The predictive power of surnames is discussed in a recent literature, including Gregory Clark's recent book*, The Son Also Rises*.) The second ingredient is to take these measures and interact them with lottery winnings. If the Moav mechanism is operative, there should be different responses to the wealth shock among people with different propensities to be on or off the only-fertility/zero-skill corner.

Results from this exercise indicate very little interaction between parental skill and the investment choice, contrary to the hypothesis just presented. These results are seen in online Appendix Table E-1. Each panel displays a different dependent variable: fertility in the top panel and school enrollment/attendance in the bottom panel. The first column replicates results from the manuscript, and the first row of each panel shows the coefficient on winning the lottery (as measured by unique match to the Smith list). The second row of each panel presents the coefficient on the additional variable that is intended to predict skill and/or fertility investments.

The even-numbered columns display the coefficient on this indicator variable in a regression of fertility or school attendance on the indicator. Column 2, for example, shows that literate fathers have 0.168 fewer children and those children, if of school age, are 13.5 percent more likely to be attending school. In Column 3, we present a model that includes father's literacy, a dummy for lottery winning, and the interaction of the two. Both of the main effects have coefficients similar to those found in other specifications. However, the interaction between these two is not statistically significant. Thus, there is no statistical evidence that the wealth effect is different for literate versus illiterate fathers in their quantity or quality decisions. The remaining columns consider the surname-averaged variables as indicators of fertility in school investments. The average literacy rate of others with the same surname also predicts lower fertility and higher school investments. (These coefficients are actually higher than those from the own-father model. This suggests some issue of measurement error or other missing variable problem. It is beyond the scope of this appendix to sort this out. Suffice it to say that these are useful proxies for the family's latent propensity to invest in quantity versus quality.) Again, however, the interaction of this characteristic with lottery winning suggests there is no difference in the wealth effect. This is inconsistent with low-skill families having a different Engel curves for fertility or school investment from high-skill families. The final measures that we consider are the average school-enrollment and fertility seen for that surname. Both measures are strongly predictive of their counterpart at the micro level. (That is, surname-averaged school predicts own-children school attendance and surname-average fertility predicts own fertility.) Nevertheless, there is no statistically significant interaction of these variables with winning the lottery.

These results, taken together with the apparently vast majority of cases being away from a corner solution with respect to child quality, suggest that the interpretation of our estimates in the manuscript is not complicated by an important heterogeneity of responses by parental differences in the propensity to quantity or quality investments.

In any case, note that the claim for the parental-skill interaction is entirely compatible with our description of alternative mechanisms by which families differ in their skill investments. Parents' skill would be just another difference in Becker's concept of 'infrastructure' that we reference in the paper.

Appendix E References:

Biehle, Reba Strickland. 1968. Edward Oxford Pioneer Farmer of Middle Georgia. *Georgia Historical Quarterly,* 187-198.

Bleakley, Hoyt and Sok Chul Hong. 2013. When the Race between Education and Technology Goes Backwards: The Postbellum Decline of White School Attendance in the Southern U.S. Unpublished manuscript, University of Chicago. April.

Clark, Gregory. 2014. *The Son Also Rises: Surnames and the History of Social Mobility: Surnames and the History of Social Mobility*. Princeton University Press.

Galor, Oded and Jospeh Zeira. 1993. Income Distribution and Macroeconomics. *Review of Economic Studies*, Vol. 60(Jan.): 35-52.

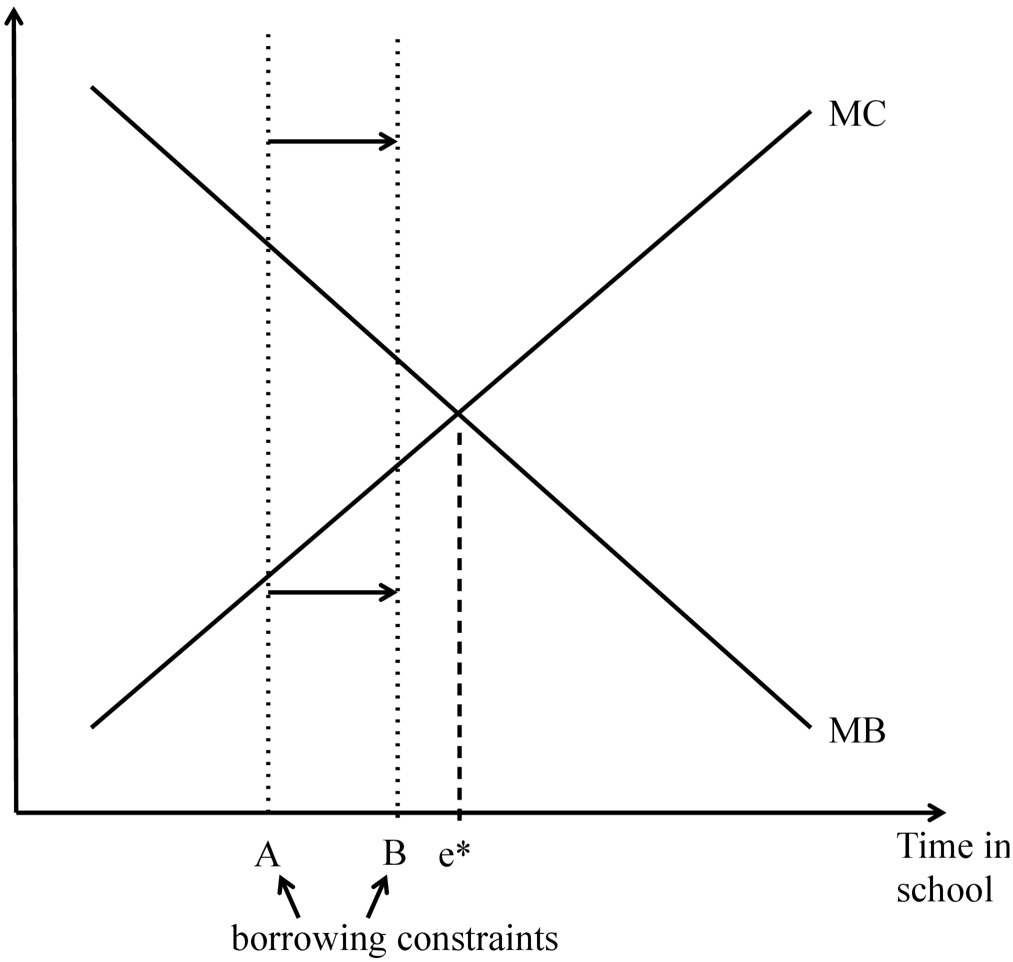
Kilpatrick, William H. 1921. The Beginnings of the Public School System in Georgia. *Georgia Historical Quarterly,* 3-19.

Moav, Omer. 2005. Cheap Children and the Persistence of Poverty. *The Economic Journal*, Vol. 115 (January), 88-110.

Olmstead, Alan L., and Paul W. Rhode. 2008. *Creating Abundance*. Cambridge Books.

Schultz, Theodore W. 1975. The value of the ability to deal with disequilibria. *Journal of economic literature,* 827-846.

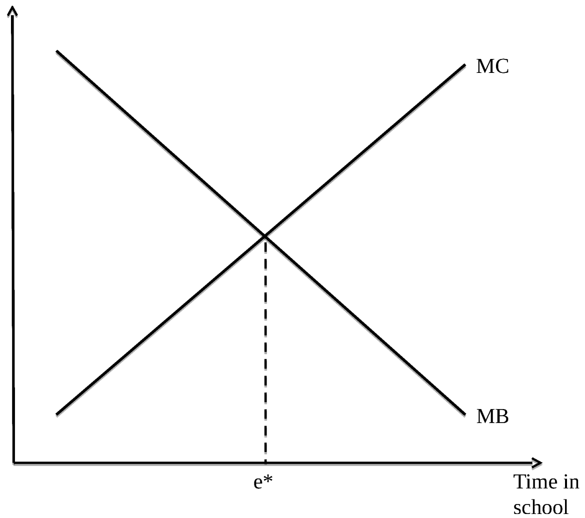
Appendix Figure E-1: Standard Model of Education, Augmented with Credit Constraint



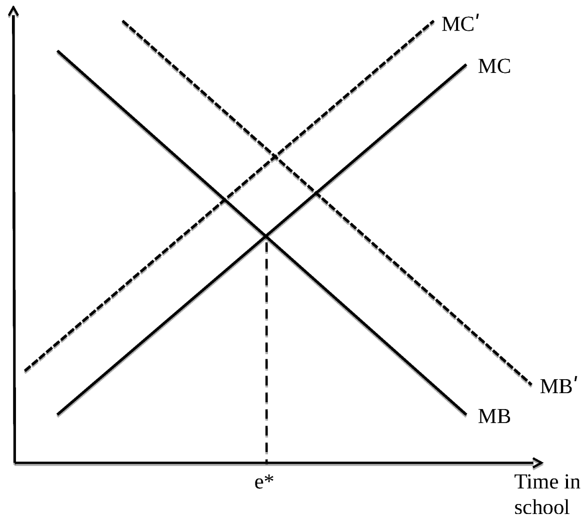
Notes: This figure displays the marginal benefit (MB) and marginal cost (MC) curves for the standard model of the optimal choice of schooling. The vertical, dotted lines (A and B) are various constraints on time in school. Winning the lottery should relax such constraints (causing movement from A to B, e.g.).

Appendix Figure E-2: Various Cases in Which Lottery Winning Affects Constraint on and Benefits and Costs of Education

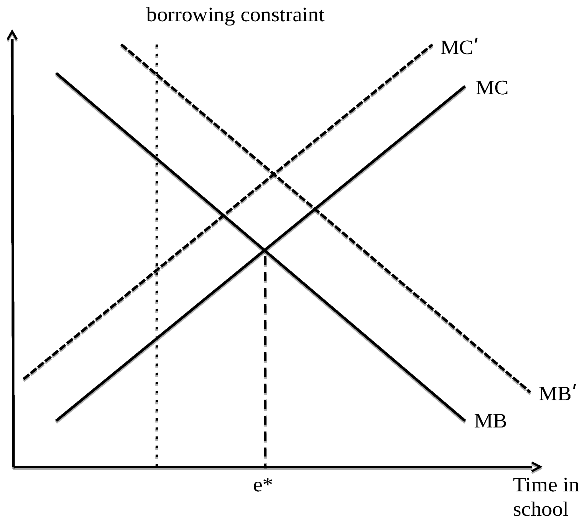
Case 2A: Unconstrained optimal choice (baseline)



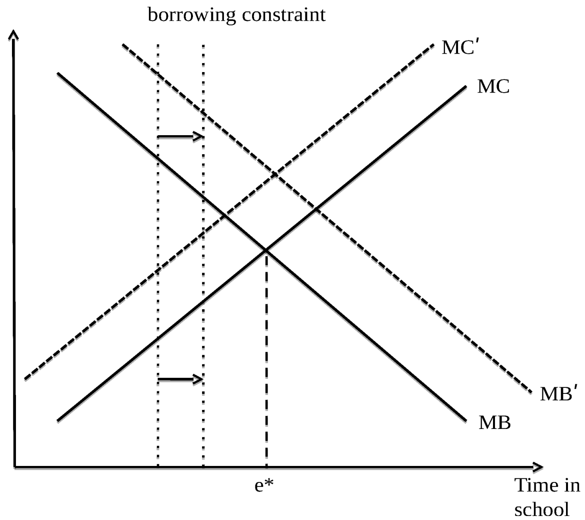
Case 2B: Possible shifts of MB and/MC curves for larger farm



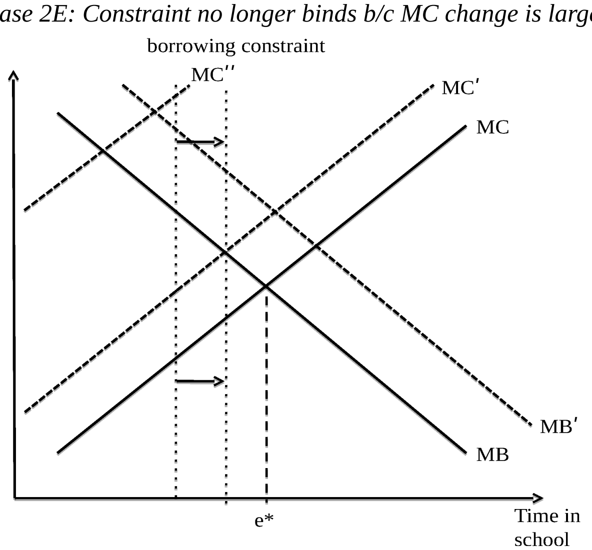
Case 2C: Constrained educational choice



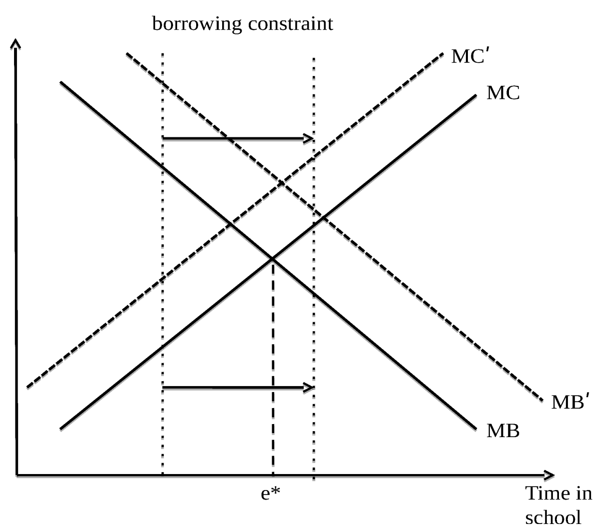
Case 2D: Wealth shock relaxes constraint



Case 2E: Constraint no longer binds b/c MC change is large



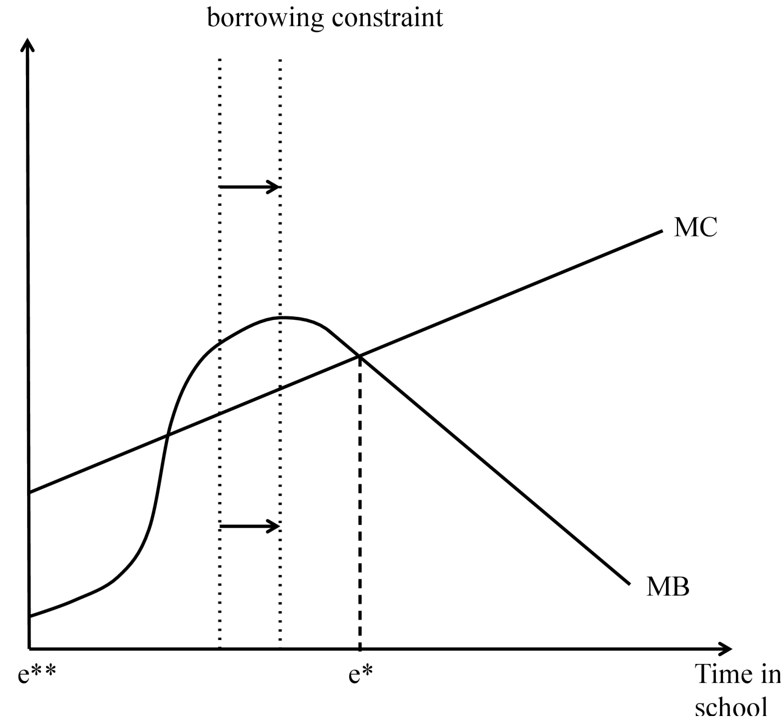
Case 2F: Constraint no longer binds b/c wealth shock is large



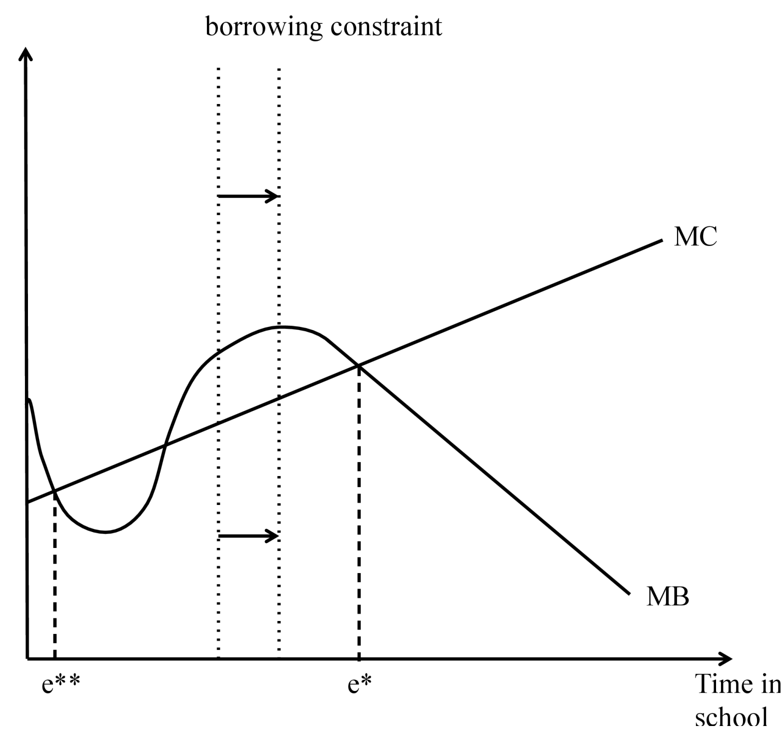
Notes: This figure displays the marginal benefit (MB) and marginal cost (MC) curves for the standard model of the optimal choice of schooling. Panels show various assumptions about how the lottery winnings affect the borrowing constraint and the MB and MC curves.

Appendix Figure E-3: Models of Education with Poverty Traps

Case 3A: Model with poverty trap at zero education



Case 3B: Model with poverty trap above zero



Notes: This figure displays the marginal benefit (MB) and marginal cost (MC) curves for models of the optimal choice of schooling as well as a possible borrowing constraint and its relaxation. Panels show various assumptions about the shape of the MB curve.

Appendix Table E-1: Interaction with Father's Illiteracy and Surname Averages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Surname-averaged variable					
	Baseline	Father's literacy		Literacy rate		In school, ages 5-15		Fertility	
Panel A: Dependent variable is number of children born post-1832 in household									
Dummy for unique match to Smith (1838) list	0.130		0.131		0.126		0.127		0.130
	(0.060) **		(0.067) **		(0.060) **		(0.060) **		(0.060) **
Additional variable		-0.168	-0.172	-1.294	-1.363	-0.198	-0.295	0.164 ***	0.175 ***
		(0.056) ***	(0.059) ***	(0.443) ***	(0.470) ***	(0.369)	(0.390)	(0.061)	(0.063)
Interaction term (using z score for surname variables)			0.025		0.033		0.044		-0.063
			(0.164)		(0.060)		(0.055)		(0.075)
Number of observations	14248	14248	14248	14024	14024	14144	14144	14248	14248
Panel B: Dependent variable is indicator for school enrollment among children aged 5-17									
Dummy for unique match to Smith (1838) list	-0.002		0.003		-0.002		-0.002		-0.002
	(0.011)		(0.012)		(0.011)		(0.011)		(0.011)
Additional variable		0.135	0.135	0.380	0.391	0.259	0.276	0.005	0.003
		(0.010) ***	(0.010) ***	(0.086) ***	(0.089) ***	(0.073) ***	(0.077) ***	(0.012)	(0.013)
Interaction term (using z score for surname variables)			-0.002		-0.005		-0.008		0.005
			(0.030)		(0.012)		(0.011)		(0.014)
Number of observations	47354	47278	47278	46673	46673	46305	46305	47050	47050

Notes: This table displays OLS estimates of equation (1) in the text, with certain exceptions. This table departs from previous ones in the use of surname-specific characteristics to proxy for differences across extended families ("dynasties") in child and wealth outcomes. We use the 1850 100% census file to construct average fertility, school attendance, and real-estate wealth among Georgia-resident households for each (soundex) surname. (Those individuals that appear in our lottery-eligible sample are excluded from the construction of these indices.) Each panel/column presents results from a separate regression. In addition to the displayed coefficients, regressions include dummies for age and for (state x county) of residence. The base sample for these regressions consists of all households in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. Households without a corresponding surname in the database of surname averages are excluded from the regressions. The dependent variables are indicated in the panel labels. A household is coded as a lottery winner if the head is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors (shown in parentheses) are heteroskedasticity robust and clustered on the surname level to account for correlation induced by the surname-averages. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Appendix F. Taxed vs. Census-Reported Wealth in Georgia

A question that naturally arises in using wealth reported in the census manuscripts as an outcome is the accuracy of that wealth measure. Steckel (1994) compares census manuscripts and property tax records for Ohio and Massachusetts and finds that census-reported wealth in 1850, 1860, and 1870 is generally lower than wealth reported in local tax records, with a closer correspondence in Massachusetts than Ohio. He also finds that household characteristics are for the most part uncorrelated with the gap between census wealth and taxed wealth.

To determine this correspondence in Georgia, the first 100 observations in the 1870 tax records of Thomas County (Ancestry.com, 2011) were examined, along with the wealth reported by the same individuals in the 1870 U.S. Census (Ancestry.com, 2009). Thomas County was chosen because its available tax records were recorded in the same year as the 1870 census. Georgia moved to an ad valorem tax system in 1852. (Wallenstein, 1985) This system provided a standard exemption of $200, and also exempted the first $300 of furniture, and did not tax crops or provisions. All other possessions were valued and taxed at a standard rate ($1.00 per $1,000 valuation).

The scatter plot in Appendix Figure 4 shows a strong linear relationship between the two wealth measures. A linear regression with taxed wealth as the dependent variable yields a coefficient on census wealth of 1.102 (*s.e.*=0.062, $p<0.001$).

Appendix F References:

Ancestry.com. 2009. *1870 United States Federal Census* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2009. Images reproduced by FamilySearch. Original data: *1870 U.S. census, population schedules*. NARA microfilm publication M593, 1,761 rolls. Washington, D.C.: National Archives and Records Administration, n.d.

Ancestry.com. 2011. *Georgia, Property Tax Digests, 1793-1892* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2011. Original data: *Georgia Tax Digests* [1890]. 140 volumes. Morrow, Georgia: Georgia Archives.

Steckel, Richard H. 1994. "Census manuscript schedules matched with property tax lists: a source of information on long-term trends in wealth inequality." *Historical Methods*, Vol 27, No. 2, pp. 71-86.

Wallenstein, Peter. 1985. 'More Unequally Taxed than any People in the Civilized World': The Origins of Georgia's Ad Valorem Tax System. *Georgia Historical Quarterly*, Vol. 69, No. 4 (Winter) , pp. 459-487.

Appendix G: Locational Choice as the Mechanism Generating the Results on Fertility & Schooling

Locational choice, at least at the county level, does not appear to be a central mechanism in driving these results. We further investigate this mechanism by examining characteristics of the 1850 county of residence in online Appendix Table 5. We begin by noting that lottery winners are slightly more likely to end up in Old Cherokee County in 1850 (Panel A, Column 1), although this difference in probabilities is quite small (2.2%). The lack of a homesteading requirement implies that there is no mechanical reason why the lottery winners should have higher rates of residence in Old Cherokee County than the lottery losers. Nonetheless, some of them may have chosen to settle on their parcel rather than flip it, and this decision apparently stuck for a small fraction. Winners were no more likely to be located in Georgia than losers (Column 2).

The remainder of Appendix Table 5 (Columns 3-17) uses county-level data (city-level or town-level data for "urban") to construct left-hand-side variables describing the local economic and demographic conditions in the 1850 place of residence. Because of the repeated data within counties, we now cluster the standard errors on county of residence. Most important for the quantity/quality results, we do not see major differences by lottery status in the average school-enrollment or fertility rates. This suggests that lottery winners were not differentially moving to areas that were more conducive to higher fertility or school attendance (the latter being perhaps because of the provision of school infrastructure). Additionally, being a lottery winner does not predict differences in county-of-residence farm values or farm sizes. While this might suggest that lottery winners bought more acreage instead of moving to counties with more valuable land, we cannot rule out that they bought the land that was more valuable within a county. We also find that winners were slightly less likely to reside in more densely settled places (Columns 12–15).[1] Additional analyses in Panel B adds surname fixed effects, Panel C adds a dummy variable for residence in Old Cherokee County, and Panel D controls for both surname fixed effects and residence in Old Cherokee County. The results are similar to Panel A throughout.

Appendix G References:

Fishman, Michael J. 1991. Population of Counties, Towns, and Cities in the United States, 1850 and 1860. ICPSR09424-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], http://doi.org/10.3886/ICPSR09424.v2.

[1] "Urban" places were those that appeared in 1850 in the U.S. Census Office's list of "Cities and Town in 1850 & 1860." (Fishman, 1991)