TABLE I
Intergenerational Mobility Estimates at the National Level

Child's outcome	Parent's Income Def.	Sample: Core sample (1)	Male children (2)	Female children (3)	Married parents (4)	Single parents (5)	1980-1985 cohorts (6)	Fixed Age at Child Birth (7)
1. Log family income (excluding zeros)	Log family income	0.344 (0.0004)	0.349 (0.0006)	0.342 (0.0005)	0.303 (0.0005)	0.264 (0.0008)	0.316 (0.0003)	0.361 (0.0008)
2. Log family income (recoding zeros to $1)	Log family income	0.618 (0.0009)	0.697 (0.0013)	0.540 (0.0011)	0.509 (0.0011)	0.528 (0.0020)	0.580 (0.0006)	0.642 (0.0018)
3. Log family income (recoding zeros to $1000)	Log family income	0.413 (0.0004)	0.435 (0.0007)	0.392 (0.0006)	0.358 (0.0006)	0.322 (0.0009)	0.380 (0.0003)	0.434 (0.0009)
4. Family income rank	Family income rank	0.341 (0.0003)	0.336 (0.0004)	0.346 (0.0004)	0.289 (0.0004)	0.311 (0.0007)	0.323 (0.0002)	0.359 (0.0006)
5. Family income rank	Family income rank (1999-2003)	0.339 (0.0003)	0.333 (0.0004)	0.344 (0.0004)	0.287 (0.0004)	0.294 (0.0007)	0.323 (0.0002)	0.357 (0.0006)
6. Family income rank	Top par. income rank	0.312 (0.0003)	0.307 (0.0004)	0.317 (0.0004)	0.256 (0.0004)	0.253 (0.0006)	0.296 (0.0002)	0.327 (0.0006)
7. Individual income rank	Family income rank	0.287 (0.0003)	0.317 (0.0004)	0.257 (0.0004)	0.265 (0.0004)	0.279 (0.0007)	0.286 (0.0002)	0.292 (0.0006)
8. Individual earnings rank	Family income rank	0.282 (0.0003)	0.313 (0.0004)	0.249 (0.0004)	0.259 (0.0004)	0.272 (0.0007)	0.283 (0.0002)	0.287 (0.0006)
9. College attendance	Family income rank	0.675 (0.0005)	0.708 (0.0007)	0.644 (0.0007)	0.641 (0.0006)	0.663 (0.0013)	0.678 (0.0003)	0.661 (0.0010)
10. College quality rank (P75-P25 gradient)	Family income rank	0.191 (0.0010)	0.188 (0.0014)	0.195 (0.0015)	0.174 (0.0014)	0.172 (0.0020)	0.198 (0.0007)	0.189 (0.0022)
11. Teenage birth (females only)	Family income rank	-0.298 (0.0006)			-0.231 (0.0007)	-0.322 (0.0016)	-0.285 (0.0004)	-0.290 (0.0011)
Number of observations		9,867,736	4,935,804	4,931,066	6,854,588	3,013,148	20,520,588	2,250,380

Notes: Each cell in this table reports the coefficient from a univariate OLS regression of an outcome for children on a measure of their parents' incomes with standard errors in parentheses. All rows report estimates of slope coefficients from linear regressions of the child outcome on the parent income measure except row 10, in which we regress college quality rank on a quadratic in parent income rank (as in Figure IVa). In this row, we report the difference between the fitted values for children with parents at the 75th percentile and parents at the 25th percentile using the quadratic specification. Column 1 uses the core sample of children, which includes all current U.S. citizens with a valid SSN or ITIN who are (1) born in birth cohorts 1980-82, (2) for whom we are able to identify parents based on dependent claiming, and (3) whose mean parent income over the years 1996-2000 is strictly positive. Columns 2 and 3 limit the sample used in column 1 to males or females. Columns 4 and 5 limit the sample to children whose parents were married or unmarried in the year the child was linked to the parent. Column 6 uses all children in the 1980-85 birth cohorts. Column 7 restricts the core sample to children whose parents both fall within a 5 year window of median parent age at time of child birth (age 26-30 for fathers; 24-28 for mothers); we impose only one of these restrictions for single parents. Child family income is the mean of 2011-12 family income, while parent family income is the mean from 1996-2000. Parent top earner income is the mean income of the higher-earning spouse between 1999-2003 (when W-2 data are available). Child's individual income is the sum of W-2 wage earnings, UI benefits, and SSDI benefits, and half of any remaining income reported on the 1040 form. Individual earnings includes W-2 wage earnings, UI benefits, SSDI income, and self-employment income. College attendance is defined as ever attending college from age 18 to 21, where attending college is defined as presence of a 1098-T form. College quality rank is defined as the percentile rank of the college that the child attends at age 20 based on the mean earnings at age 31 of children who attended the same college (children who do not attend college are included in a separate “no college” group); see Section III.B for further details. Teenage birth is defined as having a child while between age 13 and 19. In Columns 1-5 and 7, income percentile ranks are constructed by ranking all children relative to others in their birth cohort based on the relevant income definition and ranking all parents relative to other parents in the core sample. Ranks are always defined on the full sample of all children; that is, they are not re-defined within the subsamples in Columns 2-5 or 7. In Column 6, parents are ranked relative to other parents with children in the 1980-85 birth cohorts. The number of observations corresponds to the specification in row 4. The number of observations is approximately 7% lower in row 1 because we exclude children with zero income. The number of observations is approximately 50% lower in row 11 because we restrict to the sample of female children. There are 866 children in the core sample with unknown sex, which is why the number of observations in the core sample is not equal to the sum of the observations in the male and female samples.

TABLE II
National Quintile Transition Matrix

		Parent Quintile				
		1	2	3	4	5
Child Quintile	1	33.7%	24.2%	17.8%	13.4%	10.9%
	2	28.0%	24.2%	19.8%	16.0%	11.9%
	3	18.4%	21.7%	22.1%	20.9%	17.0%
	4	12.3%	17.6%	22.0%	24.4%	23.6%
	5	7.5%	12.3%	18.3%	25.4%	36.5%

Notes. Each cell reports the percentage of children with family income in the quintile given by the row conditional on having parents with family income in the quintile given by the column for the 9,867,736 children in the core sample (1980-82 birth cohorts). See notes to Table I for income and sample definitions. See Online Appendix Table VI for an analogous transition matrix constructed using the 1980-85 cohorts.

TABLE III
Intergenerational Mobility in the 50 Largest Commuting Zones

Upward Mob. Rank	CZ Name	Population	Absolute Upward Mobility	P(Child in Q5 | Parent in Q1)	Pct. Above Poverty Line	Relative Mobility Rank-Rank Slope
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Salt Lake City, Utah	1,426,729	46.2	10.8	77.3	0.264
2	Pittsburgh, Pennsylvania	2,561,364	45.2	9.5	74.9	0.359
3	San Jose, California	2,393,183	44.7	12.9	73.5	0.235
4	Boston, Massachusetts	4,974,945	44.6	10.5	73.7	0.322
5	San Francisco, California	4,642,561	44.4	12.2	72.5	0.250
6	San Diego, California	2,813,833	44.3	10.4	74.3	0.237
7	Manchester, New Hampshire	1,193,391	44.2	10.0	75.0	0.296
8	Minneapolis, Minnesota	2,904,389	44.2	8.5	75.2	0.338
9	Newark, New Jersey	5,822,286	44.1	10.2	73.7	0.350
10	New York, New York	11,781,395	43.8	10.5	72.2	0.330
11	Los Angeles, California	16,393,360	43.4	9.6	73.8	0.231
12	Providence, Rhode Island	1,582,997	43.4	8.2	73.6	0.333
13	Washington DC	4,632,415	43.2	11.0	72.2	0.330
14	Seattle, Washington	3,775,744	43.2	10.9	72.0	0.273
15	Houston, Texas	4,504,013	42.8	9.3	74.7	0.325
16	Sacramento, California	2,570,609	42.7	9.7	71.3	0.257
17	Bridgeport, Connecticut	3,405,565	42.4	7.9	72.4	0.359
18	Fort Worth, Texas	1,804,370	42.3	9.1	73.6	0.320
19	Denver, Colorado	2,449,044	42.2	8.7	73.3	0.294
20	Buffalo, New York	2,369,699	42.0	6.7	73.1	0.368
21	Miami, Florida	3,955,969	41.5	7.3	76.3	0.267
22	Fresno, California	1,419,998	41.3	7.5	71.3	0.295
23	Portland, Oregon	1,842,889	41.3	9.3	70.5	0.277
24	San Antonio, Texas	1,724,863	41.1	6.4	74.3	0.320
25	Philadelphia, Pennsylvania	5,602,247	40.8	7.4	69.6	0.393
26	Austin, Texas	1,298,076	40.4	6.9	71.9	0.323
27	Dallas, Texas	3,405,666	40.4	7.1	72.6	0.347
28	Phoenix, Arizona	3,303,211	40.3	7.5	70.6	0.294
29	Grand Rapids, Michigan	1,286,045	40.1	6.4	71.3	0.378
30	Kansas City, Missouri	1,762,873	40.1	7.0	70.4	0.365
31	Las Vegas, Nevada	1,568,418	40.0	8.0	71.1	0.259
32	Chicago, Illinois	8,183,799	39.4	6.5	70.8	0.393
33	Milwaukee, Wisconsin	1,660,659	39.3	4.5	70.3	0.424
34	Tampa, Florida	2,395,997	39.1	6.0	71.3	0.335
35	Orlando, Florida	1,697,906	39.1	5.8	71.5	0.326
36	Port St. Lucie, Florida	1,533,306	39.0	6.2	71.2	0.303
37	Baltimore, Maryland	2,512,431	38.8	6.4	67.7	0.412
38	St. Louis, Missouri	2,325,609	38.4	5.1	69.0	0.413
39	Dayton, Ohio	1,179,009	38.3	4.9	68.2	0.397
40	Cleveland, Ohio	2,661,167	38.2	5.1	68.7	0.405
41	Nashville, Tennessee	1,246,338	38.2	5.7	67.9	0.357
42	New Orleans, Louisiana	1,381,652	38.2	5.1	69.5	0.397
43	Cincinnati, Ohio	1,954,800	37.9	5.1	66.4	0.429
44	Columbus, Ohio	1,663,807	37.7	4.9	67.1	0.406
45	Jacksonville, Florida	1,176,696	37.5	4.9	68.9	0.361
46	Detroit, Michigan	5,327,827	37.3	5.5	68.5	0.358
47	Indianapolis, Indiana	1,507,346	37.2	4.9	67.5	0.398
48	Raleigh, North Carolina	1,412,127	36.9	5.0	67.3	0.389
49	Atlanta, Georgia	3,798,017	36.0	4.5	69.4	0.366
50	Charlotte, North Carolina	1,423,942	35.8	4.4	67.0	0.397

Notes: This table reports estimates of intergenerational mobility for the 50 largest commuting zones (CZs) according to their populations in the 2000 Census. The CZs are sorted in descending order by absolute upward mobility (Column 4). The mobility measures are calculated using the core sample (1980-82 birth cohorts) and the baseline family income definitions described in Table I (except for Column 5, which uses the 1980-85 birth cohorts). The measures in Columns 4 and 7 are both derived from within-CZ OLS regressions of child income rank against parent income rank. Column 7 reports the slope coefficient from this regression, which is equal to the difference in mean child income rank between children with parents in the 100th percentile and children with parents in the 0th percentile (divided by 100). Column 4 reports the predicted value at parent income rank equal to 25. Column 5 reports the percentage of children whose family income is in the top quintile of the national distribution of child family income conditional on having parent family income in the bottom quintile of the parental national income distribution. These probabilities are taken directly from Online Data Table VII. Column 6 reports the fitted values at parent rank 25 from a regression of an indicator for child family income being above the poverty line on parent income rank (see Appendix F for details). . See Online Data Table V for estimates for all CZs as well as estimates using alternative samples and income definitions.

TABLE IV
Segregation and Intergenerational Mobility

Dep. Var.:	Absolute Upward Mobility						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Racial Segregation	-0.361	-0.360					
	(0.045)	(0.068)					
Income Segregation			-0.393				-0.058
			(0.065)				(0.090)
Segregation of Poverty (<p25)				-0.508	-0.408		
				(0.155)	(0.166)		
Segregation of Affluence (>p75)				0.108	0.216		
				(0.140)	(0.171)		
Share with Commute < 15 Mins						0.605	0.571
						(0.126)	(0.165)
Urban Areas Only		x			x		
R-Squared	0.131	0.130	0.154	0.167	0.052	0.366	0.368
Observations	709	325	709	709	325	709	709

Notes: Each column reports coefficients from an OLS regression with standard errors clustered at the state level reported in parentheses. All independent and dependent variables are normalized (in the relevant estimation sample) to have mean 0 and standard deviation 1, so univariate regression coefficients equal correlation coefficients. The regressions are run using data for the 709 CZs with at least 250 children in the core sample. The dependent variable in all columns is our baseline measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile. Column 2 and 5 restrict to the sample of CZs that intersect an MSA. Racial segregation is measured by the Theil index defined in Section VI.B using racial shares at the census tract level. Income segregation is measured by a weighted average of two-group Theil indices, as in Reardon (2011). Segregation of poverty is a two-group Theil index, where the groups are defined as being above vs. below the 25th percentile of the local household income distribution. Segregation of affluence is defined analogously at the 75th percentile. Share with commute <15 minutes is the fraction of working individuals in each CZ who commute less than 15 minutes to work. See Appendix G for details on the definitions of the independent variables.

TABLE V
Income Inequality and Intergenerational Mobility: The "Great Gatsby" Curve

Dep. Var.:	Across CZs within the U.S.					Across Countries		
	Absolute Upward Mobility				Relative Mobility	Log-Log Elasticity 1985 Inequality		Log-Log Elasticity 2005 Inequality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gini Coefficient	-0.578							
	(0.093)							
Gini Bottom 99%		-0.634	-0.624		0.476	0.72	0.62	0.78
		(0.090)	(0.113)		(0.088)	(0.21)	(0.27)	(0.27)
Top 1% Income Share		-0.123	0.029		-0.032		0.17	-0.11
		(0.035)	(0.039)		(0.032)		(0.27)	(0.28)
Frac. Between p25 and p75				0.679				
				(0.111)				
Urban Areas Only			x					
R-Squared	0.334	0.433	0.380	0.462	0.224	0.518	0.536	0.531
Observations	709	709	325	709	709	13	13	12

Notes: Each column reports regression coefficients from an OLS regression with all variables normalized to have mean 0 and standard deviation 1 in the estimation sample, so univariate regression coefficients are equal to correlation coefficients. Columns 1-5 are estimated using data for the 709 CZs with at least 250 children in the core sample. The dependent variable in Columns 1-4 is our baseline CZ-level measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile in the core sample. In Column 5, the dependent variable is relative mobility, the rank-rank slope within each CZ. In Column 3, we restrict to CZs that intersect MSAs. In Columns 1-5, the Gini coefficient is defined as the Gini coefficient of family income for parents in the core sample in each CZ; the top 1% income share is defined as the fraction of total parent family income in each CZ accruing to the richest 1% of parents in that CZ; the Gini Bottom 99% is defined as the Gini coefficient minus the top 1% income share; and the fraction between p25 and p75 is the fraction of parents in each CZ whose family income is between the 25th and 75th percentile of the national distribution of parent family income for those in the core sample. In Columns 6-8, the dependent variable is the log-log IGE estimate by country from Corak (2013, Figure 1). The Gini coefficients across countries are obtained from the OECD Income Distribution Database (series "Income Distribution and Poverty: by country"). We interpret these coefficients as applying to the bottom 99% because the surveys on which they are based are typically top-coded. The top 1% income share across countries is from the World Top Income Database (series "Top 1% Income Share"). The independent variables are measured in 1985 in Columns 6 and 7 and in 2005 in Column 8.

TABLE VI
Correlates of Intergenerational Mobility: Comparing Alternative Hypotheses

Dep. Var.:	Absolute Upward Mobility			Relative Mobility		Absolute Upward Mobility		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Short Commute	0.302	0.227	0.314	-0.290	-0.325	0.331		0.319
	(0.065)	(0.077)	(0.052)	(0.061)	(0.064)	(0.070)		(0.065)
Gini Bottom 99%	-0.009	-0.017	0.060	0.006	0.343	-0.287		-0.021
	(0.053)	(0.043)	(0.097)	(0.071)	(0.095)	(0.059)		(0.054)
High School Dropout Rate	-0.147	-0.120	-0.109	0.010	0.181	-0.288		-0.140
	(0.055)	(0.038)	(0.085)	(0.064)	(0.056)	(0.059)		(0.055)
Social Capital Index	0.169	0.065	0.173	0.154	0.154	0.168		0.168
	(0.047)	(0.050)	(0.060)	(0.060)	(0.070)	(0.059)		(0.045)
Fraction Single Mothers	-0.487	-0.477	-0.555	0.591			-0.808	-0.579
	(0.062)	(0.071)	(0.089)	(0.049)			(0.085)	(0.061)
Fraction Black							0.056	0.132
							(0.073)	(0.051)
State Fixed Effects		x						
Urban Areas Only			x					
R-Squared	0.757	0.859	0.671	0.48	0.324	0.651	0.584	0.763
Observations	709	709	325	709	709	709	709	709

Notes: Each column reports coefficients from an OLS regression with standard errors clustered at the state level reported in parentheses. The regressions are run using data for the 709 CZs with at least 250 children in the core sample. The dependent variable in Columns 1-3 and 6-8 is our baseline measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile. The dependent variable in columns 4 and 5 is relative mobility, the rank-rank slope within each CZ. All independent and dependent variables are normalized (in the relevant estimation sample) to have mean 0 and standard deviation 1. Column 1 reports unweighted estimates across all CZs. Column 2 includes state fixed effects. In Column 3, we restrict to CZs that intersect MSAs. Columns 4-8 replicate the unweighted specification in Column 1 with different dependent and independent variables. The fraction with short commutes is the share of workers that commute to work in less than 15 minutes calculated using data for the 2000 Census. Gini bottom 99% is the Gini coefficient minus the top 1% income share within each CZ, computed using the distribution of parent family income within each CZ for parents in the core sample. Income-residualized high school dropout rate is the residual from a regression of the fraction of children who drop out of high school in the CZ, estimated using data from the NCES Common Core of Data for the 2000-01 school year, on mean household income in 2000. Social capital index is the standardized index of social capital constructed by Rupasingha and Goetz (2008). Fraction single mothers is the fraction of children being raised by single mothers in each CZ. Fraction black is the number of people in the CZ who are black alone divided by the CZ population. We code the high school dropout rate as 0 for 116 CZs in which dropout rate data are missing for more than 25% of the districts in the CZ, and include an indicator for having a missing high school dropout rate. We do the same for 16 CZs with missing data on social capital. See Section VI, Online Data Table IX, and Online Appendix G for additional details on the definitions of each of these variables.

ONLINE APPENDIX TABLE I
Sample Sizes vs. Vital Statistics Counts by Birth Cohort

				Base national dataset	Base CZ-level dataset
	Size of Birth Cohort (in '000s)	Percentage in DM1 database, US citizens, alive	and matched to a parent	with positive parent income in 1996-2000	and with valid parental geo information
	(1)	(2)	(3)	(4)	(5)
1977	3,327	95.9%	55.0%		
1978	3,333	97.0%	72.4%		
1979	3,494	97.6%	80.9%		
1980	3,612	99.2%	85.6%	85.2%	84.4%
1981	3,629	104.6%	91.6%	91.1%	90.3%
1982	3,681	105.5%	93.8%	93.2%	92.4%
1983	3,639	105.4%	95.4%	94.7%	93.8%
1984	3,669	105.1%	96.7%	95.8%	94.9%
1985	3,761	104.8%	97.5%	96.4%	95.4%
1986	3,757	104.7%	98.0%	96.6%	95.6%
1987	3,809	104.7%	98.4%	96.8%	95.8%
1988	3,910	104.5%	98.5%	96.8%	95.7%
1989	4,041	105.0%	98.5%	96.7%	95.6%
1990	4,158	104.7%	98.6%	96.7%	95.6%
1991	4,111	104.5%	98.5%	96.6%	95.5%
1980-1991	45,776	104.4%	96.0%	94.8%	93.8%

Notes: Column 1 reports the size of each birth cohort from 1987-1991, based on data from vital statistics obtained from the US Statistical Abstract 2012, Table 78. The remaining columns report the number of individuals in the population tax data as a percentage of the total number in the birth cohort, imposing the additional restrictions listed in the header of each column. Column 2 reports the number of individuals born in each cohort who are in the DM1 tax database, are current US citizens, and are alive in 2013. This column can differ from the birth cohort due to immigration and naturalization, emigration, and deaths before 2012. The percentage of citizens in the DM1 data rises in 1981 because citizenship status is missing for some individuals born before 1981. Column 3 further requires the individuals to be matched to parents (i.e., claimed as children dependents on individual income tax returns by a person aged 15-40 at the time of the birth of the child) in 1996 or after. Column 4, which requires in addition that parents have positive mean income between 1996-2000, is our key sample of interest for all national level statistics. Column 5 further requires valid geographical information (ZIP code) for parents. Column 5 is our key sample of interest for all local area statistics. The core sample includes the 1980-2 cohorts. The extended sample includes the 1980-91 cohorts.

ONLINE APPENDIX TABLE II
SOI Sample Counts by Birth Cohort

Cohort	Number of Observations (1)	Number of Unique Children (2)
1971	4,383	4,383
1972	7,787	5,569
1973	10,831	6,154
1974	14,330	7,065
1975	17,736	8,207
1976	17,938	8,246
1977	18,459	8,156
1978	17,756	7,958
1979	18,375	7,614
1980	19,545	7,732
1981	19,916	8,155
1982	22,331	9,929
1983	24,599	10,927
1984	28,221	12,390
1985	31,711	13,476
1986	33,221	13,540
1987	35,382	14,234
1988	38,139	15,362
1989	42,450	18,162
1990	47,768	19,805
1991	52,821	21,231
Total	523,699	228,295

Notes: This table reports the sample size for the Statistics of Income stratified random sample by birth cohort. Column 1 reports the total number of observations per cohort. Column 2 reports the number of unique children per cohort. See Appendix A for details on the construction of the SOI sample.

ONLINE APPENDIX TABLE III
Summary Statistics for Core Sample: Children Born in 1980-82

Variable	Mean (1)	Std. Dev. (2)	Median (3)
Parents:			
Family Income (1996-2000 average)	87,219	353,430	60,129
Top Earner's Income (1999-2003 average)	68,854	830,487	48,134
Fraction Single Parents	30.6%	46.1%	
Fraction Female among Single Parents	72.0%	44.9%	
Father's Age at Child Birth	28.5	6.2	28
Mother's Age at Child Birth	26.1	5.2	26
Father's Age in 1996	43.5	6.3	43
Mother's Age in 1996	41.1	5.2	41
Children:			
Family Income (2011-12 average)	48,050	93,182	34,975
Fraction with Zero Family Income	6.1%	23.9%	
Individual Income	31,441	112,394	24,931
Individual Earnings	30,345	98,692	23,811
Fraction Female	50.0%	50.0%	
Fraction Single	44.3%	49.7%	
Attend College between 18-21	58.9%	49.2%	
Fraction of Females with Teen Birth	15.8%	36.5%	
Child's Age in 2011	30.0	0.8	30
Number of Children		9,867,736	

Notes: The table presents summary statistics for the core sample (1980-82 birth cohorts); see notes to Table I for further details on the definition of the core sample. Child income is mean income in 2011-12 (when the child is approximately 30 years old), while parent family income is mean income from 1996-2000. Family income is total pre-tax household income. Top earner's income is the income of the higher-earning parent from 1999-2003 (when W-2's are available). Parents' marital status is measured in the year the parent is matched to the child. Child's individual income is the sum of W-2 wage earnings, UI benefits, and SSDI benefits, and half of any remaining income reported on the 1040 form. Individual earnings includes W-2 wage earnings, UI benefits, SSDI income, and self-employment income. A child is defined as single if he/she does not file with a spouse in 2011 and 2012. College attendance is defined as ever attending college from age 18 to 21, where attending college is defined as presence of a 1098-T form. Teenage birth is defined (for females only) as having a child while being aged 19 or less. See Section III.B and Online Appendix A for additional details on sample and variable definitions. All dollar values are reported in 2012 dollars, deflated using the CPI-U.

ONLINE APPENDIX TABLE IV

Comparison of Administrative Tax Data to CPS and ACS Survey Datasets

	Tax Data Full Sample (1)	Tax Data Core Sample (2)	2011-2012 CPS (3)	2011-2012 ACS (4)	Tax Data Full Sample (5)	Tax Data Core Sample (6)	2011-2012 CPS (7)	2011-2012 ACS (8)
Income Distribution:	*Earned Family Income*				*Total Family Income*			
% Zero	9.74%	7.32%	9.23%	12.64%	8.54%	6.11%	5.44%	8.00%
% Negative	0.00%	0.00%	0.00%	0.00%	0.33%	0.34%	0.04%	0.05%
Mean	44,278	46,805	54,313	42,382	45,406	48,050	56,438	44,845
Std. Deviation	104,528	109,667	58,556	47,879	90,594	93,182	59,145	50,072
P10	63	1,624	1,307	0	521	2,810	6,431	1,500
P25	12,724	14,984	18,843	12,000	12,842	14,919	20,414	14,000
P50	32,165	34,737	40,829	31,642	32,273	34,975	42,768	33,000
P75	62,095	65,148	75,000	57,000	62,992	66,169	76,554	60,000
P90	96,995	99,911	115,000	91,865	98,802	101,770	118,050	96,243
Demographics:								
% Married	42.43%	44.31%	49.32%	46.17%				
% Female	50.03%	49.97%	50.43%	49.98%				
% Live in South	36.83%	37.94%	38.33%	37.56%				
% College	54.62%	58.93%	66.20%	61.34%				
Observations	11,262,459	9,867,736	14,246	190,561	11,262,459	9,867,736	14,246	194,501
Sum of Samp. Weights	11,262,459	9,867,736	10,845,147	11,043,039	11,262,459	9,867,736	10,845,147	11,043,039

Notes: Columns (1) and (5) include all individuals in the Data Master-1 file from the SSA who were born in 1980-1982, are current U.S. citizens, and lived through 2012. In Columns (2) and (6), we impose the additional restriction that an individual was claimed as a dependent on a tax return in the years 1996-2012 by parents with positive income as described in the text. CPS sample consists of civilian, non-institutionalized citizens age 29-31 in the 2011 wave and 30-32 in 2012 waves of the Current Population Survey. ACS sample consists of civilian, non-institutionalized citizens born between 1980-1982 in the 2011 and 2012 American Community Surveys. Earned income refers to wages and salary plus social security and unemployment insurance plus positive self-employment income, except for the ACS measure, which does not include unemployment insurance. IRS wages and salary income is defined as the amount of all wages, tips, and other compensation before any payroll deductions (total of all amounts reported on all Forms W-2, Box 1). IRS unemployment compensation is defined as the amount of Unemployment Compensation and Railroad Retirement Board payments prior to tax withholding as reported on Form 1099-G, Box 1. IRS social security income is defined as total Social Security Administration benefits, as reported on Form SSA-1099 (as well as any Railroad Retirement Board benefits paid, as reported on Form RRB-1099, Box 3). IRS self-employment income is defined as the profit reported on Form 1040 Schedule C. In the CPS, self-employment income is business income; in the ACS, it is both farm and non-farm business income. In the tax data, total income is the sum of Adjusted Gross Income, social security, and tax exempt interest. Total income in CPS and ACS is all reported income including negative business and investment income. All dollar amounts are in 2012 dollars. Married refers to filing of joint return in 2011-2012 period for the tax data, and self-report of currently married in CPS/ACS samples. College means attended a degree granting institution between the ages of 18 and 21 in the tax data and self-report of more than high school attainment in CPS/ACS samples. South refers to filing a federal tax return in (for tax data) or being surveyed in (for ACS/CPS) one of the following states: DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX. ACS and CPS moments computed using sampling weights (inverse probability of inclusion in sample). For the ACS and CPS, the sum of the sample weights is the average of the sum of the sample weights in 2011 and in 2012.

ONLINE APPENDIX TABLE V

Estimates of Intergenerational Mobility Using Surname Means vs. Individual Incomes

Name Freq. Restriction	Number of Children	Number of Names	Rank-Rank Slope		Log-Log IGE	
			Surname	Individual	Surname	Individual
	(1)	(2)	(3)	(4)	(5)	(6)
1. No restriction	4,843,629	395,439	0.39	0.30	0.42	0.33
2. < 25	1,135,624	375,753	0.30	0.27	0.33	0.30
3. < 50	1,437,280	384,576	0.31	0.27	0.34	0.30
4. < 100	1,784,635	389,611	0.33	0.28	0.36	0.31
5. > 100	3,053,494	5,773	0.46	0.31	0.50	0.33
6. > 1,000	1,650,583	546	0.41	0.31	0.43	0.34
7. > 10,000	390,187	22	0.41	0.33	0.45	0.35
8. > 20,000	202,734	7	0.75	0.34	0.81	0.36

Notes: This table compares estimates of rank-rank slopes and log-log IGEs based on individual-level data to estimates based on surname means, as in Clark (2014). In this table, we restrict the core sample to children who have the same surname (in 2012) as their parents. The first row uses all children who satisfy this restriction. Rows 2-4 limit the sample to rare surnames: those that occur less than 25 times, 50 times, and 100 times in the sample. Conversely, rows 5-8 limit the sample to common surnames: those that occur more than 100, 1000, 10,000, and 20,000 times. Column 1 shows the number of children in each subsample (i.e., the number of observations used to estimate the rank-rank slope). Column 2 shows the number of distinct surnames in each sample. We estimate the individual-level rank-rank slopes and log-log IGE's (Columns 4 and 6) using OLS regressions on the microdata as in Table I. In Columns 3 and 5, we estimate the rank-rank slopes and log-log IGE's using OLS regressions on a dataset collapsed to surname-level means, weighting by the number of observations for each name. See Appendix D for further details.

ONLINE APPENDIX TABLE VI
National Quintile Transition Matrix: 1980-85 Cohorts

		Parent Quintile				
		1st	2nd	3rd	4th	5th
Child Quintile	1	33.1%	24.1%	17.7%	13.5%	11.7%
	2	27.7%	24.0%	19.6%	16.1%	12.6%
	3	18.7%	21.6%	21.9%	20.7%	17.0%
	4	12.7%	17.7%	21.8%	24.1%	23.7%
	5	7.8%	12.6%	18.9%	25.6%	35.1%

Notes. Each cell reports the percentage of children with family income in the quintile given by the row conditional on having parents with family income in the quintile given by the column for children in the 1980-85 birth cohorts. See notes to Table I for income and sample definitions. See Table II for an analogous transition matrix constructed using the 1980-82 birth cohorts.

ONLINE APPENDIX TABLE VII

Robustness of Spatial Variation in Intergenerational Mobility to Alternative Specifications

Change from Baseline Specification	Correlation with Baseline Mobility Estimates and Ratio of Std. Dev.			
	Upward mobility Unweighted (1)	Relative mobility Unweighted (2)	Upward mobility Pop. Weighted (3)	Relative mobility Pop. Weighted (4)
A. Alternative Samples				
1. Male children	0.99, 1.07	0.94, 1.03	0.98, 1.07	0.98, 1.01
2. Female children	0.98, 0.96	0.95, 1.09	0.97, 0.98	0.98, 1.02
3. Children of married parents	0.97, 0.96	0.89, 0.92	0.91, 1.02	0.93, 0.89
4. Children of single parents	0.97, 0.97	0.61, 1.14	0.97, 0.96	0.83, 1.02
5. Birth cohorts 1983-85	0.97, 1.00	0.84, 1.08	0.96, 0.93	0.96, 1.05
6. Birth cohorts 1986-88	0.94, 0.95	0.73, 1.11	0.82, 0.91	0.88, 0.98
7. Parent age at child birth within 5 years of median	0.98, 1.06	0.90, 1.16	0.98, 1.05	0.96, 1.02
8. Children who stay within CZ	0.94, 1.02	0.87, 1.09	0.93, 1.12	0.95, 1.04
9. Children matched to unique parents	0.99, 0.93	0.98, 0.90	0.98, 0.93	0.99, 0.93
B. Alternative Income Definitions				
10. Top parent income	1.00, 1.05	0.97, 1.00	0.99, 1.04	0.99, 0.99
11. Individual child income	0.94, 0.74	0.89, 0.80	0.83, 0.97	0.95, 0.81
12. Individual child earnings	0.93, 0.72	0.86, 0.82	0.82, 0.93	0.93, 0.82
13. Individual child income (males only)	0.96, 1.03	0.90, 1.01	0.96, 1.14	0.95, 0.97
14. Indiv child income and top parent income (males only)	0.97, 1.07	0.87, 1.02	0.97, 1.15	0.94, 0.95
15. Parent Income 1999-2003	1.00, 0.99	0.98, 0.99	1.00, 0.98	1.00, 0.99
C. Adjustments for Cost of Living and Growth Rates				
16. Cost of living adjusted income	0.98, 1.06	0.99, 0.99	0.86, 1.01	0.99, 0.97
17. Parent income measured in 2011/12	0.97, 0.90	0.92, 0.89	0.94, 0.95	0.98, 0.93
18. Controlling for growth	0.83, 0.83	0.92, 0.92	0.81, 0.83	0.96, 0.97
D. Alternative Measures of Mobility				
19. Within-CZ ranks		0.95, 0.94		0.96, 0.98
20. Prob. Child in Q5 | Parent in Q1	0.91		0.92	
21. Child income > poverty line	0.94		0.89	
E. Alternative Child Outcomes				
22. College Attendance (age 18-21)	0.71	0.68	0.53	0.72
23. College Quality (age 20)	0.71	0.51	0.55	0.65
24. Teenage Birth, females only	-0.61	-0.58	-0.64	-0.68

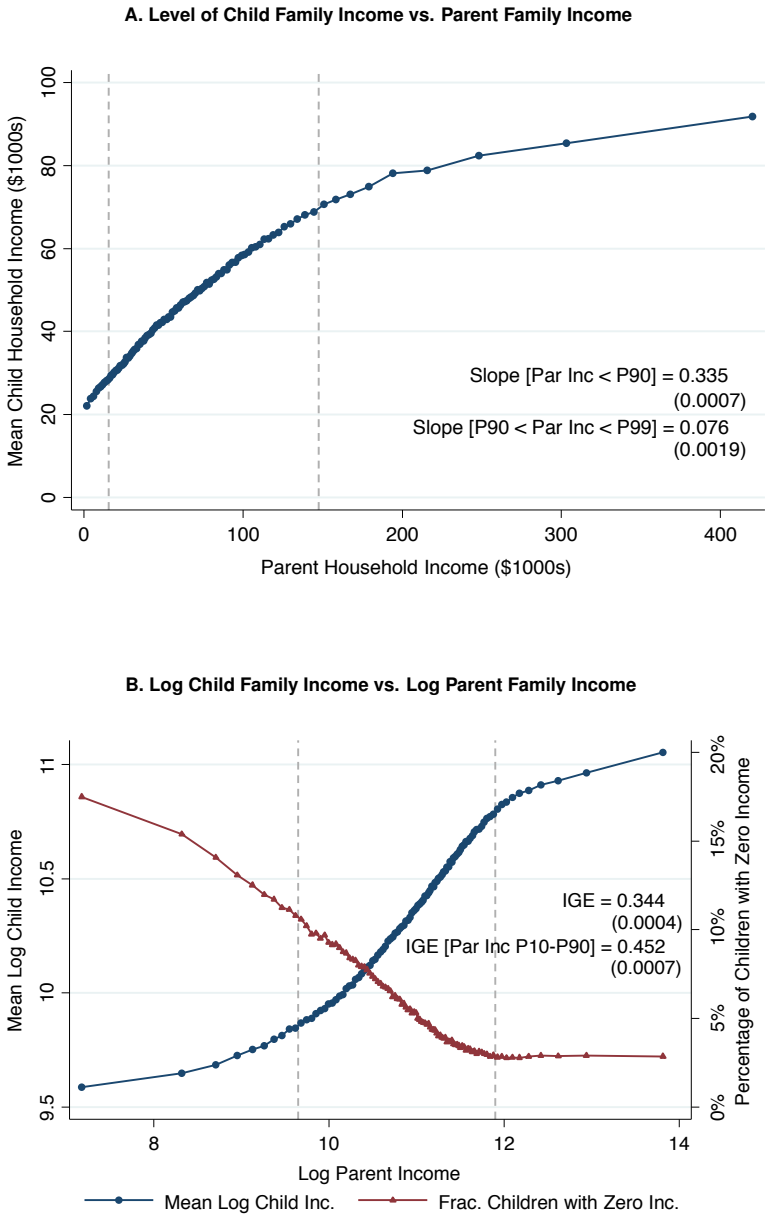
Notes: The first number in each cell of this table reports the correlation across CZs of a baseline mobility measure (using child family income rank and parent family income rank in the core sample) with an alternative mobility measure. The second number in each cell reports the ratio of the standard deviation of the alternative measure to the baseline measure. We do not report the ratio of standard deviations for statistics that are measured in different units relative to the corresponding baseline measure. The alternative mobility measures are defined either using a different sample (Panel A), a different income measure for parents or children (Panel B), adjusting for cost of living or local growth (Panel C), using alternative statistics for mobility (Panel D), or using earlier outcomes (Panel E). Column (1) reports the unweighted correlation (and SD ratio) between the alternative and baseline measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile in the core sample. Column (2) reports the unweighted correlation (and SD ratio) between the alternative and baseline measure of relative mobility, the slope of the rank-rank relationship in the core sample. Columns (3) and (4) repeat Columns (1) and (2), weighting the correlations and standard deviations by CZ population as recorded in the 2000 Census. With the exception of the transition probability in row 20, all absolute and relative mobility measures are constructed using OLS regressions of child outcomes on parent ranks as described in the text. Ranks are always defined in the full sample, prior to defining specific subsamples, except in row 19. See Appendix F for details on the definition of each measure.

ONLINE APPENDIX TABLE VIII
Correlates of Intergenerational Mobility Across Commuting Zones

	Dep. Var.:	Absolute Upward Mobility										Relative Mobility	
		Baseline		State FEs		Pop. Weighted		Urban Areas Only		Controls			
		(1)		(2)		(3)		(4)		(5)		(6)	
Race	Fraction Black Residents	-0.580	(0.066)	-0.353	(0.048)	-0.616	(0.074)	-0.673	(0.063)			0.631	(0.048)
Segregation	Racial Segregation Theil Index	-0.361	(0.045)	-0.274	(0.027)	-0.311	(0.092)	-0.360	(0.068)	-0.273	(0.046)	0.406	(0.048)
	Income Segregation Theil Index	-0.393	(0.065)	-0.260	(0.036)	-0.169	(0.105)	-0.184	(0.068)	-0.267	(0.054)	0.183	(0.063)
	Segregation of Poverty (<p25)	-0.407	(0.066)	-0.261	(0.038)	-0.216	(0.098)	-0.210	(0.066)	-0.274	(0.054)	0.218	(0.059)
	Segregation of Affluence (>p75)	-0.369	(0.064)	-0.250	(0.035)	-0.142	(0.106)	-0.155	(0.070)	-0.250	(0.052)	0.146	(0.063)
	Share with Commute < 15 Mins	0.605	(0.126)	0.342	(0.092)	0.335	(0.115)	0.548	(0.080)	0.415	(0.131)	-0.447	(0.074)
Income Distribution	Household Income per Capita for Working-Age Adults	0.050	(0.071)	-0.013	(0.075)	0.046	(0.092)	0.043	(0.076)	0.064	(0.080)	-0.145	(0.081)
	Gini coefficient for Parent Income	-0.578	(0.093)	-0.281	(0.050)	-0.236	(0.162)	-0.537	(0.120)	-0.362	(0.086)	0.346	(0.089)
	Top 1% Income Share for Parents	-0.190	(0.072)	-0.065	(0.031)	0.059	(0.094)	-0.144	(0.069)	-0.072	(0.065)	0.019	(0.063)
	Gini Bottom 99%	-0.647	(0.092)	-0.433	(0.063)	-0.416	(0.123)	-0.616	(0.114)	-0.470	(0.104)	0.473	(0.090)
	Fraction Middle Class (Between National p25 and p75)	0.679	(0.111)	0.500	(0.102)	0.293	(0.129)	0.551	(0.126)	0.458	(0.145)	-0.451	(0.109)
K-12 Education	School Expenditure per Student	0.246	(0.095)	0.026	(0.099)	0.219	(0.088)	0.236	(0.092)	0.053	(0.082)	-0.279	(0.092)
	Teacher-Student Ratio	-0.328	(0.100)	-0.213	(0.128)	0.062	(0.139)	0.024	(0.104)	-0.249	(0.088)	0.009	(0.108)
	Test Score Percentile (Controlling for Parent Income)	0.588	(0.087)	0.466	(0.074)	0.176	(0.220)	0.413	(0.147)	0.393	(0.093)	-0.317	(0.122)
	High School Dropout Rate (Controlling for Parent Income)	-0.574	(0.089)	-0.413	(0.060)	-0.433	(0.100)	-0.441	(0.108)	-0.440	(0.086)	0.328	(0.099)
Social Capital	Social Capital Index (Rupasingha and Goetz 2008)	0.641	(0.091)	0.349	(0.092)	0.299	(0.131)	0.517	(0.116)	0.478	(0.097)	-0.327	(0.085)
	Fraction Religious	0.521	(0.085)	0.357	(0.061)	0.410	(0.096)	0.417	(0.096)	0.484	(0.065)	-0.101	(0.090)
	Violent Crime Rate	-0.380	(0.146)	-0.163	(0.058)	-0.149	(0.166)	-0.367	(0.145)	-0.244	(0.062)	0.217	(0.140)
Family Structure	Fraction of Children with Single Mothers	-0.764	(0.074)	-0.571	(0.085)	-0.613	(0.129)	-0.719	(0.063)	-0.611	(0.066)	0.641	(0.046)
	Fraction of Adults Divorced	-0.486	(0.100)	-0.333	(0.085)	-0.389	(0.074)	-0.346	(0.103)	-0.569	(0.086)	0.158	(0.088)
	Fraction of Adults Married	0.571	(0.062)	0.417	(0.063)	0.221	(0.127)	0.377	(0.069)	0.365	(0.089)	-0.370	(0.078)
Tax	Local Tax Rate	0.325	(0.070)	0.135	(0.073)	0.155	(0.092)	0.182	(0.073)	0.207	(0.071)	-0.328	(0.061)
	Local Government Expenditures per Capita	0.186	(0.083)	0.074	(0.028)	0.192	(0.087)	0.085	(0.079)	0.107	(0.083)	-0.301	(0.080)
	State EITC Exposure	0.245	(0.064)			0.279	(0.076)	0.355	(0.073)	0.163	(0.073)	-0.144	(0.047)
	State Income Tax Progressivity	0.207	(0.146)			0.265	(0.070)	0.198	(0.098)	0.155	(0.133)	-0.150	(0.106)
College	Number of Colleges per Capita	0.200	(0.114)	-0.015	(0.118)	0.108	(0.088)	-0.045	(0.076)	0.060	(0.142)	-0.125	(0.052)
	Mean College Tuition	-0.018	(0.067)	-0.044	(0.039)	0.058	(0.096)	-0.015	(0.087)	-0.029	(0.066)	0.109	(0.064)
	College Graduation Rate (Controlling for Parent Income)	0.155	(0.062)	0.141	(0.052)	0.107	(0.089)	0.120	(0.095)	0.173	(0.073)	-0.025	(0.057)
Local Labor Market	Labor Force Participation Rate	0.212	(0.086)	-0.045	(0.052)	0.022	(0.090)	0.267	(0.113)	0.146	(0.073)	-0.237	(0.082)
	Fraction Working in Manufacturing	-0.261	(0.091)	0.007	(0.079)	-0.158	(0.090)	-0.128	(0.096)	0.002	(0.085)	0.393	(0.070)
	Growth in Chinese Imports 1990-2000 (Autor and Dorn 2013)	-0.175	(0.078)	0.006	(0.023)	0.001	(0.070)	0.008	(0.102)	-0.107	(0.048)	0.171	(0.083)
	Teenage (14-16) Labor Force Participation Rate	0.631	(0.087)	0.358	(0.098)	0.299	(0.153)	0.540	(0.109)	0.388	(0.090)	-0.516	(0.084)
Migration	Migration Inflow Rate	-0.258	(0.074)	-0.186	(0.049)	-0.146	(0.076)	-0.040	(0.078)	-0.285	(0.069)	-0.084	(0.067)
	Migration Outflow Rate	-0.163	(0.070)	-0.162	(0.048)	0.062	(0.094)	0.013	(0.076)	-0.145	(0.071)	-0.150	(0.070)
	Fraction of Foreign Born Residents	-0.027	(0.064)	-0.014	(0.039)	0.237	(0.083)	0.092	(0.064)	-0.004	(0.051)	-0.247	(0.055)

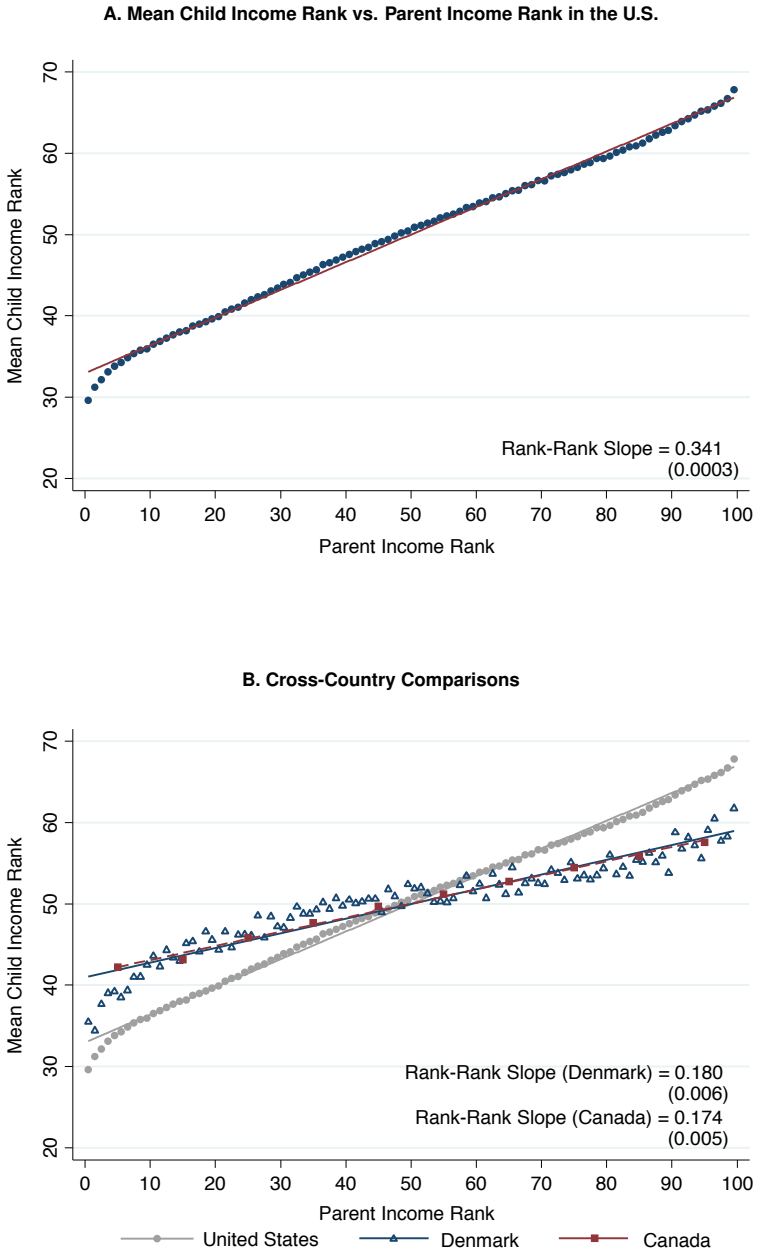
Notes: Each cell reports estimates from OLS regressions of a measure of mobility on the variable listed in each row, normalizing both the dependent and independent variables to have mean 0 and standard deviation 1 in the estimation sample, so that univariate regression coefficients equal correlation coefficients. Standard errors, reported in parentheses, are clustered at the state level. The dependent variable in Columns 1-5 is our baseline measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile. The dependent variable in Column 6 is relative mobility, the rank-rank slope in each CZ. All mobility estimates are constructed using the core sample (1980-82 cohorts) and baseline family income measures. Column 1 reports estimates from univariate unweighted regressions (raw correlation coefficients). Column 2 adds state fixed effects. Column 3 weights by Census 2000 population (and normalizes variables by weighted standard deviations). In Column 4, we restrict to CZs that intersect a Metropolitan Statistical Area. In Column 5 we control for the black share and income growth between 2000 and 2006-2010 as measured in Census data. The typical sample in Column 4 consists of 325 CZs that intersect MSAs. In the other columns the typical sample consists of the 709 CZs with at least 250 children in the core sample; however, some rows have fewer observations due to missing values for the independent variable. See Section VI, Online Data Table IX, and Appendix G for definitions of each of the correlates analyzed in this table. See Online Data Table VIII for the CZ-level data on each covariate.

FIGURE I: Association between Children's and Parents' Incomes



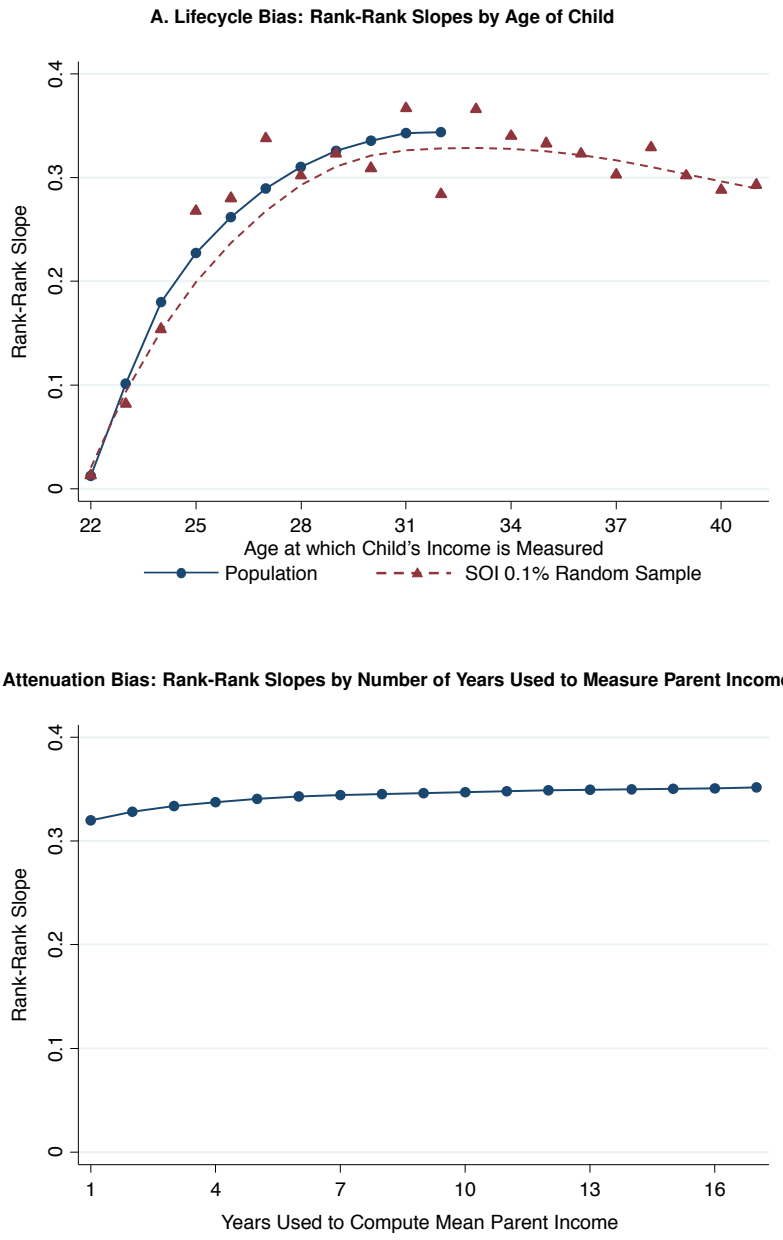
Notes: These figures present non-parametric binned scatter plots of the relationship between child income and parent income. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011-2012 family income (when the child is approximately 30 years old), while parent income is mean family income from 1996-2000. Incomes are in 2012 dollars. To construct Panel A, we bin parent family income into 100 equal-sized (centile) bins and plot the mean level of child income vs. mean level of parent income within each bin. For scaling purposes, we do not show the point for the top 1% in Panel A. In the top 1% bin, mean parent income is $1.4 million and mean child income is $114,000. In Panel B, we again bin parent family income into 100 bins and plot mean log income for children (left y-axis) and the fraction of children with zero family income (right y-axis) vs. mean parents' log income. Children with zero family income are excluded from the log income series. In both panels, the 10th and 90th percentile of parents' income are depicted in dashed vertical lines. The coefficient estimates and standard errors (in parentheses) reported on the figures are obtained from OLS regressions on the microdata. In Panel A, we report separate slopes for parents below the 90th percentile and parents between the 90th and 99th percentile. In panel B, we report slopes of the log-log regression (i.e., the intergenerational elasticity of income or IGE) in the full sample and for parents between the 10th and 90th percentiles.

FIGURE II: Association between Children's and Parents' Percentile Ranks



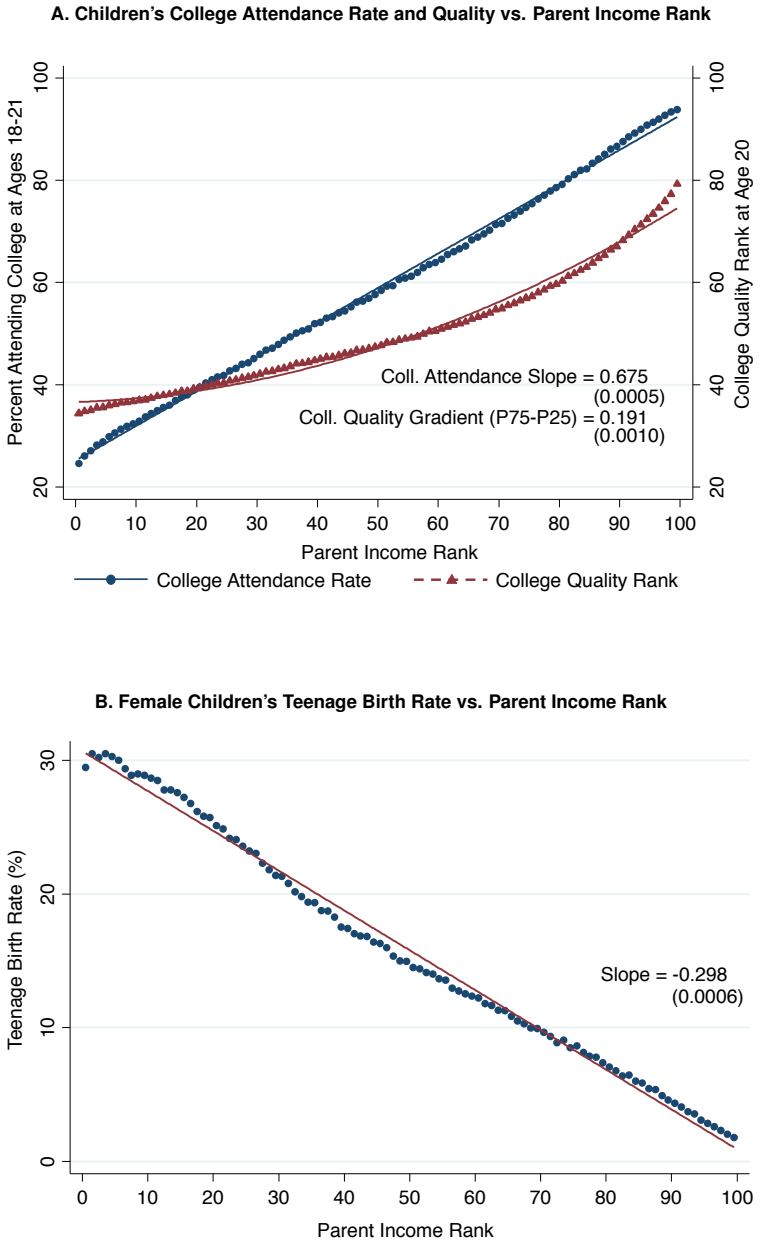
Notes: These figures present non-parametric binned scatter plots of the relationship between children's and parent's percentile income ranks. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011-2012 family income (when the child is approximately 30 years old), while parent income is mean family income from 1996-2000. Children are ranked relative to other children in their birth cohort, while parents are ranked relative to all other parents in the core sample. Panel A plots the mean child percentile rank within each parent percentile rank bin. The series in triangles in Panel B plots the analogous series for Denmark, computed by Boserup, Kopczuk, and Kreiner (2013) using a similar sample and income definitions. The series in squares plots estimates of the rank-rank series using the decile-decile transition matrix from Corak and Heisz (1999). The series in circles in Panel B reproduces the rank-rank relationship in the U.S. from Panel A as a reference. The slopes and best-fit lines are estimated using an OLS regression on the microdata for the U.S. and on the binned series (as we do not have access to the microdata) for Denmark and Canada. Standard errors are reported in parentheses.

FIGURE III: Robustness of Intergenerational Mobility Estimates



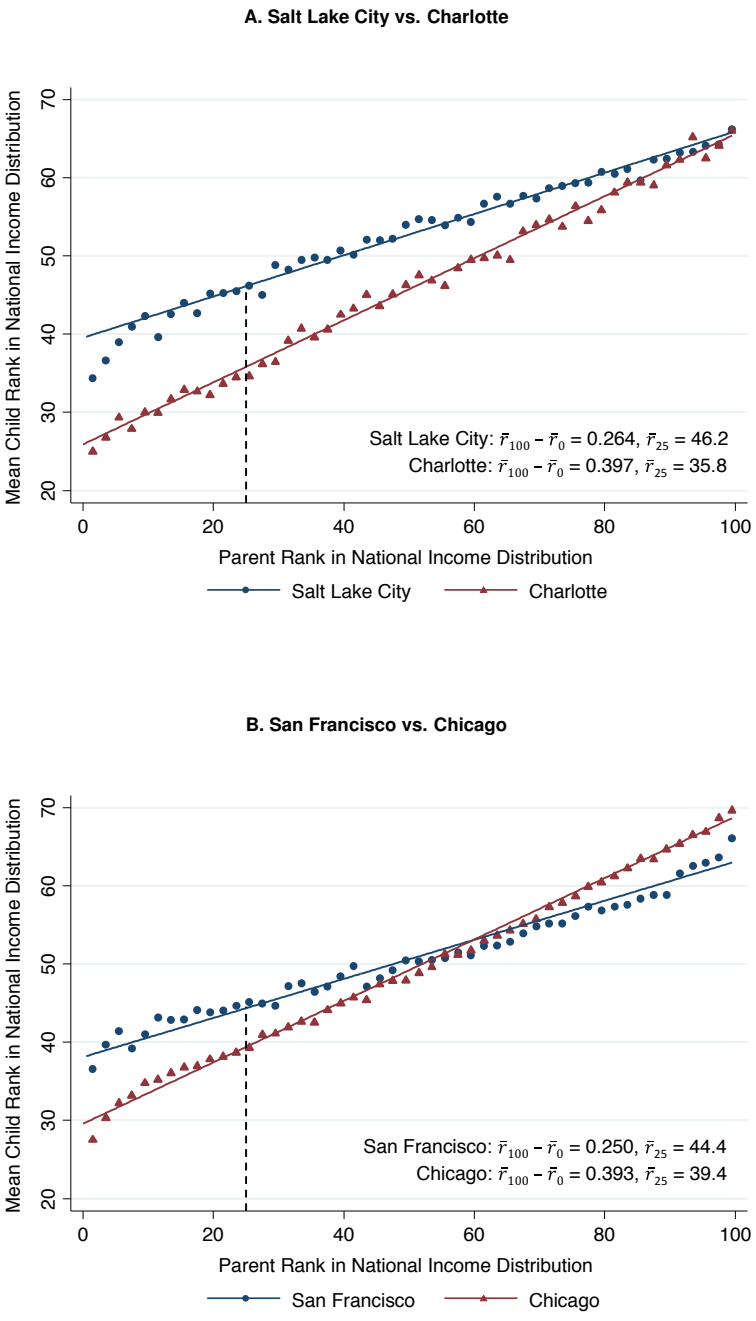
Notes: This figure evaluates the robustness of the rank-rank slope estimated in Figure IIa to changes in the age at which child income is measured (Panel A) and the number of years used to measure parents' income (Panel B). In both panels, child income is defined as mean family income in 2011-2012. In Panel A, parent income is defined as mean family income from 1996-2000. Each point in Panel A shows the slope coefficient from a separate OLS regression of child income rank on parent income rank, varying the child's birth cohort and hence the child's age in 2011-12 when the child's income is measured. The circles use the extended sample in the population data, while the triangles use the 0.1% Statistics of Income stratified random sample. The first point in Panel A corresponds to the children in the 1990 birth cohort, who are 21-22 when their incomes are measured in 2011-12 (denoted by age 22 on the figure). The last point for which we have population-wide estimates corresponds to the 1980 cohort, who are 31-32 (denoted by 32) when their incomes are measured. The last point in the SOI sample corresponds to the 1971 cohort, who are 40-41 (denoted by 41) when their incomes are measured. The dashed line is a lowess curve fit through the SOI 0.1% sample rank-rank slope estimates. In Panel B, we focus on children in the core sample (1980-82 birth cohorts) in the population data. Each point in this figure shows the coefficient from the same rank-rank regression as in Figure IIa, varying the number of years used to compute mean parent income. The first point uses parent income data for 1996 only to define parent ranks. The second point uses mean parent income from 1996-1997. The last point uses mean parent income from 1996-2012, a 17 year average.

FIGURE IV: Gradients of College Attendance and Teenage Birth by Parent Rank



Notes: These figures present non-parametric binned scatter plots of the relationship between children's college attendance rates (Panel A, circles), college quality rank (Panel A, triangles), and teenage birth rates (Panel B) vs. parents' percentile rank. Both figures are based on the core sample (1980-82 birth cohorts). Parent rank is defined based on mean family income from 1996-2000. In Panel A, the circles plot the fraction of children ever attending college between age 18-21 within each parent-income percentile bin; the triangles plot the average college quality rank at age 20 within each parent-income percentile bin. College attendance is defined as the presence of a 1098-T form filed by a college on behalf of the student. College quality rank is defined as the percentile rank of the college that the child attends at age 20 based on the mean earnings at age 31 of children who attended the same college (children who do not attend college are included in a separate "no college" group); see Section III.B for further details. Panel B plots the fraction of female children who give birth while teenagers within each parental percentile bin. Having a teenage birth is defined as ever claiming a dependent child who was born while the mother was aged 13-19. The slopes and best-fit lines for college attendance and teenage birth are estimated using linear regressions of the outcome of interest on parent income rank in the microdata. We regress college quality rank on a quadratic in parent rank to match the non-linearity of the relationship. The college quality gradient is defined as the difference between the fitted values for children with parents at the 75th percentile and parents at the 25th percentile using this quadratic specification.

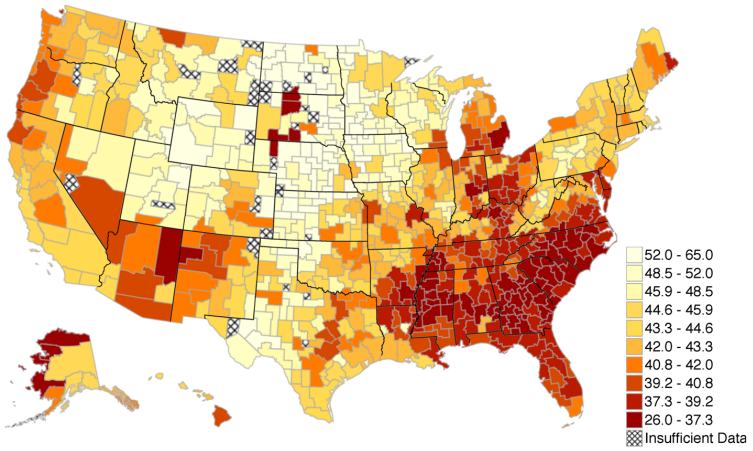
FIGURE V: Intergenerational Mobility in Selected Commuting Zones



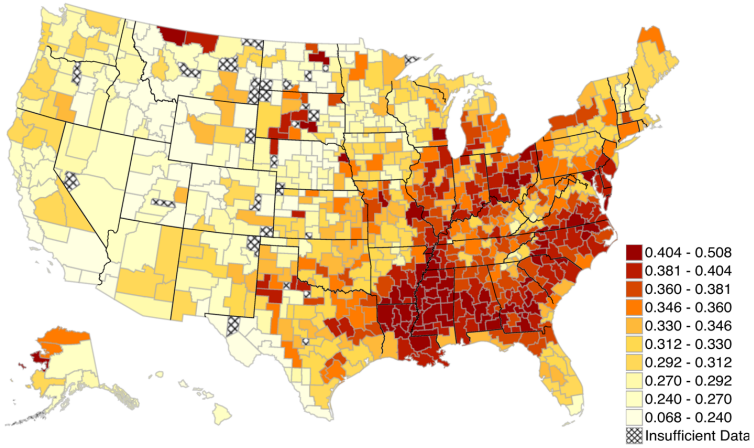
Notes: These figures present non-parametric binned scatter plots of the relationship between child and parent income ranks in selected CZs. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. Parent and child percentile ranks are always defined at the national level, not the CZ level. To construct each series, we group parents into 50 equally sized (two percentile point) bins and plot the mean child percentile rank vs. the mean parent percentile rank within each bin. We report two measures of mobility based on the rank-rank relationships in each CZ. The first is relative mobility ($\bar{r}_{100} - \bar{r}_0$), which is 100 times the rank-rank slope estimate. The second is absolute upward mobility ($\bar{r}_{25}$), the predicted child income rank at the 25th percentile of parent income distribution, depicted by the dashed vertical line in the figures. All mobility statistics and best-fit lines are estimated on the underlying microdata (not the binned means).

FIGURE VI: The Geography of Intergenerational Mobility

A. Absolute Upward Mobility: Mean Child Rank for Parents at 25th Percentile $(\bar{r}_{25})$ by CZ



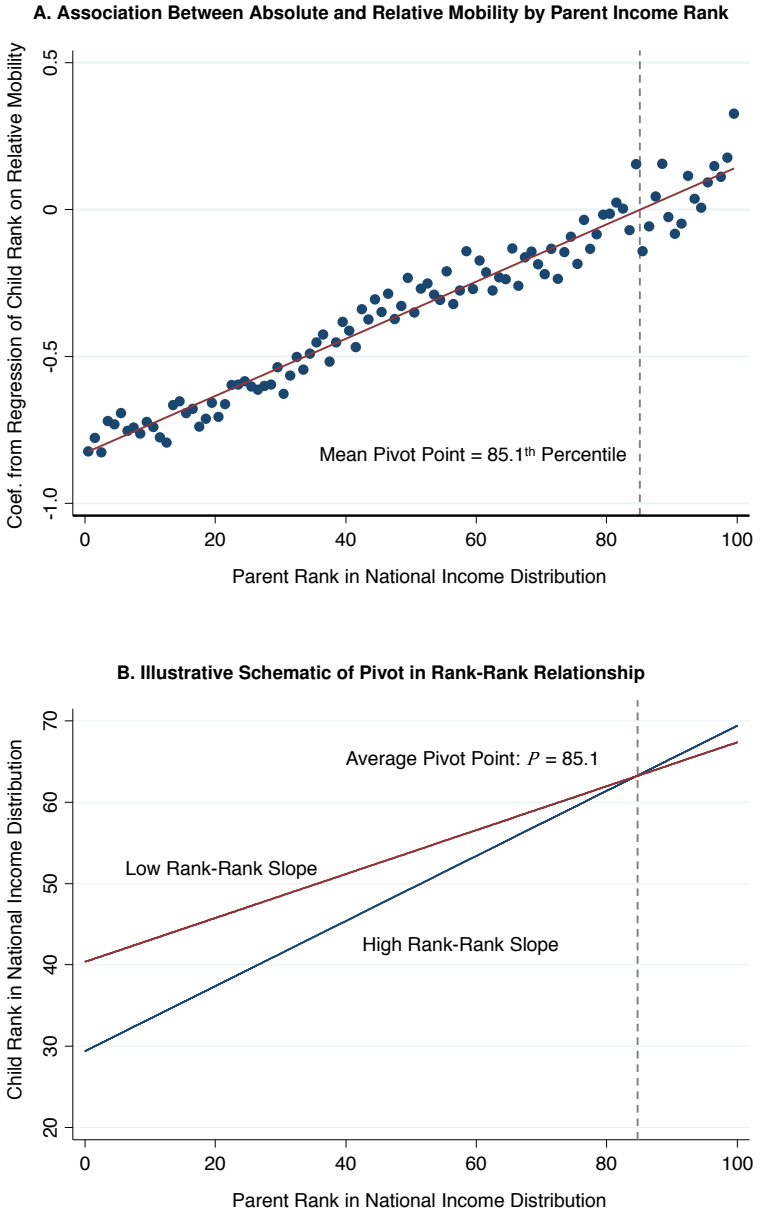
B. Relative Mobility: Rank-Rank Slopes $(\bar{r}_{100} - \bar{r}_0)/100$ by CZ



Corr. with baseline $\bar{r}_{25}$ = -0.68 (unweighted), -0.61 (pop-weighted)

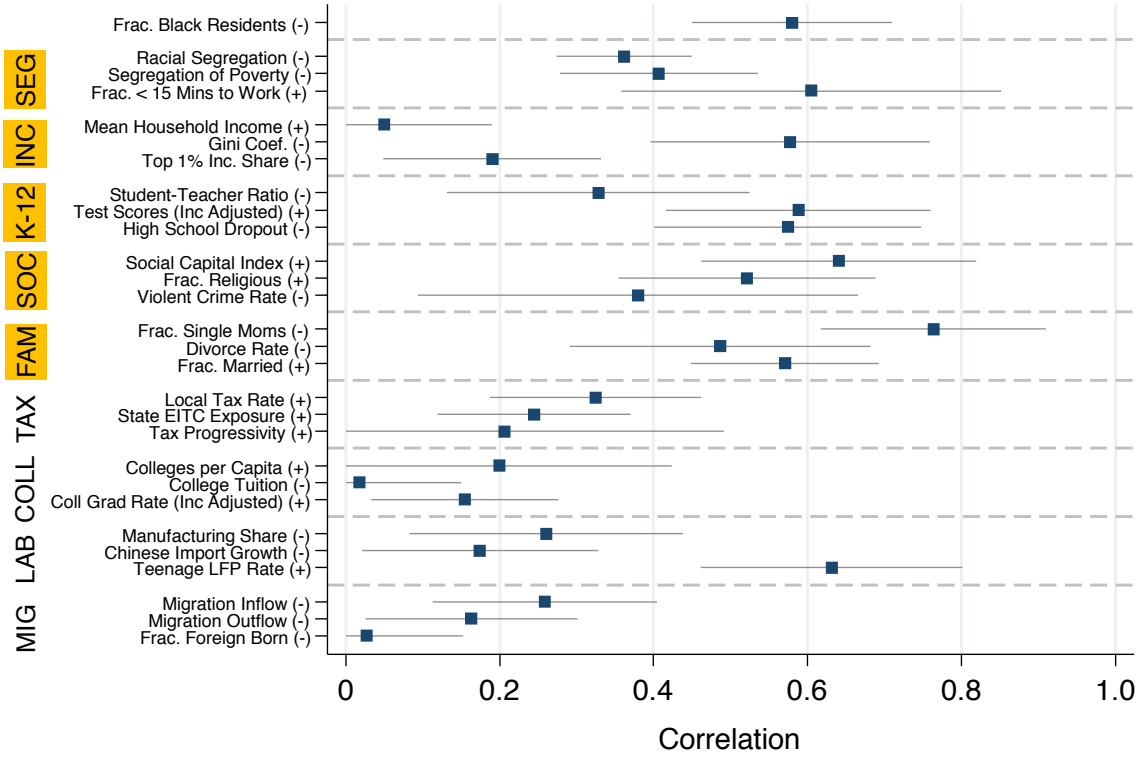
Notes: These figures present heat maps of our two baseline measures of intergenerational mobility by commuting zone (CZ). Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. In each CZ, we regress child income rank on a constant and parent income rank. Using the regression estimates, we define absolute upward mobility ($\bar{r}_{25}$) as the intercept + 25×(rank-rank slope), which corresponds to the predicted child rank given parent income at the 25th percentile (see Figure V). We define relative mobility as the rank-rank slope; the difference between the outcomes of the child from the richest and poorest family is 100 times this coefficient $(\bar{r}_{100} - \bar{r}_0)$. The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to higher absolute mobility (Panel A) and lower rank-rank slopes (Panel B). Areas with fewer than 250 children in the core sample, for which we have inadequate data to estimate mobility, are shaded with the cross-hatch pattern. In Panel B, we report the unweighted and population-weighted correlation coefficients between relative mobility and absolute mobility across CZs. The CZ-level statistics underlying these figures are reported in Online Data Table V.

FIGURE VII: Relationship Between Absolute and Relative Mobility



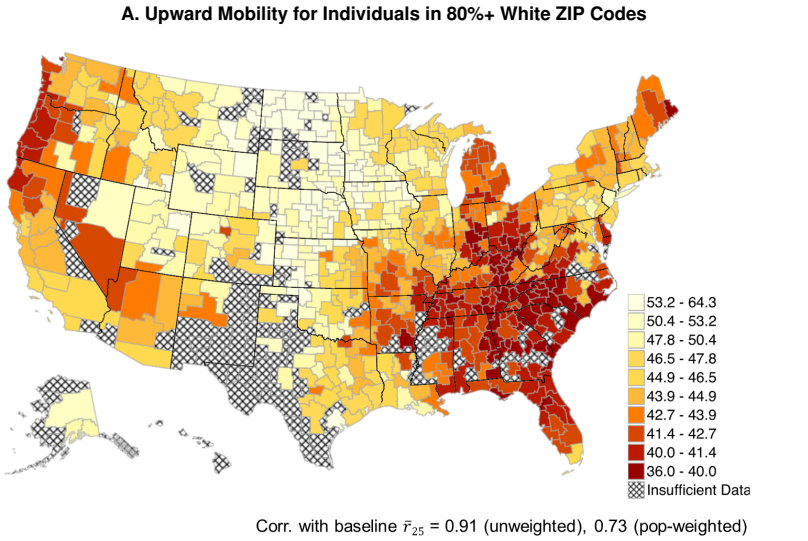
Notes: These figures illustrate the correlation between relative mobility and absolute mobility at various percentiles of the income distribution. To construct Panel A, we first calculate the mean income rank of children in CZ c with parents in (national) percentile p, denoted by $\bar{\mu}_{pc}$. We then run a CZ-level regression of $\bar{\mu}_{pc}$ on relative mobility $(\bar{r}_{100c} - \bar{r}_{0c})$ at each percentile p separately. Panel A plots the resulting regression coefficients γ_p vs. the percentile p. The coefficient γ_p can be interpreted as the mean impact of a 1 unit increase in relative mobility on the absolute outcomes of children whose parents are at percentile p. We also plot the best linear fit across the 100 coefficients. This line, estimated using an OLS regression, crosses zero at percentile $p = 85.1$. This implies that increases in relative mobility are associated with higher expected rank outcomes for children with parents below percentile 85.1 and lower expected rank outcomes for children with parents above percentile 85.1. To illustrate the intuition for this result, Panel B plots hypothetical rank-rank relationships in two representative CZs, one of which has more relative mobility than the other. Panel A implies that in such a pairwise comparison, the two rank-rank relationships cross at the 85th percentile on average, as illustrated in Panel B.

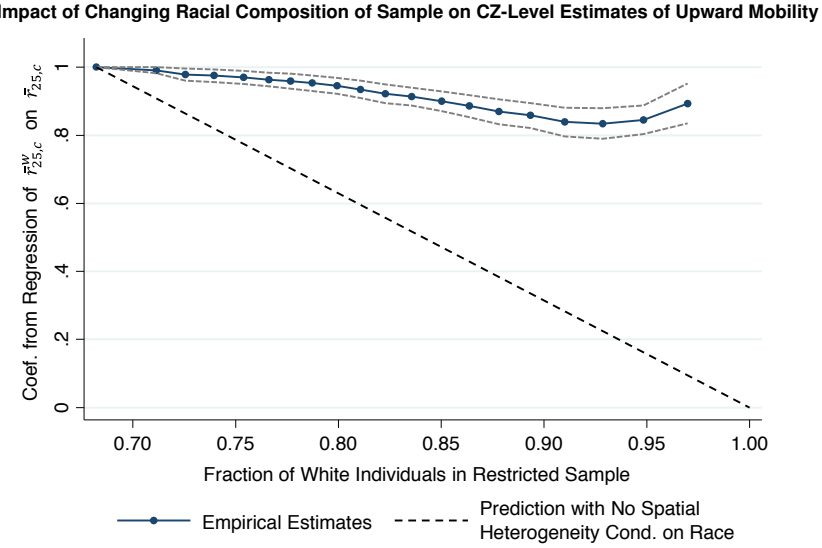
FIGURE VIII: Correlates of Spatial Variation in Upward Mobility



Notes: This figure shows the correlation of various CZ-level characteristics with absolute upward mobility ($\bar{r}_{25}$) across CZs. For each characteristic listed on the y axis, the dot represents the absolute value of the unweighted correlation of the variable with $\bar{r}_{25}$ across CZs. The horizontal bars show the 95% confidence interval based on standard errors clustered at state level. Positive correlations are shown by (+) on the y axis; negative correlations are shown by (-). We consider covariates in ten broad categories: racial demographics, segregation, properties of the income distribution, K-12 education, social capital, family structure, local tax policies, college education, labor market conditions, and migration rates. The categories with the highest correlations are highlighted. See Column 1 of Appendix Table VII for the point estimates corresponding to the correlations plotted here. See Section VI, Online Data Table IX, and Online Appendix G for definitions of each of the correlates. CZ-level data on the covariates used in this figure are reported in Online Data Table VIII.

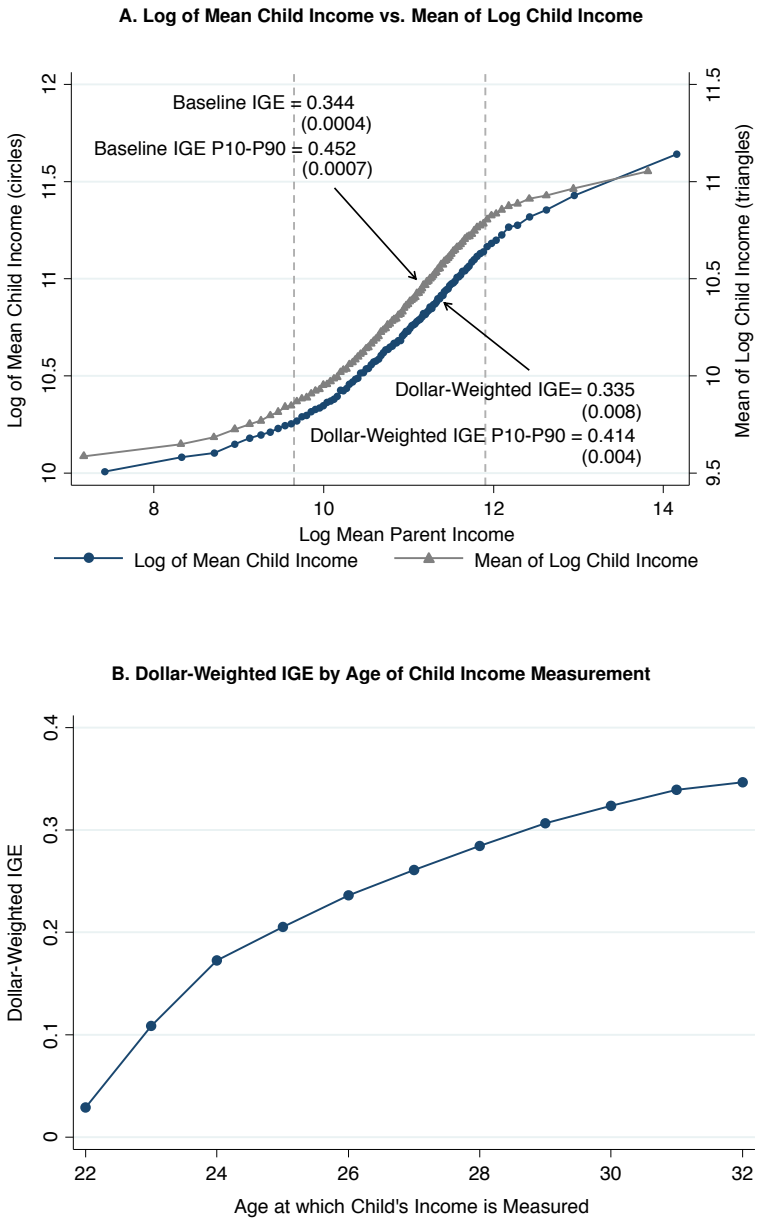
FIGURE IX: Race and Upward Mobility





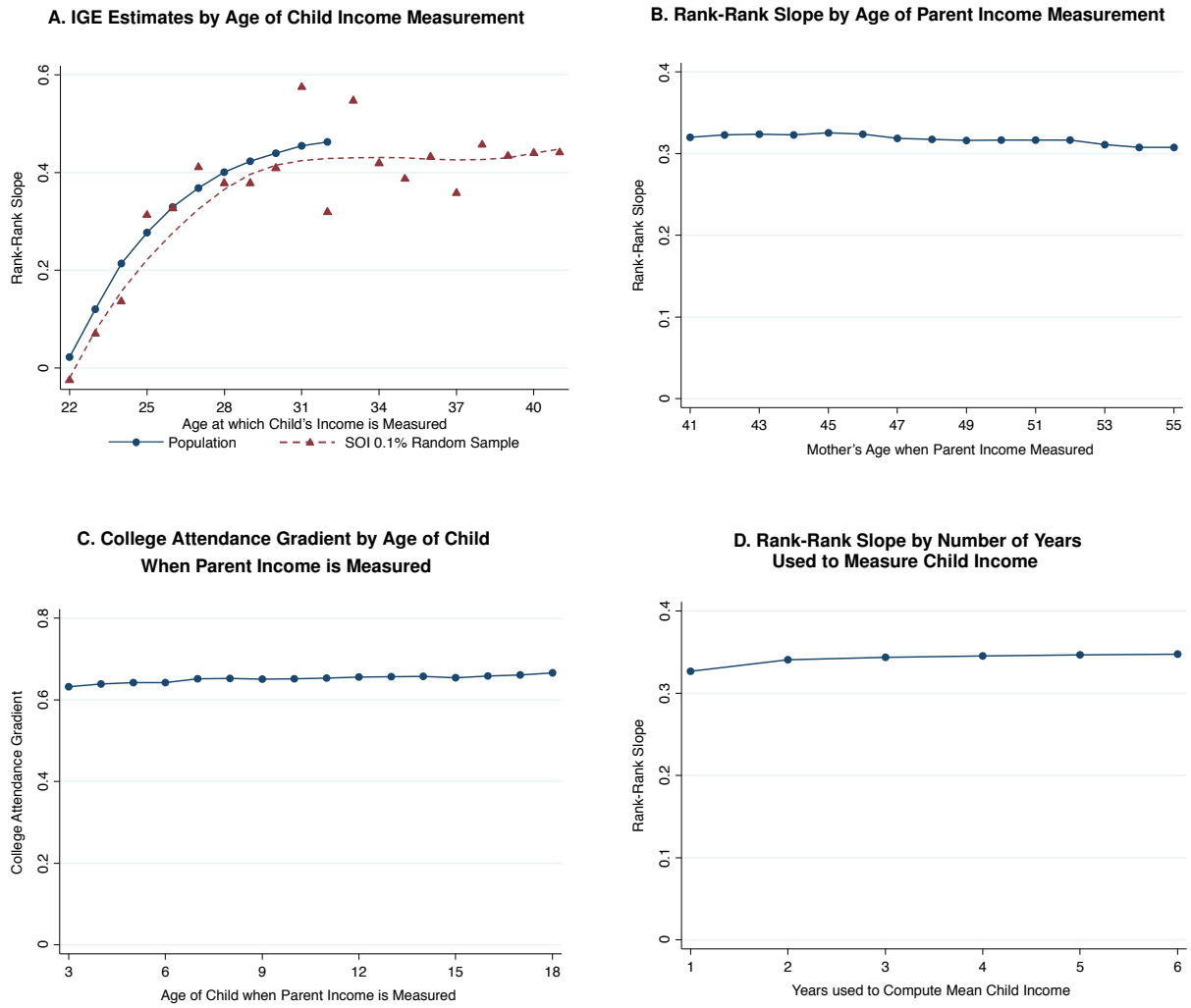
Notes: Panel A presents a heat map of absolute upward mobility for individuals living in ZIP codes with 80% or more white residents. This figure replicates Figure VIa, restricting the sample used to estimate the rank-rank regression in each CZ to parents living in ZIP codes with 80% or more white residents. Note that we color the entire CZ based on the resulting estimate of upward mobility (not just the ZIP codes used in the estimation) for comparability to other figures. CZs with fewer than 250 children living in ZIP codes with >80% white share are omitted and shaded with the cross-hatch pattern. We report the unweighted and population-weighted correlation coefficients between this measure and absolute upward mobility presented in Figure VIa across CZs. To construct Panel B, we first compute upward mobility in each CZ, restricting the sample to individuals living in ZIP codes that are more than w% white, which we denote by $\bar{r}^{w}_{25,c}$. We then regress $\bar{r}^{w}_{25,c}$ on $\bar{r}_{25,c}$, our baseline estimates of upward mobility based on the full sample, using an unweighted OLS regression with one observation per CZ with available data. We vary w from 0% to 95% in increments of 5% and plot the resulting regression coefficients against the fraction of white individuals in each of the subsamples. The confidence interval, shown by the dotted lines around the point estimates, is based on standard errors clustered at the state level. The dashed diagonal line shows the predicted relationship if there were no spatial heterogeneity in upward mobility conditional on race.

ONLINE APPENDIX FIGURE I
Dollar-Weighted vs. Traditional IGE Estimates



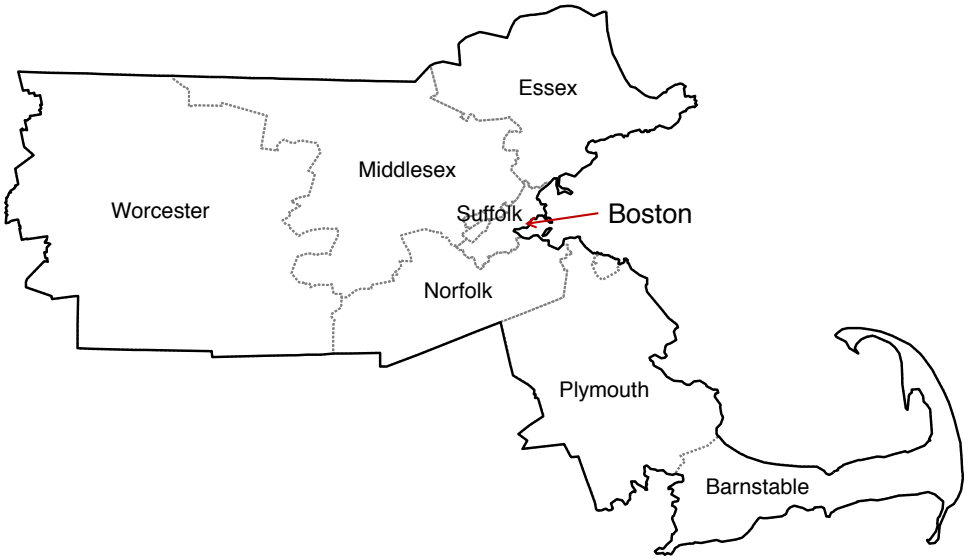
Notes: This figure compares dollar-weighted (Mitnik et al. 2014) and traditional IGE estimates. Panel A is based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. The series in circles (left axis) plots log of mean child income against log of mean parent income. The series is constructed by taking the logs of the points in Figure Ia; however, here we do not omit the top income bin. The slope coefficients, which correspond to the dollar-weighted IGE defined in Appendix C, and standard errors are estimated by OLS on the binned data. The series in triangles (right axis) reports the mean of log child income vs. the mean of log parent income (reproducing the series in Figure Ib). The slope coefficients and standard errors for the traditional IGE are estimated on the microdata. The dashed lines in Panel A show the 10th and 90th percentiles of the parent income distribution. Panel B shows how the dollar-weighted IGE varies with the age at which child income is measured. We estimate the dollar-weighted IGE by grouping parents into 100 bins based on their income rank and regressing the log of mean child income on the log of mean parent income across the 100 bins. The figure plots the slope from this regression vs. the age at which child income is measured. We measure child income in 2011-12 and analyze how the IGE varies across birth cohorts, as in Figure IIIa; see notes to that figure for further details. The first point corresponds to the children in the 1990 birth cohort, who are 21-22 when their incomes are measured in 2011-12 (denoted by age 22 on the figure). The last point corresponds to the 1980 cohort, who are 31-32 (denoted by 32) when their incomes are measured.

ONLINE APPENDIX FIGURE II
Additional Evidence on Robustness of Intergenerational Mobility Estimates



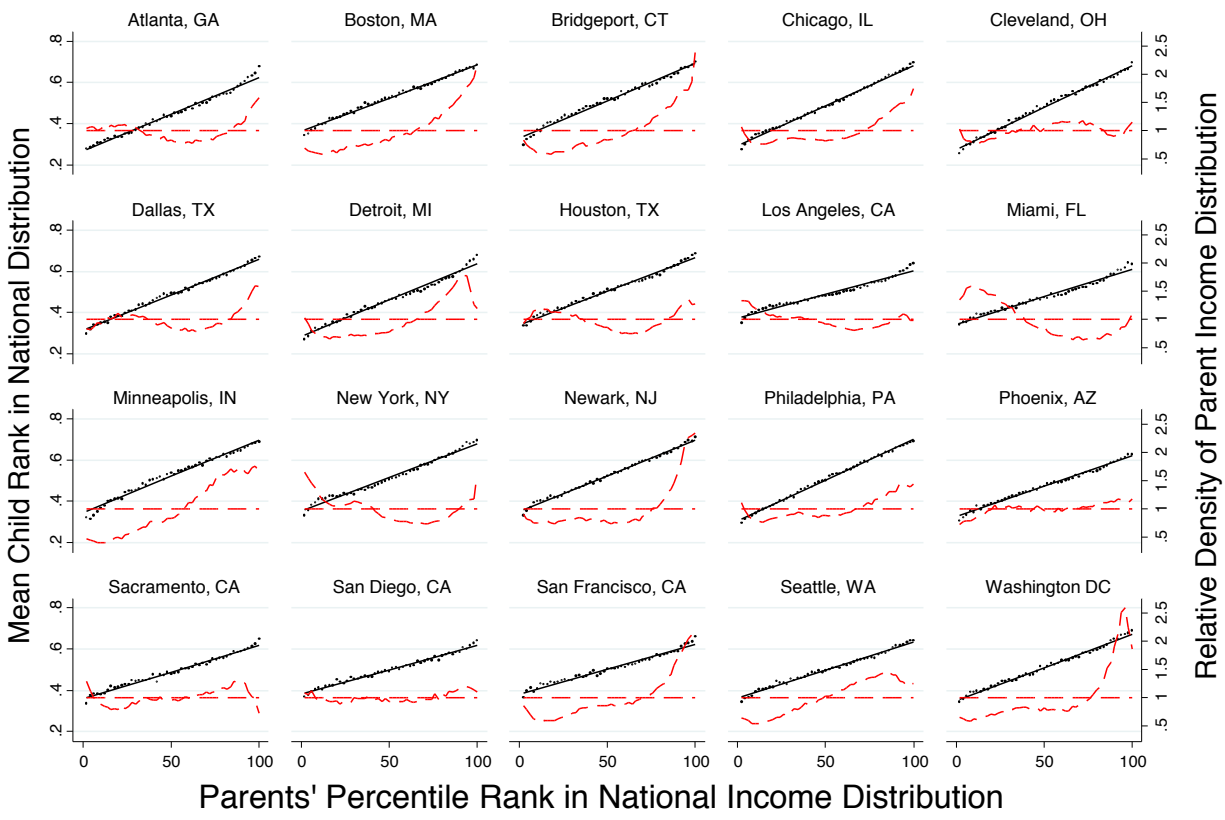
Notes: This figure evaluates the robustness of intergenerational mobility measures to lifecycle and attenuation bias. Panel A evaluates the robustness of the IGE to changes in the age at which child income is measured. Panel B evaluates the robustness of the rank-rank slope to changes in the age at which parent income is measured. Panel C evaluates the robustness of the college attendance gradient to the age of the child when parent income is measured. Panel D evaluates the robustness of the rank-rank slope to the number of years used to measure the child's income. In Panel A, we estimate the log-log IGE (excluding children with zero income), varying the age at which child income is measured. We restrict the sample to parents with income between the 10th and 90th percentile when estimating the IGE, as shown in Figure Ib. We measure child income in 2011-12 and analyze how the IGE varies across birth cohorts, as in Figure IIIa; see notes to that figure for further details. In Panel B, each point shows the slope coefficient from an OLS regression of child income rank on parent income rank (as in Figure IIa), using the core sample and varying the age at which parent income rank is measured. The first point measures parent income in 1996, when the mean age of mothers is 41. The last point measures income in 2010, when parents are 55. Panel C reproduces Appendix Figure 2b from Chetty et al. (2014). In this figure, each point shows the slope coefficient from an OLS regression of an indicator for the child attending college at age 19 on parent income rank (similar to Figure IVa), varying the year in which parent income rank is measured from 1996 to 2011. In this series, we use data from the 1993 birth cohort, which allows us to analyze parent income starting when children are 3 years old in 1996. We list the age of the child on the x axis to evaluate whether the gradient differs when children are young (although parent age is of course also rising in lockstep). In Panel D, each point shows the slope coefficient from the same rank-rank regression as in Panel B using the core sample, but here we always use a five-year (1996-2000) mean to measure parent income and vary the number of years used to compute mean child income. The point for one year measures child income in 2012 only. The point for two years uses mean child income in 2011-12. We continue adding data for prior years; the 6th point uses mean income in years 2007-2012.

ONLINE APPENDIX FIGURE III
Boston Commuting Zone



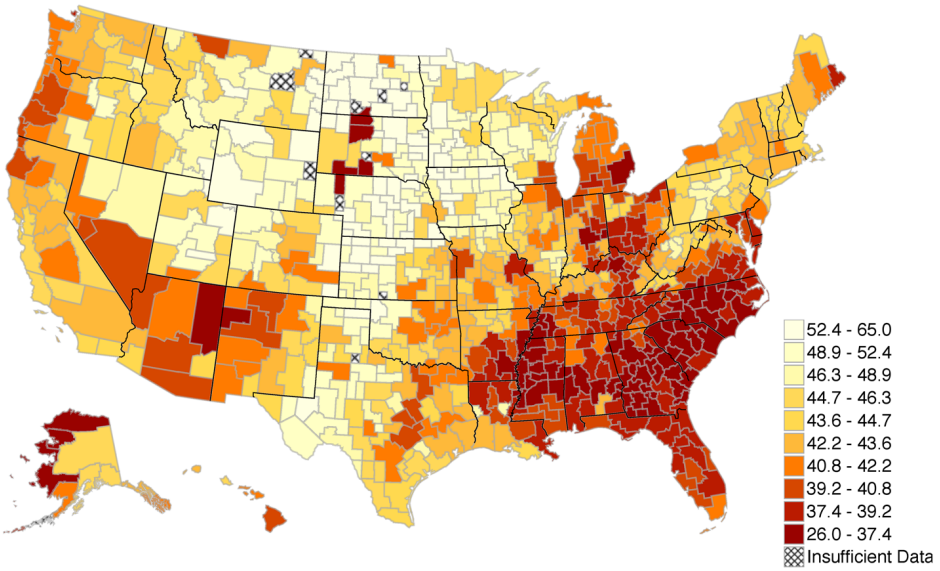
Notes: This figure shows a map of the counties that comprise the Boston Commuting Zone. The city of Boston is shown by the arrow.

ONLINE APPENDIX FIGURE IV
Rank-Rank Relationships and Income Distributions in the 20 Largest CZs



Notes: These figures present non-parametric binned scatter plots (shown by the points and solid line, left y-axis) of the relationship between child and parent income ranks in the twenty largest CZs based on population in the 2000 Census. All figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to commuting zones based on the location of their parents. Parent and child percentile ranks are always defined at the national level, not the CZ level. To construct each rank-rank series, we group parents into 50 equally sized (two percentile point) bins and plot the mean child percentile rank vs. the mean parent percentile rank within each bin. The dashed curve (right y-axis) in each panel depicts the income distribution in the CZ relative to the national distribution. This curve plots the share of parents with income in each bin in the CZ divided by the share in the same bin in the national income distribution. By construction, this curve averages to one in each CZ, shown by the horizontal dashed line in each panel.

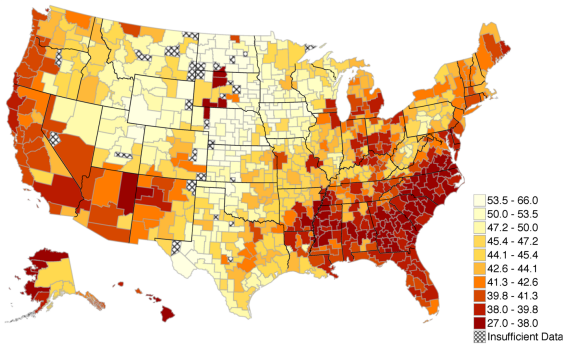
ONLINE APPENDIX FIGURE V
Estimates of Absolute Upward Mobility Pooling 1980-82 and 1980-85 Cohorts



Notes: The figure presents the map of absolute upward mobility by CZ shown on the project homepage (www.equality-of-opportunity.org). For the 709 CZs that have at least 250 children in the 1980-82 cohorts, we compute absolute upward mobility exactly as in Figure VIa. For an additional 22 CZs that have fewer than 250 children in the 1980-82 cohorts but at least 250 children in the 1980-85 cohorts, we report estimates of absolute upward mobility using the 1980-85 birth cohorts. We estimate absolute upward mobility using exactly the same procedure as described in the notes to Figure VIa. The map is constructed by grouping CZs into ten deciles based on the hybrid absolute mobility measure and shading the areas so that lighter colors correspond to higher absolute mobility. Areas with fewer that 250 children in the 1980-85 cohorts are shaded with the cross-hatch pattern.The CZ-level statistics underlying this map are reported in Online Data Table V.

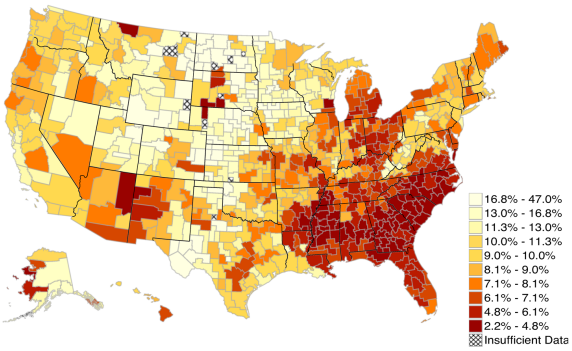
ONLINE APPENDIX FIGURE VI
Alternative Measures of Upward Mobility

A. Absolute Upward Mobility Adjusted for Local Cost-of-Living



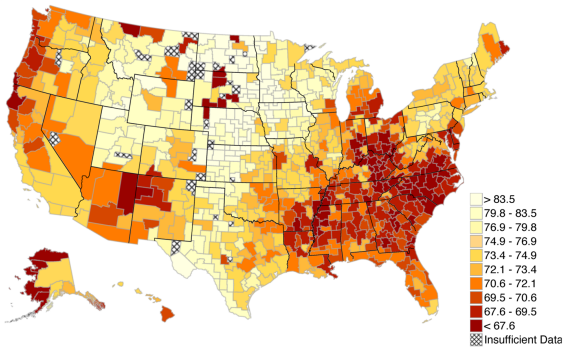
Corr. with baseline $\bar{r}_{25}$ = 0.98 (unweighted), 0.86 (pop-weighted)

B. Probability of Reaching Top Quintile from Bottom Quintile



Corr. with baseline $\bar{r}_{25}$ = 0.91 (unweighted), 0.92 (pop-weighted)

C. Fraction of Children Above Poverty Line Given Parents at 25th Percentile

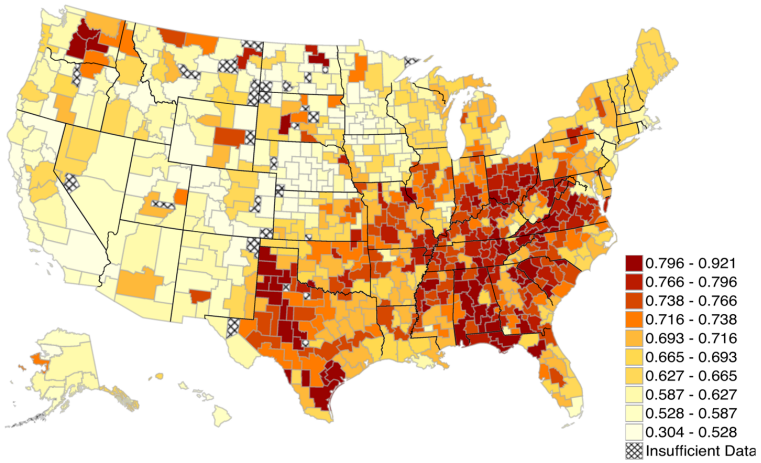


Corr. with baseline $\bar{r}_{25}$ = 0.94 (unweighted), 0.89 (pop-weighted)

Notes: These figures present heat maps for alternative measures of upward income mobility. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. All panels use baseline family income definitions for parents. Panels A and C use the core sample (1980-82 birth cohorts) and panel B uses the 1980-85 birth cohorts. Panel A replicates Figure VIa, adjusting for differences in cost-of-living across areas. To construct this figure, we first deflate parent income by a cost-of-living index (COLI) for the parent's CZ when he/she claims the child as a dependent and child income by a COLI for the child's CZ in 2012. We then compute parent and child ranks using the resulting real income measures and replicate the procedure in Figure VIa exactly. The COLI is constructed using data from the ACCRA price index combined with information on housing values and other variables as described in Appendix A. Panel B presents a heat map of the probability that a child reaches the top quintile of the national family income distribution for children conditional on having parents in the bottom quintile of the family income distribution for parents. These probabilities are taken directly from Online Data Table VI. Panel C shows the fitted values at parent rank 25 from a regression of an indicator for child family income being above the poverty line on parent income rank (see Appendix F for details). The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to higher mobility. Areas with fewer that 250 children in the core sample (or the 1980-85 cohorts for Panel B), for which we have inadequate data to estimate mobility, are shaded with the cross-hatch pattern. We report the unweighted and population-weighted correlation coefficient across CZs between these mobility measures and the baseline measure in Figure VIa. The CZ-level statistics underlying Panels A and C are reported in Online Data Table V.

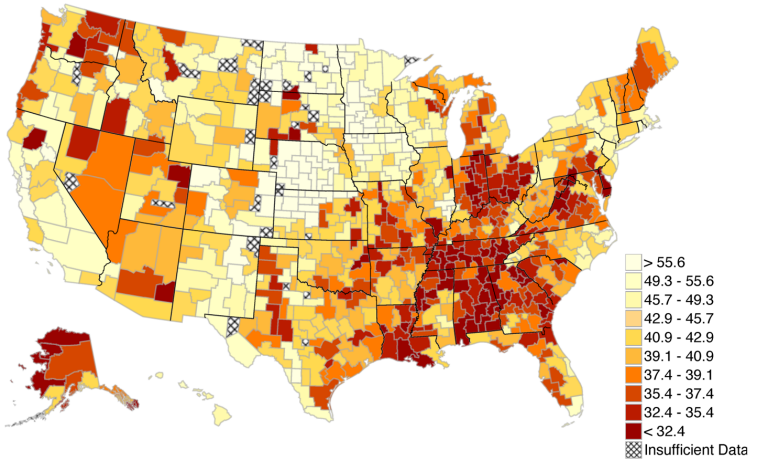
ONLINE APPENDIX FIGURE VII
The Geography of College Attendance by Parent Income Gradients

A. Slope of College Attendance-Parent Rank Gradients by CZ



Corr. with baseline $\bar{r}_{100}$- $\bar{r}_0$ = 0.68 (unweighted), 0.72 (pop-weighted)

B. College Attendance Rates for Children with Parents at the 25th Percentile by CZ

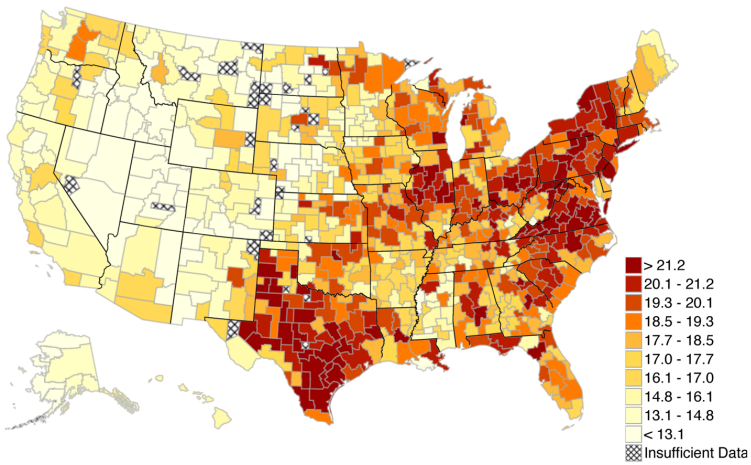


Corr. with baseline $\bar{r}_{25}$ = 0.71 (unweighted), 0.53 (pop-weighted)

Notes: To construct these figures, we regress an indicator for college attendance on parent income rank (in the national distribution) for each CZ separately. College attendance is defined by the presence of a 1098-T form filed by a college on behalf of the student. We use the core sample (1980-82 birth cohorts) and baseline family income definitions for parents. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. In Panel A, we map the slope coefficients on the college attendance indicator from the CZ-level regressions. Panel B maps the fitted values from the regressions at parent rank 25. The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to higher mobility (smaller slopes in Panel A and higher fitted values in Panel B). Areas with fewer that 250 children in the core sample, for which we have inadequate data to estimate mobility, are shaded with the cross-hatch pattern. We report the unweighted and population-weighted correlation coefficients across CZs between these mobility measures and the baseline measures in Figure VI. The CZ-level statistics underlying these figures are reported in Online Data Table V.

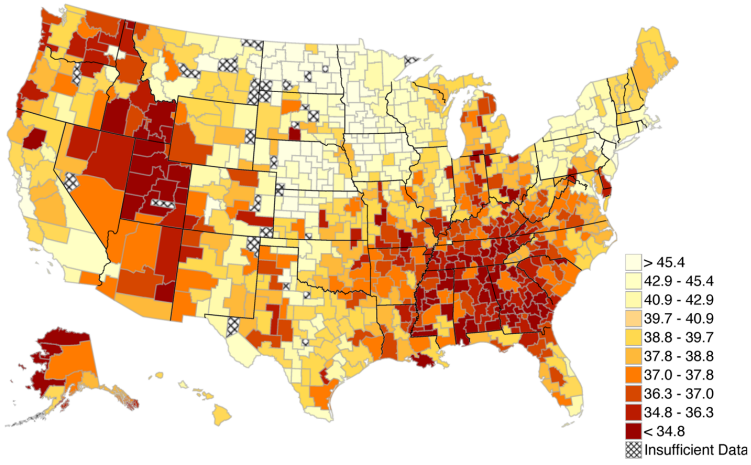
ONLINE APPENDIX FIGURE VIII
The Geography of College Quality by Parent Income Gradients

A. College Quality Gradient (P75-P25) by CZ



Corr. with baseline $\bar{r}_{100}$- $\bar{r}_0$ = 0.51 (unweighted), 0.65 (pop-weighted)

B. Mean College Quality Rank for Children with Parents at the 25th Percentile by CZ

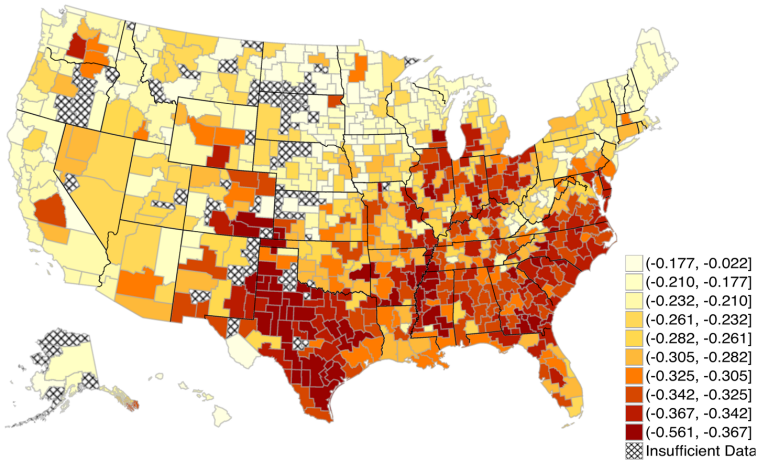


Corr. with baseline $\bar{r}_{25}$ = 0.71 (unweighted), 0.55 (pop-weighted)

Notes: To construct these figures, we regress college quality rank on a quadratic in parent income rank (in the national distribution) for each CZ separately. College quality rank is defined as the percentile rank of the college that the child attends at age 20 based on the mean earnings at age 31 of children who attended the same college (children who do not attend college are included in a separate "no college" group); see Section III.B for further details. We use the core sample (1980-82 birth cohorts) and baseline family income definitions for parents. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. In Panel A, we map the college quality income gradient, defined as the difference between the fitted values at parent rank 75 and parent rank 25 from the CZ-level regressions. Panel B maps the fitted values of college quality rank at parent rank 25 from these regressions. The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to higher mobility (smaller gradients in Panel A and higher fitted values in Panel B). Areas with fewer that 250 children in the core sample, for which we have inadequate data to estimate mobility, are shaded with the cross-hatch pattern. We report the unweighted and population-weighted correlation coefficients across CZs between these mobility measures and the baseline measures in Figure VI. The CZ-level statistics underlying these figures are reported in Online Data Table V.

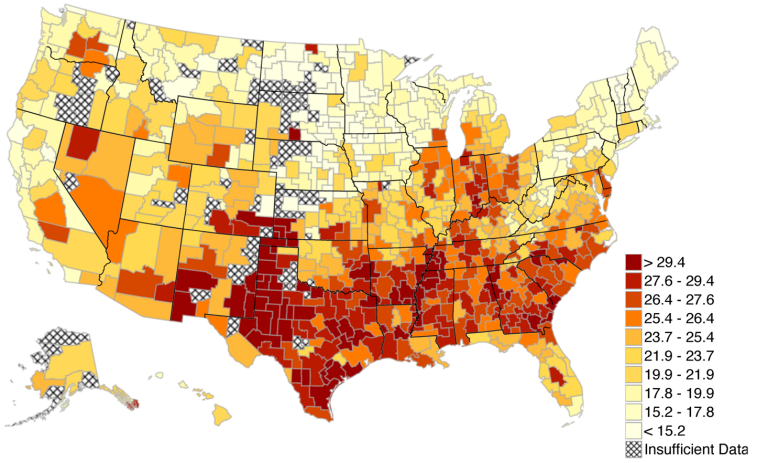
ONLINE APPENDIX FIGURE IX
The Geography of Teenage Birth by Parent Income Gradients

A. Slope of Teenage Birth-Parent Rank Gradients by CZ



Corr. with baseline $\bar{r}_{100}$-$\bar{r}_{0}$ = -0.58 (unweighted), -0.68 (pop-weighted)

B. Teenage Birth Rates for Children with Parents at the 25th Percentile by CZ

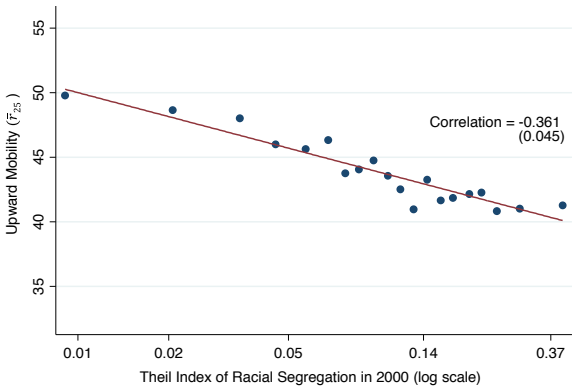


Corr. with baseline $\bar{r}_{25}$ = -0.61 (unweighted), -0.64 (pop-weighted)

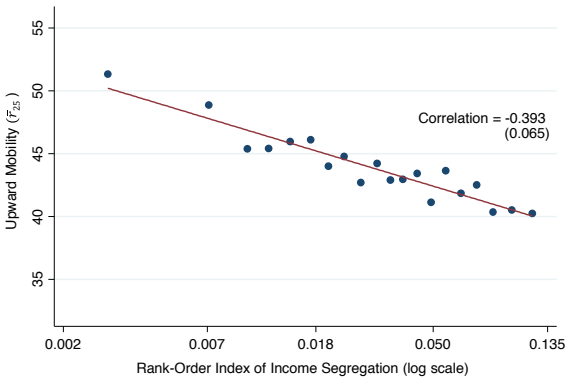
Notes: To construct these figures, we regress an indicator for teenage birth on parent income rank (in the national distribution) for each CZ separately. Teenage birth is defined as ever claiming a dependent child who was born while the mother was aged 13-19. We use female children in the core sample (1980-82 birth cohorts) and baseline family income definitions for parents. Children are assigned to commuting zones based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. In Panel A, we map the slope coefficient on the teenage birth indicator from the CZ-level regressions. Panel B maps the fitted values from these regressions at parent income rank 25. The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to smaller slopes (in magnitudes) in Panel A and smaller fitted values in Panel B. Areas with fewer that 250 female children in the core sample, for which we have inadequate data to estimate mobility measures, are shaded with the cross-hatch pattern. We report the unweighted and population-weighted correlation coefficients across CZs between these mobility measures and the baseline measures in Figure VI. The CZ-level statistics underlying these figures are reported in Online Data Table V.

ONLINE APPENDIX FIGURE X
Segregation and Upward Mobility

A. Upward Mobility vs. Theil Index of Racial Segregation in CZ

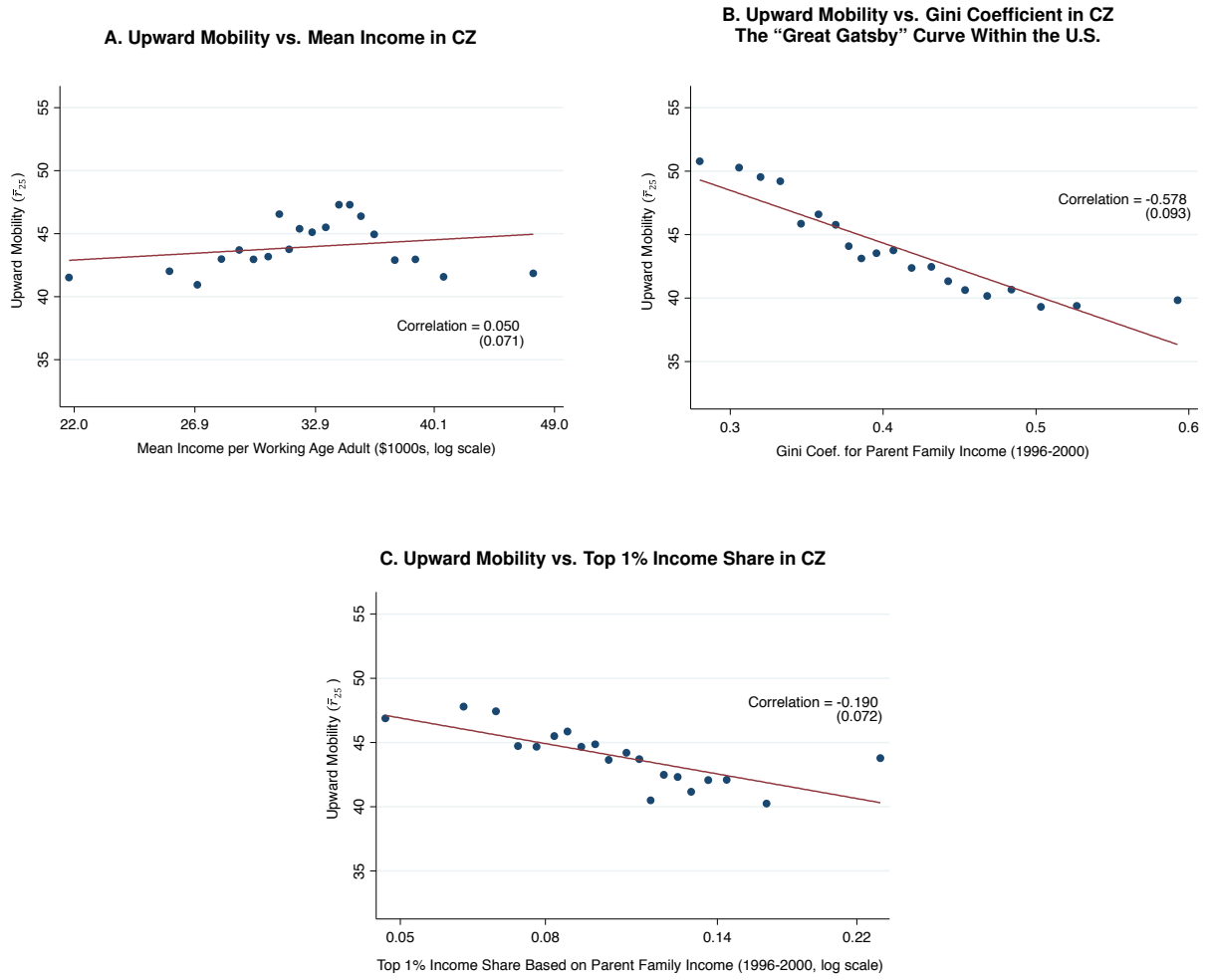


B. Upward Mobility vs. Rank-Order Index of Income Segregation in CZ



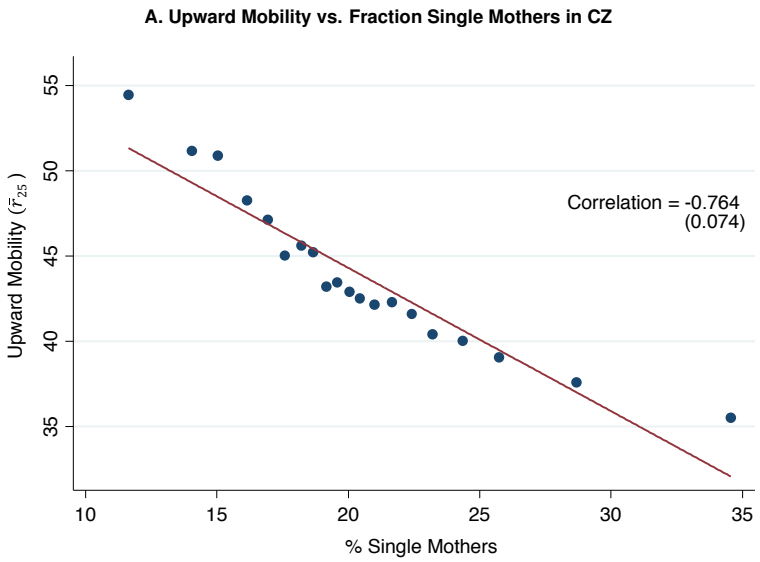
Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{r}_{25}$) vs. a multi-group Theil index of racial segregation (based on census tract level data from the 2000 Census). To construct this figure, we group CZs into twenty equally sized bins (vingtiles) based on their segregation index. We then plot the mean level of absolute upward mobility vs. the mean segregation index within each of the twenty bins (using a log scale on the x axis). Panel B presents an analogous binned scatter plot of absolute upward mobility vs. the rank-order index of income segregation from Reardon (2011). See text for details on the construction of these segregation indices. Note that these binned scatter plots provide a non-parametric representation of the conditional expectation function, but they do not show the variance in the underlying data across CZs. The correlations between the variables are estimated using the underlying CZ-level data, with standard errors (reported in parentheses) clustered by state. The correlations are estimated in levels (not logs) for consistency with Appendix Table VII.

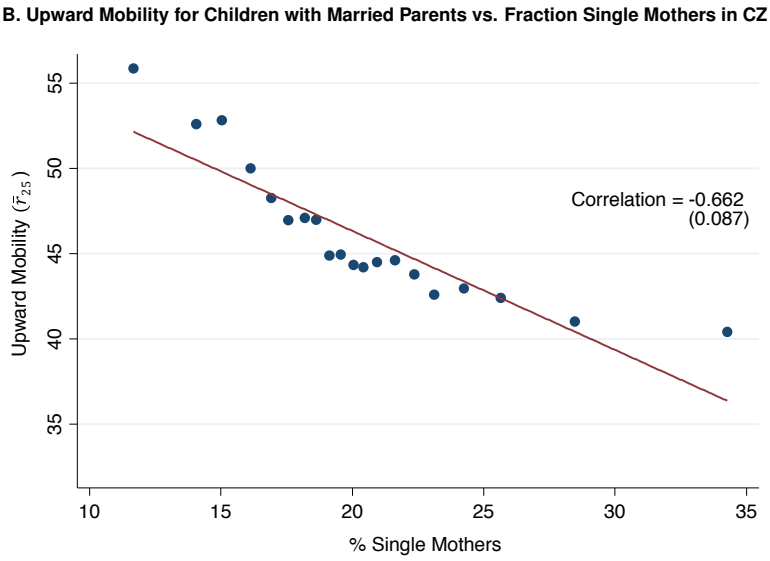
ONLINE APPENDIX FIGURE XI
Local Income Distributions and Upward Mobility



Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{r}_{25}$) vs. mean income per working age adult in the CZ (based on data from the 2000 Census). To construct this figure, we group CZs into twenty equally sized bins (vingtiles) based on mean income levels. We then plot the mean level of absolute upward mobility vs. the mean income level within each of the twenty bins (using a log scale on the x axis). Panel B presents an analogous binned scatter plot of absolute upward mobility vs. the Gini coefficient in the CZ, computed based on the core sample and mean parent income for 1996-2000. Panel C presents a binned scatter plot of absolute upward mobility vs. the fraction of income in the CZ accruing to parents in the top 1% of the local distribution (using a log scale on the x axis), again using the core sample and parents' average income for 1996-2000. Note that these binned scatter plots provide a non-parametric representation of the conditional expectation function, but they do not show the variance in the underlying data across CZs. The correlations between the variables are estimated using the underlying CZ-level data, with standard errors (reported in parentheses) clustered by state. The correlations are estimated in levels (not logs) for consistency with Appendix Table VII.

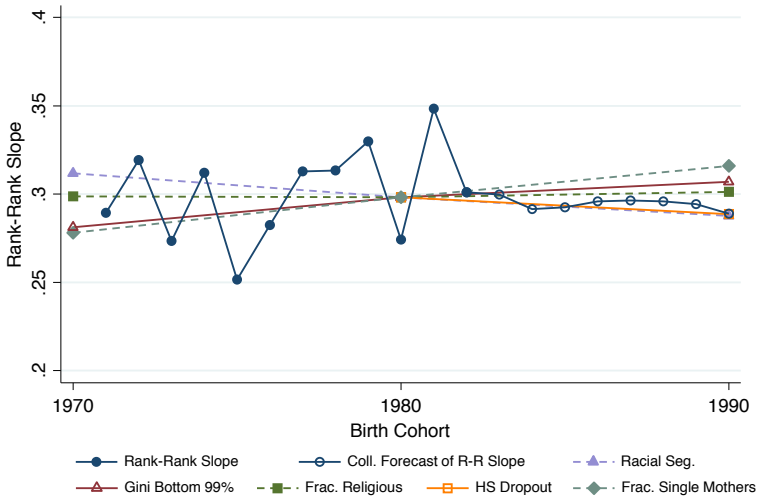
ONLINE APPENDIX FIGURE XII
Single-Parent Families and Upward Mobility





Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{r}_{25}$) vs. the fraction of children being raised by single mothers in the CZ (based on data from the 2000 Census). To construct this figure, we group CZs into twenty equally sized bins (vingtiles) based on the fraction of single parents. We then plot the mean level of absolute upward mobility vs. the mean fraction of single parents within each of the twenty bins. Panel B replicates Panel A, restricting the sample used to estimate upward mobility in each CZ to children whose own parents are married in the year they first claim the child as a dependent. Note that these binned scatter plots provide a non-parametric representation of the conditional expectation function, but they do not show the variance in the underlying data across CZs. The correlations between the variables are estimated using the underlying CZ-level data, with standard errors (reported in parentheses) clustered by state.

ONLINE APPENDIX FIGURE XIII
Predicted vs. Actual Time Trends in Relative Mobility



Notes: This figure compares actual trends in rank-rank slopes at the national level, estimated in Chetty et al. (2014), with projected changes based on trends in the five factors most strongly correlated with differences in mobility across CZs in the cross-section. The series in circles is from Chetty et al. (2014, Figure 2). The solid circles show estimates of rank-rank slopes by birth cohort using the SOI 0.1% sample. The open circles show forecasts of the rank-rank slope based on income measured at age 26 and the college attendance rates using the population data. The other series show projections of trends, each based on a different factor: (1) Theil index of racial segregation, (2) high school dropout rate, (3) Gini coefficient, (4) violent crime arrest rate, and (5) fraction of single parents. We construct these projections based on unweighted univariate CZ-level regressions of relative mobility on each factor separately. We normalize the projections (by adding a constant) so that their values match the mean observed rank-rank slope (i.e., the mean value of the series in circles) from 1971-1990. See Appendix I for further details.