Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States*

Raj Chetty, Harvard University and NBER
Nathaniel Hendren, Harvard University and NBER
Patrick Kline, UC-Berkeley and NBER
Emmanuel Saez, UC-Berkeley and NBER

January 2014

Abstract

We use administrative records on the incomes of more than 40 million children and their parents to describe three features of intergenerational mobility in the United States. First, we characterize the joint distribution of parent and child income at the national level. The conditional expectation of child income given parent income is linear in percentile ranks. On average, a 10 percentile increase in parent income is associated with a 3.4 percentile increase in a child's income. Second, intergenerational mobility varies substantially across areas within the U.S. For example, the probability that a child reaches the top quintile of the national income distribution starting from a family in the bottom quintile is 4.4% in Charlotte but 12.9% in San Jose. Third, we explore the factors correlated with upward mobility. High mobility areas have (1) less residential segregation, (2) less income inequality, (3) better primary schools, (4) greater social capital, and (5) greater family stability. While our descriptive analysis does not identify the causal mechanisms that determine upward mobility, the new publicly available statistics on intergenerational mobility by area developed here can facilitate future research on such mechanisms.

*The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Internal Revenue Service or the U.S. Treasury Department. This work is a component of a larger project examining the effects of tax expenditures on the budget deficit and economic activity. All results based on tax data in this paper are constructed using statistics originally reported in the SOI Working Paper "The Economic Impacts of Tax Expenditures: Evidence from Spatial Variation across the U.S.," approved under IRS contract TIRNO-12-P-00374 and presented at the National Tax Association meeting on November 22, 2013. We thank David Autor, Gary Becker, David Card, David Dorn, John Friedman, James Heckman, Nathaniel Hilger, Richard Hornbeck, Lawrence Katz, Sara Lalumia, Adam Looney, Jonathan Parker, Gary Solon, Danny Yagan, and numerous seminar participants for helpful comments and discussions. Sarah Abraham, Alex Bell, Shelby Lin, Alex Olssen, Evan Storms, Michael Stepner, and Wentao Xiong provided outstanding research assistance. This research was funded by the National Science Foundation, the Lab for Economic Applications and Policy at Harvard, the Center for Equitable Growth at UC-Berkeley, and Laura and John Arnold Foundation. Publicly available portions of the data and code, including intergenerational mobility statistics by commuting zone and county, are available at www.equality-of-opportunity.org.

I Introduction

The United States is often hailed as the "land of opportunity," a society in which a child's chances of success depend little on her family background. Is this reputation warranted? We show that this question does not have a clear answer because there is substantial variation in intergenerational mobility across areas *within* the U.S. The U.S. is better described as a collection of societies, some of which are "lands of opportunity" with high rates of mobility across generations, and others in which few children escape poverty.

We characterize intergenerational mobility using information from de-identified federal income tax records, which provide data on the incomes of more than 40 million children and their parents between 1996 and 2012. We organize our analysis into three parts. In the first section, we present new statistics on intergenerational mobility in the U.S. as a whole. In our baseline analysis, we focus on current U.S. citizens in the 1980-1982 birth cohorts – the oldest children in our data for whom we can reliably identify parents based on information on dependent claiming. We measure these children's income as mean total family income in 2011 and 2012, when they are approximately 30 years old. We measure their parents' income as mean family income between 1996 and 2000, when the children are between the ages of 15 and 20.

The literature has identified two econometric challenges in estimating intergenerational mobility: lifecycle bias due to measuring income at early or late ages and attenuation bias due to noise in annual measures of income (e.g., Solon 1992, Zimmerman 1992, Mazumder 2005). We show that estimates of intergenerational mobility stabilize when children reach their late twenties. Estimates of mobility are insensitive to the age of parents and children at which parent income is measured, provided that parent income is measured between age 30 and 55. We also show that using several years of data to measure parent and child income does not substantially increase estimates of mobility, perhaps because transitory measurement error is less prevalent in tax records than survey data. These results indicate that our baseline income definitions do not suffer from significant lifecycle or attenuation bias.

We begin our characterization of intergenerational mobility by regressing log child income on log parent income, as in prior work. This specification yields an intergenerational elasticity (IGE) estimate of 0.344, similar to Solon's (1992) preferred estimates based on survey data. Although the log-log specification has desirable theoretical properties (Solon 2004), it suffers from two empirical shortcomings. First, it omits observations with zero income. Since children of low income parents are much more likely to have zero income, dropping these observations overstates mobility. Second, the relationship between log child income and log parent income is highly non-linear, with much lower local IGEs below the 10th percentile and above the 90th percentile. These issues limit the suitability of the log-log specification for cross-area comparisons, as IGE estimates are sensitive to differences in local income distributions.

To address these problems, we develop a rank-rank specification similar to that used by Dahl and DeLeire (2008). We rank children based on their incomes relative to other children in the same birth cohort, including those with zero income. We rank parents of these children based on their incomes relative to other parents with children in these birth cohorts. The relationship between mean child ranks and parent ranks is almost perfectly linear. The slope of the rank-rank relationship is 0.341, i.e. a 10 percentile point increase in parent rank is associated with a 3.41 percentile increase in a child's income rank. Children's college attendance and teenage birth rates are also linearly related to parent income ranks. A 10 percentile point increase in parent income is associated with a 6.7 percentage point (pp) increase in college attendance rates and a 3 pp reduction in teenage birth rates for women.

In the second part of the paper, we use the rank-rank specification to characterize variation in intergenerational mobility across commuting zones (CZs). Commuting zones are geographical aggregations of counties that are similar to metro areas but cover the entire U.S., including rural areas (Tolbert and Sizer 1996). We assign children to commuting zones based on where they lived at age 16 – i.e., where they grew up – irrespective of whether they left that CZ afterward. When analyzing CZs, we continue to rank both children and parents based on their positions in the *national* income distribution, allowing us to measure children's absolute outcomes as we discuss below.

The relationship between mean child ranks and parent ranks remains almost perfectly linear within commuting zones, allowing us to summarize the conditional expectation of a child's rank given his parents' rank with just two parameters: a slope and intercept. The slope measures *relative mobility*: the difference in outcomes between children from top vs. bottom income families within a CZ. The intercept measures the expected rank for children from families at the bottom of the income distribution. Combining the intercept and slope for a CZ, we can calculate the expected rank of children from families at any given percentile p of the national parent income distribution. We term this measure *absolute mobility at percentile p.* Measuring absolute mobility is valuable because increases in relative mobility have ambiguous normative implications, as they may be driven by worse outcomes for the rich rather than better outcomes for the poor.

We find substantial variation in both relative and absolute mobility across CZs. Relative mobility is lowest for children who grew up in the Southeast and highest in the Mountain West and the rural Midwest.[1] Some CZs in the U.S. have relative mobility comparable to the highest mobility countries in the world, such as Denmark, while others have lower levels of mobility than any developed country for which data are available.

We find similar geographical variation in absolute mobility. We focus much of our analysis on absolute mobility at $p = 25$, which we term "absolute upward mobility." This statistic measures the mean income rank of children with parents in the bottom half of the income distribution given linearity of the rank-rank relationship. Absolute upward mobility ranges from 35.8 in Charlotte to 46.2 in Salt Lake City among the 50 largest CZs. Overall, a 1 standard deviation (SD) increase in CZ-level upward mobility is associated with a 0.2 SD improvement in a child's expected rank given parents at $p = 25$, 60% as large as the effect of a 1 SD increase in her own parents' income. Other measures of upward mobility exhibit similar spatial variation. For instance, the probability that a child reaches the top fifth of the income distribution conditional on having parents in the bottom fifth is 4.4% in Charlotte, compared with 10.8% in Salt Lake City and 12.9% in San Jose. The CZ-level mobility statistics are also robust to adjusting for differences in the local cost-of-living, shocks to local growth, and using alternative measures of income.

Absolute upward mobility is highly correlated with relative mobility: areas with high levels of relative mobility (low rank-rank slopes) tend to have better outcomes for children from low-income families. On average, children from families below percentile $p = 85$ have better outcomes when relative mobility is greater; those above $p = 85$ have worse outcomes. An important consequence of this result is that location matters most for children from low income families: the expected rank of children from low-income families varies more across CZs than the expected rank of children from high income families.

The spatial patterns of the gradients of college attendance and teenage birth rates with respect to parent income across CZs are very similar to the pattern in intergenerational income mobility. The fact that much of the spatial variation in childrens's' outcomes emerges before they enter the labor market suggests that the differences in mobility are driven by factors that affect children while they are growing up.

[1] The fact that we define location based on where children grew up is important here. Successful children who grow up in rural areas often work in a different CZ (e.g., a nearby city) as adults.

In the final part of the paper, we explore such factors by correlating the spatial variation in mobility with observable characteristics. We begin by showing that upward income mobility is significantly lower in areas with larger African-American populations. However, white individuals in areas with large African-American populations also have lower rates of upward mobility, implying that racial shares matter at the community (rather than individual) level. One mechanism for such a community-level effect of race is segregation. Areas with larger black populations tend to be more segregated by income and race, which could affect both white and black low-income individuals adversely. Indeed, we find a strong negative correlation between standard measures of racial and income segregation and upward mobility. Moreover, we also find that upward mobility is higher in cities with less sprawl, as measured by commute times to work. These findings lead us to identify segregation as the first of five broad factors that are strongly correlated with mobility.

The second factor we explore is inequality. CZs with larger Gini coefficients have less upward mobility, consistent with the "Great Gatsby curve" documented across countries (Krueger 2012, Corak 2013). In contrast, top 1% income shares are not highly correlated with intergenerational mobility both across CZs within the U.S. and across countries. Although one cannot draw definitive conclusions from such correlations, they suggest that the factors that erode the middle class hamper intergenerational mobility more than the factors that lead to income growth in the upper tail.

Third, proxies for the quality of the K-12 school system are also correlated with mobility. Areas with higher test scores (controlling for income levels), lower dropout rates, and smaller class sizes have higher rates of upward mobility. In addition, areas with higher local tax rates, which are predominantly used to finance public schools, have higher rates of mobility.

Fourth, social capital indices (Putnam 1995) – which are proxies for the strength of social networks and community involvement in an area – are very strongly correlated with mobility. For instance, high upward mobility areas tend to have higher fractions of religious individuals and greater participation in local civic organizations.

Finally, the strongest predictors of upward mobility are measures of family structure such as the fraction of single parents in the area. As with race, parents' marital status does not matter purely through its effects at the individual level. Children of married parents also have higher rates of upward mobility if they live in communities with fewer single parents.

We find modest correlations between upward mobility and local tax and government expenditure policies and no systematic correlation between mobility and local labor market conditions, rates of migration, or access to higher education. In a multiple regression, the five key factors described above generally remain statistically significant predictors of both relative and absolute upward mobility, even in specifications with state fixed effects. However, we emphasize that these factors should not be interpreted as causal determinants of mobility because all of these variables are endogenous and our analysis does not control for numerous other unobserved differences across areas.

Our results build on and contribute to an extensive empirical literature on intergenerational mobility, reviewed by Solon (1999), Grawe and Mulligan (2002), and Black and Devereux (2011). Several studies have compared mobility across countries using a log-log specification and have found that relative mobility is lower in the U.S. than in other developed countries (e.g., Bjorklund and Jäntti 1997, Jäntti et al. 2006, Corak 2013). Our estimates of the IGE in the U.S. as a whole are similar to those found in prior work, with the exception of Mazumder's (2005) widely cited estimates, which imply much lower levels of intergenerational mobility than we find here for reasons we explain in Online Appendix C. Our analysis is most closely related to contemporaneous work by Graham and Sharkey (2013), who use survey data to estimate relative mobility using log-log specifications in a subset of cities in the U.S. Their estimates are correlated with ours, but do not permit an assessment of absolute mobility and are naturally less precise due to limitations in sample size.

Our approach of within-country comparisons offers two advantages over the cross-country comparisons that have been the focus of prior comparative work. First, differences in measurement and econometric methods make it difficult to reach definitive conclusions from cross-country comparisons (Solon 2002, Black and Devereux 2011). The income measures and covariates we analyze here are all measured using the same data sources across all CZs. Second, and more importantly, we can characterize both relative and absolute mobility across CZs by using national ranks to measure children's outcomes. The cross-country literature has focused on differences in relative mobility, partly because it is difficult to compare the absolute standard of living across very different economies. Although the literature on cross-country differences in economic growth has characterized differences in mean absolute living standards across nations, much less is known about how the prospects of children from low-income families vary across countries when measured on a common absolute scale (Ray 2010).

Our analysis also relates to the literature on neighborhood effects, reviewed by Jencks and Mayer (1990) and Sampson et al. (2002). Many of the correlations we explore are motivated by hypotheses proposed in this literature, such as the impacts of concentrated poverty (Wilson 1987), residential segregation (Massey and Denton 1993, Cutler and Glaeser 1997), social capital (Coleman 1988, Putnam 1995), and local school quality (Card and Krueger 1992). However, unlike recent experimental and quasi-experimental work on neighborhood effects (e.g., Katz et al. 2001, Oreopoulos 2003), our descriptive analysis does not shed light on whether the differences in outcomes across areas are due to the causal effect of neighborhoods or differences in the characteristics of people living in those neighborhoods.

The remainder of the paper is organized as follows. Section II describes the data. Section III reports estimates of intergenerational mobility at the national level. In Section IV, we present estimates of absolute and relative mobility by commuting zone. Section V reports correlations of our mobility measures with observable characteristics of commuting zones. Section VI concludes. Statistics on intergenerational mobility and related covariates are publicly available by commuting zone, metropolitan statistical area, and county on the project website.

II Data

We use data from federal income tax records spanning 1996-2012. The data include both income tax returns (1040 forms) and third-party information returns (e.g., W-2 forms), which give us information on the earnings of those who do not file tax returns. We provide a detailed description of how we construct our analysis sample starting from the raw population data in Online Appendix A. Here, we briefly summarize the key variable and sample definitions. Note that in what follows, the year always refers to the tax year (i.e., the calendar year in which the income is earned).

II.A Sample Definitions

Our base dataset of children consists of all individuals who (1) have a valid Social Security Number or Individual Taxpayer Identification Number, (2) were born between 1980-1991, and (3) are U.S. citizens as of 2013. We impose the citizenship requirement to exclude individuals who are likely to have immigrated to the U.S. as adults, for whom we cannot measure parent income. We cannot directly restrict the sample to individuals born in the U.S. because the database only records current citizenship status.

We identify the parents of a child as the first tax filers (between 1996-2012) who claim the child as a child dependent and were between the ages of 15 and 40 when the child was born. If the child is first claimed by a single filer, the child is defined as having a single parent. For simplicity, we assign each child a parent (or parents) permanently using this algorithm, regardless of any subsequent changes in parents' marital status or dependent claiming.[2]

If parents never file a tax return, we cannot link them to their child. Although some low-income individuals do not file tax returns in a given year, almost all parents file a tax return at some point between 1996 and 2012 to obtain a tax refund on their withheld taxes and the Earned Income Tax Credit (Cilke 1998). As a result, we are able to identify parents for approximately 95% of the children in the 1980-1991 birth cohorts.[3] Because we have more opportunities to link younger children to their parents, the fraction of children linked to parents rises from approximately 90% for the early birth cohorts to nearly 99% for the most recent birth cohorts (Online Appendix Table I).[4] The fraction of children linked to parents drops sharply prior to the 1980 birth cohort because our data begin in 1996 and many children begin to the leave the household starting at age 17. This is why we limit our analysis to children born during or after 1980.

Our primary analysis sample, which we refer to as the *core sample*, includes all children in the base dataset who (1) are born in the 1980-82 birth cohorts, (2) for whom we are able to identify parents, and (3) whose mean parent income between 1996-2000 is strictly positive (which excludes 1.2% of children).[5] For some robustness checks, we use the *extended sample*, which imposes the same restrictions as the core sample, but includes all birth cohorts 1980-1991. There are approximately 10 million children in the core sample and 44 million children in the extended sample.

Statistics of Income Sample. Because we can only reliably link children to parents starting with the 1980 birth cohort in the population tax data, we can only measure earnings of children up to age 32 (in 2012) in the full sample. To evaluate whether estimates of intergenerational mobility would change significantly if earnings were measured at later ages, we supplement our analysis using annual cross-sections of tax returns maintained by the Statistics of Income (SOI) division of the Internal Revenue Service prior to 1996. The SOI cross-sections provide identifiers for dependents claimed on tax forms starting in 1987, allowing us to link parents to children back to the 1971 birth cohort using an algorithm analogous to that described above (see Online Appendix

[2] 12% of children in our core sample are claimed as dependents by different individuals in subsequent years. To ensure that this potential measurement error in linking children to parents does not affect our findings, we show that we obtain similar estimates of mobility for the subset of children who are never claimed by other individuals (row 9 of Table V).

[3] Chetty et al. (2013) present further evidence that one can identify parents for virtually all children who grew up in the U.S. by showing that 98% of children enrolled in a large school district in grades 3-8 can be linked to parents in the tax data.

[4] To ensure that our results are not biased by the missing data in the early cohorts, we also replicate our analysis restricting the sample to more recent birth cohorts (see rows 5 and 6 of Table V).

[5] We limit the sample to parents with positive income because parents who file a tax return (as required to link them to a child) yet have zero income are unlikely to be representative of individuals with zero income and those with negative income typically have large capital losses, which are a proxy for having significant wealth.

A for further details). The SOI cross-sections are stratified random samples of tax returns with a sampling probability that rises with income; using sampling weights, we can calculate statistics representative of the national distribution. After linking parents to children in the SOI sample, we use population tax data to obtain data on income for children and parents, using the same definitions as in the core sample. There are approximately 63,000 children in the 1971-79 birth cohorts in the SOI sample (Online Appendix Table II).

II.B Variable Definitions and Summary Statistics

In this section, we define the key variables we use to measure intergenerational mobility. We measure all monetary variables in 2012 dollars, adjusting for inflation using the consumer price index (CPI-U).

Parent Income. Following Lee and Solon (2009), our primary measure of parent income is total pre-tax income at the household level, which we label *parent family income.* More precisely, in years where a parent files a tax return, we define family income as Adjusted Gross Income (as reported on the 1040 tax return) plus tax-exempt interest income and the non-taxable portion of Social Security and Disability benefits. In years where a parent does not file a tax return, we define family income as the sum of wage earnings (reported on form W-2), unemployment benefits (reported on form 1099-G), and gross social security and disability benefits (reported on form SSA-1099) for both parents.[6] In years where parents have no tax return and no information returns, family income is coded as zero.

Note that this income measure includes labor earnings and capital income as well as unemployment insurance, social security, and disability benefits. It excludes non-taxable cash transfers such as TANF and SSI, in-kind benefits such as food stamps, all refundable tax credits such as the EITC, non-taxable pension contributions (e.g., to 401(k)'s), and any earned income not reported to the IRS. Income is always measured prior to the deduction of individual income taxes and employee-level payroll taxes.

In our baseline analysis, we average parents' family income over the five years from 1996 to 2000 to reduce noise due to transitory fluctuations (Solon 1992). We use the earliest years in our sample to best reflect the economic resources of parents while the children in our sample are growing up.[7]

[6] The database does not record W-2's and other information returns prior to 1999, so non-filer's income is coded as 0 prior to 1999. We verify that this is not an important source of bias by showing in Table V (row 18) that we obtain very similar results when defining parent income only using post-1999 data. We cannot observe self-employment income for non-filers and therefore code it as zero; given the strong incentives for individuals with children to file created by the EITC, most non-filers likely have very low levels of self-employment income.

[7] Formally, we define mean family income as the mother's family income plus the father's family income in each

We evaluate the robustness of our findings using a measure of individual parent income instead of family income. For single parents, parent family income coincides with individual income. For married parents, we define each parent's individual earnings as the sum of wage earnings from form W-2, unemployment benefits from form 1099-G, and Social Security and Disability benefits from form SSA-1099 for that individual. Individual earnings excludes capital and other non-labor income. To incorporate these sources of income, we add half of family non-labor income – defined as total family income minus total family earnings reported on form 1040 – to each parent's individual earnings. We divide non-labor earnings equally between spouses because we cannot identify which spouse earns non-labor income from the 1040 tax return.

Child Income. We define child family income in exactly the same way as parent family income. In our baseline analysis, we average child family income over the last two years in our data (2011 and 2012). We use the most recent two years because the children in our sample are all born after 1980, and income in the early 30's provides a better measure of lifetime income than income at earlier ages (Haider and Solon 2006). We report results using alternative years to assess the sensitivity of our findings. For children, we define household income based on current marital status rather than marital status at a fixed point in time. Because family income varies with marital status, we also report results using individual income measures for children, constructed in exactly the same way as for parents.

College Attendance. We define college attendance as an indicator for having one or more 1098-T forms filed on one's behalf when the individual is aged 18-21. Title IV institutions – all colleges and universities as well as vocational schools and other post-secondary institutions eligible for federal student aid – are required to file 1098-T forms that report tuition payments or scholarships received for every student. Because the 1098-T forms are filed directly by colleges independent of whether an individual files a tax return, we have complete records on college attendance for all children. The 1098-T data are available from 1999-2012. Comparisons to other data sources indicate that 1098-T forms capture college enrollment quite accurately overall (Chetty, Friedman, and Rockoff 2013, Section 3.1).[8]

year from 1996 to 2000 divided by 10 (or divided by 5 if we only identify a single parent). For parents who do not change marital status, this is simply mean family income over the 5 year period. For parents who are married initially and then divorce, this measure tracks the mean family incomes of the two divorced parents over time. For parents who are single initially and then get married, this measure tracks individual income prior to marriage and total family income (including the new spouse's income) after marriage. These household measures of income increase with marriage and naturally do not account for cohabitation; to ensure that these features do not generate bias, we assess the robustness of our results to using individual measures of income.

[8]Colleges are not required to file 1098-T forms for students whose qualified tuition and related expenses are waived or paid entirely with scholarships or grants. However, the forms are frequently available even for such cases,

Teenage Birth. We define a woman as having a teenage birth if she ever claims a dependent who was born while she was between the ages of 13 and 19. This measure is an imperfect proxy for having a teenage birth because it only covers children who are claimed as dependents by their mothers. Fortunately, the aggregate level and spatial pattern of teenage births in our data are closely aligned with estimates based on the American Community Survey.[9]

Summary Statistics. Table I reports summary statistics for the core sample. Median parent family income is $60,129 (in 2012 dollars). Among the 30.6% of children matched to single parents, 72.0% are matched to a female parent. Children in our core sample have a median family income of $34,975 when they are approximately 30 years old. 6.1% of children have zero income in both 2011 and 2012. 58.9% are enrolled in a college at some point between the ages of 18 and 21 and 15.9% of women have a teenage birth.

In Online Appendix Table III, we show that the total cohort size, labor force participation rate, distribution of child income, and other demographic characteristics of our core sample line up closely with corresponding estimates in the Current Population Survey and American Community Survey. This confirms that our sample covers roughly the same nationally representative population as previous survey-based research.

III National Statistics

We begin our empirical analysis by characterizing the relationship between parent and child income at the national level. We first present a set of baseline estimates of intergenerational income mobility and then evaluate the robustness of our estimates to alternative sample and income definitions.

III.A Baseline Estimates

In our baseline analysis, we use the core sample (1980-82 birth cohorts) and measure parent income as mean family income from 1996-2000 and child income as mean family income in 2011-12, when children are approximately 30 years old. Figure Ia presents a binned scatter plot of the mean family income of children versus the mean family income of their parents. To construct this figure, we divide the x axis into 100 equal-sized (percentile) bins and plot mean child income vs. mean

presumably because of automated reporting to the IRS by universities. Approximately 6% of 1098-T forms are missing from 2000-2003 because the database contains no 1098-T forms for some small colleges in these years. To verify that this does not affect our results, we confirm that our estimates of college attendance by parent income gradients are very similar for later birth cohorts (not reported).

[9] 15.8% of women in our core sample have teenage births; the corresponding number is 14.6% in the 2003 ACS. The unweighted correlation between state-level teenage birth rates in the tax data and the ACS is 0.80.

parent income in each bin. For scaling purposes, we exclude the top bin (children of parents in the top 1%) in this figure only. This binned scatter plot provides a non-parametric representation of the conditional expectation of child income given parent income. The regression coefficients and standard errors reported in this and all subsequent binned scatter plots are estimated on the underlying microdata using OLS regressions.

The conditional expectation of children's' income given parents' income is strongly concave. Below the 90th percentile of parent income, a $1 increase in parent family income is associated with a 33.5 cent increase in average child family income. In contrast, between the 90th and 99th percentile, a $1 increase in parent income is associated with only a 7.6 cent increase in child income.

Partly motivated by this non-linearity, much of the empirical literature has estimated regressions of log child income on log parent income. The slope of this regression measures the elasticity of child income with respect to parent income, commonly termed the intergenerational income elasticity (IGE). We implement this specification in the first column of row 1 of Table II, excluding children with zero income as in prior work. We obtain an IGE estimate of 0.344, similar to the estimates of Solon (1992) and Zimmerman (1992) in survey data when using multi-year income averages.

The remaining columns in the first row of Table II replicate the log-log specification for alternative samples analyzed in the prior literature. Columns 2-5 split the sample by the child's gender and the parents' marital status in the year they first claim the child. Column 6 replicates Column 1 for the extended sample of 1980-85 birth cohorts. The IGE estimates are similar for males and females, but are lower when we condition on marital status and use the extended sample for reasons we explain below in Section III.B.

While the log-log specification is a familiar and intuitive benchmark, it suffers from two shortcomings that are illustrated in Figure Ib. First, the relationship between log child income and log parent income is highly non-linear, consistent with the findings of Corak and Heisz (1999) in Canadian tax data. This is illustrated in the series in circles in Figure Ib, which plots mean log child income vs. mean log family income by percentile bin, constructed using the same method as Figure Ia. Because of this non-linearity, the IGE is sensitive to the point of measurement in the income distribution. For example, restricting the sample to observations between the 10th and 90th percentile of parent income (denoted by the vertical dashed lines in the graph) yields a considerably higher IGE estimate of 0.452.

Second, the log-log specification discards observations with zero income. The series in triangles in Figure Ib plots the fraction of children with zero income by parental income bin. This fraction varies from 18% among the poorest families to 3% among the richest families. Dropping children with zero income therefore overstates the degree of intergenerational mobility. The way in which these zeros are treated can change the IGE dramatically. For instance, including the zeros by assigning those with zero income an income of $1 (so that the log of their income is zero) raises the estimated IGE to 0.618, as shown in row 2 of Table II. If instead we treat those with 0 income as having an income of $1,000, the estimated IGE becomes 0.413. Hence, small differences in the way children's income is measured at the bottom of the distribution can produce substantial variation in IGE estimates.

To address these shortcomings of the log-log specification, we use a *rank-rank* specification similar to that proposed by Dahl and DeLeire (2008). We measure the percentile rank of the parent based on their position in the distribution of parent incomes in the core sample. Similarly, we define children's percentile ranks based on their positions in the distribution of child incomes within their birth cohorts. Importantly, this definition allows us to include zeros in child income.[10] Unless otherwise noted, we hold the definition of these ranks fixed based on positions in the aggregate distribution, even when analyzing subgroups.

Figure IIa presents a binned scatter plot of the mean percentile rank of children vs. their parents' percentile rank. The conditional expectation of a child's rank given her parents' rank is almost perfectly linear. Using an OLS regression, we estimate that a one percentage point (pp) increase in parent rank is associated with a 0.341 pp increase in the child's expected rank, as reported in row 4 of Table II.

Figure IIb compares the rank-rank relationship in the U.S. with analogous estimates for Denmark constructed using data from Boserup et al. (2013).[11] The relationship between child and parent ranks is nearly linear in Denmark as well, suggesting that the rank-rank specification can provide a good parametric summary of mobility across diverse environments. The rank-rank slope in Denmark is 0.180, nearly half that in the U.S. This corroborates prior findings that Scandinavian economies have much greater relative intergenerational mobility than the United States.

Importantly, the smaller rank-rank slope in Denmark does not necessarily mean that children from low-income families in Denmark do better than those in the U.S. in absolute terms. It could be

[10]In the case of ties, we define the rank as the mean rank for the individuals in that group. For example, if 10% of a birth cohort has zero income, all children with zero income would receive a percentile rank of 5.

[11]The Danish sample uses the 1980-81 birth cohorts and measures child income based on mean income between 2009-11. Because of differences in the structure of the administrative database, child income in the Danish sample is measured at the individual level and parents' income is the mean of the two biological parents' income from 1997-1999, irrespective of their marital status.

that children of high-income parents in Denmark have worse outcomes than children of high-income parents in the U.S., in which case the greater relative mobility in Denmark may be undesirable. One cannot distinguish between these possibilities based on Figure IIb because the ranks are defined within each country. One advantage of the within-U.S. CZ-level analysis implemented below is that it naturally allows us to study both relative and absolute outcomes by analyzing children's performance on a fixed national scale.

Transition Matrices. Although the rank-rank relationship has attractive statistical properties, there are many other measures of mobility that may be of normative interest (see e.g., Fields and Ok 1999). For example, one popular approach is to analyze transition matrices (e.g., Corak and Heisz 1999, Hertz 2006, Jäntti et al. 2006). Table III presents a quintile transition matrix: the probability that a child is in quintile m of the child income distribution conditional on his parent being in quintile n of the parent income distribution. One statistic of particular interest in this matrix is the probability of moving from the bottom quintile to the top quintile, a simple measure of "success" that we analyze below. This probability is 7.5% in the U.S. as a whole. Another notable feature of the matrix is that the rate of persistence in the bottom quintile (33.7%) is similar to the rate of persistence in the top quintile (36.5%).

To facilitate the construction of other measures of mobility, we report a 100 x 100 percentile-level transition matrix for the U.S. in Online Data Table I. Using this matrix – which characterizes the copula of the joint distribution of parent and child income – and the marginal distributions for child and parent income reported in Online Data Table II, one can construct any mobility statistic of interest for the U.S. population.

III.B Robustness of Baseline Estimates

We now evaluate the robustness of our estimates of the degree of intergenerational persistence in income to alternative specifications. We begin by evaluating two potential sources of bias emphasized in prior work: lifecycle bias and attenuation bias.

Lifecycle Bias. Prior research has shown that measuring children's income at early ages can understate intergenerational persistence in lifetime income because children with high lifetime incomes are in college and have steeper earnings profiles when they are young (Haider and Solon, 2006, Grawe, 2006, Solon 1999). To evaluate whether our baseline estimates suffer from such lifecycle bias, Figure IIIa plots estimates of the rank-rank slope by the age at which the child's income is measured. To construct this figure, we measure children's income as mean family income in 2011-

2012 and parent income as mean family income between 1996-2000, as in our baseline analysis. We then replicate the OLS regression of child income rank on parent income rank for each birth cohort between 1980-1990. For children in the 1980 birth cohort, we measure earnings in 2011-12 at age 31-32 (denoted by 32 in the figure); for the 1990 cohort, we measure earnings at age 21-22.[12] The rank-rank slope rises very steeply in the early 20's as children enter the labor force, but stabilizes around age 30. It increases by 2.1% from age 30 to 31 and 0.2% from age 31 to 32.

To obtain estimates beyond age 32, we use the SOI 0.1% random sample described in Section II.A, which contains data back to the 1971 birth cohort. The series in triangles in Figure IIIa replicates the analysis above within the SOI sample, using sampling weights to recover estimates representative of the population. The estimates in the SOI sample are very similar to those in the full population prior to age 32. After age 32, the estimates remain roughly constant. These findings indicate that rank-rank correlations exhibit little lifecycle bias provided that child income is measured after age 30, as in our baseline definition. Estimates of the IGE using the traditional log-log specification also stabilize around age 30 (not reported).

An analogous lifecycle bias can arise if parent income is measured at very old or young ages. In Online Appendix Figure Ia we plot the rank-rank slope using the core sample, varying the 5-year window used to measure parent income from a starting year of 1996 (when parents are 41 years old on average) to 2007 (when parents are 55 years old). The rank-rank estimates exhibit virtually no variation with the age of parent income measurement within this range.

A closely related concern is that parent income at earlier ages might matter more for children's outcomes, e.g. if resources in early childhood are relevant for child development (e.g., Heckman 2006, Duncan et al. 2010). While we cannot measure parent income before age 14 for children in our core sample, we can measure parent income at earlier ages for later birth cohorts. In Chetty et al. (2014), we use data from the 1993 birth cohort and regress an indicator for college attendance at age 19 on parent income rank in each year from 1996 to 2012. We reproduce the coefficients from those regressions in Online Appendix Figure Ib. The relationship between college attendance rates and parent income rank is virtually constant when children are between ages 3 and 19. Once again, this result indicates that the point at which parent income is measured (provided parents are between ages 30-55) does not significantly affect intergenerational associations, at least in administrative

[12] We vary birth cohort and hold the year of income measurement fixed to eliminate calendar year effects. We obtain very similar results if we instead track a single cohort and vary age by measuring earnings in different calendar years.

earnings records.[13]

Attenuation Bias. Income in a single year is a noisy measure of permanent income because of transitory shocks and measurement error. Solon (1992), Zimmerman (1992), and Mazumder (2005) show that using multi-year means of parent income generate significantly higher estimates of intergenerational persistence in survey data. To evaluate whether our baseline estimates suffer from such attenuation bias, Figure IIIb plots estimates of the rank-rank slope, varying the number of years used to calculate mean parent family income. To construct this figure, we measure children's income as mean family income in 2011-2012 and use the core sample of 1980-82 birth cohorts. We then replicate the OLS regression of child rank on parent rank, varying the number of years used to calculate mean parent income from one (1996 only) to 17 (1996-2012).

Consistent with the findings of Solon (1992), we find that the rank-rank slope rises when we increase the number of years used to measure parent income from one to five. However, the rank-rank slope based on five years (0.341) is only 6.5% larger than the slope based on one year of parent income (0.320). This 6.5% attenuation bias is considerably smaller than the 33% change in the IGE (from 0.3 to 0.4) reported by Solon (1992) when using a five-year average instead of one year of data. We find less attenuation bias for three reasons: (1) income is measured with less error in the tax data than in the PSID, (2) we use family income measures rather than individual income, which fluctuates more across years, and (3) we use a rank-rank specification rather than a log-log specification, which is more sensitive to income fluctuations at the bottom of the distribution.

Contrary to the findings of Mazumder (2005), the rank-rank slope is virtually unchanged by adding more years of data beyond five years: the estimated slope using 15 years of data to measure parent income (0.350) is only 2.8% larger than the baseline slope of 0.341 using 5 years of data. We believe our results differ because we directly measure parent income, whereas Mazumder imputes parent income based on race and education for up to 60% of the observations in his sample (see Online Appendix C for further details).

Prior studies have focused on measurement error in parent income rather than children's income because only the former generates attenuation bias in the standard OLS log-log regression specification, insofar as the transitory measurement error in child income is unrelated to parent income. This is not true in a rank-rank specification because the measurement error in children's

[13] While we cannot measure income before the year in which children turn 3, the fact that the college-income gradient is not declining from ages 3-19 makes it unlikely that the gradient is significantly larger prior to age 2. Parent income ranks in year t have a correlation of 0.91 with parent income ranks in year $t+1$, 0.77 in year $t+5$, and 0.65 in year $t+15$. The decay in this autocorrelation would generate a decreasing slope in the gradient in Online Appendix Figure Ib if there were a discontinuous jump in the gradient prior to age 2.

income generates misclassification error in their ranks, which can attenuate the rank-rank slope.[14] To evaluate the magnitude of this bias, we analyze the impact of varying the number of years used to measure the child's income in Online Appendix Figure Ic. The rank-rank slope increases only modestly when increasing the number of years used to compute child family income, with no detectable change once one averages over at least two years, as in our baseline measure.[15]

Alternative Income Definitions. In rows 5-7 of Table II, we explore the robustness of the baseline rank-rank estimate to alternative definitions of child and parent income. In row 5 we define the parent's rank based on the individual income of the parent with higher mean income from 1999-2003.[16] This specification eliminates the mechanical variation in family income driven by the number of parents in the household, which could overstate the persistence of income across generations if parent marital status has a direct effect of children's outcomes. The rank-rank correlation falls by approximately 10%, from 0.341 to 0.312 when we use top parent income. The impact of using individual parent income instead of family income is modest because (1) most of the variation in parent income across households is not due to differences in marital status and (2) the mean ranks of children with married parents are only 4.6 percentile points higher than those with single parents. The same logic explains why we find a similar 10% reduction when we condition on marital status in the baseline specification using family income (columns 4 and 5 of row 4).

In row 6 of Table II, we repeat this exercise for children. Here, the concern is that children of higher income parents may be more likely to marry, again exaggerating the observed persistence in family income relative to individual income. Using individual income to measure the child's rank has differential impacts by the child's gender, consistent with Chadwick and Solon (2002). For male children, using individual income instead of family income reduces the rank-rank correlation from 0.336 in the baseline specification to 0.317, a 6% reduction. For female children, using individual income reduces the rank-rank correlation from 0.346 to 0.257, a 26% reduction. These differences are likely driven by assortative mating: women with higher income parents marry men who earn more, driving up the persistence of family income across generations.

[14] Formally, measurement error in the child's rank is correlated with the parent's rank because the support of the child's rank is bounded. Intuitively, children from the highest-ranked families are more likely to be under-ranked than over-ranked due to measurement error in their income. As a result, measurement error in children's or parent's income generally leads to attenuation bias in a rank-rank specification.

[15] An ancillary implication of this result is that our estimates of intergenerational mobility are not sensitive to the specific calendar years used to measure child income.

[16] We use 1999-2003 income here because we cannot allocate earnings across spouses before 1999, as W-2 forms are available starting only in 1999. Note that top income rank differs from family income rank even for single parents because some individuals get married in subsequent years and because these individuals are ranked relative to the population, not relative to other single individuals.

Finally, in row 7 of Table II, we define a measure of child income that excludes capital and other non-labor income using the sum of individual wage earnings, UI benefits, SSDI benefits, and Schedule C self-employment income. We divide self-employment income by 2 for married individuals. This individual earnings measure yields virtually identical estimates of the rank-rank correlation.

We conclude that our baseline measure of intergenerational persistence accurately captures the degree of persistence in lifetime income across generations and focus primarily on this measure in the remainder of the paper.

III.C College Attendance and Teenage Birth Gradients

We supplement our analysis of intergenerational income mobility by studying the relationship between parent income and two additional outcomes for children: college attendance and teenage birth. Figure IV presents binned scatter plots of the college attendance rate of children (in Panel A) and the teenage birth rate for female children (in Panel B) vs. the percentile rank of parent family income using the core sample. Parent rank is defined as in the baseline specification in row 4 of Table II. College attendance is defined as attending college in one or more years between the ages 18 and 21, while teenage birth is defined (for females only) as having a child when the mother is aged 13-19; see section II.B for further details.

The relationships between both outcomes and parental income rank are again virtually linear. The slope for college attendance is 0.675. That is, moving from the lowest-income to highest-income parents increases the college attendance rate by 67.5 percentage points, similar to the estimates reported by Bailey and Dynarski (2011) using survey data. The slope of the relationship between teenage birth rates and parental income rank is -0.30. These substantial correlations suggest that much of the divergence in outcomes between children from low vs. high income families emerges well before they enter the labor market, a point we return to below when exploring spatial variation.

IV Spatial Variation in Mobility

We now turn to our central goal of characterizing the variation in intergenerational mobility across areas within the U.S. We begin by defining measures of geographic location. Next, we define two concepts of intergenerational mobility – relative mobility and absolute mobility – that we measure using a rank-rank specification. Finally, we present estimates of relative and absolute mobility by area and assess the robustness of these estimates to alternative measures.

IV.A Geographical Units

To characterize the variation in children's outcomes across areas, one must first partition the U.S. into a set of geographical areas in which children grow up. One way to conceptualize the choice of a geographical partition is using a hierarchical model in which children's outcomes depend upon conditions in their immediate neighborhood (e.g., peers or resources in their city block), local community (e.g., the quality of schools in their county), and broader metro area (e.g., local labor market conditions). To fully characterize the geography of intergenerational mobility, one would ideally estimate all of the components of such a hierarchical model.

As a first step toward this goal, we begin with a coarse partition and characterize intergenerational mobility at the level of commuting zones (CZs). CZs are aggregations of counties based on commuting patterns in the 1990 Census constructed by Tolbert and Sizer (1996) and introduced to the economics literature by Dorn (2009). Since CZs are designed to fully span the area in which people live and work, they provide a natural starting point as the coarsest partition of areas. CZs are similar to metropolitan statistical areas (MSA), but unlike MSAs, they cover the entire U.S. (including rural areas). There are 741 CZs in the U.S.; on average, each CZ contains 4 counties and has a population of 379,786.[17]

We focus on CZ-level variation both for parsimony and because mobility statistics in narrower neighborhoods may be more heavily affected by sorting within metro areas. Because property prices are typically homogeneous within narrow areas and home values are highly correlated with parent income, comparisons of individuals within a narrow neighborhood effectively condition on a proxy for parent income. As a result, the variation in parent income across individuals in a narrow area (such as a city block) must be correlated with other latent factors that may affect children's' outcomes directly, making it difficult to interpret the resulting mobility estimates.[18] Nevertheless, to obtain some insight into within-CZ variation, we also report statistics on intergenerational mobility by county in Online Data Table III. There is almost as much variance in intergenerational mobility across counties within a CZ as there is across CZs, suggesting that the total amount of geographical

[17] See Online Appendix Figure II for an example of the Boston CZ. Note that CZs in urban areas generally have much higher populations than rural CZs. To account for this variation, we always report statistics that restrict to urban areas or use population-weights in addition to unweighted measures that pool all CZs.

[18] For example, it would be difficult to assess intergenerational mobility within midtown Manhattan because there are very few low-income individuals within this homogeneously high-property-value area, and any families with low observed income in such an area would have to be latently wealthy to be able to afford to live there. Although the cross-CZ differences we document could certainly also be driven by differences in latent characteristics across individuals, we believe that understanding why mobility differs across these broader areas of the U.S. is of greater descriptive interest than understanding variation across much narrower areas that is likely to be mechanically related to differences in property values.

variation may be even greater than that documented below.[19] We defer further analysis of such within-CZ heterogeneity to future research.

We *permanently* assign each child to a single CZ based on the ZIP code from which his or her parent filed their tax return in the first year the child was claimed as a dependent. We interpret this CZ as the area where a child grew up. Because our data begin in 1996, location is measured in 1996 for 95.9% of children in our core sample. For children in our core sample of 1980-82 birth cohorts, we therefore typically measure location when children were approximately 15 years old. For the children in the more recent birth cohorts in our extended sample, location is measured at earlier ages. Using these more recent cohorts, we find that 83.5% of children live in the same CZ at age 16 as they did at age 5. Furthermore, we verify that the spatial patterns for the outcomes we can measure at earlier ages (college attendance and teenage birth) are quite similar if we define CZs based on location at age 5 instead of age 16.

Importantly, the CZ where a child grew up does not necessarily correspond to the CZ she lives in as an adult when we measure her income (at age 30) in 2011-12. In our core sample, 38% of children live in a different CZ in 2012 relative to where they grew up.

IV.B Measures of Relative and Absolute Mobility

In our baseline analysis, we measure mobility at the CZ level using the core sample (1980-82 birth cohorts) and the definitions of parent and child family income described in Section III.A. Importantly, we continue to rank both children and parents based on their positions in the *national* income distribution (rather than the distribution within their CZ), using exactly the same ranks as in Figure IIa.

We begin by examining the rank-rank relationship in selected CZs. Figure Va presents a binned scatter plot of the mean child rank vs. parent rank for children who grew up in the Salt Lake City, UT (circles) or Charlotte, NC (triangles) commuting zones. The rank-rank relationship is virtually linear in both of these CZs, as at the national level. This linearity of the rank-rank relationship is a remarkably robust property across CZs, as illustrated for the 20 largest CZs in Online Appendix Figure III.

Exploiting this approximate linearity, we summarize the conditional expectation of a child's

[19] The correlation between population-weighted CZ-level means of the county-level mobility measures with the CZ-level estimates of mobility exceeds 0.98, indicating that our approach does not suffer from aggregation bias. To further assess the robustness of our results to the geographical partition, we report statistics by MSA in Online Data Table IV. For CZs that intersect MSAs, the correlation between CZ-level and MSA-level mobility statistics exceeds 0.9.

rank given his parents' rank in each CZ using two parameters: a slope and an intercept. Formally, let y_{ic} denote the national income rank (among children in her birth cohort) of child i who grew up in CZ c. Similarly, let x_{ic} denote her parent's rank in the income distribution of parents in the core sample. For each CZ c, we estimate a slope and intercept using an OLS regression of child rank on parent rank in the microdata:

$$y_{ic} = \alpha_c + \beta_c x_{ic} + \varepsilon_{ic} \tag{1}$$

Using the linear approximation to the rank-rank relationship, let

$$\bar{y}_{pc} = \alpha_c + \beta_c p \tag{2}$$

denote the expected rank of a child whose parents' national income rank is p in CZ c.[20]

One way to measure intergenerational mobility is to ask, "What are the outcomes of children from low-income families relative to those of children from high-income families?" This question has been the focus of most prior research on intergenerational mobility (Solon 1999, Black et al. 2011).[21] To answer this question, we define the degree of *relative mobility* in CZ c as β_c, the slope of the rank-rank relationship. The difference between the expected ranks of children born to parents at the top and bottom of the income distribution is $\bar{y}_{100,c} - \bar{y}_{0,c} = 100 \times \beta_c$.

In Salt Lake City, $\beta_c = 0.264$. The expected rank of children born to the richest parents is $\bar{y}_{100} - \bar{y}_0 = 26.4$ percentiles above that of children born to the poorest parents. Charlotte exhibits much less relative mobility (i.e., much greater persistence of income across generations). In Charlotte, $\bar{y}_{100} - \bar{y}_0 = 39.7$.

A different way to measure intergenerational mobility is to ask, "What are the outcomes of children from families of a given income level in absolute terms?" We define *absolute mobility at percentile* p in CZ c as $\bar{y}_{pc}$ in (2). Given academic and public interest in the outcomes of disadvantaged youth, we focus on average absolute mobility for children from families with below-median parent income ($E\left[y_{ic}|x_{ic} < p = 50, c\right]$), which we term *absolute upward mobility.*[22] Because the rank-rank relationship is linear, the average rank of children with below-median parent income

[20] We always measure percentile ranks on a 0-100 scale and slopes on a 0-1 scale, so α_c ranges from 0-100 and β_c ranges from 0 to 1 in (2).

[21] Any measure of the extent to which parental income predicts children's' outcomes in a CZ is a measure of relative mobility. In this sense, the traditional log-log IGE is a relative mobility measure, as it measures the difference in (log) outcomes between children of high vs. low income parents.

[22] We focus on the absolute outcomes of children from low-income families both in the interest of space and because there is more variation across areas in the outcomes of children from low-income families than those from high-income families, as we show in Figure VII below. However, the CZ-level statistics in Online Data Tables V and VI can be used to analyze spatial variation in the outcomes of children from high-income families in future research.

equals the average rank of children with parents at the 25th percentile ($\bar{y}_{25,c} = \alpha_c + 25\beta_c$), illustrated by the dashed vertical line in Figure Va.

Absolute upward mobility is $\bar{y}_{25} = 46.2$ in Salt Lake City, compared with $\bar{y}_{25} = 35.8$ in Charlotte. That is, among families earning $28,800 – the 25th percentile of the national parent family income distribution – children who grew up in Salt Lake City are on average 10 percentile points higher in their birth cohort's income distribution at age 30 than children who grew up in Charlotte.

Absolute mobility is higher in Salt Lake City not just for below-median families, but at all percentiles p of the parent income distribution. The gap in absolute outcomes is largest at the bottom of the income distribution and nearly zero at the top. Hence, the greater relative mobility in this particular comparison comes purely from better absolute outcomes at the bottom of the distribution rather than worse outcomes at the top. Of course, this is not always the case. Figure Vb shows that San Francisco has substantially higher relative mobility than Chicago: $\bar{y}_{100} - \bar{y}_0 = 25.0$ in San Francisco vs. $\bar{y}_{100} - \bar{y}_0 = 39.3$ in Chicago. But part of the greater relative mobility in San Francisco comes from worse outcomes for children from high-income families. Below the 60th percentile, children in San Francisco have better outcomes than those in Chicago; above the 60th percentile, the reverse is true.

The comparisons in Figure V illustrate the importance of measuring both relative and absolute mobility. Any social welfare function based on mean income ranks that respects the Pareto principle would rate Salt Lake City above Charlotte. But normative comparisons of San Francisco and Chicago depend on the weight one puts on relative vs. absolute mobility (or, equivalently, on the weights one places on absolute mobility at each percentile p).

IV.C Baseline Estimates by CZ

We estimate (1) using OLS to calculate absolute upward mobility ($\bar{y}_{25,c} = \alpha_c + 25\beta_c$) and relative mobility (β_c) by CZ. The estimates for each CZ are reported in Online Data Table V.

Absolute Upward Mobility. Figure VIa presents a heat map of absolute upward mobility. We construct this map by dividing CZs into deciles based on their estimated value of $\bar{y}_{25,c}$. Lighter colors represent deciles with higher levels of $\bar{y}_{25,c}$.[23] Upward mobility varies significantly across

[23] We cannot estimate mobility for 32 CZs in which we have fewer than 250 children in the core sample, shown by the cross-hatched areas in the maps in Figure VI. These CZs account for less than 0.05% of the U.S. population in the 2000 Census. In Online Appendix Figure IV, we present a version of this map in which we use data from the 1980-85 cohorts to estimate mobility for the CZs that have fewer than 250 observations in the core (1980-82) sample. The estimates of mobility in the CZs with missing data are quite similar to those in neighboring CZs, consistent with the spatial autocorrelation evident in the rest of the map.

areas. CZs in the top decile have $\bar{y}_{25,c} > 52.0$, while those in the bottom decile have $\bar{y}_{25,c} < 37.4$. Note that the 37th percentile of the family income distribution for children at age 30 is \$22,900, while the 52nd percentile is \$35,500; hence, the difference in upward mobility across areas translates to substantial differences in children's' incomes.

Pooling all CZs, the unweighted standard deviation (SD) of $\bar{y}_{25,c}$ is 5.68; the population-weighted SD is 3.34. The unconditional SD of children's income ranks (which have a Uniform distribution) is $100/\sqrt{12} = 28.9$. Hence, a 1 SD improvement in CZ "quality" – as measured by its level of absolute upward mobility $\bar{y}_{25,c}$ – is associated with a $5.68/28.9 = 0.20$ SD increase in the expected income rank of children whose parents are at the 25th percentile.[24] For comparison, a 1 SD increase in parent income rank is associated with a 0.34 SD increase in a child's income rank (Figure IIa). Hence, a 1 SD improvement in CZ quality is associated with 60% as large an increase in a child's income as a 1 SD increase in her own parent's income.

There are three broad spatial patterns in upward mobility evident in Figure VIa. First, upward mobility varies substantially at the regional level. Upward mobility is lowest in the Southeast and highest in the Great Plains. The West Coast and Northeast also have high rates of upward mobility, though not as high as the Great Plains.

Second, there is substantial within-region variation as well. Using unweighted CZ-level regressions of the upward mobility estimates on Census division and state fixed effects, we estimate that 53% of the cross-CZ variance in absolute upward mobility is within the nine Census divisions and 36% is within states. For example, many parts of Texas exhibit relatively high rates of upward mobility, unlike much of the rest of the South. Ohio exhibits much lower rates of upward mobility than nearby Pennsylvania. The statistics also pick up much more granular variation in upward mobility. For example, South Dakota generally exhibits very high levels of upward mobility, with the exception of a few areas in the Southwest corner of the state. These areas are the largest Native American reservations in the U.S. and are well known to suffer from very high rates of persistent poverty.

The third generic pattern is that urban areas tend to exhibit lower levels of intergenerational mobility than rural areas on average. For instance, children from low-income families who grow

[24] An analogous calculation using the estimates of college attendance gradients by CZ in Section V.A below implies that a 1 SD increase in CZ quality is associated with a 0.19 SD (9.3 percentage point) increase in college attendance rates for children with parents at the 25th percentile. Using data from the PSID, Solon, Page and Duncan (2002, p390) estimate that a 1 SD increase in neighborhood quality is associated with a 0.32 SD increase in years of education. We find less variation in outcomes across neighborhoods presumably because commuting zones are much larger than the PSID sampling clusters analyzed by Solon, Page, and Duncan.

up in the Chicago area have significantly lower incomes at age 30 than those who grow up in rural areas in Illinois. On average, urban areas – which we define as CZs that intersect MSAs – have upward mobility of $\bar{y}_{25,c} = 41.7$, while rural areas have $\bar{y}_{25,c} = 45.8$. In interpreting this comparison, it is important to recall that our definition of geography is based on where children grew up, not where they live as adults. 44.6% of children who grow up in rural areas live in urban areas at age 30. Among those who rose from the bottom quintile of the national income distribution to the top quintile, the corresponding statistic is 55.2%.

Table IV shows statistics on intergenerational mobility for the 50 largest CZs by population. Among these cities, absolute upward mobility ranges from 46.2 in the Salt Lake City area to 35.8 in Charlotte (Column 4). There is considerable variation even between nearby cities: Pittsburgh is ranked second in terms of upward mobility among large metro areas, while Cleveland – approximately 100 miles away – is ranked in the bottom 10. Upward mobility is especially low in certain cities in the "Rust Belt" such as Indianapolis and Columbus and cities in the Southeast such as Atlanta and Raleigh. The fact that children who grow up in low-income families in Atlanta and Raleigh fare poorly is perhaps especially striking because these cities are generally considered to be booming cities in the South with relatively high rates of job growth.

In Column 5 of Table IV, we consider an alternative measure of upward mobility: the probability that a child born to a family in the bottom quintile of the national income distribution reaches the top quintile of the national income distribution. To improve precision in smaller CZs, we estimate this probability pooling the 1980-1985 birth cohorts.[25] The ranking of areas based on this statistic is similar to that based on the mean rank measure of upward mobility. The probability that a child from the lowest quintile of parental income rises to the top quintile ranges is 10.8% in Salt Lake City, compared with 4.4% in Charlotte. The city with the highest probability of moving from the bottom fifth to the top fifth is San Jose, where the probability (12.9%) is nearly three times that in Charlotte. Note that if parent income played no role in determining children's' outcomes, all quintile transition probabilities would be 20%. Hence, the variation in rates of upward mobility across areas is large relative to the maximum range of 0 to 20%.

Relative Mobility. Figure VIb presents a heat map of relative mobility. This map is constructed

[25]We verify that including more recent cohorts does not generate significant bias by showing that the national quintile transition matrix based on the 1980-85 cohorts (Online Appendix Table IV) is virtually identical to the matrix based on the 1980-82 cohorts in Table III. The complete quintile transition matrix for each CZ is reported in Online Data Table VI. Combined with the marginal distributions of parent and child income at the national level (reported in Online Data Table II), the transition matrices can be used to construct any measure of intergenerational mobility by CZ. For reference, we also provide statistics on the marginal distributions of parent and child income by CZ in Online Data Table VII.

in the same way as Panel A, dividing CZs into deciles based on the rank-rank slope β_c. In this map, lighter areas denote areas with greater relative mobility (lower β_c). Relative mobility also varies substantially across areas. The expected rank of children from the richest vs. poorest families differs by more than 40.2 percentiles in CZs in the bottom decile of relative mobility. The corresponding gap is less than 23.5 percentiles for CZs in the top decile.

The geographical patterns in relative mobility in Panel B are similar to those for absolute upward mobility in Panel A. The unweighted correlation across CZs between the two measures is -0.68; the population-weighted correlation is -0.61. This indicates that areas with greater relative mobility tend to have better absolute outcomes for children from low-income families.

To investigate the connection between absolute and relative mobility more systematically, let $\mu_{pc} = E[y_{ic}|x_{ic} = p, c]$ denote a child's expected rank given a parent rank of p in CZ c. We estimate μ_{pc} in each CZ non-parametrically as the mean value of y_{ic} for children in each percentile bin of parent income $p = 0, ..., 99$.[26] For each of the 100 values of p, we estimate an unweighted OLS regression of μ_{pc} on relative mobility β_c with one observation per CZ:

$$\mu_{pc} = a + \gamma_p \beta_c + \eta_{pc}.$$

In this equation, γ_p measures the association across CZs between a 1 unit increase in β_c (i.e., greater intergenerational persistence) and mean absolute outcomes of children whose parents were at the p^{th} percentile of the national income distribution. A negative coefficient ($\hat{\gamma}_p < 0$) implies that CZs with greater relative mobility generate better mean outcomes for children with parents at percentile p.

Figure VIIa plots the coefficients $\hat{\gamma}_p$ at each parent income percentile p along with a linear fit to the coefficients. The coefficients $\hat{\gamma}_p$ are increasing with p: CZs with greater relative mobility (lower β_c) produce better outcomes for children from lower income families. The best linear fit crosses 0 at $p = 85.1$. Hence, increases in relative mobility are associated with better outcomes for children who grow up in families below the 85th percentile on average. For families at the 85th percentile, differences in relative mobility across CZs are uncorrelated with a child's mean rank. For families in the top 15%, living in a CZ with greater relative mobility is associated with worse outcomes

[26] The expected value μ_{pc} differs from $\bar{y}_{pc}$ defined above because μ_{pc} is estimated non-parametrically using only data in percentile bin p, whereas $\bar{y}_{pc}$ is calculated based on the linear approximation to the rank-rank relationship in (2). In practice, the two estimates are extremely similar. For instance, in the 100 largest CZs, where μ_{pc} is estimated with very little error, the correlation between μ_{pc} and $\bar{y}_{pc}$ exceeds 0.99. We use the linear approximation $\bar{y}_{pc}$ in most of our analysis to obtain more precise estimates of absolute mobility in smaller CZs. However, because the goal of the exercise here is to evaluate the relationship between relative mobility β_c and absolute mobility at each percentile non-parametrically, we use μ_{pc} here.

on average for children. Observe that γ_p reaches only 0.2 for the richest families but is nearly -0.8 for the poorest families. Hence, differences in relative mobility across CZs are associated with much larger differences in absolute mobility for children from low-income families than high-income families.[27]

Figure VIIb presents a schematic that illustrates the intuition underlying the preceding results. This figure plots hypothetical rank-rank relationships in two representative CZs, one of which has more relative mobility than the other. Figure VIIa implies that in such a pairwise comparison, the rank-rank relationship "pivots" at the 85th percentile on average. This is why the spatial patterns of absolute mobility at $p = 25$ and relative mobility in Figure VI look similar.

Because the pivot point is very high in the income distribution, differences in relative mobility have a smaller effect on children's' percentile ranks in high-income families than low-income families.[28] This may be because the rich are able to insulate themselves from differences in the local environment. If the differences in relative mobility across areas are caused by differences in local policies, this result suggests that one may be able to improve the outcomes of children from poor families without hurting children from high income families significantly.

IV.D Robustness of Spatial Patterns

In Table V, we assess the robustness of the spatial patterns in mobility documented above along three dimensions: (1) changes in sample definitions, (2) changes in income measures, and (3) adjustments for factors such as differences in the cost-of-living across areas. Each cell in the table reports the correlation across CZs of our baseline mobility measure (using child family income rank and parent family income rank in the core sample) with an alternative mobility measure described in each row. Column 1 reports the unweighted correlation across CZs between our baseline measure of absolute upward mobility ($\bar{y}_{25,c}$) and the corresponding alternative measure of $\bar{y}_{25,c}$. Column 2 replicates Column 1 for relative mobility (β_c). Columns 3 and 4 replicate the correlations in Columns 1 and 2 weighting CZs by their population in the 2000 Census.

Sample Definitions. In the first section of Table V, we assess the robustness of the spatial

[27] If the rank-rank relationship were perfectly linear, the relationship plotted in Figure VIIa would be perfectly linear and $\gamma_{100} - \gamma_0 = 1$ mechanically. The slight deviation from linearity at the bottom of the distribution evident in Figure V generates the slight deviation of $\gamma_{100} - \gamma_0$ from 1.

[28] It bears emphasis that this result applies to percentile ranks rather than mean income levels. Because the income distribution has a thick upper tail, a given difference in percentile ranks translates to a much larger difference in mean incomes in the upper tail of the income distribution. The probability that children of affluent parents become very high income "superstars" may therefore differ significantly across areas, an interesting question that we defer to future research.

patterns to changes in the sample definition, as we did at the national level in Table II. Rows 1 and 2 restrict the sample to male and female children, respectively. Rows 3 and 4 consider the subsamples of married parents and single parents. The correlations of both absolute and relative mobility in these subsamples with the corresponding baseline measures is typically above 0.9.

In row 5, we replicate the baseline specifications using the 1983-85 birth cohorts (whose incomes are measured at age 27 on average in 2011-12). In row 6, we consider the 1986-88 birth cohorts instead. The intergenerational mobility estimates across CZs for these later birth cohorts are very highly correlated with the baseline estimates. This result has three implications. First, it demonstrates that the reliability of CZ-level estimates is quite high across cohorts; in particular, sampling error or cohort-specific shocks do not lead to much fluctuation in the CZ-level estimates. Second, because the later cohorts are linked to parents at earlier ages (as early as age 8), we conclude that the spatial patterns in intergenerational mobility are not sensitive to the precise age at which we link children to parents or measure their geographical location. Finally, because the earnings of later cohorts are measured at earlier ages, we conclude that one can detect the spatial differences in mobility even when measuring earnings quite early in children's' careers.

In row 7, we restrict the sample based on the age of parents at the birth of the child. We limit the sample to children whose mothers are between the ages of 24-28 and fathers are between 26-30 (a five year window around the median age of birth). The intergenerational mobility estimates in this subsample are very highly correlated with the baseline estimates, indicating that the cross-area differences in income mobility are not biased by differences in the age of child birth for low income individuals.

In row 8, we assess the extent to which the variation in intergenerational mobility comes from children who succeed and move out of the CZ as adults vs. children who stay within the CZ. To do so, we restrict the sample to the 62% of children who live in the same CZ in 2012 as where they grew up. Despite the fact that this sample is endogenously selected on an ex-post outcome, the mobility estimates remain very highly correlated with those in the full sample. Apparently, areas such as Salt Lake City that generate high levels of upward income mobility do so not just by sending successful children to other CZs as adults but also by helping children move up in the income distribution within the area.

In row 9, we restrict the sample to the 88% of children in the core sample who are not claimed as dependents by other individuals in subsequent years after they are linked to the parents we identify. We obtain very similar estimates for this "unique parent" subsample, indicating that the spatial pattern of our mobility estimates is not distorted by measurement error in linking children to their parents.

Income Definitions. In the second section of Table V, we evaluate the sensitivity of the spatial patterns to alternative definitions of income. The definitions we consider match those in the robustness analysis in Table II; see Section II.B for details on these definitions. In row 10 of Table V, we define parent income as the income of the higher earner rather than total family income to evaluate potential biases from differences in parent marital status across areas. In row 11, we measure the child's income using individual income instead of family income to assess the effects of differences in the child's marital status. In row 12, we use the child's individual earnings (excluding capital and other non-labor income). In row 13, we replicate the specification in row 8 for male children, using individual income for the child and family income for the parent. Row 14 replicates row 13, but defines the parents' income as the income of the higher earner instead. All of these definitions produce very similar spatial patterns in intergenerational mobility: correlations with the baseline measures exceed 0.9 in most cases.

Cost-of-Living, Local Growth, and Other Factors. The third section of Table V considers a set of other factors that could bias comparisons of intergenerational mobility across areas. One natural concern is that our estimates of upward mobility may be affected by differences in prices across areas. To evaluate the importance of differences in cost of living, we construct a CZ-level price index using the American Chamber of Commerce Research Association (ACCRA) price index for urban areas combined with information on housing values, population density, and CZ location (see Online Appendix A for details). We then divide parents' income by the price index for the CZ where their child grew up and the child's income by the price index for the CZ where she lives as an adult (in 2012) to obtain real income measures.

Row 15 of Table V shows that the measures of intergenerational mobility based on real incomes are very highly correlated with our baseline measures (see also Online Appendix Figure Va). The reason that cost-of-living adjustments have little effect is that prices affect *both* the parent and the child. Intuitively, in high-priced areas such as New York City, adjusting for prices reduces the child's absolute rank in the national real income distribution. But adjusting for prices also lowers the real income rank of parents living in New York City. As a result, the degree of upward mobility – i.e., the difference between the child's rank and the parent's rank – is essentially unaffected by adjusting for local prices. The preceding logic assumes that children always live in the same cities in their parents. In practice, some children move to areas with higher prices (e.g. from rural Iowa to New York City). Our measures of upward mobility are affected by the cost of living adjustment in such cases, but they are not sufficiently frequent to have a large impact on our estimates. The correlation between the cost of living in the child's CZ at age 30 and the parent's CZ is 0.77 and the correlation between a child's nominal percentile rank and the local price index is only 0.10. Because of these two factors, cost of living adjustments end up having a minor impact on the difference between child and parent income and thus have little effect on our mobility statistics.

Another potential concern with our approach is that using national ranks may misrepresent the degree of relative mobility within the local income distribution, which may better reflect a child's opportunities. To address this concern, in row 16 of Table V, we measure relative mobility using local ranks. We rank parents relative to other parents living in the same CZ and children relative to other children who grew up in the same CZ (no matter where they live as adults). We define relative mobility as the slope of the local rank-rank relationship.[29] Relative mobility based on local ranks is very highly correlated with relative mobility based on national ranks. This is because local ranks are approximately a linear transformation of national ranks.

In row 17, we measure absolute upward mobility based on the probability that the child rises from the bottom quintile of parent income to the top quintile of child income, as in Column 5 of Table IV. The spatial pattern in this measure – shown in the map in Online Appendix Figure Vb – is very similar to that in our mean-rank based measure of upward mobility, with a correlation across CZs above 0.9.

Part of the variation in upward mobility across areas could be driven by shocks to economic growth. Because we measure parent income before 2000 and child income in 2011-12, local economic growth in the intervening decade would lead the child to be ranked higher in the national income distribution than the parent. While shocks to economic growth – e.g., from the discovery of a natural resource such as oil – are a real source of upward mobility, one may be interested in isolating variation in mobility that is attributable to more stable factors that may be manipulable by policy. We assess the extent to which economic growth is responsible for the spatial variation in upward mobility in two ways. In row 18, we define parent income as mean family income in 2011-12, the same years in which we measure child income. Insofar as local economic growth raises the incomes of both parents and children, this measure nets out the effects of growth on mobility. Both the upward and relative mobility measures remain very highly correlated with the baseline

[29] We cannot study absolute mobility with local ranks because both child and parent ranks have a mean of 50 by definition: if one child moves up in the local distribution, another must move down.

measures, suggesting that differences in local economic growth drive relatively little of the spatial variation in mobility.

One limitation of using parent income in 2011-12 to net out growth effects is that growth shocks may have bigger effects on younger workers who are just entering the labor force. As an alternative approach, we regress our measures of mobility on the CZ-level growth rate from 2000-2010 and calculate residuals.[30] Row 19 of Table V shows that the correlation of the growth-adjusted relative mobility measures with the baseline measures exceeds 0.9; the correlations for absolute mobility exceed 0.8. Note that these growth-adjusted measures over-control for exogenous growth shocks insofar as growth is partly a consequence of factors that generate upward income mobility in an area. Hence, the fact that even controlling for growth rates directly does not significantly change the spatial pattern of intergenerational mobility supports the view that most of the variation in mobility across areas is not due to exogenous growth shocks in the 2000's.

IV.E College Attendance and Teenage Birth Gradients

In Figure VIII, we explore the geographical variation in the gradients of college attendance rates and teenage birth rates with respect to parent income, which we plotted at the national level in Figure IV. To construct Figure VIIIa, we regress college attendance on parent national income rank in each CZ c. We use the same specification as in (1), replacing the dependent variable with an indicator for attending college at some point between the ages of 18-21. The map plots the slope from these regressions in each CZ, dividing the CZs into deciles (with darker colors representing areas with steeper slopes), as in Figure VIa.

There is substantial variation across CZs in the association between parent income and children's college attendance rates. CZs in the top decile of college-income gradients have $\beta_c > 0.8$, i.e. children of the highest income parents are 80 percentage points more likely to attend college than children of the lowest income parents. CZs in the bottom decile have $\beta_c < 0.53$. Moreover, the spatial variation in college-income gradients is highly correlated with the spatial variation in the intergenerational income mobility.[31] The unweighted correlation across CZs of the gradients

[30] We measure income in 2000 using the Census and in 2008 using the 5-year American Community Survey, averaged over 2006-2010. We calculate household income per working age adult as aggregate income in a CZ divided by the number of individuals aged 16-64 in that CZ. Annualized income growth is calculated as the annual growth rate implied by the change in income over the 8 year period; we use 8 years because 2008 is the midpoint of 2006-2010.

[31] As above, we define absolute mobility for the college outcome as the mean college attendance rate for children whose parents are at the 25th percentile, based on the linear regression of college attendance on parent rank. As with earnings, areas with flatter college by parent income gradients (i.e., greater relative mobility) tend to have higher college attendance rates for lower income children (i.e., greater absolute upward mobility). The slopes of the college and teenage birth gradients and absolute mobility measures for each CZ are reported in Online Data Table V.

in Figure VIIIa with the corresponding relative mobility measures based on children's income in Figure VIa is $\rho = 0.68$ (row 20 of Table V).

Figure VIIIb repeats the analysis in Figure VIIIa using an indicator for having a teenage birth, defined as in Figure IVb. In this figure, we restrict the sample to females. CZs in the bottom decile (those with the steepest negative gradients) have $\beta_c < -0.37$. That is, daughters of the lowest income parents are 37 percentage points more likely to have a child while they are teenagers than daughters of the highest income parents. CZs in the top decile have $\beta_c > -0.18$. Once again, this spatial variation is quite highly correlated with the variation in intergenerational earnings mobility (row 21 of Table V).

An important implication of these results is that much of the difference in intergenerational mobility across areas emerges before children enter the labor market.[32] This suggests that the spatial variation in income mobility is driven by factors that either directly affect children at early ages – e.g., the quality of schools or social structure – or anticipatory behavioral responses to subsequent differences, such as returns to education in the local labor market. We explore mechanisms that have such properties in the next section.

V Correlates of Intergenerational Mobility

Why do some areas of the U.S. exhibit much higher rates of upward mobility than others? As a first step toward answering this question, we correlate our measures of intergenerational mobility with local area characteristics. Naturally, such correlations cannot be interpreted as causal mechanisms. Our goal here is merely to document a set of stylized facts that we hope will be helpful in guiding the search for causal determinants and the development of new models of intergenerational mobility.

We organize our analysis of such mechanisms around a set of factors that have been discussed in the sociology and economics literature: (1) race, (2) segregation, (3) income levels and inequality, (4) local public goods and tax policies, (5) primary school quality, (6) access to higher education, (7) local labor market conditions, (8) migration and networks, (9) social capital, and (10) family structure.

Because most of these factors are slow-moving and we have estimates of intergenerational income mobility for essentially one birth cohort, we focus on cross-sectional correlations rather than changes

[32]The fact that college and teenage birth gradients are similar to income mobility gradients provides further evidence that growth shocks in the 2000s do not generate the differences in mobility across areas, as college and teenage birth are measured around 2000. These results also show that the spatial patterns are unlikely to be driven by differences in reporting of taxable income.

over time. For most covariates, we use data from the 2000 Census and other publicly available datasets because many variables cannot be consistently measured in earlier years. We verify that results are similar using data from 1990 for selected variables. See Online Appendix D for details on the construction of the covariates analyzed in this section and Online Data Table VIII for CZ-level data on each of the covariates.

V.A Race

Perhaps the most obvious pattern from the maps in Figure VI is that intergenerational mobility is lower in areas with larger African-American populations, such as the Southeast. We therefore begin by exploring the role of race in upward mobility. Throughout our analysis of correlations, we focus primarily on our baseline measure of absolute upward mobility ($\bar{y}_{25,c}$). Correlations of area characteristics with relative mobility are similar, which is to be expected since absolute upward mobility and relative mobility are themselves highly correlated as shown above.

Figure IXa presents a binned scatter plot of absolute upward mobility ($\bar{y}_{25,c}$) in each CZ vs. the fraction of black individuals living in that CZ, based on data from the 2000 Census. The figure is constructed using one observation for each of the 709 CZs for which we have more than 250 parent-child pairs. To construct the binned scatter plot, we divide the variable plotted on the x-axis (% black in the CZ) into 20 equally sized bins (vingtiles) and plot the mean value of the variable plotted on the y-axis (absolute upward mobility) vs. the mean value of the x variable within each bin. We also report the unweighted correlation between the x and y variables, with standard error clustered at the state level to correct for spatial correlation across CZs. To facilitate comparisons across figures that plot the relationship between upward mobility and different factors, we always use a fixed y scale ranging from 35 to 55, approximately the 5th to 95th percentile of the distribution of $\bar{y}_{25,c}$ across CZs.

Figure IXa confirms that areas with larger African-American populations do in fact have substantially lower rates of upward mobility. The correlation between upward mobility and fraction black is -0.585. In areas that have small black populations, children born to parents at the 25th percentile can expect to reach the median of the national income distribution on average ($\bar{y}_{25,c} = 50$); in areas with large African-American populations, $\bar{y}_{25,c}$ is only 35.

The correlation in Figure IXa could be driven by two very different channels. One channel is an individual-level race effect: black children may have lower incomes than white children conditional on parent income, and hence areas with a larger black population may have lower upward mobility.

An alternative possibility is a place-level race effect: areas with large black populations might have lower rates of upward mobility for children of all races. To distinguish between these two channels, one would ideally control for race at the individual level, essentially asking whether whites have lower rates of upward mobility in areas with a larger black population.

Unfortunately, we do not observe race in our data. As an alternative, we predict race based on the parent's 5-digit ZIP code (in the year they first claim their child as a dependent). We use data from the 2000 Census to measure racial shares by ZIP code. Figure IXb replicates our measures of absolute upward mobility ($\bar{y}_{25,c}$) by CZ, restricting the sample to ZIP codes within each CZ in which at least 80% of the residents are non-hispanic whites.[33] In this subsample, 91% of individuals are white. The spatial pattern in Figure IXb is very similar to that in the original map for the full sample in Figure VIa. Most notably, even in this predominantly white sample, rates of upward mobility remain low in the Southeast and are much higher in the West. Among the 604 CZs for which we are able to compute upward mobility measures for predominantly white individuals, the unweighted correlation between upward mobility for the predominantly white sample and the full sample is 0.91.

In Figure IXc, we generalize this approach to assess how the spatial pattern of upward mobility changes as we restrict the sample to be increasingly white. To construct this figure, 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 $\bar{y}^w_{25,c}$. We then regress $\bar{y}^w_{25,c}$ on $\bar{y}_{25,c}$, 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 in Figure IXc against the fraction of white individuals in each of the subsamples. When $w = 0$, the regression coefficient is 1 by construction because $\bar{y}_{25,c} = \bar{y}^{w=0}_{25,c}$. Since 75% of the U.S. population is white, the first point on the figure is (0.75, 1). The point generated by the $w = 80\%$ threshold is (0.91, 0.84), consistent with the map in Figure IXb. The dotted lines show a 95% confidence interval for the regression coefficients based on standard errors clustered at the state level.

If the variation in upward mobility across areas were entirely driven by heterogeneity in outcomes across race at the individual level, the coefficient in Figure IXc would fall to zero as the fraction

[33]We continue to estimate $\bar{y}_{25,c}$ at the CZ level in this map, but we only include ZIP-5's within each CZ in which 80% or more of the residents are white. To facilitate comparison to Figure VI, we color the *entire* CZ based on this statistic, including ZIP-5's whose own white share is below 80%. CZs that have fewer than 250 children who grew up in ZIP codes where more than 80% of the residents are white are omitted (and shown with cross-hatch shading).

white in the sample converged to 1, as illustrated by the dashed line. Intuitively, if all of the spatial variation in Figure VIa were driven by individual-level differences in race, there would be no spatial variation left in a purely white sample. The data reject this hypothesis: even in the subsample with more than 95% white individuals, the regression coefficient remains at 0.89.

The main lesson of the analysis in this section is that both blacks and whites living in areas with large African-American populations have lower rates of upward income mobility. One potential mechanism for this pattern is the historical legacy of greater segregation in areas with more blacks. Such segregation could potentially affect both low-income whites and blacks, as racial segregation is often associated with income segregation. We turn to the relationship between segregation and upward mobility in the next section.

V.B Segregation

Prior work has argued that segregation has negative effects on economic and social outcomes through various channels: reducing exposure to successful peers and role models, decreasing funding for local public goods such as schools, or hampering access to nearby jobs (Wilson 1987, Massey and Denton 1993, Cutler and Glaeser 1997). Motivated by these theories, we explore the relationship between intergenerational mobility and various measures of segregation.

We begin by estimating the association between racial segregation and upward mobility. Following Iceland (2004), we measure segregation using a Theil (1972) index, constructed using data from the 2000 Census.[34] Let p_r denote the fraction of individuals of race r in a given CZ, with four racial groups: whites, blacks, hispanics, and others. We measure the level of racial diversity in the CZ by an entropy index: $E = \sum p_r \log_2 \frac{1}{p_r}$, with $p_r \log_2 \frac{1}{p_r} = 0$ when $p_r = 0$. Letting $j = 1, ..., N$ index census tracts in the CZ, we analogously measure racial diversity within each tract as $E_j = \sum p_{rj} \log_2 \frac{1}{p_{rj}}$ where p_{rj} denotes the fraction of individuals of race r in tract j. We define the degree of racial segregation in the CZ as

$$H = \sum_j [\frac{\text{pop}_j}{\text{pop}_{total}} \frac{E - E_j}{E}] \quad (3)$$

[34] As Iceland (2004) argues, the Theil index is an attractive measure conceptually because it captures segregation across multiple racial groups. However, we obtain similar results using alternative two-group measures of black-white segregation such as isolation indices or dissimilarity indices because alternative measures of segregation are highly correlated at the level of metro areas (Cutler et al. 1999). The segregation patterns are sufficiently stark that one can directly see the differences in segregation between the least and most upwardly mobile cities using the color-coded dot maps produced by Cable (2013) using Census data. For instance, compare Atlanta – one of the most segregated cities and one of the lowest-mobility cities in our data – to Sacramento – one of the most integrated and highest-mobility cities.

where pop_j denotes the total population of tract j and pop_{total} denotes the total population of the CZ. Intuitively, H measures the extent to which the racial distribution in each Census tract deviates from the overall racial distribution in the CZ. The segregation index H is maximized at $H = 1$ when there is no racial heterogeneity within census tracts, in which case $E_j = 0$ in all tracts. It is minimized at $H(p) = 0$ when racial diversity within each tract – as measured by entropy E_j – is the same across all tracts.

Column 1 of Table VI reports the coefficient estimate from an unweighted OLS regression of absolute upward mobility $\bar{y}_{25,c}$ on the racial segregation index, with one observation per CZ. In this and all subsequent regressions in this paper, we standardize the dependent variable and all independent variables to have mean 0 and standard deviation 1 within the estimation sample. Hence, the coefficients in the univariate regressions can be interpreted as correlation coefficients. Standard errors are clustered by state to account for spatial correlation across CZs.

More racially segregated areas have less upward mobility: the unweighted correlation between upward mobility and the racial segregation index in Column 1 is -0.361. See Online Appendix Figure VIa for the corresponding non-parametric binned scatter plot. Column 2 shows that the correlation remains at -0.360 in urban areas, i.e. CZs that overlap with metropolitan statistical areas (MSAs).

Next, we turn to the relationship between income segregation and upward mobility. Following Reardon and Firebaugh (2002) and Reardon (2011), we begin by measuring the degree to which individuals below the p^{th} percentile of the local household income distribution are segregated from individuals above the p^{th} percentile in each CZ using a two-group Theil index $H(p)$. Here, entropy in a given area is $E(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$ and the index $H(p)$ is defined using the formula in (3). Building on this measure, Reardon (2011) defines the overall level of income segregation in a given CZ as

$$\text{income segregation} = 2log(2) \int_p E(p)H(p)dp. \tag{4}$$

This measure is simply a weighted average of segregation at each percentile p, with greater weight placed on percentiles in the middle of the income distribution, where entropy $E(p)$ is maximized. We implement (4) using data from the 2000 Census, which reports income binned in 16 categories. Following Reardon (2011, Appendix 3), we measure $H(p)$ at each of these cutoffs and take a weighted sum of these values to calculate income segregation.

In Column 3 of Table VI, we regress absolute upward mobility on the income segregation index; see Online Appendix Figure VIb for the corresponding non-parametric binned scatter plot. The correlation between income segregation and upward mobility is -0.393. Interestingly, areas with a larger black population exhibit greater income segregation: the correlation between the fraction of black individuals in a CZ and the income segregation index is 0.271 (s.e. 0.077). Hence, the negative relationship between income segregation and upward mobility could partly explain why low-income white children fare more poorly in areas with large African-American populations.

In Column 4, we decompose the effects of segregation in different parts of the income distribution. Following Reardon and Bischoff (2011), we define the "segregation of poverty" as $H(p = 25)$, i.e. the extent to which individuals in the bottom quartile are segregated from those above the 25th percentile. We analogously define the segregation of affluence as $H(p = 75)$. Conditional on the degree of segregation of affluence, segregation of poverty is strongly negatively correlated with upward mobility. However, there is no significant relationship between the segregation of affluence and upward mobility conditional on segregation of poverty. Column 5 shows that the same pattern holds when restricting the sample to urban areas.

These results suggest that it is the isolation of low-income families rather than the isolation of the rich that may be most detrimental for low income children's prospects of moving up in the income distribution. One explanation of this correlation is that the separation of the middle class from the poor reduces beneficial peer effects or funding for local public goods (e.g., schools) for children from low-income families. In contrast, the separation of the affluent from the middle class may not directly harm low income individuals.

Another mechanism by which segregation may diminish upward mobility is through spatial mismatch in access to jobs (Kain 1968, Kasarda 1989, Wilson 1996). We explore this mechanism in Column 6 by correlating upward mobility with the fraction of individuals who commute less than 15 minutes to work in the CZ, based on data from the 2000 Census. Areas with less sprawl (shorter commutes) have significantly higher rates of upward mobility: the correlation between commute times and upward mobility is 0.6, higher than the univariate correlation with any of the measures of segregation.

In Column 7, we regress upward mobility on both the commute time variable and income segregation. Commute times remain strongly associated with upward mobility conditional on income segregation, but income segregation is not significantly correlated with upward mobility once we condition on commute times. These results are consistent with the view that the negative impacts of segregation may operate by making it more difficult to reach jobs or other resources that facilitate upward mobility. Note that any such spatial mismatch explanation must explain why the gradients emerge before children enter the labor market, as shown in Section IV.E. A lack of access to nearby jobs cannot directly explain why children from low-income families are also more likely to have teenage births and less likely to attend college in cities with low levels of upward mobility. However, spatial mismatch could produce such patterns if it changes children's behavior because they have fewer successful role models or reduces their perceived returns to education.

We summarize our results on segregation in the first panel of Figure X. Figure X plots the unweighted univariate correlation between absolute upward mobility and various CZ-level characteristics, using all CZs with available data for the relevant variable. The dots show the point estimate of the correlation and the horizontal lines show a 95% confidence interval, based on standard errors clustered at the state level. The sign of the correlation is shown in parentheses next to each variable.

Table VII reports each of the correlations corresponding to Figure X in Column 1 and evaluates their robustness to alternative specifications in the other columns. In Column 2, we report estimates based on within-state variation by including state fixed effects in a regression specification analogous to that in Column 1 of Table VI. Column 3 replicates Column 1, weighting each CZ by its population as recorded in the 2000 Census.[35] Column 4 restricts the sample to urban areas (CZs that intersect MSAs) and replicates Column 1. Column 5 replicates Column 1, controlling for the fraction of black individuals in the CZ and the local income growth rate from 2000-2010 (calculated as in Section IV.D using Census data) using regression specifications of the form used in Table VI. Finally, in Column 6, we correlate each covariate with relative mobility β_c.

We find a significant negative correlation between upward income mobility and each of our three primary proxies for segregation – racial segregation, income segregation, and commute times – across all the specifications in Table VII.[36] Importantly, this correlation holds even within states (Column 2), showing that the pattern is not just driven by broad regional differences across the South vs. other parts of the country. Segregation is also strongly negatively correlated with relative mobility (i.e., positively correlated with β_c).

[35]We normalize all variables by their weighted standard deviations in this and all other specifications that use weights, so that univariate regression coefficients can be interpreted as weighted correlations.

[36]We also replicated this analysis using measures of segregation from the 1990 Census and find very similar results. For example, the correlation between upward mobility and the Theil racial segregation index measured using the 1990 Census is -0.357, compared with -0.361 when measured using the 2000 Census. The correlation between upward mobility and income segregation is -0.393 using both the 1990 and 2000 Census.

V.C Income Levels and Inequality

Several studies have proposed that there may be a link between properties of the static income distribution – for instance, levels of inequality within a generation – and the level of intergenerational mobility (e.g., Corak 2013). In this subsection, we explore the correlation between properties of the local income distribution and upward income mobility.

Mean Income Levels. Figure XIa presents a binned scatter plot of absolute upward mobility vs. the mean level of household income per working age adult in the CZ, as measured in the 2000 Census.[37] There is little relationship between mean CZ-level income and upward mobility. Children in low-income families who grow up in the highest-income CZs (with mean incomes of $47,600 per year) reach almost exactly the same percentile of the national income distribution on average as those who grow up in the lowest-income areas (with mean incomes of $21,900).

The result in Figure XIa stands in contrast with the results on income segregation above, which effectively shows that low-income children who grow up in census tracts with higher levels of mean income have better outcomes. This may be because peer effects and the provision of public goods operate at a narrow geographic level. Living in a higher income metro area might not benefit low-income children if it has no effect on their peer group or resources due to income segregation within the CZ.

Income Inequality. Figure XIb presents a binned scatter plot of upward mobility vs. the Gini coefficient of parent income within each CZ. We compute the Gini coefficient for parents in our core sample within each CZ as Gini $= \frac{2}{\bar{z}_c} Cov(z_{ic}, x_{ic})$, where $\bar{z}_c$ is the mean family income (from 1996-2000) of parents in CZ c and $Cov(z_{ic}, x_{ic})$ is the covariance between the income level (z_{ic}) and the percentile rank (x_{ic}) of parents in CZ c. The correlation between the Gini coefficient and upward mobility is -0.578, implying that there is a "Great Gatsby" relationship within the U.S. similar to that documented across countries in prior work.

An alternative measure of inequality is the portion of income within a CZ that accrues to the richest households, e.g. those in the top 1%. This measure is of particular interest because the rise in inequality in the U.S. over the past three decades was driven primarily by an increase in top income shares (Piketty and Saez 2003). We calculate top 1% income shares using the distribution of parent family income within each CZ. Figure XIc shows that the relationship between upward mobility and the top 1% income share is much weaker (correlation = -0.190).

[37] We find similar results using mean income in 1990; however, we find a positive correlation with mean income in 2010, reflecting the correlation with income growth documented above.

We investigate why the Gini coefficient and top 1% share produce different results in Table VIII, which is constructed in the same way as Table VI. Column 1 replicates the regression corresponding to Figure XIb as a reference. We decompose the Gini coefficient into inequality coming from the upper tail and the rest of the income distribution by defining the bottom 99% Gini as the Gini coefficient minus the top 1% income share. The bottom 99% Gini can be interpreted as the deviation of the Lorenz curve from perfect equality amongst households in the bottom 99%. Column 2 of Table VIII shows that a 1 SD increase in the bottom 99% Gini is associated with a 0.634 SD reduction in upward mobility. In contrast, a 1 SD increase in the top 1% share is associated with only a 0.123 SD reduction in upward mobility. Column 3 replicates Column 2 for urban areas (CZs that overlap with MSAs). The pattern in urban areas is even more stark: upper tail inequality is uncorrelated with upward mobility, whereas the Gini coefficient within the bottom 99% remains very highly strongly correlated with upward mobility.

An alternative and perhaps more intuitive measure of inequality within the bottom 99% is the size of the middle class in the CZ, which we define as the fraction of parents in the CZ who have family incomes between the 25th and 75th percentiles of the national parent income distribution. Column 4 of Table VIII shows that upward mobility is strongly positively correlated with the size of the middle class.

Finally, Column 5 of Table VIII replicates Column 2 using relative mobility β_c as the dependent variable. The bottom 99% Gini coefficient is strongly positively associated with this measure, i.e. greater inequality in the bottom 99% is negatively related to relative mobility.[38] But once again, the top 1% share is uncorrelated with relative mobility.

We summarize our results on properties of the income distribution and evaluate their robustness in the second panel of Table VII and Figure X. Across all the specifications, we find (1) little association between mean income levels and mobility, (2) a robust negative relationship between the bottom 99% Gini coefficient and mobility, and (3) little relationship between the top 1% share and mobility.

Comparison to Cross-Country Evidence. The preceding analysis shows that the correlation between inequality and intergenerational mobility across areas within the U.S. is driven primarily by "middle class" inequality – i.e., inequality within the bottom 99% – rather than the concentration of income in the upper tail. Is this also the case with the original "Great Gatsby" curve

[38] Importantly, because parent and child ranks are measured in the national income distribution, there is no mechanical relationship between the level of inequality within the CZs income distribution and the rank-rank slope.

documented by Corak (2006, 2013) across countries? We answer this question using estimates of the intergenerational elasticity (IGE) from 13 developed countries compiled by Corak (2013). In Column 6 of Table VIII, we replicate Corak's (2013, Figure 1) result that there is a strong positive correlation between the Gini coefficient (as measured in survey data on income in 1985) and the IGE.[39] In Column 7, we include the top 1% income share in each country, based on statistics from the World Top Incomes Database. As in the within-U.S. analysis, there is little correlation between the top 1% income share and intergenerational mobility across countries. Column 8 shows that results are similar if one uses inequality measures from 2005 instead of 1985.

We conclude that there is a robust negative correlation between inequality within the current generation of adults and mobility across generations. However, intergenerational mobility is primarily correlated with "middle-class" inequality and not the upper tail inequality of the form that has increased dramatically in recent decades. Interestingly, this pattern parallels the results we obtained for segregation above: segregation of affluence is not significantly correlated with intergenerational mobility, while segregation of poverty (the separation of the poor from the middle class) is strongly negative associated with mobility.

V.D Local Public Goods and Tax Policies

Economic models of intergenerational mobility predict that the provision of local public goods or tax credits to low income families can relax credit constraints and increase mobility (e.g., Becker and Tomes 1979, Becker and Tomes 1986, Mulligan 1997, Ichino et al. 2011). In this subsection, we assess whether variation in local tax and expenditure policies explain the geographical variation in mobility. We report these results in the third panel of Figure X and Table VII.

We begin by correlating upward mobility with local tax rates. We measure the average local tax rate in each CZ as total tax revenue collected at the county or lower level in the CZ (based on the 1992 Census of Governments) divided by total household income in the CZ based on the 1990 Census.[40] Note that 75% of local tax revenue comes from property taxes; hence, this measure largely captures variation in property tax rates. In the baseline unweighted specification pooling all

[39] We interpret Corak's estimates of the Gini coefficient as measures that apply to the bottom 99% because surveys typically do not capture the thickness of the top tail due to top-coding.

[40] Government expenditures in the neighborhoods where low-income families live within the CZ (rather than average government expenditures) may be more relevant for upward mobility. To evaluate this possibility, we reconstructed each of the measures of public goods and school quality analyzed in this and the next subsection, weighting by the number of below-median income families living in each county or school district. The correlations between upward mobility and these measures of public goods for low-income individuals are very similar to those reported in Table VII because expenditures in low-income areas are very highly correlated with mean expenditures at the CZ level.

CZs, the correlation between absolute upward mobility is 0.32. We find a robust positive correlation between tax rates and upward mobility across the specifications in Table VII.[41]

An alternative measure of local public good provision is total local government expenditure. Tax revenue differs from local government expenditure because of inter-governmental transfers. We define mean local government expenditure as total expenditure at the county or lower level divided by household income in the CZ in 1990. The correlation between government expenditure and upward mobility is also positive, but it is smaller than that between local tax rates and upward mobility. This could potentially be because local tax rates are used primarily to finance schools, which may have a larger impact on upward mobility than expenditures funded by other sources of revenue.

Next, we evaluate whether areas that provide more transfers to low-income families through the tax system exhibit greater upward mobility. We use two state-level proxies for the progressivity of local tax policy. The first is the size of the state Earned Income Tax Credit. State EITC programs are the largest state-level cash transfer for low income earners. Because state EITC policies changed significantly over the period when children in our sample were growing up, we define a measure of mean exposure to the state EITC as the mean state EITC rate between 1981 and 2001, when the children in our sample were between the ages of 0 and 20.[42] The mean state EITC rate is positively correlated with upward mobility, with a correlation of approximately 0.25 that is fairly robust across specifications. Our second proxy for the progressivity of the local tax code is the difference between the top state income tax rate and the state income tax rate for individuals with taxable income of $20,000 in 2008 based on data from the Tax Foundation. There is a weak positive correlation between local tax progressivity and upward mobility across the various specifications in Table VII, but the correlation is not statistically significant.

In summary, we find that areas that provide more local public goods and larger tax credits for low income families tend to have higher levels of upward mobility. However, segregation and inequality are much stronger and more robust predictors of the variation in intergenerational mobility than differences in local tax and expenditure policies.

[41] We drop one small CZ (Barrow, AK) in which local government revenue exceeds mean income. We analogously exclude two outlier CZs in Alaska in which local government expenditure exceeds 50% of income when analyzing the correlation with local government expenditure below.

[42] We assign state-years without a state EITC a rate of 0 when computing this mean. See Online Appendix D for further details on the computation of state EITC rates.

V.E School Quality

The local public good that may have the most direct impact on children's outcomes is the quality of local schools (e.g., Card and Krueger 1992). We study the correlation between mobility and various proxies for school quality in the fourth panel of Figure X and Table VII.

We begin by analyzing mean public school expenditures per student, based on data from the National Center for Education Statistics (NCES) for the 1995-1996 fiscal year. The correlation between public school expenditures and upward mobility is roughly similar to that between local tax rates and upward mobility, which is to be expected given that local tax revenue is used primarily to fund schools.[43]

Since expenditures are not necessarily a good measure of the quality of education (Hanushek 1989), we next turn to one easily measurable input into the education production function: class size. Prior work has demonstrated that class size has causal effects on student achievement and long-term outcomes (Krueger 1999, Chetty et al. 2011, Fredriksson et al. 2013). We obtain data on mean student-teacher ratios from the NCES for the 1996-1997 school year.[44] When pooling all CZs, there is a strong negative correlation between class size and upward mobility (Columns 1 and 2 of Table VII). However, there is no correlation between upward mobility and class size in more urban areas (Columns 3 and 4).

One shortcoming of input-based measures of school quality is that they capture relatively little of the variation in the true quality of schools (e.g., Hanushek 2003). One common approach to addressing this problem is to use output-based measures of quality, such as the value-added of teachers. We construct simple output-based proxies for school quality based on test scores and dropout rates adjusted for differences in parent income. We obtain data on mean grade 3-8 math and English test scores by CZ from the Global Report Card, which converts scores on state-level tests to a national scale to generate test score measures by school district on a single scale. We obtain data on high school dropout rates from the NCES for the 2000-2001 school year, restricting the sample to CZs in which at least 75% of school districts have non-missing data. We regress test scores on mean parent family income (from 1996-2000) in the core sample and compute residuals to obtain an income-adjusted measure of test score gains. We construct an income-adjusted measure of dropout rates analogously.

[43]We drop observations in the top 1% of the distribution of expenditures per student to reduce the influence of outliers.

[44]We drop observations where the mean reported class size is 0 or above 100.

The income-adjusted test score and dropout rates are very highly correlated with upward mobility across all specifications, as shown in the fourth panel of Figure X and Table VII. In the baseline specification (Column 1 of Table VII), the magnitude of the correlation between both measures and upward mobility is nearly 0.6. These results are consistent with the hypothesis that the quality of schools – as judged by outputs rather than inputs – plays a role in upward mobility. At a minimum, they strengthen the view that much of the difference in intergenerational income mobility across areas emerges while children are relatively young.

V.F Access to Higher Education

Having found a correlation between measures of the quality of local public schools and upward mobility, we now turn to higher education. We construct three measures of local access to higher education using data from the Integrated Postsecondary Education Data System (IPEDS). The first measure is the number of Title IV, degree-granting colleges per capita in the CZ in 2000, which is similar to the distance-based instrument used by Card (1993). The second measure is the mean (enrollment-weighted) tuition sticker price for in-state, full-time undergraduates for colleges in the CZ, which reflects the affordability of local higher education. The third measure is the residual from an OLS regression of the mean (enrollment-weighted) graduation rate from colleges in the CZ on mean parent family income in the CZ, a rough proxy for the output of local higher education.

The correlations between all three of these measures – shown in the fifth panel of Figure X and Table VII – are small and typically statistically insignificant. We also evaluated several additional measures of access to higher education, including the mean value of institutional grants to students enrolled in colleges in the CZ, the number of low-cost (below the national median) colleges per capita in the CZ, and the mean distance to the nearest low-cost college. We found no significant relationship between any of these measures and our measures of intergenerational mobility (not reported).

We conclude that very little of the spatial variation in intergenerational mobility is explained by differences in local access to higher education. Of course, this finding does not imply that college does not play a role in upward mobility. Indeed, areas with greater upward mobility tend to have high college attendance rates for children from low-income families (Figure VIIIa), suggesting that attending college is an important pathway for moving up in the income distribution. The point here is simply that the characteristics of local colleges are not a strong predictor of children's success, perhaps because the marginal impact of improving local access to higher education on college attendance and later outcomes is small.

V.G Labor Market Structure

Some analysts have suggested that the availability of certain types of jobs (e.g., manufacturing) may provide ladders for lower-skilled workers to move up in the income distribution (e.g., Wilson 1996). To explore this possibility, we measure various characteristics of the local labor market: (1) the overall employment rate in the local labor market in 2000, (2) the fraction of workers employed in the manufacturing industry, and (3) a measure of exposure to import competition based on the growth in Chinese imports per worker from Autor et al. (2013). As shown in the sixth panel of Figure X and Table VII, all of these characteristics are weakly correlated with the variation in upward mobility, with little evidence of a clear, robust relationship across specifications. We also find no significant correlation with other indicators such as the fraction of workers employed in management or professional occupations or industry establishment shares (not reported).

One labor market indicator that is strongly correlated with upward mobility is the teenage labor force participation rate. We measure the teenage labor force participation rate as the fraction of children who have a W-2 between the ages of 14-16 in the 1985-87 birth cohorts, the earliest cohorts for which W-2 data are available at age 14 in the tax data. The unweighted correlation between the teenage labor force participation rate and absolute upward mobility is 0.629, and this correlation is highly robust across specifications. This could be because formal jobs help disadvantaged teenagers directly or because areas with good schools and other characteristics tend to have more teenagers working in the formal sector. In either case, this finding mirrors the general pattern documented above: the strongest predictors of upward mobility are factors that affect children well before they enter the labor force as adults.

V.H Migration Rates

A large literature has evaluated whether immigration rates are related to labor market outcomes (e.g., Altonji and Card 1989, Borjas et al. 1997). Here we evaluate whether there is a correlation between migration rates and children's outcomes.

We consider three measures of the extent to which a CZ is integrated with other labor markets: (1) the migration inflow rate, defined as the number of people who move into the CZ between 2004 and 2005 based on IRS Statistics of Income migration data divided by the CZ population in 2000 based on Census data, (2) the migration outflow rate, defined as the number of people who move out of the CZ between 2004 and 2005 divided by population in 2000, and (3) the fraction of foreign-born individuals living in the CZ based on the 2000 Census.

The correlations between all three of these measures – shown in the seventh panel of Figure X and Table VII – are generally quite low and statistically insignificant. In the first two specifications, migration rates are negatively correlated with upward mobility, but in the population-weighted and urban-area specifications, there are no significant relationships.

V.I Social Capital

Several studies have emphasized the importance of social capital – the strength of social networks and engagement in community organizations in local areas – for social and economic outcomes (e.g., Coleman 1988, Borjas 1992, Putnam 1995). We explore the relationship between mobility and measures of social capital used in prior work in the eighth panel of Figure X and Table VII.

Our primary proxy for social capital is the social capital index constructed by Rupasingha and Goetz (2008) and employed by Putnam (2007), which we aggregate to the CZ level using population-weighted means. This index is comprised of voter turnout rates, the fraction of people who return their census forms, and various measures of participation in community organizations in the area. We find a strong positive correlation between upward mobility and social capital of 0.641 in the baseline specification, an estimate that is quite robust across specifications.[45]

We also consider two other proxies for social capital: the fraction of religious individuals (based on data from the Association of Religion Data Archives) and the rate of violent crime (using data from the Uniform Crime Report). Religiosity is very strongly positively correlated with upward mobility, while crime rates are negatively correlated with mobility.

V.J Family Structure

Many have argued that family stability plays a key role in children's' outcomes (see e.g., Becker 1991, Murray 1984, Murray 2012). To evaluate this hypothesis, we use three measures of family structure in the CZ based on data from the 2000 Census: (1) the fraction of children living in single-parent households, (2) the fraction of adults who are divorced, and (3) the fraction of adults who are married. All three of these measures are very highly correlated with upward mobility across all the specifications we consider, as shown in the ninth panel of Figure X and Table VII.

[45] Interestingly, one of the original measures proposed by Putnam (1995) – the number of bowling alleys in an area – has an unweighted correlation of 0.562 with our measures of absolute upward mobility.

The fraction of children living in single-parent households is the strongest correlate of upward income mobility among all the variables we explored, with a raw unweighted correlation of -0.76 (see Online Appendix Figure VIIa for the corresponding non-parametric binned scatter plot). One natural explanation for this spatial correlation is an individual-level effect: children raised by a single parent may have worse outcomes than those raised by two parents (e.g., Thomas and Sawhill 2002, Lamb 2004). To test whether this individual-level effect drives the spatial correlation, we calculate upward mobility in each CZ based only on the subsample of children whose own parents are married. The correlation between upward mobility and the fraction of single parents in their CZ remains at -0.66 even in this subgroup (Online Appendix Figure VIIb). Hence, family structure correlates with upward mobility not just at the individual level but also at the community level, perhaps because the stability of the social environment affects children's outcomes more broadly. The association between mobility and family structure at the community level echoes our findings in Section V.A on the community-level effects of racial shares.

V.K Comparison of Alternative Explanations

The five factors that exhibit the strongest and most robust correlations with integenerational mobility are (1) segregation, (2) income inequality, (3) school quality, (4) social capital, and (5) family structure. We conclude our analysis by assessing which of these factors are the strongest predictors of upward mobility in multiple variable regressions.

Table IX reports estimates from regressions of absolute upward mobility or relative mobility on our preferred proxies for these five factors: the Theil index of racial segregation, the bottom 99% Gini coefficient, high school dropout rates adjusted for income differences, the social capital index, and the fraction of children with single parents. As in preceding regression specifications, the dependent and independent variables are normalized to have a standard deviation of 1 in the estimation sample in each regression in Table IX.[46]

We begin in Column 1 with an unweighted OLS regression of absolute upward mobility $\bar{y}_{25,c}$ on the five factors, pooling all CZs. All of the factors except the Gini coefficient are significant predictors of the variation in absolute upward mobility in this specification. Together, the five factors explain 70% of the variance in upward mobility across areas. Column 2 shows that the coefficients remain similar when state fixed effects are included. Column 3 reports population-

[46]We code the high school dropout rate as 0 for 126 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 that have missing data on social capital. We normalize these variables to have mean 0 and standard deviation 1 among the CZs with non-missing data.

weighted estimates, while Column 4 restricts the sample to urban areas (CZs that intersect MSAs) and reports unweighted estimates. The estimates in Columns 3 and 4 remain roughly similar to those in Columns 1 and 2, although the overall R-squared of the explanatory factors is lower in large cities. Across all the specifications, the strongest and most robust predictor is the fraction of children with single parents.

In Column 5, we use relative mobility β_c as the dependent variable instead of absolute upward mobility. The fraction of single-parents and the racial segregation index are strong predictors of differences in relative mobility across areas, but the other factors are not statistically significant.

To understand why this is the case, in Column 6 we replicate Column 5 but exclude the fraction of children with single parents. In this specification, all four of the remaining factors – including the Gini coefficient – are strong predictors of the variation in relative mobility across CZs. Column 7 replicates the specification in Column 6 using absolute upward mobility as the dependent variable. Once again, all four factors are strong predictors of upward mobility when the fraction of single parents is excluded. These results suggest that the fraction of single parents may capture some of the variation in the other factors, most notably the level of income inequality. Indeed, CZs with greater inequality have significantly higher rates of single parenthood. Hence, the results in Table IX are consistent with the view that inequality affects children's outcomes partly by degrading family structure and social ties in the community.

The last column of Table IX shows that when we regress absolute upward mobility on both the fraction of single-parent families in the CZ and the share of black residents, black shares are no longer correlated with upward mobility. This result further supports the view that the strong correlation of upward mobility with race operates through channels beyond the direct effect of race on mobility and demonstrates the tremendous explanatory power of the single-parent measure.

Overall, the results in Table IX indicate that the differences in upward mobility across areas are better explained by a combination of the factors identified above rather than any single factor. However, the regression coefficients should be interpreted with caution for two reasons. First, all of our proxies reflect latent factors that are measured with error. Because of measurement error, the regression may place greater weight on factors that are better measured rather than those that are truly the strongest determinants of mobility. Second, all of the independent variables are endogenously determined. These limitations make it difficult to identify which of the factors is the most important determinant of upward mobility.

VI Conclusion

This paper has used population-wide data to present a new portrait of intergenerational income mobility in the United States. Intergenerational mobility varies substantially across areas. For example, a child born in the bottom fifth of the income distribution has a 7.8% chance of reaching the top fifth in the U.S. as a whole. But in some places, such as Salt Lake City and San Jose, the chance of moving from the bottom fifth to the top fifth is as high as 12.9%. In others, such as Charlotte and Indianapolis, it is as low as 4.4%. The spatial variation in intergenerational mobility is strongly correlated with five factors: (1) residential segregation, (2) income inequality, (3) school quality, (4) social capital, and (5) family structure.

While we hope that these stylized facts will help guide the development of new models of intergenerational mobility, several important questions remain to be explored. Most importantly, the descriptive analysis in this paper does not identify the causal mechanisms that drive the differences in upward mobility across areas. This question can be broken into two sub-questions. First, is the variation in upward mobility across areas driven by sorting (differences in the characteristics of the residents) or causal place effects (differences in the institutions in an area)? Second, if place matters, what specific policies lead to improvements in upward mobility? In ongoing work, Chetty and Hendren (2014) study the first question by focusing on individuals who move across areas. To facilitate work on the second question, we have made our measures of upward and relative mobility as well as non-parametric quintile transition matrices and other CZ- and county-level statistics publicly available (www.equality-of-opportunity.org). For instance, these statistics could be used to study the impacts of place-based policies (Kline and Moretti 2014) on mobility.

Another dimension in which the analysis here could be expanded is by analyzing other moments of the joint distribution of parent and child income, such as persistence at the top (e.g., the top 1%) and the degree of downward mobility. The publicly available 100 x 100 percentile-level transition matrices constructed here can be used to explore these questions.

Finally, this paper has presented a snapshot of intergenerational mobility for a single set of birth cohorts. In a companion paper (Chetty et al. 2014), we use smaller random samples to study time trends in mobility. We find that the differences in mobility across areas documented here persist over time and are much larger than changes in mobility at the national level in recent decades. Understanding why some areas of the U.S. persistently generate higher rates of intergenerational mobility than others is an important challenge for future research.

ONLINE APPENDICES

A. Data Construction

Core and Extended Samples. We begin with the universe of individuals in the Death Master (also known as the Data Master-1) file produced by the Social Security Administration. This file includes information on year of birth and gender for all persons in the United States with a Social Security Number or Individual Taxpayer Identification Number. We restrict this sample to all individuals who are current US citizens as of March 2013. The Data Master-1 file does not contain historical citizenship status and thus we can only restrict to a sample who are currently US citizens as of the time at which we access the data. We further restrict to individuals who are alive through the end of 2012. The resulting dataset contains 47.8 million children across all cohorts 1980-1991 (Appendix Table I).

For each child, we define the parent(s) as the first person(s) who claim the child as a dependent on a 1040 tax form. If parents are married but filing separately, we assign the child both parents. To eliminate dependent claiming by siblings or grandparents, in the case of a potential match to married parents or single mothers, we require the mother to be age 15-40 at birth.[47] In the case of a match to a single father, we require the father to be age 15-40 at birth. If no such eligible match occurs in 1996, the first year of our data, we search subsequent years (through 2011) until a valid match is found.

Once we match a child to parent(s), we hold this definition of parents fixed regardless of subsequent dependent claims or changes in marital status. For example, a child matched to married parents in 1996 who divorce in 1997 will always be matched to the two original parents. Conversely, a child matched to a single parent in 1996 that marries in 1997 will be considered matched to a single parent, though spouse income will be included in our definition of parent income because we measure parent income at the family level in our baseline analysis.

We measure parent and child income, location, college attendance, and all other variables using data from the IRS Databank, a balanced panel covering all individuals in the United States who appear on any tax form between 1996-2012. To reduce the effects of outliers and measurement error in the upper tail of the income distribution, we use data from the IRS Statistics of Income (SOI) manually perfected cross-sectional files spanning 1996-2011 (see below for details on these files). The probability of being in the SOI sample increases with income, and approaches 1 for the highest income individuals or those whose adjusted gross income exceeds \$5 million. If an individual's adjusted gross income exceeds \$10 million, we look for the individual in the SOI sample; if present, we use the SOI measure of adjusted gross income and wage income as reported on a F1040 return. If not, we replace the adjusted gross income with the total wages reported on the filed F1040 contained in the databank. This adjustment affects 0.017% of parents in our core sample (or, equivalently, 1.7% of parents in the top 1% of the income distribution). Because the IRS Databank includes tax year 2012 whereas the SOI sample does not, we top code income at \$100 million for all individuals in 2012.

Statistics of Income Sample. Starting in 1987, the IRS Statistics of Income cross sections – which are stratified random samples of tax returns – contain dependent information, allowing us

[47] Children can be claimed as a dependent only if they are aged less than 19 at the end of the year (less than 24 if enrolled as a student) or are disabled. A dependent child is a biological child, step child, adopted child, foster child, brother or sister, or a descendant of one of these (for example, a grandchild or nephew). Children must be claimed by their custodial parent, i.e. the parent with whom they live for over half the year. Furthermore, the custodial parent must provide more than 50% of the support to the child. Hence, working children who support themselves for more than 50% cannot be claimed as dependents. See IRS Publication 501 for further details.

to link children to parents. We use the 1987-2011 SOI cross-sections to construct a sample of children born in the 1971-1991 birth cohorts and correct for errors in the upper tail of the income distribution as described above. For each SOI cross-section from 1987 to 2007, we first identify all dependent children between the ages of 12 and 16 who are alive at age 30. We then pool all the SOI cross-sections that give us information for a given birth cohort. For example, the 1971 cohort is represented by children claimed at age 16 in 1987, while the 1991 cohort is comprised by children claimed at ages 12-16 in 2003-2007. Using the sampling weights for the SOI cross-sections (see Internal Revenue Service (2013) for details), each cohort-level dataset is representative of the population of children claimed on tax returns between the ages of 12 and 16 in that birth cohort.

Unlike in the population-based samples, we do not limit the SOI sample to children who are currently citizens because citizenship data are not fully populated for birth cohorts prior to 1980 and because we begin from a sample of children claimed by parents rather than the universe of children who currently appear in the population data (which includes later immigrants). In the years where the SOI and population-based samples overlap, we obtain very similar estimates in both samples. The citizenship restriction has a minor impact because the vast majority of children claimed as dependents between the ages of 12-16 are U.S. citizens as adults. We also do not impose any age restrictions on the parents in the SOI sample. In the population-based sample, some children are claimed by different adults across years and the age restriction is useful to discriminate between these potential parents. In the SOI sample, each child can only be linked to the parents who claim her in the cross-section file, so the age restriction would not play such a role. In practice, this restriction has little impact, as the age distribution of parents in the SOI sample is very similar to that in the core sample using the population data.

Children whose parents are sampled in multiple SOI cross-sections appear multiple times in the SOI sample. There are 228,295 unique children in the SOI sample and 523,700 total observations. The SOI sample grows from 4,383 unique children in 1971 to 21,231 unique children in 1991 because we have more cross-sections to link parents to children in more recent cohorts and because the size of the SOI cross-sections has increased over time (Appendix Table II). To be consistent with the core sample definition of parent income, we define parent income as the 5-year average of parent family income from 1996-2000 in the IRS Databank.

We provide additional information on the SOI sample in Chetty et al. (2014). Using vital statistics counts, we show that the SOI sample represents roughly 85% of children in each birth cohort from 1971-1979, the cohorts we use to obtain estimates of intergenerational mobility after age 32 in Figure IIIa. We also show that summary statistics for the SOI sample are very similar to the core sample for the 1980-82 birth cohorts reported in Table I of this paper. Note that Chetty et al. (2014) compute parent income using the income of the parents in the single year of the parent-child match there, whereas we compute parent income as the five-year average over 1996-2000 here for consistency with results from the population data. Because we restrict to parents with positive income, this leads to a small difference in the SOI sample used across the two papers. For example, we have 4,384 children in the 1971 cohort, compared with 4,331 children in the sample used by Chetty et al. (2014).

Assignment of Children to Commuting Zones. Children are assigned ZIP codes of residence based on their parents' ZIP code on the form 1040 in which the parent is matched to the child. In the few cases where a parent files a F1040 claiming the child but does not report a valid ZIP code, we search information returns (such as W-2 and 1099-G forms) for a valid ZIP code in that year.

We map these ZIP codes to counties based on the 1999 Census crosswalk between ZIP codes and counties. We then aggregate counties into Commuting Zones using David Dorn's county-to-CZ crosswalk (download file E6). The counties in the U.S. Census Bureau crosswalk and in David Dorn's crosswalk are not identical because they correspond to county definitions at different points in time; in particular the U.S. Census Bureau crosswalk includes changes between 1990 and 1999. We make manual adjustments for changes that affected 200 or more people. Using this procedure, we identify the CZ of 38,839 ZIP codes. To better track individuals residing in ZIP codes that have been created since 1999, we add an additional 477 ZIP codes not valid in 1999 but appearing in the more up-to-date 2011 HUD-USPS crosswalk. For example, in 2007, Manhattan's ZIP code 10021 was split into three separate ZIP codes, resulting in the creation of new ZIP codes 10065 and 10075.

Of 9,864,965 children with non-missing ZIP codes in our core sample, 9,778,071 were assigned a childhood CZ using ZIP codes that existed in 1999; an additional 2,718 were assigned a CZ based on a ZIP code that existed in 2011 but not in 1999. For simplicity, we use the same crosswalk for all years of matching ZIP codes to CZs. We have verified that using year-specific crosswalks from ZIP codes to counties has a negligible effect on CZ assignments. All of the crosswalks we constructed are available on our project website.

Some of our specifications require tracking children's locations into adulthood using the ZIP code where they live as adults when we measure their income (e.g., for cost-of-living adjustments). We define a child's adult location using the latest non-missing ZIP code. We first search for a zipcode in their 1040 form in 2012, followed by their information returns in 2012. We then repeat this procedure for 2011 if we do not find a zipcode in 2012. This yields 9,834,975 non-missing child ZIP codes in adulthood. Of these, we match 9,537,283 to a CZ from a ZIP code using the 1999 crosswalk (i.e. this zipcode was in use in 1999) and an additional 198,317 using the later crosswalk because the ZIP code was created after 1999.

Construction of ZIP-Level Racial Shares. To construct Figure IXb and IXc, which condition on racial shares at the ZIP level, we need data on racial shares by ZIP code. The 2000 Census includes summary tables by ZIP code tabulation areas (ZCTAs) instead of ZIP code. ZCTAs are a U.S. Census Bureau geographical unit that in most cases correspond closely to ZIP codes, but sometimes do not. We use a ZCTA to ZIP Crosswalk from the John Snow Institute to assign each ZIP code a racial share based on Census 2000 ZCTA-level data from form P008.

CZ-Level Price Index. To measure real incomes, we first construct a CZ-level ACCRA price index using the 2010 ACCRA composite cost of living index (table 728) for "urbanized areas" in 2010, which we crosswalk to CZs as follows. First, we use the 2012Q1-2013Q1 correspondence (downloaded on 11/21/2013) to assign 298 out of the 325 urbanized areas to MSAs. Each county in an MSA was assigned the same value of the index. We then merge counties to CZs and calculate an unweighted mean of the index among non-missing values within the CZ. Some CZs had no counties within an MSA and were therefore assigned a missing value of the ACCRA index.

To construct a price index that covers all CZs, we regress the CZ-level ACCRA index on a quadratic in log population density (from the 2000 Census), a quadratic in log median housing values, the latitude and longitude of the CZ centroid, and state fixed effects. Housing values are the population-weighted mean of tract median housing values for owner-occupied units in the 1990 Census short form. Latitude and longitude are the mean latitude and longitude across counties within each CZ, obtained from the Census 2000 Gazetteer county-level data. The predicted values from this regression constitute our final price index that covers all CZs.

B. Comparison to Survey Datasets

In Appendix Table III, we compare selected moments of income distributions and other variables in the tax data to data from two nationally representative surveys that have been used in prior work on the income distribution: the 2011-12 CPS and the 2011-12 ACS. We restrict the ACS and CPS samples to citizens in the same birth cohorts as our core sample (1980-82). To the extent possible, we define all income variables to match the concepts in the tax data.

To assess whether our method of linking children to parents based on dependent claiming creates selection bias, we compute statistics in the tax data both on the full sample of all children in the 1980-82 birth cohorts who are current U.S. citizens and the core sample of children linked to parents. Because most children are linked to parents, the differences between these two samples is small, though children who lack valid parent matches have slightly lower earnings on average.

Overall, the tax data are very similar to the CPS and ACS. The sum of the sampling weights in our survey -based samples provide estimates of the size of the target population being sampled. This population is very similar in the tax data and the two surveys. The mean and median earnings levels are very similar, as are the fractions with non-zero income. Perhaps more surprisingly, the interquartile range (P75-P25) of earnings is also similar across the three data sources. If survey data were reported with classical measurement error, we would expect the interquartile range to be larger in survey sources. However, survey reports of income exhibit "mean reverting" measurement error which has the effect of reducing variability (Bound and Krueger 1991; Bound et al. 2001). Moreover, survey non-response tends to follow a U-shaped pattern (Kline and Santos 2013), with very high and low earning individuals being least likely to provide earnings responses, which can further reduce variability. The quantiles of family income also line up well across the three data sources, with the tax-based moments strongly resembling those from the ACS, perhaps because the ACS has a higher response rate for earnings than the CPS.

C. Comparison to Mazumder (2005)

Mazumder (2005) reports that even 5-year averages of parent earnings exhibit substantial attenuation bias because of long-lasting transitory shocks to income. This appendix provides further details on why we find much less attenuation bias than Mazumder.

Mazumder (2005, Table 4, row 1, page 246) obtains IGE estimates as high as 0.6 when using 15-year averages of parent income matched SIPP-SSA administrative data, 54% larger than his 4-year pooled estimate of 0.388. In contrast, we find little difference between IGEs based on five-year vs. fifteen-year averages of parent income both using our preferred rank-rank specification (Figure IIIb) and using a log-log IGE specification similar to that estimated by Mazumder. In particular, we obtain a log-log IGE of 0.366 using a 15-year average of parent family income, closer to the estimates of Solon (1992) than Mazumder's estimates.

We believe our results differ from Mazumder's findings because we directly observe income for all individuals in our data, whereas Mazumder imputes parent income based on race and education for up to 60% of the observations in his sample to account for top-coding in social security records. These imputations are analogous to instrumenting for parent income using race and education, an approach known to yield higher estimates of the IGE, perhaps because parents' education directly affects children's earnings (e.g., Solon, 1992, Zimmerman, 1992, Mulligan, 1997, Solon, 1999). Because the SSA earnings limit is lower in the early years of his sample, Mazumder imputes income for a larger fraction of observations when he averages parent income over more years (Mazumder 2005, Figure 3). As a result, Mazumder's estimates effectively converge toward IV estimates as he uses more years to calculate mean parent income, explaining why his estimates rise so sharply with the number of years used to measure parent income. Consistent with this explanation, when he drops imputed observations, his IGE estimates increase much less with the number of years used to measure parent income (Mazumder 2005, Table 6).

Mazumder also reports simulations of earnings processes showing that attenuation bias in the IGE should be substantial even when using five-year averages. However, he calibrates the parameters of the earnings process in his simulation based on estimates from survey data, which have much more noise than administrative records. If one replicates Mazumder's simulations using a smaller variance share for transitory shocks, one obtains results similar to ours in Figure IIIb, with little attenuation bias in estimates based on five-year averages.

To be clear, Mazumder himself acknowledges the potential bias due to imputation, as he recommends in his conclusion that "future research should attempt to verify the results here using long-term measures of permanent earnings from other sources that do not require the kind of imputations that were necessary in this study." Our analysis simply follows this recommendation.

D. Construction of CZ-Level Covariates

This appendix supplements Online Data Table IX by providing further details about our CZ covariates used in Section V. Online Data Table IX contains detailed descriptions of each CZ covariate and briefly describes the source of data for each variable. Here, we provide additional details on each data source along with links to original sources. As a reference, we provide Stata code on our website that constructs the final CZ-level covariates (data available in Online Data Table XIII) from the raw data downloaded from the links below.

Our source data are primarily at the ZIP code and county level. We map ZIP codes and counties to CZs using the procedure described in Appendix A. We compute CZ-level means of the ZIP- and county-level data using population-weighted averages, with population counts from the 2000 Census.

Race and Segregation. Racial shares are calculated from the 2000 census short form (SF1) table P008. Note that all Census data can be obtained by searching for the relevant census table on the U.S. Census Bureau's American FactFinder. The black share is defined as the number of people in a CZ that are black alone divided by the CZ population; the white share is calculated similarly. For the Hispanic share, the numerator is the number of people of any race that are Hispanic. We also calculate a residual category where the numerator is the number of people that are neither black alone nor white alone nor Hispanic.

We calculate several indices of social and economic segregation. We compute the racial segregation index using the census tract level data on racial shares from Table P008 from the 2000 Census. For segregation of affluence and poverty, we use the sample data from the 2000 census long form (SF3) on the income distribution of households in 1999 by census tract contained in table P052. Our formulas for the three income segregation measures are taken directly from Reardon (2011). We compute $H(p_k)$ for each of the 16 income groups given in table P052. We then estimate $H(p_{25})$ and $H(p_{75})$ in each CZ using the 4th order polynomial version of the weighted linear regression in equation 12 on page 23 of Reardon (2011). The overall segregation of income index is Reardon's rank-order index, which we compute from equation 13, where the δ vector is given in Appendix A4 of Reardon (2011).

To compute the commute time variable, we divide the number of workers that commute for less than 15 minutes by the total number of workers. The sample for both of these counts is restricted to workers that are at least 16 years old and do not work from home. Travel time data is from the 2000 census table P031.

Income Distributions. We compute mean income per working age adult by dividing aggregate household income in a CZ by the total number of people aged 16-64 in that CZ. These data come from the 2000 census table P054 and P008. The Gini coefficient, fraction middle class, and top 1% income share are computed using our sample of parents and the family income definitions used for the main analysis in this paper, but with family income top coded at $100 million in all years.

Taxes and Government Expenditures. We estimate local tax rates using data on tax revenue by county from the U.S. Census Bureau's 1992 Census of Government county-level summaries, which we downloaded from the ICPSR. In particular, Part 2 of the ICPSR download contains the county-level summaries. We define the tax rate in each CZ as follows. First, we calculate county tax revenue divided by the county population estimate for each county in the CZ. We then take a population-weighted mean across these counties to obtain a CZ-level mean per-person taxes. Finally, we divide mean per-person taxes by the Census 2000 estimate of nominal income per household from 2000 census table P052. We code the tax rate as missing for one CZ (Barrow, Alaska), which has a calculated tax rate that exceeds 1.

We compute total government spending per capita in each county using Census data on government expenditures by aggregating all county-level total expenditure categories and dividing by the 1992 county population estimates. We then construct a CZ-level measure by taking population-weighted means of expenditures per capita in the counties in each CZ. We code local government expenditures as missing for two CZs (Barrow, Alaska and Kotzebue, Alaska), which have unusually high expenditures per capita that exceed 50% of per capita income.

We measure state income tax progressivity as the difference between 2008 state income tax rates for incomes above $100,000 and incomes in the bottom tax bracket using data from the Tax Foundation. We obtain data on State EITC rates by year from Hotz and Scholz (2003). We calculate mean EITC rate for the years 1980-2001, setting the rate to zero for state-year pairs where there was no state EITC. Note that Wisconsin's state EITC rate depends on the number of children in a household; we use the rate for households with two children.

K-12 Education. We use the National Center for Education Statistics' Common Core of Data data for public schools for several of our K-12 covariates. School expenditures per student is taken from school-district data for the 1996-1997 fiscal year. We drop 7 CZs that are in the top 1% of the distribution of expenditures per student to reduce the influence of outliers.

We use school-level data on student-teacher ratios for the 1996-1997 school year. We drop the top 0.1% of observations, which have student-teacher ratios that exceed 100. We also drop approximately 10% of schools whose student-teacher ratios are recorded as being 0.

High school dropout rates are obtained from school-district data for the 2000-2001 school year. We code the dropout rate as missing in CZs in which more than 25% of school districts have missing data on dropout rates. We construct an income-adjusted measure of dropout rates using residuals from a CZ-level regression of the dropout rate on mean parent family income (from 1996-2000) in the core sample.

We obtain a standardized measure of grade 3-8 test scores from the National Math Percentile and National Reading Percentile series in the Global Report Card. The Global Report Card constructs national math and reading score percentiles for each school district by first calculating the district's state-level z-score for the share of students in the district scoring at the proficient level and above on statewide standardized tests, then adjusting for the relative difficulty of achieving the proficiency cutoff in each state, and finally recentering the distribution of adjusted district z-scores for each state at that state's z-score for average student scores on the National Assessment of Education Progress Exam. We calculate the student-weighted mean of the math and reading rankings over 2004, 2005, and 2007 in each CZ to arrive at our measure of mean test scores. We then construct a measure of income-adjusted test scores using the residual from a CZ-level regression of mean test scores on mean parent family income (from 1996-2000) in the core sample.

We construct enrollment-weighted means at the ZIP code level of all the school and school district level variables using the school and district ZIP codes provided in each of the data sources. We then take enrollment-weighted means across ZIP codes to construct CZ-level estimates using the ZIP to CZ crosswalk discussed in Appendix A.

Higher Education. We use the Integrated Postsecondary Education Data System (IPEDS) to construct our three measures of college access and quality. We restrict the sample to Title IV institutions that have undergraduate students, and are degree offering. The number of colleges per capita in each CZ is the number of institutions in the 2000 IPEDS in each CZ divided by the CZ population. We define college tuition as the mean in-state tuition and fees for first-time, full-time undergraduates for the institutions in each CZ. We define the enrollment-weighted mean graduation rate based on the 150% of normal time college graduation rate from IPEDS 2009, the first year for which this data is available. We construct a measure of income-adjusted graduation rates using the residual from a CZ-level regression of graduation rates on mean parent income in the core sample.

Local Labor Market Conditions. The labor force participation rate is defined as the number of people in the labor force by the total population in the sample of people that are at least 16 years old. These data are from the 2000 Census long form (SF3) in table P043. We compute the share of workers in manufacturing from the 2000 census in table P049; we divide the number of people working in manufacturing by the total number of workers.

The exposure to Chinese trade variable is the percentage change in imports per worker from China between 1990 and 2000. It is measured as the growth in imports allocated to a CZ, divided by the CZ work force in 1990 (with the growth rate defined as 10 times the annualized change). This variable was constructed by Autor et al. (2013) and provided to us by David Dorn.

The teenage labor force participation rate is defined in each CZ as the share of individuals who received one or more W-2's between the ages of 14 and 16. We calculate the teenage LFP rate using W-2 data for the 1985-1987 birth cohorts, the earliest cohorts for which we have W2 data at age 14.

Migration Rates. For inflow and outflow migration data, we use the county-to-county migration data from the Internal Revenue Service's Statistics of Income for 2004-2005. Inflow migration is the number of people moving into a CZ from counties in other CZs divided by the total CZ population; outflow migration is calculated similarly. We compute the share of each CZs population that is foreign born using sample data from the 2000 census (table P021) on the number of foreign born inhabitants divided by total CZ population. In both cases, total CZ population is the sum of county populations from the 2000 Census (table P008) over counties in the CZ.

Social Capital. For social capital, we use the 1990 county-level social capital index from Rupasingha and Goetz (2008). For religious affiliation, we use data on the self-reported number of religious adherents from the Association of Religion Data Archives at Pennsylvania State University. Data on crime rates are from the FBI's Uniform Crime Reporting program. We downloaded county-level data from the ICPSR and use the number of arrests for serious (part 1 index) violent crimes divided by the total covered population.

Family Structure. We define the share of single mothers in each county as the number of households with female heads (and no husband present) with own children present divided by the total number of households with own children present. These data from from the 2000 census long form (SF3) in table P015. We calculate the fraction married and fraction divorced in each county using the number of people that are married or divorced (in the sample of people that are 15 years and older) using data from the 2000 census in table P018.

References

Altonji, Joseph G. and David Card. 1989. "The Effects of Immigration on the Labor Market Outcomes of Natives." NBER Working Papers 3123, National Bureau of Economic Research, Inc.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The Geography of Trade and Technology Shocks in the United States." *American Economic Review*, *103* (3): 220–25.

Bailey, Martha J. and Susan M. Dynarski. 2011. "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion." NBER Working Paper 17633, National Bureau of Economic Research, Inc.

Becker, Gary S. 1991. *A Treatise on the Family*, Cambridge, Mass.: Harvard University Press.

Becker, Gary S. and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*, *87* (6): 1153–89.

Becker, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, *4* (3): S1–39.

Bjorklund, Anders and Markus Jäntti. 1997. "Intergenerational Income Mobility in Sweden Compared to the United States." *American Economic Review*, *87* (5): 1009–18.

Black, Dan, Seth Sanders, Evan Taylor, and Lowell Taylor. 2011. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *Revise and Resubmit, American Economic Review.*

Black, Sandra E. and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, chapter 16, pp. 1487–1541.

Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *The Quarterly Journal of Economics*, *107* (1): 123–50.

Borjas, George J., Richard B. Friedman, and Lawrence F. Katz. 1997. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity*, *28* (1): 1–90.

Boserup, Simon, Wojciech Kopczuk, and Claus Kreiner. 2013. "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations." Univ. if Copenhagen mimeo.

Bound, John and Alan B Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, *9* (1): 1–24.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, chapter 59, pp. 3705–3843.

Card, David. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." in "Progress on Some Persistent Econometric Problems." Econometrica 69 (September)" 1993. pp. 1127–1160.

Card, David and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, *100* (1): 1–40.

Chadwick, Laura and Gary Solon. 2002. "Intergenerational Income Mobility Among Daughters." *American Economic Review*, *92* (1): 335–344.

Chetty, Raj and Nathan Hendren. 2014. "The Value-Added of Neighborhoods: New Evidence on the Effects of Neighborhoods on Childrens' Outcomes." Harvard Univ. mimeo (in preparation).

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2013. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." NBER Working Papers 19424, National Bureau of Economic Research, Inc.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics*, *126* (4): 1593–1660.

Chetty, Raj, Nathan Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review Papers and Proceedings.* forthcoming.

Cilke, James. 1998. "A Profile of Non-Filers." U.S. Department of the Treasury, Office of Tax Analysis Working Paper No. 78.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology*, *94*, pp. S95–S120.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives*, *27* (3): 79–102.

Corak, Miles and Andrew Heisz. 1999. "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data." *Journal of Human Resources*, *34* (3): 504–533.

Cutler, David M. and Edward L. Glaeser. 1997. "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics*, *112* (3): 827–72.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, *107* (3): 455–506.

Dahl, Molly W. and Thomas DeLeire. 2008. *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*, University of Wisconsin-Madison, Institute for Research on Poverty.

Dorn, David. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." PhD dissertation, University of St. Gallen no. 3613, 2009.

Duncan, Greg J., Kathleen M. Ziol-Guest, and Ariel Kalil. 2010. "Early-Childhood Poverty and Adult Attainment, Behavior, and Health." *Child Development*, *81* (1): 306–325.

Fields, Gary S. and Efe A. Ok. 1999. "The Measurement of Income Mobility: An Introduction to the Literature." in Jacques Silber, ed., *Handbook of Income Inequality Measurement*, Vol. 71 of Recent Economic Thought Series, Springer Netherlands, pp. 557–598.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. "Long-Term Effects of Class Size." *The Quarterly Journal of Economics*, *128* (1): 249–285.

Graham, Bryan and Patrick Sharkey. 2013. "Mobility and the Metropolis: How Communities Factor Into Economic Mobility." *Washington D.C.: Economic Mobility Project, The Pew Charitable Trusts.*

Grawe, Nathan D. 2006. "Lifecycle bias in estimates of intergenerational earnings persistence." *Labour Economics*, *13* (5): 551–570.

Grawe, Nathan D. and Casey B. Mulligan. 2002. "Economic Interpretations of Intergenerational Correlations." *Journal of Economic Perspectives*, *16* (3): 45–58.

Haider, Steven and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review*, *96* (4): 1308–1320.

Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher*, *18* (4): 45–62.

Hanushek, Eric A. 2003. "The Failure of Input-Based Schooling Policies." *Economic Journal*, *113* (485): F64–F98.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science*, *312* (5782): 1900–1902.

Hertz, Tom. 2006. "Understanding mobility in America." *Center for American Progress Discussion Paper.*

Hotz, Joseph V. and John K. Scholz. "The Earned Income Tax Credit." in Rober A. Moffitt, ed., *Means-Tested Transfer Programs in the United States*, University of Chicago Press 2003. pp. 141–197.

Iceland, John. 2004. "Beyond Black and White: Metropolitan residential segregation in multi-ethnic America." *Social Science Research*, *33* (2): 248–271.

Ichino, Andrea, Loukas Karabarbounis, and Enrico Moretti. 2011. "The Political Economy Of Intergenerational Income Mobility." *Economic Inquiry*, *49* (1): 47–69.

Internal Revenue Service. 2013. "Statistics of Income: Individual Income Tax Returns, 2012." Technical Report, Washington, DC: government printing press.

Jencks, Christopher and Susan E. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." in L. Lynn and M. G. H. McGeary, eds., *Inner City Poverty in the United States*, Washington D.C.: National Academy Press, p. 111–186.

Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Eva Österbacka, Anders Björklund, and Tor Eriksson. 2006. "American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States." IZA Discussion Paper 1938, Institute for the Study of Labor (IZA).

Kain, John F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *The Quarterly Journal of Economics*, *82* (2): 175–197.

Kasarda, John D. 1989. "Urban Industrial Transition and the Underclass." *Annals of the American Academy of Political and Social Science*, *501*, pp. 26–47.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving To Opportunity In Boston: Early Results Of A Randomized Mobility Experiment." *The Quarterly Journal of Economics*, *116* (2): 607–654.

Kline, Patrick and Andres Santos. 2013. "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure." *Quantitative Economics*, *4* (2): 231–267.

Kline, Patrick and Enrico Moretti. 2014. "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics*, *6* (1): forthcoming.

Krueger, Alan. "The Rise and Consequences of Inequality in the United States." Speech at the Center for American Progress, Washington D.C. on January 12, 2012.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *The Quarterly Journal of Economics*, *114* (2): 497–532.

Lamb, Michael E. 2004. *The Role of the Father in Child Development*, Hoboken, N.J.: Wiley.

Lee, Chul-In and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics*, *91* (4): 766–772.

Massey, Douglas S and Nancy A Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*, Cambridge, Mass.: Harvard University Press.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics*, *87* (2): 235–255.

Mulligan, Casey B. 1997. *Parental priorities and economic inequality*, Chicago: University of Chicago Press.

Murray, Charles A. 1984. *Losing ground: American social policy, 1950-1980*, New York: Basic Books.

Murray, Charles A. 2012. *Coming apart: the state of white America, 1960-2010*, New York, N.Y.: Crown Forum.

Oreopoulos, Philip. 2003. "The Long-Run Consequences of Living in a Poor Neighborhood." *The Quarterly Journal of Economics*, *118* (4): 1533–1575.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics*, *118* (1): 1–41.

Putnam, Robert D. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy*, *6* (1): 65–78.

Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies*, *30* (2): 137–174.

Ray, Debraj. 2010. "Uneven Growth: A Framework for Research in Development Economics." *Journal of Economic Perspectives*, *24* (3): 45–60.

Reardon, Sean F. 2011. "Measures of income segregation." *CEPA Working Papers.* Stanford, CA: Stanford Center for Education Policy Analysis.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology*, *32* (1): 33–67.

Rupasingha, Anil and Stephan J. Goetz. 2008. "US County-Level Social Capital Data, 1990-2005." *The Northeast Regional Center for Rural Development, Penn State University, University Park, PA.*

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "Neighborhood Effects": Social Processes and New Directions in Research." *Annual Review of Sociology*, *28* (1): 443–478.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review*, *82* (3): 393–408.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3, Elsevier, pp. 1761–1800.

Solon, Gary. 2002. "Cross-Country Differences in Intergenerational Earnings Mobility." *Journal of Economic Perspectives*, *16* (3): 59–66.

Solon, Gary. 2004. "A Model of Intergenerational Mobility Variation Over Time and Place." in Miles Corak, ed., *Generational Income Mobility in North America and Europe*, Cambridge University Press, pp. 38–47.

Solon, Gary, Marianne E. Page, and Greg J. Duncan. 2000. "Correlations Between Neighboring Children In Their Subsequent Educational Attainment." *The Review of Economics and Statistics*, *82* (3): 383–392.

Theil, Henri. 1972. *Statistical decomposition analysis. With applications in the social and administrative sciences* number v. 14. In 'Studies in mathematical and managerial economics.', Amsterdam, New York: North-Holland Pub. Co.; American Elsevier Pub. Co.

Thomas, Adam and Isabel Sawhill. 2002. "For richer or for poorer: Marriage as an antipoverty strategy." *Journal of Policy Analysis and Management*, *21* (4): 587–599.

Tolbert, Charles M. and Molly Sizer. 1996. "U.S. Commuting Zones and Labor Market Areas: A 1990 update." *Economic Research Service Staff Paper*, *9614.*

Wilson, William J. 1987. *The truly disadvantaged: the inner city, the underclass, and public policy*, Chicago: University of Chicago Press.

Wilson, William J. 1996. *When work disappears: the world of the new urban poor*, 1st ed ed., New York: Knopf : Distributed by Random House, Inc.

Zimmerman, David J. 1992. "Regression toward Mediocrity in Economic Stature." *American Economic Review*, *82* (3): 409–29.

TABLE I

Summary Statistics for Core Sample: Children Born in 1980-82

Variable	Mean (1)	Std. Dev. (2)	Median (3)
Parents:			
Family Income (1996-2000 mean)	87,219	353,430	60,129
Top Earner's Income (1999-2003 mean)	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. The core sample of children 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. Child income is the average of 2011-2012 (when the child was 30) while parent family income is the average 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 II 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.

TABLE II
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)
1. Log family income (excluding zeros)	Log family income	0.344	0.349	0.342	0.303	0.264	0.316
		(0.0004)	(0.0006)	(0.0005)	(0.0005)	(0.0008)	(0.0003)
2. Log family income (recoding zeros to $1)	Log family income	0.618	0.697	0.540	0.509	0.528	0.580
		(0.0009)	(0.0013)	(0.0011)	(0.0011)	(0.0020)	(0.0006)
3. Log family income (recoding zeros to $1000)	Log family income	0.413	0.435	0.392	0.358	0.322	0.380
		(0.0004)	(0.0007)	(0.0006)	(0.0006)	(0.0009)	(0.0003)
4. Family income rank	Family income rank	0.341	0.336	0.346	0.289	0.311	0.323
		(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0002)
5. Family income rank	Top parent income rank	0.312	0.307	0.317	0.256	0.253	0.296
		(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0002)
6. Individual income rank	Family income rank	0.287	0.317	0.257	0.265	0.279	0.286
		(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0002)
7. Individual earnings rank	Family income rank	0.282	0.313	0.249	0.259	0.273	0.283
		(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0002)
8. College Attendance	Family income rank	0.675	0.708	0.644	0.641	0.663	0.678
		(0.0005)	(0.0007)	(0.0007)	(0.0006)	(0.0013)	(0.0003)
9. Teenage birth (females only)	Family income rank	-0.298			-0.231	-0.322	-0.285
		(0.0006)			(0.0007)	(0.0016)	(0.0004)
Number of observations		9,867,736	4,935,804	4,931,066	6,854,588	3,013,148	20,520,588

Notes: Each cell in this table reports the coefficient from a univariate OLS regression of the variable for children (listed in the first column) on the variable for parents (listed in the second column) for the corresponding sample (listed in columns 1-6). Column 1 uses the core sample (1980-82 birth cohorts); see notes to Table 1 for further details on the definition of the core sample. 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. 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). See notes to Table 1 for definition of child individual income and earnings. In columns 1-5, 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. In column 6, parents are ranked relative to other parents with children in the 1980-85 birth cohorts. 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 as having a child while between age 13 and 19. The number of observations corresponds to the specification in row 4. The number of observations is approximately 7% lower in row one because we exclude children with zero income. The number of observations is approximately 50% lower in row 8 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 III

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 1 for income and sample definitions. See Online Appendix Table IV for an analogous transition matrix constructed using the 1980-85 cohorts.

TABLE IV
Intergenerational Mobility in the 50 Largest Commuting Zones

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

Notes: This table reports our baseline 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 1. The measures in columns 4 and 6 are both derived from within-CZ OLS regressions of child income rank against parent income rank. Column 6 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 (Y_{25}). Under linearity of the rank-rank relationship, this is equal to the average rank of children with parents in the bottom half of the distribution. 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. See Online Data Table V for estimates for all CZs as well as estimates using alternative samples and income definitions. See Online Data Tables III and IV for estimates by county and MSA.

TABLE V
Robustness of Intergenerational Mobility Measures to Alternative Specifications

Change from Baseline Specification	Correlation with Baseline Mobility Estimates: 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	0.94	0.98	0.98
2. Female children	0.98	0.95	0.97	0.98
3. Children of married parents	0.97	0.89	0.91	0.93
4. Children of single parents	0.97	0.61	0.97	0.83
5. Birth cohorts 1983-85	0.97	0.84	0.96	0.96
6. Birth cohorts 1986-88	0.94	0.73	0.82	0.88
7. Parent age at child birth within 5 years of median	0.98	0.90	0.98	0.96
8. Children who stay within CZ	0.94	0.87	0.93	0.95
9. Children matched to unique parents	0.99	0.98	0.98	0.99
B. Alternative Income Definitions				
10. Top parent income	1.00	0.97	0.99	0.99
11. Individual child income	0.94	0.89	0.83	0.95
12. Individual child earnings	0.93	0.86	0.82	0.93
13. Individual child income (males only))	0.96	0.90	0.96	0.95
14. Indiv child income and top parent income (males only)	0.97	0.87	0.97	0.94
C. Cost of living, local growth, and other factors				
15. Cost of living adjusted income	0.98	0.99	0.86	0.99
16. Within-CZ ranks		0.95		0.96
17. Prob. Child in Q5 | Parent in Q1	0.91		0.92	
18. Parent income measured in 2011/12	0.97	0.92	0.94	0.98
19. Controlling for growth	0.83	0.92	0.81	0.96
D. Alternative Child Outcomes				
20. College Attendance (age 18-21)	0.71	0.68	0.53	0.72
21. Teenage Birth, females only	-0.61	-0.58	-0.64	-0.68

Notes: Each cell in 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, defined using a different sample (Panel A), a different income measure for parents or children (Panel B), adjusting for cost of living or other factors (Panel C), or using earlier outcomes (Panel D). Column (1) reports the unweighted correlation of the alternative and baseline measures of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile. Column (2) reports the unweighted correlation of the alternative and baseline measures of relative mobility, the slope of the rank-rank relationship. Columns (3) and (4) repeat Columns (1) and (2) weighting the correlations by CZ population as recorded in the 2000 Census. All absolute and relative mobility measures are constructed using OLS regressions of child ranks (or college or teenage birth indicators) on parent ranks as described in the text. Ranks are always defined in the full sample, prior to defining specific subsamples. See text for further details on each alternative measure.

TABLE VI

Segregation and Intergenerational Mobility

Dep. Var.:	Upward Mobility Y_{25}						
	(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. 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. 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. Column 2 and 5 restrict to the sample of CZs that intersect an MSA. Racial segregation is measured by the Theil index defined in equation (3) 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); see equation (4). 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.

TABLE VII

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)	
Segregation	Fraction Black Residents	-0.585	(0.065)	-0.358	(0.049)	-0.607	(0.077)	-0.678	(0.062)			0.627	(0.048)
	Racial Segregation Theil Index	-0.361	(0.045)	-0.274	(0.027)	-0.311	(0.092)	-0.360	(0.068)	-0.277	(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)	-0.270	(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)	-0.277	(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)	-0.253	(0.053)	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)	0.411	(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	(0.078)	-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)	-0.357	(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)	-0.070	(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)	-0.465	(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)	0.455	(0.145)	-0.451	(0.109)
Tax	Local Tax Rate	0.319	(0.073)	0.139	(0.069)	0.202	(0.085)	0.217	(0.073)	0.182	(0.076)	-0.323	(0.067)
	Local Government Expenditures per Capita	0.186	(0.083)	0.074	(0.028)	0.192	(0.087)	0.085	(0.079)	0.105	(0.083)	-0.301	(0.080)
	State EITC Exposure	0.245	(0.064)			0.279	(0.076)	0.355	(0.073)	0.160	(0.073)	-0.144	(0.047)
	State Income Tax Progressivity	0.207	(0.146)			0.261	(0.069)	0.197	(0.098)	0.155	(0.133)	-0.150	(0.106)
K-12 Education	School Expenditure per Student	0.246	(0.095)	0.026	(0.099)	0.219	(0.088)	0.236	(0.092)	0.046	(0.083)	-0.279	(0.092)
	Teacher-Student Ratio	-0.328	(0.100)	-0.213	(0.128)	0.062	(0.139)	0.024	(0.104)	-0.252	(0.088)	0.009	(0.108)
	Test Score Percentile (Controlling for Parent Income)	0.589	(0.087)	0.465	(0.074)	0.183	(0.219)	0.423	(0.147)	0.395	(0.092)	-0.322	(0.122)
	High School Dropout Rate (Controlling for Parent Income)	-0.584	(0.082)	-0.422	(0.064)	-0.452	(0.098)	-0.488	(0.098)	-0.447	(0.085)	0.337	(0.097)
College	Number of Colleges per Capita	0.194	(0.109)	-0.027	(0.111)	0.118	(0.085)	-0.033	(0.085)	0.049	(0.137)	-0.128	(0.049)
	Mean College Tuition	-0.018	(0.067)	-0.044	(0.039)	0.058	(0.097)	-0.015	(0.087)	-0.032	(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)	0.172	(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)	0.145	(0.072)	-0.237	(0.082)
	Fraction Working in Manufacturing	-0.261	(0.091)	0.007	(0.079)	-0.158	(0.090)	-0.129	(0.096)	0.006	(0.084)	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	0.629	(0.087)	0.356	(0.098)	0.299	(0.153)	0.540	(0.109)	0.380	(0.090)	-0.518	(0.084)
Migration	Migration Inflow Rate	-0.258	(0.074)	-0.185	(0.050)	-0.146	(0.076)	-0.040	(0.078)	-0.280	(0.069)	-0.085	(0.067)
	Migration Outflow Rate	-0.163	(0.070)	-0.161	(0.048)	0.062	(0.094)	0.014	(0.075)	-0.140	(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)	-0.001	(0.051)	-0.247	(0.055)
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)	0.473	(0.097)	-0.327	(0.085)
	Fraction Religious	0.522	(0.085)	0.358	(0.060)	0.410	(0.096)	0.417	(0.096)	0.484	(0.065)	-0.103	(0.090)
	Violent Crime Rate	-0.346	(0.130)	-0.168	(0.061)	-0.073	(0.159)	-0.296	(0.153)	-0.215	(0.053)	0.192	(0.129)
Family Structure	Fraction of Children with Single Mothers	-0.764	(0.074)	-0.571	(0.085)	-0.613	(0.129)	-0.719	(0.063)	-0.606	(0.069)	0.641	(0.046)
	Fraction of Adults Divorced	-0.486	(0.100)	-0.333	(0.085)	-0.389	(0.074)	-0.346	(0.103)	-0.571	(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)

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 V, Online Data Table IX, and Online Appendix D for definitions of each of the correlates analyzed in this table. See Online Data Table VII for the CZ-level data on each covariate.

TABLE VIII

Income Inequality and Intergenerational Mobility: The "Great Gatbsy" Curve

Dep. Var.:	Across CZs within the U.S. Upward mobility (1)	Upward mobility (2)	Upward mobility (3)	Upward mobility (4)	Across CZs within the U.S. Relative mobility (5)	Across Countries Log-Log Elasticity 1985 Inequality (6)	Log-Log Elasticity 1985 Inequality (7)	Across Countries Log-Log Elasticity 2005 Inequality (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)
Fraction 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; in column 5 the dependent variable is relative mobility. 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 column 6 and 7 and in 2005 in column 8.

TABLE IX
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)
Racial Segregation	-0.086	-0.111	-0.042	-0.109	0.190	0.281	-0.166	
	(0.028)	(0.020)	(0.091)	(0.045)	(0.044)	(0.047)	(0.032)	
Gini Bottom 99%	-0.042	-0.021	0.069	0.012	0.038	0.338	-0.307	
	(0.062)	(0.038)	(0.093)	(0.113)	(0.076)	(0.093)	(0.063)	
High School Dropout Rate	-0.152	-0.132	-0.262	-0.155	0.019	0.166	-0.282	
	(0.052)	(0.028)	(0.100)	(0.086)	(0.059)	(0.051)	(0.059)	
Social Capital Index	0.291	0.109	0.120	0.269	0.044	0.028	0.304	
	(0.060)	(0.054)	(0.086)	(0.068)	(0.061)	(0.074)	(0.069)	
Fraction Single Mothers	-0.489	-0.444	-0.537	-0.510	0.553			-0.791
	(0.072)	(0.073)	(0.114)	(0.103)	(0.063)			(0.088)
Fraction Black								0.035
								(0.073)
State Fixed Effects		x						
Population Weighted			x					
Urban Areas Only				x				
R-Squared	709	709	709	325	709	709	709	709
Observations	0.698	0.847	0.441	0.605	0.459	0.329	0.596	0.584

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 1-4 and 7-8 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 5 and 6 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. Column 3 weights by population (and normalizes variables by weighted standard deviations). In column 4, we restrict to CZs that intersect MSAs. Columns 5-8 replicate the unweighted specification in Column 1 with different dependent and independent variables. Racial segregation is measured by the Theil index defined in equation (3) using racial shares at the census tract level. 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. We code the high school dropout rate as 0 for 128 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. 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. See Section V, Online Data Table IX, and Online Appendix D 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	4,384	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	523,700	228,295

Notes: This table reports the sample size for the Statistics of Income stratified random sample by birth cohort. Column 2 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

Comparison of Administrative Tax Data to CPS and ACS Survey Datasets

	Tax Data Full Sample	Tax Data Core Sample	2011-2012 CPS	2011-2012 ACS	Tax Data Full Sample	Tax Data Core Sample	2011-2012 CPS	2011-2012 ACS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Distribution:	*Earned Family Income*				*Total Family Income*			
% Zero	9.74%	7.32%	9.23%	12.72%	8.54%	6.11%	5.44%	8.04%
% Negative	0.00%	0.00%	0.00%	0.00%	0.33%	0.34%	0.04%	0.05%
Mean	44,278	46,805	54,313	41,795	45,406	48,050	56,438	44,198
Std. Deviation	104,528	109,667	58,556	47,137	90,594	93,182	59,145	49,161
P10	63	1,624	1,307	0	521	2,810	6,431	1,500
P25	12,724	14,984	18,843	11,900	12,842	14,919	20,414	14,000
P50	32,165	34,737	40,829	31,000	32,273	34,975	42,768	32,800
P75	62,095	65,148	75,000	56,000	62,992	66,169	76,554	59,000
P90	96,995	99,911	115,000	90,000	98,802	101,770	118,050	95,000
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	194,501	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.

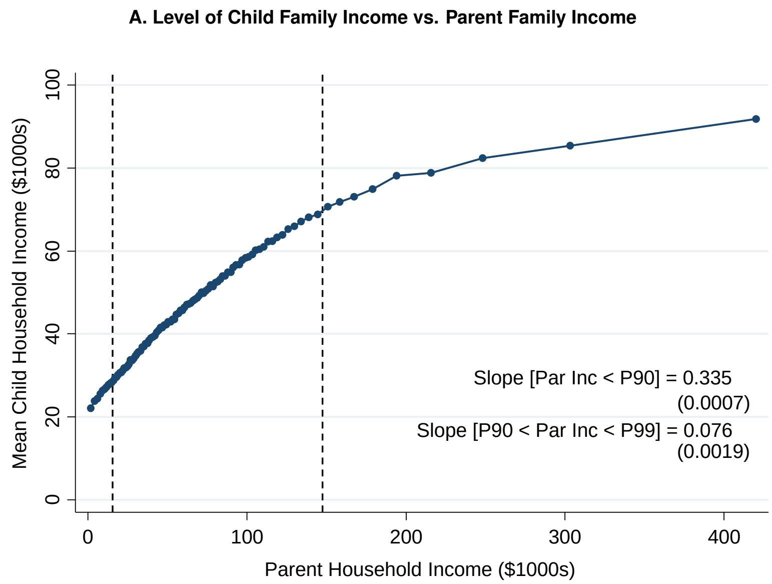
ONLINE APPENDIX TABLE IV

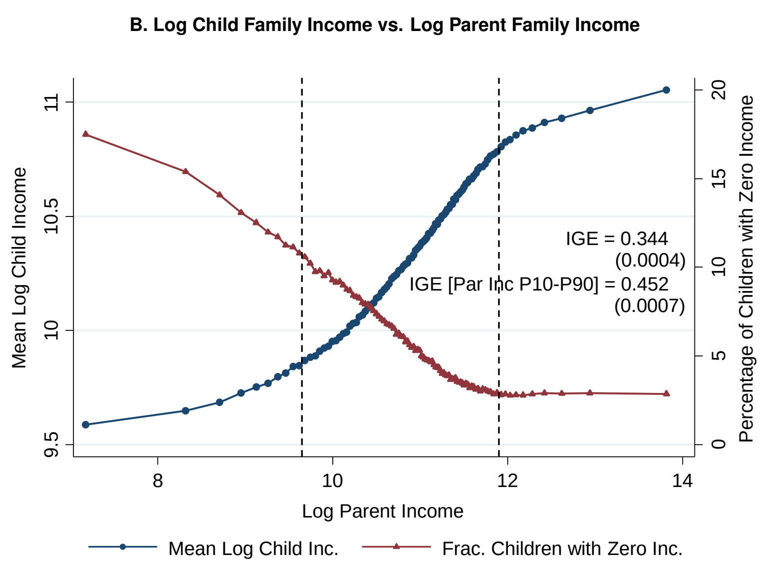
National Quintile Transition Matrix: 1980-85 Cohorts

		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 1 for income and sample definitions. See Table III for an analogous transition matrix constructed using the 1980-82 birth cohorts.

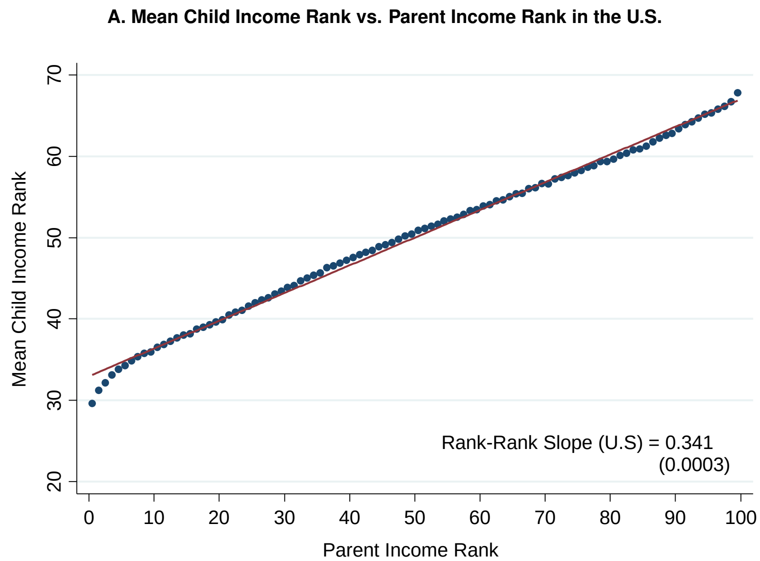
FIGURE I: Association between Children's and Parents' Income

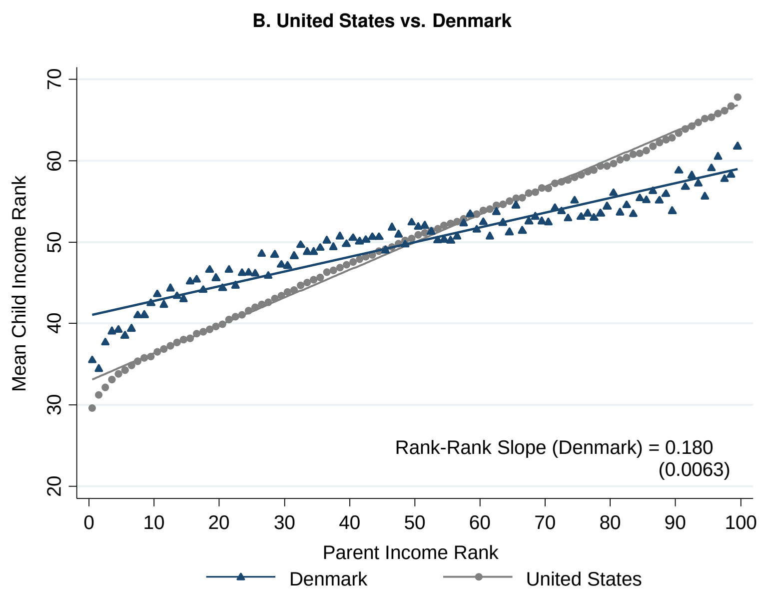




Notes: These figures present non-parametric binned scatter plots of the relationship between child income and parent income. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011-2012 family income (when the child was around 30), 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, as mean parent income in the top 1% is $1.4 million. 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 micro data. 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.

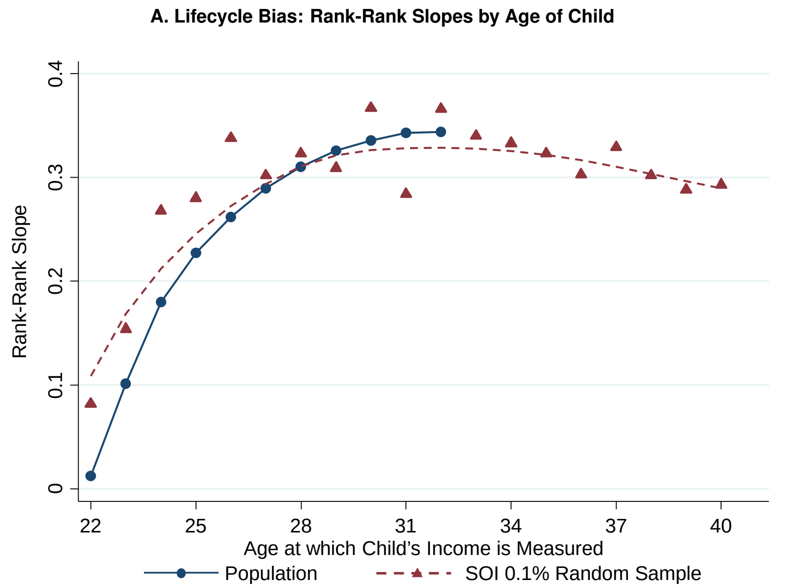
FIGURE II: Association between Children's Percentile Rank and Parents' Percentile Rank

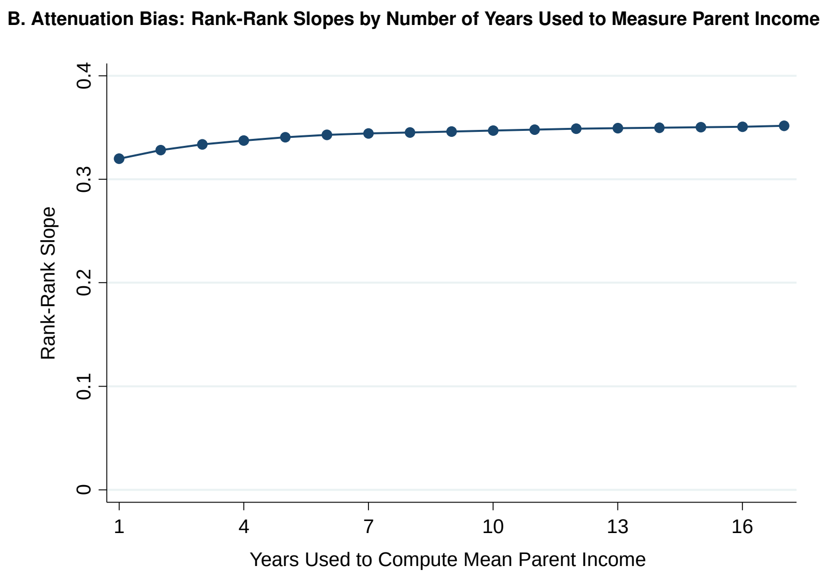




Notes: These figures present non-parametric binned scatter plots of the relationship between child and parent income ranks. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011-2012 family income (when the child was around 30), while parent income is mean family income from 1996-2000. We define a child's rank as her family income percentile rank relative to other children in her birth cohort and his parents' rank as their family income percentile rank relative to other parents of children in the core sample. Panel A plots the mean child percentile rank within each parental 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 (see text for details). The series in circles 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 micro data for the U.S. and on the binned series (as we do not have access to the micro data) for Denmark. Standard errors are reported in parentheses.

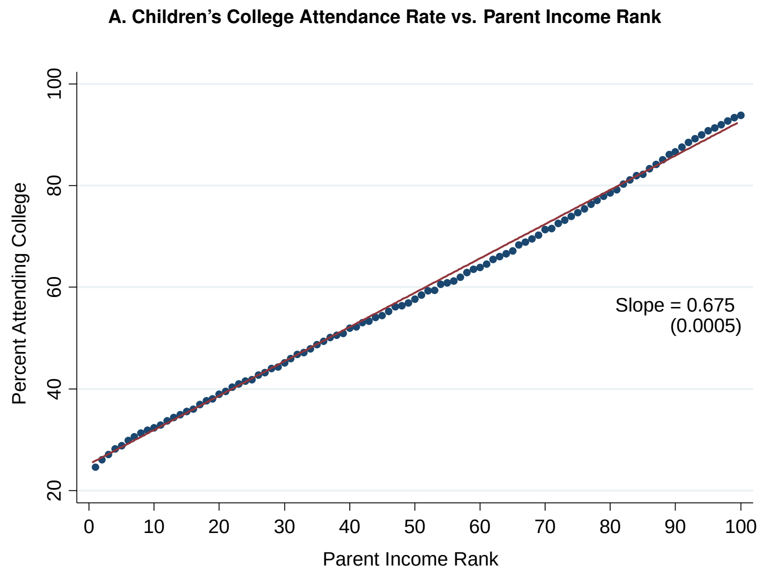
FIGURE III: Robustness of Intergenerational Mobility Estimates

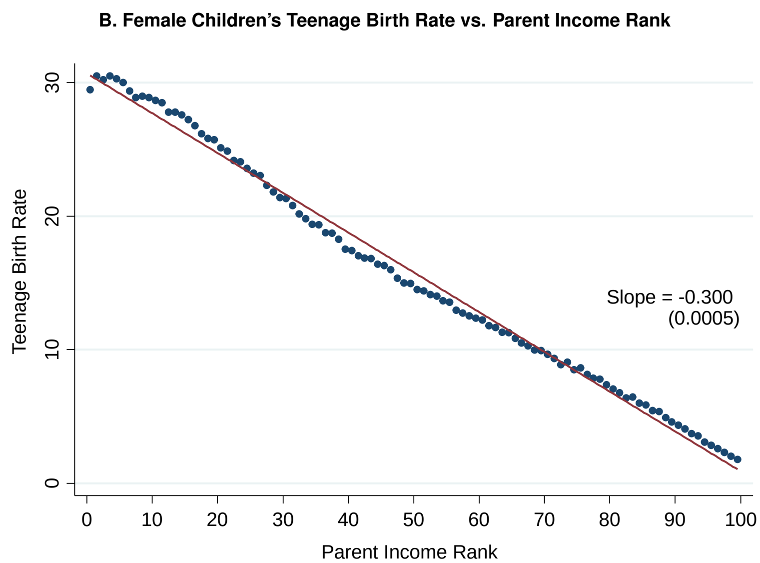




Notes: This figure evaluates the robustness of the rank-rank slope estimated in Figure IIa to changes in the age at which child income is measured (Panel A) and the number of years used to measure parents' income (Panel B). In both panels, child income is defined as mean family income in 2011-2012. In 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 age at which child income is measured in 2011-12. The blue dots use the extended sample in the population data, while the red 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 1972 cohort, who are 39-40 (denoted by 40) when their incomes are measured. The dashed red 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.

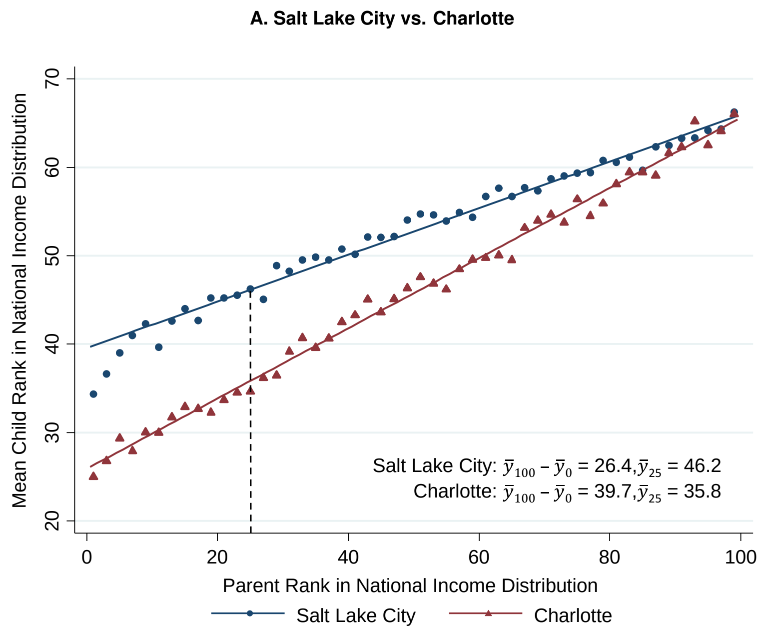
FIGURE IV: Gradients of College Attendance and Teenage Birth by Parent Rank

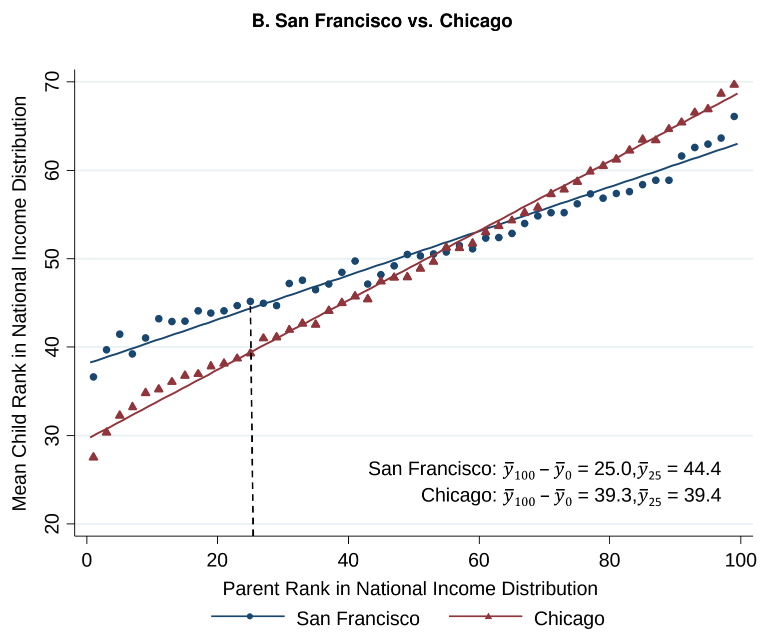




Notes: These figures present non-parametric binned scatter plots of the relationship between children's college attendance rates (Panel A) and teenage birth rates (Panel B) vs. the percentile rank of parent family income. Both figures are based on the core sample (1980-82 birth cohorts). Parent income is mean family income from 1996-2000. Panel A plots the fraction of children ever attending college between age 18-21 within each parental percentile bin. College attendance is defined by the presence of a 1098-T form filed by a college on behalf of the student. Panel B plots the fraction of female children who give birth while teenagers within each parental percentile bin. Teenage birth is defined as ever claiming a dependent child who was born while the mother was aged 13-19. The regression coefficients, standard errors, and best-fit lines are estimated on the micro data.

FIGURE V: Intergenerational Mobility in Selected Commuting Zones

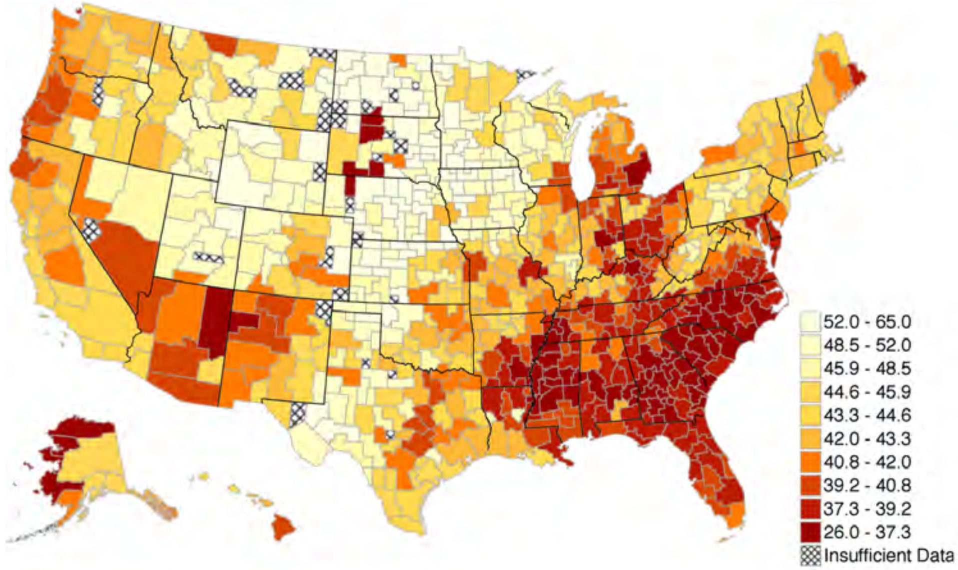




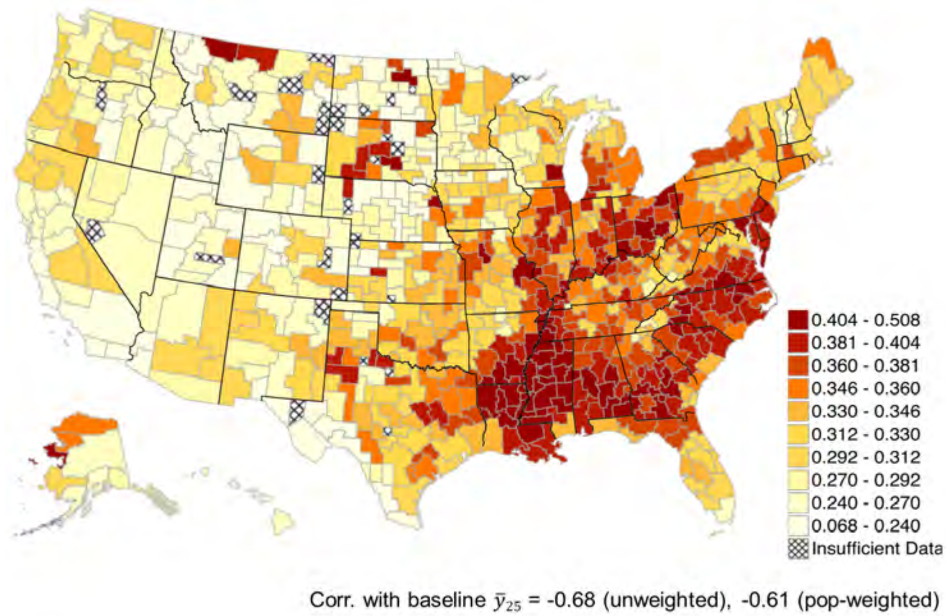
Notes: These figures present non-parametric binned scatter plots of the relationship between child and parent income ranks in selected CZs. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to commuting zones based on the location of their parents (when the child was claimed 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 ($\bar{y}_{100} - \bar{y}_0$), which is 100 times the rank-rank slope estimate. The second is absolute upward mobility ($\bar{y}_{25}$), 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 the micro data.

FIGURE VI: The Geography of Intergenerational Mobility

A. Absolute Upward Mobility: Average Child Rank for Below-Median Parents $(\bar{y}_{25})$ by CZ

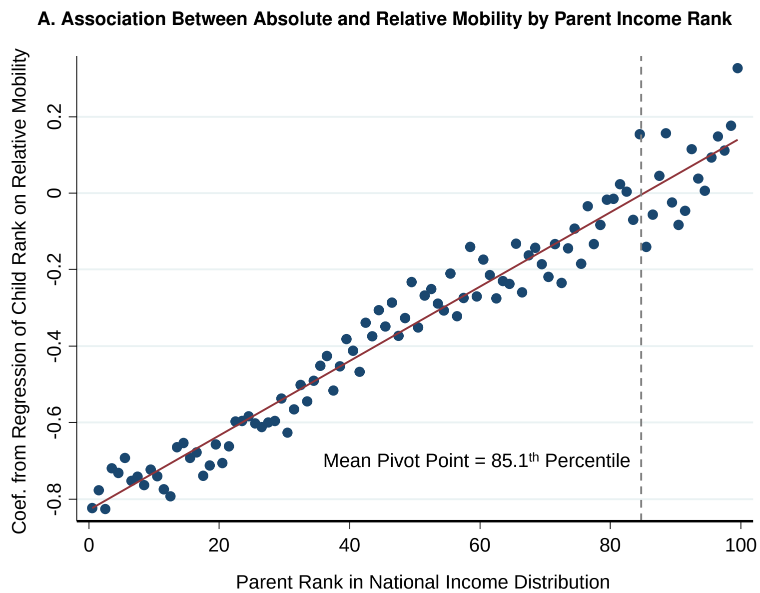


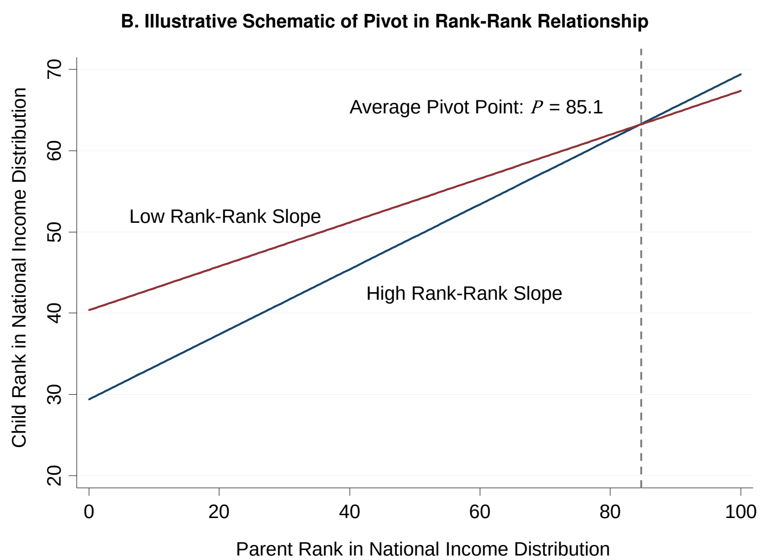
B. Relative Mobility: Rank-Rank Slopes $(\bar{y}_{100} - \bar{y}_0)/100$ by CZ



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 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 ($\bar{y}_{25}$) 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 ($\bar{y}_{100} - \bar{y}_0$). 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

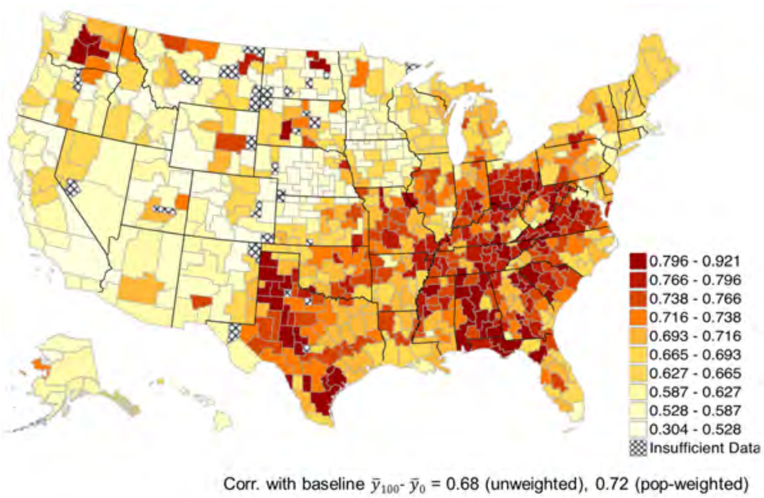




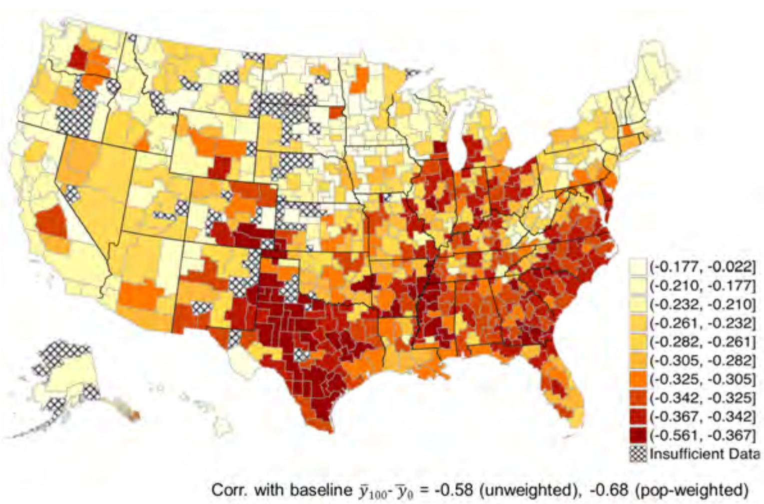
Notes: These figures illustrate the correlation between relative mobility and absolute mobility at various percentiles of the income distribution. To construct Panel A, we first calculate the mean income rank of children in CZ c with parents in (national) percentile p, denoted by $\bar{\mu}_{pc}$. We then run a CZ-level regression of $\bar{\mu}_{pc}$ on relative mobility ($\bar{y}_{100c} - \bar{y}_{0c}$) 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: The Geography of College and Teenage Birth Income Gradients

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

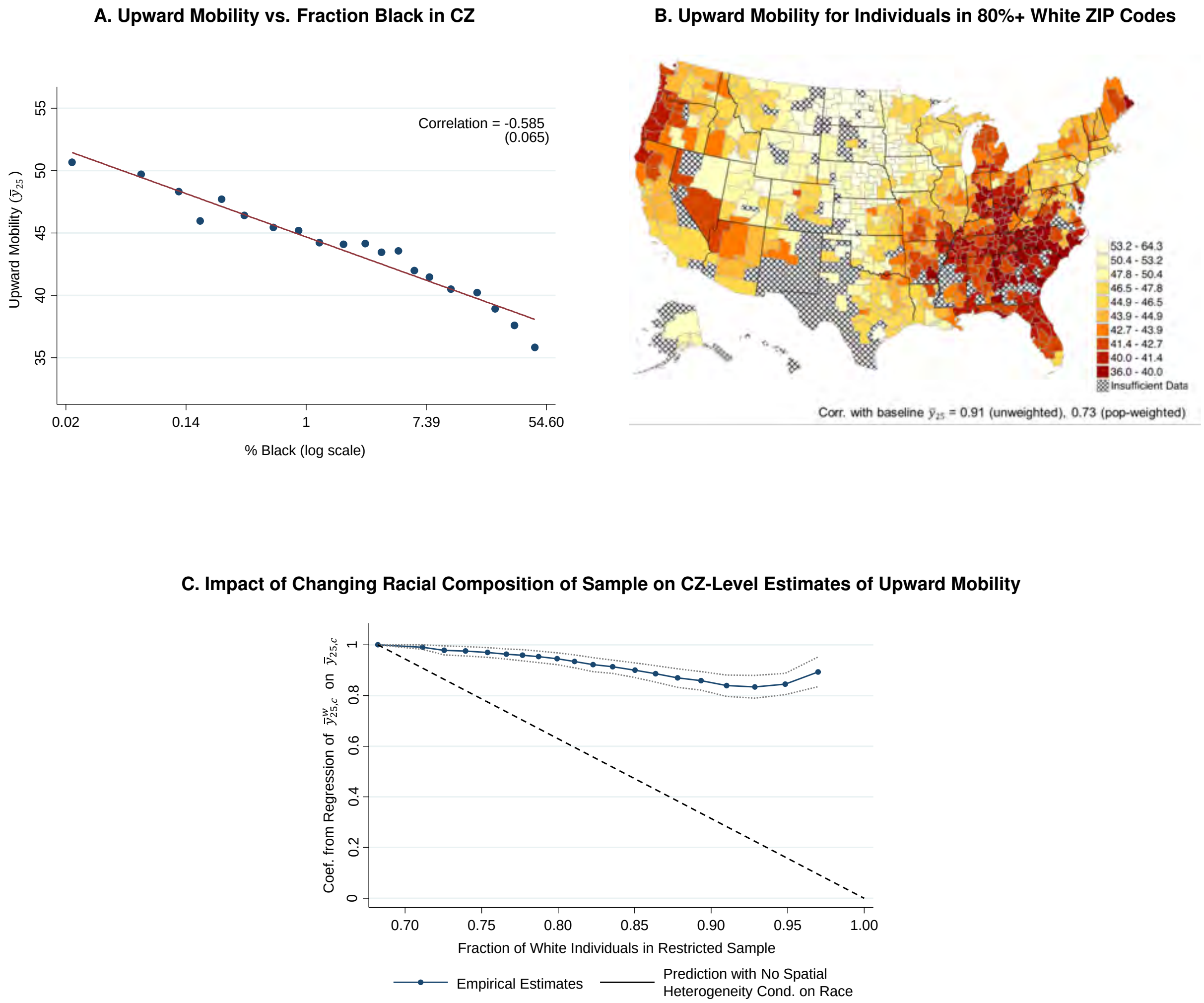


B. Slope of Teenage Birth-Parent Income Gradients by CZ



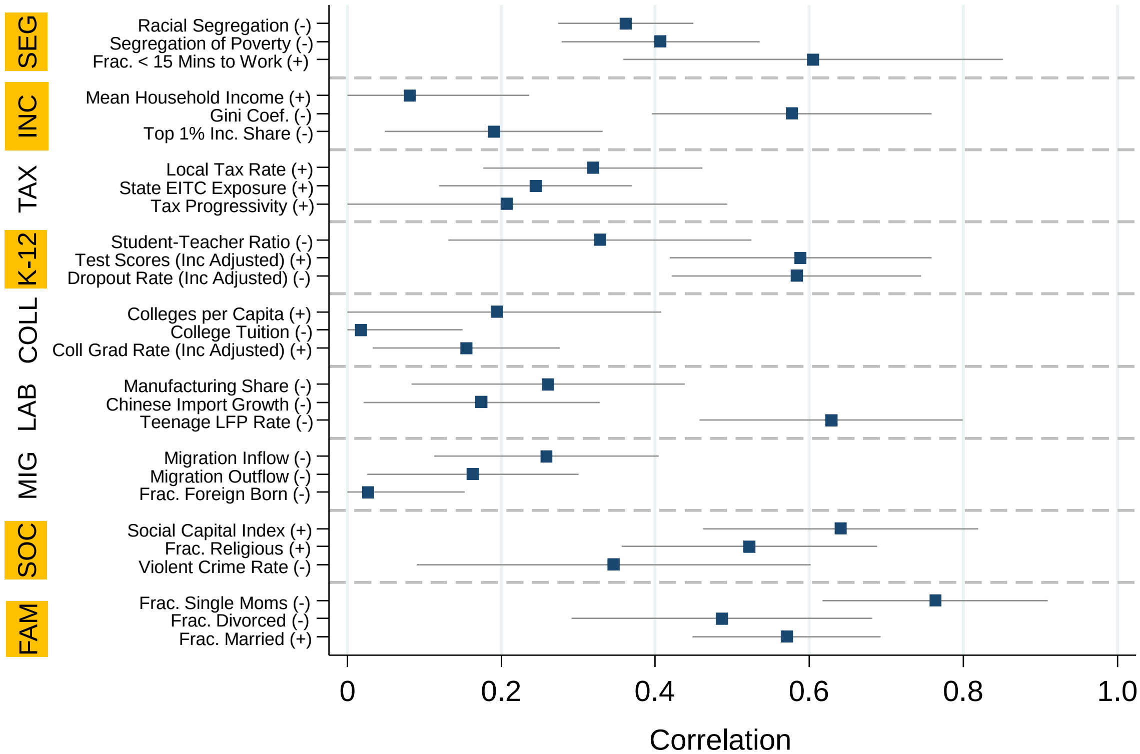
Notes: These figures present heat maps of college attendance and teenage birth rate by parent income gradients across commuting zones. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents. Children are assigned to commuting zones based on the location of their parents (when the child was claimed a dependent), irrespective of where they live as adults. In Panel A, we regress an indicator for the child ever attending college between ages 18-21 on parents' national income rank to estimate the slope by CZ, as shown in Figure IV at the national level. College attendance is defined by the presence of a 1098-T form filed by a college on behalf of the student. Panel B repeats this analysis using an indicator for teenage birth, restricting the sample to female children. Teenage birth is defined as ever claiming a dependent child who was born while the mother was aged 13-19. The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to smaller slopes in magnitude (i.e., greater relative mobility). 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 between these slopes and the relative mobility measures in Figure VIb across CZs. The CZ-level statistics underlying these figures are reported in Online Data Table V.

FIGURE IX: Race and Upward Mobility



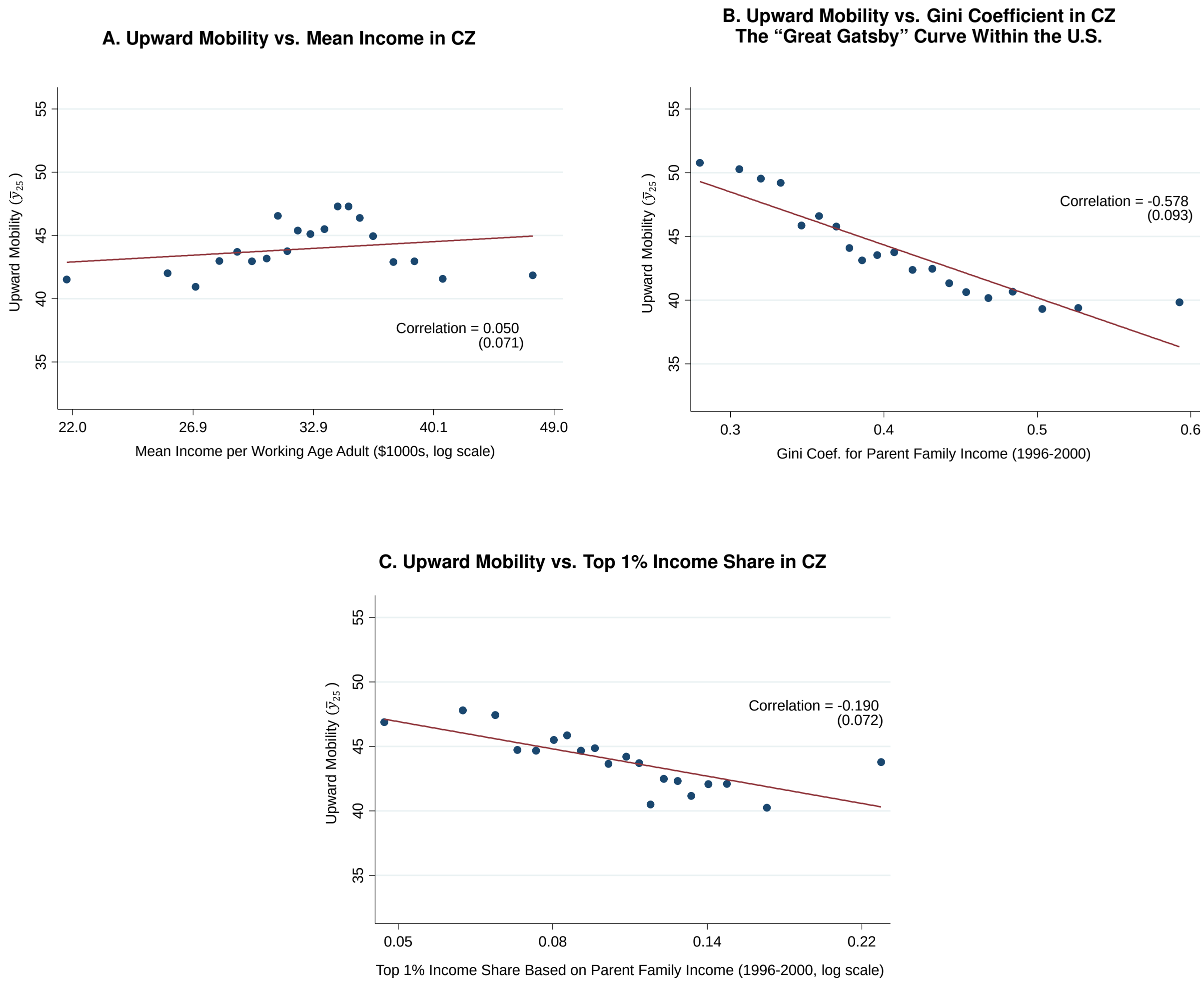
Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{y}_{25}$) vs. the fraction of black residents in a CZ (based on data from the 2000 Census). To construct this figure, we group CZs into twenty equally sized bins (vingtiles) based on their fraction of black residents. We then plot the mean level of absolute upward mobility vs. the mean black share within each of the twenty bins (log scale). The best linear fit and the correlation between the two variables are estimated using the underlying CZ-level data, with standard error (reported in parentheses) clustered by state. The correlations are in levels (not log black share) for consistency with Table VII. Panel B 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 these measures and the absolute upward mobility measures in Figure VIa across CZs. To construct Panel C, 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 $\bar{y}^w_{25,c}$. We then regress $\bar{y}^w_{25,c}$ on $\bar{y}_{25,c}$, 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.

FIGURE X: Correlates of Spatial Variation in Upward Mobility



Notes: This figure shows the correlation of various CZ-level characteristics with absolute upward mobility ($\bar{y}_{25}$) across CZs. For each characteristic listed on the y axis, the dot represents the absolute value of the unweighted correlation of the variable with $\bar{y}_{25}$ 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 nine broad categories: segregation, properties of the income distribution, local tax policies, K-12 education, college education, labor market conditions, migration rates, social capital, and family structure. The categories with the highest correlations are highlighted. See Column 1 of Table VII for the point estimates corresponding to the correlations plotted here. See Section V, Online Data Table IX, and Online Appendix D for definitions of each of the correlates. CZ-level data on the covariates used in this figure are reported in Online Data Table VIII.

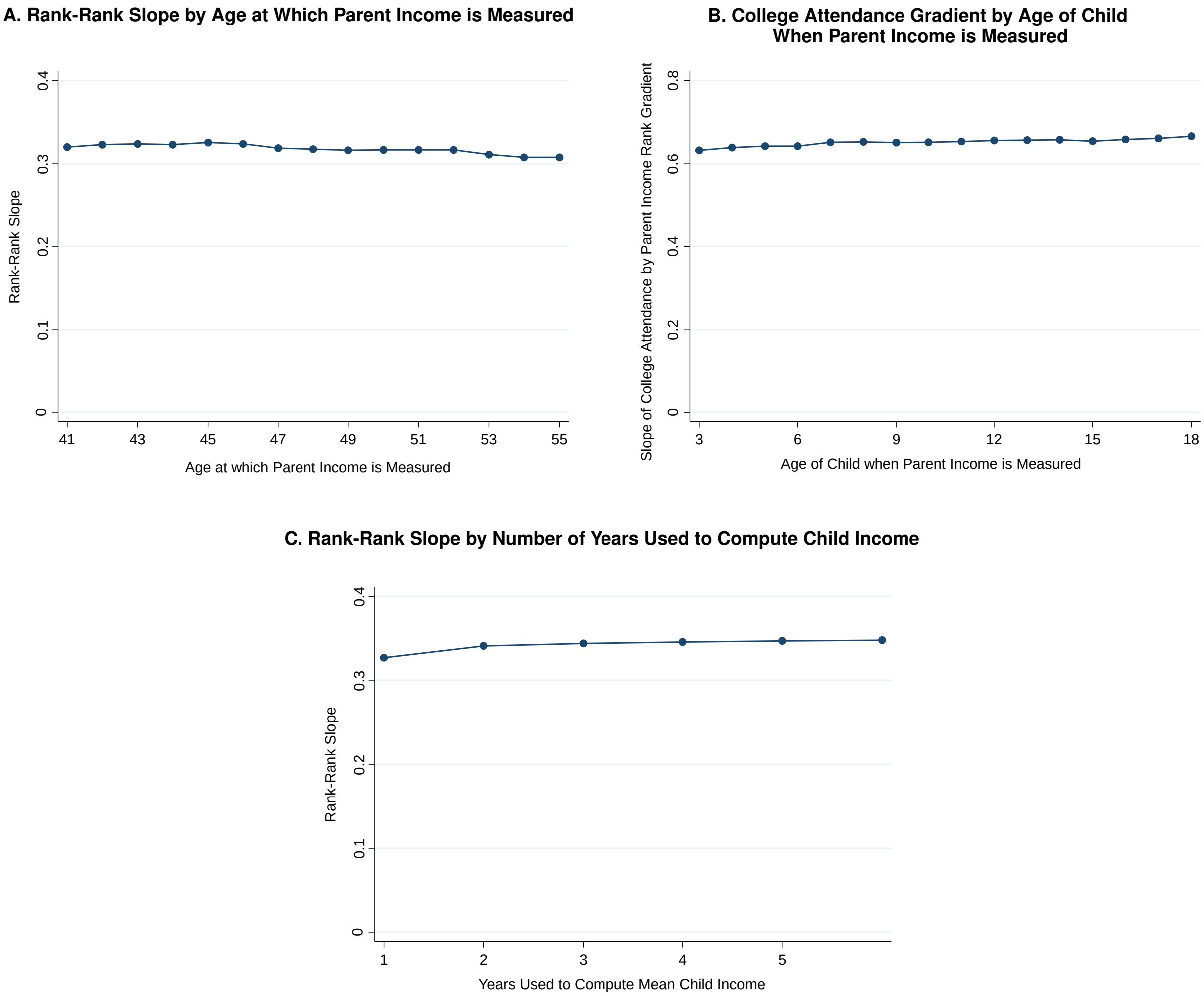
FIGURE XI: Local Income Distributions and Upward Mobility



Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{y}_{25}$) 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. The best log-linear fit and the correlation between the two variables (in levels) are estimated using the underlying CZ-level data, with standard error (reported in parentheses) clustered by state. 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 (log scale), again using the core sample and parents' average income for 1996-2000. We plot the best log-linear fit (estimated using the underlying CZ-level data) and the correlation along with its standard error, clustered by state. The correlations are in levels (not logs) for consistency with Table VII.

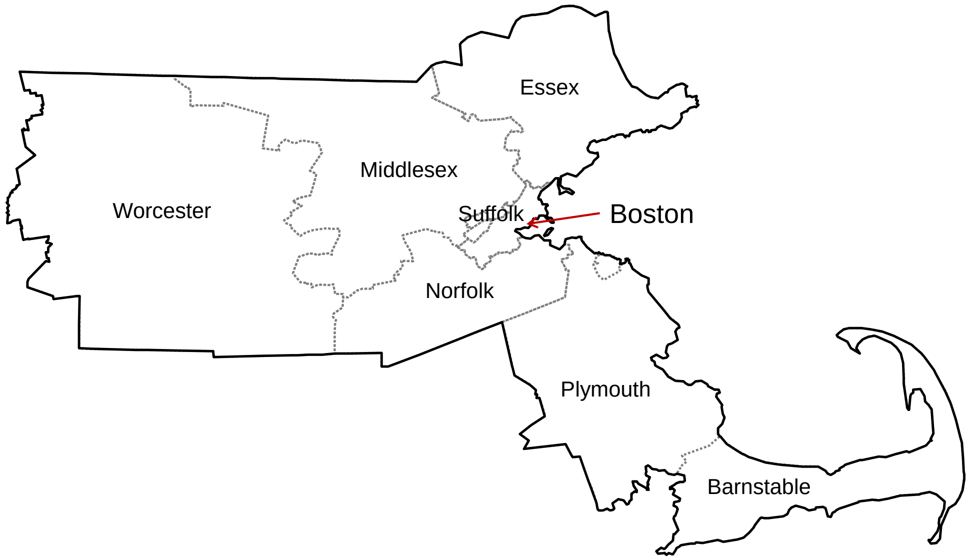
ONLINE APPENDIX FIGURE I

Additional Evidence on Robustness of Intergenerational Mobility Estimates



Notes: This figure evaluates the robustness of the rank-rank slope to changes in the age at which parent income is measured (Panel A), the age of the child when parent income is measured (Panel B), and the number of years used to measure the child's income (Panel C). Panels A and C are based on the core sample (1980-82 birth cohorts). In Panel A, each point shows the slope coefficient from an OLS regression of child income rank on parent income rank (as in Figure IIa), varying the age at which parent income rank is measured. The first point measures parent income in 1996 only, when the mean age of parents is 41. The second point measures parent income in 1997, when parents have a mean age of 42. The last point measures income in 2010, when parents are 55. Panel B 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 C, each point shows the slope coefficient from the same rank-rank regression as in Panel A 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.

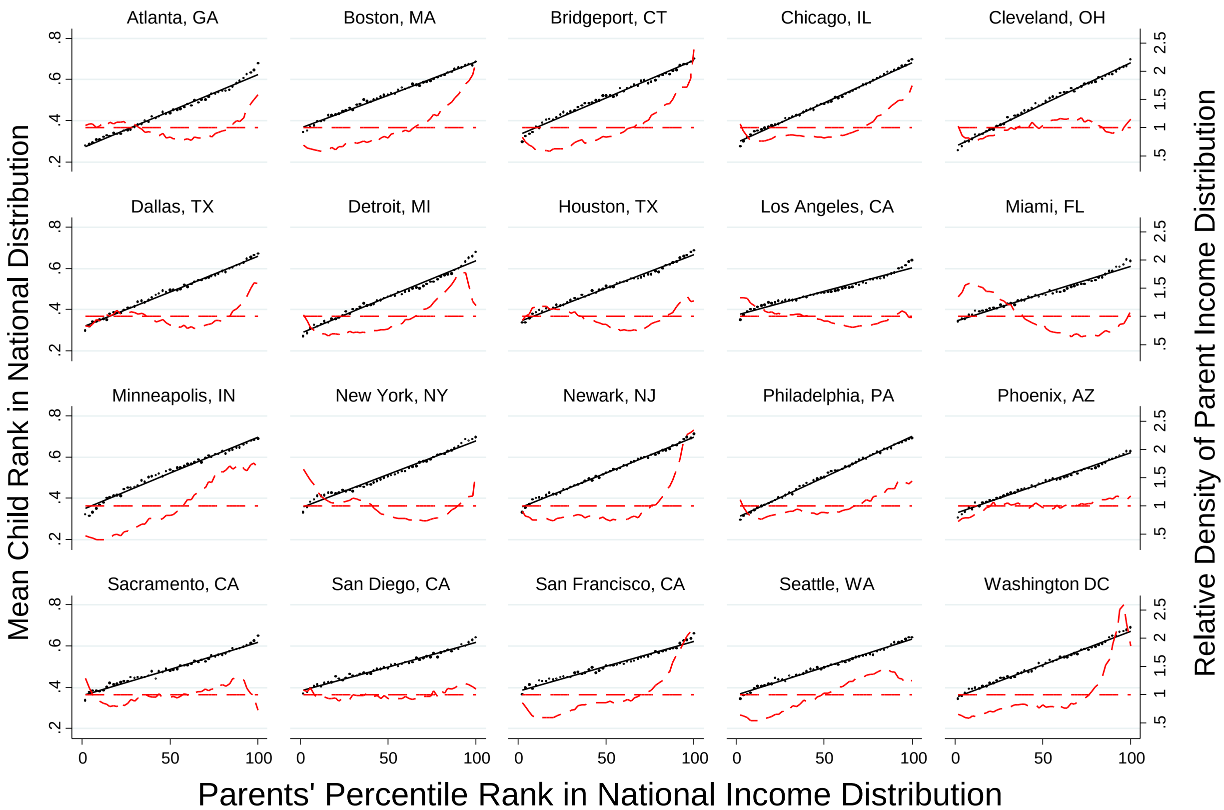
ONLINE APPENDIX FIGURE II
Boston Commuting Zone



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

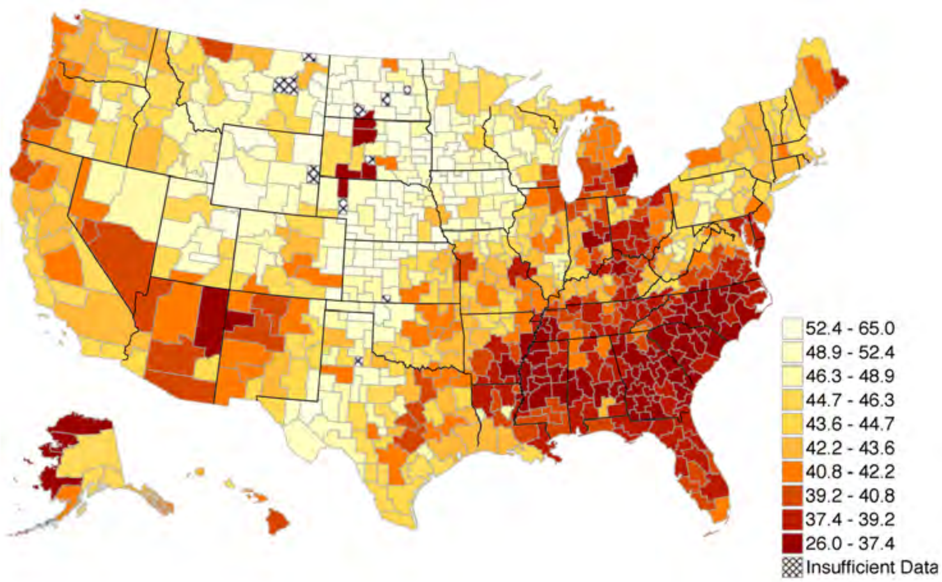
ONLINE APPENDIX FIGURE III

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 20 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. Each figure also shows the fraction of parents with income in each bin divided by the share in that bin in the national income distribution (dashed curve, right y-axis). The dashed curve averages to one (denoted by the horizontal dashed line in each panel) in each CZ by construction and depicts the income distribution in the CZ relative to the national distribution.

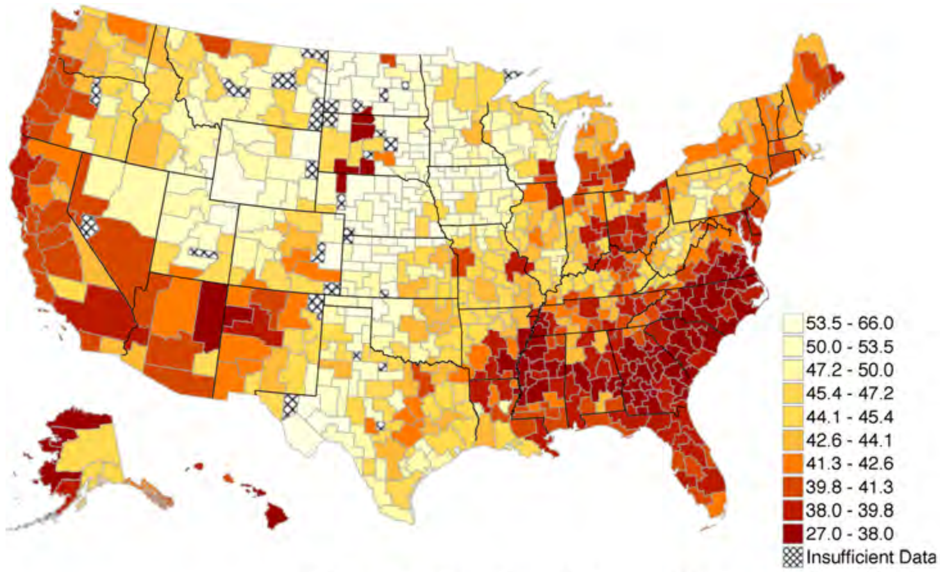
ONLINE APPENDIX FIGURE IV
Estimates of Absolute Upward Mobility Pooling 1980-82 and 1980-85 Cohorts



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.

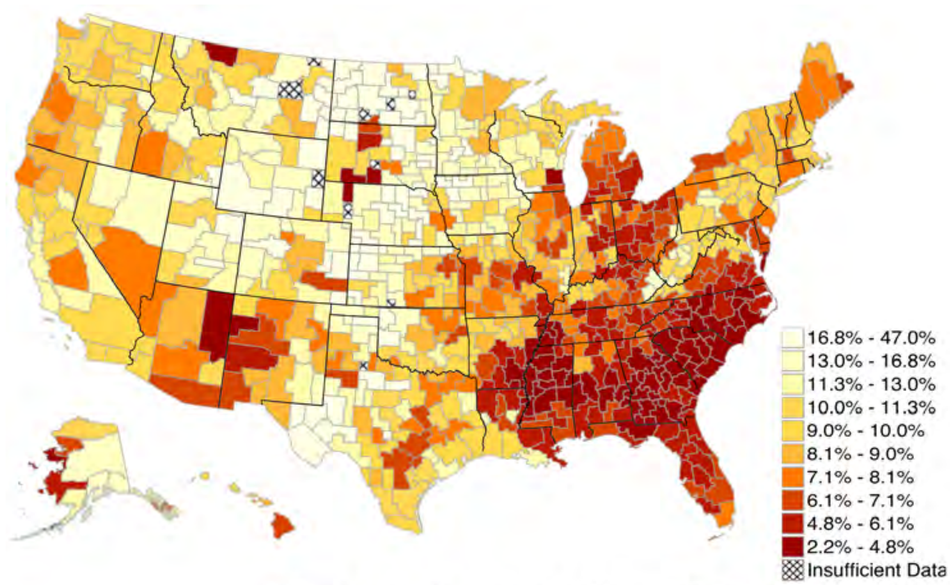
ONLINE APPENDIX FIGURE V
Alternative Measures of Upward Mobility

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



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

B. Probability of Reaching Top Quintile Given Parents in Bottom Quintile

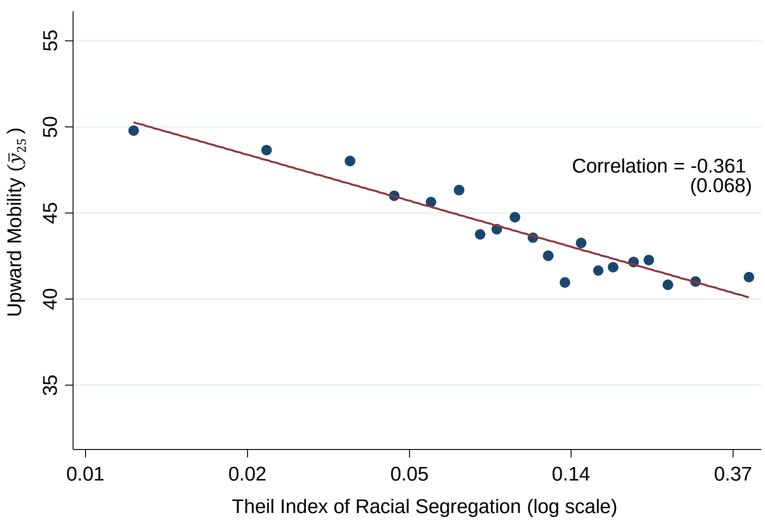


Corr. with baseline $\bar{y}_{100}$-$\bar{y}_0$ = 0.91 (unweighted), 0.92 (pop-weighted)

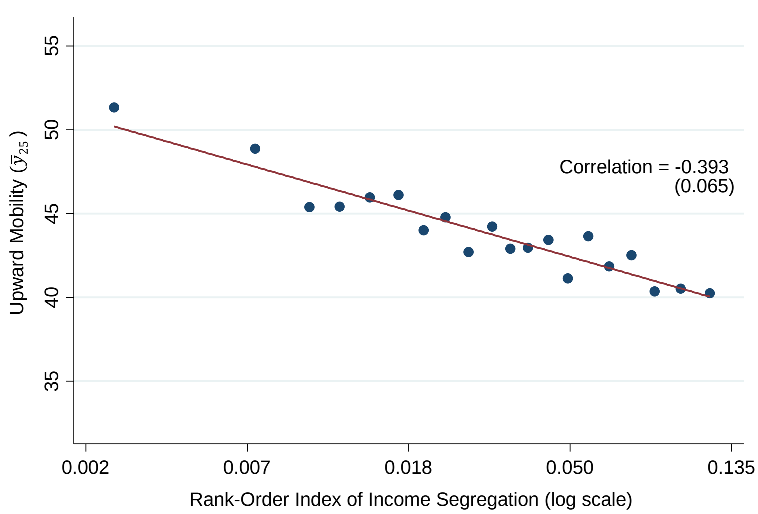
Notes: Panel A replicates Figure VIa, adjusting for differences in cost-of-living across areas. To construct this figure, we first deflate parent income by a cost-of-living index (COLI) for the parent's CZ when he/she claims the child as a dependent and child income by a COLI for the child's CZ in 2012. 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. This figure is constructed using data from the 1980-85 birth cohorts. We report the unweighted and population-weighted correlation coefficients between these measures and the absolute upward mobility measures in Figure VIa across CZs in both figures. The CZ-level statistics underlying these figures are reported in Online Data Table V.

ONLINE APPENDIX FIGURE VI
Segregation and Upward Mobility

A. Upward Mobility vs. Theil Index of Racial Segregation



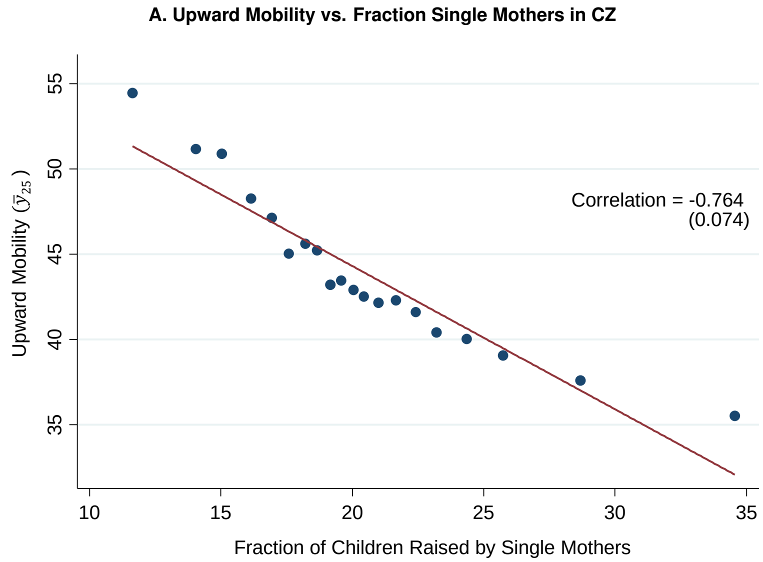
B. Upward Mobility vs. Rank-Order Index of Income Segregation

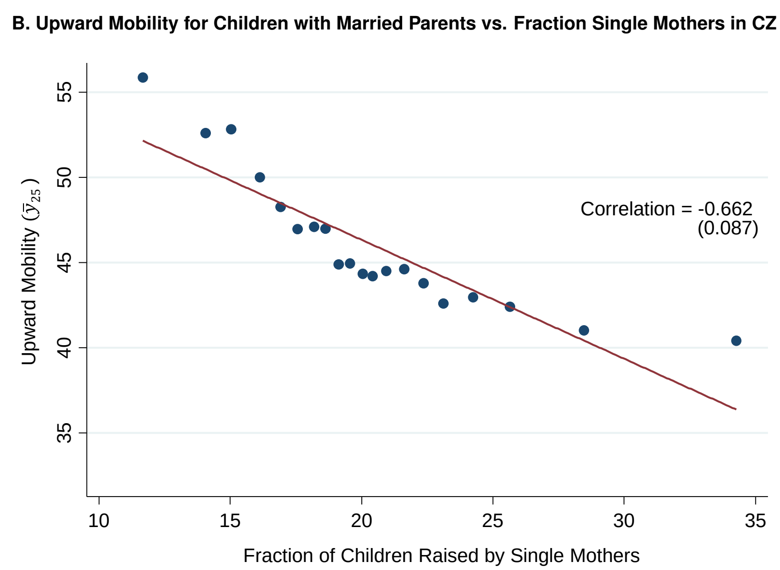


Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{y}_{25}$) vs. a multi-group Theil index of racial segregation (based on census tract level data from the 2000 Census). To construct this figure, we group CZs into twenty equally sized bins (vingtiles) based on the racial segregation index. We then plot the mean level of absolute upward mobility vs. the mean level of the segregation index within each of the twenty bins using a log scale. The best linear fit and the correlation between the two variables are estimated using the underlying CZ-level data, with standard error (reported in parentheses) clustered by state. The correlations are in levels (not logs) for consistency with Table VII. 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.

ONLINE APPENDIX FIGURE VII

Single-Parent Families and Upward Mobility





Notes: Panel A presents a binned scatter plot of absolute upward mobility ($\bar{y}_{25}$) 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. The best linear fit and the correlation between the two variables are estimated using the underlying CZ-level data, with standard error (reported in parentheses) clustered by state. 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.