THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. 129 November 2014 Issue 4

WHERE IS THE LAND OF OPPORTUNITY? THE GEOGRAPHY OF INTERGENERATIONAL MOBILITY IN THE UNITED STATES*

RAJ CHETTY
NATHANIEL HENDREN
PATRICK KLINE
EMMANUEL SAEZ

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

*The opinions expressed in this article 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 article 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, Pablo Mitnik, Jonathan Parker, Laszlo Sandor, Gary Solon, Danny Yagan, numerous seminar participants, and four anonymous referees for helpful comments. 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 http://www.equality-of-opportunity.org.

The Quarterly Journal of Economics (2014), 1553–1623. doi:10.1093/qje/qju022.
Advance Access publication on September 14, 2014.

in a child's income. Second, intergenerational mobility varies substantially across areas within the United States. 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 (i) less residential segregation, (ii) less income inequality, (iii) better primary schools, (iv) greater social capital, and (v) greater family stability. Although our descriptive analysis does not identify the causal mechanisms that determine upward mobility, the publicly available statistics on intergenerational mobility developed here can facilitate research on such mechanisms. *JEL* Codes: H0, J0, R0.

I. INTRODUCTION

The United States is often hailed as the "land of opportunity," a society in which a person's chances of success depend little on his or 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 United States. The United States 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 deidentified 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 part, we present new statistics on intergenerational mobility in the United States as a whole. In our baseline analysis, we focus on 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.[1]

1. We show that our baseline measures do not suffer from significant life cycle or attenuation bias (Solon 1992; Zimmerman 1992; Mazumder 2005) by establishing that estimates of mobility stabilize by the time children reach age 30 and are not very sensitive to the number of years used to measure parent income.

Following the prior literature (e.g., Solon 1999), we begin by estimating the intergenerational elasticity of income (IGE) by regressing log child income on log parent income. Unfortunately, we find that this canonical log-log specification yields very unstable estimates of mobility because the relationship between log child income and log parent income is nonlinear and the estimates are sensitive to the treatment of children with zero or very small incomes. When restricting the sample between the 10th and 90th percentiles of the parent income distribution and excluding children with zero income, we obtain an IGE estimate of 0.45. However, alternative specifications yield IGEs ranging from 0.26 to 0.70, spanning most of the estimates in the prior literature.[2]

To obtain a more stable summary of intergenerational mobility, we use 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. We rank parents of these children based on their incomes relative to other parents with children in these birth cohorts. We characterize mobility based on the slope of this rank-rank relationship, which identifies the correlation between children's and parents' positions in the income distribution.[3]

We find that the relationship between mean child ranks and parent ranks is almost perfectly linear and highly robust to alternative specifications. A 10 percentile point increase in parent rank is associated with a 3.41 percentile increase in a child's income rank on average. 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 article, we 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 United States,

2. In an important recent study, Mitnik et al. (2014) propose a new dollar-weighted measure of the IGE and show that it yields more stable estimates. We discuss the differences between the new measure of mobility proposed by Mitnik et al. and the canonical definition of the IGE in Section IV.A.

3. The rank-rank slope and IGE both measure the degree to which differences in children's incomes are determined by their parents' incomes. We discuss the conceptual differences between the two measures in Section II.

including rural areas (Tolbert and Sizer 1996). We assign children to CZs based on where they lived at age 16—that is, 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, which allows us to measure children's absolute outcomes as we discuss later.

The relationship between mean child ranks and parent ranks is almost perfectly linear within CZs, 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 versus 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 call 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. Some CZs in the United States have relative mobility comparable to the highest mobility countries in the world, such as Canada and 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 call "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. A 1 standard deviation increase in CZ-level upward mobility is associated with a 0.2 standard deviation improvement in a child's expected rank given parents at $p = 25$, 60% as large as the effect of a 1 standard deviation increase in his 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 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. Location matters more for children growing up in 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 birthrates with respect to parent income across CZs are very similar to the variation in intergenerational income mobility. This suggests that the spatial differences in mobility are driven by factors that affect children while they are growing up rather than after they enter labor market.

In the final part of the article, we explore such factors by correlating the spatial variation in mobility with observable characteristics. To begin, we show 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 level.

We then identify five factors that are strongly correlated with the variation in upward mobility across areas. The first is segregation: areas that are more residentially segregated by race and income have lower levels of mobility. Second, areas with more inequality as measured by Gini coefficients have less mobility, consistent with the "Great Gatsby curve" documented across countries (Krueger 2012; Corak 2013). Top 1% income shares are not highly correlated with intergenerational mobility both across CZs within the United States and across countries, suggesting that the factors that erode the middle class may 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 positively correlated with mobility.

Fourth, social capital indexes (Putnam 1995)—which are proxies for the strength of social networks and community involvement in an area—are also positively correlated with mobility. Finally, mobility is significantly lower in areas with weaker family structures, as measured, for example, by the fraction of single parents. 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 in communities with fewer single parents. Interestingly, we find no correlation between racial shares and upward mobility once we control for the fraction of single parents in an area.

We find modest correlations between upward mobility and local tax policies and no systematic correlation between mobility and local labor market conditions, rates of migration, or access to higher education. In a multivariable 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 endogenously determined and our analysis does not control for numerous other unobserved differences across areas.

Our results build on an extensive literature on intergenerational mobility, reviewed by Solon (1999) and Black and Devereux (2011). Our estimates of the level of mobility in the United States as a whole are broadly consistent with prior results, with the exception of Mazumder's (2005) and Clark's (2014) IGE estimates, which imply much lower levels of intergenerational mobility. We discuss why our findings may differ from their results in Online Appendices D and E. Our focus on within-country comparisons offers two advantages over the cross-country comparisons that have been the focus of prior comparative work (e.g., Bjorklund and Jäntti 1997; Jäntti et al. 2006; Corak 2013). First, differences in measurement and methods make it difficult to reach definitive conclusions from cross-country comparisons (Solon 2002). The variables we analyze are measured using the same data sources across all CZs. Second, and more important, we characterize both relative and absolute mobility across CZs. The cross-country literature has focused exclusively on differences in relative mobility; 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). Unlike recent experimental work on neighborhood effects (e.g., Katz, Kling, and Liebman 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. However, in a followup paper, Chetty and Hendren (2014) show that a substantial portion of the spatial variation documented here is driven by causal effects of place by studying families that move across areas with children of different ages.

The article is organized as follows. We begin in Section II by defining the measures of intergenerational mobility that we study and discussing their conceptual properties. Section III describes the data. Section IV reports estimates of intergenerational mobility at the national level. In Section V, we present estimates of absolute and relative mobility by commuting zone. Section VI reports correlations of our mobility measures with observable characteristics of commuting zones. Section VII concludes. Statistics on intergenerational mobility and related covariates are publicly available by commuting zone, metropolitan statistical area, and county on the project website (www.equality-of-opportunity.org).

II. Measures of Intergenerational Mobility

At the most general level, studies of intergenerational mobility seek to measure the degree to which a child's social and economic opportunities depend on his parents' income or social status. Because opportunities are difficult to measure, virtually all empirical studies of mobility measure the extent to which a child's income (or occupation) depends on his parents' income (or occupation).[4] Following this approach, we aim to characterize the

4. This simplification is not innocuous, as a child's realized income may differ from his opportunities. For instance, children of wealthy parents may choose not to work or may choose lower-paying jobs, which would reduce the persistence of income across generations relative to the persistence of underlying opportunities.

joint distribution of a child's lifetime pretax family income (Y_i), and his parents' lifetime pretax family income (X_i).[5]

In large samples, one can characterize the joint distribution of (Y_i, X_i) nonparametrically, and we provide such a characterization in the form of a 100×100 centile transition matrix below. However, to provide a parsimonious summary of the degree of mobility and compare rates of mobility across areas, it is useful to characterize the joint distribution using a small set of statistics. We divide measures of mobility into two classes that capture different normative concepts: relative mobility and absolute mobility. In this section, we define a set of statistics that we use to measure these two concepts empirically and compare their conceptual properties.

II.A. Relative Mobility

One way to study 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, which focuses on the relative outcomes of children from different parental backgrounds, has been the subject of most prior research on intergenerational mobility (Solon 1999; Black et al. 2011).

The canonical measure of relative mobility is the elasticity of child income with respect to parent income $\left(\frac{dE[\log Y_i | X_i = x]}{d\log x}\right)$, commonly called the intergenerational income elasticity (IGE). The most common method of estimating the IGE is to regress log child income ($\log Y_i$) on log parent income ($\log X_i$), which yields a coefficient of

$$IGE = \rho_{XY} \frac{SD(\log Y_i)}{SD(\log X_i)}, \tag{1}$$

where $\rho_{XY} = Corr(\log X_i, \log Y_i)$ is the correlation between log child income and parent income and $SD()$ denotes the standard deviation. The IGE is a relative mobility measure because it measures the difference in (log) outcomes between children of high versus low income parents.

5. If taxes and transfers do not generate rank reversals (as is typically the case in practice), using post-tax income instead of pretax income would have no effect on our preferred rank-based measures of mobility. See Mitnik et al. (2014) for a comparison of pretax and post-tax measures of the IGE of income.

An alternative measure of relative mobility is the correlation between child and parent ranks (Dahl and DeLeire 2008). Let R_i denote child i's percentile rank in the income distribution of children and P_i denote parent i's percentile rank in the income distribution of parents. Regressing the child's rank R_i on his parents' rank P_i yields a regression coefficient $\rho_{PR} = Corr(P_i, R_i)$, which we call the rank-rank slope.[6] The rank-rank slope ρ_{PR} measures the association between a child's position in the income distribution and his parents' position in the distribution.

To understand the connection between the IGE and the rank-rank slope, note that the correlation of log incomes ρ_{XY} and the correlation of ranks ρ_{PR} are closely related scale-invariant measures of the degree to which child income depends on parent income.[7] Hence, equation (1) implies that the IGE combines the dependence features captured by the rank-rank slope with the ratio of standard deviations of income across generations.[8] The IGE differs from the rank-rank slope to the extent that inequality changes across generations. Intuitively, a given increase in parents' incomes has a greater effect on the level of children's incomes when inequality is greater among children than among parents.

We estimate both the IGE and the rank-rank slope to distinguish differences in mobility from differences in inequality and to provide a comparison to the prior literature. However, we focus primarily on rank-rank slopes because they prove to be much more robust across specifications and are thus more suitable for comparisons across areas from a statistical perspective.

II.B. Absolute Mobility

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?" For example, one may be

6. The regression coefficient equals the correlation coefficient because both child and parent ranks follow a uniform distribution by construction.

7. For example, if parent and child income follow a bivariate log normal distribution, $\rho_{PR} = \frac{6ArcSin(\frac{\rho_{XY}}{2})}{\pi} \approx \frac{3\rho_{XY}}{\pi} = 0.95\rho_{XY}$ when ρ_{XY} is small (Trivedi and Zimmer 2007).

8. More generally, the joint distribution of parent and child incomes can be decomposed into two components: the joint distribution of parent and child percentile ranks (the copula) and the marginal distributions of parent and child income. The rank-rank slope depends purely on the copula, whereas the IGE combines both components.

interested in measuring the mean outcomes of children whose grow up in low-income families. Absolute mobility may be of greater normative interest than relative mobility. Increases in relative mobility (i.e., a lower IGE or rank-rank slope) could be undesirable if they are caused by worse outcomes for the rich. In contrast, increases in absolute mobility at a given income level, holding fixed absolute mobility at other income levels, unambiguously increase welfare if one respects the Pareto principle (and if welfare depends purely on income).

We consider three statistical measures of absolute mobility. Our primary measure, which we call absolute upward mobility, is the mean rank (in the national child income distribution) of children whose parents are at the 25th percentile of the national parent income distribution.[9] At the national level, this statistic is mechanically related to the rank-rank slope and does not provide any additional information about mobility.[10] However, when we study small areas within the United States, a child's rank in the national income distribution is effectively an absolute outcome because incomes in a given area have little effect on the national distribution.

The second measure we analyze is the probability of rising from the bottom quintile to the top quintile of the income distribution (Corak and Heisz 1999; Hertz 2006), which can be interpreted as a measure of the fraction of children who achieve the "American Dream." Again, when the quintiles are defined in the national income distribution, these transition probabilities can be interpreted as measures of absolute outcomes in small areas. Our third measure is the probability that a child has family income above the poverty line conditional on having parents at the 25th percentile. Because the poverty line is defined in absolute dollar terms in the United States, this statistic measures the

9. This measure is the analog of the rank-rank slope in terms of absolute mobility. The corresponding analog of the IGE is the mean log income of children whose parents are at the 25th percentile. We do not study this statistic because it is very sensitive to the treatment of zeros and small incomes.

10. We show below that the rank-rank relationship is approximately linear. Because child and parent ranks each have a mean of 0.5 by construction in the national distribution, the mean rank of children with parents at percentile p is simply $0.5 + \rho_{PR}(p - 0.5)$. Conceptually, the slope is the only free parameter in the linear national rank-rank relationship. Intuitively, if one child moves up in the income distribution in terms of ranks, another must come down.

fraction of children who achieve a given absolute living standard.[11]

It is useful to analyze multiple measures of mobility because the appropriate measure of intergenerational mobility depends on one's normative objective (Fields and Ok 1999). Fortunately, we find that the patterns of spatial variation in absolute and relative mobility are very similar using alternative measures. In addition, we provide nonparametric transition matrices and marginal distributions that allow readers to construct measures of mobility beyond those we consider here.

III. 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 (such as 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).

III.A. Sample Definitions

Our base data set of children consists of all individuals who (i) have a valid Social Security number or individual taxpayer identification number, (ii) were born between 1980 and 1991, and (iii) are U.S. citizens as of 2013. We impose the citizenship requirement to exclude individuals who are likely to have immigrated to the United States as adults, for whom we cannot measure parent income. We cannot directly restrict the sample to individuals born in the United States because the database only records current citizenship status.

We identify the parents of a child as the first tax filers (between 1996 and 2012) who claim the child as a child dependent and were between the ages of 15 and 40 when the child was born.

11. Another intuitive measure of upward mobility is the fraction of children whose income exceeds that of their parents. This statistic turns out to be problematic for our application because we measure parent and child income at different ages and because it is very sensitive to differences in local income distributions.

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.[12]

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). We are therefore able to identify parents for approximately 95 percent of the children in the 1980–1991 birth cohorts. 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 (Online Appendix Table I). 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 data set who (i) are born in the 1980–1982 birth cohorts, (ii) for whom we are able to identify parents, and (iii) whose mean parent income between 1996 and 2000 is strictly positive (which excludes 1.2% of children).[13] For some robustness checks, we use the extended sample, which imposes the same restrictions as the core sample, but includes all birth cohorts from 1980 to 1991. There are approximately 10 million children in the core sample and 44 million children in the extended sample.

1. 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

12. Twelve percent 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 Online Appendix Table VII).

13. 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.

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 (IRS) 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 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 1971–1979 birth cohorts in the SOI sample (Online Appendix Table II).

III.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).

1. Parent Income. Following Lee and Solon (2009), our primary measure of parent income is total pretax 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 nontaxable portion of Social Security and Disability (SSDI) 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 SA-1099) for both parents.[14] In years where parents have no

14. The database does not record W-2s and other information returns prior to 1999, so nonfiler's income is coded as 0 prior to 1999. Assigning nonfiling parents 0 income has little effect on our estimates because only 2.9% of parents in our core sample do not file in each year prior to 1999 and most nonfilers have very low W-2 income. For instance, in 2000, median W-2 income among nonfilers was $29.

tax return and no information returns, family income is coded as zero.[15]

Our baseline income measure includes labor earnings and capital income as well as unemployment insurance, Social Security, and disability benefits. It excludes nontaxable cash transfers such as Temporary Assistance to Needy Families and Supplemental Security Income, in-kind benefits such as food stamps, all refundable tax credits such as the EITC, nontaxable pension contributions (such as 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 obtain a proxy for parent lifetime income that is less affected by 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.[16] We evaluate the robustness of our findings using data from other years and using a measure of individual parent income instead of family income. We define individual income as the sum of individual W-2 wage earnings, unemployment insurance benefits, SSDI payments, and half of household self-employment income (see Online Appendix A for details).

Furthermore, we show below that defining parent income based on data from 1999 to 2003 (when W-2 data are available) yields virtually identical estimates (Table I, row 5). Note that we never observe self-employment income for nonfilers and therefore code it as 0; given the strong incentives for individuals with children to file created by the EITC, most nonfilers likely have very low levels of self-employment income as well.

15. Importantly, these observations are true zeros rather than missing data. Because the database covers all tax records, we know that these individuals have 0 taxable income.

16. Formally, we define mean family income as the mother's family income plus the father's family income in each year from 1996 to 2000 divided by 10 (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 five-year period. For parents who are married initially and then divorce, this measure tracks the mean family incomes of the two 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.

2. Child Income. We define child family income in 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), when children are in their early thirties. 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 the same way as for parents.

3. 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 postsecondary 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 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 to 2012. Comparisons to other data sources indicate that 1098-T forms capture college enrollment quite accurately overall (Chetty, Friedman, and Rockoff 2014, Appendix B).[17]

4. College Quality. Using data from 1098-T forms, Chetty, Friedman, and Rockoff (2014) construct an earnings-based index of "college quality" using the mean individual wage earnings at age 31 of children born in 1979–1980 based on the college they attended at age 20. Children who do not attend college are included in a separate "no college" category in this index. We assign each child in our sample a value of this college quality index based on the college in which they were enrolled at age

17. 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, presumably because of automated reporting to the IRS by universities. Approximately 6% of 1098-T forms are missing from 2000 to 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).

20. We then convert this dollar index to percentile ranks within each birth cohort. The children in the no-college group, who constitute roughly 54 percent of our core sample, all have the same value of the college quality index. Breaking ties at the mean, we assign all of these children a college quality rank of approximately $\frac{54}{2} = 27$.[18]

5. 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. Nevertheless, the aggregate level and spatial pattern of teenage births in our data are closely aligned with estimates based on the American Community Survey.[19]

6. Summary Statistics. Online Appendix Table III 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 15.8% of women have a teenage birth.

In Online Appendix B and Appendix Table IV, 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.

18. The exact value varies across cohorts. For example, in the 1980 birth cohort, 55.1% of children do not attend college. We assign these children a rank of $\frac{55.1}{2} + 0.02 = 27.7\%$ because 0.2% of children in the 1980 birth cohort attend colleges whose mean earnings are below the mean earnings of those not in college.

19. Of women in our core sample, 15.8% 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.

IV. 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 relative mobility and then evaluate the robustness of our estimates to alternative sample and income definitions.[20]

IV.A. Baseline Estimates

In our baseline analysis, we use the core sample (1980–1982 birth cohorts) and measure parent income as mean family income from 1996 to 2000 and child income as mean family income in 2011–2012, when children are approximately 30 years old. Figure I Panel A 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 horizontal axis into 100 equal-sized (percentile) bins and plot mean child income versus mean parent income in each bin.[21] This binned scatter plot provides a nonparametric representation of the conditional expectation of child income given parent income, $E[Y_i|X_i = x]$. 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.

20. We do not present estimates of absolute mobility at the national level because absolute mobility in terms of percentile ranks is mechanically related to relative mobility at the national level (see Section II). Although one can compute measures of absolute mobility at the national level based on mean incomes (e.g., the mean income of children whose parents are at the 25th percentile), there is no natural benchmark for such a statistic as it has not been computed in other countries or time periods.

21. For scaling purposes, we exclude the top bin (parents in the top 1%) in this figure only; mean parent income in this bin is $1,408,760 and mean child income is $113,846.

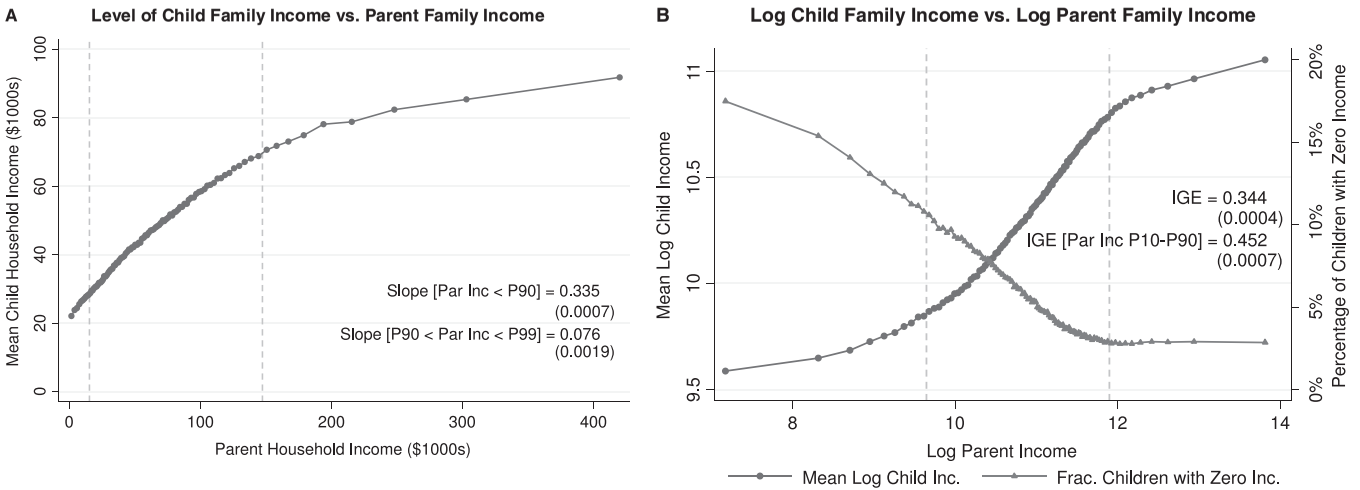

FIGURE I

Association between Children's and Parents' Incomes

These figures present nonparametric binned scatter plots of the relationship between child income and parent income. Both panels are based on the core sample (1980–1982 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011–2012 family income (when the child is approximately 30 years old), whereas parent income is mean family income from 1996 to 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 versus mean level of parent income within each bin. For scaling purposes, we do not show the point for the top 1% in Panel A. In the top 1% bin, mean parent income is $1.4 million and mean child income is $114,000. In Panel B, we again bin parent family income into 100 bins and plot mean log income for children (left y-axis) and the fraction of children with zero family income (right y-axis) versus mean parents' log income. Children with zero family income are excluded from the log income series. In both panels, the 10th and 90th percentile of parents' income are depicted in dashed vertical lines. The coefficient estimates and standard errors (in parentheses) reported on the figures are obtained from OLS regressions on the microdata. In Panel A, we report separate slopes for parents below the 90th percentile and parents between the 90th and 99th percentile. In Panel B, we report slopes of the log-log regression (i.e., the intergenerational elasticity of income or IGE) in the full sample and for parents between the 10th and 90th percentiles.

TABLE I

INTERGENERATIONAL MOBILITY ESTIMATES AT THE NATIONAL LEVEL

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sample						
Child's outcome	Parent's income def.	Core sample	Male children	Female children	Married parents	Single parents	1980–1985 cohorts	Fixed age at child birth
1. Log family income (excluding zeros)	Log family income	0.344 (0.0004)	0.349 (0.0006)	0.342 (0.0005)	0.303 (0.0005)	0.264 (0.0008)	0.316 (0.0003)	0.361 (0.0008)
2. Log family income (recoding zeros to $1)	Log family income	0.618 (0.0009)	0.697 (0.0013)	0.540 (0.0011)	0.509 (0.0011)	0.528 (0.0020)	0.580 (0.0006)	0.642 (0.0018)
3. Log family income (recoding zeros to $1,000)	Log family income	0.413 (0.0004)	0.435 (0.0007)	0.392 (0.0006)	0.358 (0.0006)	0.322 (0.0009)	0.380 (0.0003)	0.434 (0.0009)
4. Family income rank	Family income rank	0.341 (0.0003)	0.336 (0.0004)	0.346 (0.0004)	0.289 (0.0004)	0.311 (0.0007)	0.323 (0.0002)	0.359 (0.0006)
5. Family income rank	Family income rank (1999–2003)	0.339 (0.0003)	0.333 (0.0004)	0.344 (0.0004)	0.287 (0.0004)	0.294 (0.0007)	0.323 (0.0002)	0.357 (0.0006)
6. Family income rank	Top par. income rank	0.312 (0.0003)	0.307 (0.0004)	0.317 (0.0004)	0.256 (0.0004)	0.253 (0.0006)	0.296 (0.0002)	0.327 (0.0006)
7. Individual income rank	Family income rank	0.287 (0.0003)	0.317 (0.0004)	0.257 (0.0004)	0.265 (0.0004)	0.279 (0.0007)	0.286 (0.0002)	0.292 (0.0006)
8. Individual earnings rank	Family income rank	0.282 (0.0003)	0.313 (0.0004)	0.249 (0.0004)	0.259 (0.0004)	0.272 (0.0007)	0.283 (0.0002)	0.287 (0.0006)
9. College attendance	Family income rank	0.675 (0.0005)	0.708 (0.0007)	0.644 (0.0007)	0.641 (0.0006)	0.663 (0.0013)	0.678 (0.0003)	0.661 (0.0010)

TABLE I
(CONTINUED)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sample						
Child's outcome	Parent's income def.	Core sample	Male children	Female children	Married parents	Single parents	1980–1985 cohorts	Fixed age at child birth
10. College quality rank (P75–P25 gradient)	Family income rank	0.191 (0.0010)	0.188 (0.0014)	0.195 (0.0015)	0.174 (0.0014)	0.172 (0.0020)	0.198 (0.0007)	0.189 (0.0022)
11. Teenage birth (females only)	Family income rank	−0.298 (0.0006)			−0.231 (0.0007)	−0.322 (0.0016)	−0.285 (0.0004)	−0.290 (0.0011)
Number of observations		9,867,736	4,935,804	4,931,066	6,854,588	3,013,148	20,520,588	2,250,380

Notes. Each cell in this table reports the coefficient from a univariate OLS regression of an outcome for children on a measure of their parents' incomes with standard errors in parentheses. All rows report estimates of slope coefficients from linear regressions of the child outcome on the parent income measure except row 10, in which we regress college quality rank on a quadratic in parent income rank (as in Figure IV Panel A). In this row, we report the difference between the fitted values for children with parents at the 75th percentile and parents at the 25th percentile using the quadratic specification. Column (1) uses the core sample of children, which includes all current U.S. citizens with a valid SSN or ITIN who are (i) born in birth cohorts 1980–1982, (ii) for whom we are able to identify parents based on dependent claiming, and (iii) whose mean parent income over the years 1996–2000 is strictly positive. Columns (2) and (3) limit the sample used in column (1) to males or females. Columns (4) and (5) limit the sample to children whose parents were married or unmarried in the year the child was linked to the parent. Column (6) uses all children in the 1980–1985 birth cohorts. Column (7) restricts the core sample to children whose parents both fall within a 5-year window of median parent age at time of child birth (age 26–30 for fathers; 24–28 for mothers); we impose only one of these restrictions for single parents. Child family income is the mean of 2011–2012 family income, while parent family income is the mean from 1996 to 2000. Parent top earner income is the mean income of the higher-earning spouse between 1999–2003 (when W-2 data are available). Child's individual income is the sum of W-2 wage earnings, UI benefits, and SSDI benefits, and half of any remaining income reported on the 1040 form. Individual earnings includes W-2 wage earnings, UI benefits, SSDI income, and self-employment income. College attendance is defined as ever attending college from age 18 to 21, where attending college is defined as presence of a 1098-T form. College quality rank is defined as the percentile rank of the college that the child attends at age 20 based on the mean earnings at age 31 of children who attended the same college (children who do not attend college are included in a separate "no college" group); see Section III.B for further details. Teenage birth is defined as having a child while between age 13 and 19. In columns (1)–(5) and (7), income percentile ranks are constructed by ranking all children relative to others in their birth cohort based on the relevant income definition and ranking all parents relative to other parents in the core sample. Ranks are always defined on the full sample of all children; that is, they are not redefined within the subsamples in columns (2)–(5) or (7). In column (6), parents are ranked relative to other parents with children in the 1980–1985 birth cohorts. The number of observations corresponds to the specification in row 4. The number of observations is approximately 7% lower in row 1 because we exclude children with zero income. The number of observations is approximately 50% lower in row 11 because we restrict to the sample of female children. There are 866 children in the core sample with unknown sex, which is why the number of observations in the core sample is not equal to the sum of the observations in the male and female samples.

1. Log-Log Intergenerational Elasticity Estimates. Partly motivated by the nonlinearity of the relationship in Figure I Panel A, the canonical approach to characterizing the joint distribution of child and parent income is to regress the log of child income on the log of parent income (as discussed in Section II), excluding children with zero income. This regression yields an estimated IGE of 0.344, as shown in the first column of row 1 of Table I.

Unfortunately, this estimate turns out to be quite sensitive to changes in the regression specifications for two reasons, illustrated in Figure I Panel B. First, the relationship between log child income and log parent income is highly nonlinear, consistent with the findings of Corak and Heisz (1999) in Canadian tax data. This is illustrated in the series in circles in Figure I Panel B, which plots mean log child income versus mean log family income by percentile bin, constructed using the same method as Figure I Panel A. Because of this nonlinearity, 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 I Panel B plots the fraction of children with zero income by parental income bin. This fraction varies from 17% among the poorest families to 3% among the richest families. Dropping children with zero income therefore overstates the degree of intergenerational mobility. The way 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 I. If instead we treat those with 0 income as having an income of $1,000, the estimated IGE becomes 0.413. These exercises show that small differences in the way children's income is measured at the bottom of the distribution can produce substantial variation in IGE estimates.

Columns (2)–(7) in Table I replicate the baseline specification in column (1) for alternative subsamples 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 1980–1985 birth cohorts. Column (7) restricts the

sample to children whose mothers are between the ages of 24 and 28 and fathers are between 26 and 30 (a 5-year window around the median age of birth). This column eliminates variation in parent income correlated with differences in parent age at child birth and restricts the sample to parents who are younger than 50 years when we measure their incomes (for children born in 1980). Across these subsamples, the IGE estimates range from 0.264 (for children of single parents, excluding children with zero income) to 0.697 (for male children, recoding zeroes to $1).

The IGE is unstable because the income distribution is not well approximated by a bivariate log-normal distribution, a result that was not apparent in smaller samples used in prior work. This makes it difficult to obtain reliable comparisons of mobility across samples or geographical areas using the IGE. For example, income measures in survey data are typically top-coded and sometimes include transfers and other sources of income that increase incomes at the bottom of the distribution, which may lead to larger IGE estimates than those obtained in administrative data sets such as the one used here.

In a recent paper, Mitnik et al. (2014) propose a new measure of the IGE, the elasticity of *expected* child income with respect to parent income $\left(\frac{d\log\ E[Y_i|X_i=x]}{d\log\ x}\right)$, which they show is more robust to the treatment of small incomes. In large samples, one can estimate this parameter by regressing the log of mean child income in each percentile bin (plotted in Figure I Panel A) on the log of mean parent income in each bin. In Online Appendix C, we show that Mitnik et al.'s statistic can be interpreted as a dollar-weighted average of elasticities (placing greater weight on high-income children), whereas the traditional IGE weights all individuals with positive income equally. These two parameters need not coincide in general and the "correct" parameter depends on the policy question one seeks to answer. However, it turns out that in our data, the Mitnik et al. dollar-weighted IGE estimate is 0.335, very similar to our baseline IGE estimate of 0.344 when excluding children with zero income (Online Appendix Figure I Panel A).[22]

22. Mitnik et al. (2014) find larger estimates of the dollar-weighted IGE in their sample of tax returns. A useful direction for further work would be to understand why the two samples yield different IGE estimates.

In another recent study, Clark (2014) argues that traditional estimates of the IGE understate the persistence of status across generations because they are attenuated by fluctuations in realized individual incomes across generations. To resolve this problem, Clark estimates the IGE based on surname-level means of income in each generation and obtains a central IGE estimate of 0.8, much larger than that in prior studies. In our data, estimates of mobility based on surname means are similar to our baseline estimates based on individual income data (Online Appendix Table V). One reason that Clark (2014) may obtain larger estimates of intergenerational persistence is that his focus on distinctive surnames partly identifies the degree of convergence in income between racial or ethnic groups (Borjas 1992) rather than across individuals (see Online Appendix D for further details).[23]

2. Rank-Rank Estimates. Next we present estimates of the rank-rank slope, the second measure of relative mobility discussed in Section II. We measure the percentile rank of parents P_i based on their positions in the distribution of parent incomes in the core sample. Similarly, we define children's percentile ranks R_i 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.[24] Unless otherwise noted, we hold the definition of these ranks fixed based on positions in the aggregate distribution, even when analyzing subgroups.

Figure II Panel A presents a binned scatter plot of the mean percentile rank of children $E[R_i|P_i = p]$ versus their parents' percentile rank p. The conditional expectation of a child's rank given his parents' rank is almost perfectly linear. Using an OLS regression, we estimate that a 1 percentage point (pp) increase in parent rank is associated with a 0.341 pp increase in the child's mean

23. For example, Clark (2014, p. 60, Figure 3.10) compares the outcomes of individuals with the surname Katz (a predominantly Jewish name) versus Washington (a predominantly black name). This comparison generates an implied IGE close to 1, which partly reflects the fact that the black-white income gap has changed very little over the past few decades. Estimates of the IGE based on individual-level data (or pooling all surnames) are much lower because there is much more social mobility within racial groups.

24. 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.

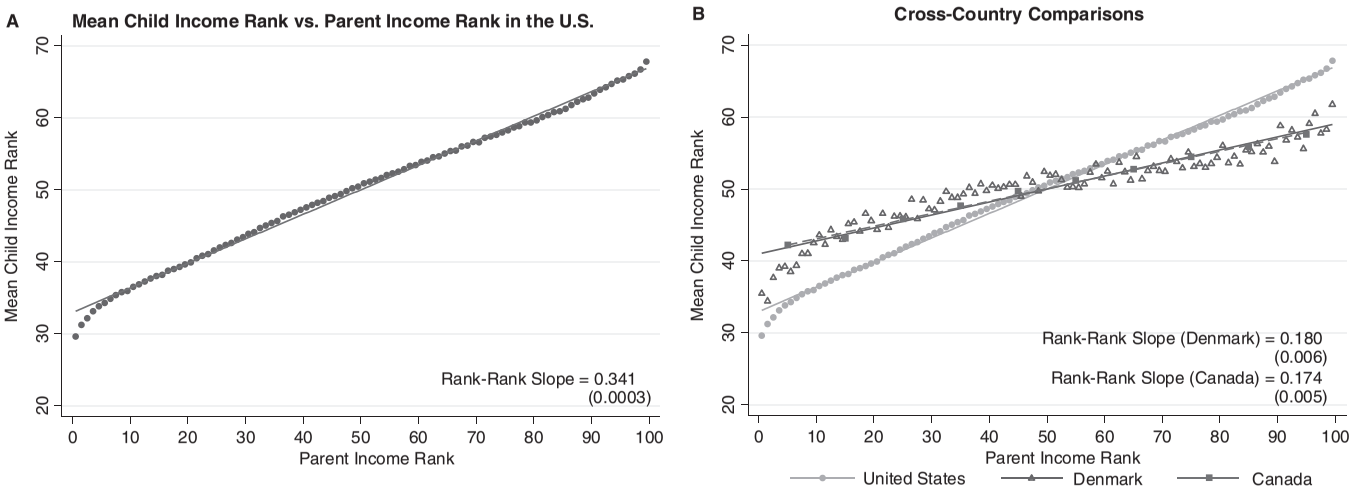

FIGURE II

Association between Children's and Parents' Percentile Ranks

These figures present nonparametric binned scatter plots of the relationship between children's and parent's percentile income ranks. Both figures are based on the core sample (1980–1982 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011–2012 family income (when the child is approximately 30 years old), and parent income is mean family income from 1996 to 2000. Children are ranked relative to other children in their birth cohort, and parents are ranked relative to all other parents in the core sample. Panel A plots the mean child percentile rank within each parent percentile rank bin. The series in triangles in Panel B plots the analogous series for Denmark, computed by Boserup, Kopczuk, and Kreiner (2013) using a similar sample and income definitions. The series in squares plots estimates of the rank-rank series using the decile-decile transition matrix from Corak and Heisz (1999). The series in circles in Panel B reproduces the rank-rank relationship in the United States from Panel A as a reference. The slopes and best-fit lines are estimated using an OLS regression on the microdata for the United States and on the binned series (as we do not have access to the microdata) for Denmark and Canada. Standard errors are reported in parentheses.

TABLE II
NATIONAL QUINTILE TRANSITION MATRIX

Child quintile	Parent quintile 1	2	3	4	5
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 (1980–1982 birth cohorts). See notes to Table I for income and sample definitions. See Online Appendix Table VI for an analogous transition matrix constructed using the 1980–1985 cohorts.

rank, as reported in row 4 of Table I. The rank-rank slope estimates are generally quite similar across subsamples, as shown in columns (2)–(7) of Table I.

Figure II Panel B compares the rank-rank relationship in the United States with analogous estimates for Denmark constructed using data from Boserup, Kopczuk, and Kreiner (2013) and estimates for Canada constructed from the decile transition matrix reported by Corak and Heisz (1999).[25] The relationship between child and parent ranks is nearly linear in Denmark and Canada as well, suggesting that the rank-rank specification provides a good summary of mobility across diverse environments. The rank-rank slope is 0.180 in Denmark and 0.174 in Canada, nearly half that in the United States.

Importantly, the smaller rank-rank slopes in Denmark and Canada do not necessarily mean that children from low-income families in these countries do better than those in the United States in absolute terms. It could be that children of high-income parents in Denmark and Canada have worse outcomes than children of high-income parents in the United States. One

25. Both the Danish and Canadian studies use administrative earnings information for large samples as we do here. The Danish sample, which was constructed to match the analysis sample in this article as closely as possible, consists of children in the 1980–1981 birth cohorts and measures child income based on mean income between 2009 and 2011. 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 to 1999, irrespective of their marital status. The Canadian sample is less comparable to our sample, as it consists of male children in the 1963–1966 birth cohorts and studies the link between their mean earnings from 1993 to 1995 and their fathers' mean earnings from 1978 to 1982.

cannot distinguish between these possibilities because the ranks are defined within each country. One advantage of the within–United States 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.

3. Transition Matrixes. Table II 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 return to later. This probability is 7.5% in the United States, compared with 11.7% in Denmark (Boserup, Kopczuk, and Kreiner 2013) and 13.4% in Canada (Corak and Heisz 1999). In this sense, the chances of achieving the American dream are considerably higher for children in Denmark and Canada than those in the United States.

In Online Data Table I, we report a 100×100 percentile-level transition matrix for the United States. Using this matrix and the marginal distributions for child and parent income in Online Data Table II, one can construct any mobility statistic of interest for the U.S. population.[26]

IV.B. Robustness of Baseline Estimates

We now evaluate the robustness of our estimates of intergenerational mobility to alternative specifications. We begin by evaluating two potential sources of bias emphasized in prior work: life cycle bias and attenuation bias.

1. Life Cycle 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 have steeper earnings profiles when they are young (Solon 1999; Grawe 2006; Haider and Solon 2006). To evaluate whether our baseline estimates suffer from such life cycle bias, Figure III Panel A plots estimates of the rank-rank slope by the age at which the child's income is measured. We construct the

26. All of the online data tables are available at http://www.equality-of-opportunity.org/index.php/data.

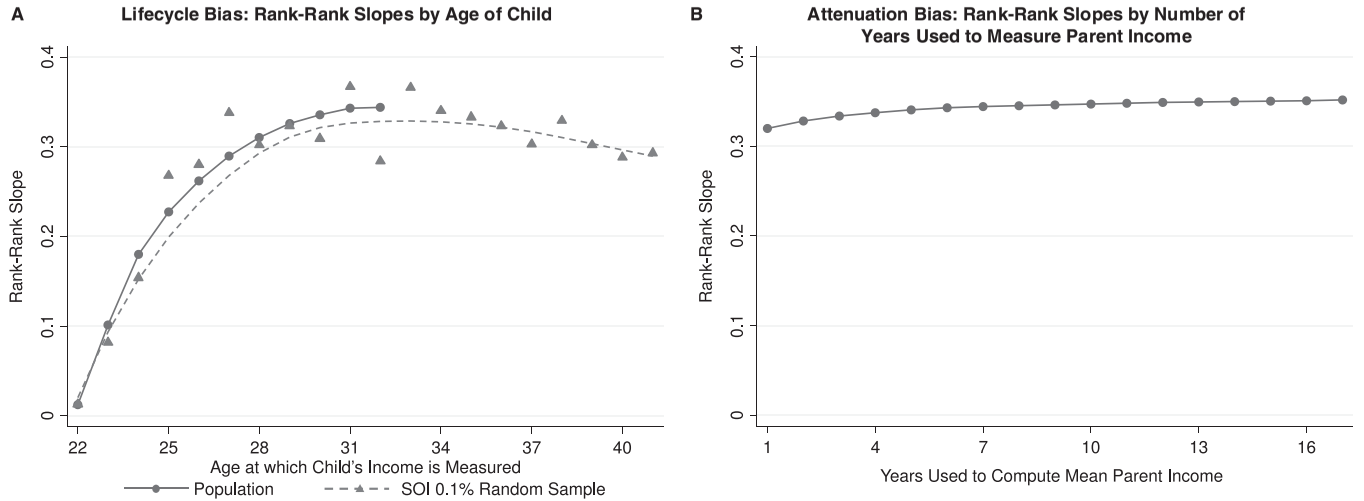

FIGURE III

Robustness of Intergenerational Mobility Estimates

This figure evaluates the robustness of the rank-rank slope estimated in Figure II Panel A 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 to 2012. In Panel A, parent income is defined as mean family income from 1996 to 2000. Each point in Panel A shows the slope coefficient from a separate OLS regression of child income rank on parent income rank, varying the child's birth cohort and hence the child's age in 2011–2012 when the child's income is measured. The circles use the extended sample in the population data, and the triangles use the 0.1% Statistics of Income (SOI) 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 2011–2012 (denoted by age 22 on the figure). The last point for which we have population-wide estimates corresponds to the 1980 cohort, who are 31–32 (denoted by 32) when their incomes are measured. The last point in the SOI sample corresponds to the 1971 cohort, who are 40–41 (denoted by 41) when their incomes are measured. The dashed line is a lowess curve fit through the SOI 0.1% sample rank-rank slope estimates. In Panel B, we focus on children in the core sample (1980–1982 birth cohorts) in the population data. Each point in this figure shows the coefficient from the same rank-rank regression as in Figure II Panel A, 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 to 1997. The last point uses mean parent income from 1996 to 2012, a 17-year average.

series in circles by measuring children's income as mean family income in 2011–2012 and parent income as mean family income between 1996 and 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 and 1990. For children in the 1980 birth cohort, we measure earnings in 2011–2012 at age 31–32 (denoted by 32 in the figure); for the 1990 cohort, we measure earnings at age 21–22.[27] The rank-rank slope rises very steeply in the early twenties 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 III.A, which contains data back to the 1971 birth cohort. The series in triangles in Figure III Panel A 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 life cycle bias provided that child income is measured after age 30, as in our baseline definition.

We also find that estimates of the IGE using the traditional log-log specification (limiting the sample between the 10th and 90th percentiles of the parent income distribution) stabilize around age 30, as shown in Online Appendix Figure II Panel A. In the population data, the IGE estimate is a strictly concave function of age and rises by only 1.7% from age 31 to 32. The SOI 0.1% sample exhibits a similar (albeit noisier) pattern.

An analogous life cycle bias can arise if parent income is measured at very old or young ages. In Online Appendix Figure II Panel B we plot the rank-rank slope using the core sample, varying the five-year window used to measure parent income from a starting year of 1996 (when mothers are 41 years old on average) to 2010 (when mothers are 55 years old). The rank-rank estimates exhibit virtually no variation with the age of parent income measurement within this range.

27. We obtain very similar results if we instead track a single cohort and vary age by measuring earnings in different calendar years.

A closely related concern is that parent income at earlier ages might matter more for children's outcomes, for example, if resources in early childhood are relevant for child development (e.g., Heckman 2006; Duncan, Ziol-Guest, and Kalil 2010). Although 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 II Panel C. 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 earnings records.[28]

2. Attenuation Bias. Income in a single year is a noisy measure of lifetime income, which attenuates estimates of intergenerational persistence (Solon (1992)). To evaluate whether our baseline estimates suffer from such attenuation bias, Figure III Panel B plots estimates of the rank-rank slope, varying the number of years used to calculate mean parent family income. In this figure, we plot the slope from an OLS regression of child rank on parent rank (as in row 4, column (1) of Table I), varying the number of years used to calculate mean parent income from one (1996 only) to 17 (1996–2012). The rank-rank slope based on five years of data (0.341) is 6.6% larger than the slope based on one year of parent income (0.320). Solon (1992) finds a 33% increase in the IGE (from 0.3 to 0.4) when using a five-year average instead of one year of data in the PSID. We find less attenuation bias for three reasons: (i) income is measured with less error in the tax data than in the PSID; (ii) we use family income measures

28. Although 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 to 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 II Panel C if there were a discontinuous jump in the gradient prior to age 2.

rather than individual income, which fluctuates more across years; and (iii) 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.

Mazumder (2005) reports that even five-year averages of parent income yield attenuated estimates of intergenerational persistence relative to longer time averages. Contrary to this result, we find that the rank-rank slope is virtually unchanged by adding more years of data beyond 5 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, with a higher imputation rate when measuring parent income using more years (see Online Appendix E for further details). Such imputations are analogous to instrumenting for income with race and education, which is known to yield upward-biased estimates of intergenerational persistence (Solon 1992).

We analyze the effect of varying the number of years used to measure the child's income in Online Appendix Figure II Panel D. The rank-rank slope increases very little 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. An ancillary implication of this result is that our estimates of intergenerational mobility are not sensitive to the calendar year in which we measure children's incomes. This finding is consistent with the results of Chetty et al. (2014), who show that estimates of intergenerational mobility do not vary significantly across birth cohorts when income is measured at a fixed age.

3. Alternative Income Definitions. In rows 5–8 of Table I, we explore the robustness of the baseline rank-rank estimate to alternative definitions of child and parent income. In row 5, we verify that the missing W-2 data from 1996 to 1998 does not create significant bias by defining parent income as mean income from 1999 to 2003. The rank-rank estimates are virtually unchanged with this redefinition.

In row 6, we define the parent's rank based on the individual income of the parent with higher mean income from 1999

to 2003.[29] 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 percent, 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 (i) most of the variation in parent income across households is not due to differences in marital status and (ii) the mean ranks of children with married parents are only 4.6 percentile points higher than those with single parents.

Next, we consider alternative income definitions for the children. Here, one 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 percent reduction. For female children, using individual income reduces the rank-rank correlation from 0.346 to 0.257, a 26 percent reduction. The change may be larger for women because women from high income families tend to marry high-income men and may choose not to work.

Finally, in row 8 of Table I, we define a measure of child income that excludes capital and other nonlabor income using the sum of individual wage earnings, UI benefits, SSDI benefits, and Schedule C self-employment income. We divide self-employment income by two for married individuals. This individual earnings measure also yields virtually identical estimates of the rank-rank slope.

29. 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.

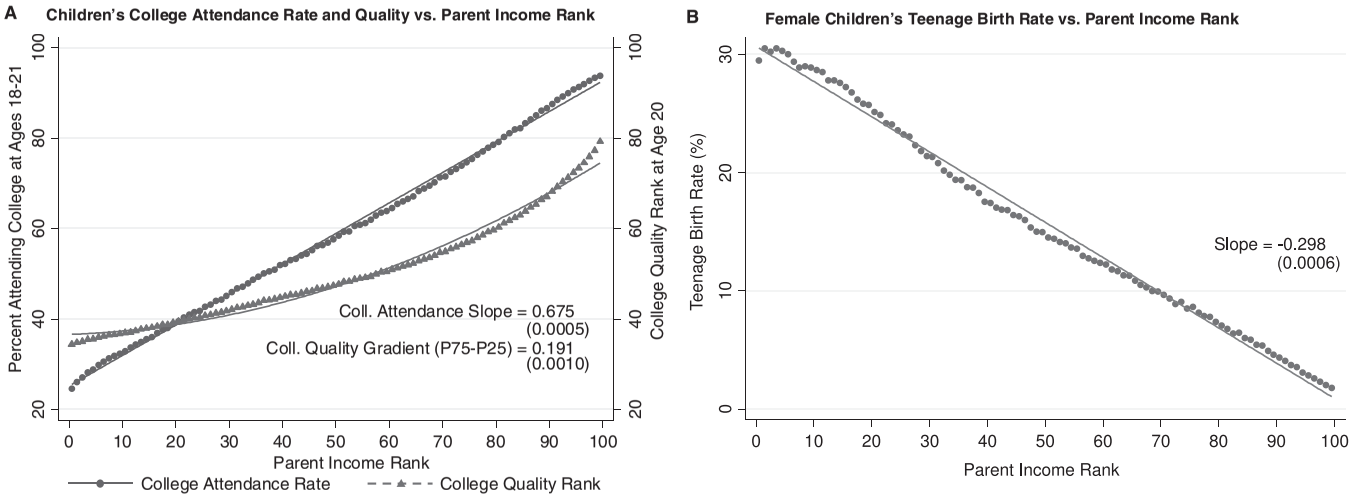

FIGURE IV

Gradients of College Attendance and Teenage Birth by Parent Rank

These figures present nonparametric binned scatter plots of the relationship between children's college attendance rates (Panel A, circles), college quality rank (Panel A, triangles), and teenage birth rates (Panel B) versus parents' percentile rank. Both figures are based on the core sample (1980–1982 birth cohorts). Parent rank is defined based on mean family income from 1996 to 2000. In Panel A, the circles plot the fraction of children ever attending college between age 18–21 within each parent-income percentile bin; the triangles plot the average college quality rank at age 20 within each parent-income percentile bin. College attendance is defined as the presence of a 1098-T form filed by a college on behalf of the student. College quality rank is defined as the percentile rank of the college that the child attends at age 20 based on the mean earnings at age 31 of children who attended the same college (children who do not attend college are included in a separate "no college" group); see Section III.B for further details. Panel B plots the fraction of female children who give birth while teenagers within each parental percentile bin. Having a teenage birth is defined as ever claiming a dependent child who was born while the mother was aged 13–19. The slopes and best-fit lines for college attendance and teenage birth are estimated using linear regressions of the outcome of interest on parent income rank in the microdata. We regress college quality rank on a quadratic in parent rank to match the nonlinearity of the relationship. The college quality gradient is defined as the difference between the fitted values for children with parents at the 75th percentile and parents at the 25th percentile using this quadratic specification.

IV.C. *Intermediate Outcomes: College Attendance and Teenage Birth*

We supplement our analysis of intergenerational income mobility by studying the relationship between parent income and two intermediate outcomes for children: college attendance and teenage birth.

The series in circles in Figure IV Panel A presents a binned scatter plot of the college attendance rate of children versus the percentile rank of parent family income using the core sample. College attendance is defined as attending college in one or more years between the ages 18 and 21. The relationship between college attendance rates and parental income rank is again virtually linear, with a slope of 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 series in triangles in Figure IV Panel A plots college quality ranks versus parent ranks. We define a child's college quality rank based on the mean earnings at age 30 of students who attended each college at age 20. The 54 percent of children who do not attend college at age 20 are included in this analysis and are assigned the mean rank for the noncollege group, which is approximately $\frac{54}{2} = 27$ (see Section III.B for details). The relationship between college quality rank and parent income rank is convex because most children from low-income families do not attend college, and hence increases in parent income have little impact on college quality rank at the bottom. To account for this non-linearity, we regress college quality ranks on a quadratic function of parent income rank and define the gradient in college quality as the difference in the predicted college quality rank for children with parents at the 75th percentile and children with parents at the 25th percentile. The P25–75 gap in college quality ranks is 19.1 percentiles in our core sample.

Figure IV Panel B plots teenage birth rates for female children versus parent income ranks. Teenage birth is defined (for females only) as having a child when the mother is aged 13–19. There is a 29.8 percentage point gap in teenage birth rates between children from the highest- and lowest-income families.

These correlations between intermediate outcomes and parent income ranks do not vary significantly across subsamples or birth cohorts, as shown in rows 9–11 of Table I. The

strength of these correlations indicates that much of the divergence between children from low- versus high-income families emerges well before they enter the labor market, consistent with the findings of prior work (such as Neal and Johnson 1996; Cameron and Heckman 2001; Bhattacharya and Mazumder 2011).

V. Spatial Variation in Mobility

We now turn to our central goal of characterizing the variation in intergenerational mobility across areas within the United States. We begin by defining measures of geographic location. We then present estimates of relative and absolute mobility by area and assess the robustness of these estimates to alternative specifications.

V.A. Geographical Units

To characterize the variation in children's outcomes across areas, one must first partition the United States 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 on conditions in their immediate neighborhood (such as peers or resources in their city block), local community (such as the quality of schools in their county), and broader metro area (such as 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 characterize intergenerational mobility at the level of commuting zones. 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 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 (MSAs), but unlike MSAs, they cover the entire United States, including rural areas. There are 741 CZs in the United States; on average, each CZ contains four counties and has a population of 380,000. See Online Appendix Figure III for an illustration of the Boston CZ.

We focus on CZ-level variation because mobility statistics in very small neighborhoods are likely to be heavily affected by sorting. Because property prices are typically homogeneous within narrow areas and home values are highly correlated with parent income, comparisons within a small neighborhood effectively condition on a proxy for parent income. As a result, the variation in parent income across individuals in a small area (such as a city block) must be correlated with other latent factors that could affect children's outcomes directly, making it difficult to interpret the resulting mobility estimates.[30] 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 variation may be even greater than that documented below.[31]

We permanently assign each child to a single CZ based on the ZIP code from which his or his 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.[32] For children in our core sample of 1980–1982 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 percent 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

30. For example, it would be difficult to estimate the degree of intergenerational mobility on Park Avenue in Manhattan because any families with low observed income in such a high-property-value area would have to be latently wealthy to be able to afford to live there.

31. We also report statistics by MSA in Online Data Table IV. For CZs that intersect MSAs, correlations between CZ-level and MSA-level mobility statistics exceed 0.9.

32. Location is measured after 1996 for approximately 3% of children because they were linked to parents based on tax returns filed after 1996. We have no information on location for the remaining 1% of children in the national sample because the ZIP code listed on the parent's tax returns is invalid or missing (see Online Appendix Table I); these children are excluded from the analysis in the remainder of the article.

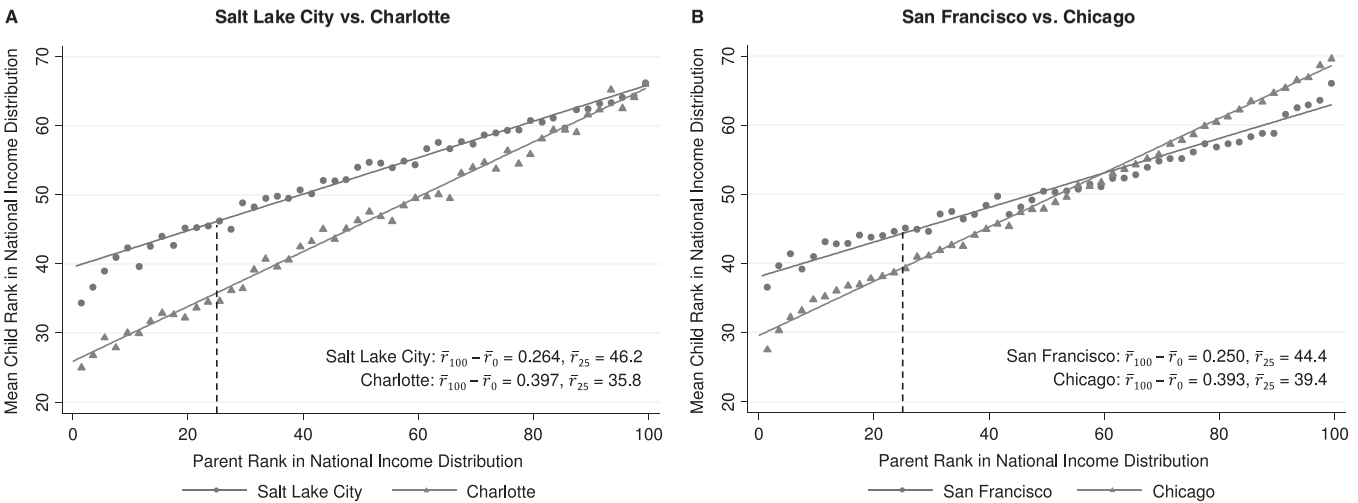

FIGURE V

Intergenerational Mobility in Selected Commuting Zones

These figures present nonparametric binned scatter plots of the relationship between child and parent income ranks in selected CZs. Both panels are based on the core sample (1980–1982 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to CZs based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. Parent and child percentile ranks are always defined at the national level, not the CZ level. To construct each series, we group parents into 50 equally sized (2 percentile point) bins and plot the mean child percentile rank versus the mean parent percentile rank within each bin. We report two measures of mobility based on the rank-rank relationships in each CZ. The first is relative mobility ($\bar{r}_{100} - \bar{r}_0$), which is 100 times the rank-rank slope estimate. The second is absolute upward mobility ($\bar{r}_{25}$), the predicted child income rank at the 25th percentile of parent income distribution, depicted by the dashed vertical line in the figures. All mobility statistics and best-fit lines are estimated on the underlying microdata (not the binned means).

attendance and teenage birth) are similar if we define CZs based on location at age 5 instead of age 16.

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

V.B. Measures of Relative and Absolute Mobility

In our baseline analysis, we measure mobility at the CZ level using the core sample (1980–1982 birth cohorts) and the definitions of parent and child family income described in Section III.B. 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).

We begin by examining the rank-rank relationship in selected CZs. Figure V Panel A presents a binned scatter plot of the mean child rank versus parent rank for children who grew up in the Salt Lake City, UT (circles), or Charlotte, NC (triangles), CZs. The rank-rank relationship is virtually linear in these CZs. The linearity of the rank-rank relationship is a remarkably robust property across CZs, as illustrated for the 20 largest CZs in Online Appendix Figure IV.

Exploiting this approximate linearity, we summarize the conditional expectation of a child's rank given his parents' rank in each CZ using two parameters: a slope and an intercept. Let R_{ic} denote the national income rank (among children in his birth cohort) of child i who grew up in CZ c. Similarly, let P_{ic} denote his parent's rank in the income distribution of parents in the core sample. We estimate the slope and intercept of the rank-rank relationship in CZ c by regressing child rank on parent rank:

$$R_{ic} = \alpha_c + \beta_c P_{ic} + \varepsilon_{ic}. \tag{2}$$

The slope of the rank-rank relationship (β_c) in equation (2) measures degree of relative mobility in CZ c, as defined in Section II. In Salt Lake City, $\beta_c = 0.264$.[33] The difference between the expected ranks of children born to parents at the top and bottom of the income distribution is $\bar{r}_{100,c} - \bar{r}_{0,c} = 100 \times \beta_c = 26.4$ in Salt Lake City. There is much less relative mobility (i.e., much greater

33. We always measure percentile ranks on a 0–100 scale and slopes on a 0–1 scale, so α_c ranges from 0 to 100 and β_c ranges from 0 to 1 in equation (3).

persistence of income across generations) in Charlotte, where $\bar{r}_{100} - \bar{r}_0 = 39.7$.

Following the discussion in Section II, we define absolute mobility at percentile p in CZ c as the expected rank of a child who grew up in CZ c with parents who have a national income rank of p:

$$\bar{r}_{pc} = \alpha_c + \beta_c p. \tag{3}$$

We focus much of our analysis on average absolute mobility for children from families with below-median parent income in the national distribution ($E[R_{ic}|P_{ic} < 50]$), which we call absolute upward mobility.[34] Because the rank-rank relationship is linear, the average rank of children with below-median parent income equals the average rank of children with parents at the 25th percentile in the national distribution ($\bar{r}_{25,c} = \alpha_c + 25\beta_c$), illustrated by the dashed vertical line in Figure V Panel A. Absolute upward mobility is $\bar{r}_{25} = 46.2$ in Salt Lake City, compared with $\bar{r}_{25} = 35.8$ 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 are 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 V Panel B shows that San Francisco has substantially higher relative mobility than Chicago: $\bar{r}_{100} - \bar{r}_0 = 25.0$ in San Francisco versus $\bar{r}_{100} - \bar{r}_0 = 39.3$ in Chicago. But part of the greater relative mobility in San

34. We integrate over the national parent income distribution rather than the local distribution when defining $E[R_{ic}|P_{ic} < 50]$ to ensure that our cross-CZ comparisons are not affected by differences in local income distributions. We focus on the absolute outcomes of children from low-income families both because the outcomes of disadvantaged youth are a central focus of policy interest 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. 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.

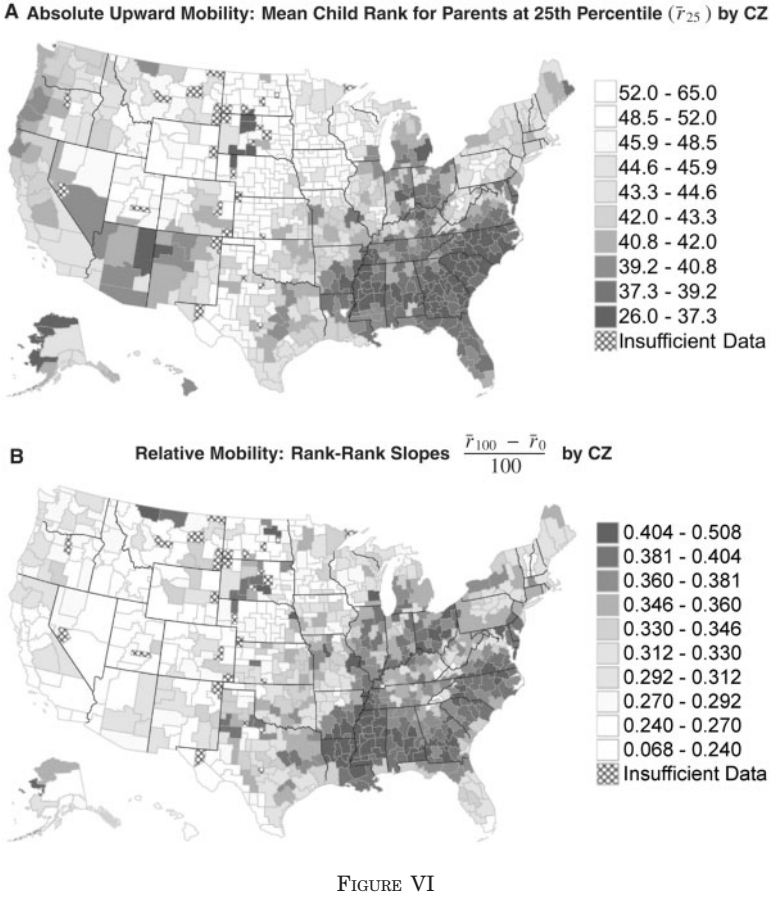

FIGURE VI

The Geography of Intergenerational Mobility

These figures present heat maps of our two baseline measures of intergenerational mobility by CZ. Both figures are based on the core sample (1980–1982 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to CZs based on the location of their parents (when the child was claimed as a dependent), irrespective of where they live as adults. In each CZ, we regress child income rank on a constant and parent income rank. Using the regression estimates, we define absolute upward mobility ($\bar{r}_{25}$) as the intercept $+ 25 \times$ (rank-rank slope), which corresponds to the predicted child rank given parent income at the 25th percentile (see Figure V). We define relative mobility as the rank-rank slope; the difference between the outcomes of the child from the richest and poorest family is 100 times this coefficient ($\bar{r}_{100} - \bar{r}_0$). The maps are constructed by grouping CZs into 10 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.

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 versus absolute mobility (or, equivalently, on the weights one places on absolute mobility at each percentile p).

V.C. Baseline Estimates by CZ

We estimate equation (2) using OLS to calculate absolute upward mobility ($\bar{r}_{25,c} = \alpha_c + 25\beta_c$) and relative mobility (β_c) by CZ. The estimates for each CZ are reported in Online Data Table V.

1. Absolute Upward Mobility. Figure VI Panel A presents a heat map of absolute upward mobility. We construct this map by dividing CZs into deciles based on their estimated value of $\bar{r}_{25,c}$. Lighter colors represent deciles with higher levels of $\bar{r}_{25,c}$.[35] Upward mobility varies significantly across areas. CZs in the top decile have $\bar{r}_{25,c} > 52.0$, whereas those in the bottom decile have $\bar{r}_{25,c} < 37.4$. Note that the 37th percentile of the family income distribution for children at age 30 is \$22,900, whereas 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 of $\bar{r}_{25,c}$ is 5.68; the population-weighted standard deviation is 3.34. The unconditional standard deviation of children's income ranks (which have a uniform distribution) is $\frac{100}{\sqrt{12}} = 28.9$. Hence, a 1 standard

35. 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 V, we present a version of this map in which we use data from the 1980–1985 cohorts to estimate mobility for the CZs that have fewer than 250 observations in the core (1980–1982) 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.

deviation improvement in CZ "quality"—as measured by its level of absolute upward mobility $\bar{r}_{25,c}$—is associated with a $\frac{5.68}{2.89} = 0.20$ standard deviation increase in the expected income rank of children whose parents are at the 25th percentile.[36] For comparison, a 1 standard deviation increase in parent income rank is associated with a 0.34 standard deviation increase in a child's income rank (Figure II Panel A). Hence, a 1 standard deviation improvement in CZ quality is associated with 60 percent as large an increase in a child's income as a 1 standard deviation increase in his own parent's income.

There are three broad spatial patterns in upward mobility evident in Figure VI Panel A. 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 percent of the cross-CZ variance in absolute upward mobility is within the nine census divisions and 36 percent 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 some of the largest Native American reservations in the United States 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

36. An analogous calculation using the estimates of college attendance gradients by CZ in Section IV.C implies that a 1 standard deviation increase in CZ quality is associated with a 0.19 standard deviation (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, p. 390) estimate that a 1 standard deviation increase in neighborhood quality is associated with a 0.32 standard deviation 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.

TABLE III

INTERGENERATIONAL MOBILITY IN THE 50 LARGEST COMMUTING ZONES

(1) Upward mobility rank	(2) CZ name	(3) Population	(4) Absolute upward mobility	(5) P(child in Q5 | parent in Q1)	(6) Pct. above poverty line	(7) Relative mobility rank-rank slope
1	Salt Lake City, UT	1,426,729	46.2	10.8	77.3	0.264
2	Pittsburgh, PA	2,561,364	45.2	9.5	74.9	0.359
3	San Jose, CA	2,393,183	44.7	12.9	73.5	0.235
4	Boston, MA	4,974,945	44.6	10.5	73.7	0.322
5	San Francisco, CA	4,642,561	44.4	12.2	72.5	0.250
6	San Diego, CA	2,813,833	44.3	10.4	74.3	0.237
7	Manchester, NH	1,193,391	44.2	10.0	75.0	0.296
8	Minneapolis, MN	2,904,389	44.2	8.5	75.2	0.338
9	Newark, NJ	5,822,286	44.1	10.2	73.7	0.350
10	New York, NY	11,781,395	43.8	10.5	72.2	0.330
11	Los Angeles, CA	16,393,360	43.4	9.6	73.8	0.231
12	Providence, RI	1,582,997	43.4	8.2	73.6	0.333
13	Washington DC	4,632,415	43.2	11.0	72.2	0.330
14	Seattle, WA	3,775,744	43.2	10.9	72.0	0.273
15	Houston, TX	4,504,013	42.8	9.3	74.7	0.325
16	Sacramento, CA	2,570,609	42.7	9.7	71.3	0.257
17	Bridgeport, CT	3,405,565	42.4	7.9	72.4	0.359
18	Fort Worth, TX	1,804,370	42.3	9.1	73.6	0.320
19	Denver, CO	2,449,044	42.2	8.7	73.3	0.294
20	Buffalo, NY	2,369,699	42.0	6.7	73.1	0.368
21	Miami, FL	3,955,969	41.5	7.3	76.3	0.267
22	Fresno, CA	1,419,998	41.3	7.5	71.3	0.295
23	Portland, OR	1,842,889	41.3	9.3	70.5	0.277
24	San Antonio, TX	1,724,863	41.1	6.4	74.3	0.320
25	Philadelphia, PA	5,602,247	40.8	7.4	69.6	0.393
26	Austin, TX	1,298,076	40.4	6.9	71.9	0.323
27	Dallas, TX	3,405,666	40.4	7.1	72.6	0.347
28	Phoenix, AZ	3,303,211	40.3	7.5	70.6	0.294
29	Grand Rapids, Michigan	1,286,045	40.1	6.4	71.3	0.378
30	Kansas City, MI	1,762,873	40.1	7.0	70.4	0.365
31	Las Vegas, NV	1,568,418	40.0	8.0	71.1	0.259
32	Chicago, IL	8,183,799	39.4	6.5	70.8	0.393
33	Milwaukee, WI	1,660,659	39.3	4.5	70.3	0.424
34	Tampa, FL	2,395,997	39.1	6.0	71.3	0.335
35	Orlando, FL	1,697,906	39.1	5.8	71.5	0.326
36	Port St. Lucie, FL	1,533,306	39.0	6.2	71.2	0.303
37	Baltimore, MD	2,512,431	38.8	6.4	67.7	0.412
38	St. Louis, MO	2,325,609	38.4	5.1	69.0	0.413
39	Dayton, OH	1,179,009	38.3	4.9	68.2	0.397
40	Cleveland, OH	2,661,167	38.2	5.1	68.7	0.405
41	Nashville, TN	1,246,338	38.2	5.7	67.9	0.357
42	New Orleans, LA	1,381,652	38.2	5.1	69.5	0.397
43	Cincinnati, OH	1,954,800	37.9	5.1	66.4	0.429
44	Columbus, OH	1,663,807	37.7	4.9	67.1	0.406

TABLE III
(CONTINUED)

(1) Upward mobility rank	(2) CZ name	(3) Population	(4) Absolute upward mobility	(5) P(child in Q5 | parent in Q1)	(6) Pct. above poverty line	(7) Relative mobility rank-rank slope
45	Jacksonville, FL	1,176,696	37.5	4.9	68.9	0.361
46	Detroit, MI	5,327,827	37.3	5.5	68.5	0.358
47	Indianapolis, IN	1,507,346	37.2	4.9	67.5	0.398
48	Raleigh, NC	1,412,127	36.9	5.0	67.3	0.389
49	Atlanta, GA	3,798,017	36.0	4.5	69.4	0.366
50	Charlotte, NC	1,423,942	35.8	4.4	67.0	0.397

Notes. This table reports estimates of intergenerational mobility for the 50 largest 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 (1980–1982 birth cohorts) and the baseline family income definitions described in Table I (except for column (5), which uses the 1980–1985 birth cohorts). The measures in columns (4) and (7) are both derived from within-CZ OLS regressions of child income rank against parent income rank. Column (7) reports the slope coefficient from this regression, which is equal to the difference in mean child income rank between children with parents in the 100th percentile and children with parents in the 0th percentile (divided by 100). Column (4) reports the predicted value at parent income rank equal to 25. Column (5) reports the percentage of children whose family income is in the top quintile of the national distribution of child family income conditional on having parent family income in the bottom quintile of the parental national income distribution. These probabilities are taken directly from Online Data Table VII. Column (6) reports the fitted values at parent rank 25 from a regression of an indicator for child family income being above the poverty line on parent income rank (see Online Appendix F for details). See Online Data Table V for estimates for all CZs as well as estimates using alternative samples and income definitions.

average. For instance, children from low-income families who grow 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 $\bar{r}_{25,c} = 41.7$, wheras rural areas have $\bar{r}_{25,c} = 45.8$. 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. Of children who grow up in rural areas, 44.6% live in urban areas at age 30. Among those who rose from the bottom quintile of the national income distribution to the top quintile, 55.2 percent of children who grew up in rural areas live in urban areas at age 30.

Table III 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 especially noteworthy because these are generally considered to be booming cities in the South with relatively high rates of job growth.

In column (5) of Table III, 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.[37] To improve precision in smaller CZs, we estimate this probability pooling the 1980–1985 birth cohorts.[38] 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 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. The chances of rising from the bottom fifth to the top fifth for children growing up in San Jose are comparable to those in Denmark and Canada (see Section IV.A). Note that if parent income played no role in determining children's outcomes, all the quintile transition probabilities would be 20%. Hence, the variation in rates of upward mobility across areas is large relative to the maximum plausible range of 0 to 20%.

In column (6) of Table III, we consider another measure of absolute upward mobility: the probability that a child has family income above the poverty line conditional on having parents at

37. In principle, differences in local income distributions within the bottom quintile could generate differences in this probability. In an earlier version of this analysis (v1.0 available on the project website, www.equality-of-opportunity.org), we accounted for these differences by calculating the chance of reaching the top quintile separately for each percentile and computed the unweighted mean across the percentiles, effectively integrating over the national parent income distribution. The adjusted CZ-level transition probabilities obtained using this approach were virtually identical to the raw transition probabilities we report in this article.

38. We verify that including more recent cohorts does not generate significant bias by showing that the national transition matrix based on the 1980–1985 cohorts (Online Appendix Table VI) is virtually identical to the matrix based on the 1980–1982 cohorts in Table II. We report the quintile transition matrix for each CZ in Online Data Table VI and provide statistics on the marginal distributions of parent and child income by CZ in Online Data Table VII.

the 25th percentile (see Online Appendix F for details on the construction of this measure). This statistic also generates very similar rankings across CZs, confirming that our results are not sensitive to the way we measure upward mobility.

2. Relative Mobility. Figure VI Panel B presents a heat map of relative mobility. This map is constructed 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 versus 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 $\mu_{pc} = E[R_{ic}|P_{ic} = p]$ denote a child's expected rank given a parent rank of p in CZ c. We estimate μ_{pc} in each CZ nonparametrically as the mean value of R_{ic} for children in each percentile bin of the parent income distribution $\mathrm{p} = 0, \ldots, 99$.[39] 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:

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

In this equation, γ_p measures the association across CZs between a 1 unit increase in β_c (i.e., greater intergenerational

39. The expected value μ_{pc} differs from $\bar{r}_{pc}$ defined above because μ_{pc} is estimated nonparametrically using only data in percentile bin p, whereas $\bar{r}_{pc}$ is calculated based on the linear approximation to the rank-rank relationship in equation (3). 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 $\bar{r}_{pc}$ exceeds 0.99. We use the linear approximation $\bar{r}_{pc}$ 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 nonparametrically, we use μ_{pc} here.

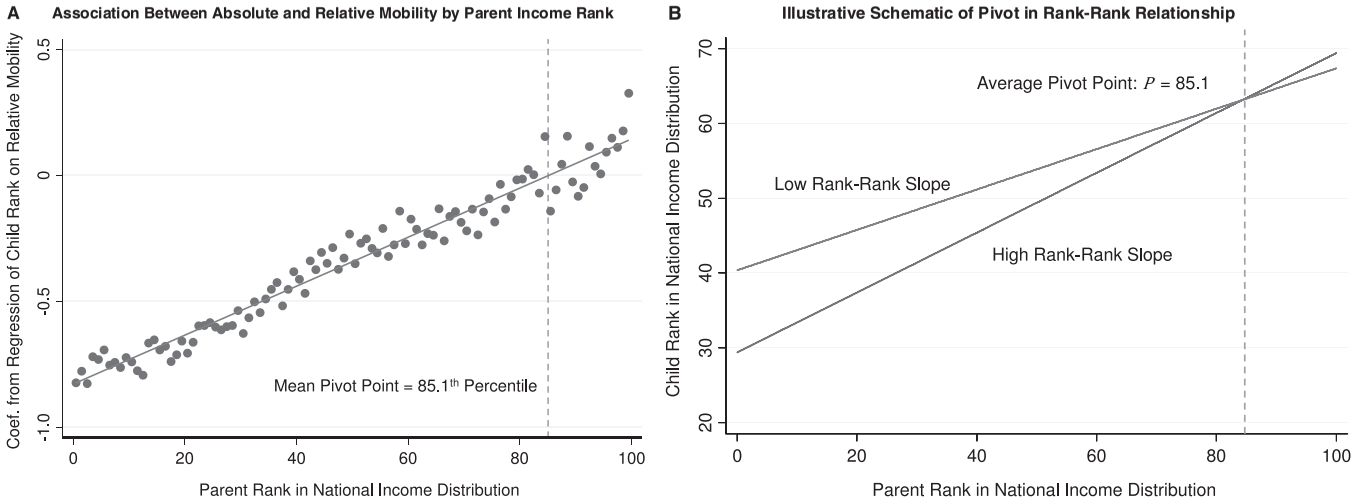

FIGURE VII

Relationship between Absolute and Relative Mobility

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 $\overline{\mu}_{pc}$. We then run a CZ-level regression of $\overline{\mu}_{pc}$ on relative mobility $(\overline{r}_{100c} - \overline{r}_{0c})$ at each percentile p separately. Panel A plots the resulting regression coefficients γ_p versus 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 0 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.

persistence) and the mean rank of children with parents 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 VII Panel A 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 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. This shows that 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.[40]

Figure VII Panel B 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 VII Panel A 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 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.[41] This may be because the rich are able to insulate themselves from differences in the local environment. If the

40. If the rank-rank relationship were perfectly linear, the relationship plotted in Figure VII Panel A 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.

41. 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

differences in relative mobility across areas are caused by differences in local policies, this result suggests that policies that improve relative mobility may be able to improve the outcomes of children from poor families without hurting children from high-income families significantly.

V.D. Robustness of Spatial Patterns

We assess the robustness of the spatial patterns in mobility along several dimensions. The results of this robustness analysis are reported in Online Appendix F and Appendix Table VII; we present a brief summary here.

We begin by considering changes in sample definitions: limiting the sample to male versus female children, married versus single parents, and later birth cohorts (for which we measure children's location at earlier ages). Measures of both absolute and relative mobility across areas in these subsamples generally have a correlation of more than 0.9 with the corresponding baseline measures reported above. Restricting the sample to hold the parents' ages at the birth of child fixed, limiting the sample to children who stay in the CZ where they grew up as adults, and limiting the sample to children linked to only one parent in all years yield very similar estimates of mobility across areas.

We also find that the spatial patterns are highly robust to using alternative measures of income used in Table I. For example, using individual income instead of family income or wage earnings instead of total income yields very spatial patterns.

We evaluate whether adjusting for differences in cost-of-living across areas affects our estimates by dividing parents' income by a local price index (based on the ACCRA survey) for the CZ where their child grew up and the child's income by the price index for the CZ where he lives in 2012 to obtain real income measures. Measures of intergenerational mobility based on real incomes are very highly correlated with our baseline measures. The degree of upward mobility—that is, the difference between the child's rank and the parent's rank—is essentially unaffected by adjusting for local prices because few children move to areas with very different levels of cost of living relative to their parents (see Online Appendix F for details).

affluent parents become very high-income "superstars" may therefore differ significantly across areas.

Because we measure parent income before 2000 and child income in 2011–2012, part of the variation in upward mobility across areas could be driven by shocks to local economic growth. While growth shocks—for example, 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 attributable to more stable factors that can be manipulated by policy. We assess the extent to which economic growth is responsible for the spatial variation in upward mobility in two ways. First, we define parent income as mean family income in 2011–2012, 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. Second, we regress upward mobility on the CZ-level growth rate from 2000 to 2010 and calculate residuals. Both of these growth-adjusted mobility measures have a correlation of more than 0.8 with our baseline measures, indicating that most of the spatial variation in upward mobility is not driven by differences in growth rates.

Finally, we consider a set of alternative statistics for relative and absolute mobility. Estimating relative mobility based on parent and child ranks in the local income distribution yields estimates that are very highly correlated with our baseline estimates based on national ranks. We also show that the two alternative measures of upward mobility analyzed in Table III—the probability of rising from the bottom fifth to the top fifth and the probability of having income above the poverty line conditional on having parents at the 25th percentile – also generate very similar spatial patterns, with correlations above 0.9 with our baseline mean rank measure of upward mobility (Online Appendix Figure VI).

V.E. Intermediate Outcomes: College Attendance and Teenage Birth

To better understand the sources of the spatial variation in intergenerational income mobility, we characterize spatial variation in the three intermediate outcomes analyzed in Figure IV: college attendance rates, college quality rank, and teenage birth rates. We first regress each of these outcomes on parent national income rank in each CZ c using specifications analogous to equation (2). We then characterize spatial variation in two

measures of mobility for each outcome using the regression estimates: the slope coefficient, which is analogous to our measure of relative mobility, and the predicted outcome for children with parents at the 25th percentile, which is analogous to our measure of absolute mobility.[42]

We present heat maps for the relative and absolute mobility measures for the three intermediate outcomes in Online Appendix Figures VII–IX; the CZ-level data underlying these maps are reported in Online Data Table V. There is substantial spatial variation in all three intermediate outcomes and the variation is highly correlated with the variation in the intergenerational income mobility. For example, college attendance rates for children with parents at the 25th percentile vary from less than 32.4% in bottom decile of CZs to more than 55.6% in the top decile of CZs. The unweighted correlation between college attendance rates at the 25th percentile and mean income ranks at the 25th percentile (absolute upward mobility) across CZs is 0.71 (Online Appendix Table VII, row 23). Similarly, teenage birth rates for female children whose parents are at the 25th percentile vary from less than 15.4% in the bottom decile of CZs to more than 29.4% in the top decile. The correlation between teen birth rates and absolute upward mobility is −0.61.

An important implication of these results is that much of the difference in intergenerational mobility across areas emerges while children are teenagers, well before they enter the labor market as adults.[43] This suggests that the spatial variation in income mobility is driven by factors that either directly affect children at early ages (such as 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.

42. Because the relationship between college quality rank and parent rank is not linear, we regress college quality ranks on a quadratic function of parent income rank and define the relative mobility measure for college quality as the difference in the predicted college quality rank for children with parents at the 75th percentile and children with parents at the 25th percentile, as in Figure IV Panel A.

43. Further supporting this claim, we find a strong positive correlation of 0.63 between teenage labor force participation rates (between the ages of 14 and 16) and upward mobility (see Figure VIII and Online Appendix H).

VI. Correlates of Intergenerational Mobility

Why do some areas of the United States 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 is merely to document a set of stylized facts to guide the search for causal determinants and the development of new models of intergenerational mobility.

We correlate our mobility statistics with various factors that have been discussed in the sociology and economics literature, such as segregation and inequality. 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 over time. For most covariates, we use data from the 2000 census and other publicly available data sets because many variables cannot be consistently measured in earlier years. See Online Appendix G 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.

Figure VIII presents a summary of our correlational results. It plots the unweighted univariate correlation between absolute upward mobility and various CZ-level characteristics, using all CZs with available data for the relevant variable. We consider several proxies for each broad factor (segregation, inequality, etc.). The dots show the point estimate of the correlation and the horizontal lines show a 95 percent confidence interval, based on standard errors clustered at the state level. The sign of the correlation is shown in parentheses next to each variable. In Online Appendix Table VIII, we report these correlations as well as estimates from several alternative specifications including state fixed effects, weighting CZs by population, restricting to urban areas, and controlling for differences in racial demographics and income growth (see Online Appendix H for details). These alternative specifications generally yield very similar results to the baseline estimates shown in Figure VIII. Most important, the correlations discussed below hold even in specifications with state fixed effects, showing that the results are not just driven by broad regional differences across the South versus other parts of the country. We also show in Online Appendix Table VIII that the factors that are positively

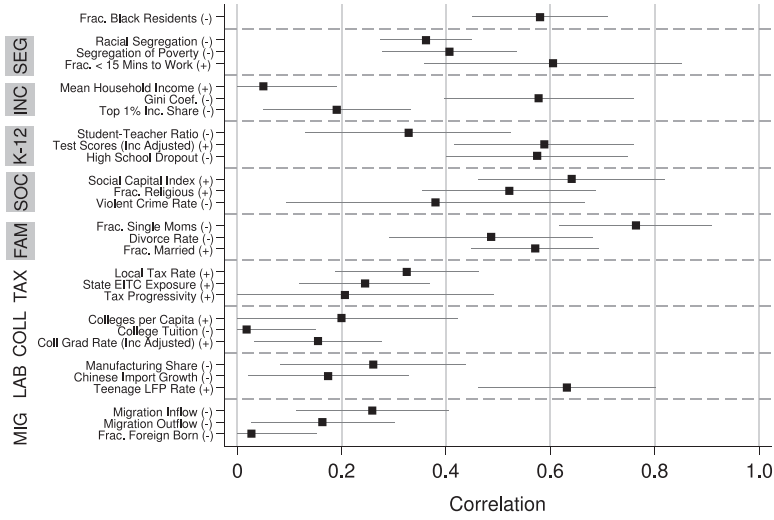

FIGURE VIII

Correlates of Spatial Variation in Upward Mobility

This figure shows the correlation of various CZ-level characteristics with absolute upward mobility ($\bar{r}_{25}$) across CZs. For each characteristic listed on the y-axis, the dot represents the absolute value of the unweighted correlation of the variable with $\bar{r}_{25}$ across CZs. The horizontal bars show the 95% confidence interval based on standard errors clustered at state level. Positive correlations are shown by (+) on the y axis; negative correlations are shown by (–). We consider covariates in 10 broad categories: racial demographics, segregation, properties of the income distribution, K–12 education, social capital, family structure, local tax policies, college education, labor market conditions, and migration rates. The categories with the highest correlations are highlighted. See column (1) of Appendix Table VII for the point estimates corresponding to the correlations plotted here. See Section VI, Online Data Table IX, and Online Appendix G for definitions of each of the correlates. CZ-level data on the covariates used in this figure are reported in Online Data Table VIII.

associated with absolute upward mobility are generally positively associated with relative mobility (i.e., are negatively correlated with rank-rank slopes).

In the remainder of this section, we discuss correlations of mobility with the categories in Figure VIII that have the strongest relationship with mobility: racial demographics, segregation, income inequality, school quality, social capital, and family structure. We discuss results for four other broad categories for which we find weaker correlations – local tax policies, higher education, labor market conditions, and migration – in Online Appendix H.

VI.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. Indeed, the unweighted correlation between upward mobility and the fraction of black residents in the CZ (based on the 2000 census) is −0.580, as shown in the first row of Figure VIII.

This correlation 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, we 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 each individual's race in our data. As an alternative, we predict race based on the parent's five-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 IX Panel A replicates the map of absolute upward mobility ($\bar{r}_{25,c}$) by CZ, restricting the sample to ZIP codes within each CZ in which at least 80% of the residents are non-Hispanic whites.[44] In this subsample, 91% of individuals are white. The spatial pattern in Figure IX Panel A is very similar to that in the original map for the full sample in Figure VI Panel A. 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.

44. We continue to estimate $\bar{r}_{25,c}$ at the CZ level in this map, but we only include ZIP-5s 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-5s 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).

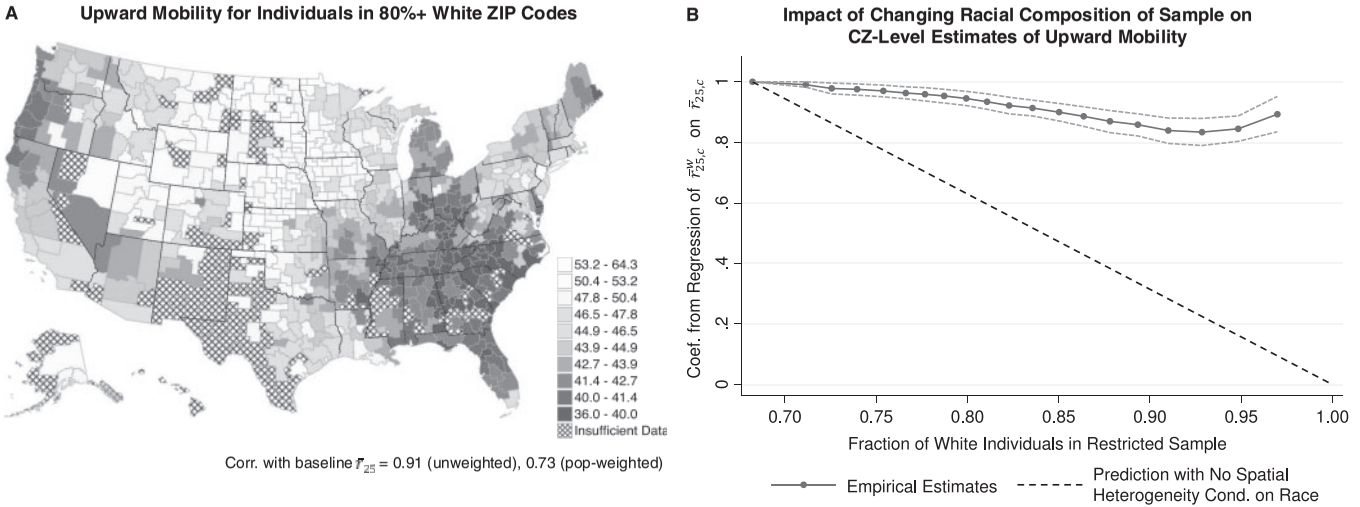

FIGURE IX

Race and Upward Mobility

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

In Figure IX Panel B, 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 $\bar{r}^{w}_{25,c}$. We then regress $\bar{r}^{w}_{25,c}$ on $\bar{r}_{25,c}$, our baseline estimates of upward mobility based on the full sample, using an unweighted OLS regression with one observation per CZ with available data. We vary w from 0% to 95% in increments of 5% and plot the resulting regression coefficients in Figure IX Panel B against the fraction of white individuals in each of the subsamples. When $w = 0$, the regression coefficient is 1 by construction because $\bar{r}_{25,c} = \bar{r}^{w=0}_{25,c}$. Since 68 percent of the U.S. population is white, the first point on the figure is (0.68, 1). The point generated by the $w = 80\%$ threshold is (0.91, 0.84), consistent with the map in Figure IX Panel A. 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 IX Panel B would fall to 0 as the fraction white in the sample converged to 1, as illustrated by the dashed line. Intuitively, if all of the spatial variation in Figure VI Panel A 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 percent white individuals, the regression coefficient remains at 0.89.

The main lesson of this analysis is that both blacks and whites living in areas with large African American populations have lower rates of upward income mobility.[45] There are many potential mechanisms for such a correlation, including differences in the institutions and industries that developed in areas with large African American populations. We are unable to distinguish between these mechanisms in our data; instead, we next turn to one such mechanism that has received the greatest attention in prior work: segregation. The United States has a historical legacy of greater segregation in areas with more blacks. Such

45. To be clear, this result does not imply that race does not matter for children's outcomes at the individual level, as shown, for example, by Mazumder (2011). Our finding is simply that there is spatial heterogeneity in upward mobility even conditional on race.

segregation could potentially affect both low-income whites and blacks, as racial segregation is often associated with income segregation.

VI.B. Segregation

Prior work has argued that segregation has harmful effects on disadvantaged individuals 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). In this section, we evaluate these hypotheses by exploring the correlation between intergenerational mobility and various measures of segregation (shown in the second panel of Figure VIII and Online Appendix Table VIII).

We begin by measuring racial segregation using a Theil (1972) index, constructed using data from the 2000 census as in Iceland (2004). Let ϕ_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: $E = \sum \phi_r \log_2 \frac{1}{\phi_r}$, with $\phi_r \log_2 \frac{1}{\phi_r} = 0$ when $\phi_r = 0$. Letting $j = 1, \ldots, N$ index census tracts in the CZ, we analogously measure racial diversity within each tract as $E_j = \sum \phi_{rj} \log_2 \frac{1}{\phi_{rj}}$ where ϕ_{rj} denotes the fraction of individuals of race r in tract j. We define the degree of racial segregation in the CZ as

$$H = \sum_j \left[\frac{\text{pop}_j}{\text{pop}_{\text{total}}} \frac{E - E_j}{E} \right], \tag{4}$$

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 all tracts have racial composition identical to the CZ as a whole, so that $E_j = E$.

Column (1) of Table IV reports the coefficient estimate from an unweighted OLS regression of absolute upward mobility $\bar{r}_{25,c}$ on the racial segregation index, with one observation per CZ. In

TABLE IV

SEGREGATION AND INTERGENERATIONAL MOBILITY

Dep. var.:	Absolute upward mobility						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Racial segregation	−0.361	−0.360					
	(0.045)	(0.068)					
Income segregation			−0.393				−0.058
			(0.065)				(0.090)
Segregation of poverty ($<p25$)				−0.508	−0.408		
				(0.155)	(0.166)		
Segregation of affluence ($>p75$)				0.108	0.216		
				(0.140)	(0.171)		
Share with commute < 15 mins						0.605	0.571
						(0.126)	(0.165)
Urban areas only		x			x		
R-squared	0.131	0.130	0.154	0.167	0.052	0.366	0.368
Observations	709	325	709	709	325	709	709

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

this and all subsequent regressions, 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.[46] Column (2) shows that the correlation remains at −0.360 in urban areas, that is, CZs that overlap with MSAs.

46. Online Appendix Figure X Panel A presents a nonparametric binned scatter plot corresponding to this regression; see Online Appendix H for details on the construction of this figure.

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 equation (4). Building on this measure, Reardon (2011) defines the overall level of income segregation in a given CZ as

$$\text{income segregation} = 2log(2)\int_{\text{p}} E(p)H(p)dp. \tag{5}$$

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 equation (5) 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 IV, we regress absolute upward mobility on the income segregation index; see Online Appendix Figure X Panel B for the corresponding nonparametric binned scatter plot. The correlation between income segregation and upward mobility is -0.393, consistent with the findings of Graham and Sharkey (2013) using survey data. 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 0.264 (std. err. 0.082). 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 $H(p = 25)$, that is, 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 $H(p = 75)$. Column (4) regresses upward mobility on both segregation of poverty and affluence. Segregation of poverty has a strong negative

association with upward mobility, whereas segregation of affluence does not. Column (5) shows that the same pattern holds when restricting the sample to urban areas. These results suggest that the isolation of low-income families (rather than the isolation of the rich) may be most detrimental for low-income children's prospects of moving up in the income distribution.

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 0.605. Column (7) shows that commute times remain a significant predictor of upward mobility in a multivariable regression but income segregation does not.

These results are consistent with the view that the negative effects of segregation may operate by making it more difficult to reach jobs or other resources that facilitate upward mobility. But any such spatial mismatch explanation must explain why the gradients emerge before children enter the labor market, as shown in Section V.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.

VI.C. Income Levels and Inequality

In this subsection, we explore the correlation between properties of the local income distribution—mean income levels and inequality—and intergenerational mobility.

1. Mean Income Levels. The third section of Figure VIII shows that the mean level of household income in a CZ (as measured in the 2000 census) is essentially uncorrelated with upward mobility (see Online Appendix Figure XI Panel A for the corresponding nonparametric binned scatter plot). Children in low-income

families who grow up in the highest-income CZs (with mean incomes of $47,600 a 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).

2. Income Inequality. Prior work has documented a negative correlation between income inequality and intergenerational mobility across countries (Corak 2013). This "Great Gatsby" curve (Krueger 2012) has attracted attention because it suggests that greater inequality within a generation could reduce social mobility. We explore whether there is an analogous relationship across areas within the United States by correlating upward mobility with the Gini coefficient of parent income within each CZ. We compute the Gini coefficient for parents in our core sample within each CZ as $\text{Gini} = \frac{2}{\overline{X}_c} Cov(X_{ic}, P_{ic})$, where $\overline{X}_c$ is the mean family income (from 1996 to 2000) of parents in CZ c and $Cov(X_{ic}, P_{ic})$ is the covariance between the income level (X_{ic}) and the percentile rank (P_{ic}) of parents in CZ c. The correlation between the Gini coefficient and upward mobility is −0.578 (see also Online Appendix Figure XI Panel B).

An alternative measure of inequality is the portion of income within a CZ that accrues to the richest households, for example, those in the top 1%. This measure is of particular interest because the rise in inequality in the United States over the past three decades was driven primarily by an increase in top income shares (Piketty and Saez 2003). We calculate top 1 percent income shares using the distribution of parent family income within each CZ. The correlation between upward mobility and the top 1 percent income share is only −0.190 (see also Online Appendix Figure XI Panel C), much weaker than that with the Gini coefficient.

We investigate why the Gini coefficient and top 1% share produce different results in Table V, which is constructed in the same way as Table IV. Column (1) replicates the regression corresponding to the raw correlation between the Gini coefficient and upward mobility 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 among households in the bottom 99%. Column (2) of Table V

TABLE V

INCOME INEQUALITY AND INTERGENERATIONAL MOBILITY: THE "GREAT GATSBY" CURVE

	Across CZs within the United States					Across countries		
Dep. var.:	Absolute upward mobility				Relative mobility	Log-log elasticity 1985 inequality		Log-log elasticity 2005 inequality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gini coefficient	−0.578							
	(0.093)							
Gini bottom 99%		−0.634	−0.624		0.476	0.72	0.62	0.78
		(0.090)	(0.113)		(0.088)	(0.21)	(0.27)	(0.27)
Top 1% income share		−0.123	0.029		−0.032		0.17	−0.11
		(0.035)	(0.039)		(0.032)		(0.27)	(0.28)
Frac. between $p25$ and $p75$				0.679				
				(0.111)				
Urban areas only			x					
R-squared	0.334	0.433	0.380	0.462	0.224	0.518	0.536	0.531
Observations	709	709	325	709	709	13	13	12

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

shows that a 1 standard deviation increase in the bottom 99% Gini is associated with a 0.634 standard deviation reduction in upward mobility. In contrast, a 1 standard deviation increase in the top 1% share is associated with only a 0.123 standard deviation reduction in upward mobility. Column (3) shows that in urban areas (CZs that overlap with MSAs), the pattern 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.

Another 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 V shows that upward mobility is strongly positively correlated with the size of the middle class.

Finally, column (5) of Table V replicates column (2) using relative mobility β_c as the dependent variable. The bottom 99% Gini coefficient is strongly positively associated with this measure. That is, greater inequality in the bottom 99% is negatively related to relative mobility. [47] But once again, the top 1% share is uncorrelated with relative mobility.

3. Comparison to Cross-Country Evidence. Next, we explore whether the size of the middle class is more strongly correlated with intergenerational mobility than upper tail inequality in the cross-country data as well. In column (6) of Table V, 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 using data from 13 developed countries compiled by Corak (2013).[48] 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–United States 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 inequality among the bottom 99% and not the extreme 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, whereas segregation of poverty is negatively associated with mobility.

47. 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.

48. We obtain estimates of the Gini coefficient by country from the OECD Income Distribution Database. We interpret these estimates as applying to the bottom 99% because surveys typically do not capture the thickness of the top tail due to top-coding.

VI.D. School Quality

In the fourth panel of Figure VIII, we study the correlation between mobility and various proxies for school quality. We first consider two proxies for inputs into school quality: mean public school expenditures per student and mean class sizes based on data from the National Center for Education Statistics (NCES) for the 1996–1997 school year. We find a positive correlation between public school expenditures and upward mobility, but the correlation is not as strong or robust as with measures of inequality or segregation. There is a strong negative correlation between class size and upward mobility (columns (1) and (2) of Online Appendix Table VIII) when pooling all CZs. 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 may capture relatively little of the variation in school quality (Hanushek 2003). To address this problem, we construct 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. The Global Report Card converts school district–level scores on statewide tests to a single national scale by benchmarking statewide test scores to scores on the National Assessment of Educational Progress (NAEP) tests. 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 percent of school districts have nonmissing data. We regress test scores on mean parent family income (from 1996 to 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.

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 VIII. In the baseline specification, 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.

VI.E. 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 (Coleman 1988; Borjas 1992; Putnam 1995). We explore the relationship between mobility and measures of social capital used in prior work in the fifth panel of Figure VIII.

Our primary proxy for social capital is the social capital index constructed by Rupasingha and Goetz (2008), 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. The correlation between upward mobility and social capital is 0.641 in our baseline specification, an estimate that is quite robust across alternative specifications. 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.

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.

VI.F. Family Structure

Many have argued that family stability plays a key role in children's outcomes (see Becker 1991; Murray 1984, 2012). To evaluate this hypothesis, we use three measures of family structure in the CZ based on data from the 2000 census: (i) the fraction of children living in single-parent households, (ii) the fraction of adults who are divorced, and (iii) the fraction of adults who are married. All three of these measures are very highly correlated with upward mobility, as shown in the sixth panel of Figure VIII.

The fraction of children living in single-parent households is the single strongest correlate of upward income mobility among all the variables we explored, with a raw unweighted correlation of -0.76 (see Online Appendix Figure XII Panel A for the corresponding nonparametric 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 (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 XII Panel B). 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 VI.A on the community-level effects of racial shares.

VI.G. Comparison of Alternative Explanations

In Table VI, we assess which of the five factors identified above—segregation, inequality, school quality, social capital, and family structure—are the strongest predictors of upward mobility in multivariable regressions that control for race and other covariates. Based on the analysis, we first identify the proxy that has the strongest and most robust univariate correlation with upward mobility in each category: the fraction of working individuals who commute less than 15 minutes to work (segregation), the bottom 99% Gini coefficient (inequality), high school dropout rates adjusted for income differences (school quality), the social capital index, and the fraction of children with single parents (family structure).[49] As in preceding regression specifications, we normalize all the dependent and independent variables to have a standard deviation of 1 in the estimation sample for each regression in Table VI.[50]

We begin in column (1) with an unweighted OLS regression of absolute upward mobility $\bar{r}_{25,c}$ 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

49. We obtain similar results if we combine the various proxies into a single index for each factor using weights from an OLS regression of absolute upward mobility on the proxies within each category.

50. 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 nonmissing data.

TABLE VI

CORRELATES OF INTERGENERATIONAL MOBILITY: COMPARING ALTERNATIVE HYPOTHESES

Dep. var.:	Absolute upward mobility			Relative mobility		Absolute upward mobility		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction short commute	0.302	0.227	0.314	−0.290	−0.325	0.331		0.319
	(0.065)	(0.077)	(0.052)	(0.061)	(0.064)	(0.070)		(0.065)
Gini bottom 99%	−0.009	−0.017	0.060	0.006	0.343	−0.287		−0.021
	(0.053)	(0.043)	(0.097)	(0.071)	(0.095)	(0.059)		(0.054)
High school dropout rate	−0.147	−0.120	−0.109	0.010	0.181	−0.288		−0.140
	(0.055)	(0.038)	(0.085)	(0.064)	(0.056)	(0.059)		(0.055)
Social capital index	0.169	0.065	0.173	0.154	0.154	0.168		0.168
	(0.047)	(0.050)	(0.060)	(0.060)	(0.070)	(0.059)		(0.045)
Fraction single mothers	−0.487	−0.477	−0.555	0.591			−0.808	−0.579
	(0.062)	(0.071)	(0.089)	(0.049)			(0.085)	(0.061)
Fraction black							0.056	0.132
							(0.073)	(0.051)
State fixed effects		x						
Urban areas only			x					
R-squared	0.757	0.859	0.671	0.48	0.324	0.651	0.584	0.763
Observations	709	709	325	709	709	709	709	709

Notes. Each column reports coefficients from an OLS regression with standard errors clustered at the state level reported in parentheses. The regressions are run using data for the 709 CZs with at least 250 children in the core sample. The dependent variable in columns (1)–(3) and (6)–(8) is our baseline measure of absolute upward mobility, the expected rank of children whose parents are at the 25th national percentile. The dependent variable in columns (4) and (5) is relative mobility, the rank-rank slope within each CZ. All independent and dependent variables are normalized (in the relevant estimation sample) to have mean 0 and standard deviation 1. Column (1) reports unweighted estimates across all CZs. Column (2) includes state fixed effects. In column (3), we restrict to CZs that intersect MSAs. Columns (4)–(8) replicate the unweighted specification in column (1) with different dependent and independent variables. The fraction with short commutes is the share of workers that commute to work in less than 15 minutes calculated using data for the 2000 census. Gini bottom 99% is the Gini coefficient minus the top 1% income share within each CZ, computed using the distribution of parent family income within each CZ for parents in the core sample. Income-residualized high school dropout rate is the residual from a regression of the fraction of children who drop out of high school in the CZ, estimated using data from the NCES Common Core of Data for the 2000–2001 school year, on mean household income in 2000. Social capital index is the standardized index of social capital constructed by Rupasingha and Goetz (2008). Fraction single mothers is the fraction of children being raised by single mothers in each CZ. Fraction black is the number of people in the CZ who are black alone divided by the CZ population. We code the high school dropout rate as 0 for 116 CZs in which dropout rate data are missing for more than 25% of the districts in the CZ, and include an indicator for having a missing high school dropout rate. We do the same for 16 CZs with missing data on social capital. See Section VI, Online Data Table IX, and Online Appendix G for additional details on the definitions of each of these variables.

specification. Together, the five factors explain 76 percent of the variance in upward mobility across areas. Column (2) shows that the coefficients remain similar when state fixed effects are included. Column (3) shows that the estimates are roughly similar when restricting the sample to urban areas (CZs that intersect MSAs). Across all the specifications, the strongest and most robust predictor is the fraction of children with single parents.

In column (4), we use relative mobility β_c as the dependent variable instead of absolute upward mobility. The fraction of

single parents and commute times 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 (5) we replicate column (4) 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 (6) replicates the specification in column (5) 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.

In the last two columns of Table VI, we explore the role of racial demographics versus the other explanatory factors. Column (7) 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 significantly correlated with upward mobility. Column (8) shows that the correlation of upward mobility with black shares is slightly positive and statistically significant when we include controls for all five explanatory factors. These results support the view that the strong correlation of upward mobility with race operates through channels beyond the direct effect of race on mobility.

Overall, the results in Table VI 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, the regression may place greater weight on factors that are measured with less error 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.

VII. Conclusion

This article has used population data to present a new portrait of intergenerational income mobility in the United States. Intergenerational mobility varies substantially across areas.

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. The spatial variation in intergenerational mobility is strongly correlated with five factors: residential segregation, income inequality, school quality, social capital, and family structure.

In this article, we have presented a cross-sectional snapshot of intergenerational mobility for a single set of birth cohorts. In a companion paper (Chetty et al. 2014), we study trends in mobility over time. We find that the level of intergenerational mobility (national rank-rank slope) has remained stable for the 1971–1993 birth cohorts in the United States, especially in comparison to the degree of variation across areas. A natural question given the results of the two papers is whether the cross-sectional correlations documented here are consistent with the time trends in mobility. To answer this question, we predict the trend in the rank-rank slope implied by changes in the five key correlates over time (see Online Appendix I and Appendix Figure XIII). The predicted changes are quite small because the factors move in opposing directions. For example, the increase in inequality and single parenthood rates in recent decades predict a small decline in mobility in recent decades. In contrast, the decline in racial segregation and high school dropout rates predict an increase in mobility of similar magnitude. Overall, the cross-sectional correlations documented here are consistent with the lack of a substantial time trend in mobility in recent decades.

The main lesson of our analysis is that intergenerational mobility is a local problem, one that could potentially be tackled using place-based policies (Kline and Moretti 2014). Going forward, a key question is why some areas of the United States generate higher rates of mobility than others. We hope that future research will be able to shed light on this question by using the mobility statistics constructed here (available online at the county by birth cohort level at www.equality-of-opportunity.org) to study the effects of local policy changes.

HARVARD UNIVERSITY AND NBER
HARVARD UNIVERSITY AND NBER
UNIVERSITY OF CALIFORNIA, BERKELEY, AND NBER
UNIVERSITY OF CALIFORNIA, BERKELEY, AND NBER

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournal.org).

REFERENCES

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

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

Bhattacharya, Debopam, and Bhashkar Mazumder, "A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States," *Quantitative Economics*, 2 (2011), 335–379.

Bjorklund, Anders, and Markus Jäntti, "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review*, 87 (1997), 1009–1018.

Black, Dan, Seth Sanders, Evan Taylor, and Lowell Taylor, "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South," Mimeo, University of Chicago, 2011.

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

Borjas, George J., "Ethnic Capital and Intergenerational Mobility," *Quarterly Journal of Economics*, 107 (1992), 123–150.

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

Cameron, Stephen V., and James J. Heckman, "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109 (2001), 455–499.

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

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104 (2014), 2633–2679.

Chetty, Raj, and Nathan Hendren, "The Value-Added of Neighborhoods: Quasi-Experimental Estimates of Neighborhood Effects on Children's Long-Term Outcomes," Mimeo, Harvard University, 2014.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review Papers and Proceedings*, 104, no. 5 (2014), 141–147.

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

Clark, Gregory, "The Son Also Rises: Surnames and the History of Social Mobility," *The Princeton Economic History of the Western World* (Princeton, NJ: Princeton University Press, 2014).

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

Corak, Miles, "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, 27, no. 3 (2013), 79–102.

Corak, Miles, and Andrew Heisz, "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data," *Journal of Human Resources*, 34, no. 3 (1999), 504–533.

Cutler, David M., and Edward L. Glaeser, "Are Ghettos Good or Bad?," *Quarterly Journal of Economics*, 112 (1997), 827–872.

Dahl, Molly W., and Thomas DeLeire, *The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative*

Data, Institute for Research on Poverty, University of Wisconsin–Madison, 2008.
Dorn, David, "Essays on Inequality, Spatial Interaction, and the Demand for Skills," PhD dissertation no. 3613, University of St. Gallen, 2009.
Duncan, Greg J., Kathleen M. Ziol-Guest, and Ariel Kalil, "Early-Childhood Poverty and Adult Attainment, Behavior, and Health," *Child Development*, 81 (2010), 306–325.
Fields, Gary S., and Efe A. Ok, "The Measurement of Income Mobility: An Introduction to the Literature," in *Handbook of Income Inequality Measurement*, Vol. 71, Jacques Silber, ed. (Dordrecht: Springer Netherlands, 1999), pp. 557–598.
Graham, Bryan, and Patrick Sharkey, "Mobility and the Metropolis: How Communities Factor Into Economic Mobility," Economic Mobility Project, Pew Charitable Trusts, Washington, DC, 2013.
Grawe, Nathan D., "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence," *Labour Economics*, 13, no. 5 (2006), 551–570.
Haider, Steven, and Gary Solon, "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96 (2006), 1308–1320.
Hanushek, Eric A., "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113, no. 485 (2003), F64–F98.
Heckman, James J., "Skill Formation and the Economics of Investing in Disadvantaged Children," *Science*, 312, no. 5782 (2006), 1900–1902.
Hertz, Tom, "Understanding Mobility in America," Center for American Progress Discussion Paper, 2006.
Iceland, John, "Beyond Black and White: Metropolitan Residential Segregation in Multi-ethnic America," *Social Science Research*, 33 (2004), 248–271.
Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Eva Österbacka, Anders Björklund, and Tor Eriksson, "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, 2006.
Jencks, Christopher, and Susan E. Mayer, "The Social Consequences of Growing Up in a Poor Neighborhood," in *Inner City Poverty in the United States*, L. Lynn, and M. G. H. McGeary, eds. (Washington, DC: National Academies Press, 1990), 111–186.
Kain, John F., "Housing Segregation, Negro Employment, and Metropolitan Decentralization," *Quarterly Journal of Economics*, 82 (1968), 175–197.
Kasarda, John D., "Urban Industrial Transition and the Underclass," *Annals of the American Academy of Political and Social Science*, 501 (1989), 26–47.
Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to Opportunity In Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 116 (2001), 607–654.
Kline, Patrick, and Enrico Moretti, "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, 6 (2014), 629–662.
Krueger, Alan, "The Rise and Consequences of Inequality in the United States," Speech at the Center for American Progress, Washington, DC, January 12, 2012.
Lamb, Michael E. *The Role of the Father in Child Development* (Hoboken, NJ: Wiley, 2004).
Lee, Chul-In, and Gary Solon, "Trends in Intergenerational Income Mobility," *Review of Economics and Statistics*, 91 (2009), 766–772.
Massey, Douglas S., and Nancy A. Denton *American Apartheid: Segregation and the Making of the Underclass* (Cambridge, MA: Harvard University Press, 1993).
Mazumder, Bhashkar, "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data," *Review of Economics and Statistics*, 87, no. 2 (2005), 235–255.

———, "Black-White Differences in Intergenerational Economic Mobility in the US," Working Paper Series WP-2011-10, Federal Reserve Bank of Chicago, 2011.
Mitnik, Pablo, Victoria Bryant, David B. Grusky, and Michael Weber, "New Estimates of Intergnerational Income Mobility Using Administrative Data," Mimeo, Statistics of Income, Internal Revenue Service, 2014.
Murray, Charles A. *Losing Ground: American Social Policy, 1950–1980* (New York: Basic Books, 1984).
———, *Coming Apart: The State of White America, 1960–2010* (New York: Crown Forum, 2012).
Neal, Derek A., and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104 (1996), 869–895.
Oreopoulos, Philip, "The Long-Run Consequences of Living in a Poor Neighborhood," *Quarterly Journal of Economics*, 118 (2003), 1533–1575.
Piketty, Thomas, and Emmanuel Saez, "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118 (2003), 1–41.
Putnam, Robert D., "Bowling Alone: America's Declining Social Capital," *Journal of Democracy*, 6 (1995), 65–78.
Ray, Debraj, "Uneven Growth: A Framework for Research in Development Economics," *Journal of Economic Perspectives*, 24, no. 3 (2010), 45–60.
Reardon, Sean F., "Measures of Income Segregation," CEPA Working Papers, 2011.
Reardon, Sean, and Kendra Bischoff, "Growth in the Residential Segregation of Families by Income, 1970–2009," US 2010 Project, 2011.
Reardon, Sean F., and Glenn Firebaugh, "Measures of Multigroup Segregation," *Sociological Methodology*, 32 (2002), 33–67.
Rupasingha, Anil, and Stephan J. Goetz, "US County-Level Social Capital Data, 1990–2005," Northeast Regional Center for Rural Development, Penn State University, 2008.
Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley, "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research," *Annual Review of Sociology*, 28 (2002), 443–478.
Solon, Gary, "Intergenerational Income Mobility in the United States," *American Economic Review*, 82, no. 3 (1992), 393–408.
———,, "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics,* Vol. 3, O. Ashenfelter, and D. Card, eds. (Amsterdam: Elsevier, 1999), 1761–1800.
———, "Cross-Country Differences in Intergenerational Earnings Mobility," *Journal of Economic Perspectives*, 16, no. 3 (2002), 59–66.
Solon, Gary, Marianne E. Page, and Greg J. Duncan, "Correlations between Neighboring Children in Their Subsequent Educational Attainment," *Review of Economics and Statistics*, 82, no. 3 (2000), 383–392.
Theil, Henri, "Statistical Decomposition Analysis. With Applications in the Social and Administrative Sciences no. v. 14," in *Studies in Mathematical and Managerial Economics* (Amsterdam: North-Holland, 1972).
Thomas, Adam, and Isabel Sawhill, "For Richer or for Poorer: Marriage as an Antipoverty Strategy," *Journal of Policy Analysis and Management*, 21 (2002), 587–599.
Tolbert, Charles M., and Molly Sizer, "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Economic Research Service Staff Paper 9614, 1996.
Trivedi, Pravin K., and David M. Zimmer, "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, 1 (2007), 1–111.
Wilson, William J. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy* (Chicago: University of Chicago Press, 1987).
———, *When Work Disappears: The World of the New Urban Poor* (New York: Knopf, 1996).
Zimmerman, David J., "Regression toward Mediocrity in Economic Stature," *American Economic Review*, 82 (1992), 409–429.

This page intentionally left blank