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

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

June 2014

Abstract

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

*The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Internal Revenue Service or the U.S. Treasury Department. This work is a component of a larger project examining the effects of tax expenditures on the budget deficit and economic activity. All results based on tax data in this paper are constructed using statistics originally reported in the SOI Working Paper "The Economic Impacts of Tax Expenditures: Evidence from Spatial Variation across the U.S.," approved under IRS contract TIRNO-12-P-00374 and presented at the National Tax Association meeting on November 22, 2013. We thank David Autor, Gary Becker, David Card, David Dorn, John Friedman, James Heckman, Nathaniel Hilger, Richard Hornbeck, Lawrence Katz, Sara Lalumia, Adam Looney, 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 www.equality-of-opportunity.org.

I Introduction

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

We characterize intergenerational mobility using information from de-identified federal income tax records, which provide data on the incomes of more than 40 million children and their parents between 1996 and 2012. We organize our analysis into three parts.

In the first part, we present new statistics on intergenerational mobility in the U.S. 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]

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 non-linear 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 percentile 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

[1] We show that our baseline measures do not suffer from significant lifecycle 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.

[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.

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 paper, 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 U.S., including rural areas (Tolbert and Sizer 1996). We assign children to commuting zones based on where they lived at age 16 – i.e., where they grew up – irrespective of whether they left that CZ afterward. When analyzing CZs, we continue to rank both children and parents based on their positions in the *national* income distribution, which allows us to measure children's absolute outcomes as we discuss below.

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

We find substantial variation in both relative and absolute mobility across CZs. Relative mobility is lowest for children who grew up in the Southeast and highest in the Mountain West and the rural Midwest. Some CZs in the U.S. 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.

[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.

We find similar geographical variation in absolute mobility. We focus much of our analysis on absolute mobility at $p = 25$, which we term "absolute upward mobility." This statistic measures the mean income rank of children with parents in the bottom half of the income distribution given linearity of the rank-rank relationship. Absolute upward mobility ranges from 35.8 in Charlotte to 46.2 in Salt Lake City among the 50 largest CZs. A 1 standard deviation (SD) increase in CZ-level upward mobility is associated with a 0.2 SD improvement in a child's expected rank given parents at $p = 25$, 60% as large as the effect of a 1 SD increase in 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 birth rates 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 paper, 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 intergenera-

tional mobility both across CZs within the U.S. 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 indices (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 e.g. 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 U.S. 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 importantly, 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).