

NBER WORKING PAPER SERIES

IS THE UNITED STATES STILL A LAND OF OPPORTUNITY? RECENT TRENDS
IN INTERGENERATIONAL MOBILITY

Raj Chetty
Nathaniel Hendren
Patrick Kline
Emmanuel Saez
Nicholas Turner

Working Paper 19844
<http://www.nber.org/papers/w19844>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2014

The views expressed in this paper are those of the authors and do not necessarily represent the views or policies of the US Treasury Department or the Internal Revenue Service or the National Bureau of Economic Research. We thank Sarah Abraham, Alex Bell, Alex Olssen, and Evan Storms for outstanding research assistance. We thank David Autor, Greg Duncan, Lawrence Katz, Alan Krueger, Richard Murnane, Gary Solon, and numerous seminar participants for helpful discussions and comments. Financial support from the Lab for Economic Applications and Policy at Harvard, the Center for Equitable Growth at UC Berkeley, and the National Science Foundation is gratefully acknowledged. The statistics reported in this paper can be downloaded from www.equality-of-opportunity.org

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility
Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner
NBER Working Paper No. 19844
January 2014, Revised May 2014
JEL No. H0,J0

ABSTRACT

We present new evidence on trends in intergenerational mobility in the U.S. using administrative earnings records. We find that percentile rank-based measures of intergenerational mobility have remained extremely stable for the 1971-1993 birth cohorts. For children born between 1971 and 1986, we measure intergenerational mobility based on the correlation between parent and child income percentile ranks. For more recent cohorts, we measure mobility as the correlation between a child's probability of attending college and her parents' income rank. We also calculate transition probabilities, such as a child's chances of reaching the top quintile of the income distribution starting from the bottom quintile. Based on all of these measures, we find that children entering the labor market today have the same chances of moving up in the income distribution (relative to their parents) as children born in the 1970s. However, because inequality has risen, the consequences of the "birth lottery" – the parents to whom a child is born – are larger today than in the past.

Raj Chetty
Department of Economics
Harvard University
1805 Cambridge St.
Cambridge, MA 02138
and NBER
chetty@fas.harvard.edu

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

Nathaniel Hendren
Harvard University
Department of Economics
Littauer Center Room 235
Cambridge, MA 02138
and NBER
nhendren@gmail.com

Nicholas Turner
Office of Tax Analysis
U.S. Department of the Treasury
1500 Pennsylvania Avenue, NW
Washington, D.C. 20220
Nicholas.Turner@treasury.gov

Patrick Kline
Department of Economics
UC, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720
and NBER
pkline@econ.berkeley.edu

There is a growing public perception that intergenerational income mobility – a child’s chance of moving up in the income distribution relative to her parents – is declining in the United States (e.g., Foroohar 2011, Zakaria 2011). However, empirical evidence on trends in intergenerational mobility is mixed. Some studies (e.g., Aaronson and Mazumder 2008, Putnam, Frederick, and Snellman 2012) find that income mobility and related indicators have declined in recent decades. But others find no trend in intergenerational income mobility over a similar time period (e.g., Hertz 2007, Lee and Solon 2009, Hauser 2010).

We present new evidence on trends in intergenerational mobility using data from de-identified tax records, building on work by Auten, Gee, and Turner (2013) and Chetty et al. (2014). These data have less measurement error and much larger sample sizes than prior survey-based studies and thus yield more precise estimates of intergenerational mobility over time.

We estimate intergenerational mobility for the 1971 to 1993 birth cohorts. For children born between 1971 and 1986, we measure mobility by estimating (1) the correlation between parent and child income percentile ranks and (2) the probability that a child reaches the top fifth of the income distribution conditional on her parents’ income quintile. For children born after 1986, we measure mobility as the correlation between parent income ranks and children’s college attendance rates, which are a strong predictor of later earnings.

We find that all of these rank-based measures of intergenerational mobility have not changed significantly over time. For example, the probability that a child reaches the top fifth of the income distribution given parents in the bottom fifth of the income distribution is 8.4% for children born in 1971, compared with 9.0% for those born in 1986. Children born to the highest-income families in 1984 were 74.5 percentage points more likely to attend college than those from the lowest-income families. The corresponding gap for children born in 1993 is 69.2 percentage points, suggesting that if anything intergenerational mobility may have increased slightly in recent cohorts. Moreover, intergenerational mobility is fairly stable over time in each of the nine census divisions of the U.S. even though they have very different levels of mobility.

Although rank-based measures of mobility remained stable, income inequality increased over time in our sample, consistent with prior work. Hence, the consequences of the “birth lottery” – the parents to whom a child is born – are larger today than in the past. A useful visual analogy is to envision the income distribution as a ladder, with each percentile representing a different rung. The rungs of the ladder have grown further apart (inequality has increased), but children’s chances of climbing from lower to higher rungs have not changed (rank-based mobility has remained stable).

This result may be surprising in light of the well known cross-country relationship between inequality and mobility, termed the “Great Gatsby Curve” by Krueger (2012). However, as we discuss in Section IV, much of the increase in inequality has come from the extreme upper tail (e.g., the top 1%) in recent decades, and top 1% income shares are not strongly associated with mobility in the cross-section across countries or metro areas within the U.S. (Chetty et al. 2014).

The paper is organized as follows. The next section presents a simple conceptual framework for measuring trends in intergenerational mobility and inequality. Section II describes the data and Section III presents the empirical results. We conclude in Section IV by discussing the findings in the context of the prior literature.

I. Measuring Intergenerational Mobility: Conceptual Issues

The study of intergenerational mobility amounts to a characterization of the joint distribution of parent and child income. Prior work (reviewed e.g., in Black and Devereux 2011) has used many different statistics to summarize this joint distribution: (1) the correlation between parent and child percentile ranks, (2) quintile transition matrices, and (3) log-log intergenerational elasticities (IGE) of child income with respect to parent income. Since each of these statistics could exhibit different time trends, we begin by formalizing how we measure intergenerational mobility.

We decompose the joint distribution of parent and child income into two components: (1) the joint distribution of parent and child ranks, formally known as the *copula* of the distribution, and (2) the *marginal distributions* of parent and child income. The marginal distributions determine the degree of inequality within each generation, typically measured by Gini coefficients or top income shares. The copula is a key determinant of mobility across generations. The first two measures of mobility described above – rank-rank correlations and quintile transition matrices – depend purely on the copula. The log-log IGE combines features of the marginal distributions and the copula.

We characterize changes in the copula and marginal distributions of income separately to distinguish changes in inequality from intergenerational mobility. We find that the copula has not changed over time: children’s chances of moving up or down in the income distribution have remained stable. However, as is well known from prior work, the marginal distributions of income have changed substantially because of growing inequality.

Together, these two facts can be used to construct various measures of mobility. For example, if one defines mobility based on relative positions in the income distribution – e.g., a child’s prospects of rising from the bottom to the top quintile – then intergenerational mobility has remained

unchanged in recent decades. If instead one defines mobility based on the probability that a child from a low-income family (e.g., the bottom 20%) reaches a fixed upper income threshold (e.g., \$100,000), then mobility has *increased* because of the increase in inequality. However, the increase in inequality has also magnified the difference in expected incomes between children born to low (e.g., bottom-quintile) vs. high (top-quintile) income families. In this sense, mobility has *fallen* because a child's income depends more heavily on her parents' position in the income distribution today than in the past.

The appropriate definition of intergenerational mobility depends upon one's normative objective. By characterizing the copula and marginal distributions separately, we allow readers to focus on the measure of mobility relevant for their objectives.

II. Data

Our data and methods build closely on our companion paper (Chetty et al. 2014, henceforth CHKS), which contains complete details on the samples and variables used below. We present a brief summary of the sample and variable definitions here as a reference.

Sample Construction. For children born during or after 1980, we construct a linked parent-child sample using population tax records spanning 1996-2012. This *population-based* sample consists of all individuals born between 1980-1993 who are U.S. citizens as of 2013 and are claimed as a dependent on a tax return filed in or after 1996. We link approximately 95% of children in each birth cohort to parents based on dependent claiming, obtaining a sample with 3.7 million children per cohort (Appendix Table 1, Column 4).

The population tax records cannot be used to link children to parents for birth cohorts prior to 1980 because they are only available starting in 1996, and our ability to link children to parents deteriorates after children turn 16 because they begin to leave home. To obtain data on earlier birth cohorts, we use the Statistics of Income (SOI) annual cross-sections. These cross-sections are stratified random samples covering approximately 0.1% of tax returns. Starting in 1987, the SOI cross-sections contain dependent information, allowing us to link children to parents.

Using the SOI cross-sections, we construct a sample of children in the 1971-82 birth cohorts, which we refer to as the *SOI sample*, as follows. We first identify all children between the ages of 12

and 16 claimed as dependents in the 1987-98 SOI cross-sections.¹ We then pool all the SOI cross-sections that give us information for a given birth cohort. For example, the 1971 cohort is comprised of children claimed at age 16 in 1987, while the 1982 cohort is comprised of children claimed at ages 12-16 in 1994-98.

The SOI sample grows from 4,331 children in 1971 to 9,936 children in 1982 (Appendix Table 1, Column 1) because we have more cross-sections to link parents to children in more recent cohorts and because the size of the SOI cross-sections has increased over time. Using the sampling weights, we estimate that the SOI sample represents 88% of children in each birth cohort (based on vital statistics counts), with slightly lower coverage rates in the early cohorts because children are less likely to be claimed as dependents as they approach age 18 and because tax credits for claiming dependents have grown over time (Appendix Table 1, Column 3).

The SOI sample is designed to be representative of the population of children claimed on tax returns between the ages of 12 and 16 in each birth cohort.² Indeed, we confirm in Appendix Table 2 that summary statistics for the SOI sample (using sampling weights) and the population-based sample are very similar for the overlapping 1980-82 birth cohorts.

Variable Definitions. We define parent family income (in real 2012 dollars) as adjusted gross income plus tax exempt interest and the non-taxable portion of social security benefits for those who file tax returns. For non-filers, we define income as the sum of wage earnings (form W-2), unemployment benefits (form 1099-G), and social security and disability benefits (form SSA-1099). In years where parents have no tax return and no information returns, family income is coded as zero.

In the population-based sample, we define parent income as mean family income over the five years when the child is 15-19 years old.³ In the SOI cross-sections, parent income is observed only in the year that the child is linked to the parent, and therefore we define parent income as family income in that year. In both the population and SOI samples, we drop observations with zero or negative parent income.

¹ We do not limit the SOI sample to current citizens because citizenship data are not fully populated for birth cohorts prior to 1980. The citizenship restriction has a minor impact on the characteristics of the sample (Appendix Table 2) because most children claimed as dependents between ages 12-16 are U.S. citizens as adults.

² Children whose parents are sampled in multiple SOI cross-sections appear multiple times in these data. There are 89,345 children in the SOI sample and 189,541 total observations. To ensure that the stratified sampling in the SOI cross sections does not bias our results, we verify that the results are very similar in the SOI Continuous Wage History subsample, a pure (unstratified) random panel that contains 10,360 children (not reported).

³ Since the data start in 1996, we use the mean from 1996-2000 (ages 16-20) for the 1980 cohort.

We define child family income in the same way as parent income, always using data from the population files. Results are similar if we use individual rather than family income measures for children (not reported).

Finally, we define college attendance at age 19 as an indicator for having a 1098-T form in the calendar year the child turns 19. Because 1098-T forms are filed directly by colleges, we have records on college attendance for all children.⁴

III. Results

Rank-Rank Specification. We begin by measuring intergenerational mobility using a *rank-rank* specification, which provides a more robust summary of intergenerational mobility than traditional log-log specifications (CHKS). We rank each child relative to others in her birth cohort based on her mean family income at ages 29-30. Similarly, we rank parents relative to other parents of children in the same birth cohort based on their family incomes.⁵

Figure 1 plots the average income rank of children (at ages 29-30) vs. parent income rank for three sets of birth cohorts in the SOI sample: 1971-74, 1975-78, and 1979-82. To reduce noise, we divide parent income ranks into 50 (rather than 100) bins and plot the mean child rank vs. the mean parent rank within each bin. The rank-rank relationship is almost perfectly linear. Its slope can be interpreted as the difference in the mean percentile rank of children from the richest families vs. children from the poorest families. The rank-rank slopes for the three sets of cohorts in Figure 1 (estimated using OLS on the binned data) are all approximately 0.30, with standard errors less than 0.01.

When interpreting the intergenerational mobility estimates in Figure 1, one must consider two potential biases that have been emphasized in prior work: lifecycle bias due to measuring income at early or late ages and attenuation bias due to noise in annual measures of income (Black and Devereux 2011). In Section III.B of CHKS, we present a detailed assessment of whether rank-rank estimates analogous to those in Figure 1 exhibit such biases. We reproduce the key lessons from that analysis in Appendix Figures 1-3, which establish three results. First, estimates of the rank-rank slope

⁴ Approximately 6% of 1098-T forms are missing from 1999-2002 because the database contains no 1098-T forms for some small colleges in those years. This creates a small jump in the college-income gradient of approximately 3 percentage points (relative to a mean of 75 percent) from 2002 to 2003. For simplicity, we use only post-2003 college attendance data here.

⁵ In the SOI sample, we always define parent and child ranks within each birth cohort and SOI cross-section year. We use sampling weights when constructing the percentiles so that they correspond to positions in the population.

stabilize fully once children reach age 30 and are within 20% of the age 30 estimate at age 26, indicating that one can obtain considerable information about mobility by age 26 (Appendix Figure 1). Second, estimates of the rank-rank slope are insensitive to the age of parents and children at which parent income is measured, provided that parents are between 30 and 55 (Appendix Figure 2). Third, using several years of data to measure parent and child income (as we do in the population-based sample) instead of one year (as we do in the SOI sample) does not increase the rank-rank slope appreciably, perhaps because transitory measurement error is less prevalent in tax records than survey data (Appendix Figure 3). These results indicate that the income definitions used in Figure 1 and in what follows do not suffer from significant lifecycle or attenuation bias.

Trends in Income Mobility. Figure 2 presents our primary estimates of intergenerational mobility by birth cohort (see Appendix Table 1 for the data plotted in this figure). The series in solid circles plots estimates of the rank-rank slope for the 1971-1982 birth cohorts using the SOI sample. Each estimate is based on an OLS regression of child rank on parent rank for the relevant cohort, weighted using inverse sampling probabilities. Consistent with Figure 1, there is no trend in these rank-rank slopes. We also find that log-log IGE estimates are stable or, if anything, falling slightly over time (Appendix Table 1).⁶

We cannot measure children's income at age 30 beyond the 1982 birth cohort because our data end in 2012. To characterize mobility for younger cohorts, we repeat the preceding analysis using income measures at age 26. The series in squares in Figure 2 plots the rank-rank slope based on child income at age 26 for the 1980-86 birth cohorts in the population-based sample. Once again, there is no trend in this series. Moreover, there is much less fluctuation across cohorts because the estimates are more precise in the population data.

Importantly, CHKS show that intergenerational mobility estimates based on income at age 26 and age 30 are highly correlated across areas within the U.S. Hence, even though the level of the rank-rank slopes at age 26 is slightly lower than the estimates at age 30, we expect trends in mobility based on income at age 26 to provide a reliable prediction of trends in mobility at age 30.

Trends in College Gradients. We cannot use income to assess mobility for children born after 1986 because many of these individuals are still completing their education or just entering the labor

⁶ The log-log IGE is stable because, as we show below, the marginal distributions of parent and child incomes have expanded at roughly similar rates. Formally, if parent and child incomes have a Bivariate Lognormal distribution and the standard deviations of parent and child log income increase by the same percentage over time, stability of the rank-rank slope implies stability of the log-log IGE.

market. We therefore use college attendance to measure intergenerational mobility for these recent birth cohorts. CHKS demonstrate that the correlation between college attendance rates and parent income is a strong predictor of differences in intergenerational income mobility across areas within the U.S. The fact that the college attendance is a good proxy for income mobility is intuitive given the strong association between higher education and subsequent earnings.

The relationship between college attendance rates and parent income ranks is approximately linear (Appendix Figure 4). We therefore summarize the association between parent income and college attendance by regressing an indicator for being enrolled in college at age 19 on parent income rank. The coefficient in this regression, which we term the college attendance gradient, can be interpreted as the gap in college attendance rates between children from the lowest- and highest-income families. The series in triangles in Figure 2 plots the college attendance gradient for the 1984-93 birth cohorts. The gap in college attendance rates between children from the lowest- and highest-income families is essentially constant at 74.5% between the 1984-89 birth cohorts. The gap falls slightly in the most recent cohorts, reaching 69.2% for the 1993 cohort. This suggests that mobility in the U.S. may be improving, although one must be cautious in extrapolating from the college gradient to the income gradient as we explain below. We find very similar results when measuring college attendance at later ages (Appendix Figure 5).

Our estimates of the college attendance gradient for the 1984 cohort are consistent with Bailey and Dynarski's (2011) estimates for the 1979-82 cohorts in survey data. Bailey and Dynarski show that the college attendance gap between children from families in the top vs. bottom quartile of the income distribution grew between the 1961-64 and 1979-82 birth cohorts. Our data show that the college attendance gap has stabilized in more recent cohorts.⁷

One can obtain a richer prediction of a child's future income using information not just on whether a child attends any college, but on which college the child attends. Using data from 1098-T forms, Chetty, Friedman, and Rockoff (2013) construct an earnings-based index of "college quality" using the mean individual wage earnings at age 31 of children born in 1979-80 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 population-based sample a value of this college

⁷ Duncan, Kalil, and Ziol-Guest (2013) show that much of the increase documented by Bailey and Dynarski is driven by the increased inequality among parents rather than an increase in the association between college attendance and the level of parent income. The slower growth of income inequality in the 1990s (Card and Dinardo 2002, Autor, Katz, and Kearney 2008) could explain why the relationship between parent income ranks and college attendance is more stable for recent cohorts.

quality index based on the college in which they were enrolled at age 19. We then convert this dollar index to percentile ranks, assigning the 52.7% of children who do not attend college a rank of 26.6.⁸

The relationship between a child's college quality rank and parent income rank is convex (Appendix Figure 6), because most children from low-income families do not attend college. To account for this non-linearity, we define the gradient in college quality as the difference in mean college quality rank between children with parents around the 75th percentile (percentiles 72 to 78) and children with parents around the 25th percentile (percentiles 22 to 28). The time series of the resulting college quality gradient is almost identical to the time series of the college attendance gradient (Appendix Figure 7). Hence, intergenerational mobility is stable (or improving slightly) not just based on college attendance rates, but also based on college quality.

Consolidated Series. We construct a consolidated series of intergenerational mobility for the 1971-93 birth cohorts by combining the age 29-30 income gradient (Appendix Table 1, Column 5), the age 26 income gradient (Column 7), and the college attendance gradient (Column 8). To do so, we multiply the age 26 income gradient by a constant scaling factor of 1.12 to match the level of the age 29-30 income gradient for the 1980-82 cohorts, when both measures are available. Similarly, we multiply the college gradients by a scaling factor of 0.40 to match the rescaled age 26 income gradients from 1984-1986.

The series in circles in Figure 2 presents the resulting consolidated series from 1971-93. The solid circles are simply the estimates based on age 29-30 income; the open circles are forecasts based on age 26 income for the 1983-86 cohorts and college attendance for the 1987-93 cohorts. This consolidated series provides a forecast of intergenerational income mobility at age 30 for recent cohorts under the assumption that the college and age 26 income gradients are always a constant multiple of the age 30 income gradient.⁹

The consolidated series is virtually flat. The estimated trend based on an OLS regression using the 23 observations in this series is -0.0006 per year and the upper bound of the 95%

⁸ The children in the no-college group all have the same value of the college quality index. Breaking ties at the mean, we assign these children a rank of $52.7/2+0.3=26.6\%$ because 0.3% of children in the sample attend colleges whose mean earnings are below the mean earnings of those not in college.

⁹ The validity of this assumption should be evaluated as data for more cohorts become available. The fact that the college gradient increased between the 1960 and 1980 birth cohorts (Bailey and Dynarski 2011) while the income gradient was unchanged (Figure 2 and Lee and Solon 2009) suggests that this assumption did not hold during that period. The college gradient might provide a better forecast for recent cohorts, as college attendance rates are more stable over the period we study (Appendix Table 5, Column 10).

confidence interval is 0.0008. This implies that intergenerational persistence of income ranks increased by at most $0.0008/0.3=0.27\%$ per year between the 1971 and 1993 birth cohorts.¹⁰

Transition Matrices. We supplement our analysis of rank-rank slopes by considering an alternative statistic that directly measures a child's chances of "success": the probability that a child reaches the top quintile of the income distribution (Auten, Gee, and Turner 2013). We define quintiles by ranking children relative to others in their birth cohort and parents relative to other parents of children in the same birth cohort.

Figure 3 plots children's probabilities of reaching the top income quintile of their cohort conditional on their parents' income quintile. Children's incomes are measured at age 26. The series in circles use the SOI sample, while those in triangles use the population-based sample. All the series exhibit little or no trend. For instance, the probability of reaching the top quintile conditional on coming from the bottom quintile of parental income is 8.4% in 1971 and 9% in 1986. Measuring child income at age 29-30 in the SOI sample yields similar results (Appendix Table 4).

Regional Differences. The trends in mobility are small especially in comparison to the variation across areas within the U.S. Using data for the 1980-85 cohorts, CHKS show that the probability that a child rises from the bottom to the top quintile is 4% in some parts of the Southeast but over 12% in other regions, such as the Mountain states. In Figure 4, we assess whether these differences across areas persist over time. This figure plots the age 26 income rank-rank slopes and college attendance gradients by birth cohort for selected Census divisions (see Appendix Table 5 for estimates for all Census divisions). We assign children to Census divisions based on where their parents lived when they claimed them as dependents and continue to rank both children and parents in the national income distribution.

The gradients are quite stable: they are consistently highest in the Southeast and lowest in the Mountain and Pacific states, with New England in the middle. There are, however, some modest differential trends across areas. For example, the age 26 income rank-rank slope fell from 0.326 to 0.307 from the 1980-1986 birth cohorts in the Southeast, but increased from 0.244 to 0.267 in New England. Studying such differential trends may be a fruitful path to understanding the causal determinants of mobility. To facilitate such work, we have publicly posted intergenerational mobility estimates by commuting zone for the 1980-1993 birth cohorts in [Online Data Table 1](#).

¹⁰ Appendix Table 3 replicates this analysis cutting the sample by the child's gender. We find no trend in mobility for males or females.

Changes in Marginal Distributions. To complement the preceding rank-based characterization of mobility, we characterize the marginal income distributions for parents and children in our sample. Appendix Table 6 presents two standard measures of inequality – Gini coefficients and top 1% income shares – for parents and children by birth cohort. Consistent with prior research, we find that inequality amongst both parents and children has increased significantly in our sample. The increase in the Gini coefficient for parents in the bottom 99% of the distribution almost exactly matches the increase observed in the Current Population Survey (see Appendix A). The increase in the Gini coefficients for children is smaller, likely because children’s income is measured at an earlier age, when the income distribution is compressed. Since the trends in the marginal distributions in our sample closely mirror those in the CPS, existing evidence on changes in marginal income distributions can be combined with the rank-based estimates of mobility presented here to construct various mobility statistics of interest.

IV. Discussion

Our analysis of new administrative records on income shows that children entering the labor market today have the same chances of moving up in the income distribution relative to their parents as children born in the 1970s.¹¹ Putting together our results with evidence from Hertz (2007) and Lee and Solon (2009) that intergenerational elasticities of income did not change significantly between the 1950 and 1970 birth cohorts, we conclude that rank-based measures of social mobility have remained remarkably stable over the second half of the twentieth century in the United States.¹² In light of the findings in our companion paper on the geography of mobility (CHKS), the key issue is not that prospects for upward mobility are declining but rather that some regions of the U.S. persistently offer less mobility than most other developed countries.

¹¹ Interestingly, rank-based measures of *intragenerational* mobility – income mobility over the lifetime for a given individual – are also stable over this period (Kopczuk, Saez, and Song 2010, Auten, Gee, and Turner 2013).

¹² As noted above, the stability of the log-log IGE documented by Hertz and Lee and Solon implies stability of the rank-rank relationship if the marginal distributions of parent and child income are both expanding similarly, which is approximately true in practice. In contrast to the findings of Hertz and Lee and Solon, Aaronson and Mazumder (2008) report evidence that mobility fell during the middle of the 20th century using Census data. However, Aaronson and Mazumder do not observe parent income in their data and therefore use the child’s state of birth as a proxy for parent income, which generates bias if there are significant place effects on income. More recently, Justman and Krush (2013) also argue that mobility declined over this period, but employ a regression specification that includes the child’s education as a control. Since education is endogenous to parent income, their regression coefficients cannot be interpreted as estimates of intergenerational income persistence.

The lack of a trend in intergenerational mobility contrasts with the increase in income inequality in recent decades. This contrast may be surprising given the well-known negative correlation between inequality and mobility across countries (Corak 2013). Based on this “Great Gatsby curve,” Krueger (2012) predicted that recent increases in inequality would increase the intergenerational persistence of income by 20% in the U.S.¹³ One explanation for why this prediction was not borne out is that much of the increase in inequality has been driven by the extreme upper tail (Piketty and Saez 2003, U.S. Census Bureau 2013). In CHKS, we show that there is little or no correlation between mobility and extreme upper tail inequality – as measured e.g. by top 1% income shares – both across countries and across areas within the U.S. Instead, the correlation between inequality and mobility is driven primarily by “middle class” inequality, which can be measured for example by the Gini coefficient among the bottom 99%. Based on CHKS’s estimate of the correlation between the bottom 99% Gini coefficient and intergenerational mobility across areas, we would expect the correlation of parent and child income ranks to have increased by only 7.5% (from 0.30 to 0.323) from the 1971 to 1993 birth cohorts (see Appendix A). From this perspective, it is less surprising that mobility has not changed significantly despite the rise in inequality.

The stability of intergenerational mobility is perhaps more surprising in light of evidence that socio-economic gaps in early indicators of success such as test scores (Reardon 2011), parental inputs (Ramey and Ramey 2010), and social connectedness (Putnam, Frederick, and Snellman 2012) have grown over time. Indeed, based on such evidence, Putnam, Frederick, and Snellman predicted that the “adolescents of the 1990s and 2000s are yet to show up in standard studies of intergenerational mobility, but the fact that working class youth are relatively more disconnected from social institutions, and increasingly so, suggests that mobility is poised to plunge dramatically.” An important question for future research is why such a plunge in mobility has not occurred.¹⁴

¹³ Krueger’s prediction is based on comparing Gini coefficients in 1985 and 2010. Children in the 1971 cohort, the first cohort in our sample, reached age 10 (roughly the midpoint of childhood) in 1981, while those in the 1993 cohort, the last cohort in our sample, reached age 10 in 2003. The increase in the Gini coefficient between 1981 and 2003 was larger than the increase between 1985 and 2010 (see Appendix A). Hence, based on Krueger’s extrapolation, we would predict that mobility would fall by more than 20% over the cohorts we study here.

¹⁴ There is a strong cross-sectional correlation across areas of the U.S. between intergenerational mobility and measures of social capital, family structure, and test scores (CHKS), making the lack of a time series relationship more surprising. One potential explanation is that other countervailing trends – such as improved civil rights for minorities or greater access to higher education – have offset these forces.

REFERENCES

- Aaronson, Daniel and Bhashkar Mazumder.** 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000," *Journal of Human Resources* 43(1): 139-172.
- Auten, Gerald, Geoffrey Gee and Nicholas Turner.** 2013. "Income Inequality, Mobility, and Turnover at the Top in the US, 1987-2010." *American Economic Review* 103(3): 168-172.
- Autor, David H., Lawrence F. Katz, and Melissa S. Kearney.** 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics* 90(2): 300-323.
- Bailey, Martha and Susan Dynarski.** 2011. "Inequality in Postsecondary Attainment." In Greg Duncan and Richard Murnane, eds., *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*: 117-132. New York: Russell Sage Foundation.
- Black, Sandra E. and Paul J. Devereux.** 2011. "Recent Developments in Intergenerational Mobility." in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, Chapter 16, pp. 1487-1541.
- Card, David and John E. DiNardo.** 2002. "Skill-Biased Technological Change And Rising Wage Inequality: Some Problems And Puzzles." *Journal of Labor Economics* 20(4): 733-783.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff.** 2013. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood" NBER Working Paper 19424, National Bureau of Economic Research, Inc.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez.** 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." NBER Working Paper 19843, National Bureau of Economic Research, Inc.
- Corak, Miles.** 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 (3): 79-102.
- Duncan, Greg J., Ariel Kalil, and Kathleen M. Ziol-Guest.** 2013. "Increasing Inequality in Parent Incomes and Children's Schooling." University of California Irvine mimeo.
- Foroohar, Rana.** 2011. "What Ever Happened to Upward Mobility?" *Time Magazine* 34(3). November 14.
- Hauser, Robert M.** 2010. "Intergenerational Economic Mobility in the United States: Measures, Differentials, and Trends." UW-Madison Center for Demography and Ecology working paper.
- Hertz, Thomas.** 2007. "Trends in the Intergenerational Elasticity of Family Income in the United States." *Industrial Relations: A Journal of Economy and Society* 46(1): 22-50.
- Justman, Moshe and Anna Krush.** 2013. "Less Equal and Less Mobile: Evidence of a Decline in Intergenerational Income Mobility in the United States." Melbourne Institute Working Paper 43/13.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937." *Quarterly Journal of Economics* 125(1): 91-128.

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

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

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118(1): 1-39.

Putnam, Robert D., Carl B. Frederick, and Kaisa Snellman. 2012. "Growing Class Gaps in Social Connectedness among American Youth." Cambridge: Harvard Kennedy School of Government. The Saguaro Seminar: Civic Engagement in America.

Ramey, Valerie and Garey Ramey. 2010. "The Rug Rat Race." *Brookings Papers on Economic Activity* Spring 2010: 129-176.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." In R. Murnane & G. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*: 91-116. New York: Russell Sage Foundation Press.

U.S. Census Bureau. 2013. "Income, Poverty, and Health Insurance Coverage in the United States: 2012." Current Population Report P60-245, Washington, DC: Government Printing Press.

Zakaria, Fareed. 2011. "The Downward Path of Upward Mobility." *Washington Post*. November 9th.

Appendix A: Changes in Inequality and Predicted Changes in Mobility

In this appendix, we first calculate the change in the Gini coefficient from 1981 to 2003 using data from the Current Population Survey and tax records. Using this estimate, we then predict the change in intergenerational mobility based on the cross-sectional relationship between mobility and inequality reported in CHKS.

Gini Coefficients: Current Population Survey. Based on data from the CPS (Table F-4 at this Census [website](#)), the Gini coefficient for after-tax income of families rose from 0.369 in 1981 (when children in the 1971 cohort were 10 years old, the mid-point of their childhood) to 0.436 in 2003 (when children in the 1993 cohort were 10 years old). There is a discontinuity of 2.1 points in the series of Gini coefficients from 1992 to 1993 due to a change in top-coding methodology. If we eliminate this jump, the Gini coefficient increases by $0.046 = 0.436 - 0.369 - 0.021$ from 1981 to 2003. We interpret this Gini coefficient as applying to the bottom 99% of the income distribution because income is top-coded in the CPS. Note that adjusting for the data break in 1993, the increase in the Gini coefficient from 1985 to 2010, the period studied by Krueger (2012), is 0.030.

Gini Coefficients: Tax Data. Using the SOI public use cross-sections, we calculate Gini coefficients and top 1% income shares using all tax filers with at least one dependent child. We measure income as pre-tax adjusted gross income including full realized capital gains for consistency between 1981 and 2003. As in CHKS, we define the Gini coefficient for the bottom 99% as the overall Gini coefficient minus the top 1% income share. We estimate that the bottom 99% Gini increases from 0.337 in 1981 to 0.382 in 2003. This increase in the bottom 99% Gini coefficient of 0.045 is nearly identical to the estimate of 0.046 from the CPS.

Predicted Change in Mobility. An unweighted OLS regression of the rank-rank slope (for the 1980-82 birth cohort) on the bottom 99% Gini coefficient with one observation per commuting zone yields a coefficient of 0.548 using the data in Online Data Tables V and VIII of CHKS. Therefore, one would predict that an increase of 0.046 in the bottom 99% Gini coefficient (the estimate based on CPS data) would increase the rank-rank correlation of parent and child income by approximately $0.548 \times 0.046 = 0.025$, 7.5% of the mean value of the rank-rank slope (0.334) in the sample analyzed by CHKS.

Appendix B: Comparison to Clark (2014)

Clark (2014) presents estimates of mobility across generations using surname averages of income, representations in elite professions, and other related outcomes. He obtains implied IGE estimates around 0.8, well above the estimates of intergenerational persistence obtained in our analysis and the prior literature (e.g., Solon 1999). In this appendix, we first replicate Clark's surname-mean approach in our data and show that estimates from surname means are generally quite similar to those obtained from conventional micro estimates of the IGE. We then provide a simple hypothesis that may explain why Clark's focus on rare surnames leads to a much higher estimated IGE.

Surname-Based Estimates. We construct estimates of intergenerational mobility across surnames as follows. We begin with all the children in our core sample and restrict attention to those whose surnames are the same as their parents' surnames. As Clark (2014, Appendix 2) notes, surname-based analyses will yield attenuated estimates of the IGE if they include parents and children who do not actually have the same surname. Consistent with this hypothesis, we find smaller estimates of rank-

rank correlations and IGE's when we use the full core sample, without limiting to children who have the same surname as their parents. We then obtain a de-identified table of surname-level means of percentile ranks (using the baseline income definition) for both parents and children. Finally, we regress the surname-level mean ranks for children on surname-level mean ranks for parents (as suggested by Clark 2014, Appendix 2), weighting by the number of individuals with each surname, to obtain a surname-level rank-rank slope. We construct surname-level estimates of the log-log IGE analogously, computing surname level means of log income (excluding zeroes) for children and parents.

Appendix Table 7 reports the results of this analysis. Each row of the table shows the estimates for different subsets of names. The first row considers all names. Rows 2-4 restrict to rare surnames, i.e. names held by fewer than 25, 50, or 100 children. Rows 5-8 conversely limit the sample to common surnames, i.e. names held by more than 100, 1000, 10000, or 20000 people. In each row, we report the number of children in the sample (Column 1), the number of unique surnames in the sample (Column 2), surname-mean based estimates of the rank-rank slope (Column 3), individual-level estimates of the rank-rank slope (Column 4), surname-based estimates of the log-log IGE (Column 5), and individual-level estimates of the log-log IGE (Column 6).

We find that the surname-based estimates are generally slightly larger than the individual-level estimates. For example, when including all names (row 1), the individual-level rank-rank slope is 0.30, compared with a surname-level rank-rank slope of 0.39. If we restrict to the rarest names (shared by fewer than 25 people), the individual-level rank-rank slope is 0.27, compared with a surname-based rank-rank slope of 0.30. The IGE estimates at the individual level are also slightly smaller than those based on surname averages. The only case in which the surname averages yield much larger implied IGE's and rank-rank correlations is in the last row of the table, where we restrict to the 7 most common names in the U.S. population. Here, the surname-based IGE is 0.81, compared with an individual-level IGE of 0.36. This implies that the rate of convergence in income across generations across these broad name groups -- which likely capture broad differences in ethnicities or race -- is much smaller than the rate of convergence within the groups. We return to this point in greater detail below. However, the general pattern that emerges from this analysis is that surname-based averages of income generally do not imply much greater intergenerational persistence than individual-level regressions unless one uses specific subsets of names for the analysis.

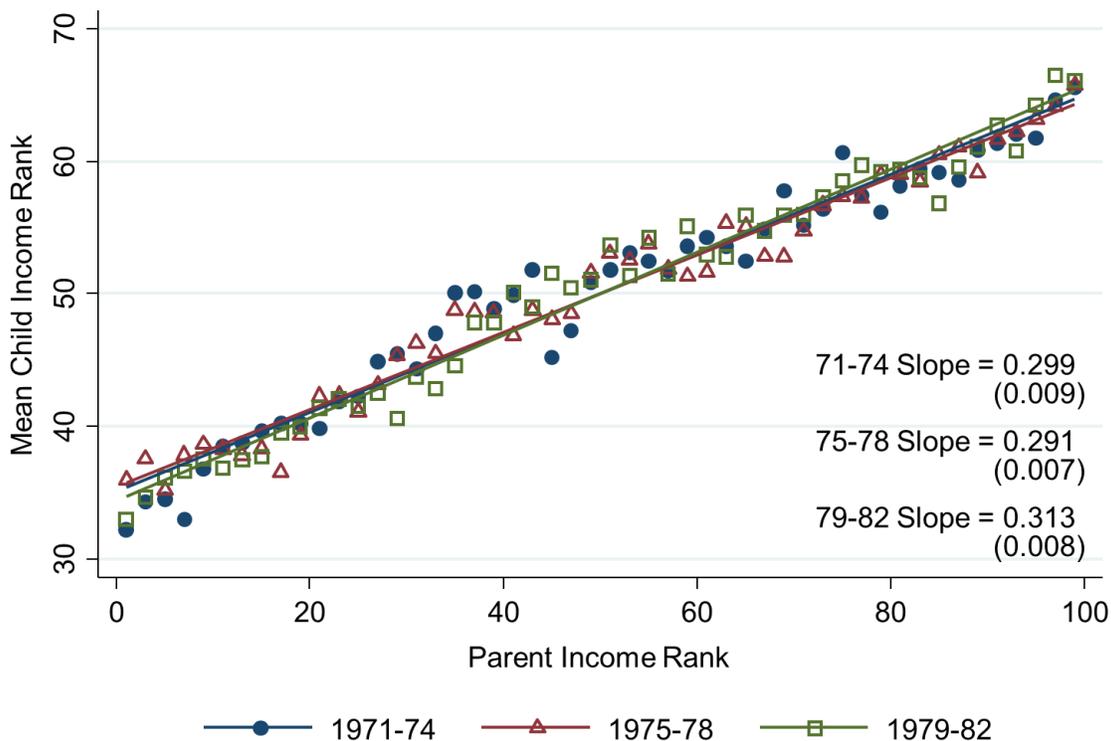
Interpretation of Clark (2014) Estimates. Why does Clark obtain much larger estimates of intergenerational persistence? There are many methodological differences between Clark's analysis and our approach above. For instance, Clark analyzes multiple generations and uses other proxies for status (such as professional occupation) rather than income. While a comprehensive analysis of the source of the difference is outside the scope of this study, we believe that one key difference is Clark's focus on distinctive surnames. For instance, one comparison Clark (204, page 60, Figure 3.10) gives is of the surname "Katz" vs. "Washington." As he notes, Katz is a common Jewish surname, while Washington is a common black surname. The comparison of intergenerational convergence in income between these two names is thus analogous to using an indicator for race as an instrument in a traditional individual-level IGE regression.

As is well known from prior work (Solon 1992), such IV estimates tend to yield much larger implied IGE's, because race may have direct effects on children's income independent of their impacts on parent income. For example, if one uses an indicator for being black as an instrument, the IGE estimate is equivalent to the proportional reduction in the black-white income gap across generations. In 1980, blacks' median earnings were 78.8% that of whites on average (Bureau of Labor Statistics

2011, Table 14, page 41). In 2010, blacks's median earnings were 79.9% that of whites. Hence, the implied between-group IGE is $78.8/79.9=0.986$, consistent with Clark's estimates. Importantly, even though there is very little convergence across racial groups during this time period, there is considerable social mobility within racial groups. This is why our estimates of the IGE based on individual-level data (or pooling all surnames) over the same period are much lower.

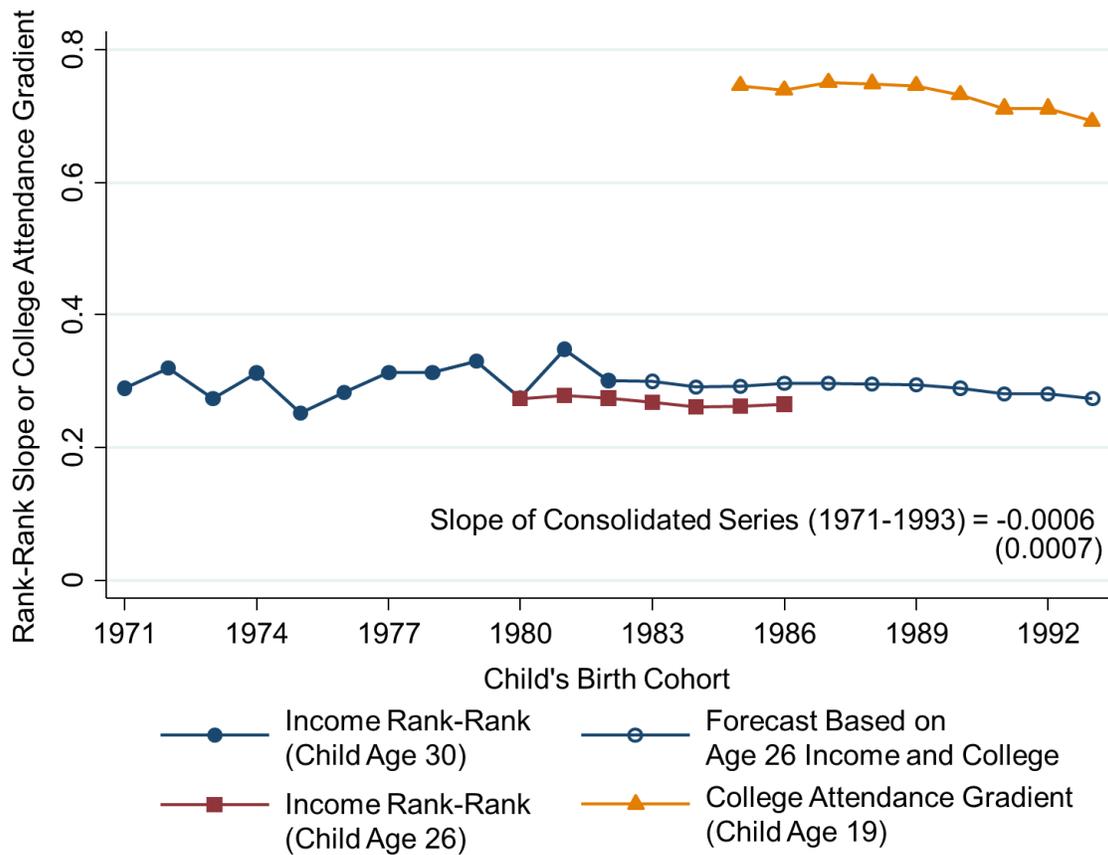
In sum, we believe that Clark's approach effectively identifies a parameter analogous to the degree of convergence in income across generations between racial or ethnic groups rather than across individuals. This is an interesting parameter, but one that differs from standard studies of intergenerational mobility that seek to measure the extent to which an individual's status is determined by his parents' idiosyncratic circumstances. This interpretation differs from that put forth by Clark, who argues that individual-level estimates do not capture latent "status" as well as surname-based averages. Our analysis of surname means suggests that the differences in the results are driven by differences in the rate of income convergence within vs. between ethnic groups rather than a downward bias in measures of intergenerational persistence based on individual data. A useful direction for future research would be to investigate why the rate of income convergence across certain ethnic groups is small even though intergenerational mobility within these groups is much higher.

Figure 1. Child Income Rank vs. Parent Income Rank by Birth Cohort



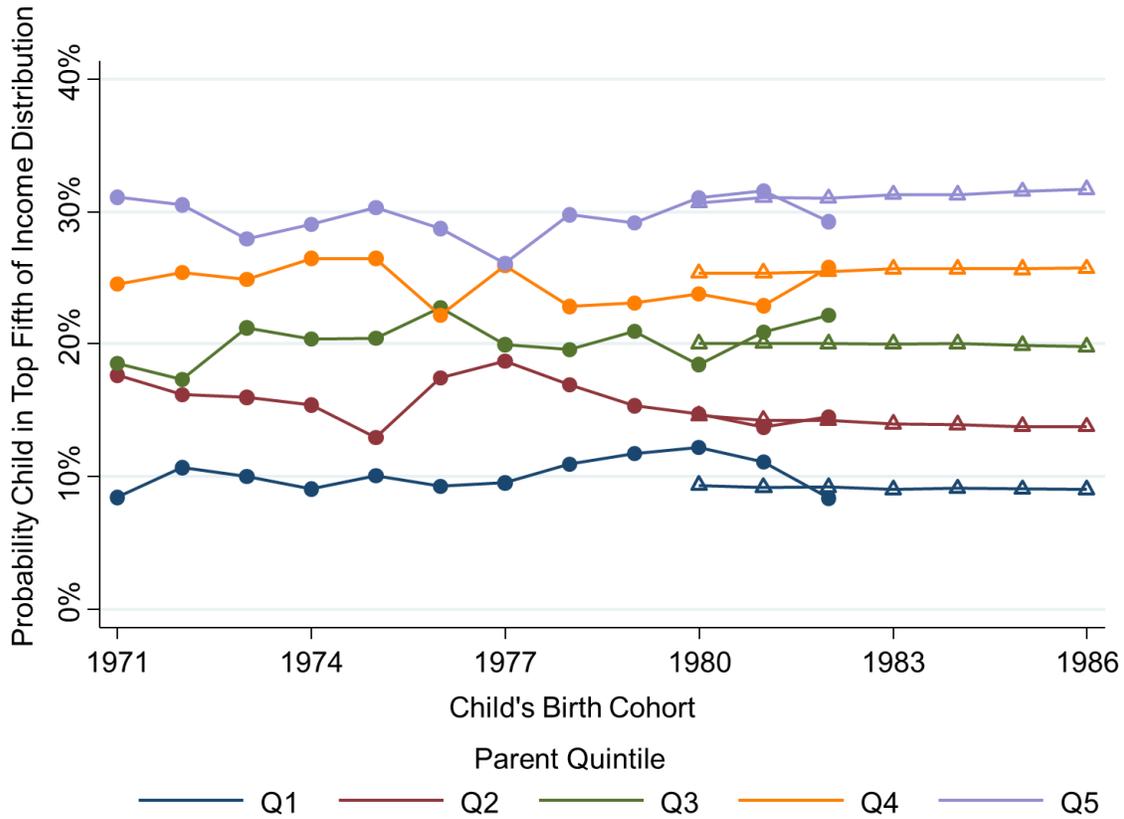
Notes: The figure plots the mean percentile income rank of children at ages 29-30 (y-axis) vs. the percentile rank of their parents (x-axis) for three groups of cohorts (1971-74, 1975-78, and 1979-82) in the SOI sample. The figure is constructed by binning parent rank into two-percentile point bins (so that there are 50 equal-width bins) and plotting the mean child rank in each bin vs. the mean parent rank in each bin. Note that the number of observations varies across bins because the SOI sample is a stratified sample. Estimates from OLS regressions on the binned data are reported for each cohort group, with standard errors in parentheses. Child income is mean family income at ages 29-30. Parent family income is measured in the year the child is claimed as a dependent (between the ages of 12 and 16). Children are ranked relative to other children in their birth cohort and SOI cross-section year. Parents are ranked relative to other parents of children in the same birth cohort and SOI cross-section year.

Figure 2. Intergenerational Mobility Estimates for the 1971-1993 Birth Cohorts



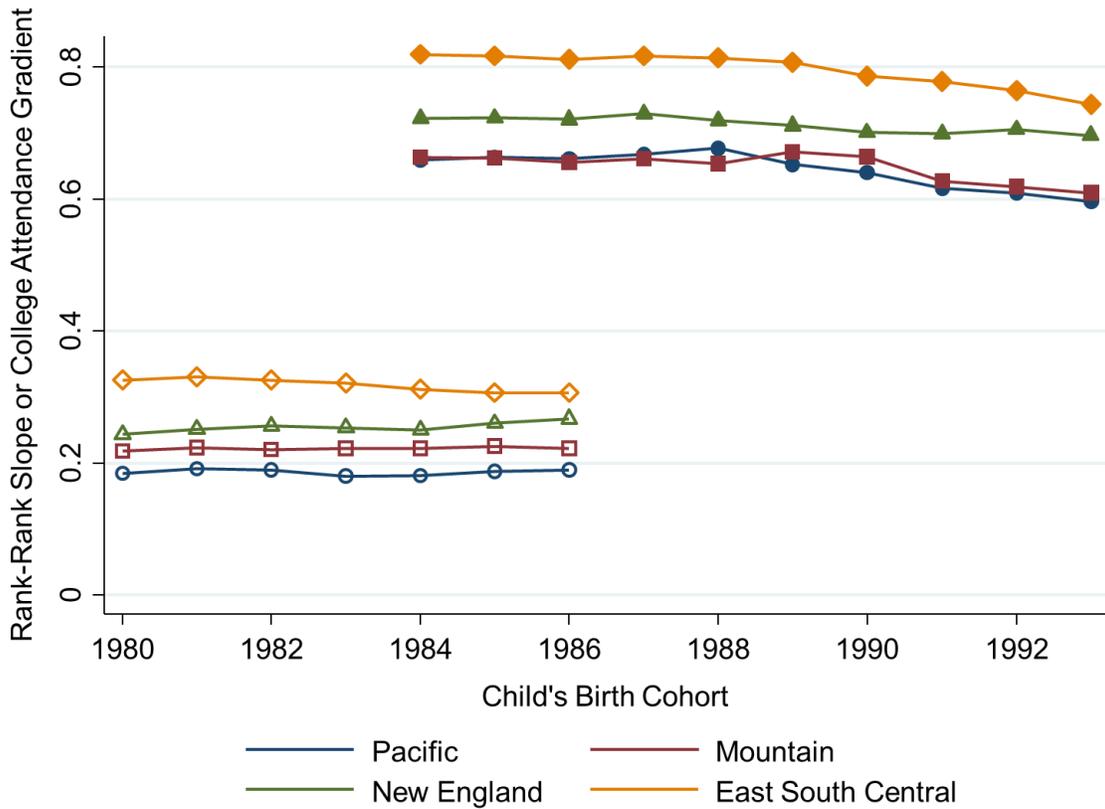
Notes: The series in solid circles plots estimates from weighted OLS regressions (using sampling weights) of child income rank at age 29-30 on parent income rank, estimated separately for each birth cohort in the SOI sample from 1971-82. The series in squares plots estimates from OLS regressions of child income rank at age 26 on parent income rank using the population-based sample for the 1980-86 birth cohorts. The series in triangles replicates the series in squares for the 1984-93 birth cohorts, changing the dependent variable to an indicator for college attendance at age 19, so that the regression coefficient measures the gradient of college attendance rates with respect to parent income rank. The series in open circles represents a forecast of intergenerational mobility based on income at age 26 for the 1983-86 cohorts and college attendance for the 1987-93 cohorts; see text for details. The slope of the consolidated series is estimated using an OLS regression, with standard error reported in parentheses. See Appendix Table 1 for the cohort-level estimates underlying this figure.

Figure 3. Probability of Reaching Top Quintile at Age 26 by Birth Cohort



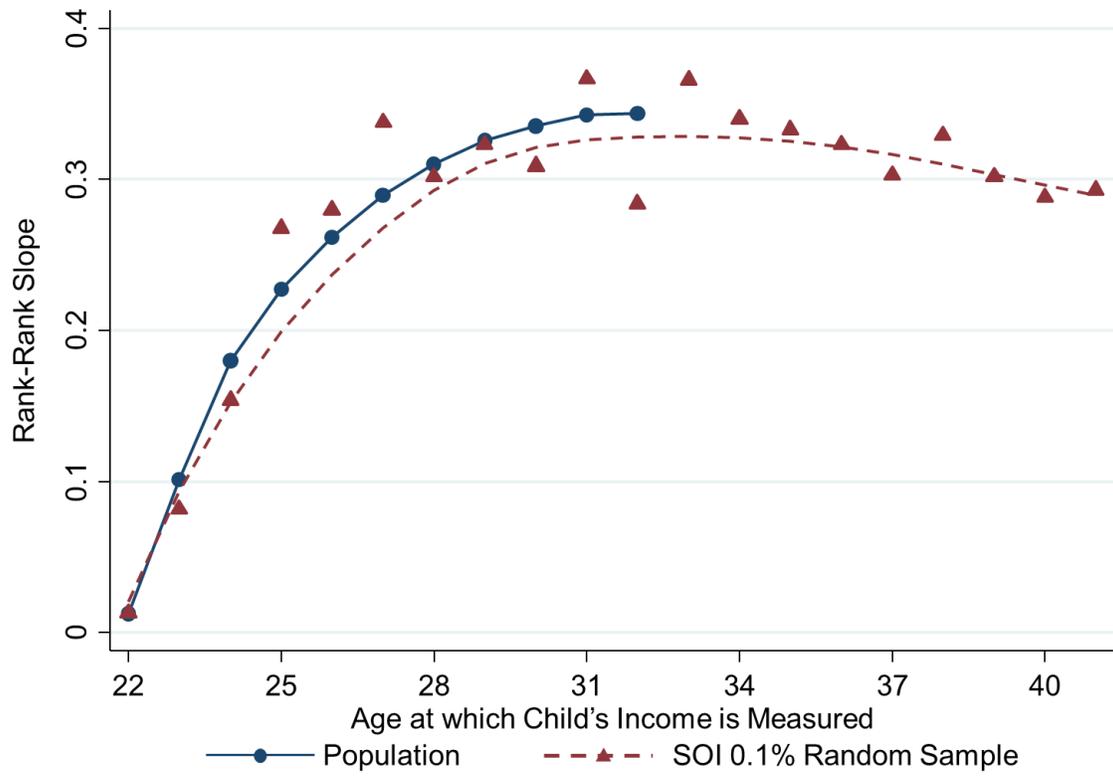
Notes: The figure plots the percentage of children who reach the top quintile of the income distribution for children in their birth cohort. We report this percentage separately for children from each parent income quintile, normalizing the five estimates to sum to one within each birth cohort. The series in circles show estimates using the SOI sample for the 1971-82 birth cohorts. The series in triangles show estimates using the population-based sample for the 1980-86 birth cohorts. Child income is measured at age 26 in both samples. In the SOI sample, parent and child quintiles are defined (using sampling weights) separately within cohort and SOI cross-section year. In the population-based sample, child and parent quintiles are defined separately within each birth cohort. See Appendix Table 4 for estimates using the SOI sample based on child income at ages 29-30.

Figure 4. Trends in Intergenerational Mobility by Census Division



Notes: The figure presents estimates of income rank-rank slopes when children are 26 (open symbols) and college attendance gradients when children are 19 (solid symbols) by birth cohort for four Census divisions. A child's Census division is defined based on the state from which parents filed their tax returns in the year they claimed the child as a dependent. Income ranks are defined nationally, not within each Census division. All estimates use the population-based sample. See Appendix Table 5 for estimates for all nine Census divisions and mean college attendance rates by Census division.

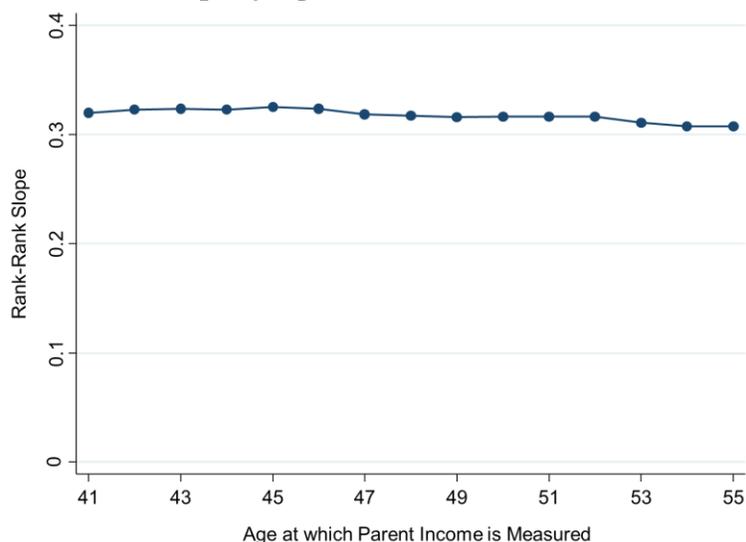
Appendix Figure 1. Lifecycle Bias: Rank-Rank Slopes by Age at which Child's Income is Measured



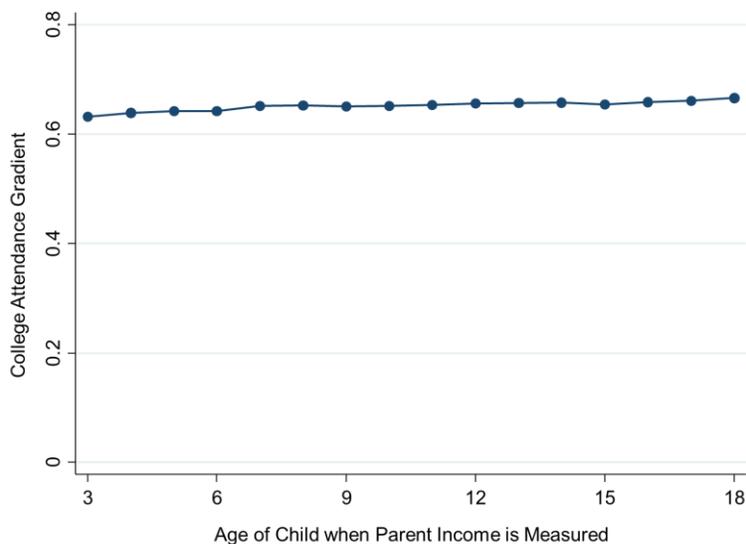
Notes: This figure (reproduced from CHKS), evaluates the robustness of the rank-rank slope to changes in the age at which child income is measured. Child income is defined as mean family income in 2011-2012. Parent income is defined as mean family income from 1996-2000. Each point shows the slope coefficient from a separate OLS regression of child income rank on parent income rank, varying the child's birth cohort and hence the age at which child income is measured in 2011-12. The blue dots use the population data, while the red triangles use the SOI sample. The first point corresponds to the children in the 1990 birth cohort, who are 21-22 when their incomes are measured in 2011-12 (denoted by age 22 on the figure). The last point for which we have population-wide estimates corresponds to the 1980 cohort, who are 31-32 (denoted by 32) when their incomes are measured. The last point in the SOI sample corresponds to the 1972 cohort, who are 39-40 (denoted by 40) when their incomes are measured. The dashed red line is a lowess curve fit through the SOI sample rank-rank slope estimates.

Appendix Figure 2. Mobility Estimates by Age of Parent Income Measurement

A. Rank-Rank Slope by Age at which Parent Income is Measured

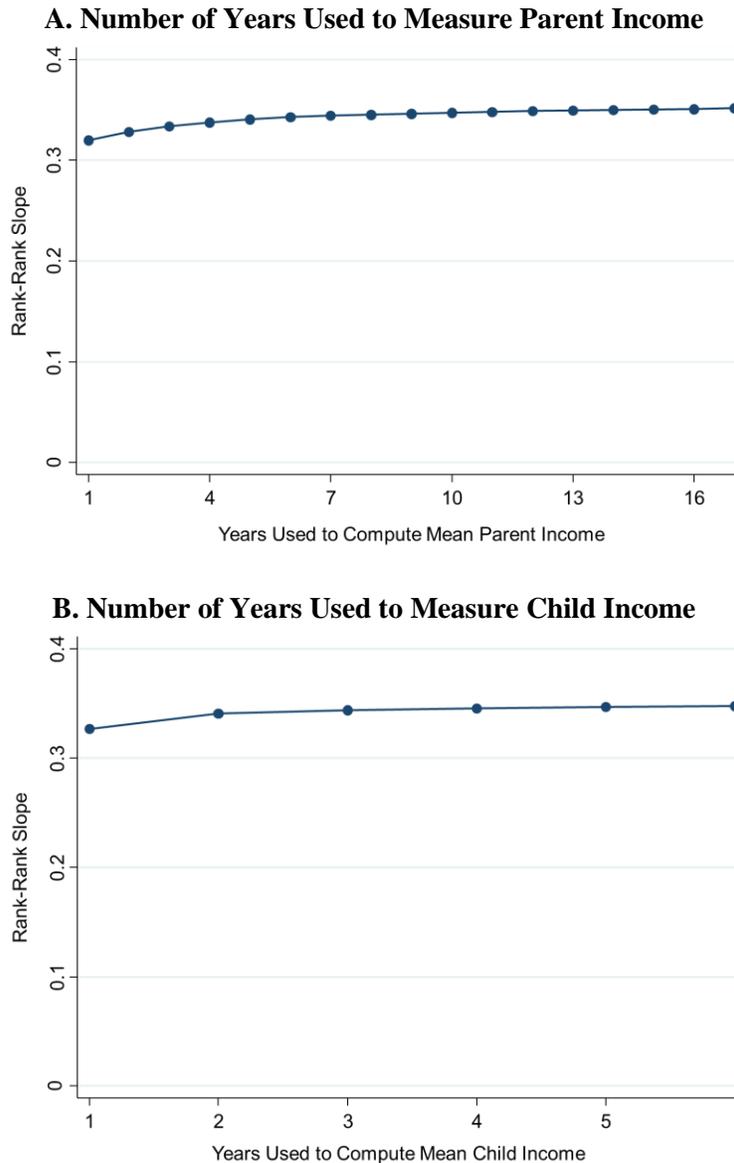


B. College Attendance Gradient by Age of Child when Parent Income is Measured



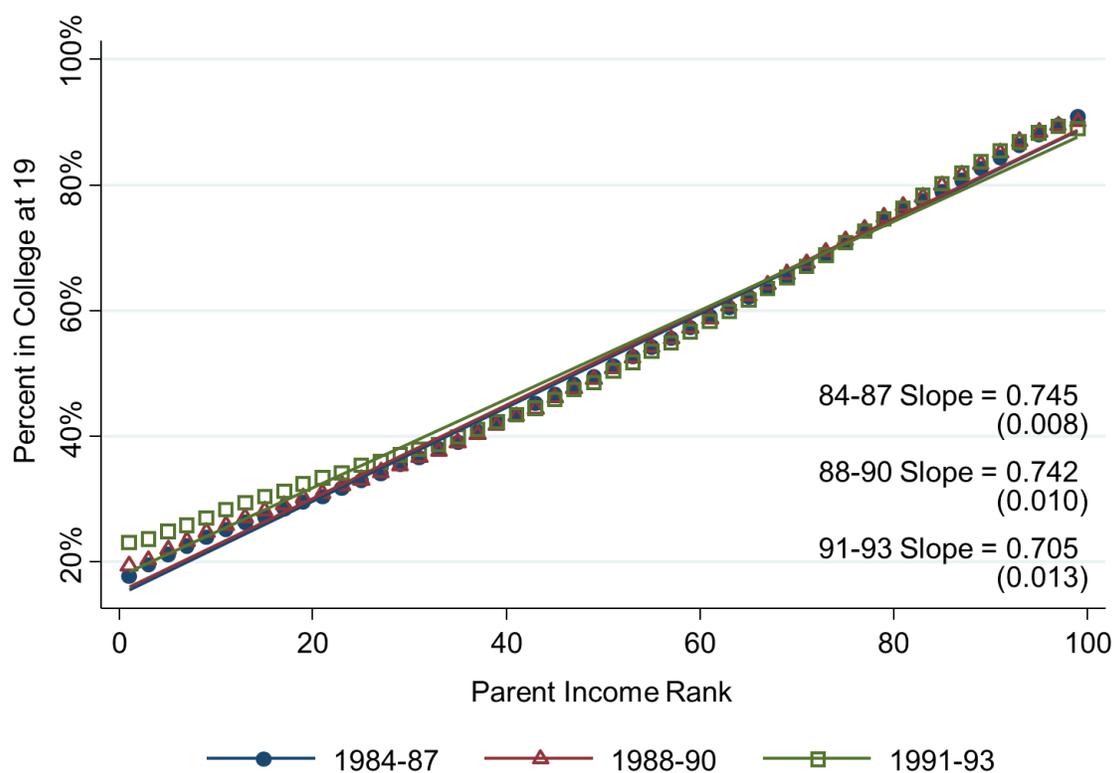
Notes: Panel A (reproduced from CHKS) evaluates the robustness of the rank-rank slope coefficient to changes in the age at which parent income is measured. Panel A is based on children born in 1980-82 in the population-based sample. Each point shows the coefficient from an OLS regression of child income rank on parent income rank, varying the age at which parent income rank is measured. The first point measures parent income in 1996 only, when the mean age of parents is 41. The second point measures parent income in 1997, when parents have a mean age of 42. The last point measures income in 2010, when parents are 55. In Panel B, we evaluate the robustness of the slope of the college-parent income gradient to the age of the child when parent income is measured. Each point shows the slope coefficient from an OLS regression of an indicator for the child attending college at age 19 on parent income rank, varying the year in which parent income rank is measured from 1996 to 2011. In this series, we use data from the 1993 birth cohort. We list the age of the child on the x axis to evaluate whether the gradient differs when children are young (although parent age is of course also rising in lockstep).

Appendix Figure 3. Attenuation Bias: Rank-Rank Slopes by Number of Years used to Compute Parent and Child Income



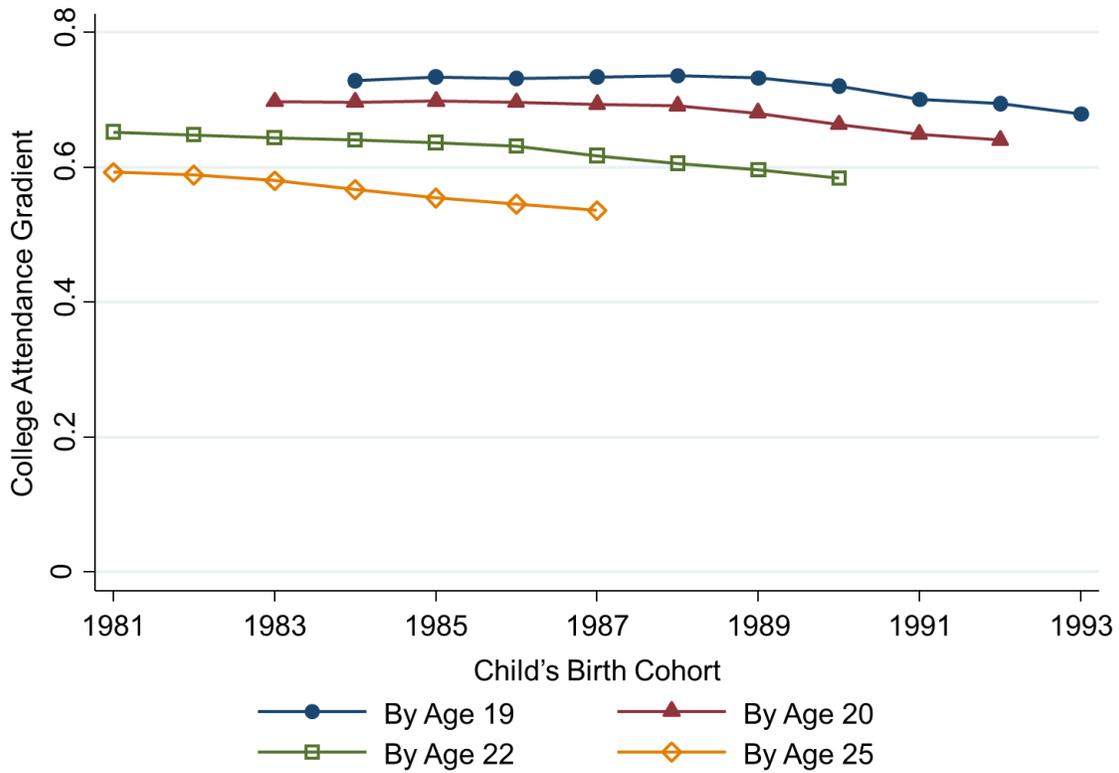
Notes: These figures (reproduced from CHKS) evaluate the robustness of the rank-rank slope estimate to changes in the number of years used compute parent income (Panel A) and child income (Panel B). The figures are based on the population sample of children in the 1980-82 cohorts. In Panel A, each point shows the slope coefficient from an OLS regression of child income rank (based on mean income in 2011-12) on parent income rank as we vary the number of years used to compute mean parent income from 1 to 17. The first point uses parent income data for 1996 only to define parent ranks. The second point uses mean parent income from 1996-1997. The last point uses mean parent income from 1996-2012, a 17 year average. In Panel B, each point shows the coefficient from the same rank-rank regression, but here we always use a five-year (1996-2000) mean to measure parent income and vary the number of years used to compute mean child income. The point for one year measures child income in 2012 only. The point for two years uses mean child income in 2011-12. We continue adding data for prior years; the 6th point uses mean income in years 2007-2012.

Appendix Figure 4. College Attendance Rates vs. Parent Income Rank by Cohort



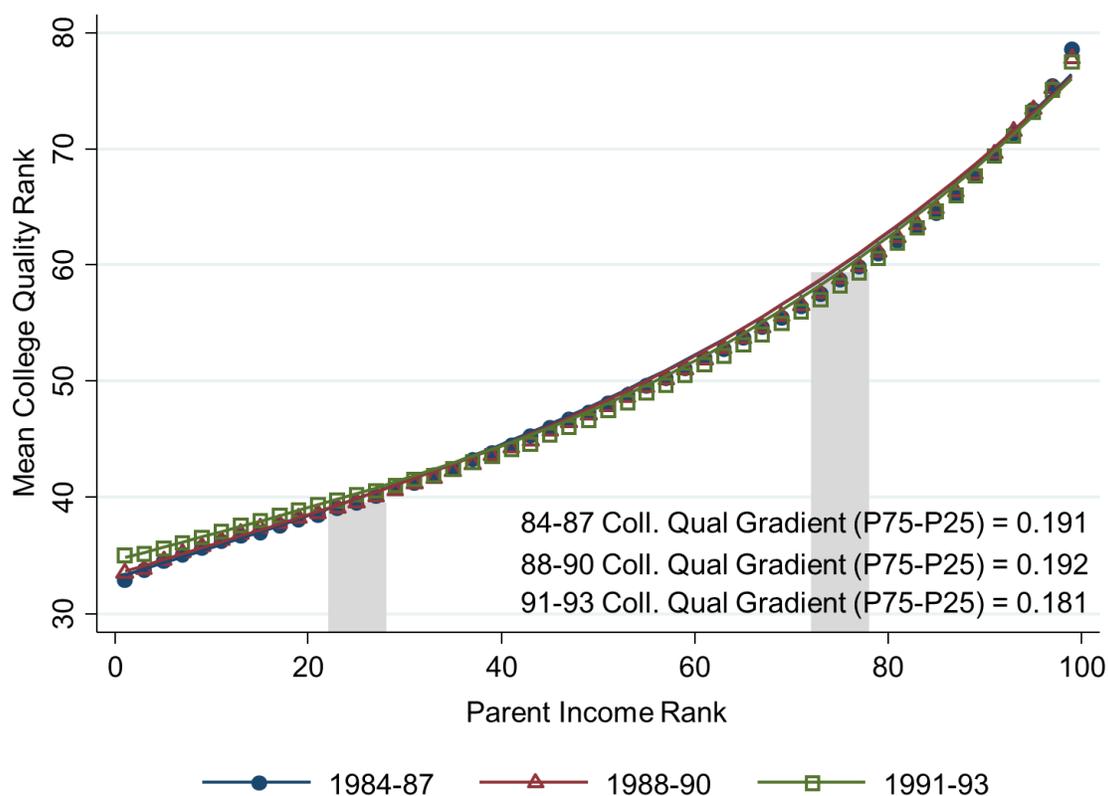
Notes: The figure plots the percentage of children in college at age 19 (y-axis) vs. the percentile rank of their parents (x-axis) for three sets of cohorts (1984-87, 1988-90, and 1991-93) in the population-based sample. The figure is constructed by binning parent rank into two-percentile point bins (so that there are 50 equal sized bins) and plotting the fraction of children attending college at 19 within each bin vs. the mean parent rank in each bin. Estimates from OLS regressions on the binned data are reported for each cohort group, with standard errors in parentheses.

Appendix Figure 5. Robustness of College Attendance Gradient to Age at which College Attendance is Measured



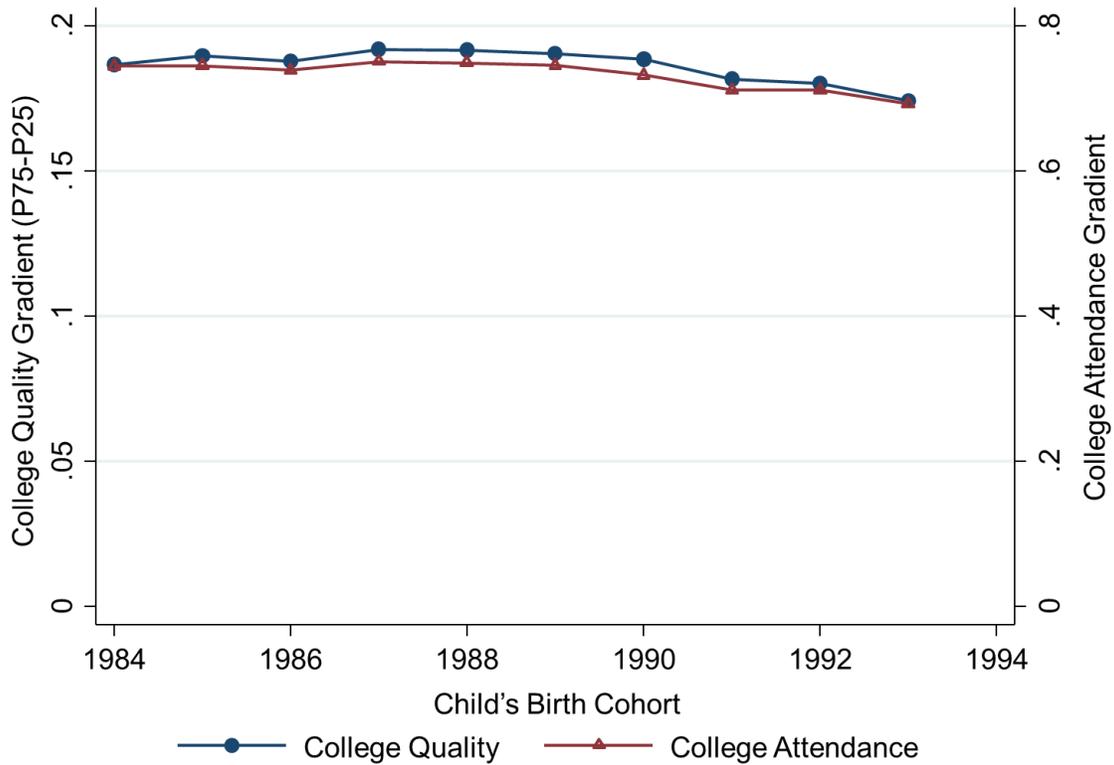
Notes: The figure evaluates the robustness of the college attendance gradient to varying the age at which college attendance is measured. Each series plots the coefficient from a regression of an indicator for college attendance on parent income rank for children in a given birth cohort, similar to the series in triangles in Figure 2. In the series in circles, college attendance is defined as an indicator for the child attending college during or before the year in which he turns 19. The college attendance indicators in the other series are defined analogously at subsequent ages. The number of cohorts covered in each series varies based on data availability; for instance, college attendance by age 25 is only observed up to the 1987 birth cohort, as our last year of data is 2012.

Appendix Figure 6. College Quality vs. Parent Income Rank by Cohort



Notes: The figure plots mean college quality rank (y-axis) vs. the income rank of parents (x-axis) for three sets of cohorts (1984-87, 1988-90, and 1991-93) in the population-based sample. The college quality index (taken from Chetty, Friedman, and Rockoff 2013) is defined as the mean individual wage earnings at age 31 of children born in 1979-80 based on the college they attended at age 20. Children who do not attend college are included in a separate “no college” category. We assign each child in our population-based sample a value of this college quality index based on the college in which they were enrolled at age 19. We then convert this dollar index to percentile ranks, assigning children who do not attend college a rank of 26.6. The figure is constructed by binning parent rank into two-percentile point bins (so that there are 50 equal sized bins) and plotting mean college quality rank in each bin vs. mean parent rank in each bin. The curves shown are lowess fits. The shaded regions correspond to parent percentiles 22-28 and 72-78. For each cohort group, we estimate the college quality gradient as the difference in mean college quality rank between these bins.

Appendix Figure 7. Trends in College Attendance vs. College Quality Gradients



Notes: This figure plots the college attendance gradient (right y axis) and college quality gradient (left y axis) for the 1984-93 birth cohorts in the population-based sample. College attendance and quality are measured at age 19. In each birth cohort, the college quality gradient is defined as the difference in mean college quality rank for children with parents around the 75th percentile (percentiles 72 to 78) vs. children with parents around the 25th percentile (percentiles 22 to 28). See Appendix Figure 6 for further details on the definition of the college quality gradient. In each birth cohort, the college attendance gradient is defined as the coefficient from an OLS regression of an indicator for college attendance on parent income rank. This college attendance gradient reproduces the series in triangles in Figure 2; see notes to that figure for further details. The data plotted in this figure are reported in Columns 8 and 9 of Appendix Table 1.

Appendix Table 1. Number of Observations and Intergenerational Mobility Statistics by Child's Birth Cohort

Birth Cohort	Sample Size		Fraction of Birth Cohort Represented		Rank-Rank Slope			College Attend.	College Quality	Log-Log	Log-Log	Log-Log	Consolidated Series
	SOI Sample	Population	SOI Sample	Population	Income at 29-30	Income at 26	Income at 26	Gradient	Gradient (P75-P25)	Age 29-30	Age 26	Age 26	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1971	4,331		81.5%		0.289	0.212				0.291	0.187		0.289
1972	5,629		83.0%		0.319	0.253				0.346	0.193		0.319
1973	6,179		88.4%		0.274	0.249				0.351	0.281		0.274
1974	7,102		83.7%		0.312	0.261				0.330	0.240		0.312
1975	8,222		85.8%		0.252	0.236				0.254	0.232		0.252
1976	8,257		87.1%		0.282	0.223				0.325	0.202		0.282
1977	8,160		90.4%		0.313	0.233				0.373	0.208		0.313
1978	7,973		88.5%		0.313	0.239				0.338	0.265		0.313
1979	7,593		89.2%		0.330	0.212				0.329	0.240		0.330
1980	7,762	3,092,647	96.1%	85.6%	0.274	0.219	0.273			0.266	0.213	0.274	0.274
1981	8,201	3,323,937	92.7%	91.6%	0.348	0.278	0.279			0.332	0.243	0.271	0.348
1982	9,936	3,448,021	95.8%	93.7%	0.301	0.247	0.274			0.289	0.205	0.257	0.301
1983		3,462,126		95.1%			0.268					0.232	0.300
1984		3,535,065		96.3%			0.261	0.745	0.187			0.219	0.291
1985		3,642,863		96.9%			0.262	0.745	0.190			0.217	0.293
1986		3,650,594		97.2%			0.265	0.739	0.188			0.217	0.296
1987		3,711,400		97.4%				0.751	0.192				0.296
1988		3,815,926		97.6%				0.749	0.192				0.296
1989		3,940,398		97.5%				0.746	0.191				0.294
1990		4,048,638		97.4%				0.732	0.189				0.289
1991		3,994,642		97.2%				0.711	0.182				0.281
1992		3,946,445		97.1%				0.711	0.180				0.281
1993		3,870,924		96.8%				0.692	0.174				0.273

Notes: Column 1 reports the number of children in the SOI sample by child's birth cohort (which is less than the total number of parent-child observations because of repeated parent sampling across years). Column 2 reports the number of children in the population-based sample. Columns 3 and 4 report the fraction of the birth cohort represented, which is equal to the total sample size (using SOI sampling weights for the SOI sample) divided by the size of the birth cohort, based on vital statistics from the Human Mortality Database at UC-Berkeley. Columns 5-8 present estimates from OLS regressions of child outcomes on parent income ranks by cohort. Columns 5-7 regress child income rank on parent income rank. In the population-based sample, ranks are defined within cohort. In the SOI sample, we define ranks by cohort and SOI cross-section year and use sampling weights in all regressions. Column 5 uses child income averaged across ages 29-30, while 6 and 7 use child income at age 26. Column 8 reports the slope of a regression of college attendance, measured as an indicator for the presence of a 1098-T form in the tax year where the child turns 19, on parent income rank. Column 9 reports the difference in average college quality percentile rank between children with parents around the 75th percentile (percentiles 72 to 78) and children with parents around the 25th percentile (percentiles 22 to 28). College quality, taken directly from Chetty, Friedman, and Rockoff (2013), is defined as the mean income at age 31 of children born in 1979-80 based on the college they attended at age 20; those not enrolled in any college are included in a separate category. Columns 10-12 report regressions similar to Columns 5-7, but regress log child income on log parent income in place of child and parent ranks used in Columns 5-7. These log-log regressions drop observations in which the child has zero income. Column 13 reports the consolidated series, where estimates for cohorts 1971-1982 are taken from Column 5 and the estimates for 1983-1993 are constructed based on Columns 7 and 8 as described in the text.

Appendix Table 2. Summary Statistics for SOI and Population-Based Samples

Sample: Variable:	SOI Sample			Population		
	Mean (1)	Median (2)	Std. Dev. (3)	Mean (4)	Median (5)	Std. Dev. (6)
<i>A. 1980-1982 Cohorts</i>						
<u>Parents:</u>						
Parent Family Income (1996-2000 mean)	86,489	54,272	417,928	87,219	60,129	353,430
Parent Income in Year Matched to Child	76,551	55,562	347,071			
Fraction Single Parents	31.4%		46.4%	30.6%		46.1%
Fraction Single Parents Female	65.9%		47.4%	72.0%		44.9%
Father's Age at Child Birth	28.8	28	6.8	28.5	28	6.2
Mother's Age at Child Birth	26.3	26	6.3	26.1	26	5.2
Father's Age When Linked to Child	42.8	42	7.0	43.5	43	6.3
Mother's Age When Linked to Child	40.3	40	6.5	41.1	41	5.2
<u>Children:</u>						
Child Family Income (2011-2012 mean)	47,696	34,146	92,397	48,050	34,975	93,182
Child Family Income (Age 29-30 mean)	45,754	32,892	128,877			
Fraction with Zero Income (Age 29-30)	6.8%		25.2%	6.1%		23.9%
Fraction Female	49.4%		50.0%	50.0%		50.0%
<i>B. 1971-1982 Cohorts</i>						
<u>Parents:</u>						
Parent Family Income (1996-2000 mean)	89,130	56,594	493,685			
Parent Income in Year Matched to Child	75,858	57,700	379,827			
Fraction Single Parents	29.2%		45.5%			
Fraction Single Parents Female	64.6%		47.8%			
Father's Age at Child Birth	28.4	28	10.4			
Mother's Age at Child Birth	25.8	25	6.3			
Father's Age When Linked to Child	42.6	42	10.4			
Mother's Age When Linked to Child	40.0	40	6.4			
<u>Children:</u>						
Child Family Income (2011-2012 mean)	58,428	40,742	102,713			
Child Family Income (Age 29-30 mean)	49,923	37,553	119,273			
Fraction with Zero Income (Age 29-30)	5.4%		22.7%			
Fraction Female	49.2%		50.0%			

Notes: The table presents summary statistics for the SOI sample (using sampling weights) in columns 1-3 and the population-based sample used in Chetty et al. (2014) in columns 4-6. Panel A restricts to children in the 1980-82 birth cohorts, while Panel B uses all cohorts in the SOI sample, 1971-1982. The SOI sample includes all individuals alive at age 30 with a valid SSN or ITIN for whom we are able to identify parents based on dependent claiming in SOI cross-sections. The population-based sample includes all current U.S. citizens with a valid SSN or ITIN for whom we are able to identify parents based on dependent claiming (at any point from 1996-2012). Family income is total pre-tax household income as defined in the text. Parents' marital status is measured in the year the parent is matched to the child. In the population-based sample, the age in which parent is linked to child is measured in 1996, the most common year in which parents are linked to children. A child is defined as single if he/she does not file with a spouse in both 2011 and 2012. All dollar values are reported in 2012 dollars, deflated using the CPI-U-RS consumer price index. See Chetty et al. (2014) for additional summary statistics for the population-based sample.

Appendix Table 3. Intergenerational Mobility Statistics by Child Gender

Birth Cohort	Rank-Rank Slope, Income at 29-30 SOI Sample		Rank-Rank Slope, Income at 26 Pop.-Based Sample		College Attendance Gradient Pop.-Based Sample		Consolidated Series	
	Female	Male	Female	Male	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1971	0.298	0.280					0.298	0.280
1972	0.316	0.322					0.316	0.322
1973	0.291	0.262					0.291	0.262
1974	0.304	0.317					0.304	0.317
1975	0.252	0.252					0.252	0.252
1976	0.314	0.251					0.314	0.251
1977	0.282	0.342					0.282	0.342
1978	0.330	0.303					0.330	0.303
1979	0.349	0.311					0.349	0.311
1980	0.276	0.273	0.281	0.266			0.276	0.273
1981	0.335	0.365	0.284	0.274			0.335	0.365
1982	0.336	0.270	0.280	0.270			0.336	0.270
1983			0.272	0.265			0.305	0.297
1984			0.267	0.255	0.739	0.752	0.300	0.286
1985			0.269	0.256	0.736	0.757	0.301	0.286
1986			0.269	0.261	0.730	0.750	0.302	0.292
1987					0.740	0.763	0.303	0.292
1988					0.736	0.763	0.301	0.292
1989					0.732	0.761	0.300	0.291
1990					0.714	0.751	0.292	0.287
1991					0.692	0.731	0.283	0.280
1992					0.692	0.732	0.283	0.280
1993					0.669	0.715	0.274	0.274

Notes: This table presents estimates of intergenerational mobility by birth cohort and gender. Ranks are defined in the full sample (pooling males and females). Columns 1-4 report coefficient estimates from OLS regressions of child income rank on parent income rank, replicating the specifications in Columns 5 and 7 of Appendix Table 1 conditioning on child gender. Columns 5-6 report coefficients from regressions of college attendance on parent income rank, replicating the specification in Column 8 on Appendix Table 1 conditioning on child gender. Columns 7 and 8 report consolidated series, which are constructed in the same way as column 13 of Appendix Table 1, with separate scaling factors by gender.

Appendix Table 4. Probabilities of Child Reaching Top Income Quintile Conditional on Parent Income Quintile

Birth Cohort	SOI Sample, Income at 26					Population Sample, Income at 26					SOI Sample, Income at 29-30				
	Parent Quintile					Parent Quintile					Parent Quintile				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	
1971	8.4%	17.7%	18.5%	24.5%	31.1%						5.9%	16.7%	15.8%	24.6%	36.9%
1972	10.7%	16.2%	17.4%	25.4%	30.5%						8.5%	13.0%	19.6%	23.7%	35.2%
1973	10.0%	16.0%	21.2%	24.9%	27.9%						9.5%	13.2%	21.1%	25.3%	30.8%
1974	9.0%	15.4%	20.4%	26.5%	29.0%						7.3%	13.2%	20.1%	26.6%	33.1%
1975	10.1%	12.9%	20.4%	26.5%	30.3%						8.6%	16.1%	19.5%	23.3%	32.6%
1976	9.3%	17.4%	22.8%	22.1%	28.7%						9.4%	15.7%	19.3%	21.4%	34.4%
1977	9.5%	18.7%	19.9%	25.9%	26.1%						9.0%	14.2%	16.3%	25.8%	34.8%
1978	10.9%	16.9%	19.6%	22.8%	29.8%						8.7%	15.9%	20.4%	21.7%	33.3%
1979	11.7%	15.4%	21.0%	23.1%	29.1%						8.5%	11.9%	20.8%	26.3%	32.8%
1980	12.2%	14.7%	18.4%	23.8%	31.0%	9.3%	14.6%	20.0%	25.3%	30.7%	8.1%	13.9%	21.4%	22.0%	34.8%
1981	11.1%	13.7%	20.9%	22.9%	31.5%	9.2%	14.3%	20.1%	25.4%	31.1%	6.1%	12.0%	20.8%	26.5%	34.8%
1982	8.4%	14.5%	22.2%	25.8%	29.3%	9.2%	14.3%	20.0%	25.5%	31.0%	8.8%	12.6%	21.7%	24.8%	32.1%
1983						9.0%	14.0%	20.0%	25.7%	31.3%					
1984						9.1%	13.9%	20.0%	25.7%	31.3%					
1985						9.1%	13.8%	19.9%	25.7%	31.5%					
1986						9.0%	13.8%	19.8%	25.7%	31.7%					

Notes: Each cell shows the percentage of children in a birth cohort who reached the top fifth of the income distribution given parents in the quintile specified in the column. Columns 1-5 and 11-15 are computed on the SOI sample using a child's income at age 26 and mean income from age 29-30, respectively. In the SOI sample, parent and child quintiles are defined (using sampling weights) separately within cohort and SOI cross-section year. Columns 6-10 use the population-based sample, measuring a child's family income at age 26 as in columns 1-5. In the population-based sample, child and parent quintiles are defined separately within each birth cohort.

Appendix Table 5. Intergenerational Mobility and College Attendance Rates by Census Division

Birth Cohort	Census Division									
	Pacific	Mountain	New England	West North Central	West South Central	Mid Atlantic	South Atlantic	East North Central	East South Central	All U.S.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. College Attendance Gradients										
1984	0.659	0.663	0.722	0.734	0.763	0.754	0.747	0.769	0.818	0.745
1985	0.663	0.662	0.723	0.741	0.754	0.756	0.738	0.774	0.816	0.745
1986	0.660	0.655	0.720	0.710	0.752	0.751	0.741	0.766	0.812	0.739
1987	0.666	0.660	0.729	0.747	0.760	0.754	0.750	0.778	0.816	0.751
1988	0.676	0.653	0.718	0.747	0.751	0.749	0.744	0.783	0.813	0.749
1989	0.652	0.671	0.711	0.756	0.740	0.743	0.750	0.775	0.807	0.746
1990	0.639	0.664	0.700	0.736	0.713	0.733	0.740	0.759	0.786	0.732
1991	0.616	0.627	0.699	0.718	0.697	0.717	0.708	0.736	0.777	0.711
1992	0.609	0.618	0.705	0.718	0.705	0.715	0.716	0.732	0.764	0.711
1993	0.597	0.609	0.696	0.696	0.680	0.691	0.713	0.718	0.742	0.692
B. College Attendance Rates										
1984	0.526	0.462	0.590	0.583	0.452	0.571	0.473	0.532	0.400	0.510
1985	0.528	0.467	0.605	0.585	0.454	0.583	0.485	0.553	0.417	0.520
1986	0.523	0.475	0.607	0.581	0.468	0.571	0.482	0.552	0.411	0.518
1987	0.524	0.484	0.611	0.598	0.464	0.593	0.483	0.558	0.413	0.525
1988	0.522	0.478	0.613	0.597	0.466	0.591	0.484	0.552	0.422	0.522
1989	0.527	0.463	0.619	0.597	0.463	0.596	0.480	0.550	0.421	0.521
1990	0.529	0.474	0.625	0.604	0.462	0.602	0.493	0.564	0.429	0.528
1991	0.525	0.490	0.633	0.599	0.480	0.607	0.489	0.562	0.431	0.530
1992	0.516	0.493	0.632	0.611	0.468	0.606	0.501	0.568	0.439	0.532
1993	0.513	0.488	0.623	0.611	0.464	0.602	0.498	0.563	0.445	0.529
C. Rank-Rank Slopes Using Income at Age 26										
1980	0.185	0.219	0.244	0.248	0.278	0.275	0.307	0.303	0.326	0.273
1981	0.192	0.224	0.251	0.262	0.283	0.285	0.312	0.307	0.331	0.279
1982	0.190	0.221	0.256	0.260	0.277	0.282	0.301	0.309	0.326	0.274
1983	0.181	0.223	0.253	0.261	0.274	0.276	0.290	0.306	0.322	0.268
1984	0.181	0.222	0.250	0.257	0.264	0.269	0.274	0.297	0.312	0.261
1985	0.188	0.226	0.261	0.263	0.262	0.274	0.268	0.299	0.307	0.262
1986	0.190	0.222	0.267	0.267	0.251	0.279	0.281	0.301	0.307	0.265

Notes: This table presents estimates of intergenerational mobility by cohort and census division using the population-based sample. We assign children to Census divisions based on where their parents lived when they claimed them as dependents. Panel A presents estimates of the college attendance gradient by Census division and cohort. For each Census division and cohort, we report the coefficient from a regression of an indicator for college attendance at age 19 on parent income rank. Panel B reports the mean college attendance rates at age 19 by Census division and cohort. Panel C presents the rank-rank slope estimate from a regression of child income rank on parent income rank, where child income is measured at age 26. In both Panel A and Panel C, income ranks are defined nationally, not within each Census division.

Appendix Table 6. Income Inequality by Cohort

Birth Cohort	Gini Coefficient for Children			Gini Coefficient for Parents		Top 1% Income Share for Parents	
	SOI Sample	SOI Sample	Population	SOI Sample	Population	SOI Sample	Population
	2-year average income	Annual income	Annual income	Annual income	5-year average income	Annual income	5-year average income
	Child Age 29-30	Child Age 26	Child Age 26	Child Age 12-16	Child Age 15-19	Child Age 12-16	Child Age 15-19
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1971	0.396	0.449		0.517		10.1%	
1972	0.453	0.478		0.503		11.5%	
1973	0.465	0.434		0.500		11.6%	
1974	0.465	0.445		0.505		11.2%	
1975	0.468	0.453		0.522		11.3%	
1976	0.475	0.476		0.515		11.4%	
1977	0.473	0.495		0.518		11.4%	
1978	0.492	0.498		0.534		11.3%	
1979	0.539	0.530		0.524		11.7%	
1980	0.557	0.515	0.487	0.524	0.497	11.8%	15.8%
1981	0.546	0.500	0.491	0.510	0.506	12.8%	16.4%
1982	0.535	0.505	0.493	0.498	0.511	13.8%	16.4%
1983			0.501		0.517		16.8%
1984			0.502		0.520		16.8%
1985			0.507		0.529		17.6%
1986			0.507		0.535		17.7%
1987					0.550		19.2%
1988					0.568		21.1%
1989					0.574		21.4%
1990					0.575		21.1%
1991					0.577		20.9%
1992					0.574		20.1%
1993					0.564		18.3%

Notes: This table presents income inequality statistics for parents and children using the income definitions and samples that we used to compute intergenerational mobility statistics in Appendix Table 1. Columns 1-3 report Gini coefficients for child family income. In Column 1, we use the SOI sample (with sampling weights) and define child income as mean family income over the 2 years when the child is aged 29-30. Column 2 replicates column 1, measuring child income at age 26 instead of 29-30. Column 3 replicates Column 2 using the population-based sample. Columns 4-5 report Gini coefficients for parent family income and columns 6-7 report top 1% income shares for parent family income. Columns 4 and 6 use the SOI sample, where parent income is measured as the family income in the year the parent is linked to the child (when the child is aged 12 to 16; see text for details). Columns 5 and 7 consider the population-based sample, where parent income is measured as the 5 year average of family income when the child is aged 15 to 19. See Appendix A for a comparison of these trends to estimates based on the CPS.

Appendix Table 7. Estimates of Intergenerational Mobility Using Surname Means vs. Individual Incomes

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

Notes: This table compares two methods of estimating rank-rank slopes and log-log IGEs. In this table, we restrict the core sample to children who have the same surname (between the ages of 30-32) as their parents. Each row imposes further sample restrictions, shown in the left-most column. We estimate the individual-level rank-rank slopes and log-log IGE's (Columns 4 and 6) as in Table I. In Columns 3 and 5, we estimate the rank-rank slopes and log-log IGE's using OLS regressions on a dataset collapsed to surname-level means, weighted by the number of observations for each name. See Appendix B for further details.