American Economic Review: Papers & Proceedings 2014, 104(5): 141–147
http://dx.doi.org/10.1257/aer.104.5.141

BIG DATA IN MACROECONOMICS: NEW INSIGHTS FROM LARGE ADMINISTRATIVE DATASETS

Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility†

By Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner*

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. We present new evidence on trends in intergenerational mobility using administrative earnings records for children born between 1971 and 1993. For the 1971–1986 birth cohorts, 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.

Although these rank-based measures of mobility have 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 paper is an abbreviated version of NBER working paper 19844 (Chetty et al. 2014b). The working paper contains a complete description of the data, additional empirical results, and further discussion of the findings in the context of the prior literature.

I. Measuring Intergenerational Mobility: Conceptual Issues

We decompose the joint distribution of parent and child income into two components: (i) the joint distribution of parent and child ranks, formally known as the *copula* of the distribution, and (ii) the *marginal distributions* of parent and child income. The marginal distributions determine the degree of inequality within each generation, typically measured by Gini coefficients

* Chetty: Harvard University, Littauer Center 226, Cambridge, MA 02138 (e-mail: chetty@fas.harvard.edu); Hendren: Harvard University, Littauer Center 235, Cambridge, MA 02138 (e-mail: nhendren@fas.harvard.edu); Kline: University of California, Berkeley, 631D Evans Hall, CA 94720, (e-mail: pkline@econ.berkeley.edu); Saez: University of California, Berkeley, 530 Evans Hall, CA 94720 (e-mail: saez@econ.berkeley.edu); Turner: US Treasury, 1500 Pennsylvania Ave, Room 1222A, Washington, DC 20220 (e-mail: nicholas.turner@treasury.gov). 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. 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 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.

† Go to http://dx.doi.org/10.1257/aer.104.5.141 to visit the article page for additional materials and author disclosure statement(s).

or top income shares. The copula is a key determinant of mobility across generations. Some commonly used measures of mobility—such as correlations between parent and child percentile ranks and quintile transition matrices—depend purely on the copula. Other measures, such as the log-log intergenerational elasticity of income, combine 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, the marginal distributions of income have widened substantially.

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 percent) 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) versus 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.

Since the appropriate definition of intergenerational mobility depends upon one's normative objective, we characterize the copula and marginal distributions separately in this paper.

II. Data

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 US citizens as of 2013 and are claimed as a dependent on a tax return filed in or after 1996. We link approximately 95 percent of children in each birth cohort to parents based on dependent claiming, obtaining a sample with 3.7 million children per cohort (online Appendix Table 1, column 4).

To obtain data on earlier birth cohorts, we use the Statistics of Income (SOI) annual cross sections since 1987, the earliest year with dependent information (Auten, Gee, and Turner 2013). These cross sections are stratified random samples covering approximately 0.1 percent of tax returns. Using the SOI cross sections, we construct a sample of children in the 1971–1982 birth cohorts, which we refer to as the *SOI sample*. We construct this sample by first identifying children between the ages of 12 and 16 claimed as dependents in the 1987–1998 SOI cross sections and then pooling all the SOI cross sections that contain information for a given birth cohort. Using the sampling weights, we estimate that the SOI sample represents 88 percent of children in each birth cohort based on vital statistics counts, with slightly lower coverage rates in the earliest cohorts (online Appendix Table 1, column 3). Summary statistics for the SOI sample (using sampling weights) and the population-based sample are very similar for the overlapping 1980–1982 birth cohorts (online Appendix Table 2).

Variable Definitions.—We define parent family income (in real 2012 dollars) as adjusted gross income plus tax exempt interest and the nontaxable portion of social security benefits for those who file tax returns. For nonfilers, 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 5 years when the child is 15–19 years old. In the SOI sample, 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 samples, we drop observations with zero or negative parent income.

We define child family income in the same way as parent income, always using data from the population files. We define a child's college attendance as an indicator for having a 1098-T form in the calendar year the child turns 19.

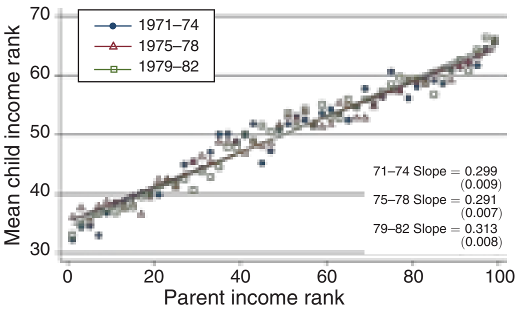

Figure 1. Child Income Rank versus Parent Income Rank by Birth Cohort

Notes: The figure plots the mean percentile income rank of children at ages 29–30 (y-axis) versus the percentile rank of their parents (x-axis) for three groups of cohorts (1971–1974, 1975–1978, and 1979–1982) in the SOI sample. The figure is constructed by binning parent rank into 2-percentile point bins (so that there are 50 equal-width bins) and plotting the mean child rank in each bin versus 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.

Because 1098-T forms are filed directly by colleges, we have records on college attendance for all children.

III. Results

Rank-Rank Specification.—We begin by measuring intergenerational mobility using a *rank-rank* specification. 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.[1] We then study the relationship between child and parent ranks. In our companion paper on the geography of mobility (Chetty et al. 2014a; henceforth CHKS), we show that such rank-rank specifications provide a more robust summary of intergenerational mobility than traditional log-log specifications.

Figure 1 plots the average income rank of children (at ages 29–30) versus parent income rank for three sets of birth cohorts in the SOI sample: 1971–1974, 1975–1978, and 1979–1982. To reduce noise, we divide parent income ranks into 50 (rather than 100) bins and plot the mean child rank versus 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 versus 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. In the working paper version, we show that these rank-rank slope estimates are robust to measuring parent and child income at different ages and using multiple years to measure income, indicating that the estimates do not suffer from significant life cycle or attenuation bias.

Trends in Income Mobility.—Figure 2 presents our primary estimates of intergenerational mobility by birth cohort (see online 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 (online Appendix Table 1).[2]

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 1980–1986 birth cohorts in the population-based sample. Once again, there is no trend in this series. Moreover, there is much

[1] 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.

[2] 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.

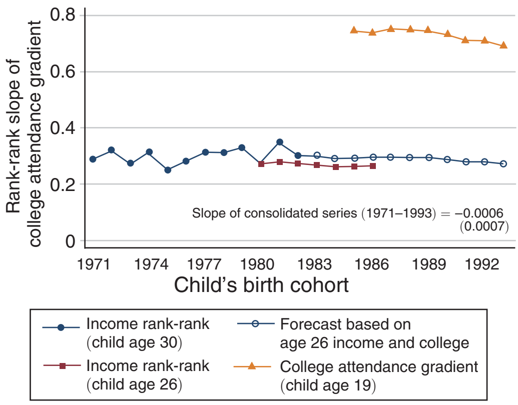

Figure 2. Intergenerational Mobility Estimates for the 1971–1993 Birth Cohorts

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 1971–1982. 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 1980–1986 birth cohorts. The series in triangles replicates the series in squares for the 1984–1993 birth cohorts, changing the dependent variable to an indicator for college attendance at age 19. The series in open circles represents a forecast of intergenerational mobility based on income at age 26 for the 1983–1986 cohorts and college attendance for the 1987–1993 cohorts; see text for details. The slope of the consolidated series is estimated using an OLS regression, with standard error reported in parentheses. See online Appendix Table 1 for the cohort-level estimates underlying this figure.

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 United States. 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.—For children born after 1986, many of whom have not yet entered the labor market, we measure intergenerational mobility based on college attendance. Naturally, college is a strong predictor of earnings. Moreover, 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 United States.

The relationship between college attendance rates and parent income ranks is approximately linear, as shown in the working paper version of this study. 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 1984–1993 birth cohorts. The gap in college attendance rates between children from the lowest- and highest-income families is essentially constant at 74.5 percent between the 1984–1989 birth cohorts. The gap falls slightly in the most recent cohorts, reaching 69.2 percent for the 1993 cohort. In the working paper version, we show that results are very similar when measuring college attendance at later ages. We also show that intergenerational mobility is stable (or improving slightly) when we analyze the correlation between parent income and the quality of the colleges their children attend, as measured by the earnings of prior graduates.

Our estimates of the college attendance gradient for the 1984 cohort are consistent with Bailey and Dynarski's (2011) estimates for the 1979–1982 cohorts in survey data. Bailey and Dynarski show that the college attendance gradient grew between the 1961–1964 and 1979–1982 birth cohorts; our data show that the college attendance gradient has stabilized more recently.

Consolidated Series.—We construct a consolidated series of intergenerational mobility for the 1971–1993 birth cohorts by combining the age 29–30 income gradient (online 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 1980–1982 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 1971–1993. 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 1983–1986 cohorts and college attendance for the 1987–1993 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.

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 percent confidence interval is 0.0008. This implies that intergenerational persistence of income ranks increased by at most $0.0008/0.3 = 0.27$ percent per year between the 1971 and 1993 birth cohorts.[3]

Transition Matrices.—As a supplement to the rank-rank correlation, Figure 3 plots children's probabilities of reaching the top income quintile of their cohort conditional on their parents' income quintile. 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. 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 percent in 1971 and 9 percent in 1986. Measuring child income at age 29–30 yields similar results (online Appendix Table 4).

Regional Differences.—The trends in mobility are small especially in comparison to the variation across areas within the United States. Using data for the 1980–1985 cohorts, CHKS show that the probability that a child rises from the bottom to the top quintile is 4 percent in some parts of the Southeast but over 12 percent in other regions, such as the Mountain states.

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

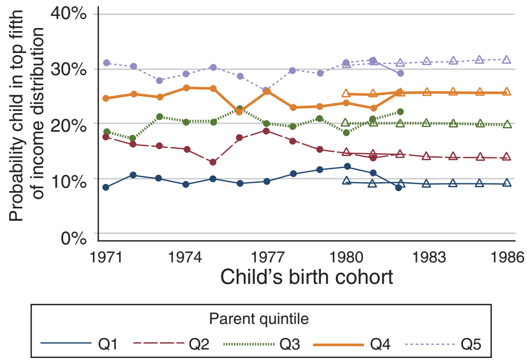

FIGURE 3. PROBABILITY OF REACHING TOP QUINTILE AT AGE 26 BY BIRTH COHORT

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 1971–1982 birth cohorts. The series in triangles show estimates using the population-based sample for the 1980–1086 birth cohorts. Child income is measured at age 26 in both samples.

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 online 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.

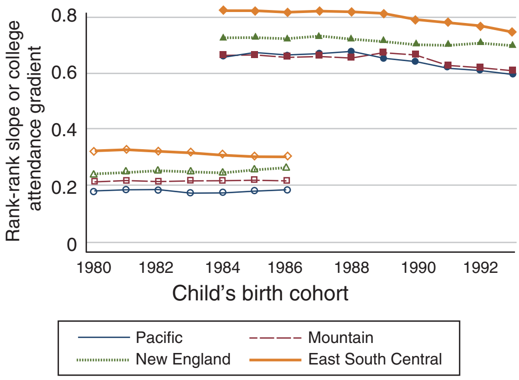

Figure 4. Trends in Intergenerational Mobility by Census Division

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. Income ranks are defined nationally, not within each Census division. All estimates use the population-based sample. See online Appendix Table 5 for estimates for all nine Census divisions and mean college attendance rates by Census division.

Changes in Marginal Distributions.—Consistent with prior research, we find that inequality amongst both parents and children—as measured by Gini coefficients and top 1 percent income shares—has increased significantly in our sample (online Appendix Table 6). The increase in the Gini coefficient for parents in the bottom 99 percent of the distribution almost exactly matches the increase observed in the Current Population Survey, as shown in Appendix A of the working paper. Hence, existing estimates of 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

Putting together our results with evidence from 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 stable over the second half of the twentieth century in the United States. The key issue in our view is not that mobility is declining but rather that some regions of the United States persistently offer less mobility than most other developed countries.

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 percent in the United States. 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). In CHKS, we show that there is little correlation between mobility and extreme upper tail inequality—as measured e.g., by top 1 percent income shares—both across countries and across areas within the United States. 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 percent. Based on CHKS' estimate of the correlation between the bottom 99 percent 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 percent (from 0.30 to 0.323) from the 1971 to 1993 birth cohorts (see the working paper for details). 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 given that socioeconomic gaps in early indicators of success such as test scores, parental inputs, and social connectedness have grown over time (Putnam, Frederick, and Snellman 2012). 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.[4]

REFERENCES

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.

Bailey, Martha, and Susan Dynarski. 2011. "Inequality in Postsecondary Attainment." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane, 117–32. New York: Russell Sage Foundation.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014a. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." National Bureau of Economic Research Working Paper 19843.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014b. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." National Bureau of Economic Research Working Paper 19844.

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

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

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91 (4): 766–72.

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

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

[4] There is a strong cross-sectional correlation across areas of the United States 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.