Intergenerational mobility: Theory, measurement, and empirics

Heidi L. Williams
MIT 14.662
Spring 2015

<u>Outline</u>

(1) Preliminaries

(2) Theory

(3) Measurement: Haider and Solon (2006), Solon (1992)

(4) Empirics: Black, Devereux and Salvanes (2005), Sacerdote (2007)

(5) Within-US geography of intergenerational mobility: Chetty et al. (2014)

1 Preliminaries

In the time since Pareto first documented skewed distributions of incomes and wealth, economists have focused attention on inequality in the distributions of earnings, income, and wealth. However, for a long time economists paid relatively more attention to inequality within generations relative to inequality within families across generations.[1]

Solon (1999)'s chapter in the *Handbook of Labor Economics* starts with the following thought experiment. Imagine two societies: society A and society B. The distribution of earnings is identical in the two societies, so in a within-generation sense the two societies are "equally unequal." But now suppose that in society A, one's relative position in the earnings distribution is exactly inherited from one's parents: if your parents were in the 90^{th} percentile of earnings in their generation it is certain that you place in the 90^{th} percentile of your own generation; if your parents were in the 5^{th} percentile in their generation you inevitably place in the 5^{th} percentile. In contrast, in society B one's relative position in the earnings distribution is completely independent of the position of one's parents: the offspring of parents in the 5^{th} percentile and the offspring of parents in the 90^{th} percentile show the same distribution of earnings. We say that society B displays high (complete) intergenerational mobility, whereas society A does not. Opinions differ about the fairness of any given society's degree of mobility, and what (if anything) should be done from a policy perspective. The sources of the intergenerational correlation in earnings may also matter for policy.

[1]The latter topic was, earlier, more of a focus of research in sociology; see for example Blau and Duncan's 1967 book *The American Occupational Structure*.

A key theoretical model in the economics literature on intergenerational mobility is the model by Becker and Tomes (1979); we will start by presenting a simplified version of this model that follows Solon (1999). We will then discuss issues that arise in estimating intergenerational mobility: over the last several decades, a tremendous amount of progress has been made in measurement, and we will review some of the key advances. Finally, in addition to empirical papers that have developed estimates of intergenerational mobility, in recent years the literature in this area has been placing increased emphasis on the causal mechanisms that underlie this relationship. Following Black and Devereux (2011)'s *Handbook of Labor Economics* chapter, we will describe several recent empirical papers in this area.

2 Theory

Two workhorse models of intergenerational mobility are those of Galton (1877) and Becker and Tomes (1979). We will start by reviewing the Becker and Tomes (1979) model, reviewing Goldberger (1989)'s critique of the Becker-Tomes model, and briefly summarizing Mulligan (1999)'s comparison of empirical support for the predictions of the Galton model and the Becker-Tomes model.

The intergenerational earnings elasticity is arguably a parameter of inherent interest, so you may be wondering what the value added of a theoretical model is in this context. The Becker-Tomes model is (not surprisingly) built on a foundation of utility-maximizing behavior: parents choose how to allocate resources between consumption and investment in their children. If some endowments (*e.g.* genetic endowments) are automatically transmitted between parents and children, but the intergenerational earnings elasticity depends not only on these endowments but also on parents' decisions about what share of their income to invest in their children, then we would like to have a theoretical model that generates predictions on what the effects of changes in policy will be on parents' decisions and in turn on childrens' outcomes. As we will see, the Becker-Tomes framework has strong predictions on whether public policies (such as the provision of public education or Head Start programs) will be effective in reducing inequality given that parents may change their investments in response to public subsidies. Many of the strong predictions that come out of the Becker-Tomes model are an artifact of specific assumptions that are not empirically tested in their paper, and relatively small changes in these assumptions lead to dramatically different predictions. My goal in walking through the Becker-Tomes model is not to give you a sense that this is the "right" model – indeed, as we will discuss, subsequent research has in general not provided strong empirical support for the predictions of the Becker-Tomes framework – but rather to give you a sense of why economic theory can be useful in this area.

2.1 Becker and Tomes (1979)

This simplified version of Becker and Tomes (1979) follows Solon (1999). A family consisting of one parent (generation $t-1$) and one child (generation t) must allocate the parent's lifetime earnings y_{t-1} between the parent's own consumption C_{t-1} and investment I_{t-1} in the child's earnings capacity. The budget constraint is:

$$y_{t-1} = C_{t-1} + I_{t-1} \tag{1}$$

Note that this assumes that parents can't borrow to invest in their children: there are imperfect capital markets. The technology which translates investment I_{t-1} into the child's lifetime earnings y_t is:

$$y_t = (1+r)I_{t-1} + E_t \tag{2}$$

where r is a return to parents' human capital investment I_{t-1} and E_t represents the combined effect of all other determinants of the child's lifetime earnings. The parent divides y_{t-1} between C_{t-1} and I_{t-1} to maximize a Cobb-Douglas utility function which takes as arguments their consumption C_{t-1} and their child's lifetime earnings y_t. Substituting in the above expressions for C_{t-1} (from the budget constraint) and y_t gives the following:

$$\begin{aligned} U &= (1-\alpha)\log(C_{t-1}) + \alpha\log(y_t) && (3)\\ &= (1-\alpha)\log(y_{t-1} - I_{t-1}) + \alpha\log((1+r)I_{t-1} + E_t) && (4) \end{aligned}$$

where knowledge of E_t is assumed and the parameter $\alpha \in (0,1)$ indexes the parent's taste for y_t relative to C_{t-1}. Solving for the first-order condition and re-arranging terms, we can write the optimal choice of I_{t-1} as:

$$\begin{aligned} \frac{\partial U}{\partial I_{t-1}} &= 0 && (5)\\ \Rightarrow I_{t-1} &= \alpha y_{t-1} - \frac{(1-\alpha)E_t}{1+r} && (6) \end{aligned}$$

Once we know the level of investment that parents will choose given their earnings, we can solve for the child's earnings as a function of her parent's earnings as follows. Substituting the expression for I_{t-1} into the equation for the child's lifetime earnings y_t, we can solve for y_t as:

$$\begin{aligned} y_t &= (1+r)I_{t-1} + E_t && (7)\\ &= (1+r)\left(\alpha y_{t-1} - \frac{(1-\alpha)E_t}{1+r}\right) + E_t && (8)\\ &= \beta y_{t-1} + \alpha E_t && (9) \end{aligned}$$

where $\beta = (1+r)\alpha$.

If the variance of earnings is the same in each generation and if E_t is orthogonal to y_{t-1}, then β will measure the correlation between the child's and the parent's lifetime earnings. However, Becker and Tomes argue this condition will generally not hold. They decompose E_t as:

$$E_t = e_t + u_t \tag{10}$$

where e_t is the child's 'endowment' of earnings capacity (aside from the part resulting from the parent's conscious investment I_{t-1}) and u_t is the child's 'market luck,' assumed to be independent of y_{t-1} and e_t. The endowment e_t represents the combined effect of many child attributes influenced by nature, nurture, or both. Becker and Tomes describe e_t as *"determined by the reputation and 'connections' of their families, the contribution to the ability, race, and other characteristics of children from the genetic constitutions of their families, and the learning, skills, goals, and other 'family commodities' acquired through belonging to a particular family culture."* Given this characterization of the endowment, Becker and Tomes assume that the child's endowment e_t is positively correlated with the parent's endowment e_{t-1}; in particular, they assume e_t follows a first-order autoregressive process:

$$e_t = \lambda e_{t-1} + v_t \tag{11}$$

where $0 \leq \lambda < 1$ and v_t is serially uncorrelated with variance σ_v^2. Solon here switches to suppress intercepts by expressing all variables in deviation-from-mean form. As long as $\lambda > 0$, E_t will be positively correlated with y_{t-1} because both depend on the parent's endowment e_{t-1}; in this case, the intergenerational earning correlation is not simply β. Substituting from the equations outlined above, we have:

$$\begin{aligned} y_t &= \beta y_{t-1} + \alpha E_t && (12)\\ &= \beta y_{t-1} + \alpha(e_t + u_t) && (13)\\ &= \beta y_{t-1} + \alpha e_t + \alpha u_t && (14) \end{aligned}$$

Becker and Tomes make a series of assumptions that assure stationarity in the process for y: $0 < \beta < 1$, the population variance of e_t is $\sigma_e^2 = \frac{\sigma_v^2}{1-\lambda^2}$ for all t, and the population variance of u_t is σ_u^2 for all t. They then derive an expression for the intergenerational earnings correlation generated by this model:

$$corr(y_t, y_{t-1}) = \delta\beta + (1-\delta)\frac{\beta+\lambda}{1+\beta\lambda} \tag{15}$$

where $\delta = \frac{\alpha^2\sigma_u^2}{(1-\beta^2)\sigma_y^2}$ is the proportion of the variance in y originating from innovations in the u series rather than in the v series. I won't work through this derivation in detail here, but working through the math - particularly for some special cases - is helpful in building intuition for this expression.

Solon (1999) highlights several limitations of the Becker and Tomes (1979) framework. First, it ignores the intergenerational transmission of assets other than human capital; this distinction is explored in Becker and Tomes (1986). Second, the model of course relies on functional form assumptions, such as the specific Cobb-Douglas form of the utility function. Third, by assuming single-parent families, it ignores the role of assortative mating in intergenerational mobility. Fourth, by assuming single-child families, it ignores the role of division of family resources among multiple children. However, despite these limitations, the model is very useful in illustrating several key aspects of the intergenerational transition of earnings. The model clarifies that we should expect intergenerational transmission to occur through multiple processes: investment in human capital, parental earnings, and the child's endowed capacity (which is influenced via some combination of nature and nurture by the parent's endowment). These processes depend on a number of other parameters, such as the returns to human capital and the relative magnitudes of variances in market luck and endowment luck. Loury (1981) presents a related model in which he explores education-specific tax policies in a similar framework.

Becker and Tomes use this model to examine a number of comparative statics, a main theme of which is to highlight the potential for "offsetting effects." For example, their model predicts that public education and other programs to aid the young (*e.g.* Head Start) may not significantly reduce inequality (and in theory, could widen inequality) if there are compensating decreases in parental expenditures. However, this type of prediction is quite fragile - if the effect of luck on child's income were multiplicative rather than additive, then no offsetting would occur. Intuitively, what matters is whether public subsidies through programs such as Head Start are complements or substitutes with parents' investments (as well as assumptions about capital market imperfections).

Gelber and Isen (2011) provide an empirical analysis of this offset prediction of the Becker-Tomes model. Using data from the Head Start Impact Study - a randomized study of Head Start - they find that in response to children's Head Start access, parents are substantially and statistically significantly *more* involved with their children along a wide variety of dimensions. For example, parents read to their children more often, and for a longer amount of time, when their children have access to Head Start than when they do not. This increased investment appears to persistent even when the children are no longer attending Head Start. Table 3 summarizes their main estimates. These results are inconsistent with the prediction of the Becker-Tomes model that an increase in publicly-provided Head Start should cause a decrease in parental investment in children.

Table 3. *Effect of HS on particular parent involvement outcomes. The table shows coefficients and standard errors on the treatment dummy from probit, ordered probit, or IV regressions of parent involvement on HS enrollment. Dependent variable: measures of parent involvement (listed in column headings)*

	(1) Number of Times Read	(2) How long read	(3) Minutes reading	(4) Days with father	(5) Practiced math	(6) Visited art gallery	(7) Track child's learning	(8) Learning materials available
Panel A: During								
HS enrollment	0.21	2.78	18.71	0.91	0.26	0.12	0.37	0.10
	(0.09)***	(0.82)***	(4.79)***	(0.85)	(0.09)***	(0.07)*	(0.09)***	(0.07)
R-squared	--	0.00	0.00	--	--	--	--	--
Log-likelihood	-1956648	--	--	-249374	-1161974	-744096	-548854	-5679
N	7257	7211	7232	1729	3678	7257	3693	3097
Panel B: After								
HS enrollment	0.01	1.46	7.04	1.38	0.15	0.03	--	0.09
	(0.06)	(0.81)*	(4.78)	(0.69)***	(0.07)*	(0.06)		(0.06)
R-squared	--	0.00	0.00	--	--	--	--	--
Log-likelihood	-2003609	--	--	0.00	-1092651	-857878	--	-10212
N	7075	7035	7051	3336	3539	7072	--	5908

Notes: The table shows the results of regressions in which measures of parent involvement are related to HS enrollment or access. The dependent variable in question is listed in each column heading. The regression is a probit in Columns 5 and 6; an ordered probit in Columns 1, 4, and 8; and two-stage least squares in Columns 2, 3, and 7. In Columns 1, 4, 5, 6, and 8, we form the Wald estimate by dividing the coefficient estimate by the first stage (0.68). The table shows coefficient estimates and standard errors; since the regressions are estimated using different methods, these coefficient estimates must be interpreted accordingly. Panel A shows results for the During period, while children are potentially enrolled in HS, and Panel B shows results for the period after which children are potentially enrolled. The standard deviation of the dependent variable in the During (After) period is 15.97 (16.13), 103.0 (101.3), and 9.53 (8.41) in Columns 2, 3, and 4 respectively. Standard errors are clustered at the level of the program. All observations are weighted by the final parent weights. In Column 1, the dependent variable is the number of times a parent read to the child per week (ordered in categories). In Column 2, the dependent variable is how long in minutes the parent read to the child at each sitting. In Column 3, the dependent variable is how many minutes per week the parent reads to the child, constructed by multiplying the dependent variables from Columns 1 and 2. The sample size is slightly bigger for the constructed variable in Column 3 than in Column 2 because those occasional individuals who report that they do not read to their child (as recorded in Column 1) do not answer the question about how many minutes per sitting they read to their child (in Column 2), but we infer in constructing the dependent variable in Column 3 that they read to their child 0 minutes per week. In Column 4, the dependent variable is "In the past month, on about how many days has [CHILD] seen (his/her) father." In Column 5, the dependent variable is "Use dance or act out stories to practice math ideas such as numbers or size." In Column 6, the dependent variable is "Visited an art gallery, museum, or historical site." In Column 7, the dependent variable is "Track how child learns and grows by keeping notes about (his/her) behavior or progress"; this variable is not available in the After period. In Column 8, the dependent variable is "A variety of learning materials are available," which is reported by the HSIS interviewer. The results are extremely similar when controlling for the covariates included in Columns 2 and 5 of Table 2. *** denotes significance at the 1% level; ** at the 5% level; and * at the 10% level.

2.2 Galton (1877) and Mulligan (1999)

Goldberger (1989) famously critiqued the Becker-Tomes model and related work. Although his critique is nuanced and touches on many points, one of his major criticisms was to ask whether the Becker-Tomes model has added value relative to an existing, 'non-optimizing' model by Francis Galton (Galton, 1877).

The key idea of the Galton model is that a person's characteristics are positively correlated with those of her parents, but also "regress to mediocrity" so that on average the personal characteristics of a child are less extreme than those of her parents. Goldberger (1989) represents Galton's (Galton, 1877) model as:

$$e_t \quad = \quad (1-c)k + ce_{t-1} + v_t \tag{16}$$

where e_t is the child's height, e_{t-1} is her parent's height, k is the population mean height, c is the inheritability parameter, and v_t is the disturbance (assumed to be independent of past e's). In subsequent work, Galton collected data on the heights of thousands of adults and their parents, cross-tabulated them, plotted the regression of child height on parental height, and found the regression slope to be $\frac{2}{3}$ (where parental height was measured as an average of maternal and paternal height: *"The deviates of the children are to those of the midparent as 2 to 3"*).

In Mulligan (1999) and in previous work, Mulligan clarifies (building in part on Goldberger's critique) that the Galton and Becker-Tomes models indeed give many similar predictions: for example, economic status regresses to the mean across generations in both models. Without adding additional auxiliary assumptions, it is difficult to distinguish predictions of the two models. Mulligan describes five auxiliary assumptions that can be added to the Becker-Tomes model (for example, that preferences do not vary "too much" across families and that "few enough" families are borrowing constrained) that generate empirically testable distinct implications of the Becker-Tomes model. Based on evidence from the existing literature as well as his own estimates in Mulligan (1999), he finds that these distinct implications receive very limited empirical support. Given the lack of strong empirical support for the Becker-Tomes framework, Mulligan concludes that the challenge facing economists is to produce a model of intergenerational mobility with predictions that are (a) distinct from Galton's, and (b) true.

3 Measurement

Black and Devereux (2011) lay out a basic framework for empirical estimation of intergenerational mobility. Consider the benchmark regression:

$$log(Y_1) \quad = \quad \alpha + \beta log(Y_0) + \varepsilon \tag{17}$$

Using lowercase for logs, and taking deviations from population means to remove the intercept, we can re-write this regression equation as:

$$y_1 = \beta y_0 + e \tag{18}$$

where subscript 1 refers to the child, subscript 0 refers to the parent, and y is a measure of permanent earnings. The parameter β is the intergenerational earnings elasticity, and $(1 - \beta)$ is a measure of intergenerational mobility.

The intergenerational correlation ρ is an alternative to the elasticity that has also been widely used in the literature. The correlation between the log earnings of parent and child equals the elasticity provided that the standard deviation of log earnings is the same for both generations:

$$\rho_{y_0,y_1} = \frac{cov(y_0, y_1)}{\sigma_0 \sigma_1} \tag{19}$$

$$= \frac{cov(y_0, y_1)}{\sigma_0^2} \cdot \sigma_0^2 \cdot \frac{1}{\sigma_0 \sigma_1} \tag{20}$$

$$= \beta \frac{\sigma_0}{\sigma_1} \tag{21}$$

where σ is the standard deviation of log earnings. The correlation therefore factors out the cross-sectional dispersion of earnings in the two generations.

In practice, a number of issues arise when attempting to estimate intergenerational mobility. We have defined y as a measure of permanent earnings, yet we do not observe permanent earnings in many (or any) datasets. The estimate of β will be biased if father's permanent earnings (the right-hand-side variable) are measured with error, but not if the son's earnings are subject to classical measurement error.

Early estimates tended to use earnings in one year for both fathers and sons, and tended to find quite small intergenerational earnings correlations for US data - 0.2 or less. From this, many commentators concluded that the US was a highly mobile society. Solon (1992) investigated whether measurement error was responsible for these low estimates using data from the Panel Study of Income Dynamics (PSID), a nationally representative sample with panel data on wages and income for father-son pairs.

Table 2 illustrates that estimates of ρ are much smaller when estimated based on one year of data relative to when ρ is estimated on two- to five-year averages.

TABLE 2—OLS ESTIMATES OF ρ FROM LOG EARNINGS DATA

Year of father's log earnings	Measure of father's log earnings				
	Single-year measure	Two-year average	Three-year average	Four-year average	Five-year average
1967	0.386				
	(0.079)				
	[322]	0.425			
		(0.090)			
1968	0.271	[313]	0.408		
	(0.074)		(0.087)		
	[326]	0.365	[309]	0.413	
		(0.081)		(0.088)	
1969	0.326	[317]	0.369	[301]	0.413
	(0.073)		(0.083)		(0.093)
	[320]	0.342	[309]	0.357	[290]
		(0.078)		(0.088)	
1970	0.285	[312]	0.336	[298]	
	(0.073)		(0.084)		
	[318]	0.290	[301]		
		(0.082)			
1971	0.247	[303]			
	(0.073)				
	[307]				

Notes: Standard-error estimates are in parentheses, and sample sizes are in brackets.

Table 4 reports estimates that use father's years of education as an instrument for father's income. While acknowledging caveats about the exclusion restriction, the results are consistent with the across-year average OLS estimates in suggesting higher estimates of ρ.

TABLE 4—OLS AND IV ESTIMATES OF ρ FOR VARIOUS SINGLE-YEAR INCOME MEASURES IN 1967

Income measure	OLS	IV	Sample size
Log earnings	0.386	0.526	322
	(0.079)	(0.135)	
Log wage	0.294	0.449	316
	(0.052)	(0.095)	
Log family income	0.483	0.530	313
	(0.069)	(0.123)	
Log (family income/poverty line)	0.476	0.563	313
	(0.060)	(0.103)	

Note: Standard-error estimates are in parentheses.

Solon's estimates suggested that the intergenerational earnings correlations in the US were at least 0.4, twice as high as the previous 'consensus' estimates of 0.2. As we discuss below, subsequent research focused on obtaining better estimates of permanent earnings by averaging over even more years of data (to allow for persistent transitory shocks), and by paying careful attention to the ages of both fathers and sons at the time earnings are measured (to address 'lifecycle bias').

This line of research spurred a realization among economists that this was an area where the sociologists had gotten it "right" and the economists had gotten it "wrong." Economists had been estimating very low intergenerational earnings correlations (0.2) whereas sociologists' estimates based on occupation data were more on the order of 0.4-0.5. In retrospect we understand that occupation is a better measure of permanent income.

3.1 Persistent transitory shocks

Mazumder (2005) notes that the use of short-term averages of earnings may be problematic given that the literature on earnings dynamics has suggested that transitory shocks to earnings are persistent, so that averages of earnings taken over 4 to 5 years (as in Solon (1992)) will provide rather poor measures of permanent economic status. He uses a simulation to demonstrate that 5-year averages of fathers' earnings may be expected to yield estimates that are biased down by approximately 30%.

To address this issue, Mazumder uses a new data source - the 1984 Survey of Income and Program Participation (SIPP) matched to the Social Security Administration's Summary Earnings Records (SER) to generate new empirical estimates. Although the matched SIPP-SER data has some disadvantages (such as high rates of censored earnings in some years), it provides longer-term earnings histories for both parents and children with no problem of sample attrition.

Mazumder's estimates in Table 4 suggest that when earnings are averaged over a 15-year period, intergenerational earnings correlations are on the order of 0.6.

TABLE 4.—INTERGENERATIONAL ELASTICITIES USING SER FOR FATHERS' EARNINGS

Fathers	Elasticity (Standard Error) N														
	Sons					Daughters					Pooled				
Log Avg. Earn.	84–85	82–85	79–85	76–85	70–85	84–85	82–85	79–85	76–85	70–85	84–85	82–85	79–85	76–85	70–85
	Father Earnings Must Be Positive Each Year														
Drop	0.253	0.349	0.445	0.553	0.613	0.363	0.425	0.489	0.557	0.570	0.308	0.388	0.470	0.559	0.600
noncovered	(0.043)	(0.059)	(0.079)	(0.099)	(0.096)	(0.065)	(0.087)	(0.110)	(0.140)	(0.159)	(0.039)	(0.052)	(0.067)	(0.084)	(0.093)
fathers	1262	1218	1160	1111	1063	1178	1124	1070	1031	982	2440	2342	2230	2142	2045
Impute	0.289	0.313	0.376	—	—	0.350	0.395	0.422	—	—	0.322	0.358	0.404	—	—
noncovered	(0.050)	(0.052)	(0.062)			(0.062)	(0.081)	(0.096)			(0.039)	(0.048)	(0.056)		
fathers	1485	1462	1433			1360	1339	1310			2845	2801	2743		
Drop	0.273	0.419	0.474	0.533	0.652	0.526	0.563	0.635	0.750	0.754	0.393	0.487	0.553	0.643	0.707
government &	(0.060)	(0.082)	(0.096)	(0.111)	(0.135)	(0.089)	(0.137)	(0.150)	(0.173)	(0.192)	(0.057)	(0.077)	(0.086)	(0.100)	(0.118)
self-employed	844	825	801	779	746	782	758	736	719	690	1626	1583	1537	1498	1436
	Allow Some Years of Zero Father Earnings*														
Drop	0.234	0.334	0.434	—	—	0.312	0.423	0.506	—	—	0.269	0.377	0.472	—	—
noncovered	(0.043)	(0.057)	(0.069)			(0.060)	(0.065)	(0.091)			(0.034)	(0.043)	(0.056)		
fathers	1295	1268	1227			1201	1168	1127			2496	2436	2354		
Impute	0.238	0.342	0.403	—	—	0.295	0.384	0.474	—	—	0.266	0.365	0.441	—	—
noncovered	(0.042)	(0.057)	(0.059)			(0.055)	(0.061)	(0.080)			(0.033)	(0.042)	(0.049)		
fathers	1534	1550	1571			1394	1406	1424			2928	2956	2995		
Drop	0.242	0.355	0.441	0.523	0.575	0.400	0.504	0.600	0.731	0.847	0.304	0.422	0.570	0.622	0.703
government &	(0.059)	(0.080)	(0.084)	(0.101)	(0.109)	(0.084)	(0.083)	(0.113)	(0.130)	(0.145)	(0.046)	(0.061)	(0.073)	(0.081)	(0.087)
self-employed	874	869	862	895	917	803	794	785	825	831	1677	1663	1647	1720	1748

Dependent variable is children's log average earnings, 1995–1998. All results use tobit specification.

Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For fathers, earnings for those identified as noncovered are either dropped or imputed for the years 1979–1985 as indicated. For the years before 1979, no adjustment is attempted. Earnings for topcoded fathers are imputed using March CPS data for 1970 to 1980 and using 1984 SIPP for 1981 to 1985. Standard errors are adjusted for within family correlation when more than one sibling is present.

*Required years of positive earnings are: 1 for 2-year averages; 2 for 4-year averages; 3 for 7-year averages; 7 for 10-year averages; and 11 for 16-year averages.

3.2 Lifecycle bias

The association between current and lifetime income variables change over the life cycle. This issue is important in practice because fathers' earnings are often measured relatively late in life whereas sons' earnings are typically measured at young ages. Haider and Solon (2006) address this issue by regressing log earnings at age a on the log of the present value of lifetime earnings:

$$y_{0a} = \mu_a y_0 + v \tag{22}$$

$$y_{1a} = \lambda_a y_1 + u \tag{23}$$

Assume that the error terms are uncorrelated with each other and with lifetime earnings (a strong assumption). Black and Devereux (2011) characterize the following expression for the probability limit of the intergenerational elasticity:

$$\text{plim } \hat{\beta} = \beta \frac{\lambda_a \mu_a var(y_0)}{\mu_a^2 var(y_0) + var(v)} \tag{24}$$

$$= \beta \lambda_a \theta_a \text{ where } \theta_a = \frac{\mu_a var(y_0)}{\mu_a^2 var(y_0) + var(v)} \tag{25}$$

The key thing to note from the above expression is the the θ_a term is an additional term contributing to the bias in $\hat{\beta}$.

A major innovation of the Haider-Solon paper is to use new data that allows them to observe nearly career-long earnings histories based on 1951-1991 Social Security Administration (SSA) earnings histories of the members of the Health and Retirement Study (HRS) sample. Using this data, Figure 2 presents estimates of how λ_a (the correlation between earnings at age a and lifetime earnings) and θ_a (the additional term measuring the bias in the estimate $\hat{\beta}$) varies over the life cycle.

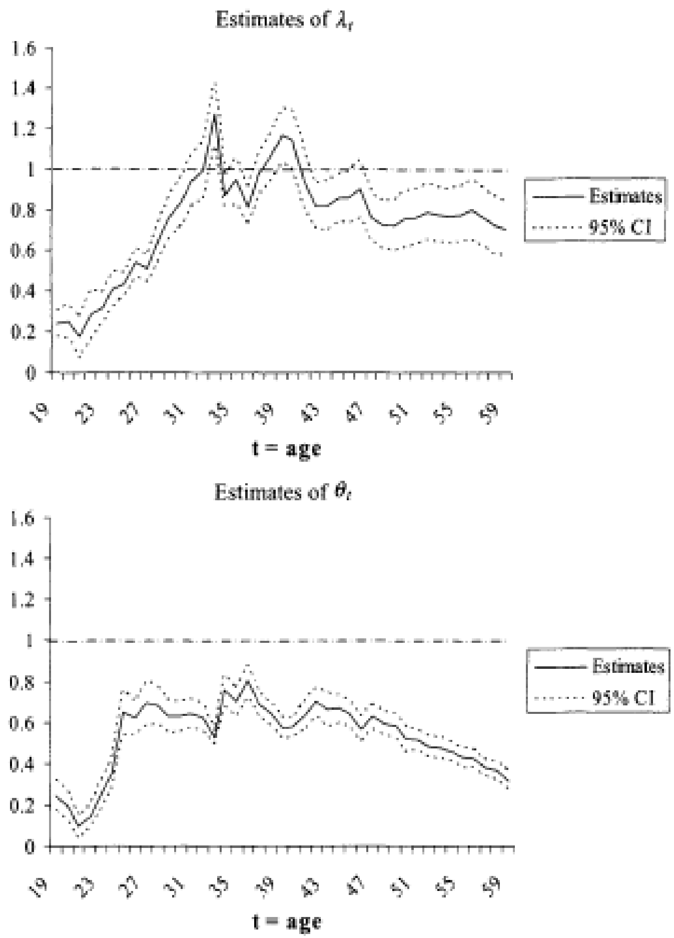


FIGURE 2. MAIN ESTIMATES OF λ_t AND θ_t

Note: The solid lines graph the parameter estimates, and the dotted lines are 1.96 estimated standard errors above and below the solid lines.

Haider and Solon's results suggest that the correlation between current and lifetime earnings is low when men are in their twenties (as low as 0.2 in the US before age 25), and are close to 1 once men reach their 30s - remaining high until their late forties. After the late fifties, estimates of this correlation decline to around 0.6. These results suggest there could be large attenuation bias if earnings of sons age 30 or younger are included in the analysis.

3.3 Cross-country estimates of the intergenerational earnings elasticity

Given a growing number of estimates of the intergenerational earnings elasticity, researchers have begun to analyze how this elasticity varies across countries. Solon (2002) reviews estimates and concludes that the US and the UK appear to be less mobile than are Canada, Finland, and Sweden. Black and Devereux (2011) review some more recent research which has attempted to generate a more comparable set of international estimates (as well as correct for lifecycle bias) which finds similar conclusions: intergenerational earnings elasticities appears to be highest in the US (0.5-0.6), lower in the UK (0.3), and lower still in Nordic countries (less than 0.3).

Solon (2004) develops one model of why the intergenerational earnings elasticity might vary across time and space as a function of governmental investment in children's human capital that may be progressive in the sense that the ratio of government investment to parental income decreases with parental income.

4 Empirics

In their update of Solon (1999)'s chapter in the *Handbook of Labor Economics*, Black and Devereux (2011) note that in recent years researchers' focus has largely shifted away from estimating intergenerational earnings elasticities, and shifted towards trying to better understand the causal mechanisms underlying these correlations. One strand of this research has aimed to quantify the relative importance of genetic influences versus environmental influences versus the interaction of the two in explaining childrens' outcomes. A second strand of this research has aimed to establish the effect of individual parent attributes (such as parents' education and income) on childrens' outcomes. Much of the interest in both of these strands of research stems from a desire to inform public policy: without understanding the mechanisms behind the intergenerational correlations, developing appropriate public policies is difficult.

Black and Devereux (2011) frame their review of this literature in terms of different methodological approaches that researchers have used to shed light on mechanisms. I will follow their discussion, but primarily emphasize two approaches that have seen particularly important advances in recent years:

1. Adoptee studies: particularly Sacerdote (2007) and Bjorklund, Lindahl and Plug (2006)

2. Natural experiment/IV estimates: particularly Black, Devereux and Salvanes (2005)

4.1 Sibling and neighborhood correlations

Sibling correlations in earnings provide one measure of intergenerational influences: positive correlations imply that shared genetic and environmental factors cause siblings to be more similar than two randomly chosen members of society. Solon (1999) reviews this literature and cites a consensus value of 0.4 for the correlation of log earnings between brothers in the US. Black and Devereux (2011) review some more recent estimates from Nordic countries that find lower correlations - on the order of 0.15-0.2 - consistent with the idea that family background factors may be more important in the US than in Nordic countries.[2]

Researchers have also decomposed the sibling correlation in earnings into the intergenerational earnings elasticity and s, where s is a measure of all variables shared by siblings that are unrelated to parental earnings (see Solon (1999) for details on this decomposition). One component of s that has been the focus of empirical studies is neighborhood characteristics. For example, Page and Solon (2003) use the PSID to examine correlations in adult earnings between brothers and between unrelated boys in the same neighborhood. They estimate that the correlation for unrelated boys is 0.16 - about half the size of the brothers correlation. The estimates of neighborhood effects from this type of study are likely an upper bound on true neighborhood effects if other family traits are also correlated within neighborhoods. Like the sibling studies, these neighborhood studies have not aimed to distinguish causal mechanisms.

The (still ongoing) Moving to Opportunity project is a large-scale experiment in Baltimore, Boston, Chicago, LA, and NYC that is providing experimental evidence on neighborhood effects (relevant research papers are posted here: `http://www.nber.org/mtopublic/`). Families were eligible for participation if they had children and resided in public housing or project-based Section 8 assisted housing in a high-poverty census tract. Interested eligible families selected from a waiting list were randomly assigned to one of three program groups:

1. Experimental group: receive a restricted Section 8 voucher that can only be used in a 'low poverty' area (census tract with under a 10 percent poverty rate in 1990)
2. Section 8 group: receive a regular (non-restricted) Section 8 voucher
3. Control group: not offered additional housing assistance

Although results are still coming out from ongoing survey rounds, the initial evidence from Moving to Opportunity has suggested benefits for daughters and adverse effects for sons. Subsequent research by the interdisciplinary Moving to Opportunity team has investigated some potential mechanisms for these gender differences.

4.2 Regression analysis using adoptees

Psychologists and sociologists have long used adoption data as a way to examine the effects of family environment, largely focusing on measuring the heritability of IQ. If (1) adopted children are randomly assigned to families as infants, and (2) adopted children are treated exactly the same as biological children, then adoption can be considered an experimental intervention that randomly assigns children to families. Major contributions of recent economics papers in this area have been to contribute *much* larger samples, to identify contexts where it is possible to exploit quasi-random assignment of children to adoptive families, to investigate a wider range of

[2] Black and Devereux (2011) discuss some more recent research that has examined sibling correlations across different sibling types (identical twins, fraternal twins, full siblings, half siblings, *etc.*) in an attempt to distinguish the roles of nature and nurture.

outcome variables, and to adopt a "treatment effects" framework that relies on fewer assumptions than the traditional behavioral genetics framework.

Three types of empirical approaches have been applied to adoptee data:

1. Bivariate regression approach. A first approach is to estimate a bivariate regression separately for adopted children and for non-adopted siblings of the following form: $y_1 = \alpha + \lambda y_0 + \varepsilon$, where y_1 is a child outcome (say, log earnings), and y_0 is the analogous variable for the adoptive parent. The comparison is then made between the value of λ for adoptees and non-adoptees. If nurture is unimportant, we would expect λ to be 0 for adoptees and positive for non-adoptees (because of the genetic correlation between parent and child). If genetics and endowments in infancy are unimportant, we would expect λ to be positive and equal for adoptees and non-adoptees. Therefore, the relative value of λ for adoptees and non-adoptees gives an indication of the importance of nature versus nurture.

2. Multivariate regression approach. A second approach uses multivariate regressions on a sample of adoptees to attempt to determine which particular parental characteristics matter most: $y_1 = \alpha + \lambda_1 S_0^m + \lambda_2 S_0^f + \lambda_3 Z + \varepsilon$, where S_0^m and S_0^f refer to education of mother and father, and z refers to some other characteristics such as family income and family size. Note that in general this approach cannot be used to identify the causal effects of specific environmental factors on child outcomes because it is impossible to control for enough variables to hold "all else" equal and isolate the causal effect of, say, mothers' education.

3. Combining information on biological and adoptive parents. A third approach uses data on both biological and adoptive parents to estimate regressions on the sample of adopted children such as: $y_1 = \alpha + \lambda_a y_{0a} + \lambda_b y_{0b} + \varepsilon$, where a references adoptive parents and b references biological parents. This model allows a direct comparison of the influence of the characteristics of biological and adoptive parents.

4.2.1 Sacerdote (2007)

Data and research design

Sacerdote (2007) analyzes a new data set of Korean-American adoptees who were quasi-randomly assigned to adoptive families. The adoptees were placed by Holt International Children's Services during 1964-1985, and were quasi-randomly assigned to families conditional on the family being certified by Holt to adopt: Holt uses a queuing (first-come first-served) policy to assign Korean adoptees to families. As a result, assignment of children to families is effectively random condition on the adoptee's cohort and gender. Sacerdote validates this argument by showing empirical evidence that adoptee's pre-treatment characteristics are uncorrelated with adoptive family characteristics.

This data collection was a major undertaking, involving a collaborative effort by Sacerdote and Holt to survey adoptees and their families during 2004-2005. A public-use version of the data is publicly available: http://www.dartmouth.edu/~bsacerdo/holt_adoption_public_use2006.dta. Sacerdote highlights two disadvantages of his data: first, the response rate to the initial survey was low (34%); and second, he relies on parental reports of adult adoptee outcomes. To deal with these issues, Sacerdote and Holt undertook two additional efforts. First, they re-surveyed a sample of the non-respondents and documented that responses were not significantly correlated with child outcomes. Second, they directly surveyed a smaller sample of the adoptees/non-adoptees and found a high degree of correspondence between their responses and their parents' reports.

Empirical framework #1: Variance decomposition

Sacerdote (2007) outlines a standard behavioral genetics model which guides his first empirical approach. Suppose that child outcomes Y are produced by a linear and additive combination of genetic inputs G, shared family environment F and unexplained factors S. Then we can write an outcome like years of education as:

$$Y = G + F + S \tag{26}$$

Note the strong assumptions here: nature (G) and family environment (F) enter linearly and additively. Assume further that G and F are not correlated. Taking the variance of both sides then gives:

$$\sigma_Y^2 = \sigma_G^2 + \sigma_F^2 + \sigma_S^2 \tag{27}$$

Dividing both sides by the variance in the outcome (σ_Y^2) and defining $h^2 = \frac{\sigma_G^2}{\sigma_Y^2}$ (heritability), $c^2 = \frac{\sigma_F^2}{\sigma_Y^2}$ (family environment), and $e^2 = \frac{\sigma_S^2}{\sigma_Y^2}$ (error term) yields the standard behavioral genetics equation:

$$1 = h^2 + c^2 + e^2 \tag{28}$$

That is, the variance of child outcomes is the sum of the variance from genetic inputs, the variance from family environment, and the variance from non-shared environment (the residual).

Empirical framework #2: Treatment effects

Sacerdote also takes a different, "treatment effects" approach in order to investigate the effect of being assigned to particular family types on adoptee outcomes. Interpreting these effects as causal requires that assignment to the treatment group is quasi-random. Sacerdote defines three different types of adoptive families based on their observables:

1. Type one (27% of the sample): highly educated, small families (3 or fewer children) where both parents have four years of college education

2. Type three (12% of the sample): neither parent has four years of college education and there are four or more children in the family

3. Type two (61% of the sample): families not in either of the two extreme groups

Sacerdote calculates the treatment effects from assignment to type one versus three and type two versus three. He does this by taking the set of adoptees in his sample and estimating regressions of the form:

$$E_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 \text{Male}_i + \gamma A_i + \rho C_i + \varepsilon_i \tag{29}$$

where E_i is educational attainment for child i, $T1_i$ is an indicator for being assigned to a type one family, $T2_i$ is an indicator for being assigned to a type two family, A_i is a full set of single year of age indicators (included because, *e.g.*, educational attainment varies with age), and C_i is a full set of cohort indicators (needed for random assignment; defined as the year in which the child initially entered the Holt system). The less educated, larger adoptive families (type three) are the omitted category. The gender indicator is included because adoptive families are sometimes able to request the adoptee's gender (hence, this is needed for random assignment). Because of the quasi-random assignment, β_1 can be interpreted as the causal effect of assignment to a highly educated, small family relative to assignment to a less educated, large family; however, education and family size are not necessarily the relevant channels.

Empirical framework #3: Estimation of transmission coefficients

In part for comparability between his results and the existing literature, Sacerdote calculates transmission coefficients for a variety of outcome variables:

$$E_i = \alpha + \delta_1 E_{M_i} + \beta_3 \text{Male}_i + \gamma A_i + \rho C_i + \varepsilon_i \tag{30}$$

where E_{M_i} is adoptive mother's years of education. The quasi-random assignment ensures that δ_1 is not biased by selection of adoptees into families: this estimate measures the transmission that takes place purely through nurture. Sacerdote then estimates analogous regressions for non-adoptees:

$$E_j = \alpha + \delta_2 E_{M_j} + \beta_3 \text{Male}_j + \gamma A_j + \rho C_j + \varepsilon_j \tag{31}$$

where E_{M_j} represents mother's education (instead of adoptive mother's education). A comparison of δ_1 and δ_2 is an estimate of how much of the transmission of education (or other outcomes) works through nurture, as opposed to through nature and nurture combined.

Descriptive results

Figures 1, 2, and 3 display some of the raw means graphically. Figure 1 shows the probability of graduating from college by family size, separately for the adoptees and non-adoptees. Both groups show a steep decline in college graduation rates associated with each additional child added to the family. Because this fact survives a host of robustness checks, it suggests that either there is a direct impact of family size on educational attainment, or that family size proxies for something important and unobserved about the family.

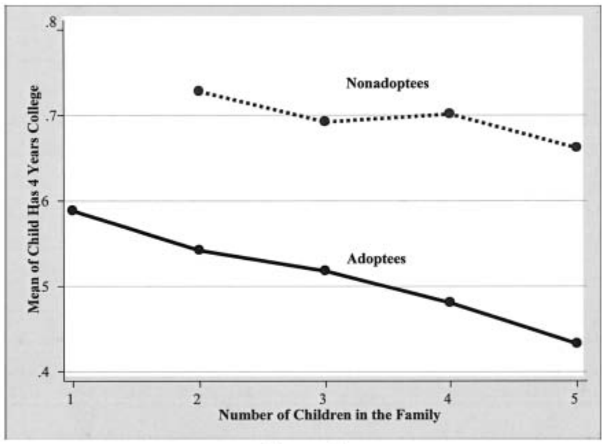


FIGURE I
Mean (College Attendance) By Family Size
Dashed line is for nonadoptees (higher line), solid line is for adoptees.

Figure 2 shows the mean of children's years of education for both the adoptees and non-adoptees for each level of mother's education. There appears to be strong transmission of education from mothers to children, but the upward sloping line is steeper for nonadoptees.

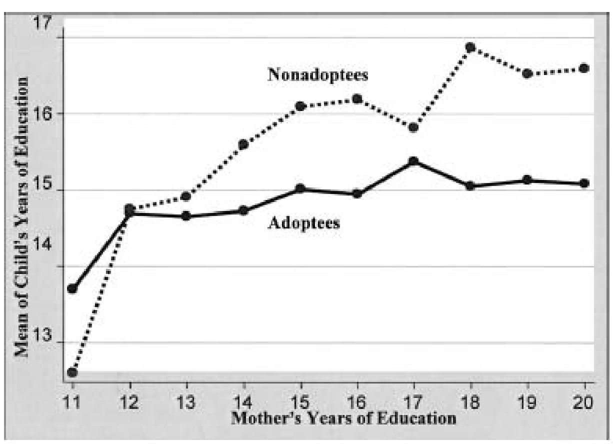


FIGURE II
Mean Child's Years of Education vs. Mother's
Dashed line is for nonadoptees. Solid line is for adoptees.

Figure 3 shows that income follows a different pattern: income transmission is almost non-existent for adoptees but strongly positive for non-adoptees.

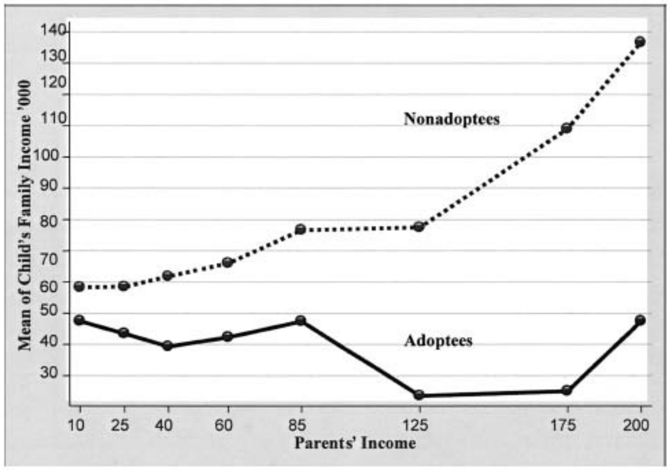


FIGURE III
Mean of Child's Family Income By Parents' Income at Adoption
Dashed line is for nonadoptees (higher line). Solid line is for adoptees.

Variance decomposition

Table 4 and Figure 4 (below) show the correlations in outcomes among sibling pairs after removing age, cohort, and gender effects. For educational attainment, biological siblings have a correlation of 0.34 - 2.4 times larger than the correlation of 0.14 for adoptive siblings. In contrast drinking behavior is almost as correlated for adoptive siblings as for biological siblings. Table 5 uses the behavioral genetics framework to translate these correlations into the percent explained by nature, shared family environment, and the residual.

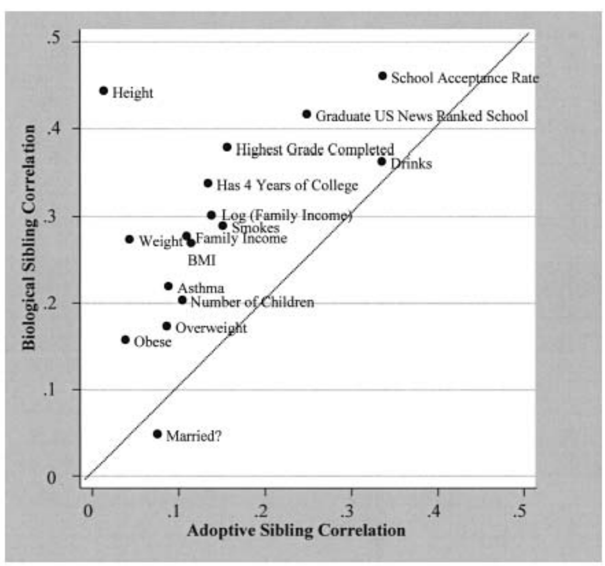


FIGURE IV
Comparison of Adoptive and Nonadoptive Sibling Correlations for Various Outcomes
This graph displays the results in Table IV.

Treatment effects and multiple regression results

With the caveat that it is impossible to definitively separate out root causal mechanisms in this framework, Table 6 presents multiple regression results that attempt to identify which aspects of family environment are most important for adoptees' outcomes. Mother's education and the number of children in the family stand out as having large estimated effects on adoptees' educational attainment.

Table 7 estimates the treatment effects from assignment to specific family types as outlined above. Assignment to a small, highly educated family relative to a lesser educated, large family increases educational attainment by 0.75 years and raises the probability of graduating from college by 16.1 percentage points. The probability of graduating from a US News Ranked college is increased by 23.1 percentage points relative to a mean of 37.3. These are very large estimated effects of family environment. The point estimates for child's family income is also large (11.3 percent increase) but not statistically significant.

TABLE VII

TREATMENT EFFECTS FROM ASSIGNMENT TO HIGH EDUCATION, SMALL FAMILY

	Treatment effect "middle group" of families vs. large, less educated	Treatment effect high education small family vs. large, less educated	Nonadoptees: High education small family vs. large, less educated	Effect from a 1 standard deviation change in family environment index
Child's years of education	0.314 (0.226)	0.749 (0.245)**	2.157 (0.264)**	0.845
Child has 4+ years college	0.060 (0.056)	0.161 (0.057)**	0.317 (0.031)**	0.179
Log child's household income	0.071 (0.081)	0.113 (0.089)	0.210 (0.089)*	0.263
Child four-year college ranked by US News	0.082 (0.052)	0.231 (0.060)**	0.365 (0.052)**	0.224
Acceptance rate of child's college	−0.007 (0.035)	0.016 (0.036)	−0.053 (0.032)	0.098
Child drinks (yes/no)	0.099 (0.050)*	0.178 (0.049)**	0.229 (0.041)**	0.280
Child smokes (yes/no)	0.013 (0.044)	−0.006 (0.048)	−0.075 (0.024)**	0.162
Child's BMI	−0.509 (0.460)	−0.941 (0.468)*	−0.929 (0.498)	1.224
Child overweight	−0.030 (0.047)	−0.077 (0.045)	−0.088 (0.048)	0.121
Child obese	−0.020 (0.023)	−0.044 (0.018)*	−0.037 (0.018)*	0.047
Child has asthma	−0.005 (0.028)	0.013 (0.031)	−0.005 (0.034)	0.085
Number of children	−0.070 (0.099)	−0.199 (0.103)*	−0.580 (0.132)**	0.267
Child is married	0.014 (0.050)	0.000 (0.056)	−0.092 (0.053)	0.123

I split the sample into three groups: High education small families are defined as those with three or fewer children in which both the mother and father have a college degree (Type 1). Twenty-seven percent of adoptees are assigned to such a family. Large lesser educated families are defined as those with four or more children and where neither parent has a college degree (Type 3). Thirteen percent of adoptees are assigned to such a family. The remaining families (which are either small or have a parent with a college degree) are Type 2. Column (1) shows the coefficient on the dummy for assignment to Type 2 relative to Group 3. Column (2) shows the coefficient on the dummy for assignment to Type 1 (small high education) relative to Type 3 (large less educated).

Column (3) shows this Type 1 versus 3 "effect" for the non-adoptees. In a each row, the effects in Columns (1) and (2) are estimated together with a single regression while Column (3) uses a separate regression. Column (4) shows the effect for the adoptees from a one standard deviation move in an index of shared family environment. This is calculated by taking the square root of the variance share explained by shared family environment in the previous table and multiplying by the standard deviation of the outcome variable: that is, $R \times \sigma_y = \sigma_{yhat}$ = predicted effect on the outcome from a one standard deviation change in an index of family environment. Standard errors are corrected for within family correlation (1 cluster by family).

Transmission coefficients

Table 8 shows estimated transmission coefficients. Each additional year of mother's education is associated with 0.09 years of education for adoptees and 0.32 years for non-adoptees; this suggests that roughly 28 percent of measured transmission of education from mothers to children is working directly through nurture. Body mass index and height exhibit no transmission for adoptees, but drinking appears to be transmitted equally well to adoptees and nonadoptees.

TABLE VIII
TRANSMISSION COEFFICIENTS FROM PARENTS TO CHILDREN FOR ADOPTEES AND NONADOPTEES

	(1) Adoptees' Transmission coefficient	(2) Nonadoptees' transmission coefficient
Years of education (mother to child)	0.089 (0.029)a**	0.315 (0.038)**
Has 4+ years college (mother to child)	0.102 (0.034)**	0.302 (0.037)**
Log household income (parents to child)	0.186 (0.111)	0.246 (0.080)**
Height inches (mother to child)	−0.004 (0.034)	0.491 (0.049)**
Is obese (mother to child)	0.003 (0.020)	0.108 (0.034)**
Is overweight (mother to child)	−0.026 (0.029)	0.174 (0.037)**
BMI (mother to child)	0.002 (0.025)	0.221 (0.045)**
Smokes (0–1) (mother to child)	0.132 (0.088)	0.108 (0.115)
Drinks (0–1) (mother to child)	0.210 (0.033)**	0.244 (0.038)**

I regress the child's outcome on the corresponding outcome for the mother (or in the case of income, the parents). Each cell is from a separate regression which also includes age dummies, dummies for year of admission to Holt, and a dummy for the child being male. For income and education regressions I restrict the sample to children ages 25+. For log (income), I attempt to correct for measurement error in parents' income by instrumenting for the survey measure of parents' income using the parents' income measure reported in Holt records.

a Robust standard errors in parentheses: I cluster at the family level.

* significant at 5%;

** significant at 1%.

In his conclusion, Sacerdote observe that the US black-white gap in years of schooling and college completion could - based on his results - be produced by a one standard deviation change in family environment. If black-white family gaps are one standard deviation, they would suffice to explain black-white differences in educational attainment.

4.2.2 Bjorklund, Lindahl and Plug (2006)

Bjorklund, Lindahl and Plug (2006) use administrative data from Statistics Sweden on a large sample of adoptees that includes data on both biological and adoptive parents for each adoptee. Using this data, they can estimate linear models of the form:

$$Y_i^{ac} \quad = \quad \alpha_0 + \alpha_1 Y_j^{bp} + \alpha_2 Y_i^{ap} + v_i^{ac} \tag{32}$$

where j subscripts the family in which the child is born, i subscripts the family in which the child is adopted and raised, and Y represents a characteristic like log income. To interpret α_1 and α_2, it must be that adoptees are randomly assigned to adoptive families; the authors acknowledge that this assumption is often violated in their data and investigate potential biases.

Table 2 reports intergenerational transmission estimates for education and income using a linear model. Panel 1 reports results for own-birth children, and Panel 2 reports results for adoptees.

TABLE II
ESTIMATED TRANSMISSION COEFFICIENTS IN LINEAR MODELS

	Years of schooling			University			Earnings	Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own-birth children								
Bio father	.240**		.170**	.339**		.237**	.235**	.241**
	(.002)		(.002)	(.004)		(.004)	(.005)	(.004)
Bio mother		.243**	.158**		.337**	.246**		
		(.002)	(.002)		(.004)	(.004)		
Adopted children								
Bio father	.113**		.094**	.184**		.148**	.047	.059*
	(.016)		(.016)	(.036)		(.036)	(.034)	(.028)
Bio mother		.132**	.101**		.261**	.229**		
		(.017)	(.017)		(.034)	(.034)		
Adoptive father	.114**		.094**	.165**		.102**	.098**	.172**
	(.013)		(.014)	(.024)		(.026)	(.038)	(.031)
Adoptive mother		.074**	.021		.145**	.097**		
		(.014)	(.015)		(.024)	(.026)		
Sum of estimates for bio and adoptive fathers	.227**		.188**	.349**		.249**	.145**	.231**
	(.019)		(.029)	(.040)		(.059)	(.049)	(.040)
Sum of estimates for bio and adoptive mothers		.207**	.122**		.406**	.326**		
		(.021)	(.016)		(.039)	(.029)		

Standard errors are shown in parentheses; * indicates significance at 5 percent level, and ** at 1 percent level. All specifications include controls for the child's gender, 4 birth cohort dummies for the child, 8 birth cohort dummies for biological/adoptive father/mother, and 25 region dummies of where the biological/adoptive family lived in 1965. The numbers of observations in the second panel for own-birth and adopted children are 94,079/2,125 in columns (1)–(6), 87,079/1,780 in column (7) and 91,932/1,976 in column (8).

The authors draw four conclusions from these results:

1. Biological parents matter
2. Adoptive parents matter
3. On the basis of a comparison of biological and adoptive parents, most of the mother's influence on children takes place through pre-birth factors; for fathers, pre- and post-birth factors appear to be equally important for education, whereas post-birth factors are more important for earnings and income
4. The total impact of adoptive and biological parents' resources on the outcomes of adoptive children is remarkably similar to the impact of the biological parent's outcomes for that of biological children

Where comparable estimates are available, these estimates line up reasonably well with Sacerdote's estimates.

The authors also estimate a model that allows for interactions:

$$Y_i^{ac} = \alpha_0 + \alpha_1 Y_j^{bp} + \alpha_2 Y_i^{ap} \alpha_3 Y_j^{bp} Y_i^{ap} + v_i^{ac} \tag{33}$$

The coefficient α_3 will be positive if children with beneficial birth family backgrounds benefit more from good adoptive family backgrounds.

Table 4 reports nonlinear intergenerational transmission estimates that include the square of parental characteristics for non-adoptees, and include the interaction of birth and adoptive parent characteristics for adoptees. The estimated coefficients on the quadratic terms are positive and statistically significant, suggesting that intergenerational associations are stronger in families with higher education and income. Interacting the adoptive and biological parent measures, the authors find evidence of a positive interaction for mother's education and father's earnings/income, but not for father's education.

TABLE IV
ESTIMATED TRANSMISSION COEFFICIENTS IN NONLINEAR MODELS WITH INTERACTIONS

	Years of schooling				University		Earnings		Income	
	Fathers		Mothers		Fathers	Mothers	Fathers		Fathers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Own-birth children										
Bio parent		−.009		−.058**				−.807**		−.938**
		(.015)		(.017)				(.075)		(.064)
Bio parent squared		.011**		.014**				.069**		.077**
		(.001)		(.001)				(.005)		(.004)
Adopted children										
Bio parent	.050	−.222	−.055	−.472**	.199**	.166**	−.187	−.403	−1.164*	−1.342*
	(.051)	(.127)	(.055)	(.139)	(.045)	(.041)	(.108)	(.502)	(.525)	(.670)
Bio parent squared		.015*		.023**				.017		.015
		(.006)		(.006)				(.037)		(.034)
Adoptive parent	.061	−.003	−.097	−.310**	.170**	.108**	−.293*	−.076	−.995*	−.998
	(.043)	(.090)	(.050)	(.121)	(.025)	(.026)	(.125)	(.648)	(.501)	(.710)
Adoptive parent squared		.004		.012*				−.003		.003
		(.004)		(.005)				(.043)		(.035)
Bio parent * Adoptive parent	.006	.003	.018**	.013*	−.041	.286**	.043**	.034**	.156*	.151*
	(.004)	(.005)	(.005)	(.005)	(.074)	(.071)	(.015)	(.010)	(.067)	(.068)

Standard errors are shown in parentheses; * indicates significance at 5 percent level, and ** at 1 percent level. All specifications include controls for the child's gender, 4 birth cohort dummies for the child, 8 birth cohort dummies for biological/adoptive father/mother, and 25 region dummies of where the biological/adoptive family lived in 1965. The numbers of observations in the second panel for own-birth and adopted children are 94,079/2,125 in columns (1)–(6), 87,079/1,780 in column (7), and 91,932/1,976 in column (8).

4.3 Natural experiment/instrumental variable estimates

One way to estimate causal effects of specific channels underlying intergenerational transmission is to identify variation in *e.g.* parental education or income that is plausibly unrelated to other parental characteristics. A number of papers have looked at how income provided through welfare programs affect children's outcomes; see Almond and Currie (2011) for a review of this literature. Other papers have focused on the causal relationship between patents' education and childrens' education. We here focus on Black, Devereux and Salvanes (2005) as one recent example; Black and Devereux (2011) review other related papers in this literature.

4.3.1 Black, Devereux and Salvanes (2005)

Parents with higher levels of education have children with higher levels of education. Why is this? There are a number of potential explanations, including selection (the type of parents who obtain more education has the type of child who will do well) and causation (obtaining more education makes one a different type of parent, and leads to children having higher educational outcomes). Black, Devereux and Salvanes (2005) examine this question in the context of a (drastic) change in compulsory schooling laws in Norway in the 1960s.

Pre-reform, children were required to attend school through seventh grade; after the reform, this was extended to ninth grade. Implementation of the reform occurred in different municipalities at different times, providing regional as well as time-series variation. Using registry data from Norway, the authors use this reform as an instrument to examine the causal relationship between parent's education and children's education. Their empirical model is summarized by the following two equations:

$$S_1 = \beta_0 + \beta_1 S_0 + \beta_2 \text{AGE}_1 + \beta_3 \text{AGE}_0 + \beta_4 \text{MUNICIPALITY}_0 + \varepsilon \tag{34}$$

$$S_0 = \alpha_0 + \alpha_1 \text{REFORM}_0 + \alpha_2 \text{AGE}_1 + \alpha_3 \text{AGE}_0 + \alpha_4 \text{MUNICIPALITY}_0 + \nu \tag{35}$$

where S is the number of years of education, AGE refers to a full set of age indicators, MUNICIPALITY refers to a full set of municipality indicators, and REFORM equals 1 if the individual was affected by the education reform (0 otherwise). Subscript 0 refers to the parent, and subscript 1 refers to the child. The authors estimate the model separately by parent gender and child gender using 2SLS, where the second equation is the first stage with REFORM_0 serving as an instrument for S_0.

As had been documented previously, this Norwegian reform resulted in a significant change in the educational attainment of individuals at the bottom of the educational distribution. Table 2 shows the distribution of education averaged over the two years prior to the reform ant the two years immediately following the reform, and Table 3a shows the first stage estimates.

TABLE 2—DISTRIBUTION OF EDUCATION TWO YEARS BEFORE AND AFTER THE REFORM

Years of education	Before	After
7	3.5%	1.2%
8	8.9%	1.6%
9	3.4%	12.9%
10	29.6%	26.6%
11	8.5%	8.8%
12	17.2%	19.1%
13	6.7%	6.7%
14	5.4%	5.8%
15	2.7%	3.4%
16+	14.2%	14.1%
N	89,320	92,227

Notes: Before indicates education distribution of cohorts in the two years prior to the reform, while After indicates the distribution of those two years post reform. Note that because the reform occurred in different municipalities at different times, the actual year of the reform varies by municipality.

TABLE 3A—FIRST-STAGE RESULTS

	Full sample of parents		Parents' education <10 years	
	Mother's education	Father's education	Mother's education	Father's education
All	0.142*	0.192*	0.749*	0.795*
	(.029)	(.042)	(.017)	(.024)
Sons	0.127*	0.196*	0.742*	0.814*
	(.035)	(.051)	(.019)	(.029)
Daughters	0.161*	0.197*	0.755*	0.779*
	(.036)	(.050)	(.019)	(.027)

Notes: Each estimate represents the coefficient from a different regression. Robust standard errors in parentheses. First stage also includes dummies for parent's age, parent's municipality, and child's age.

* Significant at 5-percent level.

Table 3 shows that, as expected, OLS estimates find a positive relationship between parents' and childrens' education. The 2SLS estimates are imprecisely estimated and statistically insignificant, in part because the first stage is relatively weak in the full sample. In order to focus on the part of the distribution where the reform had bite, the authors show estimates in columns (3) and (4) that restrict the sample to parents with nine or fewer years of education. The OLS and 2SLS point estimates are quite similar to the full sample estimates in columns (1) and (2), but the 2SLS estimates are much more precise.

TABLE 3—RELATIONSHIP BETWEEN PARENTS' AND CHILDREN'S EDUCATION

	Dependent variable: Children's education			
	Full sample		Parent's education <10	
	OLS	IV	OLS	IV
Mother–all	0.237*	0.076	0.211*	0.122*
	(0.003)	(0.139)	(0.017)	(0.043)
	N = 143,579		N = 39,605	
Mother–son	0.212*	0.199	0.197*	0.176*
	(0.004)	(0.185)	(0.021)	(0.054)
	N = 73,663		N = 20,135	
Mother–daughter	0.264*	−0.029	0.225*	0.066
	(0.004)	(0.186)	(0.023)	(0.063)
	N = 69,916		N = 19,470	
Father–all	0.217*	0.030	0.200*	0.041
	(0.003)	(0.132)	(0.021)	(0.062)
	N = 96,275		N = 22,148	
Father–son	0.209*	0.029	0.151*	0.008
	(0.004)	(0.171)	(0.027)	(0.071)
	N = 49,492		N = 11,235	
Father–daughter	0.226*	0.022	0.244*	0.081
	(0.004)	(0.186)	(0.033)	(0.094)
	N = 46,783		N = 10,913	

Notes: Sample includes children aged 25–35. Robust standard errors in parentheses. Each estimate represents the coefficient from a different regression. All specifications include dummies for parent's age, parent's municipality and child's age.

* Significant at 5-percent level.

For fathers, the estimates are all close to zero and statistically insignificant. For mothers, there is a positive effect of maternal education on the education of sons, but no such relationship for daughters. The authors' interpretation of these results is that the positive correlation between parents' education and children's education largely represents positive relationships between other factors that are correlated with education (such as ability, family background, or income): the true causal effect of parental education on child education appears to be weak.

Figure 1 presents a visual representation of the estimates for the restricted sample.

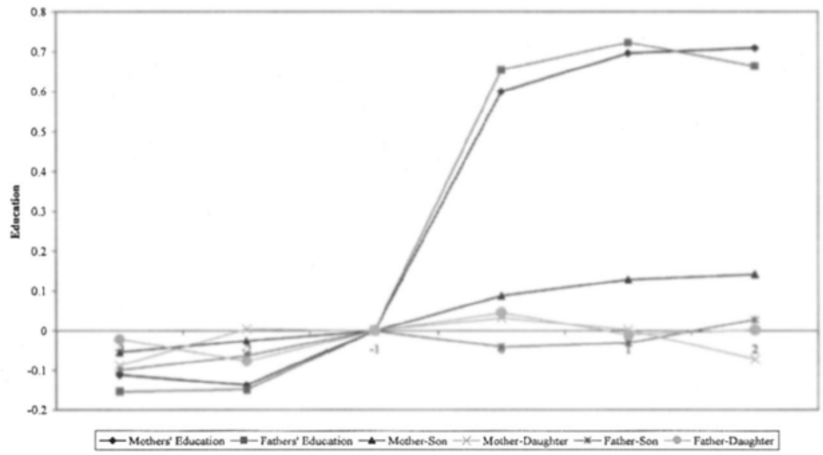


FIGURE 1. EFFECT OF NORWEGIAN EDUCATION REFORM ON EDUCATION FIRST STAGE (EFFECT ON PARENTS) REDUCED FORM (EFFECT ON CHILDREN)

Notes: Estimated on the restricted sample. Lines represent average education for each group with cohort and municipality effects taken out; time zero represents the year of the reform.

Although not included in the final version of the paper, in the conclusion the authors note they find no evidence that women who received more education due to the reform married better educated or wealthier men, nor that these women have fewer children. Overall, the authors conclude that their results provide limited support for intergenerational spillovers as a compelling argument for compulsory schooling laws.

4.3.2 Parental education and infant health

Several studies have examined the effect of parental education on infant health. While only studying infant outcomes, as we will discuss next week these estimates can still be relevant to intergenerational transmission because several studies have suggested better infant health has a positive causal effect on later adult outcomes.

McCrary and Royer (2011) use a regression discontinuity design centered around school entry dates in California and Texas. Women born just before the school entry date start school a year earlier than women born just after the school entry date, and subsequently end up with higher education on average. Using this variation in an instrumental variables framework, McCrary and Royer find no effect of education on fertility or age-at-first-birth, and estimate very small and statistically insignificant effects on infant health as proxied by birth weight.

Currie and Moretti (2003) use an instrumental variables strategy that relies on variation at a higher level of the educational distribution than does the McCrary and Royer (2011) paper, and find evidence that higher maternal education induced by college openings reduces fertility and improves infant health as proxied by birth weight. Taken at face value, the results of these two studies suggest there may be important heterogeneity in the effects of maternal education on child outcomes.

5 Within-US geography of intergenerational mobility: Chetty et al. (2014)

Chetty et al. (2014) investigate intergenerational mobility for the 1980-1982 birth cohorts in the United States.

5.1 Data

1. Sample definition: The baseline sample includes U.S. citizens in the 1980-1982 birth cohorts. The authors identify parents of a child as the first tax filers who claim the child as a child dependent and were between the ages of 15 and 40 when the child was born. Note that if the child was first claimed by a single filer, the child is defined as having a single parent, which is how they measure family stability.

2. Income measures and intermediate outcomes: Parents' income is collected from federal income tax records. The definition of income includes labor earnings and capital income as well as unemployment insurance, Social Security, and disability benefits reported to the IRS. Family income is averaged over the five years from 1996 to 2000 to obtain a proxy for parents' lifetime income, when the children are between the ages of 15 and 20. Child income is defined similarly, and averaged over the last two years of available data (2011 and 2012) when children are in their early thirties. Besides children's income, the authors collect data on college attendance, college quality, and teenage births.

   Appendix Table III. Appendix Table III presents summary statistics for the parents and the children in the baseline sample.

ONLINE APPENDIX TABLE III
Summary Statistics for Core Sample: Children Born in 1980-82

Variable	Mean (1)	Std. Dev. (2)	Median (3)
Parents:			
Family Income (1996-2000 average)	87,219	353,430	60,129
Top Earner's Income (1999-2003 average)	68,854	830,487	48,134
Fraction Single Parents	30.6%	46.1%	
Fraction Female among Single Parents	72.0%	44.9%	
Father's Age at Child Birth	28.5	6.2	28
Mother's Age at Child Birth	26.1	5.2	26
Father's Age in 1996	43.5	6.3	43
Mother's Age in 1996	41.1	5.2	41
Children:			
Family Income (2011-12 average)	48,050	93,182	34,975
Fraction with Zero Family Income	6.1%	23.9%	
Individual Income	31,441	112,394	24,931
Individual Earnings	30,345	98,692	23,811
Fraction Female	50.0%	50.0%	
Fraction Single	44.3%	49.7%	
Attend College between 18-21	58.9%	49.2%	
Fraction of Females with Teen Birth	15.8%	36.5%	
Child's Age in 2011	30.0	0.8	30
Number of Children		9,867,736	

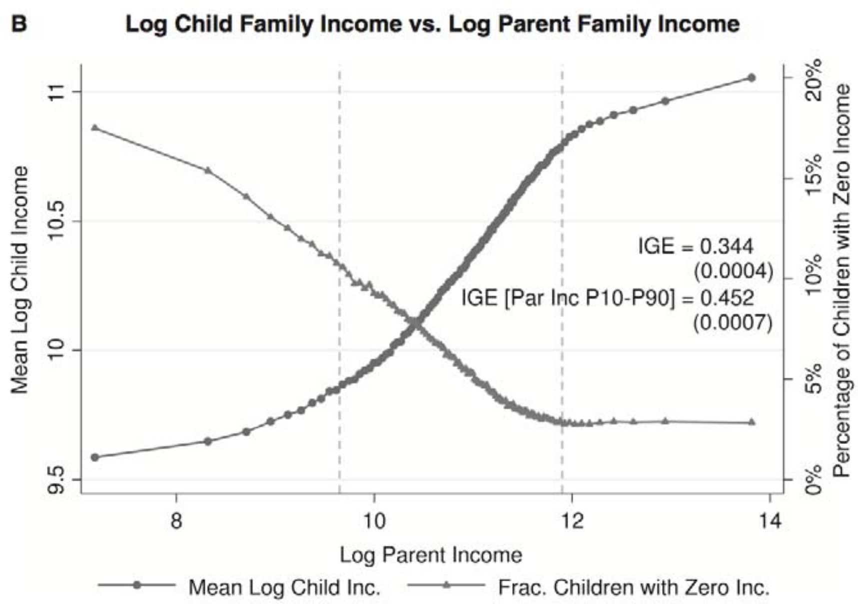
5.2 National level analysis

The authors estimate the intergenerational income elasticity (IGE) as:

$$IGE = \rho_{XY} \frac{SD(\log Y_i)}{SD(\log X_i)}, \tag{36}$$

where ρ_{XY} is the correlation between log child income and parents' income and $SD()$ denotes the standard deviation.

Figure I Panel B. Figure I Panel B plots the relationship between log child income and log parent income. Note that the log-log specification discards observations with zero income, which accounts for as high as 17% of the poorest families.

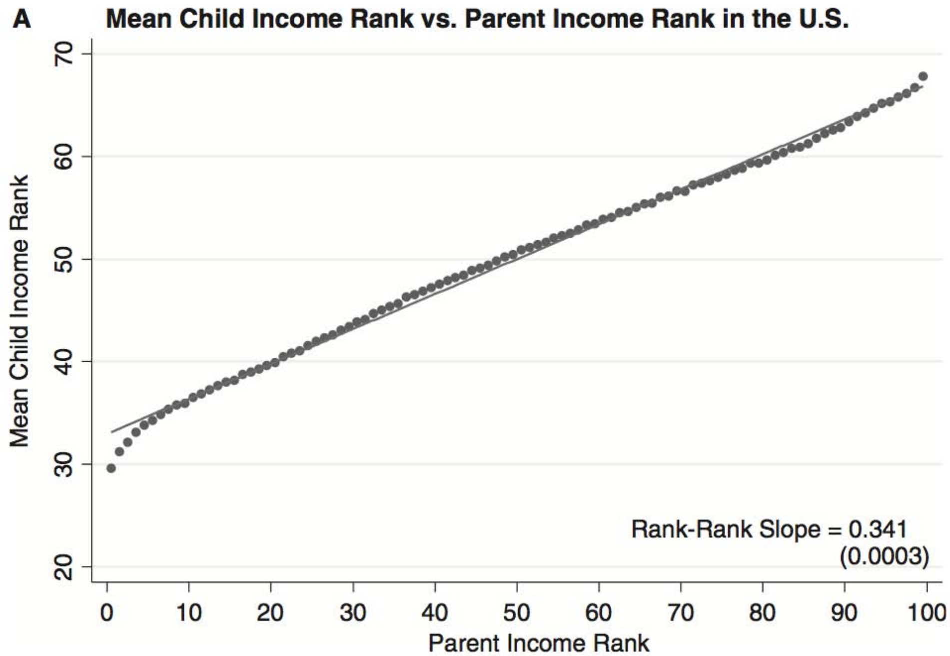


The authors argue that a more stable model is a rank-rank specification, which identifies the correlation between children's and parents' positions in the income distribution:

$$\frac{dE[\log R_i | P_i = p]}{d \log p} = \rho_{PR} \frac{SD(\log R_i)}{SD(\log P_i)}, \tag{37}$$

where ρ_{PR} is called the rank-rank slope and measures the association between a child's position in the income distribution and her parents' position in the distribution.

Figure II Panel A. Figure II Panel A plots the relationship between child's rank and parents' rank in their respective income distributions. The relationship is roughly linear - on average, a 10 percentile increase in parent income rank is associated with a 3.4 percentile increase in a child's income rank. The authors note that the slope is much more moderate in Denmark and Canada.



5.3 Spatial variation

The paper characterizes the variation in intergenerational mobility across commuting zones (CZs). Children were assigned to CZs based on where they lived at age 16 as reflected by the ZIP code on their parents' tax return, irrespective of whether they left that CZ afterward.

The authors decompose the mobility measures in each CZ into relative and absolute mobility.

- Relative mobility: For a child with the R_i percentile and parents in the P_i percentile in their respective income distributions, who grows up in CZ c, the authors estimate

$$R_{ic} = \alpha_c + \beta_c P_{ic} + \epsilon_{ic}, \tag{38}$$

  where β_c measures the relative mobility in CZ c. This is the difference in outcomes between children from top versus bottom income families within a CZ.

- Absolute mobility: Using the α_c and β_c from previous estimates, consider

$$\bar{r}_{pc} = \alpha_c + \beta_c p, \tag{39}$$

  where $\bar{r}_{pc}$ is the expected rank of children with parents at p percentile in CZ c, which measures the absolute mobility in CZ c at percentile p. The authors also define the "absolute upward mobility" as the absolute mobility at the 25th percentile (i.e. the $\bar{r}_{25,c}$ from $\bar{r}_{25,c} = \alpha_c + 25\beta_c$.)

Figure VI. Figure VI presents heat maps of absolute upward mobility and relative mobility. To illustrate the difference between the two measures, the authors note that San Francisco has substantially higher relative mobility than Chicago (lower in β_c) but lower absolute mobility at the 60th percentile. The contrast means that part of the greater relative mobility in San Francisco comes from worse outcomes for children from high income families. Three broad spatial patterns found in Panel A are: large variation at the regional level (lowest in the Southeast); large variation within regions; and lower levels of intergenerational mobility in urban than rural areas on average.

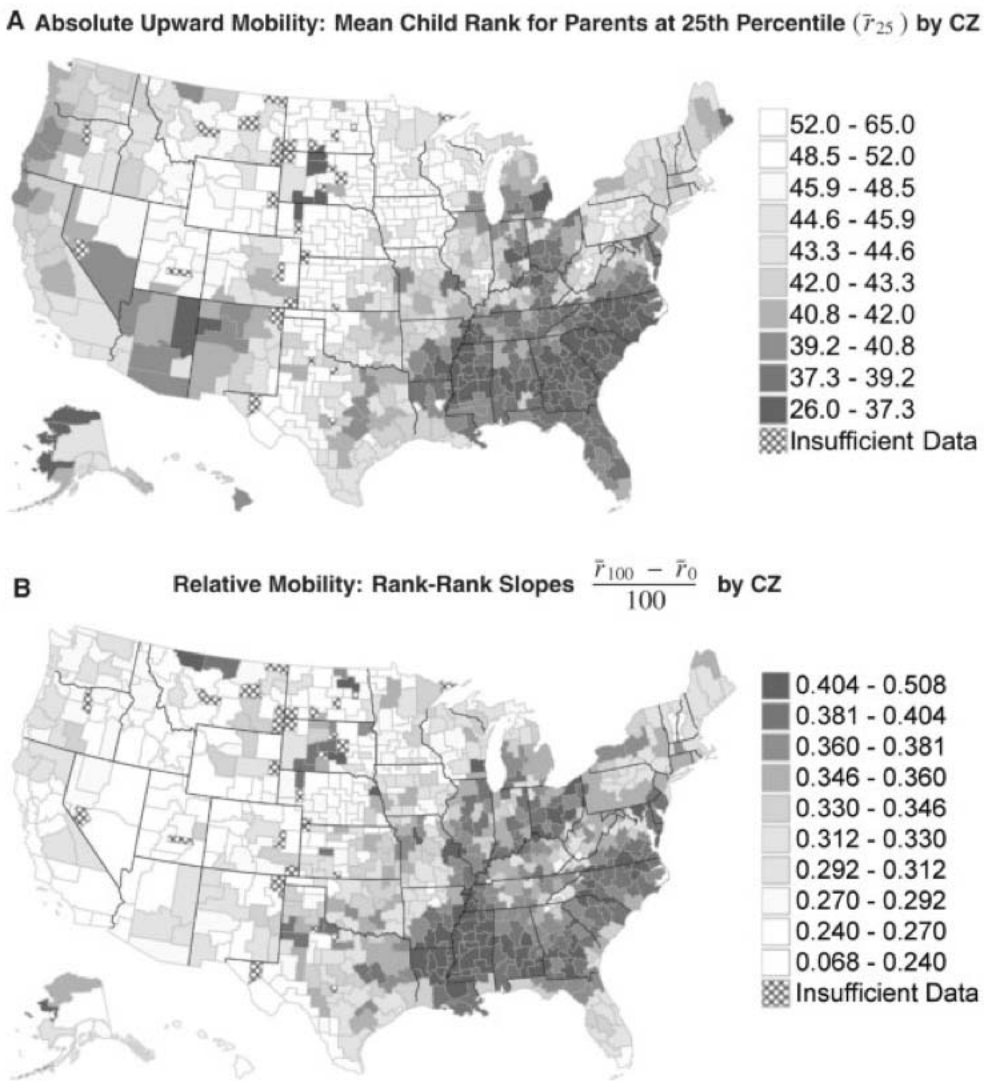


FIGURE VI

The Geography of Intergenerational Mobility

5.4 Correlates of intergenerational mobility

Figure VIII. Figure VIII presents a summary of the correlational results, plotting the unweighted univariate correlation between absolute upward mobility and various CZ-level characteristics, which are proxies for each broad factor. The dots show the point estimate of the correlation and the horizontal lines show a 95 percent confidence interval.

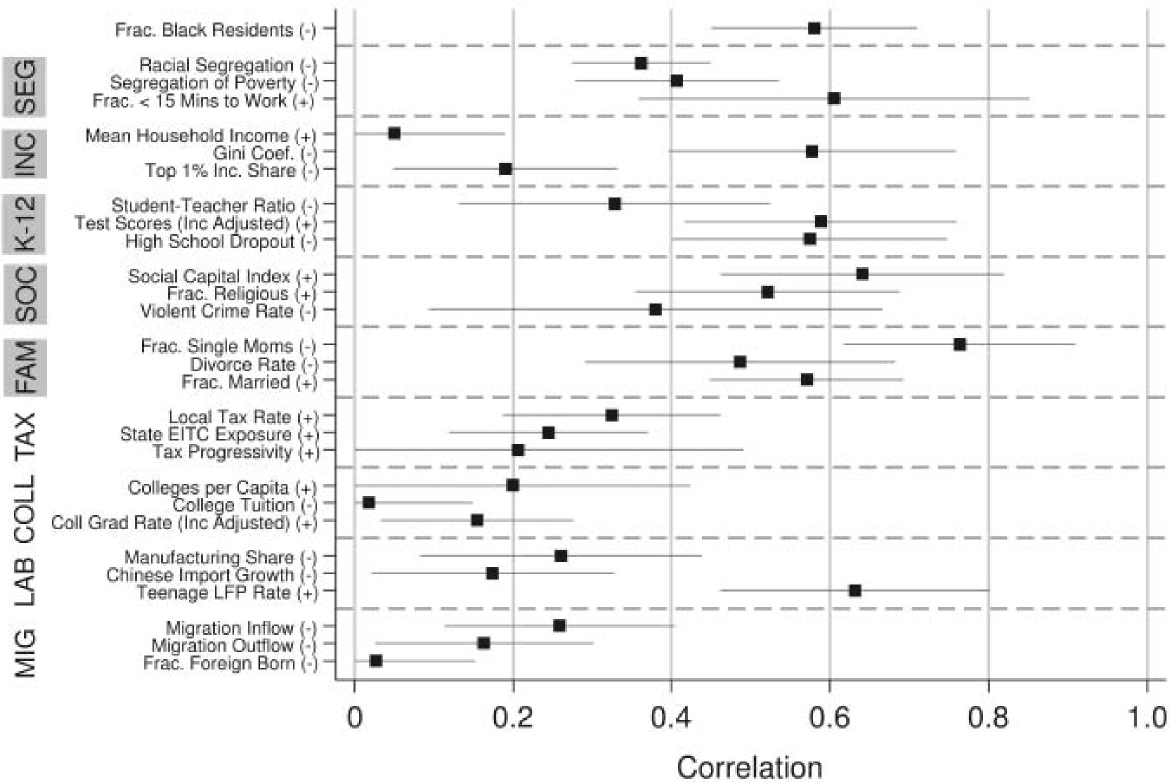


FIGURE VIII

Correlates of Spatial Variation in Upward Mobility

References

Almond, Douglas and Janet Currie, "Human capital development before age five," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 4B*, 2011, pp. 1487–1541.

Becker, Gary and Nigel Tomes, "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, 1979, *87* (6), 1153–1189.

_ **and** _ , "Human capital and the rise and fall of families," *Journal of Labor Economics*, 1986, *4* (3, Part 2: The Family and the Distribution of Economic Rewards), S1–S39.

Bjorklund, Anders, Mikael Lindahl, and Erik Plug, "The origins of intergenerational associations: Lessons from Swedish adoption data," *Quarterly Journal of Economics*, 2006, *121* (3), 999–1028.

Black, Sandra and Paul J. Devereux, "Recent developments in intergenerational mobility," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 4B*, 2011, pp. 1487–1541.

—, Paul Devereux, and Kjell Salvanes, "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital," *American Economic Review*, 2005, *95* (1), 437–449.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Currie, Janet and Enrico Moretti, "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *Quarterly Journal of Economics*, 2003, *118* (4), 1495–1532.

Galton, Francis, "Typical laws of heredity," *Proceedings of the Royal Institution of Great Britain*, 1877, *8*, 282–301.

Gelber, Alexander and Adam Isen, "Children's schooling and parents' investment in children: Evidence from the Head Start Impact Study," 2011. NBER working paper 17704.

Goldberger, Arthur, "Economic and mechanical models of intergenerational transmission," *American Economic Review*, 1989, *79* (3), 504–513.

Haider, Steven and Gary Solon, "Life-cycle variation in the association between current and lifetime earnings," *American Economic Review*, 2006, *96* (4), 1308–1320.

Loury, Glenn, "Intergenerational transfers and the distribution of earnings," *Econometrica*, 1981, *49* (4), 843–867.

Mazumder, Bhashkar, "Fortunate sons: New estimates of intergenerational mobility in the United States using Social Security earnings data," *Review of Economics and Statistics*, 2005, *87* (2), 235–255.

McCrary, Justin and Heather Royer, "The effect of maternal education on fertility and infant health: Evidence from school entry policies using exact date of birth," *American Economic Review*, 2011, *101* (1), 158–195.

Mulligan, Casey, "Galton versus the human capital approach to inheritance," *Journal of Political Economy*, 1999, *107* (S6), S184–S224.

Page, Marianne and Gary Solon, "Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban," *Journal of Labor Economics*, 2003, *21* (4), 831–855.

Sacerdote, Bruce, "How large are the effects from changes in family environment? A study of Korean American adoptees," *Quarterly Journal of Economics*, 2007, *122* (1), 119–157.

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

—, "Intergenerational mobility in the labor market," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 3A*, 1999, pp. 1761–1800.

—, "Cross-country differences in intergenerational earnings mobility," *Journal of Economic Perspectives*, 2002, *16* (3), 59–66.

—, "A model of intergenerational mobility variation over time and place," in Miles Corak, ed., *Generational income mobility in North American and Europe*, 2004, pp. 38–47.