sociological science

Supplement to:

Mitnik, Pablo A., Victoria Bryant, and Michael Weber. 2019. "The Intergenerational Transmission of Family-Income Advantages in the United States." Sociological Science 6: 380-415.

Contents

Appendix A. Did Mazumder (2005) overestimate economic persistence?

Chetty et al.'s (2014) criticism of Mazumder (2005) boils down to the claim that Mazumder's approach for dealing with missing parental data was tantamount to instrumental variable (IV) estimation and therefore led to overly high IGE estimates.[1] Mazumder (2016) has disputed this contention. In particular, he reiterated an argument he had made in his 2005 paper, according to which (a) IV estimates may well be downward biased when both right-side and left-side lifecycle downward biases are present, and (b) this was most likely the case with his data, given that children were "too young" while parents were "too old" at the time their income was measured (Mazumder 2016:92).[2] This argument is important because it entails that Chetty et al.'s contention should be interpreted as a plausible hypothesis that cannot be adjudicated without additional evidence.

Unfortunately, the only additional empirical evidence available on this issue is a robustness analysis reported in Table 6 of Mazumder (2005) that, despite Mazumder's protestation to the contrary (Mazumder 2016:92-93), is quite inconclusive. According to Mazumder (2016), his robustness analysis suggests that his 2005 estimates are not upward biased. In our view, however, the results in that analysis are not inconsistent with the claim that his estimates are upward biased. In particular, it is important in this regard that the IGE estimates in his Table 6 change very little when switching from a seven- to a ten-year measure of parental income once topcoded parents (i.e., those parents for whom imputation was needed) are dropped. Mazumder does provide a possible "selection-effect" explanation for this fact, but it is not clear that this hypothesis should have the upper hand vis-à-vis Chetty et al.'s (2014) hypothesis.

Therefore, our view is that the arguments articulated by Chetty et al. (2014) and Mazumder (2016), respectively, for and against the notion that Mazumder (2005) overestimated persistence are inconclusive.

Appendix B. Correct interpretation of the conventionally estimated IGE

Mitnik and Grusky (2017) have shown that the conventionally estimated IGE has been widely misinterpreted: While mobility scholars have assumed that they estimated the elasticity of the *expectation* of children's income, in the general case they actually estimated the elasticity of the *geometric mean* of children's income. That the latter is the case follows immediately from exponentiating and taking natural logarithm on the left-hand side of Equation [1]. Recalling that $GM(W) \equiv \exp E(\ln(W))$, for W any random variable, it is then easy to see that Equation [1] is equivalent to [1′]. It immediately follows that β_1 is the percentage differential in the geometric mean of children's long-run income with respect to a marginal percentage differential in parental long-run income.

The parameter β_1 is (also) the IGE of the expectation when the population error term satisfies very special conditions (see Santos Silva and Tenreyro 2006; Petersen 2017; Wooldridge 2002:17), but not otherwise. The parameter $\beta_1$would (also) be the elasticity of the conditional expectation of the children's income in the general case if it were the case that $E(\ln Y|x) = \ln E(Y|x)$, but this equality does not hold (due to Jensen's inequality).

Appendix C. Share interpretation of IGEs

In the main text we indicated that Equation [5] follows from Equation [2]. Here we provide the derivation. Using subscripts to indicate explicitly the distributions to which the operators variance, standard deviation and expectation refer, we may write:

$$\ln E_{Y|x}(Y|x) = \alpha_0 + \alpha_1 \ln x \qquad [2]$$

$$\ln E_{Y|X}(Y|X) = \alpha_0 + \alpha_1 \ln X$$

$$Var_X\big(\ln E_{Y|X}(Y|X)\big) = Var_X(\alpha_0 + \alpha_1 \ln X)$$

$$Var_X\big(\ln E_{Y|X}(Y|X)\big) = [\alpha_1]^2\, Var_X(\ln X)$$

$$SD_X\big(\ln E_{Y|X}(Y|X)\big) = \{[\alpha_1]^2\, Var_X(\ln X)\}^{1/2}$$

$$SD_X\big(\ln E_{Y|X}(Y|X)\big) = \alpha_1\, SD_X(\ln X)$$

$$\alpha_1 = \frac{SD_X(\ln E_{Y|X}(Y|X))}{SD_X(\ln X)}. \qquad [5]$$

In the main text we also argued that the nonparametric IGE_e provides an approximation to the ratio between the inequality in children's opportunities and the inequality in parental income (the quantity in the right-hand side of Equation [5]). Dropping the explicit subscripting for the probability distributions to which the operators variance, standard deviation and expectation refer, which should now be clear, we show here why that is the case.

From Equation [6]:

$$SD(\ln E(Y|X)) = \mathrm{SD}\big(\mathrm{F}(\ln \mathrm{X})\big).$$

A first-order Taylor-series approximation to the variance of the random function $F(.)$ centered around $E(\ln X)$ gives:

$$\mathrm{Var}\big(\mathrm{F}(\ln \mathrm{X})\big) \approx \left(\left.\frac{\partial F(t)}{\partial t}\right|_{\mathrm{t = E(ln\ X)}}\right)^2 Var(\ln X),$$

so

$$\frac{\mathrm{SD}\left(\mathrm{F}(\ln \mathrm{X})\right)}{SD(\ln X)} \approx \left.\frac{\partial F(t)}{\partial t}\right|_{\mathrm{t} = \mathrm{E}(\ln \mathrm{X})}.$$

Intergenerational curves are well approached by third-degree polynomials. We may then write:

$$\ln E(Y|X) = F(\ln X) \approx \gamma_0 + \gamma_1 \ln X + \gamma_2 (\ln X)^2 + \gamma_3 (\ln X)^3.$$

We then have:

$$\left.\frac{\partial F(t)}{\partial t}\right|_{\mathrm{t} = \mathrm{E}(\ln \mathrm{X})} \approx \gamma_1 + 2\,\gamma_2 E(\ln X) + 3\,\gamma_3 [E(\ln X)]^2. \qquad [C1]$$

Let's denote the nonparametric $\mathrm{IGE_e}$ by $\bar{\eta}$. We then have:

$$\bar{\eta} = E\left(\frac{\partial F(\ln X)}{\partial \ln X}\right) \approx E\left(\frac{\partial [\gamma_0 + \gamma_1 \ln X + \gamma_2 (\ln X)^2 + \gamma_3 (\ln X)^3]}{\partial \ln X}\right)$$

$$\approx \gamma_1 + 2\,\gamma_2 E(\ln X) + 3\,\gamma_3 E((\ln X)^2). \qquad [C2]$$

Given that $E((\ln X)^2) = [E(\ln X)]^2 + Var(\ln X)$, it follows from Equations [C1] and [C2] that

$$\left.\frac{\partial F(t)}{\partial t}\right|_{\mathrm{t} = \mathrm{E}(\ln \mathrm{X})} \approx \bar{\eta} - 3\,\gamma_3 Var(\ln X).$$

As, empirically, γ_3 is very small (if not statistically indistinguishable from zero), $\bar{\eta}$ provides an approximation to $\left.\frac{\partial F(t)}{\partial t}\right|_{\mathrm{t} = \mathrm{E}(\ln \mathrm{X})}$ and therefore to $\frac{SD(\ln E(Y|X))}{SD(\ln X)}$. Of course, the closer to quadratic the intergenerational curve (i.e., the closer to zero γ_3 is), the better this approximation will be (and vice versa). With our data, $\frac{SD(\ln E(Y|X))}{SD(\ln X)}$ is approximately six percent smaller than $\bar{\eta}$.

A similar analysis applies in the case of the nonparametric $\mathrm{IGE_g}$**.**

Appendix D. Estimation and statistical inference

We provide here a detailed discussion of the estimators employed to estimate the four IGEs in Figure 1, which were briefly described in the main text. We also provide details on the computation and interpretation of the confidence intervals reported in the main text, and explain why we did not follow the customary approach of including polynomials on children's and parents' age as control variables when estimating IGEs.

Estimators

Figure 2, in the main text, provides a summary of our discussion in this section.

The estimates of the constant IGE_g are all the result of estimating the PRF of Equation [1] by OLS, i.e., they were produced with the "OLS log-log estimator." This applies to the estimate reported by Chetty et al. (2014) and to the estimates based on the SOI-M data we obtained.

The estimates of the constant IGE_e based on the SOI-M data we produced were generated by estimating Equation [2] with the PPML estimator (Santos Silva and Tenreyro 2006). Like the OLS log-log estimator, the PPML estimator is semiparametric, that is, it makes no assumption regarding the distribution of the dependent variable and is consistent as long as the mean function is correctly specified (Gourieroux, Monfort, and Trognon 1984).[3]

The estimate of the constant IGE_e reported by Chetty et al. (2014) is based on a two-step estimator of Equation [2]. In the first step, nonparametric estimates of $\ln E(Y|\ln x)$ are generated by binning the children's parental income into 100 equal-sized (centile) bins, computing the mean income of parents and children within each bin, and taking the natural logarithm of those means. In the second step, an estimate of α_1 is generated by running an OLS regression of the estimates of $\ln E(Y|\ln x)$ on the corresponding $\ln x$ values.

As indicated in the main text, the estimators of the nonparametric IGEs are all two-step estimators: The first step produces nonparametric estimates of a number of points in the relevant intergenerational curve—i.e., the curve defined by either Equation [6] or Equation [7]—while the second step estimates the average slope of the curve through a numerical approximation based on the estimated points. Across datasets, the estimators only differ on the nonparametric approach used to estimate the points of the intergenerational curves and on the number of points that are estimated and employed in the numerical approximations.

The estimates of the nonparametric IGEs based on Chetty et al.'s (2014) data we produced for this article rely on estimates of the points of the intergenerational curves they obtained by (a) binning the children's parental income into centile bins, and (b) computing the log mean income of parents and the log mean income and mean log income of children within each bin.[4] These points pertain, approximately, to the 0.5, 1.5, …, 98.5, 99.5 percentiles of parental income. The estimates based on the SOI-M data we produced rely on estimates of 196 points pertaining to the 1,1.5, 2, . . ., 98, 98.5, 99 percentiles of parental income. We obtained these estimates by resorting to local polynomial regressions (Cleveland, Devlin, and Grosse 1988; Cleveland and Grosse 1991) of $\ln Y$ on $\ln X$ (in the case of the nonparametric IGE_g) and of Y on X (in the case of the nonparametric IGE_e). In all cases we used degree 1 polynomials and a tricube weight function. The smoothing parameter, however, is specific to each estimated curve, as it is the global minimizer of the AICc—a version of the Akaike Information Criterion specifically tailored to nonparametric regression (Hurvich, Simonoff, and Tsai 1998)—within the range of [.08, 1].

Regardless of dataset and IGE concept, in the second step the nonparametric IGE is estimated as follows. First, the expected point elasticity (or average slope of the curve) between any two contiguous points estimated in the first step is approximated by computing the corresponding ratio of finite differences. For instance, in the case of the nonparametric IGE_e, for two contiguous values of parental income x_1 and x_2, the approximate average elasticity between those points is computed by dividing $(\ln E(Y|x_2) - \ln E(Y|x_1))$ by $(\ln x_2 - \ln x_1)$ (where all quantities refer to estimates generated in the first step). Second, the nonparametric IGE estimate is computed as a simple average of the estimated elasticities between all pairs of contiguous points. Importantly, all segments defined by contiguous points cover the same share of children in the population: 0.5 percent with the points estimated with the SOI-M data and about 1 percent with the points estimated with Chetty et al.'s (2014) data.[5] Therefore, the final average is based on a set of self-weighting partial averages.[6]

Computation and interpretation of confidence intervals

In Table 3, we report 95 percent confidence intervals for the IGE estimates. In the case of the IGE estimates from Chetty et al. (2014), we simply report the confidence intervals implicit in the provided standard errors. In the case of the new estimates of constant IGEs based on the SOI-M data we obtained, we construct confidence intervals based on robust standard errors (which is mandatory with a PML estimator); the standard errors also take into account the clustering of children into families (see, e.g., Rogers 1993).

In the case of the estimates of nonparametric IGEs based on the SOI-M data we produced, statistical inference is based on the nonparametric bootstrap with 2,000 bootstrap samples. In the nonparametric context, use of the bootstrap-based percentile method (Efron and Tibshirani 1986) produces "variability bounds" (Racine 2008) or "confidence bands" (Wasserman 2006), which can be interpreted as approximations to true confidence intervals. For simplicity of terminology, in Table 3 we refer to the resulting intervals as "confidence intervals." We cannot compute similar (approximate) confidence intervals for the nonparametric IGE estimates based on Chetty et al.'s (2014) data we report, as we do not have access to the underlying microdata (and therefore cannot generate the needed bootstrap results).

Age controls

The relationship between long-run income measures (for instance, lifetime average income) and short-run proxy measures varies with the age at measurement (both for parents and children). For this reason, when estimating the constant IGE_g with proxy measures, it has been customary to include polynomials on children's and parents' age as control variables, each indexing the age at which the

income measurements were taken. However, Chetty et al.'s (2014) estimates of constant IGEs and our estimates of constant and nonparametric IGEs were produced without including such controls. In both cases, the variability in children's age in the samples is very minor, so controlling for age is unnecessary. In addition, because the age at which parents have their children is not exogenous to their income and the former's age may affect the latter's life chances, Mitnik et al. (2018: Online Appendix K) argued that controlling for parental age is inconsistent with the objective of measuring the gross association between parents' and children's income. In line with this argument and with the comparative purposes of this article, all new estimates we produced were generated without including age controls.

Appendix E. Chetty et al.'s (2014) evidence on lifecycle bias: Critical examination

Chetty et al. (2014) claimed that both the (constant) IGE_g and the (constant) IGE_e stabilize around age 30 and that this entails that their IGE estimates are free of left-side lifecycle bias.[7] However, their evidence for this claim is very problematic. As also argued by Mazumder (2016:113-14), in order to estimate the (constant) IGE_g for children older than 32, Chetty et al. resorted to auxiliary samples covering cohorts born earlier than those in their core sample. As the parental income of the children from these auxiliary samples is measured in 1996-2000 (i.e., the same period employed to measure parental income in their core sample), when children's age increases the age at which parental income is measured increases as well (e.g., for children age 41, parents' income is measured when parents are about ten years older than for children age 31). It is well-known that measuring parents' income when they are in their fifties depresses estimates (e.g., Grawe 2006; Haider and Solon 2006; Mazumder 2001). This means that, even if there is a reduction in left-side lifecycle bias as children get older than 32, in Chetty et al.'s analysis this would tend to be masked by the effect of simultaneously increasing the age at which parental income is measured (as the parents of core-sample children are already in their mid-40s in 1996-2000).[8]

On its part, their evidence for the (constant) IGE_e, which is based on the core sample only, does not include estimates for children older than 32. In this case Chetty et al. reported that estimates increase at a decreasing rate as children move from ages 22 to 32, and based on the fact that the estimated IGE is 2.1 percent higher at age 32 than at age 31, they concluded that it stabilizes around age 30. However, even with a decreasing growth rate, a 2.1 percent increase in one year (of age) is not necessarily that small. For instance, if the rate of growth each year is 0.9 of what it was in the previous year (and given their estimate of 0.343 by age 32), we should expect an IGE_e of about 0.38 by age 40, or close to 20 percent higher than by age 30. More generally, the "argument by extrapolation" they offer is not very persuasive, as the observed IGE-age points do not constrain that much the ways we may reasonably image the IGE-age curve continues past age 32. Chetty et al.'s argument is also inconsistent with the lifecycle-bias analyses

of Mitnik et al. (2018: Online Appendix I) and Mitnik (2017a), which suggest that estimates of the IGE_e of total family income do not stabilize around age 30.[9]

Appendix F. Attenuation bias and the advantage of tax income data

Chetty et al. (2014) claimed that income IGE estimates based on tax data should be less affected by attenuation bias than previously reported in the literature. They provided three arguments: (a) family income fluctuates less than individual earnings across years, (b) income is measured with less error in tax data than in survey data, and (c) the approach Mazumder (2005) employed to deal with missing parental information led him to overestimate the magnitude of attenuation bias. We consider these arguments in turn.

It is generally accepted that family income fluctuates less over time than father's earnings (e.g., Mazumder 2005:250), so the first argument seems unproblematic. On its part, the second argument can be strengthened considerably. The reason is that it's not really necessary that tax data be measured with less error than survey data for attenuation bias to be less of a problem with the former data. As tax data cover much better than survey data the upper tail of the parental-income distribution, which makes the "signal" larger, it should be enough that tax data do not include more "noise" than survey data (Mitnik 2017a:29-30).

This is easiest to see in the case of the constant IGE_g. Assuming no lifecycle bias, the standard analysis of attenuation bias in the OLS estimation of this elasticity (e.g., Solon 1992), is that $plim\ \widehat{\beta_1} = \beta_1 \frac{Var(\ln S)}{Var(\ln S)+Var(\varepsilon)}$, where ε is a zero-expectation additive noise in the logarithm of the short-run measure of parental income S with respect to the logarithm of the long-run measure X. This entails that even if $Var(\varepsilon)$ were the same with survey and tax data, the "attenuation factor" multiplying β_1 would still be closer to one with tax data due to their better coverage of the upper tail of the parental-income distribution, as this can be expected to lead to a larger value of $Var(\ln S)$. Mitnik (2017a) has shown that a similar analysis applies in the case of the constant IGE_e.

The third argument is, of course, closely related to the argument we considered in our Appendix A (see above). There is, however, a subtle but consequential difference between the two arguments. In that appendix, the focus was on the estimates Mazumder (2005) produced with the measure of father's earnings based on the maximum number of years of information he had available, with Chetty et al. (2014) identifying a feature of the data (imputation of father's earnings) that can be expected to push the estimates up and Mazumder (2016) identifying a different feature of the data (children's and fathers' ages) that can be expected to push them down. In contrast, here the focus is on the full series of estimates Mazumder obtained when increasing the years of information used to compute the measure of father's

earnings. As Mazumder (2005) imputes earnings for a larger share of fathers when he uses more years of information, the estimates increase not only because the variance of the error $Var(\varepsilon)$ falls but also because the estimates are moving toward IV estimates. In this context (a) the children's ages remain constant over the analysis, and (b) the fathers' ages decrease when more years of information are used, which in this case can be expected to reduce *downward* right-side lifecycle bias (as the average age of parents is 47 with the parental measure based on the fewest years of information). Therefore, it is quite likely that the estimates of attenuation bias computed from Mazumder's results are upward biased.[10]

We therefore conclude that Chetty et al.'s (2014) claim that, for estimates based on tax income data, attenuation bias should be a less serious problem than previously reported in the literature has clear merit.

Appendix G. Chetty et al.'s (2014) evidence on attenuation bias: Critical examination

Chetty et al. (2014) claimed that their estimates of the constant IGE_g based on tax data nearly stabilize once five years are employed, and that this entails that they are (nearly) free of attenuation bias. To support their claim, they reported that the constant IGE_g of family income hardly increased (i.e., from 0.344 to 0.366, or six percent) when they used 15 years of parental information instead of five.[11]

Mazumder (2016:114-115) has forcefully criticized this evidence. He pointed out that when Chetty et al. constructed their measure of parental income based on 15 years of information, all of the additional years pertained to when the parents were older. For instance, if for a specific child the five-year measure of parental income pertains to when her parents are 42-46 years old (in 1996-2000), the fifteen-year measure pertains to when they are 42-56 years old (in 1996-2010). But, as we mentioned in Appendix E (see above), measuring parents' income when they are old depresses estimates; it is for this reason that it has long been argued that attenuation bias is best reduced by adding parental information from parents' prime-age period, not by adding information when they are in their fifties (Mazumder 2005:247-248). Moreover, Mazumder (2016:114-115) has provided survey-data evidence consistent with his argument, which suggest that adding years of parental information "in the wrong direction" may even *reduce* estimates if the added years pertain to when the parents are old enough. Mazumder concluded that the results that Chetty et al. (2014) reported were most likely distorted by the increasing noisiness of the additional years of parental information they used in their analysis.

Chetty et al. did consider this possibility, but rejected it on the argument that they had provided evidence that "estimates of mobility are not sensitive to varying the age in which parent income is measured over the range observed in our dataset" (2014: Online Appendix E, ftn. 9). The evidence in question, however, pertains to the rank-rank slope. Parents' ranks may remain the same as they get older

even as the differences between their incomes increase. As the latter can be expected to raise IGEs, the argument does not hold much water.

In Mitnik et al.'s (2018: Online Appendix H) gender-specific attenuation-bias analyses with tax data, they focused on the IGE_e rather than the IGE_g. They computed the constant IGE_e with parental income measures based on one to nine years of information, and reached three conclusions: (a) attenuation bias is greatly reduced by using nine years of parental information, (b) using five years instead of nine years would result in a non-negligible increase in bias, and (c) although estimates appear to be reaching a plateau once nine years of parental information are employed, it is not possible to rule out that some (downward) bias remains. Mitnik et al.'s evidence is therefore inconsistent with the notion that, with tax-based data, five years of information are enough to eliminate the bulk of attenuation bias.

Appendix H. Values used in the mean imputation of income and log income for nonadmin children

The Annual Social and Economic Supplement of the Current Population Survey (CPS-ASEC) identifies likely nonfilers using a tax simulation model. Although this information is available for the entire period covered by the SOI-M Panel, the CPS-ASEC data after 2003 have serious inconsistencies and cannot be used. We therefore use pooled CPS-ASEC data from 1999 to 2003 to estimate the mean income of nonadmins by gender-age group. The resulting mean values, which we used for income imputation, are as follows (all in 2010 dollars): 26-30 year-old men: $4,910; 31-35 year-old men: $5,815; 36-40 year-old men: $6,706; 26-30 year-old women: $5,372; 31-35 year-old women: $6,574; 36-40 year-old women: $7,560.

Approximately one-third of CPS nonadmins have zero family income. Therefore, simply computing their mean log income by gender-age group is not feasible. As discussed in the main text, we computed two sets of values. While in one case CPS nonadmins with zero income are dropped, in the other they are assigned an income of $1. The resulting mean values are the following (the first figure in each pair pertains to the computation in which CPS nonadmins with zero income are dropped): 26-30 year-old men: 8.25 and 3.84; 31-35 year-old men: 8.40 and 4.72; 36-40 year-old men: 8.62 and 5.33; 26-30 year-old women: 8.42 and 4.80; 31-35 year-old women: 8.49 and 5.67; 36-40 year-old women: 8.63 and 6.04. These mean log income figures provide upper and lower imputation values leading, respectively, to lower and upper estimates of the IGE_g.

Appendix I. Additional attenuation-bias results

Given Chetty et al.'s (2014) strong denial that their IGE_g estimates are affected by attenuation bias, it seems imperative that we examine whether our empirical findings with regards to attenuation bias are robust to the various ways in which attenuation-bias analyzes can be conducted. This is what we do

here, focusing on the constant IGEs (as in the previous literature on attenuation bias) for men and women pooled. We also provide and discuss evidence relevant for assessing whether the estimates based on the SOI-M data we produced using nine-year parental-income measures are likely to be (nearly) free of attenuation bias.

In Table I1 we present estimates of the constant IGE_e and IGE_g obtained with parental measures based on five, eight and nine years of information. We generated these estimates with children's income measured either in 2004 or 2010—that is, when the children were in their early or late thirties—and using either the "common-rules" or the "common-sample" approach. Briefly, the common-rules approach uses (nearly) the same sample inclusion rules regardless of the number of years of information employed to compute the parental income measure; in contrast, the common-sample approach uses (nearly) the same sample to generate all estimates, i.e., the sample that is selected when the inclusion rules are applied with a particular *n*-year parental variable, regardless of the number of years of information actually employed to compute the parental variable used for estimation (for details, see Mitnik et al. 2018: Online Appendix H).

Here we implemented this approach for $n = 9$ (common-sample approach I) and $n = 5$ (common-sample approach II).[12] In addition, in the case of the IGE_g, we generated the estimates shown in Table I1 after assigning nonadmin children \$0 income (i.e., dropping them from the analysis, as in Chetty et al. 2014), or CPS-based mean log income values (computed after dropping CPS nonadmins with zero income, as in the lower-bound estimates in the second column of Table 3).

The results for the three approaches are displayed in Table I1 in contiguous horizontal panels, each of which includes five columns. We start by focusing on the first and third columns of each panel—which show IGE estimates based on five- and nine-year parental-income variables—and on the fourth column—which shows the percent difference between these two estimates. The results in these columns are uniformly inconsistent with Chetty et al.'s (2014) contention that five years of parental information are enough to nearly eliminate attenuation bias. Regardless of approach, IGE concept, children's ages, and treatment of nonadmin children, the estimates increase substantially—between 7.5 and 9.1 percent for the IGE_e, and between 15.7 and 33.0 percent for the IGE_g—when switching from the five-year to the nine-year parental measure. Therefore, we can conclude that our evidence that IGE estimates based on five-year parental measures are affected by substantial attenuation bias is very robust.

The second and third columns in each panel allow to compare IGE estimates based on eight- and nine-year parental-income variables, while the fifth column shows the corresponding percent differences. The comparison of estimates based on eight and nine years of information provides some evidence on whether the latter are likely to be (nearly) free of attenuation bias; if this is the case, we should observe

Table I1. IGE estimates with five, eight and nine years of parental information

	Common-sample approach I					Common-sample approach II					Common-rules approach				
	Years of par. inf.			% Δ 5 to 9 years	% Δ 8 to 9 years	Years of par. inf.			% Δ 5 to 9 years	% Δ 8 to 9 years	Years of par. inf.			% Δ 5 to 9 years	% Δ 8 to 9 years
	5	8	9			5	8	9			5	8	9		
Constant IGE_e, CPS-based mean imputation															
Year 2010, children in their later 30s	0.429	0.457	0.461	7.5	0.8	0.427	0.453	0.456	7.0	0.8	0.428	0.460	0.461	7.7	0.3
Year 2004, children in their early 30s	0.368	0.397	0.401	9.1	1.0	0.371	0.396	0.400	7.8	0.8	0.371	0.399	0.401	8.1	0.5
Constant IGE_g, $0 imputation															
Year 2010, children in their later 30s	0.308	0.387	0.393	27.6	1.7	0.318	0.387	0.396	24.5	2.2	0.321	0.386	0.393	22.6	1.8
Year 2004, children in their early 30s	0.271	0.317	0.324	19.6	2.0	0.277	0.323	0.329	18.5	1.8	0.279	0.319	0.324	16.1	1.5
Constant IGE_g, CPS-mean imputation (CPS nonadmins dropped)															
Year 2010, children in their later 30s	0.343	0.434	0.456	33.0	5.1	0.383	0.459	0.470	22.7	2.4	0.385	0.435	0.456	18.5	4.9
Year 2004, children in their early 30s	0.324	0.397	0.416	28.5	4.7	0.359	0.417	0.425	18.5	1.9	0.360	0.399	0.416	15.7	4.3

Notes:

The common-sample approach keeps the sample (nearly) fixed when estimating IGEs using five and nine years of parental information. There are two ways of implementing this approach. In the columns under "Common-sample I" the sample inclusion rules are applied with the nine-year parental variable. In the columns under "Common-sample II" the sample inclusion rules are applied with the five-year variable. The common rules approach keeps the sample inclusion rules (nearly) fixed, and applies them with the five- or the nine-year parental variable, as relevant. "% Δ" denotes "percent difference."

that the differences between estimates are very small. The results on this regard differ across IGE concepts, and also across treatments of nonadmin children when estimating the IGE_g.

In the case of the IGE_e, no difference in estimates is larger than one percent and the differences are half of that or less in the case of the common-rule approach. This is exactly what we would expect if the nine-year estimates had (almost) converged to the long-run estimates of interest, so these findings are a reason for optimism. At the same time, as Mitnik et al. (2018: Online Appendi H) have also pointed out, evidence like this does not allow to rule out that a non-negligible amount of attenuation bias still remains.[13]

In the case of the IGE_g with CPS-based mean imputation for nonadmin children, estimates increase between 1.9 and 5.1 percent when we use nine instead of eight years of parental information. When nonadmin children are instead assigned $0 and therefore dropped from the analyses, estimates increase significantly less, between 1.5 and 2.2 percent—but still significantly more than in the case of the IGE_e. It therefore seems safe to conclude that, in both cases, the IGE_g estimates based on nine years of parental information are affected by residual attenuation biases of a non-negligible magnitude. This provides further support for our interpretation of the IGE_g estimates obtained by dropping CPS nonfilers with zero income (when computing values for mean imputation) as *lower-bound* estimates, and therefore for our claim that Chetty et al.'s (2014) IGE_g estimates greatly understated true economic persistence.

Appendix J. Shapley decompositions of total bias in IGE_g estimation

In the main text we indicated that to properly capture the effects of lifecycle, attenuation, functional-form and selection bias on the estimates based on Chetty et al. (2014) data and methodological decisions, all estimates should be obtained with two-year samples, but that proceeding this way was not feasible. Here, we explain why this is the case as well as the alternative approach we used.

The first problem is that, with two-year samples, (a) the relevant mean values for imputation are not the mean annual log income values employed to produce the estimates shown in Table 3, due to the dependencies between children's income across years, and (b) the auxiliary data that would be needed to compute mean values that take into account those dependencies are not available (the CPS-ASEC data are unsuitable for this purpose). The second problem is that some two-year measures would require using 2009 information to compute them. However, as reported by Mitnik et al. (2018: Online Appendix I), income data for 2008 and 2009 appear to be seriously affected by the Great Recession, which greatly compressed the income distribution. Using the 2009 data to compute income IGEs is therefore unadvisable.

As an alternative, we proceeded as follows. As we did in the case of the IGE_e, we computed 16 estimates, using in all cases one-year samples. However, before computing the Shapley decompositions, we adjusted all estimates based on samples where nonadmin children's income was assumed to be zero. We did this adjustment separately for estimates of the constant and the nonparametric IGE_g, under the assumption that they all underestimate the IGE_g by the same amounts that the corresponding estimates based on the SOI-M (one-year) all-biases sample underestimate it (compared to the estimates based on the SOI-M two-year all-biases sample; see Figure 6). After introducing these adjustments, we computed the Shapley decompositions as before.

Importantly, with this approach, the total biases are exactly what is wanted in all cases. The total biases we would like to decompose are in all cases differences between the IGE_g estimates based on the SOI-M best sample and the SOI-M two-year all-biases sample. As the *adjusted* estimates based on the SOI-M (one-year) all-biases sample are identical to the estimates based on the SOI-M two-year all-biases sample, the total biases we actually decompose (differences between the estimates based on the SOI-M best sample and the *adjusted* estimates based on the SOI-M all-biases sample) are equal to the total biases of interest. We make use of this equivalence in defining total biases in the notes to Table 6 and 7.

Notes

[1] For the reason why IV estimates are usually expected to be upward biased in the intergenerational-mobility context, see Solon (1992: Appendix) and Mitnik (2017b:8-10).

[2] See Mitnik (2017b: Eqs. 13 and 14) for a measurement-error model consistent with this argument.

[3] See Mitnik (2017c) for how to estimate the constant IGE_e using the PPML estimator and the statistical package Stata.

[4] The estimated points have been made publicly available by Chetty et al. (2014); see the sources in Table 3 for details.

[5] The latter under the assumption that, within centiles of parental income, mean and median parental income are approximately the same.

[6] In the case of estimates based on the SOI-M data, the estimation of nonparametric IGEs ignores the intergenerational curves' final left and right segments (each covering 1 percent of children). Because the curve is estimated less precisely at the boundaries, these "trimmed estimators" are often more efficient. The point estimates from the trimmed and untrimmed estimators are, however, very similar. With Chetty et al.'s (2014) data, the estimation of the nonparametric IGEs also ignores the final left and right segments (each covering approximately 0.5 percent of children).

[7] For the IGE_g, see Chetty et al. (2014:1580 and Online Appendix Figure IIa). For the IGE_e, see Chetty et al. (2014: Online Appendix C and Figure Ib). Chetty et al. (2014) also claimed that their rank-rank slope estimates are unaffected by lifecycle bias. Our discussion here is only concerned with, and is only meant to apply to, lifecycle bias in the estimation of IGEs.

[8] An additional potential issue is that, in producing their evidence on lifecycle-bias, Chetty et al. (2014) estimated the IGE_g not with the full sample (as is standard), but only with children with parental income between the 10^{th} and the 90^{th} percentiles (see notes to their Online Appendix Figure IIa). Proceeding this way could have had an impact on their results.

[9] Mitnik et al.'s (2018: Online Appendix I) results are not conclusive, as some of their IGE estimates appear to have been driven downward by the Great Recession. Nevertheless, overall they do suggest that IGE estimates based on a sample of men and women that are 29-32 years old when their income is measured should be affected by lifecycle bias.

[10] This could still not be the case if, for instance, negative lifecycle biases are substantially larger in absolute value with the IV than with the OLS estimator. There is no evidence, however, that this is so.

[11] See Chetty et al. (2014: Table 1 and Online Appendix E). Chetty et al. (2014) also claimed that their rank-rank slope estimates are unaffected by attenuation bias. Our discussion here is only concerned with, and is only meant to apply to, attenuation bias in the estimation of IGEs.

[12] In our Shapley decompositions we used the common-rules approach. The attenuation-bias evidence provided by Chetty et al. (2014; see our Appendix G, above) is likely based on the common-sample approach (with $n = 5$). Mitnik et al. (2018: Online Appendix H) and Mazumder (2005) used both approaches in their attenuation-bias analyses.

[13] First, the evidence in Table I1 is not inconsistent with the estimates converging very slowly, past year nine, to the long-run estimates; if this were the case, the long-run estimates of interest could still be significantly larger than those based on nine years. Second, the argument advanced by Mazumder (2016) in his criticism of Chetty et al.'s (2014) attenuation-bias evidence is relevant here as well: It is possible that the nine-year estimates only appear to have (almost) converged to the long-run estimates of interest because the years of income information that are being added as we move from the five-year measure to the nine-year measure pertain to the "wrong parental ages" (see our Appendix G, above).

References

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

Cleveland, William, Susan Devlin, and Eric Grosse. 1988. "Regression by Local Fitting: Methods, Properties, and Computational Algorithms." *Journal of Econometrics* 37(1): 87-114.

Cleveland, William and Eric Grosse. 1991. "Computational Methods for Local Regression." *Statistics and Computing* 1: 47-62.

Efron, Bradley and Robert Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science* 1(1): 1-154.

Gourieroux, C, A. Monfort and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52(3): 681-700.

Grawe, Nathan. 2006. "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence." *Labour Economics* 13: 551-570.

Haider, Steven and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96(4): 1308-1320.

Hurvich, Clifford, Jeffrey Simonoff, and Chih-Ling Tsai. 1998. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion*." Journal of the Royal Statistical Society* B, 60: 271–293.

Mazumder, Bhashkar. 2001. "The Miss-measurement of Permanent Earnings: New Evidence from Social Security Earnings Data." Federal Reserve Bank of Chicago Working Paper 2001-24.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics* 87(2): 235-255.

Mazumder, Bhashkar. 2016. "Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitation of Tax Data." In *Inequality: Causes and Consequences*, edited by Lorenzo Cappellari, Solomon Polacheck, and Konstantinos Tatsiramos. Bingley: Emerald.

Mitnik, Pablo. 2017a. "Estimating the Intergenerational Elasticity of Expected Income with Short-Run Income Measures: A Generalized Error-in-Variables Model." Stanford Center on Poverty and Inequality Working Paper.

Mitnik, Pablo. 2017b. "Intergenerational Income Elasticities, Instrumental Variable Estimation, and Bracketing Strategies." Stanford Center on Poverty and Inequality Working Paper.

Mitnik, Pablo. 2017c. "Estimators of the Intergenerational Elasticity of Expected Income: A Tutorial." Stanford Center on Poverty and Inequality Working Paper.

Mitnik, Pablo, Victoria Bryant, Michael Weber and David Grusky. 2018. “A Very Uneven Playing Field: Economic Mobility in the United States.” Stanford Center on Poverty and Inequality Working Paper.

Mitnik, Pablo and David Grusky. 2017. The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility.” Stanford Center on Poverty and Inequality Working Paper.

Racine, Jeffrey. 2008. “Nonparametric Econometrics: A Primer.” *Foundations and Trends in Econometrics* 3(1): 1-88.

Rogers, Williams. 1993. “Regression Standard Errors in Clustered Samples.” *Stata Technical Bulletin* 13: 19-23.

Santos Silva, J. M. C. and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88(4): 641-658.

Solon, Gary. 1992. “Intergenerational Income Mobility in the United States.” *American Economic Review* 82(3): 393-408.

Petersen, Trond. 2017. “Multiplicative Models for Continuous Dependent Variables: Estimation on Unlogged versus Logged Form.” *Sociological Methodology* 47:113-164.

Wasserman, Larry. 2006. *All of Nonparametric Statistics*. New York: Springer.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass: The MIT Press.