Intergenerational Mobility in India:
New Measures and Estimates Across Time and Social Groups*

Sam Asher†

Paul Novosad‡

Charlie Rafkin§

September 2022

Appendix: For Online Publication Only

*We are thankful for useful discussions with Alberto Abadie, David Autor, Emily Blanchard, Raj Chetty, Eric Edmonds, Shahe Emran, Francisco Ferreira, Amy Finkelstein, Nate Hilger, Larry Katz, David Laibson, Ethan Ligon, Erzo Luttmer, Whitney Newey, Elias Papaioannou, Nina Pavcnik, Bruce Sacerdote, Frank Schilbach, Na'ama Shenhav, Forhad Shilpi, Andrei Shleifer, Gary Solon, Bob Staiger, Doug Staiger, Chris Snyder and Elie Tamer, among others. Annaka Balch, Ali Campion, Toby Lunt, Ryu Matsuura, and Taewan Roh provided excellent research assistance. This project received financial support from the IZA GLM-LIC program. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. This paper contains some material previously contained in the retired paper, "Getting Signal from Interval Data: Theory and Applications to Mortality and Intergenerational Mobility." An earlier version of this paper had the subtitle "Across Time, Space, and Communities."

†Imperial College London, sam.asher@imperial.ac.uk

‡Dartmouth College, paul.novosad@dartmouth.edu, corresponding author

§Massachusetts Institute of Technology, crafkin@mit.edu

A Appendix A: Additional Tables and Figures

Figure A1

Coresidence Rates by Age and Gender

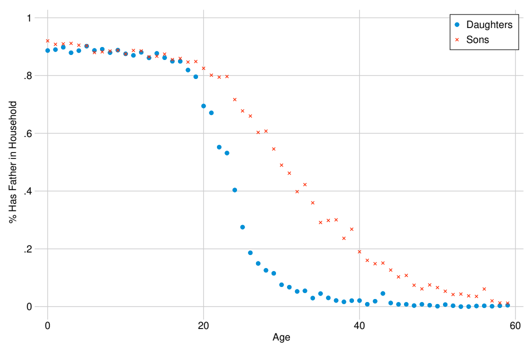


Figure A1 shows the share of individuals who live in the same household as their father as a function of gender and age. Source: IHDS (2012).

Figure A2

Bias in Mobility Estimates When Sample is Limited to Coresident Pairs

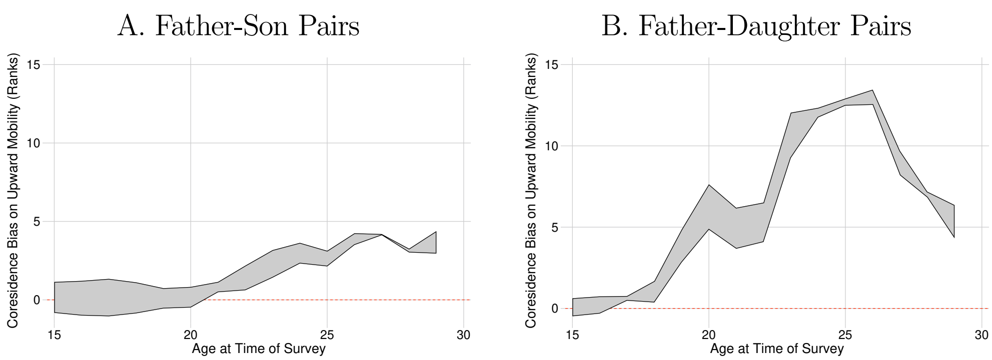
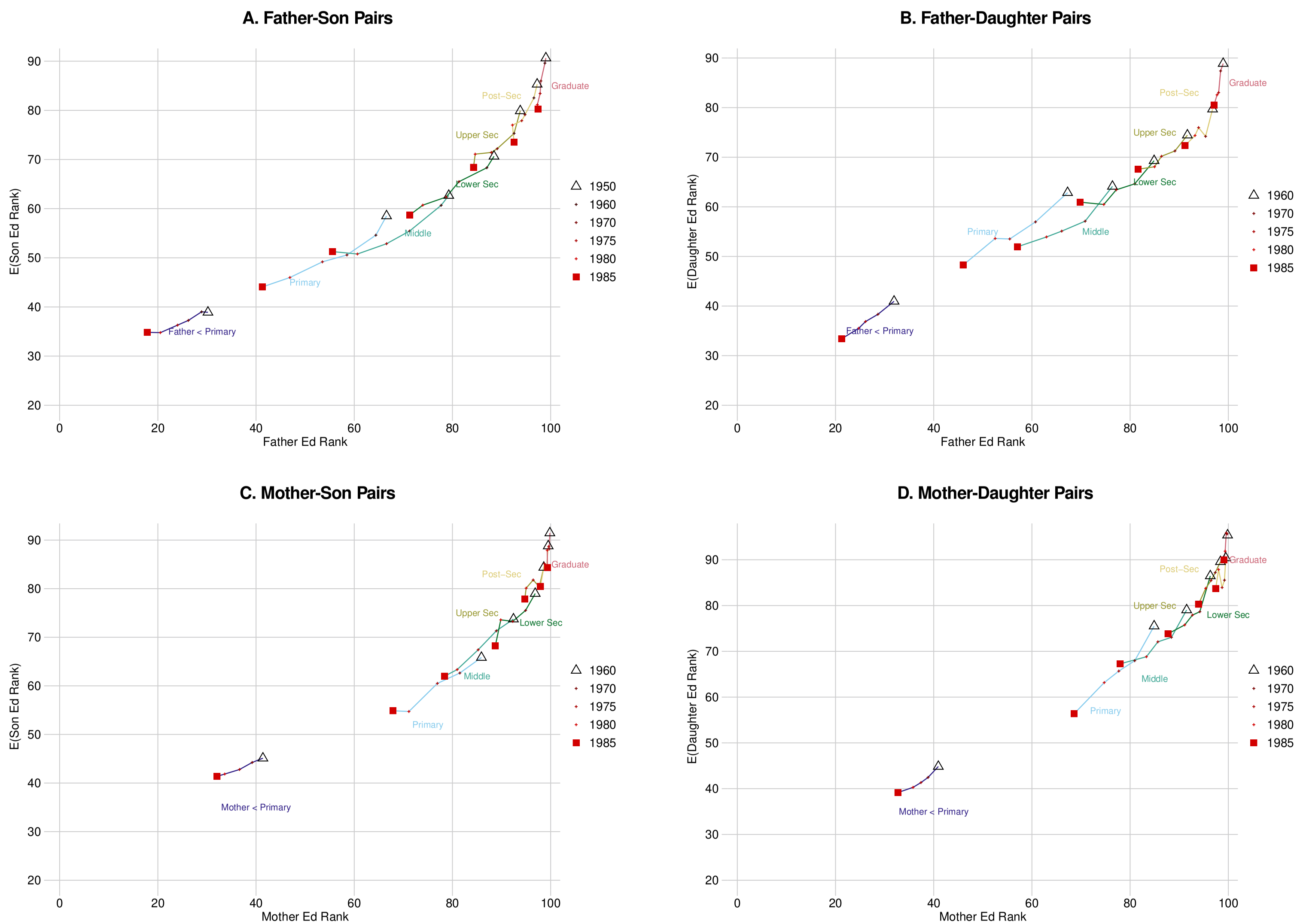


Figure A2 shows the bias in a measure of upward mobility when children who do not live with their parents are excluded from the sample. The bias is shown as a function of child age. The mobility measure is bottom half mobility (μ_0^{50}), which is the expected child rank conditional on being born to a parent in the bottom half of the education distribution. Bias is calculated as the coresident-only measure minus the full sample measure. Source: IHDS (2012).

Figure A3

Joint Parent-Child Education Moments Over Time



The figure shows the conditional expectation function of child education rank, given parent education rank for six cohorts. Each set of connected points corresponds to a time series for a different *level* of education; the time series (which moves from the triangle to the square) shows changes in a child's expected rank (Y axis) given a parent at some fixed level of education, along with the change in the father's education rank (X axis) in the cohort of fathers. Points move to the southwest because low-education fathers have lower ranks as education rises (moving left), and thus have children with lower education ranks (moving down). Ranks are calculated as the midpoint rank of a given education bin. Source: IHDS (2012).

Figure A4

Robustness of Upward Mobility to Survivorship Bias

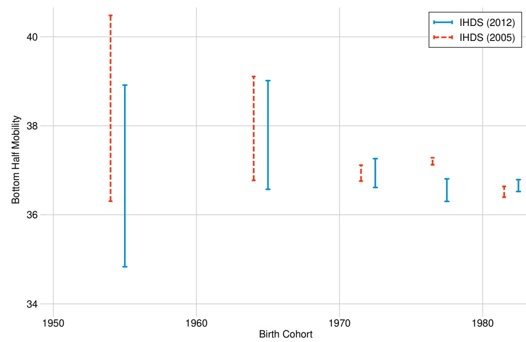


Figure A4 shows a test of survivorship bias in estimates of bottom half mobility. The figure shows estimates of bottom half mobility calculated for the 1950s to 1980–85 birth cohorts, measured separately in the 2005 and 2012 rounds of the IHDS. If there was substantial survivorship bias in the mobility measures, we would expect the estimates to differ across the two surveys because of the deaths of some of the respondents.

Figure A5

Bottom Half Mobility (μ_0^{50}) for Mother-Son and Mother-Daughter Pairs

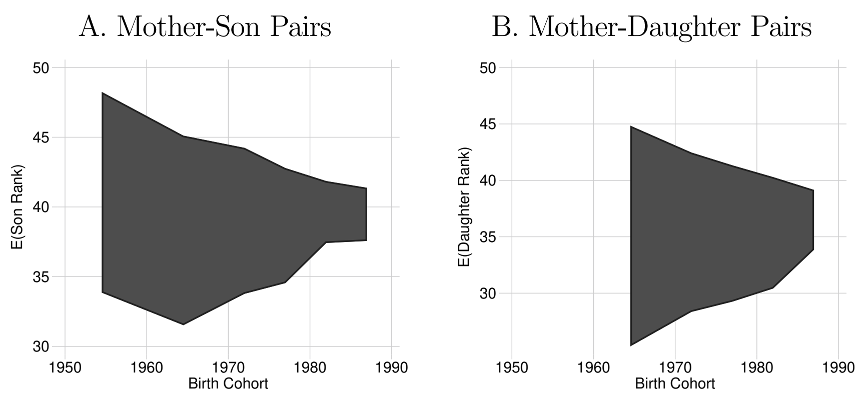


Figure A5 shows bounds on aggregate trends in intergenerational mobility, using cohorts born from 1950–59 through 1985–89, focusing on mother-son and mother-daughter links. The measure used is bottom half mobility (μ_0^{50}), which is the average rank attained by children born to parents who are in the bottom half of the education distribution. The bounds are very wide because of the large share of mothers who report bottom-coded education levels. Source: IHDS (2012).

Figure A6

Trends in Mobility by Subgroup, 1950–1989 Birth Cohorts
Education Level Outcomes

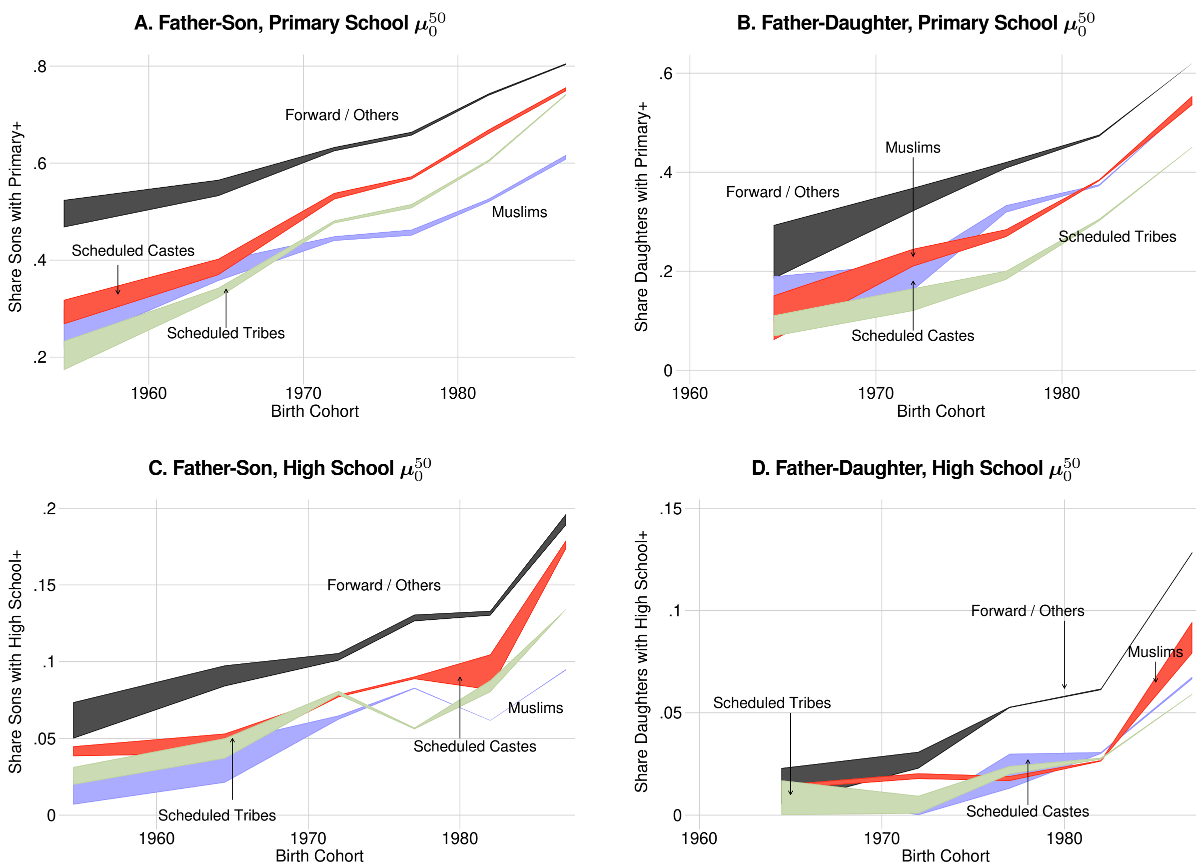
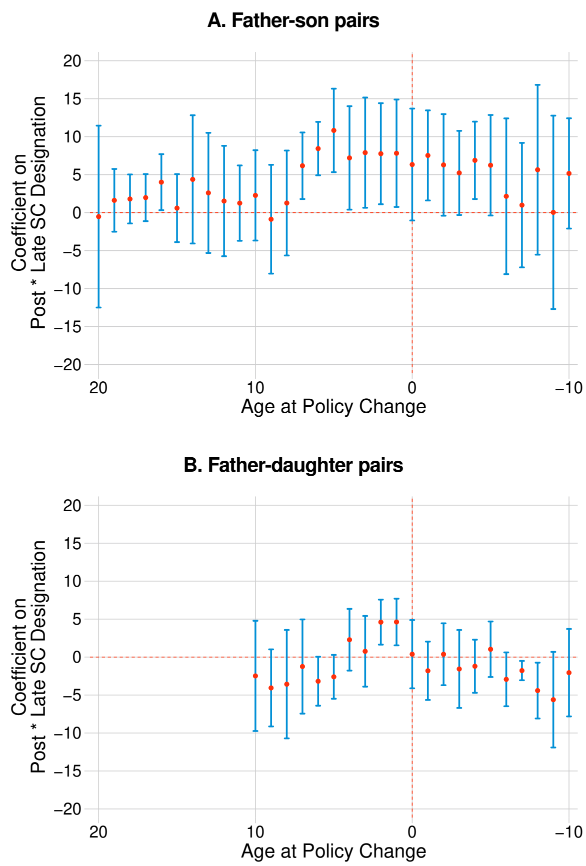


Figure A6 presents bounds on intergenerational mobility, stratified by four prominent social groups in India: Scheduled Castes, Scheduled Tribes, Muslims, and Forward Castes/Others. The figure is analogous to Figure 6, but shows the expected probability that a child attains a given education level (primary in Panels A and B, and secondary in Panels C and D), conditional on having a father in the bottom half of the father education distribution. Linked father-daughter education data are not available for the 1950–59 birth cohort. Source: IHDS (2012).

Figure A7

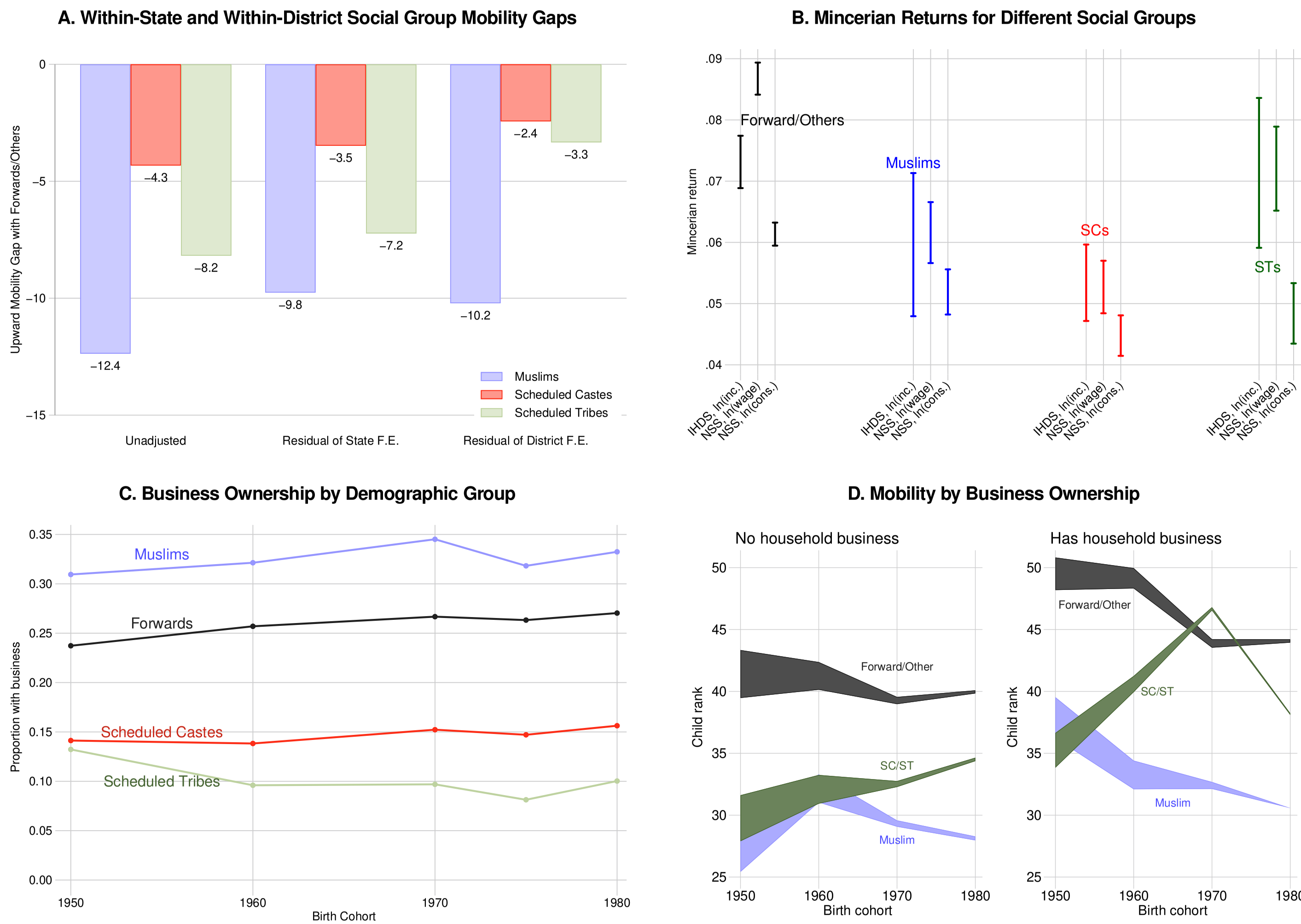
Effect of Scheduled Caste Designation on Upward Mobility
Robustness to Alternate Post Years



The figure shows point estimates from Equation 6.1, with a range of definitions of "post", the first year at which post-policy-change cohorts are modeled as exposed to the new policy regime. The X axis shows the child's age at the time of the policy change in 1977; negative ages describe children born after 1977. All outcomes are measured in 2012. The Y axis shows a regression coefficient which describes the relative gains in rank points to "Late" Scheduled Caste children born after their castes were added to the Scheduled Caste list (see Section 6.1). Source: IHDS (2012).

Figure A8

Other Candidate Mechanisms



Panel A presents the bottom-half mobility disadvantage relative to Forwards/Others faced by Muslims, Scheduled Castes and Scheduled Tribes. The first set of bars shows the average national mobility ranks. Upward mobility for the Forward/Others reference group is 42. Upward mobility is partially identified; we show the midpoint of the bounds, which in all cases span less than a single rank point. The following two sets of bars are calculated using within-state and within-district education ranks; they describe mobility gaps that are entirely within states and districts respectively. Panel B shows 95% confidence intervals for the Mincerian return to household log income (IHDS 2012), individual log wages (NSS 2012), and household log per capita income (NSS 2012). Panel C shows the share of individuals who report that they work in their own business, by social group and time. Panel D shows bottom half mobility for the major social groups, separated by individual business ownership. Scheduled Castes and Tribes are pooled to increase power, since few members of either group own businesses. Source for Panels C and D: NSS 2012.

Table A1

Transition Matrices for Father and Son Education in India

A. Sons Born 1950-59

	Son highest education attained						
Father ed attained	< 2 yrs. (31%)	2-4 yrs. (11%)	Primary (17%)	Middle (13%)	Sec. (13%)	Sr. sec. (6%)	Any higher (8%)
<2 yrs. (58%)	0.46	0.12	0.16	0.11	0.09	0.03	0.03
2-4 yrs. (12%)	0.10	0.18	0.21	0.19	0.16	0.09	0.07
Primary (13%)	0.07	0.08	0.31	0.16	0.19	0.08	0.10
Middle (5%)	0.06	0.05	0.09	0.30	0.18	0.14	0.18
Secondary (5%)	0.03	0.01	0.04	0.13	0.38	0.11	0.30
Sr. secondary (2%)	0.02	0.00	0.02	0.10	0.11	0.36	0.38
Any higher ed (2%)	0.01	0.01	0.01	0.03	0.09	0.13	0.72

B. Sons Born 1960-69

	Son highest education attained						
Father ed attained	< 2 yrs. (27%)	2-4 yrs. (10%)	Primary (16%)	Middle (16%)	Sec. (14%)	Sr. sec. (7%)	Any higher (10%)
<2 yrs. (56%)	0.40	0.12	0.16	0.14	0.10	0.04	0.03
2-4 yrs. (13%)	0.11	0.16	0.18	0.23	0.15	0.07	0.08
Primary (13%)	0.10	0.05	0.25	0.18	0.19	0.09	0.14
Middle (5%)	0.06	0.04	0.08	0.28	0.21	0.13	0.19
Secondary (6%)	0.03	0.02	0.08	0.11	0.35	0.17	0.25
Sr. secondary (2%)	0.02	0.01	0.03	0.07	0.20	0.25	0.42
Any higher ed (2%)	0.01	0.00	0.02	0.02	0.08	0.11	0.75

C. Sons Born 1970-79

	Son highest education attained						
Father ed attained	< 2 yrs. (20%)	2-4 yrs. (7%)	Primary (16%)	Middle (17%)	Sec. (16%)	Sr. sec. (10%)	Any higher (14%)
<2 yrs. (49%)	0.34	0.09	0.19	0.16	0.13	0.05	0.04
2-4 yrs. (11%)	0.10	0.13	0.18	0.24	0.18	0.08	0.10
Primary (12%)	0.08	0.06	0.22	0.19	0.18	0.12	0.14
Middle (6%)	0.05	0.01	0.07	0.27	0.21	0.17	0.21
Secondary (8%)	0.04	0.01	0.05	0.11	0.28	0.19	0.32
Sr. secondary (2%)	0.01	0.01	0.02	0.10	0.17	0.21	0.48
Any higher ed (4%)	0.00	0.00	0.02	0.03	0.10	0.14	0.70

D. Sons Born 1980-89

	Son highest education attained						
Father ed attained	< 2 yrs. (13%)	2-4 yrs. (6%)	Primary (16%)	Middle (23%)	Sec. (15%)	Sr. sec. (11%)	Any higher (16%)
<2 yrs. (35%)	0.25	0.10	0.21	0.24	0.10	0.05	0.04
2-4 yrs. (10%)	0.09	0.11	0.18	0.28	0.15	0.10	0.09
Primary (14%)	0.05	0.05	0.24	0.26	0.17	0.10	0.13
Middle (9%)	0.03	0.03	0.07	0.32	0.19	0.17	0.19
Secondary (9%)	0.01	0.00	0.05	0.16	0.25	0.22	0.30
Sr. secondary (4%)	0.01	0.01	0.04	0.09	0.16	0.22	0.47
Any higher ed (5%)	0.01	0.00	0.02	0.08	0.14	0.14	0.62

Table A1 shows transition matrices by decadal birth cohort for Indian fathers and sons. These data are visualized in Figure A3 for all father/mother-son/daughter dyads. Source: IHDS (2012).

Table A2

Internal Consistency of Reports of Parents' Education

	Father-Son		Father-Daughter		Mother-Daughter	
	(1)	(2)	(3)	(4)	(5)	(6)
Age		-0.000		-0.018		-0.008
		(0.008)		(0.016)		(0.007)
Child years of education		0.008		0.037*		0.003
		(0.013)		(0.021)		(0.011)
Log household income		-0.005		-0.051		-0.026
		(0.029)		(0.058)		(0.036)
Constant	0.053	0.054	-0.002	0.912	0.006	0.545
	(0.056)	(0.431)	(0.103)	(0.841)	(0.052)	(0.466)
N	1258	1255	440	440	726	725
r2	0.00	0.00	0.00	0.01	0.00	0.00

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Table A2 shows measures of internal consistency when there are multiple reports of an individual's father in the IHDS. We calculate the difference between a person's report of their parent's education and the parent's own reporting of it when in the same household. We then regress this difference on a constant (which provides the average difference, in Columns 1, 3, and 5), and on a series of household characteristics (Columns 2, 4, and 6). Source: IHDS (2012).

Table A3

Characteristics of Top and Bottom Half Individuals and Households

		Age 20–29		Age 50–59	
		Bottom Half	Top Half	Bottom Half	Top Half
Individual	Any wage	0.712	0.417	0.636	0.561
		(0.453)	(0.493)	(0.481)	(0.496)
	Log(wage)	2.875	3.237	2.969	3.785
		(0.510)	(0.699)	(0.609)	(0.947)
	Rural	0.738	0.548	0.762	0.483
		(0.440)	(0.498)	(0.426)	(0.500)
	Years of Education	4.988	12.167	1.850	10.487
		(3.017)	(1.647)	(2.170)	(2.254)
	Muslim	0.176	0.097	0.127	0.060
		(0.381)	(0.296)	(0.333)	(0.238)
	SC	0.259	0.196	0.240	0.130
		(0.438)	(0.397)	(0.427)	(0.336)
	ST	0.120	0.053	0.121	0.047
		(0.325)	(0.225)	(0.326)	(0.211)
Household	Log(income)	10.809	11.372	11.102	11.810
		(0.833)	(0.978)	(0.942)	(1.062)
	Log(per capita consumption)	9.654	10.133	9.764	10.301
		(0.554)	(0.641)	(0.606)	(0.677)

Table A3 shows summary statistics describing individuals from the bottom and top half of the education distribution, respectively. The individual statistics describe men born in 1983–92, and in 1953–62. The household statistics describe the households where those men reside. Standard deviations are in parentheses.

Table A4

Bottom Half Mobility Calculated Using Binned vs. Granular Education

Panel A: Binned Education

Group	1960–69	1980–89
All	[36.6, 39.0]	[37.1, 37.2]
Forward/Other	[41.8, 44.0]	[41.3, 41.3]
Muslim	[31.3, 33.6]	[28.9, 29.0]
Scheduled Castes	[32.9, 35.2]	[36.9, 37.0]
Scheduled Tribes	[29.4, 31.3]	[33.1, 33.1]

Panel B: Granular Education

Group	1960–69	1980–89
All	[36.5, 38.9]	[36.3, 37.2]
Forward/Other	[41.6, 43.7]	[41.1, 41.1]
Muslim	[31.2, 33.6]	[28.1, 29.3]
Scheduled Castes	[33.0, 35.2]	[36.5, 37.0]
Scheduled Tribes	[29.3, 31.3]	[33.4, 33.5]

Table A4 compares national and subgroup bottom half mobility when calculated using IHDS data downcoded to match standard education categories (Panel A, identical to Table 1) and using IHDS data with unadjusted granular years of education (Panel B). The results are similar because there are few individuals with education levels which were both in the bottom 50% and needed to be downcoded.

Table A5

Relationship Between Fertility and Subgroup Upward Mobility

	(1)	(2)	(3)
Muslim	-13.476***	-12.338***	-9.287***
	(0.976)	(1.697)	(1.721)
Scheduled Caste	-4.163***	-2.608**	-1.901
	(0.749)	(1.281)	(1.268)
Scheduled Tribe	-9.075***	-8.040***	-8.291***
	(1.076)	(1.851)	(1.829)
Urban	3.881***	3.812***	3.514***
	(0.782)	(1.276)	(1.261)
Number of Siblings			-2.359***
			(0.304)
N	6345	2347	2347
r2	0.11	0.15	0.18

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Table A5 shows estimates from regressions of child education rank on social group indicators and an individual's number of siblings, a proxy for mother's fertility. The sample is limited to individuals born in 1985–89 to fathers with two or fewer years of education. The outcome variable thus corresponds to μ_0^{51}, a close analog of bottom half mobility (μ_0^{50}). Column 1 shows the estimation without the fertility measure for the full sample. Column 2 limits the data to the set of individuals for whom mother's fertility can be measured, and Column 3 adds the fertility variable. The effect of fertility on subgroup mobility gaps is understood as the change in the subgroup coefficient from Column 2 to Column 3. All regressions control for state fixed effects. Source: IHDS (2012).

B Appendix B: Formalization of Bounds on CEF-based Mobility Measures

This Appendix provides details about the analytical and numerical procedures used to bound the CEF and functions of the CEF. These methods are straightforward applications of Novosad et al. (2022). In Appendix B.1 and Appendix B.2, we reproduce the text of several propositions contained in Novosad et al. (2022) for ease of reference, but relegate the proofs to Novosad et al. (2022). In Appendix B.3, we explain the simple procedure to adapt the numerical techniques in Novosad et al. (2022) to this setting.

Relationship to Novosad et al. (2022). Novosad et al. (2022) is concerned with estimating bounds on $E(y|x=i)$ and various functions of that CEF, where x is an interval-censored adult education rank and y is that same adult's mortality rate. This paper is concerned with the same mathematical problem, where x is an interval-censored parent education rank and y is a measure of child socioeconomic status. Note that the monotonicity condition here is similar to that in Novosad et al. (2022). Here, we assume child status is *increasing* in parent education rank; Novosad et al. (2022) assumes adult survivalship is *increasing* in adult education rank.

B.1 Formal Statement of Proposition 1

Let the function $Y(x)=E(y|x)$ be defined on [0,100]. Form the set of non-overlapping intervals $[x_k,x_{k+1}]$ that cover [0,100] for $k\in\{1,...,K\}$. We seek to bound $E(y|x)$ when x is known to lie in the interval $[x_k,x_{k+1}]$; there are K such intervals. Suppose that

$$x\sim U(0,100), \qquad \text{(Assumption U)}$$

and define

$$r_k:=\frac{1}{x_{k+1}-x_k}\int_{x_k}^{x_{k+1}}Y(x)dx.$$

Adopt the following assumptions from Manski and Tamer (2002):

$$\text{Prob}(x\in[x_k,x_{k+1}])=1. \qquad \text{(Assumption I)}$$

$$E(y|x) \text{ must be weakly increasing in } x. \qquad \text{(Assumption M)}$$

$$E(y|x,\, x \text{ is interval censored})=E(y|x). \qquad \text{(Assumption MI)}$$

Proposition 1. *Let x be in bin k. Under assumptions M, I, MI (Manski and Tamer, 2002) and*

U, and without additional information, the following bounds on $E(y|x)$ are sharp:

$$\begin{cases} r_{k-1} \leq E(y|x) \leq \frac{1}{x_{k+1}-x}((x_{k+1}-x_k)r_k-(x-x_k)r_{k-1}), & x < x_k^* \\ \frac{1}{x-x_k}((x_{k+1}-x_k)r_k-(x_{k+1}-x)r_{k+1}) \leq E(y|x) \leq r_{k+1}, & x \geq x_k^* \end{cases}$$

where

$$x_k^* = \frac{x_{k+1}r_{k+1}-(x_{k+1}-x_k)r_k-x_k r_{k-1}}{r_{k+1}-r_{k-1}}.$$

B.2 Formal Statement of Analytical Bounds on μ_a^b

We now state a proposition, also contained in Novosad et al. (2022), that permits us to bound μ_a^b.

Define

$$\mu_a^b = \frac{1}{b-a}\int_a^b E(y|x)di.$$

Let Y_x^{min} and Y_x^{max} be the lower and upper bounds respectively on $E(y|x)$ given by Proposition 1. We seek to bound μ_a^b when x is only known to lie in some interval $[x_k, x_{k+1}]$.

Proposition 2. *Let $b \in [x_k, x_{k+1}]$ and $a \in [x_h, x_{h+1}]$ with $a < b$. Let assumptions M, I, MI (Manski and Tamer, 2002) and U hold. Then, if there is no additional information available, the following bounds are sharp:*

$$\begin{cases} Y_b^{min} \leq \mu_a^b \leq Y_a^{max}, & h=k \\ \frac{r_h(x_k-a)+Y_b^{min}(b-x_k)}{b-a} \leq \mu_a^b \leq \frac{Y_a^{max}(x_k-a)+r_k(b-x_k)}{b-a}, & h+1=k \\ \frac{r_h(x_{h+1}-a)+\sum_{\lambda=h+1}^{k-1} r_\lambda(x_{\lambda+1}-x_\lambda)+Y_b^{min}(b-x_k)}{b-a} \leq \mu_a^b \leq \frac{Y_a^{max}(x_{h+1}-a)+\sum_{\lambda=h+1}^{k-1} r_\lambda(x_{\lambda+1}-x_\lambda)+r_k(b-x_k)}{b-a}, & h+1<k. \end{cases}$$

B.3 Bounding Functions of the CEF

We now describe our numerical procedure for bounding arbitrary functions of the CEF. The key simplification is to partition the CEF into a step function with M steps; this gives us a highly flexible shape for the CEF but lets us iterate over a finite set of possible CEFs. We describe the process for M=100.

We conduct the following process.

1. Consider the set of CEFs that can: (a) match the observed mean levels of child rank within each parent rank bin, and (b) are consistent with any additional assumptions (e.g., monotonocity and/or smoothness assumptions).
2. For every CEF in this set, generate a function of the CEF. Report the maximum and minimum value of this function, collecting values over all CEFs in this set.

Formally, index interval-censored bins by k: define the non-overlapping intervals $[x_k, x_{k+1}]$ that cover [0,100] for $k \in \{1,...,K\}$. Then define $\{r_k\}_{k=1}^K$ as the set of observed mean values of y over each bin $k \in \{1,...,K\}$. Further define $S(\{r_k\}_{k=1}^K)$ to be the collection of CEFs that is consistent with these bin means and any desired auxiliary assumptions. For example, noting that x is uniformly distributed, we can put:

$$S\big(\{r_k\}_{k=1}^K\big) = \Big\{Y(x) \mid Y(x) \text{ is weakly increasing}\Big\} \\ \bigcap \left\{Y(x) \middle| \frac{1}{x_{k+1}-x_k}\int_{x_k}^{x_{k+1}} (Y(x)-r_k(x))dx = 0, \text{ for all } k\right\}. \tag{B.1}$$

Our objective is to bound $\gamma = \gamma(Y)$, some function of the CEF. In particular, we face the following constrained optimization problem to obtain the maximum and minimum values of γ:

$$\gamma^{\min} = \min_{Y \in S\left(\{r_k\}_{k=1}^K\right)} \tilde{\gamma}(Y) \tag{B.2}$$

$$\gamma^{\max} = \max_{Y \in S\left(\{r_k\}_{k=1}^K\right)} \tilde{\gamma}(Y). \tag{B.3}$$

Novosad et al. (2022) provide details on the numerical techniques used to solve this problem. The bounds we report are the set $[\gamma^{\min}, \gamma^{\max}]$. For the case of the rank-rank gradient (the only time the numerical optimization is needed in the paper), we let γ represent the slope of the linear approximation to the CEF. That is, fixing a CEF $Y(x)$, define

$$(\gamma, b) := \operatorname*{argmin}_{\gamma', b' \in \mathbb{R}} \int_0^{100} (Y(x) - \gamma' x + b')^2 dx.$$

We then use Equations B.2 and B.3 to calculate the minimum and maximum γ' that can be generated from the set of valid CEFs. These form the bounds on the rank-rank gradient for a given set of moments.

C Appendix C: Robustness to Alternate Assumptions

C.1 Robustness to Non-Uniform Within-Bin Subgroup Distributions

Our bounds on the full sample CEF $E(y|x)$ (See Section 3.2 and Appendix C) rely on the uniformity of the rank distribution, which is given when working with a national sample. However, when working with population subsamples (e.g. Muslims), uniformity is not guaranteed. Take the example of the 1960s, where 57% of fathers are in the lowest education bin. Conditional on being in the bottom bin, the distribution of latent ranks of Muslim fathers is *not* necessarily uniform.

This lack of uniformity creates a potential bias. For example, approximately 10% of the fathers in the bottom education bin are Muslims. If the latent ranks of these fathers were all concentrated at the bottom of the bin, and the latent ranks of Hindus were concentrated at the top of the bin, then the mobility gap between Hindus and Muslims would be biased upward. In other words, the gap in son outcomes between Hindus and Muslims could be driven not only by a difference in outcomes conditional on father education rank, but also by unobserved differences in the latent father ranks.

The extent of bias is determined by the extent to which the within-bin latent education rank distribution for each subgroup differs from the uniform distribution and how that difference changes over time. In this section, we present three pieces of evidence that these departures do not substantially bias our primary results.

First, we examine whether there is substantial socioeconomic divergence between SCs and Muslims in the parent generation, using additional data. The evidence rejects a large enough change to explain the SC/Muslim mobility divergence. Second, we show that the divergence of upward mobility between Scheduled Castes and Muslims is found even when we rank parents according to their position in the education distribution of their own subgroup—given this ranking, the latent rank distribution within each bin is guaranteed to be uniform, eliminating the bias threat (at the cost of calculating a slightly less useful mobility statistic). Third, we use parametric assumptions to estimate the latent rank distribution suggested by the distribution of education completion across bins. We show that the maximal bias under a range of parametric assumptions is very small and unlikely to affect our conclusions.

The issues addressed in this appendix are not unique to our analysis, but are implicit in any comparison of groups that conditions on education levels. However, our discussion of latent education ranks makes this concern particularly visible.

C.1.1 Socioeconomic Changes for Muslims and SCs in Parent Generations

Figure C1 shows time series plots with various socioeconomic indicators representing the parent generation of our sample. We focus on SC and Muslim outcomes, as we aim to test the hypothesis that changes in relative positions at the bottom of the socioeconomic distribution drive the relative mobility changes documented for these groups in the paper. Panel A and B show education levels and ranks of the *parents* of the 1950–89 birth cohorts. Muslim parents have higher education in all years; there is a partial convergence of about one rank point — equivalent to less than half a year of education — between the two groups. A convergence of this size is far too small to explain the 7 rank point (or $\sim$1.5 years of education) that has opened between bottom-half children in these groups.

However, these estimates are from the full distribution of parents; perhaps Muslims at the bottom of the distribution have done relatively worse than those at the top. We cannot, of course, compare education changes in the bottom half of the distribution, since they are entirely bottom-coded. We therefore turn to household consumption, looking at individuals aged 40–60 in NSS samples from 1983–2012.[58] We limit the sample to individuals in the bottom half of the education distribution in their cohort/year. Panel C shows the log consumption gap between Muslims and Forwards/Others, and the same gap between SCs and Forwards/Others, from 1983–2012. Panel D shows the same result in terms of consumption ranks. The gaps are largely stable over the sample period; there is no evidence that bottom-half Muslims have lost ground to members of Scheduled Castes over the sample period.

In short, there is little evidence to suggest that Muslim bottom-half parents of the 1980s were particularly negatively selected as compared to Muslim bottom-half parents of the 1950s or 1960s cohorts.

[58]To our knowledge, the 1983 NSS is the earliest electronically available NSS with per capita consumption recorded. 2012 is the last NSS survey year available. If fathers are 20–30 when their children is born, this set of surveys covers the parents of birth cohorts ranging from 1943–1992, thus approximately covers the parents of children in our main sample.

Figure C1

Trends in Socioeconomic Status for Muslims and SCs in Parent Generations

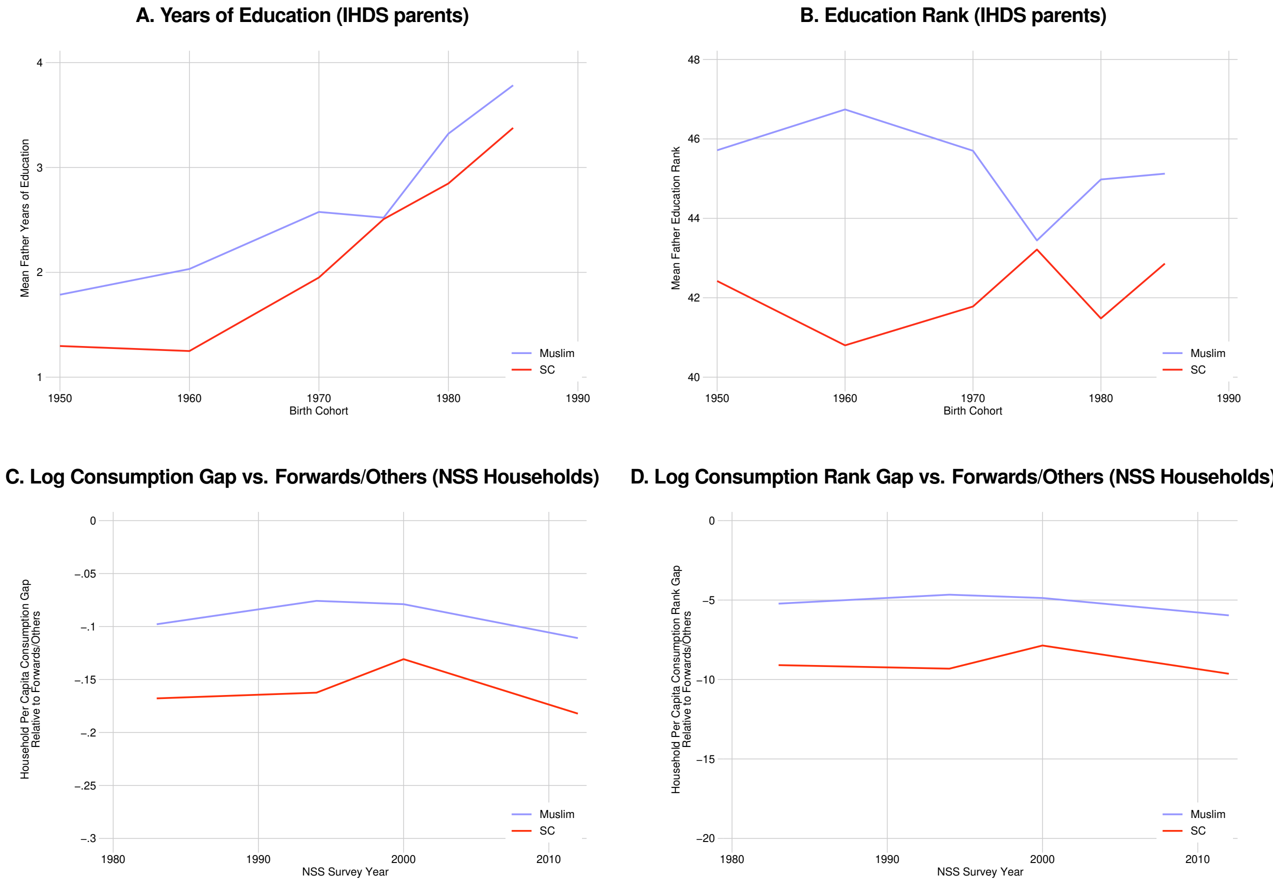


Figure C1 shows trends in socioeconomic status of SCs and Muslims. Panels A and B present education levels (years) and within-cohort education ranks for the *parents* of male respondents in the birth cohorts on the X axis. Ranks are calculated using the midpoint rank of education bins. Panel C uses household-level NSS data to show the log consumption gap between Muslims and Forwards/Others (blue) and SCs and Forwards/Others (red), for households where the head is aged 30–50 and *is in the bottom half of the education distribution.* The X axis shows the NSS survey year. Panel D shows the consumption rank gaps for the same surveys/groups. Source: IHDS (2012), NSS (1983–2012).

C.1.2 Using Within-Subgroup Rank Distributions which are Uniform by Construction

We show here that the Muslim-SC mobility divergence is robust to calculating parent ranks *within* subgroups. Under this rank definition, latent parent ranks within each subgroup are uniform by construction—the latent ranks of SCs in the bottom 50% of SCs must be uniformly distributed. This fully resolves the non-uniform bias problem, but the cost of making this assumption is that we are no longer comparing groups with similar levels of education—the least educated 50% of SCs have a lower level of education than the least educated 50% of Forwards and thus cannot be expected to attain the same outcomes even if there are no cross-group outcome differences after controlling for parent education. For this reason, we use national ranks in the body of the paper. Note that SC and Muslim parents have similar levels of education (much more similar than their children, see Figure C1), making the latter concern less important here as well.

Figure C2 shows the result. Panel A repeats the result of Figure 6 for Forwards/Others, Muslims, and SCs, using national ranks, showing changes in upward mobility (μ_0^{50}) over time for each group. Panel B shows the same result, with parent ranks calculated within their own subgroups. The bounds in Panel B are too wide to distinguish mobility changes between SCs and Muslims, because the within-rank bottom-coding problem is more severe among the marginalized groups, where the parent generation is less educated. More than 70% of SC parents in the 1960s report a bottom-coded education level, resulting in wide bounds on μ_0^{50} for this rank definition.

To tighten the bounds, we instead estimate μ_0^{70}: the expected child outcome given a parent in the bottom 70% of the parent education distribution. Panel C shows μ_0^{70} calculated using national ranks, as in the body of the paper. Panel D shows μ_0^{70} calculated using own-subgroup ranks, as in Panel B. The divergence of SCs and Muslims, and the convergence of SCs and Forwards/Others is sustained in both of these panels. The level gap between SCs/Muslims and Forwards/Others is higher in Panels B and D because the bottom x% of SCs/Muslims represent lower levels of education than the bottom x% of Forwards/Others, whereas Panels A and C hold parent education constant.

The consistency of these results with parent ranks calculated within subgroups (which are uniform by construction) strongly suggests that our primary results are not driven by differential unobserved changes in the latent parent rank distributions of the individual subgroups.

Figure C2

Subgroup Upward Mobility (Fathers/Sons):
National Ranks vs. Within-Subgroup Ranks

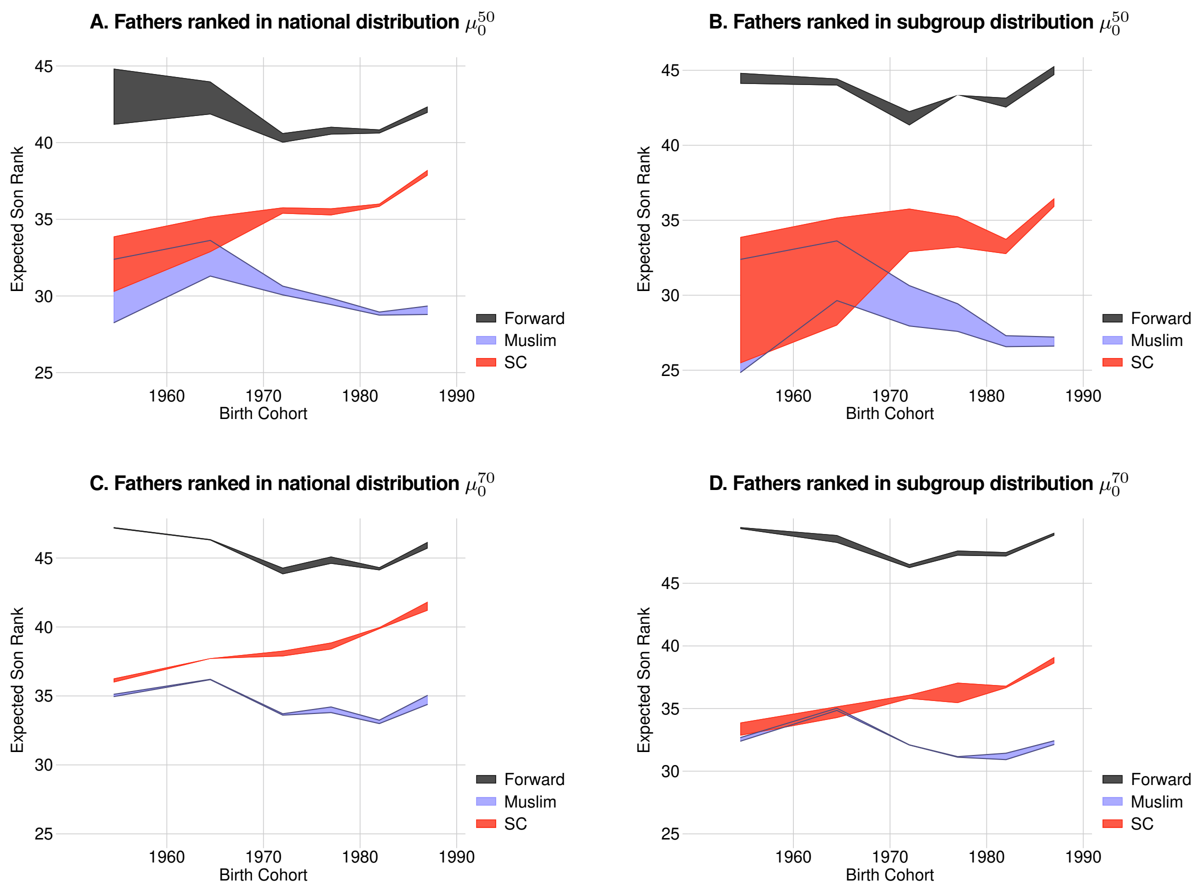


Figure C2 shows trends in upward mobility for Forward/Others, Muslims, and SCs. Panel A presents bottom-half mobility by ranking fathers in the national distribution, as in the body of the paper. Panel B ranks fathers within each subgroup, recovering uniformity by construction. Panels C and D are similar to A and B, except they present bounds on μ_0^{70} (i.e., average son rank, conditional on being born to a father in the bottom 70%) rather than μ_0^{50}.

C.1.3 Inferring Latent Education Ranks Using Parametric Assumptions

This section takes an alternate route to modeling the unobserved variation at the bottom of the education distribution. We assume that the latent rank distribution takes a parametric form (normal or lognormal); we then estimate the full distribution using the cross-bin education distribution for every group.

These parameterized distributions let us produce a continuous latent rank distribution for each subgroup. In the body of the paper, we assumed this distribution was uniform for each group within each bin. The parameterizations let us predict separate within-bin latent rank distributions for each group, based on each group's full distribution of education. We can compare these predicted latent distributions to the uniform distribution used in the paper to gauge the extent of bias that could arise from the uniformity assumption.

For each population subgroup, we fit a normal and a lognormal distribution to the sample distribution of years of education. We then create a simulated population that has the same proportion of each subgroup as the true population, and for each individual, we draw their years of education from the fitted parametric distribution. This gives us a continuous education distribution that matches the moments from the discrete sample distribution. Finally, we transform the years of education variable into ranks with respect to the entire population. This gives us a simulated population with continuous ranks.

We focus on the 1960–69 and 1985–89 cohorts, as we aim to check the validity of our conclusion that Muslim and SC mobility have diverged over this period.

Table C1 compares the moments from the IHDS sample with the moments from the simulated distributions. Here, we are ignoring children and only examining how far the parent latent rank distribution *across groups* differs from what we would get from the uniformity assumption.

For both the 1960–69 and the 1985–89 birth cohorts, the simulated moments are close matches to the raw data. The group ordering and approximate gaps between groups is preserved; the standard deviation of the simulated data is slightly higher than that of the true binned data, which is to be expected, given the truncation of the binned data.

In Table C2, we use the simulated data to examine the distribution of the latent variable *within* the bins where our method assumed uniformity in the main analysis. Specifically, we examine the mean parent education rank conditional on being in the bottom 50%. As expected, parents from less educated social groups have lower latent ranks even after conditioning on being in the bottom 50%.[59] However, the differences are very small, and they do not change much from the 1960–69 to the 1985–89 birth cohorts, under any of the distributional assumptions. Even in the worst case scenario (the lognormal distribution with constant variance), the gap between Muslim and SC parents in the bottom 50% shrinks from 2.5 to 1, a 1.5 percentage point change.

Given the average CEF slope of 0.5, this suggests that changing latent parental status within the coarse education bins can explain at most 0.75 rank points of the growing difference between

[59] If the subgroup distributions were all uniform within this bin, then all groups would have a mean rank of 25.

Table C1

Actual and Simulated Moments from the Education Rank Distribution

A. 1960-1969 Birth Cohort

	Binned Data		Simulated Distributions			
			Normal		Lognormal	
	Mean	SD	Mean	SD	Mean	SD
Forward / Other	55.2	26.8	55.8	30.1	55.9	29.4
Muslim	46.7	24.5	46.6	27.4	46.2	28.1
Scheduled Castes	40.8	21.0	39.6	23.0	39.7	24.6
Scheduled Tribes	39.1	20.2	38.5	23.3	37.7	23.9

B. 1985-1989 Birth Cohort

	Binned Data		Simulated Distributions			
			Normal		Lognormal	
	Mean	SD	Mean	SD	Mean	SD
Forward / Other	56.5	27.6	56.5	28.6	56.6	27.9
Muslim	45.1	26.7	44.8	27.5	45.4	28.2
Scheduled Castes	42.9	26.7	42.8	27.5	42.6	28.1
Scheduled Tribes	35.6	25.0	36.1	24.9	34.9	26.1

Table C1 shows the mean and standard deviation of the true data (IHDS), compared with the mean and standard deviation of simulated distributions, split by demographic subgroup.

Muslims and SC/STs. Under other distributional assumptions the potential bias is even smaller. In comparison, our midpoint estimate of this change from 1960–69 to 1985–89 in the body of the paper is 7.4 rank points. This is not a strict upper bound, because child outcomes could be correlated with latent ranks, something we do not address here. But the scale of the effect on the average rank difference conditional on being in the bottom half strongly suggests that non-uniformity within bins is not a major factor explaining our results.

In short, we consider it unlikely that changing parental position within observed rank bins can explain the growing mobility gap between SCs and Muslims. All the evidence brought to bear suggests that the relative positions of these groups within the bottom education bins has not changed enough to substantially bias our group-level estimates.

Table C2

Simulated Average Parent Rank Conditional on Rank ≤ 50

A. 1960-1969 Birth Cohort

	Group-level Variance		Constant Variance	
	Normal	Lognormal	Normal	Lognormal
Forward / Other	24.4	25.4	27.0	27.2
Muslim	24.9	24.1	24.3	24.5
Scheduled Castes	26.0	24.9	22.6	22.2
Scheduled Tribes	25.2	24.3	21.8	21.9

B. 1985-1989 Birth Cohort

	Group-level Variance		Constant Variance	
	Normal	Lognormal	Normal	Lognormal
Forward / Other	26.6	27.6	27.5	27.2
Muslim	24.4	24.0	24.1	24.4
Scheduled Castes	23.7	23.2	23.2	23.3
Scheduled Tribes	22.8	21.1	21.0	21.1

Table C2 presents the simulated average parent rank conditional on being in the bottom half of the distribution under two parametric distributions (normal and lognormal). The left panel estimates distribution mean and variance separately for each demographic subgroup; the right panel uses the same variance for each distribution, estimated from all the data.

D Appendix D: Data Construction

This section describes the data sources and data construction in detail.

D.1 IHDS

The India Human Development Survey (IHDS) is a nationally representative survey of 41,554 households, with rounds in 2004–05 and 2011–12. Definitions of social groups are described in the body of the paper. This section focuses on linking parents to children.

The primary module of IHDS records the education of the father of the household head. A secondary module, the women's survey, records the education of the father and mother of the female respondent, as well as the father and mother of her husband if she is married. The women's survey is given to one or two women aged 15–49 in each household. Because of the upper age restriction on the women's survey, the oldest daughter in our sample is born in 1962; we therefore do not have any links from mothers or links to daughters for the 1950–59 birth cohort.

Finally, we created additional parent-child links using information from the relationship field in the household roster. Specifically, we linked the household head to their children and parents. We linked the spouse of the household head to their children. We linked grandchildren of the household head to the child of the household only in cases where there was no possible ambiguity about the parents of the grandchildren. In cases with no possible ambiguity, we linked nieces/nephews of the household head to brothers of the household head. We did not link individuals on the basis of in-law relationships, because of the ambiguity in the definition of the sibling-in-law (i.e. sibling of spouse vs. spouse of sibling).

In many cases, a parent's education is recorded in multiple ways, allowing us to cross-check the validity of the responses. For example, the household head's father's education may be obtained from (i) the household roster (if he is coresident); (ii) from the household head's response to the father education question; and (iii) from his wife's responses to the husband's father's education question. The average correlation between parent education measured across different sources is 0.9. Appendix Table A2 shows that the discrepancies between measures are not correlated with household characteristics.

D.2 Data from other countries

We refer in the paper to mobility data from several other countries. Data from Denmark, Sweden, and Norway were generously shared with us by Boserup et al. (2014) and Bratberg et al. (2015). Income mobility estimates for the U.S. were drawn from Chetty et al. (2014b) and Chetty et al. (2020). Educational mobility estimates from the U.S. were calculated from a parent-child education transition matrix describing children in the 2005–2015 ACS and parents in the 2000 Census, from the data package of Chetty et al. (2020).