

# Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities\*

Sam Asher<sup>†</sup>

Paul Novosad<sup>‡</sup>

Charlie Rafkin<sup>§</sup>

January 2021

## Abstract

We study intergenerational mobility in India over time, across groups, and across space. We show that the modern set of rank-based mobility measures can be at best partially identified with education data. We develop a new measure of upward mobility that works well under data constraints common in developing countries. We find that intergenerational mobility in India has been constant and low since before liberalization. Among boys, rising mobility for Scheduled Castes is almost exactly offset by declining mobility among Muslims, a comparably sized group with few constitutional protections. Mobility among girls is lower, with less cross-group variation over time. Mobility is highest in places that are southern, urban, and have high average education levels. A natural experiment suggests that affirmative action for Scheduled Castes has substantially improved their mobility. Our measures are relevant for the study of mobility in poorer countries and in historical contexts.

---

\*We are thankful for useful discussions with Alberto Abadie, David Autor, Emily Blanchard, Raj Chetty, Eric Edmonds, Shahe Emran, Francisco Ferreira, Nate Hilger, Larry Katz, David Laibson, Ethan Ligon, Erzo Luttmer, Whitney Newey, Elias Papaioannou, Nina Pavnik, Bruce Sacerdote, 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.”

<sup>†</sup>Johns Hopkins University, sasher2@jhu.edu

<sup>‡</sup>Dartmouth College, paul.novosad@dartmouth.edu, corresponding author

<sup>§</sup>MIT, crafkin@mit.edu

## 1 Introduction

There are two widely held narratives regarding access to opportunity in India. On the one hand, economic liberalization, rapid economic growth, and urbanization have vastly expanded the set of opportunities available to Indians, leading to the emergence of a large middle class. The political sphere has also opened, with the emergence of a wide range of parties organized around caste, region, and ideology. Decades of affirmative action programs have targeted government benefits at historically disadvantaged groups. On the other hand, some of India's entrenched inequalities seem as persistent as ever. Marriage across religious, caste, and class lines is exceedingly rare. Elites in business, government and civil society are still largely from upper classes and castes. Inequality has risen, and religious cleavages may be deepening (Chancel and Piketty, 2019). In this paper, we shed light on changing access to opportunity in India by studying the intergenerational transmission of economic status (Solon, 1999; Black and Devereux, 2011; Chetty et al., 2014a; Chetty et al., 2020), with a particular emphasis on the 478 million people (39%) who belong to India's major disadvantaged groups (Muslims, Scheduled Castes, and Scheduled Tribes).

We focus on measuring the persistence of socioeconomic *rank* across generations, isolating intergenerational mobility from changes in inequality and growth, in the spirit of Solon (1999). We develop a set of methods that makes it possible to apply modern rank-based measurements of upward mobility in contexts where coarse educational outcomes are the only viable measure of socioeconomic status that can be linked across generations. These methods may be useful in studies of intergenerational mobility in other developing countries, as well as in historical contexts in richer countries.

Because of data quality and availability, as well as the challenge of measuring individual income in households with joint production, studies of intergenerational mobility in developing countries (and in historical contexts) often use education as a proxy for social status.<sup>1</sup> Moreover, canonical intergenerational mobility models like Loury (1981), Becker and Tomes (1986), and Galor and Zeira (1993) often emphasize the role of human capital investment. A key challenge with educational

---

<sup>1</sup>Linked parent-child education data are also much more widely available than linked income data. Recent studies of intergenerational mobility focusing on education include Black et al. (2005), Güell et al. (2013), Wantchekon et al. (2015), Card et al. (2018), Derenoncourt (2019), and Alesina et al. (2021). More are summarized in Black and Devereux (2011).

mobility is that education data are often coarsely measured; for instance, for the 1960–69 birth cohort in India, over 50% of fathers and 80% of mothers report a bottom-coded level of education. This makes it difficult to use rank-based measures of mobility, such as *absolute mobility*, which require observing parents at specific percentiles in the socioeconomic status distribution.<sup>2</sup> Studies of educational mobility have instead focused on estimators like the correlation coefficient between parents' and children's educational outcomes. These linear estimators have several limitations (discussed in Section 3.1), the most important of which is that they are not meaningful for subgroup analysis, because they measure individuals' progress only against other members of their own group (Hertz, 2005).<sup>3</sup>

In this paper, we show that with education data, the new generation of rank-based estimators of intergenerational mobility can at best be partially identified. Intuitively, when income mobility estimators are applied directly to educational mobility, they do not account for the loss of information associated with coarse measurement of ranks; they instead rely on implicit and untested assumptions about the latent rank distribution.<sup>4</sup> We treat the estimation of child outcomes conditional on latent parent ranks as an interval data problem; for each mobility measure, we calculate the *set* of values that are consistent with a latent conditional expectation function (CEF) that generates the coarsely observed moments.<sup>5</sup>

We introduce a new measure of upward mobility, *bottom half mobility*, which is the expected rank of a child born to a parent in the bottom half of the education distribution. Bottom half mobility has a similar interpretation to other measures of upward mobility, but it can be bounded

---

<sup>2</sup>Chetty et al. (2014a) define absolute mobility at percentile  $i$  as the expected income rank of a child, conditional on that child being born to a parent at the  $i$ th income percentile.

<sup>3</sup>For example, the parent-child rank-rank gradient among U.S. Blacks is nearly identical to that among U.S. Whites, even though Black children obtain considerably worse outcomes in expectation than White children at every percentile of the parent income distribution (Chetty et al., 2020).

<sup>4</sup>The notion of a latent distribution of education ranks arises directly out of the standard human capital model (Card, 1999); individuals who are close to the margin of obtaining the next discrete level of education are those with high latent ranks in each bin. Note that because we use education as a proxy for socioeconomic status, the latent education rank is the parameter of interest rather than the precise number of years of education obtained. We discuss this concept further in Section 3.3.

<sup>5</sup>We use a standard approach to interval data established by Manski and Tamer (2002) and extended to the problem of estimating outcomes conditioning on coarse education ranks in Novosad et al. (2020). We generate this set by using best- and worst-case assumptions about the underlying data generating process that are consistent with the data we do observe.

tightly even in contexts with extreme interval censoring.<sup>6</sup> In contrast, once prior measures (absolute upward mobility and the rank-rank gradient) are adjusted to account for the underlying uncertainty associated with interval data, their bounds become too wide to be meaningful. This paper thus relaxes the hidden assumptions underlying most mobility estimators in settings with education data and still obtains precise (if partially-identified) mobility estimates. To our knowledge, bottom half mobility is the first measure of intergenerational educational mobility that can be meaningfully compared across time and space, across countries, and across population subgroups.<sup>7,8</sup>

We apply our measure to India, using data from the 2012 India Human Development Survey (IHDS) and the 2012 Socioeconomic and Caste Census (SECC).<sup>9</sup> We document trends in educational mobility from the 1950–59 to the 1985–89 birth cohorts.<sup>10</sup> We focus on measuring mobility from fathers to sons and daughters; mobility from mothers to children cannot be bounded tightly because the bottom-coding of mothers' education is so severe.

We present three main findings. First, upward mobility has remained constant for the past several decades, despite dramatic gains in average levels of education and income. An Indian son born in the bottom half of the parent education distribution in 1985–89 (our youngest cohort) can expect to obtain the percentile 37.7; a daughter obtains percentile 35.6.<sup>11</sup> A similar child in the U.S., which has low intergenerational mobility by OECD standards, on average attains education percentile

---

<sup>6</sup>Bottom half mobility describes the expected outcome of children conditional on having a parent in the bottom half of the distribution. Absolute upward mobility Chetty et al. (2014a) conditions on the *median* parent in the bottom half of the distribution. If the conditional expectation function is linear in parent rank, the two measures are identical. If the CEF is concave, then bottom half mobility puts more weight on the outcomes of the least privileged children. Note that the linear parent-child income rank CEF in the United States is an exception rather the rule; in most countries, these CEFs are non-linear (Bratsberg et al., 2007; Boserup et al., 2014; Bratberg et al., 2015; Connolly et al., 2019).

<sup>7</sup>Hertz (2008) provides a decomposition of intergenerational mobility that permits comparisons across subgroups. Like work before Chetty et al. (2014), this is a linear estimator that is difficult to use with interval-censored data.

<sup>8</sup>Our approach can also be used to calculate other measures based on the CEF, such as the expected outcome a child born in the bottom 40%, or the median outcome of a child born in the bottom half; our results focus on bottom half mobility because it is a very close analog to the widely used measure of absolute upward mobility.

<sup>9</sup>The former is a sample survey, and the latter is a socioeconomic census with high geographic resolution covering all individuals in the country.

<sup>10</sup>Our main estimates are not subject to concerns about coresidence bias, because they link children to parents even when children are not in the same household. Our geographically precise estimates are restricted to coresident father/son pairs due to data limitations; we restrict this sample to sons aged 20–23 and show that the sample selection bias for this group is likely to be very small.

<sup>11</sup>Following convention (Chetty et al., 2014b; Chetty et al., 2020), we always rank children in the own-gender distribution.

41.7.<sup>12,13</sup> This suggests that India's decades of economic growth have lifted the economic status of individuals in the bottom half of the socioeconomic distribution without substantially changing their likelihood of moving to a higher socioeconomic *rank*.<sup>14</sup>

Second, we show that there are significant changes in the cross-group distribution of upward mobility over time, particularly among sons. We divide the population into Scheduled Castes (SCs), Scheduled Tribes (STs), Muslims, and Forwards/Others.<sup>15</sup> Consistent with prior work (Hnatkovska et al., 2012; Emran and Shilpi, 2015), we find that sons from India's constitutionally protected marginalized groups, the Scheduled Castes and Tribes, have closed respectively 50% and 30% of the mobility gap to Forwards/Others. In contrast, upward mobility for Muslim sons has steadily declined from the 1960s to the present. The expected educational rank of a Muslim boy born in the bottom half of the parent distribution has fallen from between percentiles 31 and 34 to a dismal 29. Muslim sons have considerably worse upward mobility today than both Scheduled Castes (38) and Scheduled Tribes (33), a striking finding given that, compared to Muslims, STs tend to live in much more remote and low mobility areas. The comparable figure for U.S. Black men is 35.<sup>16</sup> Higher caste groups have experienced constant and high upward mobility over time, a result that contradicts a popular notion that it is increasingly difficult for higher-caste Hindus to get ahead.

Our measures for father-daughter mobility are less precise, but the subgroup patterns appear

---

<sup>12</sup>In a society where children's outcomes are independent of parents (i.e. perfect mobility), a child born in the bottom half of the distribution obtains the 50th percentile on average. In a society with no upward mobility, (i.e. where all children obtain the same percentile as their parents) the same child attains the 25th percentile.

<sup>13</sup>All of our mobility estimates are robust to different data construction methods, and we show that survivorship bias, migration, or bias in estimates from coresident parent-child households do not substantially affect our results. We also show that unobserved changes in the latent rank distribution of population subgroups within education bins cannot drive the secular changes that we document.

<sup>14</sup>A naive application of the canonical rank-rank gradient estimator would have suggested that mobility has improved significantly. This finding is driven entirely by the top 10 parent education percentiles, and is not robust to accounting for interval censoring in the education rank data.

<sup>15</sup>We include non-Muslims in Other Backward Castes (OBC) in the "Others" category. Measuring OBC mobility is challenging because OBC definitions are less stable over time, are sometimes inconsistently classified between federal and state lists, and may be reported inconsistently by the same individual over time. These concerns apply to SC and ST groups, but at a considerably smaller scale. OBCs also did not gain affirmative action benefits analogous to those targeted at SCs and STs until 1992, after the births of our final cohorts. The very small number of Muslim SC/STs are categorized as Muslims; reclassifying them as SCs or STs, or excluding Sikhs, Jains and Christians from the "others" category do not affect our results.

<sup>16</sup>This was calculated using the methodology in this paper and education data from Chetty et al. (2020). Bottom half *income* mobility for U.S. Black men is 39 (Chetty et al., 2020).

to be different. Daughters from poor Muslim, SC, and ST households all have persistently lower mobility than Forwards/Others, and there is minimal convergence over the sample period.

Third, we describe substantial variation in upward mobility across 5,600 rural subdistricts and 2,000 cities and towns. Paralleling results from Chetty et al. (2014b), we find substantial heterogeneity even within small geographic regions. Upward mobility is highest in urban areas, and in places with high consumption, education, school supply, and manufacturing employment, which are broad correlates of development. High mobility is negatively correlated with caste segregation and land inequality. Geography-subgroup interactions are important; for instance, daughters have higher mobility than sons in urban areas, but lower in rural areas.

The final section of the paper examines several potential mechanisms for the divergence of Scheduled Castes from Muslims over the last 30 years. We show that this divergence cannot be explained by differential returns to education, occupational patterns, geography, or differential fertility. However, we find suggestive evidence that the basket of affirmative action policies targeted to India's scheduled groups (but not to Muslims) may have had a substantial impact on their mobility. Following Cassan (2019), we study a natural experiment that added many castes to the Scheduled Caste lists in 1977. We show that when a caste group gets assigned to Scheduled Caste status, it experiences on average a 7–8 rank point increase in upward mobility over the next twenty years. This is the same size as the rank mobility gap that has opened between Muslims and Scheduled Castes over the same period. Our findings are thus consistent with the claim that educational quotas, government job reservations, and other affirmative action policies may drive the upward mobility gap that has opened between Scheduled Castes and Muslims. However, because we are limited to birth cohort  $\times$  demographic group variation, these results are suggestive and not dispositive.

**Contributions and literature review.** Our paper's contributions are both methodological and empirical. While bounding methods along the lines of Manski and Tamer (2002) and Novosad et al. (2020) are familiar in the econometrics literature, we are the first to apply these methods to the setting of educational mobility. We both identify and address methodological challenges that make it difficult to interpret research on educational mobility. In particular, we define a new parameter,

bottom half mobility, which is the first educational mobility statistic that is valid for comparing population subgroups across different contexts. As a result, our tools have the potential to be useful in other settings, for instance in cross-country comparisons of educational mobility.

Prior researchers have used CEF-based mobility measures to examine subgroup outcomes, but the coarse measurement problem has forced them to use inconsistent measures over time and across contexts. For example, Card et al. (2018) and Derenoncourt (2019) define upward mobility in the 1920s as the 9th grade completion rate of children whose parents have 5–8 years of school (or approximately parent percentiles 45–70).<sup>17</sup> Relatedly, Alesina et al. (2021) define upward mobility in Africa as the likelihood that a child born to a parent who has not completed primary school manages to do so. While these measures capture the ability of children to exceed the education levels of their parents, they do not distinguish between average educational gains and changes in the ability of individuals to move up the socioeconomic distribution in *relative* terms, the latter being central to the mobility definitions of Solon (1999) and Chetty et al. (2014b). These measures thus combine upward mobility and economic growth.<sup>18</sup> The key advantages of our measure are that it isolates upward mobility from both aggregate growth and changes in inequality, and it can be used to compare groups from similar points in the parent rank distribution.

Empirically, we present several previously unknown facts about upward mobility in India. Our findings imply that virtually all of the upward mobility gains in India over recent decades have accrued to Scheduled Castes and Tribes, groups that have constitutional protections, reservations in politics and education, and who have been targeted by many development policies. There is no evidence that any of these gains have come at the expense of higher-caste groups. On the other hand, mobility has declined for Muslims. We are not aware of studies of intergenerational mobility for Indian Muslims, even though they number almost 200 million people (higher than the number of Scheduled Castes).<sup>19</sup>

---

<sup>17</sup>Card et al. (2018) then compare this measure with absolute upward mobility (i.e., children of parents at the 25th percentile) in the present.

<sup>18</sup>They also condition on substantially different parts of the education distribution in different times and places. Card et al. (2018) measure  $E(y > 50 | x \in [45, 70])$ . Alesina et al. (2021) measure, for example,  $E(y > 52 | x \in [0, 76])$  in Mozambique (where 76% of parents and 48% of children have not completed primary), but to  $E(y > 18 | x \in [0, 42])$  in South Africa, where  $y$  is child rank and  $x$  is parent rank. See Section 3.5 for more detail.

<sup>19</sup>Other economics papers on Indian Muslims include Khamis et al. (2012) and Bhalotra and Zamora (2010), who note poor education outcomes among Muslims. The Sachar Committee Report (2006) and Basant et al. (2010)

These empirical estimates contribute to a burgeoning literature on intergenerational mobility, especially in the developing world. Black and Devereux (2011) provide a review of cross-country estimates of intergenerational educational mobility, and recent work has examined IM in the United States (Chetty et al., 2014b; Chetty et al., 2014a; Chetty et al., 2020), western Europe (Bratberg et al., 2015), or Africa (Alesina et al., 2021), among many other regions. The literature on India has: (i) emphasized absolute outcomes (such as consumption), which are rising for all groups due to India's substantial economic growth (Maitra and Sharma, 2009; Hnatkovska et al., 2013); or (ii) compared subgroups using the parent-child outcome correlation or regression coefficient, which describes the outcomes of subgroup members relative to their own group, rather than to the national population (Hnatkovska et al., 2013; Emran and Shilpi, 2015; Azam and Bhatt, 2015).<sup>20</sup> Studies of affirmative action in India have found impacts on educational attainment of SC/STs (Frisancho Robles and Krishna, 2016; Bagde et al., 2016; Cassan, 2019; Khanna, 2020), but have not examined intergenerational mobility; our findings suggest that the children of less educated SCs benefit from affirmative action, in contrast to the critique arguing that benefits from affirmative action in India accrue only to the already prosperous members of targeted groups. We have posted code online to calculate all of our mobility measures, along with bottom half mobility estimates with high geographic granularity.<sup>21</sup>

More broadly, our paper relates to broader work on religion and economic development (McCleary and Barro, 2006; Becker and Woessmann, 2009). Alesina et al. (2020) and Platas (2018) find low mobility and educational outcomes for Muslims in Sub-Saharan Africa, where Islam plays a very different cultural, political and social role from India.<sup>22</sup>

Our paper proceeds as follows. Section 2 provides background on India's social groups. Section 3 describes our methodological innovation in the context of prior measures of intergenerational educational mobility. Section 4 describes the data sources. Section 5 presents results on national and cross-group mobility trends, and the geographic distribution of intergenerational mobility. Section 6 summarizes some recent research on Muslims in India, none of which addresses intergenerational mobility.

<sup>20</sup>Note that there is a parallel literature examining the persistence of income within an individual lifetime in India; this is sometimes described as *income* or *economic mobility*. That literature is focused largely on measurement error in income over the course of an individual's lifetime and is thus not directly related to our work (Azam, 2016; Li et al., 2019).

<sup>21</sup><https://github.com/devdatalab/paper-anr-mobility-india/>

<sup>22</sup>See also Kurian (2018) for a summary of the literature on Islam and economic performance.

presents our analysis of mechanisms. Section 7 concludes.

## **2 Context on Intergenerational Mobility in India**

India's rapid economic transformation and caste system make it a particularly important setting for understanding intergenerational mobility for at least two reasons. First, Indian society has undergone a large transformation over the last forty years. Economic liberalization, starting in the 1980s, dismantled many parts of India's post-Independence socialist experiment. Decades of sustained economic growth have resulted in substantial reductions in poverty and the rise of a large middle class. This setting thus permits us to examine intergenerational mobility against the backdrop of rapid economic growth in a large developing nation. Figure 1 presents the considerable increase in average educational attainment across Forward castes, Scheduled Castes, Scheduled Tribes, and Muslims from the 1950–1990 birth cohorts.

Second, India's caste system is characterized by a set of informal rules that inhibit intergenerational mobility by preventing individuals from taking up work outside of their caste's traditional occupation and from marrying outside of their caste. While some have argued that economic growth is reducing the influence of old social and economic divisions on economic opportunities, caste and religion remain important predictors of economic status (Munshi and Rosenzweig, 2006; Ito, 2009; Hnatkovska et al., 2013; Mohammed, 2019). Since independence in 1947, the government has systematically implemented policies intended to reduce the disadvantage of communities that are classified as Scheduled Castes or Scheduled Tribes. These groups are targeted by a range of government programs and benefit from reservations in educational and political institutions.

India's Muslims constitute a similar population share as the Scheduled Castes and Scheduled Tribes (14% for Muslims vs. 17% for SCs and 14% for STs). While Muslim disadvantage has been widely noted, including by the well-known federal Sachar Report (2006), there are few policies in place to protect them and there has not been an effective political mobilization in their interest. On the contrary, a large-scale social movement (the Rashtriya Swayamsevak Sangh, or RSS) and several major political parties have rallied around pro-Hindu platforms and policies which arguably discriminate against Muslim religious, economic, and cultural practices. Violent anti-Muslim riots have been closely

tied to political parties and political movements (Wilkinson, 2006; Berenschot, 2012; Blakeslee, 2018).

Understanding how mobility has changed for these population groups is important even if all social groups are becoming better off. As noted by Chetty et al. (2020), intergenerational mobility governs the steady-state distribution of outcomes across social groups. Characterizing intergenerational mobility for India’s disadvantaged minority groups therefore has important consequences for equity.

### 3 Methods: Measuring Mobility in Developing Countries

#### 3.1 Background: Measurement of Intergenerational Mobility

This paper focuses on measuring intergenerational mobility, which we define as the persistence of socioeconomic rank across generations. This approach was popularized by Solon (1999) and extended by Chetty et al. (2014b) and Chetty et al. (2020), all of whom emphasize isolating the persistence of socioeconomic rank from changes in economic growth and inequality.<sup>23</sup>

**Desirable properties of mobility estimators.** Our aim is to develop a mobility estimator with three desirable properties relative to the prior literature.<sup>24</sup> First, the estimator should distinguish mobility from aggregate economic growth or inequality. Several statistics that have been used to measure intergenerational mobility describe combinations of these processes. The canonical mobility estimator, the coefficient from a regression of child income on parent income, depends on both mobility and changes in inequality (Chetty et al., 2014b). The probability that a child obtains a higher socioeconomic outcome than their parent (Chetty et al., 2017; Alesina et al., 2021) simultaneously measures intergenerational mobility and economic growth. Inequality and growth in India have been widely studied; our focus is on the persistence of socioeconomic rank as distinct from these other phenomena. Second, we seek a measure that is valid for subgroup analysis; notably, the rank-rank regression coefficient does not pass this test, as noted below. Third, we seek a measure that is comparable across different times and contexts.

**Prior measures of intergenerational mobility.** The first generation of intergenerational mobility studies summarized the joint parent-child outcome distribution with a single linear parameter,

---

<sup>23</sup>For review papers on intergenerational mobility, see Corak (2013), Black and Devereux (2011), and Roemer (2016).

<sup>24</sup>This is not meant to be an exhaustive list; rather, we merely highlight two properties that many mobility measures do not have, and which our new measure does.

such as the correlation coefficient between children’s earnings and parents’ earnings (Solon, 1999; Black and Devereux, 2011). Such gradient measures are easy to calculate and interpret and they form the basis of studies in dozens of countries (Hertz et al., 2008). These measures have two main limitations: (i) they do not distinguish between changes in opportunity at the top and the bottom of the distribution;<sup>25</sup> and (ii) they are not well-suited for between-group comparisons. The parent-child outcome gradient in a population subgroup effectively compares children’s outcomes against more advantaged members of their own group. A subgroup can therefore have a lower gradient (suggesting higher mobility) and yet worse outcomes at every point in the parent distribution.<sup>26</sup>

More recent studies have analyzed the entire joint parent-child income distribution, and in particular, the conditional expectation function of child rank given parent rank (Boserup et al., 2014; Chetty et al., 2014a; Bratberg et al., 2015; Hilger, 2016). The most widely used measure in the new generation of studies is *absolute upward mobility*, defined as the expectation of a child’s income rank, conditional on having a parent at the 25th income percentile, or  $p_{25} = E(y|x=25)$ , where  $y$  is the child rank and  $x$  is the parent rank. Proposed by Chetty et al. (2014a), this measure describes the expected rank of a child born to the median parent in the bottom half of the parent rank distribution.<sup>27</sup>

The measure  $p_{25}$  has the desirable properties that we highlight above: it isolates the persistence of social rank across generations, it can be used to study population subgroups, and it is comparable across contexts. It has the disadvantage of being difficult to use when education is the best available measure of socioeconomic status, as in developing or historical contexts, as we explain in the following section. Our proposed measure, *bottom half mobility*, is similar in interpretation to  $p_{25}$ , but works well in these contexts.

---

<sup>25</sup>This weakness is consequential only if the conditional expectation of the child outcome given the parent outcome is non-linear. But many studies find non-linear CEFs (Bratsberg et al., 2007; Boserup et al., 2014; Bratberg et al., 2015; Connolly et al., 2019); the linearity of the income rank CEF described by Chetty et al. (2014a) is an exception. The Indian education rank CEF that we study below is clearly non-linear.

<sup>26</sup>For a striking example, see Chetty et al. (2020), who show that Black sons suffer the same large rank disadvantage at every point in the parent rank distribution, giving them a nearly identical rank-rank gradient to White sons.

<sup>27</sup>Chetty et al. (2014a) use the term “absolute mobility” because this measure does not depend on the value of the CEF at any other point in the parent rank distribution, distinguishing it from the rank-rank correlation which they describe as a *relative* mobility measure. Other authors use the term “absolute mobility” to describe the set of mobility measures which use child levels as outcomes rather than child ranks. To avoid confusion, we use the term “absolute mobility” only when making reference to the specific measure used by Chetty et al. (2014a).

### 3.2 Educational Mobility and Income Mobility

In the study of upward mobility in developing countries, education is often the preferred proxy of social status, for three reasons. First, matched parent-child education data are more widely available than matched income data. Because education is fixed early in life and parents' education is generally known by the next generation, it is possible to record a parent's education even when the parent is no longer alive.

Second, income is subject to significant measurement error in developing countries. Transitory incomes are noisy estimates of lifetime income, especially in agriculture, and depend on an individual's place in the life cycle. Subsistence consumption is difficult to measure, and many individuals report zero income. Measurement error in income is consequential because it biases mobility estimates upward (Zimmerman, 1992). In contrast, education levels are measured more precisely than income, and rarely change in adulthood, mitigating life-cycle bias.

Third, individual permanent income is difficult to assign in households with joint production processes, which are common among the rural poor. When these households span multiple generations, it can be impossible to distinguish between parent income and child income. Education, in contrast, is directly ascribed to an individual.

Studies of intergenerational mobility in developing countries, as well as in historical periods in developed countries, have thus often focused on the intergenerational persistence of social status as measured by educational attainment (Solon, 1999; Güell et al., 2013; Wantchekon et al., 2015; Card et al., 2018; Derenoncourt, 2019; Alesina et al., 2021). But studies of educational mobility have thus far not developed measures that isolate mobility from growth and are valid across subgroups.

We cannot directly estimate  $p_{25}$  with education data, because coarse measurement of educational completion makes it impossible to identify a parent at a precise education percentile. We highlight this challenge in Figure 2A, which shows the average child education rank in each parent education rank bin, for two Indian birth cohorts: 1960–69 (circles) and 1985–89 (x's).<sup>28</sup> The solid and dashed vertical lines respectively show the boundaries for the bottom-coded education bin in the two cohorts.

---

<sup>28</sup>Data are from the Indian Human Development Survey (2012), described in Section 4.

In the 1960–69 birth cohort, a full 57% of fathers report a bottom-coded education level; in the 1985–89 cohort, this figure is 36%. How does one identify the expected child rank given a parent at the 25th percentile in these birth cohorts?<sup>29</sup>

Figure 2B shows two conditional expectation functions that are both perfectly consistent with the 1960–69 moments. The data available cannot distinguish between these two functions, but they have very different implications for upward mobility: the flatter of the two functions implies a much higher expected rank for a child growing up at the bottom of the distribution. This example demonstrates that the CEF of child rank given parent rank can be at best partially identified from education data; Section 3.3 below shows that we can nevertheless recover useful estimates of upward mobility from the data.

Coarse measures of parental status like these are extremely common in the mobility literature. In older generations in developing countries, bottom-coding rates in excess of 50% are widespread (Narayan and Van der Weide, 2018); researchers encounter a similar problem when studying older generations in richer countries. Internationally comparable censuses often report education in as few as four or five categories. Income mobility is also often based on censored estimates; in the well-known British Cohort Study, one income bin contains more than 30% of the data. Table A1 reports the number of parent education bins used in a set of recent studies of intergenerational mobility from several rich and poor countries. Many studies observe education in fewer than ten bins; the population share in the bottom bin is often above 20%, and in many developing countries it is above 50%.

Comparing transition matrices over time poses a similar problem to that of estimating  $p_{25}$ : when education bin boundaries change over time (due to changes in the education distribution), transition matrix cells cannot be directly compared in different years. Using a quantile transition matrix would resolve the problem of comparability, but calculating the quantile transition matrix from coarsely binned rank data poses exactly the same challenge as calculating  $p_{25}$ .

---

<sup>29</sup>The graph shows why we cannot assume that the expected child rank given a parent at the 25th percentile is equivalent to the expected child rank given a parent in the bottom education bin. The size of the bottom bin is a function of the granularity of the education coding and the national distribution of education; when this bin is smaller, children in the bin are more negatively selected and will have mechanically lower education. In the case of Figure 2A, the CEFs of the two birth cohorts look highly similar in rank space, but the expected child rank in the bottom bin is mechanically lower in the 1985–89 birth cohort because of the smaller bin size.

### 3.3 Estimating Bounds on a CEF with Censored Education Data

We aim to construct a measure of intergenerational mobility which satisfies the desirable properties above and can be tightly estimated with education data. To address the non-observability of granular parent education ranks, we use a partial identification framework suggested in the example above and derived in Novosad et al. (2020), which we describe here.<sup>30</sup>

Formally, we begin with the goal of measuring  $p_i = E(y|x=i)$ , where  $y$  is a child outcome and  $x$  is a parent education rank. We observe only that  $x$  lies within some interval  $[x_k, x_{k+1}]$ , where  $k$  indexes rank bins. The mean value of  $E(y|x)$  in bin  $k$  is observed.<sup>31</sup>

We require only two substantive assumptions to derive bounds on  $E(y|x)$ . First, we assume that there is a latent continuous parent education rank; this implies a meaningful but unobserved ranking of parent educations *within* each observed education category. Formally:

$$E(y|x=i) \text{ has support for all values of } i \in [0, 100]. \quad (\text{Assumption 1})$$

This assumption arises directly from a standard human capital model where observed differences in education levels reflect individual differences in costs and benefits of seeking education (Card, 1999). The latent education rank  $x$  reflects the amount that the marginal benefit or cost of obtaining the next level of education (e.g., “Middle School”) would need to change in order for a given individual to progress to the next level. Individuals who are at the margin of obtaining the next level of education (*i.e.* they would need only a small increase in marginal benefit in order to do so) have the highest educational ranks within their rank bin. Individuals who would not advance further even if the net benefit increased substantially have the lowest ranks in the bin. Alternatively,  $x$  represents what an individual’s education rank would be if education choices were fully continuous and observed as such. The latent

---

<sup>30</sup>Novosad et al. (2020) take the interval data framework of Manski and Tamer (2002) and extend it to identify  $E(y|x=i)$ , where  $y$  is adult mortality and  $x$  is adult education rank. The methodological contribution of the present paper is to frame the educational mobility question as an interval data problem, and to use these partial identification tools to provide a measure with desirable properties.

<sup>31</sup>For parsimony, our framework considers a setting in which only the parent rank is interval-censored — *i.e.*, we take the child rank variable as *not* censored. We address extensions with censored child ranks and potential bias from our approach in Section 5.4.2.

rank thus reflects the underlying factors that shift individuals' demand for education, which can be expected to be correlated with socioeconomic status (Card, 1999).<sup>32,33</sup> Assumption 1 lets us treat the estimation of  $E(y|x=i)$  as an interval censoring problem (Manski and Tamer, 2002; Novosad et al., 2020).

We also assume that the expectation of a child's education level is weakly increasing in the latent parent education rank: having a more advantaged parent cannot make a child worse off. Formally:

$$E(y|x) \text{ must be weakly increasing in } x. \quad (\text{Assumption 2})$$

Empirically, average socioeconomic outcomes of children are strongly monotonic in parent socioeconomic outcomes across many socioeconomic measures and countries (Dardanoni et al., 2012).<sup>34</sup> Average child education is also monotonic in parent education in every birth cohort and almost every subgroup-cohort that we study in India (see Appendix Table A2).<sup>35</sup> Note that the conventional linear estimation of educational mobility also implicitly imposes monotonicity.

Given Assumptions 1 and 2, we can obtain sharp bounds on the child CEF. The analytical formulation of the bounds is derived in Novosad et al. (2020) (for the context of estimating individual mortality by education rank) and presented in Appendix B for the context of upward mobility. Intuitively, these bounds leverage best- and worst-case assumptions about the CEFs that are still consistent with the empirical averages within each bin, as well as monotonicity, given the uniformity of the rank distribution.

Figure 3A shows the bounds on the Indian CEF for the 1960–69 and the 1985–89 birth cohorts.

---

<sup>32</sup>Note that like other papers on intergenerational educational mobility, we use education strictly as a *proxy* for socioeconomic status. Our interpretation of the latent rank holds even if individuals at different latent ranks within the same bin have obtained *exactly* the same number of years of education. All things equal, the individual with a higher latent rank is understood in expectation to have socioeconomic advantages in dimensions other than years of education.

<sup>33</sup>While the transformation of income to income rank is common, the same transformation in education has been rare, perhaps because of the coarse data problem that we identify in this paper. See, for example, Rosenbaum et al. (2000).

<sup>34</sup>In theory, a means-tested welfare program with a sharp discontinuation of benefits at a rank boundary could result in a non-monotonic CEF. We consider it unlikely that any of India's (or any other developing country's) welfare programs are both sufficiently effective and sufficiently precisely targeted to the cumulative set of advantages encapsulated by the parent's latent education rank to break the monotonicity assumption. Affirmative action for SCs is unlikely to cause a non-monotonic CEF since both SCs and non-SCs are widely dispersed throughout the socioeconomic status distribution.

<sup>35</sup>In practice, when education data are highly granular, non-monotonicity may emerge from monotonic distributions due to sampling error. While this occurs for a minority of subgroup-cohorts in our data, it only occurs at the very top of the distribution where bins are very small. These non-monotonicities do not affect our calculations of upward mobility, which only use information from bins adjacent to the bottom half of the parent distribution. To use our analytical method in cases where non-monotonicity is likely to result from sampling error, we advise pooling small bins into larger bins until monotonicity is restored.

While the bounds are tight in parts of the CEF where the education bins are small, they are very wide in the bottom half of the distribution where the data is heavily interval-censored. Absolute upward mobility ( $p_{25}$ ) can be read directly from this graph by examining the bounds of the CEF at  $x=25$ . In this case, the bounds are far too wide to be meaningful for either the 1960–69 or the 1985–89 cohort.<sup>36</sup>

We can obtain tighter bounds by imposing a constraint on the curvature of the CEF; this is equivalent to assuming that a marginal change in parent education rank cannot result in a discontinuous change in the slope or level of the CEF. Figure 3B shows bounds on the CEF under a conservative curvature constraint. The bounds are tighter, but at  $p_{25}$  they remain too wide to be meaningful even with constrained curvature. The limit case of the curvature constraint restricts curvature all the way to zero, resulting in a linear estimation that is equivalent to calculating the rank-rank gradient, which in this case is a poor fit to the data. There is therefore no curvature constraint that both fits the data and generates tight bounds on  $p_{25}$ . Because the bounds on our primary mobility measures are only marginally improved by the curvature constraint, we take a parsimonious approach and present all results below with unconstrained curvature. Although we do not use them in this paper, our framework admits the imposition of arbitrary shape restrictions on the CEF.<sup>37</sup>

**Bottom half mobility.** A similar approach lets us bound functions of the CEF. We focus on a function which we call *bottom half mobility*, which describes the expected outcome of a child born to parents in the bottom half of the parent distribution. Bottom half mobility is thus similar to the widely used measure absolute upward mobility, the latter of which describes the expected outcome of a child born to the median parent in the bottom half of the parent distribution.<sup>38,39</sup>

The measures  $p_{25}$  and  $\mu_0^{50}$  both describe the rate of convergence of groups in the bottom half of the socioeconomic distribution. For each of these measures, a value of 50 implies convergence to

---

<sup>36</sup>The bounds on  $p_{25}$  are 18.5, 54.6 in the 1960–69 cohort and [30.9, 44.1] for the 1985–89 cohort.

<sup>37</sup>The procedure and choice of curvature constraints are described in Appendix B. Appendix B.3 describes the framework for arbitrary shape restrictions.

<sup>38</sup>If the CEF is linear,  $p_{25}$  is equal to  $\mu_0^{50}$ ; if the CEF is concave at the bottom of the parent distribution, then Jensen's Inequality implies that  $\mu_0^{50}$  will be lower than  $p_{25}$ , reflecting the greater persistence of bad outcomes at the bottom of the distribution. For non-linear CEFs,  $\mu_0^{50}$  thus effectively puts more weight on outcomes at the very bottom of the parent distribution than  $p_{25}$ .

<sup>39</sup>Our framework can further be used to bound many other measures, such as  $\mu_0^{20}$  (the expected rank of a child born in the bottom 20%),  $E(y > 80|x < 20)$  (the probability that a child born in the bottom 20% makes it into the top 20%), or the conditional median function of child rank given parent rank.

the mean rank in one generation, and a value of 25 implies no convergence at all. Both measures are useful, just as the mean and median of a given distribution have advantages and disadvantages as summary statistics. If the CEF is non-linear, then  $\mu_0^{50}$  incorporates information about the mobility of individuals below the 25th percentile. The most important advantage of  $\mu_0^{50}$  is that, as we show below, it can be tightly bounded in contexts where  $p_{25}$  cannot.<sup>40</sup>

One limitation of this paper is that, like prior work on educational mobility, we do not incorporate information on educational quality, which is not available in our data. While educational quality data would be desirable, note that child outcomes are likely to remain monotonic in parent education levels *in expectation* even if there is heterogeneity in education quality. Further, because all of our ranks are calculated within cohorts, changes in education quality over time would not affect our conclusions. However, our geographic estimates of mobility could be biased if regional educational quality growth is not correlated with the growth in the average years of education obtained.

### 3.4 Comparing Bounds on Different Functions of the CEF

Panels A through C of Figure 4 respectively show bounds on the rank-rank gradient (denoted  $\beta$ ), absolute upward mobility ( $p_{25}$ ), and bottom half mobility ( $\mu_0^{50}$ ) for the 1960–69 and the 1985–89 birth cohorts in India. As benchmarks, we also show similar measures for USA and Denmark.<sup>41</sup>

The bounds on the conventional measures  $\beta$  and  $p_{25}$  are not informative either in levels or in changes; they are consistent with both major decreases and major increases in mobility from 1960–85. In contrast, our proposed measure  $\mu_0^{50}$  can be bounded tightly in the 1960s and can be nearly point estimated in the 1980s. On the basis of  $\mu_0^{50}$ , we can clearly distinguish upward mobility in India from the U.S. and Denmark (India is about as far below the U.S. as the U.S. is below Denmark), and we can reject substantial changes in upward mobility over the sample period.

Figure 4 shows the key advantage of bottom half mobility: it can be tightly bounded even with

---

<sup>40</sup>Appendix B.2 provides a formal statement of analytical bounds on  $\mu_a^b$  (derived in Novosad et al. (2020) for the case of mortality estimation). Appendix B.3 provides the framework for bound arbitrary functions of the CEF, including the best linear approximator to the CEF (or the rank-rank gradient)—the linear estimator used in the bulk of the prior educational mobility literature.

<sup>41</sup>The rank-rank gradient is benchmarked against educational mobility estimates from Hertz et al. (2008). For  $p_{25}$  and  $\mu_0^{50}$ , we use income mobility estimates from Chetty et al. (2014a).

severely interval-censored rank data. For intuition behind the tight bounds on  $\mu_0^{50}$ , note that  $\mu_a^b$  is point-identified when  $a$  and  $b$  correspond to bin boundaries in the data. If  $X\%$  of parents are in the bottom-coded rank bin, then  $\mu_0^X$  is the expected child rank, conditional on having parents in the bottom-coded bin, and is point-identified. In general,  $\mu_a^b$  is tightly bounded when  $a$  and  $b$  are close to bin boundaries in the data, by virtue of the continuity of the CEF and uniformity of the rank distribution. In contrast, absolute mobility at percentile  $i$  ( $E(y|x=i)$ ) cannot be point-identified for any value of  $i$ . For further clarification, Figure 5 presents a graphical example demonstrating how the relatively tight bounds on  $\mu_0^{50}$  are obtained.

Note that the wide uncertainty around changes in mobility as measured by  $\beta$  and  $p_{25}$  reflect a strength of our approach. When rank data are highly censored, we should indeed have less certainty over the ability of individuals to move up from the bottom of the rank distribution. By delivering precise point estimates regardless of the coarseness of the data, conventional methods use hidden assumptions (such as linearity of the CEF, in the case of the rank-rank gradient) and convey excess precision.

The rank-rank gradient, absolute mobility, and bottom half mobility are all scalar statistics that capture different characteristics of the intergenerational persistence of rank, and they may all be of independent policy interest. However, only bottom half mobility can be measured informatively given the type of education data typically available in developing countries. This makes bottom half mobility the first measure of intergenerational educational mobility that can be compared meaningfully across population subgroups, across countries, and across time.

### 3.5 Comparison with Other Approaches

This section briefly contrasts our approach to measuring intergenerational educational mobility with other recent approaches in the literature.

Card et al. (2018) and Derenoncourt (2019) use education data to compare geographic patterns in upward mobility between the 1920s and the 1980s. Both papers define upward mobility in the 1920s birth cohort as the 9th grade completion rate of children whose parents have 5–8 years of school, described by Card et al. (2018) as “roughly in the middle of the parental education distribution.” They then compare this measure with  $p_{25}$  for a birth cohort in the 1980s. Translating this into

our framework, where  $x$  is a parent rank and  $y$  is a child outcome, these papers are comparing  $E(y \geq 50 | x \in [45, 70])$  for the 1920s birth cohort to  $E(y | x = 25)$  in the present. The 1920s measure therefore is a good statistic for children getting more education than their parents, but is difficult to interpret as a measure of churn across socioeconomic ranks. The advantage of the 1920s measure is that it is easily calculated, since it corresponds to bin boundaries in the education data. Our approach broadens the scope of possible measures: we can consistently measure  $\mu_0^{50} = E(y | x \in [0, 50])$  in both periods (or  $\mu_a^b$  for any  $a, b$ ), regardless of the bin boundaries available in the data.

Alesina et al. (2021), who study intergenerational mobility across Africa, face the same problem of variable rank bin boundaries across countries and time for both parent and child education. They focus on education *levels*, defining upward mobility as the probability that a child born to a parent who has not completed primary school manages to do so. This is a combined measure of aggregate increases in education (*i.e.* growth) and the persistence of socioeconomic rank. It also conditions on different groups in different times and countries; in rank terms, this measure approximately describes  $E(y > 52 | x \in [0, 76])$  in Mozambique (where 76% of parents and 48% of children have not completed primary school) and  $E(y > 18 | x \in [0, 42])$  in South Africa. As above, our approach makes it possible to distinguish intergenerational mobility (as defined in Solon (1999) and Chetty et al. (2014b)) from overall growth, and to condition on constant ranks across countries and time.

Finally, when constructing transition matrices from data that are not subdivided by exact quantiles, researchers often take the ad hoc approach of randomly reassigning individuals across bins to create the desired quantile bins. This approach is taken by the World Bank's recent flagship report on intergenerational mobility (Narayan and Van der Weide, 2018). While this approach may seem innocuous, it in fact implicitly assumes that the CEF is a step function with zero slope between bin boundaries. This can result in biased estimates that are misleadingly precise. For example, Narayan and Van der Weide (2018) find virtually identical outcomes for children growing up in the bottom three quartiles of the parent distribution in Ethiopia — a result that is a mechanical artifact of over 80% of parents reporting the bottom-coded education level.

## 4 Data

We draw on two datasets that report matched parent-child educational attainment. Our main results rely on the India Human Development Survey (IHDS), a nationally representative survey of 41,554 households, with rounds in 2004–05 and 2011–12. The IHDS identifies religion and Scheduled Tribe or Scheduled Caste status. We classify SC/ST Muslims, who make up less than 2% of SC/STs, as Muslims.<sup>42</sup> About half of Muslims are Other Backward Castes (OBCs); we classify these as Muslims. We do not consider OBCs as a separate category in this paper because OBC status is inconsistently reported across surveys, due to both misreporting and changes in the OBC schedules. Analysis of mobility of OBCs will therefore require detailed analysis of subcaste-level descriptors and classifications which are beyond the scope of the current work. We pool Christians, Sikhs, Jains and Buddhists, who collectively make up less than 5% of the population, with higher-caste Hindus (*i.e.* forward castes and OBCs); we describe this group as “Forward/Other.” We find broadly similar results if we exclude these other religions from the sample.

Crucially, the IHDS solicits information on the education of parents of the majority of respondents, even if those parents have died or are not resident in the household. Estimates using the IHDS data are thus not subject to concerns about coresidence bias. Parent-child coresidence rates decline rapidly with child age, as shown in Appendix Figure A1. Appendix Figure A2 shows the bias that arises from estimating mobility from coresident parent-child pairs—the difference between bottom half mobility estimated using only coresident parent-child pairs and the same measure estimated using the full information on the parents of non-coresident children. The bias rises substantially for sons over age 25 and daughters over the age of 18. When looking at older cohorts, it is thus essential to include children who no longer live with their parents as we do in the IHDS sample; earlier Indian mobility estimates based on coresident children as old as 40 should be treated with caution.

We estimate mobility in the past by studying children from older birth cohorts, also in the 2011–2012 IHDS. We address potential survivorship bias by showing that our results are consistent

---

<sup>42</sup>Classifying this group as SC/ST or excluding this group does not affect any of the results because overall they represent less than 0.4% of the population.

with an analysis of the *same* birth cohorts using the 2004–05 IHDS, described in Section 5.1. We pool the data into 10-year birth cohorts for 1950–69, and 5-year birth cohorts for 1970–1989 where we have more power. The data do not contain links for mothers or daughters for the 1950–59 birth cohort.

The sample size of the IHDS is too small to study geographic variation in any detail. We therefore draw on the 2011–12 Socioeconomic and Caste Census (SECC), an administrative socioeconomic database covering all individuals in the country that was collected to determine eligibility for various government programs. The household roster describes age, gender, education, and Scheduled Caste or Scheduled Tribe status. Assets and income are reported at the household rather than the individual level and thus cannot be used to estimate mobility. Religion and subcaste were recorded but not included with the publicly posted data, and are therefore not available in the geographic analysis.<sup>43</sup>

The SECC allows us to construct parent-child links only when parents and children reside in the same household. To minimize bias from exclusion of non-coresident children, we limit our analysis of the SECC to sons aged 20 to 23, a set of children for whom schooling is largely complete, but parent coresidence rates are still high. We do not study daughters using the SECC, because many daughters have already left home at ages when other daughters are still completing their education.<sup>44</sup> The SECC sample thus consists of 31 million young men and their fathers. For the coresident father-son pairs that are observed in both datasets, IHDS and SECC produce similar point estimates for upward mobility.

Given the strengths and limitations of each dataset, we use the SECC to study cross-sectional geographic variation in mobility, and we use the IHDS to study mobility differences across groups and across time.

Education is reported in seven categories in the SECC.<sup>45</sup> IHDS records completed years of education. To use a consistent definition of education and mobility in the two parts of the paper, we recode the SECC into years of education, based on prevailing schooling boundaries, and we

---

<sup>43</sup>Additional details of the SECC and the scraping process are described in Asher and Novosad (2020) and in Appendix D.

<sup>44</sup>Appendix Figure A2 shows that upward mobility estimates of daughters living with their fathers are already biased upward by 5 rank points—the mobility difference between the U.S. and Finland—by age 20.

<sup>45</sup>The categories are (i) illiterate with less than primary; (ii) literate with less than primary (iii) primary; (iv) middle; (v) secondary (vi) higher secondary; and (vii) post-secondary. These are standard categories used in many of India's surveys, including the National Sample Surveys.

downcode the IHDS so that it reflects the highest level of schooling completed, *i.e.*, if someone reports thirteen years of schooling in the IHDS, we recode this as twelve years, which is the level of senior secondary completion.<sup>46</sup> The loss in precision by downcoding the IHDS is minimal, because most students exit school at the end of a completed schooling level.<sup>47</sup>

The oldest cohort of children that we follow was born in the 1950s and would have finished high school before the beginning of the liberalization era in the 1980s. The cohorts born in the 1980s would have completed much of their schooling during the liberalization era. The youngest cohort in this study was born in 1989; cohorts born in the 1990s may not have completed their education at the time that they were surveyed and are therefore excluded.

Details on construction of parent-child links, as well as additional data sources and information on variables used in the geographic analysis, can be found in Appendix D.

## 5 Results: Intergenerational Mobility in India

This section outlines our three main empirical results. First, we find that intergenerational educational mobility has remained constant, once we account for the interval-censoring of the education data. Second, we find an increase in IM for SCs and STs, and a large reduction for Muslims. We find no evidence of reductions in IM for Forward/Others. Third, our spatial estimates suggest that IM is higher in cities, and there is substantial variation across space at both the micro and macro scale.

### 5.1 National Estimates

Figure 6 shows our main measure of upward mobility (bottom half mobility, or  $\mu_0^{50} = E(y|x \in [0,50])$ , where  $y$  is the child education rank). Panel A shows the father-son relationship. Upward mobility has been largely static over time, moving from [36.6,39.0] for the 1960–69 birth cohort to [37.5,37.9] for the 1980–85 birth cohort.<sup>48</sup> The bounds on the 1950–59 estimates are wider, leaving open the

---

<sup>46</sup>We code the SECC category “literate without primary” as two years of education, as this is the number of years that corresponds most closely to this category in the IHDS data, where we observe both literacy and years of education. Results are not substantively affected by this choice.

<sup>47</sup>The bottom-coded education group is virtually unaffected by this downcoding, since fewer than 0.5% of individuals in any cohort are downcoded to the bottom bin; this downcoding thus has almost no effect on our mobility results, as we show in Appendix Table A4.

<sup>48</sup>The measures are very tightly bounded for the more recent birth cohorts, because there is a rank boundary close to 50 in the parent distribution. When the distance between upper and lower bounds is less than 0.3, we

possibility of some gains from the 1950s to the 1960s birth cohorts.<sup>49</sup> The comparable measure for the 1980s birth cohort in the United States is 41.7, which is low by OECD standards.<sup>50</sup> Note that a naive application of the rank-rank gradient to these data would instead suggest substantial and precise improvements in mobility over the sample period; Figure 4 shows that the data do not support a precise estimate of this statistic once the interval censoring is taken into account.

Panel B of the same figure describes mobility from fathers to daughters. This sample does not go back to the 1950s since there are no respondents from that birth cohort in the women's surveys. We cannot reject a broadly similar pattern to the father-son results, though the bounds in the 1960s are wider than for sons, leaving open the possibility of mobility losses over this period. In the youngest birth cohort, father-daughter mobility is 35.6, about two rank points lower than father-son mobility. Daughters are thus less likely to escape low relative socioeconomic status than sons.

Obtaining informative mobility estimates for the mother-child relationships is more difficult, because the distribution of women's education is much more right-skewed (and thus censored in rank terms). Among mothers of the 1960s birth cohort, 82% had less than two years of education. For the 1985–89 birth cohort, this number was 65%. Under such severe censoring, we cannot estimate  $\mu_0^{50}$  with any precision. Even in the most recent 1985–89 birth cohort, we estimate bottom half mobility to be [37.5,41.4] for sons and [33.8,39.1] for daughters.<sup>51</sup> We thus focus on estimates of mobility based on fathers.

We can also calculate  $\mu_0^{50}$  with child education *levels* as the  $y$  variable. For example,  $E(\text{child years} \geq 12 | x \in (0,50))$  describes the likelihood that a child attains high school or greater, conditional on having a parent in the bottom half. Panels C and D of Figure 6 show this measure for father-son and father-daughter links respectively. The graphs also show  $E(\text{child years} \geq 12 | x \in (50,100))$ , which is the likelihood of child high school attainment conditional on having a parent in the top half of the report the midpoint as a point estimate.

---

<sup>49</sup> Appendix Figure A3 shows that these results are unlikely to be affected by survivorship bias. We estimate upward mobility for the same birth cohorts using the IHDS 2004–05; if mobility estimates for older cohorts were affected by differential mortality of high mobility groups, we would find different estimates from the earlier data, but the bounds are highly similar and show the same lack of change over time.

<sup>50</sup>Source: our calculations using  $\mu_0^{50}$ , based on data from Chetty et al. (2020). There is not yet a wide set of internationally comparable estimates of rank-based educational mobility, in part because of the methodological challenges described and addressed in this paper.

<sup>51</sup>Appendix Figure A4 shows the admittedly uninformative graph of this measure over time.

education distribution. These graphs show the secular increase in high school attainment over time for children from privileged and underprivileged backgrounds. Daughters from bottom half families have experienced the least gains, while daughters born in the top half of the distribution have almost closed the gap with well-off sons. For both sons and daughters, gains in high school attainment have accrued almost entirely to children from the top half of the distribution, a reflection of the stagnant overall upward rank mobility seen in panels A and B. However, these estimates confound intergenerational mobility with aggregate increase in education, which is why we focus on  $\mu_0^{50}$  in rank terms.

To summarize, children born to less privileged families in post-liberalization India have very similar prospects for moving up in the rank distribution as they did in the pre-liberalization era. To be clear, living standards have improved for individuals across the rank distribution; it is the probability of making progress in rank terms which is unchanged. This result thus contradicts the narrative of India becoming a land of greater churn in terms of relative social status.

## 5.2 Changing Mobility Across Social Groups

We next examine how these levels and trends differ across groups. Figure 7 presents results analogous to those above but subdivided into Muslims, Scheduled Castes, Scheduled Tribes, and all others. Panel A shows bottom half mobility ( $\mu_0^{50}$ ) from the 1950s to the 1980s for father-son pairs, revealing substantial trend differences across groups. As noted by other researchers, upward mobility for Scheduled Tribes, and especially for Scheduled Castes, has improved substantially (Hnatkovska et al., 2012; Emran and Shilpi, 2015). SCs born in the bottom half of the parent distribution in the 1960s could expect to obtain between the 33rd and 35th percentile; the comparable group born in 1985–89 obtains the 38th percentile, closing approximately half of the mobility gap with upper castes. Upward mobility for members of Scheduled Tribes rises from [29,31] to 33 over the same period.

In contrast with SCs and STs, Muslim intergenerational mobility declines substantially, falling from [31,34] in the 1960s to 29 in the 1985–89 birth cohort. These changes not only constitute a major decline in mobility, but make Muslim men the least upwardly mobile group in modern India. Mobility for Muslim sons is lower even than for ST sons, who are often thought of as having benefited very little from Indian growth. The fact that a Muslim boy born to a family in the bottom half of the distribution

can expect to obtain the 29th percentile implies that Muslims born into low status are very likely to remain low status. Finally, the “Forward/Others” group, which predominantly consists of higher-caste Hindus, shows little change, with mobility shifting from [42,44] to 42. The flat trend in upward mobility for sons can therefore be decomposed into gains for SCs and STs and losses for Muslims.

Panel B shows downward mobility ( $\mu_{50}^{100}$ ) for father-son links over the same period; this measure reflects the persistence of high status among each group. We see a small amount of convergence between the three marginalized groups and the Forward/Others group, chiefly from the 1970s to the 1980s birth cohort. But there is no sign of the dramatic divergence between SCs and Muslims that was found for upward mobility.<sup>52</sup>

To interpret these results, we note that one rank point is associated with about 0.15 years of education on average in 1985. The 9 rank point mobility difference between Muslims and SCs thus corresponds to 1.3 years of education in 1985, when children of bottom half parents on average attained 6.5 years of education.

Panels C and D of Figure 7 show the same results for father-daughter pairs. Among daughters, with the exception of recent minor gains for SCs from top half families, none of the marginalized groups have made substantial gains relative to Forwards/Others. There is also little sign of the divergence between SCs and Muslims that was observed among sons. Table 2 summarizes the changes over time for the full sample and all the population subgroups, along with bootstrap confidence sets to account for sampling variation, calculated following Chernozhukov et al. (2007).<sup>53</sup> Table 3 shows confidence sets for mobility differences between groups for the youngest (1985–89) birth cohort.<sup>54</sup>

Note that estimating  $\mu_0^{50}$  for population subgroups implicitly assumes that the subgroup parent

---

<sup>52</sup> Appendix Figure A5 shows analogous results to Figure 7, but with education levels (at least primary, and at least high school) as outcomes, rather than education ranks. The advantage of these graphs is that they present outcomes for children that are not subject to interval censoring; the parent variable remains interval censored. The results are consistent with the rank-based estimates, confirming that the separation between Scheduled Caste and Muslim sons is not driven by unobserved changes in interval-censored ranks for *children* from these groups. The upward trend in the levels graphs reflects the overall rise in educational attainment seen in Figures 6C and 6D.

<sup>53</sup> Appendix Table A4 shows similar estimates with ranks calculated from the granular years of education in the IHDS; the mobility estimates and subgroup differences are nearly identical.

<sup>54</sup>The confidence sets in Tables A4 and 3 are wider than mobility confidence intervals from prior studies because they reflect both statistical variation and uncertainty due to coarse measurement of education, the latter of which has not been addressed by prior studies. Nevertheless, the cross-group differences in the youngest birth cohort are all highly significant, as is the trend difference between SCs and Muslims.

populations are uniformly distributed across the bottom half of the rank distribution. This assumption is unlikely to hold, but mobility *changes* will be biased only if the parent density functions change substantially in the same period. We verify robustness to alternate distributional assumptions in Subsection 5.4.1.

To summarize, we observe a sharp divergence between upward mobility for sons from Scheduled Caste and Muslim groups. Muslim sons from poor families have declining mobility and very little opportunity to improve their relative social status. This low mobility may also adversely affect female Muslims, since marriage is nearly universal in India and almost entirely within social group, and female labor force participation is very low. Understanding how marriage ties interact with the upward mobility of sons and daughters is of interest but beyond the scope of the present paper.

### 5.3 Upward Mobility Across Geographic Areas

We next describe the geographic variation in upward mobility. The limited sample size of the IHDS only allows us to examine geographic patterns of subgroup mobility at a low resolution. Appendix Figure A6 shows bottom half mobility, disaggregated by child gender, social group, and rural-urban status. Mobility is systematically higher in urban areas, but subgroup disadvantage varies substantially by geography. The urban-rural gap is much higher for daughters than for sons, such that urban daughters in fact have about five rank points higher upward mobility than urban sons. Muslims and SCs on average have higher mobility in cities (not shown), but their relative position in cities with respect to Forwards/Others is worse.

The remainder of this section uses the SECC, which has a large sample and precise location identifiers that make it possible to generate mobility estimates in very small geographic areas. As note above, due to data limitations, the SECC sample is limited to 20–23-year-old coresident sons, and does not include data on religion. We do not explore time series patterns across geography, both because we do not observe where children grew up and because we do not observe the parents of children from earlier birth cohorts.

Figure 8A presents a heat map of upward mobility across 4000 subdistricts and 2000 major towns

across all of India.<sup>55</sup> The geographic variation is substantial. The interquartile range of upward mobility is [31.8,42.8] across cities/towns and [30.3,45.1] across rural subdistricts; for comparison, the IQR of  $p_{25}$  across commuting zones in the U.S. is slightly smaller: [39.9,47.1] (Chetty et al., 2014a). Upward mobility is consistently highest in southern India—Tamil Nadu and Kerala—and is also noticeably high in the mountainous states of the North. Parts of the Hindi-speaking belt—especially the state of Bihar—and the Northeast have some of the lowest mobility in India. Gujarat is noteworthy as a state with high economic growth but relatively low mobility.

In broad regions of high mobility, there are low mobility islands, such as the rugged region between Andhra Pradesh and Karnataka. Cities and towns for the most part stand out as islands of higher mobility. However, there is not a single subdistrict or town in Bihar with higher average mobility than the southern states.

Figure 8B shows a ward-level mobility map of Delhi, which shows substantial variation in mobility based on neighborhood of residence even within a single city. The 90th percentile ward has 52% higher upward mobility than the 10th percentile ward (47.9 vs. 31.4). Children in the dense and industrial areas of Northeast Delhi have the least opportunity; the average child from a bottom half family in these areas can expect to obtain the 32nd percentile nationally. Children from similarly-ranked families in Southwest Delhi, about 5km away, can expect to obtain the 44th percentile.

To explore some of the potential drivers of geographic variation in upward mobility, Figure 9 presents the association between bottom half mobility and several correlates identified by the earlier literature on India and other countries. Panel A presents coefficients from a set of bivariate regressions of bottom half mobility on characteristics of rural subdistricts across India, with X variables normalized to mean zero and standard deviation one. Panel B presents analogous results for the town sample.<sup>56</sup>

At the rural level, the traditional markers of economic development — consumption, and average levels of education — are the strongest correlates of upward mobility. Local public goods and

---

<sup>55</sup>There are 8000 towns in the 2011 Population Census, on which the SECC is based. While our coverage of rural areas is almost complete, the data posted online described only 2000 of the towns. The town sample is broadly representative of the urban population in demographics and income. The graph shows the midpoint of the bounds; in 99% of cases, the bound width is less than two rank points.

<sup>56</sup>All variables in these figures are defined in Appendix D.

manufacturing employment are also positively correlated with upward mobility. Availability of primary schools is the least important of these, but availability of high schools is highly correlated with mobility. Surprisingly, the share of SCs is positively correlated with upward mobility even though SCs have lower mobility on average. Segregation and land inequality are negatively correlated with upward mobility, a parallel result to that found in the United States (Chetty et al., 2014a).

In urban places, we find that upward mobility is higher in towns that have (i) higher population; (ii) more SCs; (iii) more educated populations; and (iv) more high schools per capita. As in rural areas, SC/ST segregation is negatively associated with mobility.

Place of residence is thus a strong predictor of upward mobility. These local mobility estimates have two limitations. First, they are based on the educational outcomes of children born between 1989 and 1992, the majority of whom finished their education by 2010. They therefore reflect the circumstances that drove education choices in the period 2000–2010, which may be different from the present. Second, the estimates do not account for migration, as we do not observe respondents' location of birth. Low mobility in Northeast Delhi could in part be the effect of immigration from poor parts of rural North India. The more local the mobility estimate, the greater is the potential bias from migration. Ideally, we would have local surveys that record both location of origin and parental education; to our knowledge, there are no such surveys with high geographic precision. The rural (subdistrict-level) estimates are less likely to be biased by migration, because permanent migration in rural areas in India is extremely low.<sup>57</sup>

## 5.4 Robustness of Bottom Half Mobility to Alternate Assumptions

### 5.4.1 Non-Uniform Within-Bin Subgroup Distributions

Our bounds on the full sample CEF  $E(y|x)$  (explained in Section 3.3) use the uniformity of the rank distribution, which is given when working with the 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. In particular, groups

---

<sup>57</sup>Foster and Rosenzweig (2007) find decadal rural-to-urban migration rates for 15–24 year old males of about 3% in 1961–2001; Munshi and Rosenzweig (2016) present confirmatory evidence in IHDS and the DHS.

with lower average levels of education might occupy lower average ranks within every education bin.

To address this point, we make three remarks. First, note that the general framework can be used with *any* distribution of conditioning data; in Appendix B, we provide general bounds on  $\mu_0^{50}$  with an arbitrary parametric distribution. We use uniformity as a natural benchmark.

Second, we demonstrate in Appendix C.1 that any bias from assuming uniformity within subgroups is too small to affect our conclusions. We show that both the divergence between Muslims and SCs and the convergence between SCs and Forwards/Others are robust to defining parent education ranks *within* social groups rather than within the national distribution. By this definition, latent ranks are uniform by construction.<sup>58</sup>

Third, in Appendix C.1, we perform another exercise that imputes continuous latent rank distributions from the binned subgroup education distributions under various parametric assumptions. We can then point-estimate bottom half mobility directly from these imputed distributions. The dispersion of these imputed latent ranks across social groups provides a measure of bias in bottom half mobility as measured in the censored data. Under the worst-case assumptions, at most 10% of the growing gap between Muslims and SCs can be explained by changes in latent parent ranks within education bins.

To understand why the bias is small, note that while Scheduled Castes and Muslims are certainly disadvantaged on average, their education distributions still have significant overlap with higher-caste Hindus. Moreover, young SCs and Muslims have similar parent education distributions from earlier generations, so within-bin variation is unlikely to affect comparisons of these two groups against each other.

#### 5.4.2 Interval Censoring of Child Ranks

In our results thus far, we have assumed that child ranks are directly observed and are equal to the midpoint of each child's rank bin. But child ranks are also interval-censored in our context. To address the concern that this issue materially affects our conclusions, we make three remarks.

First, when we use uncensored measures of child outcomes, such as primary or high school comple-

---

<sup>58</sup>We do not use this measure for our primary results because it no longer conditions on parents with comparable socioeconomic status; a parent in the least educated 50% of forward castes has more education than a parent in the least educated 50% of Scheduled Castes. Nevertheless, the consistency of this finding implies that the trends observed are unlikely to be driven strictly by non-uniform latent rank distributions within education bins.

tion (in Figures 6C, 6D, and Appendix Figure A5), we continue to find substantial divergence of SCs and Muslims from bottom-half families, as well as the relative lack of progress of individuals from the bottom half as a whole. Note that these measures still hold *parent rank* fixed across generations using our partial identification approach, so they are valid for cross-group comparisons. They have the disadvantage of combining mobility with growth, since all groups have had large increases in average educational attainment over time. As a result, we prefer the outcome measures based on child rank because these isolate intergenerational mobility from growth and are internationally comparable. But when considering cross-group comparisons alone, the education-level outcome measures are equally valid, because all groups experience the same change in the relationship between national education ranks and education levels over time. The robustness of the Muslim-SC divergence using level outcomes therefore demonstrates that the divergence using rank outcomes cannot be driven by censoring of child ranks.

Second, note that the censoring problem for child ranks is much smaller than the censoring problem for parent ranks, because children are on average more educated than their parents. This makes their education bins more evenly distributed. In the 1960–69 birth cohort, the bottom child education bin (and the largest) contains 26.5% of the population; in 1985–89, it contains only 9%. The bias from censored child ranks is thus expected to be considerably smaller than that from censored parental ranks, where the largest bins can contain 60% of the data.

Finally, in Appendix C.2 we provide two approaches to estimate the bias in our main estimates from son censoring. First, we estimate the maximal extent of this bias by examining the effects of the best- and worst-case assumptions regarding the latent child rank distribution on our mobility estimates. This puts an upper bound on the potential bias from latent child ranks. Second, we impute these latent ranks using data on children’s wage ranks within education bins; doing so, we find virtually the same mobility estimates as in our results above, suggesting the actual bias is small.<sup>59</sup> These results suggest that using the midpoint of a child’s rank bin is capturing most of the meaningful variation in child ranks.

---

<sup>59</sup>Specifically, we define an individual’s rank *within* an education bin based on their wages, yielding a continuous distribution. Results are similar if we use household income.

## 6 Potential Mechanisms for Subgroup Mobility Differences

In this section, we explore a series of mechanisms that could potentially explain the upward mobility differences and changes across social groups in India. We focus on explaining our most striking finding, namely the growing mobility gap between India's two largest marginalized groups: Scheduled Castes and Muslims. Over the last 30 years, SCs have closed half of the mobility gap with upper castes, while Muslims have fallen further behind. Understanding this divergence is both important in its own right in explaining mobility trends for a third of India's population, and to understand the drivers of mobility among marginalized groups in India and other developing countries. We examine the extent to which these group differences can be explained by: (i) affirmative action for Scheduled Castes; (ii) differential fertility across social groups; (iii) patterns in location of residence; and (iv) different occupational patterns causing differential returns to education. These analyses are suggestive, but they point toward affirmative action as a key mechanism for the SC/Muslim divergence, and largely reject the other three mechanisms.

### 6.1 Affirmative Action for Scheduled Groups

First, we consider the hypothesis that the basket of programs and policies targeted to Scheduled Castes has driven the increase in upward mobility of SCs relative to Muslims since the 1950s. Affirmative action for SCs is primarily comprised of reservations (positions only available to SCs) in higher education admissions, political office (federal, state, and local legislatures), and public sector employment. There are also direct educational benefits such as scholarships and dedicated schools for SCs. To study how affirmative action affects bottom half mobility, we exploit a change in the population groups eligible for Scheduled Caste status that was made in 1977, previously studied by Cassan (2019). Between 1956 and 1977, the list of social groups that was eligible for Scheduled Caste status varied across regions within states; this inconsistency was left over from the reorganization of states along linguistic lines in 1956. In 1977, a federal law harmonized these lists within states, arbitrarily moving many additional groups into the SC designation, making them eligible for SC-targeted benefits. This policy change makes it possible to examine the impact of Scheduled Caste

status, while controlling for a group's ethnicity, historical experience, and narrow geographic region. This policy change is a good natural experiment because a clear federal rule dictated which groups would change status, and it was not possible for groups to lobby to be included on the new list.

We use data from Cassan (2019) to test whether groups newly added to scheduled lists experienced gains in upward mobility. We divide SCs into two groups: (i) those classified as SCs at the time when states were reorganized on linguistic lines in 1956 (whom we call early SCs); and (ii) a group which obtained protected status only in 1977 (whom we call late SCs). Following Cassan (2019), we assume that individuals needed to be 6 or younger in 1977 to benefit in terms of education from the change in status.<sup>60</sup> We therefore treat individuals born later than 1970 as being in the late SC group.

We use the IHDS data, which is not affected by concerns about coresidence bias. We assign individuals to early and late SC groups using the IHDS's *jati*-level group identifiers.<sup>61</sup> Figure 10 demonstrates the upward mobility trajectory of the early and late SC groups, showing bounds on  $\mu_0^{50}$  over time. In the 1950s and 1960s, the early SC group experiences rapid relative increases in mobility and diverges from the late SC group, which has not yet obtained protected status. Beginning with the 1970 birth cohorts when the late SC group obtains protected status, it begins to close the mobility gap. The mobility gap peaks right before the late SC group gains SC status and the gap then steadily closes through the remainder of the sample period.

We can formally estimate the impact of SC status in a regression based on the specification of Cassan (2019):

$$Y_{i,j,r,c} = \beta_0 + \beta_1 \text{LateSC}_{j,r} + \beta_2 \text{post}_c + \beta_3 (\text{post}_c \times \text{LateSC}_{j,r}) + \nu_{j,r} + \eta_{r,c} + \zeta_{j,c} + \epsilon_{i,j,r,c}, \quad (6.1)$$

where  $Y_{i,j,r,c}$  is the education rank of child  $i$  in jati  $j$ , region  $r$ , and birth cohort  $c$ . We include fixed effects for jati  $\times$  region ( $\nu$ ), region  $\times$  cohort ( $\eta$ ), and jati  $\times$  cohort ( $\zeta$ ). These fixed effects exploit the fact that the same jati group could be an early SC or a late SC depending on its region within a state. The coefficient of interest,  $\beta_3$ , therefore compares individuals in late SC groups born after

---

<sup>60</sup>We find similar results if we use a cutoff of 11 or younger, also used in Cassan (2019).

<sup>61</sup>A jati is a caste identifier that is more granular than the broad Scheduled Caste category, which includes many jatis.

1970 to those born earlier in the same narrow social group and the same region, controlling for outcomes of individuals from early SC groups in the same region, and for outcomes of members of the same jati group in other regions. Regressions are clustered at both the jati and the region level.

Using bottom half mobility as a  $Y$  variable in this specification is challenging because, for each group, it can at best be bounded. Instead, we proxy bottom half mobility by restricting the sample to a set of father ranks that is as close as possible to the bottom 50% but can also be identified as consistently as possible across the different decades. To do this, we define a  $Y$  variable equivalent to  $\mu_0^{59}$  in the 1950s birth cohort,  $\mu_0^{57}$  in the 1960s, and  $\mu_0^{58}$  in the 1970s and 1980s.

The result of this regression is shown in Table 4. Column 1 shows the specification above. The cohorts exposed to the basket of affirmative action policies experience an 8 rank-point change in upward mobility. Column 2 shows robustness to a specification where we limit the sample to sons of fathers with less than two years of education in all years instead of the sons of fathers in the approximate bottom 60% as noted above. The point estimate is similar. Column 3 (using the same sample as Column 1) shows that both the 1970s and 1980s birth cohorts of late SCs benefited relative to the earlier cohorts. These results accord with Cassan (2019), who found that gaining SC status led to a 10 percentage point increase in literacy and a 7 percentage point increase in secondary school attainment. However, it was not necessarily the case that these gains would accrue to children of low education parents; indeed, affirmative action's critics in India often suggest that it is captured by a “creamy layer” of the already prosperous among targeted groups, a story that our findings reject.

These results are consistent with affirmative action having induced a large effect on upward mobility for Scheduled Caste groups. The rank-point gain of late SCs is comparable in magnitude to the upward mobility gap in the 1980s between Muslims and SCs, and it emerged over only 20 years of affirmative action. Note that this specification does not directly test for the effect of affirmative action on SCs as a whole vis-à-vis Muslims. However, if these treatment effects for early and late SCs are externally valid for the potential effect of affirmative action on Muslims, then affirmative action could explain the entire contemporary mobility gap between Scheduled Castes and Muslims. On the other hand, because we are limited to relatively coarse variation in demographic groups

over time, these results are suggestive but not decisive.

## 6.2 Group Differences and Fertility

Muslims on average have higher fertility than either Scheduled Castes, Scheduled Tribes, or Forward Castes and other groups. In this section, we consider whether higher fertility could cause lower mobility for Muslims, perhaps through a household expenditure channel where children with many siblings received fewer educational inputs. We explore this question in a regression framework. We estimate the number of siblings of each individual based on their mothers' responses to the IHDS women's survey, which has a question about the number of live births. This variable differs from total fertility by excluding children who have died. We only have information on mothers' fertility for children who live with their mothers; we therefore focus on sons under the age of 30, for whom the coresidence rate is highest.<sup>62</sup> The average number of siblings for Muslims is 4.1, compared with 3.0 for both SCs and STs, and 2.6 for Forwards/Others.

As in Section 6.1, we require a point estimate of upward mobility to use in a regression. We use  $\mu_0^{51}$ , which can be point estimated as the education of children whose fathers completed two or fewer years of education.<sup>63</sup> We regress this mobility measure on a set of group indicators (Muslim, SC, ST), an urban indicator, and a set of state fixed effects, showing the results in Table 5. Column 1 shows the Muslim mobility gap in the full sample and Column 2 shows the same gap in the set of sons whose coresident mother answered the women's survey. The Muslim upward mobility disadvantage is 12 rank points, without adjusting for fertility. Column 3 adds a control for the number of siblings, which brings the Muslim mobility gap down by 25%.<sup>64</sup>

High fertility can thus explain at most 25% of the Muslim mobility disadvantage relative to SCs. This is likely to be an upper bound, because household income is a direct cause of both children's education and parental fertility (Schultz, 2003). Higher fertility can thus explain at most a small share of the present-day mobility disadvantage experienced by Muslims.

---

<sup>62</sup>For daughters, coresidence begins to fall rapidly as soon as schooling is finished, leaving too little sample to estimate mobility among coresiders. Restricting the sample to individuals aged 20–23 as we did for the SECC would cut our sample too much to obtain informative estimates.

<sup>63</sup>We find similar estimates if we use children of fathers with strictly less than 2 years of education.

<sup>64</sup>An additional sibling is associated with 2.4 fewer ranks in the outcome distribution.

### 6.3 Geography and Subgroup Differences

To describe the extent to which Muslim disadvantage in upward mobility can be explained by geography, we examine here whether Muslims live in low mobility places, or whether they have low mobility after conditioning on place. Because the SECC does not record religion, we use the IHDS and explore cross-state and cross-district variation.

SCs, STs and Muslims are unevenly distributed across the country; the 25th-percentile district in SC population is only 8% SC. The equivalent numbers for STs and Muslims (0.4% and 2.7%) reflect the greater geographic concentration of these groups.

To examine the relationship between place and subgroup outcomes, we regenerate father and son education ranks within states and within districts. Mobility estimates generated in this way thus describe the ability of disadvantaged children to increase their relative rank within their own district. If low overall mobility for Muslims is a function of living in districts where everyone has low opportunity, then their within-district mobility gap with Forwards/Others should be substantially smaller than the national mobility gap. We focus on the static father-son mobility gap for the most recent birth cohort, corresponding to individuals born in 1985–89.

The results are shown in Figure 11. The first set of bars shows the mobility gap between Forwards/Others and each of the three marginalized groups in the 1985–89 birth cohort. For simplicity, we show the midpoint of the bounds; the width of the bounds is less than one rank point in all cases. Upward mobility for the Forward/Others reference group is 42.

The following two sets of bars show the same gaps for within-state and within-district ranks. The extent to which group disadvantages in upward mobility can be explained by location differ substantially by group. District of residence explains about 18% of the Muslim upward mobility gap, 44% of the Scheduled Caste upward mobility gap, and 60% of the Scheduled Tribe mobility gap.<sup>65</sup> The result for Scheduled Tribes is consistent with the fact that STs disproportionately live in remote areas of the country with low levels of public goods and educational attainment. Given

---

<sup>65</sup>IHDS districts are not representative so these results should be treated with caution; however, the ordering of the changes is the same when we use only within-state ranks—the middle set of bars in Figure 11.

the uneven distribution of SCs and Muslims throughout India, the unimportance of district as an explanation for their mobility differences is noteworthy. Muslim disadvantage cannot be explained by the broad regions in which Muslims live. However, these results do not rule out the possibility that finer geographic definitions (such as urban neighborhoods) could explain a greater share of the mobility gap; unfortunately, higher resolution analysis is not possible with the data available at this time. To be clear, these results show that location is not a major mediator of Muslim disadvantage; Section 5.3 shows that location is an important predictor of mobility in the aggregate.

#### **6.4 Occupations, Returns to Education, and Subgroup Differences**

We next examine whether occupational choices and returns to education can explain the low and falling upward mobility of Muslims. If the returns to education are lower in Muslims' occupations, then low education outcomes of Muslim children born into poor families would reflect a different expected career path but not necessarily a lack of opportunity.

To evaluate this hypothesis, we first examine the Mincerian returns to education for the different social groups. Figure 12 shows point estimates from Mincerian return regressions for each social group, calculated in three ways: (i) household log income on household head education (IHDS); (ii) individual log wages on individual education (National Sample Survey or NSS); and (iii) household log consumption on household head education (NSS).<sup>66</sup> Across all three measures, there is no evidence that Muslims have lower returns to education than Scheduled Castes or Tribes. The point estimates for Muslims are higher than for SCs in all cases, though both Muslims and SCs have lower returns to education than Forward/Others. Mincerian returns may not reflect the causal effect of education on income and consumption, but there is no evidence here that Muslims are choosing less education because their returns are lower.

Muslims are more likely to work as small-scale entrepreneurs than the other major social groups (Figure 13A). Even if returns to education for Muslims are not lower overall, they could be low for

---

<sup>66</sup>In each case, we regress the outcome variable on individual years of education, age, and age squared. We restrict the data to men aged 18–64 to avoid concerns about selection into labor markets. Results are similar if we restrict to young ages (reflecting education of the youngest birth cohorts), if we include women and men, and if we include additional controls.

Muslims who choose to run small businesses. The data rejects this hypothesis as well. Figure 13B divides the IHDS sample into individuals who own their own business (right panel) and individuals who do not (left panel). We pool SCs and STs for this graph because very few SCs and STs own businesses, leading to small samples. The divergence in upward mobility between SC/STs and Muslims is sustained and of similar magnitude both among business- and non-business owning families.

In conclusion, the evidence does not support the idea that the decline in Muslim upward mobility is an artifact of Muslim occupational choice or returns to education.

## 7 Conclusion

In this paper, we present a set of tools that are well-suited to measuring intergenerational educational mobility in developing countries and other contexts where high-quality income data linked across generations are unavailable. Our partial identification approach takes seriously the loss of information given education that is reported in coarse rank bins. Our proposed measure, bottom half mobility, (i) isolates intergenerational mobility from growth and inequality; (ii) is analogous to the popular *absolute upward mobility* measure; (iii) is informative about intergenerational mobility even when education data are very coarse; and (iv) is easy to calculate. Bottom half mobility is also the first measure of intergenerational educational mobility that is meaningful for cross-group analysis across contexts; the absence of such a measure has made it difficult to study subgroup mobility in developing countries.

In our analysis of India, we find that in spite of enormous economic and political changes, upward mobility has barely changed from the 1950s to the 1980s birth cohorts. This lack of change overall can be decomposed into substantial gains for SC/STs and substantial losses for Muslims. The falling mobility of Muslims has not previously been noted in part because there has been no prior methodology for creating comparable rank bins across cohorts. Our estimate of the causal effect of India's basket of affirmative action policies targeted to SCs suggests the effect of these policies may be large enough to explain the entire SC / Muslim divergence. However, more research is needed to elucidate the factors behind low Muslim mobility in India.

Our work has only begun to describe the wide geographic and cross-group variation in intergenerational mobility in India. As in the U.S., individuals growing up in different parts of India, even

conditional on similar economic conditions in the household, can expect vastly different opportunities and outcomes throughout their lives.

Intergenerational mobility is a crucial policy objective even in the context of high economic growth and improvements in income for all social groups in modern India. The best opportunities remain scarce, and debates regarding who will be eligible for social programs and positions in universities that provide access to those opportunities are extremely heated. Moreover, even if there is growth on average, the extent of intergenerational mobility across groups determines whether Muslims, Scheduled Castes, and Scheduled Tribes will occupy a position of permanent disadvantage in the long run.

Our high geographic resolution mobility datasets are posted online along with the code used to construct these estimates. Future work investigating the causes of geographic and social group differences in upward mobility has the potential to inform policies that expand the equality of opportunity in India.

## References

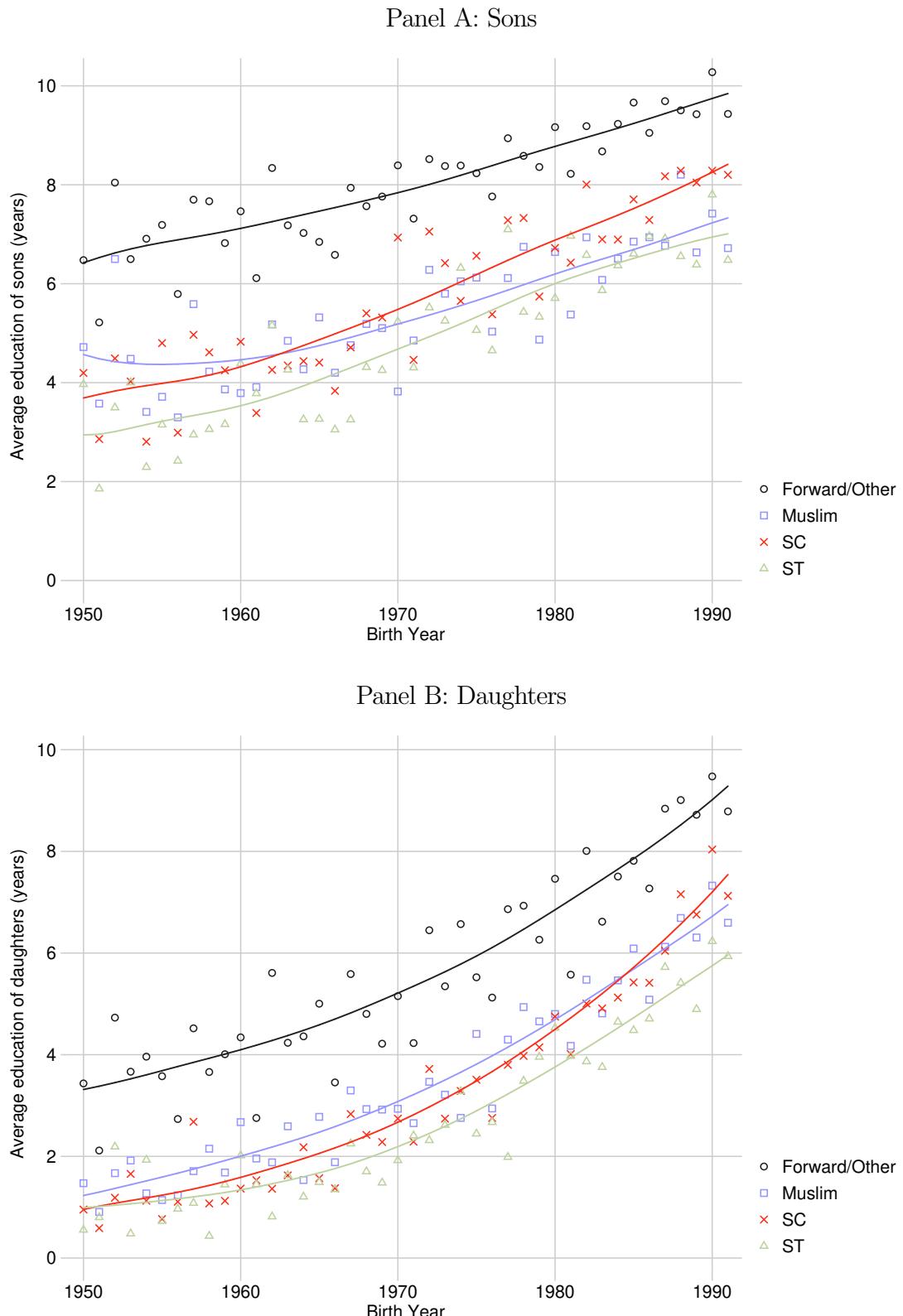
- Alesina, Alberto, Sebastian Hohmann, Elias Papaioannou, and Stelios Michalopoulos,** “Intergenerational Mobility in Africa,” *Econometrica*, 2021, 89 (1), 1–35.
- , — , **Stelios Michalopoulos, and Elias Papaioannou**, “Religion and Educational Mobility in Africa,” 2020. National Bureau of Economic Research Working Paper No. 28270.
- Asher, Sam and Paul Novosad**, “Rural Roads and Local Economic Development,” *American Economic Review*, 2020, 110 (3).
- Azam, Mehtabul**, “Household Income Mobility in India: 1993–2011,” 2016. IZA Discussion Paper No. 10308.
- and **Vipul Bhatt**, “Like Father, Like Son? Intergenerational Educational Mobility in India,” *Demography*, 2015, 52 (6).
- Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor**, “Does affirmative action work? Caste, gender, college quality, and academic success in India,” *American Economic Review*, 2016, 106 (6).
- Basant, Rakesh, Abusaleh Shariff et al.**, *Handbook of Muslims in India: Empirical and policy perspectives*, Oxford University Press, 2010.
- Becker, Gary S. and Nigel Tomes**, “Human capital and the rise and fall of families,” *Journal of Labor Economics*, 1986, 4 (3, Part 2), S1–S39.
- Becker, Sascha O. and Ludger Woessmann**, “Was Weber Wrong? A Human Capital Theory of Protestant Economic History\*,” *The Quarterly Journal of Economics*, 2009, 124 (2).
- Berenschot, Ward**, *Riot politics: Hindu-Muslim violence and the Indian state*, Columbia University Press, 2012.
- Bhalotra, Sonia and Bernarda Zamora**, “Social divisions in education in India,” in Rakesh Basant and Abusaleh Shariff, eds., *Handbook of Muslims in India*, New Delhi: Oxford University Press, 2010.
- Black, S, P Devereux, and Kjell G Salvanes**, “Why the Apple Doesn’t Fall: Understanding Intergenerational Transmission of Human Capital,” *American Economic Review*, 2005, 95 (1).
- Black, Sandra E. and Paul J. Devereux**, “Recent Developments in Intergenerational Mobility,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam: North Holland Press, 2011.
- Blakeslee, David S**, “The Rath Yatra Effect: Hindu Nationalist Propaganda and the Rise of the BJP,” 2018. Working Paper.
- Boserup, Simon Halphen, Wojciech Kopczuk, and Claus Thustrup Kreiner**, “Stability and persistence of intergenerational wealth formation: Evidence from Danish wealth records of three generations,” 2014. Working paper.
- Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel Schnitzlein, and Kjell Vaage**, “A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden and the U.S.,” *The Scandinavian Journal of Economics*, 2015, 119 (1).
- Bratsberg, Bernt, Knut Røed, Oddbjørn Raaum, Robin Naylor, Markus Jäntti, Tor Eriksson, and Eva Österbacka**, “Nonlinearities in intergenerational earnings mobility: Consequences for cross-country comparisons,” *Economic Journal*, 2007, 117 (519).
- Card, David**, “The causal effect of education on earnings,” in “Handbook of Labor Economics,” Vol. 3, Elsevier, 1999, pp. 1801–1863.
- , **Ciprian Domnisoru, and Lowell Taylor**, “The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility,” 2018. NBER Working Paper No.25000.
- Cassan, Guilhem**, “Affirmative Action, Education and Gender: Evidence from India,” *Journal of Development Economics*, 2019, 136.
- Chancel, Lukas and Thomas Piketty**, “Indian income inequality, 1922–2015: From British Raj to

- Billionaire Raj?,” *The Review of Income and Wealth*, 2019, 65 (S1).
- Chernozhukov, Victor, Han Hong, and Elie Tamer**, “Estimation and confidence regions for parameter sets in econometric models,” *Econometrica*, 2007, 75 (5).
- Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang**, “The fading American dream: trends in absolute income mobility since 1940,” *Science*, 2017, 356 (6336).
- , **Nathaniel Hendren, Maggie R Jones, and Sonya R Porter**, “Race and Economic Opportunity in the United States: An Intergenerational Perspective,” *Quarterly Journal of Economics*, 2020, 135 (2).
- , — , — , — , **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.
- , — , — , — , **and Nicholas Turner**, “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility,” *American Economic Review: Papers & Proceedings*, 2014, 104 (5), 141–147.
- Connolly, Marie, Miles Corak, and Catherine Haeck**, “Intergenerational Mobility between and within Canada and the United States,” *Journal of Labor Economics*, 2019, 37 (S2).
- Corak, Miles**, “Income Inequality, Equality of Opportunity, and Intergenerational Mobility,” *Journal of Economic Perspectives*, 2013, 27 (3).
- Dardanoni, Valentino, Mario Fiorini, and Antonio Forcina**, “Stochastic Monotonicity in Intergenerational Mobility Tables,” *Journal of Applied Econometrics*, 2012, 27.
- Derenoncourt, Ellora**, “Can you move to opportunity? Evidence from the Great Migration,” 2019. Working paper.
- Dunn, Christopher E.**, “The intergenerational transmission of lifetime earnings: Evidence from Brazil,” *B.E. Journal of Economic Analysis and Policy*, 2007, 7 (2).
- Elbers, Chris, Jean Lanjouw, and Peter Lanjouw**, “Micro-level Estimation of Poverty and Inequality,” *Econometrica*, 2003, 71 (1), 355–364.
- Emran, M.S. and F. Shilpi**, “Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam,” *Journal of Human Resources*, 2011, 46 (2).
- and **Forhad Shilpi**, “Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India,” *World Development*, 2015, 72.
- Foster, Andrew D. and Mark R. Rosenzweig**, “Economic Development and the Decline of Agricultural Employment,” in “Handbook of Development Economics,” Vol. 4 2007, chapter 47, pp. 3051–3083.
- Frisancho Robles, Veronica C. and Kala Krishna**, “Affirmative Action in Higher Education in India: Targeting, Catch Up and Mismatch,” *Higher Education*, 2016, 71 (5).
- Galor, Oded and Joseph Zeira**, “Income distribution and macroeconomics,” *The review of economic studies*, 1993, 60 (1), 35–52.
- Güell, Maia, José V Rodríguez Mora, and Christopher I. Telmer**, “The informational content of surnames, the evolution of intergenerational mobility, and assortative mating,” *Review of Economic Studies*, 2013, 82 (2).
- Guest, Avery M, Nancy S Landale, and James C Mccann**, “Intergenerational Occupational Mobility in the Late 19th Century United States,” *Social Forces*, 1989, 68 (2).
- Hertz, Tom**, “Rags, riches and race: The intergenerational economic mobility of black and white families in the United States,” in Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves, eds., *Unequal Chances: Family Background and Economic Success*, Princeton University Press, 2005.
- , “A group-specific measure of intergenerational persistence,” *Economics Letters*, 2008, 100 (3), 415–417.

- \_\_\_\_\_, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina, "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends," *The B.E. Journal of Economic Analysis & Policy*, 2008, 7 (2).
- Hilger, Nathaniel G**, "The Great Escape: Intergenerational Mobility Since 1940," 2016. NBER Working Paper No. 21217.
- Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh B. Paul**, "Castes and Labor Mobility," *American Economic Journal: Applied Economics*, 2012, 4 (2).
- \_\_\_\_\_, \_\_\_, and \_\_\_, "Breaking the caste barrier: intergenerational mobility in India," *The Journal of Human Resources*, 2013, 48 (2).
- Ito, Takahiro**, "Caste discrimination and transaction costs in the labor market: Evidence from rural North India," *Journal of Development Economics*, 2009, 88 (2), 292–300.
- Khamis, Melanie, Nishith Prakash, and Zahra Siddique**, "Consumption and social identity: Evidence from India," *Journal of Economic Behavior & Organization*, 2012, 83 (3), 353–371.
- Khanna, Gaurav**, "Does Affirmative Action Incentivize Schooling? Evidence from India," *Review of Economics and Statistics*, 2020, 102 (2).
- Knight, John, Terry Sicular, and Ximing Yue**, "Educational Inequality in China: The Intergenerational Dimension," 2011. Center for Human Capital and Productivity Working Paper.
- Kuran, Timur**, "Islam and economic performance: Historical and contemporary links," *Journal of Economic Literature*, 2018, 56 (4), 1292–1359.
- Li, Hao, Daniel Millimet, and Punarjit Roychowdhury**, "Measuring Economic Mobility in India Using Noisy Data: A Partial Identification Approach," 2019. Working Paper.
- Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren**, "Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations," *Journal of Human Resources*, 2012, 50 (1).
- Long, Jason and Joseph Ferrie**, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, 2013, 103 (4).
- Loury, Glenn C**, "Intergenerational transfers and the distribution of earnings," *Econometrica: Journal of the Econometric Society*, 1981, pp. 843–867.
- Maitra, Pushkar and Anurag Sharma**, "Parents and Children: Education Across Generations in India," 2009. Working Paper.
- Manski, Charles F. and Elie Tamer**, "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 2002, 70 (2), 519–546.
- McCleary, Rachel M. and Robert J. Barro**, "Religion and Economy," *Journal of Economic Perspectives*, 2006, 20 (2).
- Mohammed, AR Shariq**, "Does a good father now have to be rich? Intergenerational income mobility in rural India," *Labour Economics*, 2019, 60.
- Munshi, Kaivan and M Rosenzweig**, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 2016, 106 (1), 46–98.
- \_\_\_\_\_, and **Mark Rosenzweig**, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 2006, 96 (4), 1225–1252.
- Narayan, Ambar and Roy Van der Weide**, *Fair Progress? Economic Mobility Across Generations Around the World*, World Bank Group, 2018.
- Novosad, Paul, Charlie Rafkin, and Sam Asher**, "Mortality Change Among Less Educated Americans," 2020. Working Paper.
- Piraino, Patrizio**, "Intergenerational Earnings Mobility and Equality of Opportunity in South Africa," *World Development*, 2015, 67 (March), 396–405.

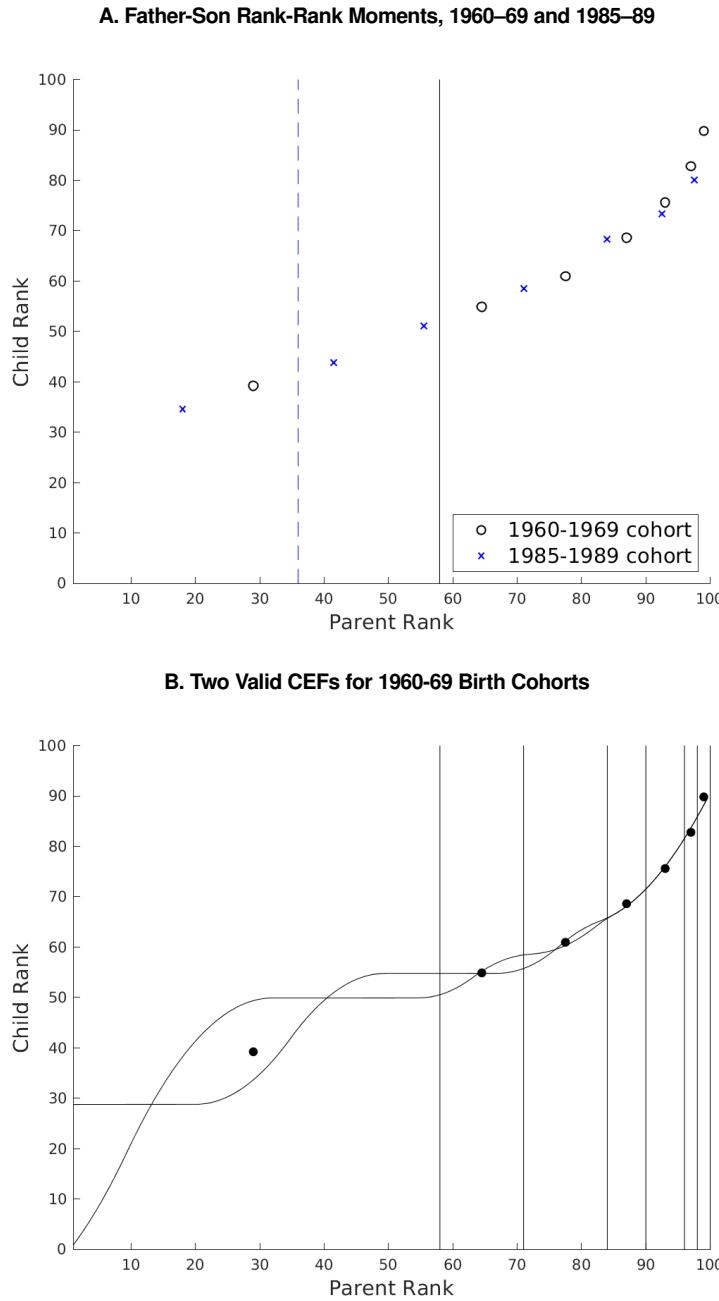
- Platas, Melina R.**, "Culture and the Persistence of Educational Inequality: Lessons from the Muslim–Christian Education Gap in Africa," 2018. New York University Abu Dhabi Working Paper.
- Roemer, John.**, "Equality of Opportunity: Theory and Measurement," *Journal of Economic Literature*, 2016, 54 (4).
- Rosenbaum, Dan T et al.**, "Ability, educational ranks, and labor market trends: The effects of shifts in the skill composition of educational groups," 2000. Working paper.
- Sachar Committee Report**, "Social, Economic and Educational Status of the Muslim Community of India," Technical Report, Government of India 2006.
- Schultz, T Paul**, "School subsidies for the poor: evaluating the Mexican Progresa poverty program," *Journal of Development Economics*, 2003, 74, 199–250.
- Solon, Gary**, "Intergenerational Mobility in the Labor Market," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam: North Holland Press, 1999, pp. 1761–1800.
- Wantchekon, Leonard, Marko Klašnja, and Natalija Novta**, "Education and Human Capital Externalities: Evidence from Benin," *The Quarterly Journal of Economics*, 2015, 130 (2), 703–757.
- Wilkinson, Steven I.**, *Votes and violence: Electoral competition and ethnic riots in India*, Cambridge University Press, 2006.
- Zimmerman, David J.**, "Regression toward mediocrity in economic stature," *The American Economic Review*, 1992.

**Figure 1**  
Educational Attainment in India



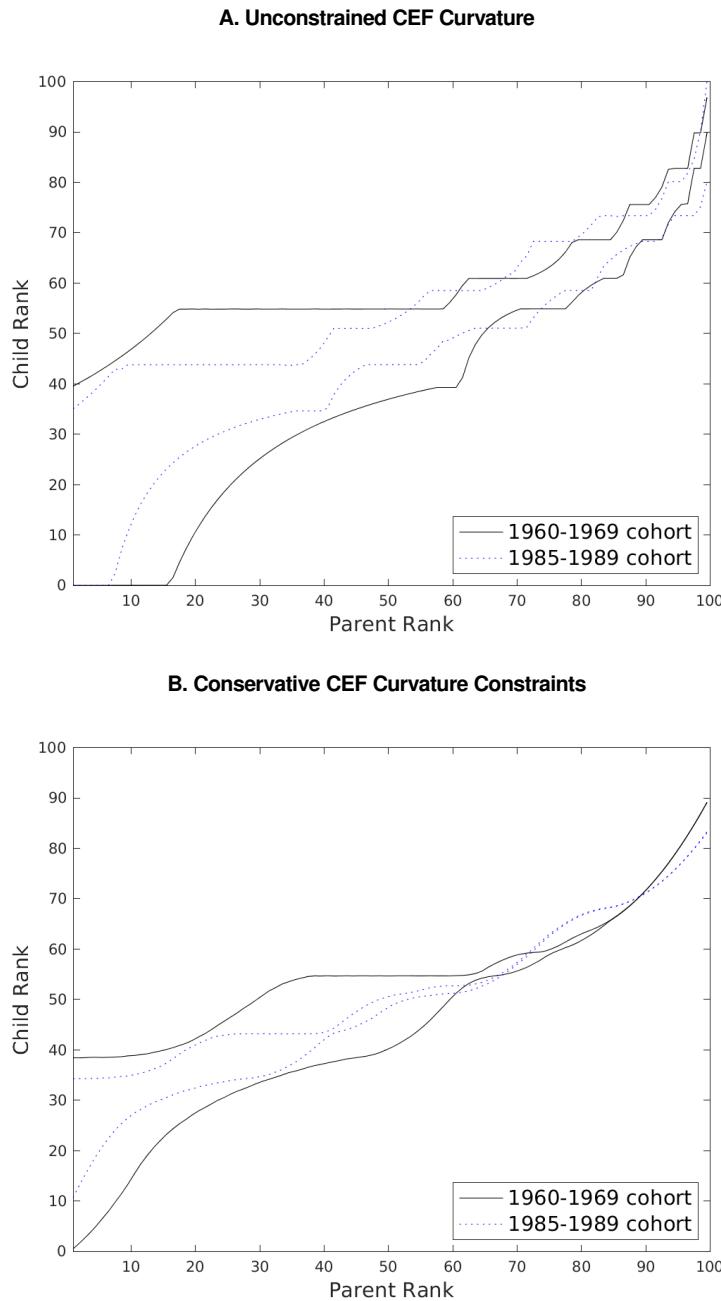
This figure presents average educational attainment by birth cohort in India, using the IHDS samples described in Section 4. Lines are lowess fits.

**Figure 2**  
Father-Son Mobility: Raw Moments and Example CEFs



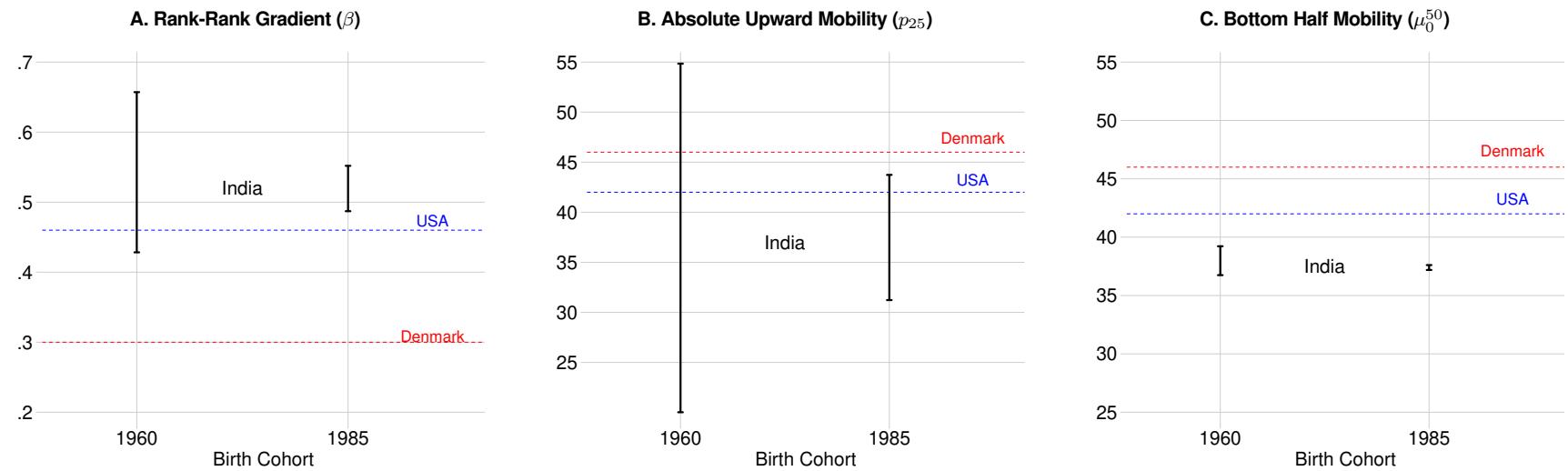
Panel A of the figure shows the average child education rank in each parent education rank bin for the 1960–69 and 1985–89 birth cohorts. The vertical lines show the boundaries for the bottom parent bin, which corresponds to less than two years of education. The solid line corresponds to the 1960–69 birth cohort and the dashed line to the 1985–89 birth cohort. Points are displayed at the midpoint of each parent rank bin. Panel B shows the 1960–69 moments again, along with two simulated conditional expectation functions which are equally good fits to the moments. Source: IHDS 2012.

**Figure 3**  
 Bounded Father-Son Rank CEFs:  
 1960–69 and 1985–89 Birth Cohorts



The figure presents the change over time in the rank-rank relationship between Indian fathers and sons born in the 1960–69 and the 1985–89 birth cohorts. The graph shows bounds on the expected child education rank, given a parent at a given rank in the parent education distribution. Panel A shows bounds on father-son CEFs with unconstrained curvature. Panel B shows bounds using a conservative curvature constraint equal to double the maximum curvature of parent-child *income* rank-rank CEFs from Sweden, Denmark, Norway, and the United States (Chetty et al., 2014b; Boserup et al., 2014; Bratberg et al., 2015). Source: IHDS 2012.

**Figure 4**  
 Bounds on Mobility Measures in India:  
 1960–69 and 1985–89 Birth Cohorts



The figure shows bounds on three mobility statistics for the 1960–69 and 1985–89 birth cohorts, estimated on father-son pairs in India. For reference, we display estimates of similar statistics from USA and Denmark. Data on rank-rank education gradients for USA and Denmark are from Hertz et al. (2008). For  $p_{25}$  and  $\mu_0^{50}$ , the USA and Denmark references are *income* mobility estimates from Chetty et al. (2014a). The Indian measures are all based on education data. The rank-rank gradient is the slope coefficient from a regression of son education rank on father education rank.  $p_{25}$  is absolute upward mobility, which is the expected rank of a son born to a father at the 25th percentile.  $\mu_0^{50}$  is bottom half mobility, which is the expected rank of a son born to a father below the 50th percentile. Source: IHDS 2012.

**Figure 5**  
Sample Calculation of  $\mu_0^{50}$  for 1960–69 Birth Cohort

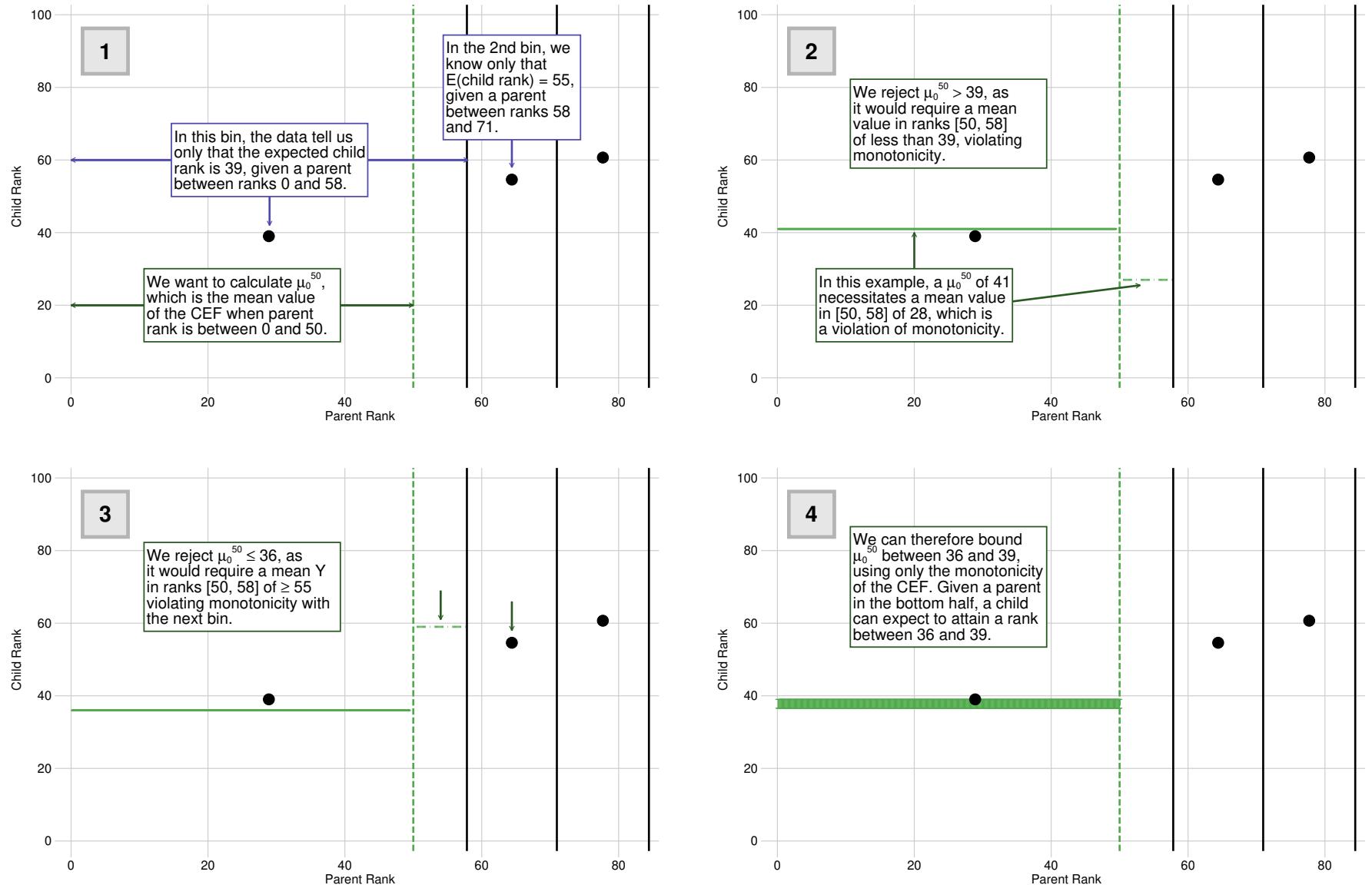


Figure 5 walks through the process of calculating bounds on  $\mu_0^{50} = E(y|x \in [0,50])$  using data from the 1960–69 birth cohort in India. Source: IHDS (2012).

**Figure 6**  
Bottom Half Mobility, Fathers to Sons and Daughters

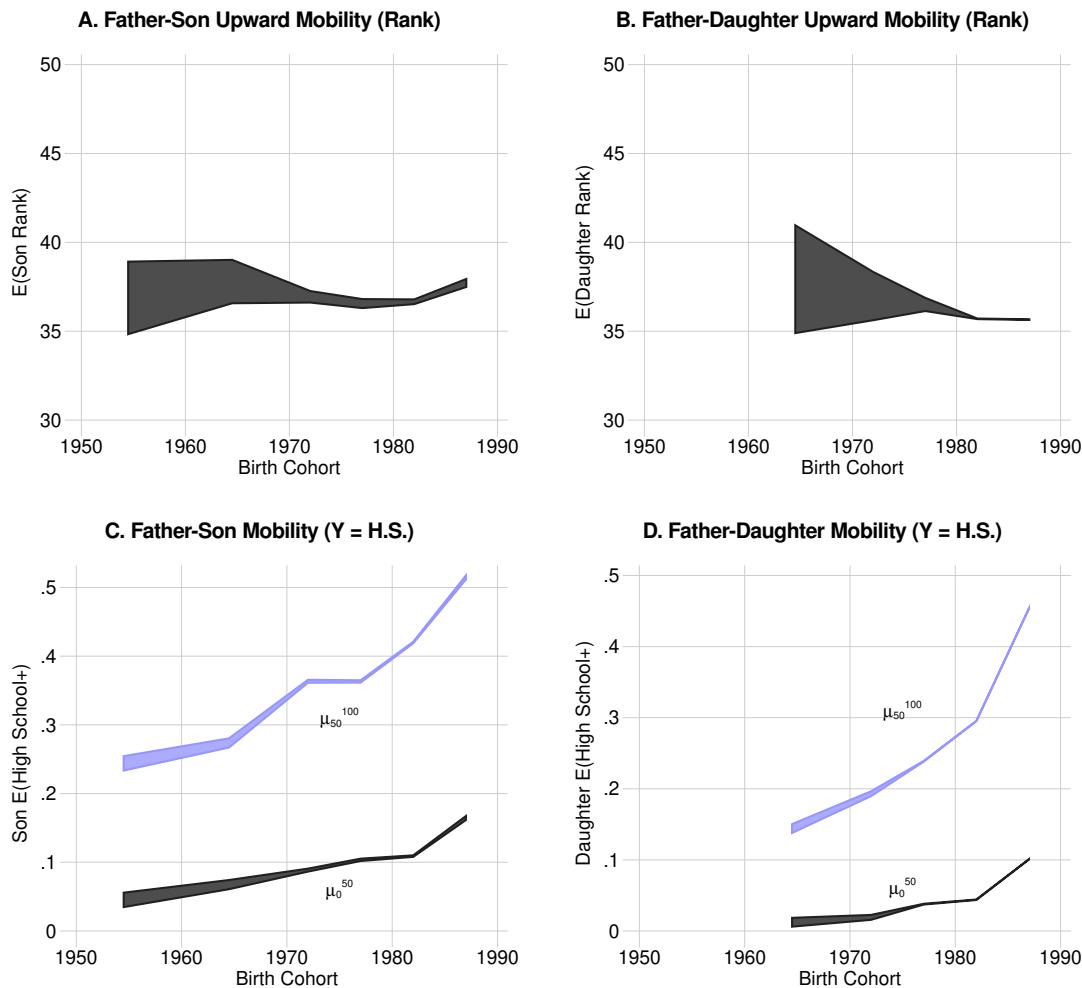


Figure 6 presents bounds on national intergenerational mobility, using cohorts born from 1950 through 1989. Panels A and B show bottom half mobility ( $\mu_0^{50} = E(y|x \in [0,50])$ ), where  $x$  is parent rank and  $y$  is child rank. This is the average rank attained by children born to parents who are in the bottom half of the education distribution, respectively for sons and daughters. Panels C and D show an analogous measure,  $E(HS|x \in [0,50])$  (gray) and  $E(HS|x \in [50,100])$  (blue). The first (gray) is the share of children completing high school, conditional on having parents in the bottom half of the education distribution. The second (blue) is the share of children completing high school, conditional on parents in the top half of the parent distribution. Source: IHDS 2012.

**Figure 7**  
Trends in Mobility by Subgroup, 1950–1989 Birth Cohorts

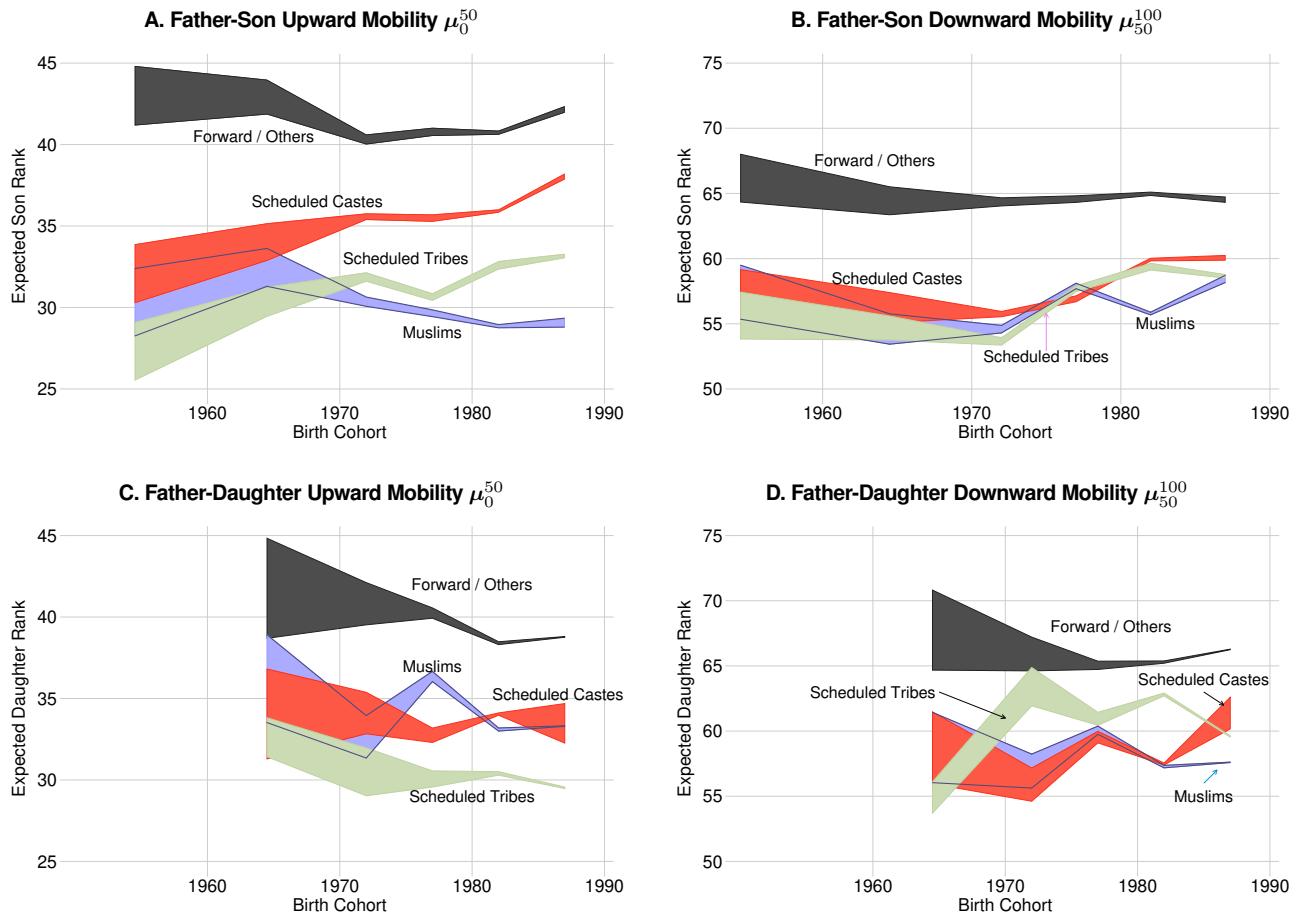


Figure 7 presents bounds on trends in intergenerational mobility, stratified by four prominent social groups in India: Scheduled Castes, Scheduled Tribes, Muslims, and Forward Castes/Others. The mobility measure in Panels A and B is bottom half mobility ( $\mu_0^{50}$ ), or the average rank among children born to fathers in the bottom half of the father education distribution. The measure in Panels C and D is top half mobility ( $\mu_{50}^{100}$ ), or the average rank among children born to fathers in top half of the father education distribution. Linked father-daughter education data are not available for the 1950–59 birth cohort. Source: IHDS 2012.

**Figure 8**  
Upward Mobility by Geographic Location: National and Neighborhood Estimates

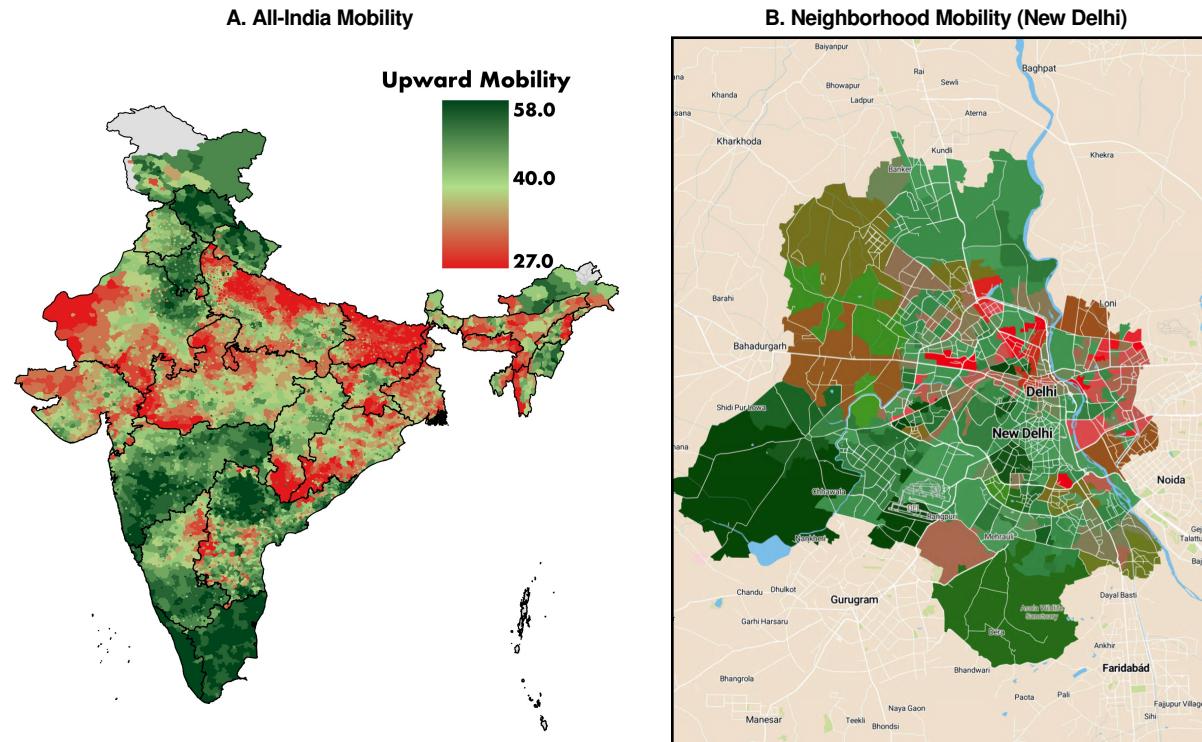


Figure 8 Panel A presents a map of the geographic distribution of upward mobility across Indian subdistricts and towns. Panel B shows a map of the geographic distribution of upward mobility across the wards of Delhi. Upward mobility ( $\mu_0^{50}$ ) is the average education rank attained by sons born to fathers who are in the bottom half of the father education distribution. Green areas have the highest mobility and red areas the lowest. The heat map legend applies to both panels of the figure. Source: SECC 2012.

**Figure 9**  
Correlates of Upward Mobility, 1985–1989 Birth Cohort

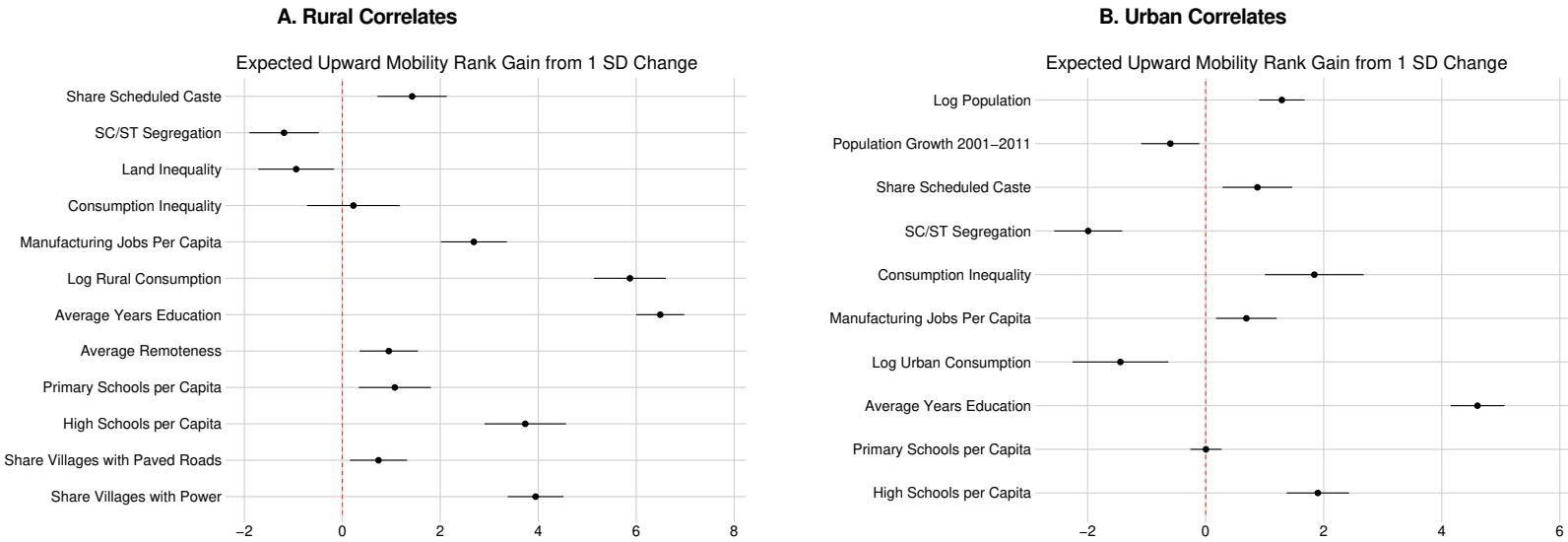


Figure 9 shows the cross-sectional relationship between local upward mobility and various characteristics of locations. Bottom half mobility ( $\mu_0^{50}$ ) is the average rank attained by sons born to fathers who are in the bottom half of the education distribution. Each point in the graph is a coefficient from a bivariate regression of bottom half mobility on the standardized variable on the Y axis. The Y variables are standardized to mean 0 and standard deviation 1 to make them comparable. Source: SECC 2012.

**Figure 10**  
Jati Redesignation and Intergenerational Mobility

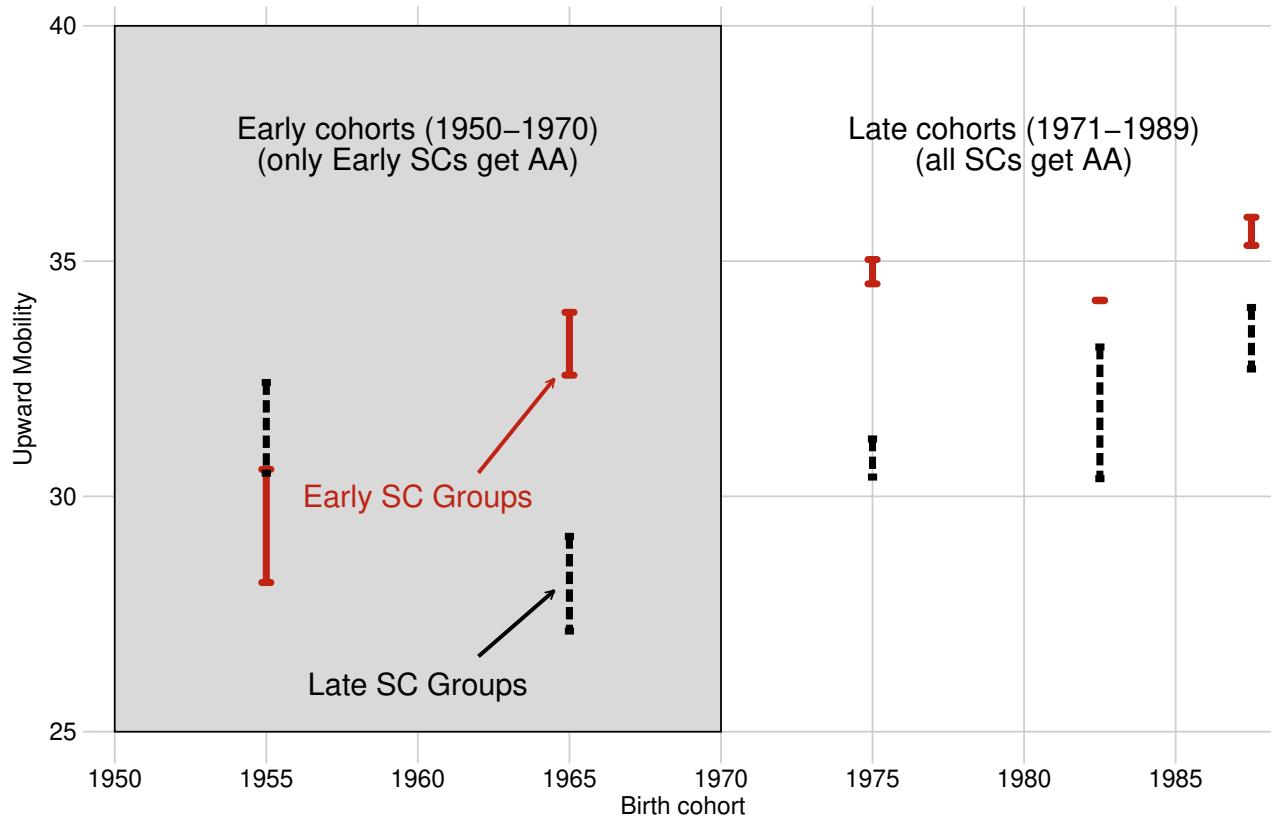


Figure 10 shows bounds on bottom half mobility  $\mu_0^{50}$  for two social groups in India. The dark red series shows upward mobility for groups that were designated as Scheduled Castes beginning in the 1950s. The black series shows upward mobility for groups that were not designated as Scheduled Castes until 1977; birth cohorts later than 1970 (outside the grey box) are those who were young enough to benefit. Source: IHDS 2012.

**Figure 11**  
Within-State and Within-District Social Group Mobility Gaps

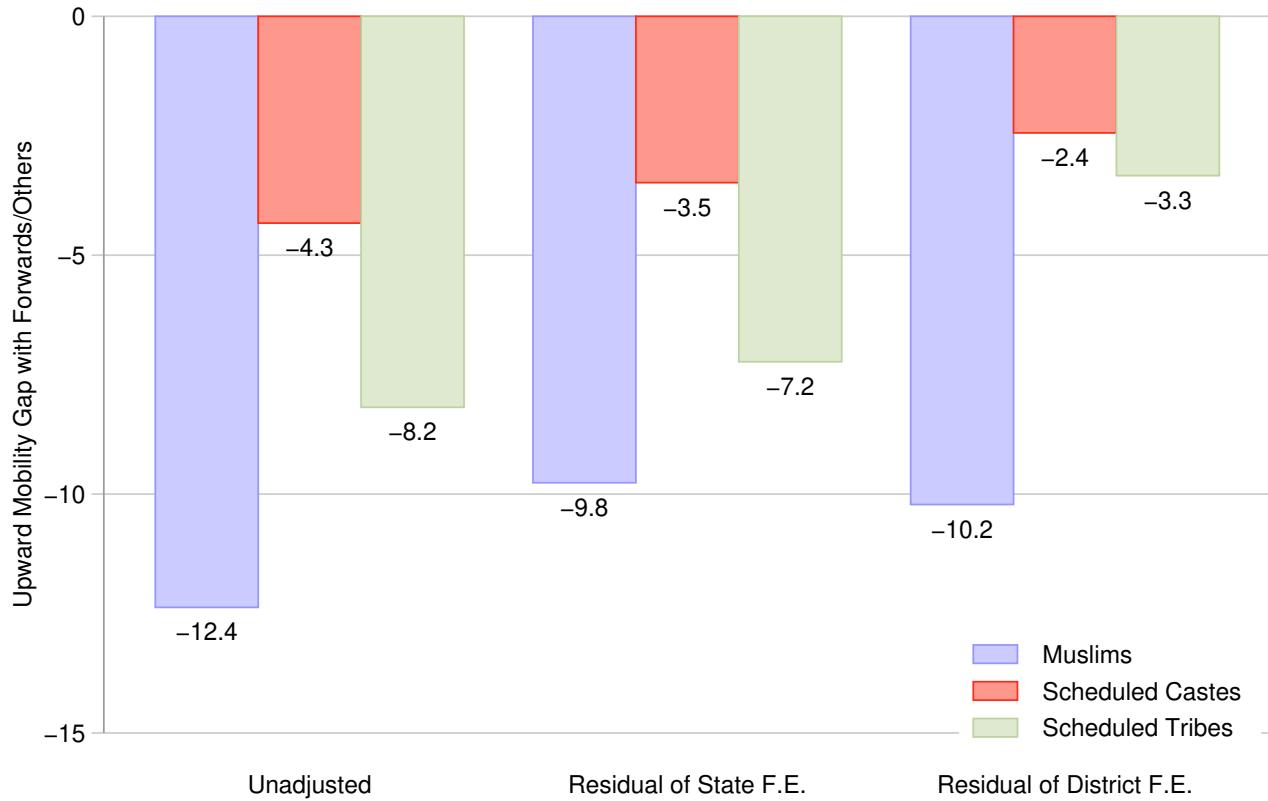


Figure 11 presents the upward mobility disadvantage relative to Forwards/Others faced by Muslims, Scheduled Castes and Scheduled Tribes. The first set of bars shows the aggregate mobility disadvantage of each group in rank terms. The second set of three bars shows the gaps calculated using within-state father and son education ranks. The third set shows gaps calculated using within-district father and son education ranks. Upward mobility in all cases is defined as  $\mu_0^{50} = E(y|x \in [0,50])$ , where  $x$  is parent rank and  $y$  is child rank. Upward mobility is partially identified; for simplicity, we show the midpoint of the bounds, which in all cases span less than a single rank point. Source: IHDS 2012.

**Figure 12**  
Mincerian Returns for Different Social Groups

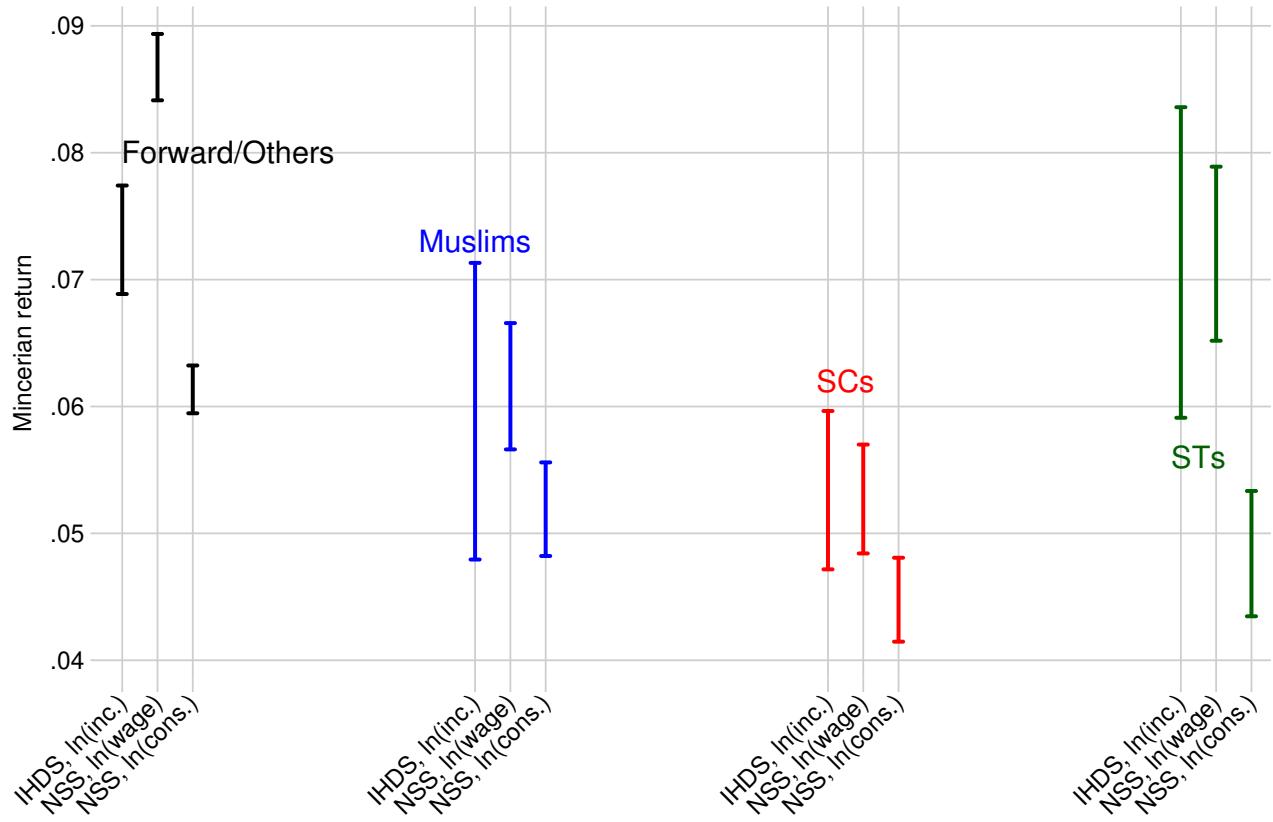
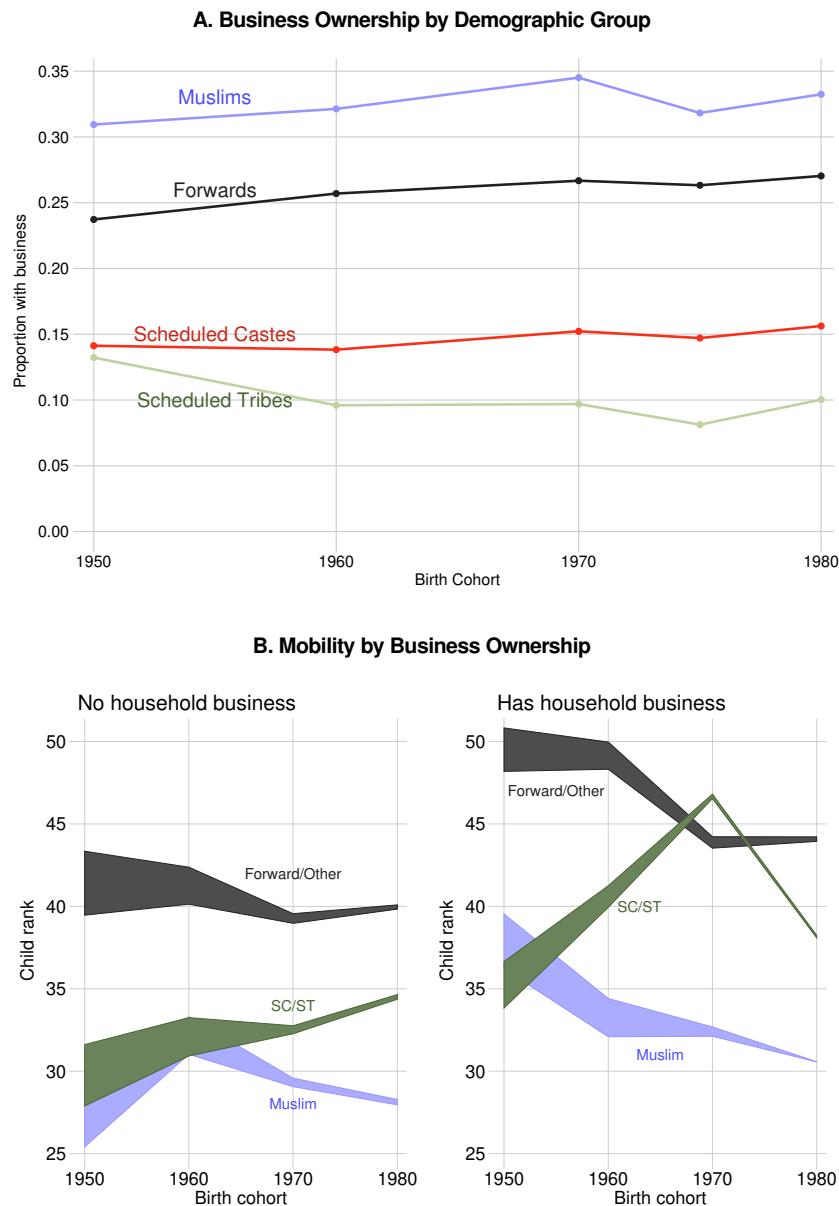


Figure 12 shows Mincerian returns to education for each major social group. Each band shows 95% confidence intervals for the Mincerian return to household log income (IHDS 2012), individual log wages (NSS 2012), and household log per capita income (NSS 2012).

**Figure 13**  
Business Ownership and Intergenerational Mobility



Panel A of Figure 13 shows the share of individuals who report that they work in their own business, by social group and time. Panel B shows bottom half mobility ( $\mu_0^{50}$ ) 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: NSS 2012.

**Table 1**  
Bin Sizes in Studies of Intergenerational Mobility

Study	Country	Birth Cohort of Son	Number of Parent Outcome Bins	Population Share in Largest Bin
Alesina et al. (2021)	Many countries in Africa <sup>67</sup>	1960–2005	5	83%
Card et al. (2018), Derenoncourt (2019)	USA	1920s	≥ 17	41% <sup>68</sup>
Dunn (2007)	Brazil	1972–1981	> 18	20% <sup>69</sup>
Emran and Shilpi (2011)	Nepal, Vietnam	1992–1995	2	83%
Güell et al. (2013)	Spain	~ 2001	9	27% <sup>70</sup>
Guest et al. (1989)	USA	~ 1880	7	53.2%
Hnatkovska et al. (2013)	India	1918–1988	5	Not reported
Knight et al. (2011)	China	1930–1984	5	29% <sup>71</sup>
Lindahl et al. (2012)	Sweden	1865–2005	8	34.5%
Long and Ferrie (2013)	Britain	~ 1850	4	57.6%
	Britain	~ 1949–55	4	54.2%
	USA	~ 1850–51	4	50.9%
	USA	~ 1949–55	4	48.3%
Piraino (2015)	South Africa	1964–1994	6	36%

Table A1 presents a review of papers analyzing educational and occupational mobility. The sample is not representative; we focus on papers where interval censoring may be a concern. The column indicating number of parent outcome bins refers to the number of categories for the parent outcome used in the main specification. The outcome is education in all studies with the exception of Long and Ferrie (2013) and Guest et al. (1989), in which the outcome is occupation.

<sup>67</sup>Many countries are studied; the table shows illustrative statistics for Ethiopia, one of the largest countries in the sample.

<sup>68</sup>Source: Census Bureau (1940).

<sup>69</sup>Includes all people born after about 1990.

<sup>70</sup>Includes all people born after about 1960.

<sup>71</sup>This is the proportion of sons in 1976 who had not completed one year of education — an estimate of the proportion of fathers in 2002 with no education, which is not reported.

<sup>72</sup>Estimate is from the full population rather than just fathers.

<sup>73</sup>This reported estimate does not incorporate sampling weights; estimates with weights are not reported.

**Table 2**  
Changes in Upward Mobility Over Time

A. Father/son pairs

	All groups	Forward/Others	Muslims	SCs	STs
1960–1969	[36.6, 39.0] {35.7, 39.8}	[41.8, 44.0] {40.6, 45.2}	[31.3, 33.6] {29.4, 35.5}	[32.9, 35.2] {31.5, 36.6}	[29.4, 31.3] {27.1, 33.6}
1980–1989	[37.1, 37.2] {36.4, 37.9}	[41.3, 41.3] {40.2, 42.4}	[28.9, 29.0] {27.5, 30.3}	[36.9, 37.0] {35.4, 38.6}	[33.1, 33.1] {31.1, 35.1}
Change over time	[-1.9, 0.6] {-2.9, 1.6}	[-2.7, -0.5] {-4.4, 1.1}	[-4.7, -2.3] {-7.1, 0.1}	[1.8, 4.1] {-0.4, 6.3}	[1.8, 3.7] {-1.3, 6.8}
Fraction overlapping bounds	0.818	0.322	0.050	0.102	0.118

B. Father/daughter pairs

	All groups	Forward/Others	Muslims	SCs	STs
1960–1969	[34.9, 41.0] {34.1, 41.8}	[38.7, 44.8] {37.6, 46.0}	[33.5, 38.9] {31.7, 40.7}	[31.3, 36.8] {29.8, 38.3}	[31.4, 33.8] {29.0, 36.2}
1980–1989	[35.4, 35.5] {34.6, 36.2}	[38.0, 38.2] {36.8, 39.3}	[32.0, 33.5] {31.0, 34.6}	[32.9, 34.2] {31.6, 35.5}	[30.4, 30.5] {28.4, 32.4}
Change over time	[-5.6, 0.6] {-6.7, 1.7}	[-6.9, -0.5] {-8.5, 1.1}	[-6.9, -0.0] {-8.9, 2.0}	[-3.9, 2.9] {-5.9, 4.9}	[-3.4, -0.9] {-6.5, 2.1}
Fraction overlapping bounds	0.802	0.244	0.446	0.982	0.344

Table 2 shows estimates of full sample and subgroup bottom half mobility ( $\mu_0^{50}$ ) for the 1960–69 and 1980–89 birth cohorts for father-son (Panel A) and father-daughter (Panel B) pairs. We show both bounds (in square brackets) and 90% confidence sets (in curly braces) on those bounds. The table also reports the bounds and 90% confidence sets on the change in bottom half mobility between these two time periods. We obtain confidence sets by generating 1,000 bootstrap draws, estimating bounds on each bootstrap draw, and following the framework in Chernozhukov et al. (2007) to form 90% confidence sets from bootstrapped bounds. Because these are confidence sets rather than confidence intervals, instead of  $p$ -values we show the fraction of bootstraps in which the 1960–69 and 1980–89 bounds are overlapping. Source: IHDS (2012).

**Table 3**  
Group Differences in Upward Mobility

	F/O minus SC	F/O minus Muslim	SC minus Muslim
Father/son ( $\mu_0^{50}$ )	[4.6, 5.0] {2.8, 6.9}	[11.6, 12.1] {9.8, 13.9}	[6.9, 7.3] {4.5, 9.7}
Fraction overlapping bounds	0.000	0.000	0.000
Father/daughter ( $\mu_0^{50}$ )	[4.2, 4.5] {1.8, 6.9}	[5.1, 5.5] {3.0, 7.6}	[0.8, 1.1] {-2.1, 4.0}
Fraction overlapping bounds	0.000	0.000	0.509
Father/son ( $\mu_{50}^{100}$ )	[4.8, 5.3] {3.3, 6.8}	[9.0, 9.4] {5.7, 12.7}	[3.9, 4.3] {0.7, 7.6}
Fraction overlapping bounds	0.000	0.000	0.012
Father/daughter ( $\mu_{50}^{100}$ )	[7.7, 8.0] {4.0, 11.7}	[7.8, 8.2] {5.2, 10.8}	[0.0, 0.4] {-3.9, 4.3}
Fraction overlapping bounds	0.000	0.000	0.329

Table 3 shows estimates of cross-group differences in bottom half mobility ( $\mu_0^{50}$ ) and top half mobility ( $\mu_{50}^{100}$ ) in the 1980–89 birth cohorts. We show both bounds (in square brackets) and 90% confidence sets (in curly braces) on those bounds. We obtain confidence sets by generating 1,000 bootstrap draws, estimating bounds on each bootstrap draw, and following the framework in Chernozhukov et al. (2007) to form 90% confidence sets from bootstrapped bounds. Because these are confidence sets rather than confidence intervals, instead of  $p$ -values we show the fraction of the bounds for the two social groups that are overlapping. For example, the value of 0.509 in the final column indicates that 50.9% of the bootstraps generate overlapping bounds for the two groups. Source: IHDS (2012).

**Table 4**  
Effect of Caste Redesignation on Scheduled Caste Upward Mobility

	(1)	(2)	(3)
Post * Late SC	8.432*** (1.794)	6.764*** (1.555)	
1970-79 * Late SC			6.739** (2.025)
1980-89 * Late SC			9.649*** (2.580)
N	4502	3746	4502
r2	0.32	0.34	0.32

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 4 shows estimates from Equation 6.1, which describes the impact of Scheduled Caste redesignation on upward mobility. The dependent variable is the child education rank. The sample consists of SC sons of fathers with less than two years of education in the 1960s and 1970s, and SC sons of fathers with 2 or fewer years of education in the 1980s. The dependent variable thus corresponds to  $\mu_0^{59}$  in the 1950s birth cohort,  $\mu_0^{57}$  in the 1960s, and  $\mu_0^{58}$  in the 1970s and 1980s. Late SC is an indicator for jati groups that were added to Scheduled Caste lists in the caste redesignation of 1977. All estimations control for region\*cohort, jati\*region, jati\*cohort, and birth year, and are clustered at the jati and the region levels. Source: IHDS (2012).

**Table 5**  
Relationship Between Fertility and Subgroup Upward Mobility

	(1)	(2)	(3)
Muslim	-13.476*** (0.976)	-12.338*** (1.697)	-9.287*** (1.721)
Scheduled Caste	-4.163*** (0.749)	-2.608** (1.281)	-1.901 (1.268)
Scheduled Tribe	-9.075*** (1.076)	-8.040*** (1.851)	-8.291*** (1.829)
Urban	3.881*** (0.782)	3.812*** (1.276)	3.514*** (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 5 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  $\mu_0^{51}$ , a close analog of bottom half mobility ( $\mu_0^{50}$ ). 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).

## A Appendix A: Additional Tables and Figures

**Figure A1**  
Coresidence Rates by Age and Gender

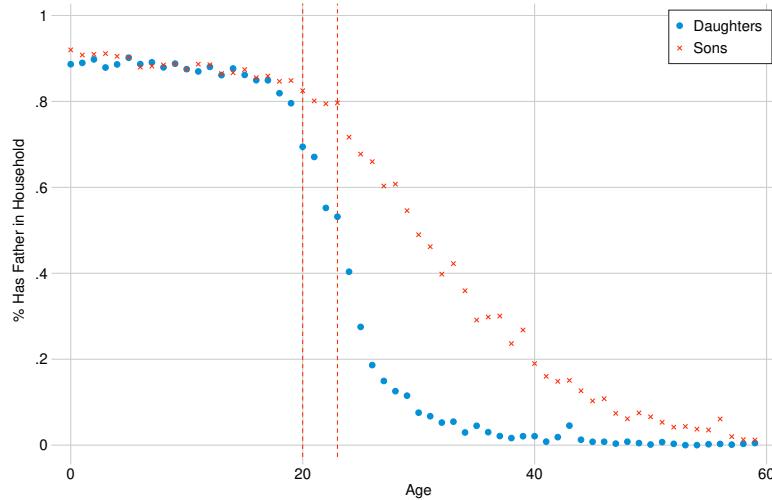


Figure A1 shows the share of individuals who live in the same household as their father as a function of gender and age. The vertical lines indicate ages 20–23, which is the sample restriction used in the SECC for the geographic analysis. Source: IHDS (2012).

**Figure A2**  
Bias in Mobility Estimates When Sample is Limited to Coresident Pairs

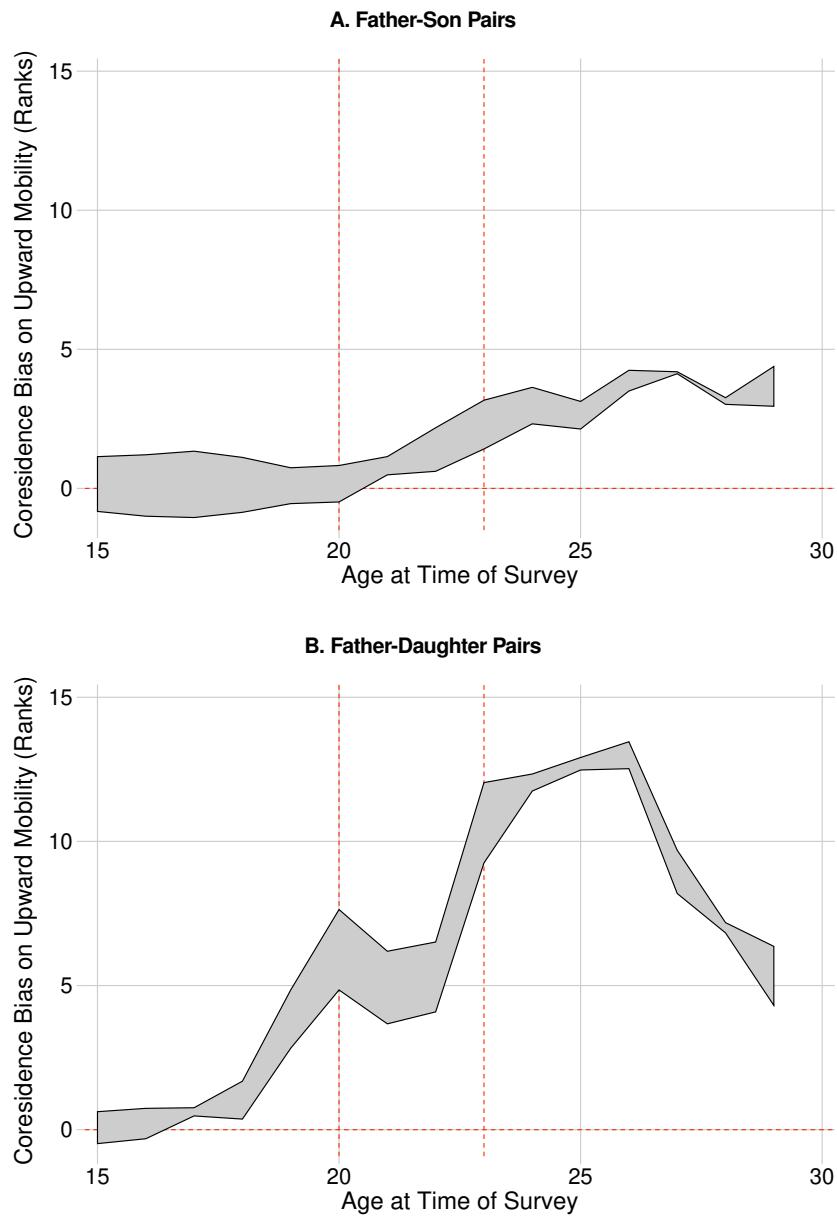


Figure A2 shows the bias in a measure of upward mobility when children who do not live with their parents are excluded. The bias is shown as a function of child age. The mobility measure is bottom half mobility ( $\mu_0^{50}$ ), which is the expected child rank conditional on being born to a parent in 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**  
Robustness of Upward Mobility to Survivorship Bias

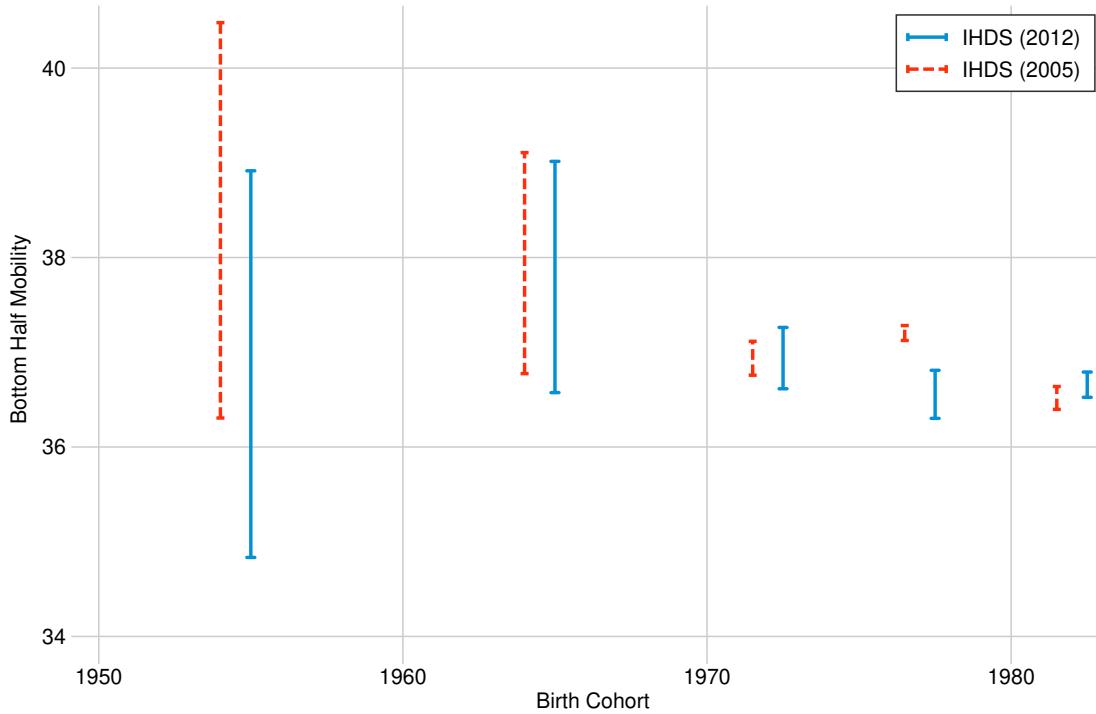


Figure A3 shows a test of survivorship bias in estimates of bottom half mobility. The figure shows estimates of bottom half mobility calculated for the 1950s to 1980–85 birth cohorts, measured separately in the 2005 and 2012 rounds of the IHDS. If there was substantial survivorship bias in the mobility measures, we would expect the estimates to differ across the two surveys because of the deaths of some of the respondents.

**Figure A4**  
Bottom Half Mobility ( $\mu_0^{50}$ ) for Mother-Son and Mother-Daughter Pairs

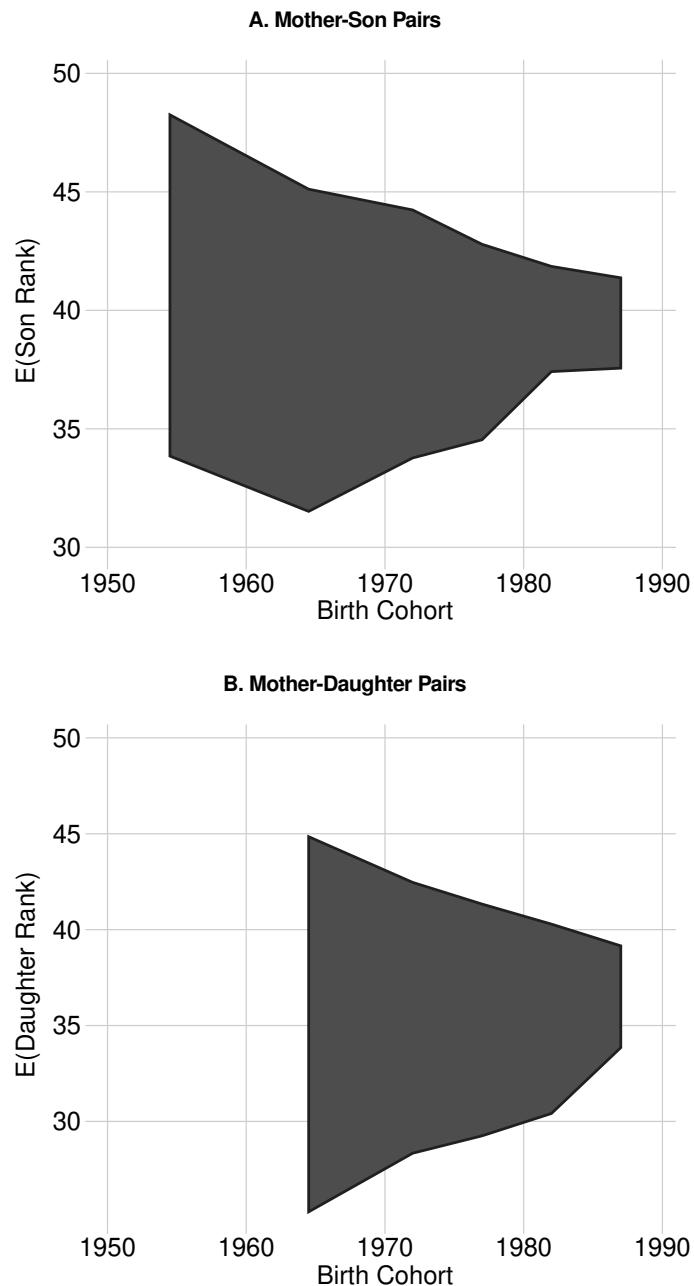


Figure A4 shows bounds on aggregate trends in intergenerational mobility, using cohorts born from 1950–59 through 1985–89, focusing on mother-son and mother-daughter links. The measure used is bottom half mobility ( $\mu_0^{50}$ ), which is the average rank attained by children born to parents who are 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 A5**  
Trends in Mobility by Subgroup, 1950–1989 Birth Cohorts  
Education Level Outcomes

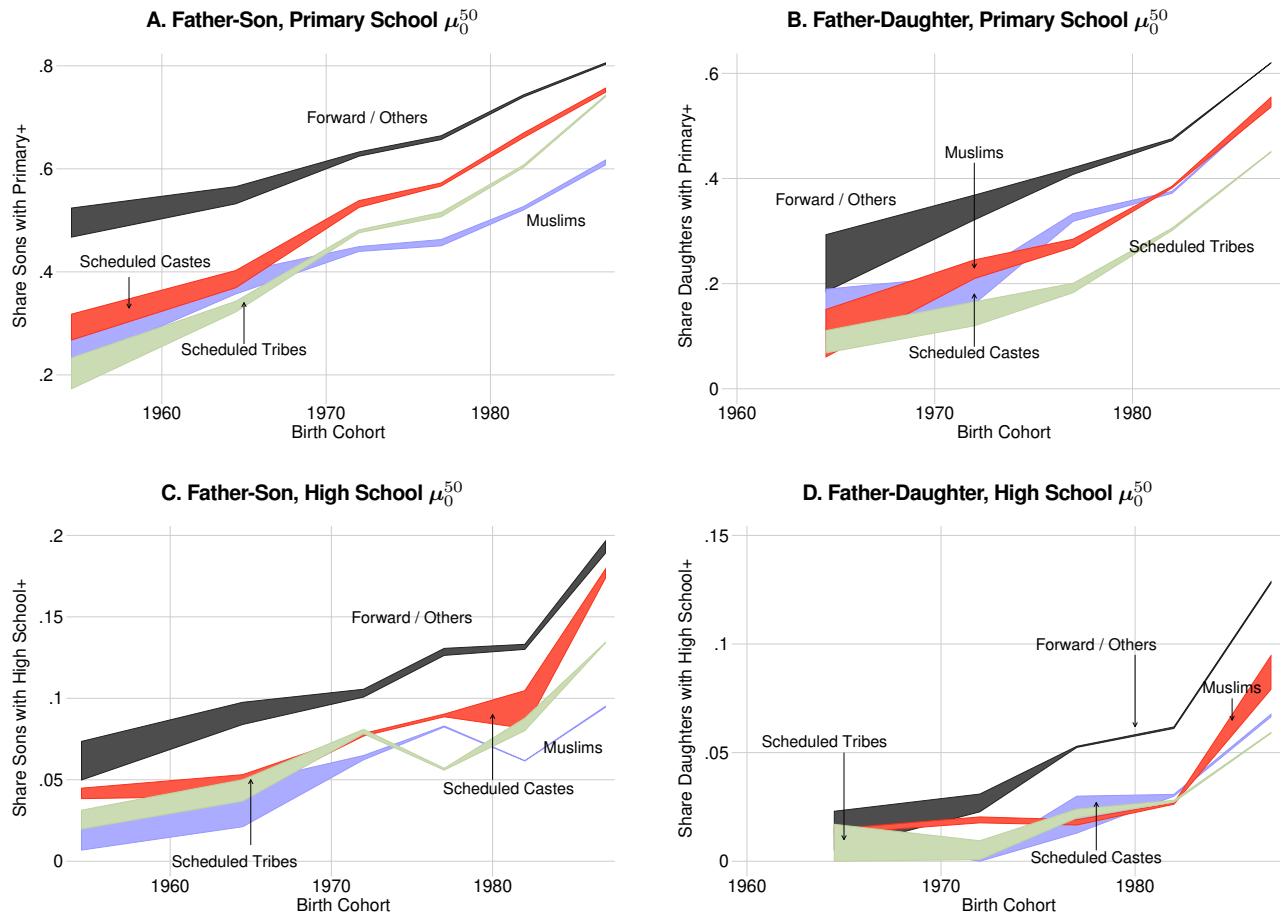


Figure A5 presents bounds on intergenerational mobility, stratified by four prominent social groups in India: Scheduled Castes, Scheduled Tribes, Muslims, and Forward Castes/Others. The figure is analogous to Figure 7, but shows the expected probability that a child attains a given education level (primary 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 A6**  
Bottom Half Mobility, Separated by Urban/Rural Population

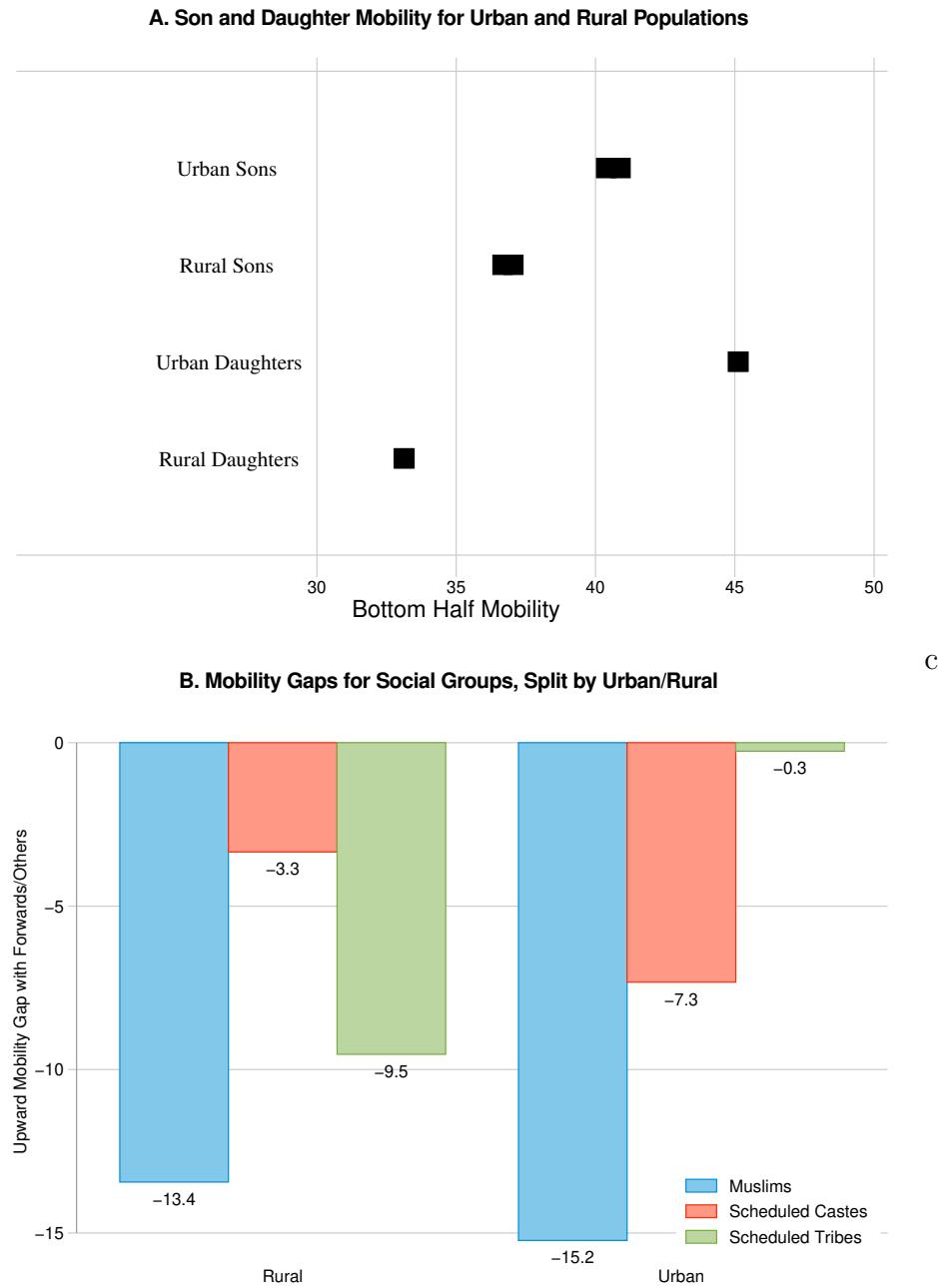


Figure A6A shows estimates of bottom half mobility ( $\mu_0^{50}$ ) for the 1985–89 birth cohort, disaggregated by gender and by urban/rural residence at the time of the survey. Panel B of the figure shows the gap in upward mobility between each population subgroup and the Forward/Other group, disaggregated by urban and rural residence. For example, the first bar shows that Muslim upward mobility is 13.4 rank points lower than Forward/Other upward mobility. Source: IHDS (2012).

**Table A1**  
Bin Sizes in Studies of Intergenerational Mobility

Study	Country	Birth Cohort of Son	Number of Parent Outcome Bins	Population Share in Largest Bin
Alesina et al. (2021)	Many countries in Africa <sup>74</sup>	1960–2005	5	83%
Card et al. (2018), Derenoncourt (2019)	USA	1920s	≥ 17	41% <sup>75</sup>
Dunn (2007)	Brazil	1972–1981	> 18	20% <sup>76</sup>
Emran and Shilpi (2011)	Nepal, Vietnam	1992–1995	2	83%
Güell et al. (2013)	Spain	~ 2001	9	27% <sup>77</sup>
Guest et al. (1989)	USA	~ 1880	7	53.2%
Hnatkovska et al. (2013)	India	1918–1988	5	Not reported
Knight et al. (2011)	China	1930–1984	5	29% <sup>78</sup>
Lindahl et al. (2012)	Sweden	1865–2005	8	34.5%
Long and Ferrie (2013)	Britain	~ 1850	4	57.6%
	Britain	~ 1949–55	4	54.2%
	USA	~ 1850–51	4	50.9%
	USA	~ 1949–55	4	48.3%
Piraino (2015)	South Africa	1964–1994	6	36%

Table A1 presents a review of papers analyzing educational and occupational mobility. The sample is not representative; we focus on papers where interval censoring may be a concern. The column indicating number of parent outcome bins refers to the number of categories for the parent outcome used in the main specification. The outcome is education in all studies with the exception of Long and Ferrie (2013) and Guest et al. (1989), in which the outcome is occupation.

<sup>74</sup>Many countries are studied; the table shows illustrative statistics for Ethiopia, one of the largest countries in the sample.

<sup>75</sup>Number of parent bins refers to the most granular years of education available, from Card et al. (2018) Figure 1. Other analyses in Card et al. (2018) include coarser bins. Source for population share in largest bin: Census Bureau (1940).

<sup>76</sup>Includes all people born after about 1990.

<sup>77</sup>Includes all people born after about 1960.

<sup>78</sup>This is the proportion of sons in 1976 who had not completed one year of education — an estimate of the proportion of fathers in 2002 with no education, which is not reported.

<sup>79</sup>Estimate is from the full population rather than just fathers.

<sup>80</sup>This reported estimate does not incorporate sampling weights; estimates with weights are not reported.

**Table A2**  
Transition Matrices for Father and Son Education in India

**A. Sons Born 1950-59**

Father ed attained	Son highest education attained						
	< 2 yrs. (31%)	2-4 yrs. (11%)	Primary (17%)	Middle (13%)	Sec. (13%)	Sr. sec. (6%)	Any higher (8%)
<2 yrs. (60%)	0.47	0.12	0.17	0.11	0.09	0.03	0.03
2-4 yrs. (12%)	0.10	0.18	0.22	0.19	0.16	0.09	0.06
Primary (13%)	0.07	0.08	0.31	0.16	0.19	0.08	0.10
Middle (6%)	0.06	0.05	0.09	0.30	0.17	0.14	0.18
Secondary (5%)	0.03	0.02	0.04	0.12	0.37	0.11	0.30
Sr. secondary (2%)	0.02	0.00	0.03	0.11	0.11	0.35	0.38
Any higher ed (2%)	0.01	0.01	0.01	0.03	0.08	0.13	0.72

**B. Sons Born 1960-69**

Father ed attained	Son highest education attained						
	< 2 yrs. (27%)	2-4 yrs. (10%)	Primary (16%)	Middle (16%)	Sec. (14%)	Sr. sec. (7%)	Any higher (10%)
<2 yrs. (57%)	0.41	0.12	0.16	0.14	0.09	0.04	0.04
2-4 yrs. (13%)	0.12	0.17	0.18	0.22	0.15	0.08	0.08
Primary (14%)	0.09	0.05	0.26	0.18	0.20	0.09	0.13
Middle (6%)	0.06	0.04	0.09	0.29	0.21	0.13	0.19
Secondary (6%)	0.03	0.02	0.08	0.12	0.35	0.16	0.25
Sr. secondary (2%)	0.02	0.02	0.03	0.07	0.19	0.25	0.41
Any higher ed (2%)	0.01	0.01	0.02	0.03	0.09	0.11	0.73

**C. Sons Born 1970-79**

Father ed attained	Son highest education attained						
	< 2 yrs. (20%)	2-4 yrs. (8%)	Primary (17%)	Middle (18%)	Sec. (16%)	Sr. sec. (10%)	Any higher (12%)
<2 yrs. (50%)	0.33	0.10	0.19	0.17	0.12	0.05	0.04
2-4 yrs. (11%)	0.11	0.16	0.20	0.22	0.15	0.08	0.08
Primary (15%)	0.08	0.06	0.24	0.23	0.18	0.11	0.11
Middle (8%)	0.05	0.03	0.09	0.29	0.21	0.17	0.16
Secondary (9%)	0.03	0.02	0.06	0.12	0.31	0.19	0.27
Sr. secondary (3%)	0.01	0.01	0.02	0.08	0.17	0.29	0.42
Any higher ed (4%)	0.00	0.00	0.02	0.05	0.10	0.17	0.66

**D. Sons Born 1980-89**

Father ed attained	Son highest education attained						
	< 2 yrs. (12%)	2-4 yrs. (7%)	Primary (16%)	Middle (20%)	Sec. (16%)	Sr. sec. (12%)	Any higher (17%)
<2 yrs. (38%)	0.26	0.10	0.21	0.20	0.12	0.06	0.05
2-4 yrs. (11%)	0.08	0.17	0.19	0.24	0.15	0.09	0.08
Primary (17%)	0.05	0.04	0.22	0.23	0.20	0.13	0.13
Middle (12%)	0.03	0.02	0.10	0.28	0.20	0.17	0.20
Secondary (11%)	0.02	0.01	0.05	0.13	0.23	0.24	0.32
Sr. secondary (5%)	0.02	0.01	0.04	0.09	0.15	0.24	0.46
Any higher ed (5%)	0.01	0.01	0.02	0.05	0.10	0.16	0.65

Table A2 shows transition matrices by decadal birth cohort for Indian fathers and sons in the study.  
Source: IHDS (2012).

**Table A3**  
Internal Consistency of Reports of Parents' Education

	Father-Son (1)	Father-Daughter (2)	Father-Daughter (3)	Mother-Daughter (4)	Mother-Daughter (5)	Mother-Daughter (6)
Age	-0.000 (0.008)		-0.018 (0.016)		-0.008 (0.007)	
Child years of education	0.008 (0.013)		0.037* (0.021)		0.003 (0.011)	
Log household income	-0.005 (0.029)		-0.051 (0.058)		-0.026 (0.036)	
Constant	0.053 (0.056)	0.054 (0.431)	-0.002 (0.103)	0.912 (0.841)	0.006 (0.052)	0.545 (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 A3 shows measures of internal consistency when there are multiple reports of an individual's father in the IHDS. Each column is a regression of the level difference between two different measures of a parent's education. The constant term in Columns 1, 3, and 5 thus shows the average differences, and Columns 2, 4, and 6 effectively regress that difference on individual characteristics to measure the extent to which they predict the discrepancy. Source: IHDS (2012).

**Table A4**  
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 the SECC (Panel A, identical to Table 2) and using IHDS data with unadjusted granular years of education (Panel B). The results are similar because there are very few individuals with education levels which were both in the bottom 50% and needed to be downcoded.

## B Appendix B: Details on Methodology

This section provides details about the analytical and numerical procedures to bound the CEF and functions of the CEF. These methods are straightforward applications of Novosad et al. (2020). In Appendix B.1 and Appendix B.2, we reproduce the text of several propositions contained in Novosad et al. (2020) for ease of reference, but relegate the proofs to Novosad et al. (2020). In Appendix B.3, we explain the simple procedure to adapt the numerical techniques in Novosad et al. (2020) to this setting.

**Relationship to Novosad et al. (2020).** Novosad et al. (2020) 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. (2020). Here, we assume child status is *increasing* in parent education rank; Novosad et al. (2020) 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  $k \in \{1, \dots, K\}$ . 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

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

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):

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

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

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

**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:

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

where

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

### B.2 Formal Statement of Analytical Bounds on $\mu_a^b$

We now state a proposition, also contained in Novosad et al. (2020), that permits us to bound  $\mu_a^b$ .

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  $\mu_a^b$  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 procedure for bounding arbitrary functions of the CEF. 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., monotonicity 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  $k \in \{1, \dots, K\}$ . Then define  $\{r_k\}_{k=1}^K$  as the set of observed mean values of  $y$  over each bin  $k \in \{1, \dots, K\}$ . 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:

$$\begin{aligned} S(\{r_k\}_{k=1}^K) = & \left\{ Y(x) \mid Y(x) \text{ is weakly increasing} \right\} \\ & \cap \left\{ Y(x) \mid \frac{1}{x_{k+1}-x_k} \int_{x_k}^{x_{k+1}} (Y(x)-r_k(x)) dx = 0, \text{ for all } k \right\}. \end{aligned} \quad (\text{B.1})$$

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  $\gamma$ :

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

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

Novosad et al. (2020) provide details on the numerical techniques used to solve this problem. The bounds we report are the set  $[\gamma^{\min}, \gamma^{\max}]$ . We now describe how we apply this process in the case of the rank-rank gradient and with curvature constraints.

**Rank-rank gradient.** In the case of the rank-rank gradient, we let  $\gamma$  represent the slope of the linear approximation to the CEF. That is, fixing a CEF  $Y(x)$ , define

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

**Curvature constraints.** In the case of reporting the CEF with curvature constraints (e.g., Figure 3B), we simply define  $p_x(Y)$  to be the value of the CEF at a given  $x$ . We define  $S$  to be the set of CEFs that are consistent with monotonicity and a second derivative that lies below a given magnitude in absolute value. In the case where there are no CEFs that precisely match the bin means (e.g., for a small enough curvature constraint), we solve a modified problem described formally in Novosad et al. (2020). Define  $T(\{r_k\}_{k=1}^K)$  to be the set of CEFs that (a) minimize some distance metric between the bin means and the CEF, and (b) are consistent with the observed bin means and extra assumptions (monotonicity and curvature). In particular, for distance metric  $\|\cdot\|$ , define

$$M(Y) := \int_0^{100} \|Y(x) - r_k(x)\| dx,$$

where  $r_k(x) := r_k$  if  $x \in [x_k, x_{k+1}]$ .  $M(Y)$  is the weighted distance between a given CEF  $Y$  and the bin means  $\{r_k\}$ . Then define

$$T(\{r_k\}_{k=1}^K) = P \cap \left\{ Y(x) \mid \left( \int_0^{100} \|Y(x) - r_k(x)\| dx \right) \leq \underline{M}(P) \right\}, \quad (\text{B.4})$$

for

$$\underline{M}(P) := \min_{Y \in P} M(Y)$$

and

$$P := \left\{ Y(x) \mid Y(x) \text{ is weakly increasing and has second derivative less than } \underline{C} \right\}.$$

Put otherwise, we find the minimum distance between the CEFs and observed bin means, as long as these CEFs obey certain properties. Then, we find the set of CEFs that obey these properties and satisfy this minimum distance. If there are CEFs that precisely meet the bin means, then  $T = S$ .

Finally, we report:

$$p_x^{\min} = \min_{Y \in T(\{r_k\}_{k=1}^K)} \tilde{p}_x(Y) \quad (\text{B.5})$$

$$p_x^{\max} = \max_{Y \in T(\{r_k\}_{k=1}^K)} \tilde{p}_x(Y). \quad (\text{B.6})$$

In practice, we choose the mean-squared error as the distance metric.

## C Appendix C: Robustness to Alternate Assumptions

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

Our bounds on the full sample CEF  $E(y|x)$  (explained in Section 3.3) use the uniformity of the rank distribution, which is given when working with the 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 not be driven by a difference in outcomes conditional on father education rank, but could be driven 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 or how it changes over time. In this section, we present two pieces of evidence that these departures do not bias our primary results.

First, 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. Second, 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 present in any comparison of groups that conditions on education levels. However, our discussion of latent education ranks makes this concern particularly transparent.

### C.1.1 Using Within-Subgroup Rank Distributions which are Uniform by Construction

We focus in this section on the finding that Scheduled Caste and Muslim upward mobility have diverged from the 1960s to the 1980s birth cohorts (Figure 7). If the latent ranks of Muslim parents fell relative to SC parents, conditional on being in each education bin, then this result could be spurious—a mechanical result of our assumption that Muslim and SC parent latent ranks are uniformly distributed within each rank bin.

We show here that this result is robust to calculating parent ranks *within* subgroups. Under this rank definition, the latent parent ranks are uniform by construction—the latent ranks of SCs in the bottom 50% of SCs must be uniformly distributed. 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.

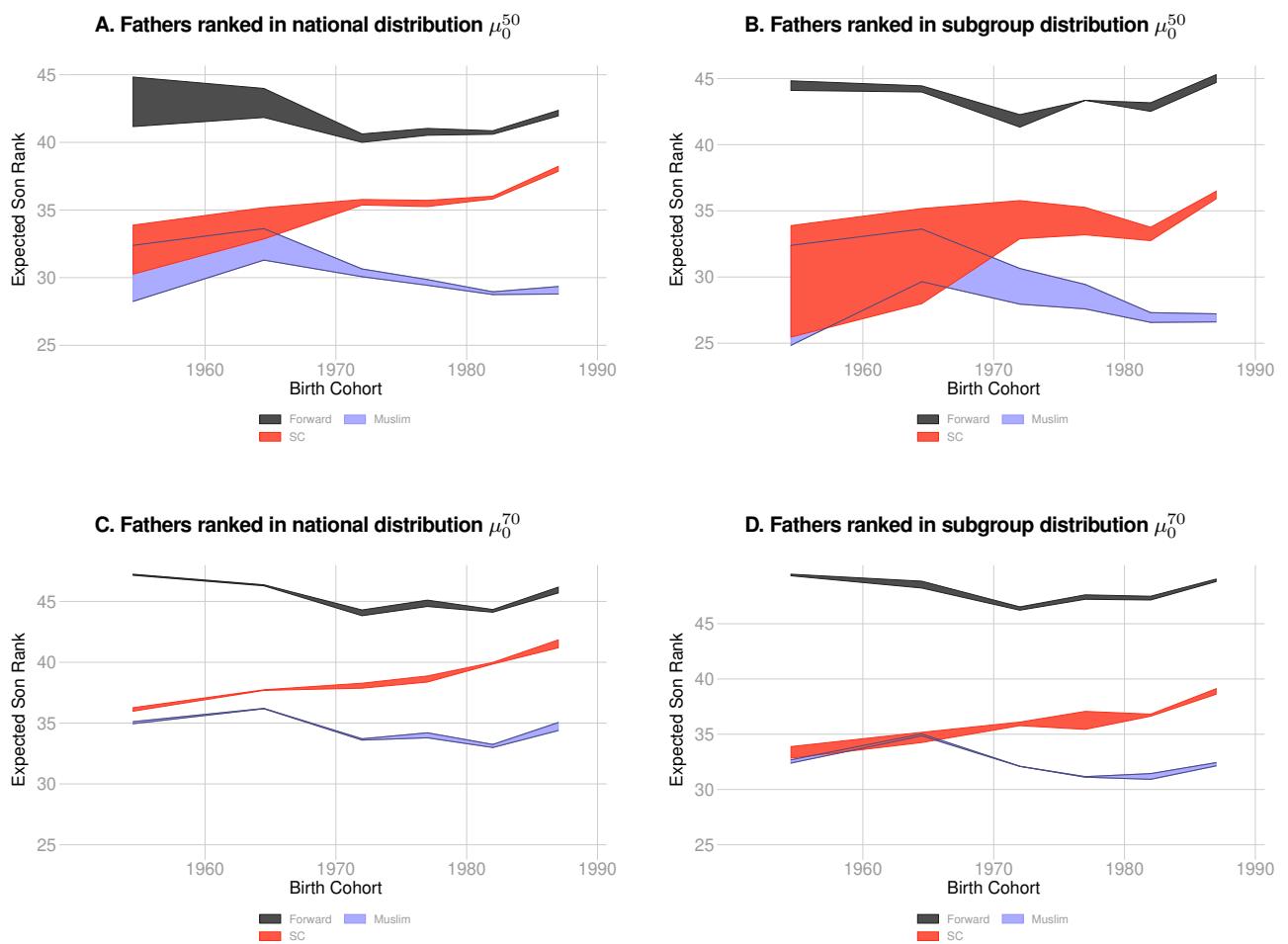
Figure C1 shows the result. Panel A repeats the result of Figure 7 for Forwards, Muslims, and SCs, using national ranks, showing changes in upward mobility ( $\mu_0^{50}$ ) over time for each group. Panel B shows the same result, with parent ranks calculated within their own subgroups. The bounds in 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. More than 70% of SC parents in the 1960s report a bottom-coded education level, resulting in wide bounds on  $\mu_0^{50}$  for this rank definition.

To tighten the bounds, we instead estimate  $\mu_0^{70}$ : the expected child outcome given a parent in the bottom 70% of the parent education distribution. Panel C shows  $\mu_0^{70}$  calculated using national ranks, as in the body of the paper. Panel D shows  $\mu_0^{70}$  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.

Our claim in the body of the paper is that SCs and Muslims with parents in the least educated 50–70% (nationally) have divergent outcomes. We show here that SCs and Muslims in the least educated 50–70% of SCs and Muslims (respectively) have similarly divergent outcomes. The consistency of our results when we calculate parent ranks 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 C1**  
Subgroup Upward Mobility (Fathers/Sons):  
National Ranks vs. Within-Subgroup Ranks



### C.1.2 Inferring Latent Education Rank from Parametric Assumptions

We impute the within-bin latent education rank distribution for each population subgroup by fitting a parametric function to the entire subgroup education distribution, using the binned data. From these parameterized distributions, we then predict a continuous latent rank distribution for each subgroup. We can compare that predicted continuous latent distribution to the uniform distribution to determine the extent of bias arising from the assumption of uniformity.

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. 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.

**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		55.5	30.2	56.0	29.3	
Muslim	46.7	24.5		47.3	27.6	46.4	27.8	
Scheduled Castes	40.8	21.0		39.9	22.9	39.3	24.5	
Scheduled Tribes	39.1	20.2		39.0	23.4	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	28.5	56.5	27.9	
Muslim	45.1	26.7		44.9	27.7	45.4	28.4	
Scheduled Castes	42.9	26.7		42.7	27.4	42.7	28.1	
Scheduled Tribes	35.6	25.0		35.9	25.2	35.1	26.1	

In Table C2, we use the simulated data to examine the distribution of the latent variable *within* the bins where our method assumes uniformity. In particular, 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%.<sup>81</sup> However, the differences are very small, and they do not change much from the 1960–69 to the 1985–89 birth

<sup>81</sup>If the subgroup distributions were all uniform within this bin, then all groups would have a mean rank of 25.

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.

**Table C2**  
Simulated Average Parent Rank Conditional on Rank  $\leq 50$

A. 1960-1969 Birth Cohort

	Group-level Variance		Constant Variance	
	Normal	Lognormal	Normal	Lognormal
Forward / Other	24.2	25.4	27.0	27.1
Muslim	24.7	24.7	24.1	24.7
Scheduled Castes	26.3	24.6	22.5	22.3
Scheduled Tribes	25.5	24.4	22.6	22.3

B. 1985-1989 Birth Cohort

	Group-level Variance		Constant Variance	
	Normal	Lognormal	Normal	Lognormal
Forward / Other	26.8	27.6	27.4	27.4
Muslim	24.2	24.0	24.0	24.2
Scheduled Castes	23.8	23.1	23.3	23.5
Scheduled Tribes	22.4	21.2	21.2	20.7

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 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.

We therefore consider it unlikely that changing parental position within observed rank bins can explain the growing mobility gap between SCs and Muslims. The relative positions of other groups within their bins has similarly not changed enough to substantially bias our group-level estimates.

## C.2 Robustness to Adjusting for Censored Child Ranks

In the main part of the paper, we focus on bounding a function  $Y(x) = E(y|x)$  when  $y$  is observed without error, but  $x$  is observed with interval censoring. In this section, we modify the setup to consider simultaneous interval censoring in the conditioning variable  $x$  and in observed outcomes  $y$ . In the mobility context, this double-censoring setup arises when the  $y$  variable is a child rank (as in Figure 6A and 6B), but not when the  $y$  variable is a child level of education (as in Figure 6C and 6D). As noted in the paper, all of our results are consistent whether we use levels or ranks as outcomes. We nevertheless proceed here to examine the potential bias from ignoring the censoring of child ranks.

One approach to this problem would be to use a problem setup similar to that presented in Section 3, that takes into account the information on child ranks that is lost by binning. We developed a numerical optimization that could bound the CEF by searching over all possible joint distributions of latent  $x$  and  $y$  variables, but with  $n^2$  the number of parameters compared with the single-censoring model, it proved computationally infeasible.

We therefore present two alternate approaches to the problem here. First, we define the latent distributions of  $y$  variables that would generate the maximal and minimal mobility statistic in theory. We can then bound the mobility statistic following our standard method under these best- and worst-case assumptions. This union of these bounds is a very conservative bound on the mobility statistic given censoring in both the  $y$  and  $x$  variables.

Second, we can shed light on the distribution of the true average value of  $y$  in each  $x$  bin if other data is available. This approach is feasible whenever more information is available about children than about their parents, as is the case in our context (and in many others). Specifically, we use data on child wages to predict whether the true latent child rank distribution ( $y$ ) is better represented by the best- or worst-case mobility scenario. The joint wage distribution suggests that the true latent distribution of  $y$  in each bin is very close to the best case distribution, which we used in Section 5.

### C.2.1 Best and Worst Case Mobility Distributions

In this section, we take a sequential approach to the double-censoring problem. We first calculate the set of child CDFs for each level of parent education that would correspond to the highest and lowest possible intergenerational mobility. From these CDFs, we can calculate the average latent rank of children in each parent bin. These may be different from the latent rank implied by assigning each child the midpoint of their bin. With these latent ranks, we can then follow the estimation procedure outlined in Section 3. This gets us a wider set of bounds that takes the censoring of child ranks into account.

To make the example concrete, consider two children who have less than 2 years of education; one is from a rich family and one is from a poor family. Assume that 20% of children in the population have less than 2 years of education. In the body of the paper, we would assume that each of these children has a rank of 10 (i.e. the midpoint of the bottom bin). But it is possible that the child from a poor family has a latent rank of 7 and the child from a rich family has a latent rank of 13. In this case, mobility would be lower than what we have measured in the paper.

Given that child rank is known only to lie in one of  $h$  bins, there are two hypothetical scenarios that describe the best and worst cases of intergenerational mobility. Mobility will be lowest if child outcomes are sorted perfectly according to parent outcomes *within* each child bin, and highest if there is no additional sorting within bins.<sup>82</sup>

Note that the case of perfect sorting within bins fits very poorly with the standard human capital model. In this model, the binned education data reflects a continuous demand for education with a lumpy number of years available for purchase. It is difficult to theorize a distribution where there is a large mass of rich children bunched just below a bin boundary, and no rich children just above that boundary. Given that children of rich and poor parents appear in every bin in the child distribution, the true distribution is likely to be closer to the uniform case than the perfectly ordered case. Note also that we do not consider the case of perfectly reversed sorting, where the children of the least educated parents occupy the highest ranks within each child rank bin, as it would violate

---

<sup>82</sup>Specifically, these scenarios respectively minimize and maximize both the rank-rank gradient and  $\mu_0^x$  for any value of  $x$ . To minimize and maximize  $p_x$ , a different within-bin arrangement is required for every  $x$ . We leave this out for the sake of brevity, and because bounds on  $p_x$  are minimally informative even with uncensored  $y$ .

the stochastic dominance condition (and is implausible).

Appendix Figure C2 shows two set of CDFs that correspond to these two scenarios for the 1960–69 birth cohort. In Panel A, children’s ranks are perfectly sorted according to parent education within bins. Each line shows the CDF of child rank, given some father education. The points on the graph correspond to the observations in the data—the value of each CDF is known at each of these points and thus the CDFs must pass through them. Children below the 27th percentile are in the lowest observed education bin. Within this bin, the CDF for children with the least educated parents is concave, and the CDF for children with the most educated parents is convex—indicating that children from the best off families have the highest ranks within this bin. This pattern is repeated within each child bin. The implausibility of this scenario is reflected by the kinked nature of these CDFs, which are unlikely to appear in the real world. Panel B presents the high mobility scenario, where children’s outcomes are uniformly distributed within child education bins, and are independent of parent education within child bin.

Each of these CDFs can be collapsed to a single mean child rank for each parent bin. From these expected child ranks, we can then use the method from Section 3 to calculate bounds on any mobility statistic. The top two rows of Table C3 shows bounds on  $\mu_0^{50}$  and on the rank-rank gradient for the high and low mobility scenarios. Taking censoring in the child distribution into account widens the bounds on all parameters. The effect is proportionally larger for bottom half mobility, because it was so precisely estimated before—the bounds on  $\mu_0^{50}$  approximately double in width when censoring of son data is taken into account.

These bounds are very conservative, as the worst case scenario is implausible, as noted above. In the next subsection, we draw on additional data on children, which suggests that the best case mobility scenario is close to the true joint latent distribution.

### C.2.2 Estimating the Child Distribution Within Censored Bins

Because we have additional data on children, we can estimate the shape of the child CDF within parent-child education bins using rank data from other outcome variables that are not censored. Under the assumption that latent education rank is correlated with other measures of socioeconomic rank, this exercise sheds light on whether Panel A or Panel B in Figure C2 better describes the true latent distribution.

Figure C3 shows the result of this exercise using wage data from men in the 1960s birth cohort. To generate this figure, we calculate children's ranks first according to education, and then according to wage ranks within each education bin.<sup>83</sup> The solid lines depict this uncensored rank distribution for each father education; the dashed gray lines overlay the estimates from the high mobility scenario in Panel B of Figure C2.

If parent education strongly predicted child wages *within* each child education bin, we would see a graph like Panel A of Figure C2. The data clearly reject this hypothesis. There is some additional curvature in the expected direction in some bins, particularly among the small set of college-educated children, but the distribution of child cumulative distribution functions is strikingly close to the high mobility scenario, where father education has only a small effect on child wage ranks after child education is taken into account. The last row of Table C3 shows mobility estimates using the within-bin parent-child distributions that are predicted by child wages; the mobility estimates are nearly identical to the high mobility scenario. This result suggests that our assumption in Section 5 that the latent child rank is the midpoint of the rank bin for all parent groups is not affecting our estimates very much.

Note that there is no comparable exercise that we can conduct to improve upon the situation when parent ranks are interval censored, because we have no information on parents other than their education, as is common in mobility studies. If we had additional information on parents, we could conduct a similar exercise. The closest we can come to this is by observing the parent-child rank distribution in countries with more granular parent ranks, as we did in Section 3. The results in that section suggest that interval censoring of parent ranks does indeed mask important features of the mobility distribution.

Note finally that the potential bias from assuming uniformity within child rank bins is increasing in the size of the rank bins. Because children are more educated than parents in every cohort, this bias is smaller for children than it would be for parents. It is also smaller for the younger cohorts of children born in the 1980s than it is for the example we used here.

---

<sup>83</sup>We limit the sample to the 50% of men who report wages. Results are similar if we use household income, which is available for all men. Household income has few missing observations, but in the many households where fathers are coresident with their sons, it is impossible to isolate the son's contribution to household income from the father's, which biases mobility estimates downward.

**Figure C2**  
 Best- and Worst-Case Son CDFs  
 by Father Education (1960-69 Birth Cohort)

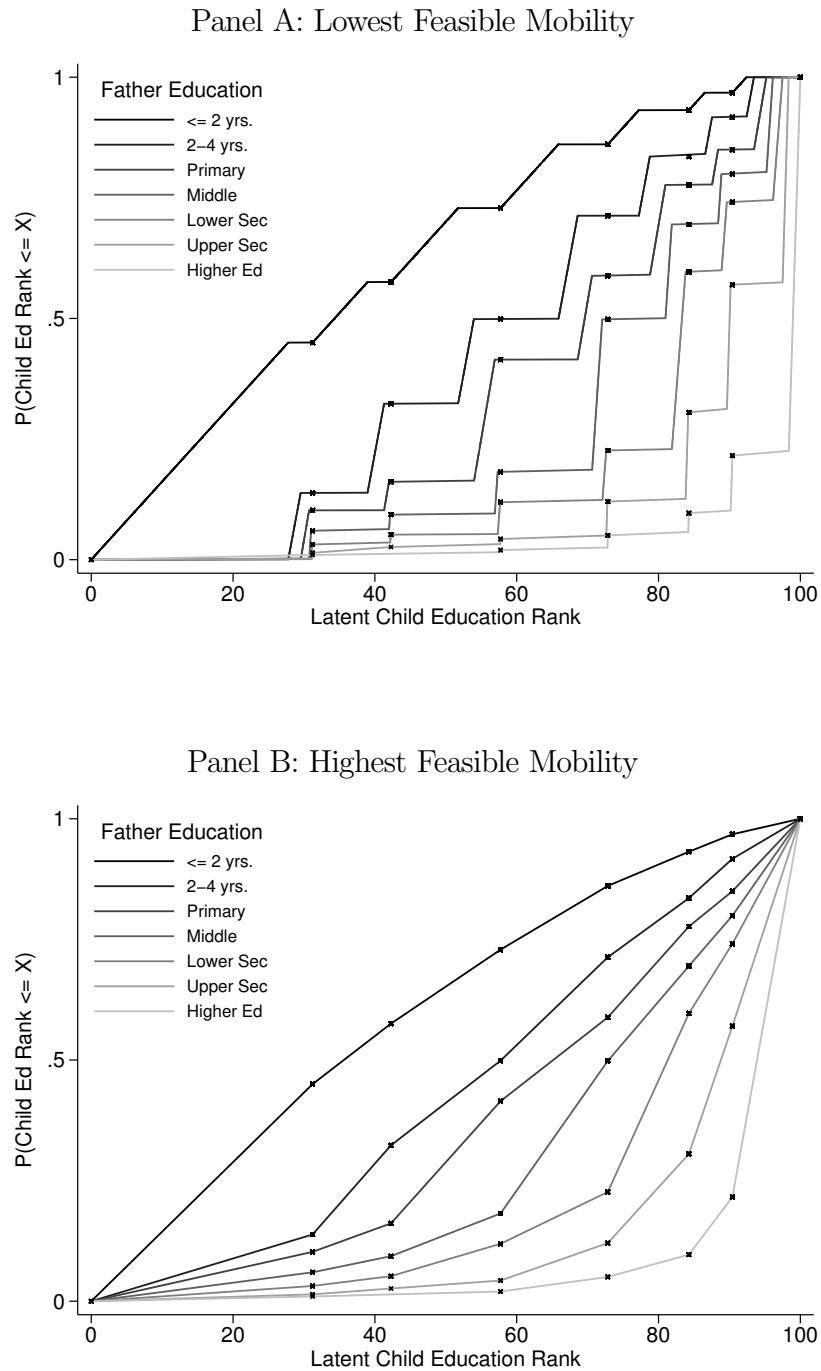


Figure C2 shows a set child CDFs conditional on each level of father education that correspond to the best and worst case scenarios for intergenerational mobility. The lines index father types. Each point on a line shows the probability that a child of a given father type obtains an education rank less than or equal to the value on the X axis in the national education distribution. The large markers show the points observed in the data.

**Figure C3**  
**Son Outcome Rank CDF**  
**by Father Education (1960-69 Birth Cohort)**  
**Joint Education/Wage Estimates**

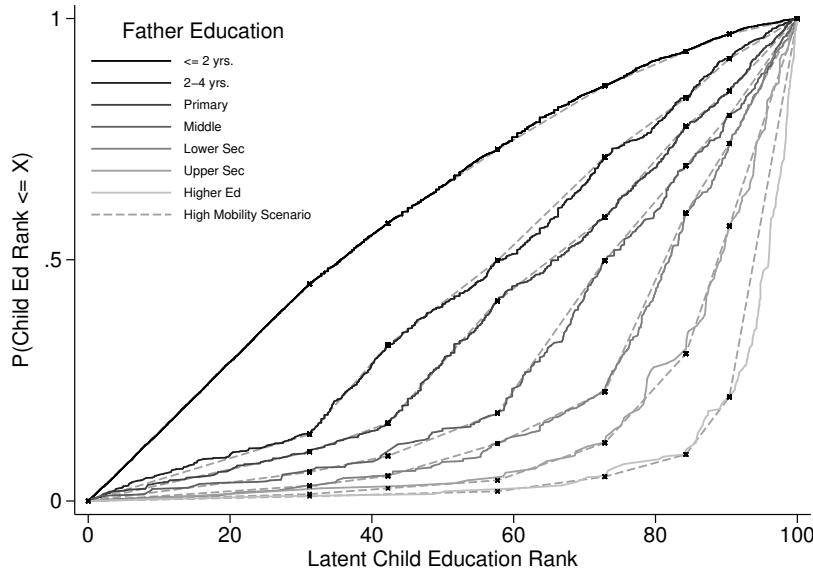


Figure C3 plots a son rank CDF separately for each father education level, for sons born in the 1960s in India. Sons are ranked first in terms of education, and then in terms of wages. Sons not reporting wages are dropped. For each father type, the graph shows a child's probability of attaining less than or equal to the rank given on the X axis.

**Table C3**  
**Mobility Estimates under Double-Censored CEF**

	Upward Interval Mobility ( $\mu_0^{50}$ )	Rank-Rank Gradient ( $\beta$ )
Low mobility scenario	[32.33, 35.90]	[0.55, 0.80]
High mobility scenario	[35.86, 38.80]	[0.45, 0.67]
Wage imputation scenario	[35.79, 38.70]	[0.46, 0.67]

Table C3 presents bounds on  $\mu_0^{50}$  and the rank-rank gradient  $\beta$  under three different sets of assumptions about child rank distribution within child rank bins. The low mobility scenario assumes children are ranked by parent education within child bins. The high mobility scenario assumes parent rank does not affect child rank after conditioning on child education bin. The wage imputation predicts the within-bin child rank distribution using child wage ranks and parent education.

## 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 A3 shows that the discrepancies between measures are not correlated with household characteristics.

### D.2 SECC

The 2011–12 Socioeconomic and Caste Census (SECC), an administrative socioeconomic database covering all individuals in the country that was collected to determine eligibility for various government programs.

The data underlying SECC were posted to the internet by the government, with each village and urban neighborhood represented by hundreds of pages in PDF format. Each town/village was posted for only ninety days. Over a period of two years, we scraped over two million files, parsed the embedded data into text, and translated the text from twelve different Indian languages into English. This process is described in more detail in Asher and Novosad (2020).

At the end of this process, we have individual data from approximately 450,000 villages and 2000 towns. This

covers 90% of villages and 25% of towns in India; the town data are less complete because many towns had only partial data posted on the SECC web site, or were not posted at all. The set of towns in the data cover all major states of the country and have a very similar population distribution to the full distribution of towns.

We use SECC to create linked data on father and son education. As noted in the body of the paper, we focus on ages 20–23, and do not look at girls, who are much less likely to be coresident with their parents at that age. We also do not create mother-child links, because of the substantial censoring of mothers' education ranks, described in Section 5.1.

When SECC records the relationship between individuals in a household, we create parent-child links following the same algorithm as used in the IHDS, as described above. For records where SECC did not provide family relationships, we impute the identity of the father based on the household structure. As noted, our child sample consists of men aged 20–23. We assume that a coresident man aged between 15 and 50 years older is the father. In most cases, there is only one such individual, and the father identity is directly assigned. We exclude observations where there is more than one potential father by this definition; results are virtually identical if we assume the father's education is the mean of the candidate fathers.

To validate this algorithm, we replicated the algorithm in the IHDS, where we observe the identity of the individual's father. In 5% of cases, the algorithm identified a father where there was none. In an additional 5% of cases, the algorithm did not identify a father (due to ambiguity) when there was one. In the 90% of cases where the algorithm identified a father and there was a father present, the correlation between the education measures was 99.7%. It is therefore unlikely that significant bias arises from the subset of SECC observations where we do not observe the relationship field.

### D.3 Other data sources

This section describes several other data sources which we drew upon to calculate the correlates of upward mobility in Section 5.3.

The 2001 and 2011 Population Censuses provides basic demographic information for villages and towns, include the Scheduled Caste share. The village and town directories in the same census describe local public goods, including number of primary and high schools, and access to paved roads and electricity. Consumption and consumption inequality were calculated from the SECC using small area estimates following Elbers et al. (2003); the specific process we used is described in more detail in Asher and Novosad (2020). Average years of education in a location were calculated from the education variable in the SECC. Manufacturing jobs per capita were calculated by dividing the number of non-farm jobs recorded in the 2013 Economic Census by the 2011 census population. SC/ST segregation is the dissimilarity measure of segregation. We calculated this at the subdistrict level for rural areas, describing dissimilarity across villages. At the town level, it describes dissimilarity across enumeration blocks, which are units of about 200 households.

#### D.4 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).