Intergenerational Mobility in India: Estimates from New Methods and Administrative Data*

Sam Asher†

Paul Novosad‡

Charlie Rafkin§

September 2018
First version: February 2016

Abstract

Estimating intergenerational mobility in developing countries is difficult because matched parent-child income records are rarely available and education is measured very coarsely. In particular, there are no established methods for comparing educational mobility for subsamples of the population when the education distribution is changing over time. We resolve these problems using new methods in partial identification and new administrative data, and study intergenerational mobility across groups and across space in India. Intergenerational mobility for the population as a whole has remained constant since liberalization, but cross-group changes have been substantial. Rising mobility among historically marginalized Scheduled Castes is almost exactly offset by declining intergenerational mobility among Muslims, a comparably sized group that has few constitutional protections. These findings contest the conventional wisdom that marginalized groups in India have been catching up on average. We also explore heterogeneity across space, generating the first high-resolution geographic measures of intergenerational mobility across India, with results across 5600 rural subdistricts and 2300 cities and towns. On average, children are most successful at exiting the bottom of the distribution in places that are southern, urban, or have higher average education levels.

*We are thankful for useful discussions with Emily Blanchard, Raj Chetty, Eric Edmonds, Shahe Emran, Francisco Ferreira, Nate Hilger, David Laibson, Ethan Ligon, Erzo Luttmer, Nina Pavcnik, Bruce Sacerdote, Na'ama Shenhav, Forhad Shilpi, Gary Solon, Doug Staiger and Chris Snyder. Annaka Balch, Toby Lunt, Ryu Matsuura, and Taewan Roh provided excellent research assistance. This project received financial support from the IZA GLM-LIC program and from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374.

†World Bank, sasher@worldbank.org

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

§MIT, crafkin@mit.edu

1 Introduction

The intergenerational transmission of economic status, a proxy for equality of opportunity, has implications for inequality, allocative efficiency and subjective welfare (Solon, 1999; Black and Devereux, 2011; Chetty et al., 2014; Chetty et al., 2017). Because intergenerational mobility (IM) captures changes in status across generations, it is a useful measure for describing changes in access to opportunity over the long run. Studies of intergenerational mobility (IM) in developing countries have had less influence than in rich countries because of methodological challenges and an absence of data comparable to the administrative tax records that are the basis of studies in richer countries. In this paper, we develop a set of measures of intergenerational educational mobility that perform well under the data constraints often faced in developing countries. Using new administrative data, we apply them to the analysis of IM in India across time, across major population subgroups, and across space.

From before liberalization to the present, we find that intergenerational educational mobility for India as a whole has remained constant on average, but with considerable cross-group heterogeneity. While our results reinforce earlier findings of rising opportunity among Scheduled Castes and Scheduled Tribes that have been noted by other authors, we find that these changes are almost exactly offset in the population as a whole by declining upward mobility among Muslims.

Estimating IM in developing countries presents several challenges. Datasets linking parent and child wages or household incomes are rare, and administrative tax databases cover a small and unrepresentative share of the population. Even when linked parent and child wages or income are recorded, they generate biased measures of intergenerational mobility because of measurement error, non-wage income, and the inability to distinguish parent and child income when both reside together.

Researchers have therefore focused on intergenerational educational mobility: the dependence of children's educational outcomes on their parents' education.[1] But existing ap-

[1]Recent studies of intergenerational mobility focusing on education and occupation include Black et al.

proaches to the measurement of educational mobility have several limitations. There is no established method for comparing educational mobility across subgroups. The most common educational mobility estimator, the correlation between parents' and children's education, is not meaningful for subgroups because it measures children's progress only against other members of their own group (Hertz, 2005). Rank-based estimates are comparable across time, but the coarse measurement of education makes many of them impossible to point estimate (Asher et al., 2018). For instance, Chetty et al. (2014) focus on absolute upward mobility, defined as the expected child rank given a parent at the 25th percentile of the parent rank distribution. How can this be calculated for the 1950s birth cohort in India, for whom the bottom 60% of parents report bottom-coded levels of education? Point estimating this measure requires strong and questionable implicit assumptions. Similar challenges bedevil the construction of quantile transition matrices from education data.

We resolve this problem with a partial identification approach, drawing on methods from Asher et al. (2018), where we showed that rank-based mobility measures can at best be bounded. Our analysis focuses on two measures: (i) the expected outcome of a child born into the bottom half of the parent outcome distribution (upward interval mobility, henceforth referred to as upward mobility); and (ii) the expected outcome of a child born into the *top* half of the parent distribution (downward interval mobility).

These measures have several desirable properties. First, they have similar policy relevance to absolute upward and absolute downward mobility, suggested by Chetty et al. (2014), but our measures can be bounded tightly given coarsely measured education data, while the absolute mobility measures cannot.[2] Second, they are valid for subsets of the population. Third, these measures are comparable across time and space, as well as across countries. Finally, the partial identification approach explicitly takes into account the coarse measure-

(2005), Long and Ferrie (2013), Güell et al. (2013), and Wantchekon et al. (2015).

[2]Absolute upward mobility describes the expection of a child outcome, conditional on that child being born to a parent at the 25th percentile. Upward interval mobility describes the expectation of a child outcome, conditional on that child being born to a parent between percentiles 0 and 50. If the conditional expectation function is linear in parent rank, the two measures are identical.

ment of education. The bounds on our measures widen as education is measured more coarsely, reflecting the loss of information on granular ranks. Conventional point estimation methods ignore this source of error.

We use two data sources to study educational mobility in India. To study changes over time, we use the India Human Development Survey (IHDS), which stands out among India's major household surveys in obtaining information about the education of the fathers of male respondents. To study mobility across space, we use the Socioeconomic and Caste Census (SECC), an administrative census undertaken in 2012, which documents education for all people in India at a fine geographic level. In the SECC, we can only match parent and child education when both are in the same household. We show using the IHDS that the bias from limiting the sample to coresident households becomes substantial above age 24. We therefore limit the SECC sample to children ages 20–23, an extreme sample limitation possible only with comprehensive administrative data. As we discuss in Section 3, data availability requires us to focus on the relationship between fathers and sons; transmission of status to daughters or from mothers is an important topic for future work.

We document trends in educational mobility from the 1950s to the 1980s birth cohorts. In the pooled population, we find that both upward and downward mobility have remained constant for the past 30 years, in spite of dramatic educational gains. An Indian child born in the bottom half of the parent education distribution can expect to obtain the 37th percentile. A similar child in the U.S., which has low intergenerational mobility by OECD standards, on average attains the 39th education percentile.[3]

We next examine mobility across time for four of India's major population subgroups: Scheduled Castes (SCs), Scheduled Tribes (STs), Muslims, and Forward Castes/Others.[4]

[3]In a society where children's outcomes were totally independent of parents (i.e. total mobility), a child born in the bottom half of the distribution would obtain 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 would attain the 25th percentile.

[4]We include non-Muslim OBCs 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.

Group identity is a strong predictor of both levels and changes in intergenerational mobility. Consistent with prior work (Hnatkovska et al., 2012; Emran and Shilpi, 2015), we find that India's constitutionally protected marginalized groups, the Scheduled Castes and Tribes, have lower intergenerational mobility than upper caste groups, and have closed respectively 50% and 30% of the mobility gap to Forwards/Others. Our most novel and striking finding is that upward mobility has fallen substantially among Muslims in the last twenty years. The expected educational rank of a Muslim child born in the bottom half of the parent distribution has fallen from between 31 and 34 to a dismal 29. Muslims now have considerably worse upward mobility today than both Scheduled Castes (37.4–37.8) and Scheduled Tribes (32.5–32.7). The comparable figure for African Americans is 34. 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.[5]

Absolute outcomes for all social groups have been improving systematically due to India's substantial economic growth over the study period. We focus on rank outcomes because they allow us to study the extent to which marginalized groups have closed the upward mobility gap with Forward / Other groups, holding the level shift in outcomes constant. We find a similar divergence across groups when we focus on absolute outcomes like a child's probability of attaining a high school education, an essential prerequisite for a white collar job in government or the private sector. In fact, access to high school and college for Muslims from bottom half families has stagnated for the last fifteen years. Turning to downward mobility, we find that the mobility gap for all three marginalized groups has shrunk substantially; Muslims born in the upper half of the distribution have lost some ground relative to SCs and STs, but much less starkly than was the case with upward mobility.

The extent to which cross-group differences in mobility are driven by location varies substantially by group. Among STs, district of residence explains 59% of the upward mobility gap with the Forward/Other reference group. Place matters considerably less for SCs and

[5]We find similar results if we exclude Sikhs, Jains and Christians from the "others" category. Collectively, these groups account for less than 5% of the population.

Muslims: district of residence explains 14% of upward mobility for SCs and only 9% for Muslims. These results likely reflect the fact that many members of Scheduled Tribes live in remote areas and continue to practice traditional economic activities, while Muslims and Scheduled Castes are better integrated with the modern economy. Children in India's only Muslim-majority state (Jammu and Kashmir) in fact have higher mobility than the Indian average.

Finally, we examine the correlates of intergenerational mobility at a fine geographic level: rural subdistricts (n=5600) and cities and towns (n=2000). Urban mobility is consistently higher than rural mobility: children in cities can expect a 4-5 percentile higher rank in the outcome distribution than children in villages, at every parent rank. This advantage varies by subgroup—for instance, we find that the Muslim disadvantage in cites is actually three points larger than the Muslim disadvantage in villages. In rural areas, average income and education are the strongest predictors of upward mobility. Public goods, manufacturing jobs, the share of Scheduled Caste individuals, low segregation and low land inequality are also statistically significant predictors of high upward mobility. In cities, average education, size and low segregation are strongly predictive of upward mobility, but levels of consumption are not. All of our mobility estimates are robust to different data construction methods, and we take care to show that survivorship bias, migration or bias in estimates from coresident parent-child households are unlikely to substantially affect our results.

We make three main contributions. First, we develop methods for the measurement of intergenerational mobility that address the data limitations often faced in developing country contexts. To our knowledge, these are the first estimates of intergenerational educational mobility that are valid for population subgroups and are comparable across time and space. Second, we show that Muslims are losing substantial ground in intergenerational mobility, and currently have lower mobility than either Scheduled Castes or Scheduled Tribes. Given the large amount of work studying access to opportunity among Scheduled Castes, our results highlight the importance of studying economic outcomes of Muslims in India, especially from poor families. Despite a population share comparable to Scheduled Castes, this group is overlooked by much of the economic literature on marginalized groups in India.[6] Third, we provide the first geographically precise estimates of intergenerational mobility in India. Nearly all of the geographic variation in intergenerational mobility is below the state level, though there are substantial differences across states. This is also the first work to measure or document the importance of residential segregation at the neighborhood level 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 with constitutional protections, reservations in politics and education, and who have been targeted by many development policies. Further, none of those gains have come at the expense of higher caste groups, notwithstanding political mobilization against affirmative action by higher caste groups in recent years. For other groups, there is little evidence that economic liberalization has substantially increased opportunities for those in the lower half of the rank distribution to attain higher relative social position, and for Muslims these opportunities have substantially deteriorated.

These patterns have to our knowledge not been identified because earlier studies have either (i) focused on 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 describe 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). Studies on affirmative action in India have identified improvements in SC/ST access to higher education but have not examined impacts on Muslims (Frisancho Robles and Krishna, 2016; Bagde et al., 2016); our findings point to the importance of studying the effects of such policies on a wider set of marginalized groups.[7]

[6] Notable exceptions 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) summarize some recent research on Muslims on India, none of which addresses intergenerational mobility.

[7] In an analogous finding, Bertrand et al. (2010) find that when Indian colleges intentionally select lower caste students, they end up admitting fewer women.

Our paper proceeds as follows. Section 2 describes current approaches to estimating intergenerational mobility, with a focus on estimates using coarse education data. Section 3 describes the Indian data. Section 4 describes our measures of upward and downward mobility. Section 5 presents results on national mobility trends, cross-group mobility trends, and the geographic distribution of intergenerational mobility. Section 6 concludes.

2 Background: Theory and Empirics of Mobility Measurement

When intergenerational mobility is low, the social status of individuals is highly dependent on the social status of their parents (Solon, 1999). In more mobile societies, individuals are less constrained by their circumstances at birth. If ability is not perfectly correlated with birth circumstances, then there may be substantial efficiency losses resulting from the lack of opportunity for those who grow up poor. There is also a widespread normative belief that all individuals should have equal opportunity to develop their abilities. There is a growing literature on the variation in intergenerational mobility across countries, across groups within countries, and across time.[8]

The first generation of intergenerational mobility studies described matched parent-child outcome distributions with a single linear parameter, or gradient, such as the correlation coefficient between children's earnings and parents' earnings (Solon, 1999; Black and Devereux, 2011). Gradient measures are easy to calculate and interpret and they form the basis of studies in dozens of countries. The two main limitations of gradient measures are that they (i) do not distinguish between changes in opportunity at the top and the bottom of the distribution;[9] and (ii) they are not well-suited for between-group comparisons, because the subgroup gradient measures children's outcomes against better off members of their own group. A subgroup can therefore have a lower gradient (suggesting more mobility) in spite

[8]See Hertz et al. (2008) for cross-country comparisons, and Solon (1999), Corak (2013), Black and Devereux (2011), and Roemer (2016) for review papers.

[9]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); the linearity of the income rank CEF described by Chetty et al. (2014) is an exception. The Indian education rank CEF that we study below is also non-linear.

of having worse outcomes at every point in the parent distribution.

Recent studies have instead drawn on rich administrative data such as tax records to analyze the entire joint parent-child income *rank* distribution, and in particular, the conditional expectation of child rank given parent rank (Boserup et al., 2014; Chetty et al., 2014; Hilger, 2016; Bratberg et al., 2015). Several useful measures can be drawn from this CEF. The slope of the best linear approximator to the CEF, also called the rank-rank gradient, is analogous to the parent-child income correlation. Absolute mobility at percentile i, which we denote p_i, describes the value of the CEF at percentile i of the parent distribution (Chetty et al., 2014).[10] Absolute upward mobility is p_{25}, which describes the expected rank of a child born to the median parent in the bottom half of the parent rank distribution. Absolute downward mobility, or p_{75}, describes the persistence of privileged ranks across generations.

These measures have several strengths. They are valid for cross-group comparisons: p_{25} lets us compare the outcomes of children in different population subsets who are born to parents with similar incomes. The rank-based measures are also valid for cross-country analysis, because they are invariant to changes in the parent or child income distributions. However, as we describe below, there is no established method for calculating these rank based measures from education data.

2.1 Educational Mobility and Income Mobility

Education data have been widely used for mobility studies for three reasons.[11] First, representative matched parent-child income data are frequently unavailable. In developing countries, there are virtually no administrative income records that cover a large share or representative share of the population.

[10] Chetty et al. (2014) 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. (2014).

[11] Some recent studies examining educational mobility include Solon (1999), Black et al. (2005) and Restuccia and Urrutia (2004). More are summarized in Black and Devereux (2011).

Second, even when matched parent and child income data are available, measurement error problems are substantial. Transitory incomes are noisy estimates of lifetime income, subsistence consumption is difficult to measure, and many individuals report zero income; these problems are exacerbated among the rural poor.[12] Measurement error in income is consequential because it biases mobility estimates upward.

Third, income measures are subject to life cycle bias if parent and child income are not measured at the same age; individuals with high permanent income may spend more time in school and have lower incomes than their peers when young.

For these reasons, researchers have often turned to intergenerational educational mobility. Educational mobility, or the dependence of child education on parent education, is highly correlated with intergenerational income mobility (Solon, 1999). In developing countries in particular, educational attainment may be a better proxy for lifetime income than a single observation of transitory income. Focusing on education is also mostly free of life cycle bias, because education levels rarely change later in life. Third, matched parent-child education data are more widely available than matched parent-child income data.

Estimates of educational mobility do have one important drawback: outcomes are typically reported in a small number of categories. Even when years of education are specifically measured, they are bunched at school completion levels. Individuals who have not completed primary school, who exceed half of the population in many poor countries, are frequently bottom-coded. The parent rank distribution is thus observed coarsely, making rank-based mobility estimates challenging to calculate.

To make the problem concrete, consider the parent-child education transition matrix for children born in the 1950s in India, shown in Appendix Table A1. 60% of parents are coded in the bottom bin. How does one measure p_{25} in this setting? Existing work has either

[12]Intergenerational income mobility studies either focus on individual wages or total household earnings, both of which are problematic. Given that transitioning out of agricultural work and into wage work is a central predictor of consumption, restricting the sample to wage earners has obvious deficiencies. But for the set of households where parents and children are coresident, the share of total household earnings attributable to the parent is not known.

assumed that the CEF of child rank is linear across all parent ranks, or that the CEF is constant within each parent bin. As we will show below, the first assumption is rejected by the Indian data. The second assumption is unlikely to be true: the child of a parent at the 59th percentile of the rank distribution is likely to have a better outcome than the child of a parent in the bottom 1%.[13] This assumption also generates measures of mobility that are not invariant to the percentile boundaries of education bins. Changes in the granularity of education data would lead to changes in measures of mobility, which is clearly undesirable.[14]

Similar challenges arise when comparing transition matrices over time. Given that the size of education bins are changing over time, there are few cells of the matrix that be directly compared in different years. Using a quantile transition matrix would resolve the problem, but calculating the quantile transition matrix from coarsely binned rank data poses the same challenge as calculating p_{25}.

We address these challenges directly. In Section 4, we describe how methods from the interval data literature allow us to calculate bounds on rank-based mobility measures under a minimal and reasonable set of structural assumptions.

2.2 Context: Educational Mobility in India

While intergenerational mobility is of interest around the world, India's caste system and high levels of inequality make it a particularly important setting for such work. 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. 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 and politi-

[13]We elaborate on the exact definition of a parent at a granular percentile in this context in Section 4.

[14]For an extreme example, consider the case of Ethiopia, where 85% of parents have bottom-coded levels of education in older cohorts. Assuming equal outcomes within parents bins would suggest a quartile transition matrix where children born in the 1st, 2nd and 3rd quartile all have equal opportunities of success. Clearly one should be cautious when comparing this transition matrix to one calculated from fully supported rank data.

cal institutions. If effective, these policies would substantially improve the intergenerational mobility of targeted groups.

India's Muslims constitute a similar population share as the Scheduled Castes and Scheduled Tribes (14% for Muslims vs. 16.6% for SCs and 14% for STs). Muslims have also experienced discrimination and have worse socioeconomic outcomes than the general population (Sachar Committee Report, 2006). While Muslim disadvantage has been widely noted, including by the well-publicized 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. In fact, a large scale social movement (the RSS) and several major political parties and have successfully rallied around pro-Hindu platforms and policies which have been argued to discriminate against Muslim religious and cultural practices. The ruling federal coalition at the time of writing (the BJP) arose out of the Hindu nationalist movement. Muslims have also frequently been targeted by mob violence.

The last 30 years have seen tremendous growth in market opportunities in India as well as in the availability and level of education. While some have argued that economic growth is making old social and economic divisions less important to the economic opportunities of the young, caste remains an important predictor of economic opportunity (Munshi and Rosenzweig, 2006; Ito, 2009; Hnatkovska et al., 2013; Mohammed, 2016). Understanding how mobility has changed for these population groups is thus an essential component of understanding secular trends in intergenerational mobility in India. Further, whether economic progress can overcome traditional hierarchies of social class and religion is a central question for both India and the broader world.

3 Data

To estimate intergenerational educational mobility in India, we draw on two databases that report matched parent-child educational attainment. The India Human Development Survey (IHDS) is a nationally representative survey of 41,554 households, with rounds in 2004-05 and 2011-12. Crucially, the IHDS solicits information on the education of fathers of household heads, even if the fathers are not resident, allowing us to match father to son education for almost all men of all ages. The IHDS also distinguishes between Scheduled Tribes, Scheduled Castes, Muslims, Other Backward Castes (OBCs) and forward castes. We classify SC/ST Muslims, who make up less than 2% of SC/STs, as Muslims. About half of Muslims are 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 very complex), and is 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; we describe this group as "Forward/Other."

Our main dataset is the 2012 round of the IHDS. Drawing on the education of older birth cohorts allows us to estimate a time series in mobility. To allay concerns that differential mortality across more or less educated fathers and sons might bias our estimates, we replicate our analysis on the *same* birth cohorts using the IHDS 2005. By estimating mobility on the same cohort at two separate time periods, we identify a small survivorship bias for the 1950-59 birth cohort (reflecting attrition of high mobility dynasties), but zero bias for the cohorts from the 1960s forward. Our results of interest largely describe trends from the 1960s forward, so survivorship bias among the oldest cohorts does not influence any of our conclusions. Data are pooled into decadal cohorts.

The strengths of the IHDS for our study are its documentation of religion and its recording of the education of non-coresident parents. However, the IHDS sample has too small a sample to study geographic variation in much detail.

To study geographic variation in mobility, we draw on the 2011-12 Socioeconomic and Caste Census (SECC), an administrative socioeconomic database that was collected to determine eligibility for various government programs. The data was posted on the internet by the government, with each village and urban neighborhood represented by hundreds of pages in PDF format. 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. The data include age, gender, education, an indicator for Scheduled Tribe or Scheduled Caste status, and relationship with the household head. Assets and income are reported at the household rather than the individual level, and thus cannot be used to estimate mobility.[15] The SECC provides the education level of every parent and child residing in the same household. Sons who can be matched to fathers through coresidence represent about 85% of 20-year-olds and 7% of 50-year-olds. Education is reported in seven categories.[16] For some results, we work with a 1% sample of the SECC, stratified across India's 640 districts.

The size of the SECC makes it possible to calculate mobility with high geographic precision, but the limitation is that we do not observe parent-child links for children who no longer live with their parents. We therefore focus on children aged 20 to 23, a set of ages where education is likely to be complete but coresidence is high enough that the bias from excluding non-coresident pairs is low. We selected these ages by examining the size of the coresidence bias in the IHDS. In Appendix Figure A1, we plot the difference in upward and downward mobility between the full sample and sample of coresident father-son pairs only. For the pooled 20–23 age group, we can rule out a bias of more than two percentage points in the child rank. For higher ages, the bias rapidly exceeds five percentage points—more than the mobility difference between the United States and Finland. Earlier Indian mobility estimates which were based on coresident children as old as 40 should thus be treated with caution. The sample thus consists of 31 million young men and their fathers. The age restriction in the SECC and the absence of data on religion limits our fine geographic analysis to the modern cross-section. We further verified that IHDS and SECC produce similar point estimates for the coresident father-son pairs that are observed in both datasets.

IHDS records completed years of education. To make the two data sources consistent,

[15] Additional details of the SECC and the scraping process are described in Asher and Novosad (2018).

[16] 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.

we recode the SECC into years of education, based on prevailing schooling boundaries, and we 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.[17] The loss in precision by downcoding the IHDS is minimal, because most students exit school at the end of a completed schooling level.

The oldest cohort of sons that we follow was born in the 1950s and would have finished high school well before the beginning of the liberalization era. 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 were excluded.

The transmission of status from mothers to children or from fathers to daughters are both important topics of study, but is more difficult to measure due to data limitations. Young women leave households at an earlier age than boys — there is thus no age at which most women have completed their education and at which most women continue to reside with their parents. This makes it difficult to measure mobility for young women in the SECC. In the IHDS, close to 75% of mothers report a bottom-coded education; as we discuss in the next section, this implies that our primary measure of upward mobility will have extremely wide bounds. Intergenerational mobility among women is thus beyond the scope of this study, but is an important topic for future work.

4 Methods: Rank-Based Mobility Estimates with Education Data

This section summarizes and expands upon methods described in Asher et al. (2018). We are interested in generating estimates of intergenerational mobility that can be derived from the conditional expectation of child rank given parent rank. We define this CEF as $E(Y|X = i)$, where Y is the child rank and X is the parent rank. Scalar mobility measures are functions of the CEF. For example, the rank-rank gradient is the slope of the best linear approximator

[17]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.

to the CEF, and absolute upward mobility is $E(Y|X = 25)$.

These measures are challenging to estimate with education data, because education ranks are observed only in coarse bins. For the 1950s birth cohort, for example, we may observe that a parent's rank is somewhere in the bottom 60%, but we have no additional information about that parent.[18] How should one calculate a mobility measure such as p_{25} when we cannot identify any parent at the 25th percentile?

We treat this as an interval data problem and take a very general approach requiring minimal assumptions. First, we assume that there is a latent continuous education rank distribution, which we observe with interval censoring. This assumption arises directly from a standard human capital model where education has convex costs and individual-specific differences in costs and benefits are constant shifters in the utility function. The latent rank 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, while individuals who would not advance further even if the net benefit changed a great deal will have the lowest ranks in the bin. This is a relatively weak assumption; the alternative approach of assuming that all individuals in the bottom bin were equally distant from completing primary school strikes us as implausible.

Second, we assume that the expectation of a child's education rank is weakly increasing in the latent parent education rank. This is also a very mild assumption, given that average socioeconomic outcomes of children are strongly monotonic in parent socioeconomic outcomes across many socioeconomic measures and countries (Dardanoni et al., 2012). Average child education is also monotonic in parent education across all birth cohorts that we study in India. Note that this set of assumptions is considerably weaker than the implicit assumptions made in other studies of educational mobility, which either assume that the CEF is totally

[18] Data like these are common in studies of intergenerational educational mobility, because information on parents often comes from a limited set of questions asked of children.

linear or is piecewise constant.

Under these assumptions, the child education rank CEF can be partially identified at any point, as can scalar functions of the CEF (Manski and Tamer, 2002; Asher et al., 2018).[19] A partial identification approach allows us to take seriously the fact that granular ranks are not precisely observed; bounds will be wider when ranks are censored more substantially. This contrasts with conventional point estimation approaches, where increased censoring may cause bias that is not reflected in standard errors.

The process is depicted in Figure 1, with the 1950s birth cohort in Panel A and the 1980s birth cohort in Panel B. The solid vertical lines indicate boundaries of observed parent rank bins and the points indicate the expected child rank in each parent bin. The functions plot the upper and lower bound on the child CEF at each parent rank.

The figure shows that bounds on the CEF are tight at high parent ranks, where the parent rank is observed with high granularity (because fewer parents have obtained high levels of education). In the bottom half of the rank distribution, the bounds are very wide. The bounds on p_{25} ($[28.4, 44.9]$, indicated by the dashed vertical line) are particularly wide given the interval censoring in the 1950s birth cohort. It is difficult to say anything meaningful about this measure of upward mobility for this birth cohort. In Asher et al. (2018), we show that bounds on the rank-rank gradient are also too wide to be informative, and propose a new set of mobility measures that can be tightly bounded even when interval censoring is severe, which we call interval mobility.

Interval mobility, or μ_a^b, is the expected outcome of a child whose parents are between percentiles a and b in the parent education rank distribution, $E(Y|X \in [a, b])$. Asher et al. (2018) show that interval mobility can often be tightly bounded even when bounds on other intergenerational mobility measures such as the rank-rank correlation or absolute mobility are too wide to be useful. Interval mobility can be bounded particularly tightly because it is known for some intervals. For instance, when 60% of parents are in the bottom bin, then

[19] Stata code to generate the bounds is provided at `https://github.com/paulnov/anr-bounds`.

the mean child rank among these parents is the sample analog to $E(y|x \in [0, 60])$, or μ_0^{60}. Interval mobility estimates with similar boundaries will be tightly bounded by virtue of the continuity of the CEF and uniformity of the rank distribution. In contrast, absolute mobility at percentile i cannot be point identified for any value of i. All of these measures capture slightly different characteristics of the rank-rank CEF, and could be of policy interest. However, interval mobility is the only one of these measures that can be measured informatively given the type of education data typically available in developing countries.

We focus on two specific measures. First, we consider *upward interval mobility* μ_0^{50}, which is the expected outcome of a child whose parents are in the bottom half of the parent education distribution. Second, we consider *downward interval mobility* μ_{50}^{100}, the expected outcome of a child whose parents are in the *top* half of the parent education distribution. These measures have very similar policy relevant to absolute upward mobility and absolute downward mobility, and are the same if the CEF is linear. The advantage of interval mobility measures is that they can be tightly bounded even when rank data have a high degree of censoring. For instance, under the same assumptions as those used in Figure 1, we can bound upward interval mobility between $[34.9, 37.1]$ for the 1950s birth cohort. For parsimony, we henceforth refer to the interval measures as upward and downward mobility. Note that a high mobility society will have a high value of upward mobility and a low value of downward mobility, reflecting lower persistence of both low and high socioeconomic ranks across generations.

Note that child ranks are also interval censored in our context. We show in Asher et al. (2018) that interval censoring in the child distribution is unlikely to cause substantial bias for two reasons. First, because of increasing education over time, child rank bins are more evenly distributed than parent rank bins, which decreases potential bias from censoring. Second, when we impute within-bin ranks using additional data on children's wages (not available for parents), we find virtually the same mobility estimates. This suggests that using the midpoint of a child's rank bin is capturing most of the meaningful variation in child ranks. Finally, note that we can calculate all of our cross-group and cross-sectional estimates using an uncensored measure of child outcomes, such as an indicator function for high school completion; we find similar results when we do so.

5 Results: Intergenerational Mobility in India

5.1 National Estimates and Group Differences

Figure 2 presents our main measures of mobility, upward and downward interval mobility, over time. Upward mobility is $E(Y|X \in [0,50])$ and downward mobility is $E(Y|X \in [50,100])$, where Y is the child education rank and X is the parent education rank. Neither measure has changed substantially over time. Upward mobility has gone from an interval of [34.9, 37.1] to 36.8, while downward mobility has gone from [61, 63] to 62.5.[20] Sons born to fathers in the either half of the education distribution in 1950 have the same expected outcomes as those in 1980. This puts upward mobility in India approximately two percentage points behind the U.S. ($\mu_0^{50} \in [38.9, 39.3]$), which is considered to have low mobility by international standards.[21]

The gap between upward and downward mobility describes the expected penalty faced by children from bottom half families relative to children from top half families. This gap is about 26 percentage points, and has been persistent for birth cohorts from the 1950s to the present.[22] The Y axis in these graphs is the expected education rank of the son, which is a relative outcome. The results indicate that the likelihood of children moving up in the education distribution relative to their parents is persistent and low.

In Figure 3, we again present measures of upward and downward mobility (μ_0^{50} and μ_{50}^{100}), but with an absolute outcome on the Y axis: primary school attainment in Panel A and high school attainment in Panel B. The light blue line shows the probability that a son attains primary school or higher, conditional on having a father in the bottom 50% of the rank

[20]The measures are very tightly bounded for the 1980s birth cohort, because there is a rank boundary close to 50 in the parent distribution. When the distance between upper and lower bound is less than 0.2, we report the midpoint as a point estimate.

[21]p_{25} in the U.S. income distribution is 42 (Chetty et al., 2014); educational outcomes are thus more persistent across generations than income.

[22]This number can be compared to a 20 rank point gap between children from the bottom and top halves in the U.S. income distribution.

distribution, or $E(\text{school} \geq \text{Primary}|\text{parent rank} \in [0, 50])$. The dark gray line shows the same outcome, conditional on a father in the top half of the distribution. The trend lines are increasing in both panels, reflecting that Indians at all levels of the socioeconomic distribution are now obtaining more education than they were in the past. The gap between the upward and downward mobility estimates for primary school has closed substantially; this is in part because children from top half families nearly universally attain primary school, so catchup is mechanical. For high school education, the gap has widened substantially, from 20 percentage points to 37 percentage points.[23] High school degrees are nearly a necessary condition for desirable white-collar or government jobs. High school completion is booming, but nearly all of the gains have accrued to children from families in the top half of the distribution.

We next examine how these levels and trends differ across groups. Figure 4 presents results analogous to those above but subdivided into Muslims, Scheduled Castes, Scheduled Tribes, and all others. Panel A shows upward interval mobility (μ_0^{50}) from the 1950s to the 1980s, revealing substantial differences across groups. As noted by other researchers, upward mobility for Scheduled Tribes, and especially for Scheduled Castes, has improved substantially. SCs born in the bottom half of the parent distribution in the 1950s could expect to obtain between the 30th and 34th percentile; the comparable group in the 1980s obtains the 38th percentile, closing approximately half of the mobility gap with upper castes. Upward mobility for members of Scheduled Tribes rises from $[25, 29]$ to 32 over the same period.

In contrast with SCs and STs, Muslim intergenerational mobility rises weakly from the 1950s to the 1960s, but then declines substantially, falling from $[31, 34]$ in the 1960s to 29 in the 1980s birth cohort. These changes not only constitute a substantial decline in mobility, but make Muslims the least upwardly mobile group in present-day India, lower even than the Scheduled Tribes who are often thought of as having benefited very little from Indian industrialization. The fact that a Muslim born in a family in the bottom half of the distribution

[23] It may be noted that the top half had five times the high school completion rate as the bottom half in 1950s, and a rate only three times higher in the 1980s. However, the completion rate was so low for children of parents in the lower half of the distribution that this convergence does not strike us as a useful measure.

can expect to obtain the 29th percentile implies almost no reversion to the national mean among this group. Finally, the "Forward/Others" group, which predominantly consists of higher caste Hindus, shows little change, with mobility shifting from $[41, 45]$ to 41. In the period since the 1970s where the bounds are less than a percentage point in width, this group shows no change in mobility. The static trend in mobility can therefore be decomposed into gains for SCs and STs and losses for Muslims.

Panel B shows downward interval mobility (μ_{50}^{100}) over the same period. Among children of parents in the top half of the distribution, members of all three historically marginalized groups are more similar. The bounds in the 1950s are too broad to be interpreted, but since the 1960s, Scheduled Castes and Scheduled Tribes have each risen by about four expected ranks, while Muslims have risen by slightly less. Downward mobility among higher caste Hindus has remained static at about 65, considerably higher than the outcomes among the other groups.

Panels C and D show upward mobility results with education levels as outcomes, comparable to Figure 3. The graphs show the share of son bottom half families who respectively obtained primary school (Panel C) and high school (Panel D). Children with a high school education represent 15% of the 1950s birth cohort and 25% of the 1980s birth cohort. The probability of finishing primary or high school conditional on beginning in the bottom half of the parent distribution is rising for all groups, reflecting higher national education rates. The cross-group differences and changes are similar to those in Panel A; for both primary and high school attainment, SCs and STs have experienced substantially larger gains than Muslims. A particular concerning finding is that among children of bottom half parents, Muslim high school attainment did not rise at all from the 1970s to the 1980s birth cohorts.

5.2 The Geography of Group Differences

This section explores the geography of group differences using the IHDS. The finest geographic identifier in the IHDS is the district, but the survey is representative only at the state level.

First, we explore whether the group differences described above can be explained by the places where SCs, STs and Muslims live. All three groups are unevenly distributed; the 25th percentile district in SC population is only 8% SC. The equivalent numbers for STs and Muslims (0.4% and 2.7%) reflect that these groups are even more geographically concentrated.

To examine the effect of place on group differences, we regenerate father and son education ranks within states and within districts. Mobility estimates generated in this way thus describe the ability of children to increase their relative rank within their own district. If low overall mobility for Muslims is a function of their living in districts where everyone has low educational opportunity, then their within-district mobility gap with Forwards / Others should be substantially smaller than the national mobility gap.

These results are shown in Figure 5. The first set of bars shows the relative mobility gap to Forwards / Others for the three marginalized groups for the 1980s birth cohort using the national ranks; these gaps correspond to the differences between groups in Figure 4A in the 1980s. Upward mobility for the Forward / Others reference group is 41. The following two sets of bars show the same gaps for within-state and within-district ranks.

The extent to which group differences in upward mobility can be explained by location differ substantially by group. District of residence explains about 9% of the Muslim upward mobility gap, 14% of the Scheduled Caste upward mobility gap, and a full 59% of the Scheduled Tribe mobility gap.[24] The result for Scheduled Tribes is consistent with the continued dependence of many STs on traditional livelihoods in remote areas. Given the uneven distribution of SCs and Muslims throughout India, the unimportance of district as an explanation for their lower mobility is worthy of note. A striking fact is that Jammu & Kashmir, the only majority Muslim state, has the fourth highest level of upward mobility in India. Note that 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, but these

[24]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 5.

geographic identifiers are not available in the IHDS.[25] To be clear, these results show that location is not a major mediator of SC and Muslim disadvantage; however, we show below that location is an important predictor of mobility in the aggregate in Section 5.3.

Figure 6 disaggregates the same mobility gaps into residents of urban and rural areas, with the national father/son ranks used in the main analysis. Overall upward mobility is on average five points higher in urban areas (not shown in figure); this difference varies substantially across groups. In particular, Muslims are considerably more disadvantaged in terms of upward mobility in cities. Conditional on birth to parents in the bottom half of the education distribution, Muslims expect to obtain a rank 15 points lower than higher caste Hindus in cities, compared to a 12 point disadvantage in rural areas. That said, mobility is still slightly higher for Muslims in cities than in rural areas. SCs and STs have similar disadvantages in rural and urban areas, even though they obtain ranks about six points lower in rural than in urban areas.

5.3 Geographic Analysis of Mobility at a Finer Level

The Socioeconomic and Caste Census allows us to go to a finer level of geographic detail, but as noted above, does not document Muslim identity. We therefore focus in this section on the geographic predictors of intergenerational mobility, pooling results across all population groups. As noted in Section 3, the sample consists of individuals aged 20–23 (or born between 1989 and 1992, as this is the sample for which we can accurately measure educational mobility.

We begin by mapping the distribution of intergenerational mobility across India. Figure 7 presents a heat map of subdistrict- and town-level upward interval mobility for all of India. The geographic variation is substantial. Upward mobility in consistently highest in the southernmost part of India—Tamil Nadu and Kerala—and is also noticeably high in the hilly states of the North. Parts of the Hindi-speaking belt—especially the state of Bihar—and

[25] These identifiers are available in the SECC, but we cannot identify Muslims in the SECC, making it impossible to replicate this exercise.

the Northeast are among the lowest mobility parts of India. Cities and towns for the most part stand out as islands with higher mobility. Gujarat is interesting to note as a state with very high economic growth but relatively low mobility.

The local variation in mobility remains substantial. In broad regions of high mobility, there are low mobility islands, such as the hilly region between Andhra Pradesh and Karnataka, and vice versa. However, there is not a single subdistrict or town in Bihar with higher average mobility than the southern states.

In fact, there is substantial variation in mobility based on neighborhood of residence even within a single city. Figure 8 shows a ward-level mobility map of Delhi. The highest mobility wards have upward mobility that is 38% higher than the lowest mobility wards. Children who grow up in the dense and industrial areas of Northeast Delhi have the least opportunity; the average child from a bottom half family in this area can expect to obtain the 32nd percentile nationally. Children from similarly-ranked families who grow up in Southwest Delhi, less than 40km away, can expect to obtain the 44th percentile.

One caveat to these maps is that they are based on the educational outcomes of children born between 1989 and 1992, the majority of whom finished their education by 2010. The maps thus reflect the circumstances that drove education choices in the period 2000–2010. The maps also do not reflect migration, as people are coded according to their residence in 2012. Part of the low mobility in Northeast Delhi may thus arise from its substantial immigration from rural North India. The more local the mobility estimate, the greater the potential bias from migration. To better unpack these local estimates, it will be necessary to conduct surveys that record location of origin; to our knowledge, there are no such surveys that are systematically available for urban neighborhoods of India. The subdistrict-level estimates are less likely to be biased by migration, because the vast majority of permanent migration is within subdistricts.

We next aim to describe the characteristics of places with high and low mobility. Figure 9 presents the association between upward interval mobility and several correlates identified by the earlier literature. Panel A presents bivariate correlations between upward mobility and location characteristics across all rural subdistricts in India (n=5000). Panel B presents analogous results across the 2000 largest towns. The indicators cover four broad areas: subgroup distribution (specifically, presence of SCs and STs, and the residential segregation of SCs and STs); inequality (consumption and land inequality); development (manufacturing jobs per capita, average consumption, average education, and remoteness); and local public goods (schools, paved roads, and electricity). All measures are standardized to mean zero and standard deviation one so that they can be meaningfully compared.

At the rural level, the traditional markers of economic development — manufacturing jobs, monthly consumption, and average levels of education — are the strongest correlates of upward mobility. Local public goods are also positively correlated with upward mobility. Interestingly, availability of primary schools is the least important of these, but availability of high schools is highly correlated with mobility. The share of SCs and STs is surprisingly positively correlated with upward mobility, again reflecting the weak role of geography in describing group differences in mobility. Segregation and land inequality are negatively correlated with upward mobility, though the effect size is smaller than that for the core development indicators above. The negative association between segregation and mobility echoes the relationship between segregation and intergenerational mobility in the United States (Chetty et al., 2014).

In urban places, we find that more populous districts, those with higher SC shares, those with higher average years of education, and those with more high schools per capita enjoy more upward interval mobility. As in rural areas, SC/ST segregation is negatively associated with mobility. By contrast, consumption inequality is positively associated with mobility — a fact which does not accord with the cross-country findings or the finding in rural areas, as explained above.

6 Conclusion

The methodological approach in this paper is likely to be useful for analyses of mobility in many developing countries. Given the large number of less educated parents in older generations in many developing countries, conventional mobility estimates will suffer from the weaknesses noted in Section 4 and by Asher et al. (2018). We have shown that interval mobility is a measure that is easy to calculate and highly informative regarding intergenerational mobility even when education data are very coarse. We also provide the first measure of intergenerational educational mobility that is meaningful for cross-group analysis; the absence of such a measure has prevented researchers from studying subgroup mobility in developing countries.

We have shown that upward interval mobility in India has barely changed from the 1950s to the 1980s. This lack of change overall can be decomposed into substantial gains for SC/STs and substantial losses for Muslims. The latter result has not previously been noted in part because there has previously been no methodology for creating comparable rank bins across cohorts.

We also find that urban areas are significantly more mobile than rural areas — for example, the mobility gap between urban and rural locations is about equal to today's gap between higher caste Hindus and SCs. Using granular geographic data, we find that village assets like roads and schools are associated with more upward interval mobility; on the other hand, SC/ST segregation is associated with lower levels of upward interval mobility.

Our work has only begun to describe the wide geographic and cross-group variation in intergenerational mobility. As in the U.S., location is a very strong predictor of intergenerational mobility, even if cross-group difference appear to be largely invariant to location for Muslims and SCs. Individuals growing up in different parts of India, even conditional on similar economic conditions in the household, can expect to experience vastly different opportunities and outcomes throughout their lives. Future work describing the geographic variation in mobility in more detail, and moving toward causal estimates of the impact of place, will be important in providing a basis for policies that create opportunities for those who are currently being left behind.

References

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," 2018. Working paper.

___ , ___ , **and Charlie Rafkin**, "Getting Signal from Interval Data: Theory and Applications to Mortality and Intergenerational Mobility," 2018.

Azam, Mehtabul 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.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan, "Affirmative action in education: Evidence from engineering college admissions in India," *Journal of Public Economics*, 2010, *94* (1-2), 16–29.

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.

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.

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).

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility," 2017. NBER Working Paper No. 23618.

___ , **Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

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.*

Emran, M.S. and Forhad Shilpi, "Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India," *World Development*, 2015, *72.*

Frisancho Robles, Veronica C. and Kala Krishna, "Affirmative Action in Higher Education in India: Targeting, Catch Up and Mismatch," *Higher Education*, 2016, *71* (5).

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).

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.

____, 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.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, 2013, *103* (4).

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.

Mohammed, A R Shariq, "Does A Good Father Now Have To Be Rich? Intergenerational Income Mobility in Rural India," 2016. Working paper.

Munshi, Kaivan 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.

Restuccia, Diego and Carlos Urrutia, "Intergenerational Persistence of Earnings: The Role of Early and College Education," *American Economic Review*, 2004, *94* (5).

Roemer, John., "Equality of Opportunity: Theory and Measurement," *Journal of Economic Literature*, 2016, *54* (4).

Sachar Committee Report, "Social, Economic and Educational Status of the Muslim Community of India," Technical Report, Government of India 2006.

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.

Figure 1

Raw Data and Bounds on Mobility CEF by Birth Cohort

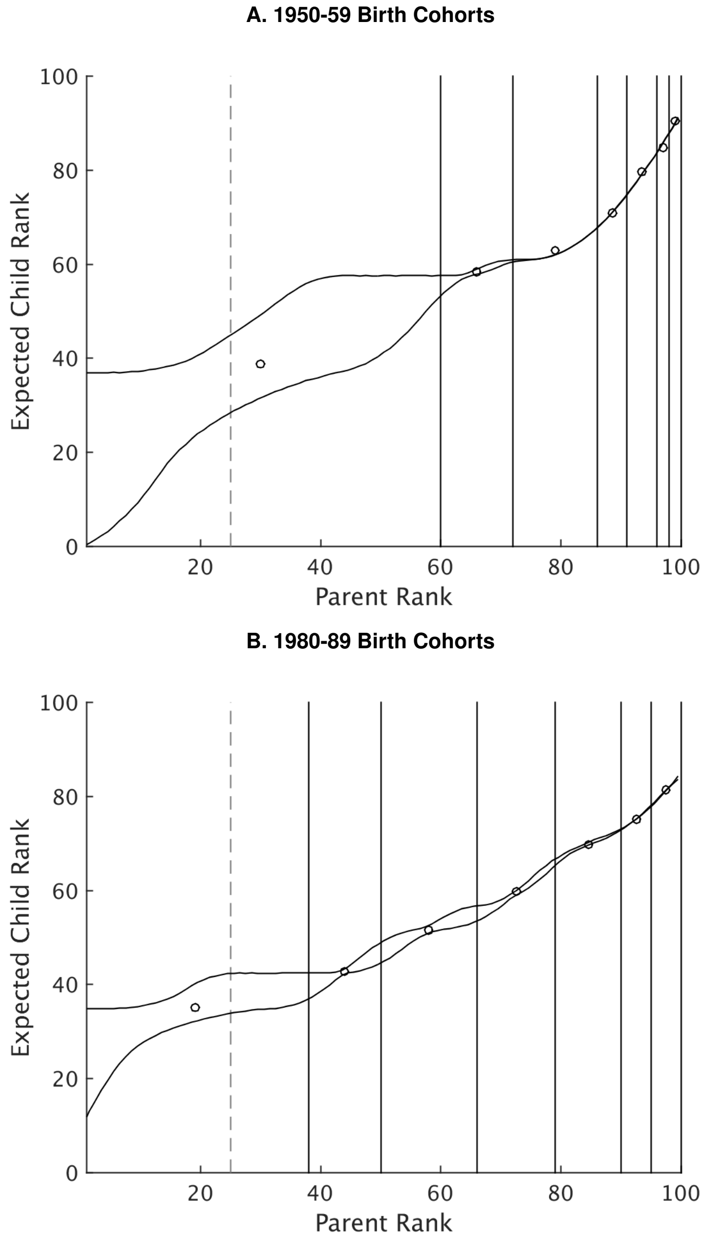


Figure 1 shows the raw data and bounds on the conditional expectation of son education rank given father education rank. Panel A is for sons born in 1950-59 and Panel for sons born 1980-89. The points show the raw data—the average son education rank and the midpoint of the father education rank for each of the seven observed levels of father education. The vertical lines show the rank bin boundaries for the levels of father education. No information about fathers is observed except which of the seven rank bins they are in. The solid lines show the bounds on the CEF, calculated assuming only monotonicity, following Asher et al. (2018). The dashed vertical line indicates the 25th father percentile; the bounds at this point are the bounds on absolute upward mobility (p_{25}).

Figure 2

Trends in Mobility, 1950–1985 Birth Cohorts

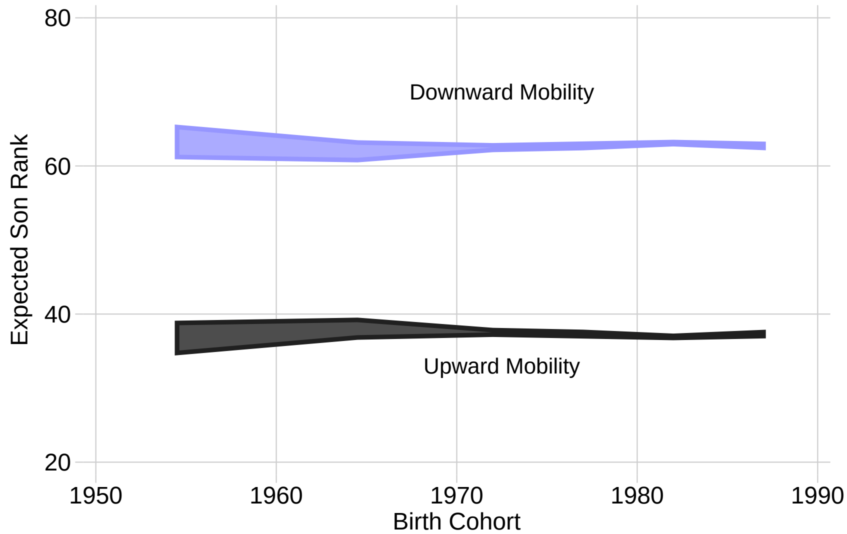


Figure 2 presents bounds on aggregate trends in intergenerational mobility, using cohorts born in 1950 through 1985. The data source is IHDS. The light blue line shows downward interval mobility (μ_{50}^{100}) while the dark gray line shows upward interval mobility (μ_0^{50}). Downward interval mobility is the average rank attained by children born to fathers who are in the top half of the education distribution. Upward interval mobility is the average rank attained by children born to fathers who are in the bottom half of the education distribution. Bounds are calculated from the coarse education data following Asher et al. (2018).

Figure 3

Trends in Mobility with Absolute Outcomes, 1950–1985 Birth Cohorts

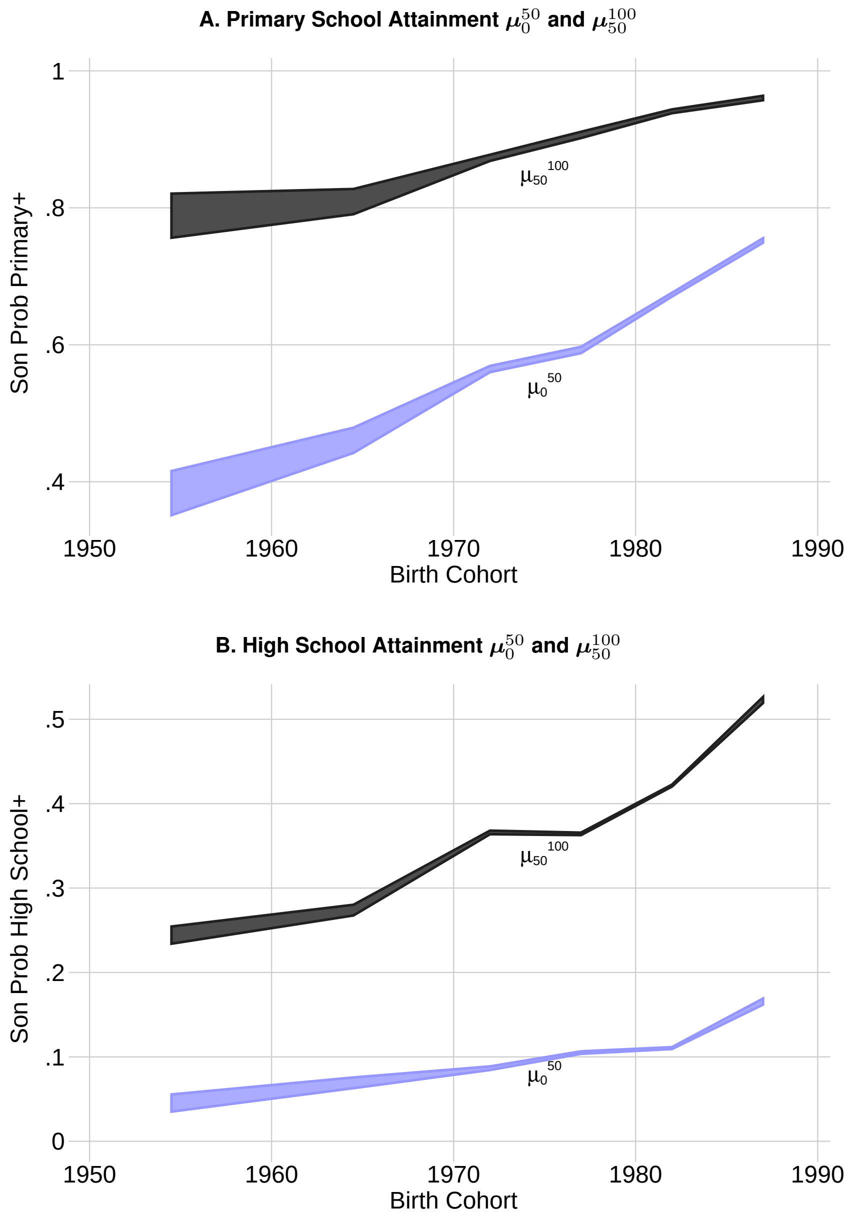


Figure 3 presents bounds on aggregate trends in intergenerational mobility, using cohorts born in 1950 through 1985. The data source is IHDS. The outcome measure is the probability that a child completes either primary (Panel A) or high school (Panel B). Upward mobility here (the dark gray lines) is measured as the probability that a child obtains the given level of education, conditional on having a father in the bottom 50% of the parent education distribution. Downward mobility (the light blue lines) is the same probability, conditional on having a father in the top 50% of the parent education distribution. Bounds are calculated from the coarse education data following Asher et al. (2018).

Figure 4
Trends in Mobility by Subgroup, 1950–1985 Birth Cohorts

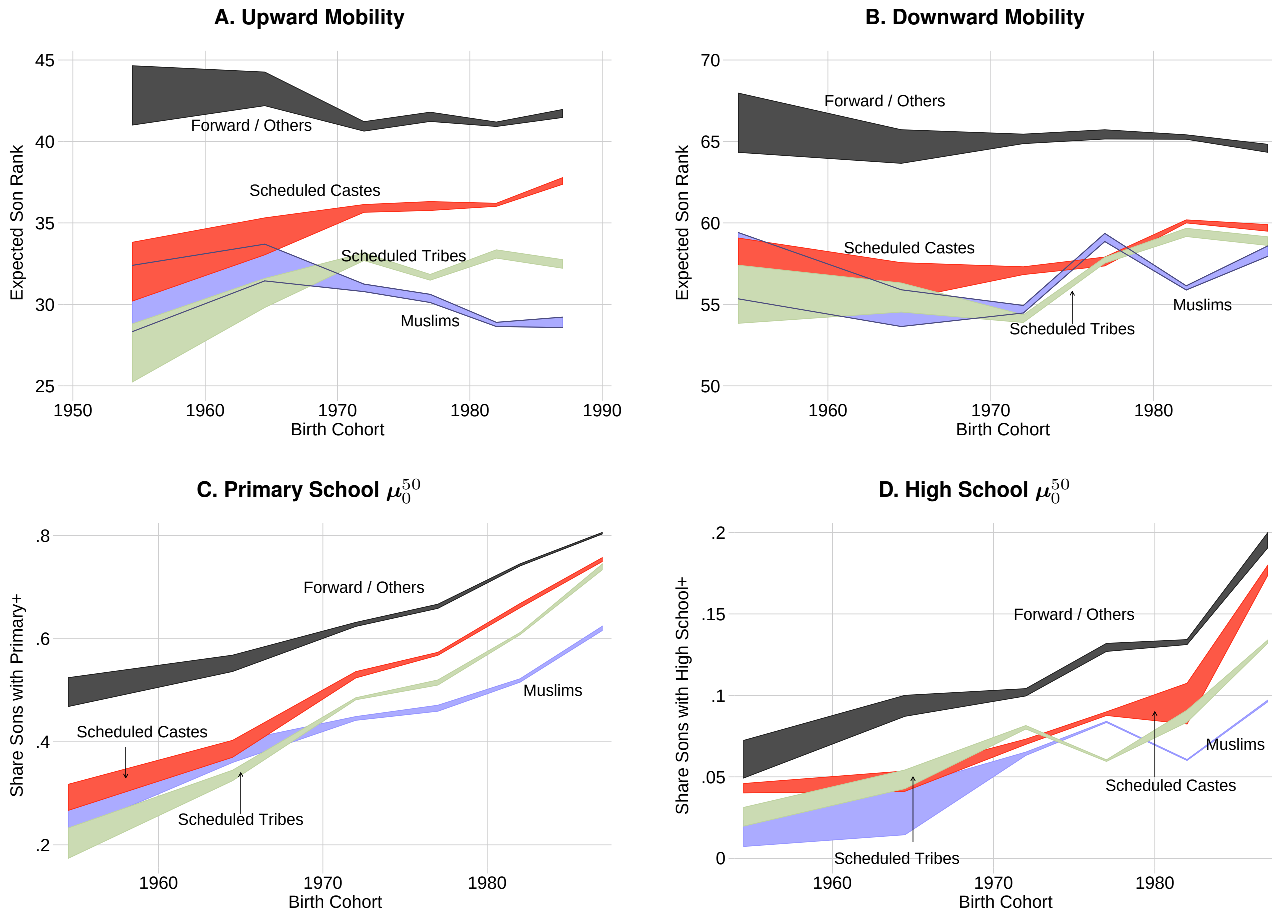


Figure 4 presents bounds on trends in intergenerational mobility, stratified by four prominent social groups in India: Scheduled Castes, Scheduled Tribes, Muslims, and Forward Castes/Others. The data source is IHDS. Panel A presents bounds on trends in upward interval mobility (μ_0^{50}), or the average rank among sons born to fathers in the bottom half of the father education distribution. Panel B presents bounds on trends in downward interval mobility (μ_{50}^{100}), or the average education rank among sons born to fathers in the top half of the father education distribution. Panel C presents bounds on trends in the proportion of sons completing primary school, conditional on being born to a father in the bottom half of the education distribution (μ_0^{50}). Panel D presents bounds on trends in the proportion of sons attaining a high school degree, conditional on being born to a father in the bottom half of the education distribution (again μ_0^{50}). Bounds are calculated from coarse education data following Asher et al. (2018).

Figure 5
Within-State and Within-District Mobility Gaps

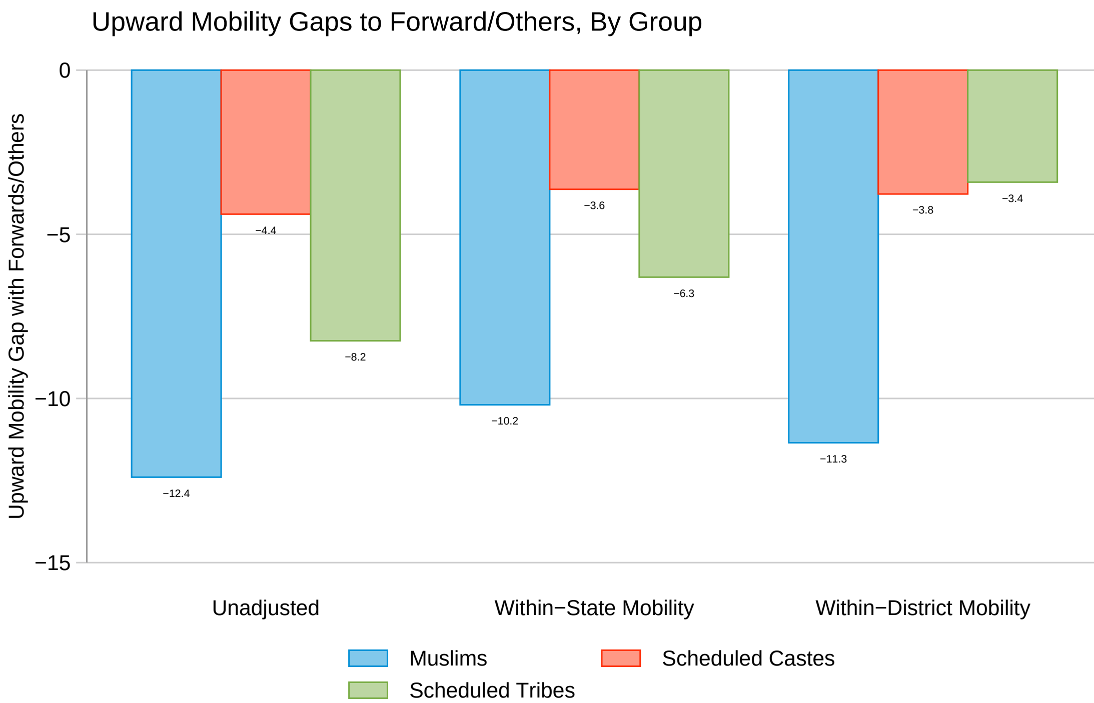


Figure 5 presents the gaps in upward mobility for Muslims, Scheduled Castes and Scheduled Tribes relative to other groups (mainly higher caste Hindus). The data source is IHDS. The first set of bars shows the gaps calculated using within-state father and son education ranks. The third set shows gaps calculated using within-district father and son education ranks. Upward mobility is defined as μ_0^{50}, with son education rank as the outcome variable. Upward mobility is only partially identified following Asher et al. (2018); for simplicity, we show the midpoint of the bounds, which in all cases span less than a single rank.

Figure 6
Urban and Rural Mobility Gaps

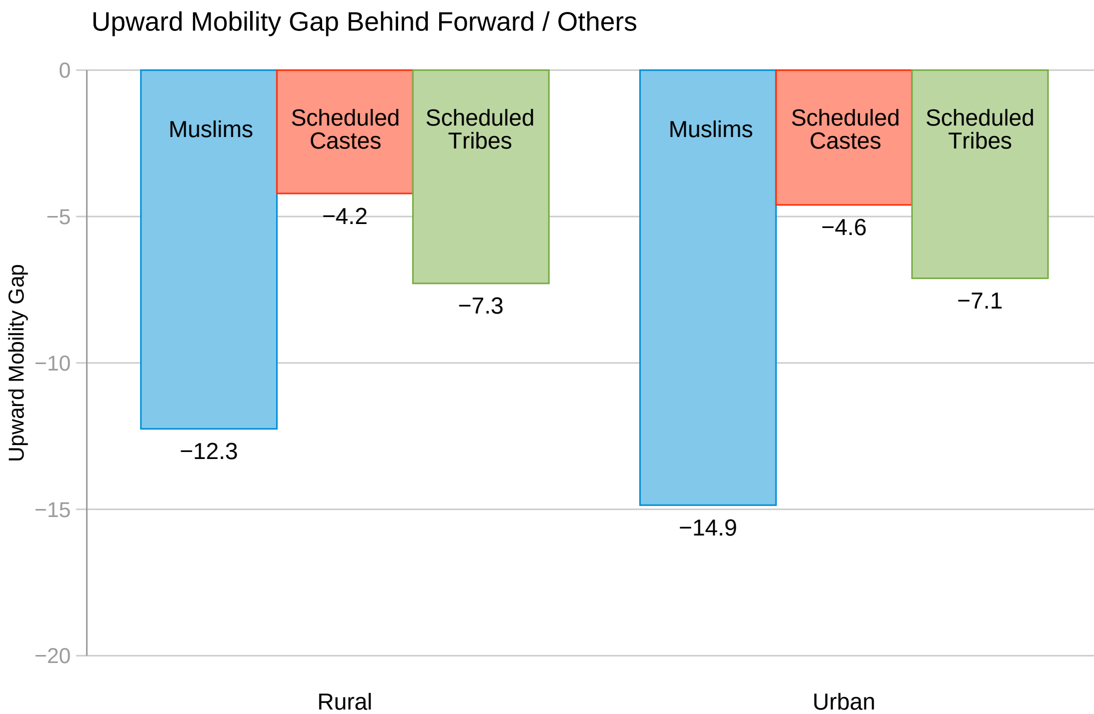


Figure 6 presents gaps in upward mobility for Muslims, Scheduled Castes and Scheduled Tribes relative to other groups (mainly higher caste Hindus), separately for those residing in rural (left bars) and urban (right bars) areas. The data source is IHDS. Upward mobility is defined as μ_0^{50}, with son education rank as the outcome variable. Upward mobility is only partially identified following Asher et al. (2018); for simplicity, we show the midpoint of the bounds, which in all cases span less than a single rank.

Figure 7

Mobility by Geographic Location: National Estimates

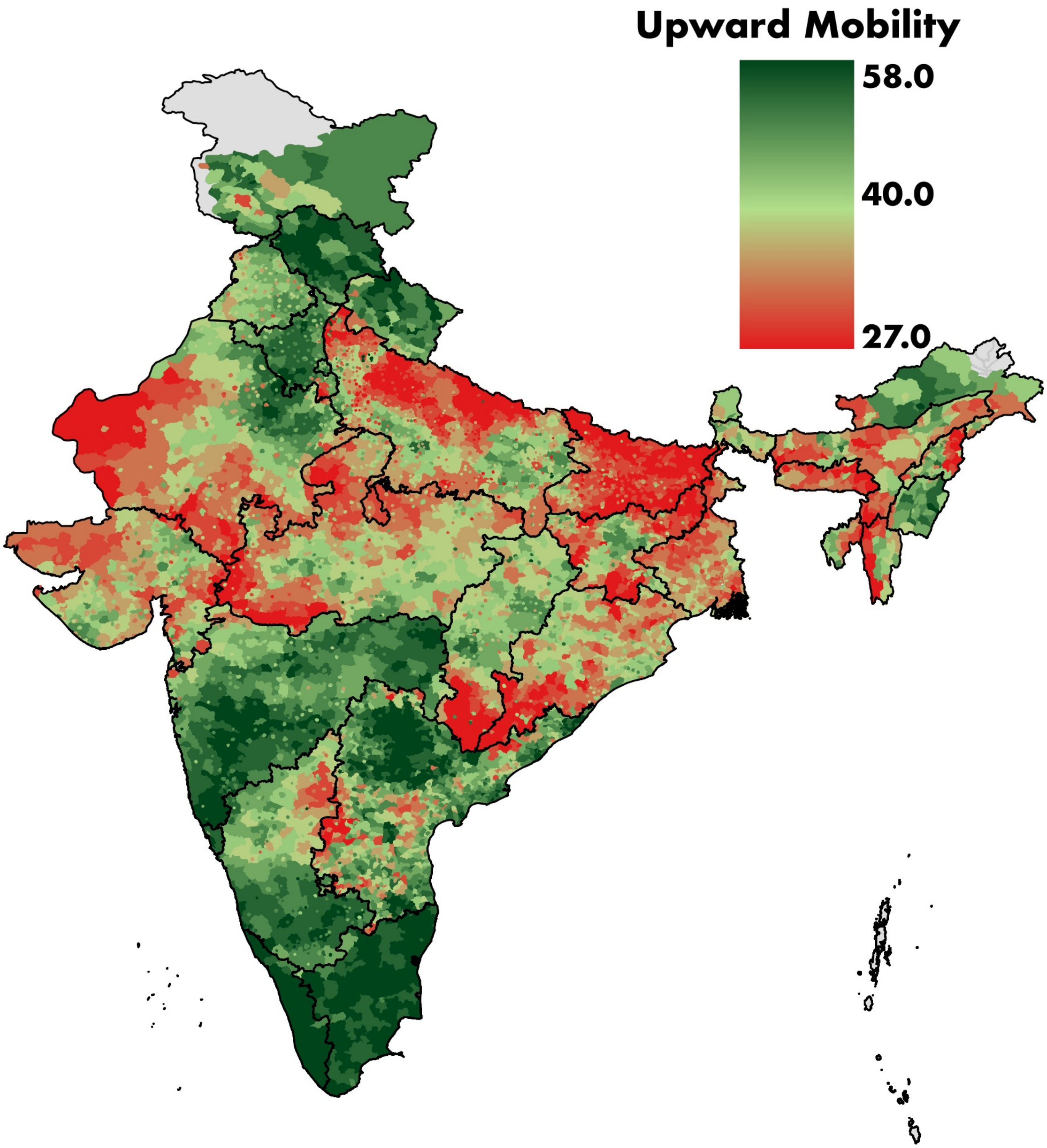


Figure 7 presents a map of the geographic distribution of upward mobility across Indian subdistricts and towns. Upward mobility (μ_0^{50}) is the average education rank attained by sons born to fathers who are in the bottom half of the father education distribution.

Figure 8

Mobility by Geographic Location: Neighborhood Estimates from Delhi

Figure 8 presents a map of the geographic distribution of upward mobility across the wards of Delhi. Upward mobility (μ_0^{50}) is the average education rank attained by sons born to fathers who are in the bottom half of the father education distribution.

Figure 9

Correlates of Upward Mobility, 1985 Birth Cohort

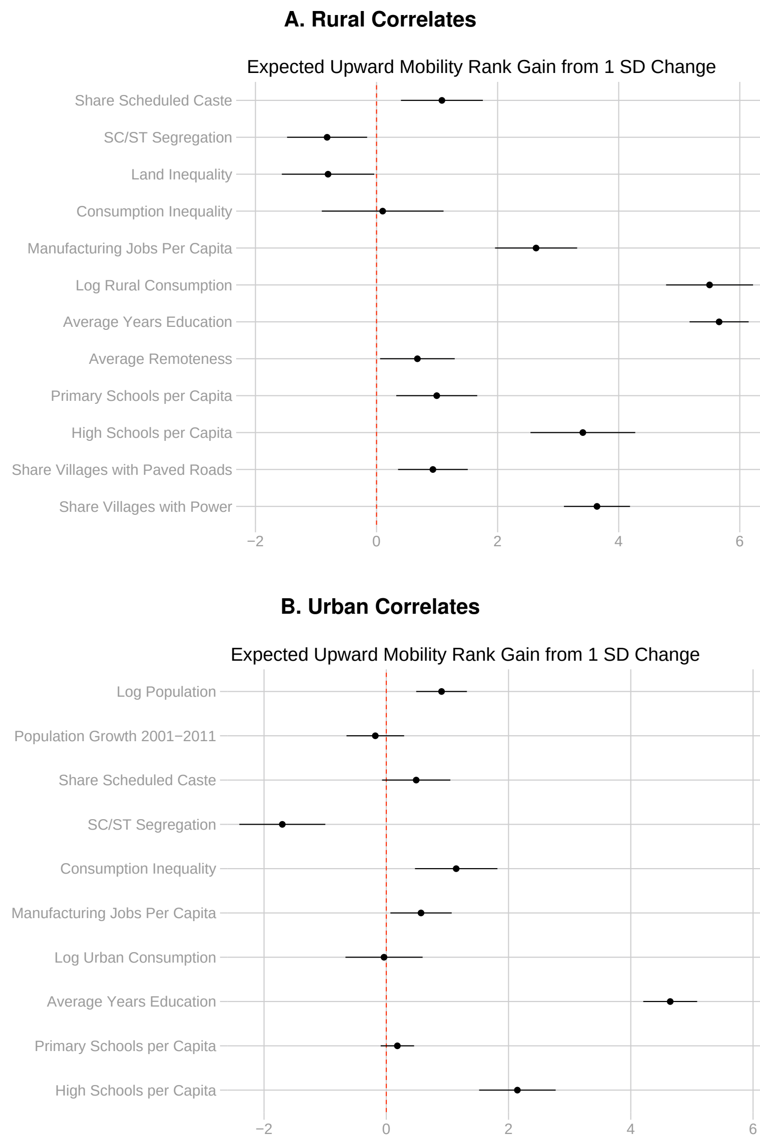


Figure 9 presents coefficients that illustrate the relationship between district-level characteristics and upward mobility. Upward interval mobility (μ_0^{50}) is the average rank attained by children born to fathers who are in the bottom half of the education distribution. Panel A presents correlations for rural areas, while Panel B presents correlations for urban areas.

A Appendix A: Additional Tables and Figures

Table A1

Transition Matrices for Father and Son Education in India

A: Sons Born 1950-59

	Son highest education attained						
Father ed attained	< 2 yrs. (31%)	2-4 yrs. (11%)	Primary (17%)	Middle (13%)	Sec. (13%)	Sr. sec. (6%)	Any higher (8%)
<2 yrs. (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

	Son highest education attained						
Father ed attained	< 2 yrs. (27%)	2-4 yrs. (10%)	Primary (16%)	Middle (16%)	Sec. (14%)	Sr. sec. (7%)	Any higher (10%)
<2 yrs. (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

	Son highest education attained						
Father ed 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

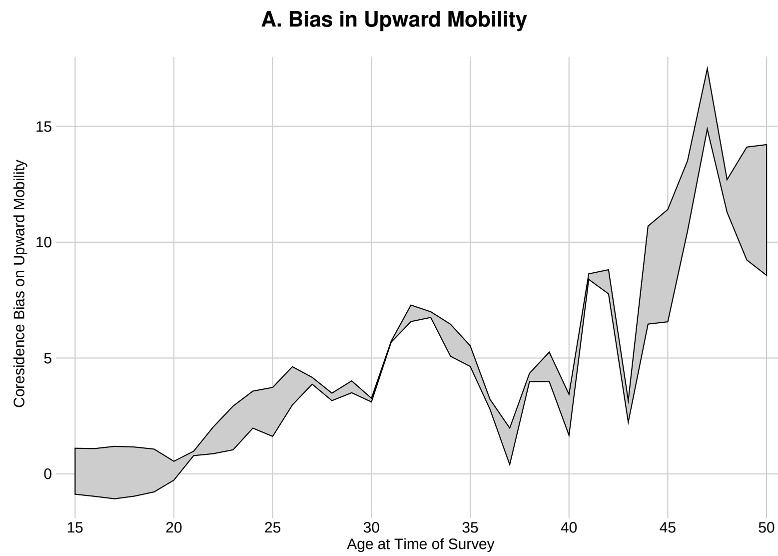
D: Sons Born 1980-89

	Son highest education attained						
Father ed 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 A1 shows transition matrices by decadal birth cohort for Indian fathers and sons in the study.

Figure A1

Bias in Coresidence Estimates of Upward and Downward Mobility by Age



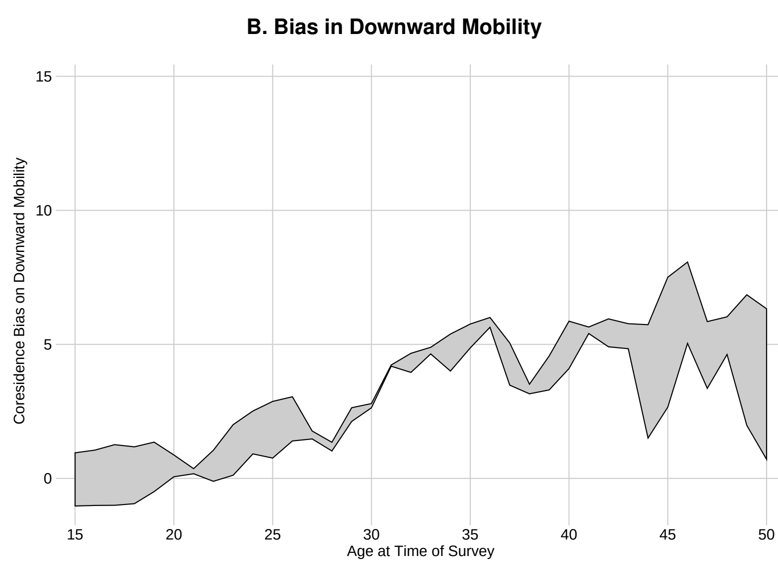


Figure A2

Trends in Mobility with Absolute Outcomes, 1960–1985 Birth Cohorts

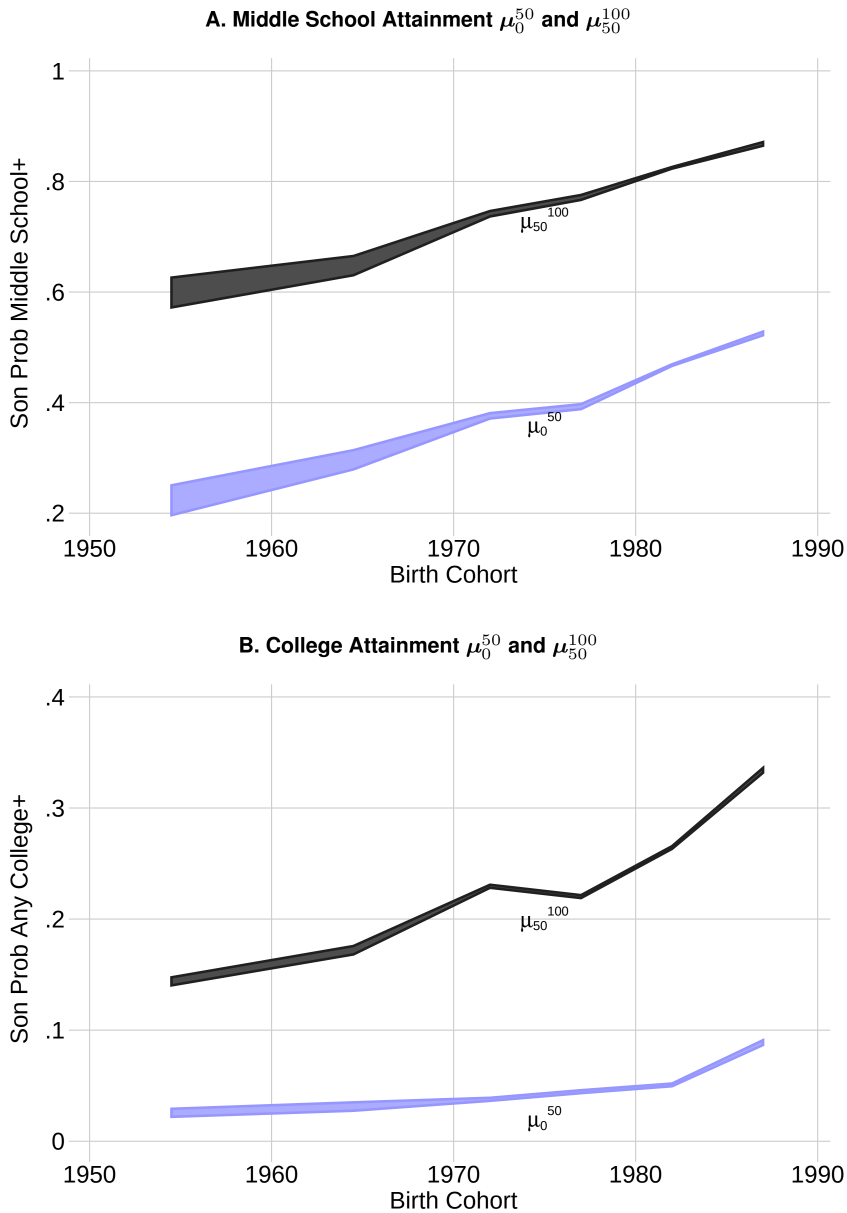


Figure A2 presents bounds on aggregate trends in intergenerational mobility, using cohorts born in 1950 through 1985. The data source is IHDS. The outcome measure is the probability that a child completes either middle school (Panel A) or university (Panel B). Upward mobility here (the dark gray lines) is measured as the probability that a child obtains the given level of education, conditional on having a father in the bottom 50% of the parent education distribution. Downward mobility (the light blue lines) is the same probability, conditional on having a father in the top 50% of the parent education distribution. Bounds are calculated from the coarse education data following Asher et al. (2018).

Figure A3

Correlates of Downward Mobility, 1985 Birth Cohorts

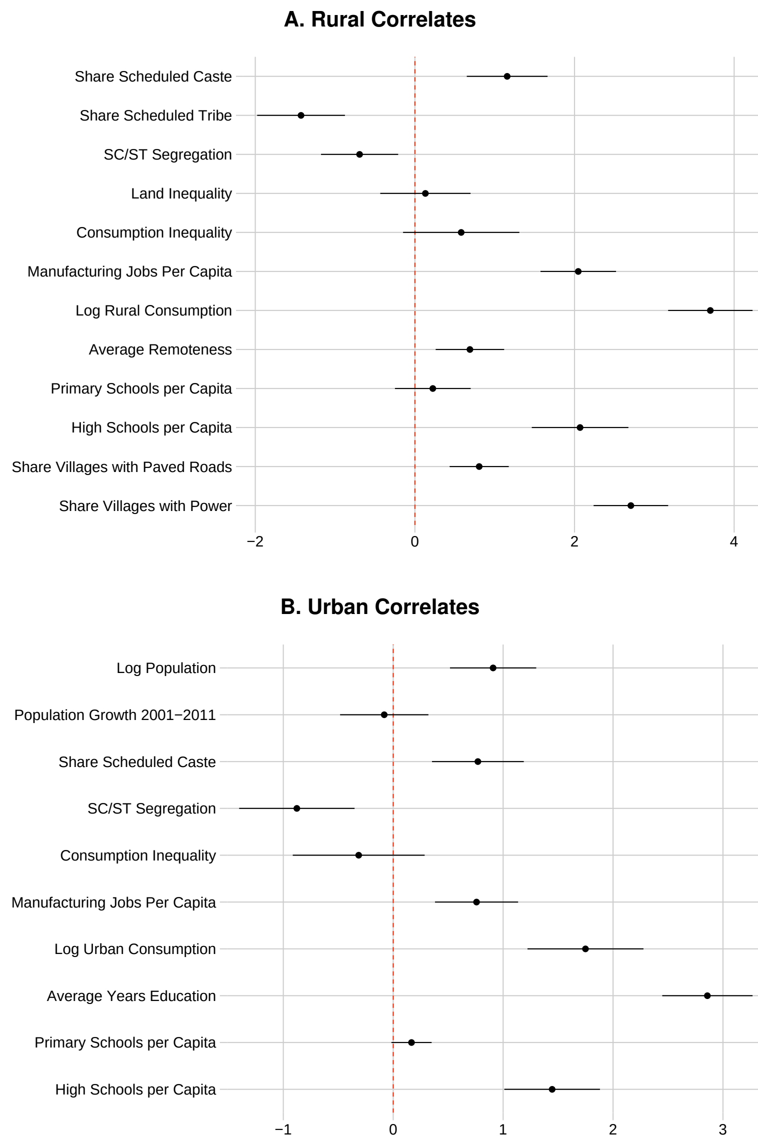


Figure A3 presents coefficients that illustrate the relationship between district-level characteristics and downward interval mobility (μ_{50}^{100}). Downward interval mobility is the average rank attained by sons born to fathers who are in the top half of the education distribution. Panel A presents correlations for rural areas, while Panel B presents correlations for urban areas. The methodology to generate the correlations is described in the text.