Estimating Intergenerational Mobility with Coarse Data: A Nonparametric Approach
Sam Asher∗
Paul Novosad†
Charlie Rafkin‡

May 2017
First version: February 2016

Abstract

Estimates of intergenerational mobility in developing countries usually focus on the persistence of education across generations, due to lack of other matched parent-child outcome data. These estimates may be significantly biased because education is observed only at discrete intervals and in large bins. We present a new nonparametric method that, given interval-censored data, delivers informative bounds on the conditional expectation of child rank given parent rank under minimal structural assumptions. Drawing on new administrative census data, we apply our method to India, we recover precise bounds in the top half of the distribution and wider bands at the bottom, where education is most coarsely observed. We show that rank persistence has declined at the very top of the education distribution, and that the gap between marginalized groups and the general population appears to be shrinking. Our nonparametric method may be useful in other contexts with interval-censored or top-coded data.

∗World Bank Research Group
†Dartmouth College, paul.novosad@dartmouth.edu
‡National Bureau of Economic Research, crafkin@nber.org
I Introduction

The intergenerational transmission of economic status, a proxy for equality of opportunity, is of increasing interest (Solon, 1999; ?, Chetty et al., 2014). High dependence of children’s outcomes on the outcomes of their parents is generally considered inferior to low dependence, and has implications for allocative efficiency and subjective welfare.

Intergenerational mobility estimates from the developing world typically use education as a proxy for social rank, because other matched parent-child outcomes are rarely available.\textsuperscript{1} A significant disadvantage of education as a measure of social rank is that it is interval censored, that is, it is measured in broad bins. Years of education are often lumped at school completion thresholds and surveys often report only the highest level attained, rather than exact years. More importantly, a large share of the population in developing countries has completed zero years of education, making it impossible to distinguish between outcomes in this bottom group. For example, for Indian children born in the 1960s, over 50% of their parents report zero years of education, providing little information about the latent distribution of social status among uneducated parents. Almost every study of intergenerational educational mobility implicitly assumes that the conditional expectation of children’s outcomes are linear in latent parent outcomes in this large bottom bin, but this assumption is contradicted in nearly every study where parent and child outcomes are observed with higher granularity.

In this paper, we derive estimates of intergenerational mobility that account for interval censoring in education data. We do this by calculating bounds on the set of nonparametric conditional expectation functions that can fit the observed distribution of parent and child outcomes. We can then bound any of the statistics of intergenerational mobility that can be calculated from the conditional expectation function.

For each parent rank between 1 and 100, we calculate bounds on the expected rank of the child under a minimal set of structural assumptions. First, we assume that the condi-

\textsuperscript{1}Recent studies have measured intergenerational educational mobility in Brazil (Dunn, 2007), China (Knight et al., 2011), India (Hnatkovska et al., 2013), South Africa (Piraino, 2015) and Turkey (Aydemir and Yazici, 2016).
tional expectation function of children’s outcomes knowing parent outcomes (henceforth, the 
CEF) is monotonically increasing, or that an increase in the parent’s rank cannot lower the 
expected value of the child’s rank. With just this fundamental assumption, we can obtain 
meaningful bounds on the CEF, and thus on many statistics of intergenerational mobility. 
These bounds are not analogous to confidence intervals; every parameter estimate with the 
bounds can precisely match the empirical moments of the parent-child distribution. We 
estimate confidence intervals on the bounds using bootstrapping.

Next, we assume that the curvature of the CEF has some upper bound. This assumption 
is justified by examination of a wide set of empirical parent-child rank distributions that are 
estimated on data with full support in the rank distribution, typically from OECD coun-
tries. As we decrease the allowed curvature threshold, the bounds on the mobility estimates 
necessarily shrink. In the limit case where restrict the CEF curvature to zero, we recover the 
parent-child rank elasticity, whose slope is one of the most common summary statistics of 
intergenerational mobility. Our method thus provides a generalization of the classical linear 
approach to the estimation of intergenerational mobility.

We propose a rule of thumb for a better curvature restriction, based on existing nonpara-
metric mobility estimates. Specifically, we propose that the estimated curvature should be 
less than the double the upper limit of curvature found in existing empirical studies that are 
estimated on the full support of the rank distribution.$^{2}$

In an example from Denmark, where the fully supported parent-child income rank distribu-
tion is known, we show that our method can generate reasonable bounds from artificially 
censored data. We then turn to India, where the fully supported parent-child distribution is 
not known.$^{3}$ We draw on new administrative census data on all matched father-son pairs in

---

$^{2}$We do not use the exact limit, because the full distribution has only been estimated in developed 
countries, and we do not know what the full distribution looks like in poor countries.

$^{3}$Studies focusing on regression elasticity estimates of mobility have reported decreasing elasticities in 
India, suggesting higher mobility over time (Maitra and Sharma, 2009; Azam and Bhatt, 2012; Hnaatkowska et 
al., 2013). However, these changes are primarily driven by changes in the variance of the education distribu-
tion; estimates using the correlation coefficient have suggested little change in mobility over recent decades.$^{7}$
The administrative data allow us to estimate precisely the average rank of sons born to fathers at every level of education, but as is common in studies in developing countries, we observe fathers’ education only in one of seven bins. We can attain tight bounds on the CEF in the top half of the distribution, where the education bins are relatively small. However, because the bottom bin is very large (36% of children born in the 1990s have fathers with less than 2 years of education, and 59% of children born in the 1950s), the bounds on the CEF in the bottom of the distribution are considerably wider. Using our rule of thumb curvature restriction, we cannot reject important positive or negative changes in intergenerational mobility in the bottom half of the distribution. However, we can show that mobility has increased at the very top of the distribution even under conservative assumptions.

The data clearly reject linearity of the CEF. Nevertheless, for comparison with the existing literature, we estimate bounds on the linear rank-rank elasticity. The traditional assumption of linearity would generate rank elasticity estimates of 0.59 for the 1950s cohort and 0.52 for the 1990s cohort. Once we allow for the possibility that many nonparametric CEFs can deliver the same observed child outcomes, we find that the rank-rank elasticity can at best be bounded by the values (0.49, 0.71) for the older cohort and (0.49, 0.53) for the younger cohort. By virtue of their linearity, even these bounds fail to capture the importance difference in certainty between mobility at the top of the distribution and mobility at the bottom.

Finally, we examine the widely discussed mobility gap between India’s general population and its historically marginalized scheduled castes and scheduled tribes (SC/STs) (Munshi and Rosenzweig, 2006; Kijima, 2006; Ito, 2009; Hnatkovska et al., 2012). To measure mobility, we focus on absolute mobility at the 25th percentile, which describes the expected outcome of children born to parents in the bottom half of the distribution. In both the younger and the older cohort, we find that the upper and lower bounds are significantly lower for lower castes in both periods. However, the bounds are overlapping, indicating that there are feasible pairs

---

4While mobility with respect to daughters and mothers is equally relevant, we are constrained because the only matched pairs that we can observe in our data are fathers and sons.

5Because the fully supported rank distribution is uniform, the rank correlation coefficient is equivalent to the rank elasticity.
of SC/ST and General CEFs in which the mobility gap is zero. In order to generate a test statistic of the difference between SC/STs and Generals, it is necessary to assume a prior distribution for the value of absolute mobility. Under the assumption of uniformity, which we argue is appropriate, we can reject equality of SC/STs and Generals in either period, but we cannot reject that the SC/ST-General gap has remained unchanged over 40 years.

Our key insight is that when there is interval censoring in the data, there may be a wide set of conditional expectation functions that can deliver precisely the same set of moments, functions which can be bounded by structural assumptions. In the context of intergenerational mobility, the conventional assumption of linearity in the distribution is typically rejected by the data. Even if a linear CEF is desired (for example, for parsimony), there may be multiple linear estimators that perfectly fit an underlying non-linear CEF that meets all the moment restrictions. Our application shows that estimates of intergenerational mobility in developing countries are considerably more precise at the top of the outcome distribution than they are at the bottom, especially for older cohorts. This is a concerning finding, given that many studies of intergenerational mobility are motivated by concern for the disadvantaged.

Our method is particularly well-suited to the context of intergenerational mobility, because the monotonicity assumption is both reasonable and significantly constrains the set of feasible functions. However, the approach of estimating structurally constrained nonparametric bounds may be useful in any context where researchers are working with interval-censored data with coarse bins.

Our paper proceeds as follows. Section II describes the theory and estimation of intergenerational mobility, with a focus on estimation in developing countries. Section III describes the Indian data. Section IV describes our method for calculating nonparametric bounds on the conditional expectation of child rank as a function of father’s rank. Section V applies our method to the estimation of changes in mobility over time and estimating differences in mobility between social groups in India. Section VI concludes.
II Background: Theory and Empirics of Mobility Measurement

Borrowing an example from Solon (1999), consider two societies with equal levels of talent at birth, but where Society A has children whose economic outcomes are perfectly determined by parents’ outcomes and Society B has children whose outcomes are not predicted at all by parents’ outcomes. Societies with greater equality of opportunity, like Society B, can be expected to have greater allocative efficiency, and may have higher subjective welfare and tolerance of inequality. The extent of mobility in a society is an increasingly studied aspect of inequality; there is a growing literature on the variation in intergenerational mobility across countries and across groups within countries, as well as how it has changed over time.\(^6\)

The first generation of intergenerational mobility studies described the matched parent-child outcome distribution with a single linear parameter, typically using either: (i) the correlation coefficient between children’s earnings and parents’ earnings; or ii) the parent-child earnings elasticity, \(i.e.\) the regression coefficient obtained by regressing children’s earnings on parents’ earnings (Solon, 1999; ?). Easy to calculate and interpret, these measures have been widely used and are the basis of studies in dozens of countries.

Recent studies often transform the parent-child outcome distribution into a rank distribution (??; Chetty et al., 2014), with both parent and child ranks calculated within the children’s cohorts. Since the parent and child rank distributions are uniform, the parent-child rank correlation coefficient and rank elasticity are now identical, and the mobility measure is robust to changes in the variance of the parent-child outcome distribution. The rank distributions are also more easily compared across countries than level distributions.\(^8\)

Studies able to draw on richer data have examined the entire nonparametric conditional expectation of child outcomes given parent outcomes. These studies have highlighted that

---

\(^6\)See (Hertz et al., 2008) for cross-country comparisons, and Solon (1999), Corak (2013), ?, and ? for review papers.

\(^7\)These measures are related through the variance of parents’ and children’s outcomes: in the absence of additional regressors, \(\beta = \rho \frac{\mu_{children}}{\mu_{parents}}\), where \(\sigma\) denotes the standard deviation of lifetime earnings, \(\rho\) is the correlation coefficient and \(\beta\) is the parent-child outcome elasticity.

\(^8\)The rank distribution also facilitates handling of outcome data with zero observations, such as individuals who report zero income.
the CEF is often non-linear, and both mobility and changes in mobility may differ at the top and the bottom of the distribution.\(^9\) Chetty et al. (2014) propose a numerical measuring capturing the value of the nonparametric distribution at any specific rank, which they call absolute mobility. Absolute mobility at the \(i\)-th percentile, which we denote by \(p_i\), is the expectation of a child’s rank if the child’s parent is at the \(i\)-th percentile.\(^{10}\) Absolute mobility at any rank may be interesting; Chetty et al. (2014) focus on \(p_{25}\), which describes the expected rank for the median child in the bottom half of the distribution.

II.A Educational Mobility and Income Mobility

In developing countries, data on matched parents and children are difficult to obtain; the administrative data that underlies studies in developed countries does not exist in poorer countries. Even when matched parent and child income data are available (for example, for the set of parents and children who are coresident), measurement error problems are large. Transitory incomes are very 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.\(^{11}\)

For these reasons, studies of intergenerational mobility in developing countries often proxy lifetime income and opportunity with the level of education. This approach has been validated in countries where both approaches are possible; intergenerational educational mobility 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. It also does not face the problem of life cycle bias. The income or income rank of a young person may not be a good proxy for her lifetime income.

---

\(^9\) The CEF of child income given parent income is largely linear in the United States (Chetty et al., 2014), but has important nonlinearities in Denmark (\(?\)), Norway and Sweden (\(?\)), especially in the tails. We are not aware of any studies in developing countries that have been able to report a fully supported rank CEF.

\(^{10}\) Chetty et al. (2014) use a linear predictor of this value because the CEF of child income given parent income is largely linear in the United States. Faced with nonlinear data, we will use a nonparametric predictor.

\(^{11}\) Intergenerational mobility studies often limit samples to the set of individuals who report a wage. See, for example, Hnatkovska et al. (2013) and Azam (2016). Given that transitioning out of agricultural work and into wage work is a central predictor of consumption, this measure has obvious deficiencies.
income, because high permanent income individuals may spend more time in school and thus have lower income than their peers when young. Educational attainment does not face this problem; for most individuals, the level of education will not change after their twenties.

Estimates of educational mobility do have one important drawback, which is the focus of this paper. Education is typically reported in a small number of categories. Even when years of education are specifically measured, they are bunched at school completion levels. In the poorest countries, a large share of individuals report zero years of education completed. As a result, large swaths of the distribution must be assigned the same education rank, even though the latent social rank of these individuals is not the same; child outcomes are likely increasing in parent social status, even in this bottom bin. Given this data constraint, researchers typically assume that the average child value at the mean parent value is equal to the average child value in the entire bin, an assumption that we formalize below. This assumption will be valid if the child CEF is linear, but in the many cases when the CEF is non-linear, this assumption leads to biased mobility estimates due to Jensen’s Inequality. In this paper, we show that intergenerational mobility can be calculated without the assumption of linearity. Because the latent CEF is unobserved, we can at best bound the true CEF, but in some cases the bounds are tight enough to make useful inferences.

Table 1 describes a set of recent studies of intergenerational educational mobility from several rich and poor countries. For each study, we report the range of birth cohorts of sample children, the number of education bins used (where reported), and the share of the parents reporting zero education. Several of the studies observe education in fewer than ten bins, the population share in the bottom bin is often above 20%, and sometimes it is above 50%. Even in the youngest cohorts, a large share of parents have not attained primary education.

---

12To be clear, our method is relevant under the condition that the mobility researcher is using educational attainment as a proxy for some latent continuous social rank. For applications where the level of education is the outcome of interest rather than a proxy of opportunity (for example, the estimation of the returns to education), the level of education is the measure of interest, and the binning problem may not create bias. This said, if there is meaningful unobserved quality variation at each level of education, our method again becomes relevant.

13The studies that we found from Africa did not report the size of the bottom bin, but estimates of primary completion suggest that it is large. In 1980, the interquartile range of the primary enrollment rate was (0.48, 0.8).
school. In our data on India, 36% of fathers of sons in the youngest cohort (born 1983-92) have no schooling. For sons born between 1952-1963, that number rises to 59%. By contrast, in developed countries, the size of the top bin may be too large to develop tight bounds on the CEF. For example, Björklund et al. (2006) study the education of adopted children and report that the size of the top education category for adoptive parents is 40%. In short, interval censoring may be a concern in many studies of intergenerational educational mobility.

II.B Context: Educational Mobility in India

Intergenerational mobility is of particular interest in India. India’s caste system can be considered an institution that expressly prevents intergenerational mobility. While some have argued that growth is making old social and economic divisions less important to the economic opportunities of the young (Hnatkovska et al., 2013), caste remains an important predictor of economic opportunity. The last 30 years have seen tremendous growth in market opportunities in India as well as in the availability and level of education, but many have argued that this growth has been unequal and is leaving many behind (Dreze and Sen, 2013; Field et al., 2016). Whether economic progress can overcome traditional hierarchies of social class is a central question for both India and the broader world.

Increasing intergenerational mobility in India has also been a major objective of the Indian government since independence, leading to policies such as reservations of educational and political positions for members of India’s disadvantaged groups. Equalizing access to education in order to improve mobility has also been an explicit policy goal. Understanding how mobility has changed over time for the general population as well as for historically marginalized groups is central in evaluating these policies.

III Data

To estimate intergenerational educational mobility in India, we draw on two databases that report matched parent-child educational attainment. First, we use administrative microdata...
from the national Socioeconomic Census conducted in 2012. Beginning in 1992, the Government of India has conducted multiple household censuses in order to determine eligibility for various government programs. In 1992, 1997 and 2002, these were referred to as Below Poverty Line (BPL) censuses; the fourth such census, the Socioeconomic and Caste Census (SECC), was launched in 2011 but primarily conducted in 2012.\footnote{It is often referred to as the 2011 SECC, as the initial plan was for the survey to be conducted between June and December 2011. However, various delays meant that the majority of surveying was conducted in 2012. We therefore use 2012 as the relevant year for the SECC.}

The Government of India posted data from the SECC to the internet to enable individuals to verify their information. Each village and urban neighborhood was 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 India languages into English. The individual-level data that we use contain variables describing age, gender, an indicator for Scheduled Tribe or Scheduled Caste status, disability, marital status and relationship with the household head. At the household level, the SECC reports assets and income; we do not use these as they cannot be assigned to a specific individual in the household.\footnote{Additional details of the SECC and the scraping process are described in Asher and Novosad (2016).} The SECC tells us the education level of every parent and child residing in the same household. Because our second dataset reports only matched data on fathers and sons, we focus on the same group in the SECC. Sons who can be matched to fathers through coresidence represent about 85% of 20-year-olds and 7% of 50-year-olds. To ease the computational burden of the analysis, we work with a 1% sample of the SECC, stratified across India’s 640 districts.

The SECC allows us to estimate the parent-child education CEF very precisely for coresident pairs of fathers and sons. This estimation will be biased if the pattern of sons living away from their fathers changes with the fathers’ education.\footnote{Emran et al. (2017) argue that the coresidence bias for intergenerational mobility depends on how mobility is measured, and that the father-son correlation coefficient has only a small bias. In our own estimates using a similar powered test using the IHDS, we found a similarly small point estimate to Emran et al. (2017) of 2 ranks out of 100, but the 95\% confidence interval for the coresidence bias is (-4, 8), leaving the coresident estimates with meaningful uncertainty.} We therefore supplement the
SECC with data from the 2011-2012 round of the India Human Development Survey (IHDS). The IHDS is a nationally representative survey of 41,554 households in 1,503 villages and 971 urban neighborhoods across India. The IHDS is India’s major panel household survey, and collects data on demographic, economic, educational and health characteristics. Crucially, the IHDS solicits information on the education of fathers of household heads, even if the fathers are not resident, allowing us to fill in the gaps in the SECC data. Since SECC contains data on all coresident fathers and sons, we use the IHDS strictly for non-coresident fathers and sons. IHDS contains household weights to make the data nationally representative; we assign constant weights to SECC, given our use of a 1% sample. By combining the two datasets, we can obtain an unbiased and nationally representative estimate of the joint parent-child education distribution.

IHDS records completed years of education, while the SECC records the highest level of education attained, in seven categories.\(^{18}\) To make the two data sources consistent, we re-code 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 six years of schooling in the IHDS, we recode this as five years, which is the level of primary completion.\(^{19}\) The loss in precision by downcoding the IHDS is minimal, because most students exit school at the end of a completed schooling level. Appendix Figure A1 shows the proportion of sons who coreside with their fathers in the SECC. This is also the proportion of each birth cohort for which we get data from the SECC.\(^{20}\)

We estimate mobility over time by examining the joint distribution of fathers’ and sons’

\(^{18}\)The categories are (i) illiterate; (ii) literate without primary (iii) primary; (iv) middle; (v) secondary (vi) higher secondary; and (vii) post-secondary.

\(^{19}\)We code the SECC category “literate without primary” as 2 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. Presented results are minimally affected by this choice; downcoding this category to zero years results in a slightly larger bottom bin, and a less precise estimate of mobility at the bottom of the distribution.

\(^{20}\)Our strategy of combining the IHDS and SECC means that the standard errors on conditional means will be smallest for younger birth cohorts and larger for older birth cohorts, because older children are less likely to live at home. In practice, the uncertainty in measurement of the mean will prove to be much smaller than the uncertainty due to interval censoring, so this will not prove important.
educational attainment for individuals in different birth cohorts. All outcomes are measured in 2012, but because education levels only rarely change in adulthood, these measures capture educational choices made decades earlier. We use decadal cohorts reflecting individuals’ ages at the time of surveying. The oldest cohort of sons that we follow was born in 1952 and would have finished high school by 1970, before the beginning of the liberalization era. The cohorts born after 1985 would have begun school after the major liberalization reforms of 1991, and spent most of their schooling period in the post-liberalization era. The youngest cohort in this study was born in 1992; any younger cohorts may not have completed their education at the time that they were surveyed and thus were excluded. Appendix Figure A2 shows the educational attainment of fathers and sons by cohort; both groups show dramatic gains in educational attainment over the last forty years.

IV Estimating bounds on Intergenerational Mobility

IV.A Interpreting Empirical Moments of the Parent-Child Rank Distribution

In this section, we describe a method to generate nonparametric bounds on the conditional expectation of a child’s rank (for example, in the permanent income distribution) given the parent’s rank. This conditional expectation function (the CEF henceforth) is also referred to as the absolute mobility function. We use notation from Chetty et al. (2014), and describe the expected rank of the child of an \( i \)-th percentile parent as \( p_i \). In this notation, the CEF is a function satisfying the following:

\[
CEF(i) : I \rightarrow [0, 100] \\
i \mapsto p_i
\]

where \( I \) is the set of parent ranks, and child ranks are any real number in \([0, 100]\).

Panel A of Figure 1 demonstrates the problem that we are faced with, using data from Indian men born between 1973 and 1982. Parent education is observed in 7 bins, representing the highest level of schooling attained by each father. The vertical lines show the bin boundaries. The bottom bin is the largest, comprising the 46% of fathers who reported zero
education. The points in the graph indicate the mean rank of sons in the education distribution, conditional on being born into each father rank bin, or \( \hat{r}_k \); these are the empirical moments that a model must fit. The two functions drawn in Panel A show two possible nonparametric CEFs that each *perfectly* fit the empirical moments. Note that neither of these functions passes directly through the point at (23, 37), but the mean value in the first bin is 37 for both functions. There is also a valid CEF that goes through (23, 37) but we have deliberately chosen functions that do not, to highlight that passing through this point is not required. In contrast, the conventional linear approach implicitly assumes that the true CEF passes through (23, 37). Our goal is to describe the set of all CEFs that meet these moment restrictions under some set of assumptions.

We summarize the problem as the calculation of bounds on 100 discrete levels of the expected child rank given parent rank. We work with the discrete problem of calculating the CEF at each integer parent rank for computational tractability; given monotonicity, the discrete CEF will be a very close approximation of the continuous CEF. We write these expected child outcomes as \( p_1, p_2, p_3, \ldots, p_{100} \), the vector of which we denote \( \mathbf{p} \). The set of parent ranks \( I \) is the set of integers from 1 to 100. If the parent rank distribution is uncensored and fully supported, then the sample analog to \( p_i \) is just the mean rank of children born to parents at the \( i \)-th percentile. However, if the parent distribution is interval censored, such that child ranks are only observed in a small number of parent rank bins, there may be a set of vectors \( \mathbf{\hat{p}} \) that are consistent with the observed values.

We assume that the researcher observes parent outcomes in \( K \) bins, indexed by \( k \). Bin \( k \) contains parents ranked from \( b_k^- \) to \( b_k^+ \). Its mean and median both are denoted by \( \bar{b}_k = \frac{b_k^- + b_k^+}{2} \), owing to the uniformity of the parent rank distribution. We write \( \mathbf{b} \) to denote the vector of mean parent rank in each bin, \( \bar{b}_1, \bar{b}_2, \bar{b}_3, \ldots, \bar{b}_K \).

The researcher also observes the mean child rank in each parent bin \( k \), which we denote \( \bar{r}_k \). The value \( \bar{r}_k \) is the sample analog of the expected child rank, given a parent in bin \( k \), or
the average value of the CEF in this range:

\[ \bar{\tau}_k = \mathbb{E}(p_i| i \in (b^-_k, b^-_k + 1, b^-_k + 2, \ldots, b^+_k)) = \frac{1}{b^+_k - b^-_k} \sum_{i=b^-_k}^{b^+_k} p_i. \]

The values \( b \) and \( r \), both of which are comprised of \( K \) elements, are the empirical moments available to the researcher. The points in Figure 1A thus represent \((\bar{\tau}_k, \bar{b}_k)\) pairs; \( \bar{\tau}_1 = 37 \) and \( \bar{b}_1 = 23 \).

Now, let the mean child rank in bin \( k \) for any vector \( \hat{p} \) over the support of the parent ranks 1 to 100 be denoted by \( \hat{\tau}^k \). A conditional expectation function \( \hat{p} = \hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{100} \) is said to be \textit{MSE-minimizing} under a set of constraints \( \{C\} \) if it: i) minimizes the sum of the squared distance between the mean child rank within each bin \( k \) and the values in \( r \), and ii) the CEF satisfies all the constraints. Formally, a CEF \( \hat{p} \) is MSE-minimizing under the set of constraints \( \{C\} \) if the function \( \hat{p} \) satisfies \( \{C\} \) and:

\[
\hat{p} = \arg\min_{\mathbf{z} \in [0,100]^{100}} \left[ \sum_{k=1}^{K} \left( \left( \frac{1}{b^+_k - b^-_k} \sum_{i=b^-_k}^{b^+_k} z_i \right) - \bar{\tau}_k \right)^2 \right] \tag{1}
\]

\[
= \arg\min_{\mathbf{z} \in [0,100]^{100}} \left[ \sum_{k=1}^{K} (z^k - \bar{\tau}_k)^2 \right] \tag{2}
\]

where \([0,100]^{100}\) represents the set of all 100-element real vectors where each element lies in \([0,100]\). Note that in the expression above, \( \mathbf{z} \) is taken to be a candidate vector, and \( z^k \) is its within-bin mean in bin \( k \). \( \hat{p} \) is thus any vector that minimizes MSE distance between each of its bin means \( \hat{\tau}^k \) and the observed conditional mean within that bin, \( \bar{\tau}_k \). There may be many MSE-minimizing CEFs; we seek the set of all MSE-minimizing CEFs, denoted by \( P = \{\hat{p}\} \). With reference to Figure 1A, a MSE-minimizing CEF minimizes the error between the points and the mean value of the candidate CEF in each bin.

Researchers in this setting often fit models directly to the pairs \((\bar{\tau}_k, \bar{b}_k)\). This implicitly
assumes that the average of the CEF for children within a bin is equal to the value of the CEF at the bin midpoint $\bar{b}_k$, that is, they assume that:

$$\hat{r}_k = p_{\bar{b}_k}$$

There exists $\hat{p} \in P$ such that this assumption is valid, for example, if the CEF is linear in each bin. But it is trivial that for any set of moment conditions $\{b, r\}$, there exists $\hat{p} \in P$ such that $\hat{r}_k \neq \hat{p}_{\bar{b}_k}$; Figure 1A shows two such examples. By Jensen’s inequality, if the CEF is convex in bin $k$, $p_{\bar{b}_k}$ must lie below the average rank for children whose parents are in bin $k$; conversely, concavity implies $p_{\bar{b}_k} > r_k$. In settings where bins are large and the CEF is nonlinear, the value at $p_{\bar{b}_k}$ could differ substantially from the average rank within bin $k$.

Many of the major intergenerational mobility statistics can be expressed as functions of the CEF. Given the set of MSE-minimizing CEFs, we can calculate the set of feasible values for any mobility statistic; we will show results for the rank-rank elasticity and absolute mobility at the 25th percentile, but the method is more broadly applicable.

Panel B of Figure 1 provides an illustration. The solid line depicts the predicted values from the frequently estimated parent-child rank elasticity on the same data as in Panel A. These values are generated by a regression on grouped data with the values $\{(r_k, \bar{b}_k), k \in \{1, ..., K\}\}$. The dashed lines in the figure show predicted values from estimating the rank-rank regression on the two nonparametric CEFs presented in Panel A. The conventional estimation implies a unique MSE-minimizing elasticity that best fits the data — but in fact all three sets of predicted values in Panel B are MSE-minimizing linear approximations to CEFs that perfectly fit the seven moment conditions. The conventional elasticity estimation incorrectly conveys a single solution because of its interpretation of the moments as $p_{\bar{b}_k}$ rather than

\[\text{We weight each observation by the number of ranks in each bin, so that each rank in the latent distribution is equally weighted.}\]

\[\text{In passing, it is clear in this case that the assumption of linearity that underlies the elasticity estimation is not warranted by the evidently convex data.}\]
$p^k$, i.e., it assumes $p_{23} = 37$, whereas the correct interpretation of this moment condition is:

$$E(p_i | i \in (1, 2, ..., b_1^+)) = \frac{1}{b_1^+ - b_1^-} \sum_{b_1^-}^{b_1^+} p_i = 37.$$  

IV.B Calculating Bounds on the Nonparametric Parent-Child Rank Distribution

There is an infinite set of CEFs that fit the empirical moments $r$ and $b$ perfectly. In this section, we show how we can use a minimal set of structural assumptions (monotonicity and a second-derivative constraint) to bound the value of the CEF at every value in the parent rank distribution.

Our goal is to estimate numerical bounds on the set of MSE-minimizing vectors $P$. We begin by calculating bounds on each element of $P$. That is, for every integer $i$, we wish to obtain the largest and smallest $\hat{p}_i$ such that $\hat{p}_i$ is the $i$-th element of some vector $\hat{p} \in P$. Put formally, let $P_i$ represent the set of all elements $\hat{p}_i$ such that $\hat{p}_i$ is the $i$-th element of some MSE-minimizing vector $\hat{p} \in P$. Then we seek the bounds $\hat{p}_i^{\text{min}} = \inf \{ P_i \}$ and $\hat{p}_i^{\text{max}} = \sup \{ P_i \}$.

We make two structural assumptions. The first states that the child’s expected rank is a monotonically increasing function of the parent’s rank:

$$\hat{p}_i > \hat{p}_{i-1} \quad \text{(Assumption 1: Monotonicity)}$$

In other words, a child born to a higher ranked parent will on average attain an equal or higher rank in the outcome distribution.\footnote{Estimating a non-linear function to the group means would not solve the underlying problem of interval censoring. A quadratic function fits the bin means very closely, but generates estimates that are biased upward due to Jensen’s Inequality: in this case, the average value of the quadratic function in each bin will be larger than the estimated bin mean.}

We also assume that the curvature of the CEF may be constrained. We formalize this as

\footnote{Even a system that obtains perfect equality of opportunity will have this characteristic, and it is observed in every empirical observation of the parent-child income and education distributions that we are aware of.}
a constraint on the absolute magnitude on how quickly the slope of the CEF can change:

\[ |(\hat{p}_i - \hat{p}_{i-1}) - (\hat{p}_{i-1} - \hat{p}_{i-2})| < C \quad \text{(Assumption 2: Bound on Second Derivative)} \]

This assumption is analogous to imposing that the first derivative is Lipschitz.\textsuperscript{25,26} Depending on the value of $C$, this constraint may or may not bind.

This parameterization of the curvature constraint has the desirable property that its limits generate two useful scenarios. If the threshold $C$ is set very high, the constraint does not bind and we recover the unrestricted bounds on the CEF. If $C$ is set to zero, the MSE-minimizing $\hat{p}$ vector will correspond to the set of predicted values from the parent-child rank regression, the slope of which is the widely used parent-child rank elasticity. Our nonparametric approach thus generalizes the widely used linear rank-rank elasticity.

The CEF bounds can be calculated through a two-step process. First, we construct the set $P$ of all MSE-minimizing 100-element vectors $\hat{p}$:

\[
P = \left\{ \arg\min_{z \in [0,100]} \left[ \sum_{k=1}^{K} (z_k - \bar{r}_k)^2 \right] \right\} \quad (4)
\]

such that:

\begin{align*}
    z_j &> z_{j-1}, \quad \text{(C1)} \\
    (z_j - z_{j-1}) - (z_{j-1} - z_{j-2}) &< C. \quad \text{(C2)}
\end{align*}

\textsuperscript{25}Because we are working with CEFs over discrete values, we use the terms “first derivative” and “slope” to mean the slope of the line connecting the CEF when evaluated at parent rank $i$ and $i + 1$. We use the term “second derivative at $i$” to describe the change in the slope from the line segment $i - 1$ to $i$ to the line segment from $i$ to $i + 1$.

\textsuperscript{26}Let $X$ be a metric space with metric $d_X$. A function $f : X \rightarrow X$ is said to be Lipschitz continuous if there exists $K \in \mathbb{R}$ such that for all $x_1, x_2 \in X$,

\[ d_X(f(x_1), f(x_2)) \leq K d_X(x_1, x_2). \]

Under our assumption, there is some constant $\overline{C}$ such that the second derivative is less than $\overline{C}$. Because the CEFs are not continuous, it is not strictly accurate to say that we impose that the first derivative of the CEF is Lipschitz. But our assumption is similar — we assume that there is a bound on how quickly the slopes can grow.
The bounds on each element are then given by:

\[
\hat{p}_i^{\text{min}} = \inf \{ \hat{p}_i \mid \hat{p}_i \in P_i \}
\]

\[
\hat{p}_i^{\text{max}} = \sup \{ \hat{p}_i \mid \hat{p}_i \in P_i \}
\]

(5)

To solve this problem numerically, we require a slightly different framing, because the set \( P \) has an infinite number of elements and thus cannot be calculated. First, we use a numerical optimization tool to calculate the minimum MSE from Equation 5. We then run a second set of optimizations to calculate the minimum and maximum value for each \( \hat{p}_i \). The second optimization is given by Equation 6 with the additional constraint that the MSE is at most the minimum value obtained in the first step:

\[
\hat{p}_i^{\text{min}} = \min \limits_{z \in [0,100]} \min_{100} z_i
\]

such that

\[
z_j > z_{j-1},
\]

\[
(z_j - z_{j-1}) - (z_{j-1} - z_{j-2}) < C,
\]

\[
\left[ \sum_{k=1}^{K} (z^k - \tau_k)^2 \right] \leq \text{MSE}
\]

(6)

(7)

(8)

(9)

We run this second optimization 200 times, to calculate the minimum and maximum (replacing min with max) for each of the 100 elements of \( \hat{p}_i \).

IV.C Application to Sample Data

In this section, we explain the process of calculating nonparametric bounds on the parent-child rank distribution using sample data from India, continuing with the cohort of sons born between 1973 and 1982.

To calculate the unrestricted bounds, we run the optimization above without the curva-
ture constraint. With no restriction on the curvature of the CEF, there are many solutions that perfectly fit the moments, so the MSE of all candidate functions is zero. Figure 2 Panel A shows the full set of upper and lower bounds for each value $p_i$. To reiterate, these bounds do not represent uncertainty in the measurement of the moments: all of the CEFs underlying these bounds precisely match the empirical moments, and are thus equally valid as estimates of the latent CEF, even if the moments are very precisely estimated. We discuss the calculation of standard errors for these bounds in Section IV.D.

Even with just the assumption of monotonicity, this figure shows that we can obtain meaningful bounds on the father-son CEF in parts of the distribution. Because a minority of parents in this cohort attained higher levels of education, the rank bins are smaller and the bounds on the child CEF are considerably tighter at the top of the distribution. Under just the assumption of monotonicity, we can infer that the son born to the 25th percentile father attains an expected rank between 27 and 48, or $\hat{p}_{25} \in (27, 48)$. The child born to the 75th percentile parent obtains an expected rank between 53 and 64, or $\hat{p}_{75} \in (53, 64)$. For reference, $p_{25} = 46$ and $p_{75} = 55$ using Danish data on incomes. The bound on $p_{25}$ is considerably wider, owing to the wide bottom parent rank bin.

The bounds on the individual elements $p_i$ are codependent. Thus the full CEF cannot take the form of the lower envelope in Panel A of Figure 2. To clarify this point, in Panel B of Figure 2 we graph a pair of CEFs that respectively minimize and maximize $p_{25}$, which are used to bound the value of $p_{25}$ in Panel A. Because the CEF is constrained by the mean son outcome in the bottom bin, a low value of $p_{25}$ implies a high value of $p_1$, and vice versa. Panel B also highlights that the curvature of the CEF implied by either the upper or lower bound on $p_{25}$ is extremely sharp; as we will see below, it is sharper than that seen in any of the fully supported rank distributions from other countries, motivating estimations with stricter curvature restrictions.

\footnote{Numerically, we do this by setting $C$ to a very large value.}

\footnote{Note that there are multiple functions that respectively minimize and maximize $p_{25}$, and this figure only shows one such pair of functions.}
We next examine how the bounds on the CEF change as we tighten the curvature restriction. The four panels of Figure 3 show the bounds on the CEF with the magnitude of the 2nd derivative progressively constrained, with the limit $\overline{C} = 0$ in Panel D. In Panels A and B, we see that the precision of the predicted CEF remains highest when the bin size is very small and the function is bounded on both sides. This will be true in every developing country where the size of the bottom education bin is very large. The bounds appear smaller in Panels C and D, but this apparent precision comes at the cost of bias. Panel B is the last figure where the MSE of the moment function is zero; in Panels C and D, the parameterization of the CEF makes it impossible to meet all the moment conditions. The narrow bounds in the latter two figures come primarily from the restriction on the functional form of the CEF. In the linear case (Panel D), the curvature restriction generates a CEF that is clearly a poor fit to the data. We present an alternate representation of these results in Appendix Figure A3, which shows the bounds on $p_{10}$ and $p_{35}$ as a function of the curvature constraint.

While it may be desirable to summarize the parent-child distribution with a scalar statistic, the assumption of linearity in the latent CEF is unnecessary. Given that there are multiple MSE-minimizing CEFs (such as the two shown in Figure 1A), we propose instead that the linear parent-child estimator be calculated and bounded from the full set of feasible CEFs. We can calculate bounds on the linear rank-rank elasticity (or indeed any mobility statistic) by recognizing that the elasticity is a single-valued function of the CEF, which we call $\hat{\beta}(\hat{p})$. Since each candidate CEF in the set $P$ corresponds to an elasticity, the elasticity is bounded by the minimum and maximum elasticity in the set. We calculate this numerically as above, by first calculating the minimum MSE under a given set of constraints, and then minimizing and maximizing $\hat{\beta}(\hat{p})$ subject to the constraint set and an additional constraint that the MSE not be above the value calculated in the first step.

Figure 4 shows the estimated upper and lower bounds on the father-son rank elasticity under a range of curvature restrictions. The unrestricted bounds range from 0.48 to 0.60. The naive elasticity of 0.54 falls in the middle of this range, but unlike the bounds, does not
reflect the inherent uncertainty about the shape of the CEF within the large bins where it is not precisely observed.

IV.D Bootstrapping Standard Errors on the Bounds

The bounds thus far have been presented as though the average child rank in each bin is known with certainty, but in fact these means are measured with sampling error. We can use bootstrapping to obtain confidence intervals on the CEF bounds that reflect uncertainty in measurement of the bin means.

To do this, we draw 100 bootstrap samples from the underlying datasets. Figure 5 shows the bootstrap 95% confidence interval around the upper and lower bound estimates of the Indian father-son CEF analogous to Figure 3B. The confidence intervals are tight because the census dataset from which these father and son education values are drawn is very large. In settings where the empirical distribution of parent and child education are based on smaller samples or with more underlying variation, measurement uncertainty may be a more important factor.

IV.E Choosing the Curvature Parameter

In all of these applications, the width of the bounds is heavily dependent on the curvature threshold \( C \). Without data on the latent rank distribution, it is difficult to obtain an estimate of an appropriate curvature restriction from the sample being studied. To provide an empirical basis for a reasonable curvature threshold, we therefore examined as many studies as we could find that report uncensored parent-child outcome rank distributions. Where possible, we obtained the underlying data, and where we could not obtain the underlying data we estimated the rank-rank matrix by digitizing the underlying data from the paper’s graphics. In each of these studies, we obtained nonparametric fits to the data using 5-piece

---

29Because the IHDS is weighted, we take weighted random draws that take into account the higher representation of some observations in the population. The bootstrap sample can then be treated as an unweighted sample from the underlying dataset. Because the total value of weights is 10 million, we do not have sufficient computational power to calculate bootstrap samples with the same size as the underlying sample. We chose a bootstrap sample size of 10,000 which allowed us to generate 100 bootstrap samples in about 24 hours of computation.
cubic splines with equally spaced knots.\textsuperscript{30} We calculated the maximum value of the 2nd derivative in each of these studies.

Figure 6 shows the splines that we fitted to each dataset and the corresponding range of the 2nd-derivative in each of these functions. Given that many of these studies exhibited significantly more curvature in the tails of the distribution, we also report a second set of maxima from the 10th to the 90th percentile of these distributions. The magnitude of the 2nd derivative is highest in Swedish data, at 0.067 in the full sample and 0.046 in the [10,90] interval. Given that all these estimates come from developed countries (because uncensored paired parent-child ranks are generally unavailable in developing countries), as a conservative rule of thumb, we propose setting a curvature limit to twice these upper limits. In that case, we obtain that $C = 0.12$ for the full sample and $C = 0.092$ in the [10,90] interval. Panel B in Figure 3 corresponds to the midpoint of these two estimates; this is the constraint that we use going forward, recognizing that it may underestimate the size of the bounds at the tails of the distribution.

\textbf{IV.F Simulation: Bounds on Mobility in Denmark}

In this section, we simulate our method by taking data from the fully supported child CEF from the Danish income distribution, interval-censoring that data, and then recovering bounds on the true CEF from that interval-censored data. We begin with data reporting the average child income rank for every integer father rank in 1–100. We divide the data into seven bins that approximate the bin sizes in the Indian data used above. We then calculate bounds on the CEF using only the binned data.

Panels A–D of Figure 7 show the result of this exercise, with curvature limits corresponding to Panels A–D of Figure 3. The gray diamonds show the underlying uncensored child ranks at each parent rank and the solid black circles show the constructed bin means of the censored data which are the empirical moments for the calculation of bounds. The solid lines show the

\textsuperscript{30} Splines with more knots tended to overfit the data (for example, by predicting nonmonotonic segments) and splines with fewer knots often failed to account for curvature in the data, especially in the tails. These estimates are minimally affected by marginal changes in the number of knots.
upper and lower nonparametric bounds on the Danish CEF. As above, the bounds are tighter at the top of the distribution where we have artificially imposed smaller bins, and wider at the bottom. Compared with the Indian exercise, the bounds are tighter across the entire distribution because of the higher level of mobility in Denmark: since the expected ranks of sons are more similar across fathers ranks, the monotonicity constraint significantly constrains the shape of the CEF. In Panels A and B, the true data are almost entirely contained by the bounds. It is important to note that the true data is not always centered by the bounds; from $p_{30}$ to $p_{40}$ the true CEF is near the upper bound, and from $p_{1}$ to $p_{10}$ it is nearer the lower bound. Under the tighter curvature constraint of Panel C, we begin to see errors in the tails, and the linear Panel D significantly overestimates mobility at the bottom of the distribution.

Almost any parametric best fit to the censored data would miss the concavity at the lower tail of the distribution. The strength of our method is that it makes clear the uncertainty about the CEF in the domain where the interval censoring is most severe. The severity of the bias of parametric methods is increasing in the non-linearity of the underlying CEF. The Danish example may thus understate the bias if the true CEF in developing countries has greater curvature.

V Applications: Changes in Intergenerational Mobility in India

In this section, we apply the methodology above to the Indian data to answer two questions. First, how has intergenerational educational mobility changed over the last forty years? Second, is there a mobility disadvantage for members of India’s historically marginalized groups, the Scheduled Castes and Scheduled Tribes, and has this disadvantage changed over time?

Because estimates of educational ranks are interval-censored in India, we present all mobility estimates as nonparametric bounds. Since these bounds depend on the curvature restriction, we will present three sets of estimates for each result: (i) our preferred estimates based on the curvature restriction proposed in Section IV.E, or $\overline{C} = 0.12$; (ii) a less conservative restriction of $\overline{C} = 0.06$; and (iii) a set of linear estimates, where $\overline{C} = 0$. We present less rather than more conservative estimates as they have been more widely used in the literature.
and allow us to relate our preferred estimates to those obtained by previous researchers.

**V.A Indian Intergenerational Mobility Over Time**

To examine intergenerational mobility over time in India, we compare the joint distribution of father and son ranks for an older cohort, born between 1953 and 1962 and a younger cohort born between 1983 and 1992; we refer to these as the 1950s and 1980s cohorts respectively.

Panel A of Figure 8 displays the raw data that are available to the researcher: the conditional mean of son ranks in the education distribution, as a function of father ranks. Since the fathers’ education ranks are interval-censored, we plot father rank at the median value in each interval where education is reported. The dashed vertical line shows the bin boundary for the least educated fathers in the old cohort; 59% of fathers are in this bin. The solid line shows the equivalent bin boundary for fathers in the young cohort; 36% of them report fewer than two years of education.

The remaining three panels of Figure 8 show bounds on the father-son rank distribution for both cohorts under the three curvature restrictions. Panel B shows our preferred bounds estimate, parameterized with the constraint described in Section IV.E. Panel C shows less conservative bounds and Panel D shows the fully restricted linear estimate. We focus first on our preferred estimates in Panel B. The very large share of fathers with zero education makes it difficult to describe changes in intergenerational mobility at the bottom of the distribution. The bounds on the older generation are largely uninformative, and the tighter CEF bounds of the young cohort are almost fully contained in the bounds of the older cohort. We can rule out neither zero nor large changes in mobility in the lower half of the distribution. Our bounds on absolute mobility at the 25th percentile $\hat{p}_{25}$ are (34, 42) for the older cohort and (28, 45) for the younger cohort. However, we can bound the father-son rank CEF more precisely at the top of the distribution due to smaller education rank bins. We detect a

---

\[^{31}\]Appendix Figures A4 and A5 show similar estimates with an old cohort defined respectively from 1963 to 1972, and from 1973 to 1982. The results presented here do not change substantially when considering the intermediate cohorts; we present only the end-points of our sample for graph legibility.

\[^{32}\]For legibility, we omit bootstrap standard errors. As demonstrated in Figure 5, these imply a slight widening of the bounds across the distribution and do not change our conclusions.
small increase in mobility at the very top of the distribution. Absolute mobility at the 95th percentile declined from (81,82) to (76,77), implying a gain in mobility: children of parents at the top of the distribution can now expect outcomes closer to the median rank.

Because the bins are small at the top of the distribution (reflecting that only 20% of even the youngest cohort of Indians advance beyond middle school) and large at the bottom, marginal changes in the curvature parameters in fact make little difference to the precision of these conclusions. In Panel C, we employ a tighter curvature restriction, allowing half as much curvature. The conclusions about $\hat{p}_{25}$ and $\hat{p}_{95}$ are sustained.\textsuperscript{33}

Panel D shows predicted values from the linear estimates (\textit{i.e.} $C = 0$), which would be obtained by the naive calculation of the father-son rank elasticity. These estimates suggest an unambiguous increase in mobility: the slope of the father-son linear CEF has clearly fallen. Because the estimated relationship is constrained to be linear, no distinction is made between changes at the top of the distribution and changes at the bottom. The graph thus understates the change in mobility at the top of the distribution (showing $\hat{p}_{95}$ falling from 76 to 73), and significantly overstates the precision of mobility estimates at the bottom of the distribution (showing $\hat{p}_{25}$ rising from 35 to 37). Also note that the predicted values for $\hat{p}_{95}$ for both young and old cohorts fall outside of the feasible bounds of our preferred curvature restriction. The imposition of linearity necessitates a CEF with significant bias at the top of the distribution.

We can conclude from these estimates that mobility at the very top of the distribution has improved. This represents a reduction in the reproduction of elites; the ordering of ranks in the top 10% of the distribution is less persistent than it was in the past. However, these estimates show that detecting all but the largest mobility changes at the bottom of the distribution is difficult given the significant interval censoring that we observe.

\textsuperscript{33}At lower levels of the parent rank distribution, the CEF of the young cohort is bounded near the top of the bounds on the old cohort. However, the bounds are still overlapping. The empirical CEFs observed in Figure 6 suggest that curvature may be greatest at the tails, making us hesitant to draw conclusions about the tails of the Indian distribution using a more restrictive $C$. 

25
V.B Intergenerational Mobility of Scheduled Castes and Scheduled Tribes

We next examine the extent to which we can separately measure intergenerational mobility differences between the general population and India’s historically marginalized groups, the scheduled castes and tribes (SC/STs).

Figure 9 shows estimates of the bounds on the absolute mobility distribution, calculated separately for generals and SC/STs, using our preferred curvature restrictions. Panel A shows bounds on estimates of the absolute mobility CEF for the oldest cohort, 1953-1962, calculated separately for generals and for SC/STs. For both groups, we rank both fathers and sons according to their position in the national education distribution, because we are interested in the ability of SC/STs to move up in the national distribution, not to move up only relative to other SC/STs. Panel B of the figure shows equivalent estimates for the youngest cohort, born between 1983 and 1992. As with the previous figure, we obtain tighter bounds at the top of the rank distribution than at the bottom for both cohorts. The bounds at the bottom are tighter for the young cohorts than for the older cohorts, owing to the smaller share of zero education fathers in the young cohort.

We focus first on the top half of the distribution, where estimates are considerably more precise. We draw three conclusions from these estimates. First, among older cohorts, there is a significant caste disadvantage in terms of absolute mobility in the top half of the distribution. Sons of SC/STs at the same rank in the parent distribution can be expected to end up eight to ten percentiles lower than sons of generals. The bounds are tight and do not overlap (nor do the bootstrapped bounds, omitted for legibility), allowing us to make this statement with confidence. Over the course of the next thirty years, this SC/ST disadvantage diminishes substantially but is not totally erased. In the younger cohort, sons of SC/STs face a three to five percentage point disadvantage compared to sons in the general population with equally educated fathers. In both the SC/ST and the general distribution, we can also see the flattening of the CEF at the highest parent ranks, reflecting the decreased persistence of elite outcomes.
In the bottom half of the distribution, the results are less clear. For both young and old cohorts, the bounds on SC/STs and generals are overlapping, with both upper and lower bounds lower for SC/STs. For the 1950s cohort, \( \hat{p}_{25} \) is in the interval \((24, 35)\) for SC/STs, and \((34, 45)\). For the 1980s cohort, \( \hat{p}_{25} \) is bounded by \((32, 37)\) for SC/STs and \((36, 41)\). Recall that these bounds all represent values of \( p_{25} \) from CEFs that fit the moments perfectly; it is thus feasible that the true value \( \hat{p}_{25} \) is identical for SC/STs and generals, with all moments perfectly satisfied.

How can we compare these two bounded estimates? While lowering \( \tilde{C} \) would yield tighter bounds, it would imply a curvature restriction that is excessively tight when compared with data from other countries. Instead, we examine the distribution of the difference in estimates under the prior that \( p_{25} \) is uniformly distributed between the bounds.\(^{34}\) Under a uniform prior, the 1950s cohort SC/ST disadvantage (defined as \( p_{25, \text{general}} - p_{25, \text{SC/ST}} \)) is distributed according to \( U(34, 45) - U(24, 35) \) and the 1980s SC/ST disadvantage according to \( U(36, 41) - U(32, 37) \). We plot these two distributions in Figure 10. The solid line shows the estimate of the SC/ST absolute mobility disadvantage for the 1950s cohort, and the dashed line shows the same for the 1980s cohort. For each of these density functions, we can calculate the CDF of the distribution at zero, which is analogous to a p-value from a one-tailed test of the null hypothesis that the SC/ST disadvantage is zero. This value is 0.01 for the older cohort and 0.02 for the younger. Under the uniform prior, we can thus say with 95% confidence that SC/STs face a mobility disadvantage in both periods. We can run an analogous test of whether the SC/ST disadvantage has shrunk, as suggested by the point estimates in Figure 10. The p-value for this test is 0.13, indicating that we can have less confidence in this claim.\(^{35}\) These p-values are only valid under the uniform prior; other

\(^{34}\)Other prior distributions could be justified based on the expected convexity or concavity of the CEF around \( p_{25} \). An assumption of a linear latent CEF would suggest a prior that weights the midpoint of the bounds, but linearity is rejected both by the observed moments and data from other countries.

\(^{35}\)Figure A6 shows the same sets of CEF bounds, but paired by population group rather than by cohort. We can see from these graphs that any convergence at the bottom of the distribution has come from gains for SC/STs rather than losses for the general population, but the bounds are too wide to make this claim decisively.
prior distributions would imply different distributions of the SC/ST disadvantage.

VI Conclusion

Education is a useful proxy for social rank, as it is easy to measure, data are widely available, and it is not subject to life cycle bias. The interval censoring that is inherent to the measurement of educational outcomes can affect estimates of intergenerational mobility, but nonparametric methods can be used to recover bounds on the fully supported latent rank distribution. This method offers a way forward for the study of intergenerational mobility.

Using data from India, we find that mobility estimates are more precise at the top of the rank distribution, and more precise for recent cohorts, results which are driven by the large number of parents sharing the bottom rung of the education distribution. These findings are likely to be replicated in other developing countries, nearly all of which have low levels of education among older cohorts.

Our ability to examine differences across groups or changes across time from bounds on the conditional expectation function of child rank given parent rank depends to some extent on the prior knowledge that we have on the shape of the CEF in the most interval-censored parts of the distribution. These priors could be informed by studies examining the rank distributions in developing countries for variables which are not as severely interval-censored as education. Improving our understanding of the relationship between parent and child outcomes at the bottom of the distribution in developing countries is essential to the intergenerational mobility research agenda.
References


Munshi, Kaivan and Mark Rosenzweig, “Traditional institutions meet the modern world:


<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Birth Cohort of Son</th>
<th>Number of Parent Education Bins</th>
<th>Population Share in Bottom Bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aydemir and Yazici (2016)</td>
<td>Turkey</td>
<td>1990(^{36})</td>
<td>15</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>Turkey</td>
<td>1960(^{37})</td>
<td>15</td>
<td>78%</td>
</tr>
<tr>
<td>Hilger (2015)</td>
<td>USA, White</td>
<td>1911–1914</td>
<td>&gt; 15</td>
<td>6.1%</td>
</tr>
<tr>
<td></td>
<td>USA, Black</td>
<td>1911–1914</td>
<td>&gt; 15</td>
<td>11.9%</td>
</tr>
<tr>
<td></td>
<td>USA, White</td>
<td>1971–1974</td>
<td>&gt; 15</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>USA, Black</td>
<td>1971–1974</td>
<td>&gt; 15</td>
<td>0.7%</td>
</tr>
<tr>
<td>Hnatkovska et al. (2013)</td>
<td>India</td>
<td>1918 – 1988</td>
<td>5</td>
<td>Not reported</td>
</tr>
<tr>
<td>Knight et al. (2011)</td>
<td>China</td>
<td>1930–1984</td>
<td>5</td>
<td>29%(^{39})</td>
</tr>
</tbody>
</table>

\(^{36}\)Includes all people born after about 1990.

\(^{37}\)Includes all people born after about 1960.

\(^{38}\)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.

\(^{39}\)This reported estimate does not incorporate sampling weights; estimates with weights are not reported.
Figure 1
Joint Distribution of Father-Son Education Ranks and Sample Conditional Expectation Functions

Panel A

Panel B

The graphs show sample conditional expectation functions of child education ranks given parent education ranks. The sample is Indian son born between 1973-82 and their fathers. The circles represent the mean ranks of fathers and sons in each fathers’ bin in the education distribution. The vertical lines show the boundaries between parent education rank bins. In Panel A, the solid lines show two different conditional expectation functions that match the moments (the bin means) perfectly. Panel B shows linear approximations to the two functions in Panel A.
Figure 2
Conditional Expectation Function Bounds of Son Rank Given Father Rank

Panel A

Panel B
Figure 3
Father-Son Rank Conditional Expectation Function Bounds
Under Curvature Constraints
Figure 4
Bounds on Father-Son Rank Elasticity Under Different Curvature Constraints
Figure 5
Father-Son
Rank Conditional Expectation Functions Under Curvature Constraints: Bootstrap Standard Errors
Figure 6
Spline Approximations to Empirical Rank Distributions in the Literature

Panel A: U.S.A.

Panel B: Denmark

Panel C: Sweden

Panel D: Norway
Figure 7
Nonparametric Bounds using Danish Mobility Data
Figure 8
Changes in Intergenerational Educational Mobility from 1953-1962 Cohort to 1983-1992 Cohort

Panel A

Panel B

Panel C

Panel D
**Figure 9**

Intergenerational Educational Mobility: Schedule Castes/Tribes vs. General Population

**Panel A**

**Panel B**
Figure 10
General vs. SC/ST gap in absolute mobility at 25th percentile
Figure A1 presents the proportion of sons who coreside with their fathers by birth cohort. We define coresidence as a variable that encodes whether we can match the sons to fathers; for some sons who inhabit the same household as their father, we are unable to match the sons and fathers’ educations due to ambiguity. Shaded areas plot 95% confidence intervals. Data source: Socioeconomic and Caste Census (2012).
Figure A2
Educational attainment by birth cohort and caste

The figure presents average years of education by birth cohort, stratified by members scheduled castes and tribes, and the general population, for all men in India. The figure uses the combined SECC and IHDS dataset, along with sampling weights, as described in section III. Birth cohorts are grouped into 10 year bins, where the birth cohort coordinates plotted at \{1957,1967,1977,1987\} correspond to the years \{1953-1962,1963-1972,1973-1982,1983-1992\} respectively. Shaded areas plot 95% confidence intervals.
Figure A3
Bounds on $p_{10}$ and $p_{25}$ Under Different Curvature Constraints
Figure A4
Changes in Intergenerational Educational Mobility
from 1963-1972 Cohort to 1983-1992 Cohort

Panel A

Panel B

Panel C

Panel D
Figure A5

Panel A

Panel B

Panel C

Panel D
Figure A6
Intergenerational Educational Mobility: Schedule Castes/Tribes vs. General Population

Panel A: Generals, young and old cohort

Panel B: SC/STs, young and old cohort