Partial Identification of Expectations with Interval Data*

Sam Asher†

Paul Novosad‡

Charlie Rafkin§

November 2017

Abstract

A conditional expectation function (CEF) can at best be partially identified when the conditioning variable is interval censored. With a small number of bins, existing methods often yield minimally informative bounds. We prove novel nonparametric bounds for contexts where the distribution of the censored variable is known. We next introduce a new class of measures that describe the conditional mean across a fixed interval of the conditioning space. These measures can be bounded more tightly than the CEF itself (and in some cases point estimated), making meaningful inference possible even when the bounds on the CEF are very wide. We show the performance of the method in simulations and then present two applications. First, we resolve a known problem in the estimation of mortality as a function of education; current estimates of mortality of less educated women in the U.S. are known to be biased because high school educated women occupy a decreasing educational rank over time. Our method makes it possible to hold the rank bin constant, revealing that current estimates of rising mortality for women with high school or less are biased upward in some cases by a factor of three. The method is also applicable to the estimation of education gradients in fertility and marriage patterns, where similar compositional problems exist. Second, we apply the method to the estimation of intergenerational mobility: researchers frequently use coarsely measured education data in the many contexts where matched parent-child income data is unavailable. We show that conventional measures like the rank-rank correlation may be uninformative once interval censoring of educational rank is taken into account; our interval mean measures provide tight bounds on parameters of interest.

*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, Chris Snyder and Elie Tamer. Taewan Roh provided excellent research assistance. We are grateful to Bratberg et al. (2017) for sharing data. All errors are our own.

†World Bank, sasher@worldbank.org

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

§National Bureau of Economic Research, crafkin@nber.org

I Introduction

The value of a conditional expectation function (CEF) can at best be partially identified when the conditioning variable is interval censored (Manski and Tamer, 2002). When the number of observed intervals is small or the slope of the CEF changes very quickly, existing methods can yield bounds that are minimally informative. In this paper, we develop three innovations that can yield narrower bounds on parameters of interest, and apply them in two policy-relevant settings.

First, we show that using information on the distribution of the conditioning variable can give tighter CEF bounds. We prove sharp bounds on the conditional expectation of a variable given interval-censored data from a known latent distribution. The method is broadly applicable, because distributions are known or commonly assumed for many economic variables. For some conditioning variables (such as ranks), no additional assumptions are required because the rank distribution is uniform by construction. For others (such as income), distributional assumptions on the variable of interest are common and reasonable, and results under alternative assumptions can be tested.

Second, we derive a class of measures that describe the CEF mean across a fixed interval of the conditioning variable. We show that this conditional mean can be bounded at least as tightly as any point on the CEF for many of these intervals. This makes meaningful inference possible for policy-relevant parameters even when the bounds on the CEF itself are very wide.

Third, we show that imposing a Lipshitz condition on the curvature of the CEF can further improve bounds. Once a curvature constraint is imposed, the problem can no longer be solved analytically, but we show that bounds on the CEF or on any function of the CEF can be calculated as the solution to a numerical constrained optimization problem. This curvature constraint can also substitute for the requirement of monotonicity that is commonly required for inference with interval data.

Although we focus on conditional expectation functions, our method can be directly applied to any function or moment of the variable of interest. For example, we can bound any percentile of the conditional outcome distribution.

We validate our method and explore the properties of the conditional mean measures by simulating interval censoring in a context where the full distribution of the latent variable is observed. We then present two applications, focusing on two problems where researchers have struggled with inference due to interval-censored data.

Application 1: Mortality as a Function of Education

In our first application, we resolve a long-standing problem in the estimation of mortality as a function of education. Researchers have recently noted increases in the mortality of less-educated individuals in the U.S. (Meara et al., 2008; Cutler and Lleras-Muney, 2010b; Cutler et al., 2011; Olshansky et al., 2012; Case and Deaton, 2015; Case and Deaton, 2017). For example, mortality among women aged 50–54 with high school education or less has risen from 459 deaths per 100,000 people in 1992 to 587 deaths in 2015. A known concern with these estimates is that education levels are rising over time, particularly among women and minorities. For example, women with a high school degree or less represented 64% of women in 1992 and only 39% of women in 2015. Hence women with only a high school degree represent a smaller and more negatively selected share of the population over time (Bound et al., 2015). Figure 1 shows mortality for 50–54 year old U.S. women as a function of the median education rank in each of three educational categories, illustrating the simultaneous changes in mortality and in the distribution of education. In the bottom bin, mortality rises over time, but the median education rank declines. Estimates of mortality within fixed education quantiles would solve the problem, but there is no established method to generate quantiles when intervals in the data do not correspond to quantile boundaries.[1] Some studies have argued that the bias is likely to be zero, while others have suggested that it may explain all of the recent mortality increases.[2]

[1] Mortality data typically report education in a small number of coarse categories. Most studies on mortality and education use only two or three categories of education.

[2] Recent high profile work by Case and Deaton (2015, 2017) has focused on individuals with high school

To make progress on this problem, we make two assumptions. First, we assume that the observed education rank represents a latent, continuous rank that is observed only in coarse intervals, a common assumption in this literature (Goldring et al., 2016). Second, we assume that mortality is decreasing in latent educational rank. This assumption holds across bins for every year between 1992 and 2015, for every population group, and across every income ventile (Chetty et al., 2016). Under just these assumptions, we can calculate bounds on the expectation of mortality at any rank or within an arbitrary rank interval using our new method in set identification.

We focus on women age 50-54, because their increasing education over time makes estimation of mortality within any educational group difficult. We focus on estimating average mortality among the set of latent education ranks that correspond to women with no more than a high school education in 1992—the bottom 64% of the education distribution for women. In 1992, mortality for this set of women can be point estimated, as it is directly observed in the data. For other years, it can be bounded at best, but we show that the bounds are tight and informative. We establish that the mortality increase for women in this part of the education distribution is between 29 and 38 more deaths per 100,000. The unadjusted estimate (which compares the bottom 64% in 1992 to the bottom 39% in 2015) suggests an increase of 128 deaths, which is more than three times greater than the upper bound that we calculate.

For some population groups, the mortality increases noted in the recent literature are sustained once we focus on constant rank groups, while for others, the unadjusted estimates have the incorrect sign. For example, the unadjusted mortality of men ages 50–54 with less than college education has risen by 1.2%, but when we hold the rank bin constant at 1992 levels, we can bound the change between a 7.5% *decline* in mortality and a 0.1%

education or less, arguing that their average school completion has not substantially changed from 1990 to the present. This claim is defensible for men but less so for women. From 1992 to the present, high school educated men have gone from 54% to 44% of the population in 2015; women in the same education group have gone from 64% of the population to 39%. Changes are even larger for other age groups over other time periods.

increase for the group (men in percentiles 0-54 in the education distribution). Unadjusted estimates of women's mortality are considerably more biased upward than men's, because women's education has increased more than men's over the 24-year sample period, making the compositional changes larger. Finally, we confirm the middle-age increase in both men and women's mortality from suicide, poisoning and liver disease, termed "deaths of despair" by Case and Deaton (2017); compositional changes only marginally reduce these estimates for less educated people because i) these deaths were considerably more rare, and ii) the slope of the best linear approximation to the mortality-education rank for this class of death was small in 1992.

Similar compositional issues arise in any context where the researcher is interested in changes in the relationship between education and some outcome variable over time. Our method may be useful, for example, in the study of birth outcomes, marriage patterns or disability as a function of education (Cutler and Lleras-Muney, 2010a; Aizer and Currie, 2014; Bertrand et al., 2016).[3]

Application 2: Intergenerational Mobility

The methods presented here can also resolve several problems in the estimation of intergenerational educational mobility. Estimates of intergenerational mobility are most often derived from the conditional expectation of some child status measure as a function of some parent status measure. Researchers often use education as a measure of parent status, because data on educational attainment are widely available and may be less subject to measurement error than parent income data. Again assuming that education is a latent variable that is observed in coarse bins, we focus on an example from a developing country where the interval data problem is extreme. In older generations in India, over 50% of parents report less than two years of education; the dependence of child status on parent status, conditional on being in

[3] In a context where education is strictly considered as an input to the production function (such as estimating the returns to education), unadjusted estimates may be preferred. But if education is intended to proxy for socioeconomic status, adjusting for the population share and rank position of a given educational subgroup will prevent bias from changes in the distribution of education over time.

the bottom half of the parent distribution, is thus unobserved.

Many mobility measures are difficult to calculate if the researcher only has data on discrete educations, because these measures require knowledge of the CEF at specific points in the parent distribution. Absolute upward mobility is defined by Chetty et al. (2014b) as the expected outcome of a child who is born at the 25^{th} percentile of the parent rank distribution. Cross-group comparisons typically use the expected outcome of a child in a group, conditional on beginning in a given quantile of the national distribution (Hertz, 2005).[4]

With interval censoring, there is no established method to calculate an outcome in a given rank bin if the ranks of interest lie within the bins observed in the data. Some researchers have generated rank mobility measures by imposing CEF linearity, but this assumption is rejected by many empirical parent-child rank distributions. Our method makes it possible to bound these measures with only the assumption of monotonicity.

We propose a new measure of mobility, *upward interval mobility*, which is the expected outcome of a child born in the bottom half of the parent distribution. This method is of similar policy value to absolute upward mobility (and identical if the CEF is linear) but it can be tightly bounded in many contexts where absolute upward mobility cannot.[5]

We apply the method to the Indian data, and we show that the rank-rank gradient and absolute upward mobility measures are uninformative about the mobility of older cohorts once we account for interval censoring. In contrast, upward interval mobility can be tightly bounded, and we argue that it is likely to be tightly bounded in other educational mobility contexts which have similar censoring patterns. Using this measure, we show that upward mobility in India has been static since the 1950s. If the coarse data problem were ignored, the

[4]The rank-rank best linear approximation and other linear estimators of the parent-child outcome function are not informative about subgroup mobility, because they compare children of low-ranked parents with children of high-ranked parents from the same subgroup, which can be misleading (Aaronson and Mazumder, 2008). For example, the rank-rank gradient in a minority group may be lower than the rank-rank gradient for a majority population, but that only suggests that low-ranked minority children are mobile relative to high-ranked minority children.

[5]Absolute upward mobility measures the expected outcome of the median child in the bottom half of the parent distribution; upward interval mobility measures the mean expected outcome in the bottom half of the parent distribution.

naive approach to estimating the rank-rank gradient would suggest small but unambiguous mobility gains over the same period, which are in fact driven by changes in the CEF in the top half of the distribution.

The interval problem for educational mobility is most severe in developing countries, but is important in other contexts as well. Internationally comparable censuses frequently report education in as few as four categories. In wealthier countries, it is common for a large share of the population to be in a topcoded education bin.[6]

Our paper proceeds as follows. Section II describes our method for calculating nonparametric bounds on a conditional expectation function given an interval-censored conditioning variable. We first present analytical bounds on the CEF and on the mean of the CEF across an arbitrary interval. We then present the setup of a numerical optimization problem that allows the researcher to impose a curvature constraint, either in place of or in addition to the monotonicity condition. Section III presents results from simulated interval censored data, demonstrating circumstances where we can achieve tight and loose bounds. Section IV applies the method to the estimation of changing U.S. mortality, calculating mortality estimates for constant-rank education groups. Section V applies our method to the estimation of intergenerational educational mobility, and presents estimates of mobility changes in India from 1950 to the present. Section VI concludes. Stata and Matlab code to implement our methods are available on the corresponding author's web site.

II Bounds on CEFs with a Known Conditioning Distribution

This section describes the main contribution of the paper. We describe a method to calculate analytical bounds on a CEF when the conditioning variable is observed under interval censoring. The bounds are sharp and depend only on the assumption of a weakly monotonic

[6] In one mobility study from Sweden, for example, 40% of adoptive parents were topcoded with 15 or more years of education (Björklund et al., 2006). Studies on the persistence of occupation across generations also frequently use a small number of categories and face a similar challenge when the occupational structure changes significantly over time, as it has with farm work in the United States. See, for example, Long and Ferrie (2013), Xie and Killewald (2013) and Guest et al. (1989).

CEF and a known distribution for the conditioning variable. Imposing an additional restriction on the CEF curvature can yield tighter bounds but requires numerical optimization to solve. Given CEF bounds, we can also bound any statistic that can be derived from the CEF, including the best-fit linear approximator.

We describe the method by working through an example motivated by Figure 1. We wish to estimate some statistic that describes mortality in 1992 and 2015 for a group of people occupying a constant rank or set of ranks in the population. The challenge is that the rank bin boundaries change between 1992 and 2015. In 1992, 64% of women had less than or equal to a high school education, while in 2015, this number was 39%. To create comparable groups, we assume that the observed education rank represents a latent, continuous rank, and that mortality is weakly decreasing in latent rank. We focus on women aged 50–54, because i) this age group has been highlighted in other recent research, and ii) the change in education for this group has been large over the sample period.

Note that for consistency with the prior literature on interval data, we assume in the theory that that CEF is weakly monotonically increasing in the conditioning variable. The CEF of mortality risk is monotonically *decreasing* in education. We calculated results using survival rates (equal to one minus mortality rates), which are monotonically increasing, but we display graphs of mortality which are the object of interest.

II.A Nonparametric Inference with Interval Data

Define the outcome as y and the conditioning variable as i; the conditional expectation function is $Y(i) = E(y|i)$. Let the function $Y(i)$ be defined on $i \in [0, 100]$, and assume $Y(i)$ is integrable. We also assume throughout that $\underline{Y} \leq Y(i) \leq \overline{Y}$, that is, the function is bounded absolutely.[7]

With interval data, we do not observe i directly, but only that it lies in one of K bins.

[7]In most applications, parameters of interest are likely to have upper and lower bounds either in theory or in practice. Loosening the absolute upper and lower bound restriction would result in wider bounds for the CEF in the bottom or top intervals, but informative inference is still possible even in these outer bins. In the case of mortality, we will impose that the upper bound is a mortality rate of 100%.

Let $f_k(i)$ be the probability density function of i in bin k. Define the expected outcome in the k^{th} bin as

$$r_k = E\left(y|i \in [i_k, i_{k+1}]\right) = \int_{i_k}^{i_{k+1}} Y(i)f(i)di,$$

where i_k and i_{k+1} define the bin boundaries of bin k. Note that this expression holds due to the law of iterated expectations. The limits of the conditioning variable are assumed to be known, and are denoted by i_1 and i_{K+1}. Further define the expected outcomes in the intervals directly above and below the intervals of interest as $r_{k+1} = E\left(y|i \in [i_{k+1}, i_{k+2}]\right)$ and $r_{k-1} = E\left(y|i \in [i_{k-1}, i_k]\right)$, if they exist. Define $r_0 = \underline{Y}$ and $r_{K+1} = \overline{Y}$. We observe the sample analogs to these expressions in the data, and denote the observed mean outcome in bin k with $\overline{r}_k$.

Panel A of Figure 2 depicts the setup, using data on the total mortality of women ages 50–54 in the United States in 2015. We use the three education bins proposed by Case and Deaton (2015); the vertical lines show the rank bin boundaries. The bottom bin comprises 39% of women, all of whom report less than or equal to a high school education. The middle bin comprises women who report some college, and the third bin comprises women who report a BA or more. The points in the graph plot the average total mortality in each bin (deaths out of 100,000), plotted at the median rank in the bin. The lines plot two (of many) possible nonparametric CEFs, each of which fit the sample means with zero error. Note that both of these functions have the same mean in the first bin.[8]

Manski and Tamer (2002) derive bounds on $E(y|i)$ given interval measurement of i. The essential structural assumption that constrains the outcome is Monotonicity (M):

$E(y|i)$ must be either weakly increasing or weakly decreasing in i. (Assumption M)

In our context, this assumption states that a person with higher education will on average

[8] A naive polynomial fit to the midpoints in the graph would be a biased fit to the data because of Jensen's Inequality.

have weakly lower mortality risk.[9] Manski and Tamer (2002) also introduce the following Interval (I) and Mean Independence (MI) assumptions. For i which appears in the data as lying in bin k,

$$i \text{ in bin } K \implies P(i \in [i_k, i_{k+1}]) = 1. \tag{Assumption I}$$

$$E(y|i,\ i \text{ lies in bin } K) = E(y|i). \tag{Assumption MI}$$

Assumption I states that the rank of all people who report education ranks in category k are actually in bin k. Assumption MI states that censored observations are not different from non-censored observations; this always holds in our context because all of the data are interval censored.

Under these assumptions, Manski and Tamer (2002) provide the following sharp bounds on $E\left(y|i\right)$, given our setup with fully interval censored data:

$$r_{k-1} \leq E(y|i) \leq r_{k+1} \tag{Manski-Tamer bounds}$$

We can improve upon these bounds if the distribution of i is known. In some cases, as with ranks, the distribution is given by the definition of the variable. In other cases, conventional distributions are frequently assumed (such as lognormal or Pareto for income data). Alternatively, data could be transformed into a known distribution, for example, by taking the rank of the data. The analytical results are particularly parsimonious given the simplicity of the uniform distribution; we therefore consider the following assumption (U):

$$i \sim U(i_0, i_{K+1}) \tag{Assumption U}$$

where U is the uniform distribution. All of our results can be analytically derived under a general known distribution.

[9]Mortality is decreasing in educational attainment for every group and time period in the CDC data; it is also a monotonically decreasing function of income (Chetty et al., 2016).

If i is uniformly distributed, we know that:

$$E\left(i|i \in [i_k, i_{k+1}]\right) = \frac{1}{2}\left(i_{k+1} - i_k\right). \tag{II.1}$$

We derive the following proposition.

Proposition 1. *Let i be in bin k. Under assumptions IMMI and U, and without additional information, the following bounds on $E(y|i)$ are sharp:*

$$\begin{cases} r_{k-1} \leq E(y|i) \leq \frac{1}{i_{k+1}-i}\left(\left(i_{k+1} - i_k\right) r_k - \left(i - i_k\right) r_{k-1}\right), & i < i_k^* \\ \frac{1}{i-i_k}\left(\left(i_{k+1} - i_k\right) r_k - \left(i_{k+1} - i\right) r_{k+1}\right) \leq E(y|i) \leq r_{k+1}, & i \geq i_k^* \end{cases}$$

where

$$i_k^* = \frac{i_{k+1} r_{k+1} - \left(i_{k+1} - i_k\right) r_k - i_k r_{k-1}}{r_{k+1} - r_{k-1}}.$$

At the interval endpoints ($i = i_k$ or $i = i_{k+1}$), these bounds reduce to the bounds from Manski and Tamer (2002).

The proposition is obtained from the insight that the value of $E(Y|i)$ at a point I in bin k (below the midpoint) will only be minimized if all points in bin k to the left of I have the same value. Since all points to the right of I are constrained by the outcome value in the subsequent bin $k+1$, $E(y|i = I)$ will need to rise above the Manski and Tamer (2002) lower bound as I increases, in order to meet the bin mean. Intuitively, consider the point $E(Y|i_{k+1} - \epsilon)$. In order for this point to take on a value well below the bin mean r_k, it needs to be the case that a very large share of the density in bin k lies between $i_{k+1} - \epsilon$ and i_{k+1}. This is ruled out by the uniform distribution, and indeed by most common distributions, implying that the Manski and Tamer (2002) is too wide for most distributions. Appendix B provides additional intuition and a proof.

We generalize the proposition to obtain the following result for a known distribution of i:

Proposition 2. *Let i be in bin k. Let $f_k(i)$ be the probability density function of i in bin k. Under assumptions IMMI, and without additional information, the following bounds on $E(y|i)$ are sharp:*

$$\begin{cases} r_{k-1} \leq E(y|i) \leq \frac{r_k - r_{k-1}\int_{i_k}^{i} f_k(s)ds}{\int_{i}^{i_{k+1}} f_k(s)ds}, & i < i_k^* \\ \frac{r_k - r_{k+1}\int_{i}^{i_{k+1}} f_k(s)ds}{\int_{i_k}^{i} f_k(s)ds} \leq E(y|i) \leq r_{k+1}, & i \geq i_k^* \end{cases}$$

where i_k^* satisfies:

$$r_k = r_{k-1}\int_{i_k}^{i_k^*} f_k(s)ds + r_{k+1}\int_{i_k^*}^{i_{k+1}} f_k(s)ds.$$

A proof of the proposition is in Appendix B.

Figure 3 compares Manski and Tamer (2002) bounds to those obtained under the additional assumption of uniformity, using the mortality data. The new bounds are a significant improvement, especially in places where the data are particularly coarse and near the bin boundaries. For example, without using information on the distribution type, one could not reject that mortality for people in the first bin is 100,000/100,000 until just before the first bin boundary. The improvements in the other bins are less extreme but still substantial.

In addition to bounding the value of the CEF $E(Y|i)$ at any given point, we can also bound many analytical functions of the CEF, which we represent in the form $m(E(Y|i))$. One function of interest is the slope of the best linear approximation of the CEF; this is difficult to bound analytically, but we bound this numerically in Section II.B. Here, we focus on the average value of the CEF over an arbitrary interval of the conditioning space, or $\mu_a^b = E(Y|i \in [a,b])$. This function has several desirable properties: (i) it is bounded analytically; (ii) it is frequently bounded more tightly than $E(Y|i)$; and (iii) it has a similar interpretation to $E(Y|i)$ and is thus likely to be policy-relevant.

Let $f_{a,b}(i)$ represent the probability density function of i between a and b. Define μ_a^b as

$$\mu_a^b = \int_a^b E(y|i) f_{a,b}(i) di. \tag{II.2}$$

We now state analytical bounds on μ_a^b given uniformity. Let Y_i^{max} be the analytical upper bound on $E(y|i)$, given by Proposition 1. Let Y_i^{min} be the analytical lower bound on $E(y|i)$. The following proposition defines sharp bounds on μ_a^b under the assumption that i is uniformly distributed:

Proposition 3. *Let $a \in [i_h, i_{h+1}]$ and $b \in [i_k, i_{k+1}]$. Let assumptions IMMI and U hold. Then, if no additional information is available, the following bounds are sharp:*

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

The proof is in Appendix B.

We note two special cases. First, if $a = b$, then $\mu_a^b = E(Y|i = a)$. Second, if a and b correspond exactly to bin boundaries, then the bounds on μ_a^b collapse to a point: in this case, μ_a^b is just a weighted average of the bin means between a and b.

In fact, μ_a^b can be very tightly bounded whenever a and b are close to bin boundaries. For intuition, consider the following examples. If $\delta \in [a, b]$, μ_a^b can be written as a weighted mean of the two subintervals $\frac{\delta-a}{b-a}\mu_a^\delta + \frac{b-\delta}{b-a}\mu_\delta^b$.[10] If μ_a^δ is known (because there are bin boundaries at a and δ), then any uncertainty about the value of the CEF in the range $[a, \delta]$ is not consequential for the bounds on μ_a^b. If b is close to δ, the weight on the unknown value μ_δ^b is very small, and μ_a^b can be tightly bounded. Similarly, if instead μ_a^b is known, and b is again close to δ, then μ_a^δ can be tightly estimated even if μ_δ^b has wide bounds.

Bounds on other functions of the CEF are more difficult to calculate analytically, but can be defined as the set of solutions to a pair of minimization and maximization problems that take the following structure. We write the conditional expectation function in the form $Y(i) = s(i, \gamma)$, where γ is a finite-dimensional vector that lies in parameter space G and

[10] The weights on each subcomponent here assume that i is uniformly distributed. A different distribution would use different weights.

serves to parameterize the CEF through the function s. For example, we could estimate the parameters of a linear approximation to the CEF by defining $s(i, \gamma) = \gamma_0 + \gamma_1 * i$. The $s(i, \gamma)$ parameterization can also serve simply to discretize the CEF to aid with numerical optimization. In our numerical optimization below, we will define γ as a vector of 100 discrete values of the CEF in each single rank interval. Then $s(i, \gamma_{50})$ represents $E(y|i = 50)$, for instance. Define Γ as the set of feasible parameterizations of the CEF that obey monotonicity and minimize mean squared error with respect to the observed interval data:

$$\Gamma = \underset{g \in G}{\operatorname{argmin}} \left[\sum_{k=1}^{K} \frac{\int_{i_k}^{i_{k+1}} f_k(i) di}{i_{K+1} - i_1} \left(\left(\int_{i_k}^{i_{k+1}} s(i, g) f(i) di \right) - \overline{r}_k \right)^2 \right] \tag{II.3}$$

$$\text{such that}$$

$$E(y|i) \text{ is weakly increasing in } i. \tag{Monotonicity}$$

The $\frac{\int_{i_k}^{i_{k+1}} f(i) di}{i_{K+1} - i_1}$ term weights the mean squared error by the density contained in each bin.[11] Recall that for the rank distribution, $i_1 = 0$ and $i_{K+1} = 100$.

Any statistic m that is a single-valued function of the CEF, such as the average value of the CEF in an interval as described above, or the slope of the best fit line to the CEF, can be defined as $m(\gamma)$.[12] The bounds on $m(\gamma)$ are therefore:

$$\begin{aligned} m^{min} &= \inf\{m(\gamma) \mid \gamma \in \Gamma\} \\ m^{max} &= \sup\{m(\gamma) \mid \gamma \in \Gamma\}. \end{aligned} \tag{II.4}$$

For example, bounds on the best-fit linear approximation to the CEF are computed as follows. We consider the set of all CEFs that satisfy monotonicity and minimize mean-squared error with respect to the observed bin means. In many cases, there will exist

[11] While we choose to use a weighted mean squared error penalty, in principle Γ could use other penalties.

[12] We obtain the outcome-rank best linear approximation as the slope parameter which numerically minimizes the following:

$$\underset{\hat{\beta}, \hat{\beta}_0}{\operatorname{argmin}} \left\{ \int_I E(y|i) - \left(\hat{\beta} i + \hat{\beta}_0 \right) di \right\}.$$

monotonic CEFs that exactly match the observed data and the minimum mean-squared error will be zero. We then compute the slope of the best-fit linear approximation to each CEF. The largest and smallest slope constitute m^{min} and m^{max}.[13]

CEF Bounds Under Constrained Curvature

The candidate CEFs that are used to create the upper and lower envelopes in Figures 3 have large discontinuities. For example, the candidate function that generates the lower bound at $i = 25$ takes a very high value in the range $[0, 25]$, and then discontinuously falls to the value in the next bin over the range $[0, 39]$. If a function like this is an implausible description of the data, then the researcher may wish to impose an additional constraint on the curvature of the CEF, which will generate tighter bounds on the CEF and functions of the CEF. For example, examination of the mortality-income relationship (which can be estimated at each of 100 income ranks) displayed in Figure A1) reveals no such discontinuities.[14]

We consider a curvature restriction with the following structure:

$$s(i,\gamma) \text{ is twice-differentiable and } |s''(i,\gamma)| \leq \overline{C}. \qquad \text{(Curvature Constraint)}$$

This is analogous to imposing that the first derivative is Lipshitz.[15] Depending on the value of $\overline{C}$, this constraint may or may not bind.

Note that the most restrictive curvature constraint, $\overline{C} = 0$, is analogous to the assumption that the CEF is linear. In the context of mortality, an intermediate curvature restriction is

[13]Note that bounds on arbitrary functions of the CEF obtained in this way may not necessarily be sharp in the following sense. While the outer m^{min} and m^{max} are valid values of the parameter of interest, it could be possible that an intermediate value within $[m^{min}, m^{max}]$ could not be obtained for any CEF that matches the restrictions.

[14]More complex structural restrictions can also be imposed. For example, the CEF might be continuous within education bins, but there could be large discontinuities due to sheepskin effects at the education bin boundaries (Hungerford and Solon, 1987).

[15]Let X, Y be metric spaces with metrics d_X, d_Y respectively. The function $f : X \to Y$ is *Lipschitz continuous* if there exists $K \geq 0$ such that for all $x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2).$$

therefore less restrictive than the default practice in many current studies of mortality which seek to estimate the mortality-education best linear approximation (e.g., Cutler et al. (2011) and Goldring et al. (2016)). We discuss the choice of curvature restriction below.

In the results that follow, we primarily emphasize the bounds under $\overline{C} = \infty$. Our results are not sensitive to choice of $\overline{C}$. We primarily introduce this curvature restriction as we suspect it may be useful in other settings, and we wish to provide guidance on how to incorporate such an assumption should it be used.

II.B Numerical Calculation of CEF Bounds

This section describes a method to numerically solve the constrained optimization problem in Equations II.3 and II.4. We take a nonparametric approach for generality: explicitly parameterizing an unknown CEF with limited data is unsatisfying and could yield inaccurate results if the interval censoring conceals a non-linear within-bin CEF. In the context of mortality (and mobility, Section V), many CEFs of interest do not appear to obey a familiar parametric form (see Figures A1 and A2).

To make the problem numerically tractable, we solve the discrete problem of identifying the feasible values taken by $E(y|i)$ at each of N discrete values of i. We thus assume $E(y|i) = s(i, \gamma)$, where γ is a vector that defines the value of the CEF between each of the N bins. We use $N = 100$ in our analysis, corresponding to integer ranks, but other values may be useful depending on the application. Given continuity in the latent function, the discretized CEF is a very close approximation of the continuous CEF; increasing the value of N does not change any of our results, but increases computation time.

We solve the problem through a two-step process. Define a N-valued vector $\hat{\gamma}$ as a candidate CEF. First, we calculate the minimum MSE from the constrained optimization problem given by Equation II.3. We then run a second pair of constrained optimization problems that respectively minimize and maximize the value of $m(\hat{\gamma})$, with the additional constraint that the MSE is equal to the value obtained in the first step, denoted $\underline{\text{MSE}}$. Equation II.5 shows the second stage setup to calculate the lower bound on $m(\hat{\gamma})$. Note that this particular setup is specific to the uniform rank distribution, but setups with other distributions would be similar.

$$m^{min} = \min_{\hat{\gamma} \in [0,100]^N} m(\hat{\gamma}) \tag{II.5}$$

such that

$$s(i, \hat{\gamma}) \text{ is weakly increasing in } i \qquad \text{(Monotonicity)}$$

$$|s''(i, \hat{\gamma})| \leq \overline{C}, \qquad \text{(Curvature)}$$

$$\left[\sum_{k=1}^{K} \frac{\|I_k\|}{100} \left(\left(\frac{1}{\|I_k\|} \sum_{i \in I_k} s(i, \hat{\gamma}) \right) - \overline{r}_k \right)^2 \right] = \underline{\text{MSE}} \qquad \text{(MSE Minimization)}$$

where I_k is the set of discrete values of i between i_k and i_{k+1} and $\|I_k\|$ is the width of bin k. The complementary maximization problem obtains the upper bound on $m(\hat{\gamma})$.

Note that setting $m(\gamma) = \gamma_i$ obtains bounds on the value of the CEF at rank i. Calculating this for all ranks i from 1 to 100 generates analogous bounds to those derived in proposition 1, but satisfying the additional curvature constraint. Similarly $m(\gamma) = \frac{1}{b-a} \sum_{i=a}^{b} \gamma_i$ obtains bounds on μ_a^b.

II.C Example with Sample Data

In this section, we demonstrate the bounding method using data from mortality in the United States, continuing with the mortality of 50–54 year-old women in 2015.

Panel A of Figure 4 graphs the analytical upper and lower bounds on $E(y|i)$ at each value of i. A continuum of functions within these bounds obtain MSE of zero and can be considered equally plausible fits to the data. These bounds do not reflect statistical uncertainty but uncertainty about the CEF in the unobserved parts of the latent rank distribution.[16] Panels B and C of Figure 4 show the effect on the CEF bounds of progressively tightening the curvature constraint, with all bounds calculated numerically. Panel D shows the limit case

[16] We do not present standard errors because we are working with the universe of deaths in a large country and statistical imprecision is very small in this context. We discuss and present bootstrap confidence sets in Section V where statistical imprecision is more important.

with $\overline{C} = 0$; the CEF in this figure is identical to the predicted values from a regression of mortality on median education rank. While a stricter curvature restriction will typically tighten the bounds, this may come at the expenses of ruling out a plausible CEF, even if the MSE remains zero.[17]

The mortality-education data are minimally informative over the question of which curvature restriction to choose. One approach that could be taken is to examine the curvature on a closely related function that is not interval censored. We do this by examining the CEF of mortality given income rank which we show in Figure A1, using data from Chetty et al. (2016). In the income rank data for 52-year-old women, we calculate a maximum $\overline{C}$ of approximately 1.6. This suggests that Panel C of Figure 4 may represent a conservative curvature restriction on the mortality-education data. However, it is possible that the education CEF has greater curvature than the income CEF; it is thus advisable to explore the sensitivity of results to different curvature constraints.

In Table 1, we present estimates of $E(Y|i)$ and $E(Y|i \in (a,b))$ for women ages 50–54 in 2015, for various values of i, a and b, under different curvature constraints. We first highlight the statistics p_{32} and μ_0^{64}. p_{32} describes the expected mortality of the person with the median rank in the bottom education group (high school education or less) in 1992. μ_0^{64} describes the mean expected mortality among individuals occupying the same rank positions as those with high school or less in 1992. We present these statistics for both 1992 and 2015, but the rank bins are defined based on education ranks in 1992 to allow for constant-rank comparisons.

The table highlights the precision of the CEF interval mean estimates over the CEF point estimates. μ_0^{64} is nearly point identified in 1992 because 0 and 64 are very close to bin boundaries in 1992, and it is tightly bounded in 2015 as well.[18] In contrast, p_{32} has

[17]The curvature restriction does not always have to tighten the bounds. If $\underline{MSE}$ for $\overline{C}_1$ is smaller than the $\underline{MSE}$ corresponding to $\overline{C}_2$, then permitting the CEF to move farther away from the observed could yield larger bounds in some contexts. In practice, this does not occur very regularly.

[18]We have used integer approximations to these parameters for convenience; if we used the average mortality for the precise proportion of women with less than or equal to a high school degree (63.658%), then the parameter would be precisely point identified.

moderate bounds in both periods, which shrink substantially as we constrain the curvature. These two statistics are both useful summaries of mortality among the less educated, but μ_0^{64} is estimated with at least 22 times more precision than p_{32}. Note also that the bounds on μ_0^{64} do not substantially tighten as we constrain the curvature; this motivates us to work with unconstrained curvature in the applications in Sections IV and V.

The increased precision of interval estimates over estimates of the CEF at a point is greatest when the desired interval is close to a set of boundaries present in the data. We present several other pairs of measures in the table for which this is less true: i) p_{10} and μ_0^{20}, and ii) p_{25} and μ_0^{50}. In both cases, the interval mean measure can be estimated more tightly, but the advantage is less stark. Finally, because it is a frequently estimated parameter, we report bounds on the mortality-education rank best linear approximation.

In the final column of the table, we highlight the point estimates that would be derived from an assumption of a linear CEF of mortality given education. This approach would obscure the potential for very large increases in mortality at the bottom of the distribution. With no curvature constraint, p_{10} could be as large as 1,314.2 deaths/100,000, while a linear fit suggests that p_{10} is just 640.8 deaths/100,000. Note that the linear estimates do not necessarily lie at the midpoints of the bounds.

III Simulation: Bounds on the U.S. Mortality-Income CEF

In this section, we validate our method in a simulation by taking data from the fully supported mortality-income CEF from the US income distribution (Chetty et al., 2016), interval censoring that data, and then recovering bounds on the true CEF from that interval censored data. The exercise shows that our approach works in practice. It also illustrates that studying partially identified bounds on the CEF, rather than a simple best-fit to the interval-censored data, can permit the researcher to recover important features of the CEF that she might miss if she attempted simply to fit a parametric form to the observed bins.

Chetty et al. (2016) collect tax-linked income-mortality records from people in the United States. They make publicly available mortality rates by income percentile for men and women by age and year. We again focus on women aged 52 in 2014, the group most comparable to the group of women aged 50–54 in 2015 that we have examined so far.

First, we generate an estimate of the true CEF from the mortality-income data by fitting a cubic spline with four knots to the data. This spline is the same spline used to obtain an estimate of $\overline{C}$, and is plotted for men and women in Appendix Figure A1.[19]

Next, we simulate interval censoring by obtaining the mean of the true CEF within income rank bins that cover the same ranks as the education bins observed in our 2014 mortality-education data. In this simulation, there are 39% of people in the bottom bin, 29% of people in the middle bin, and the remaining 33% of people in the top bin.[20] After applying this interval censoring, we have a dataset with three income-rank bins and average mortality in each bin. This dataset is structurally identical to the mortality-education dataset from Section II. We compute bounds on the CEF using only the binned data.

Panels A–D of Figure 5 present bounds on the binned data, with curvature limits that vary from ∞ to 1. The dashed lines show the underlying data. The solid circles show the constructed bin means of the censored data; these are the only data that we use for the optimization. The solid lines show the upper and lower envelopes that we calculate for the nonparametric CEF.

The exercise illustrates that conservative curvature constraints yield bounds that contain the true CEF at every point, but it also shows that imposing an excessive curvature constraint may yield bounds that do not contain the true CEF. In Panels A through C, the true data are always contained by the bounds, because the true CEF has a second derivative of less than 5. It is important to note that the true CEF is not always centered within the bounds; from ranks 25 to 40, the true CEF is near the bottom bound, and from ranks 90 to 100, it is nearer to the upper bound. Panel D imposes a curvature constraint of 1, which is smaller

[19] We use a spline approximation rather than the raw data because the variation across neighboring rank bins is most likely idiosyncratic given the small number of deaths in an age bin defined by a single year. By using information from neighboring points, the splines is likely a better estimate of mortality risk than the individual rank bin means.

[20] We round to the nearest integer, since we only observe integer percentiles in the data from Chetty et al. (2016).

than the true curvature of the CEF. As a result, the CEF briefly crosses the bounds, at around rank 20.

The exercise also illustrates that assuming a parametric form for the underlying CEF can yield misleading results. The three data points observed by the econometrician are almost perfectly linear. One might be tempted, therefore, to assume the CEF is given by the MSE-minimizing line fitted to the bin means, estimated using OLS. That procedure would fail to identify the convexity at the bottom of the distribution. The strength of our method is that in the domains where interval censoring is most severe, it makes clear that we can at best obtain bounds on the partially identified CEF.

Table 2 shows estimates of a range of statistics of interest under different curvatures, as well as the true estimate. We highlight three results. First, the interval mean measures (μ) generate tighter bounds than the CEF values p, with no greater propensity for error. Second, $\overline{C}$ is consequential for p, but relatively unimportant for μ. Third, the linear estimates generated with $\overline{C} = 0$ are systematically biased relative to the true estimates, and often deliver estimates that are outside the bounds even of CEFs with unconstrained curvature.

IV Estimating U.S. Mortality in Constant Education Rank Bins

In this section, we examine changes in U.S. mortality for individuals at constant ranks in the education distribution. Many researchers have noted that mortality is rising for individuals in less educated groups; however, the changing composition of these groups over time has made this finding difficult to interpret. For example, women with a high school education or less represented the least educated 64% of the population in 1992, and the least educated 39% of the population in 2015. If mortality is decreasing in education in the cross-section, as is suggested by Figure 1, then the changing size and composition of the group of women without college would account for at least some of this mortality increase. This bias has been frequently noted in the literature, but different authors have reached widely different conclusions regarding its size and importance.[21] In this section, we show that the methods above can put tight bounds on mortality change estimates for many constant education rank bins of interest.

Mortality by education records come from the U.S. Center for Disease Control's WONDER database and total population by age, gender and education come from the Current Population Survey, as in Case and Deaton (2017). As is standard in this literature, when deaths are missing education information, we assign them to bins according to education proportions in the population. Our estimates differ very slightly from Case and Deaton (2017) because (i) we include the institutionalized population (who are counted in death records) in our population counts and they do not; and (ii) Case and Deaton (2017) drop states with inconsistent education reporting.[22] We obtain the number of institutionalized people from the U.S. Census or American Communities Survey and linearly impute the number of institutionalized people in each cell for years when neither survey is observed. More details on data construction are available in the data appendix.

As above, we assume that the observed mortality data describe a monotonic relationship between mortality and latent education rank, the latter of which is observed only in coarse bins. We bound the value of the CEF at constant education ranks, without the need to observe these ranks directly in the data. This makes it possible to generate mortality estimates that describe a constant size group occupying a constant set of education ranks over time. We focus in this section on women aged 50–54, because this is a group whose education

[21]Cutler et al. (2011) adjust for compositional shifts by predicting propensity to attend college using region, marital status and income, and then using this propensity as a conditioning variable. They argue that compositional shifts are not important for mortality changes from the 1970s to the 1990s. This approach is limited by the extent to which these variables can predict education, and they are in many cases unavailable. Dowd and Hamoudi (2014) and Bound et al. (2015) perform analytical exercises that suggest that compositional shifts explain most or all of recent mortality changes. Bound et al. (2015) estimates mortality for the bottom quartile of the education distribution, implicitly assuming that mortality is constant within each interval-censored mortality rank bin. Goldring et al. (2016) derive a one-tailed test for changes in the mortality-education gradient, but they do do not calculate the bias in existing mortality estimates or estimate mortality in constant rank bins. Our method requires no additional covariates, and bounds mortality at an arbitrary education rank under only the assumption of monotonicity.

[22]We establish in Appendix C that dropping these states has virtually no effect on point estimates in the years where these data are public, and we have applied for the restricted access data that will allow us to do this for all years.

composition has shifted substantially over time—the share with high school or less falls from 64% to 39% from 1992-2015—making unadjusted mortality estimates difficult to interpret.[23]

Panel A of Figure 6 plots mean total mortality for women age 50-54 in each education group in 1992 and in 2015, along with analytical bounds on CEFs with unconstrained curvature. The bounds are largely overlapping across the entire education distribution, and too wide to infer very much about changes in mortality. In Panel B, we restrict curvature to the approximately three times the maximum curvature from the income-mortality data, as discussed in Section II. Under constrained curvature, we can identify a clear decline in mortality at the top of the education distribution. In the less educated half of the distribution, the bounds are improved relative to the case with unconstrained curvature, but we cannot reject mortality gains or losses at any percentile.

We next turn to the interval mean measures, which permit tighter inference. For women aged 50–54, we focus on estimating mean mortality in the bottom 64%, denoted by μ_0^{64}. This measure describes mean mortality for women who occupied positions in the rank distribution that would give them high school education or less in 1992. In 2015, the group of women occupying these rank positions includes all women with high school or less, and some women with some college education, but none with bachelor's degrees or higher. For men, we focus on the bottom 54%, which is the population share with high school or less in 1992; this share falls to 44% by 2015. By ranking men and women against members of their own gender, we estimate mortality for a given percentile group of men or women; that is, the least educated group can be interpreted as the "the 64% of least educated women," rather than "women in the bottom 64% of the population education distribution." We chose own-gender reference points because women's and men's labor market opportunities and choices are often different and because women and men often share households and incomes. However, alternate choices could be considered and estimated.

Panel A of Figure 7 shows bounds on total mortality for women aged 50-54, in a constant

[23]We use 5-year bins for ages rather than larger bins to ensure that the average age in the bin does not change over time (Gelman and Auerbach, 2016).

rank bin that corresponds to the ranks of women with high school education or less in 1992. Note that mortality in 1992 can be point estimated, because the 0-64 rank bin interval is exactly that observed in the data. As education levels diverge from those in 1992, the bounds progressively widen. The "x" markers in the figure plot the unadjusted estimates of mortality among women with less than or equal to high school education; these mortality estimates describe a group occupying a shrinking and more negatively selected share of the population over time. The unadjusted estimates, which are the object of study in most earlier work in U.S. mortality, significantly overestimate mortality increases for the constant rank group. The upper bound on mortality gain from 1992-2015 for this group of women is 8.5%, compared to the unadjusted estimate of 28%. Panel B shows the same figure for men. The unadjusted estimates are closer to the bounds here because men have gained less education than women over this period. We can bound the mortality change for men in the set [-7.1%, +0.3%], compared with the unadjusted estimate of +1.2%.[24] Panels C and D present analogous results for combined deaths from suicide, poisoning and liver disease, described by Case and Deaton (2017) as "deaths of despair." The unadjusted mortality estimates continue to overstate the constant rank mortality changes, but the difference is small here because (i) deaths of despair have increased substantially among all groups; and (ii) the education gradient in deaths of despair was small in 1992.

Table 3 shows unadjusted and constant-rank estimates of women's mortality changes from 1992-2015 for age groups from 20 to 69, for all education categories. The unadjusted estimates systematically overestimate mortality increases for all groups, because the mean rank in each group has declined over this period. In some cases (e.g. women with less than high school, age 55-59), the unadjusted estimate has the wrong sign.

The extent of the bias on the naive estimates is increasing in the magnitude of the

[24]The previous literature, especially Case and Deaton (2015, 2017) have focused on *white* men, whose unadjusted mortality is rising. We find this in our data as well, but this is balanced by large declines in mortality of black men, leading to declining mortality among the bottom 54%. Estimating separate bounds for different racial groups requires additional assumptions about the relative positions of these groups in the unobserved part of the latent education distribution, and we leave this exercise for future work.

mortality-education gradient, and in the magnitude of the shift in bin boundaries. Given the significant variation across age groups and genders, blanket assumptions about the existence or lack of bias in such estimates are therefore unlikely to be useful. Unadjusted estimates of men's mortality changes from 1992-2015 are close to the constant rank bounds, as are unadjusted estimates of deaths of despair for both men and women. For women's total mortality, however, the naive estimates overstate mortality increases in many cases by a factor of three or more, and in some cases they show a different sign.

V Application: Intergenerational Educational Mobility

The study of intergenerational mobility is another context where researchers may wish to study the conditional expectation function when the conditioning variable is interval censored.[25] In this application, we examine a case where interval censoring is severe. We illustrate that common measures, including both the rank-rank gradient and absolute upward mobility (p_{25}), are either inaccurate or uninformative once we account for interval censoring. We introduce upward interval mobility (μ_0^{50}) which can be tightly bounded. In our example using father-son education Indian data, using the gradient without incorporating interval censoring would yield the inaccurate conclusion that mobility has improved. By contrast, using upward interval mobility shows that intergenerational educational mobility has remained stagnant.

Studies of intergenerational mobility typically rely upon some measure of rank in the social hierarchy which can be observed for both parents and children (Chetty et al., 2014a; Chetty et al., 2017b). In many contexts, the only measure of social rank available for parents is their level of education. In richer countries, this frequently arises for studies of mobility in eras that predate the availability of administrative income data.[26] In developing countries, matched parent-child data are considerably more rare, making educational mobility the only

[25]For a review of intergenerational mobility, see Solon (1999), Hertz (2008), Corak (2013), Black and Devereux (2011), and Roemer (2016).

[26]See, for example, Black et al. (2005), Long and Ferrie (2013), and Güell et al. (2013).

feasible object of study.[27]

The coarseness of educational data creates several problems for the study of intergenerational educational mobility.[28] Many useful measures of mobility, such as absolute mobility (Chetty et al., 2014a), require information on the CEF of child rank given parent rank at a given percentile. But the CEF of educational rank is typically not observed at any given percentile, because each percentile in the parent distribution lies within some larger bin. Studies on educational mobility therefore typically focus on linear estimators of the parent-child outcome relationship, such as the slope of the best linear approximator to the CEF, i.e. the rank-rank gradient. While this is a useful summary statistic for the entire distribution, it is not useful for cross-group comparison. The within-group rank-rank 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 of having worse outcomes than other groups at every point in the parent distribution.[29] A second limitation of the rank-rank gradient is that it aggregates information about mobility at the top and at the bottom of the parent distribution; it is not directly informative about upward mobility in the bottom half of the distribution. Transition matrices, another useful tool in mobility studies, also require observation of the CEF at fixed parent intervals.

There is no established methodology for calculating constant rank mobility measures from educational data, even though such measures are central to current research on mobility. This paper demonstrates how to obtain bounds on i) the parent-child CEF at percentiles where the parent variable is interval censored, and ii) the average value of the child CEF across an

[27] See, for example, Wantchekon and Stanig (2015), Hnatkovska et al. (2013) or Emran and Shilpi (2015).

[28] Table A1 reports the number of bins used in a set of recent studies of intergenerational mobility from several rich and poor countries. 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%. We specifically selected a set of studies where coarse data is likely to be an important factor. This list notably excludes several recent studies using tax data from developed countries, where the parent rank distribution is observed without substantial interval censoring.

[29] An extreme example makes this clear. Suppose children in some population subgroup A all end up at the 10th percentile of the outcome distribution with certainty. The rank-rank gradient for this group would be zero (assuming some variation in parent outcomes), implying perfect mobility. But in fact the group would have virtually no upward mobility.

arbitrary percentile range of the parent distribution.[30] We work through an example using data from India, where the coarse data problem is severe, and show that our method can recover parameters of interest that are difficult to obtain otherwise. We combine data from two data sources, including administrative data on the education of every person in India in 2012, to obtain a representative sample of every father-son pair in India.[31] The details of the data construction are described in Appendix D.B.

Education for both fathers and sons is observed in seven categories.[32] Because sons' education levels are also reported categorically, we do not directly observe the expected child outcome in each parent education bin. Extending the method in this paper to the joint problem of interval-censored X and Y data is difficult and beyond the scope of the current work. In Appendix C, we present a constrained optimization setup that solves the joint problem, but it is a 10,000 parameter numerical problem that is computationally difficult to solve. We instead assign to children the midpoint of their rank bin, and we show in Appendix C that other data on children (specifically, children's wages, for which the rank distribution is uncensored) suggests that this is a very good approximation to the true latent joint parent-child distribution.[33] We can then calculate the expected rank of sons for every education level of fathers, and use the method presented above. An alternate approach would be to use a socioeconomic rank measure for sons that can be observed continuously, such as wages or household income, which may also be an object of interest in the study of mobility (Chetty et al., 2014a).

[30]Our method is loosely related to Chetty et al. (2017a), who use a numerical procedure to bound absolute mobility at the 25th percentile, given just the marginal distributions of children's and parents' incomes. We bound the value of the CEF analytically, beginning from an empirical CEF with an interval-censored conditioning variable, but our optimization setup imposes similar constraints.

[31]We are restricted to the study of fathers and sons because the data do not match daughters to parents or children to mothers when they do not live in the same household.

[32]The categories are (i) less than two years of education; (ii) at least two years but no primary; (iii) primary; (iv) middle school; (v) secondary; (vi) senior secondary; and (vii) post-secondary or higher.

[33]Because parental education is often obtained by asking children, it is common to have data on many child outcomes, but only the education level of parents, as we do here. The midpoint rank is a good approximation because the residual correlation of father education and son wages is very small once son's education is controlled for. This, of course, may not hold in other contexts. In Appendix C we also present a method that generates best and worst case bounds under joint censoring for the case of intergenerational mobility.

Panel A of Figure 8 shows the raw data for cohorts born in the 1950s and in the 1980s. Each point plots the midpoint of a father education rank bin against the expected child rank in that bin. The vertical lines plot the boundary for the lowest education bin for each cohort, which corresponds to fathers with less than two years of education. In the 1950s birth cohort (solid line), this group represents 60% of the population; it represents 38% for the 1980s cohort (dashed line). When we estimate the rank-rank gradient directly on these bin means, we find small but unambiguous mobility gains over this 30-year period. The graph makes clear that the decrease in the gradient is driven by changes in mobility in the top half of the distribution. If we are interested in a measure of upward mobility, or a measure that is valid for subgroup comparisons, we will need information on the latent CEF underlying the points in this panel. Note also that the CEF is evidently non-linear. A naive nonlinear parametric fit to the bin midpoints would be biased due to Jensen's Inequality.

Panel B of Figure 8 shows the bounds on the parent-child CEFs for these birth cohorts; we select a curvature constraint of 0.1, which is approximately 1.5 times the maximum curvature observed in uncensored parent-child *income* data from the United States, Denmark, Sweden and Norway.[34] The bounds on the CEF are widest at the bottom of the distribution where interval censoring is most severe, and are worse for the older generation with the larger bottom rank bin. The bounds on the CEF in bottom half of the distribution are consistent with both large positive and large negative changes in mobility. Absolute upward mobility (the value of the CEF at the 25th percentile) can evidently not be bounded informatively.[35]

We can make meaningful progress by focusing on an interval-based measure such as $\mu_a^b = E(Y|i \in (a,b))$. We call this measure *interval mobility*, and focus in particular on $\mu_0^{50} = E(Y|i \in (0,50))$, which we call *upward interval mobility.* This statistic is closely

[34]We selected these countries because we were able to obtain precise uncensored parent-child income rank data for them from Chetty et al. (2014b), Boserup et al. (2014) and Bratberg et al. (2015). Graphs for the spline estimations used to calculate the curvature constraints are displayed in Appendix Figure A2. Results are substantively similar under different curvature constraints.

[35]Note that a more restrictive curvature constraint would narrow the bounds, but at the expense of imposing excessive structure that would rule out plausible CEFs, especially given the evident nonlinearity in the data.

related to absolute upward mobility. The latter describes the outcome of the median child born to a parent in the bottom half of the parent distribution, whereas upward interval mobility describes the *mean* outcome in the bottom half of the parent distribution. These measures are likely to be of similar policy interest and economic significance, but we show here that upward interval mobility can be bounded tightly in contexts with severe interval censoring, while absolute upward mobility cannot.

Figure 9 shows bounds on the three mobility statistics discussed for each decadal cohort: the rank-rank gradient, absolute upward mobility (p_{25}), and upward interval mobility (μ_0^{50}). The figure illustrates that using the rank-rank gradient or absolute upward mobility would yield inaccurate or incomplete conclusions. For reference, we plot recent estimates of similar educational mobility measures from USA and Denmark.[36] Once we allow expected child outcomes to vary within the bottom parent education bin, the rank-rank gradient becomes largely uninformative. Absolute upward mobility has similarly wide bounds, which are consistent with both large mobility gains and will small mobility losses over the sample period. In contrast, upward interval mobility is estimated with tight bounds in all periods. According to this measure, upward mobility has changed very little over the four decades studied; there is a small gain from the 1950s to the 1960s, followed by a small decline from the 1960s to the 1980s. On average, Indian mobility is as far below that in the United States as mobility in the United States is below Denmark. Table 4 reports the bounds for each measure and cohort with bootstrap confidence sets under a range of curvature restrictions.[37] Moderate curvature restrictions generate substantial improvements on the estimation of the value of the CEF (e.g., p_{25}), but are considerably less important for the interval mean measures (e.g., μ_0^{50}), which are tightly bounded even with unconstrained curvature.

In conclusion, the most widely used mobility estimator, the rank-rank gradient, presents

[36] Rank-rank correlations of education are from Hertz (2008), which are equal to the slope of the rank-rank regression coefficient if estimated on uncensored rank data. For absolute mobility, we calculate p_{25} for the U.S. and Denmark from the distributions shown in Figure A2, with data from Chetty et al. (2014a).

[37] We calculate bootstrap confidence sets using 1,000 bootstrap samples from the underlying datasets, following methods described in Imbens and Manski (2004) and Tamer (2010).

an incomplete picture and potentially biased picture of intergenerational educational mobility. Upward interval mobility, in contrast, yields informative estimates even without using a curvature constraint, making it feasible to study upward mobility in the lower-ranked parts of the distribution, even in a context with extreme interval censoring. In this application, we find that intergenerational mobility in India has not changed from 1960–1990.

VI Conclusion

We have proposed methods that generate useful bounds on a conditional expectation function when the conditioning variable is censored. Tight bounds on parameters of interest are possible because of three innovations. First, CEF bounds are substantially improved when the distribution of the conditioning variable is known, and many economic contexts have distributions that are either known with certainty or are assumed by convention. Second, we show that the conditional mean across some interval can often be bounded as or more tightly than the CEF itself. Third, bounds can be improved by imposing a constraint on the curvature of the CEF, which may be justified in many empirical contexts; a curvature constraint can further substitute for the assumption of monotonicity. In our applications, the first two innovations prove sufficient to bound parameters of interest informatively, but any of these alone is insufficient. Simulations of interval censoring indicate that the methods perform well in common empirical scenarios.

We have shown that our method can be used to solve known problems in the study of mortality and of intergenerational mobility. Generating bounds on some outcome variable by education quantile is an application with many other potential uses, given the large number of contexts where education is of interest as a dependent variable but available only in a small number of bins. Other useful applications may be found where the conditioning variable takes the form of interval-censored income data, or Likert scale responses, among others.

References

Aaronson, Daniel and Bhashkar Mazumder, "Intergenerational Economic Mobility in the United States, 1940 to 2000," *The Journal of Human Resources*, 2008, *43* (1), 139–172.

Aizer, Anna and Janet Currie, "The intergenerational transmission of inequality: Maternal disadvantage and health at birth," *Science*, 2014, *344* (6186).

Asher, Sam and Paul Novosad, "Rural Roads and Structural Transformation," 2017.

Aydemir, Abdurrahman and Hakki Yazici, "Intergenerational Education Mobility and the Level of Development: Evidence from Turkey," 2016.

Bertrand, Marianne, Patricia Cortés, Claudia Olivetti, and Jessica Pan, "Social norms, labor market opportunities, and the marriage gap for skilled women," 2016.

Björklund, Anders, Mikael Lindahl, and Erik Plug, "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *Quarterly Journal of Economics*, 2006, *121* (3), 999–1028.

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.

Bound, John, Arline T Geronimus, Javier M Rodriguez, and Timothy A Waidmann, "Measuring Recent Apparent Declines In Longevity: The Role of Increasing Educational Attainment," *Health Affairs*, 2015, *34* (12).

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).

Case, Anne and Angus Deaton, "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences*, 2015, *112* (49).

—— **and** —— , "Mortality and morbidity in the 21st century," *Brookings Papers on Economic Activity*, 2017.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang, "The fading American dream: trends in absolute income mobility since 1940," *Science*, 2017, *356* (6336).

___ **, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility," 2017.

___ **, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler**, "The Association Between Income and Life Expectancy in the United States, 2001-2014," *JAMA*, 2016, *315* (16).

___ **, 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.

___ **,** ___ **,** ___ **,** ___ **, and Nicholas Turner**, "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review: Papers & Proceedings*, 2014, *104* (5), 141–147.

Corak, Miles, "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, 2013, *27* (3).

Cutler, David and Adriana Lleras-Muney, "The Education Gradient in Old Age Disability," in David A. Wise, ed., *Research Findings in the Economics of Aging*, The University of Chicago Press, 2010.

___ **and** ___ , "Understanding differences in health behaviors by education," *Journal of Health Economics*, 2010, *29* (1).

___ **, Fabian Lange, Ellen Meara, Seth Richards-Shubik, and Christopher Ruhm**, "Rising educational gradients in mortality: The role of behavioral risk factors," *Journal of Health Economics*, 2011, *30* (6).

Dardanoni, Valentino, Mario Fiorini, and Antonio Forcina, "Stochastic Monotonicity in Intergenerational Mobility Tables," *Journal of Applied Econometrics*, 2012, *27.*

Dowd, Jennifer and Amar Hamoudi, "Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality," *International Journal of Epidemiology*, 2014, *43* (4).

Dunn, Christopher E., "Intergenerational Economic Mobility around the World," *B.E. Journal of Economic Analysis and Policy*, 2007, *7* (2), 1–40.

Emran, M.S. and F. Shilpi, "Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam," *Journal of Human Resources*, 2011, *46* (2).

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

Gelman, Andrew and Jonathan Auerbach, "Age-aggregation bias in mortality trends," *Proceedings of the National Academy of Sciences*, 2016, *113* (7).

Goldring, Thomas, Fabian Lange, and Seth Richards-Shubik, "Testing for changes in the SES-mortality gradient when the distribution of education changes too," *Journal of Health Economics*, 2016, *46* (2).

Güell, Maia, José V Rodríguez Mora, and Christopher I. Telmer, "The informational content of surnames, the evolution of intergenerational mobility, and assortative mating," *Review of Economic Studies*, 2013, *82* (2).

Guest, Avery M, Nancy S Landale, and James C Mccann, "Intergenerational Occupational Mobility in the Late 19th Century United States," *Social Forces*, 1989, *68* (2).

Hertz, Tom, "Rags, riches and race: The intergenerational economic mobility of black and white families in the United States," in Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves, eds., *Unequal Chances: Family Background and Economic Success*, Princeton University Press, 2005.

——, "A group-specific measure of intergenerational persistence," *Economics Letters*, 2008, *100* (3), 415–417.

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh B. Paul, "Breaking the caste barrier: intergenerational mobility in India," *The Journal of Human Resources*, 2013, *48* (2).

Hungerford, Thomas and Gary Solon, "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 1987, *69* (1).

Imbens, Guido W. and Charles F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72* (6).

Knight, John, Terry Sicular, and Ximing Yue, "Educational Inequality in China: The Intergenerational Dimension," 2011.

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren, "Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations," *Journal of Human Resources*, 2012, *50* (1).

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

Manski, Charles F. and Elie Tamer, "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 2002, *70* (2).

Meara, Ellen, Seth Richards, and David Cutler, "The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981-2000," *Health Affairs*, 2008, *27* (2).

Olshansky, Jay, Toni Antonucci, Lisa Berkman, Robert Binstock, Axel Boersch-Supan, John Cacioppo, Bruce Carnes, Laura Carstensen, Linda Fried, Dana Goldman, James Jackson, Martin Kohli, John Rother, Yuhui Zheng, and John Rowe, "Differences In Life Expectancy Due To Race And Educational Differences Are Widening, And Many May Not Catch Up," *Health Affairs*, 2012, *31* (8).

Piraino, Patrizio, "Intergenerational Earnings Mobility and Equality of Opportunity in South Africa," *World Development*, 2015, *67* (March), 396–405.

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

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.

Tamer, Elie, "Partial Identification in Econometrics," *Annual Review of Economics*, 2010, *2* (1), 167–195.

Wantchekon, Leonard and Piero Stanig, "The Curse of Good Soil? Land Fertility, Roads, and Rural Poverty in Africa," 2015.

Xie, Yu and Alexandra Killewald, "Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment," *The American Economic Review*, 2013, *103* (5).

Figure 1

Women's Total Mortality by Education Group,
Age 50-54, 1992-2015

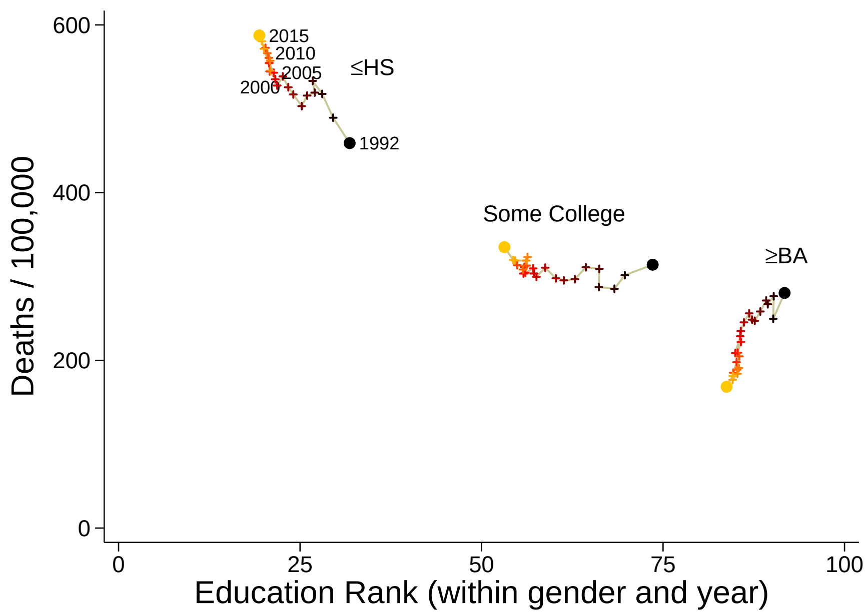


Figure 1 plots mortality rates vs. the mean education rank for three groups: people who have less than or equal to a high school degree, people who have some college education, and people who have a BA or more. Each point represents a rate within a year and education rank. The lighter colored points correspond to later years. The population displayed includes all women in the United States, who are ranked within their gender and year.

Figure 2

Candidate Functions for Conditional Expectation of Mortality
given Latent Education Rank

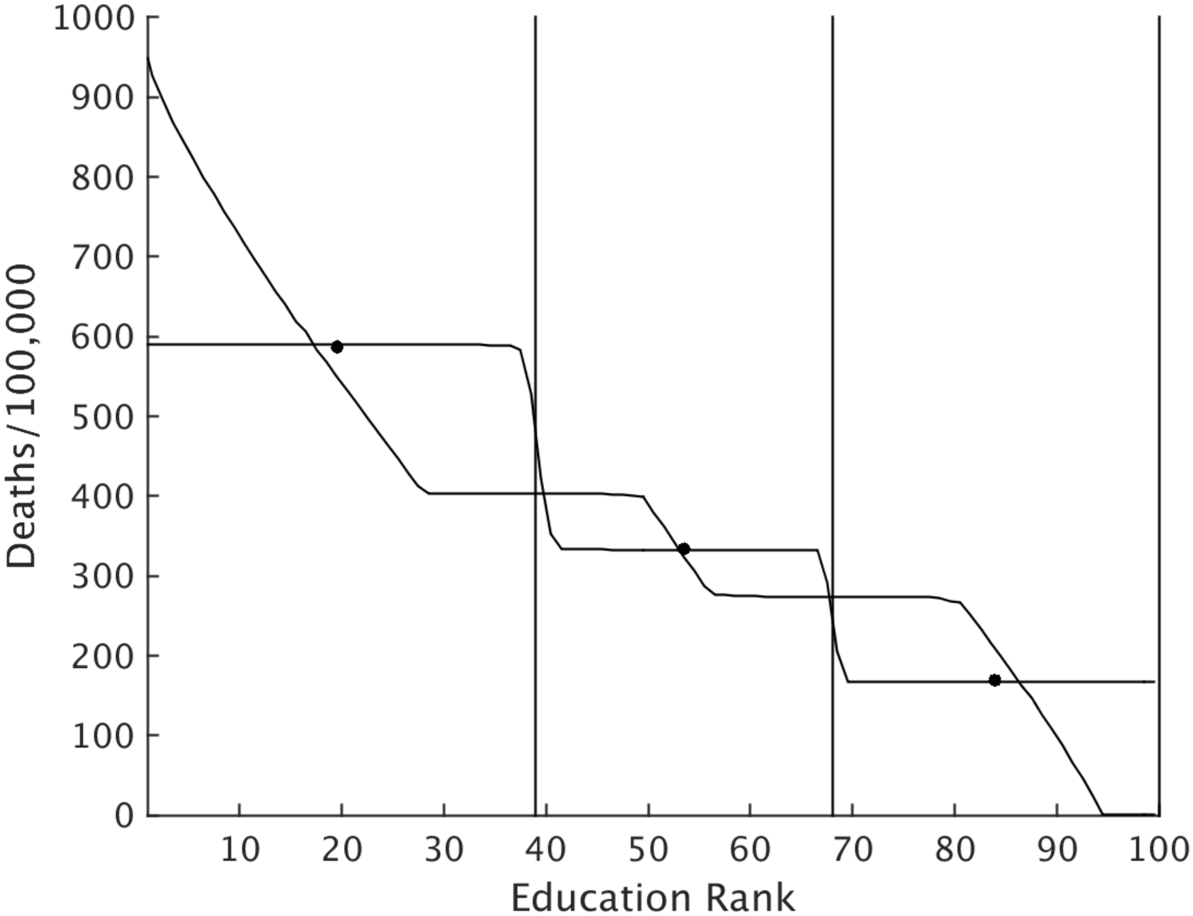


Figure 2 shows candidate conditional expectation functions of mortality given education rank for women aged 50–54 in the United States in 2015. The vertical lines show the bin boundaries and the points show the mean value of child rank in each bin. The points are centered at the rank bin means on the X axis.

Figure 3

Analytical Bounds on the CEF of Mortality given Education Rank

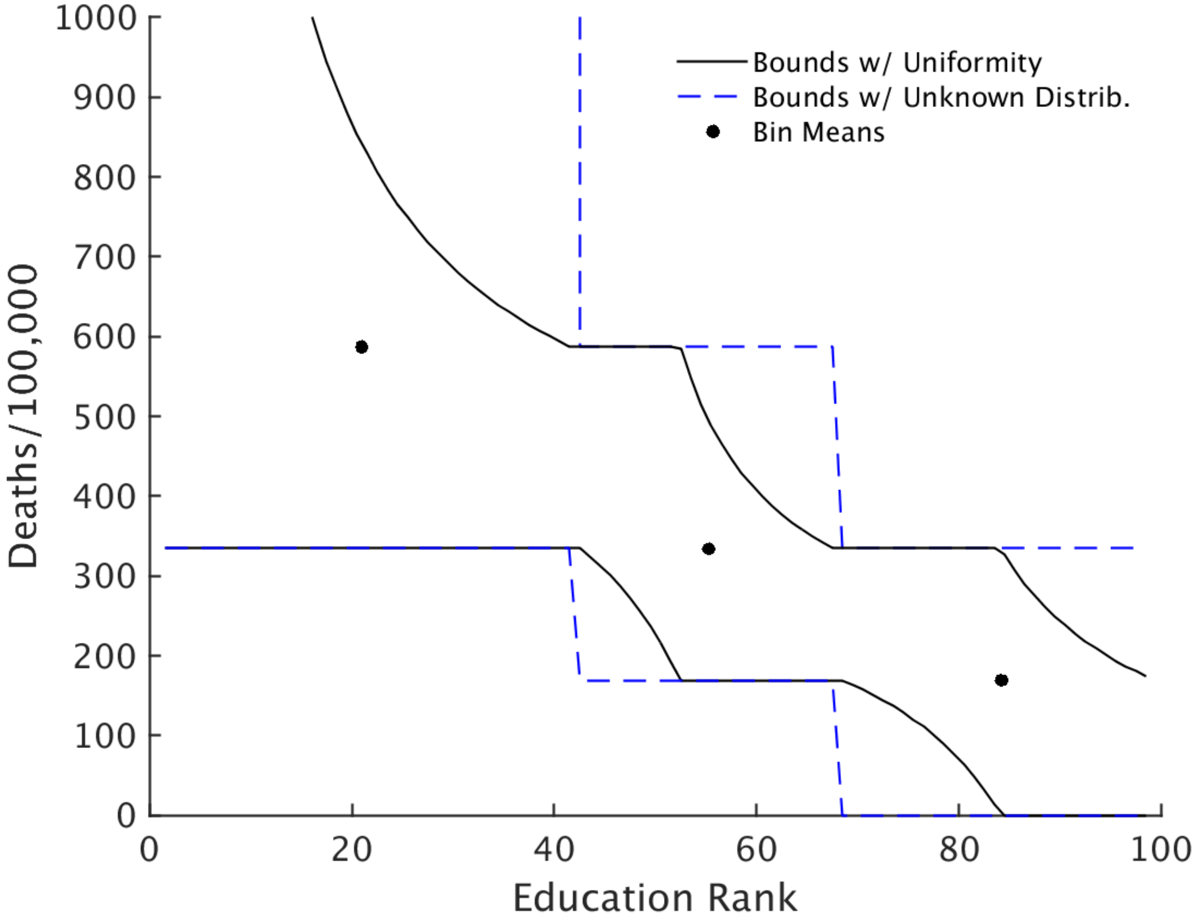


Figure 3 shows bounds on the conditional expectation of mortality given education rank for women aged 50–54 in the United States in 2015. The vertical lines show the bin boundaries and the points show the mean value of child rank in each bin. The points are centered at the rank bin means on the X axis. The dashed lines show analytical bounds when the distribution of the X variable is unknown (Manski and Tamer, 2002). The solid line shows analytical bounds when the distribution of the X variable is uniform.

Figure 4

CEF of Mortality given Education Rank:
Bounds Under Curvature Constraints

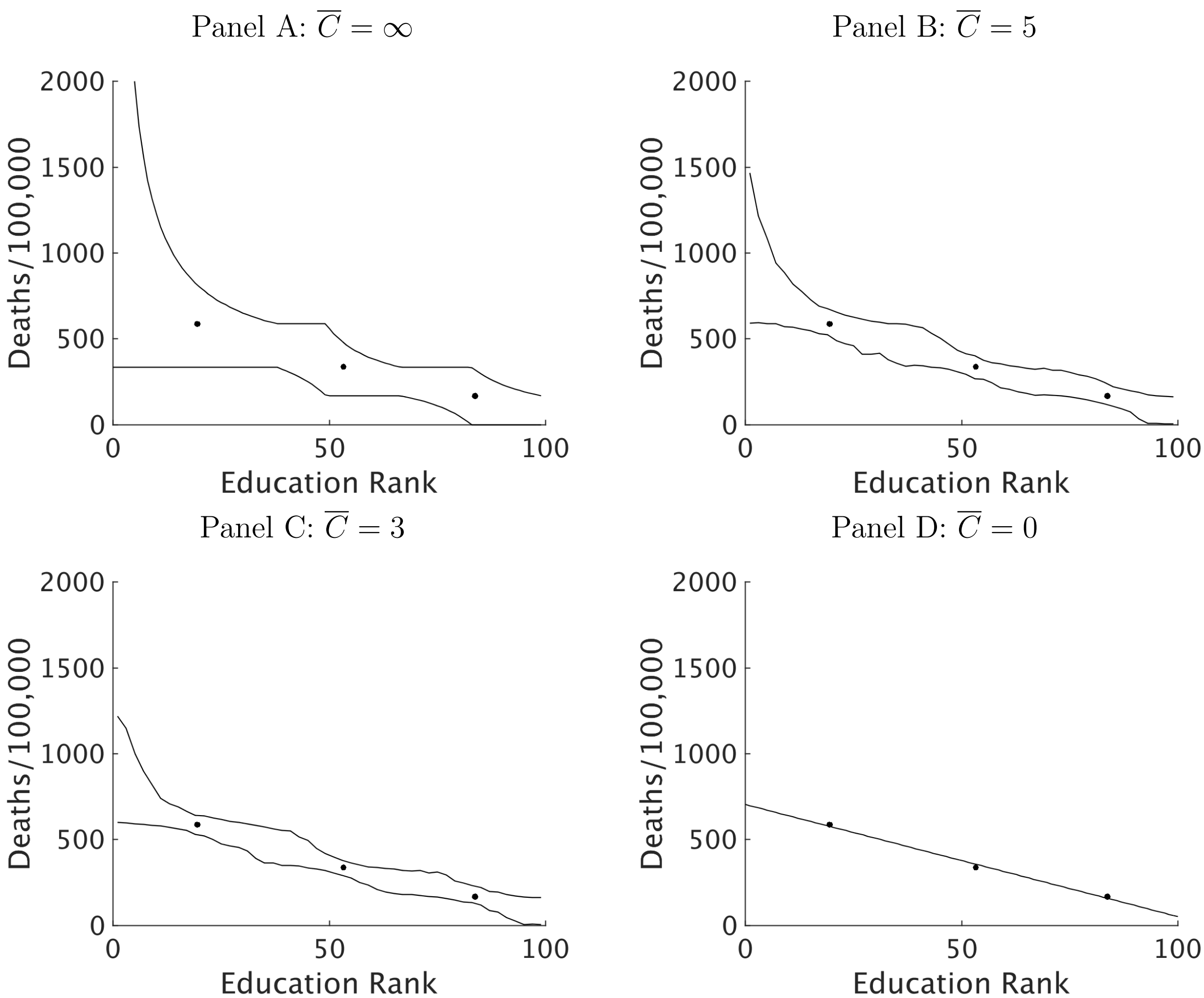


Figure 4 presents bounds on conditional expectation functions of mortality given education ranks for women ages 50–54 in 2015. Education rank is measured relative to the set of all women aged 50-54. The lines in each panel represent the upper and lower bounds on the CEF at each rank, obtained under different curvature restrictions. All four panels impose monotonicity, as well as the constraint on the second derivative given in the panel title. The circles represent the mean mortality of women in each bin in the education distribution.

Figure 5

Simulated Interval Censoring and Bounds using U.S. Mortality-Income Data

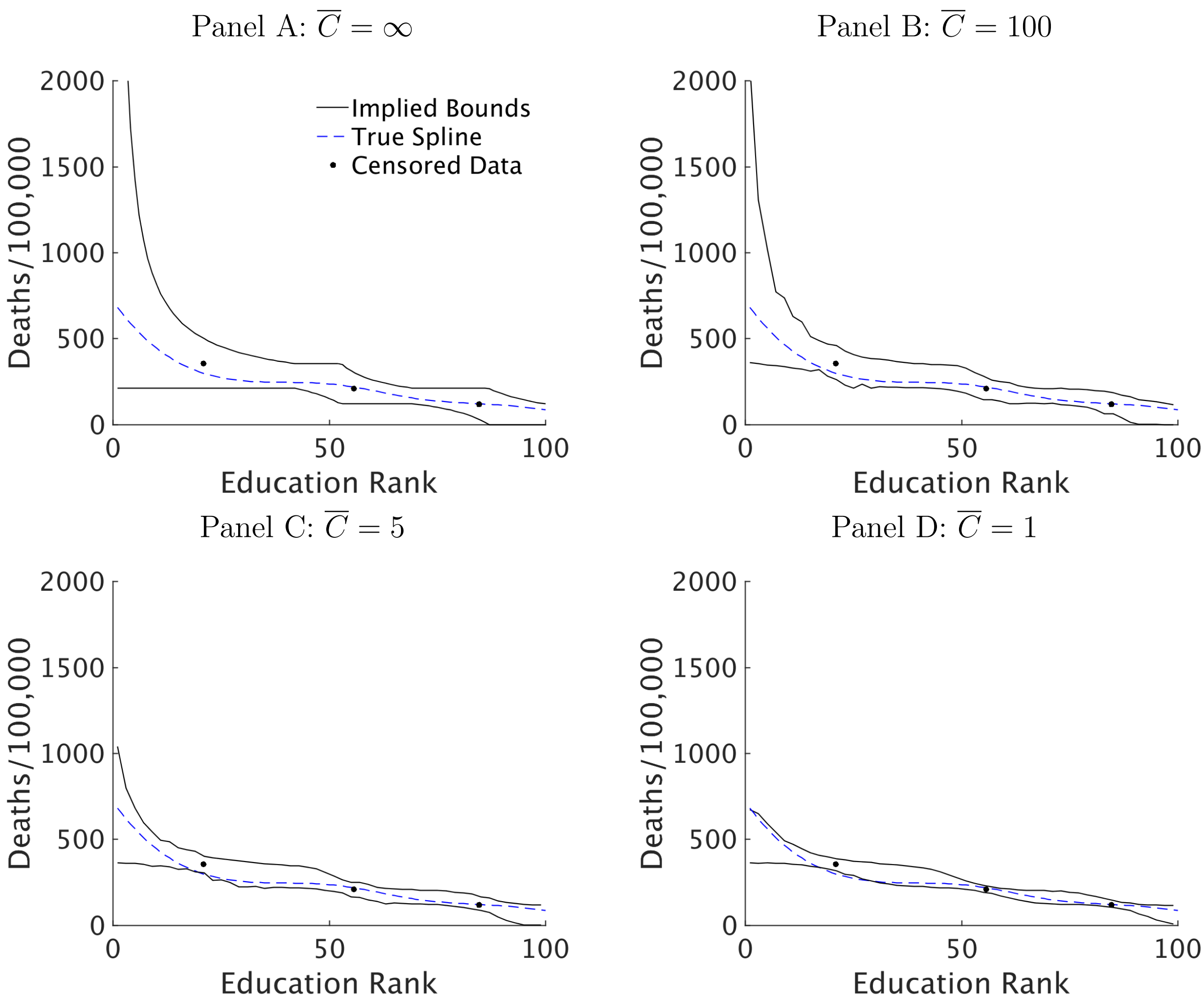
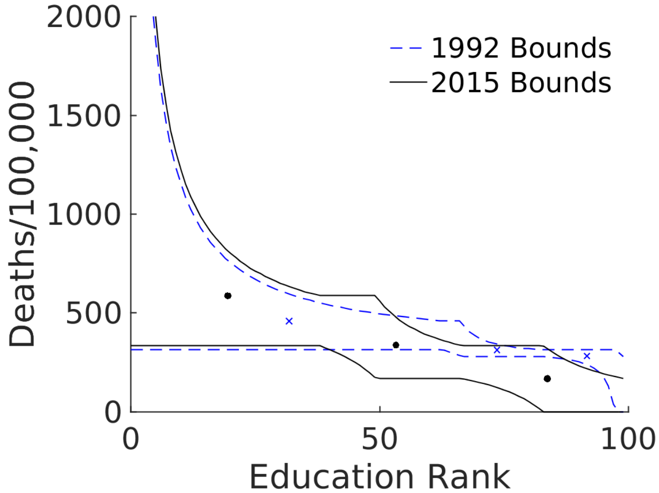


Figure 5 shows results from a simulation using matched mortality-income rank data from Chetty et al. (2016) in 2014 for women aged 52. We simulated interval censoring using the censored mortality-education data for women aged 50–54 in 2014, so that the only observable data were the points in the graphs, indicating the mean mortality in each education bin. Education rank is measured relative to the set of all women aged 50-54 in 2014. We then calculated bounds under four different curvature constraints, indicated in the graph titles. The solid lines show the upper and lower bound of the CEF at each point in the parent distribution, and the dashed line shows the spline fit to the underlying data (described in Figure A1).

Figure 6

Change in Total Mortality of U.S. Women, Age 50-54
Bounds on Conditional Expectation Functions

Panel A: Monotonicity Only



Panel B: $\overline{C} = 5$

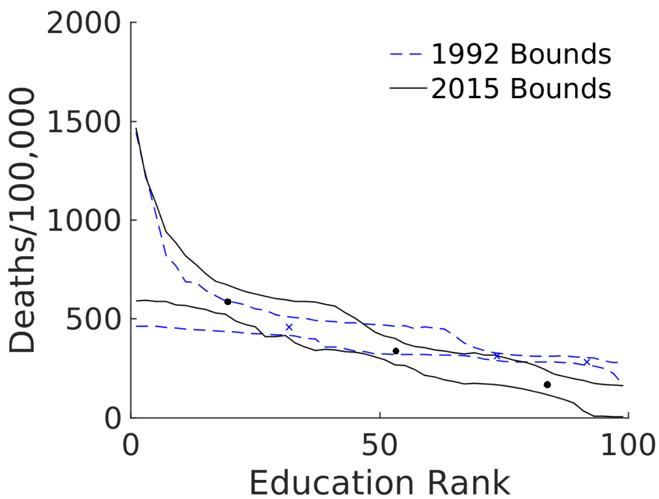


Figure 6 shows bounds on the conditional expectation function of mortality as a function of latent educational rank. The sample consists of U.S. women aged 50-54; mortality is measured in deaths per 100,000 women, and the graph shows the mean across the sample years 1992 and 2015. Panel A shows analytical bounds with no curvature constraint. Panel B uses the rule of thumb curvature constraint suggested by Section II. Education rank is measured relative to the set of all women aged 50-54.

Figure 7

Changes in U.S. Mortality, Age 50-54, 1992-2015:
Constant Rank Interval Estimates (High School or Less in 1992)

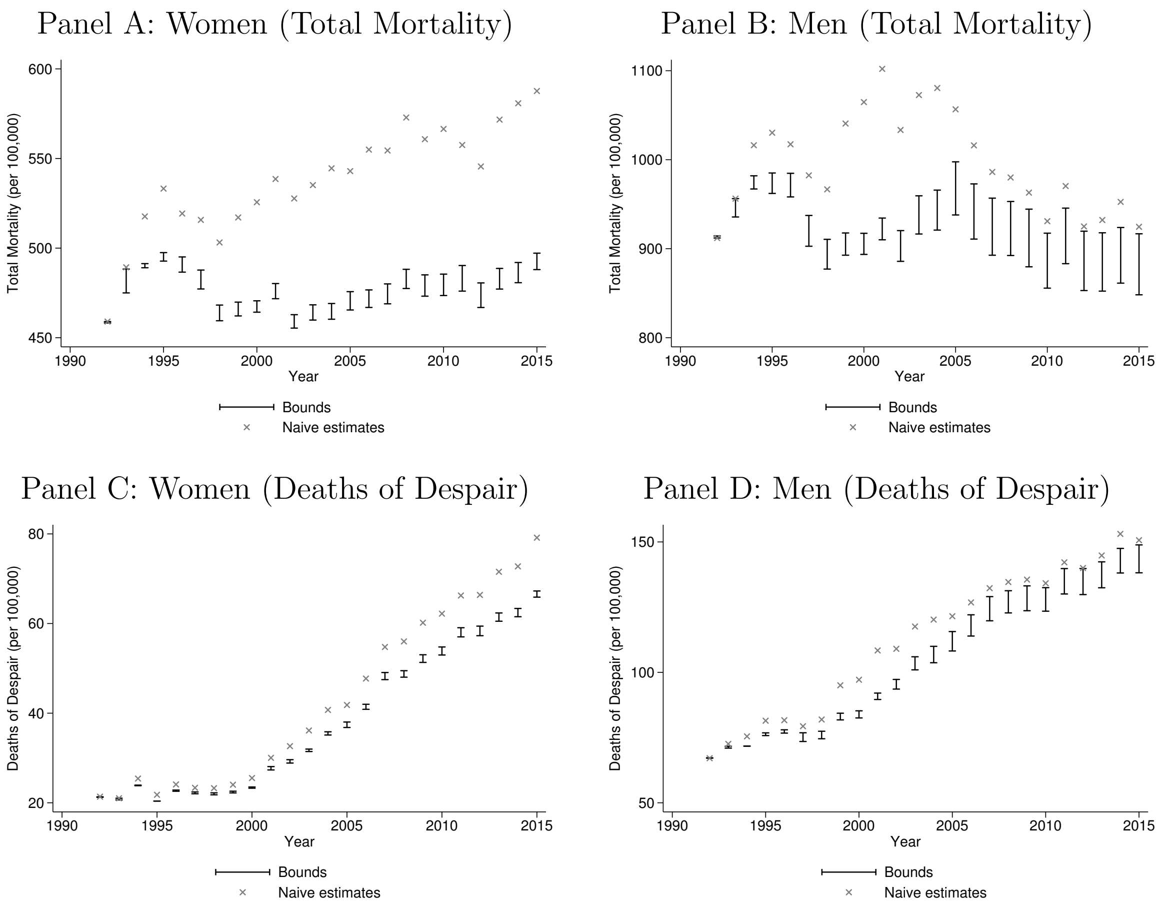
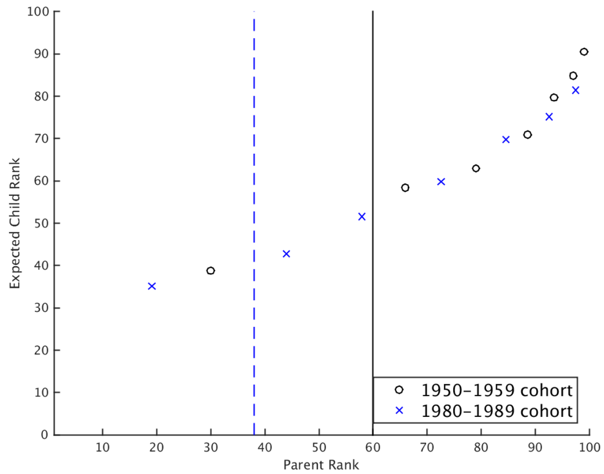


Figure 7 shows bounds on estimates of mortality for men and women aged 50–54 over time. The sample is defined by the set of latent education ranks corresponding to a high school education or less in 1992. This is ranks 0-64 for women, and 0-54 for men. Panel A shows total mortality for women age 50-54, and Panel B shows total mortality for men age 50–54. Panels C and D show mortality from deaths of despair for both groups. The figure presents analytical bounds, and curvature is unconstrained.

Figure 8

Changes in Intergenerational Educational Mobility in India
from 1950s to 1980s Birth Cohorts

Panel A: Rank Bin Midpoints



Panel B: CEF Bounds

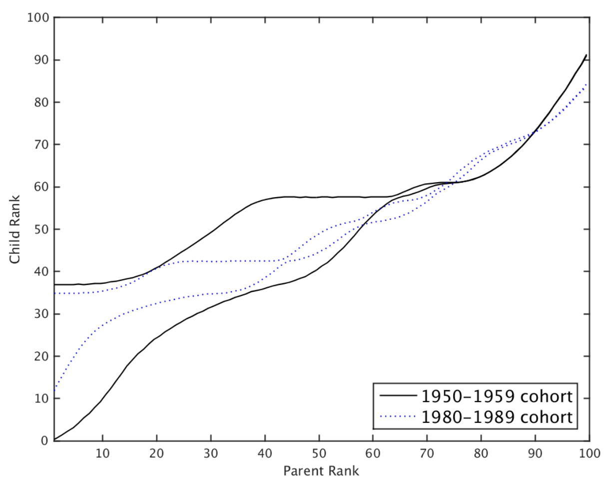
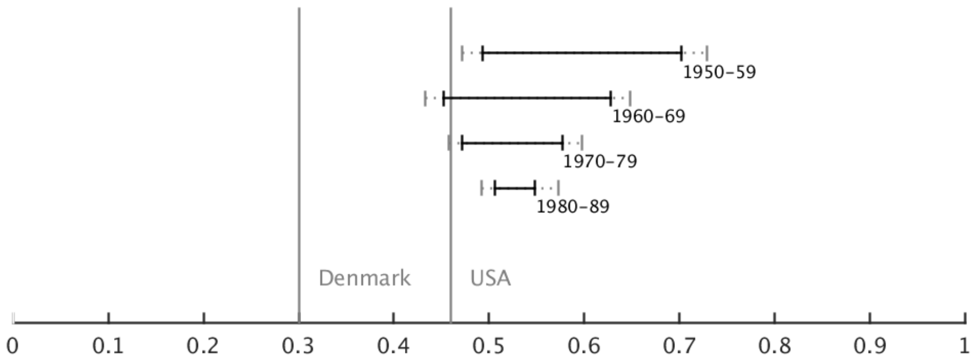


Figure 8 presents the change over time in the rank-rank relationship between Indian fathers and sons born in the 1950s and the 1980s. Panel A presents the raw rank bin means in the data. The vertical lines indicate the size of the lowest parent education rank bin, representing fathers with less than two years of education; the solid line shows this value for the 1950s cohort, and the dashed line for the 1980s cohort. Panels B presents the bounds on the child rank conditional expectation function at each parent rank, under our rule-of-thumb curvature constraint $\overline{C} = 0.10$.

Figure 9
Mobility Bounds for 1950s to 1980s Birth Cohorts

Panel A: Rank-Rank Gradient



Panel B: Absolute and Interval Mobility: p_{25} and μ_0^{50}

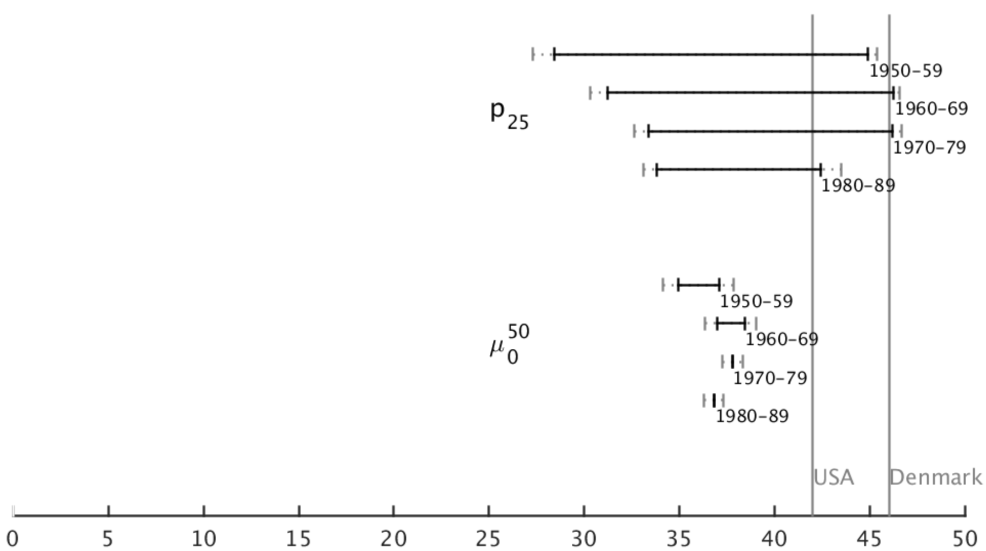


Figure 9 shows bounds on three mobility statistics, estimated on four decades of matched Indian father-son pairs. The solid lines show the estimated bounds on each statistic and the gray dashed lines show the 95% bootstrap confidence sets, based on 1000 bootstrap samples. Each of these statistics was calculated using our preferred curvature constraint of $\overline{C} = 0.10$. For reference, we display the rank-rank education gradient for USA and Denmark (from Hertz (2008)), and p_{25} for USA and Denmark (from Chetty et al. (2014a)). The rank-rank gradient is the slope coefficient from a regression of son education rank on father education rank. p_{25} is absolute upward mobility, which is the expected rank of a son born to a family at the 25th percentile. μ_0^{50} is upward interval mobility, which is the expected rank of a son born below to a family below the 50th percentile.

Table 1

Key Statistics for Mortality of Less Educated Women Ages 50–54

Panel A: 1992

Statistic	$\overline{C} = \infty$	$\overline{C} = 5$	$\overline{C} = 3$	$\overline{C} = 0$
p_{10}: First Quintile Median	[314.02, 1236.14]	[454.58, 841.77]	[455.05, 804.06]	526.35
p_{25}: Bottom Half Median	[314.02, 683.09]	[387.37, 593.07]	[419.37, 556.95]	479.25
p_{32}: Median $\leq$ High School (1992)	[314.02, 602.38]	[325.63, 556.13]	[344.08, 538.52]	457.28
p_{19}: Median $\leq$ High School (2015)	[314.02, 799.57]	[427.79, 660.63]	[404.32, 650.63]	498.09
μ_0^{20}: First Quintile Mean	[459.02, 775.54]	[460.31, 755.43]	[461.26, 760.20]	526.35
μ_0^{50}: Bottom Half Mean	[459.02, 498.63]	[465.75, 495.99]	[467.11, 498.46]	479.25
μ_0^{64}: Mean $\leq$ High School (1992)	[458.24, 459.02]	[459.02, 459.02]	[459.02, 459.02]	457.28
μ_0^{38}: Mean $\leq$ High School (2015)	[459.02, 556.92]	[466.21, 555.47]	[466.76, 549.86]	498.09

Panel B: 2015

Statistic	$\overline{C} = \infty$	$\overline{C} = 5$	$\overline{C} = 3$	$\overline{C} = 0$
p_{10}: First Quintile Median	[335.15, 1313.50]	[573.58, 810.45]	[581.27, 805.93]	640.81
p_{25}: Bottom Half Median	[335.15, 726.72]	[391.60, 644.25]	[417.14, 626.64]	542.58
p_{32}: Median $\leq$ High School (1992)	[335.15, 641.09]	[341.81, 597.27]	[359.63, 592.98]	496.74
p_{19}: Median $\leq$ High School (2015)	[335.15, 850.31]	[479.03, 698.19]	[489.73, 678.46]	581.87
μ_0^{20}: First Quintile Mean	[587.66, 824.81]	[588.46, 825.06]	[593.25, 814.47]	640.81
μ_0^{50}: Bottom Half Mean	[531.01, 587.66]	[535.08, 569.87]	[535.78, 565.08]	542.58
μ_0^{64}: Mean $\leq$ High School (1992)	[488.17, 497.33]	[490.82, 498.08]	[490.25, 496.40]	496.74
μ_0^{38}: Mean $\leq$ High School (2015)	[587.66, 592.87]	[589.59, 592.30]	[589.10, 592.93]	581.87

Table 1 presents bounds on mobility statistics under progressively tighter curvature constraints. The last column in each panel presents point estimates obtained from the best linear approximation to the CEF given the mean mortality observed in each bin. p_i is the value of the CEF at i; μ_a^b is the average value of the CEF between points a and b. Panel A presents the statistics for women in 1992. Panel B presents the statistics for women in 2015.

Table 2

Mortality at Different Income Ranks:
Simulation Results

Statistic	Value from Linear Fit	True Value	$\overline{C} = \infty$	$\overline{C} = 3$
p_{10}	393.4	444.9	$[213.7, 797.3]$	$[352.6, 519.5]$
p_{25}	337.1	272.3	$[213.7, 447.1]$	$[285.6, 393.7]$
p_{32}	310.8	251.0	$[213.7, 396.1]$	$[236.3, 367.3]$
p_{19}	359.6	315.4	$[213.7, 520.9]$	$[323.8, 429.5]$
μ_0^{20}	393.4	456.7	$[361.4, 505.5]$	$[362.2, 500.1]$
μ_0^{50}	337.1	335.3	$[330.4, 361.4]$	$[332.2, 353.7]$
μ_0^{64}	310.8	307.6	$[304.9, 312.5]$	$[306.4, 311.7]$
μ_0^{39}	357.7	361.4	$[361.4, 363.3]$	$[361.9, 364.9]$
β	-3.8	-4.1	$[-3.3, -4.2]$	$[-3.4, -4.1]$

Table 2 presents bounds on mortality statistics computed in a simulation exercise. We begin with mortality-income rank data on women aged 52 in 2014 from Chetty et al. (2016) and compute the best-fit spline to the data to obtain a close estimate of the true CEF for this distribution. We then simulate interval censoring according to the education bins for women aged 50–54 in 2014, where women are ranked within their own group. We then compute bounds on mortality statistics obtained data with simulated censoring. p_i is the value of the CEF at i; μ_a^b is the average value of the CEF between points a and b; β is the slope of the best linear approximation of the CEF.

Table 3

Unadjusted and Constant Rank Estimates of Mortality Changes for Women 50-54, 1992 to 2015

Age	≤ High School		Some College		B.A. or Higher	
	Unadjusted Estimate	Constant Rank Bounds	Unadjusted Estimate	Constant Rank Bounds	Unadjusted Estimate	Constant Rank Bounds
25-29	42.8	[16.1, 24.1]	9.6	[-28.0, 6.3]	-8.8	[-28.6, -8.8]
30-34	46.7	[17.6, 27.4]	22.2	[-22.6, 21.3]	-9.0	[-42.5, -9.0]
35-39	33.1	[8.9, 26.3]	24.4	[-39.0, 24.4]	-17.0	[-53.9, -17.0]
40-44	47.6	[16.4, 44.2]	35.8	[-55.7, 35.8]	-34.5	[-82.9, -34.5]
45-49	74.9	[22.3, 46.6]	33.7	[-80.4, 33.7]	-56.9	[-127.5, -56.9]
50-54	128.6	[30.0, 40.1]	21.1	[-145.5, 21.1]	-111.8	[-272.8, -111.8]
55-59	84.6	[-47.4, -37.2]	39.3	[-168.6, 39.3]	-158.4	[-398.9, -158.4]
60-64	5.0	[-177.8, -168.5]	-89.7	[-372.7, -89.7]	-242.1	[-625.3, -242.1]
65-69	-82.7	[-266.8, -293.7]	-135.9	[-475.9, -328.0]	-627.1	[-1093.1, -627.1]

Table 3 presents bias in estimates of the mortality increase. We present the original estimate of the change in mortality in the bottom bin, computed by taking the mean mortality in the group of women receiving a high school education or less in 2015 and subtracting the mean mortality in 1992. We then obtain the largest rank in the bottom bin, which we denote b; for example, for 1992 women aged 50–54, this would be 64. We present the set-identified interval of bounds on changes in μ_0^b, the average value of the CEF from ranks 0 to b. These estimates are computed with analytical bounds and no curvature constraint. Panel A presents changes in total mortality. Panel B presents changes in deaths of despair Case and Deaton (2017) — deaths from suicide, poisoning and liver disease.

Table 4

Bounds on Intergenerational Educational Mobility in India

	$\overline{C} = \infty$			$\overline{C} = 0.20$		
Cohort	Gradient	p_{25}	μ_0^{50}	Gradient	p_{25}	μ_0^{50}
1950-59	[0.457, 0.742]	[13.0, 58.3]	[34.8, 38.7]	[0.474, 0.722]	[25.5, 48.1]	[34.8, 37.9]
	(0.447, 0.763)	(10.2, 59.8)	(34.0, 39.0)	(0.464, 0.745)	(24.2, 48.3)	(34.0, 38.4)
1960-69	[0.436, 0.655]	[22.2, 54.5]	[37.0, 39.1]	[0.444, 0.639]	[29.3, 49.3]	[37.0, 38.8]
	(0.421, 0.677)	(19.7, 54.8)	(36.3, 39.5)	(0.429, 0.661)	(28.0, 49.5)	(36.3, 39.3)
1970-79	[0.463, 0.595]	[29.0, 48.6]	[37.8, 37.8]	[0.468, 0.584]	[32.2, 48.3]	[37.8, 37.8]
	(0.455, 0.616)	(26.8, 49.7)	(37.3, 38.0)	(0.461, 0.603)	(31.9, 49.1)	(37.3, 38.0)
1980-89	[0.500, 0.565]	[32.3, 42.3]	[36.8, 36.8]	[0.505, 0.556]	[33.3, 42.8]	[36.8, 36.8]
	(0.488, 0.591)	(30.2, 43.6)	(36.4, 37.3)	(0.492, 0.582)	(32.8, 43.6)	(36.4, 37.3)

	$\overline{C} = 0.10$			$\overline{C} = 0$		
Cohort	Gradient	p_{25}	μ_0^{50}	Gradient	p_{25}	μ_0^{50}
1950-59	[0.492, 0.702]	[28.4, 44.9]	[34.9, 37.1]	0.587	35.6	35.6
	(0.480, 0.727)	(27.2, 45.3)	(34.0, 37.8)	(0.577, 0.595)	(35.4, 35.9)	(35.4, 35.9)
1960-69	[0.452, 0.629]	[31.2, 46.3]	[37.0, 38.5]	0.538	36.8	36.8
	(0.436, 0.629)	(31.1, 46.5)	(36.9, 38.9)	(0.530, 0.553)	(36.5, 37.0)	(36.5, 37.0)
1970-79	[0.472, 0.577]	[33.4, 46.2]	[37.8, 37.8]	0.534	36.9	36.9
	(0.465, 0.597)	(32.7, 46.5)	(37.3, 38.0)	(0.524, 0.549)	(36.6, 37.2)	(36.6, 37.2)
1980-89	[0.506, 0.548]	[33.8, 42.4]	[36.8, 36.8]	0.537	36.8	36.8
	(0.494, 0.575)	(33.3, 43.1)	(36.3, 37.3)	(0.523, 0.551)	(36.5, 37.2)	(36.5, 37.2)

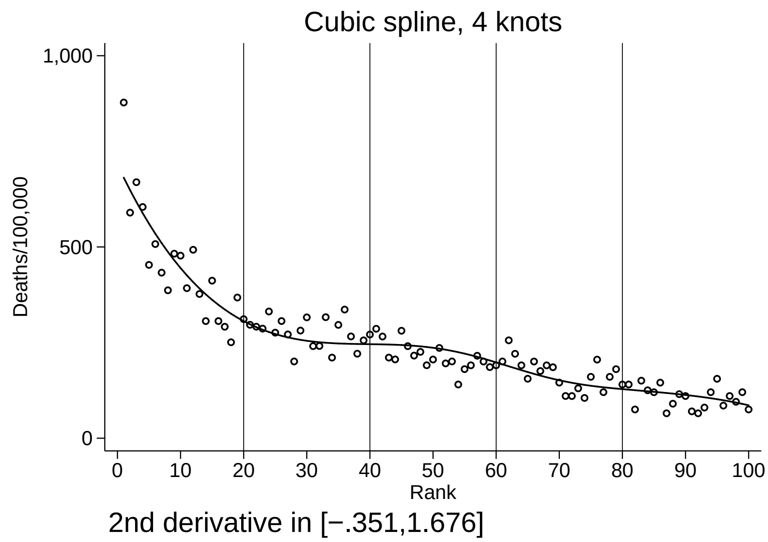
The table shows estimates of bounds on three scalar mobility statistics, for different decadal cohorts and under different restrictions $\overline{C}$ on the curvature of the child rank conditional expectation function given parent rank. The rank-rank gradient is the slope coefficient from a regression of son education rank on father education rank. p_{25} is absolute upward mobility, which is the expected rank of a son born to a family at the 25th percentile. μ_0^{50} is upward interval mobility, which is the expected rank of a son born below to a family below the 50th percentile. When $\overline{C} = 0$, the bounds shrink to point estimates. Bootstrap 95% confidence sets are displayed in parentheses below each estimate based on 1000 bootstrap samples.

A Appendix A: Additional Tables and Figures

Figure A1

Spline Approximations to Empirical Mortality-Income CEFs

Panel A: 52 Year-Old Women in 2014



Panel B: 52 Year-Old Men in 2014

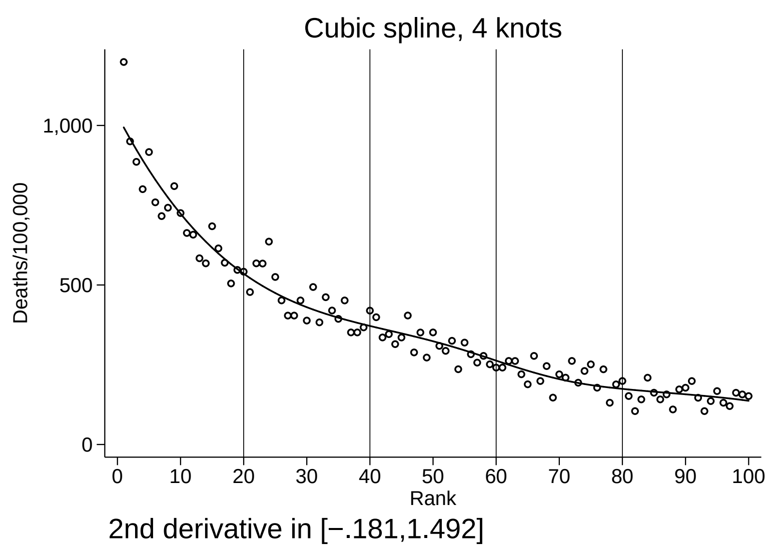


Figure A1 presents estimates of the conditional expectation functions obtained from fully supported mortality income-rank distribution in the United States. The data come from Chetty et al. (2016). The CEFs were fitted using cubic splines, with knots at every ten quantiles (as indicated by the vertical lines). The functions plot the best cubic spline fit to each series, and the circles plot the underlying data.

Figure A2

Spline Approximations to Empirical Parent-Child Rank Distributions

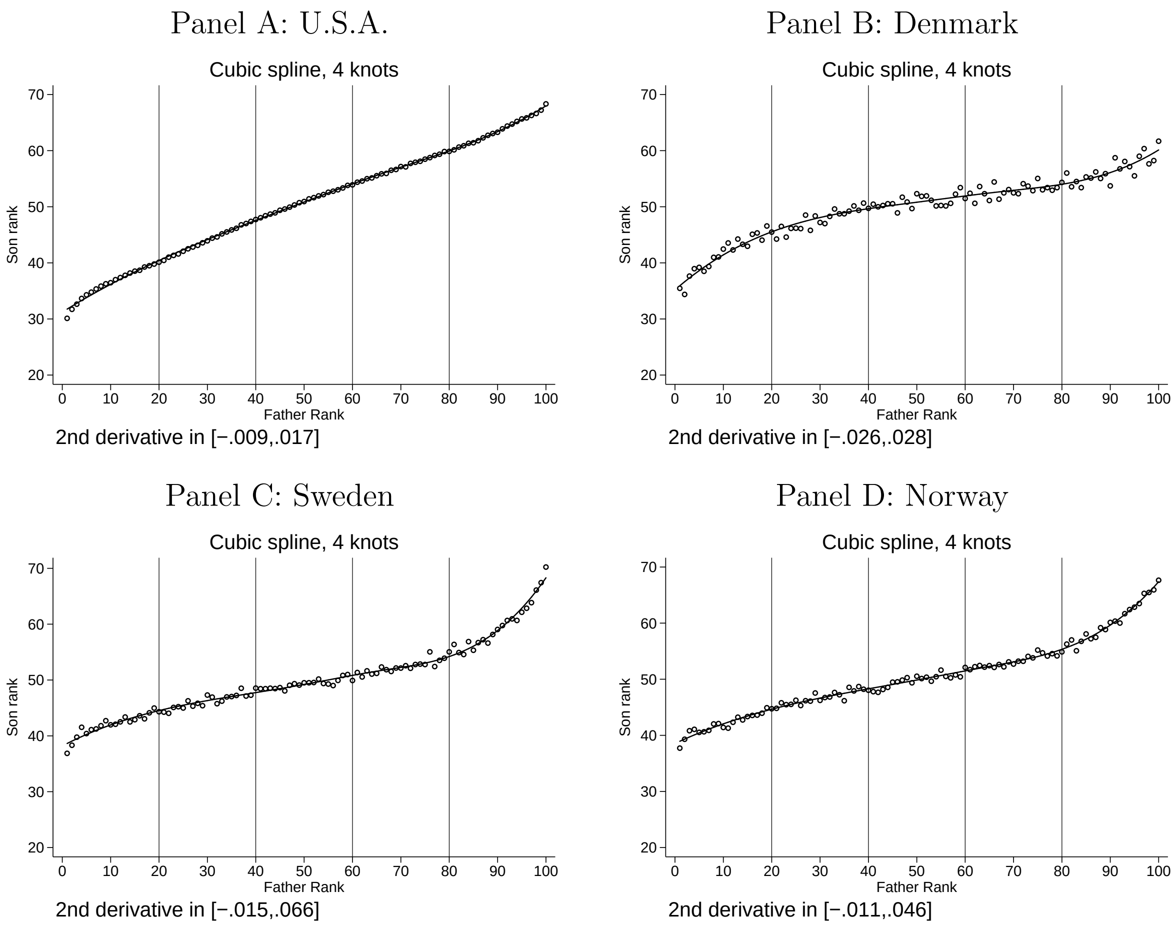


Figure A2 presents estimates of the conditional expectation functions obtained from fully supported parent-child rank-rank income distribution in several developed countries. The data for U.S.A. and Denmark come from Chetty et al. (2014a), who obtained the Denmark data from Boserup et al. (2014). The data for Sweden and Norway come from Bratberg et al. (2015). The CEFs were fitted using cubic splines, with knots at 20, 40, 60, and 80 (as indicated by the vertical lines). The functions plot the best cubic spline fit to each series, and the circles plot the underlying data. Underneath each graph, we present the range of the second derivative across the support and in the domain 10–90.

Table A1

Bin Sizes in Studies of Intergenerational Mobility

Study	Country	Birth Cohort of Son	Number of Parent Outcome Bins	Population Share in Largest Bin
Aydemir and Yazici (2016)	Turkey	1990[38]	15	39%
	Turkey	1960[39]	15	78%
Dunn (2007)	Brazil	1972–1981	> 18	20%[40]
Emran and Shilpi (2011)	Nepal, Vietnam	1992-1995	2	83%
Güell et al. (2013)	Spain	$\sim$ 2001	9	27%[41]
Guest et al. (1989)	USA	$\sim$ 1880	7	53.2%
Hnatkovska et al. (2013)	India	1918-1988	5	Not reported
Knight et al. (2011)	China	1930–1984	5	29%[42]
Lindahl et al. (2012)	Sweden	1865-2005	8	34.5%
Long and Ferrie (2013)	Britain	$\sim$ 1850	4	57.6%
	Britain	$\sim$ 1949-55	4	54.2%
	USA	$\sim$ 1850-51	4	50.9%
	USA	$\sim$ 1949-55	4	48.3%
Piraino (2015)	South Africa	1964–1994	6	36%

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

[39] Includes all people born after about 1990.

[40] Includes all people born after about 1960.

[41] 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.

[42] Estimate is from the full population rather than just fathers.

[43] This reported estimate does not incorporate sampling weights; estimates with weights are not reported.

Table A2

Transition Matrices for Fathe 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 A2 shows transition matrices by decadal birth cohort for Indian fathers and sons in the study.

B Appendix B: Proofs

Proof of Proposition 1. Define the conditional expectation function of child rank given parent rank as $Y(i) = E(y|i)$. Let the function $Y(i)$ be defined on $i \in [0, 100]$, and assume $Y(i)$ is integrable. We want to bound $E(y|i)$ when i is known to lie in the interval $[i_k, i_{k+1}]$; there are K such intervals. Define the expected child outcome in bin k as

$$r_k = \int_{i_k}^{i_{k+1}} Y(i) f_k(i) di.$$

Note that

$$r_k = E(y|i \in [i_k, i_{k+1}])$$

via the law of iterated expectations. Define $r_0 = 0$ and $r_{K+1} = 100$.

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

$$P(i \in [i_k, i_{k+1}]) = 1. \tag{Assumption I}$$

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

$$E(y|i \text{ is interval censored }) = E(y|i). \tag{Assumption MI}$$

From Manski and Tamer (2002), we have:

$$r_{k-1} \leq E(y|i) \leq r_{k+1} \tag{Manski-Tamer bounds}$$

Suppose also that

$$i \sim U(0, 100). \tag{Assumption U}$$

In that case,

$$r_k = \frac{1}{i_{k+1} - i_k} \int_{i_k}^{i_{k+1}} Y(i)$$

, substituting the probability distribution function for the uniform distribution within bin k. Then we derive the following proposition.

Proposition 1. *Let i be in bin k. Under assumptions IMMI (Manski and Tamer, 2002) and U, and without additional information, the following bounds on $E(y|i)$ are sharp:*

$$\begin{cases} r_{k-1} \leq E(y|i) \leq \frac{1}{i_{k+1}-i} \left((i_{k+1} - i_k) \, r_k - (i - i_k) \, r_{k-1} \right), & i < i_k^* \\ \frac{1}{i-i_k} \left((i_{k+1} - i_k) \, r_k - (i_{k+1} - i) \, r_{k+1} \right) \leq E(y|i) \leq r_{k+1}, & i \geq i_k^* \end{cases}$$

where

$$i_k^* = \frac{i_{k+1} r_{k+1} - (i_{k+1} - i_k) \, r_k - i_k r_{k-1}}{r_{k+1} - r_{k-1}}.$$

The intuition behind the proof is as follows. First, find the function z which meets the bin mean and is defined as r_{k-1} up to some point j. Because z is a valid CEF, the lower bound on $E(y|i)$ is no larger than z up to j; we then show that j is precisely i_k^* from the statement. For points $i > i_k^*$, we show that the CEF which minimizes the value at point i must be a horizontal line up to i and a horizontal line at r_{k+1} for points larger than i. But there is only one such CEF, given that the CEF must also meet the bin mean, and we can solve analytically for the minimum value the CEF can attain at point i. We focus on lower bounds for brevity, but the proof for upper bounds follows a symmetric structure.

Part 1: Find i_k^*. First define $\mathcal{V}_k$ as the set of weakly increasing CEFs which meet the bin mean. Put otherwise, let $\mathcal{V}_k$ be the set of $v : [i_k, i_{k+1}] \to \mathbb{R}$ satisfying

$$r_k = \frac{1}{i_{k+1} - i_k} \int_{i_k}^{i_{k+1}} v(i) di.$$

Now choose $z \in \mathcal{V}_k$ such that

$$z(i) = \begin{cases} r_{k-1}, & i_k \leq i < j \\ r_{k+1}, & j \leq i \leq i_{k+1}. \end{cases}$$

Note that z and j both exist and are unique (it suffices to show that just j exists and is unique, as then z must be also). We can solve for j by noting that z lies in $\mathcal{V}_k$, so it must meet the bin mean. Hence, by evaluating the integrals, j must satisfy:

$$\begin{aligned} r_k &= \frac{1}{i_{k+1} - i_k} \int_{i_k}^{i_{k+1}} z(i) di \\ &= \frac{1}{i_{k+1} - i_k} \left(\int_{i_k}^{j} r_{k-1} di + \int_{j}^{i_{k+1}} r_{k+1} di \right) \\ &= \frac{1}{i_{k+1} - i_k} \left((j - i_k) \, r_{k-1} + (i_{k+1} - j) \, r_{k+1} \right). \end{aligned}$$

Note that these expressions invoke assumption U, as the integration of $z(i)$ does not require any adjustment for the density on the i axis. For a more general proof with an arbitrary distribution of i, see section B.

With some algebraic manipulations, we obtain that $j = i_k^*$.

Part 2: Prove the bounds. In the next step, we show that i_k^* is the smallest point at which no $v \in \mathcal{V}_k$ can be r_{k-1}, which means that there must be some larger lower bound on $E(y|i)$ for $i \geq i_k^*$. In other words, we prove that

$$i_k^* = \sup \Big\{ i | \text{ there exists } v \in \mathcal{V}_k \text{ such that } v(i) = r_{k-1}. \Big\}.$$

We must show that i_k^* is an upper bound and that it is the least upper bound.

First, i_k^* is an upper bound. Suppose that there exists $j' > i_k^*$ such that for some $w \in \mathcal{V}_k$, $w(j') = r_{k-1}$. Observe that by monotonicity and the bounds from Manski and Tamer (2002), $w(i) = r_{k+1}$ for $i \leq j'$; in other words, if $w(j')$ is the mean of the mean of the prior bin, it can be no lower or higher than the mean of the prior bin up to point j'. But since $j' > j$, this means that

$$\int_{i_k}^{j'} w(i) di < \int_{i_k}^{j'} z(i) di,$$

since $z(i) > w(i)$ for all $h \in (j, j')$. But recall that both z and w lie in $\mathcal{V}_k$ and must therefore meet the bin mean; i.e.,

$$\int_{i_k}^{i_{k+1}} w(i) di = \int_{i_k}^{i_{k+1}} z(i) di.$$

But then

$$\int_{j'}^{i_{k+1}} w(i) di > \int_{j'}^{i_{k+1}} z(i) di.$$

That is impossible by the bounds from Manski and Tamer (2002), since $w(i)$ cannot exceed r_{k+1}, which is precisely the value of $z(i)$ for $i \geq j$.

Second, j is the least upper bound. Fix $j' < j$. From the definition of z, we have shown that for some $h \in (j', j)$, $z(h) = r_{k-1}$ (and $z \in \mathcal{V}_k$). So any point j' less than j would not be a lower bound on the set — there is a point h larger than j' such that $z(h) = r_{k-1}$.

Hence, for all $i < i_k^*$, there exists a function $v \in \mathcal{V}_k$ such that $v(i) = r_{k-1}$; the lower bound on $E(y|i)$ for $i < i_k^*$ is no greater than r_{k-1}. By choosing z' with

$$z'(i) = \begin{cases} r_{k-1}, & i_k \leq i \leq j \\ r_{k+1}, & j < i \leq i_{k+1}, \end{cases}$$

it is also clear that at i_k^*, the lower bound is no larger than r_{k-1} (and this holds in the proposition itself, substituting in i_k^* into the lower bound in the second equation).

Now, fix $i' \in (i_k^*, i_{k+1}]$. Since i_k^* is the supremum, there is no function $v \in \mathcal{V}_k$ such that $v(i') = r_{k-1}$. Thus for $i' > i_k^*$, we seek a sharp lower bound larger than r_{k-1}. Write this lower bound as

$$Y_{i'}^{min} = \min \Big\{ v(i') \text{ for all } v \in \mathcal{V}_k \Big\},$$

where $Y_{i'}^{min}$ is the smallest value attained by any function $v \in \mathcal{V}_k$ at the point i'.

We find this $Y_{i'}^{min}$ by choosing the function which maximizes every point after i', by attaining the value of the subsequent bin. The function which minimizes $v(i')$ must be a horizontal line up to this point.

Pick $\tilde{z} \in \mathcal{V}_k$ such that

$$\tilde{z}(i) = \begin{cases} \underline{Y}, & i_k \leq i' \\ r_{k+1}, & i' < i_{k+1} \leq i_{k+1} \end{cases}.$$

By integrating $\tilde{z}(i)$, we claim that $\underline{Y}$ satisfies the following:

$$\frac{1}{i_{k+1} - i_k} \left((i' - i_k)\, \underline{Y} + (i_{k+1} - i')\, r_{k+1} \right) = r_k.$$

As a result, $\underline{Y}$ from this expression exists and is unique, because we can solve the equation. Note that this integration step also requires that the distribution of i be uniform, and we generalize this argument in B.

By similar reasoning as above, there is no $Y' < \underline{Y}$ such that there exists $w \in \mathcal{V}_k$ with $w(i') = Y'$. Otherwise there must be some point $i > i'$ such that $w(i') > r_{k+1}$ in order that w matches the bin means and lies in $\mathcal{V}_k$; the expression for $\underline{Y}$ above maximizes every point after i', leaving no additional room to further depress $\underline{Y}$.

Formally, suppose there exists $w \in \mathcal{V}_k$ such that $w(i') = Y' < \underline{Y}$. Then $w(i') < \tilde{z}(i')$ for all $i < i'$, since w is monotonic. As a result,

$$\int_{i_k}^{i'} \tilde{z}(i) di > \int_{i_k}^{i'} w(i) di.$$

But recall that

$$\int_{i_k}^{i_{k+1}} w(i) di = \int_{i_k}^{i_{k+1}} \tilde{z}(i) di,$$

so

$$\int_{i'}^{i_{k+1}} w(i) di > \int_{i'}^{i_{k+1}} \tilde{z}(i) di.$$

This is impossible, since $\tilde{z}(i) = r_{k+1}$ for all $i > i'$, and by Manski and Tamer (2002), $w(i) \leq r_{k+1}$ for all $w \in \mathcal{V}_k$. Hence there is no such $w \in \mathcal{V}_k$, and therefore $\underline{Y}$ is smallest possible value at i', i.e. $\underline{Y} = Y_{i'}^{min}$.

By algebraic manipulations, the expression for $\underline{Y} = Y_i^{min}$ reduces to

$$Y_i^{min} = \frac{(i_{k+1} - i_k)\, r_k - (i_{k+1} - i) r_{k+1}}{i - i_k}, \; i \geq i_k^*.$$

The proof for the upper bounds uses the same structure as the proof of the lower bounds.

Finally, the body of this proof gives sharpness of the bounds. For we have introduced a CEF $v \in \mathcal{V}_k$ that obtains the value of the upper and lower bound for any point $i \in [i_k, i_{k+1}]$. For any value y within the bounds, one can generate a CEF $v \in \mathcal{V}_k$ such that $v(i) = y$. □

Proof of Proposition 2. Suppose we relax assumption U and merely characterize i by some known probability density function. Then we can derive the following bounds.

Proposition 2. *Let i be in bin k. Let $f_k(i)$ be the probability density function of i in bin k. Under assumptions IMMI (Manski and Tamer, 2002), and without additional information, the following bounds on*

$E(y|i)$ are sharp:

$$\begin{cases} r_{k-1} \leq E(y|i) \leq \frac{r_k - r_{k-1}\int_{i_k}^{i} f_k(s)ds}{\int_i^{i_{k+1}} f_k(s)ds}, & i < i_k^* \\ \frac{r_k - r_{k+1}\int_i^{i_{k+1}} f_k(s)ds}{\int_{i_k}^{i} f_k(s)ds} \leq E(y|i) \leq r_{k+1}, & i \geq i_k^* \end{cases}$$

where i_k^* satisfies:

$$r_k = r_{k-1}\int_{i_k}^{i_k^*} f_k(s)ds + r_{k+1}\int_{i_k^*}^{i_{k+1}} f_k(s)ds.$$

The proof follows the same argument as in proposition 1. With an arbitrary distribution, $\mathcal{V}_k$ now constitutes the functions $v : [i_k, i_{k+1}] \to \mathbb{R}$ which satisfy:

$$\int_{i_k}^{i_{k+1}} v(s)f_k(s)ds = r_k.$$

As before, choose $z \in \mathcal{V}_k$ such that

$$z(i) = \begin{cases} r_{k-1}, & i_k \leq i < j \\ r_{k+1}, & j \leq i \leq i_{k+1}. \end{cases}$$

Because the distribution of i is no longer uniform, in that case, j must satisfy

$$\begin{aligned} r_k &= \int_{i_k}^{i_{k+1}} z(s)f_k(s)ds \\ &= r_{k-1}\int_{i_k}^{j} f_k(s)ds + r_{k+1}\int_{j}^{i_{k+1}} f_k(s)ds. \end{aligned}$$

This implies that $j = i_k^*$, precisely.

The rest of the arguments follow identically, except we now claim that for $i > i_k^*$, $\underline{Y} = Y_i^{min}$ satisfies the following:

$$r_k = \int_{i_k}^{i} Y_i^{min} f_k(s)ds + \int_{i}^{i_{k+1}} r_{k+1}f_k(s)ds.$$

By algebraic manipulations, we obtain:

$$Y_i^{min} = \frac{r_k - r_{k+1}\int_i^{i_{k+1}} f_k(s)ds}{\int_{i_k}^{i} f_k(s)ds}$$

and the proof of the lower bounds is complete. As before, the proof for upper bounds follows from identical logic. □

Proof of Proposition 3. Define

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

Let Y_i^{min} and Y_i^{max} be the lower and upper bounds respectively on $E(y|i)$ given by proposition 1. We seek to bound μ_a^b when i is observed only in discrete intervals.

Proposition 3. *Let $b \in [i_k, i_{k+1}]$ and $a \in [i_h, i_{h+1}]$. Let assumptions IMMI (Manski and Tamer, 2002) and U hold. Then, if there is no additional information available, the following bounds are sharp:*

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

The order of the proof is as follows. If a and b lie in the same bin, then μ_a^b is maximized only if the CEF is minimized prior to a. As in the proof of proposition 1, that occurs when the CEF is a horizontal line at Y_i^{min} up to a, and a horizontal line Y_i^{max} at and after a. If a and b lie in separate bins, the value of the integral in bins that are contained between a and b is determined by the observed bin means. The portions of the integral that are not determined are maximized by a similar logic, since they both lie within bins. We prove the bounds for maximizing μ_a^b, but the proof is symmetric for minimizing μ_a^b.

Part 1: Prove the bounds if a and b lie in the same bin. We seek to maximize μ_a^b when $a, b \in [i_k, i_{k+1}]$. This requires finding a candidate CEF $v \in \mathcal{V}_k$ which maximizes $\int_a^b v(i)di$. Observe that the function $v(i)$ defined as

$$v(i) = \begin{cases} Y_a^{min}, & i_k \leq i < a \\ Y_a^{max}, & a \leq i \leq i_{k+1} \end{cases}$$

has the property that $v \in \mathcal{V}_k$. For if $a \geq i_k^*$, $v = \tilde{z}$ from the second part of the proof of proposition 1. If $a < i_k^*$, the CEF in $\mathcal{V}_k$ which yields Y_a^{max} is precisely v (by a similar argument which delivers the upper bounds in proposition 1).

This CEF maximizes μ_a^b, because there is no $w \in \mathcal{V}_k$ such that

$$\frac{1}{b-a}\int_a^b w(i)di > \frac{1}{b-a}\int_a^b v(i)di.$$

Note that for any $w \in \mathcal{V}_k$, $\frac{1}{i_{k+1}-i_k}\int_{i_k}^{i_{k+1}} w(i)di = \frac{1}{i_{k+1}-i_k}\int_{i_k}^{i_{k+1}} v(i)di = r_k$. Hence in order that $\int_a^b w(i)di > \int_a^b v(i)di$, there are two options. The first option is that

$$\int_{i_k}^{a} w(i)di < \int_{i_k}^{a} v(i)di.$$

That is impossible, since there is no room to depress w given the value of v after a. If $a < i_k^*$, then it is clear that there is no w giving a larger μ_a^b, since $r_{k-1} \leq w(i)$ for $i_{k-1} \leq i \leq a$, so w is bounded below by v. If $a \geq i_k^*$, then $v(i) = r_{k+1}$ for all $a \leq i \leq i_{k+1}$. That would leave no room to depress w further; if $\int_{i_k}^a w(i)di < \int_{i_k}^a v(i)di$, then $\int_a^{i_{k+1}} w(i)di > \int_a^{i_{k+1}} v(i)di$, which cannot be the case if $v = r_{k+1}$, by the bounds given in Manski and Tamer (2002).

The second option is that

$$\int_b^{i_k} w(i)di < \int_b^{i_k} v(i)di.$$

This is impossible due to monotonicity. For if $\int_a^b w(i)di > \int_a^b v(i)di$, then there must be some point $i' \in [a, b)$ such that $w(i') > v(i')$. By monotonicity, $w(i) > v(i)$ for all $i \in [i', i_{k+1}]$ since $v(i) = Y_a^{max}$ in that interval. As a result,

$$\int_b^{i_k} w(i)di > \int_b^{i_k} v(i)di,$$

since $b \in (i', i_{k+1})$. (If $b = i_{k+1}$, then only the first option would allow w to maximize the desired μ_a^b.)

Therefore, there is no such w, and v indeed maximizes the desired integral. Integrating v from a to b, we obtain that the upper bound on μ_a^b is $\frac{1}{b-a}\int_a^b Y_a^{max} di = Y_a^{max}$. Note that there may be many functions which maximize the integral; we only needed to show that v is one of them.

To prove the lower bound, use an analogous argument.

Part 2: Prove the bounds if a and b do not lie in the same bin. We now generalize the set up and permit $a, b \in [0, 100]$. Let $\mathcal{V}$ be the set of weakly increasing functions such that $\frac{1}{i_{k+1}-i_k}\int_{i_k}^{i_{k+1}} v(i)di = r_k$ for all $k \leq K$. In other words, $\mathcal{V}$ is the set of functions which meet the means of every bin. Now observe that for all $v \in \mathcal{V}$,

$$\begin{aligned}\mu_a^b &= \frac{1}{b-a}\int_a^b v(i)di \\ &= \frac{1}{b-a}\left(\int_a^{i_{h+1}} v(i)di + \int_{i_{h+1}}^{i_k} v(i)di + \int_{i_k}^{b} v(i)di\right),\end{aligned}$$

by a simple expansion of the integral.

But for all $v \in \mathcal{V}$,

$$\int_{i_{h+1}}^{i_k} v(i)di = \sum_{\lambda=h+1}^{k-1} r_\lambda(i_{\lambda+1} - i_\lambda)$$

if $h+1 < k$ and

$$\int_{i_{h+1}}^{i_k} v(i)di = 0$$

if $h+1 = k$. For in bins completely contained inside $[a, b]$, there is no room for any function in $\mathcal{V}$ to vary; they all must meet the bin means.

We proceed to prove the upper bound. We split this into two portions: we wish to maximize $\int_a^{i_{h+1}} v(i)di$ and we also wish to maximize $\int_{i_k}^{b} v(i)di$. The values of these objects are not codependent. But observe that the CEFs $v \in \mathcal{V}_k$ which yield upper bounds on these integrals are the very same functions which yield upper bounds on $\mu_a^{i_{h+1}}$ and $\mu_{i_k}^b$, since $\mu_s^t = \frac{1}{t-s}\int_s^t v(i)di$ for any s and t. Also notice that a and i_{h+1} both lie in bin h, while b and i_k both lie in bin k, so we can make use of the first portion of this proof.

In part 1, we showed that the function $v \in \mathcal{V}$, $v : [i_h, i_{h+1}] \to \mathbb{R}$, which maximizes $\mu_a^{i_{h+1}}$ is

$$v(i) = \begin{cases} Y_a^{min}, & i_h \le i < a \\ Y_a^{max}, & a \le i \le i_{h+1}. \end{cases}$$

As a result

$$\max_{v\in\mathcal{V}}\left\{\int_a^{i_{h+1}} v(i)di\right\} = \int_a^{i_{h+1}} Y_a^{max}di = Y_a^{max}(i_{h+1} - a).$$

Similarly, observe that i_k and b lie in the same bin, so the function $v : [i_k, i_{k+1}] \to \mathbb{R}$, with $v \in \mathcal{V}$ which maximizes $\int_{i_k}^b v(i)di$ must be of the form

$$v(i) = \begin{cases} Y_{i_k}^{min}, & i_k \le i < a \\ Y_{i_k}^{max}, & b \le i \le i_{k+1}. \end{cases}$$

With identical logic,

$$\max_{v\in\mathcal{V}}\left\{\int_{i_k}^{b} v(i)di\right\} = \int_{i_k}^{b} Y_{i_k}^{max}di = Y_{i_k}^{max}(b - i_k).$$

And by proposition 1, $i_k \le i_k^*$ so $Y_{i_k}^{max} = r_k$. (Note that if $i_k = i_k^*$, substituting i_k^* into the second expression of proposition 1 still yields that $Y_{i_k}^{max} = r_k$.)

Now we put all these portions together. First let $h+1 = k$. Then $\int_{i_{h+1}}^{i_k} v(i)di = 0$, so we have that we maximize μ_a^b by

$$\frac{1}{b-a}\left(Y_a^{max}(i_{h+1} - a) + r_k(b - i_k)\right).$$

Similarly, if $h+1 < k$ and there are entire bins completely contained in $[a, b]$, then we maximize μ_a^b by

$$\frac{1}{b-a}\left(Y_a^{max}(i_{h+1} - a) + \sum_{\lambda=h+1}^{k-1} r_\lambda(i_{\lambda+1} - i_\lambda) + r_k(b - i_k)\right).$$

The lower bound is proved analogously. Sharpness is immediate, since we have shown that the CEF which delivers the endpoints of the bounds lies in $\mathcal{V}$. As a result, there is a function delivering any intermediate value for the bounds. $\square$

C Appendix C: CEF Bounds When X and Y are Interval Censored

In the main part of the paper, we focus on bounding a function $Y(i) = E(y|i)$ when y is observed without error, but i is observed with interval censoring. In this section, we modify the setup to consider simultaneous interval censoring in the conditioning variable i and in observed outcomes y. This arises, for example, in the study of educational mobility, where latent education ranks of both parents and children are interval censored.

We first present a setup that takes a similar approach to the bounding method presented in Section II. We can define bounds on the CEF $E(y|i)$ when both Y and i are interval-censored as a solution to a constrained optimization problem. The problem has a number of parameters that is an order of magnitude higher than that in Section II, and proved too computationally intensive to solve in the Indian test case (where interval censoring is severe). We therefore present a sequential approach that yields theoretical bounds on the double-censored CEF for the case of intergenerational mobility.

Specifically, we develop the theoretical best- and worst-case distributions of censored y variables for a given intergenerational mobility statistic. Our best- and worst-case assumptions yield two different sets of joint parent-child bin means. We then use the method in Section II to calculate bounds on the mobility statistic under each case. The union of these bounds is a conservative bound on the mobility statistic given censoring in both the y and i variables.

Finally, we can shed light on the distribution of the true parameter within these bounds if other data is available. In the context of intergenerational mobility, and in our specific empirical context, it is frequently the case that more information is available about children than about their parents. We use data on child wages to predict whether the true latent child rank distribution (y) is better represented by the best- or worst-case mobility scenario. The joint wage distribution suggests that the true latent distribution is very close to the best case distribution, which we used in Section V, because there is little effect of parent education on child wages after conditioning on child education.

C.A Solution Definition for CEF Bounds with Double Censoring

We are interested in bounding a function $E(y|i)$, where y is known only to lie in one of H bins defined by intervals of the form $[y_h, y_{h+1}]$, and i is known only to lie in one of K bins defined by intervals of the form $[i_k, i_{k+1}]$. For simplicity, we focus on the case where both y and i are uniformly distributed on the interval $[0, 100]$.[44]

We can define the joint distribution of y and i by allowing the latent cumulative distribution function

[44]Taking a different known distribution into account would require imposing different weights on the mean-squared error function and budget constraint below, but would otherwise not be substantively different.

of y to depend on the conditioning variable i:

$$F(r,i) = P(y \leq r | I = i) \tag{C.1}$$

The CEF $E(y|i)$ is thus given by:

$$E(y|i) = \int_0^{100} r f(i,r) dr \tag{C.2}$$

where $f(i,r)$ is the probability density function corresponding to the cumulative distribution function in Equation C.1, when the conditioning variable takes the value i. Note that r in this case represents a child rank. This expression simply denotes that $E(y|i)$ is the average value from 0 to 100 on the y-axis, holding i fixed.

We do not observe the sample analog of $F(i,r)$ directly. Rather, we observe the sample analog of the following expression for each of $H * K$ bin combinations:

$$P\left(y \leq y_{h+1} \,\middle|\, i \in [i_k, i_{k+1}]\right) = \frac{1}{i_{k+1} - i_k} \int_{q=i_k}^{i_{k+1}} F(q, y_{h+1}) dq \tag{C.3}$$

We denote this sample analog $\hat{P}\left(y \leq y_{h+1} \,\middle|\, i \in [i_k, i_{k+1}]\right)$ as $\hat{R}(k,h)$. Equation C.3 states that the probability that y is less than y_{h+1} is the average value of the cumulative distribution function in that bin. Since i is uniform, we can write its probability distribution function within the bin as $\frac{1}{i_{k+1}-i_k}$.

We parameterize each cumulative distribution function as $F(i,r) = S(i,r,\gamma_i)$, where r is the outcome variable, i is the conditioning variable, and γ_i is a parameter vector in some parameter space G_i. Similarly let $f(i,r) = s(i,r,\gamma_i)$. In our numerical calculation, we define G_i as $[0,1]^{100}$, a vector which gives the value of the cumulative distribution function at each of 100 conditioning variable percentiles on the y-axis. Put otherwise, holding i fixed, we seek the 100-valued column vector γ_i which contains the value of the cumulative distribution function at each of the 100 possible y values: $y = 1, y = 2, \dots, y = 100$. As a result, γ_i must lie within $[0,1]^{100}$. Note that there are as many vectors γ_i as there are possible values for the conditioning variable i. If we discretize also i as $1, 2, \dots, 100$, then the we define the matrix of 100 cumulative distribution functions, indexed by i, as $\boldsymbol{\gamma^{100}} = [\gamma_1 \ \gamma_2 \ \dots \gamma_{100}]$. To be explicit, $\boldsymbol{\gamma^{100}}$ is a 100×100 matrix constructed by setting its i^{th} column as γ_i. We write that $\boldsymbol{\gamma^{100}} \in G^{100}$.

We also introduce a new monotonicity condition for this context. In this set up, monotonicity implies that the outcome distribution for any value of i first-order stochastically dominates the outcome distribution at any lower value of i. Put otherwise,

$$s(i, r, g_i) \text{ is weakly decreasing in } i \qquad \text{(Monotonicity)}$$

In the mobility context, this statement implies that the child rank distribution of a higher-ranked parent stochastically dominates the child rank distribution of a lower-ranked parent.[45,46]

In that case, the following minimization problem defines the set of feasible values of γ_i for each value i:

$$\Gamma = \underset{\boldsymbol{g} \in \boldsymbol{G^{100}}}{\text{argmin}} \left[\sum_{k=1}^{K} \sum_{h=1}^{H} \left(\int_{q=i_k}^{ik+1} S(q, h, g_k) dq - \hat{R}(k, h) \right)^2 \right] \tag{C.4}$$

$$\text{such that}$$

$$s(i, r, g_i) \text{ is weakly decreasing in } i \qquad \text{(Monotonicity)}$$

$$\sum_{k=1}^{K} (i_{k+1} - i_k) S(k, r, g_k) = r \qquad \text{(Budget Constraint)}$$

$$S(i, 0, g_i) = 0 \qquad \text{(End Points)}$$

$$S(i, 100, g_i) = 1.$$

In the above minimization problem, $\boldsymbol{g}$ is a candidate vector satisfying the conditions, each g_i refers to the candidate cumulative distribution function holding i fixed. A valid set of cumulative distribution functions is one that minimizes error with respect to all of the observed data points and obeys the monotonicity condition. The budget constraint requires that the weighted sums of cumulative distribution functions across all conditioning groups must add up to the population cumulative distribution function. For example, $J\%$ of children must on average attain less than or equal to the J^{th} percentile. The constraints on the end points of the cumulative distribution function are redundant given the other constraints, but are included to highlight how the end points constrain the set of possible outcomes. For simplicity, we have not included a curvature constraint, but such a constraint would be a sensible further restriction on the feasible parameter space in many contexts.

Once a set of candidate cumulative distribution functions have been identified, they have a one-to-one correspondence with the CEF given an interval censored conditioning variable, (described by Equation C.2),

[45] Dardanoni et al. (2012) find that a similar conditional monotonicity holds in almost all mobility tables in 35 countries.

[46] A stronger monotonicity assumption would require that the hazard function is decreasing in i. This is equivalent to stating that the cumulative distribution function must be weakly decreasing in i *conditional upon i being above some value.*

and thus with any function of the CEF. These statistics can be numerically bounded as in Section II.

This problem is computationally more challenging than the problem of censoring only in the conditioning variable dealt with in Section II. In the case of the rank distribution, if we discretize both outcome and conditioning variables into 100 separate percentile bins, then the problem has 10,000 parameters and 10,000 constraints, and an additional 9800 curvature constraint inequalities if desired. This problem proved computationally too difficult to resolve. Restricting the set of discrete bins (e.g. to deciles) is unsatisfying because it requires significant rounding of the raw data which could substantively affect results. We proceed by taking advantage of additional information about mobility distributions.

C.B Best and Worst Case Mobility Distributions

Our goal is to bound some kind of mobility function given interval censored data on both parent and child ranks. In this section, we take a sequential approach to the double-censoring problem. We use additional information about the structure of the mobility problem to obtain worst- and best-case parent cumulative distribution functions for intergenerational mobility. From these cumulative distribution functions, we can obtain worst- and best-case CEFs using Equation C.2. First, we calculate bounds on the average value of the child rank in each child rank * parent rank cell. We then apply the methods from Section II on the best and worst case bounds; the union of resulting bounds describes the bounds on the mobility statistic of interest.

In the body of the paper, we obtained the average son rank within each parent rank bin to obtain the moments. In this appendix, we obtain two new sets of bin means — one corresponding to the best-case, and one corresponding to the worst-case. By taking the union of our bounds on the CEF for each of these moments, we can obtain a new, conservative, set of bounds on the CEF that takes into account censoring in the y variable.

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

Consider a simple 2x2 case. In 1960s India, among the set of boys who attained less than two years of education (the bottom 27% of the child distribution) 55% have fathers with less than two years of education, and 45% have fathers with two or more years of education. We do not observe how the children of each parent group are distributed *within* the bottom 27%. For this case, mobility will be lowest if children of the least educated parents occupy the bottom ranks of this bottom bin, or ranks 0 through 15, and children of more educated parents occupy ranks 16 through 27. Mobility will be maximized if parental education has no relationship with rank, conditional on the child rank bin. We do not consider the case of perfectly reversed sorting, where the children of the least educated parents occupy the highest ranks within each child rank bin, as it would violate the stochastic dominance condition (and is implausible).

Appendix Figure C1 shows two set of cumulative distribution functions that correspond to these two scenarios for the 1960–69 birth cohort. In Panel A, children's ranks are perfectly sorted according to parent education within bins. Each line shows the cumulative distribution function of child rank, given some father education. The points on the graph correspond to the observations in the data—the value of each cumulative distribution function is known at each of this points. Children below the 27th percentile are in the lowest observed education bin. Within this bin, the cumulative distribution function for children with the least educated parents is concave, and the CDF for children with the most educated parents is convex—indicating that children from the best off families have the highest ranks within this bin. This pattern is repeated within each child bin. Panel B presents the high mobility scenario, where children's outcomes are uniformly distributed within child education bins, and are independent of parent education within child bin.

For each of these cumulative distribution functions, we can calculate the expected value of child rank given parent rank (Equation C.2). From these expected values, we can then calculate bounds on any mobility statistic, as in Sections II and V. Table C1 shows the expected child rank by parent education for the high and low mobility scenario, as well as bounds on the mobility statistics discussed above. Taking censoring in the child distribution into account widens the bounds on all parameters. The effect is proportionally the greatest on the interval mean measure, because it was so precisely estimated before—the bounds on μ_0^{50} approximately double in width when censoring of son data is taken into account.

Note that these bounds are very conservative, as the worst case scenario is unlikely to reflect the true uncensored joint parent-child rank distribution, due to the number and sharpness of kinks in the cumulative distribution functions in Panel A of Figure C1. A curvature constraint on the cumulative distribution function would move the set of feasible solutions closer to those in the uniform case, narrowing the mobility bounds toward the high mobility scenario. We next draw on additional data on children, which suggests that the best case mobility scenario is close to the true joint distribution.

C.C Estimating the Child Distribution Within Censored Bins

Because we have additional data on children, we can estimate the shape of the child cumulative distribution function within parent-child education bins using rank data from other outcome variables that are not censored. Under the assumption that the latent education rank distribution is correlated with other socioeconomic rank distributions, we can shed light on whether Panel A or Panel B in Figure C1 better describes the true solution.

Figure C2 tests this hypothesis using wage data from men in the 1960s birth cohort. To generate this figure, we calculate children's ranks first according to education, and then according to wage ranks within each education bin.[47] The solid lines depict this uncensored rank distribution for each father education; the dashed gray lines overlay the estimates from the high mobility scenario in Panel B of Figure C1.

If parent education strongly predicted child wages *within* each child education bin, we would see a graph like Panel A of Figure C1. The data clearly reject this hypothesis. There is some additional curvature in the expected direction in some bins, particularly among the small set of college-educated children, but the distribution of child cumulative distribution functions is strikingly close to the high mobility scenario, where father education has little predictive power over child outcomes after child education is taken into account. The last row of Table C1 shows mobility estimates using the within-bin parent-child distributions that are predicted by child wages; the mobility estimates are nearly identical to the high mobility scenario. This result supports the assumption made in Section V that latent child rank within a child rank bin is uncorrelated with parent rank.

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

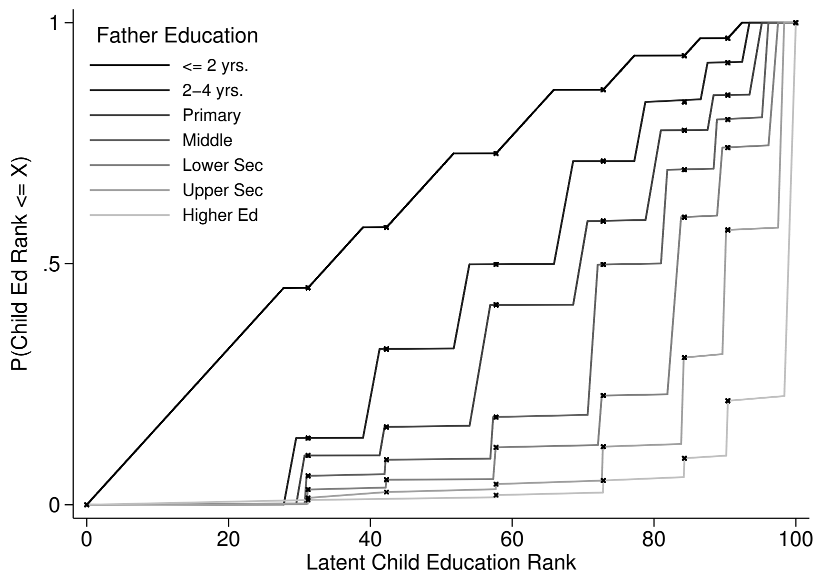
One additional factor that makes censoring in the child distribution a smaller concern is the fact that children are more educated than parents in every cohort, and thus the size of the lowest education bin is smaller for children than for parents. This result is likely to be true in many other countries where education is rising. Of course, in other contexts, we may lack additional information about the distribution of the i and y variable within bins, and researchers may prefer to work with conservative bounds as described in C.B.

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

Figure C1

Best- and Worst-Case Son Cumulative Distribution Functions
by Father Education (1960-69 Birth Cohort)

Panel A: Lowest Feasible Mobility



Panel B: Highest Feasible Mobility

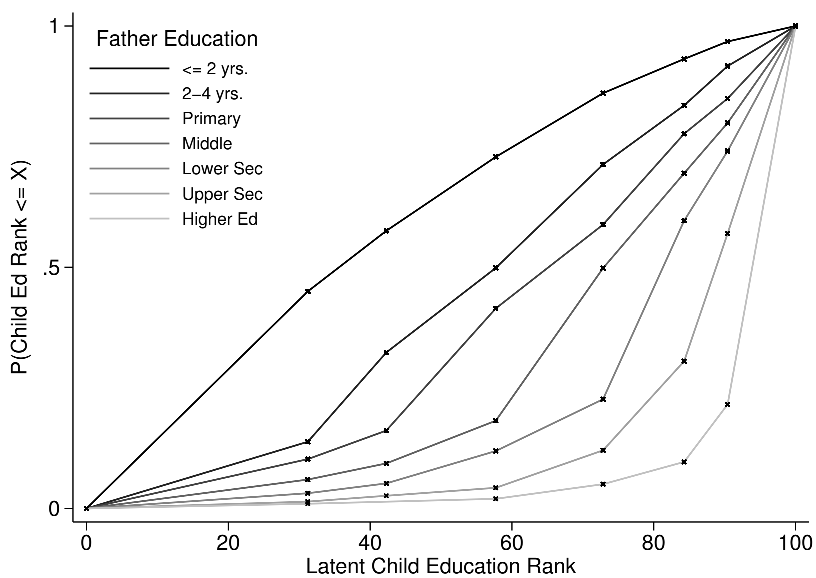


Figure C1 shows bounds on the child education rank cumulative distribution function, separately for each father education group. The lines index father types. Each point on a line shows the probability that a child of a given father type obtains an education rank less than or equal to the value on the X axis in the national education distribution. The large markers show the points observed in the data.

Figure C2

Son Outcome Rank Cumulative Distribution Function
by Father Education (1960-69 Birth Cohort)

Joint Education/Wage Estimates

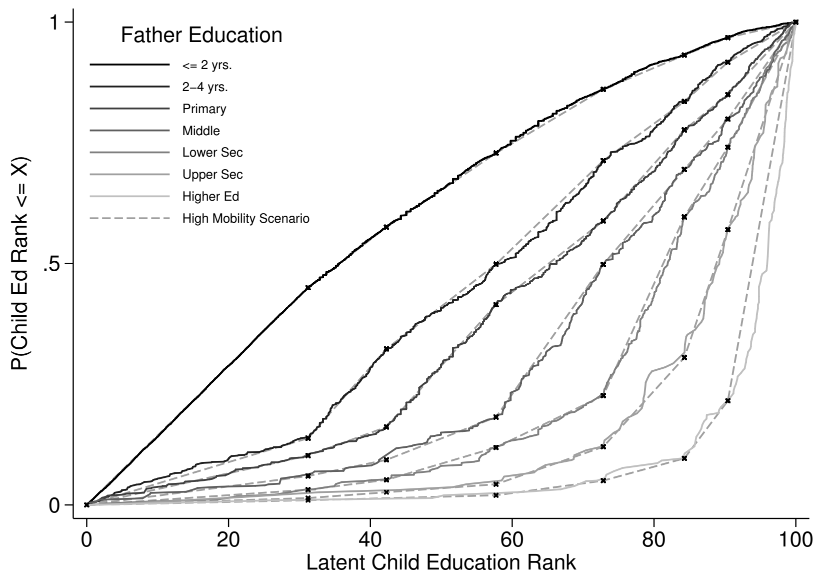


Figure C2 plots separate son rank cumulative distribution functions separately for each father education group, for sons born in the 1960s in India. Sons are ranked first in terms of education, and then in terms of wages. Data are from the IHDS 2011-12. Sons not reporting wages are dropped. For each father type, the graph shows a child's probability of attaining a given rank percentile in the child education rank distribution.

Table C1

Mobility Estimates under Double-Censored CEF

	Upward Interval Mobility (μ_0^{50})	Absolute Upward Mobility (p_{25})	Rank-Rank Gradient (β)
Low mobility scenario	[32.33, 35.90]	[25.42, 43.08]	[0.55, 0.80]
High mobility scenario	[35.86, 38.80]	[29.69, 45.96]	[0.45, 0.67]
Wage imputation scenario	[35.79, 38.70]	[29.67, 45.78]	[0.46, 0.67]

Table C1 presents bounds on μ_0^{50}, p_{25}, and the rank-rank gradient β under three different sets of assumptions about child rank distribution within child rank bins. The low mobility scenario assumes children are ranked by parent education within child bins. The high mobility scenario assumes parent rank does not affect child rank after conditioning on child bin. The wage imputation predicts the within-bin child rank distribution using child wage ranks and parent education.

D Appendix D: Data Sources

D.A Data on Mortality in the United States

For comparability, we attempt to follow very closely the data construction procedure used in Case and Deaton (2017). We are grateful that the authors share software for data construction on their paper's website to simplify this process.

Death records come from the CDC WONDER database. We have deaths counts by race, gender, and education from 1992–2015, as well as information on cause of death. To obtain mortality rates by year, we obtain the number of people in each age-race-gender-education cell from the Current Population Survey.

The mortality records are characterized by some data with missing education. We follow standard practice in assuming that the education data are missing at random; we assign the missings the educations of the observed educations in the age-race-gender cell whose deaths we observe in that year. For example, suppose that in a given age-race-gender-year cell which contains death counts, 20% of the data are missing, 40% have a high school degree or less, 30% have some college, and 10% have a BA or more. We assign 10% of the missings to high school or less, 7.5% to some college, and 2.5% to a BA or more.

Case and Deaton (2017) drop several states that inconsistently report education. Geographies are not provided in the public mortality data for after 2005, so Case and Deaton must obtain private geographic identifiers from the National Center for Health Statistics. We have not obtained such identifiers. Hence we simply apply the imputation procedure described above for all people.

The death records contain the universe of deaths in the U.S. The CPS only interviews people who are not institutionalized — e.g., not in a prison or health institution. As a result, the denominator used by Case and Deaton (2017) is slightly smaller than the true denominator. To account for people who are institutionalized, we obtain the number of institutionalized people missing from the CPS in the U.S. Census for 1990 and 2000, and the American Communities Survey for 2005–2015. For non-Census years prior to 2005, we linearly impute the number of institutionalized people in each age-race-gender-education cell; e.g., for 1995, we take the midpoint of the observed number of institutionalized people in 1990 and 2000. For instance, among women ages 50–54 in 1992, just under 0.4% with a high school degree or less are institutionalized. Among that group, mortality falls from 460.8 to 459.0 once we include institutionalized people in the denominator.

D.B Intergenerational Mobility: Matched Parent-Child Data from India

To estimate intergenerational educational mobility in India, we draw on two databases that report matched parent-child educational attainment. The first is an administrative census dataset describing the education level of all parents and their coresident children. Because coresidence-based intergenerational mobility estimates may be biased, we supplement this with a second dataset with a representative sample of non-coresident father-son pairs. We focus on fathers and sons because we do not have data on non-coresident mothers and/or daughters. This section describes the two datasets.

The Socioeconomic and Caste Census (SECC) was conducted in 2012, to collect demographic and socioeconomic information determining eligibility for various government programs.[48] 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 individual-level data that we use describe age, gender, 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.[49] The SECC provides the education level of every parent and child residing in the same household. Sons

[48] 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.

[49] Additional details of the SECC and the scraping process are described in Asher and Novosad (2017).

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.[50] To ease the computational burden of the analysis, we work with a 1% sample of the SECC, stratified across India's 640 districts.

We 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. Crucially, the IHDS solicits information on the education of fathers of household heads, even if the fathers are not resident, allowing us to fill the gaps in the SECC data. Since the SECC contains data on all coresident fathers and sons, our main mobility estimates 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 appending the two datasets, we can obtain an unbiased and nationally representative estimate of the joint parent-child education distribution.[51] IHDS reports neither the education of non-coresident mothers nor of women's fathers, which is why our estimates are restricted to fathers and sons.

IHDS records completed years of education. To make the two data sources consistent, 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.[52] The loss in precision by downcoding the IHDS is minimal, because most students exit school at the end of a completed schooling level.

[50]The categories are (i) illiterate; (ii) literate without primary (iii) primary; (iv) middle; (v) secondary (vi) higher secondary; and (vii) post-secondary.

[51]We verified that IHDS and SECC produce similar point estimates for the coresident father-son pairs that are observed in both datasets. Point estimates from the IHDS alone (including coresident and non-coresident pairs) match our point estimates, albeit with larger standard errors.

[52]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.

We estimate changes in mobility over time by examining the joint distribution of fathers' and sons' educational attainment for sons in different birth cohorts. All outcomes are measured in 2012, but because education levels only rarely change in adulthood, these measures capture educational investments made decades earlier. We use decadal cohorts reflecting individuals' ages at the time of surveying. To allay concerns that differential mortality across more or less educated fathers and sons might bias our estimates, we replicated our analysis on the *same* birth cohorts using the IHDS 2005. By estimating mobility on the same cohort at two separate time periods, we identified 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 (in part because standard errors are largest for the 1950s cohort, making inference more difficult), so survivorship bias among the oldest cohorts does not influence any of our conclusions.

Appendix Table A2 shows the transition matrix across the seven levels of education that can be consistently coded across these datasets, for decadal birth cohorts from the 1950s to the 1980s. Note that these transition matrices are not directly comparable, because the population shares in each row and column are different in each period. As above, we assume that there is an underlying latent education that we observe only in these coarse rank bins.