

A nonparametric analysis of black–white differences in intergenerational income mobility in the United States

DEBOPAM BHATTACHARYA

Department of Economics, University of Oxford

BHASHKAR MAZUMDER

Federal Reserve Bank of Chicago

Lower intergenerational income mobility for blacks is a likely cause behind the persistent interracial gap in economic status in the United States. However, few studies have analyzed black–white differences in intergenerational income mobility and the factors that determine these differences. This is largely due to the absence of appropriate methodological tools. We develop nonparametric methods to estimate the effects of covariates on two measures of mobility. We first consider the traditional transition probability of movement across income quantiles. We then introduce a new measure of upward mobility which is the probability that an adult child’s relative position exceeds that of the parents. Conducting statistical inference on these mobility measures and the effects of covariates on them requires nontrivial modifications of standard nonparametric regression theory since the dependent variables are nonsmooth functions of marginal quantiles or relative ranks. Using National Longitudinal Survey of Youth data, we document that blacks experience much less upward mobility across generations than whites. Applying our new methodological tools, we find that most of this gap can be accounted for by differences in cognitive skills during adolescence.

KEYWORDS. Intergenerational mobility, upward mobility, nonparametric regression, Hadamard differentiability, black–white mobility gap.

JEL CLASSIFICATION. C14, D31.

1. INTRODUCTION

A topic of long standing interest among social scientists is the persistent disadvantage in economic status faced by blacks in the United States many generations after the end of slavery and several decades after the elimination of state sanctioned segregation. The fact that the racial gap is so highly persistent suggests that blacks in the United States may have low rates of upward intergenerational economic mobility. While there is a vast literature on the black–white earnings gap and a growing number of studies on intergenerational mobility (IGM) in the United States, very few studies have examined differences in the rates of IGM between racial groups in the United States and the underlying

Debopam Bhattacharya: debobhatta@gmail.com

Bhashkar Mazumder: bmazumder@frbchi.org

Copyright © 2011 Debopam Bhattacharya and Bhashkar Mazumder. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](http://creativecommons.org/licenses/by-nc/3.0/). Available at <http://www.qeconomics.org>.

DOI: 10.3982/QE69

sources behind those differences. Such an analysis could potentially provide insight into why there is such persistence in racial inequality and whether this may reflect greater inequality of opportunity for blacks. A better understanding of the factors behind differences in IGM could also inform policies intended to address racial gaps such as early life interventions or affirmative action policies. This paper studies racial differences in relative income mobility over generations, using U.S. data from the National Longitudinal Survey of Youth (NLSY), which contains large intergenerational samples of both blacks and whites.

The dearth of studies on racial differences in IGM is due, in part, to the fact that most recent research in this area has focused on using one particular measure, the intergenerational elasticity (IGE), which is simply the regression coefficient obtained by regressing (log) child's permanent income on (log) parents' permanent income.¹ The IGE provides a measure of income persistence, and 1 minus the IGE is widely used as a measure of relative mobility. However, the IGE cannot be used to compare mobility differences between population subgroups with respect to the *entire* distribution. For example, the IGE for blacks only describes the rate at which earnings among black children regress to the black mean—not the mean of the entire distribution.

An alternative approach is to calculate transition probabilities to describe the rates of movement across specific quantiles of the distribution over a generation. Since transition probabilities can measure the movements of blacks across the income distribution of the entire population comprising both blacks and whites, one can make meaningful statements concerning racial differences in mobility.

However, a difficulty arises with transition probabilities if one wants to estimate rates of IGM *conditional* on (continuous) covariates like test scores. For the IGE, it is straightforward to measure effects of covariates: one simply needs to include them along with their interactions with parents' income as additional regressors and the statistical theory is straightforward. In contrast, a formal statistical method for using covariates in transition matrices is lacking. The development of such a methodology would allow one to investigate the underlying mechanisms behind black–white differences in IGM. For example, it is often hypothesized that inadequate parental investment in children's human capital could lead to reduced mobility. Therefore, it is natural to consider the association between IGM and measures of human capital such as education and test scores. Previous research using the NLSY has shown that cognitive skills during adolescence as measured by percentile scores on the Armed Forces Qualifying Test (AFQT) can account for black–white gaps in educational attainment and adult wages (Cameron and Heckman (2001), Neal and Johnson (1996)); so it would be useful to examine whether this result extends to measures of intergenerational income mobility.

To address the above-cited void in the literature, we develop in this paper a nonparametric statistical methodology for analyzing conditional transition probabilities. The relevant inference theory for *marginal* transition probabilities was previously

¹The intergenerational correlation (IGC), used by many researchers, is qualitatively similar and the two measures are equivalent when the variance in income is unchanging across generations.

developed by Formby, Smith, and Zheng (2004). When the relevant covariates are discrete, one can simply apply their results within each covariate category to conduct inference on conditional transitions. But with continuously distributed covariates, the parameters of interest are infinite dimensional and thus nonparametric smoothing methods are warranted.^{2,3} Using a standard parametric model for conditional mobility (e.g., a probit) is problematic here because it is unclear what type of joint distribution of errors will imply a probit form for transition probabilities; in particular, a bivariate normal error distribution does not. Furthermore, a probit is shown below to produce misleading qualitative conclusions in our empirical analysis.

While transition probabilities can be effectively used to compare relative mobility across subgroups, its substantive drawback is its overtly disaggregate nature, that is, there are an infinite number of transition probabilities that depend on which quantile is chosen to be the threshold for the sons, and a summary measure of mobility across *relative* income positions would be useful to consolidate the information provided in transition matrices. Therefore, we introduce a new alternative measure of upward mobility. Specifically, we condition on families where the father's income is at or below a particular percentile (say, the median) and then estimate the probability that the income rank of a son randomly picked from such a family exceeds his parents' rank in the prior generation. It is a single easily interpretable summary measure and its value does not depend on arbitrary discretization of income distributions. Further, as explained below, it is more robust than the transition probability to heterogeneity of income within the reference group of families that we started with. We conduct statistical inference on both the level of this mobility measure and the effects of covariates on that level, using NLSY data. Establishing the relevant distribution theory requires some methodological innovation because IGM measures involve outcome variables that contain nonsmooth functions of initially estimated rank-type functionals. Standard nonparametric regression theory is therefore inadequate for this purpose.

Our first set of empirical results documents the large racial disparity in transition probabilities out of the bottom of the income distribution. For example, we find that blacks are 26 percent less likely to move out of the bottom quartile than are whites. We then show that black–white differences in mobility are much smaller when based on our measure of upward mobility. This is because (i) our measure captures the fact that many blacks exhibit small upward movements in relative ranks which are ignored by transition probabilities and because (ii) black income levels below any given threshold tend to be smaller than white income levels, blacks need much larger absolute gains to surpass a common percentile threshold, leading them to have lower transition probability.

²The AFQT percentile scores take on values from 1 to 99. Given our sample size, analysis by racial group within each percentile of AFQT is very imprecise. Furthermore, small differences in AFQT percentiles, unlike differences in race, are unlikely to imply big changes in functional relationships. Therefore, treating AFQT score as a continuously distributed covariate is the natural and correct approach.

³Trede (1998) showed how continuous covariates can be included in analyzing a class of inequality reducing functionals that have been used to measure *intragenerational* mobility but not for measures based on transition probabilities.

Our most striking finding, however, is that when we (nonparametrically) control for AFQT scores, most of the racial gap in IGM disappears—whether measured via transition probability or upward mobility. We are careful in our interpretation of this finding. First, existing studies suggest that these scores largely reflect accumulated differences in family background and other influences, and are not primarily a measure of innate endowments; indeed, there is evidence that black–white differences in tests scores can be strongly affected by environmental influences (e.g., Neal and Johnson (1996), Hansen, Heckman, and Mullen (2004), Chay, Guryan, and Mazumder (2009)). Second, like most of the existing IGM literature, our analysis is essentially descriptive in nature and one should be cautious in attaching causal interpretations to them. Third, all measures used in the present paper are based on relative positions and do not pertain to movements out of absolute thresholds such as from under the poverty line. However, our finding that there is no racial gap in mobility, conditional on cognitive test scores in adolescence, is new and merits further analysis to establish possible policy implications, such as the importance of early-life interventions relative to, say, access to affirmative action later in life.

The rest of the paper proceeds as follows: Section 2 presents a review of some related papers on black–white differences in IGM; Section 3 describes and discusses the parameters of interest; Section 4 states the asymptotic distribution theory required for conducting statistical inference on these measures; Section 5 describes the NLSY data; Section 6 presents the empirical results; Section 7 concludes. All technical proofs as well as a technical lemma (Lemma 1) on Hadamard differentiability are collected in the [Appendixes](#). Throughout the paper, the symbol $:=$ denotes equality by definition.

2. RELATED LITERATURE

We now provide a brief discussion of some of the literature related to black–white differences in IGM. The broader literature on black–white differences in economic status is beyond the scope of this paper. Fryer (2010) provided a review of this literature. Similarly, Black and Devereux (2010) and Solon (1999) provided extensive surveys of research on IGM. For the most part, studies on black–white earnings differentials have had relatively little to say about IGM. As we discuss above, this is in part due to the lack of appropriate measures for group differences. However, the fact that intergenerational samples of blacks are small has also likely inhibited research. A few studies have addressed the racial dimension of the intergenerational transmission of status, but have not produced direct estimates of *income* mobility. For example, Datcher (1981), using the PSID, regressed adult outcomes on family background characteristics, separately by race and sex and found that black families are not as successful as white families in translating parental economic gains into offspring achievement. Corcoran and Adams (1997) also used the PSID and compared how the probability of escaping poverty differs by race and sex, and how these estimates are affected by adding covariates. They found that black children living in poverty are substantially more likely to be poor as adults and that structural economic factors during childhood (e.g., local unemployment rates) play a key role.

Hertz (2005) was the first investigator to use transition probabilities to directly measure intergenerational income mobility by race. Using the PSID, Hertz estimated very

large racial differentials in the probability of leaving the bottom quartile, with blacks substantially less upwardly mobile. Hertz also found evidence of greater downward mobility among blacks in the probability of leaving the top quartile. Using probit models, Hertz found that these racial differences could not be explained by parental income or education. Given the relatively smaller samples in the PSID and possible concerns about the quality of the data on blacks, it would be useful to produce similar estimates using the NLSY.⁴ In his study, Hertz provided some motivation for our construction of a summary measure of upward mobility. He pointed out that with transition matrices “the problem... is that there is no best way to summarize their content.”

Hertz (2008) developed an alternative estimator for comparing IGM for different groups that takes into account both the within—and the between—group effects, and found that blacks have four time greater earnings persistence using this measure. However, it is not clear how to interpret the scale of this measure, that is, whether it exhibits generally desirable properties and whether it lends itself to conditional estimates.

3. PARAMETERS OF INTEREST

We first describe the parameters of interest based on transition probabilities and then those related to our new measure of upward mobility.

3.1 Conditional transition probabilities

Let $F_0(\cdot)$ and $F_1(\cdot)$ denote the cumulative distribution function (c.d.f.) of the overall income distribution for fathers and sons, respectively. Then the transition probability measures the probability that a son is at or above the s th quantile of F_1 , conditional on his father being between the s_1 th and s_2 th quantiles of $F_0(\cdot)$, that is,

$$\theta(s, (s_1, s_2)) = \frac{\text{Prob}[F_1(Y_1) \geq s, s_1 \leq F_0(Y_0) \leq s_2]}{\text{Prob}[s_1 \leq F_0(Y_0) \leq s_2]}. \quad (1)$$

Notice that $\theta(s, (s_1, s_2))$ can be decomposed by levels of discrete and continuous covariates X such as age and education of the father and/or the son as

$$\begin{aligned} \theta(s, (s_1, s_2)) &= \int \text{Prob}[F_1(Y_1) \geq s | s_1 \leq F_0(Y_0) \leq s_2, X = x] dF(x | s_1 \leq F_0(Y_0) \leq s_2) \\ &:= \int \theta(x; s, (s_1, s_2)) dF(x | s_1 \leq F_0(Y_0) \leq s_2), \end{aligned}$$

where

$$\theta(x; s, (s_1, s_2)) = \text{Prob}[F_1(Y_1) \geq s | s_1 \leq F_0(Y_0) \leq s_2, X = x] \quad (2)$$

⁴Although Hertz made maximum use of the PSID by including the oversample of poorer households, there is some concern about the validity of the sample due to issues related to availability of the initial sampling frame (Lee and Solon (2009)). In addition, there has been substantial attrition of black families in the PSID.

is labeled the *conditional transition probability*. Notice that F_1 and F_0 in (2) still refer to unconditional income distributions in the two generations. To interpret (2) directly, consider a numerical example. Suppose X denotes whether a father is black, $s = 0.5$, $s_1 = 0.4$, and $s_2 = 0.6$. Suppose there is a total of 100 father–son pairs within the (0.4, 0.6) quantiles of the entire income distribution of the fathers' generation. Of these, suppose 40 are black and 60 are white. Suppose 10 of the 40 black sons earn above the s th percentile (in the sons' generation) and 30 of the 60 white sons do the same. Then, the overall transition probability is $\theta(s, (s_1, s_2)) = (10 + 30)/100 = 0.4$. The conditional transition probability among blacks is $\theta(\text{black}; 0.5, (0.4, 0.6)) = 10/40 = 0.25$ and that among whites is $\theta(\text{white}; 0.5, (0.4, 0.6)) = 30/60 = 0.5$. The white minus black difference is 0.25, suggesting higher mobility among whites.

One can analogously define transition probability conditional on more covariates, for example, race and fathers' college attainment, and use it to measure the black–white difference in transition at each value of fathers' education,

$$\tilde{\theta}_B(z; s, (s_1, s_2)) - \tilde{\theta}_W(z; s, (s_1, s_2)),$$

where

$$\begin{aligned} \tilde{\theta}_B(z; s, (s_1, s_2)) &= \text{Prob}[F_1(Y_1) \geq s | s_1 \leq F_0(Y_0) \leq s_2, Z = z, \text{black} = 1], \\ \tilde{\theta}_W(z; s, (s_1, s_2)) &= \text{Prob}[F_1(Y_1) \geq s | s_1 \leq F_0(Y_0) \leq s_2, Z = z, \text{black} = 0]. \end{aligned} \quad (3)$$

To interpret (3), refer to the previous numerical example and let $Z = 1$ denote that the father went to college. Now suppose that among the 40 black fathers, 10 went to college and 30 did not. Of the 10 black sons whose fathers went to college, suppose 6 surpassed the s th quantile, and suppose that of the 30 black fathers who did not attend college, only 4 sons surpassed the s th percentile. Similarly, suppose that among the 60 white fathers, 50 went to college of whom 25 sons surpassed the s th quantile and among 10 white fathers who did not attend college, 5 sons surpassed the s th percentile. Then $\tilde{\theta}_W(\text{college}; 0.5, (0.4, 0.6)) = 25/50 = 0.5$ and $\tilde{\theta}_B(\text{college}; 0.5, (0.4, 0.6)) = 6/10 = 0.6$, and the difference is -0.1 . These hypothetical numbers suggest that although blacks in the (0.4, 0.6) class are less mobile than whites on average, the college educated blacks among them are, in fact, more mobile than college educated whites. In Section 6 we explore racial differences in conditional transition probabilities nonparametrically using actual data on sons' education and AFQT percentile scores.

3.2 Upward mobility

We formally introduce our new measure of upward mobility in this section. We first present the analytic expressions and then discuss the substantive features which make our measure both intuitively appealing and analytically different from measures based on transition probabilities.

Our direct measure of upward mobility is simply the probability that the son's percentile rank in the overall income distribution of his generation exceeds that of his parents' in the income distribution of the parents' generation by a fixed amount. Let Y_0 and

Y_1 denote father's and son's income with respective marginal c.d.f.'s F_0 and F_1 . Then for fixed $0 < s_1 < s_2 < 1$, we define upward mobility for families between the s_1 th and s_2 th quantiles by an extent $\tau \in [0, 1 - s_2]$ as

$$v(\tau, s_1, s_2) = \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | s_1 \leq F_0(Y_0) \leq s_2). \quad (4)$$

One can alternatively define upward mobility by conditioning on $F_0(Y_0) = s$ rather than $s_1 \leq F_0(Y_0) \leq s_2$, and thus avoid aggregation bias due to income heterogeneity within the interval (s_1, s_2) . However, conditioning on $F_0(Y_0) = s$ requires additional smoothing since $F_0(Y_0)$ is continuously distributed. By making the length of the interval (s_1, s_2) small, we both avoid this smoothing and yet remove some of the aggregation bias. Furthermore, defining the reference group to be $s_1 \leq F_0(Y_0) \leq s_2$ is also consistent with how transition probabilities have been traditionally defined in the literature, and thus enables direct comparison and contrast between the two measures.⁵

In analogy with (2) above, one may introduce covariates X into the analysis and define conditional upward mobility at $X = x$ as

$$\begin{aligned} v_c(x; \tau, s_1, s_2) \\ = \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | s_1 \leq F_0(Y_0) \leq s_2, X = x). \end{aligned} \quad (5)$$

The idea is that we start with all families where the father was between the s_1 th and s_2 th percentile. This ensures that all the corresponding sons have equal "space to move up." With these families constituting our population, we then evaluate the extent of upward mobility for different groups defined by values of X . Below, we derive the statistical distribution theory for estimates of $v(\tau, s_2, s_1)$ and $v_c(x; \tau, s_2, s_1)$. In Section 6, we contrast overall upward mobility among blacks versus whites and then analyze how controlling for relevant covariates affects this difference.

Contrasting upward mobility with transition probability A key feature of $v(\tau, s_1, s_2)$ is that it counts the sons' small upward movements in relative positions from their fathers' movements, which are ignored by transition probabilities. Comparing upward mobility

$$v(\tau, s_1, s_2) = \text{Prob}(F_1(Y_1) > F_0(Y_0) + \tau | s_1 \leq F_0(Y_0) \leq s_2)$$

and the transition probability

$$\theta(s_2, (s_1, s_2)) = \text{Prob}[F_1(Y_1) > s_2 | s_1 \leq F_0(Y_0) \leq s_2],$$

one can see that unlike transition probability, $v(\tau, s_1, s_2)$ is counting those sons whose ranks exceeded their fathers' by τ but did not necessarily exceed s_2 . The resulting magnitude of difference between the two measures, however, is an empirical question and depends on the joint distribution of (Y_0, Y_1) .

⁵Conditioning on $F_0(Y_0) = s$, requires averaging over a subjectively determined bandwidth around s , thus entertaining a certain amount of income heterogeneity in finite samples. In addition, the inference theory becomes far more complicated, so we do not pursue this option here.

An advantage of our upward mobility measure can be readily seen when comparing mobility between, say, whites versus blacks. Suppose we condition on fathers with incomes at or below the median income in the fathers' generation, that is, $s_1 = 0$ and $s_2 = 0.5$. Suppose this interval is $(0, \$40,000)$. Then it is reasonable to expect that black incomes are concentrated in the lower end of $(0, \$40,000)$ and white incomes are concentrated in the upper end (see Figure 5 below which compares the c.d.f. of white and black parental income in the bottom quintile). Now, the transition probability counts only those sons who exceed the median income for the sons' generation so that black sons have to make a much larger income gain than white sons to be counted as having made progress. Our upward mobility measure corrects this by requiring all sons to have advanced by (at least) the same amount τ with respect to their fathers' percentile rank for them to be counted as having progressed.

The extent of movement τ controls how much we want to include small movements in relative position. With $\tau = 0$, every positive movement is counted, however small. As τ rises, we count only larger movements in relative position but keep this *extent* of movement the same for all subgroups of the population. This implies, however, that we are applying different *absolute* thresholds to sons of different subgroups, for example, black sons are counted as having moved up even if their rank has not crossed an absolute threshold as long as they have made sufficient progress relative to their own fathers. In contrast, the transition probability keeps the absolute threshold constant and hence counts wealthier subgroups with smaller extent of movements, but ignores some poorer subgroups with larger extent of movement from their parents' rank.

As we show in Section 6, black–white differences in mobility (cf. Figure 4) are in fact, quite different, depending on which measure is used. Specifically, we find that whites appear to be much more upwardly mobile relative to blacks when measured by the transition probability of moving out of a given quantile. The white–black difference in mobility, however, is much smaller when measured in terms of our upward mobility index. The key reason for this is that many black sons make relatively small upward movements which are missed by $\theta(s_2, (s_1, s_2))$ but are captured by $v(\tau, s_1, s_2)$. Incomes of white fathers tend to be larger than those of black fathers (Figure 5) below virtually all fixed percentile thresholds, so black sons need a larger increase in absolute income to be counted by the transition probability measure.

Alternative definition of mobility One can alternatively define overall mobility based on transition matrices after incorporating effects of covariates. Consider a transition matrix based on an arbitrary M -class discretization of the marginal distributions of Y_0 and Y_1 : $\tilde{\Theta} = \{\tilde{\theta}(j, k)\}_{j,k=1,\dots,M}$. Then Shorrocks's (1978) measure of *unconditional* mobility is given by

$$M_1 = \frac{K - \text{trace}(\tilde{\Theta})}{K - 1} = 1 - \frac{\sum_{j=1}^K \tilde{\theta}(j, j) - 1}{K - 1}.$$

One can incorporate covariates into the above formula and define

$$M_1(x) = 1 - \frac{\sum_{j=1}^K \tilde{\theta}(j, j; x) - 1}{K - 1}, \quad (6)$$

where

$$\tilde{\theta}(j, j; x) = \text{Prob}(\zeta_j \leq Y_1 \leq \zeta_{j+1} | \xi_j \leq Y_0 \leq \xi_{j+1}, X = x),$$

and ζ_j and ξ_j denote the j th marginal quantiles of Y_1 and Y_0 , respectively. Given the simple linear relation (6), inference on $M_1(x)$ follows straightforwardly from inference on $\tilde{\theta}(j, j; x)$. However, this measure depends crucially on the discretization employed, which is clearly an undesirable feature. Altering the above formulas to allow for a continuous transition matrix seems complicated⁶ and we leave that to future research.

3.3 Measurement error

Researchers working on earnings mobility have paid particular attention to measurement error in sons' and fathers' earnings in the context of intergenerational regressions (cf. Haider and Solon (2006)).⁷ It is interesting to note that all our measures of mobility are based on the relative positions of individuals in the population, so if ranks of individuals are preserved despite measurement errors, then our measures will not be affected by the fact that we have erroneous earnings measures. One specific example of this is where reported earnings are a monotone function of true earnings, that is, if for two people denoted 1 and 2, true incomes satisfy $y_1^* > y_2^*$, then their reported incomes satisfy $y_1 > y_2$. This can be easily consistent with nonclassical measurement error, that is, $y - y^*$ being negatively correlated with true earnings y^* (Bound, Brown, Duncan, and Rodgers (1994)). In this case, all our measures based on reported y will be identical to those based on y^* . With more general types of measurement error, using time averaged incomes or earnings, as is common in the IGM literature, can partially mitigate the effect of purely random measurement error in addition to providing more reliable estimates for permanent income. This is the approach we follow in the application below. Finally our measures of sons' earnings are taken around the age of 40 when life-cycle bias (Haider and Solon (2006)) is minimized.

Based on the parameters (2) and (5), one can define the corresponding marginal effects by differentiating them with respect to the regressor values and/or can summarize them by density-weighted average derivatives, à la Powell, Stoker, and Stock (1989). For brevity, we do not pursue these quantities here.

⁶The problem is that $\int_0^1 \theta(s, s) ds$ is not a probability, unlike $\sum_{j=1}^K \tilde{\theta}(j, j)$.

⁷We are aware of only one study that has examined the effect of measurement error in the context of transition probabilities. O'Neill, Sweetman, and Van de Gaer (2007) showed that measurement error can induce a modest bias in transition probabilities compared to regressions and that this bias may vary at different points of the distribution.

4. ESTIMATION AND DISTRIBUTION THEORY

We now turn to estimation of the parameters and derivation of their asymptotic properties. Note that we have defined four parameters above, namely, (1), (2), (4), and (5). Formby, Smith, and Zheng (2004) analyzed only (1) and so, in what follows, we derive the distribution theory for the other three. The analysis of (2) requires slight modification of standard kernel regression theory. We only provide an outline of the proof by pointing out the modifications needed. The estimators of (4) and (5) are fundamentally harder to analyze owing to the presence of the terms $\hat{F}_1(\cdot) - \hat{F}_0(\cdot) > \tau$ in the definition of the dependent variables. These cannot be inverted in the same way that, say, $s_1 \leq F_0(Y_0) \leq s_2$ in (1) can be inverted as $F_0^{-1}(s_1) \leq Y_0 \leq F_0^{-1}(s_2)$ and analyzed using well known results for marginal quantile estimation. Our analysis of (4) and (5) therefore relies crucially on the idea of Hadamard differentiability, and we use Hoeffding’s inequality to control the errors involved in the estimation of $\hat{F}_1(\cdot)$ and $\hat{F}_0(\cdot)$.

Given that the support of income variables can be taken to be bounded below, without loss of generality, we assume that the supports of Y_0 and Y_1 are subsets of $[1, \infty)$. Note that all our mobility measures are based on quantiles, so fixed location shifts in either variable do not affect any of the measures. For fixed s, t , denote the s th quantile of Y_1 by ζ_1 and the t th quantile of Y_0 by ζ_0 with corresponding estimates by $\hat{\zeta}_1, \hat{\zeta}_0$. For transition probability, W denotes the indicator $1\{Y_1 \leq \zeta_1, Y_0 \leq \zeta_0\}$, and for upward mobility, W denotes $1\{F_1(Y_1) - F_0(Y_0) > \tau\}$; V denotes $1\{F_0(Y_0) \leq s\} := 1\{Y_0 \leq \zeta_0\}$ in both cases.

4.1 Conditional transition probability

We first state the distribution theory for estimating the conditional transition probability (cf. (2)):

$$\begin{aligned} \theta(x; s, t) &= \text{Prob}[Y_1 \geq \zeta_1 | Y_0 \leq \zeta_0, X = x] \\ &= 1 - \frac{\text{Prob}[Y_1 \leq \zeta_1, Y_0 \leq \zeta_0 | X = x]}{\text{Prob}[Y_0 \leq \zeta_0 | X = x]}. \end{aligned} \tag{7}$$

Now (7) can be estimated by

$$\hat{\theta}(x; s, t) = 1 - \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0)}{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(Y_{0i} \leq \hat{\zeta}_0)}, \tag{8}$$

where $K(\cdot)$ is a d -dimensional kernel and σ_n is a sequence of bandwidths. Let $f(y_0, y_1 | x)$ denote the density of (Y_0, Y_1) conditional on $X = x$ and define

$$\phi(x, \zeta_0, \zeta_1) = \text{Prob}\{Y_1 \leq \zeta_1, Y_0 \leq \zeta_0 | X = x\} = \int_1^{\zeta_1} \int_1^{\zeta_0} f(y_0, y_1 | x) dy_0 dy_1,$$

$$\hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1) = \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0)}{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)},$$

$$\hat{\phi}(x, \zeta_0, \zeta_1) = \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0)}{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)}.$$

The asymptotic distribution for conditional (on covariates) transition probabilities is based on the following proposition. We state this and subsequent propositions in terms of a d -dimensional X , all of whose components are continuously distributed. For discrete covariates, the analysis is identical to that for the marginal (i.e., unconditional) measures.

PROPOSITION 1. *Suppose that conditions NW1–5 in Appendix A are satisfied. Assume further that for $X = x$, (Y_0, Y_1) admits a nonnegative joint density with respect to the Lebesgue measure everywhere on the joint support. Then we have*

$$(n\sigma_n^d)^{1/2}(\hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1) - \hat{\phi}(x, \zeta_0, \zeta_1)) = o_p(1).$$

For the proof, see Appendix A.

The implication is that the distribution of $\hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1)$ and of the infeasible estimator $\hat{\phi}(x, \zeta_0, \zeta_1)$ are identical. The argument is based on the observation that $(\hat{\zeta}_0, \hat{\zeta}_1)$ converges at the parametric \sqrt{n} rate, but $\hat{\phi}(x, \zeta_0, \zeta_1)$ converges to $\phi(x, \zeta_0, \zeta_1)$ slower than the \sqrt{n} rate and a standard equicontinuity argument can then be used to handle the nonsmoothness of $1(Y_1 \leq \zeta_1, Y_0 \leq \zeta_0)$ in the ζ 's. Note also that through our assumptions, we have used an “undersmoothed” estimator to achieve bias reduction and omitted bounded moment assumptions on the errors because the dependent variable and $\phi(\cdot, \cdot, \cdot)$ lie in $[0, 1]$.

An exactly analogous proposition with a virtually identical proof applies to

$$\hat{\psi}(x, \hat{\zeta}_0) = \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(Y_{0i} \leq \hat{\zeta}_0)}{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)}.$$

Returning to (8), we note that $\hat{\theta}(x; s, t) = \hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1) / \hat{\psi}(x, \hat{\zeta}_0)$ and its asymptotic distribution follows by the standard delta method and the Cramer–Wold device.

4.2 Marginal upward mobility

We now consider the notationally simpler version of $v(\tau, s)$ defined in (4),⁸

$$v(\tau, s) = \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s), \quad (9)$$

which can be estimated by

$$\hat{v}(\tau, s) = 1 - \frac{\frac{1}{n} \sum_{i=1}^n 1(\hat{F}_1(y_{1i}) \geq \hat{F}_0(y_{0i}) + \tau, \hat{F}_0(y_{0i}) \leq s)}{s}, \quad (10)$$

where

$$\hat{F}_1(y_{1i}) = \frac{1}{n} \sum_{j \neq i} 1(y_{1j} \leq y_{1i}).$$

We now state the asymptotic distribution of $\hat{v}(\tau, s)$. Let $F(\cdot, \cdot)$ denote the joint c.d.f. of (Y_0, Y_1) with corresponding joint density $f(\cdot, \cdot)$. Then for fixed s, τ , one may view $v(\tau, s)$ as a functional $v(F)$. We can, therefore, estimate it by $v(\hat{F})$, where \hat{F} denotes the usual empirical c.d.f. We obtain a large sample distribution of $v(\hat{F})$. The key step is to show that the functional $F \mapsto v(F)$ is smooth in the Hadamard sense, with a derivative at F given by a linear functional $v'_F(\cdot)$.⁹ This is done in Lemma 1 in Appendix B, and the relevant tail conditions and the proof are stated there as well. If one assumes that the joint density of (Y_0, Y_1) is bounded away from zero on a compact support, then the proof is considerably simpler. This assumption, however, excludes families with “abnormally” high and low earnings in either generation—which is typically where the density is close to zero—and this is clearly undesirable. So we establish Hadamard differentiability under more general tail conditions on the joint density and its partial derivatives.

Given this lemma, the asymptotic distribution of $\hat{v} = v(\hat{F})$ can be derived as follows. Let \rightsquigarrow denote standard weak convergence of distribution functions and define the Gaussian process \mathcal{G} by $\sqrt{n}(\hat{F} - F) \rightsquigarrow \mathcal{G}$. Then from Lemma 1 and the functional delta method, we have that

$$\sqrt{n}(\hat{v} - v_0) \xrightarrow{d} v'_F(\mathcal{G}),$$

whence $v'_F(\mathcal{G})$ is distributed as a univariate zero-mean normal; see technical Appendix B for the exact form of this distribution.

It is well known that the bootstrap provides consistent approximations to the asymptotic distribution of the sample c.d.f. process $\sqrt{n}(\hat{F} - F)$.¹⁰ Using the Hadamard differen-

⁸One can move from (9) to (4) using simple subtractions.

⁹The concept of Hadamard differentiability has been used before in the context of analyzing features of *univariate* income distributions (cf. Bhattacharya (2007) and Barrett and Donald (2009)). The results obtained here involve more complicated functionals of bivariate distribution functions and are not related to the results in the above papers.

¹⁰For a textbook treatment, see Theorem 3.6.1, part (iii) in Van der Vaart and Wellner (1996) and its discussion on page 346 of the same text.

tiability result of our Lemma 1, it follows, via the functional delta method for the bootstrap in probability (cf. Van der Vaart and Wellner (1996, Theorem 3.9.11)), that bootstrapping will lead to consistent approximation of the distribution of the estimator of \hat{v} . We summarize the above discussion in the following proposition, where “regularity Conditions A and B” refer to those in Lemma 1 in Appendix B and \hat{v}^* denotes the bootstrap version of \hat{v} .

PROPOSITION 2. *Under regularity Conditions A and B, the bootstrap distribution of $\sqrt{n}(\hat{v}^* - \hat{v})$ will consistently estimate the distribution of $\sqrt{n}(\hat{v} - v_0)$.*

The proof follows from Hadamard differentiability of the map $F \mapsto v(F)$ (Lemma 1 in Appendix B) by applying the functional delta method.

In the application discussed below, we use the bootstrap to approximate standard errors for the marginal upward mobility by race (Table 2), and for mobility by race and parent income (Table 4). We also provide a histogram for the bootstrap distribution (Figure 3) and summarize some descriptive measures pertaining to the distribution, such as moments, skewness, and kurtosis (Table 3). Standard tests fail to reject normality of the distribution, as is to be expected, given the Gaussian form of the ingredients of $v'_F(\mathcal{G})$.

4.3 Conditional upward mobility

Recall from (5) that conditional upward mobility is given by

$$v_c(\tau, s; x) = \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s, X = x).$$

Its natural estimates then is

$$\begin{aligned} \hat{v}_c(\tau, s; x) &= \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s)}{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(\hat{F}_0(Y_{0i}) \leq s)}. \end{aligned}$$

Recall the definitions $W := 1(F_1(Y_1) - F_0(Y_0) > \tau)$ and $V := 1\{F_0(Y_0) \leq s\}$.

PROPOSITION 3. *Suppose the data (X_i, Y_{1i}, Y_{0i}) for $i = 1, \dots, n$, are independent and identically distributed (i.i.d.) and assumptions NW1–4, NW5' in Appendix C hold. Then we have that*

$$\begin{aligned} & (n\sigma_n^d)^{1/2} [\hat{v}_c(\tau, s; x) - v_c(\tau, s; x)] \\ &= \frac{1}{E(V|X = x)} \times (n\sigma_n^d)^{1/2} \times \{\hat{E}(W|X = x) - E(W|X = x)\} \\ & \quad - \frac{E(W|X = x)}{\{E(V|X = x)\}^2} \times (n\sigma_n^d)^{1/2} \times \{\hat{E}(V|X = x) - E(V|X = x)\} \\ & \quad + R_n(x), \end{aligned}$$

where $\hat{E}(W|X = x)$ and $\hat{E}(V|X = x)$ denote, respectively, the Nadaraya–Watson regression estimates of W and V on X , and the remainder $R_n(x)$ is $o_p(1)$.

Asymptotic normality now follows by standard arguments for Nadaraya–Watson regressions, for example, Bierens (1994, Theorem 10.2.1) whose conditions are implied by assumptions NW1–4 and NW5'.

See Appendix C for the proof.

In our empirical application, we want to compare the entire curve of the AFQT-conditioned mobility for blacks with that of whites. Consequently, we need to construct uniform confidence bands on (the difference in) these regression curves using the decomposition in Proposition 3. This corresponds to testing the null hypothesis that the black–white difference in mobility conditional on AFQT = x is zero for every x . Based on Proposition 3 and strengthening $R_n(x) = o_p(1)$ to $\sup_x |R_n(x)| = o_p(1)$ (see Appendix D), we can apply Theorem 4.3.1 of Hardle, (1990) to construct uniform 95% confidence bands which, asymptotically, contain the true curve $v_c(\tau, s; x)$ with probability 95%. The details of this construction and the applicability of Hardle's theorem are outlined in Appendix D.

5. DATA

We use data from the National Longitudinal Survey of Youth (NLSY). The NLSY started with a sample of individuals who were between the ages of 14 and 21 as of December 31, 1978 and has subsequently tracked these individuals into adulthood. Respondents were interviewed annually from 1979 through 1994 and thereafter interviews occurred every other year. The survey collects data on income obtained in the prior year. We use the data from the 1998–2004 interviews when sample members were around the age of 40. To avoid having to deal with issues related to labor force participation, we focus only on men (i.e., sons) in this study.

During the first few years of the survey, the NLSY also interviewed the *parents* of the sample members and collected data on total family income. We subtract from this measure any recorded earnings of the sons and average this over 1978, 1979, and 1980. This serves as our measure of income for the parents' generation. To measure the income of sons, we use the NLSY respondents' own average annual earnings when they were adults in 1997, 1999, 2001, and 2003. In contrast to using only a single year of sons' earnings or parent income, the time averaging provides a better measure of permanent income for each generation (Solon (1992)).¹¹ All income variables are deflated to 1978 dollars using the CPI-U. Our sample restrictions lead to a sample of 2766 white and black men.

A key covariate for our analysis is the Armed Forces Qualifying Test (AFQT). All individuals in the NLSY were given the AFQT test in 1980 as part of the renorming of the test. The U.S. military views the AFQT score as “a general measure of trainability and

¹¹We do not require individuals (or their parents) to have nonmissing data on income in all years. Instead, we average income over all available years within the specified range of years. We also include instances of zero earnings for sons and zero income for parents.

predictor of on-the-job performance.”¹² Following Neal and Johnson (1996), we use the 1989 version of the percentile score. We also estimate effects conditional on educational attainment. For this measure, we use years of completed schooling by age 26.

6. RESULTS

In this section, we produce empirical estimates and standard errors of IGM for black and white men using the two measures described earlier: transition probabilities and upward mobility. For each measure, we show two sets of results: unconditional estimates and conditional on AFQT test scores.

6.1 Marginal probabilities

6.1.1 Upward transition probabilities We begin by showing estimates for upward IGM using transition probabilities. We have simplified the notation from (1) to use a common cutoff, s , in both generations. To facilitate comparisons with the upward mobility measure we have introduced in this paper, we also consider transition probabilities where the son must surpass the quantile by the amount τ , namely $\text{Prob}[F_1(Y_1) > s + \tau | F_0(Y_0) \leq s]$. Confidence intervals for these are calculated using analogs of the analytical formulae in Formby, Smith, and Zheng (2004).

The results are shown in Table 1. In the first set of three columns, we produce separate estimates for whites, blacks, and the white–black difference for the baseline case where $\tau = 0$. In the subsequent sets of columns, we allow τ to vary from 0.1 to 0.3. In each row, we condition on parent income being below the s percentile where s goes from 5 to 50 in increments of 5. It is immediately evident that the white–black differences are dramatic. For example, the baseline transition probability out of the bottom quartile is 71 percent for whites, but only 45 percent for black, or a 26 percentage point difference. We plot the transition probabilities for whites and blacks along with the pointwise 95 percent confidence intervals in Figure 1. The figure plots the results in intervals of 5 percentile points as parent percentile varies from 5 to 50. As is evident in the chart, except for those at the very bottom of the distribution (below the 5th percentile), blacks are significantly less likely to surpass the quantile thresholds.

This is an important finding because most previous research on IGM has used measures such as the intergenerational elasticity, which do not allow for comparisons of group differences in mobility with respect to the *entire* population.¹³

Interestingly, the white–black difference in the transition probability out of the bottom quartile does not change very much as we allow τ to vary. For example, the racial gap in the probability of rising from the bottom quartile to at least the 45th percentile (i.e., $\tau = 0.2$) is 23 percentage points. When we condition on parents who are at or below the median and allow τ to be large (0.2–0.3), then we find that the interracial mobility gap begins to narrow to a smaller, but still significant, 10 percentage point difference.

¹²http://www.defenselink.mil/prhome/poprep2002/chapter2/c2_recruiting.htm.

¹³See our discussion of the previous literature in Section 2.

TABLE 1. Transition probability estimates by race $\theta = \text{Prob}(F_1(Y_1) > s + \tau, F_0(Y_0) < s) / \text{Prob}(F_1(Y_0) < s)$.^a

s	$\tau = 0$			$\tau = 0.1$			$\tau = 0.2$			$\tau = 0.3$		
	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B
0.05	0.978 (0.030)	0.891 (0.025)	0.087 (0.041)	0.849 (0.057)	0.579 (0.043)	0.270 (0.073)	0.704 (0.070)	0.407 (0.044)	0.297 (0.084)	0.593 (0.084)	0.280 (0.040)	0.312 (0.093)
0.10	0.917 (0.030)	0.702 (0.027)	0.216 (0.042)	0.760 (0.046)	0.458 (0.030)	0.302 (0.055)	0.632 (0.053)	0.340 (0.028)	0.292 (0.061)	0.555 (0.054)	0.249 (0.025)	0.306 (0.059)
0.15	0.812 (0.030)	0.616 (0.026)	0.196 (0.042)	0.692 (0.035)	0.423 (0.026)	0.269 (0.045)	0.542 (0.037)	0.309 (0.023)	0.232 (0.046)	0.459 (0.038)	0.212 (0.020)	0.247 (0.046)
0.20	0.752 (0.028)	0.524 (0.025)	0.228 (0.041)	0.618 (0.033)	0.389 (0.025)	0.229 (0.044)	0.496 (0.035)	0.281 (0.022)	0.215 (0.041)	0.379 (0.037)	0.192 (0.019)	0.187 (0.043)
0.25	0.708 (0.025)	0.447 (0.024)	0.261 (0.036)	0.558 (0.026)	0.326 (0.021)	0.232 (0.035)	0.459 (0.027)	0.234 (0.019)	0.225 (0.034)	0.342 (0.029)	0.156 (0.017)	0.186 (0.035)
0.30	0.646 (0.024)	0.403 (0.020)	0.244 (0.034)	0.539 (0.026)	0.290 (0.018)	0.249 (0.033)	0.418 (0.024)	0.200 (0.016)	0.217 (0.028)	0.305 (0.021)	0.131 (0.013)	0.174 (0.025)
0.35	0.583 (0.023)	0.349 (0.020)	0.234 (0.034)	0.478 (0.023)	0.254 (0.018)	0.224 (0.032)	0.366 (0.022)	0.173 (0.015)	0.193 (0.028)	0.257 (0.020)	0.120 (0.013)	0.136 (0.024)
0.40	0.544 (0.019)	0.311 (0.018)	0.233 (0.029)	0.427 (0.019)	0.223 (0.017)	0.203 (0.027)	0.315 (0.018)	0.148 (0.013)	0.167 (0.024)	0.228 (0.018)	0.105 (0.011)	0.122 (0.022)
0.45	0.494 (0.017)	0.262 (0.017)	0.232 (0.024)	0.372 (0.018)	0.180 (0.015)	0.192 (0.025)	0.264 (0.016)	0.123 (0.012)	0.141 (0.020)	0.190 (0.014)	0.080 (0.011)	0.109 (0.020)
0.50	0.428 (0.015)	0.226 (0.015)	0.202 (0.023)	0.320 (0.016)	0.152 (0.012)	0.168 (0.021)	0.227 (0.015)	0.107 (0.011)	0.119 (0.020)	0.147 (0.011)	0.065 (0.009)	0.082 (0.015)

^aSee text for a description of the estimator. Data are from the NLSY. We use multiyear averages of sons' income over 1997–2003 and parent income measured over 1978–1980. Standard errors are in parentheses.

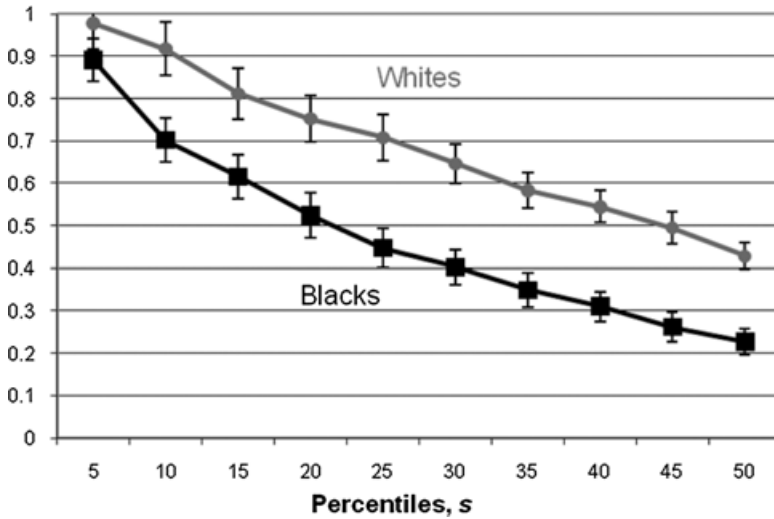


FIGURE 1. Transition probabilities conditional on parent percentile.

6.1.2 *Upward mobility* We now show an analogous set of estimates of our upward mobility measure

$$\text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s)$$

for whites, blacks, and the white–black difference in Table 2. We find much smaller racial differences in our baseline case ($\tau = 0$). For example, among white men whose family income during their youth was below the 25th percentile, 84 percent achieved a higher percentile than their parents. The comparable figure for black men is 76 percent, implying a difference of about 8 percentage points. The results are plotted in intervals of 5 percentile points along with pointwise 95 percent confidence bands in Figure 2.

To calculate pointwise confidence intervals for mobility v , we compute the sample analog \hat{v} and then draw 200 bootstrap resamples from our sample. The use of the bootstrap is justified via the functional delta method, discussed above.¹⁴ For each bootstrap resample, we calculate the corresponding estimate v^* and the statistic $t^* = \sqrt{n}|v^* - \hat{v}|$. We then calculate z^* , the 95th percentile of t^* , and use $(\hat{v} - \frac{z^*}{\sqrt{n}}, \hat{v} + \frac{z^*}{\sqrt{n}})$ as our confidence interval. We calculate the standard errors σ_v shown in the table by taking the standard deviation of v^* . The histogram for the bootstrap distribution of $(v^* - \hat{v})$ is plotted in Figure 3 for the case of whites, where $s = 0.25$ and $\tau = 0.2$. We report the summary statistics (e.g., skewness, kurtosis, and tests for normality) for various values of s and τ in Table 3. Because the histograms do not look perfectly symmetric, we also calculated equal-tailed confidence intervals. Since we found no consistent pattern in the relative size of the confidence intervals between the symmetric and the equal-tailed, we chose to report the symmetric confidence intervals.

¹⁴While studentization may be preferable before bootstrapping for higher order refinements, it is quite challenging to simulate the distribution stated in Theorem 1, so we simply use the unstudentized version.

TABLE 2. Upward mobility estimates by race $v = \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s)$.^a

s	$\tau = 0$			$\tau = 0.1$			$\tau = 0.2$			$\tau = 0.3$		
	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B
0.05	0.977 (0.024)	0.950 (0.018)	0.027 (0.033)	0.904 (0.047)	0.635 (0.044)	0.270 (0.066)	0.745 (0.065)	0.420 (0.045)	0.325 (0.083)	0.614 (0.073)	0.312 (0.040)	0.303 (0.084)
0.10	0.947 (0.022)	0.883 (0.022)	0.065 (0.032)	0.840 (0.035)	0.574 (0.032)	0.266 (0.051)	0.698 (0.047)	0.377 (0.031)	0.321 (0.059)	0.595 (0.053)	0.288 (0.025)	0.307 (0.061)
0.15	0.909 (0.021)	0.835 (0.020)	0.074 (0.029)	0.786 (0.031)	0.567 (0.027)	0.219 (0.042)	0.629 (0.040)	0.390 (0.025)	0.240 (0.047)	0.519 (0.040)	0.281 (0.025)	0.238 (0.048)
0.20	0.871 (0.021)	0.796 (0.017)	0.075 (0.027)	0.755 (0.029)	0.556 (0.024)	0.198 (0.039)	0.592 (0.030)	0.387 (0.022)	0.205 (0.037)	0.485 (0.032)	0.285 (0.020)	0.200 (0.039)
0.25	0.838 (0.021)	0.762 (0.019)	0.076 (0.030)	0.724 (0.024)	0.537 (0.024)	0.187 (0.038)	0.575 (0.028)	0.373 (0.024)	0.202 (0.036)	0.463 (0.028)	0.274 (0.019)	0.188 (0.034)
0.30	0.821 (0.018)	0.734 (0.019)	0.087 (0.027)	0.715 (0.021)	0.521 (0.021)	0.193 (0.033)	0.568 (0.026)	0.360 (0.020)	0.208 (0.036)	0.447 (0.025)	0.262 (0.019)	0.185 (0.035)
0.35	0.786 (0.019)	0.717 (0.017)	0.069 (0.026)	0.668 (0.020)	0.514 (0.023)	0.154 (0.030)	0.537 (0.021)	0.360 (0.019)	0.178 (0.031)	0.415 (0.023)	0.263 (0.016)	0.153 (0.028)
0.40	0.757 (0.018)	0.704 (0.016)	0.052 (0.025)	0.641 (0.017)	0.506 (0.020)	0.135 (0.028)	0.513 (0.020)	0.357 (0.019)	0.156 (0.027)	0.393 (0.019)	0.254 (0.018)	0.139 (0.027)
0.45	0.731 (0.015)	0.687 (0.017)	0.044 (0.024)	0.605 (0.021)	0.495 (0.021)	0.110 (0.032)	0.484 (0.019)	0.350 (0.018)	0.134 (0.028)	0.367 (0.019)	0.248 (0.017)	0.119 (0.026)
0.50	0.695 (0.014)	0.668 (0.018)	0.028 (0.025)	0.578 (0.016)	0.481 (0.020)	0.097 (0.028)	0.457 (0.016)	0.342 (0.018)	0.115 (0.025)	0.342 (0.017)	0.242 (0.015)	0.100 (0.024)

^aSee text for a description of the estimator. Data are from the NLSY. We use multiyear averages of sons' income over 1997–2003 and parent income measured over 1978–1980. Bootstrapped standard errors are in parentheses.

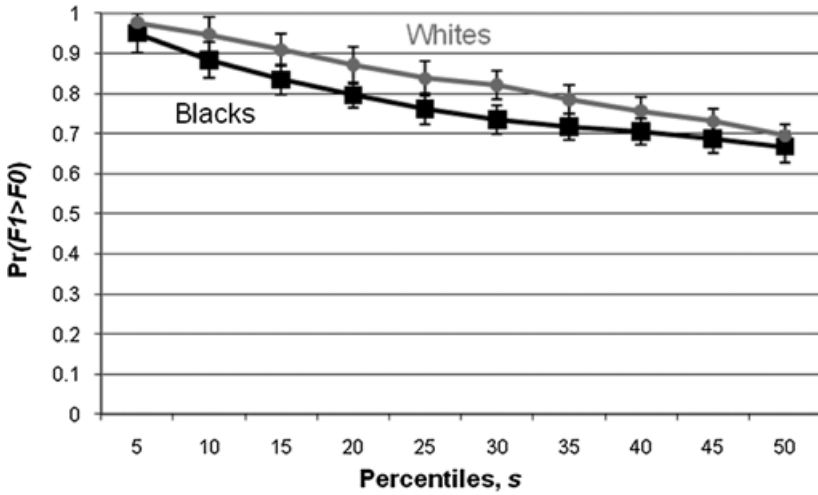


FIGURE 2. Upward mobility conditional on parent percentile $\text{Prob}[F_1(Y_1) > F_0(Y_0) | F_0(Y_0) \leq s]$.

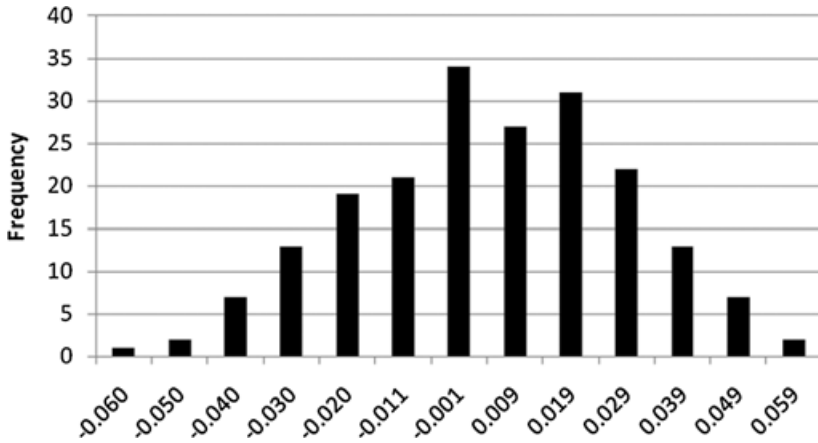


FIGURE 3. Histogram of $v^* - v$.

TABLE 3. Summary statistics of bootstrapped values of $v^* - v$.^a

	Mean	Median	Skewness	Kurtosis	p -Value Skew Test	p -Value Kurt. Test	p -Value Joint (Chi sq)
For $s = 0.25$							
$t = 0$	-0.009	-0.010	0.100	2.787	0.550	0.637	0.746
$t = 0.1$	0.000	0.000	-0.091	2.700	0.587	0.417	0.618
$t = 0.2$	0.001	0.001	0.118	2.916	0.483	0.982	0.780
For $s = 0.5$							
$t = 0$	-0.001	-0.002	0.283	3.301	0.097	0.297	0.143
$t = 0.1$	-0.001	-0.002	0.066	2.638	0.695	0.285	0.519
$t = 0.2$	0.001	0.001	-0.017	2.543	0.918	0.131	0.314

^aIn all cases $N = 200$. p -values are from using the sktest command in STATA v10.1.

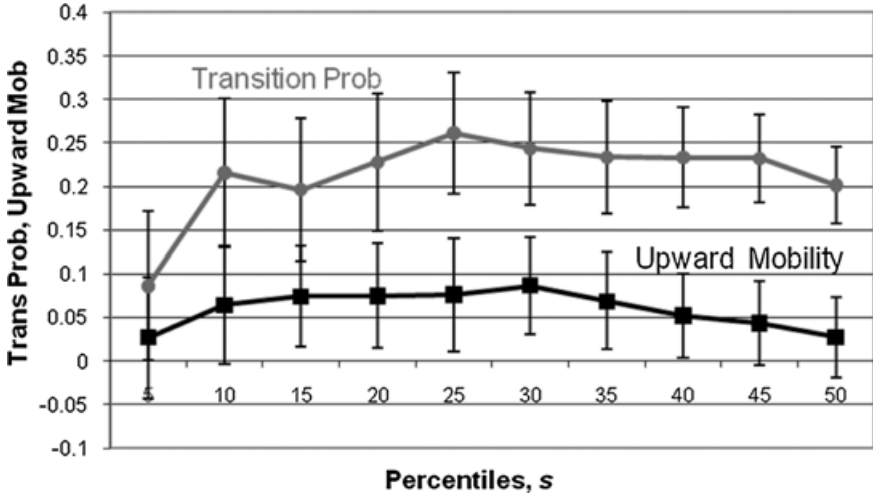


FIGURE 4. Transition probabilities versus upward mobility (whites–blacks) conditional on parent percentile.

As Figure 2 makes clear, aside from those whose family income was at or below the 5th percentile, whites appear to experience greater upward mobility than blacks, but not nearly as much as implied by the difference in the transition probabilities. The gap in most cases, however, is statistically significant as is shown in Figure 4 where we plot the white minus black difference for both the transition probability and the upward mobility along with confidence bands.

Clearly, among poorer families there are many blacks who exceed their parents rank in the distribution but do not surpass them by enough to move across specific quantiles. As discussed in Section 3, the fact that the white distribution of parent income lies to the

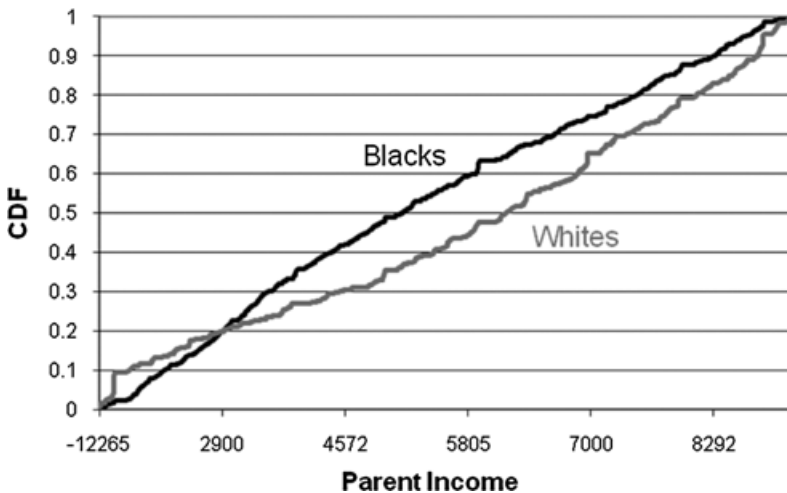


FIGURE 5. c.d.f. of parent income conditional on being in the bottom quintile: whites versus blacks.

right of blacks over most of the support makes it more likely that whites surpass the quantile thresholds more easily. This is illustrated in Figure 5, which plots the c.d.f.'s of the parent income distribution for both blacks and whites. This implies that if blacks and whites below the threshold experienced equal-sized percentile gains, then the transition probabilities would generally be higher for whites.¹⁵

The remaining columns of Table 2 show the comparable results as τ varies from 0.1 to 0.3. In each case, the magnitude of the black–white difference is generally between 15 and 25 percentage points and does not change too much as s changes. These results are comparable to the upward transition probability results in Table 1 and suggest that the two measures produce roughly similar results for larger values of τ .

Thus far the IGM measures presented have used progressively larger samples that have added more families as s is increased. This “cumulative” approach could obscure patterns that might arise if we focused more finely on upward economic mobility for individuals coming from specific parts of the income distribution. In addition, the fact that the white distribution lies to the right of the black distribution suggests that blacks may have a built in advantage with respect to upward mobility using cumulative samples since they have more “room” to rise. To address this, we recalculated measures by using nonoverlapping ranges (s_1 to s_2) for parent income that move progressively up the income distribution. Table 4, which presents these results, demonstrates that much of the rapid upward mobility experienced by blacks is concentrated at the very bottom of the distribution. For example, among those whose parents were between the 21st and 25th percentile, upward mobility is 28 percentage points more rapid for whites than blacks. Overall, these results suggest that by most measures, the *extent* of upward mobility among blacks is vastly lower than for whites.

6.2 Conditional probabilities

The underlying mechanisms by which economic status is transmitted across generations is not yet well understood and is clearly a question of great importance. Estimates of IGM conditional on key covariates can potentially shed light on this question. Understanding the source of the black–white mobility gap in particular is of great policy interest.

As noted in Section 1, previous studies using the NLSY have taken advantage of the availability of scores on the AFQT as a measure of cognitive skills to identify a source of interracial inequality. For example, Neal and Johnson (1996) showed that the black–white wage gap among adults can largely be explained by pre-market skills as proxied by AFQT scores during adolescence. O’Neill, Sweetman, and Van de Gaer (2006) showed that equalization of cognitive skill gaps does not fully account for the black–white gap at the low end of the distribution. Cameron and Heckman (2001) showed that the sizable gap in college enrollment between whites and blacks can largely be accounted for by AFQT scores. These previous findings suggest the possibility that the average black–white IGM gap might be accounted for by inclusion of AFQT scores but that there might

¹⁵However, in other results (not shown) we also find that the magnitude of the percentile gains for blacks are actually much lower than for whites.

TABLE 4. Upward mobility estimates by race using intervals of parent income $v = \text{Prob}[F_1(Y_1) - F_0(Y_0) > \tau \mid s_1 \leq F_0(Y_0) \leq s_2]$.^a

s_1 to s_2	$\tau = 0$			$\tau = 0.1$			$\tau = 0.2$			$\tau = 0.3$		
	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B	Whites	Blacks	W – B
0.01 to 0.05	0.977 (0.024)	0.950 (0.018)	0.027 (0.033)	0.904 (0.047)	0.635 (0.044)	0.270 (0.066)	0.745 (0.065)	0.420 (0.045)	0.325 (0.083)	0.614 (0.073)	0.312 (0.040)	0.303 (0.084)
0.06 to 0.10	0.915 (0.048)	0.813 (0.035)	0.102 (0.059)	0.770 (0.067)	0.511 (0.048)	0.259 (0.083)	0.647 (0.079)	0.332 (0.043)	0.315 (0.093)	0.573 (0.079)	0.263 (0.035)	0.311 (0.090)
0.11 to 0.15	0.847 (0.047)	0.708 (0.051)	0.138 (0.070)	0.698 (0.062)	0.547 (0.053)	0.151 (0.083)	0.518 (0.075)	0.423 (0.051)	0.095 (0.093)	0.395 (0.068)	0.263 (0.050)	0.132 (0.089)
0.16 to 0.20	0.780 (0.058)	0.645 (0.048)	0.134 (0.079)	0.679 (0.067)	0.516 (0.053)	0.162 (0.089)	0.501 (0.070)	0.376 (0.050)	0.124 (0.082)	0.404 (0.066)	0.300 (0.049)	0.104 (0.087)
0.21 to 0.25	0.751 (0.052)	0.473 (0.070)	0.278 (0.092)	0.645 (0.058)	0.376 (0.062)	0.269 (0.089)	0.532 (0.057)	0.256 (0.060)	0.275 (0.082)	0.404 (0.058)	0.186 (0.056)	0.218 (0.083)
0.26 to 0.30	0.755 (0.049)	0.534 (0.059)	0.221 (0.077)	0.677 (0.061)	0.406 (0.061)	0.271 (0.088)	0.542 (0.065)	0.265 (0.051)	0.277 (0.083)	0.388 (0.062)	0.173 (0.050)	0.215 (0.077)
0.31 to 0.35	0.639 (0.072)	0.495 (0.073)	0.144 (0.104)	0.471 (0.063)	0.420 (0.075)	0.051 (0.098)	0.408 (0.060)	0.358 (0.076)	0.050 (0.102)	0.282 (0.061)	0.272 (0.071)	0.010 (0.104)
0.36 to 0.40	0.613 (0.055)	0.489 (0.092)	0.124 (0.113)	0.510 (0.061)	0.371 (0.090)	0.139 (0.117)	0.392 (0.056)	0.313 (0.087)	0.079 (0.113)	0.282 (0.053)	0.110 (0.068)	0.172 (0.090)
0.41 to 0.45	0.578 (0.060)	0.258 (0.096)	0.320 (0.116)	0.385 (0.071)	0.220 (0.090)	0.165 (0.111)	0.307 (0.063)	0.162 (0.088)	0.145 (0.094)	0.213 (0.047)	0.085 (0.057)	0.128 (0.072)
0.46 to 0.50	0.450 (0.055)	0.311 (0.080)	0.138 (0.094)	0.393 (0.064)	0.225 (0.079)	0.168 (0.114)	0.275 (0.053)	0.195 (0.068)	0.080 (0.089)	0.166 (0.041)	0.135 (0.064)	0.031 (0.071)

^aSee text for a description of the estimator. Data are from the NLSY. We use multiyear averages of sons' income over 1997–2003 and parent income measured over 1978–1980. Bootstrapped standard errors are in parentheses.

be differences that remain at the low end. However, it is important to note that unlike these previous studies, our measures of mobility capture movements in the distribution *relative to the parent generation*, so it is not obvious whether mobility gaps will be eliminated the same way that level gaps are. Below, we produce estimates of upward transition probabilities and our measure of upward mobility for black and white men where we now include AFQT scores nonparametrically as a covariate.

We employ Nadaraya–Watson (NW) regressions. To do so, we first normalized the regressor to lie between 0 and 1, using maximum and minimum possible values of the AFQT, namely 99 and 1, and then estimated the regressions at 100 points with spacing of 0.01. We used an Epanechnikov kernel and chose the bandwidth σ_n in accordance with assumption NW4 (Appendix A), where $d = 1$. We experimented with bandwidths around the range $n^{-1/4}$ (moving from $n^{-1/5}$ to $n^{-1/3}$), where n denotes the size of the effective sample (this size varies depending on which parent percentile and race are conditioned on). Our results for conditional mobility were quite stable over this range, so we report the results at the $\sigma_n = n^{-1/4}$ value. For inference, we calculated uniform bands using the analytical formulas from Hardle (1990, Algorithm 4.3.2), which are based on Bickel and Rosenblatt (1973).¹⁶ The latter steps are reproduced in Appendix D. Uniform, rather than pointwise, confidence bands are necessary because here we are making inference on the entire conditional mobility curve as a function of the conditioning variable AFQT to see how mobility differences vary with AFQT. Therefore, we need data-based bands which contain the *entire* true curves with at least a pre-assigned probability. Joining pointwise confidence limits reduces the coverage probability arbitrarily below the nominal level, leading to wrong confidence statements.

6.2.1 Conditional transition probabilities We estimate the effect of AFQT scores on upward transition probabilities separately by race. Our dependent variable is the probability of leaving the bottom quintile. Figure 6 shows the result of this exercise. We find that conditional on AFQT scores, whites have only slightly higher likelihood of exiting the bottom quintile and that this gap does not vary a great deal across the AFQT distribution. For example, at the 25th interval of our normalized AFQT scores, the transition probability for whites is 0.63 and for blacks is 0.61, or a difference of just 2 percentage points. At the 10th interval, the gap is about 7 percentage points and at the 75th interval, the gap is about 15 percentage points. At no point in the AFQT distribution can we reject the hypothesis that the transition probabilities are the same.

The shapes of the regression lines are also similar between blacks and whites for the bottom half of the distribution. In the upper half of the AFQT distribution, however, the slopes differ and the lines fan apart. It is important to note, however, that there are relatively few data for blacks in the upper end of the AFQT distribution as is evidenced by the widening confidence intervals.

This finding of similar point estimates of conditional transition probabilities using AFQT scores can be contrasted with results using years of education. In Figure 7, we do a

¹⁶Hardle (1990, Theorem 4.3.1) justified the validity of this algorithm. In the [Appendixes](#), we show that our conditions NW1–NW4, NW5', and NW6 imply the sufficient Conditions A1–A5 of Hardle (1990), p. 116 for this theorem to hold.

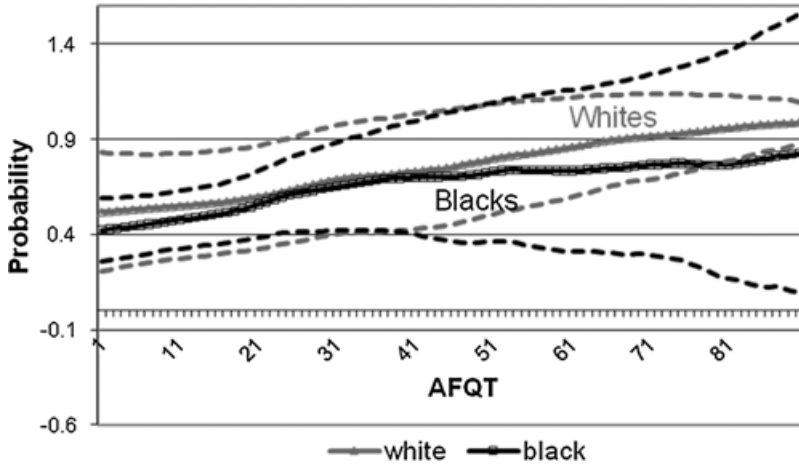


FIGURE 6. Probability of leaving the bottom quintile conditional on AFQT: whites versus blacks.

similar exercise where we instead use 20 intervals of the sons' years of completed schooling (normalized) as a covariate in estimating the transition probabilities by race. Here we find sharp differences in the transition probability, even conditional on years of schooling, throughout much of the distribution. For example among those in the 10th interval, with roughly 10 years of schooling, the transition probability out of the bottom quintile for whites is 67 percent, while for blacks it is just 45 percent. At the higher end of the education distribution, however, the racial gap converges, and at the very top of the distribution, black mobility is actually higher. Our confidence intervals are quite large, however, so although the differences are quite large over much of the distribution, they are not statistically significant. In similar exercises using measures of parent education (not shown), we find broadly similar results. Therefore, like Hertz (2005), we find that par-

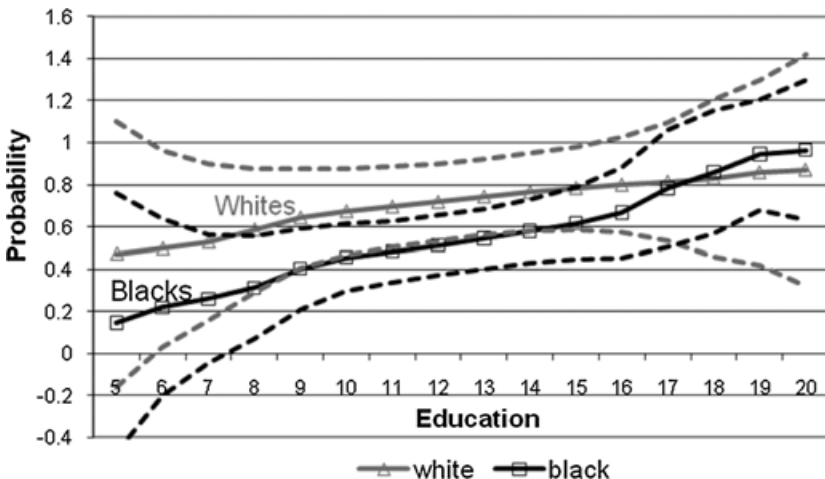


FIGURE 7. Transition probability of leaving the bottom quintile conditional on ed: whites versus blacks.

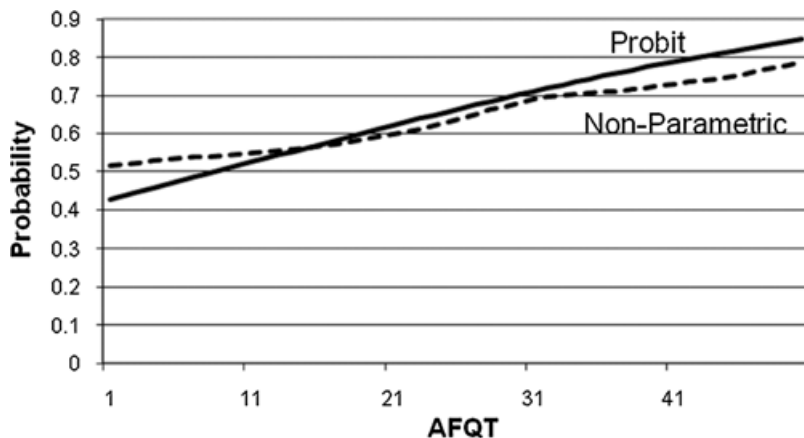


FIGURE 8. Comparison of probit and nonparametric estimates of the transition probability of whites leaving the bottom quintile conditional on AFQT.

ent education cannot explain the black–white mobility gap for most of the distribution. However, we do find that accounting for AFQT scores does appear to account for the gap.

Finally, we also find that using our nonparametric approach produces some important substantive differences compared to simply estimating a probit with AFQT as a covariate, that is,

$$\text{Prob}[Y_1 \leq \zeta_1, Y_0 \leq \zeta_0 | \text{AFQT} = x] = \Phi(\beta_0 + \beta_1 x).$$

This is particularly true for whites at the bottom of the distribution and for blacks at the top of the distribution. In Figure 8, we compare the transition probability results for whites in the bottom of the distribution with the results from simply using a probit. As the chart shows, moving from the first percentile of the AFQT distribution to the median nearly doubles the transition probability of leaving the bottom quintile for whites, from 0.43 to 0.85, when using the probit. In contrast, the nonparametric estimator implies an increase of only about 27 percentage points, from 0.52 to 0.79. This is not surprising because the probit estimate at a point is affected by the outcome at far-off regressor values unlike the nonparametric estimates.

6.2.2 Conditional upward mobility We also estimate the effect of AFQT scores on our measure of upward mobility separately by race. For this exercise, we condition on parent income being at or below the 20th percentile and set $\tau = 0$. The results are shown in Figure 9. In this case, the effects on the black–white gap are even more striking, as the point estimates imply that upward mobility is virtually identical for blacks and whites in the bottom half of the distribution.

6.2.3 Discussion of AFQT results We wish to be careful to point out that we do not think that the finding that AFQT scores can account for the black–white IGM gap lends itself to any simple interpretation or any obvious policy remedy. The development of cognitive skills that we measure in adolescence can be due to a range of factors, including

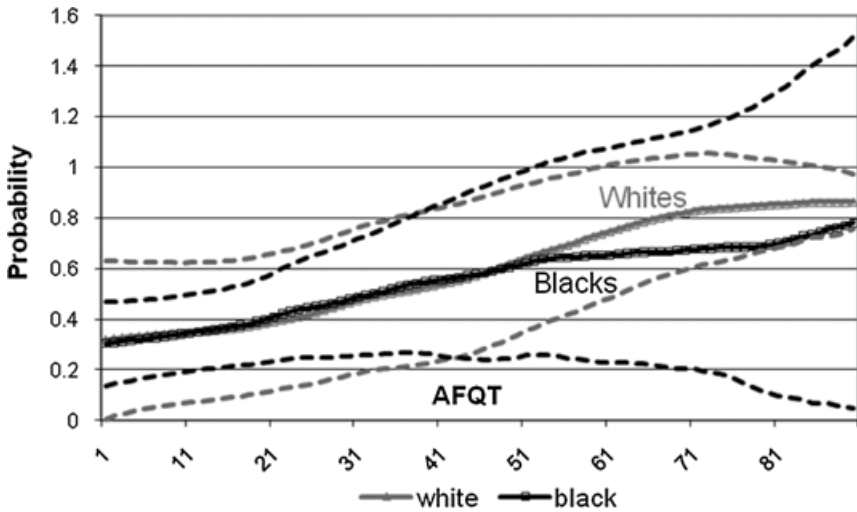


FIGURE 9. Upward mobility conditional on AFQT scores: whites versus blacks.

health endowments, parental investment, peer influences, and school quality. Our results suggest that whatever the underlying causes of the gap in cognitive skills, it appears to translate into significant differences in IGM.

Understanding the formation of the black–white skills gap has been, and will likely continue to be, an area of intense research activity. For example, recent work by Chay, Guryan, and Mazumder (2009) using military applicant data showed that much of the apparent narrowing of the black–white test score gap during the 1980s can be attributed to improvements in infant health arising from greater access by southern blacks to hospitals during the 1960s. Dobbie and Fryer (forthcoming) found that the Harlem Children’s Zone, which combines community programs with charter schools, can significantly close black–white achievement gaps. These results among others suggest that there is potential for policy to address the sharp black–white differences in upward mobility highlighted here.

7. CONCLUDING THOUGHTS

In this paper, we develop new analytic tools that allow for an investigation of interracial differences in IGM and its underlying sources. Using large intergenerational samples from the NLSY, we document that upward transition probabilities for blacks in the bottom of the income distribution are sharply lower than for whites. We introduce a new measure of upward mobility that overcomes some of the limitations of the transition probability. The new measure is simply the probability that the sons’ rank in the distribution exceeds the parents’ rank in the prior generation. The baseline upward mobility measure shows a much smaller interracial gap in IGM partially because it reflects the fact that many blacks make small gains in rank over generations that are missed by the transition probability. On the other hand if we adjust our upward mobility measure to require rank gains of a certain amount, then the two measures paint a more similar picture of low upward mobility for blacks.

We also investigate how the interracial differences in upward mobility are impacted by incorporating the effects of cognitive skills during adolescence as measured by AFQT scores, using nonparametric methods. Remarkably, we find that AFQT scores can account for virtually the entire black–white difference in the ability to rise out of the bottom quintile. Many factors can potentially impact the development of cognitive skills, and understanding the source of these differences remains an important topic for future research.

There are many other aspects of interracial differences in IGM which we have not considered. For example, there may be important differences by gender. An analysis of other covariates, such as measures of health, family structure, wealth, and noncognitive skills, are also important areas for examination. For example, Heckman, Stixrud, and Urzua (2006) demonstrated the importance of noncognitive skills (e.g., dependability, persistence) on socioeconomic outcomes. We also limited our outcome of interest to labor market earnings and it may be fruitful to analyze patterns in mobility with respect to other measures such as hours worked, wages, and total family income. Finally, we have limited our analysis to upward mobility, but there may be important interracial differences with respect to downward mobility as well.

The interpretation of mobility measures as indices of “equality of opportunity” has been critiqued by several authors (e.g., Van de Gaer, Schokkaert, and Martinez (2001), Roemer (2004), Swift (2005), Jencks and Tach (2006)). Roemer (2004), in particular, emphasized that a society with high equality of opportunity is one that lets children from varying backgrounds *exerting the same effort* to reach similar economic status. To the extent educational attainment is an index of effort, our mobility analysis conditional on educational status of the sons would address this concern. This is an issue that we intend to pursue in detail in future research.

The methodological innovations of the present paper were primarily motivated by nonparametric empirical analysis of IGM, but they have potentially more general applicability. For example, one can use these methods to analyze the persistence in relative rankings of mutual funds over time and to analyze what factors (e.g., fund size, manager ability) lead to these changes and how they differ across sectors or fund categories. For instance, Chevalier and Ellison (1999) examined how characteristics of mutual fund managers affects their performance in the cross section, but one could extend this analysis to look at how these characteristics impact their ranks over longer periods. Similarly, one could extend the analysis of Kopczuk, Saez, and Song (2010), who estimated rates of *intragenerational* upward mobility by using the probability of an individual moving into the top quintile from the bottom two quintiles by adding covariates or comparing differences by industry or occupation. More generally, whenever the parameter of interest is a nonparametric regression or a functional thereof but the dependent variable involves preliminary components estimated from the same data set, possibly in a nonsmooth way, the methods developed here can be utilized to get the respective distribution theories.

APPENDIXES WITH PROOFS

In the statements of the results and in their proofs, c denotes a generic positive constant not always having the same value and whenever moments, derivatives or Lebesgue densities are defined, they are implicitly assumed to exist.

APPENDIX A

Discussion of Proposition 1

We now state a set of general regularity conditions which imply zero-mean asymptotic normality for the Nadaraya–Watson estimated regression of the unobserved random variable

$$W := 1\{Y_1 \leq \zeta_1, Y_0 \leq \zeta_0\}$$

on X , evaluated at $X = x$. These conditions are standard (for textbook treatments, see Bierens (1994, Theorem 10.2.1) or Pagan and Ullah (1999, Theorems 3.5 and 3.6), but we state them here to make the subsequent proposition and lemma statements self-contained.

CONDITION NW.

NW1. X is a d -dimensional continuously distributed random variable with Lebesgue density $f(\cdot)$ which is positive at x . For all our applications, $d = 1$.

NW2. The data (X_i, Y_{1i}, Y_{0i}) are i.i.d.

NW3. $K(\cdot)$ is a Borel-measurable, bounded, and real-valued kernel function with d -dimensional argument satisfying (i) $\int K(a) da = 1$, $\int aK(a) da = 0$, $\int a_i^2 \times K(a) da < \infty$ for $i = 1, \dots, d$, (ii) $\int |K(a)| da < \infty$, (iii) for some $\delta > 0$, $\int |K(a)|^{2+\delta} da < \infty$.

NW4. The bandwidth sequence σ_n satisfies $\lim_{n \rightarrow \infty} \sigma_n = 0$, $\lim_{n \rightarrow \infty} n\sigma_n^d = \infty$, and $\lim_{n \rightarrow \infty} \sigma_n^2(n\sigma_n^d)^{1/2} = 0$; $\frac{1}{\sigma_n^d} K\left(\frac{a-b}{\sigma_n}\right)$ is uniformly bounded for $a, b \in \text{support}(X)$.

NW5. The functions $f(\cdot)$ and $f(\cdot) \times \phi(\cdot, \zeta_0, \zeta_1)$ and their derivatives up to order 2 are continuous and uniformly bounded.

Then we have

$$(n\sigma_n^d)^{1/2}(\hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1) - \phi(x, \zeta_0^0, \zeta_1^0)) \xrightarrow{d} N\left(0, \frac{v^2(x)}{f(x)} \int K^2(u) du\right),$$

where $v^2(x) = \phi(x, \zeta_0^0, \zeta_1^0) \times (1 - \phi(x, \zeta_0^0, \zeta_1^0))$.

Proof outline for Proposition 1

First note that the function $\phi(x, \cdot, \cdot)$ is Lipschitz with respect to the Euclidean norm $\|\cdot\|$,

$$|\phi(x, \zeta_0, \zeta_1) - \phi(x, \tau_0, \tau_1)| \leq \|(\zeta_0, \zeta_1) - (\tau_0, \tau_1)\| \delta(x), \quad (11)$$

with $\delta(\cdot)$ uniformly bounded on the support of X . To see this, note that $\nabla\phi(x, a, b) = f_{Y_0, Y_1|X}(a, b|x)$, so that applying the mean-value theorem to the left hand side of (11) in conjunction with NW5 yields the result.

Now consider the expression

$$\bar{m}(\zeta, x) = \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0)$$

whose expectation is given by

$$\begin{aligned} \bar{m}^*(\zeta, x) &= E_{X_i} \left(\frac{1}{\sigma_n^d} K\left(\frac{X_i - x}{\sigma_n}\right) \phi(X_i, \zeta_0, \zeta_1) \right) \\ &= \int K(u) f(x + u\sigma_n) \phi(x + u\sigma_n, \zeta_0, \zeta_1) du \\ &= f(x) \phi(x, \zeta_0, \zeta_1) + O(\sigma_n^2). \end{aligned} \tag{12}$$

So

$$\begin{aligned} \bar{m}^*(\hat{\zeta}, x) &= f(x) \phi(x, \hat{\zeta}_0, \hat{\zeta}_1) + O(\sigma_n^2) \\ &= f(x) [\phi(x, \zeta_0^0, \zeta_1^0) + \phi_0(x, \tilde{\zeta}_0, \tilde{\zeta}_1)(\hat{\zeta}_0 - \zeta_0) + \phi_1(x, \tilde{\zeta}_0, \tilde{\zeta}_1)(\hat{\zeta}_1 - \zeta_1)] \\ &\quad + O(\sigma_n^2), \end{aligned}$$

where $\tilde{\zeta}_1$ denotes value intermediate between $\hat{\zeta}_1$ and ζ_1^0 and similarly, $\tilde{\zeta}_0$. Now,

$$\begin{aligned} &\hat{\phi}(x, \hat{\zeta}_0, \hat{\zeta}_1) - \phi(x, \zeta_0^0, \zeta_1^0) \\ &= \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0) \\ &\quad \underbrace{\hspace{10em}}_{\hat{f}(x)} - \phi(x, \zeta_0^0, \zeta_1^0) \\ &= \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{\sigma_n}\right) \{1(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0) - 1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0)\} \\ &\quad \underbrace{\hspace{10em}}_{\hat{f}(x)} \\ &\quad + \underbrace{\left\{ \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0) \right\}}_{\hat{f}(x)} - \phi(x, \zeta_0^0, \zeta_1^0) \\ &\quad \underbrace{\hspace{10em}}_{=S_n, \text{ say}} \\ &= \underbrace{\{[\bar{m}(\hat{\zeta}, x) - \bar{m}^*(\hat{\zeta}, x)] - [\bar{m}(\zeta^0, x) - \bar{m}^*(\zeta^0, x)]\}}_{\hat{f}(x)} \\ &\quad \underbrace{\hspace{10em}}_{T_{1n}} \end{aligned} \tag{13}$$

$$+ \underbrace{\frac{\bar{m}^*(\hat{\zeta}, x) - \bar{m}^*(\zeta^0, x)}{\hat{f}(x)}}_{T_{2n}} + S_n.$$

Now $(n\sigma_n^d)^{1/2}S_n$, under the assumptions NW, will be $O_p(1)$ and zero-mean normal (cf. Bierens (1994, Theorem 10.2.1)), namely

$$(n\sigma_n^d)^{1/2}S_n \xrightarrow{d} N\left(0, \frac{v^2(x)}{f(x)} \int K^2(u) du\right). \tag{14}$$

Next, using (12) and the fact that $(\hat{\zeta}_0 - \zeta_0)$ and $(\hat{\zeta}_1 - \zeta_1)$ have parametric convergence rates, we get that

$$\begin{aligned} & (n\sigma_n^d)^{1/2}T_{2n} \\ &= \frac{f(x) \left[\begin{array}{l} \phi_0(x, \tilde{\zeta}_0, \tilde{\zeta}_1) \times \underbrace{(n\sigma_n^d)^{1/2} \times (\hat{\zeta}_0 - \zeta_0)}_{o_p(1)} \\ + \phi_1(x, \tilde{\zeta}_0, \tilde{\zeta}_1) \times \underbrace{(n\sigma_n^d)^{1/2} \times (\hat{\zeta}_1 - \zeta_1)}_{o_p(1)} \end{array} \right] + \underbrace{O_p((n\sigma_n^d)^{1/2} \times \sigma_n^2)}_{o_p(1), \text{ by NW4}}}{\underbrace{\hat{f}(x)}_{=O_p(1)}} \tag{15} \\ &= o_p(1). \end{aligned}$$

The nonstandard term in (13) is T_{1n} and we now demonstrate a stochastic equicontinuity property for it. Letting

$$v_n(\zeta, x) = (n\sigma_n^d)^{1/2} \{ \bar{m}(\zeta, x) - \bar{m}^*(\zeta, x) \},$$

the first term satisfies

$$(n\sigma_n^d)^{1/2}T_{1n} = \frac{v_n(\hat{\zeta}, x) - v_n(\zeta^0, x)}{\hat{f}(x)}, \tag{16}$$

and we now show that the numerator of (16) is $o_p(1)$, for each x . Now, for fixed x , the class of functions

$$g_n(W, \zeta) := \frac{1}{\sigma_n^d} K\left(\frac{X-x}{\sigma_n}\right) 1(Y_1 \leq \zeta_1, Y_0 \leq \zeta_0)$$

form a type IV class (cf. Andrews (1994, equation 5.3)) with $p = 2$. This follows from the Lipschitz property (11) and the uniform boundedness of $\frac{1}{\sigma_n^d} K\left(\frac{X-x}{\sigma_n}\right)$. This, in turn, implies that the sequence $v_n(\zeta, x)$ is stochastically equicontinuous. Now, using the same steps as Andrews (1994), leading to his equation (3.8), we conclude that

$$v_n(\hat{\zeta}, x) - v_n(\zeta^0, x) = o_p(1).$$

Put (13), (14), (15), and (16) together with $\hat{f}(x) = O_p(1)$ to conclude.

APPENDIX B

Hadamard differentiability of upward mobility

Let $f(y_0, y_1)$ and $f^{(0)}(y_0, y_1)$ denote, respectively, the joint density of (Y_0, Y_1) and its derivative with respect to (w.r.t.) the first argument, evaluated at the point (y_0, y_1) . Let $f_1(y_1)$ denote the marginal density of Y_1 and let c denote a generic positive constant. Since all our measures are robust to monotone transformation of the income variables, we continue to assume that the support of the income variables is contained in $[1, \infty)$.

CONDITION A. (i) For some $\alpha > 1$, we have $f_1(x) \geq \frac{c}{x^\alpha}$ for x large enough, which also implies that $F_1^{-1}(u) < c(1-u)^{1/(1-\alpha)}$. (ii) $f^{(0)}(y_0, F_1^{-1}(F_0(y_0) + \tau)) \leq \frac{c}{y_0^{\alpha_0}}$ for some $\alpha_0 > 0$. (iii) For some $\varepsilon > 0$, $1 - F_0(y_0) > cy_0^{(1+\varepsilon-\alpha_0)(\alpha-1)/\alpha}$ and (iv)¹⁷

$$\int_1^\infty (1 - F_0(y_0))^{\alpha/(\alpha-1)} f(y_0, F_1^{-1}(F_0(y_0) + \tau)) dy_0 < \infty.$$

It is interesting to note that if the tails of (Y_0, Y_1) have a joint Pareto distribution, then all of these conditions are automatically satisfied. To see this, assume that (Y_0, Y_1) satisfy

$$\text{Prob}(Y_0 \geq y_0, Y_1 \geq y_1) = \frac{1}{(1 + (y_0 - 1) + (y_1 - 1))^\gamma}$$

for all $y_0, y_1 \geq 1$ for some $\gamma > 0$. Then their joint density is given by

$$f(y_0, y_1) = \frac{\gamma(\gamma + 1)}{(1 + (y_0 - 1) + (y_1 - 1))^{\gamma+2}}.$$

Then one may verify that Conditions A(i)–(iv) are satisfied with $\alpha = \gamma + 1$, $\alpha_0 = \gamma + 2$, and $\varepsilon = 1 + \gamma + \gamma(\gamma + 1)$.

An exactly symmetric set of conditions is assumed to hold for the marginal density $f_0(\cdot)$ of Y_0 as well.

CONDITION B. (i) For some $\beta > 1$, we have $f_0(x) \geq \frac{c}{x^\beta}$ for x large enough, which also implies that $F_0^{-1}(u) < c(1-u)^{1/(1-\beta)}$. (ii) $f^{(0)}(F_0^{-1}(s), y_1) \leq \frac{c}{y_1^{\beta_0}}$ for some $\beta_0 > 0$. (iii) For some $\delta > 0$, we have $1 - F_1(y_1) > cy_1^{(1+\delta-\beta_0)(\beta-1)/\beta}$ and (iv) $\int_1^\infty (1 - F_1(y_1))^{\beta/(\beta-1)} f(F_0^{-1}(s), y_1) dy_1 < \infty$.

To show that the map $F \mapsto \nu(F)$ is Hadamard differentiable, let $\bar{D}[1, \infty)$ denote the space of bivariate c.d.f.'s on $[1, \infty)$, satisfying Conditions A(i)–(iv) and B(i)–(iv). Denote by D_0 the space of sample paths corresponding to the composite Brownian bridge $[\mathbb{G}_\lambda \circ F]$, where G_λ is a standard Brownian bridge and F is any c.d.f. in $\bar{D}[1, \infty)$. Let $D = \bar{D}[1, \infty) \cup D_0$, equipped with the supremum norm. We want to show Hadamard

¹⁷Condition A(iv) is like a moment condition. Recall that for a positive random variable X with marginal c.d.f. $G(\cdot)$ and support A , the quantity $\int_A (1 - G(x)) dx$ equals $E(X)$.

differentiability of the map $F \mapsto v(F)$ as a map from the normed vector space D to \mathbb{R} . Consider perturbations $F_t(y_0, y_1) = F(y_0, y_1) + tH_t(y_0, y_1) \in \tilde{D}[1, \infty)$ with $H_t \rightarrow H \in D_0$ uniformly as $t \rightarrow 0$. We want to show that

$$\left| \frac{v(F_t) - v(F)}{t} - v'_F(H) \right| \rightarrow 0 \quad \text{as } t \rightarrow 0$$

for a linear functional $v'_F(\cdot)$, which is a map from $\tilde{D}[1, \infty)$ to \mathbb{R} .

LEMMA 1. *Under Conditions A(i)–(iv) and B(i)–(iv), the map $F \mapsto v(F)$ from $D \rightarrow \mathbb{R}$, defined as*

$$v(F) = \int_1^{F_0^{-1}(s)} \int_1^{F_1^{-1}(F_0(y_0) + \tau)} f(y_0, y_1) dy_1 dy_0$$

for any fixed $s, \tau \in (0, 1)$, is Hadamard differentiable at F tangentially to D_0 . The derivative at F in the direction H is given by the linear functional $v'_F(\cdot)$ defined as

$$\begin{aligned} v'_F(H) &= \frac{H_0(F_0^{-1}(s))}{f_0(F_0^{-1}(s))} \int_1^{F_1^{-1}(F_0(y_0) + \tau)} f(F_0^{-1}(s), y_1) dy_1 \\ &\quad + \int_1^{F_0^{-1}(s)} \frac{H_0(y_0) - H_1(F_1^{-1}(F_0(y_0) + \tau))}{f_1(F_1^{-1}(F_0(y_0) + \tau))} f(y_0, F_1^{-1}(F_0(y_0) + \tau)) dy_0 \\ &\quad + \int_1^{F_0^{-1}(s)} \int_1^{F_1^{-1}(F_0(y_0) + \tau)} dH(y_0, y_1), \end{aligned}$$

where

$$H_0(a) = \lim_{x \rightarrow \infty} H(a, x) \quad \text{and} \quad H_1(a) = \lim_{x \rightarrow \infty} H(x, a). \tag{17}$$

PROOF. Consider perturbations $F_t(y_0, y_1) = F(y_0, y_1) + tH_t(y_0, y_1)$ with $F_{0t}(y_0) = F_0(y_0) + tH_{0t}(y_0)$ and $F_{1t}(y_1) = F_1(y_1) + tH_{1t}(y_1)$ denoting the corresponding marginals. Let $H_t \rightarrow H$ uniformly as $t \rightarrow 0$, and let H_0 and H_1 denote its marginals. We want to show that for a linear functional $v'_F(\cdot)$,

$$\left| \frac{v(F_t) - v(F)}{t} - v'_F(H) \right| \rightarrow 0 \quad \text{as } t \rightarrow 0. \tag{18}$$

Define

$$\begin{aligned} z_{1t}(y_0) &= F_1^{-1}(F_0(y_0) + \tau), & z_{1t}(y_0) &= F_{1t}^{-1}(F_{0t}(y_0) + \tau), \\ z_0 &= F_0^{-1}(s), & z_{0t} &= F_{0t}^{-1}(s). \end{aligned}$$

So we need to show

$$\left| \frac{\int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} f_t(y_0, y_1) - \int_1^{z_0} \int_1^{z_1(y_0)} f(y_0, y_1) dy_1 dy_0}{t} - v'_F(H) \right| \rightarrow 0 \quad \text{as } t \rightarrow 0.$$

Note that the first term inside $|\cdot|$ can be expanded as

$$\begin{aligned}
 & \int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} f(\bar{z}_{1t}(y_0), y_0) dy_0 + \frac{z_{0t} - z_0}{t} \times \int_1^{z_1(y_0)} f(\bar{z}_{0t}, y_1) dy_1 \\
 & + \int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH_t(y_0, y_1) \\
 = & \underbrace{\left[\int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} [f(y_0, \bar{z}_{1t}(y_0)) - f(y_0, z_1(y_0))] dy_0 \right]}_{T_{10t}} \\
 & + \underbrace{\frac{z_{0t} - z_0}{t} \times \int_1^{z_1(y_0)} [f(\bar{z}_{0t}, y_1) - f(z_0, y_1)] dy_1}_{T_{11t}} \\
 & + \underbrace{\int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} f(z_0, y_1) dy_1}_{T_{2t}} + \underbrace{\frac{z_{0t} - z_0}{t} \times \int_1^{z_1(y_0)} f(z_0, y_1) dy_1}_{T_{3t}} \\
 & + \underbrace{\int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH_t(y_0, y_1) - \int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH(y_0, y_1)}_{T_{4t}} \\
 & + \underbrace{\int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH(y_0, y_1) - \int_1^{z_0} \int_1^{z_1(y_0)} dH(y_0, y_1)}_{T_{5t}} \\
 & + \underbrace{\int_1^{z_0} \int_1^{z_1(y_0)} dH(y_0, y_1)}_{T_6}.
 \end{aligned}$$

We show that as $t \rightarrow 0$, there are four steps:

Step 1. $|T_{1t}| \rightarrow 0$.

Step 2.

$$T_{2t} \rightarrow \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1) dy_1.$$

Step 3.

$$T_{3t} \rightarrow \int_1^{z_0} \frac{H_0(y_0) - H_1(z_1(y_0))}{f_1(z_1(y_0))} f(y_0, z_1(y_0)) dy_0.$$

Step 4. $|T_{4t}| \rightarrow 0, |T_{5t}| \rightarrow 0$.

Then we have shown (18) with

$$\begin{aligned}
 v'_F(H) &= \frac{H_0(F_0^{-1}(s))}{f_0(F_0^{-1}(s))} \int_1^{F_1^{-1}(F_0(y_0)+\tau)} f(F_0^{-1}(s), y_1) dy_1 \\
 &+ \int_1^{F_0^{-1}(s)} \frac{H_0(y_0) - H_1(F_1^{-1}(F_0(y_0) + \tau))}{f_1(F_1^{-1}(F_0(y_0) + \tau))} f(y_0, F_1^{-1}(F_0(y_0) + \tau)) dy_0 \\
 &+ \int_1^{F_0^{-1}(s)} \int_1^{F_1^{-1}(F_0(y_0)+\tau)} dH(y_0, y_1),
 \end{aligned}$$

which is linear in H .

For Steps 1 and 2, we need the derivation

$$\begin{aligned}
 F_1(z_{1t}(y_0)) + tH_{1t}(z_{1t}(y_0)) &= F_{1t}(z_{1t}(y_0)) = F_{0t}(y_0) + \tau \\
 &= F_0(y_0) + \tau + tH_{0t}(y_0) = F_1(z_1(y_0)) + tH_{0t}(y_0),
 \end{aligned}$$

implying that

$$\begin{aligned}
 tH_{1t}(z_{1t}(y_0)) - tH_{0t}(y_0) &= F_1(z_1(y_0)) - F_1(z_{1t}(y_0)) \\
 &= [z_1(y_0) - z_{1t}(y_0)]f_1(\tilde{z}_{1t}(y_0)),
 \end{aligned}$$

where for any y_0 and t , $\tilde{z}_{1t}(y_0)$ lies in between $z(y_0)$ and $z_t(y_0)$. Therefore,

$$\frac{z_{1t}(y_0) - z_1(y_0)}{t} = \frac{H_{0t}(y_0) - H_{1t}(z_t(y_0))}{f_1(\tilde{z}_{1t}(y_0))}. \tag{19}$$

Similarly, $F_0(z_0) = s = F_{0t}(z_{0t}) = F_0(z_{0t}) + tH_{0t}(z_{0t})$, whence

$$\frac{z_{0t} - z_0}{t} = \frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})}. \tag{20}$$

Below, c denotes a generic constant, not always of the same value.

Proving Step 1 By a mean-value theorem argument,

$$\begin{aligned}
 |T_{10t}| &\leq \int_1^{z_{0t}} \left| \frac{z_{1t}(y_0) - z_1(y_0)}{t} [f(y_0, \tilde{z}_{1t}(y_0)) - f(y_0, z_1(y_0))] \right| dy_0 \\
 &\leq \int_1^{z_{0t}} \left| \frac{[z_t(y_0) - z(y_0)]^2}{t} f^{(1)}(y_0, \tilde{z}_{1t}(y_0)) \right| dy_0,
 \end{aligned}$$

where $f^{(1)}(\cdot, \cdot)$ denotes the derivative w.r.t. the second argument and $\tilde{z}_{1t}(y_0)$ lies between $z(y_0)$ and $z_t(y_0)$. Since $z_{0t} < \infty$, using (19) we get that

$$|T_{10t}| \leq t \int_1^\infty \left| \frac{[H_{0t}(y_0) - H_{1t}(z_{1t}(y_0))]^2}{f_1^2(\tilde{z}_{1t}(y_0))} f^{(1)}(y_0, \tilde{z}_{1t}(y_0)) \right| dy_0.$$

We show that (i) $[H_{0t}(y_0) - H_{1t}(z_{1t}(y_0))]^2$ is uniformly bounded, (ii) $f_1^2(\bar{z}_{1t}(y_0)) \geq \frac{c}{\bar{z}_{1t}(y_0)^{2\alpha}} \geq c(1 - F_0(y_0))^{2\alpha/(1-\alpha)}$ for y_0 large enough and t small enough, and (iii) $f^{(1)}(y_0, \bar{z}_{1t}(y_0)) \leq \frac{c}{y_0^{\alpha_0}}$ for some $\alpha_0 > 1$. Then we have

$$|T_{10t}| \leq ct \int_1^\infty \left| \frac{1}{y_0^{\alpha_0} (1 - F_0(y_0))^{2\alpha/(1-\alpha)}} \right| dy_0 \leq ct \int_1^\infty \frac{1}{y_0^{1+\varepsilon}} dy_0 \rightarrow 0$$

by A(iii).

To see (i), note that $\{[H_{0t}(y_0) - H_{1t}(z_{1t}(y_0))] - [H_0(y_0) - H_1(z_{1t}(y_0))]\}$ converges uniformly to 0, and $H_0(\cdot)$ and $H_1(\cdot)$ are uniformly bounded.

Next,

$$\begin{aligned} z_{1t}(y_0) &= F_{1t}^{-1}(F_{0t}(y_0) + \tau) \stackrel{(1)}{\leq} c(1 - F_0(y_0) - \tau)^{1/(1-\alpha)} \\ &= c(1 - F_0(y_0) - tH_{0t}(y_0) - \tau)^{1/(1-\alpha)} \\ &\leq c'(1 - F_0(y_0) - \tau)^{1/(1-\alpha)} \\ &\leq c(1 - F_0(y_0))^{1/(1-\alpha)} \end{aligned} \tag{21}$$

for small enough t , since $\alpha > 1$ and $tH_{0t}(\cdot)$ converges uniformly to 0. Inequality (1) comes from Condition A(i). Similarly,

$$z_1(y_0) \leq c(1 - F_0(y_0))^{1/(1-\alpha)} \tag{22}$$

and, therefore, (ii) follows. Finally, (iii) follows from (22), (22), and Condition A(ii).

Next, for $|T_{11t}|$, we have that

$$\begin{aligned} |T_{11t}| &\leq \left| \frac{z_{0t} - z_0}{t} \times \int_1^{z_1(y_0)} [f(\bar{z}_{0t}, y_1) - f(z_0, y_1)] dy_1 \right| \\ &\leq \left| \frac{[z_{0t} - z_0]^2}{t} \times \int_1^{z_1(y_0)} f^{(0)}(\bar{z}_{0t}, y_1) dy_1 \right| \\ &\leq t \left| \left[\frac{H_{0t}(z_{0t})}{f_0(\bar{z}_{0t})} \right]^2 \times \int_1^{z_1(y_0)} f^{(0)}(\bar{z}_{0t}, y_1) dy_1 \right| \\ &\stackrel{(2)}{\leq} ct \int_1^{z_1(y_0)} \frac{1}{y_0^{1+\delta}} dy_1 \leq ct \int_1^\infty \frac{1}{y_0^{1+\delta}} dy_1 \end{aligned}$$

for t small enough and some $\delta > 0$. Inequality (2) follows from Conditions B(i)–(iii) using arguments analogous to those for T_{10t} . This implies that $|T_{11t}| \rightarrow 0$.

Proving Step 2 We have

$$\begin{aligned} &\left| T_{2t} - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1) dy_1 \right| \\ &= \left| \int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} f(z_0, y_1) dy_0 - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1) dy_1 \right| \end{aligned}$$

$$\begin{aligned}
&= \left| \int_1^{z_{0t}} \frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})} f(z_0, y_1) dy_1 - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1) dy_1 \right| \\
&\leq \int_1^{z_{0t}} \left| \frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})} - \frac{H_0(z_0)}{f_0(z_0)} \right| f(z_0, y_1) dy_1 \\
&\stackrel{(1)}{\leq} c \int_1^{z_{0t}} \left| H_{0t}(z_{0t}) - H_0(z_0) \right| (1 - F_1(y_1))^{\beta/(\beta-1)} f(z_0, y_1) dy_1 \\
&\leq c \left(\sup_u |H_{0t}(u) - H_0(u)| + |H_0(z_{0t}) - H_0(z_0)| \right) \\
&\quad \times \int_1^\infty (1 - F_1(y_1))^{\beta/(\beta-1)} f(z_0, y_1) dy_1 \\
&\rightarrow 0
\end{aligned}$$

as $t \rightarrow 0$, by **B(iv)**. Inequality (1) is a consequence of **B(i)–B(iii)**.

Proving Step 3 We have

$$\begin{aligned}
&\left| T_{3t} - \int_1^{z_{0t}} \frac{H_0(y_0) - H_1(z_1(y_0))}{f_1(z_1(y_0))} f(y_0, z_1(y_0)) dy_0 \right| \\
&= \left| \int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} f(z_1(y_0), y_0) dy_0 \right. \\
&\quad \left. - \int_1^{z_{0t}} \frac{H_0(y_0) - H_1(z_1(y_0))}{f_1(z_1(y_0))} f(y_0, z_1(y_0)) dy_0 \right| \\
&\leq \int_1^{z_{0t}} \left| \frac{z_{1t}(y_0) - z_1(y_0)}{t} - \frac{H_0(y_0) - H_1(z_1(y_0))}{f_1(z_1(y_0))} \right| f(y_0, z_1(y_0)) dy_0 \\
&= \int_1^\infty \left| \frac{H_{0t}(y_0) - H_{1t}(z_{1t}(y_0))}{f_1(\tilde{z}_{1t}(y_0))} - \frac{[H_0(y_0) - H_1(z_1(y_0))]}{f_1(z_1(y_0))} \right| f(y_0, z_1(y_0)) dy_0 \\
&\stackrel{(1)}{\leq} c \int_1^\infty \left\{ |H_{0t}(y_0) - H_{1t}(z_{1t}(y_0)) - [H_0(y_0) - H_1(z_1(y_0))]| \right. \\
&\quad \left. \times (1 - F_0(y_0))^{\alpha/(\alpha-1)} f(y_0, z_1(y_0)) \right\} dy_0 \\
&\leq c \left\{ \sup_{y_0} |H_{0t}(y_0) - H_{1t}(z_t(y_0)) - [H_0(y_0) - H_1(z(y_0))]| \right. \\
&\quad \left. \times \int_1^\infty (1 - F_0(y_0))^{\alpha/(\alpha-1)} f(y_0, z_1(y_0)) dy_0 \right\},
\end{aligned}$$

which goes to zero if $\int_1^\infty (1 - F_0(y_0))^{\alpha/(\alpha-1)} f(z(y_0), y_0) dy_0 < \infty$, which is Condition **A(iv)**.

Note that the inequality $\stackrel{(0)}{\leq}$ follows from step (ii) in the proof of Step 1, above. Finally, since $\int_1^{z_{0t}} \frac{H_0(y_0) - H_1(z_1(y_0))}{f_1(z_1(y_0))} f(y_0, z_1(y_0)) dy_0$ is continuous in z_{0t} , the conclusion follows.

Proving Step 4 $T_{4t} \rightarrow 0$ since $H_t \rightarrow H$ uniformly and T_{5t} goes to zero by the continuous mapping theorem since paths of an F -Brownian bridge are everywhere continuous with probability 1. \square

Functional delta method and Proposition 3

Since $\sqrt{n}(\hat{F} - F) \rightsquigarrow \mathcal{G}$, we have from Lemma 1 and the functional delta method that

$$\sqrt{n}(\hat{v} - v_0) \xrightarrow{d} v'_F(\mathcal{G}),$$

whence $v'_F(\mathcal{G})$ is distributed as a univariate zero-mean normal given by

$$\begin{aligned} v'_F(\mathcal{G}) &= \frac{\mathcal{G}_0(F_0^{-1}(s))}{f_0(F_0^{-1}(s))} \int_1^{F_1^{-1}(F_0(y_0)+\tau)} f(F_0^{-1}(s), y_1) dy_1 \\ &\quad + \int_1^{F_0^{-1}(s)} \frac{\mathcal{G}_0(y_0) - \mathcal{G}_1(F_1^{-1}(F_0(y_0) + \tau))}{f_1(F_1^{-1}(F_0(y_0) + \tau))} f(y_0, F_1^{-1}(F_0(y_0) + \tau)) dy_0 \\ &\quad + \int_1^{F_0^{-1}(s)} \int_1^{F_1^{-1}(F_0(y_0)+\tau)} d\mathcal{G}(y_0, y_1), \end{aligned}$$

where \mathcal{G}_0 and \mathcal{G}_1 are stochastic processes defined from \mathcal{G} , analogous to (17), for example, $\mathcal{G}_0(a)$ is a univariate normal with mean zero and variance $F_0(a) \times [1 - F_0(a)]$. Now we can apply the functional delta method argument to justify consistency of the bootstrap via Van der Vaart and Wellner (1996), Theorem 3.9.11.

APPENDIX C

PROPOSITION 3. *Recall that*

$$\begin{aligned} v_c(\tau, s; x) &= \text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s, X = x) \\ &= \frac{\text{Prob}(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s | X = x)}{\text{Prob}(F_0(Y_0) \leq s | X = x)} := \frac{A(F_0, F_1, x)}{B(F_0, x)}, \end{aligned}$$

is estimated by

$$\begin{aligned} \hat{v}_c(\tau, s; x) &= \left(\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s) \right. \\ &\quad \left. / \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) \right) \\ &\quad \left/ \left(\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(\hat{F}_0(Y_{0i}) \leq s) / \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) \right) \right) \\ &:= \frac{\hat{A}_n(\hat{F}_0, \hat{F}_1, x)}{\hat{B}_n(\hat{F}_0, x)}, \end{aligned}$$

where for $l = 0, 1$, $\hat{F}_l(Y_{li}) = \frac{1}{n-1} \sum_{j \neq i} 1(Y_{lj} \leq Y_{li})$ and $K(\cdot)$ is a d -dimensional kernel function with a bandwidth sequence σ_n , satisfying the NW conditions specified in the Appendixes. The object whose distribution is needed is given by

$$\hat{v}_c(\tau, s; x) - v_c(\tau, s; x) = \frac{\hat{A}_n(\hat{F}_0, \hat{F}_1, x)}{\hat{B}_n(\hat{F}_0, x)} - \frac{A(F_0, F_1, x)}{B(F_0, x)}.$$

Let

$$\hat{f}(x) = \frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right),$$

$$\hat{A}_n(F_0, F_1, x) = \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)}{\hat{f}(x)},$$

$$\hat{B}_n(F_0, x) = \frac{\frac{1}{n\sigma_n^d} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1(F_0(Y_{0i}) \leq s)}{\hat{f}(x)},$$

$$W := 1(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s),$$

$$V = 1(F_0(Y_0) \leq s).$$

Then

$$\begin{aligned} T_n(x) &:= \hat{A}_n(\hat{F}_0, \hat{F}_1, x) - A(F_0, F_1, x) \\ &= \underbrace{\{\hat{A}_n(F_0, F_1, x) - A(F_0, F_1, x)\}}_{\tilde{T}_{1n}(x)} + \underbrace{\{\hat{A}_n(\hat{F}_0, \hat{F}_1, x) - \hat{A}_n(F_0, F_1, x)\}}_{\tilde{T}_{2n}(x)}. \end{aligned}$$

We show that $\tilde{T}_{2n}(x)$ is of smaller order of magnitude than $\tilde{T}_{1n}(x)$. This implies that, asymptotically, the distribution of T_n will be that of \tilde{T}_{1n} , which is simply the Nadaraya-Watson regression of the unobserved random variable W on X , evaluated at $X = x$ and denoted by $\hat{E}(W|X = x)$. The formal result is stated below and its proof appears in the appendix under Theorem 2. An exactly analogous result holds for the denominator $\hat{B}_n(\hat{F}_0, x)$.

The following additional assumption is used.

ASSUMPTION NW5'. The functions $f(x)$, $E(W|X = x)$, and $f(x) \times E(W|X = x)$ are twice differentiable, and the functions and their derivatives up to order 2 are continuous and uniformly bounded.

CLAIM. Suppose the data (X_i, Y_{1i}, Y_{0i}) for $i = 1, \dots, n$, are i.i.d. and assumptions NW1-4 and NW5' hold. Then we have that

$$(n\sigma_n^d)^{1/2} \tilde{T}_{2n}(x) = o_p(1).$$

PROOF OF CLAIM AND THUS PROPOSITION 3. Let $T_{2n}(x) = \tilde{T}_{2n}(x)\hat{f}(x)$. We now show that $E(\sqrt{n\sigma_n^d}|T_{2n}(x)|) \rightarrow 0$, which implies that $|T_{2n}(x)| = o_p((n\sigma_n^d)^{-1/2})$, and thus establish the result since $\hat{f}(x)$ is bounded in probability by assumption.

First observe that

$$\begin{aligned}
 & E\left[|1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s) \right. \\
 & \quad \left. - 1(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)\right] \\
 &= \text{Prob}\left[|1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s) \right. \\
 & \quad \left. - 1(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)\right] \neq 0 \\
 &= \text{Prob}\left[\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s\} \right. \\
 & \quad \left. \cap (F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)^c\right] \\
 & \quad + \text{Prob}\left[\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s\}^c \right. \\
 & \quad \left. \cap (F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)\right] \\
 &\leq \text{Prob}\left[\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s\} \cap (F_1(Y_{1i}) - F_0(Y_{0i}) \leq \tau)\right] \\
 & \quad + \text{Prob}\left[\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s\} \cap (F_0(Y_{0i}) > s)\right] \\
 & \quad + \text{Prob}\left[\{F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s\} \cap (\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) \leq \tau)\right] \\
 & \quad + \text{Prob}\left[\{F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s\} \cap (\hat{F}_0(Y_{0i}) > s)\right] \\
 &\leq \text{Prob}\left[\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, F_1(Y_{1i}) - F_0(Y_{0i}) \leq \tau\right] \\
 & \quad + \text{Prob}\left\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) \leq \tau, F_1(Y_{1i}) - F_0(Y_{0i}) > \tau\right\} \\
 & \quad + \text{Prob}\left[\hat{F}_0(Y_{0i}) \leq s, F_0(Y_{0i}) > s\right] + \text{Prob}\left[F_0(Y_{0i}) \leq s, \hat{F}_0(Y_{0i}) > s\right].
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 & E(\sqrt{n\sigma_n^d}|T_{2n}(x)|) \\
 &= \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E\left(K\left(\frac{X_i - x}{\sigma_n}\right) \left|1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s) \right. \right. \\
 & \quad \left. \left. - 1(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)\right)\right) \\
 &= \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}, Y_{1i}} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) E\left(|1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s) \right. \right. \\
 & \quad \left. \left. - 1(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \leq s)\right) \middle| X_i, Y_{0i}, Y_{1i} \right\} \\
 &\leq \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}, Y_{1i}} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \right. \tag{23}
 \end{aligned}$$

$$\begin{aligned}
 & \times \text{Prob}\left\{\left(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, F_1(Y_{1i}) - F_0(Y_{0i}) \leq \tau\right) \mid X_i, Y_{0i}, Y_{1i}\right\} \\
 & + \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}, Y_{1i}} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \right. \\
 & \times \text{Prob}\left\{\left(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) \leq \tau, F_1(Y_{1i}) - F_0(Y_{0i}) > \tau\right) \mid X_i, Y_{0i}, Y_{1i}\right\} \\
 & + \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}} E \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \text{Prob}\{\hat{F}_0(Y_{0i}) \leq s, F_0(Y_{0i}) > s \mid X_i, Y_{0i}\} \right\} \\
 & + \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}} E \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \text{Prob}\{\hat{F}_0(Y_{0i}) > s, F_0(Y_{0i}) \leq s \mid X_i, Y_{0i}\} \right\} \\
 & := S_{1n} + S_{2n} + S_{3n} + S_{4n}, \text{ say.}
 \end{aligned}$$

We show that $S_{1n} \rightarrow 0$ and an exactly analogous proof shows that S_{2n}, S_{3n} , and S_{4n} are also $o(1)$.

Now, for fixed X_i, Y_{0i} , and Y_{1i} , and the fact that, for example, $\hat{F}_1(Y_{1i}) = \frac{1}{n-1} \times \sum_{j \neq i} 1(Y_{ij} \leq Y_{1i})$, we have that

$$\begin{aligned}
 & \text{Prob}(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, F_1(Y_{1i}) - F_0(Y_{0i}) < \tau \mid X_i, Y_{0i}, Y_{1i}) \\
 & = \text{Prob}(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) - (F_1(Y_{1i}) - F_0(Y_{0i})) > \tau - (F_1(Y_{1i}) - F_0(Y_{0i})), \\
 & \quad F_1(Y_{1i}) - F_0(Y_{0i}) < \tau \mid X_i, Y_{0i}, Y_{1i}) \\
 & \leq \exp(-2(n-1)(\tau - (F_1(Y_{1i}) - F_0(Y_{0i})))^2) \times 1(F_1(Y_{1i}) - F_0(Y_{0i}) < \tau)
 \end{aligned}$$

by Hoeffding’s inequality (note that conditional on $Y_{i1}, \hat{F}_1(Y_{1i}) = \frac{1}{n-1} \sum_{j \neq i} 1(Y_{ij} \leq Y_{1i})$ is an average of independent, binary (0, 1) random variables, thus satisfying the hypothesis of Hoeffding’s inequality). Thus, we have that

$$\begin{aligned}
 S_{1n} & \leq \frac{1}{\sqrt{n\sigma_n^d}} \sum_{i=1}^n E_{X_i, Y_{0i}, Y_{1i}} \left[K\left(\frac{X_i - x}{\sigma_n}\right) \right. \\
 & \quad \left. \times \exp(-2(n-1)(\tau - (F_1(Y_{1i}) - F_0(Y_{0i})))^2) 1(F_1(Y_{1i}) - F_0(Y_{0i}) < \tau) \right] \\
 & = \frac{n}{\sqrt{n\sigma_n^d}} E_{X, Y_0, Y_1} \left[K\left(\frac{X - x}{\sigma_n}\right) \right. \\
 & \quad \left. \times \exp(-2(n-1)(\tau - (F_1(Y_1) - F_0(Y_0)))^2) 1(F_1(Y_1) - F_0(Y_0) < \tau) \right] \\
 & = \frac{n}{\sqrt{n\sigma_n^d}} E_X \left[K\left(\frac{X - x}{\sigma_n}\right) G_n(X) \right],
 \end{aligned}$$

where

$$G_n(x) = E_{Y_0, Y_1|X}[\exp(-2(n-1)(\tau - (F_1(Y_1) - F_0(Y_0))))^2 \times 1(F_1(Y_1) - F_0(Y_0) < \tau)|X = x].$$

Continuing with the previous display, we have

$$\begin{aligned} S_{1n} &\leq \frac{n}{\sqrt{n\sigma_n^d}} E_X \left[K \left(\frac{X-x}{\sigma_n} \right) G_n(X) \right] \\ &= \frac{n\sigma_n^d}{\sqrt{n\sigma_n^d}} \int [K(u)G_n(x + \sigma_n u)f(x + \sigma_n u)] du \\ &= \sqrt{n\sigma_n^d} \int [K(u)G_n(x + \sigma_n u)f(x + \sigma_n u)] du \tag{24} \\ &= f(x)\sqrt{n\sigma_n^d} \int K(u)G_n(x) du + \text{terms of smaller order} \\ &= f(x)\sqrt{n\sigma_n^d} G_n(x) + \text{terms of smaller order.} \end{aligned}$$

Now, notice that defining $Z = \tau - (F_1(Y_{1i}) - F_0(Y_{0i}))$, then $G_n(x)$ is of the form

$$\begin{aligned} G_n(x) &= E_{Z|X}[\exp(-2(n-1)Z^2) \times 1(Z > 0)|X = x] \\ &\leq c \int \exp(-2(n-1)z^2)f(z|x) dz \\ &\leq c' \int \exp(-2(n-1)z^2) dz \tag{25} \\ &= O(n^{-1/2}) \end{aligned}$$

by the normal (Gaussian) integral formula. From (24) and (25), it follows that

$$E(\sqrt{n\sigma_n^d}|T_{2n}(x)|) = O(n^{-1/2} \times \sqrt{n\sigma_n^d}) = O(\sqrt{\sigma_n^d}) = o(1).$$

Together with analogous proofs for S_{2n} , S_{3n} , and S_{4n} , this implies that $\sqrt{n\sigma_n^d}T_{2n}(x) = o_p(1)$, which is the desired result. □

APPENDIX D

Uniform confidence bands for conditional mobility (following Hardle (1990, Algorithm 4.3.2))

First note that the inequality in line 3 of (25) is uniform in x because X has bounded support. Consequently the “terms of smaller order” in (24) are also small uniformly in x . Given that $f(\cdot)$ is also uniformly bounded by $NW5'$, the conclusion $\sqrt{n\sigma_n^d}T_{2n}(x) = o_p(1)$ can be strengthened to $\sup_x \sqrt{n\sigma_n^d}|T_{2n}(x)| = o_p(1)$. This shows that $(n\sigma_n^d)^{1/2}[\hat{v}_c(\tau, s; x) - v_c(\tau, s; x)]$, as an empirical process, is asymptotically equivalent to linear combinations

of the Nadaraya–Watson regression residual processes

$$(n\sigma_n^d)^{1/2}\{\hat{E}(W|X=x) - E(W|X=x)\}$$

and

$$(n\sigma_n^d)^{1/2}\{\hat{E}(V|X=x) - E(V|X=x)\},$$

which converge weakly to zero-mean Gaussian processes. Consequently, they are amenable to the treatment in Hardle (1990), based on Bickel and Rosenblatt (1973), which develops uniform confidence bands for NW regression curves. We now outline Hardle's construction.

For each sample value x of the conditioning variable X , bandwidth σ_n , and kernel $K(\cdot)$, denote estimated density at $X = x$ by

$$\hat{f}(x) = \frac{1}{n\sigma_n} \sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right).$$

Consider dependent variables $W_i = 1(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s)$ for upward mobility and $W_i = 1(Y_{1i} \leq \hat{\xi}_1, Y_{0i} \leq \hat{\xi}_0)$ for transition probability. Denote regression estimate (predicted value) at $X = x$ by

$$\hat{\mu}(x) = \frac{\frac{1}{n\sigma_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\sigma_n}\right) W_i}{\hat{f}(x)}.$$

Corresponding to the Epanechnikov kernel, set

$$c_K = \int_{-1}^1 K^2(u) du = \int_{-1}^1 \frac{9}{16}(1 - u^2)^2 du = \frac{3}{5} = 0.6,$$

$$C_2 = \frac{\int_{-1}^1 \{K'(u)\}^2 du}{2c_K} = 1.25,$$

$$\delta = \sqrt{2 \ln\left(\frac{1}{\sigma_n}\right)},$$

$$d_n = \sqrt{\left(2 \ln\left(\frac{1}{\sigma_n}\right)\right)} + \frac{1}{2\sqrt{\left(2 \ln\left(\frac{1}{\sigma_n}\right)\right)}} \ln\left(\frac{C_2}{2\pi^2}\right)$$

$$= \delta + \frac{1}{2\delta} \ln\left(\frac{C_2}{2\pi^2}\right),$$

$$c_\alpha = -\ln(-0.5 \times \ln(1 - 0.05)) = 3.66,$$

$$\omega^2(x) = \frac{1}{n\sigma_n \hat{f}(x)} \sum_{i=1}^n \{W_i - \hat{\mu}(X_i)\}^2 K\left(\frac{X_i - x}{\sigma_n}\right).$$

Then lower and upper limit of uniform confidence integrals are given by

$$C_{\text{LO}}(x) = \hat{\mu}(x) - \left\{ \frac{c_\alpha}{\delta} + d_n \right\} \times \frac{\sqrt{\frac{\omega^2(x)}{n\sigma_n \times \hat{f}(x)}}}{c_K},$$

$$C_{\text{UP}}(x) = \hat{\mu}(x) + \left\{ \frac{c_\alpha}{\delta} + d_n \right\} \times \frac{\sqrt{\frac{\omega^2(x)}{n\sigma_n \times \hat{f}(x)}}}{c_K}.$$

Use the following additional assumption:

NW6. The density of X is bounded away from zero on its support.

Justification of the above algorithm comes from theorem 4.3.1 of [Hardle \(1990\)](#). Our Assumptions NW implies conditions A1–A5 of [Hardle \(1990\)](#). To see this, note that NW5' implies A1 (note that for a binary outcome Y , the variance $\text{var}(Y|x) = E(Y|x) \times (1 - E(Y|x))$) so that differentiability of $E(Y|x)$ implies differentiability of $\text{var}(Y|x)$; the Epanechnikov kernel satisfies Condition A2, the outcome variables lie in $[0, 1]$ with probability 1, implying A3; Condition NW6 implies A4 and choice of bandwidth $\sigma_n = n^{-1/4}$ implies A5.

REFERENCES

- Andrews, D. W. K. (1994), “Empirical process methods in econometrics.” In *Handbook of Econometrics*, Volume 4, 2247–2294, Elsevier, Amsterdam. [364]
- Barrett, G. F. and S. G. Donald (2009), “Statistical inference with the generalized Gini indices of inequality, poverty and welfare.” *Journal of Business and Economic Statistics*, 27 (1), 1–17. [346]
- Bhattacharya, D. (2007), “Inference on inequality from household survey data.” *Journal of Econometrics*, 137 (2), 674–707. [346]
- Bickel, P. J. and M. Rosenblatt (1973), “On some global measures of the deviations of density function estimates.” *The Annals of Statistics*, 1 (6), 1071–1095. [357, 376]
- Bierens, H. (1994), *Topics in Advanced Econometrics*. Cambridge University Press, Cambridge, UK. [348, 362, 364]
- Black, S. and P. Devereux (2010), “Recent developments in intergenerational mobility.” In *Handbook of Labor Economics*, Volume 4 (D. Card and O. Ashenfelter, eds.), Elsevier, Amsterdam. [338]
- Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994), “Evidence on the validity of cross-sectional and longitudinal labor market data.” *Journal of Labor Economics*, 12 (3), 345–368. [343]
- Cameron, S. V. and J. J. Heckman (2001), “The dynamics of educational attainment for black, Hispanic and white males.” *Journal of Political Economy*, 109 (3), 455–499. [336, 355]

- Chay, K., J. Guryan, and B. Mazumder (2009), "Birth cohort and the black–white achievement gap: The role of health soon after birth." Working paper, Federal Reserve Bank of Chicago. [338, 360]
- Chevalier, J. and G. Ellison (1999), "Are some mutual fund managers better than others? Cross sectional patterns in behavior and performance." *Journal of Finance*, 54 (3), 875–899. [361]
- Corcoran, M. and T. Adams (1997), "Race, sex, and the inter-generational transmission of poverty." In *Consequences of Growing up Poor* (G. Duncan and J. Brooks-Gunn, eds.), 461–517, Russell Sage Foundation, New York. [338]
- Datcher, L. (1981), "Race/sex differences in the effects of background on achievement." In *Five Thousand American Families: Patterns of Economic Progress*, Volume IX (M. S. Hill, D. H. Hill, and J. N. Morgan, eds.), 359–390, Institute for Social Research, University of Michigan, Ann Arbor. [338]
- Dobbie, W. and R. Fryer (forthcoming), "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem children's zone." *American Economic Journal: Applied Economics*. [360]
- Formby, J., W. J. Smith, and B. Zheng (2004), "Mobility measurement, transition matrices and statistical inference." *Journal of Econometrics*, 120, 181–205. [337, 344, 349]
- Fryer, R. (2010), "Racial inequality in the 21st century: The declining significance of discrimination." In *Handbook of Labor Economics*, Volume 4 (D. Card and O. Ashenfelter, eds.), Elsevier, Amsterdam. [338]
- Haider, S. J. and G. Solon (2006), "Life cycle variation in the association between current and lifetime earnings." *American Economic Review*, 96 (4), 1308–1320. [343]
- Hansen, K., J. Heckman, and K. Mullen (2004), "The effect of schooling and ability on achievement test scores." *Journal of Econometrics*, 121 (1–2), 39–98. [338]
- Hardle, W. (1990), *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK. [348, 357, 376, 377]
- Heckman, J., J. Stixrud, and S. Urzua (2006), "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24 (3), 411–482. [361]
- Hertz, T. (2005), "Rags, riches and race: The intergenerational economic mobility of black and white families in the United States." In *Unequal Chances: Family Background and Economic Success* (S. Bowles, H. Gintis, and M. O. Groves, eds.), Princeton University Press, Princeton, NJ. [338, 358]
- Hertz, T. (2008), "A group specific measure of intergenerational persistence." *Economic Letters*, 100 (3), 415–417. [339]
- Jencks, C. and L. Tach (2006), "Would equal opportunity mean more mobility?" In *Mobility and Inequality: Frontiers of Research From Sociology and Economics* (S. Morgan and D. Grusky, eds.), Stanford University Press, Palo Alto, CA. [361]

- Kopczuk, W., E. Saez, and J. Song (2010), “Earnings inequality and mobility in the United States: Evidence from social security data since 1937.” *Quarterly Journal of Economics*, 125 (1), 91–128. [361]
- Lee, C.-I. and G. Solon (2009), “Trends in intergenerational income mobility.” *Review of Economics and Statistics*, 91 (4), 766–772. [339]
- Neal, D. A. and W. R. Johnson (1996), “The role of premarket factors in black–white wage differences.” *Journal of Political Economy*, 104 (5), 869–895. [336, 338, 349, 355]
- O’Neill, D., O. Sweetman, and D. Van de Gaer (2006), “The impacts of cognitive skills on the distribution of the black–white wage gap.” *Labour Economics*, 13, 343–356. [355]
- O’Neill, D., O. Sweetman, and D. Van de Gaer (2007), “The effects of measurement error and omitted variables when using transition matrices to measure inter-generational mobility.” *Journal of Economic Inequality*, 5, 159–178. [343]
- Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*. Cambridge University Press. [362]
- Powell, J., J. Stock, and T. Stoker (1989), “Semiparametric estimation of index coefficients.” *Econometrica*, 57, 1403–1430. [343]
- Roemer, J. E. (2004), “Equal opportunity and inter-generational mobility: Going beyond inter-generational income transition matrices.” In *Generational Income Mobility in North America and Europe* (M. Corak, ed.), Cambridge University Press, Cambridge, UK. [361]
- Shorrocks, A. F. (1978), “The measurement of mobility.” *Econometrica*, 46 (5), 1013–1024. [342]
- Solon, G. (1992), “Intergenerational income mobility in the United States.” *American Economic Review*, 82 (3), 393–408. [348]
- Solon, G. (1999), “Intergenerational mobility in the labor market.” In *Handbook of Labor Economics*, Volume 3A (O. C. Ashenfalter and D. Card, eds.), Elsevier, Amsterdam. [338]
- Swift, A. (2005), “Justice, luck and the family.” In *Unequal Chances: Family Background and Economic Success* (S. Bowles, H. Gintis, and M. O. Groves, eds.), Princeton University Press, Princeton, NJ. [361]
- Trede, M. (1998), “The age profile of mobility measures: An application to earnings in West Germany.” *Journal of Applied Econometrics*, 13, 397–409. [337]
- Van De Gaer, D., E. Schokkaert, and M. Martinez (2001), “Three meanings of intergenerational mobility.” *Economica*, 272, 519–538. [361]
- Van der Vaart, A. and J. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. [346, 347, 371]