Nonparametric Analysis of Intergenerational Income Mobility with Application to the United States

Debopam Bhattacharya
Dartmouth College

Bhashkar Mazumder
Federal Reserve Bank of Chicago

Sept 26, 2007

Abstract

This paper concerns the problem of inferring the effects of covariates on intergenerational income mobility, i.e. on the relationship between the incomes of parents and future earnings of their children. We focus on two different measures of mobility- (i) traditional transition probability of movement across income quantiles over generations and (ii) a new direct measure of upward mobility, viz. the probability that an adult child's relative position exceeds that of the parents. We estimate the effect of possibly continuously distributed covariates from data using nonparametric regression and average derivatives and derive the distribution theory for these measures. The analytical novelty in the derivation is that the dependent variables involve nonsmooth functions of estimated components- marginal quantiles for transition probabilities and relative ranks for upward mobility- thus necessitating nontrivial modifications of standard nonparametric regression theory. We use these methods on US data from the National Longitudinal Survey of Youth to study black-white differences in intergenerational mobility, a topic which has received scant attention in the literature. We document that whites experience greater intergenerational mobility than blacks. Estimates of conditional mobility using nonparametric regression reveal that most of the interracial mobility gap can be accounted for by differences in cognitive skills during adolescence. The methods developed here have wider applicability to estimation of nonparametric regression and average derivatives where the dependent variable either involves a preliminary finite-dimensional estimate in a nonsmooth way or is a nonsmooth functional of ranks of one or more random variables.

1 Introduction

Intergenerational mobility (IGM, henceforth) refers to the extent of movement in economic status across generations. A society with high mobility offers greater "equality of opportunity" at birth because parent's economic status in such a society has a smaller effect on the eventual economic status of their children. Most of the empirical economic literature on IGM has focused on a single parameter as a summary measure of mobility viz. the elasticity of children's earnings with respect to their parents' earnings.[1] However, there is clearly reason to think that mobility patterns might differ markedly across subgroups of the population. In the US, for instance, it is particularly interesting to know whether the prospects for upward mobility differ across racial groups, given the legacy of slavery and segregation. From a policy perspective, it is also important to understand the channels through which intergenerational income persistence arises. For example, if parents face borrowing constraints that impede human capital investment this could induce lower mobility for parents with talented children but limited means (Becker and Tomes, 1979). Therefore, it would be useful to know how mobility measures are affected by the inclusion of covariates that measure levels of human capital such as schooling or test scores.

This paper develops a theory of nonparametric inference for estimating the effects of covariates on a set of statistics based on the joint distribution of parent and child income. This distribution theory allows us to investigate differences in IGM across population subgroups and to examine how these differences are affected by covariates without making restrictive distributional or functional form assumptions on the data generating process. The methodology is then applied to data from the National Longitudinal Survey of Youth (NLSY) where we investigate differences in mobility patterns between black and white men in the US.

Theoretical analysis of intergenerational income mobility has traditionally been based on two broad and complementary measures- (i) the intergenerational elasticity (IGE)[2] which is simply the regression coefficient obtained by regressing (log) child's permanent income on (log) parents' permanent income and (ii) matrices of transition probabilities

[1]See Corak (2006) or Bowles and Gintis (2002) for recent surveys of the literature.

[2]The intergenerational correlation (IGC) has also been used by many researchers. The IGC is qalitatively similar to the IGE and the two measures are equivalent when the variance in in income is unchanging across generations.

which measure the rates of movement across quantiles of the distribution over subsequent generations. The IGE, therefore, relates *mean* economic status in one generation to economic status in the previous generation,[3] while transition probabilities capture relationships between relative positions across generations including positions far away from the median. The IGE, for instance, tells us nothing about a son's probability of staying in the bottom quintile of the overall income distribution of his cohort, given his father's (relative) position among his generation.

Another important limitation of the IGE is that one cannot use it to analyze mobility differences across 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. In contrast, transition probabilities can be used to make statements concerning the movements of blacks across the income distribution of the entire population comprising both blacks and whites.

On the other hand, it is straightforward to measure effects of covariates on the IGE, one simply needs to include the covariates and their interactions with parents' income as additional regressors and the statistical theory is straightforward. In contrast, a formal statistical theory for using covariates in transition matrices seems to be lacking. The development of such a theory therefore would combine the advantages of transition matrix based approaches together with the ability to investigate the effects of covariates on mobility. For instance, using PSID data from the US, Hertz (2005) has previously shown that mobility among blacks is low at the low end of the income distribution and high at the high end relative to whites. Conditional transition probabilities will let us infer to what extent this difference can be attributed to differences across say, education groups rather than differences within education levels. In other words, it can be used to answer questions like: would an additional year of education increase a black son's mobility by more or less than that for a white son. In addition, we can also investigate black-white differences in the mobility "levels" at different points in the distribution of parent income. In that sense, compared to the IGE, *conditional* transition probabilities allow us to answer a richer set of questions that are of more direct relevance for policy design.

The distribution theory for estimated *marginal* transition probabilities was previously developed in Formby, Smith and Zheng (2004), henceforth FSZ. When the relevant covari-

[3] One may use quantile regressions to calculate IGE to avoid problems of outliers in and top-coding of earnings data.

ates are discrete, one can simply apply the results of FSZ 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.[4] The theory of inference here is distinct from that of usual nonparametric regression because the dependent variable involves nonsmooth functions of the relevant quantiles which are also typically estimated from the same data. One relevant covariate of significant interest is the son's percentile score on Armed Forces Qualifying Test (often taken to be an indicator of cognitive skills in adolescent years- see discussion below) which runs from 1 to 100 in the data. Given the size of typical household surveys, analysis within each percentile will be very imprecise, particularly when broken up by racial subgroups. 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 will be the natural and correct approach.

However, one drawback of using transition probabilities is its overtly disaggregate nature (i.e. there are an infinite number of transition probabilities depending on which quantiles are compared) and a summary measure of mobility across *relative* income positions is useful for consolidating the information provided in transition matrices. In this paper, we introduce a new and intuitively simple measure of overall upward mobility in relative terms and derive distribution theories for inferring both its level and the effects of covariates on it from data. Our measure is simply the probability that a son's income rank in his generation exceeds his parents' rank in the prior generation. A nice feature of this measure is that it probably captures what most policymakers actually think of as mobility, namely to what extent are people doing better than their parents? It is a single summary measure and therefore easy to interpret. At the same time, its value does not depend on arbitrary discretization of income distributions, unlike existing measures of overall mobility.[5] In our application with US data, we show that black-white differences in mobility are much smaller when based on our measure of upward mobility

[4]Parametric analysis is problematic here because it is unclear what type of joint distributions imply, e.g. a probit form for conditional transition probabilities. In the application, we show that restrictive parametric inference, as an approximation, can produce misleading qualitative conclusions (see fig. 7 and discussion in section 4.2.1, below).

[5]Several overall mobility measures based on arbitrary discretization of income exist in the literature (see section 2.2 below for more on this point).

compared to traditional (upward) transition probabilities. This is because (i) our measure captures small upward movements in relative positions of a large number of Blacks which are ignored by transition probabilities and (ii) blacks typically need much larger upward mobility to surpass a common percentile threshold than whites because white incomes first-order stochastically dominate blacks everywhere (c.f. section 2.2 below) and, in particular, below any common threshold.

1.1 Contributions

The key methodological contribution of this paper is to develop a nonparametric theory of inference when the outcome variables of interest involve nonsmooth functions of initially estimated parameters or functionals. For transition probabilities, the outcome variable involves a nonsmooth function of marginal quantiles of income. For estimates of upward mobility, the outcome variable involves nonsmooth functions of estimated ranks. To our knowledge, no method currently exists in the econometric or statistics literature for deriving the distribution theory for such estimates. In the present paper, we develop precisely such a methodology.

For analysis of conditional transition probabilities and its derivatives, we control first stage estimation errors using stochastic equicontinuity type arguments. But owing to the peculiar forms of the parameters here, one cannot use sufficient conditions typically used in the parametric or semiparametric literature (c.f. Andrews (1994), section 4). We therefore establish equicontinuity properties directly using the U-statistic type forms of the relevant error processes. For analysis of upward mobility involving estimated ranks in the first stage, we show and use Hadamard differentiability of the map from the joint c.d.f. of father and son's income to upward mobility and for controlling first stage estimation errors, we use Hoeffding's inequality.

For both measures, the initial estimation error affects the asymptotic distribution of both the unconditional mobility estimate and its density-weighted average derivative (w.r.t. covariates) but not that of the level effects of covariates on it. The intuition behind this result is that the initial estimates have parametric or exponential rates of convergence while estimated level effects of covariates typically converges at slower rates.

Although the methods developed here are motivated by and are crucial for nonparametric analysis of IGM, they have more general applicability. For one, these tools can be directly used to analyze economic mobility for an individual over her own lifetimei.e. how eventual position in the income distribution of her cohort (say, at retirement) is related to her starting position. Another example would be the relation between initial and eventual positions of a firm in the size distribution of firms in its industry. 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 dataset, the methods developed here can be utilized to get the respective distribution theories. For instance, consider estimating the effect of covariates on the probability that a household lies below the poverty line. The poverty line is often estimated from the same dataset[6] and therefore the methods developed here will be applicable to that case directly. Alternatively, consider estimating the effects of covariates like study hours and IQ score on the probability that a high school senior is in the top 5% of his graduating class. High school class rank often has important implications for college admissions beyond high-school GPA on which this rank is based. They reveal a student's relative performance with respect to his peers in the same school and so implicitly control for inter-school variations in absolute grades. The methodology of this paper can be used to conduct statistical inference on regressions of such *relative* outcomes on covariates. Note well that such relative regressions tell us something very different from quantile regressions. The former captures effects of covariates on the relative position in the *marginal* distribution of the dependent variable but the latter pertains to relative position in its *conditional* (on the covariate) distribution.

We also provide a number of novel substantive contributions to the empirical literature on IGM. We are the first to use a large and nationally representative sample of blacks and whites to estimate interracial differences in IGM.[7] We document that there are sizable differences in upward transition probabilities between blacks and whites. We apply nonparametric methods to investigate how a continuous measure of cognitive skill, the Armed Forces Qualifying Test (AFQT), affects the transition probabilities for blacks and whites. AFQT scores have been previously used to account for other aspects of black-white inequality (Neal and Johnson, 1996; Cameron and Heckman, 2001) but, to our knowledge, have not been used in studies of IGM. Interestingly, we show that this variable can account for most of the gap between blacks and whites in their ability to

[6] Zheng (2001) performs inference on the marginal rather than the conditional distribution of estimated poverty with an estimated poverty line.

[7] Hertz (2005) also shows large black-white differences using the PSID but for reasons we discuss in the paper it is unclear whether the PSID sample of blacks is representative.

rise from the bottom quintile of the income distribution. We also find similar results with respect to inter-racial differences in our new measure of upward mobility. We show that alternative parametric methodology, such as running probit regressions for conditional transitions lead to misleading conclusions. The reader may note that in accordance with virtually all of the existing literature on economic mobility including all the papers cited above and below, what we propose here are essentially descriptive measures which are suggestive and one should be cautious in attaching causal interpretations to them.

1.2 Plan of the paper

The plan of the paper is as follows. Section 2 describes the parameters of interest, section 3 discusses the asymptotic distribution theory, section 4 presents the application using NLSY data. Finally, section 5 concludes. All proofs are collected in section 6. In the statements of the theorems and in the proofs, c will denote a generic positive constant not always having the same value and whenever derivatives (or Lebesgue densities) are defined, they are implicitly assumed to exist.

2 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.

2.1 Conditional transition probabilities and derivatives

Let (ζ_1, ζ_0) denote the bottom tth and sth percentile of the overall income distribution for sons and fathers, respectively. Then, the transition probability measures the probability that a son is at or below ζ_1, conditional on his father being at or below ζ_0, i.e.

$$\theta(s,t) = \frac{\Pr[Y_1 \leq \zeta_1, Y_0 \leq \zeta_0]}{\Pr[Y_0 \leq \zeta_0]}. \tag{1}$$

$\theta(s,t)$ can be decomposed by level of discrete and continuous covariates X such as age and education of the father and/or the son as

$$\begin{aligned}\theta(s,t) &= \int \frac{\Pr[Y_1 \leq \zeta_1, Y_0 \leq \zeta_0 | X = x]}{\Pr[Y_0 \leq \zeta_0]} dF(x) \\ &= \int \theta(x; s, t)\, dF(x),\end{aligned}$$

where

$$\theta(x;s,t) = \frac{\Pr[Y_1 \le \zeta_1, Y_0 \le \zeta_0 | X = x]}{\Pr[Y_0 \le \zeta_0]}. \qquad (2)$$

Note that the denominator is not conditioned on X. So when X denotes, say, a son's education, the derivative of $\theta(x;s,t)$ w.r.t. x represents the change in the son's probability of being stuck at or below ζ_0 when x increases by 1 unit, where the population of interest is all families whose fathers' incomes were at or below ζ_0. When X denotes the father's education, $\theta(x;s,t)$ can be used to infer how effective is a highly educated father in improving his son's condition relative to a less educated father who was in the same quantile of income as him.

Note also that from the transition probabilities as defined above, one can derive the ones defined in, e.g. FSZ (2004) or Shorrocks (1978). The latter, upon conditioning, would be defined as

$$\pi(x;(s_1,s_2),(t_1,t_2)) = \frac{\Pr\left[\zeta_{1s_1} \le Y_1 \le \zeta_{1s_2}, \zeta_{0t_1} \le Y_0 \le \zeta_{0t_2} | X = x\right]}{\Pr\left[\zeta_{0t_1} \le Y_0 \le \zeta_{0t_2}\right]}, \qquad (3)$$

where $0 < s_1 < s_2 < 1$ and $0 < t_1 < t_2 < 1$ and ζ_{1s_j}, ζ_{0t_j} denote the s_j and t_jth quantiles of Y_0 and Y_1, respectively. The numerator equals

$$\begin{aligned} &\Pr\left[Y_1 \le \zeta_{1s_2}, Y_0 \le \zeta_{0t_2} | X = x\right] \\ &- \Pr\left[Y_1 \le \zeta_{1s_2}, Y_0 \le \zeta_{0t_1} | X = x\right] \\ &- \Pr\left[Y_1 \le \zeta_{1s_1}, Y_0 \le \zeta_{0t_2} | X = x\right] \\ &+ \Pr\left[Y_1 \le \zeta_{1s_1}, Y_0 \le \zeta_{0t_1} | X = x\right] \end{aligned}$$

and the denominator is $\Pr\left[Y_0 \le \zeta_{0t_2}\right] - \Pr\left[Y_0 \le \zeta_{0t_1}\right]$. Below, we derive distribution theory for estimates of $\Pr[Y_1 \le \zeta_{1s}, Y_0 \le \zeta_{0t} | X = x]$ for a generic pair of quantiles (ζ_{1s}, ζ_{0t}) from which one can easily get the distribution of the estimates of the numerator of $\pi(x;(s_1,s_2),(t_1,t_2))$. The distribution theory for the estimate of $\pi(x;(s_1,s_2),(t_1,t_2))$ follows by the usual delta-method.

We choose to focus both our methodology and application on definition (2) rather than (3) for two reasons. First, definition (2) with, say, $s = t = 0.2$ measures a son's chances of remaining stuck in the lowest quintile if his father was in the lowest quintile. This probability seems to be of greater and more immediate policy appeal than the somewhat pedantic figure measuring the chances of the son being between the first and second quintile given that the father was between the third and the fourth quintile. Secondly, derivation of the properties of the estimates of (2) are notationally less messy. As we showed above, moving from one definition to the other is easy and does not involve any serious technical challenge. Also, it should be obvious how to refine the definition of $\theta(x; s, t)$ to make it, say, race-specific.

Parameters of interest based on the above definitions are:

(i) Conditional mobility $\theta_B(x; s, t) - \theta_W(x; s, t)$, measuring the black-white difference in transition at each value of x, where e.g.

$$\theta_B(x; s, t) = \frac{\Pr[Y_1 \leq \zeta_1, Y_0 \leq \zeta_0 | X = x, Black = 1]}{\Pr[Y_0 \leq \zeta_0 | Black = 1]}$$

and $\zeta_1(\zeta_0)$ is still the marginal quantile of the Y_1 (Y_0) distribution comprising both blacks and whites.

(ii) Marginal effects $\frac{\partial}{\partial x_j}\{\theta_B(x; s, t) - \theta_W(x; s, t)\}$, measures the black-white difference in the effect of increases in X (analogous to black-white difference in returns to education)

(iii) (Weighted) average derivatives based on (ii), i.e. $\delta^*_B(s, t) - \delta^*_W(s, t)$ where for $k = W, B$

$$\delta^*_k(s, t) = \int w_k(x) \frac{\partial}{\partial x_j} \theta_k(x; s, t) \, dF_k(x) \Big/ \int w_k(x) \, dF_k(x) \tag{4}$$

using weights $w_k(.)$ which are typically the marginal density of X and $F_W(.)$ denotes the c.d.f. of X for whites.

While (i) describes differences in mobility levels by values of x, averaged over other covariates, (ii) and (iii) shed light on the effectiveness of policy in changing existing differences by influencing values of X. As is well-known, derivatives typically converge slower than the original estimates which makes inference on (ii) very imprecise. We have a relatively smaller number of observations on blacks in our data and estimating the marginal effects for blacks is an important substantive parameter for us. In view of these facts, we concentrate on average derivative type estimates which summarize the black-white gap in the marginal effect of changing a continuous X. One would expect such an estimate to have a parametric rate of convergence (and thus enable more precise inference), which is what we rigorously show below. The technical innovation here is that the dependent variable involves estimated quantiles. In the case of averaged derivatives, the estimation error in the quantiles will affect the asymptotic distribution of the averaged marginal effects but in the case of conditional probabilities, the effect of this initial estimation error will be shown to be negligible.

2.2 Upward mobility

We now formally introduce our new measure of upward mobility. 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. We believe that this measure more closely conforms to what most people think of as economic mobility. Indeed much of the recent attention in the popular press concerning IGM has been couched in terms of the prospects for *upward* mobility for those starting in the bottom of the distribution.[8] Policymakers also tend to be more concerned about mobility with respect to what it signifies about the prospects for economic gains among the poor and disadvantaged groups and whether government interventions are necessary to foster greater upward mobility.

Let Y_0, Y_1 denote parent and son's income with respective marginal CDF's F_0 and F_1. Then for a fixed $s \in [0,1]$, we define upward mobility for families under percentile s by an extent $\tau \in [0, 1-s]$ as

$$\begin{aligned} \upsilon(\tau, s) &= \Pr\left(F_1(Y_1) - F_0(Y_0) > \tau | F_0(Y_0) \leq s\right) \\ &= 1 - \Pr\left(F_1(Y_1) - F_0(Y_0) < \tau | F_0(Y_0) \leq s\right) \\ &= 1 - \frac{\Pr\left(F_1(Y_1) - F_0(Y_0) < \tau, F_0(Y_0) \leq s\right)}{s} \\ &= 1 - \frac{1}{s}\int_1^{F_0^{-1}(s)}\int_1^{F_1^{-1}(F_0(y_0)+\tau)} f(y_0, y_1)\, dy_1 dy_0. \end{aligned} \tag{5}$$

Observe that $F_1(Y_1)$, $F_0(Y_0)$ are identically distributed uniform (0,1) variates and so the distribution of $F_1(Y_1) - F_0(Y_0)$ is symmetric around 0. This means that $\upsilon(0,1) = 0.5$ and $\upsilon(\tau, 1) = 1 - \upsilon(-\tau, 1)$, no matter what the joint distribution of (Y_0, Y_1) is. However, for an arbitrary $s \in (0,1)$, the value of $\upsilon(\tau, s)$ will depend on the joint distribution of

[8] For example, the Wall Street Journal began a front page article about class mobility as follows: "The notion that the U.S is a special place where any child can grow up to be president, a meritocracy where smarts and ambition matter more than parenthood and class, dates to Benjamin Franklin. ... The promise that a child born in poverty isn't trapped there remains a staple of America's self-portrait." (Wessel, 2005) Similarly, in a frontpage article in the New York Times on social mobility (Scott and Leonhardt, 2005) the public's beliefs concerning mobility are described in a poll asking "Is it possible to start out poor, work hard, and become rich?"

(Y_0, Y_1) for every τ, including $\tau = 1$. For the purpose of this paper, the leading case of interest is where $\tau = 0$ which gives the probability that the son's relative position exceeds that of the father. But we develop inference theory for a general τ and for a generic value of s.

Introducing covariates X into the analysis, define conditional upward mobility at values of $X = x$ as

$$\begin{aligned} \upsilon_c(\tau, s; x) &= \frac{\Pr(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s | X = x)}{\Pr(F_0(Y_0) \leq s)} \\ &= \frac{\Pr(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s | X = x)}{s}. \end{aligned} \tag{6}$$

This measure is analogous to (2) above. The idea is that we start with all families where the father was below the sth percentile. This ensures that all the corresponding sons have equal "space to move up". With these families constituting our population, we evaluate the extent of upward mobility for children at various values x of X. Denote the corresponding average derivative by

$$\delta_\upsilon(\tau, s) = E_X\left[f(X)\frac{\partial}{\partial x_j}\upsilon_c(\tau, s; X)\right] \tag{7}$$

which (when normalized by $E_X[f(X)]$) measures the average effect of the jth component of X on the probability of improving relative status.

Below, we will derive the statistical distribution theory for $\upsilon(\tau, s)$, $\upsilon_c(\tau, s; x)$ and $\delta_\upsilon(\tau, s)$. In the application, we will contrast overall upward mobility among blacks versus whites and then analyze how inclusion of relevant covariates affects this difference.

It is useful to note that 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: $\tilde{\Theta} = \left\{\tilde{\theta}(j,k)\right\}_{j,k=1,\ldots M}$. Then Shorrock's measure of *unconditional* mobility is given by

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

One can incorporate covariates into the above formula and define

$$M_1(x) = 1 - \frac{\sum_{j=1}^{K}\tilde{\theta}(j,j;x) - 1}{K-1} \tag{8}$$

where

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

and ζ_j, ξ_j denote the jth marginal quantiles of (Y_1, Y_0) respectively. Given the simple linear relation (8), inference on $M_1(x)$ will follow straightforwardly from inference on $\tilde{\theta}(j, j; x)$. However, this measure will depend crucially on the discretization employed which is clearly an undesirable feature. Altering the above formulas to allow for a continuous transition matrix seem complicated[9] and we leave that to future research. Instead, we focus on our measure $\upsilon(0, s)$, which, we believe, is much closer to what is commonly understood as mobility and whose enhancement appears to be a stated goal of liberal policy. This single summary measure does not employ any discretization, has an immediate intuitive interpretation and, unlike the IGE, is based on a direct comparison of the relative *positions* between fathers and son.

Another notable feature of $\upsilon(0, s)$ is that it counts small upward movements in relative positions which are ignored by transition probabilities. Comparing $\upsilon(0, s) = \Pr(F_1(Y_1) > F_0(Y_0) | F_0(Y_0) \leq s)$ and $1 - \theta(s, s) = \Pr[F_1(Y_1) > s | F_0(Y_0) \leq s]$, one can see that unlike $1 - \theta(s, s)$, $\upsilon(0, s)$ is counting all sons whose ranks exceeded their fathers' but did not exceed s. In our application, this makes a substantial impact on black-white differences in mobility (see figure 3 below). 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 difference between whites and blacks is much smaller when measured in terms of our upward mobility index. The first reason for this is that many black sons make relatively small upward movements which are missed by $\theta(s, s)$ but captured by $\upsilon(0, s)$. The second reason is that incomes of white fathers first order stochastic dominates that of black fathers (figure 4) everywhere. Therefore, for any overall percentile $F_0^{-1}(s)$, sons born to black fathers below $F_0^{-1}(s)$ need a larger increase in absolute income to surpass it compared to sons of white fathers below $F_0^{-1}(s)$. This suggests that even if rates of upward mobility are similar across groups, transition probabilities are likely to be much larger for whites.

3 Estimation and distribution theory

We now turn to estimation of the parameters outlined above and derivation of asymptotic properties of these estimates. Note that we have defined 6 parameters above, viz. (1),

[9] The problem is that $\int_0^1 \theta(s, s) \, ds$ is not a probability, unlike $\sum_{j=1}^{K} \tilde{\theta}(j, j)$ and one needs to replace $\theta(s, s)$ with a density type analog before integrating. What that analog should be is not obvious.

(2), (4), (5), (6), (7). FSZ (2004) had analyzed only (1) and so, in what follows, we will derive the distribution theory for the other five. The derivations for (2) and (4) involve nontrivial modifications of the respective methods for kernel-based conditional mean and Powell, Stock and Stoker's (1989) (PSS, henceforth) density-weighted average derivatives (dwad). The modifications are necessary because the dependent variable in the final estimates involve initial estimates of quantiles which are estimated from the same dataset. We will show below that this first-stage estimation has no effect on the asymptotic distribution of conditional mean estimates but has nontrivial effects for the dwad. This is intuitively natural because estimates of marginal quantiles converge at parametric rates just like the dwad measures but the conditional mean has a slower rate of convergence. However, lack of smoothness of the dependent variable in the quantiles makes the formal justifications nontrivial.

The estimates of (5), (6) and (7) also require an altogether different type of analysis from standard nonparametric regression theory owing to the presence of $\hat{F}_1(.)$ and $\hat{F}_0(.)$ in the definition of the dependent variables. Our derivations will rely crucially on the idea of Hadamard differentiability and will use Hoeffding's inequality to control the errors involved in the estimation of $\hat{F}_1(.)$ and $\hat{F}_0(.)$.

3.1 Conditional transition probability

Recall from (2), that the conditional transition probabilities and their estimates are given respectively by

$$\begin{aligned}\theta(x;s,t) &= \frac{\Pr\left[Y_1 \le \zeta_1, Y_0 \le \zeta_0 | X = x\right]}{\Pr\left[Y_0 \le \zeta_0\right]} \\ \hat{\theta}(x;s,t) &= \frac{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1\left(Y_{1i} \le \hat{\zeta}_1, Y_{0i} \le \hat{\zeta}_0\right)}{\left(\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)\right) \times \left(\frac{1}{n}\sum_{i=1}^n 1\left(Y_{0i} \le \hat{\zeta}_0\right)\right)},\end{aligned} \tag{9}$$

where $K(.)$ is a standard d-dimensional kernel and σ_n is a sequence of bandwidths. Also, let

$$\begin{aligned}\phi(x,\zeta_0,\zeta_1) &= \Pr\left\{Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0 | X_i = x\right\} = \int_1^{\zeta_1}\int_1^{\zeta_0} f(y_0, y_1|x)\, dy_0 dy_1 \\ \phi_1(x,\zeta_0,\zeta_1) &= \int_1^{\zeta_0} f(y_0,\zeta|x)\, dy_0,\ \phi_0(x,\zeta_0,\zeta_1) = \int_1^{\zeta_1} f(\zeta_0, y_1|x)\, dy_1.\end{aligned}$$

We now state the main theorem of this section which describes the asymptotic distribution for conditional (on covariates) transition matrices. The key point is that the first-stage estimation of (ζ_0, ζ_1) has no effect on that final distribution. In other words, the distribution of

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

and that of the infeasible estimator $\hat{\phi}(x, \zeta_0, \zeta_1)$ are identical. The intuition behind this result is that $\left(\hat{\zeta}_0, \hat{\zeta}_1\right)$ 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 $\sqrt{n}$-rate. The proof is nonstandard due to the facts that (i) $\hat{\phi}(x, \zeta_0, \zeta_1)$ is not smooth in (ζ_0, ζ_1) and (ii) the problem here is the reverse of standard semiparametric problems (c.f. Andrews (1994), section 3.4 and Pakes and Pollard (1989)) in that the initial estimator here, viz. $\left(\hat{\zeta}_0, \hat{\zeta}_1\right)$, is finite-dimensional and the ultimate parameter of interest, viz. $\phi(x, \zeta_0, \zeta_1)$ is infinite-dimensional, in contrast to the standard case where it is the other way round.

We state this and subsequent theorems 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.

Theorem 1 *Let X be continuously distributed with dimension d and the data (X_i, Y_{1i}, Y_{0i}) be i.i.d. Suppose that $K(.)$ and σ_n in (9) satisfy standard conditions for mean-zero asymptotic normality of conditional means (c.f. Pagan and Ullah (1999) theorem 3.5, 3.6). Assume further that for $X = x$, (Y_0, Y_1) admits a nonnegative joint density w.r.t. the Lebesgue measure everywhere on the joint support. Further, the function $\phi(x, ., .)$ is Lipschitz with respect to the Euclidean norm $\|.\|$:*

$$|\phi(x, \zeta_0, \zeta_1) - \phi(x, \tau_0, \tau_1)| \leq \|(\zeta_0, \zeta_1) - (\tau_0, \tau_1)\| \, \delta(x) \tag{10}$$

with $\delta(.)$ uniformly bounded on the support of X.[10] *Then one will have*

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

where $\sigma^2(x) = Var\left(1\left(Y_{1i} \leq \zeta_1^0, Y_{0i} \leq \zeta_0^0\right)|X\right) = \phi\left(x, \zeta_0^0, \zeta_1^0\right) \times \left(1 - \phi\left(x, \zeta_0^0, \zeta_1^0\right)\right).$

Proof. See appendix ■

[10] When conditional on $X = x$, the vector (Y_0, Y_1) is distributed with a density bounded above, uniformly in x, then this condition is automatically satisfied.

Returning to (9), by the usual delta method,

$$\left(n\sigma_n^d\right)^{1/2}\left\{\hat{\theta}_B\left(x;s,t\right)-\theta_B^0\left(x;s,t\right)\right\}=\left(n\sigma_n^d\right)^{1/2}\frac{\left(\hat{\phi}_B\left(x,\hat{\zeta}_0,\hat{\zeta}_1\right)-\phi_B\left(x,\zeta_0^0,\zeta_1^0\right)\right)}{\pi_B}+o_p\left(1\right)$$

where

$$\begin{aligned}\phi_B\left(x,\zeta_0,\zeta_1\right) &= \Pr\left\{Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0|X_i=x,B=1\right\},\\ \pi_B &= \Pr\left[Y_0\leq\zeta_0|B=1\right].\end{aligned}$$

This follows because the estimate of π_B will converge at parametric rates and will not affect the distribution of $\hat{\theta}_B\left(x;s,t\right)$. One thus gets that

$$\left(n\sigma_n^d\right)^{1/2}\left\{\hat{\theta}_B\left(x;s,t\right)-\theta_B^0\left(x;s,t\right)\right\}\xrightarrow{d}N\left(0,\frac{\sigma_B^2\left(x\right)}{\pi_B^2 f_B\left(x\right)}\int K^2\left(u\right)du\right)\tag{11}$$

where $\sigma_B^2\left(x\right)=\phi_B\left(x,\zeta_0^0,\zeta_1^0\right)\times\left(1-\phi_B\left(x,\zeta_0^0,\zeta_1^0\right)\right)$ and $f_B\left(x\right)$ is simply the density of X for blacks at x.

3.2 Distribution of the gap in levels

In this paper, our substantive interest is focused around the black-white gap in conditional transition probabilities. In that context, it is interesting to note that the estimates $\left(n\sigma_n^d\right)^{1/2}\left\{\hat{\theta}_B\left(x;s,t\right)-\theta_B^0\left(x;s,t\right)\right\}$ for blacks and $\left(n\sigma_n^d\right)^{1/2}\left\{\hat{\theta}_W\left(x;s,t\right)-\theta_W^0\left(x;s,t\right)\right\}$ for whites will be asymptotically uncorrelated. The intuition for this result is as follows. There are two ways in which these two quantities could be correlated- through (i) the common estimates $\hat{\zeta}_0,\hat{\zeta}_1$ which appear in the numerators of both $\hat{\theta}_B\left(x;s,t\right)$ and $\hat{\theta}_W\left(x;s,t\right)$ and (ii) because the denominators are respectively $\frac{\frac{1}{n}\sum_{i=1}^n B_i 1\left(Y_{0i}\leq\hat{\zeta}_0\right)}{\frac{1}{n}\sum_{i=1}^n B_i}$ and $\frac{\frac{1}{n}\sum_{i=1}^n W_i 1\left(Y_{0i}\leq\hat{\zeta}_0\right)}{\frac{1}{n}\sum_{i=1}^n W_i}$. Using the same logic as above, the preliminary estimates $\hat{\zeta}$ will have no effect on the asymptotic distributions and this removes one source of dependence between $\hat{\theta}_B\left(x;s,t\right)$ and $\hat{\theta}_W\left(x;s,t\right)$. Secondly, even $\frac{\frac{1}{n}\sum_{i=1}^n B_i 1\left(Y_{0i}\leq\zeta_0\right)}{\frac{1}{n}\sum_{i=1}^n B_i}$ and $\frac{\frac{1}{n}\sum_{i=1}^n W_i 1\left(Y_{0i}\leq\zeta_0\right)}{\frac{1}{n}\sum_{i=1}^n W_i}$ will be asymptotically uncorrelated since they are sample averages of the random variable $1\left(Y_0\leq\zeta_0\right)$ among different subgroups. To see this formally, let $y_i=1\left(Y_{0i}\leq\zeta_0\right)$ and let

$$\hat{g}_B=\frac{\frac{1}{n}\sum_{i=1}^n B_i y_i}{\frac{1}{n}\sum_{i=1}^n B_i}\equiv\frac{\bar{y}_B}{\bar{d}_B}\text{ and }g_B=\frac{E\left(\bar{y}_B\right)}{E\left(\bar{d}_B\right)}=\frac{\mu_B}{\pi_B}$$

Asymptotically,

$$\begin{aligned}\sqrt{n}\left(\hat{g}_B - g_B\right) &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(B_i y_i - \mu_B\right)}{\pi_B} - \frac{\mu_B}{\pi_B^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(B_i y_i - \mu_B\right) + o_p\left(1\right) \\ &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_i y_i}{\pi_B} - \frac{\mu_B}{\pi_B^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_i y_i + o_p\left(1\right) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_i}{\pi_B}\left(y_i - \frac{\mu_B}{\pi_B}\right) + o_p\left(1\right).\end{aligned}$$

So the asymptotic covariance is given by,

$$E\left[\frac{B_i}{\pi_B}\left(y_i - \frac{\mu_B}{\pi_B}\right)\frac{W_i}{\pi_W}\left(y_i - \frac{\mu_W}{\pi_W}\right)\right] - E\left[\frac{B_i}{\pi_B}\left(y_i - \frac{\mu_B}{\pi_B}\right)\right]E\left[\frac{W_i}{\pi_W}\left(y_i - \frac{\mu_W}{\pi_W}\right)\right] = 0$$

since $B_i W_i \equiv 0$ and

$$E\left[\frac{W_i}{\pi_W}\left(y_i - \frac{\mu_W}{\pi_W}\right)\right] = E\left[\frac{W_i y_i}{\pi_W} - \frac{\mu_W}{\pi_W^2}W_i\right] = E\left[\frac{W_i y_i}{\pi_W}\right] - \frac{\mu_W}{\pi_W} = 0.$$

Thus, we get that

$$\begin{aligned}&\left(n\sigma_n^d\right)^{1/2}\left\{\hat{\theta}_B\left(x;s,t\right) - \theta_B^0\left(x;s,t\right) - \hat{\theta}_W\left(x;s,t\right) + \theta_W^0\left(x;s,t\right)\right\} \\ &\xrightarrow{d} N\left(0, \left\{\frac{\sigma_B^2\left(x\right)}{\pi_B^2 f_B\left(x\right)} + \frac{\sigma_W^2\left(x\right)}{\pi_W^2 f_W\left(x\right)}\right\}\int K^2\left(u\right)du\right) \qquad (12)\end{aligned}$$

where $\sigma_W^2\left(x\right) = \phi_W\left(x,\zeta_0^0,\zeta_1^0\right) \times \left(1 - \phi_W\left(x,\zeta_0^0,\zeta_1^0\right)\right)$ and $f_W\left(x\right)$ is simply the density of X for whites at x. This result simplifies the calculation of asymptotic variances of the black-white gap in conditional transition probabilities since no covariance term needs to be calculated.

3.3 Density weighted average derivative (dwad) for transition probabilities

Consider the quantity

$$\delta^*\left(\zeta^0\right) = \int_{\mathcal{X}} \phi_j'\left(x,\zeta_0^0,\zeta_1^0\right) f^2\left(x\right) dx$$

which is the density weighted average derivative of $\phi\left(x,\zeta_0,\zeta_1\right)$ (defined above) w.r.t. the jth component of X, first introduced in econometrics by PSS (1989). Using a standard integration by parts formula and noting that $\phi\left(x,\zeta_0^0,\zeta_1^0\right)$, being a probability, is uniformly bounded in x, one gets that

$$\delta^*\left(\zeta^0\right) = -2\int_{\mathcal{X}} f_j'\left(x\right)\phi\left(x,\zeta_0^0,\zeta_1^0\right) f\left(x\right) dx = -2E\left\{1\left\{Y_1 \le \zeta_1, Y_0 \le \zeta_0\right\} f_j'\left(X\right)\right\}.$$

So let

$$\delta^*\left(\zeta^0\right) = -2E\left\{1\left\{Y_1 \le \zeta_1^0, Y_0 \le \zeta_0^0\right\} f_j'(X)\right\}$$

The natural estimate of this is $\hat{\delta}^*\left(\hat{\zeta}\right)$ where

$$\begin{aligned}
&\hat{\delta}^*\left(\hat{\zeta}\right) \\
= &-\frac{2}{n}\sum_{i=1}^{n} 1\left(Y_{1i} \le \hat{\zeta}_1, Y_{0i} \le \hat{\zeta}_0\right) \frac{1}{(n-1)\,\sigma_n^{d+1}} \sum_{l \neq i} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right) \\
= &-\frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} 1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} \frac{1}{\sigma_n^{d+1}} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right) \\
&-\frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} \left[\begin{array}{c} 1\left(Y_{1i} \le \hat{\zeta}_1, Y_{0i} \le \hat{\zeta}_0\right) \\ -1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} \end{array}\right] \frac{1}{\sigma_n^{d+1}} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right).
\end{aligned}$$

Thus, ignoring the factor 2 for now,

$$\begin{aligned}
&-\left(\hat{\delta}^*\left(\hat{\zeta}\right) - \delta^*\left(\zeta^0\right)\right) \\
= &\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} 1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} \frac{1}{\sigma_n^{d+1}} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right) - \delta^*\left(\zeta^0\right) \\
&+\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} \left[\begin{array}{c} 1\left(Y_{1i} \le \hat{\zeta}_1, Y_{0i} \le \hat{\zeta}_0\right) \\ -1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} \end{array}\right] \frac{1}{\sigma_n^{d+1}} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right). \qquad (13)
\end{aligned}$$

The first term in the previous display can be handled directly by the PSS method whence one gets that

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} 1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} \frac{1}{\sigma_n^{d+1}} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right) - \delta^*\left(\zeta^0\right)\right) \\
= &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\begin{array}{c} f(X_i) \frac{\partial}{\partial x_j}\phi\left(X_i, \zeta_0^0, \zeta_1^0\right) - \frac{\partial}{\partial x_j} f(X_i) \\ \times\left(1\left\{Y_{1i} \le \zeta_1^0, Y_{0i} \le \zeta_0^0\right\} - \phi\left(X_i, \zeta_0^0, \zeta_1^0\right)\right) \end{array}\right) + o_p(1) \\
\equiv &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_i + o_p(1). \qquad (14)
\end{aligned}$$

The following theorem describes the behavior of the second term T_n in (13) by expressing it as (asymptotically equivalent to) an empirical process, when scaled by $\sqrt{n}$.

Theorem 2 *Under assumptions of theorem 3.3 of PSS (1989), with their Y replaced by $1\left\{Y_1 \le \zeta_1^0, Y_0 \le \zeta_0^0\right\}$, and the conditions of theorem 1, we have that*

$$\sqrt{n}T_n = \bar{m}^{*\prime}(\zeta_0) \times \sqrt{n}\left(\hat{\zeta} - \zeta^0\right) + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\bar{m}^{*\prime}(\zeta_0) \times \varphi_i\right] + o_p(1), \qquad (15)$$

where

$$\bar{m}^{*\prime}(\zeta_0)\times\varphi_i = \frac{\partial}{\partial\zeta_0}\left[E\left\{\phi_j'\left(X,\zeta_0^0,\zeta_1^0\right)f(X)\right\}\right]\times\frac{s-1\left(Y_{0i}\leq\zeta_0^0\right)}{f_{Y_0}\left(\zeta_0^0\right)} + \frac{\partial}{\partial\zeta_1}\left[E\left\{\phi_j'\left(X,\zeta_0^0,\zeta_1^0\right)f(X)\right\}\right]\times\frac{t-1\left(Y_{1i}\leq\zeta_1^0\right)}{f_{Y_1}\left(\zeta_1^0\right)}.$$

Now (14), which follows from PSS theorem 3.3, plus (15) imply that

$$\begin{aligned}\sqrt{n}\left[\hat{\delta}^*\left(\hat{\zeta}\right)-\delta^*\left(\zeta^0\right)\right] &= \bar{m}^{*\prime}(\zeta_0)\times\sqrt{n}\left(\hat{\zeta}-\zeta^0\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^n\Psi_i+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\Psi_i+\bar{m}^{*\prime}(\zeta_0)\times\varphi_i\right\}+o_p(1).\end{aligned}$$

Proof. See appendix ■

Putting back the factor 2, finally, we have that

$$\sqrt{n}\left[\hat{\delta}^*\left(\zeta^0\right)-\delta^*\left(\hat{\zeta}\right)\right] = -\frac{1}{\sqrt{n}}\sum_{i=1}^n\gamma_i+o_p(1),$$

with

$$\begin{aligned}\gamma_i &= 2f(X_i)\frac{\partial}{\partial x_j}\phi\left(X_i,\zeta_0^0,\zeta_1^0\right)-2\frac{\partial}{\partial x_j}f(X_i)\times\left(1\left\{Y_{1i}\leq\zeta_1^0,Y_{0i}\leq\zeta_0^0\right\}-\phi\left(X_i,\zeta_0^0,\zeta_1^0\right)\right)\\ &\quad+2\frac{\partial}{\partial\zeta_0}\left[E\left\{\phi_j'\left(X,\zeta_0^0,\zeta_1^0\right)f(X)\right\}\right]\times\frac{s-1\left(Y_{0i}\leq\zeta_0^0\right)}{f_{Y_0}\left(\zeta_0^0\right)}\\ &\quad+2\frac{\partial}{\partial\zeta_1}\left[E\left\{\phi_j'\left(X,\zeta_0^0,\zeta_1^0\right)f(X)\right\}\right]\times\frac{t-1\left(Y_{1i}\leq\zeta_1^0\right)}{f_{Y_1}\left(\zeta_1^0\right)}.\end{aligned} \tag{16}$$

3.3.1 Normalization

To be able to interpret the weighted average derivative estimator, it is necessary to normalize its scale. We consider the natural normalization

$$\delta\left(\zeta^0\right)=\delta^*\left(\zeta^0\right)/E\left(f(X)\right).$$

Then the natural estimate of this is $\hat{\delta}\left(\hat{\zeta}\right) = \hat{\delta}^*\left(\hat{\zeta}\right)/\hat{E}\left(f\left(.\right)\right)$ which will satisfy

$$\begin{aligned}
& \sqrt{n}\left[\hat{\delta}\left(\hat{\zeta}\right) - \delta\left(\zeta^0\right)\right] \\
= & \frac{1}{E\left(f\left(X\right)\right)}\sqrt{n}\left[\hat{\delta}^*\left(\zeta^0\right) - \delta^*\left(\hat{\zeta}\right)\right] - \delta^2\left(\zeta^0\right)\sqrt{n}\left[\hat{E}\left(f\left(.\right)\right) - E\left(f\left(X\right)\right)\right] \\
& +\sqrt{n}\left[\hat{\delta}^*\left(\zeta^0\right) - \delta^*\left(\hat{\zeta}\right)\right] \times \left[\frac{1}{\hat{E}\left(f\left(.\right)\right)} - \frac{1}{E\left(f\left(X\right)\right)}\right] \\
& +\sqrt{n}\left[\hat{E}\left(f\left(.\right)\right) - E\left(f\left(X\right)\right)\right] \times \delta\left(\zeta^0\right) \times \left[\frac{1}{E\left(f\left(X\right)\right)} - \frac{1}{\hat{E}\left(f\left(.\right)\right)}\right] \\
= & \frac{1}{E\left(f\left(X\right)\right)}\sqrt{n}\left[\hat{\delta}^*\left(\zeta^0\right) - \delta^*\left(\hat{\zeta}\right)\right] - \delta^2\left(\zeta^0\right)\sqrt{n}\left[\hat{E}\left(f\left(.\right)\right) - E\left(f\left(X\right)\right)\right] + o_p\left(1\right)
\end{aligned}$$

provided $\hat{E}\left(f\left(.\right)\right) \xrightarrow{p} E\left(f\left(X\right)\right) >> 0$. Now, one can get an asymptotically linear form of $\sqrt{n}\left[\hat{E}\left(f\left(.\right)\right) - E\left(f\left(X\right)\right)\right]$ by using the derivation in PSS. To see this, assume $d = 2$ for notational simplicity. Note that

$$\begin{aligned}
E\left(f\left(X_1, X_2\right)\right) &= \int\int 1.f^2\left(x_1, x_2\right)dx_1dx_2 \\
&= -2\int\int x_1\frac{\partial f\left(x_1, x_2\right)}{\partial x_1}f\left(x_1, x_2\right)dx_1dx_2 \\
&= -2E\left(x_1\frac{\partial f\left(x_1, x_2\right)}{\partial x_1}\right).
\end{aligned}$$

This is exactly like the density-weighted average derivative estimator, except that the dependent variable is now x_1. Applying the PSS results (apply their their equation 3.1 and 3.17 with $y = g\left(x\right) = x_1$), it now follows that

$$\sqrt{n}\left[\hat{E}\left(f\left(.\right)\right) - E\left(f\left(X\right)\right)\right] = -\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(f\left(X_i\right) - E\left(f\left(X\right)\right)\right) + o_p\left(1\right).$$

Consequently,

$$\sqrt{n}\left[\hat{\delta}\left(\hat{\zeta}\right) - \delta\left(\zeta^0\right)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_i - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(f\left(X_i\right) - E\left(f\left(X\right)\right)\right) + o_p\left(1\right)$$

where γ_i is defined in (16). The asymptotic distribution of $\hat{\delta}\left(\hat{\zeta}\right)$ follows.

3.4 Marginal upward mobility

The natural estimate of $\upsilon\left(\tau, s\right)$ defined in (5) is given by

$$\hat{\upsilon}\left(\tau, s\right) = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} 1\left(\hat{F}_1\left(y_{1i}\right) \geq \hat{F}_0\left(y_{0i}\right) + \tau\right) 1\left(\hat{F}_0\left(y_{0i}\right) \leq s\right)}{s} \qquad (17)$$

where

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

We will now derive the asymptotic distribution of $\hat{v}(\tau, s)$. Let $F(.,.)$ denote the joint c.d.f. of (Y_0, Y_1) with corresponding joint density $f(.,.)$. Then for fixed s, τ, one may view $v(\tau, s)$ as a functional $v(F)$. We can therefore estimate it by $v\left(\hat{F}\right)$, where $\hat{F}$ denotes the usual empirical c.d.f.. We will obtain a large sample distribution of $v\left(\hat{F}\right)$ using the functional delta method by showing that the functional $F \mapsto v(F)$ is Hadamard-differentiable.

If one assumes that the joint density of (Y_0, Y_1) is bounded away from zero on a compact support, then the proof of Hadamard differentiability is considerably simpler. This assumption may be tenable if the population of interest excludes families with "abnormally" high and low earnings in either generation which is typically where the density will be close to zero. However, for the sake of greater generality, we dispense with this assumption and establish Hadamard differentiability under more general tail conditions on the joint density and its partial derivatives.

Since all our measures are robust to monotone transformation of the income variables, we will assume that the support of the income variables is contained in $[1, \infty)$.

We are now ready to state a formal theorem. 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 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.

Condition : **(Ai)** for some $\alpha > 1$, we have $f_1(x) \geq \frac{c}{x^\alpha}$ for x large enough which also implies that $F_1^{-1}(u) < c(1-u)^{\frac{1}{1-\alpha}}$, **(Aii)** $f^{(0)}\left(y_0, F_1^{-1}(F_0(y_0)+\tau)\right) \leq \frac{c}{y_0^{\alpha_0}}$ for some $\alpha_0 > 0$, **(Aiii)** for some $\varepsilon > 0$, $1 - F_0(y_0) > c y_0^{\frac{(1+\varepsilon-\alpha_0)(\alpha-1)}{\alpha}}$ and **(Aiv)**

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

Note that conditions A(i)-A(iii) concern the tail behavior of the densities and their derivatives evaluated at the marginal quantiles. They are more general than the more convenient but less realistic condition that the density is bounded away from zero uniformly on a compact support. Condition A(iv) is like a moment condition. Recall that for a positive random variable X with marginal c.d.f. $G(.)$ and support A, the quantity $\int_A (1 - G(x))\, dx$ equals $E(X)$.

It is interesting to note that if (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

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

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)-A(iv) are satisfied with $\alpha = \gamma + 1$, $\alpha_0 = \gamma + 2$ and $\varepsilon = 1 + \gamma + \gamma(\gamma + 1)$.

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

(Bi) for some $\beta > 1$, we have $f_0(x) \geq \frac{c}{x^{\beta}}$ for x large enough which also implies that $F_0^{-1}(u) < c(1-u)^{\frac{1}{1-\beta}}$, **(Bii)** $f^{(0)}\left(F_0^{-1}(s), y_1\right) \leq \frac{c}{y_1^{\beta_0}}$ for some $\beta_0 > 0$, **(Biii)** for some $\delta > 0$, we have $1 - F_1(y_1) > c y_1^{\frac{(1+\delta-\beta_0)(\beta-1)}{\beta}}$ and

$$\textbf{(Biv)} \int_1^{\infty} (1 - F_1(y_1))^{\frac{\beta}{\beta-1}} f\left(F_0^{-1}(s), y_1\right) dy_1 < \infty.$$

Theorem 3 *The map $F \mapsto \upsilon(F)$ from $\bar{D}[1,\infty) \to \mathbb{R}$, defined as*

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

for any $s \in (0,1)$ is Hadamard differentially tangentially to D_0 where $\bar{D}[1,\infty)$ is the space of cdf's satisfying conditions (Ai)-(Aiv) and (Bi)-(Biv). D_0 is the space of sample paths corresponding to the composite Brownian bridge $\mathbb{G}_{\lambda} \circ F$ where $\mathbb{G}_{\lambda}$ is a standard Brownian bridge and F is any c.d.f. in $\bar{D}[1,\infty)$. The derivative at $H = (H_0, H_1)$ is given by the linear functional $\upsilon'_F(.)$ defined as

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

Proof. in Appendix ■

Now it follows, via the functional delta method (c.f. van der Vaart and Wellner (1996), theorem 3.9.11), that bootstrapping will lead to consistent approximation of the distribution of the estimate of $\upsilon(\tau, s)$ and hence of $1 - \frac{1}{s}\upsilon(\tau, s)$.

3.5 Conditional upward mobility

Recall from (6) that conditional upward mobility is given by

$$\upsilon_c(\tau, s; x) = \frac{\Pr\left(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \le s | X = x\right)}{s}.$$

Ignoring the fixed s in the denominator, $\upsilon_c(\tau, s; x)$ is estimated by

$$\hat{\upsilon}_c(\tau, s; x) = \frac{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1\left(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \le s\right)}{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)},$$

where $\hat{F}_1(Y_{1i}) = \frac{1}{n-1}\sum_{j \neq i} 1(Y_{ij} \le Y_{1i})$ and $K(.)$ is a standard d-dimensional kernel function with a bandwidth sequence σ_n, satisfying the standard conditions for asymptotic normality of conditional means (with undersmoothing).Therefore,

$$\begin{aligned}
& \hat{\upsilon}_c(\tau, s; x) - \upsilon_c(\tau, s; x) \\
= & \left\{ \frac{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) 1\left(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \le s\right)}{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)} - \upsilon_c(\tau, s; x) \right\} \\
& + \frac{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right) \times \left[\begin{array}{c} 1\left(\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \le s\right) \\ -1\left(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \le s\right) \end{array}\right]}{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{x_i - x}{\sigma_n}\right)} \\
\equiv & \; T_{1n} + \frac{T_{2n}}{\hat{f}(x)}, \text{ say.}
\end{aligned} \tag{18}$$

We would like to show that $\frac{T_{2n}}{\hat{f}(x)}$ has a smaller order of magnitude than T_{1n}. This will imply that asymptotically, the distribution of $\hat{\upsilon}_c(\tau, s; x)$ will be that of a standard Nadaraya-Watson regression function of the unobserved random variable $1\left(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \le s\right)$ on X. The following proposition states this result formally and its proof appears in the appendix.

Proposition 1 *Let the data (X_i, Y_{1i}, Y_{0i}) be i.i.d. Let the kernel function $K(.)$ and the bandwidth sequence σ_n satisfy standard conditions for the asymptotic zero mean normality of the (infeasible) Nadaraya-Watson regression estimate of* $1\left(F_1(Y_{1i}) - F_0(Y_{0i}) > \tau, F_0(Y_{0i}) \le s\right)$

on X. Then, we have that

$$
\begin{aligned}
& \sqrt{n\sigma_n^d}\left(\hat{\upsilon}_c(\tau, s; x) - \upsilon_c(\tau, s; x)\right) \\
= & \sqrt{n\sigma_n^d}\left\{\frac{\hat{E}\left(1\left(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s\right)|X = x\right)}{s} - \upsilon_c(\tau, s; x)\right\} + o_p(1) \\
& \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{s^2 f(x)}\int K^2(u)\,du\right),
\end{aligned}
$$

where $\sigma^2(x) = s \times \upsilon_c(\tau, s; x) \times (1 - s \times \upsilon_c(\tau, s; x))$ and $f(x)$ is the marginal density of X.

3.6 Density Weighted Average Derivative (dwad) for upward mobility

Let F denote the joint distribution function of (X, Y_0, Y_1) and consider the dwad parameter based on (6)

$$
\begin{aligned}
\delta_\upsilon(\tau, s; F) &= E_X\left[f(X)\frac{\partial}{\partial x_j}\upsilon_c(\tau, s; X)\right] \\
&= -2E_{X,Y_0,Y_1}\left[1\left\{F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s\right\}\frac{\partial}{\partial x_j}f(x)\right] \\
&= -2E_X\left[\Pr\left\{F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s|X\right\}\frac{\partial}{\partial x_j}f(X)\right] \\
&= -2\int_{x \in Supp(X)}\left[\int_1^{F_{0|x}^{-1}(s)}\int_1^{F_{1|x}^{-1}\left(F_{0|x}(y_0)+\tau\right)} f(y_0, y_1|x)\,dy_1 dy_0\right] f(x)\frac{\partial}{\partial x_j}f(x)\,dx,
\end{aligned}
$$

where we have dropped the fixed s from the denominator for now. Its estimate is given by

$$
\begin{aligned}
& \hat{\delta}_\upsilon\left(\tau, s; \hat{F}\right) \\
= & -\frac{2}{n}\sum_{i=1}^{n} 1\left\{\hat{F}_1(Y_{1i}) - \hat{F}_0(Y_{0i}) > \tau, \hat{F}_0(Y_{0i}) \leq s\right\}\frac{1}{(n-1)\sigma_n^{d+1}}\sum_{l \neq i} K_j'\left(\frac{X_i - X_l}{\sigma_n}\right).
\end{aligned}
$$

The analysis of this estimator is exactly analogous to that of the dwad in section 3.3 above. The only difference is that ordinary differentiability with respect to ζ is now replaced by Hadamard differentiability with respect to the CDF's (F_0, F_1), as in theorem 3. The specific steps are as follows.

Observe that

$$
\begin{aligned}
& \sqrt{n}\left(\hat{\delta}_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right) \\
= & \sqrt{n}\left(\hat{\delta}_\upsilon\left(\tau,s;F^0\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right) \\
& +\sqrt{n}\left(\delta_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right) \\
& +\sqrt{n}\left[\left(\hat{\delta}_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;\hat{F}\right)\right)-\left(\left(\hat{\delta}_\upsilon\left(\tau,s;F^0\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right)\right)\right] \\
= & \sqrt{n}\left(\hat{\delta}_\upsilon\left(\tau,s;F^0\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right)+\sqrt{n}\left(\delta_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right) \\
& +\left(\upsilon_n\left(\hat{F}\right)-\upsilon_n\left(F^0\right)\right),
\end{aligned}
$$

where $\upsilon_n(F)=\sqrt{n}\left[\hat{\delta}_\upsilon(\tau,s;F)-\delta_\upsilon(\tau,s;F)\right]$. The first term in the above display can be analyzed exactly as in PSS (1989) to get an asymptotically normal law. Next consider the second term. Using steps analogous to those in the proof of theorem 3, one can show that

$$
\delta_\upsilon(\tau,s;F)=\int_{x\in Supp(X)}\left[\int_1^{F_{0|x}^{-1}(s)}\int_1^{F_{1|x}^{-1}\left(F_{0|x}(y_0)+\tau\right)} f(y_0,y_1|x)\,dy_1dy_0\right] f(x)\frac{\partial}{\partial x_j}f(x)\,dx,
$$

where $F_{0|x}, F_{1|x}$ denote the marginal cdf's of Y_0 and Y_1 and $f(.,.|x)$ is their joint density conditional on $X=x$, is Hadamard differentiable w.r.t. $F=(F_0,F_1)$. Since $\sqrt{n}\left(\hat{F}-F^0\right)$ has a Gaussian limit, the distribution of the second term in the previous display will be a mean-0 Gaussian. Finally, using steps analogous to theorem 2 one can analogously show that $\upsilon_n(.)$ is stochastically equicontinuous. The key adjustment in the proof is to note that for $F\neq G$, $\{\delta_\upsilon(\tau,s;F)-\delta_\upsilon(\tau,s;G)\}$ will be the same order of magnitude as $\|F-G\|$ which follows from the Hadamard differentiability of $\delta_\upsilon^*(\tau,s;F)$ in F, with a uniformly bounded derivative.

Putting all of this together, the asymptotic distribution of $\sqrt{n}\left(\hat{\delta}_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right)$ is the same as that of

$$
\sqrt{n}\left(\hat{\delta}_\upsilon\left(\tau,s;F^0\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right)+\sqrt{n}\left(\delta_\upsilon\left(\tau,s;\hat{F}\right)-\delta_\upsilon\left(\tau,s;F^0\right)\right)
$$

and thus mean 0 Gaussian.

4 Application

In this section we produce empirical estimates of IGM for black and white men using the two measures described in the previous sections: transition probabilities and upward mobility. For each measure we show three sets of results: unconditional estimates, estimates conditional on AFQT test scores, and density weighted average derivatives with respect to AFQT test scores. For data, we use a sample of 2766 individuals from the National Longitudinal Survey of Youth (NLSY) who were between the ages of 14 and 22 in 1979. We measure the average family income of these individuals when they were living at home with their parents in 1978, 1979 and 1980. We also measure their average annual earnings as adults in 1996, 1998, 2000 and 2002.[11]

4.1 Marginal probabilities

4.1.1 Upward transition probabilities

We begin by showing estimates for upward IGM using transition probabilities. These represent the probability that son's income, (Y_1) surpasses a given quantile conditional on parent income (Y_0) having been at or below the same quantile in the parent generation (i.e. $s = t$ in (1)). We also consider transition probabilities where the son must surpass the quantile by a certain amount, τ, to facilitate comparisons with the upward mobility estimator we introduce in this paper.

$$\theta_\tau(s) = \frac{\Pr[Y_1 > \zeta_1 + \tau, Y_0 \leq \zeta_0]}{\Pr[Y_0 \leq \zeta_0]}.$$

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 95 percent confidence intervals in Figure 1. 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

[11] Specifically, we use the time average over any years during the relevant time period in which data are available. This allows us to include individuals even if data is missing in some years. The time averaging also provides a better measure of permanent income in both generations. All income variables are deflated to 1978 dollars using the CPI-U.

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. We are only aware of one previous study, Hertz (2005) that has documented differences between blacks and whites in intergenerational transition probabilities. However, Hertz relied on PSID data where there is some concern about the representativeness of intergenerational samples to identify black-white differences.[12]

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 that are at or below the median and allow τ to be large (0.2 to 0.3) then we find that the interracial mobility gap begins to narrow to a smaller, but still significant, 10 percentage point difference.

4.1.2 Upward mobility

We now show an analogous set of estimates of our measure of upward mobility for whites and blacks and the white-black difference in Table 2. We now 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 along with confidence bands in Figure 2. As the figure makes clear, aside from those whose family income was at or below the fifth 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 3 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 quantiles. As discussed

[12] Lee and Solon (2006), for example raise concerns about the usefulness of the oversample of poorer households in the PSID. In addition, there has been significant attrition among black families in the nationally representative portion of the PSID since the sample began in 1968 (Solon, 1992).

in section 2, the fact that the white distribution of parent income lies to the right of blacks will make it more likely that whites will surpass the quantile thresholds more easily. This is illustrated in Figure 4 which plots the CDF's of the parent income distribution for both blacks and whites and shows that the white distribution stochastically dominates the black distribution. This implies that if blacks and whites below the threshold experienced equal sized percentile gains, then the transition probabilities would be higher for whites. However, in other results (not shown) we also find that the magnitude of the percentile gains for blacks are actually much lower than for whites. In any case, these results provide some additional descriptive facts that are useful to consider when discussing IGM differences between whites and blacks. So while it is true that blacks do experience rates of upward mobility that are only modestly lower than whites, the *extent* of this mobility is vastly lower.

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.

4.2 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.

Previous studies using the NLSY have taken advantage of the availability of scores on the Armed Forces Qualifying Test (AFQT) as measure of cognitive skills to identify this as a source of interracial inequality.[13] For example, Neal and Johnson (1996) have shown that the black-white wage gap among adults can largely be explained by pre-market skills as proxied by AFQT scores during adolescence. Similarly, Cameron and Heckman (2001) have shown that the sizable gap in college enrollment between whites

[13] All individuals in the NLSY were given the AFQT test in 1980 as part of the renorming of the test. Following Neal and Johnson (1996) we use the 1989 version of the percentile score. The U.S. military views the AFQT score as "a general measure of trainability and predictor of on-the-job performance". (http://www.defenselink.mil/prhome/poprep2002/chapter2/c2_recruiting.htm)

and blacks can largely be accounted for by AFQT scores. Cameron and Heckman view the AFQT score as capturing the cumulative effects of family background influences in making students academically prepared for college. Therefore, it would not be surprising if the IGM gap might also be accounted for by inclusion of AFQT scores. We produce estimates of upward transition probabilities and our measure of upward mobility for black and white men where we now include AFQT scores as a covariate.

4.2.1 Conditional Transition Probabilities

We estimate the effect of AFQT scores on upward transition probabilities separately by race by using Nadaraya-Watson regression. Our dependent variable is the probability of leaving the bottom quintile.[14] Figure 5 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 percentile of AFQT scores, the transition probability for whites is 0.65 and for blacks is 0.63, or a difference of just 3 percentage points. At the 5th percentile the gap is about 5 percentage points and at the 75th percentile 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 shape of the regression lines are also similar between blacks and whites for the bottom half of the distribution. At the very low end of the AFQT distribution an increase in the AFQT percentile leads to a sizable rise in the probability of exiting the bottom quintile for both blacks and whites. 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 is relatively little data for blacks in the upper end of the AFQT distribution as is evidenced by the wide confidence intervals.

This finding of similar conditional transition probabilities using AFQT scores can be contrasted with results using years of education. In figure 6, we do a similar exercise where we instead use the sons' years of completed schooling 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. For example among those with 10

[14] We used an Epanchnikov kernel and 0.8 times the sample size as the bandwidth. We experimented with the leave-one-out cross validation approach but found that it gave little useful guidance. Our main results are stable for a reasonably wide choice of bandwidths around this value.

years of schooling, the transition probability out of the bottom quintile for whites is 60 percent while for blacks it is just 36 percent. The difference of 24 percentage points is statistically significant. In similar exercises using measures of parent education (not shown) we find broadly similar results. Therefore, like Hertz (2005), we find that parent education cannot explain the black-white mobility gap. However, we 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 running a probit with AFQT as a covariate. This is particularly true for blacks. In figure 7 we compare the transition probability results for blacks with the results from simply using a probit. As the chart shows, moving from the fifth percentile of the black AFQT distribution to the median nearly doubles the transition probability of leaving the bottom quintile from 0.28 to 0.54 when using the non-parametric estimator. In contrast, the probit implies an increase of only 10 percentage points from 0.40 to 0.50.

4.2.2 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 below the bottom quintile and set $\tau = 0$. The results are shown in Figure 8. In this case the effects on the black-white gap are even more striking as the point estimates imply that upward mobility is actually higher for blacks than whites once we condition on AFQT scores. We also find a much flatter relationship between the AFQT score and upward mobility than we did with the transition probability. For example, for blacks moving from the 25th percentile of the AFQT distribution to the 50th percentile raises the probability of surpassing one parent's by just 6 percentage points, from 85 percent to 91. Similarly for whites the analogous gain is also 6 percentage points, going from 81 percent to 87 percent.

4.2.3 Discussion of 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 health endowments, parental investment, peer influences or school quality. Understanding the formation of the black-white skills gap has been, and will likely continue to be, an area of intense research activity. Our results suggest that whatever the underlying causes of the gap in cognitive skills it appears to translate into significant differences in IGM.

4.3 Density Weighted Average Derivatives

4.3.1 Transition probabilities

The density weighted average derivative (DWAD) is used as a summary measure of the effect of the sons' AFQT on his probability of leaving the bottom quintile. The DWAD gives us a single number which is a weighted average of the slope in a nonparametric regression of transition probability on AFQT scores. The weight on the slope at a specific value of AFQT is proportional to the density of AFQT at that point. The formula used to compute the average derivative for blacks is the sample counterpart of

$$\frac{-2E\left\{1\left\{Y_1 \leq \zeta_1, Y_0 \leq \zeta_0\right\} \frac{\partial}{\partial(AFQT)} f\left(AFQT|B=1\right)\right\}}{\Pr\left[Y_0 \leq \zeta_0 | B=1\right]},$$

where $\frac{\partial}{\partial(AFQT)} f\left(AFQT|B=1\right)$ represents the derivative of the marginal density of AFQT for blacks. The choice of bandwidth was guided by the discussion in Powell and Stoker (1996). In this case, it is proportional to $n^{-2/7}$ and the constant of proportionality was chosen so that the DWAD are consistent with slopes of the conditional transition probabilities (see also the discussion in Pagan and Ullah (1999), page 190).

In figure 9 we show the average derivative, together with the 95% confidence interval, corresponding to transition probabilities out of the tth quantile as t varies from 15 to 50 in increments of 5. For both whites and blacks the average derivative gradually falls as we increase t. We find for all the transition probabilities that the DWAD is larger for blacks than whites but the difference is not statistically significant. It is not so obvious a priori what one might expect to have found. On the one hand as Figure 5, shows, the slope of the conditional transition probability appears to be steeper for whites than for blacks through much of the distribution. On the other hand, the slope is steeper for blacks at the bottom of the AFQT distribution and since these observations receive much greater weight due to their higher density (see fig. 10), this would favor blacks having a higher DWAD. It is worth emphasizing that the difference between the DWAD for blacks and that for whites cannot be directly interpreted as the difference in return to AFQT. Rather, it is the difference in return to AFQT at the current levels of AFQT of the two groups. We are also cautious about interpreting these results too strongly for two reasons. First, we have found that the results for the DWAD are more sensitive to the choice of bandwidth than the conditional probabilities. Second the data on black AFQT scores are very thin once we get to the upper half of the AFQT distribution. The difference in the DWAD is consistent with the notion that it might be much more important to improve the mobility prospects for blacks at the low end of the AFQT distribution where the returns are clearly very high.

4.3.2 Upward mobility

We show the results for calculating the DWAD for our upward mobility measure in Figure 11. For this exercise, we again condition on parent income being below the tth percentile with t running from 15 to 50 in increments of 5 and set $\tau = 0$. As was the case with the transition probabilities we find that blacks appear to have a higher average derivative although again the differences are not statistically significant. Again we caution reading too much into this result given the sensitivity to bandwidth choice and the limited data on blacks in the upper half of the AFQT distribution.

5 Conclusion

In this paper, we have developed the analytic tools for investigating levels of IGM and effects of covariates on it, based on sample data. We have focused on nonparametric regression of transition probabilities and a new direct measure of upward mobility on continuously distributed covariates and average derivatives thereof. Available statistical techniques cannot be used to derive the sampling distribution theory of these estimates because the dependent variables here are nonsmooth functions of a separate set of initial estimates. Therefore, we have developed the relevant asymptotic distribution theory which allows us to investigate the difference in the nature and causes of intergenerational mobility across population subgroups using survey data.

Applying our techniques to micro data from the NLSY, we have demonstrated that most of the black-white difference in ability to rise out of the bottom quintile can be accounted for by differences in cognitive skills during adolescence as measured by the AFQT score.

Although our analytical methods are applied in the context of intergenerational mobility here, they are applicable to any problem involving nonparametric regression and average derivative estimation where the dependent variable involves nonsmooth functions of preliminary finite-dimensional estimates or estimated ranks.

References

[1] Andrews, D. W. K. (1994): Empirical Process Methods in Econometrics, in Handbook of Econometrics, vol 4, pp. 2247-94, Elsevier.

[2] Becker, Gary S. and Nigel Tomes (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87 (1979), 1153-1189.

[3] Bowles Samuel and Herbert Gintis (2002) The Inheritance of Inequality. *Journal of Economic Perspectives* 16:3-30

[4] Cameron, Stephen V. and James J. Heckman (2001), "The Dynamics of Educational Attainment for Black, Hispanic and White Males," *Journal of Political Economy* 109:3, 455-499.

[5] Corak, Miles (2006). "Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility." IZA Discussion Paper No. 1993.

[6] Formby, J., Smith W. J. & Zheng, B. (2004): Mobility measurement, transition matrices and statistical inference, *Journal of Econometrics*, vol. 120, pp. 181-205.

[7] Hertz, Tom, 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*. Ed. Samuel Bowles, Herbert Gintis and Melissa Osborne Groves. Princeton University Press.

[8] Kesavan, S. (1989): Topics in functional analysis and applications, New York, Wiley.

[9] Neal, Derek A. and William R. Johnson (1996): "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy* 104:5, 860-895.

[10] Pagan, A. and Aman Ullah (1999): Nonparametric Econometrics, Cambridge University Press.

[11] Pakes, A. and D. Pollard (1989): Simulations and the asymptotics of optimization estimators, *Econometrica*, vol. 57, pp. 1027-57.

[12] Powell, J. Stock, J and Stoker, T. (1989): Semiparametric Estimation of index coefficients, *Econometrica*, vol. 57, pp. 1403-30.

[13] Powell, J and Stoker, T. (1996): Optimal bandwidth choice for density weighted averages, *Journal of Econometrics*, vol. 75, pp. 391-416.

[14] Scott, Janny and David Leonhardt (2005): "Shadowy Lines that Still Divide," *New York Times*, May 15, 2005.

[15] Shorrocks (1978): The measurement of mobility, *Econometrica*, vol. 46, number 5, pp. 1013-1024.

[16] Solon, Gary (1992), "Intergenerational Income Mobility in the United States," *American Economic Review* 82(3), pp. 393-408.

[17] Wessel, David (2005), "As rich-poor gap widens in U.S., class mobility stalls," *Wall Street Journal*, May 13, 2005.

[18] Zheng, B. (2001): Statistical inference for poverty measures with relative poverty lines, *Journal of Econometrics*; 101(2), pages 337-56.

6 Appendix with Proofs

Proof of theorem 1:

Proof. Consider the expression

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

whose expectation is given by

$$\begin{aligned}
\bar{m}^*(\zeta, x) &= E\left\{\frac{1}{\sigma_n^d} K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0\right)\right\} \\
&= E_{X_i}\left(\frac{1}{\sigma_n^d} K\left(\frac{X_i - x}{\sigma_n}\right) E\left\{1\left(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0\right) | X_i\right\}\right) \\
&= E_{X_i}\left(\frac{1}{\sigma_n^d} K\left(\frac{X_i - x}{\sigma_n}\right) \phi\left(X_i, \zeta_0, \zeta_1\right)\right) \\
&= \int K(u) f(x + u\sigma_n) \phi(x + u\sigma_n, \zeta_0, \zeta_1) \, du \\
&= f(x) \phi(x, \zeta_0, \zeta_1) + O\left(\sigma_n^2\right).
\end{aligned}$$

So

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

where $\tilde{\zeta}_1$ denote value intermediate between $\hat{\zeta}_1$ and ζ_1^0 and similarly, $\tilde{\zeta}_0$. The numerator of our estimate (20) satisfies

$$\begin{aligned}
& \frac{1}{n\sigma_n^d}\sum_{i=1}^{n} K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0\right) - \phi\left(x, \zeta_0^0, \zeta_1^0\right) f(x) \\
= & \ \bar{m}\left(\hat{\zeta}, x\right) - \phi\left(x, \zeta_0^0, \zeta_1^0\right) f(x) \\
= & \ \bar{m}\left(\hat{\zeta}, x\right) - \bar{m}^*\left(\hat{\zeta}, x\right) + \bar{m}^*\left(\hat{\zeta}, x\right) - \bar{m}^*\left(\zeta^0, x\right) + \bar{m}^*\left(\zeta^0, x\right) - \phi\left(x, \zeta_0^0, \zeta_1^0\right) f(x) \\
= & \ \left[\bar{m}\left(\hat{\zeta}, x\right) - \bar{m}^*\left(\hat{\zeta}, x\right)\right] + f(x)\left[\begin{array}{c} \phi\left(x, \zeta_0^0, \zeta_1^0\right) + \phi_0\left(x, \tilde{\zeta}_0, \tilde{\zeta}_1\right)\left(\hat{\zeta}_0 - \zeta_0\right) \\ +\phi_1\left(x, \tilde{\zeta}_0, \tilde{\zeta}_1\right)\left(\hat{\zeta}_1 - \zeta_1\right) \end{array}\right] \\
& +O\left(\sigma_n^2\right) \\
= & \ \left\{\left[\bar{m}\left(\hat{\zeta}, x\right) - \bar{m}^*\left(\hat{\zeta}, x\right)\right] - \left[\bar{m}\left(\zeta_0, x\right) - \bar{m}^*\left(\zeta^0, x\right)\right]\right\} \\
& +\bar{m}\left(\zeta^0, x\right) - \bar{m}^*\left(\zeta^0, x\right) \\
& +f(x)\left[\phi_0\left(x, \tilde{\zeta}_0, \tilde{\zeta}_1\right)\left(\hat{\zeta}_0 - \zeta_0\right) + \phi_1\left(x, \tilde{\zeta}_0, \tilde{\zeta}_1\right)\left(\hat{\zeta}_1 - \zeta_1\right)\right] \\
& +O\left(\sigma_n^2\right).
\end{aligned}$$

The third term will have a parametric rate of convergence and therefore will be asymptotically negligible, relative to the second term. The second term is standard and will have the variance of standard conditional mean type estimators. The nonstandard term is the first one and we now demonstrate a stochastic equicontinuity property for this expression. Letting

$$\begin{aligned}
& v_n(\zeta, x) \\
= & \left(n\sigma_n^d\right)^{1/2}\left\{\bar{m}(\zeta, x) - \bar{m}^*(\zeta, x)\right\} \\
= & \frac{1}{\left(n\sigma_n^d\right)^{1/2}}\sum_{i=1}^{n}\left\{\begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0\right) \\ -E\left(K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0\right)\right) \end{array}\right\}
\end{aligned}$$

the the first term is

$$\begin{aligned}
& v_n\left(\hat{\zeta}, x\right) - v_n\left(\zeta^0, x\right) \\
= & \frac{1}{n\sigma_n^d}\sum_{i=1}^{n}\left\{\begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right)\left(1\left(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0\right) - 1\left(Y_{1i} \leq \zeta_1^0, Y_{0i} \leq \zeta_0^0\right)\right) \\ -E\left(K\left(\frac{X_i - x}{\sigma_n}\right)\left(1\left(Y_{1i} \leq \hat{\zeta}_1, Y_{0i} \leq \hat{\zeta}_0\right) - 1\left(Y_{1i} \leq \zeta_1^0, Y_{0i} \leq \zeta_0^0\right)\right)\right) \end{array}\right\}.
\end{aligned}$$

Let us define

$$\begin{aligned}
&\tilde{v}_n(\zeta, x) \\
=\ & \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ \begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right) E\left\{1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) | X_i\right\} \\ -E\left(K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right)\right) \end{array} \right\} \\
=\ & \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ \begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right) \phi\left(X_i, \zeta_0, \zeta_1\right) \\ -E\left(K\left(\frac{X_i - x}{\sigma_n}\right) 1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right)\right) \end{array} \right\},
\end{aligned}$$

so that

$$\begin{aligned}
&v_n(\zeta_0, \zeta_1, x) - \tilde{v}_n(\zeta_0, \zeta_1, x) \\
=\ & \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) - \phi\left(X_i, \zeta_0, \zeta_1\right)\right\} \right\}.
\end{aligned}$$

Then for any $\zeta_1 > \eta_1$ and $\zeta_0 > \eta_0$, and $\zeta = (\zeta_0, \zeta_1)$, $\eta = (\eta_0, \eta_1)$, we have that

$$\begin{aligned}
v_n(\zeta, x) - v_n(\eta, x) &= v_n(\zeta, x) - \tilde{v}_n(\zeta, x) + \tilde{v}_n(\eta, x) - v_n(\eta, x) \\
&\quad + \tilde{v}_n(\zeta, x) - \tilde{v}_n(\eta, x).
\end{aligned}$$

We will show that $v_n(\zeta, x)$ is stochastically equicontinuous by showing that $|v_n(\zeta, x) - v_n(\eta, x)|$ is asymptotically uniformly small in probability when $\|\zeta - \eta\|$ is small enough. This equivalence follows from Andrews (1994), equation (2.3). Continuing from the previous display, first observe that

$$\begin{aligned}
&E\left(v_n(\zeta, x) - \tilde{v}_n(\zeta, x) + \tilde{v}_n(\eta, x) - v_n(\eta, x)\right)^2 \\
=\ & E\left(\begin{array}{c} \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) - \phi\left(X_i, \zeta_0, \zeta_1\right)\right\} \right\} \\ -\frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{1\left(Y_{1i} \le \eta_1, Y_{0i} \le \eta_0\right) - \phi\left(X_i, \eta_0, \eta_1\right)\right\} \right\} \end{array} \right)^2 \\
=\ & E\left(\frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^{n} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{ \begin{array}{c} 1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) - 1\left(Y_{1i} \le \eta_1, Y_{0i} \le \eta_0\right) \\ -\phi\left(X_i, \zeta_0, \zeta_1\right) + \phi\left(X_i, \eta_0, \eta_1\right) \end{array} \right\} \right\} \right)^2 \\
=\ & \frac{1}{n\sigma_n^d} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[\begin{array}{c} \left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{ \begin{array}{c} 1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) - 1\left(Y_{1i} \le \eta_1, Y_{0i} \le \eta_0\right) \\ -\phi\left(X_i, \zeta_0, \zeta_1\right) + \phi\left(X_i, \eta_0, \eta_1\right) \end{array} \right\} \right\} \\ \times \left\{ K\left(\frac{X_j - x}{\sigma_n}\right) \left\{ \begin{array}{c} 1\left(Y_{1j} \le \zeta_1, Y_{0j} \le \zeta_0\right) - 1\left(Y_{1j} \le \eta_1, Y_{0j} \le \eta_0\right) \\ -\phi\left(X_j, \zeta_0, \zeta_1\right) + \phi\left(X_j, \eta_0, \eta_1\right) \end{array} \right\} \right\} \end{array} \right] \\
=\ & \frac{1}{n\sigma_n^d} \sum_{i=1}^{n} E\left\{ K\left(\frac{X_i - x}{\sigma_n}\right) \left\{ \begin{array}{c} 1\left(Y_{1i} \le \zeta_1, Y_{0i} \le \zeta_0\right) - 1\left(Y_{1i} \le \eta_1, Y_{0i} \le \eta_0\right) \\ -\phi\left(X_i, \zeta_0, \zeta_1\right) + \phi\left(X_i, \eta_0, \eta_1\right) \end{array} \right\} \right\}^2,
\end{aligned}$$

since all terms with $i \neq j$ have zero mean. Since $\zeta_1 > \eta_1$ and $\zeta_0 > \eta_0$, the random variable $1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0) - 1(Y_{1i} \leq \eta_1, Y_{0i} \leq \eta_0)$ has a Bernoulli distribution conditional on X_i with mean equal to $\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)$ and variance $[1 - \phi(X_i, \zeta_0, \zeta_1) + \phi(X_i, \eta_0, \eta_1)] \times [\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]$. Therefore, the previous display equals

$$
\begin{aligned}
& \frac{1}{\sigma_n^d} E\left\{K^2\left(\frac{X_i - x}{\sigma_n}\right)[1 - \phi(X_i, \zeta_0, \zeta_1) + \phi(X_i, \eta_0, \eta_1)] \times [\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]\right\} \\
\leq \; & \|(\zeta_0, \zeta_1) - (\eta_0, \eta_1)\| \int K^2(u)\, \delta(x + u\sigma_n)\, f(x + u\sigma_n)\, du
\end{aligned}
$$

which, for large enough n, will be bounded above by $c\|(\zeta_0, \zeta_1) - (\eta_0, \eta_1)\|$ for some constant c by the dominated convergence theorem. Further,

$$
\begin{aligned}
& \tilde{v}_n(\zeta, x) - \tilde{v}_n(\eta, x) \\
= \; & \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^n \left\{K\left(\frac{X_i - x}{\sigma_n}\right)\phi(X_i, \zeta_0, \zeta_1) - E\left(K\left(\frac{X_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \zeta_1, Y_{0i} \leq \zeta_0)\right)\right\} \\
& - \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^n \left\{K\left(\frac{X_i - x}{\sigma_n}\right)\phi(X_i, \eta_0, \eta_1) - E\left(K\left(\frac{X_i - x}{\sigma_n}\right) 1(Y_{1i} \leq \eta_1, Y_{0i} \leq \eta_0)\right)\right\} \\
= \; & \frac{1}{(n\sigma_n^d)^{1/2}} \sum_{i=1}^n \left\{\begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)] \\ -E\left\{K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]\right\} \end{array}\right\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
& E[\tilde{v}_n(\zeta, x) - \tilde{v}_n(\eta, x)]^2 \\
= \; & \frac{1}{n\sigma_n^d} \sum_{i=1}^n \sum_{j=1}^n E\left[\begin{array}{c} \left\{\begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)] \\ -E\left\{K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]\right\} \end{array}\right\} \\ \times \left\{\begin{array}{c} K\left(\frac{X_j - x}{\sigma_n}\right)[\phi(X_j, \zeta_0, \zeta_1) - \phi(X_j, \eta_0, \eta_1)] \\ -E\left\{K\left(\frac{X_j - x}{\sigma_n}\right)[\phi(X_j, \zeta_0, \zeta_1) - \phi(X_j, \eta_0, \eta_1)]\right\} \end{array}\right\} \end{array}\right] \\
= \; & \frac{1}{\sigma_n^d} E\left\{\begin{array}{c} K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)] \\ -E\left\{K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]\right\} \end{array}\right\}^2 \\
= \; & E\left\{\frac{1}{\sigma_n^d} K^2\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]^2\right\} \\
& -\sigma_n^d \left(\frac{1}{\sigma_n^d} E\left\{K\left(\frac{X_i - x}{\sigma_n}\right)[\phi(X_i, \zeta_0, \zeta_1) - \phi(X_i, \eta_0, \eta_1)]\right\}\right)^2.
\end{aligned}
$$

The second term will go to 0 and the first term, by the Lipschitz property, will be at most

$$\begin{aligned}
& \|(\zeta_0,\zeta_1)-(\eta_0,\eta_1)\|^2 E\left\{\frac{1}{\sigma_n^d}K^2\left(\frac{X_i-x}{\sigma_n}\right)\delta^2(X_i)\right\} \\
= & \|(\zeta_0,\zeta_1)-(\eta_0,\eta_1)\|^2\left\{\delta^2(x)f(x)\int K^2(u)\,du+O(\sigma_n)\right\} \\
= & O\left(\|(\zeta_0,\zeta_1)-(\eta_0,\eta_1)\|^2\right).
\end{aligned}$$

Putting all of the together, it follows that $|v_n(\zeta,x)-v_n(\eta,x)|$ is $O_p(\|(\zeta_0,\zeta_1)-(\eta_0,\eta_1)\|)$, whence the equicontinuity follows. Now, using the same steps as Andrews (1994) leading to his equation (3.8), we conclude that $v_n\left(\hat{\zeta},x\right)-v_n\left(\zeta^0,x\right)=o_p(1)$. Thus we get

$$\begin{aligned}
& \left(n\sigma_n^d\right)^{1/2}\left\{\frac{1}{n\sigma_n^d}\sum_{i=1}^n K\left(\frac{X_i-x}{\sigma_n}\right)1\left(Y_{1i}\le\hat{\zeta}_1,Y_{0i}\le\hat{\zeta}_0\right)-\phi\left(x,\zeta_0^0,\zeta_1^0\right)f(x)\right\} \\
= & \left(n\sigma_n^d\right)^{1/2}\left\{\bar{m}\left(\zeta^0,x\right)-\bar{m}^*\left(\zeta^0,x\right)\right\}+o_p(1) \\
= & \frac{1}{\left(n\sigma_n^d\right)^{1/2}}\sum_{i=1}^n\left\{\begin{array}{c} K\left(\frac{X_i-x}{\sigma_n}\right)1\left(Y_{1i}\le\zeta_1^0,Y_{0i}\le\zeta_0^0\right) \\ -E\left(K\left(\frac{X_i-x}{\sigma_n}\right)1\left(Y_{1i}\le\zeta_1^0,Y_{0i}\le\zeta_0^0\right)\right)\end{array}\right\}+o_p(1).
\end{aligned}$$

Thus we conclude that estimating the quantiles has no effect on the asymptotic distribution of the estimate of $\phi(x,\zeta_0,\zeta_1)$. Therefore using standard theory for conditional mean estimates (under undersmoothing), one would get that

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

where $\sigma^2(x)=Var\left(1\left(Y_{1i}\le\zeta_1^0,Y_{0i}\le\zeta_0^0\right)|X\right)=\phi\left(x,\zeta_0^0,\zeta_1^0\right)\times\left(1-\phi\left(x,\zeta_0^0,\zeta_1^0\right)\right)$. ■

Proof of theorem 2:

Proof. The second term T_n in (13) equals

$$\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{l\neq i}\left[\begin{array}{c}1\left(Y_{1i}\le\hat{\zeta}_1,Y_{0i}\le\hat{\zeta}_0\right) \\ -1\left\{Y_{1i}\le\zeta_1^0,Y_{0i}\le\zeta_0^0\right\}\end{array}\right]\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right). \tag{19}$$

Let

$$
\begin{aligned}
\bar{m}(\zeta) &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i} 1\left(Y_{1i}\leq\zeta_1, Y_{0i}\leq\zeta_0\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\
\bar{m}^*(\zeta) &= E\left[1\left(Y_{1i}\leq\zeta_1, Y_{0i}\leq\zeta_0\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right]\\
&= E\left[\Pr\left(Y_{1i}\leq\zeta_1, Y_{0i}\leq\zeta_0|X_i,X_l\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right]\\
&= E\left[\phi\left(X_i,\zeta_0,\zeta_1\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right]\\
&= E_{X_l}\left\{E_{X_i|X_l}\left[\phi\left(X_i,\zeta_0,\zeta_1\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)|X_l\right]\right\}\\
&= \int\int\phi\left(x+\sigma_n u,\zeta_0,\zeta_1\right)K_j'(u)\frac{1}{\sigma_n}f\left(x+\sigma_n u\right)du f(x)\,dx\\
&= \int\frac{1}{\sigma_n}\left[\phi\left(x+\sigma_n u,\zeta_0,\zeta_1\right)K(u)|_1^{\infty}\right]f\left(x+\sigma_n u\right)dx\\
&\quad -\int\int K(u)\,\phi_j'\left(x+\sigma_n u,\zeta_0,\zeta_1\right)du f\left(x+\sigma_n u\right)f(x)\,dx\\
&= -\int\phi_j'\left(x,\zeta_0,\zeta_1\right)f^2(x)\,dx+O\left(\sigma_n^2\right)\\
&= -E\left\{\phi_j'\left(X,\zeta_0,\zeta_1\right)f(X)\right\}+O\left(\sigma_n^2\right).
\end{aligned}
$$

Then $\sqrt{n}$ times (19) equals

$$
\begin{aligned}
&\sqrt{n}\left(\bar{m}\left(\hat{\zeta}\right)-\bar{m}\left(\zeta^0\right)\right)\\
=\;&\sqrt{n}\left(\bar{m}^*\left(\hat{\zeta}\right)-\bar{m}^*\left(\zeta^0\right)\right)+\sqrt{n}\left\{\left[\bar{m}\left(\hat{\zeta}\right)-\bar{m}^*\left(\hat{\zeta}\right)\right]-\left[\bar{m}\left(\zeta^0\right)-\bar{m}^*\left(\zeta^0\right)\right]\right\}\\
=\;&\bar{m}^{*\prime}\left(\tilde{\zeta}\right)\times\sqrt{n}\left(\hat{\zeta}-\zeta^0\right)+\sqrt{n}\left\{\left[\bar{m}\left(\hat{\zeta}\right)-\bar{m}^*\left(\hat{\zeta}\right)\right]-\left[\bar{m}\left(\zeta^0\right)-\bar{m}^*\left(\zeta^0\right)\right]\right\}\\
=\;&\bar{m}^{*\prime}\left(\tilde{\zeta}\right)\times\sqrt{n}\left(\hat{\zeta}-\zeta^0\right)+\left(v_n\left(\hat{\zeta}\right)-v_n\left(\zeta_0\right)\right)
\end{aligned}
$$

where $v_n(\zeta)=\sqrt{n}\left[\bar{m}(\zeta)-\bar{m}^*(\zeta)\right]$ and $\tilde{\zeta}$ lies on the line joining $\hat{\zeta}$ and ζ^0. If we can show that $v_n(\zeta)$ is stochastically equicontinuous, then we are done.

Now we want to show the stoch equicontinuity property of v_n. Let us define

$$
\begin{aligned}
&\tilde{v}_n(\zeta)\\
=\;&\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\left\{\begin{array}{c}\Pr\left\{Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0|X_i,X_l\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\ -E\left(\Pr\left\{Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0|X_i,X_l\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right)\end{array}\right\}\\
=\;&\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\left\{\begin{array}{c}\phi\left(X_i,\zeta_0,\zeta_1\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\ -E\left(\phi\left(X_i,\zeta_0,\zeta_1\right)\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right)\end{array}\right\}
\end{aligned}
$$

so that

$$\begin{aligned}
& v_n(\zeta_0,\zeta_1)-\tilde{v}_n(\zeta_0,\zeta_1) \\
= \; & \frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\left\{\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\left\{1(Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0)-\phi(X_i,\zeta_0,\zeta_1)\right\}\right\}.
\end{aligned}$$

Then for any $\zeta_1>\eta_1$ and $\zeta_0>\eta_0$, and $\zeta=(\zeta_0,\zeta_1)$, $\eta=(\eta_0,\eta_1)$, we have that

$$\begin{aligned}
v_n(\zeta)-v_n(\eta) &= v_n(\zeta)-\tilde{v}_n(\zeta)+\tilde{v}_n(\eta)-v_n(\eta) \\
&\quad+\tilde{v}_n(\zeta,x)-\tilde{v}_n(\eta,x).
\end{aligned}$$

We will show that $v_n(\zeta)$ is stochastically equicontinuous by showing that $|v_n(\zeta)-v_n(\eta)|$ is asymptotically small in probability when $\|\zeta-\eta\|$ is small enough. Now,

$$\begin{aligned}
& E\left(v_n(\zeta)-\tilde{v}_n(\zeta)+\tilde{v}_n(\eta)-v_n(\eta)\right)^2 \\
= \; & \frac{1}{n(n-1)^2}E\left(\begin{array}{c}\sum_{i=1}^{n}\sum_{l\neq i}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\ \times\left[\begin{array}{c}\{1(Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0)-\phi(X_i,\zeta_0,\zeta_1)\}\\ -\{1(Y_{1i}\leq\eta_1,Y_{0i}\leq\eta_0)-\phi(X_i,\eta_0,\eta_1)\}\end{array}\right]\end{array}\right)^2 \\
= \; & \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{l\neq i}\sum_{i'=1}^{n}\sum_{l'\neq i'}E\left(\begin{array}{c}\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_{i'}-X_{l'}}{\sigma_n}\right)\\ \left[\begin{array}{c}\{1(Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0)-\phi(X_i,\zeta_0,\zeta_1)\}\\ -\{1(Y_{1i}\leq\eta_1,Y_{0i}\leq\eta_0)-\phi(X_i,\eta_0,\eta_1)\}\end{array}\right]\\ \left[\begin{array}{c}\{1(Y_{1i'}\leq\zeta_1,Y_{0i'}\leq\zeta_0)-\phi(X_{i'},\zeta_0,\zeta_1)\}\\ -\{1(Y_{1i'}\leq\eta_1,Y_{0i'}\leq\eta_0)-\phi(X_{i'},\eta_0,\eta_1)\}\end{array}\right]\end{array}\right) \\
= \; & \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{l\neq i}\sum_{l'\neq i}E\left(\begin{array}{c}\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\\ \left[\begin{array}{c}\{1(Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0)-\phi(X_i,\zeta_0,\zeta_1)\}\\ -\{1(Y_{1i}\leq\eta_1,Y_{0i}\leq\eta_0)-\phi(X_i,\eta_0,\eta_1)\}\end{array}\right]^2\end{array}\right) \\
\stackrel{(1)}{=} \; & \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{l\neq i}\sum_{l'\neq i}E\left(\begin{array}{c}\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\\ \left[1-\phi(X_i,\zeta_0,\zeta_1)+\phi(X_i,\eta_0,\eta_1)\right]\\ \times\left[\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right]\end{array}\right) \\
\leq \; & \|\zeta-\eta\|\,E_{X_i}\left\{E\left(\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)|X_i\right)\right\} \\
= \; & \|\zeta-\eta\|\times\left\{O(1)+O\left(\sigma_n^2\right)\right\}=O(\|\zeta-\eta\|).
\end{aligned}$$

where equality $\stackrel{(1)}{=}$ follows from the fact that since $\zeta_1>\eta_1$ and $\zeta_0>\eta_0$, the random variable $1(Y_{1i}\leq\zeta_1,Y_{0i}\leq\zeta_0)-1(Y_{1i}\leq\eta_1,Y_{0i}\leq\eta_0)$ has a Bernoulli distribution conditional on X_i

with mean equal to $\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)$ and variance $[1-\phi(X_i,\zeta_0,\zeta_1)+\phi(X_i,\eta_0,\eta_1)]\times[\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)]$.

Next,

$$
\begin{aligned}
& E\left[\tilde{v}_n(\zeta)-\tilde{v}_n(\eta)\right]^2 \\
= \; & \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{l\neq i}\sum_{i'=1}^{n}\sum_{l'\neq i'} E\left[\begin{array}{c}\left[\begin{array}{c}\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\ -E\left\{\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\}\end{array}\right]\\ \times\left[\begin{array}{c}\left\{\phi(X_{i'},\zeta_0,\zeta_1)-\phi(X_{i'},\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_{i'}-X_{l'}}{\sigma_n}\right)\\ -E\left\{\left\{\phi(X_{i'},\zeta_0,\zeta_1)-\phi(X_{i'},\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_{i'}-X_{l'}}{\sigma_n}\right)\right\}\end{array}\right]\end{array}\right]
\end{aligned}
$$

The terms with $i\neq i'$, $l\neq l'$ will have expectation 0. Terms with $i=i'$ and $l=l'$ will be

$$
\begin{aligned}
& \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{l\neq i} E\left[\begin{array}{c}\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\\ -E\left\{\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\}\end{array}\right]^2 \\
= \; & \frac{1}{(n-1)}E\left[\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right]^2 \\
& -\frac{1}{(n-1)}\left(E\left\{\left\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\right\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\}\right)^2
\end{aligned}
$$

The first term is at most

$$
\begin{aligned}
& \|\zeta-\eta\|^2\frac{1}{(n-1)}E_{X_l}\left(E\left[\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right]^2|X_l\right) \\
= \; & \|\zeta-\eta\|^2\frac{1}{(n-1)}\int\int\frac{1}{\sigma_n^{2(d+1)}}\left\{K_j'\left(\frac{x-y}{\sigma_n}\right)\right\}^2 f(x)f(y)\,dydx \\
= \; & \|\zeta-\eta\|^2\frac{1}{(n-1)}\int\int\frac{1}{\sigma_n^{2+d}}\left\{K_j'(u)\right\}^2 f(y+u\sigma_n)f(y)\,dudy \\
= \; & \|\zeta-\eta\|^2\times O\left(\frac{1}{n\sigma_n^{d+2}}\right)
\end{aligned}
$$

So if $n\sigma_n^{d+2}\to\infty$, which is an assumption in PSS (theorem 3.3), then first term is at most $o\left(\|\zeta-\eta\|^2\right)$. The second term is bounded by the first term by Cauchy-Schwartz and so is at most $o\left(\|\zeta-\eta\|^2\right)$.

Terms with $i = i'$ and $l \neq l'$ will be the same order as

$$\begin{aligned} & E\left[\begin{array}{c} \left[\begin{array}{c} \{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right) \\ -E\left\{\{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\} \end{array}\right] \\ \times\left[\begin{array}{c} \{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right) \\ -E\left\{\{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\right\} \end{array}\right] \end{array}\right] \\ = & E\left[\{\phi(X_i,\zeta)-\phi(X_i,\eta)\}^2\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\right] \\ & -E\left\{\{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\} \\ & \times E\left\{\{\phi(X_i,\zeta)-\phi(X_i,\eta)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\right\}. \end{aligned}$$

The first term will be $O\left(\|\zeta-\eta\|^2\right)$ because

$$\begin{aligned} & E\left[\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_i-X_{l'}}{\sigma_n}\right)\right] \\ = & \frac{1}{\sigma_n^2}\int\left[\int\int K_j'(x+\sigma_n u)\,K_j'(x+\sigma_n v)\,f(x+\sigma_n u)\,f(x+\sigma_n v)\,dudv\right]f(x)\,dx \\ = & -\int\left[\int\int f'(x+\sigma_n u)\,f'(x+\sigma_n v)\,K(u)\,K(v)\,dudv\right]f(x)\,dx \\ < & \ \infty. \end{aligned}$$

The second term will also be at most $O\left(\|\zeta-\eta\|^2\right)$ by Cauchy-Schwartz.

Terms with $i \neq i'$ but $l = l'$ give

$$\begin{aligned} & E\left[\begin{array}{c} \{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\}\{\phi(X_{i'},\zeta_0,\zeta_1)-\phi(X_{i'},\eta_0,\eta_1)\} \\ \times\frac{1}{\sigma_n^{2(d+1)}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)K_j'\left(\frac{X_{i'}-X_l}{\sigma_n}\right) \end{array}\right] \\ & -\left[\begin{array}{c} E\left\{\{\phi(X_{i'},\zeta_0,\zeta_1)-\phi(X_{i'},\eta_0,\eta_1)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_{i'}-X_l}{\sigma_n}\right)\right\} \\ \times E\left\{\{\phi(X_i,\zeta_0,\zeta_1)-\phi(X_i,\eta_0,\eta_1)\}\frac{1}{\sigma_n^{d+1}}K_j'\left(\frac{X_i-X_l}{\sigma_n}\right)\right\} \end{array}\right] \end{aligned}$$

which is $O\left(\|\zeta-\eta\|^2\right)$ by an analogous argument. This establishes that $v_n(\zeta)$ is stochastically equicontinuous. It follows that

$$\sqrt{n}T_n = \bar{m}^{*\prime}(\zeta_0)\times\sqrt{n}\left(\hat{\zeta}-\zeta^0\right)+o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{m}^{*\prime}(\zeta_0)\times\varphi_i\right]+o_p(1).$$

■

Proof of theorem 3:

Condition : (Ai) for some $\alpha > 1$, we have $f_1(x) \geq \frac{c}{x^{\alpha}}$ for x large enough which also implies that $F_1^{-1}(u) < c(1-u)^{\frac{1}{1-\alpha}}$, **(Aii)** $f^{(1)}\left(y_0, F_1^{-1}(F_0(y_0)+\tau)\right) \leq \frac{c'}{y_0^{\alpha_0}}$ for some $\alpha_0 > 0$, **(Aiii)** For some $\varepsilon > 0$, $1 - F_0(y_0) > \frac{c}{y_0^{\frac{2(1+\varepsilon-\alpha_0)\alpha}{\alpha-1}}}$ and **(Aiv)** $\int_1^{\infty}(1-F_0(y_0))^{\frac{\alpha}{\alpha-1}} f\left(y_0, F_1^{-1}(F_0(y_0)+\tau)\right) d$ ∞.

(Bi) for some $\beta > 1$, we have $f_0(x) \geq \frac{c}{x^{\beta}}$ for x large enough which also implies that $F_0^{-1}(u) < c(1-u)^{\frac{1}{1-\beta}}$, **(Bii)** $f^{(0)}\left(F_0^{-1}(s), y_1\right) \leq \frac{c'}{y_1^{\beta_0}}$ for some $\beta_0 > 0$, **(Biii)** for some $\delta > 0$, $1 - F_1(y_1) > \frac{c}{y_0^{\frac{2(1+\delta-\beta_0)\beta}{\beta-1}}}$ and **(Biv)** $\int_1^{\infty}(1-F_1(y_1))^{\frac{\beta}{\beta-1}} f\left(F_0^{-1}(s), y_1\right) dy_1 < \infty$.

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 $F_{1t}(y_0) = F_1(y_1) + tH_{1t}(y_{1t})$ denoting the corresponding marginals. Let $H_t \to H$ uniformly as $t \to 0$ and let H_0 and H_1 denote its marginals. We want to show that for a linear functional $\upsilon_F'(.)$,

$$\left|\frac{\upsilon(F_t) - \upsilon(F)}{t} - \upsilon_F'(H)\right| \to 0 \text{ as } t \to 0. \tag{20}$$

Define

$$\begin{aligned} z_1(y_0) &= F_1^{-1}(F_0(y_0)+\tau),\ z_{1t}(y_0) = F_{1t}^{-1}(F_{0t}(y_0)+\tau) \\ z_0 &= F_0^{-1}(s),\ \ z_{0t} = F_{0t}^{-1}(s). \end{aligned}$$

So we need to show

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

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

$$\int_1^{z_{0t}} \frac{z_{1t}(y_0)-z_1(y_0)}{t} f(\bar{z}_{1t}(y_0), y_0)\, dy_0 + \frac{z_{0t}-z_0}{t} \times \int_1^{z_1(y_0)} f(\bar{z}_{0t}, y_1)\, dy_1$$

$$+ \int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH_t(y_0, y_1)$$

$$= \underbrace{\left[\begin{array}{c} \underbrace{\int_1^{z_{0t}} \frac{z_{1t}(y_0)-z_1(y_0)}{t} \left[f(y_0, \bar{z}_{1t}(y_0)) - f(y_0, z_1(y_0)) \right] dy_0}_{T_{10t}} \\ + \underbrace{\frac{z_{0t}-z_0}{t} \times \int_1^{z_1(y_0)} \left[f(\bar{z}_{0t}, y_1) - f(z_0, y_1) \right] dy_1}_{T_{11t}} \end{array} \right]}_{T_{1t}}$$

$$+ \underbrace{\int_1^{z_{0t}} \frac{z_{1t}(y_0)-z_1(y_0)}{t} f(z_0, y_1)\, dy_1}_{T_{2t}}$$

$$+ \underbrace{\frac{z_{0t}-z_0}{t} \times \int_1^{z_1(y_0)} f(z_0, y_1)\, dy_1}_{T_{3t}}$$

$$+ \underbrace{\int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH_t(y_0, y_1) - \int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH(y_0, y_1)}_{T_{4t}}$$

$$+ \underbrace{\int_1^{z_{0t}} \int_1^{z_{1t}(y_0)} dH(y_0, y_1) - \int_1^{z_0} \int_1^{z_1(y_0)} dH(y_0, y_1)}_{T_{5t}}$$

$$+ \underbrace{\int_1^{z_0} \int_1^{z_1(y_0)} dH(y_0, y_1).}_{T_6}$$

We will show that as $t \to 0$,

Step 1: $|T_{1t}| \to 0$

Step 2:

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

Step 3:

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

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

Then we will have shown (20) with

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

which is linear in H.

For steps 1 and 2, we will need the following 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, $\tilde{z}_t(y_0)$ lies in between $z(y_0)$ and $z_t(y_0)$. Therefore,

$$\frac{z_{1t}(y_0) - z_1(y_0)}{t} = \frac{H_{0t}(y_0) - H_{1t}(z_t(y_0))}{f_1(\tilde{z}_{1t}(y_0))}. \tag{21}$$

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

$$\frac{z_{0t} - z_0}{t} = \frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})}. \tag{22}$$

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

Step 1: By a mean-value theorem argument,

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

where $f^{(1)}(.,.)$ denotes derivative w.r.t. the second argument and $\breve{z}_{1t}(y_0)$ lies in between $z(y_0)$ and $z_t(y_0)$. Using (21), we get

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

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

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

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 $H_0(.)$ and $H_0(.)$ are uniformly bounded.

Next,

$$\begin{aligned} & z_{1t}(y_0) \\ = \; & F_{1t}^{-1}(F_{0t}(y_0) + \tau) \overset{(1)}{<} c(1 - F_{0t}(y_0) - \tau)^{\frac{1}{1-\alpha}} = c(1 - F_0(y_0) - tH_{0t}(y_0) - \tau)^{\frac{1}{1-\alpha}} \\ \leq \; & c'(1 - F_0(y_0) - \tau)^{\frac{1}{1-\alpha}} \\ \leq \; & c(1 - F_0(y_0))^{\frac{1}{1-\alpha}} \end{aligned} \tag{23}$$

for small enough t, since $\alpha > 1$ and $tH_{0t}(.)$ converges uniformly to 0. Inequality (1) comes from condition **Ai**. Similarly,

$$z_1(y_0) \leq c(1 - F_0(y_0))^{\frac{1}{1-\alpha}} \tag{24}$$

and therefore (ii) follows. Finally (iii) follows from (23), (24) and condition Aii.

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

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

for t small enough and some $\delta > 0$. Inequality (2) follows from conditions Bi, Bii, Biii using arguments analogous to those for T_{10t}. This implies that $|T_{11t}| \to 0$.

Step 2:

$$
\begin{aligned}
& \left|T_{2t} - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1)\, dy_1\right| \\
= & \left|\int_1^{z_{0t}} \frac{z_{1t}(y_0) - z_1(y_0)}{t} f(z_0, y_1)\, dy_0 - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1)\, dy_1\right| \\
= & \left|\int_1^{z_{0t}} \frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})} f(z_0, y_1)\, dy_1 - \frac{H_0(z_0)}{f_0(z_0)} \int_1^{z_1(y_0)} f(z_0, y_1)\, dy_1\right| \\
\leq & \int_1^{z_{0t}} \left|\frac{H_{0t}(z_{0t})}{f_0(\tilde{z}_{0t})} - \frac{H_0(z_0)}{f_0(z_0)}\right| f(z_0, y_1)\, dy_1 \\
& \overset{(1)}{\leq} c \int_1^{z_{0t}} \left|H_{0t}(z_{0t}) - H_0(z_0)\right| (1 - F_1(y_1))^{\frac{\beta}{\beta-1}} f(z_0, y_1)\, dy_1 \\
\leq & c \left(\sup_u \left|H_{0t}(u) - H_0(u)\right| + \left|H_0(z_{0t}) - H_0(z_0)\right|\right) \int_1^{\infty} (1 - F_1(y_1))^{\frac{\beta}{\beta-1}} f(z_0, y_1)\, dy_1 \\
\to & \; 0
\end{aligned}
$$

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

Step 3:

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

which goes to zero if $\int_1^{\infty} (1 - F_0(y_0))^{\frac{\alpha}{\alpha-1}} f(z(y_0), y_0)\, dy_0 < \infty$, which is condition (**Aiv**). Note that the inequality $\overset{(0)}{\leq}$ 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.

Step 4:

$T_{4t} \to 0$ since $H_t \to H$ 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. ■

Proof of proposition 1:

Proof. Recall display (18) and note that we want to show that T_{2n} is of smaller order of magnitude than T_{1n}. We will show that $E\left(\sqrt{n\sigma_n^d}\left|T_{2n}\right|\right) \to 0$ which will imply that $\left|T_{2n}\right| = o_p\left(\left(n\sigma_n^d\right)^{-1/2}\right)$. Together with the fact that $T_{1n} = O_p\left(\left(n\sigma_n^d\right)^{-1/2}\right)$, this would establish the proposition.

First observe that

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

Therefore,

$$
\begin{aligned}
& E\left(\sqrt{n\sigma_n^d}\left|T_{2n}\right|\right) \\
= & \frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E\left(K\left(\frac{X_i-x}{\sigma_n}\right)\left|\begin{array}{c} 1\left(\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)>\tau,\hat{F}_0\left(Y_{0i}\right)\leq s\right) \\ -1\left(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)>\tau,F_0\left(Y_{0i}\right)\leq s\right)\end{array}\right|\right) \\
= & \frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E_{X_i,Y_{0i},Y_{1i}}\left\{\begin{array}{c} K\left(\frac{X_i-x}{\sigma_n}\right)\times \\ E\left(\left|\begin{array}{c} 1\left(\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)>\tau,\hat{F}_0\left(Y_{0i}\right)\leq s\right) \\ -1\left(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)>\tau,F_0\left(Y_{0i}\right)\leq s\right)\end{array}\right| |X_i,Y_{0i},Y_{1i}\right)\end{array}\right\} \\
\leq & \frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E_{X_i,Y_{0i},Y_{1i}}\left\{K\left(\frac{X_i-x}{\sigma_n}\right)\Pr\left\{\left(\begin{array}{c}\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)>\tau, \\ F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)\leq\tau\end{array}\right)|X_i,Y_{0i},Y_{1i}\right\}\right\} \\
& +\frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E_{X_i,Y_{0i},Y_{1i}}\left\{K\left(\frac{X_i-x}{\sigma_n}\right)\Pr\left\{\left(\begin{array}{c}\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)\leq\tau, \\ F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)>\tau\end{array}\right)|X_i,Y_{0i},Y_{1i}\right\}\right\} \\
& +\frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E_{X_i,Y_{0i}}E\left\{K\left(\frac{X_i-x}{\sigma_n}\right)\Pr\left\{\hat{F}_0\left(Y_{0i}\right)\leq s,F_0\left(Y_{0i}\right)>s|X_i,Y_{0i}\right\}\right\} \\
& +\frac{1}{\sqrt{n\sigma_n^d}}\sum_{i=1}^{n} E_{X_i,Y_{0i}}E\left\{K\left(\frac{X_i-x}{\sigma_n}\right)\Pr\left\{\hat{F}_0\left(Y_{0i}\right)>s,F_0\left(Y_{0i}\right)\leq s|X_i,Y_{0i}\right\}\right\} \\
\equiv & S_{1n}+S_{2n}+S_{3n}+S_{4n},\text{ say.}
\end{aligned}
\tag{25}
$$

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

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

$$
\begin{aligned}
& \Pr\left(\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)>\tau,F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)<\tau|X_i,Y_{0i},Y_{1i}\right) \\
= & \Pr\left(\begin{array}{c}\hat{F}_1\left(Y_{1i}\right)-\hat{F}_0\left(Y_{0i}\right)-\left(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)\right)>\tau-\left(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)\right), \\ F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)<\tau|X_i,Y_{0i},Y_{1i}\end{array}\right) \\
\leq & \exp\left(-2\left(n-1\right)\left(\tau-\left(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)\right)\right)^2\right)\times 1(F_1\left(Y_{1i}\right)-F_0\left(Y_{0i}\right)<\tau),
\end{aligned}
$$

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

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

Continuing with the previous display, we have

$$\begin{aligned}
S_{1n} &\leq \frac{n}{\sqrt{n\sigma_n^d}} E_X \left[K\left(\frac{X - x}{\sigma_n}\right) G_n(X) \right] \\
&= \frac{n\sigma_n^d}{\sqrt{n\sigma_n^d}} \int \left[K\left(u\right) G_n(x + \sigma_n u) f\left(x + \sigma_n u\right) \right] du \\
&= \sqrt{n\sigma_n^d} \int \left[K\left(u\right) G_n(x + \sigma_n u) f\left(x + \sigma_n u\right) \right] du \\
&= f\left(x\right) \sqrt{n\sigma_n^d} \int K\left(u\right) G_n(x) du + \text{terms of smaller order} \\
&= f\left(x\right) \sqrt{n\sigma_n^d} G_n(x) + \text{terms of smaller order.} \qquad (26)
\end{aligned}$$

Now, notice that $G_n(x)$ is of the form

$$\begin{aligned}
G_n(x) &= E_{Z|X} \left[\exp\left(-2\left(n-1\right) Z^2\right) \times 1(Z > 0) | X = x \right] \\
&\leq c \int \exp\left(-2\left(n-1\right) z^2\right) f\left(z|x\right) dz \\
&\leq c' \int \exp\left(-2\left(n-1\right) z^2\right) dz \\
&= O\left(n^{-1/2}\right) \qquad (27)
\end{aligned}$$

by the normal (Gaussian) integral formula. From (26) and (27), it follows that

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

Together with analogous proofs for S_{2n}, S_{3n}, S_{4n}, this implies that in (18), $\sqrt{n\sigma_n^d}T_{2n} = o_p(1)$, but under standard assumptions for Nadaraya-Watson regressions, (c.f. Pagan and Ullah (1999) theorem 3.5, 3.6), $\sqrt{n\sigma_n^d}T_{1n} = O_p(1)$ and $\hat{f}(x) = f(x) + o_p(1)$. This shows that

$$\begin{aligned} & \sqrt{n\sigma_n^d}\left(\hat{\upsilon}_c(\tau, s; x) - \upsilon_c(\tau, s; x)\right) \\ = \; & \sqrt{n\sigma_n^d}\left\{\frac{\hat{E}\left(1\left(F_1(Y_1) - F_0(Y_0) > \tau, F_0(Y_0) \leq s\right)|X = x\right)}{s} - \upsilon_c(\tau, s; x)\right\} + o_p(1) \\ & \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{s^2 f(x)} \int K^2(u)\, du\right) \end{aligned}$$

where $\sigma^2(x) = s \times \upsilon_c(\tau, s; x) \times (1 - s \times \upsilon_c(\tau, s; x))$ and $f(x)$ is the marginal density of X. ■

Table 1: Transition Probability Estimates by Race

$\mathrm{Prob}(Y1>\xi+\tau,\ Y0<\xi)/\mathrm{Prob}(Y0<\xi)$

		τ=0			τ=0.1			τ=0.2			τ=0.3	
ξ	Whites	Blacks	W-B	Whites	Blacks	W-B	Whites	Blacks	W-B	Whites	Blacks	W-B
5	0.978	0.891	0.087	0.849	0.579	0.270	0.704	0.407	0.297	0.593	0.280	0.312
	(0.034)	(0.032)	(0.049)	(0.056)	(0.048)	(0.076)	(0.072)	(0.047)	(0.089)	(0.074)	(0.047)	(0.091)
10	0.917	0.702	0.216	0.760	0.458	0.302	0.632	0.340	0.292	0.555	0.249	0.306
	(0.031)	(0.027)	(0.044)	(0.044)	(0.030)	(0.053)	(0.048)	(0.031)	(0.056)	(0.051)	(0.029)	(0.057)
15	0.812	0.616	0.196	0.692	0.423	0.269	0.542	0.309	0.232	0.459	0.212	0.247
	(0.030)	(0.031)	(0.044)	(0.037)	(0.030)	(0.050)	(0.040)	(0.028)	(0.049)	(0.039)	(0.024)	(0.048)
20	0.752	0.524	0.228	0.618	0.389	0.229	0.496	0.281	0.215	0.379	0.192	0.187
	(0.027)	(0.025)	(0.039)	(0.031)	(0.025)	(0.040)	(0.033)	(0.023)	(0.039)	(0.036)	(0.019)	(0.039)
25	0.708	0.447	0.261	0.558	0.326	0.232	0.459	0.234	0.225	0.342	0.156	0.186
	(0.025)	(0.024)	(0.038)	(0.026)	(0.020)	(0.035)	(0.027)	(0.019)	(0.034)	(0.029)	(0.015)	(0.032)
30	0.646	0.403	0.244	0.539	0.290	0.249	0.418	0.200	0.217	0.305	0.131	0.174
	(0.023)	(0.023)	(0.033)	(0.024)	(0.020)	(0.031)	(0.025)	(0.018)	(0.032)	(0.023)	(0.015)	(0.028)
35	0.583	0.349	0.234	0.478	0.254	0.224	0.366	0.173	0.193	0.257	0.120	0.136
	(0.020)	(0.020)	(0.031)	(0.021)	(0.018)	(0.030)	(0.022)	(0.016)	(0.028)	(0.020)	(0.013)	(0.025)
40	0.544	0.311	0.233	0.427	0.223	0.203	0.315	0.148	0.167	0.228	0.105	0.122
	(0.019)	(0.019)	(0.029)	(0.019)	(0.016)	(0.027)	(0.018)	(0.014)	(0.025)	(0.016)	(0.011)	(0.022)
45	0.494	0.262	0.232	0.372	0.180	0.192	0.264	0.123	0.141	0.190	0.080	0.109
	(0.015)	(0.020)	(0.027)	(0.016)	(0.015)	(0.023)	(0.015)	(0.013)	(0.020)	(0.015)	(0.010)	(0.019)
50	0.428	0.226	0.202	0.320	0.152	0.168	0.227	0.107	0.119	0.147	0.065	0.082
	(0.015)	(0.015)	(0.024)	(0.014)	(0.013)	(0.022)	(0.014)	(0.010)	(0.018)	(0.013)	(0.008)	(0.016)

Notes: See text for a description of the estimator. Data is from the NLSY and uses multiyear averages of son's income over 1996-2002 and parent income measured over 1978-1980. Standard errors are in parentheses.

Table 2: Upward Mobility Estimates by Race

Prob($F_1(Y_1)-F_0(Y_0)>\tau | Y_0<=s$)

	τ=0			τ=0.1			τ=0.2			τ=0.3		
s	Whites	Blacks	W-B	Whites	Blacks	W-B	Whites	Blacks	W-B	Whites	Blacks	W-B
5	0.977	0.950	0.027	0.904	0.635	0.270	0.745	0.420	0.325	0.614	0.312	0.303
	(0.024)	(0.018)	(0.033)	(0.047)	(0.044)	(0.066)	(0.065)	(0.045)	(0.083)	(0.073)	(0.040)	(0.084)
10	0.947	0.883	0.065	0.840	0.574	0.266	0.698	0.377	0.321	0.595	0.288	0.307
	(0.022)	(0.022)	(0.032)	(0.035)	(0.032)	(0.051)	(0.047)	(0.031)	(0.059)	(0.053)	(0.025)	(0.061)
15	0.909	0.835	0.074	0.786	0.567	0.219	0.629	0.390	0.240	0.519	0.281	0.238
	(0.021)	(0.020)	(0.029)	(0.031)	(0.027)	(0.042)	(0.040)	(0.025)	(0.047)	(0.040)	(0.025)	(0.048)
20	0.871	0.796	0.075	0.755	0.556	0.198	0.592	0.387	0.205	0.485	0.285	0.200
	(0.021)	(0.017)	(0.027)	(0.029)	(0.024)	(0.039)	(0.030)	(0.022)	(0.037)	(0.032)	(0.020)	(0.039)
25	0.838	0.762	0.076	0.724	0.537	0.187	0.575	0.373	0.202	0.463	0.274	0.188
	(0.021)	(0.019)	(0.030)	(0.024)	(0.024)	(0.038)	(0.028)	(0.024)	(0.036)	(0.028)	(0.019)	(0.034)
30	0.821	0.734	0.087	0.715	0.521	0.193	0.568	0.360	0.208	0.447	0.262	0.185
	(0.018)	(0.019)	(0.027)	(0.021)	(0.021)	(0.033)	(0.026)	(0.020)	(0.036)	(0.025)	(0.019)	(0.035)
35	0.786	0.717	0.069	0.668	0.514	0.154	0.537	0.360	0.178	0.415	0.263	0.153
	(0.019)	(0.017)	(0.026)	(0.020)	(0.023)	(0.030)	(0.021)	(0.019)	(0.031)	(0.023)	(0.016)	(0.028)
40	0.757	0.704	0.052	0.641	0.506	0.135	0.513	0.357	0.156	0.393	0.254	0.139
	(0.018)	(0.016)	(0.025)	(0.017)	(0.020)	(0.028)	(0.020)	(0.019)	(0.027)	(0.019)	(0.018)	(0.027)
45	0.731	0.687	0.044	0.605	0.495	0.110	0.484	0.350	0.134	0.367	0.248	0.119
	(0.015)	(0.017)	(0.024)	(0.021)	(0.021)	(0.032)	(0.019)	(0.018)	(0.028)	(0.019)	(0.017)	(0.026)
50	0.695	0.668	0.028	0.578	0.481	0.097	0.457	0.342	0.115	0.342	0.242	0.100
	(0.014)	(0.018)	(0.025)	(0.016)	(0.020)	(0.028)	(0.016)	(0.018)	(0.025)	(0.017)	(0.015)	(0.024)

Notes: See text for a description of the estimator. Data is from the NLSY and uses multiyear averages of son's income over 1996-2002 and parent income measured over 1978-1980. Standard errors are in parentheses.

Figure 1: Transition Probabilities Conditional On Parent Percentile

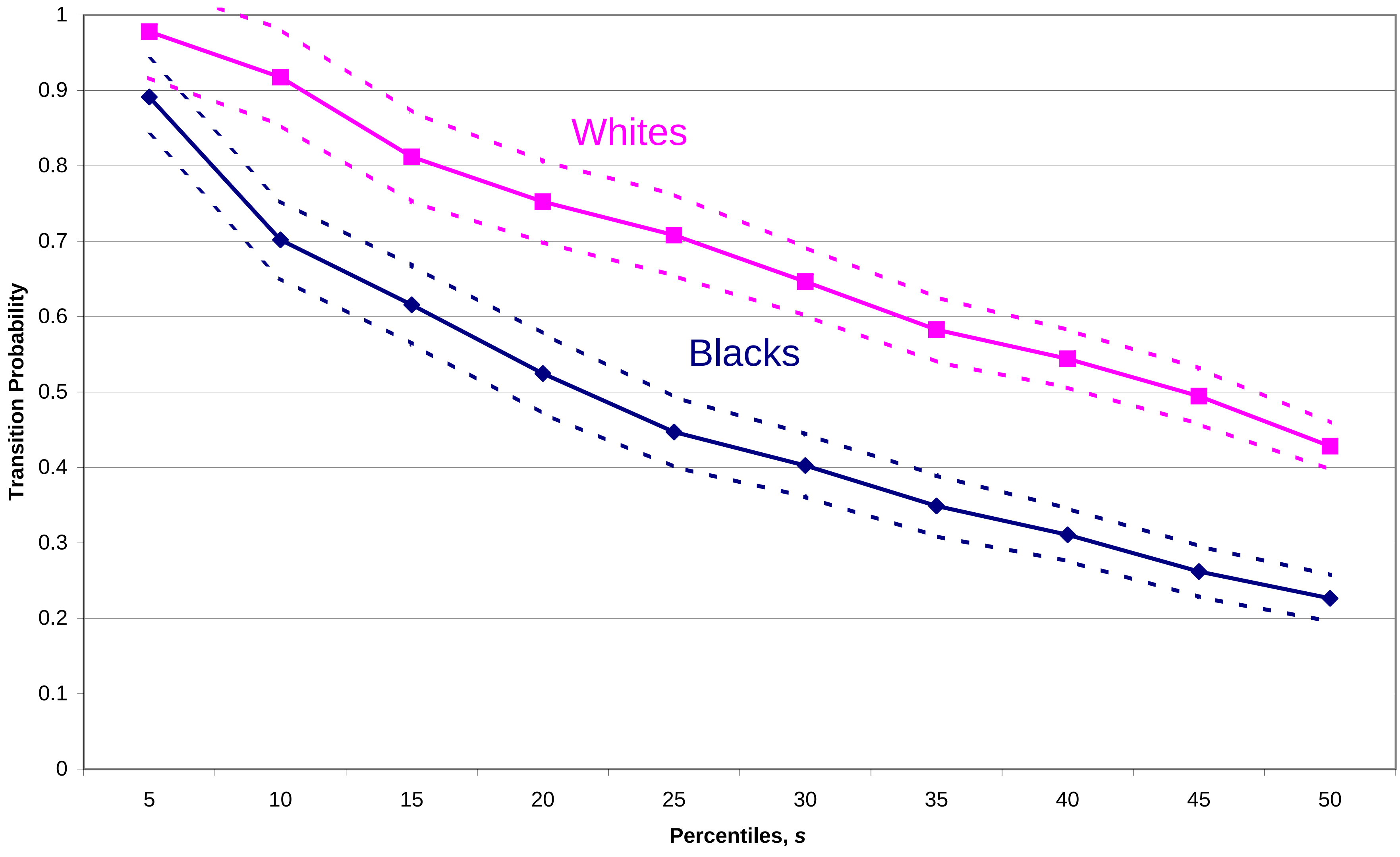

Figure 2: Upward Mobility Conditional On Parent Percentile
Pr (*F1>F0*|*F0<=s*)

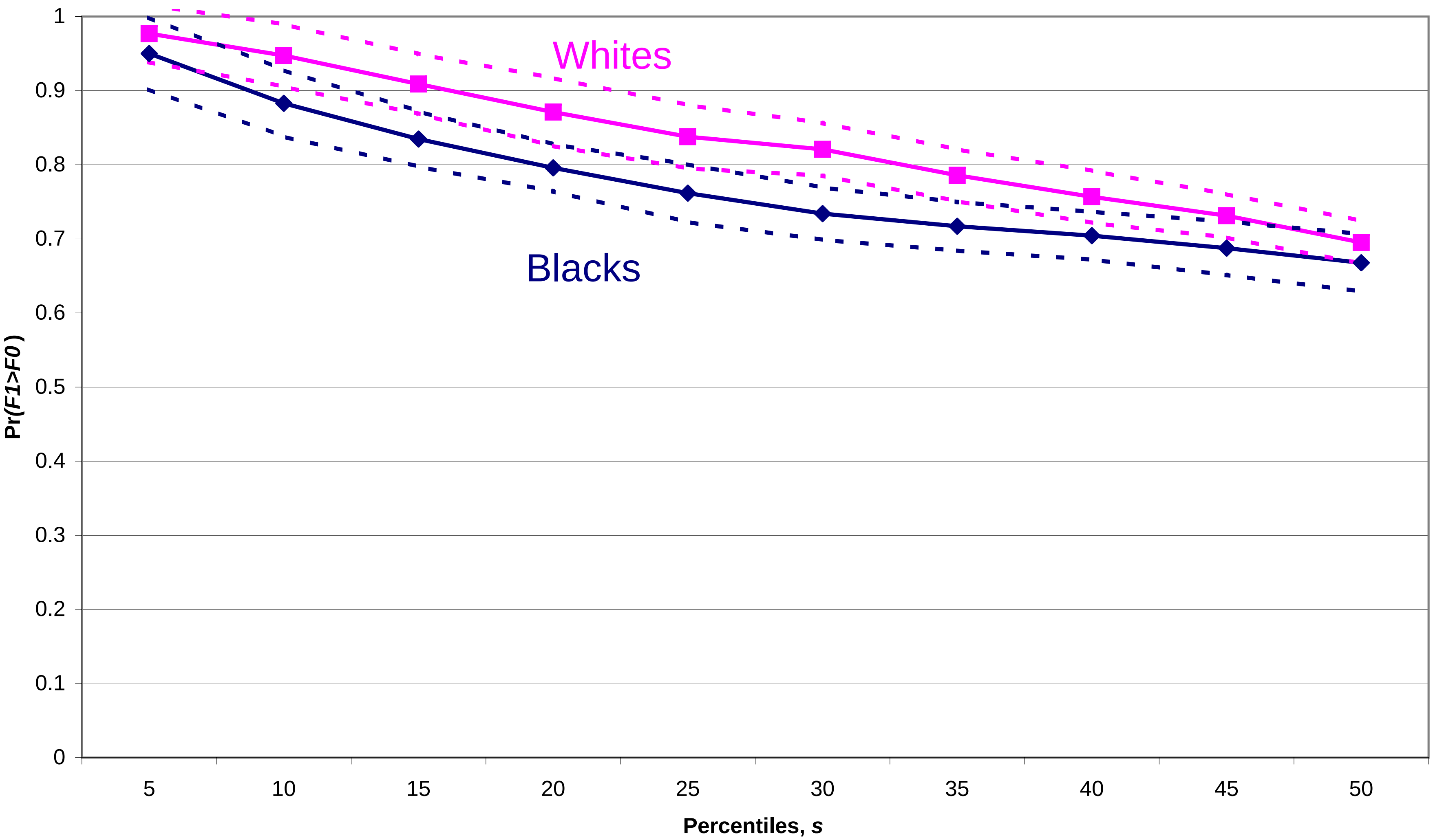

Figure 3: Transition Probabilities vs Upward Mobility, Whites Minus Blacks Conditional On Parent Percentile

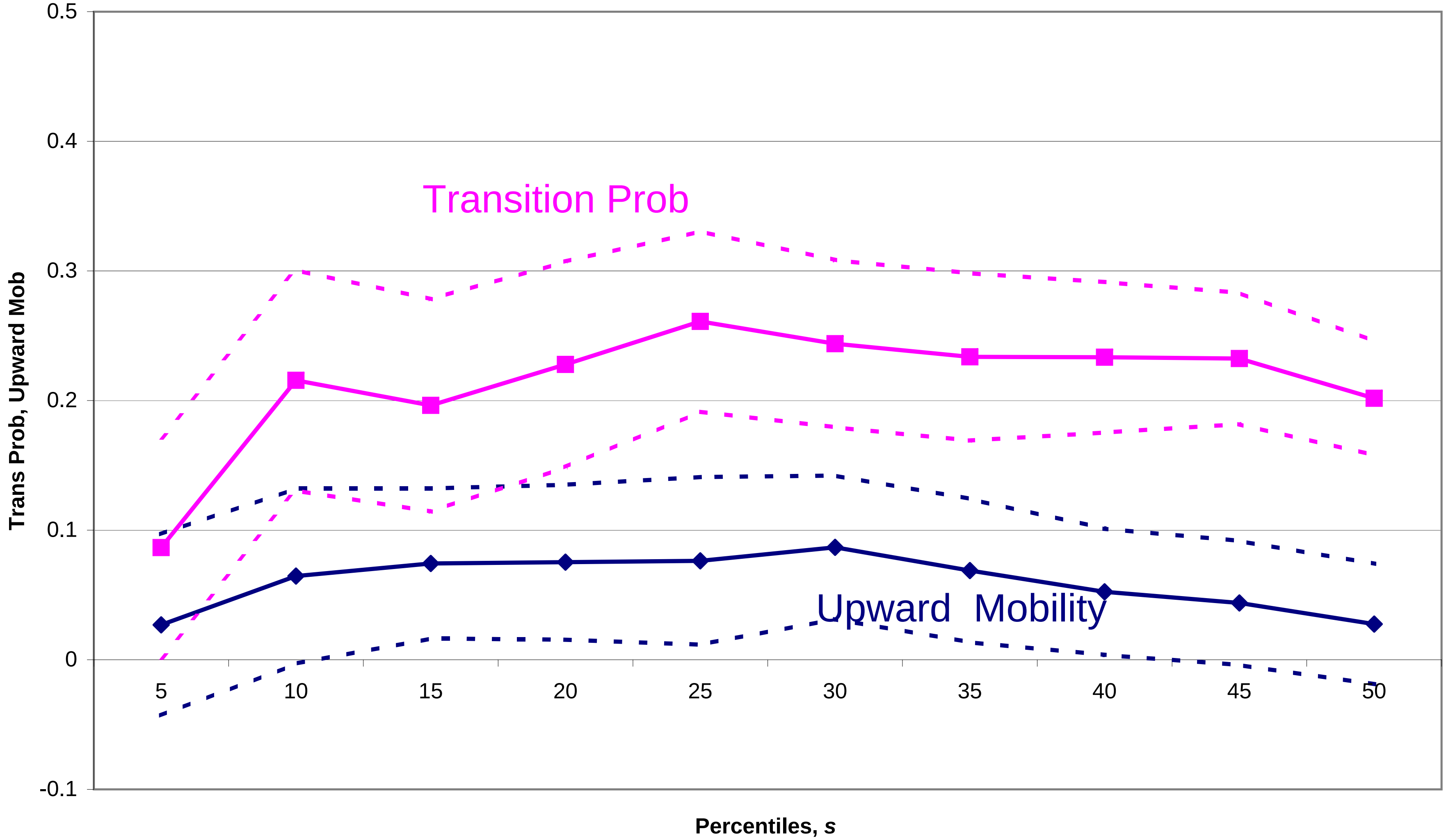

Figure 4: CDF of Parent Income Conditional on Being in the Bottom Quintile
Whites vs Blacks

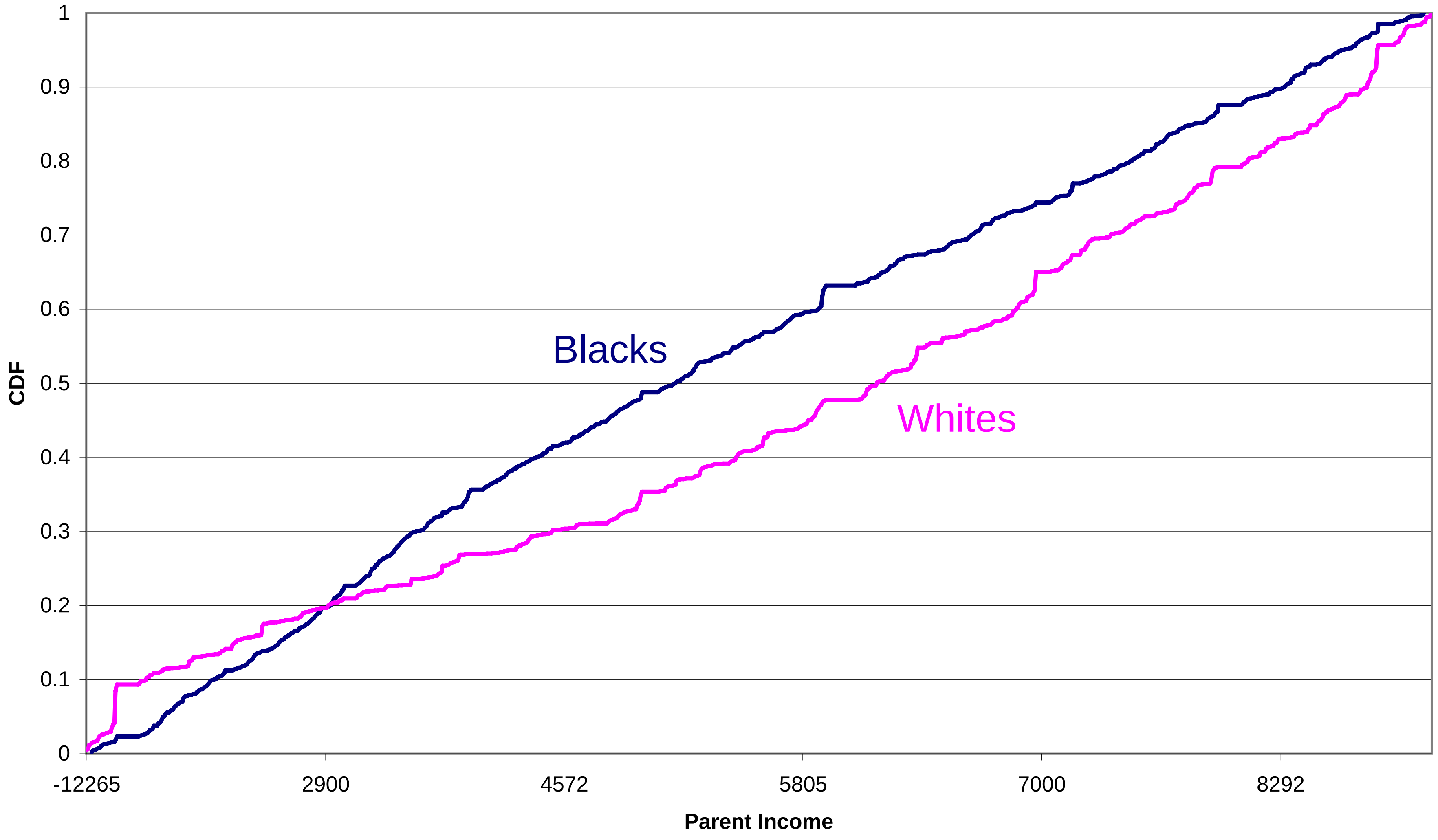

Figure 5: Probability of Leaving Bottom Quintile Conditional on AFQT
Whites vs Blacks

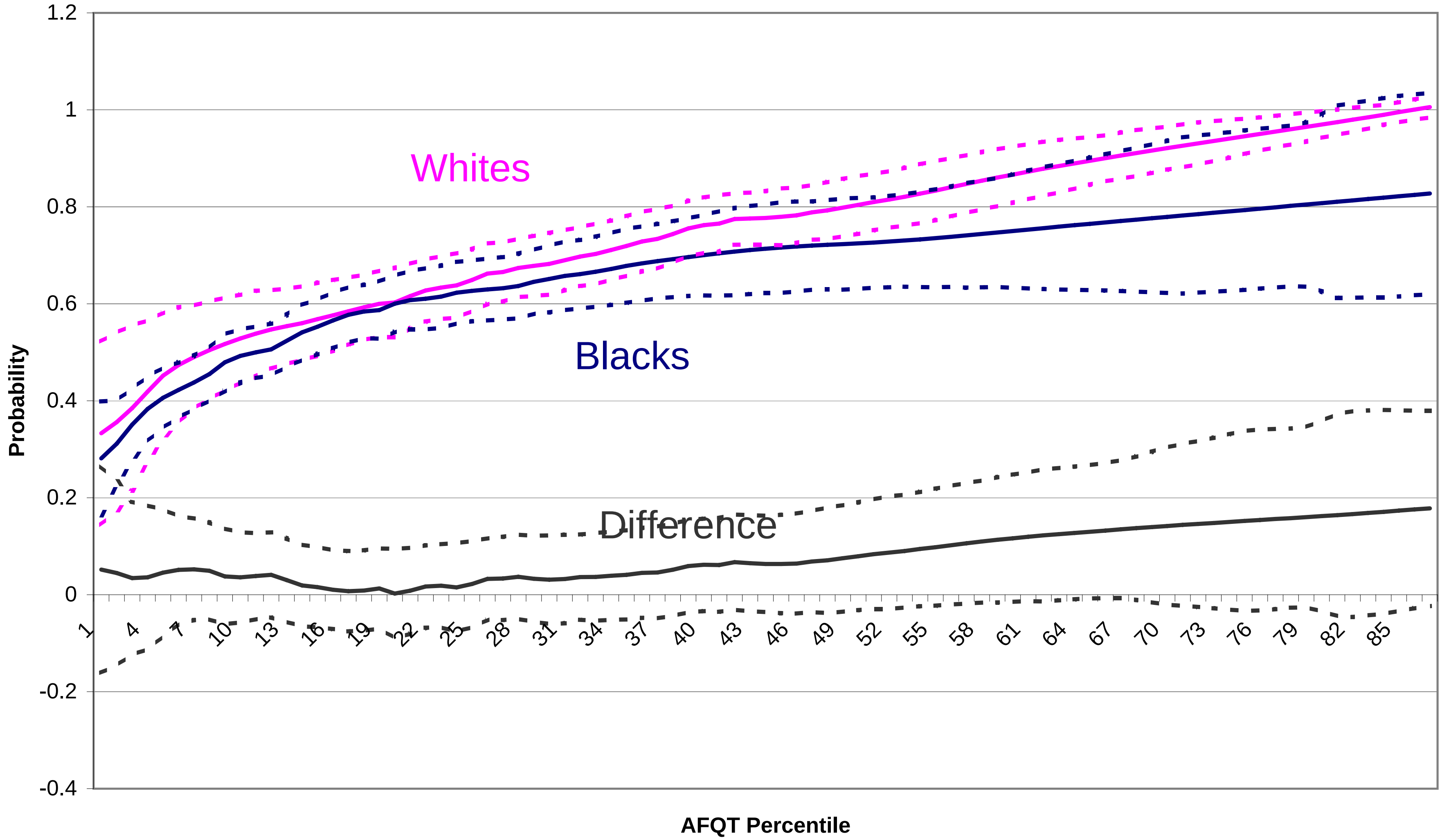

Figure 6: Transition Probability of Leaving Bottom Quintile Conditional on Own Education
Whites vs Blacks

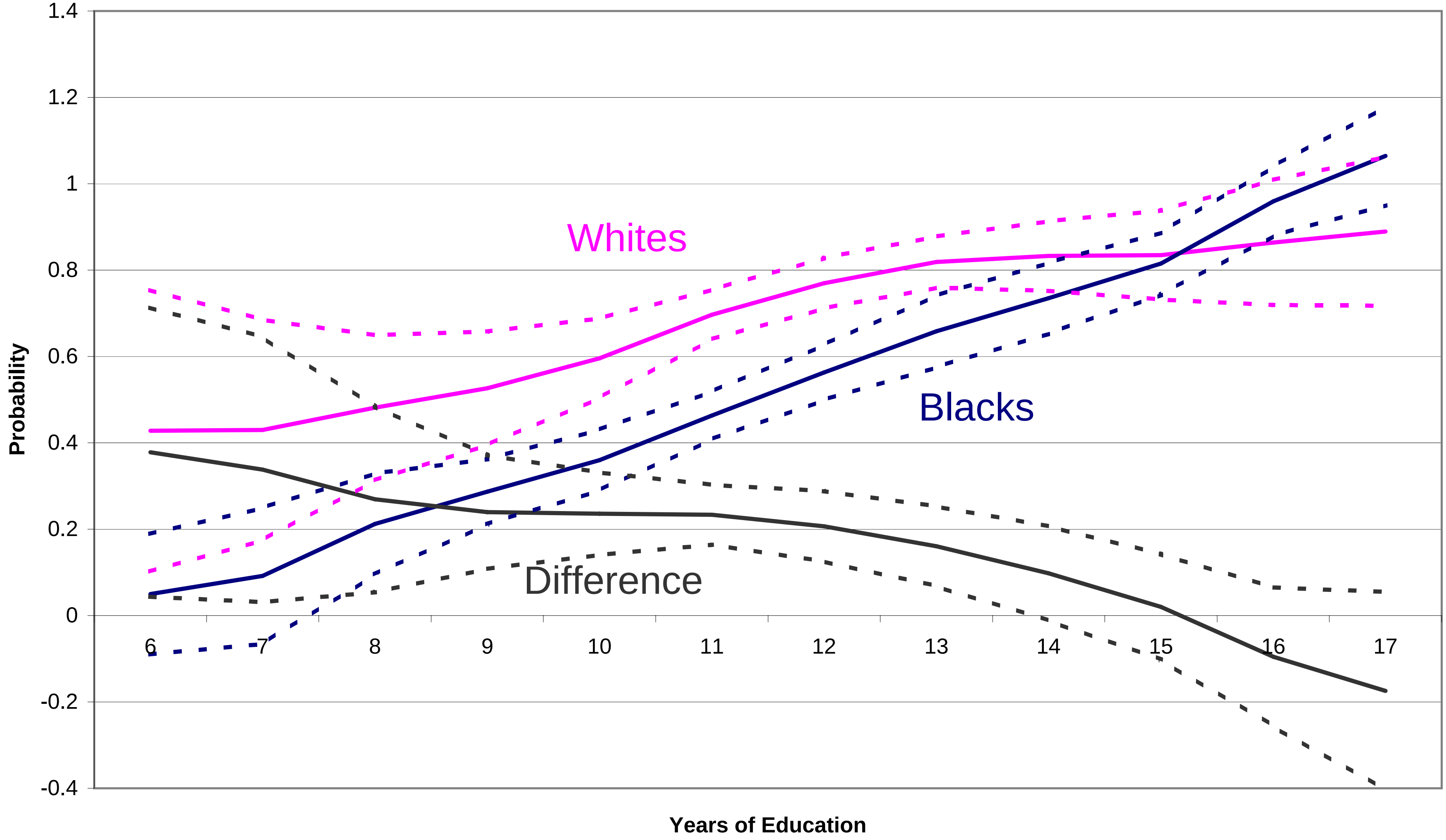

Figure 7: Comparison of Probit and Non-Parametric Estimates of Transition Probability of Blacks Leaving Bottom Quintile Conditional on AFQT

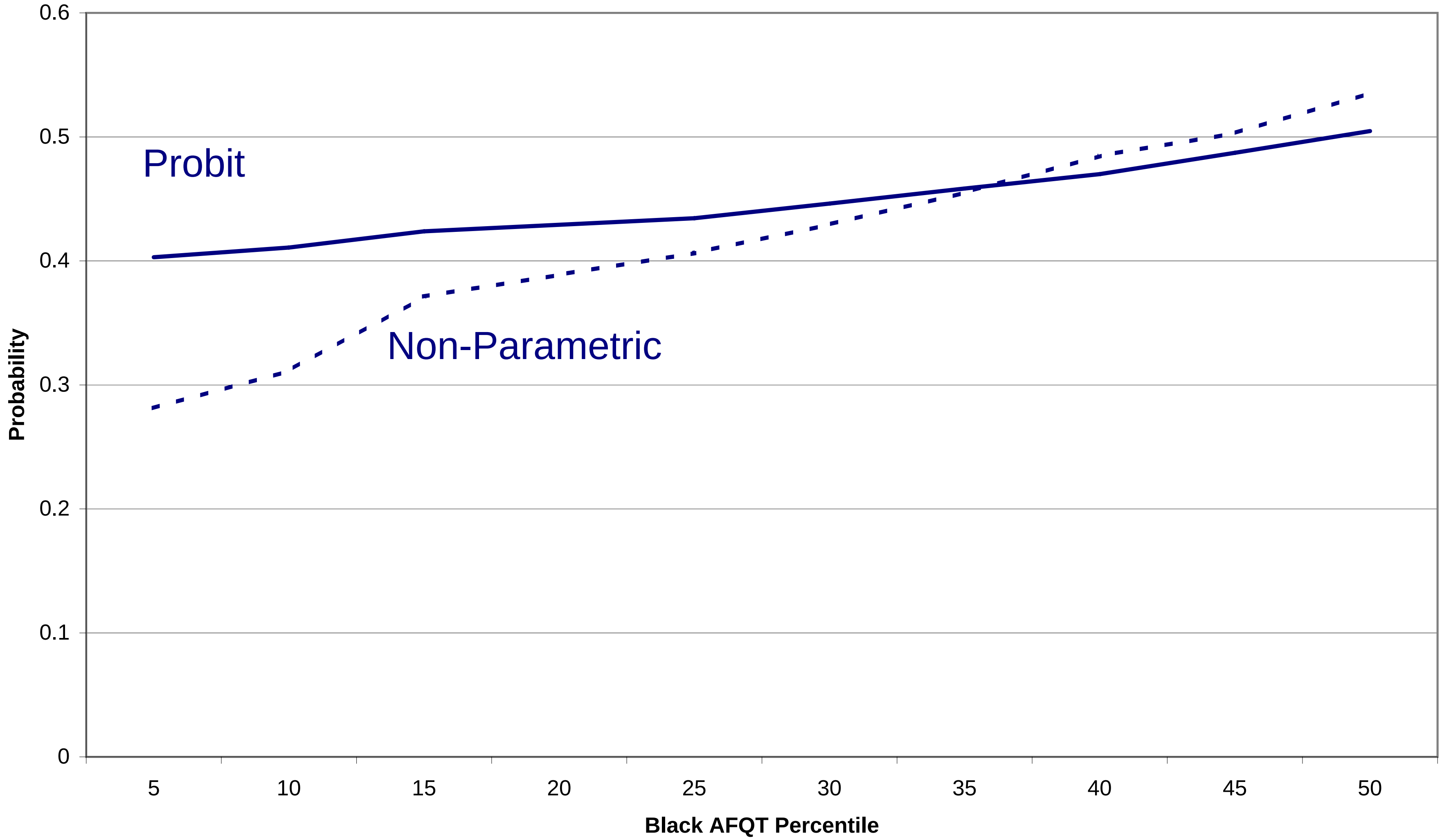

Figure 8: Upward Mobility Conditional on AFQT Scores
Whites vs Blacks

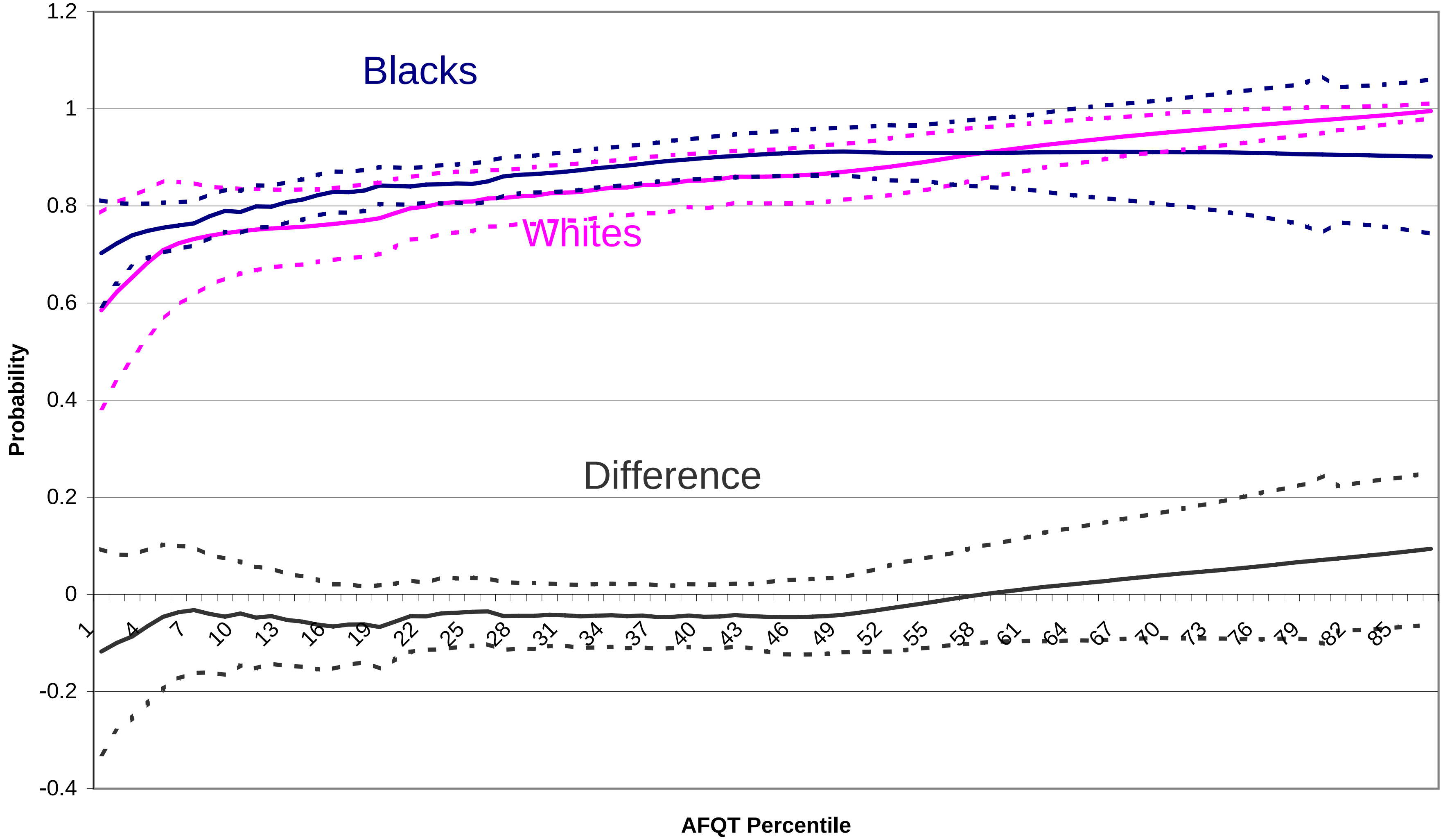

Figure 9: Density Weighted Average Derivative of Transition Probability (Son>Q|Dad<Q) With Respect to AFQT, Blacks vs Whites

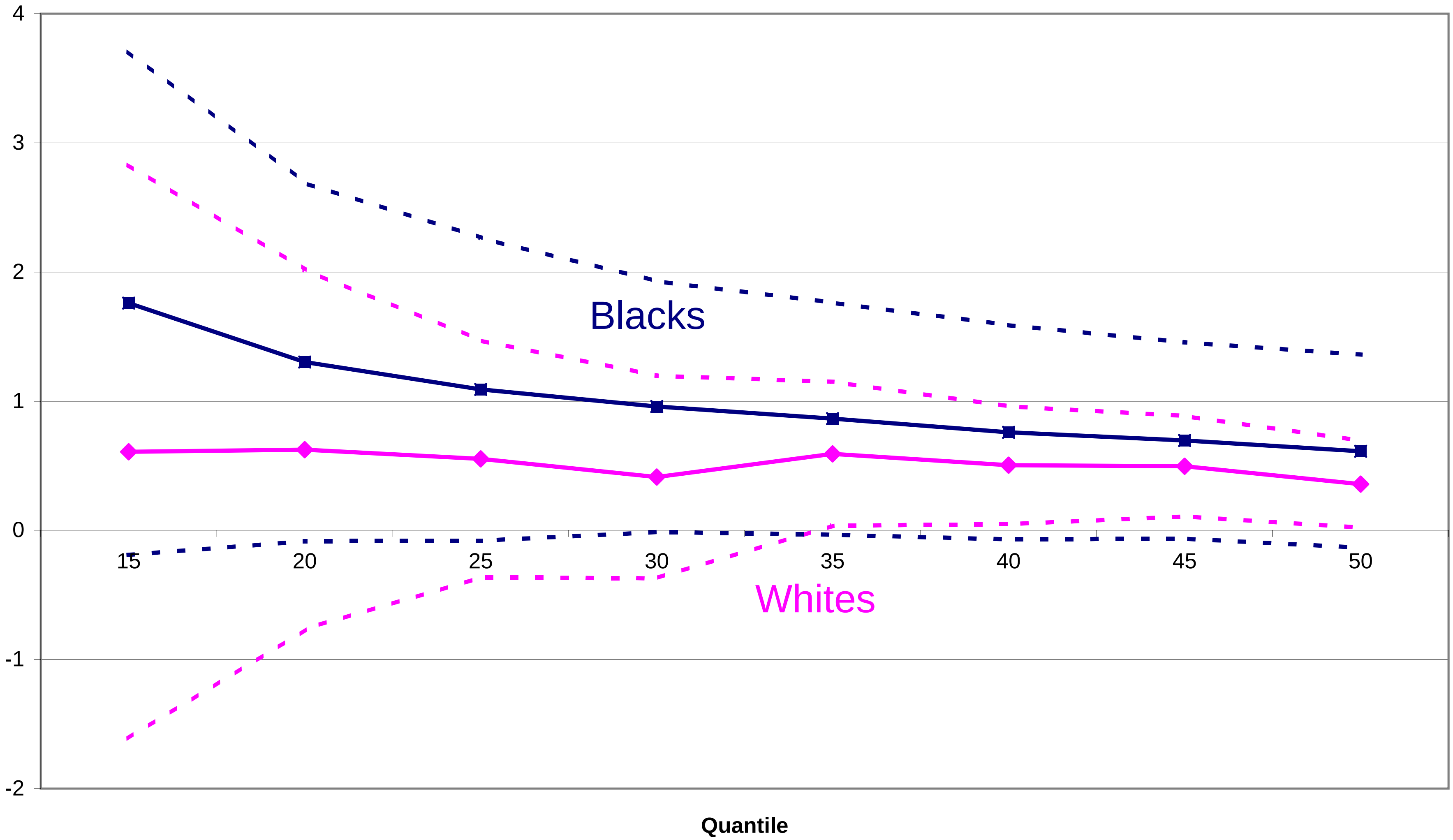

Figure 10: Kernel Density Estimates of AFQT Percentiles

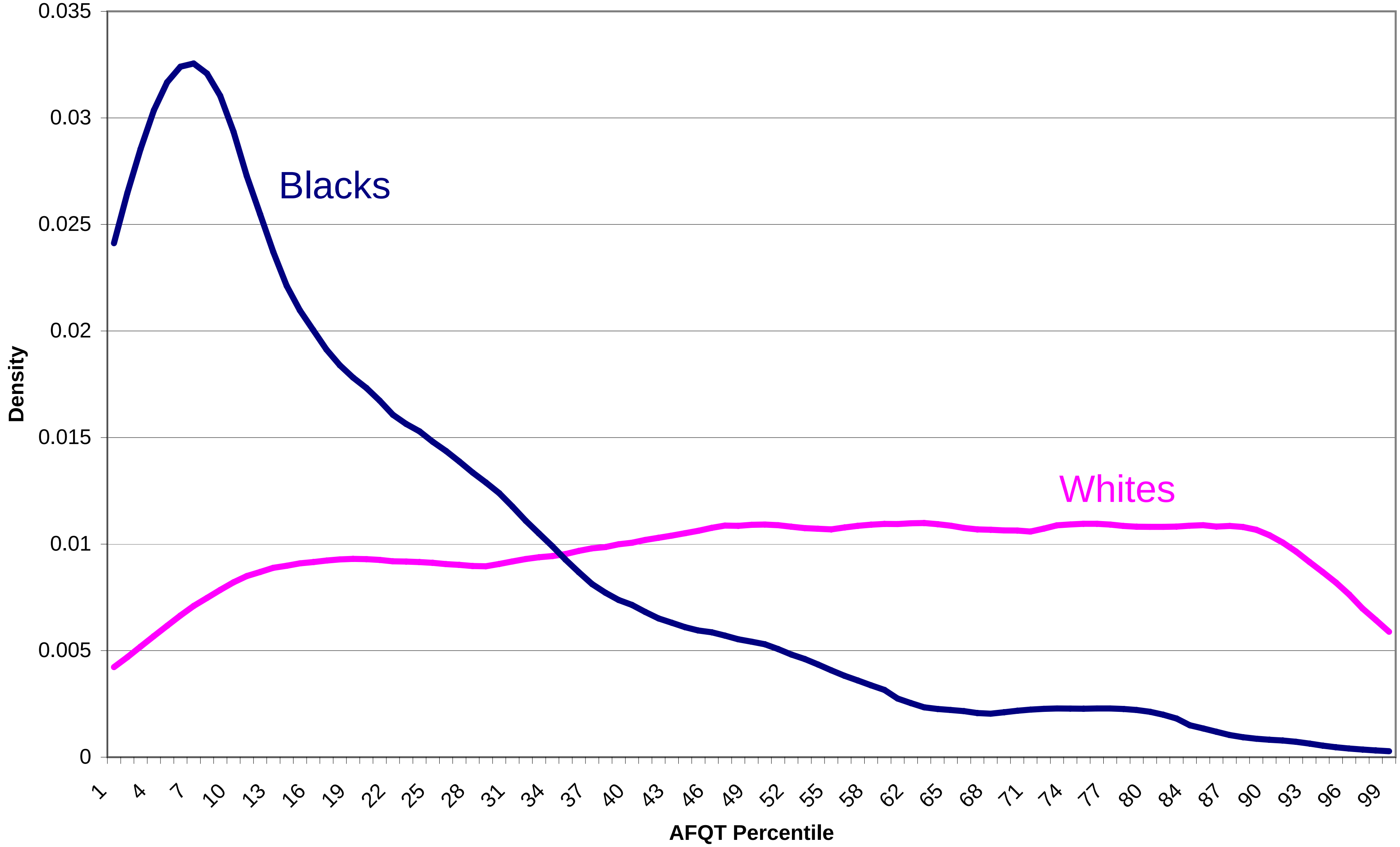

Figure 11: Density Weighted Average Derivative of Upward Mobility With Respect to AFQT, Blacks vs Whites