Head Start and the distribution of long term education and labor market outcomes*

Monique De Haan† Edwin Leuven‡

April 2019

Abstract

We investigate the effect of Head Start on education and wage income for individuals in their 30s in the NLSY79. We contribute to the existing literature by examining effects across the outcome distributions, using an approach that relies on two weak stochastic dominance assumptions that can be checked using pre-Head Start cohorts. We find that Head Start has positive and statistically significant effects on years of education and wage income. We also uncover important heterogeneity in the effectiveness of the program; the effects are concentrated at the lower end of the distribution, and the effects are strongest for women, blacks and Hispanics.

JEL-codes: H52, I24, I28, J13, J24, J31.

Keywords: Head Start, Preschool, Long term outcomes, Partial identification

*We thank David Deming, Jim Heckman, seminar participants, the editor and three anonymous referees for valuable feedback and suggestions. Rita Ginja kindly provided Head Start participation effect estimates for Carneiro and Ginja (2014).

†Department of Economics, University of Oslo. Also affiliated with CESifo, ESOP and Statistics Norway. moniqued@econ.uio.no

‡Department of Economics, University of Oslo. Also affiliated with IZA, CESifo, CEPR and Statistics Norway. edwin.leuven@econ.uio.no

1 Introduction

Head Start is a major federally funded preschool program in the U.S. It is targeted at children from low-income parents and provides these children and their parents with schooling, health, nutrition, and social welfare services. Although many studies argue that investments in early childhood, including preschool, are crucial for many outcomes later in life (Knudsen et al., 2006; Elango et al., 2016), there are concerns about the effectiveness of Head Start. Many of the recent concerns are based on results from the Head Start Impact Study (HSIS), which randomly assigned about 5,000 eligible 3- and 4-year old children either to a treatment group that was allowed to enroll in a participating Head Start center or to a control group that did not have access to any of the participating Head Start centers. The results from this randomized experiment show positive effects of Head Start on cognitive outcomes immediately after the program, but these positive effects quickly fade out (Puma et al., 2010). Recently Kline and Walters (2016) and Feller et al. (2016) show that the finding of fade out is sensitive to the choice of counterfactual treatment. In addition, as argued by Gibbs et al. (2011), fade out in cognitive test scores does not necessarily imply that Head Start is ineffective. In fact, a few recent studies that have evaluated Head Start using quasi-experimental designs find positive effects on medium and longer term outcomes such as crime and health outcomes (Carneiro and Ginja, 2014; Currie and Thomas, 1995, 2000; Deming, 2009; Garces et al., 2002; Ludwig and Miller, 2007).

A disadvantage of these quasi-experimental studies is that they rely on stronger assumptions than the randomized experiment of the Head Start Impact Study. In addition, these studies observe individuals in their teens or early 20s. For certain outcomes, such as crime, these may be the appropriate ages to measure the outcome variable, while measuring education in people's early 20s could lead to truncation because individuals might not have finished their education. Similarly, labor market outcomes are better measured when individuals are in their 30s if one wants to reduce life-cycle bias (Bhuller et al., 2017; Böhlmark and Lindquist, 2006; Haider and Solon, 2006). While these existing long run effect studies broadly agree on how Head Start participation affects health outcomes, there is less agreement regarding the effect on educational attainment and very little evidence on the effect on subsequent earnings.[1]

In this study we use the NLSY to investigate the effect of Head Start on long term education and labor market outcomes and contribute to the existing literature in three ways. First, we investigate the impact of Head Start on outcomes observed for individuals in their 30s. Observing individuals in their 30s allows us to look at the effect of Head Start on wage income measured when everyone has completed formal education, something which previous studies have not been able to do. Second, we use a partial identification approach that relies on two weak stochastic dominance assumptions. The major advantage of using the NLSY is that it allows us to check the validity of these assumptions using data on pre-Head Start cohorts (born between 1957-1959) who did not have the opportunity to enroll in Head Start. Third, in contrast to previous studies that have estimated (local) average treatment effects, we estimate upper and lower bounds around entire cumulative potential outcome distributions. By focusing on cumulative distributions we can investigate whether the impact of Head Start differs between the top and bottom end of the outcome distribution. To our knowledge we are the first to investigate the impact of Head Start across the distribution of long term outcomes. Bitler et al. (2014) also estimate distributional impacts of Head Start, but they estimate quantile treatment effects on cognitive and non-cognitive outcomes in preschool through 1st grade while we focus on long term education and labor market outcomes.

The empirical analysis in this paper follows a partial identification approach based on two assumptions. Since Head Start is targeted at disadvantaged children we assume that the potential outcome distributions of Head Start participants are weakly stochastically dominated by those of non-participants. This assumption is motivated by the eligibility criteria of Head Start and is consistent with observed selection into the program (Schnur

[1]Using the PSID, Garces et al. (2002) and Grosz et al. (2016), report impact estimates of Head Start participation on earnings for 23 to 25-year-olds, but find no evidence of such a relationship. Section 2 gives a more detailed overview of the literature.

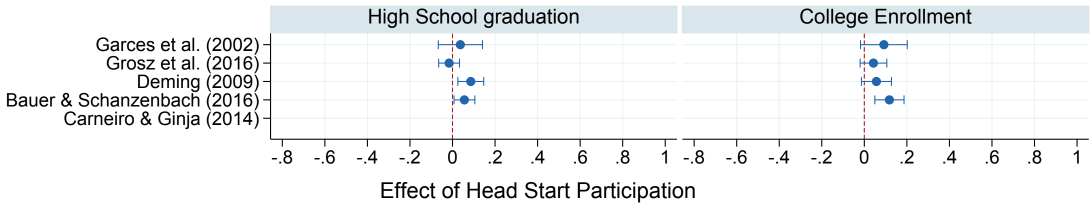
et al., 1992). In addition we assume that the potential outcome distributions of individuals with low educated parents are weakly stochastically dominated by those of individuals with high educated parents. The first assumption is a variant of a monotone treatment selection (MTS) assumption, while the second implies that we use parental education as a monotone instrumental variable (MIV) following Manski and Pepper (2000). By performing Kolmogorov-Smirnov tests using data on parental background and on outcomes of pre-Head Start cohorts we find strong support for the validity of these two identifying assumptions.

Combining the two stochastic dominance assumptions results in lower bounds that show that Head Start has a positive and statistically significant effect on years of education and on wage income. We also find that there is important heterogeneity in the effectiveness of the program. The significant positive effects are concentrated at the lower end of the distribution, and the effects are strongest for women, blacks and Hispanics. In line with Kline and Walters (2016) and Feller et al. (2016) we find evidence indicating that the counterfactual matters: the lower bounds are higher when the counterfactual is only informal care compared to a counterfactual which is a mixture of informal care and other preschool.

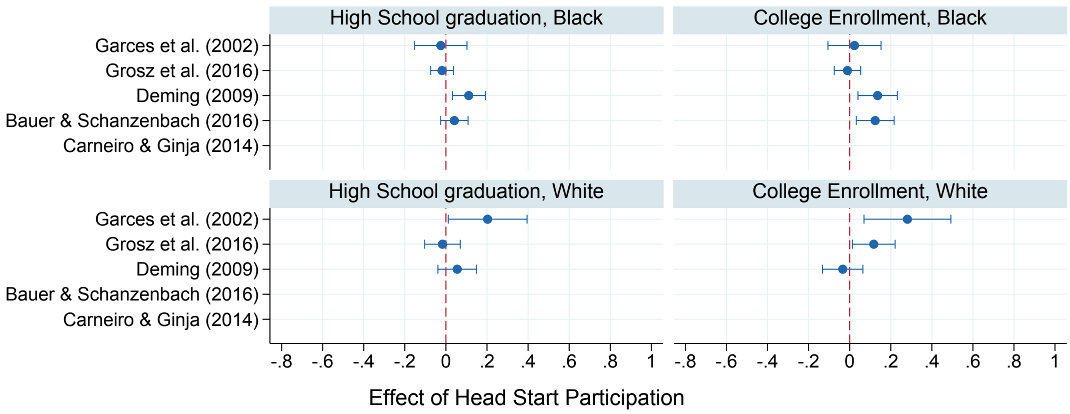
2 Background and literature

Head start was launched in 1965 by the Office of Economic Opportunity (OEO), with the goal to prepare children from disadvantaged backgrounds for compulsory schooling. It started as an eight-week summer program, but from 1966 onwards it continued as a year-round program. Head Start is targeted at children from low-income families, more specifically, children from families with income below or on the poverty line are eligible to participate in Head Start.

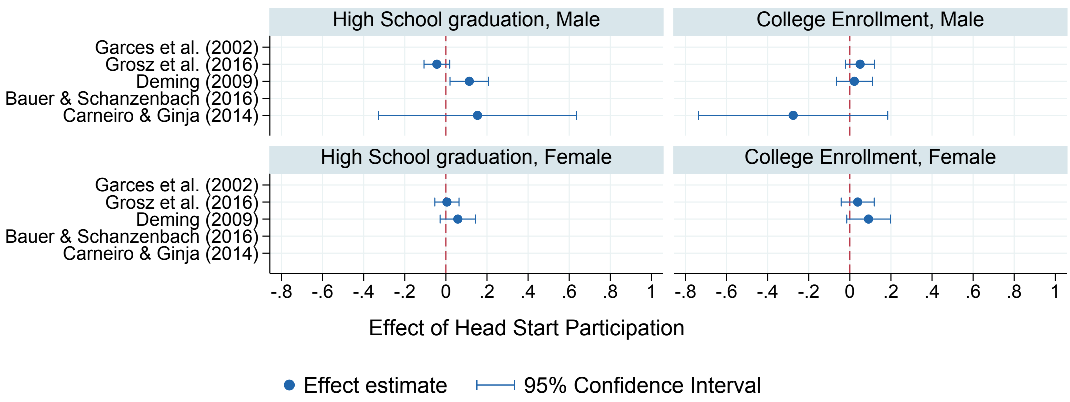
Starting with the Westinghouse Study in 1969 there have been numerous evaluations of the short term impacts of Head Start. The literature on long term effects of Head Start



(a) Average effects



(b) Effects by race



(c) Effects by gender

Figure 1. Quasi-experimental estimates of Head Start participation on long-run schooling outcomes

is however much smaller.[2] Figure 1 summarizes the available estimates of the effect of Head Start participation on long-run schooling outcomes.[3] As shown in the top panel of this figure, there are only four studies, Deming (2009); Garces et al. (2002); Grosz et al. (2016); Bauer and Schanzenbach (2016), that report estimates for a population that contains individuals from both genders and all races.[4] All four studies estimate family-fixed effect models and thus rely on variation in Head Start participation between siblings. The middle and bottom panels of Figure 1 show their effect estimates by race and by gender. The bottom panel also shows the estimate of a fifth study, Carneiro and Ginja (2014), that uses a (fuzzy) regression discontinuity design based on income eligibility rules to estimate the causal effect of Head Start participation.[5] This study only report results for men.[6]

As can be seen in Figure 1, most of these individual quasi-experimental studies on long-term outcomes find some positive effects of Head Start participation, but they differ substantially in the specific long-term education outcomes that are affected as well as the subgroups that are found to benefit from Head Start. For example, while Garces et al. (2002) find large positive and statistical significant effects on high school graduation and college enrollment for whites and no effects for blacks, Deming (2009) finds estimates close to zero for whites and positive and statistically significant effects for blacks. As pointed out by Elango et al. (2016), it is unclear whether the lack of consistency between these studies is due to differences in (counterfactual) treatment, differences in population,

[2] Although these studies also look at other outcomes such as health and crime, we focus our discussion on the results for schooling and earnings in light of the outcomes in the current paper.

[3] Figure 1 only reports estimates on the effect of Head Start participation on long run schooling outcomes. Both Garces et al. (2002) and Grosz et al. (2016) report impact estimates of Head Start participation on earnings for individuals who are 23-25 years old, but they do not find evidence of such a relationship.

[4] Ludwig and Miller (2007) exploit a discontinuity in Head Start funding rates at the OEO cutoff for grant-writing assistance. They report evidence of positive effects on high school completion and college attendance. Recently Thompson (2018) estimated intention-to-treat effects of average county Head Start funding per child aged three to six in the early years of the program by exploiting geographic variation in the timing of Head Start funding. We do not report their estimates in Figure 1 because we focus on the effect of Head Start participation and it is not clear that these estimates of Head Start funding can be interpreted as the effect of Head Start participation because the treatment, receipt of Head Start grants and county funding levels, could have also affected spending per participant.

[5] Carneiro and Ginja (2014) do not report the effect of Head Start participation, but only first-stage (Table 2) and reduced form effects (Table 8), Rita Ginja kindly provided the IV-probit effect estimates and bootstrapped standard errors shown in Figure 1.

[6] They are unable to estimate effects for women because their first-stages are insignificant.

or problems related to the empirical approach. This is also highlighted by Grosz et al. (2016), who show that the local average treatment effects obtained in the family fixed effect approach rely on families that differ from other Head Start families in size and other observable dimensions. They also show that this heterogeneity with respect to family size explains half of the difference between the OLS and family fixed effect estimate.

3 Data

Our analysis uses data from the National Longitudinal Study of Youth 1979 (NLSY79) which is a sample of 14 to 22-year-olds living in the U.S. in 1979 who were interviewed annually up to 1994 and every other year after. Although the oldest individuals in the NLSY79 were born in 1957, the first cohort to become eligible for Head Start was born in 1960, and we thus base our analysis on the 1960–64 cohorts. We use the full NLSY, also including the supplemental black and Hispanic samples, because the method used in the analysis is non-parametric and requires sufficient data to avoid empty cells. The supplemental samples are also necessary to have large enough sample sizes in the analyses that stratify by race.[7]

As outcomes in our analysis, we use individual's highest observed years of education as well as yearly wage income both reported in 1994, when the individuals were in their early 30s.[8,9] Information on Head Start participation was also collected in 1994, when respondents were asked whether they attended the Head Start program as a child, as well as whether they attended any type of preschool.[10]

We restrict the main sample to Head Start participants and individuals who did not

[7] We estimate lower bounds using sample weights in Section 6 below.

[8] In 1994 the respondents were between 30 and 34 years old.

[9] For each of the survey years information about the highest completed grade is available. We use the maximum of the reported highest completed grade over the years 1979-1994 as our measure of years of education. Yearly wage income is measured by the question "During 1993, how much did you receive from wages, salary, commissions, or tips from all (other) jobs, before deductions for taxes or anything else?".

[10] The actual Head Start question asked "Now think back to when you were a child. To your knowledge, did you ever attend a Head Start program when you were a preschooler?". In Section 6 below we show that self-reported Head Start participation is broadly consistent with historical data, but that there is probably some underreporting. We also show that misreporting is unlikely to seriously bias our estimates.

Table 1. Descriptive statistics

	All	Head Start		Race		
		Yes	No	White	Black	Hispanic
Head Start	0.23			0.08	0.49	0.21
Age	32.1	32.0	32.1	32.1	32.1	32.0
Female	0.50	0.52	0.50	0.49	0.51	0.51
Race:						
- White	0.49	0.16	0.59			
- Black	0.31	0.66	0.21			
- Hispanic	0.20	0.17	0.20			
Parental Education:						
- Less than High School	0.21	0.26	0.19	0.10	0.19	0.50
- Some High School	0.15	0.22	0.13	0.11	0.25	0.11
- High School	0.40	0.38	0.41	0.47	0.40	0.24
- College (1-3 years)	0.12	0.07	0.13	0.14	0.09	0.08
- College (4+ years)	0.12	0.07	0.14	0.18	0.06	0.06
Family income 1978	16,303	11,603	17,759	21,096	10,946	13,077
Years of education	12.8	12.6	12.8	13.1	12.6	12.1
Wage income	22,633	19,637	23,456	25,226	19,057	20,790
N	4,876	1,132	3,744	2,404	1,518	954

Note: Sample sizes for wage income are: 3,781; 815; 2,966; 1,985; 1,060 and 736.

participate in Head Start nor any type of preschool. This means that in the main analysis we estimate effects of Head Start relative to informal care and not relative to other types of preschool. We also show results where we include individuals who attended another type of preschool in the estimation sample.

Basic background information such as age (birth year), gender and race is available in the data. The respondents also provided information on parental education. Since education is more often missing for the father than for the mother, the main analysis uses the highest reported completed grade of either the mother or father as a measure of parental education which is recoded in the following categories: less than high school, some high school, high school, 1–3 years of college and 4 or more years of college.

Table 1 reports descriptive statistics on the variables that we use below. First, about one out of four respondents in our sample attended Head Start. The average respondent was 32 years old in 1994. Thirty-one percent of respondents is black, 20 percent is Hispanic, and the remaining half is white. About 20 percent of the individuals in our dataset have parents whose highest completed education is less than high school, while 15 percent of parents attended and 40 percent completed high school. Of the remaining 24 percent of parents with some college education, half completed 4 years or more.

The final two rows of Table 1 report years of education and yearly wage income (in 1994 USD). We see that by 1994 respondents had attained, on average, about 13 years of education, or slightly more than high school. Reported wage income is on average about 23,000 USD.[11]

4 Empirical approach

4.1 Non-parametric bounds

Let $Y_i(h)$ be individual i's potential outcome if her Head Start status is h, where $h = 1$ if she participates in Head Start and $h = 0$ otherwise. Let D_i equal 1 if individual i actually participated in Head Start and equal 0 otherwise. The link between the observed outcome Y and the potential outcomes is given by $Y_i \equiv Y_i(1) \cdot D_i + Y_i(0) \cdot (1 - D_i)$.

Many studies focus on estimating a specific parameter of the potential outcome distributions, such as the mean. Instead, we focus on the entire cumulative distribution of potential education and labor market outcomes. The causal effect of interest is then the effect of Head Start participation on the probability of obtaining an education or labor market outcome greater than γ:[12]

$$\triangle(\gamma) = \Pr(Y(1) > \gamma) - \Pr(Y(0) > \gamma) = F_{Y(0)}(\gamma) - F_{Y(1)}(\gamma) \tag{1}$$

We estimate (1) for values of γ over the whole support of $Y(h)$.

The causal effect is the difference between two cumulative potential outcome distri-

[11] Sample size is smaller for wage income which is mostly due to non-employment.
[12] To economize on notation we omit the individual subscript i from hereon.

bution functions (CDFs); the CDF we would observe with no Head Start as potential treatment, $F_{Y(0)}(\gamma)$, and the CDF we would observe with Head Start as potential treatment, $F_{Y(1)}(\gamma)$. By using the law of iterated expectations we can decompose these two cumulative potential outcome distributions as follows

$$F_{Y(1)}(\gamma) = F(\gamma|D=1) \cdot \Pr(D=1) + F_{Y(1)}(\gamma|D=0) \cdot \Pr(D=0) \tag{2}$$

$$F_{Y(0)}(\gamma) = F(\gamma|D=0) \cdot \Pr(D=0) + F_{Y(0)}(\gamma|D=1) \cdot \Pr(D=1) \tag{3}$$

Equations (2) and (3) highlight the identification problem; we observe the cumulative outcome distributions for Head Start participants, $F(\gamma|D=1)$, and for non-participants, $F(\gamma|D=0)$. We also observe the proportion of participants, $\Pr(D=1)$, and non-participants $\Pr(D=0)$. However, we do not observe the cumulative potential outcome distribution for the participants had they not participated in Head Start, $F_{Y(0)}(\gamma|D=1)$, nor the cumulative potential outcome distribution for the non-participants had they participated in Head Start, $F_{Y(1)}(\gamma|D=0)$.

The starting point of our analysis is based on a simple fact: CDFs are bounded between 0 and 1. We can therefore replace the unobserved cumulative potential outcome distributions, $F_{Y(1)}(\gamma|D=0)$ and $F_{Y(0)}(\gamma|D=1)$, by 0 to get lower bounds and by 1 to get upper bounds on $F_{Y(1)}(\gamma)$ and $F_{Y(0)}(\gamma)$. This implies that we can obtain the following bounds without adding assumptions (Manski, 1989, 1990):

$$F(\gamma|D=1) \cdot \Pr(D=1) \leq F_{Y(1)}(\gamma) \leq F(\gamma|D=1) \cdot \Pr(D=1) + \Pr(D=0) \tag{4}$$

$$F(\gamma|D=0) \cdot \Pr(D=0) \leq F_{Y(0)}(\gamma) \leq F(\gamma|D=0) \cdot \Pr(D=0) + \Pr(D=1) \tag{5}$$

To further tighten these No-Assumption (NOA) bounds we continue by imposing two nonparametric weak stochastic dominance assumptions, proposed by Manski (1997); Manski and Pepper (2000), which we discuss in turn.

The first assumption is a Monotone Instrumental Variable (MIV) assumption, which is a weak stochastic dominance assumption with respect to potential outcome distributions as a function of a so-called monotone instrumental variable. We use the maximum level of parental education as a monotone instrumental variable:

Assumption 1. *Monotone Instrumental Variable (MIV) – The potential outcome distributions of children with parents of a given education level are weakly stochastically dominated by those of children with more educated parents:*

$$F_{Y(h)}(\gamma|X=x_2) \leq F_{Y(h)}(\gamma|X=x_1) \quad \forall\gamma, \forall h, \forall x_2 > x_1 \tag{6}$$

The MIV assumption states that if everyone would receive the same treatment – either Head Start ($h=1$) or no Head Start ($h=0$) – then the probability of obtaining at most γ years of education would, on average, not be higher for individuals with high educated parents ($X=x_2$) compared to individuals with low educated parents ($X=x_1$). Note that, unlike classical IV estimation, this allows for a direct effect of parents' level of education on the potential education and labor market outcomes as long as this effect is not negative.

We can exploit this weak stochastic dominance assumption to tighten the No-Assumption bounds in the following way. We first compute upper and lower bounds on the cumulative potential outcome distributions $F_{Y(h)}(\gamma|X=x)$ for each level of parent's education x. Under the MIV assumption $F_{Y(h)}(\gamma|X=x^*)$ is no lower than any of the lower bounds on $F_{Y(h)}(\gamma|X=x)$ for all $x > x^*$. We can therefore obtain the MIV lower bound on $F_{Y(h)}(\gamma|X=x^*)$ by taking the maximum of the lower bounds on $F_{Y(h)}(\gamma|X=x)$ for $x \geq x^*$. Similarly we can obtain the MIV upper bound on $F_{Y(h)}(\gamma|X=x^*)$ by taking the minimum of the upper bounds on $F_{Y(h)}(\gamma|X=x)$ for $x \leq x^*$.

Suppose parent's level of education can take on three values; low middle and high. Figure 2 shows illustrative upper and lower bounds around the cumulative distribution of a potential education or labor market outcome for a sample of individuals with middle educated parents, $F_{Y(h)}(\gamma|X=mid)$. Under the MIV assumption $F_{Y(h)}(\gamma|X=mid) \leq F_{Y(h)}(\gamma|X=low)$ which implies that $F_{Y(h)}(\gamma|X=mid)$ should also be smaller than the upper bound on $F_{Y(h)}(\gamma|X=low)$. If the upper bound on $F_{Y(h)}(\gamma|X=low)$ is more

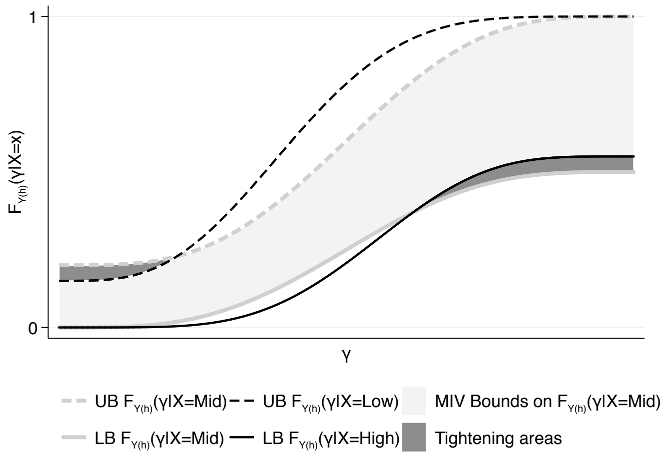


Figure 2. Example of how an MIV can tighten the bounds

informative (and thus smaller) than the upper bound on $F_{Y(h)}(\gamma|X = mid)$ then we can tighten the upper bound on $F_{Y(h)}(\gamma|X = mid)$ by replacing it by the upper bound on $F_{Y(h)}(\gamma|X = low)$. In Figure 2 this happens for low values of γ and the dark shaded area shows where the bounds on $F_{Y(h)}(\gamma|X = mid)$ become sharper.

Under a similar reasoning, we can use the lower bound on $F_{Y(h)}(\gamma|X = high)$ to tighten the lower bound on $F_{Y(h)}(\gamma|X = mid)$. By the MIV assumption $F_{Y(h)}(\gamma|X = mid) \geq F_{Y(h)}(\gamma|X = high)$ which implies that $F_{Y(h)}(\gamma|X = mid)$ should also be higher than the lower bound on $F_{Y(h)}(\gamma|X = high)$. Figure 2 illustrates this tightening on the lower bound of $F_{Y(h)}(\gamma|X = mid)$ for high values of γ and the corresponding dark shaded area shows where this bound is sharpened. Finally note that for the lowest value of X the MIV can only sharpen the lower bound, while for the highest value of X the MIV can only sharpen the upper bound.

By applying the logic illustrated in Figure 2 to the bounds on each $F_{Y(h)}(\gamma|X = x^*)$ and then taking the weighted average of the MIV bounds over all $x^* \in X$, we obtain the following aggregate MIV-bounds on $F_{Y(h)}(\gamma)$.

$$\sum_{x^* \in X} \left(\max_{x \geq x^*} \mathrm{LB}_{F_{Y(h)}(\gamma|X=x)} \right) \Pr(X = x^*)$$
$$\leq F_{Y(h)}(\gamma) \leq$$
$$\sum_{x^* \in X} \left(\min_{x \leq x^*} \mathrm{UB}_{F_{Y(h)}(\gamma|X=x)} \right) \Pr(X = x^*) \quad \forall \gamma, h \quad (7)$$

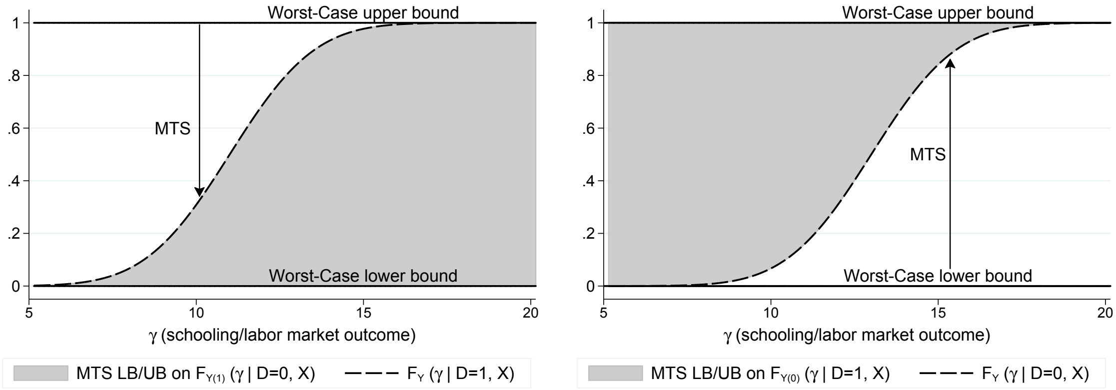
The second weak stochastic dominance assumption that we use to tighten the bounds is the Monotone Treatment Selection assumption, which is motivated by the eligibility criteria of Head Start as described in Section 2. Equation (8) shows the MTS assumption.

Assumption 2. *Monotone Treatment Selection (MTS) – The distribution of potential outcomes of Head Start participants are weakly stochastically dominated by those of non-participants:*

$$F_{Y(h)}(\gamma|D = 0, X) \leq F_{Y(h)}(\gamma|D = 1, X) \quad \forall \gamma, h \qquad (8)$$

The MTS assumption implies that if all individuals would receive the same treatment – either Head Start ($h = 1$) or no Head Start ($h = 0$) – the probability of obtaining an education or labor market outcome smaller or equal than some value γ should, on average, be weakly higher for the participants ($D = 1$) compared to the non-participants ($D = 0$). Note that for the MTS assumption to hold it is *not* required that for *each* Head Start participant the potential probability of obtaining an education or labor market outcome smaller or equal than γ is weakly higher than this potential probability for *any* of the non-participants, instead this should hold *on average*.

Figure 3 illustrates how this MTS assumption can be used to tighten the bounds. Panel (a) shows how to tighten the bounds around the cumulative potential outcome distribution in case of Head Start as potential treatment for non-participants; $F_{Y(1)}(\gamma|D = 0, X)$. All we know without imposing additional assumptions is that it lies between the worst-case lower- and upper-bounds 0 and 1. However, under the MTS assumption the potential outcome



(a) Non-participants: Bounds around $F_{Y(1)}(\gamma|D=0,X)$

(b) Participants: Bounds around $F_{Y(0)}(\gamma|D=1,X)$

Figure 3. Illustration of the MTS assumption

distribution of non-participants weakly stochastically dominates the potential outcome distribution of the participants. This means that we can use the observed cumulative distribution of the participants, $F_Y(\gamma|D=1,X)$, as an upper bound on the unobserved cumulative potential outcome distribution for the non-participants, $F_{Y(1)}(\gamma|D=0,X)$. Panel (b) shows that under a similar reasoning we can use the observed cumulative distribution of the non-participants, $F_Y(\gamma|D=0,X)$ as a lower bound on the unobserved cumulative potential outcome distribution for the participants, $F_{Y(0)}(\gamma|D=1,X)$. Equation (9) show these MTS bounds.

$$
\begin{aligned}
&F_Y(\gamma|D=1,X)\cdot \Pr(D=1|X) \leq F_{Y(1)}(\gamma|X) \leq F_Y(\gamma|D=1,X) \\
&\qquad\qquad F_Y(\gamma|D=0,X) \leq F_{Y(0)}(\gamma|X) \leq F_Y(\gamma|D=0,X)\cdot \Pr(D=0|X)+\Pr(D=1|X)
\end{aligned}
\tag{9}
$$

In the analysis we combine the MTS and MIV assumptions by first calculating MTS upper and lower bounds on $F_{Y(h)}(\gamma|X)$ for each level of parents' education and then use these in equation (7) to obtain the combined MTS-MIV bounds. This implies that the MTS assumption should hold conditional on the level of parents' education X.

So far we used the MTS and MIV assumptions to tighten the bounds around the two cumulative potential outcome distribution functions, $F_{Y(1)}(\gamma)$ and $F_{Y(0)}(\gamma)$. To obtain a lower bound on the causal effect, $\triangle(\gamma) = F_{Y(0)}(\gamma) - F_{Y(1)}(\gamma)$, we subtract the upper bound on $F_{Y(1)}(\gamma)$ from the lower bound on $F_{Y(0)}(\gamma)$.[13]

While all bounds are consistent under the maintained assumptions, they may have finite-sample biases when they are obtained by taking maxima and minima over collections of nonparametric estimates. All bounds using the MIV-assumption are therefore corrected for finite sample bias using the bootstrap following Kreider and Pepper (2007).[14] Finally we use the methods from Imbens and Manski (2004) to obtain 90% and 95% confidence intervals around the bounds based on 999 bootstrap replications.[15]

4.2 *Combining two monotone instrumental variables*

The MIV-assumption described in Assumption 1 combines the education of the father and the mother in one monotone instrumental variable by taking the highest reported completed grade of either the mother or the father. We also report results where we use the highest reported completed grade of both the mother (X^M) and the father (X^F) as two separate MIV's, both recoded in the following 3 categories: less than high school, high school and more than high school. In this case we use the following semi-monotone instrumental

[13]The upper bounds on the causal effects are never small enough to be informative.

[14]Kreider and Pepper (2007) suggest to estimate the finite sample bias as $\widehat{\text{bias}} = \left(\frac{1}{K}\sum_{k=1}^{K}\theta_k\right) - \widehat{\theta}$, where $\hat{\theta}$ is the initial estimate of the upper or lower bound and θ_k is the estimate of the k^{th} bootstrap replication. The bias-corrected MIV-bounds are subsequently obtained by subtracting the estimated biases from the estimated upper and lower bounds.

[15]Equation (10) gives their formula for a 95-percent confidence interval:

$$CI_{0.95} = \left(\widehat{lb} - c_{IM}\cdot\hat{\sigma}_{lb},\ \widehat{ub} + c_{IM}\cdot\hat{\sigma}_{ub}\right) \tag{10}$$

where $\widehat{lb}$ and $\widehat{ub}$ are the estimated upper and lower bounds and $\hat{\sigma}_{lb}$ and $\hat{\sigma}_{ub}$ are the estimated standard errors of the estimated lower and upper bounds, obtained by 999 bootstrap replications. The parameter c_{IM} depends on the width of the bounds and is obtained by solving equation (11).

$$\Phi\left(c_{IM} + \frac{\left(\widehat{ub} - \widehat{lb}\right)}{\max\left\{\hat{\sigma}_{lb}, \hat{\sigma}_{ub}\right\}}\right) - \Phi(-c_{IM}) = 0.95 \tag{11}$$

variable assumption

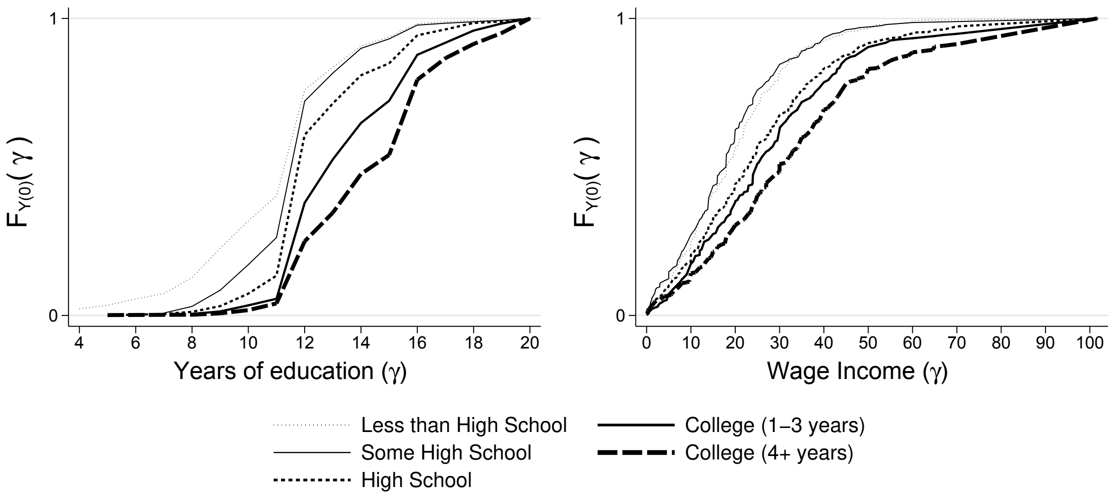
$$F_{Y(h)}\left(\gamma|X^M=x_2^M,X^F=x_2^F\right)\leq F_{Y(h)}\left(\gamma|X^M=x_1^M,X^F=x_1^F\right)$$
$$\forall\gamma,\ \forall h,\ \forall x_2^M\geq x_1^M \text{ and } x_2^F\geq x_1^F \tag{12}$$

The MIV assumption states that if everyone would receive the same treatment – either Head Start ($h=1$) or no Head Start ($h=0$) – then the probability of obtaining at most γ years of education would, on average, not be higher for individuals with a high educated father and a high educated mother compared to individuals whose mother, father or both parents have a lower education level. The assumption states nothing about the stochastic dominance of the potential outcome distributions if we compare individuals who have a high educated mother and a low educated father with individuals who have a high educated father and a low educated mother. The computation of the bounds using two monotone instruments is very similar to the MIV bounds in equation (7) except that the maxima and minima are taken over pairs of values of father's and mother's education that are ordered.

4.3 Assumption check

The MIV assumption The MTS and MIV assumptions are untestable since they involve counterfactual outcomes that are not observed for everyone. However, since the pre-Head Start cohorts in the NLSY79 (i.e. those born from 1957-1959) did not have the opportunity to enroll in Head Start, the counterfactual outcome without Head Start ($Y(0)$) is observed for all these individuals. This allows us to check whether the weak stochastic dominance assumption of our MIV holds in this sample of pre-Head Start cohorts.

Figure 4 plots the cumulative distribution functions of the long-term outcomes we consider – education and wage income – by parental education. The distribution functions need to be weakly ordered for Assumption 1 to hold, with those of individuals with more educated parents shifted uniformly to the right compared to those of individuals with less educated parents. The left panel shows these cumulative distributions for years of education. As can be seen in the figure, there is a clear and strict ordering, which is



Note: Figures are based on data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations equal 4873 (education) and 2153 (wage income).

Figure 4. MIV Check – Stochastic dominance of outcomes among pre-Head Start cohorts

consistent with our MIV assumption. The right panel shows the results for wage income. The cumulative distribution functions of individuals with parents who attained less than or some high school overlap, and the first column of Table 2 shows we cannot reject that they are equal using a one-side Kolmogorov-Smirnov test (McFadden, 1989). Note that this is consistent with our MIV assumption since that only requires weak first order stochastic dominance. The remaining distribution functions show again strict first-order stochastic dominance and are therefore consistent with the MIV assumption.

When we estimate bounds on the effect of Head Start separately by gender and by race and MIV assumption also needs to hold conditional on gender and race. Figures A1 and A2 in the appendix show the cumulative distributions of education and wage income for the pre-Head start cohorts for each level of the MIV separately for men, women, blacks, whites and Hispanics. Although not all distributions show a strict stochastic dominance ordering, the Kolmogorov-Smirnov tests in Table 2 show that for none of the sub-samples the null hypothesis is rejected, which is consistent with the validity of the MIV assumption conditional on gender and conditional on race.

Figures A6 and A7 and Table A1 in the appendix show the MIV-assumption check described in Section 4.3 for the case of two MIV's. For years of education as outcome we

Table 2. Test of the conditional MIV assumption — p-values for $\mathscr{H}_0 : F_j = F_{j-1}$ vs $\mathscr{H}_1 : F_j > F_{j-1}$

	Sample					
	All	Men	Women	White	Black	Hispanic
A. Education, j :						
2- Some High School	1.000	1.000	0.944	0.978	1.000	0.986
3- High School	1.000	0.998	1.000	1.000	1.000	1.000
4- College (1-3 years)	1.000	1.000	1.000	1.000	0.991	1.000
5- College (4+ years)	0.999	1.000	0.998	0.999	1.000	0.964
B. Wage Income, j :						
2- Some High School	0.229	0.132	0.822	0.545	0.648	0.679
3- High School	0.999	0.999	0.984	0.999	0.999	0.980
4- College (1-3 years)	0.996	0.884	0.873	0.995	0.498	0.291
5- College (4+ years)	0.835	0.978	0.611	0.583	0.993	0.936

Note: Reported p-values are from one sided Kolmogorov-Smirnov tests, using data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations for education equal 4,873 (all), 2,425 (men), 2,448 (women), 3,172 (white), 1,044 (black), 657 (Hispanic). Number of observations for wage income equal 2,153 (all), 1,099 (men), 1,054 (women), 1,189 (white), 582 (black), 382 (Hispanic).

observe a strict ordering with the cumulative distributions of those with higher educated fathers/ mothers shifted uniformly to the right. For wage income we do not always observe this strict ordering, but the one-sided Kolmogorov-Smirnov tests in Table A1 show that for none of the sub-samples the null hypothesis is rejected, which is consistent with the validity of the two-MIV assumption.

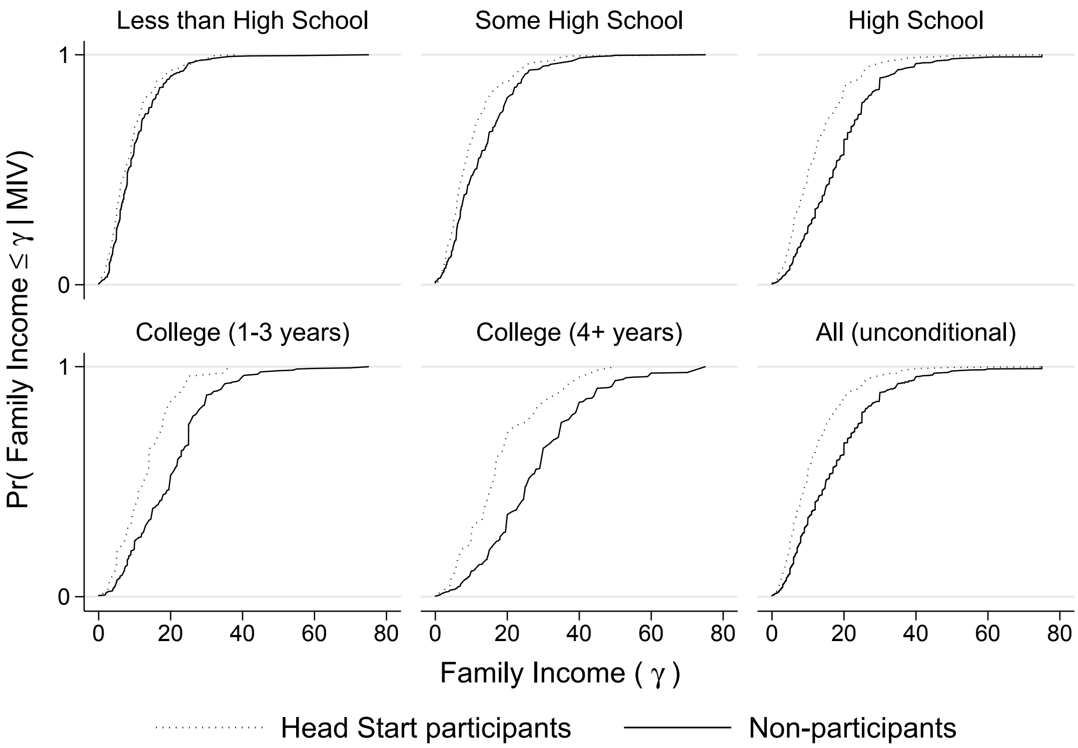
The MTS assumption The main motivation for using the MTS assumption ultimately comes from the eligibility rules which make that Head Start participants come disproportionally from disadvantaged backgrounds. Schnur et al. (1992) study selection into Head Start and their findings support the validity of the MTS assumption. They use data from the Educational Testing Service Head Start Longitudinal Study (HSLS) which followed over 1,300 children living in poor neighborhoods in three regions in the U.S, during 1969-1970 and preceding possible Head Start enrollment. Children who ultimately attended Head Start were at a disadvantage on virtually every background familial characteristics and cognitive measure compared to both i) children who ultimately did not attend preschool and ii) children who attended other preschools. Head Start participants were also less cognitively advanced than children attending other preschools and similar to children attending no preschool conditional on race, site, and family characteristic variables.

If not all eligible children enroll in Head Start then it is theoretically possible that on some margin the average potential outcome for the non-participants compared to the participants would violate the MTS assumption. This requires two things. First, selection into Head Start conditional on eligibility must be positive. Second, such negative selection out of Head Start among the eligibles must be large enough to reverse the overall positive selection out of Head Start through eligibility.

While we argue that the non-participating eligible children will typically be a small share of the non-participants (making reversion of the MTS assumption unlikely), we do not have data on eligibility to verify this. However, Schnur et al. also show that conditional on eligibility selection into Head Start is negative and not positive. In particular, for the eligible children they find that those "who attended Head Start had significantly lower cognitive scores, had mothers with lower education, and had fewer rooms per person than those who attended no preschool. Father absence and maternal education expectations, although lower in the Head Start group, were not significantly different than in the no preschool group".

Both the eligibility rules and the evidence of Schnur et al. (1992) therefore support the (conditional) MTS assumption. In addition we can investigate the validity of the MTS in our data by checking whether background characteristics of the Head Start participants are indeed weakly stochastically dominated by those of non-participants for the different sub-samples in which the MTS must hold. Figure 5 shows cumulative distributions of family income measured in 1978 when the individuals were between 14 and 18 years old.[16] For each of the values of the MIV the distribution of family income for the Head

[16]Family income could potentially be used as a MIV, but we do not do this for the following reasons. First, information on family income is not available when the individuals are of preschool age, it is only collected from 1978 and onwards. In addition, eligibility is determined by family income which implies that there are no or very few Head Start participants for certain values of a MIV that is based on family income. Finally, the MTS assumption should hold conditional on the MIV, which we think is a stronger assumption when using family income as MIV compared to using parental education as MIV.



Note: Number of observations equals 861 (less than high school), 614 (some high school), 1,619 (high school), 473 (college 1-3 years), 461 (college 4+ years) and 4028 (all).

Figure 5. MTS Check – Conditional (on MIV) CDFs of Family income at age 14-18 for Head Start participants and non-participants

Table 3. Test of MTS assumption — p-values for $\mathscr{H}_0 : F_{j,h=0} = F_{j,h=1}$ vs $\mathscr{H}_1 : F_{j,h=0} > F_{j,h=1}$

	Sample					
	All	Men	Women	White	Black	Hispanic
j:						
1- Less than High School	0.978	0.985	0.921	0.550	0.651	0.303
2- Some High School	0.875	0.914	0.953	0.872	0.344	0.868
3- High School	1.000	1.000	1.000	1.000	0.970	0.941
4- College (1-3 years)	0.995	0.995	0.957	0.940	0.832	0.975
5- College (4+ years)	0.997	0.999	0.960	0.962	0.966	0.718
Unconditional	0.999	0.984	1.000	0.997	0.944	0.845

Note: Reported p-values are from one sided Kolmogorov-Smirnov tests, using data on family income in 1978 for the Head Start cohorts (born between 1960–1965). Number of observations equal 4,028 (all), 2,018 (men), 2,010 (women), 1,957 (white), 1,268 (black), 803 (Hispanic)

Start participants is stochastically dominated by the distribution of non-participants, which is in line with the MTS assumption. The first column in Table 3 shows indeed that the assumption that the distribution of family income of the Head Start participants is weakly stochastically dominated by that of the non-participants is not rejected at conventional significant levels. Figures A3 and A4 in the appendix report the cumulative distributions of family income for the participants and non-participants, separately by gender and by race. Although in some sub-samples there is no strict stochastic dominance for some of the values of the MIV, Table 3 shows that the null hypothesis is not rejected in any of the sub-samples, which implies that we do not reject the MTS assumption conditional on gender nor conditional on race.

Although not complete, the evidence of Schnur et al. (1992) as well as the checks in Tables 2 and 3 all support our identifying assumptions.

5 The effects of Head Start on long term outcomes

5.1 A simple example: The effect of Head Start on high school graduation

Before we present our main results under the combined MTS-MIV assumption we set out to illustrate how the MIV, the MTS and the combined MTS-MIV assumption tighten the bounds and to clarify which of the assumptions has most identifying power in our data. We do this for the average treatment effect (ATE) of Head Start on the probability of high school graduation:

$$ATE = E\left[HS(1)\right] - E\left[HS(0)\right]$$

where $HS(h)$ equals one if someone completes high school under treatment h and is zero otherwise.

To estimate the average causal effect we need to estimate the mean potential high

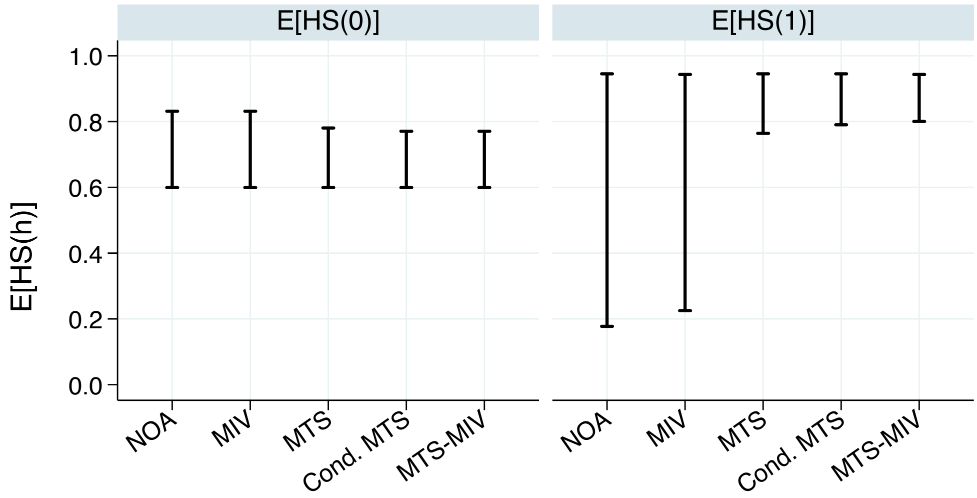


Figure 6. Bounds around the mean potential probabilities of high school graduation

school completion rate $E[HS(h)]$ with Head Start ($h=1$) and without Head Start ($h=0$)

$$E[HS(h)] = E[HS(h)|D=h]\Pr(D=h) + \underbrace{E[HS(h)|D=1-h]}_{\text{unobserved}}\Pr(D=1-h)$$

which involves the unobserved mean counterfactual high school completion rate for Head Start participants ($D=1$) and non-participants ($D=0$). Since high school completion rates are bounded between zero and one, so are the mean counterfactual high school completion rates $E[HS(h)|D=1-h]$. This gives the following No-Assumption (NOA) bounds

$$E[HS|D=h]\cdot\Pr(D=h) \leq E[HS(h)] \leq E[HS|D=h]\cdot\Pr(D=h) + \Pr(D=1-h) \quad (13)$$

which are shown by the first vertical bars in the panels of Figure 6.

Next we can exploit the MIV assumption which implies that if everybody were to be assigned to the same Head Start treatment then the high school completion rate would *on average* not be lower for children whose parents are more educated ($X=x_2$) than the high school completion rate of children whose parents have less education ($X=x_1$):

$$E[HS(h)|X=x_2] \geq E[HS(h)|X=x_1] \qquad \forall x_2 > x_1, h=0,1 \quad (14)$$

To tighten the bounds using the MIV assumption we start out by computing the No-Assumption bounds for the two mean potential outcomes separately by parental education. These bounds are shown by the gray vertical bars in the top two panels of Figure 7. Because the MIV assumption implies that the average potential probability of high school graduation is nondecreasing in parents' education, the MIV lower bound for a given level of parental education $X = x$ is obtained by taking the maximum over all the No-Assumption lower bounds where the level of parents' education is less than x. The top-right panel in Figure 7 shows we can tighten the lower bounds around $E\left[HS(1)\,|X = x\right]$ for the three highest levels of parents' education by the lower bound for children whose parents have some high school.

The MIV upper bounds are obtained in a similar fashion, but now by taking the minimum over all upper bounds in the sub-samples where parents' level of education is higher or equal to the level in the particular sub-sample. As can be seen in the top two panels in Figure 7, in this particular case the MIV assumption does not result in tighter bounds around $E\left[HS(0)\,|X = x\right]$ for any level of parents' education (x). The MIV bounds around the unconditional potential outcomes – the second set of vertical bars in the panels of Figure 6 – are obtained by averaging the conditional-MIV bounds over the distribution of parental schooling.

We can also construct bounds using the MTS assumption which assumes that *on average* Head Start participants do not have higher potential high school graduation rates than non-participants. Above we bounded the counterfactual high school graduation rate for non-participants $E[HS(1)|D = 0]$ from below by zero. Because the MTS assumes that non-participants on average do not do worse than participants we can now use the average high school graduation rate of participants $E[HS|D = 1]$ as a lower bound instead. Similarly, where before we bounded the counterfactual high school graduation rate for participants $E[HS(0)|D = 1]$ from above by one, we can now use the observed high school graduation rate of non-participants $E[HS|D = 0]$ as an upper bound. The third set of vertical bars in Figure 6 shows that the unconditional MTS assumption substantially tightens the

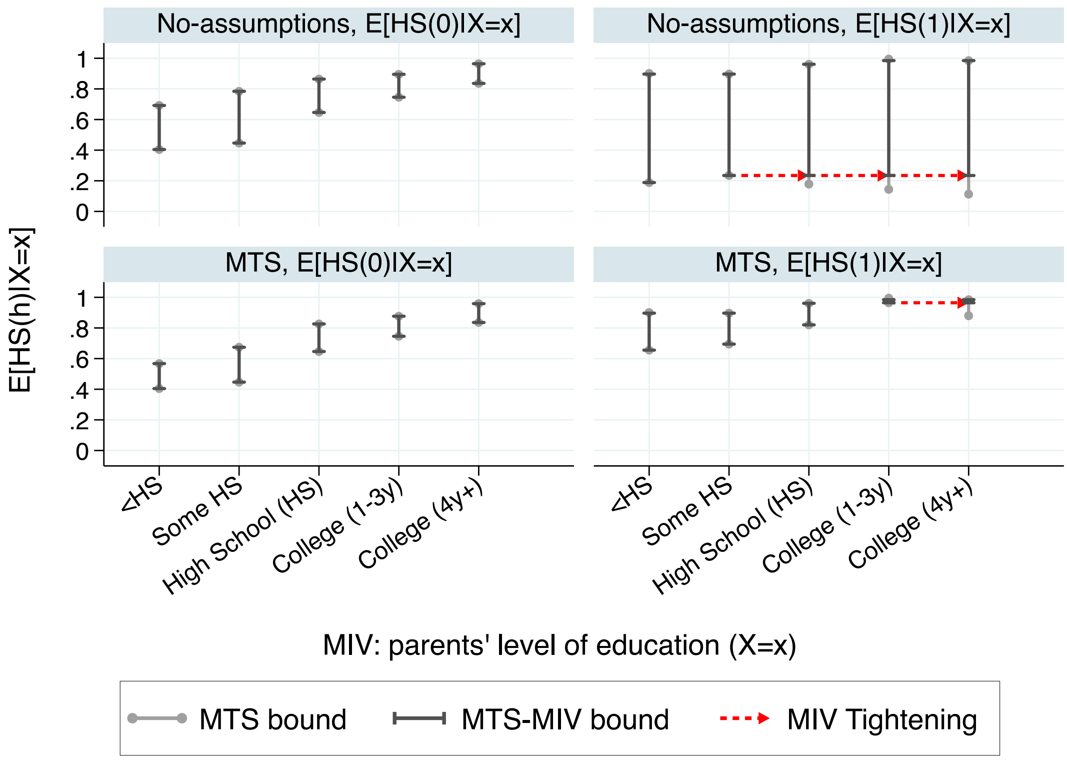


Figure 7. Bounds on the mean potential probabilities of high school graduation by MIV

bounds around the two mean potential high school graduation rates.

We can also impose the MTS and MIV assumptions simultaneously. In this case we first construct the MTS bounds around the two mean potential outcomes separately for each subsample defined by the MIV and then use these to construct the MIV bounds. In this case the MTS assumption must hold conditional on parents' level of education. The vertical bars in the bottom two panels of Figure 7 show the MTS bounds around the average potential high school graduation rate conditional on parental education.

Comparing the top and bottom panels of Figure 7 shows that the identifying power of the MTS assumption varies with parent's level of education. Especially for $E\left[HS(1)\,|X=x\right]$ we see that the tightening of the bounds by the MTS assumption is much stronger for higher levels of parents' education. The reason becomes clear if we compare the No-Assumption lower bound on $E\left[HS(1)\,|X=x\right]$, $\left(E\left[HS\,|D=1,X=x\right]\cdot\Pr\left(D=1|X=x\right)\right)$ with the MTS lower bound $\left(E\left[HS\,|D=1,X=x\right]\right)$. The difference between the No-Assumption lower bound and the MTS lower bound depends on the the share of Head Start participants in the subsample defined by parents' level of education, $\Pr(D=1|X=x)$. Since the share of participants varies with parents' level of education we see that the identifying power of the MTS assumption varies with the values of our MIV and this is the first reason for obtaining tighter bounds when combining the MTS and MIV assumptions. If we next take the weighted average over the sub-sample MTS bounds we get conditional MTS bounds around $E[HS(1)]$ and $E[HS(0)]$, which are shown by the fourth set of vertical bars in Figure 6.

We can tighten the bounds further by exploiting the MIV assumption to obtain MTS-MIV upper and lower bounds on $E[HS(1)|X=x]$ and $E[HS(0)|X=x]$. These MTS-MIV bounds are shown by the black capped bars in the bottom two panels of Figure 7 and are obtained by taking the maximum over all MTS lower bounds where the level of parents' education is lower or equal to the level in the particular sub-sample and the MIV upper bounds are obtained by taking the minimum over all MTS upper bounds in the sub-samples where parents' level of education is higher or equal to the level in the particular sub-sample. The dashed lines with arrows in Figure 7 indicate where this tightening occurs.

If we next take the weighted average over these sub-sample MTS-MIV bounds we get the MTS-MIV bounds around $E[HS(1)]$ and $E[HS(0)]$, which are shown by the final set of vertical bars in Figure 6. If we compare the MTS, conditional MTS, and the MTS-MIV bounds in Figure 6, we can see that both steps in the process of combining the MTS and MIV assumptions have identifying power in the sense that the conditional MTS bounds are tighter than the unconditional MTS bounds and the MTS-MIV bounds are tighter than the conditional MTS bounds.

Equation (15) shows how we can obtain bounds around the parameter of interest, the average causal effect of Head Start on high school graduation

$$LB_{E[HS(1)]} - UB_{E[HS(0)]} \leq E[HS(1)] - E[HS(0)] \leq UB_{E[HS(1)]} - LB_{E[HS(0)]} \tag{15}$$

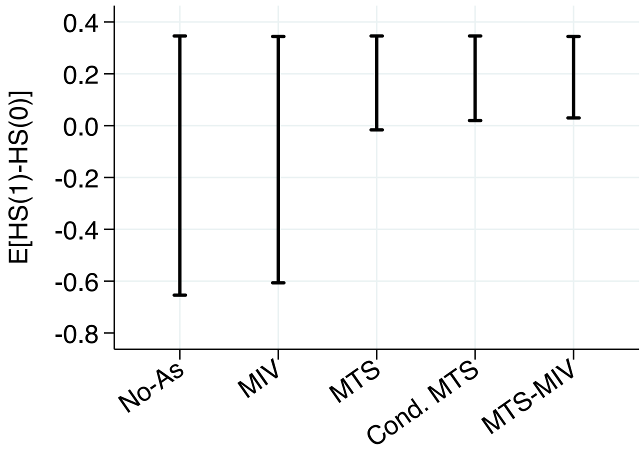


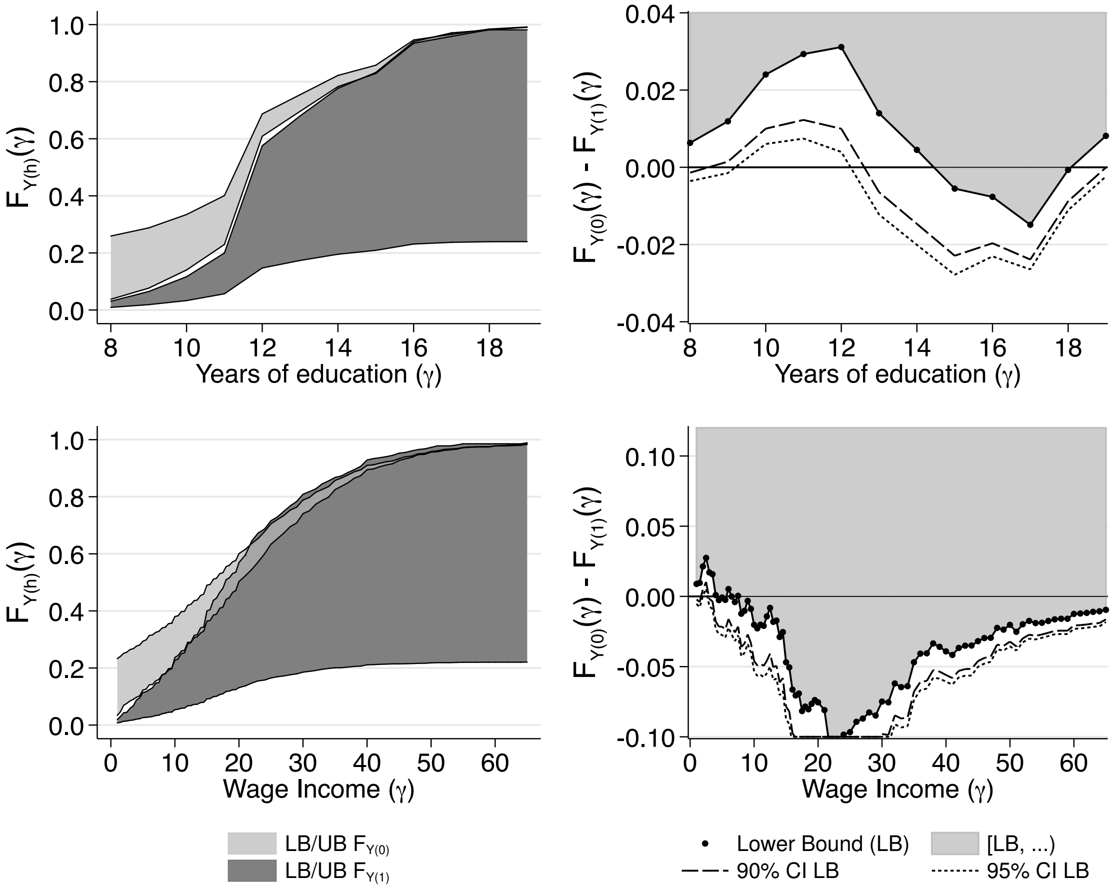
Figure 8. Bounds around the ACE of Head Start on high school graduation

Figure 8 displays these bounds around the ACE. The tightest bounds, obtained by combining the MTS and MIV assumptions, shows that Head Start participation increases the probability of high school graduation by at least 3 and at most 34 percentage points. These bounds on the ACE are not corrected for potential finite sample bias and Figure 8 also does not report confidence intervals. The bias-corrected MTS-MIV lower bound on $E[HS(1)] - E[HS(0)]$ as well as the lower bound of the 95 percent confidence interval are shown in Figure 9 below at $\gamma = 11$ (since $E[HS(1)] - E[HS(0)] = F_{Y(0)}(11) - F_{Y(1)}(11)$). Bias correction leaves the bounds essentially unchanged. The average causal effect of Head Start participation on high school graduation is significantly different from zero as the lower bound of the 95 percent confidence interval equals 0.01.

5.2 *Overall effects*

The top left panel of Figure 9 shows the MTS-MIV bounds on the cumulative potential outcome distribution of education for the main sample.[17] The light gray area bounds the cumulative potential outcome distribution without Head Start ($F_{Y(0)}(\gamma)$), while the dark gray area bounds the cumulative potential outcome distribution with Head Start ($F_{Y(1)}(\gamma)$). This figure shows that the bounds are informative in the sense that there are points on the

[17] In Figure A8 in the appendix we show results where we use no assumptions, only the MTS and only the MIV assumption.



Note: Number of observations equals 4876 (years of education) and 3787 (wage income). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 9. MTS-MIV bounds on the effect of Head Start on education and earnings

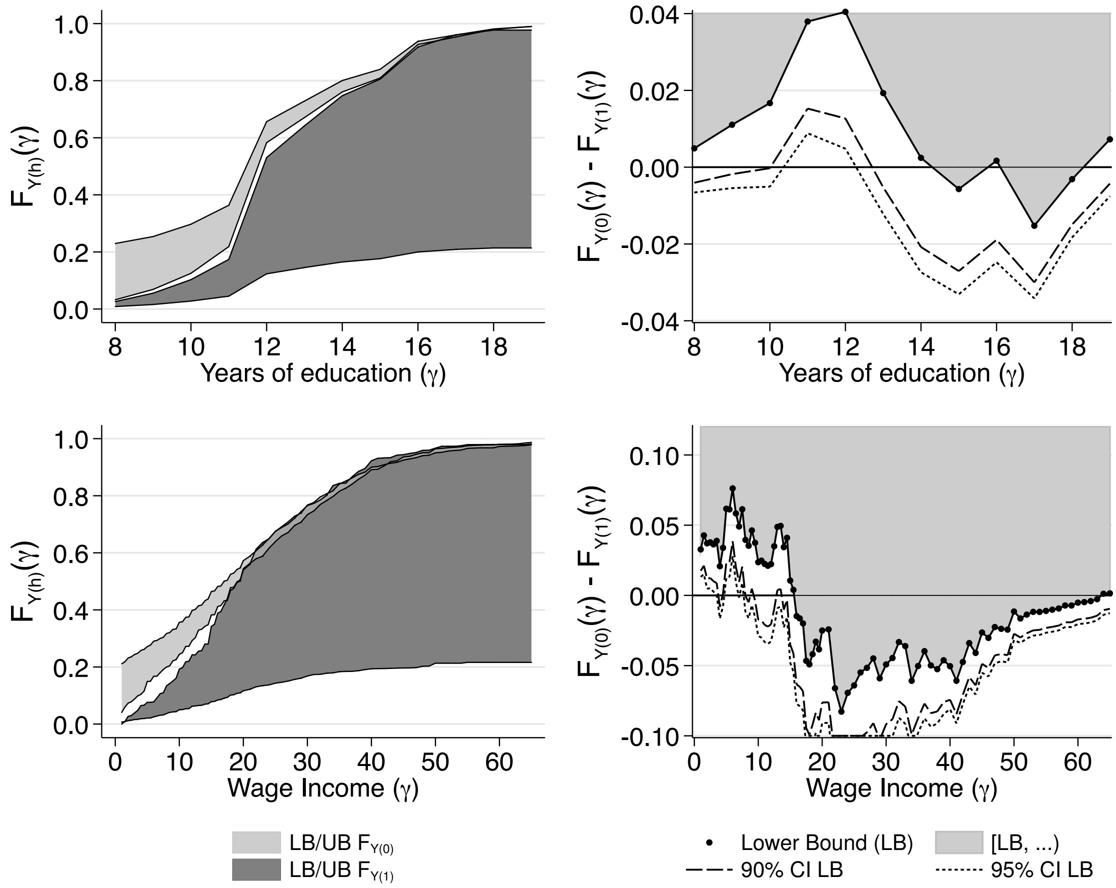
support of education where the lower bound on the cumulative distribution function of $Y(0)$ is larger than the upper bound on the cumulative distribution function of $Y(1)$.

As explained above, to calculate the lower bound on the effect of Head Start on achieving at least γ years of education we subtract the upper bound on $F_{Y(1)}(\gamma)$ from the lower bound on $F_{Y(0)}(\gamma)$. This is the white area in between the shaded areas in Figure 9 where we bound the cumulative potential outcome distributions. The top right panel in Figure 9 shows the lower bound on this causal effect at the different education margins. As can be seen in the figure, for γ up to 14 years of education there is a positive lower bound on the effect of Head Start on obtaining more than γ years of education. The top right panel in Figure 9 also shows the (lower bound of the) 90 and 95 percent confidence intervals. We find statistically significant lower bounds on the probability of obtaining more than 10, 11 and 12 years of education.

The bottom left panel of Figure 9 shows the bounds on the cumulative potential outcome distributions of wage income. As can be seen from the figure, the lower bound on $F_{Y(0)}(\gamma)$ and the upper bound on $F_{Y(1)}(\gamma)$ are only separated at the lower end up to values of γ of about 5,000 USD. The bottom right panel of Figure 9 plots the corresponding lower bounds on the effect of Head Start on obtaining different levels of income, as well as the lower bounds of the 90 and 95 percent confidence intervals. It shows that there is a statistically significant effect of Head Start on wage income but only at the very bottom end of the distribution.

5.3 Combining two monotone instruments

As described in Section 4.2, it is possible to use mother's and father's level of education as two separate MIV's instead of combining the two into one monotone instrument. An advantage of using two separate MIV's is that it can give more informative bounds. A disadvantage is that we have to drop 18% of the observations because we can only include individuals in the sample if we have information on the education of both the mother and the father.



Note: Number of observations equals 4022 (years of education) and 3183 (wage income). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 10. MTS- two MIV bounds on the effect of Head Start on education and earnings

The top two panels of Figure 10 show the results for years of education when we use mother's and father's education as two separate MIV's. The top left panel looks very similar to the top left panel of Figure 9, but from the top right panel it becomes clear that the bounds using two monotone instruments are tighter than when we combine parent's education into one MIV. The results show that Head Start increases high school graduation (more than 11 years of education) rates by at least 4 percentage points. This is a substantial effect as 22 percent of the complete sample and 24 percent of the Head Start participants did not complete high school (obtained less than 12 years of education).

The bottom two panels of Figure 10 report the bounds for wage income as outcome variable. These bounds are clearly tighter than in Figure 9 and indicate that there is a substantial and statistically significant positive effect of Head Start on wage income at the bottom end of the distribution. The biggest effects are found around the 1993 single person poverty threshold (7,518 USD); the estimated lower bound shows for example that Head Start increases the probability of earning 7,500 dollar or more by at least 6 percentage points.

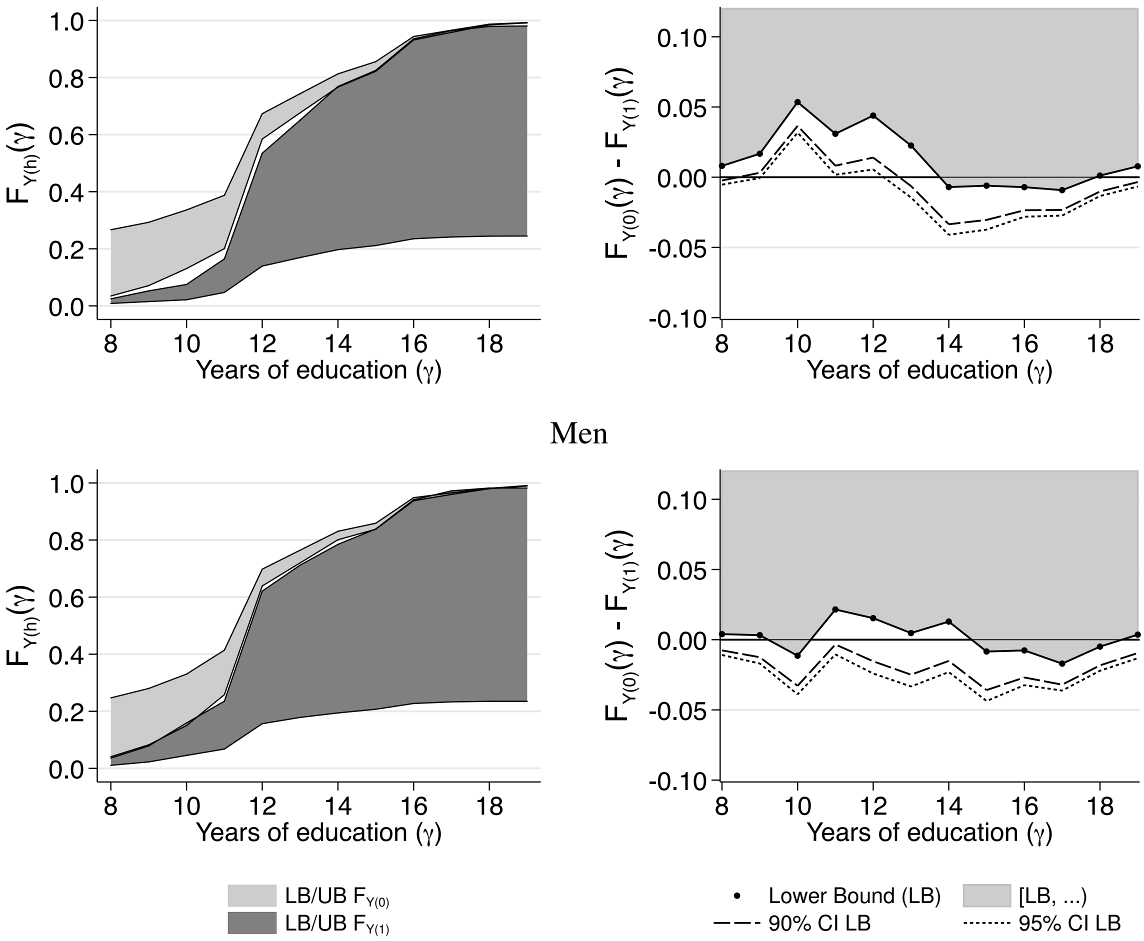
5.4 Effects by gender

Many studies have documented that early childhood interventions affect men and women differently and also found substantial differences across race. Following these results and other studies of Head Start we therefore investigate treatment effects for these different subgroups.

The top right panel of Figure 11 reports the lower bounds on the effect on education for women. This shows that Head Start increases the probability of completing more than 10 years of education by at least 5 percentage points and high school completion by at least 3 percentage points. The figure also shows a positive lower bound for the year following high school, but the cumulative potential outcome distributions are not separated at higher levels of education. Around high school the lower bounds are however all significant at the 5 percent level. To compare, the bottom right panel of Figure 11 reports the lower

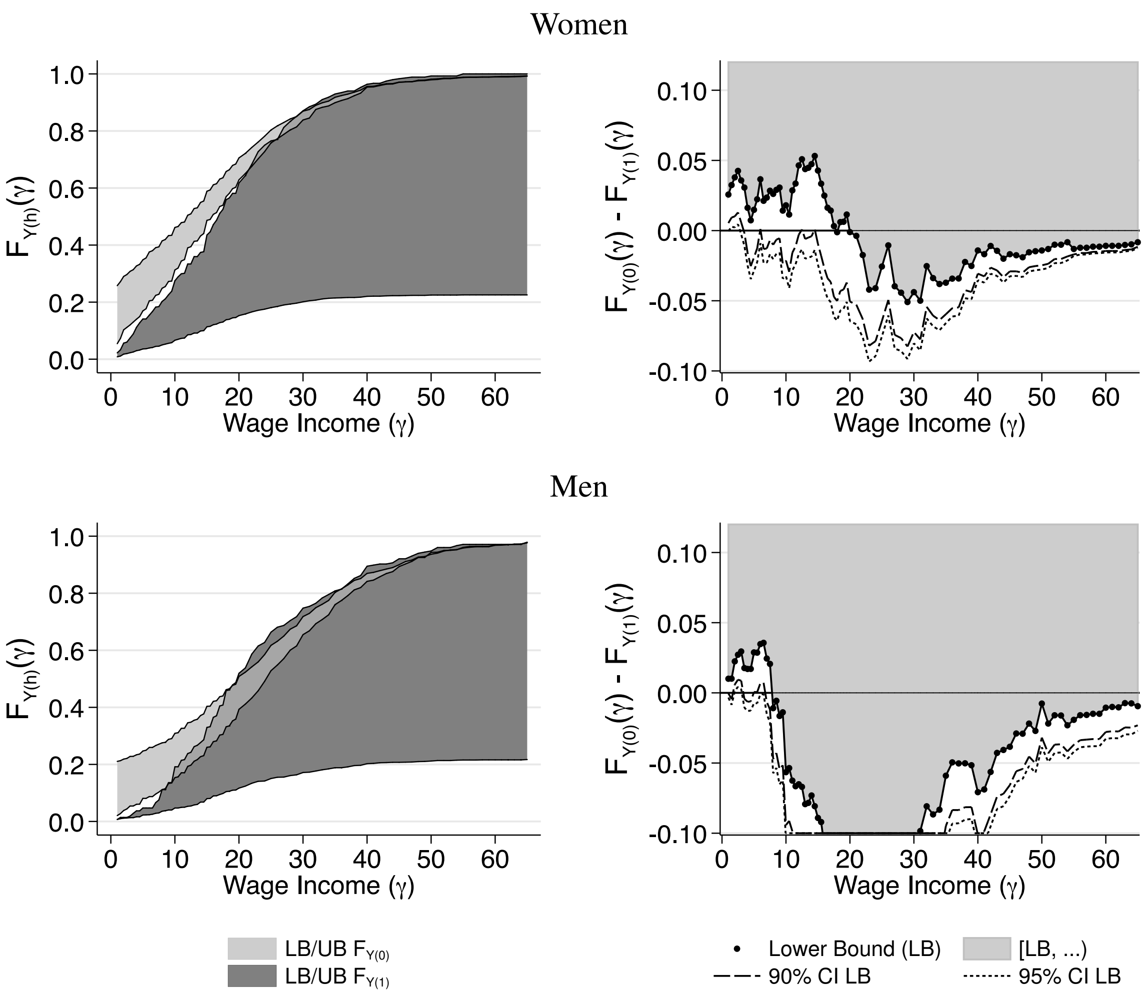
Women

Men



Note: Number of observations equals 2452 (women) and 2424 (men). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 11. The effect of Head Start on years of education – By gender



Note: Number of observations equals 1780 (women) and 2007 (men). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 12. The effect of Head Start on wage income in 1993– By gender

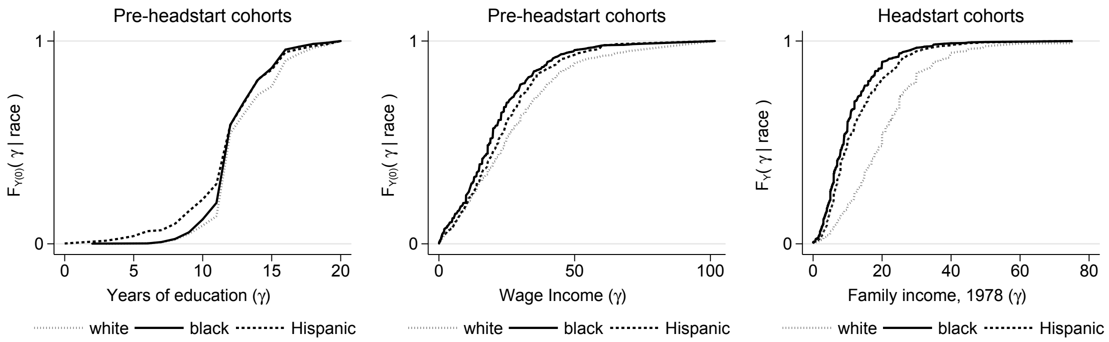
bounds on the effect on education for men. While the lower bounds on the effect of Head Start are positive from 11 to 14 years of education, they are smaller than for women and not statistically significant (the lower bound on the impact on high school completion is close to being significant at the 10 percent level). We therefore find informative bounds for women, but not for men.

The top panels of Figure 12 report the bounds on the effect on wage income for women. We estimate positive lower bounds on the effect of Head Start increasing income beyond γ, for levels of γ up to 20,000 USD, and up to 15,000 USD the lower bounds amount to about 3 percentage points. Although these bounds are systematically positive at the lower end of the distribution, they are relatively imprecise. They are only significant at the 5 percent level for very low values of γ. For men we see in the bottom panel of Figure 12 positive lower bounds on the effect of Head Start increasing earnings beyond levels up to 7,000 USD, which tend to be statistically significant at the 10 percent level. Although imprecise these results suggest that Head Start may successfully raise income for women up to relatively high levels, while for men the bounds suggest some impact around single person poverty lines.

5.5 *Effects by race*

Investigating the impact separately by race is in particular relevant in the context of Head Start since its eligibility criteria target the poor, and consequently a disproportionate share of Head Start participants are black and to a lesser extent Hispanic. So although there are hardly any participation disparities by gender, the probability of being exposed to Head Start is markedly different for children from white, black or Hispanic families as could be seen in Table 1.

There are also reasons to expect heterogenous effect by race because we find the largest lower bounds at the bottom end of the distribution, which indicates that those with low ability and/or low background characteristics tend to benefit the most from participating in Head Start. That blacks and Hispanics are overrepresented at the lower end of the



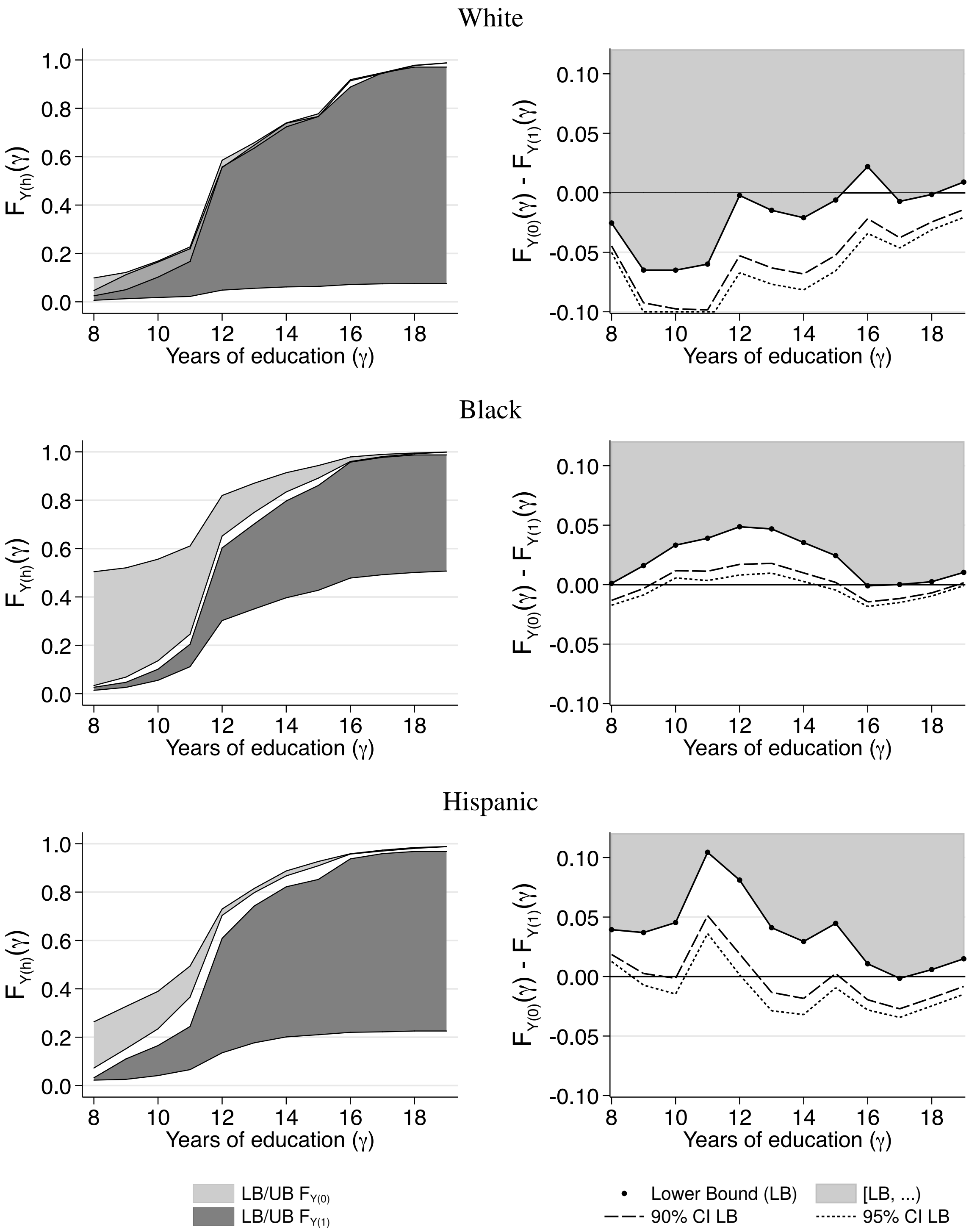
Note: Number of observations equals 4,873 (years of education), 2,153 (wage income, 1993), and 4,028 (family income, 1978)

Figure 13. CDF's of Y(0) in the pre-Head start cohorts and family income by race

distribution is illustrated by Figure 13, which shows the CDFs of education and wage income for the pre-Head Start cohorts, $F_{Y(0)}$, as well as the CDF of family income in 1978 (for the Head Start cohorts) by race. The distributions of $Y(0)$ and family income of whites stochastically dominate those of blacks and Hispanics, which suggest that we would expect larger effects of Head Start for blacks and Hispanics.

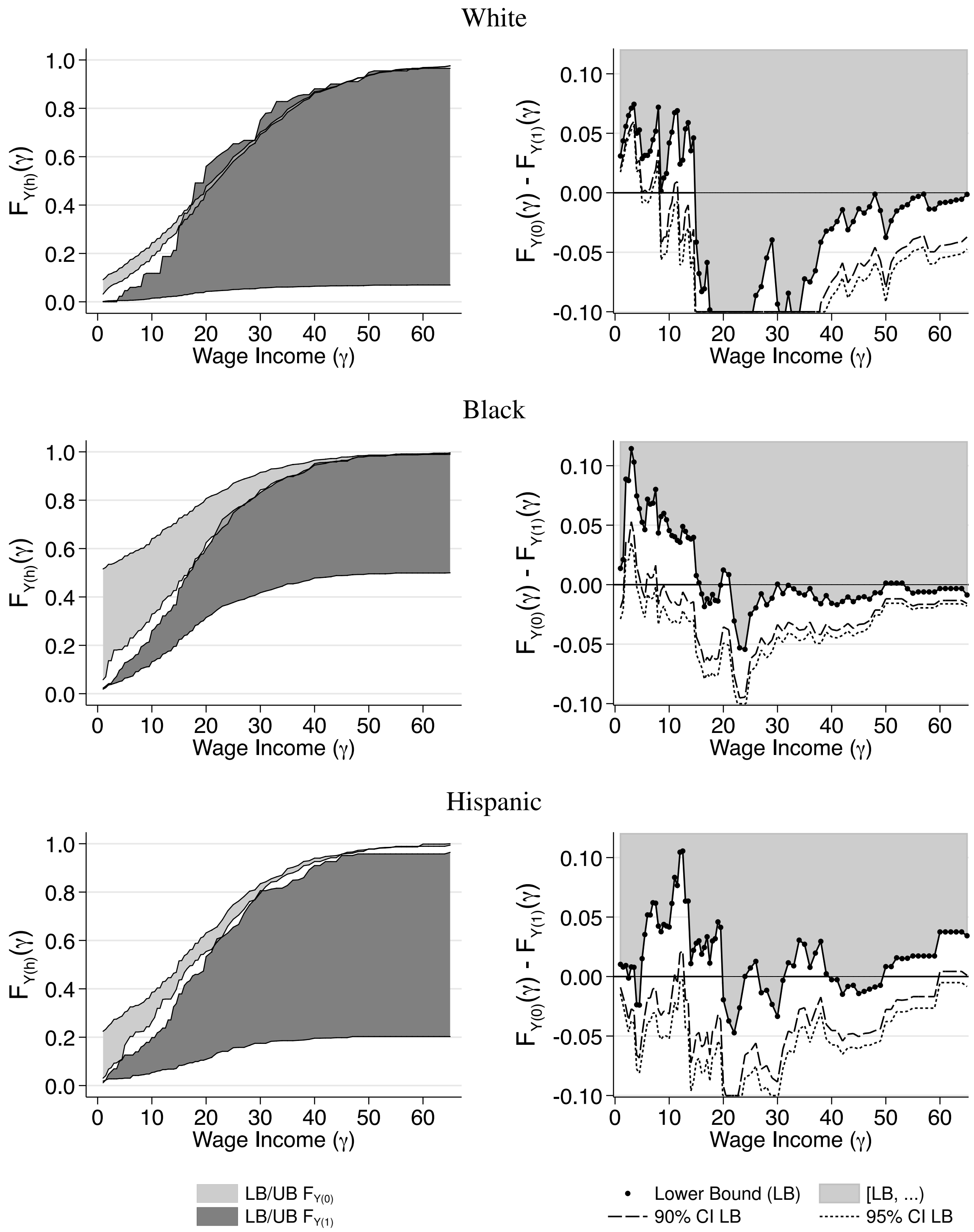
First consider the top panel in Figure 14 which shows that the bounds on the cumulative potential outcome distributions overlap and that the lower bounds on the effects on education for whites are essentially all negative and thus not informative. The middle panel shows the estimated bounds for blacks. Here we see a substantial gap between the bounds on the cumulative potential outcome distributions which translates into a positive lower bound on the effect of Head Start for a wide range of education levels. These lower bounds imply that Head Start increases completed years of education for blacks at all margins from 9 to 15 years of education. Around high school graduation these lower bounds are around 5 percentage points and statistically significant at the 5 percent level.

The bottom panel in Figure 14 presents the results for Hispanics. Here we find positive lower bounds for a similar wide margin of completed education as for blacks. The lower bound is particularly high at the high school completion margin (i.e. having more than 11 years of education) where we find that Head Start increases the probability of having a high school diploma or more by at least 10 percentage points.



Note: Number of observations equals 2404 (white), 1518 (black) and 954 (Hispanic). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 14. The effect of Head Start on years of education– By race



Note: Number of observations equals 1988 (white), 1061 (black) and 738 (Hispanic). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 15. The effect of Head Start on wage income in 1993– By race

Figure 15 reports the results for wage income. The top panel shows the results for whites. Although the lower bounds on the impact on education were uninformative, we do see positive and statistically significant lower bounds on the impact of Head Start at the bottom of the wage income distribution where the lower bound on $F_{Y(0)}(\gamma)$ and the upper bound on $F_{Y(1)}(\gamma)$ are separated for values of γ up to 15,000 USD. The middle panel panel shows the results for blacks. Here we also see that the lower bound on $F_{Y(0)}(\gamma)$ and the upper bound on $F_{Y(1)}(\gamma)$ are separated over a similar range as for whites. The lower bounds tend to be statistically significant around the poverty thresholds. Finally, the bottom panel reports the estimated bounds for Hispanics. While the estimates show that the cumulative potential outcome distributions are systematically separated up to 20,000 USD, the lower bounds are mostly imprecisely estimated.

To summarize, these results show that Head Start has a statistically significant positive effect on years of education, in particular for blacks and Hispanics. For wage income we also find evidence that Head Start has beneficial impacts, with effects located at the lower end of the distribution.

5.6 *Effects of Head Start on the treated*

In this paper we estimate bounds on the cumulative potential outcome distributions, $F_{Y(0)}(\gamma)$ and $F_{Y(1)}(\gamma)$, as well as lower bounds on the causal effect of Head Start which we define as the difference between these two cumulative potential outcome distributions; $\triangle(\gamma) = F_{Y(0)}(\gamma) - F_{Y(1)}(\gamma)$. Although our estimated bounds show how the effects of Head Start vary over the outcome distribution, it is also of interest to know how the effects for the treated $(D = 1)$ vary over the outcome distribution; $\triangle(\gamma|D = 1) = F_{Y(0)}(\gamma|D = 1) - F_{Y(1)}(\gamma|D = 1)$. The causal effect that we focus on in this paper is a weighted average of the causal effect on the treated and the causal effect on the non-treated:

$$\triangle(\gamma) = \triangle(\gamma|D = 1)P(D = 1) + \triangle(\gamma|D = 0)P(D = 0)$$

which implies that if the effect of Head Start on the probability of obtaining an education or labor market outcome bigger than γ for the non-treated is not higher than the effect for the treated ($\triangle(\gamma|D=1) \geq \triangle(\gamma|D=0)$), the lower bounds reported in this paper can be interpreted as (conservative) lower bounds on the effects on the treated. Our subsample analysis suggests that this is indeed the case, because the estimated lower bounds are highest for the subsamples with the highest shares of Head Start participants (blacks and Hispanics).

6 Robustness

6.1 The importance of the counterfactual

So far we reported results where we compare the effectiveness of Head Start with informal care. To see whether the results are sensitive to the choice of the counterfactual, Figure 16 shows results where we include individuals that attended another non-Head-Start preschool program in the group of non-participants. This means that we compare Head Start with a counterfactual that is a mixture of informal care and alternative center-based preschool programs. It also implies that our sample size increases (by 16 percent) and that our MTS assumption changes a bit because we include the respondents that attended another preschool in the group of non-participants. Figure A5 in the appendix shows that, for each of the values of the MIV, the distribution of family income for the Head Start participants is stochastically dominated by the distribution of the group that includes the non-participants and those that attended another preschool program. This is in line with the MTS assumption.

Figure 16 shows that the results are qualitatively very similar to the results in Figure 9. The lower bounds on the effect of Head Start are however lower in Figure 16, for example, Head Start increases the probability of high school graduation by at least 3 percentage points when the counterfactual is informal care compared to 2 percentage points when the counterfactual is a mixture of informal care and center-based preschool. These results

Note: Number of observations equals 5659 (years of education) and 4439 (wage income). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 16. MTS-MIV bounds on the effect of Head Start – sample including other preschool

confirm that it is important to be explicit about the counterfactual and that the effects of Head Start seem strongest when informal (home-based) care is the alternative treatment and are in line with Kline and Walters (2016) and Feller et al. (2016) who find that the fading out of the effect of Head Start on cognitive test scores is sensitive to the choice of the counterfactual treatment.

6.2 Survey reporting

As the NLSY is a survey, Head Start participation is self-reported and there might be misreporting. One way to check this is to compare Head Start enrollment in the NLSY with national enrollment statistics. This is however not straightforward due to data limitations. National enrollment counts are by calendar year, while in the NLSY we do not know when people attended Head Start and can only compute enrollment rates at the cohort level. It is possible to map annual enrollment to cohort enrollment by making some assumptions. Using national enrollment statistics gives an upper bound on the national participation rate of 17 percent in 1969/1970. This assumes that children enroll in Head Start for only 1 year. If children would have enrolled twice (for example when they were four and when they were five) the implied cohort level enrollment rate would half to about 9 percent. The enrollment rate for the 1964 cohort (who would have enrolled around 1969/70) in the NLSY is 13.4 percent and is therefore consistent with possible enrollment rates implied by the national enrollment statistics.

Across all eligible cohorts in the NLSY we similarly estimate an average national participation rate of 12.3 percent which is also consistent with what we know about Head Start participation from the national enrollment data from this period. Self-reported enrollment in the NLSY is also higher for blacks than for whites (43.2 percent versus 6.4 percent) and among the children of parents who did not attend college than among children with such a parent (15.4 percent versus 5.8 percent), again patterns that are broadly consistent with historical data on the characteristics of early Head Start participants (DHEW, 1968, 1970, 1972).

Even though self-reported Head Start participation is consistent with historical data, it seems likely that there is some degree of misreporting. Studies on survey reporting in other contexts such as the reporting of transfer programs suggest that such programs are typically under-reported, and that false positives tend to be quantitatively unimportant (see f.e. Meyer and Mittag, 2019).

It is not possible to correct our estimates from reporting bias without information on the extend, direction and correlates of misreporting. In our context we would for example need to know how reporting bias varies by parental education, gender and age. To understand the potential implication of reporting error for our overall estimates we derive the implications for the No-Assumption and MTS bounds in Appendix A.2 and we summarize here.

Remember that the effect is defined as

$$\triangle(\gamma) = \Pr(Y(1) > \gamma) - \Pr(Y(0) > \gamma) = F_{Y(0)}(\gamma) - F_{Y(1)}(\gamma) \tag{16}$$

and that we focus on the lower bound of $\triangle(\gamma)$:

$$LB_{\triangle(\gamma)} = LB_{F_{Y(0)}(\gamma)} - UB_{F_{Y(1)}(\gamma)}$$

Let D^* be true Head Start participation, and D reported Head Start participation. We assume that

$$\Pr(\text{false positive}) = P(D = 1, D^* = 0) = 0$$

and use the following notation

$$\begin{aligned}
\phi &\equiv \Pr(\text{false negative}) = P(D = 0, D^* = 1) \\
p^* &\equiv \Pr(D^* = 1) \\
p &\equiv \Pr(D = 1) = \Pr(D = 1, D^* = 1) + \Pr(D = 1, D^* = 0) = p^* - \phi
\end{aligned}$$

In Appendix A.2 we show that the No-Assumption bounds on $F_{Y(1)}(\gamma)$ do not change when we allow for false negative misreporting, while the bounds on $F_{Y(0)}(\gamma)$ do change. The No-Assumption bounds that allow for false negatives are shown below

$$F_Y(\gamma|D=1)p \leq F_{Y(1)}(\gamma) \leq F_Y(\gamma|D=1)p + (1-p)$$
$$F_Y(\gamma|D=0)(1-p) - \phi \leq F_{Y(0)}(\gamma) \leq F_Y(\gamma|D=0)(1-p) + p + \phi$$

and imply that the No-Assumption lower bound on the effect goes down by ϕ.

Characterizing the nature of selection into misreporting can help to tighten the bounds. A natural benchmark is ignorability which assumes that misreporting among the treated is at random:

$$F_{Y(h)}(\gamma|D^*=1, D=0) = F_{Y(h)}(\gamma|D^*=1, D=1) = F_{Y(h)}(\gamma|D^*=1)$$

Studies on survey reporting such as Meyer and Mittag (2019) suggest that among the treated the mis-reporters are weakly positively selected:

$$F_{Y(h)}(\gamma|D^*=1, D=1) \geq F_{Y(h)}(\gamma|D^*=1) \geq F_{Y(h)}(\gamma|D^*=1, D=0)$$

Assuming that misreporting is ignorable or positive, implies that the potential outcome distributions of those that report that they did not participate in Head Start while in fact they did are not stochastically dominated by the outcome distributions of those that correctly reported to have participated in Head Start. This would be consistent with that those for whom Head start was not so important are more likely to forget that they participated.

We show in Appendix A.2 that the misreporting corrected MTS lower bound on the effect assuming that selection into misreporting is ignorable or positive equals

$$LB\triangle(\gamma) = \left(1 + \frac{\phi}{(1-p^*)}\right) \cdot \left(F_Y(\gamma|D=0) - F_Y(\gamma|D=1)\right)$$

which implies that current lower bounds are off by a factor $\frac{\phi}{1-p^*}$. Taking the upper bound on the national employment rate as the true participation rate $p^* = 0.170$ and the estimated participation rate in the NLSY as the reported participation rate $p = 0.134$ we find that the maximum relative bias because of misreporting is about 4% and therefore relatively minor.

6.3 Survey weighting

As noted above, the NLSY over-samples blacks and Hispanics which means that the results based on the NLSY data including over-samples do not necessarily carry over to the U.S. population. We can obtain bounds around the cumulative potential outcome distributions of the U.S. population by using the sampling weights provided in the NLSY. It is however not possible to obtain sampling-design corrected confidence intervals around these bounds because these need to be bootstrapped for which we need primary sampling units and strata which are not available in the public use dataset. What we can do is compare the weighted and unweighted estimates of the lower bounds.[18]

Both for education and income we find that when stratifying by race using the NLSY sampling weights gives lower bound estimates which are very similar to the bounds obtained without sampling weights. Correlation coefficients are about 0.99 for whites, 0.96 for blacks and 0.90 for Hispanics. As the NLSY under-samples whites and we do not find positive effects of Head Start on education for whites, the bounds based on sampling weights are not positive with the exception of attaining at least 12 years of schooling where the lower bound equals 0.02. For wage income, the estimated lower bounds with sampling weights are very similar or even higher than the estimates obtained without sampling weights.

In summary, weighting does not appear to affect the race specific lower bounds on the effect of Head Start on schooling and income. When we compute bounds on the effects of Head Start using sampling weights these tend to to be uninformative for schooling because of the uninformative bounds for whites. For income the bounds with and without sampling weights turn out to be very similar.

[18] Figure A9 and Figure A10 report estimated lower bounds on the effects of Head Start both with and without the use of sampling weights.

7 Conclusion and discussion

Assessing the effect of Head Start on long term outcomes has turned out to be challenging for at least two reasons. First, long run outcomes are often not observed. Second, it is difficult to find exogenous variation in Head Start participation that can be exploited to estimate relevant treatment effects. The few available studies that focus on longer term outcomes rely on quasi-experimental evidence, and tend to find positive impacts. This evidence is however scattered and the studies disagree on who benefits and what outcome margins are affected.

The current paper contributes to this small literature and is the first to consider the effect of Head Start across the distribution of long term outcomes. It estimates these long term impacts without relying on quasi-experimental variation in Head Start participation, but instead relies on two weak stochastic dominance assumptions. This approach results in bounds around the cumulative potential outcome distributions of education and wage income. While previous studies on the long term effects of Head Start estimate (local) average treatment effects, our focus on the distribution of outcomes paints a richer picture of how Head Start participation affects schooling and earnings, also within subgroups defined by race or gender. This allows us to assess whether effect heterogeneity across groups is consistent with effect heterogeneity along the outcome distribution.

The tightest bounds show that Head Start increases high school graduation by at least 4 percentage points and the probability of earning more than the (one-person) poverty threshold by at least 6 percentage points. The positive lower bounds are concentrated at the bottom end of the distribution, which suggests that Head Start offers the highest benefits to those with low skills and/or social background. This is confirmed by our sub-sample analyses where we find large lower bounds on the payoffs to Head Start for blacks and Hispanics. Our results therefore indeed show a consistent pattern of effect heterogeneity and suggest that Head Start benefits those that need it the most.

References

Bauer, Lauren and Diane Whitmore Schanzenbach. 2016. The long-term impact of the Head Start program. Tech. rep., The Hamilton Project, Brookings.

Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes. 2017. Life cycle earnings, education premiums and internal rates of return. *Journal of Labor Economics* 35, no. 4:993–1030.

Bitler, Marianne P., Hilary W. Hoynes, and Thurston Domina. 2014. Experimental evidence on distributional effects of Head Start. Working Paper 20434, National Bureau of Economic Research, Cambridge, MA.

Böhlmark, Anders and Matthew J. Lindquist. 2006. Life-cycle variations in the association between current and lifetime income: Replication and extension for sweden. *Journal of Labor Economics* 24, no. 4:879–896.

Carneiro, Pedro and Rita Ginja. 2014. Preventing behavior problems in childhood and adolescence: evidence from Head Start. *American Economic Journal: Economic Policy* 6, no. 4:135–73.

Currie, Janet and Duncan Thomas. 1995. Does Head Start make a difference? *The American Economic Review* 85, no. 3:341–364.

———. 2000. School quality and the longer-term effects of Head Start. *Journal of Human Resources* 35, no. 4:755–74.

Deming, David. 2009. Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics* 1, no. 3:111–134.

DHEW. 1968. Project head start 1965–1966: A descriptive report of programs and participants. Tech. rep., Office of Child Development, U.S. Department of Health, Education and Welfare (DHEW), Washington, DC.

———. 1970. Project head start 1968: A descriptive report of programs and participants. Tech. rep., Office of Child Development, U.S. Department of Health, Education and Welfare (DHEW), Washington, DC.

———. 1972. Project head start 1969–1970: A descriptive report of programs and participants. Tech. rep., Office of Child Development, U.S. Department of Health, Education and Welfare (DHEW), Washington, DC.

Elango, Sneha, James J. Heckman, Andres Hojman, and Jorge Luis Garcia. 2016. Early childhood education. In *Means-tested transfer programs in the united states II*, ed. R. Moffit. University of Chicago Press.

Feller, Avi, Todd Grindal, Luke W. Miratrix, and Lindsay Page. 2016. Compared to what? variation in the impact of early childhood education by alternative care-type settings. *Annals of Applied Statistics* 110, no. 3:1245–1285.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. Longer-term effects of Head Start. *American Economic Review* 92, no. 4:999–1012.

Gibbs, Chloe, Jens Ludwig, and Douglas L. Miller. 2011. Does Head Start do any lasting good? In *Legacies of the war on poverty*, eds. Martha Bailey and Sheldon Danziger. New York: Russell Sage Foundation, 36–65.

Grosz, Michel Z., Douglas L. Miller, and Na'ama Shenhav. 2016. Long-term effects of Head Start: New evidence from the PSID. Tech. rep.

Haider, Steven and Gary Solon. 2006. Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review* 96, no. 4:1308–1320.

Imbens, Guido W. and Charles F. Manski. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72, no. 6:1845–1857.

Kline, Patrick and Christopher Walters. 2016. Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics* 131, no. 4:1795–1848.

Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonko. 2006. Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences* 103, no. 27:10155–10162.

Kreider, Brent and John V. Pepper. 2007. Disability and employment: reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association* 102, no. 478:432–441.

Ludwig, Jens and Douglas L. Miller. 2007. Does Head Start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics* 122, no. 1:159–208.

Manski, Charles F. 1989. Anatomy of the selection problem. *Journal of Human resources* 24, no. 3:343–360.

———. 1990. Nonparametric bounds on treatment effects. *The American Economic Review, Papers & Proceedings* 80, no. 2:319–323.

———. 1997. Monotone treatment response. *Econometrica* 65, no. 6:1311–1334.

Manski, Charles F. and John V. Pepper. 2000. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 68, no. 4:997–1010.

McFadden, Daniel. 1989. Testing for stochastic dominance. In *Studies in the economics of uncertainty*, eds. T. Fomby and T.K. Seo. Springer, 113–134.

Meyer, Bruce D. and Nikolas Mittag. 2019. Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net. *American Economic Journal: Applied Economics* 9, no. 2:176–204.

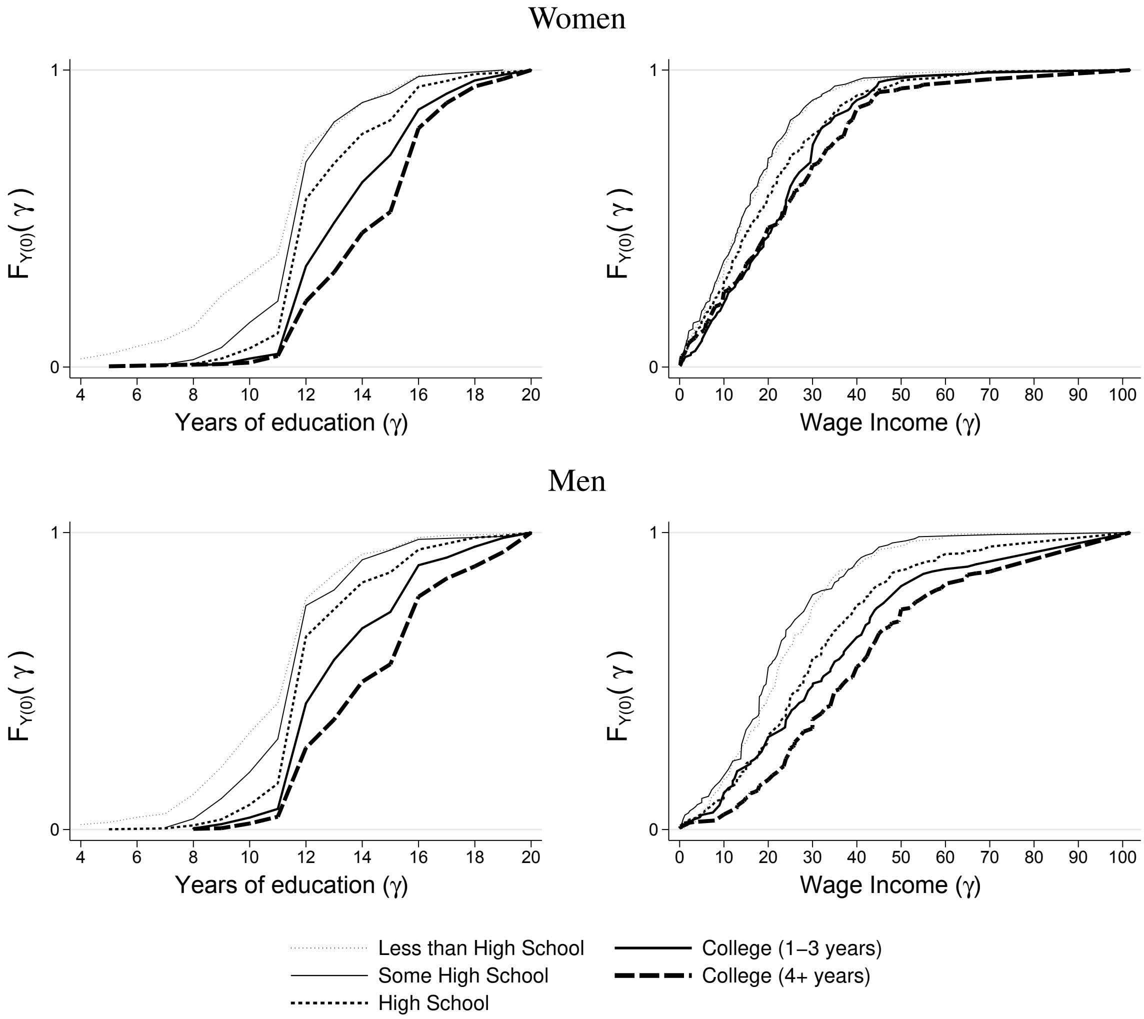
Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid et al. 2010. *Head Start impact study. final report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Schnur, Elizabeth, Jeanne Brooks-Gunn, and Virginia C. Shipman. 1992. Who attends programs serving poor children?: The case of Head Start attendees and nonattendees. *Journal of Applied Developmental Psychology* 13, no. 3:405–421.

Thompson, Owen. 2018. Head start's impact in the very long-run. *Journal of Human Resources* 53, no. 4:1100–1139.

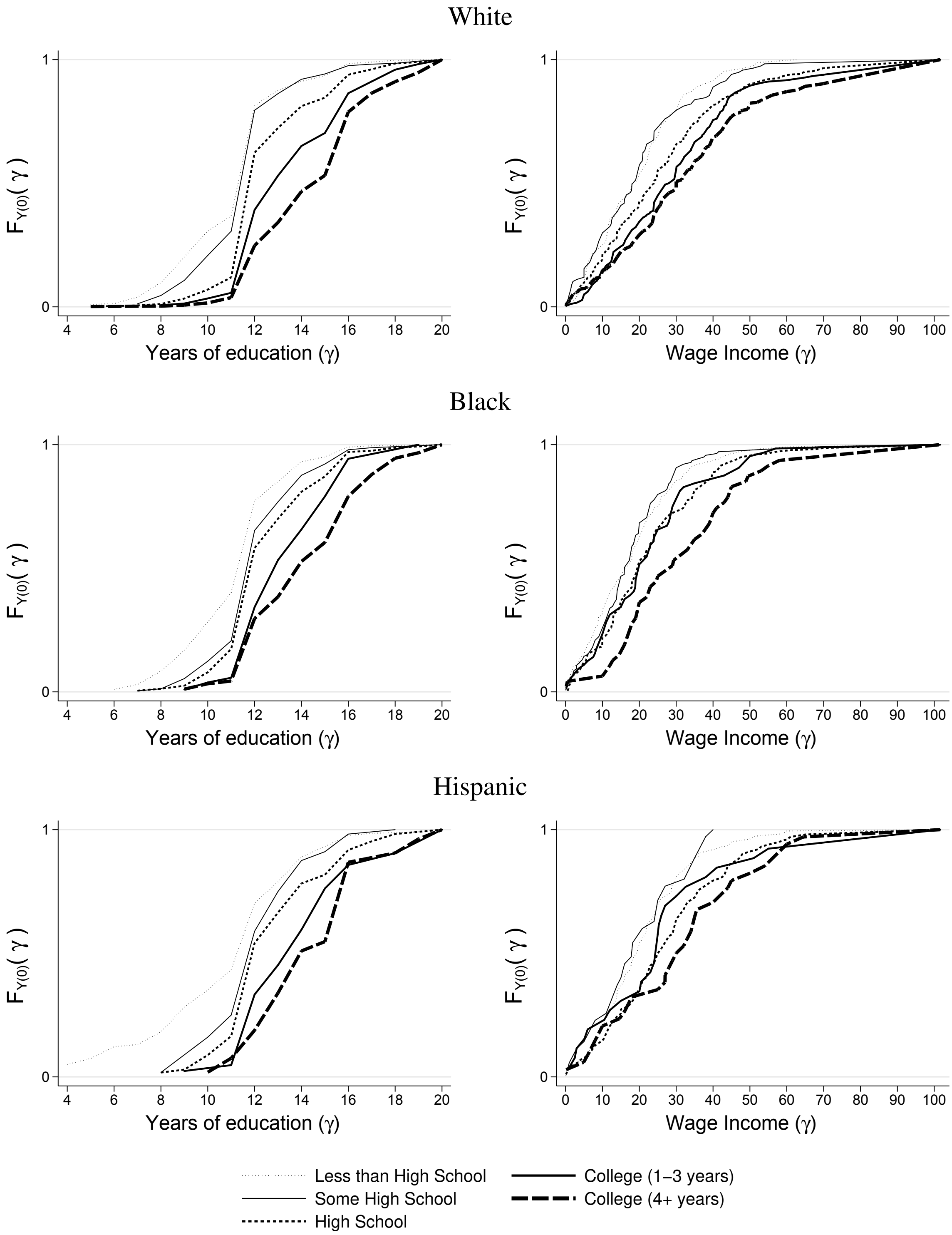
A Appendix – Head Start and the distribution of long term education and labor market outcomes

A.1 Extra results



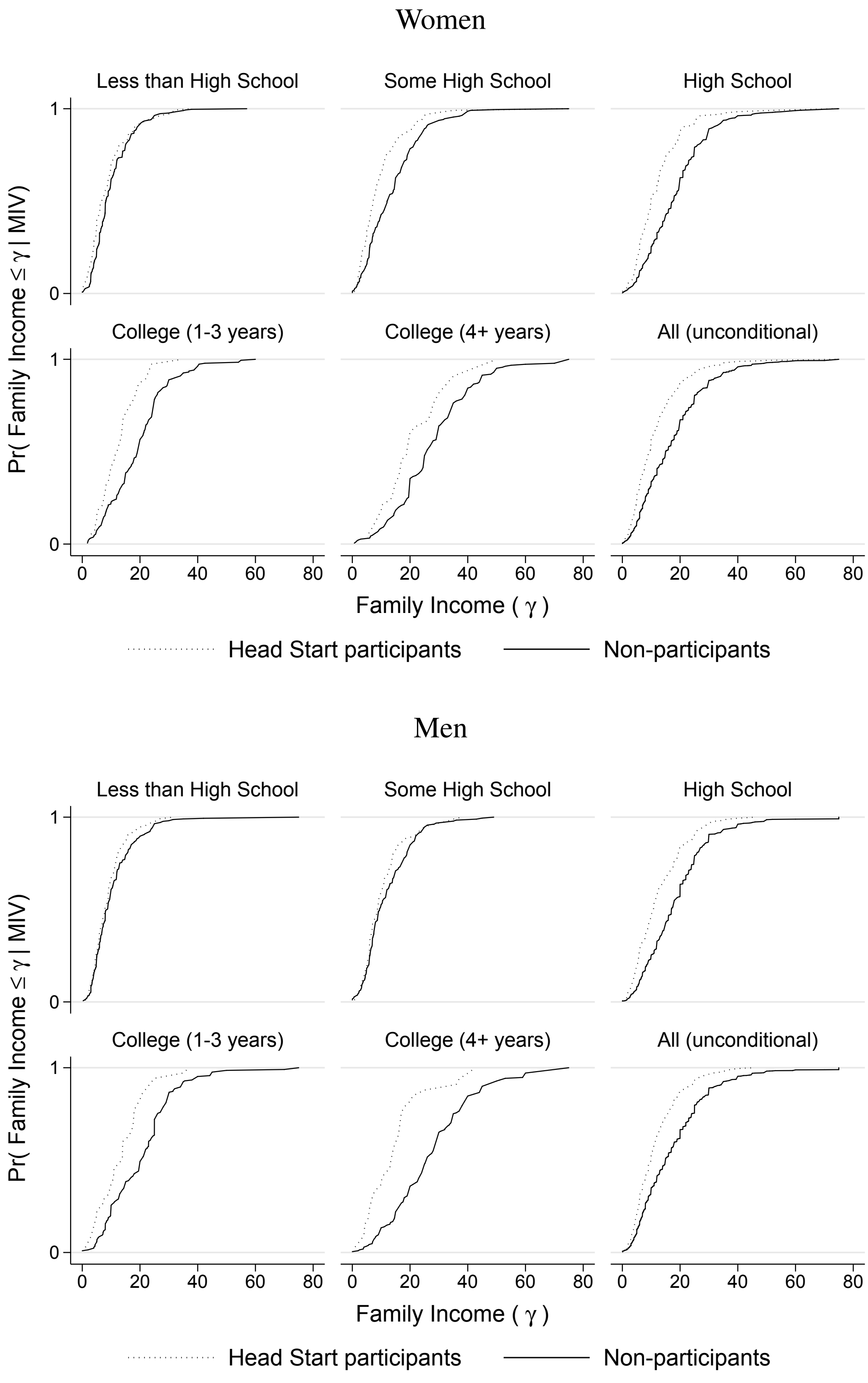
Note: Figures based on data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations for education equal 2425 (men) and 2448 (women). Number of observations for wage income equal 1099 (men) and 1054 (women).

Figure A1. MIV Assumption check – by gender



Note: Figures based on data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations for education equal 3172 (white), 1044 (black) and 657 (Hispanic). Number of observations for wage income equal 1189 (white), 582 (black) and 382 (Hispanic).

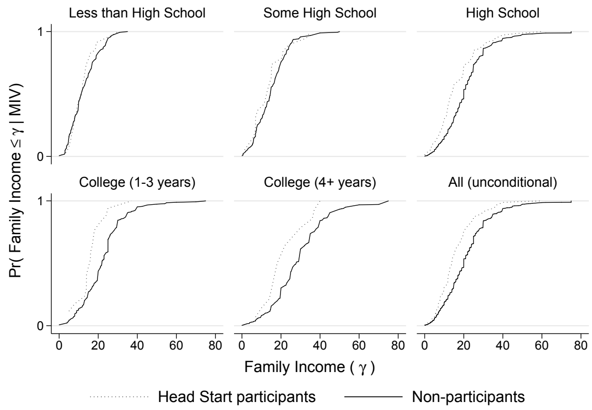
Figure A2. MIV Assumption Check – by race



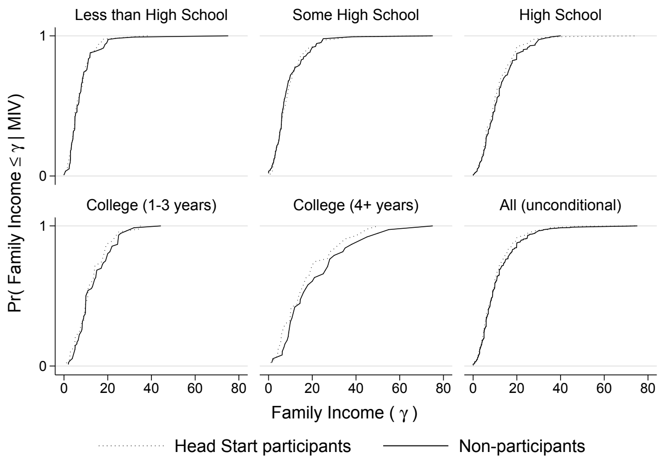
Note: Number of observations for women equal 439 (less than high school), 334 (some high school), 791 (high school), 227 (college 1-3 years), 219 (college 4+ years) and 2010 (all). Number of observations for men equal 422 (less than high school), 280 (some high school), 828 (high school), 246 (college 1-3 years), 242 (college 4+ years) and 2018 (all).

Figure A3. Family income and the MIV, for participants and non-participants– By gender

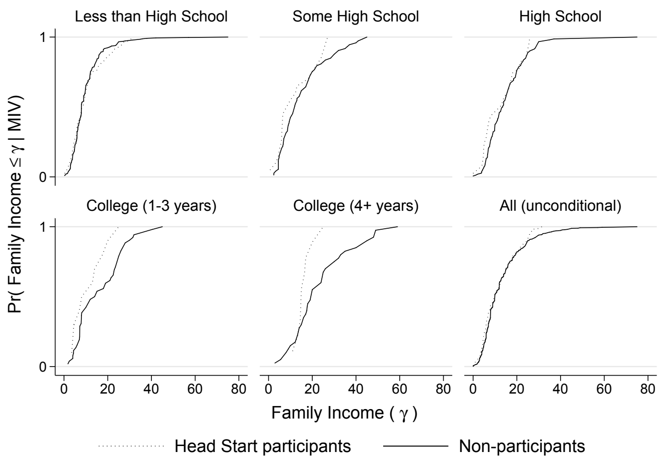
White



Black

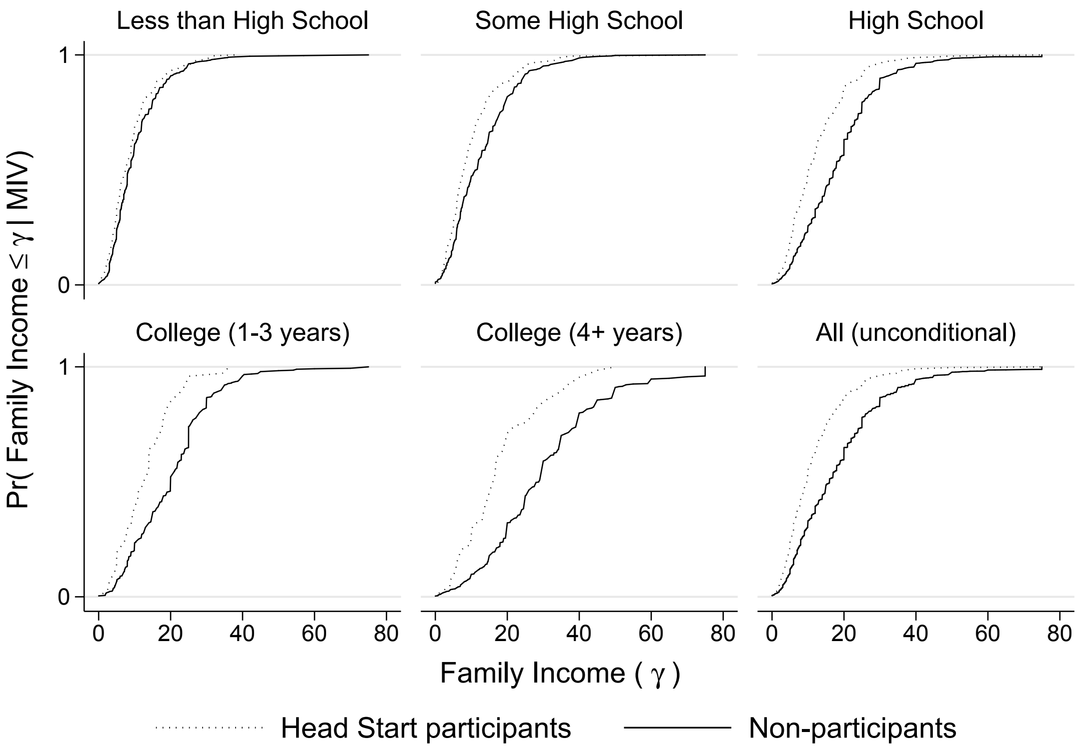


Hispanic



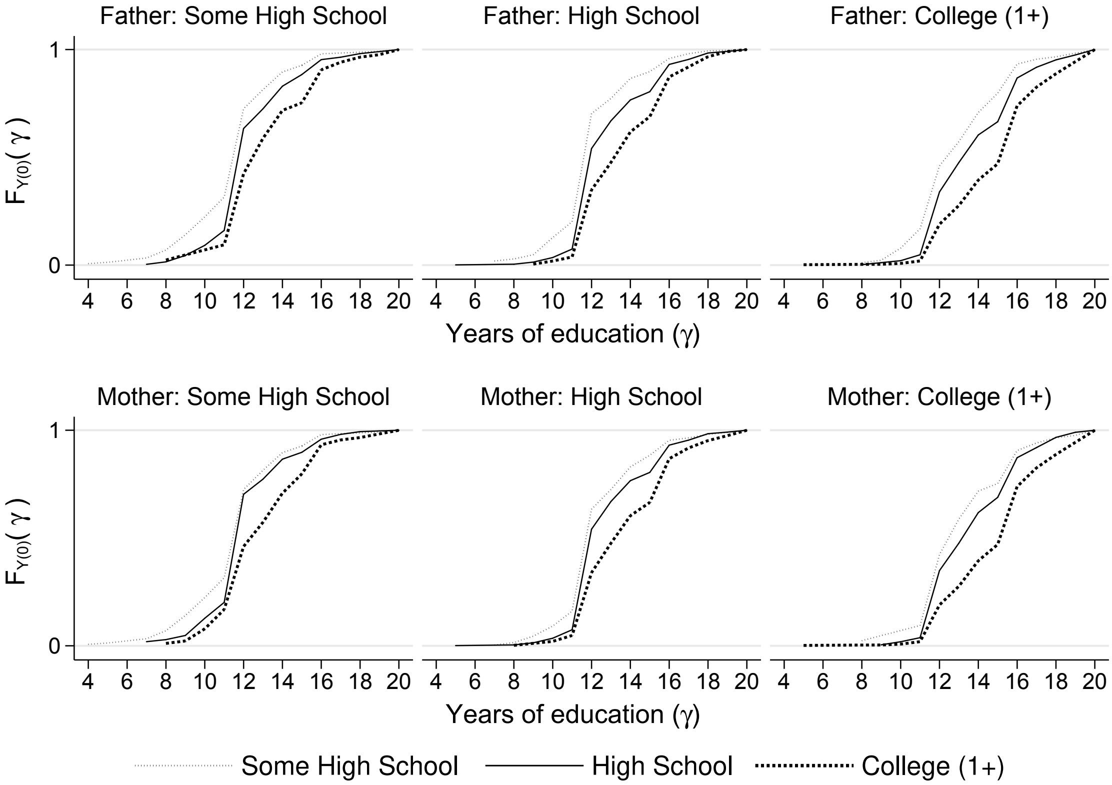
Note: Number of observations for white equal 201 (less than high school), 208 (some high school), 928 (high school), 289 (college 1-3 years), 331 (college 4+ years) and 1957 (all). Number of observations for black equal 252 (less than high school), 313 (some high school), 500 (high school), 122 (college 1-3 years), 81 (college 4+ years) and 1268 (all). Number of observations for Hispanic equal 408 (less than high school), 93 (some high school), 191 (high school), 62 (college 1-3 years), 49 (college 4+ years) and 803 (all).

Figure A4. Family income and the MIV, for participants and non-participants– By race



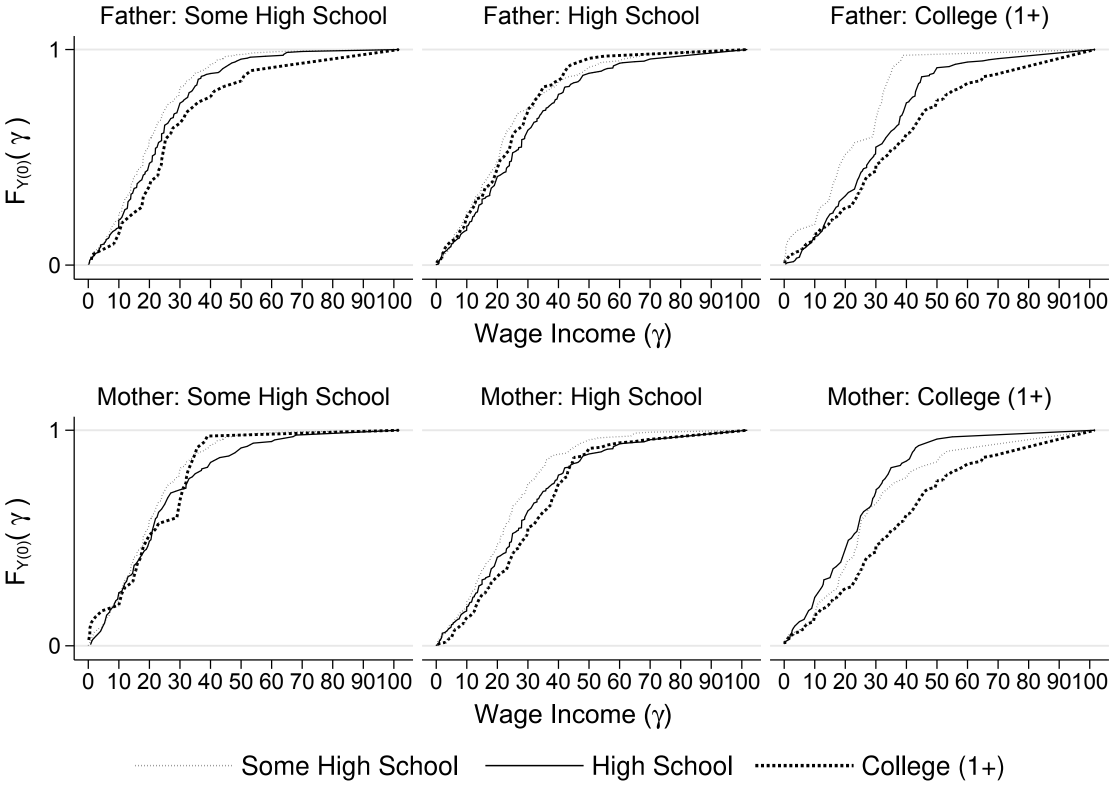
Note: Number of observations equal 947 (less than high school), 663 (some high school), 1825 (high school), 554 (college 1-3 years), 669 (college 4+ years) and 4658 (all).

Figure A5. Family income at age 14-18 and the MIV, for Head Start participants and non-participants (including other preschool)



Note: Number of observations equals 4132.

Figure A6. MIV Assumption Check, Years of education – Two MIV's



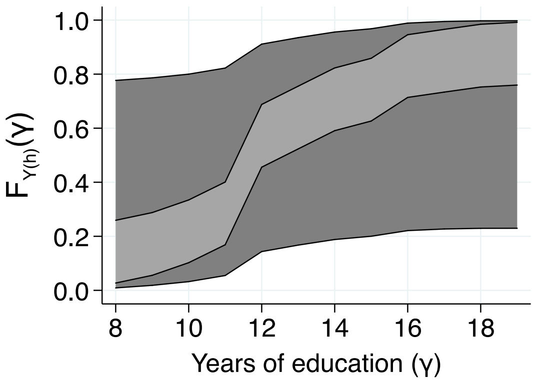
Note: Number of observations equals 1814.

Figure A7. MIV Assumption Check, Wage Income – Two MIV's

Table A1. Check 2 MIV's – p-values of tests for $\mathscr{H}_0 : F_j = F_{j-1}$vs $\mathscr{H}_1 : F_j > F_{j-1}$

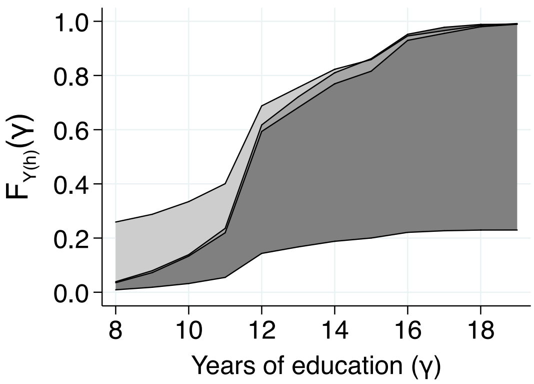
		Education	Wage income
Mother (k):	Father (j):		
- Some High School	High School	0.985	0.971
	College (1+)	1.000	0.297
- High School	High School	0.932	0.943
	College (1+)	1.000	0.819
- College (1+)	High School	0.976	0.413
	College (1+)	0.999	0.972
Father (k):	Mother (j):		
- Some High School	High School	1.000	0.964
	College (1+)	0.989	0.979
- High School	High School	0.999	0.912
	College (1+)	1.000	0.147
- College (1+)	High School	0.989	1.000
	College (1+)	0.998	0.708

Note: Reported p-values are from one sided Kolmogorov-Smirnov test ($\mathscr{H}_0 : F_j = F_{j-1}$ vs $\mathscr{H}_1 : F_j > F_{j-1}$) separately for sub-samples defined by the values of the other parents education (k), using data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations equals 4132 (education) and 1814 (wage income).



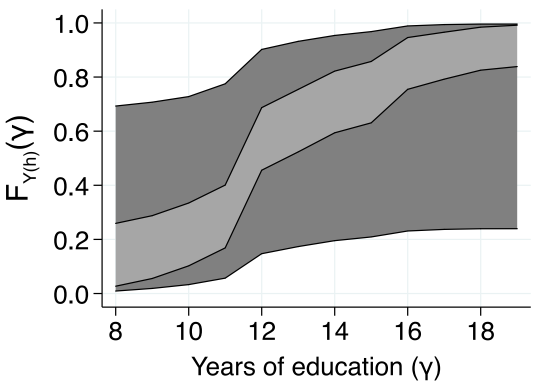
(a) Wage income - No assumption bounds

(b) Wage income - No assumption bounds



(c) Wage income - MTS + No assumption bounds

(d) Wage income - MTS + No assumption bounds



(e) Wage income - MIV + No assumption bounds

(f) Wage income - MIV + No assumption bounds

LB/UB $F_{Y(0)}$
LB/UB $F_{Y(1)}$

Figure A8. Bounds on potential outcome distributions – Alternative assumptions

Table A2. Descriptive statistics, Using sampling weights

		Head Start		Race		
	All	Yes	No	White	Black	Hispanic
Head Start	0.14	1.00	0.00	0.08	0.49	0.20
Age	32.0	32.0	32.1	32.1	32.0	32.0
Female	0.49	0.51	0.49	0.49	0.51	0.49
Race:						
- White	0.79	0.41	0.85	1.00	0.00	0.00
- Black	0.15	0.50	0.09	0.00	1.00	0.00
- Hispanic	0.07	0.09	0.06	0.00	0.00	1.00
Parental Education:						
- Less than High School	0.16	0.08	0.17	0.18	0.07	0.07
- Some High School	0.13	0.08	0.14	0.14	0.10	0.09
- High School	0.44	0.42	0.45	0.47	0.41	0.24
- College (1-3 years)	0.13	0.20	0.12	0.11	0.25	0.12
- College (4+ years)	0.13	0.22	0.12	0.10	0.17	0.48
Family income 1978	19,321	13,070	20,414	21,363	11,230	13,554
Years of education	13.0	12.6	13.1	13.2	12.6	12.1
Wage income	24,362	20,598	24,929	25,422	19,288	20,883
N	4,876	1,132	3,744	2,404	1,518	954

Note: Sample sizes for wage income are: 3,781; 815; 2,966; 1,985; 1,060 and 736.

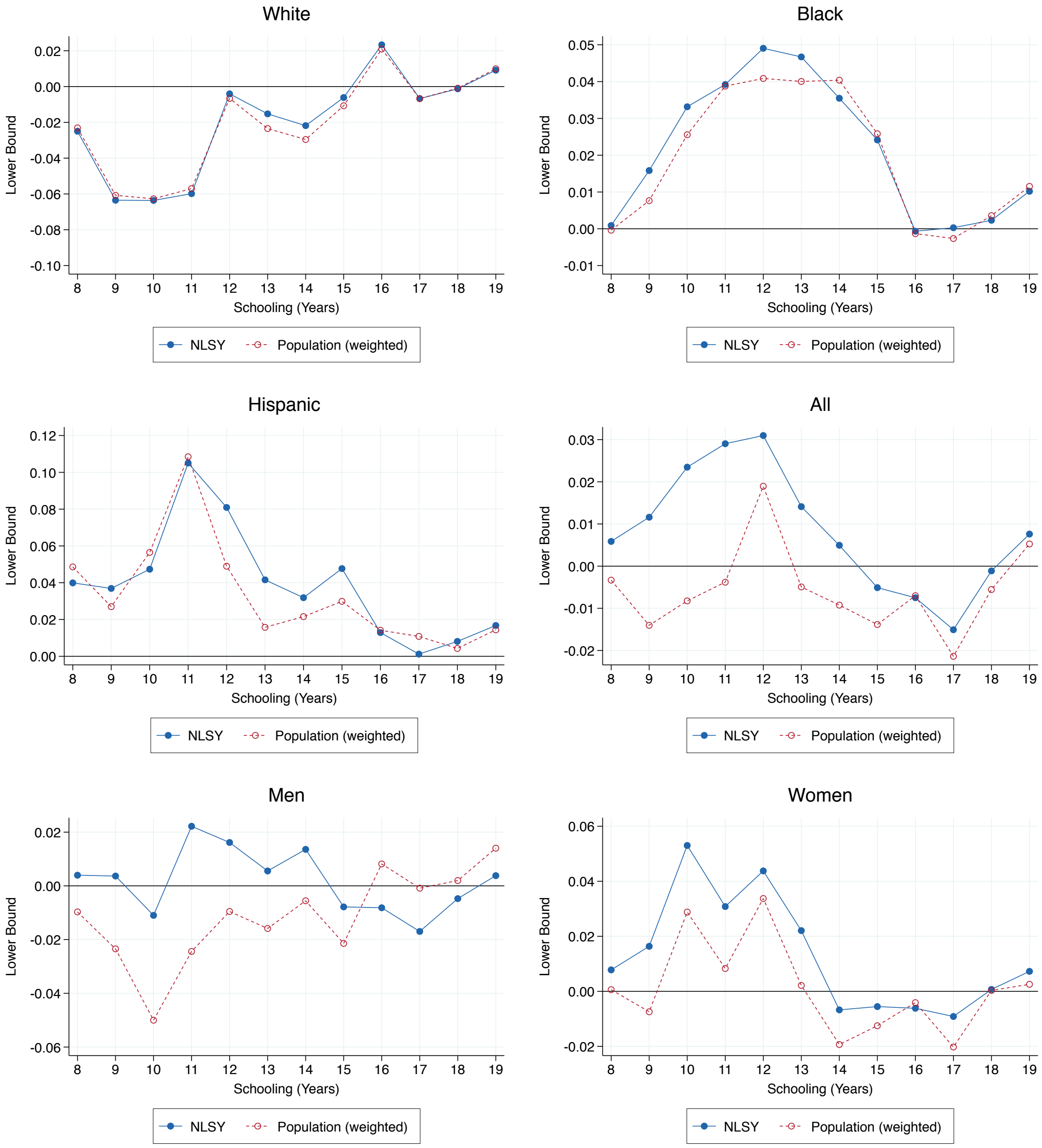


Figure A9. MTS-MIV bounds on the effect of Head Start on Education – Baseline and weighted estimates

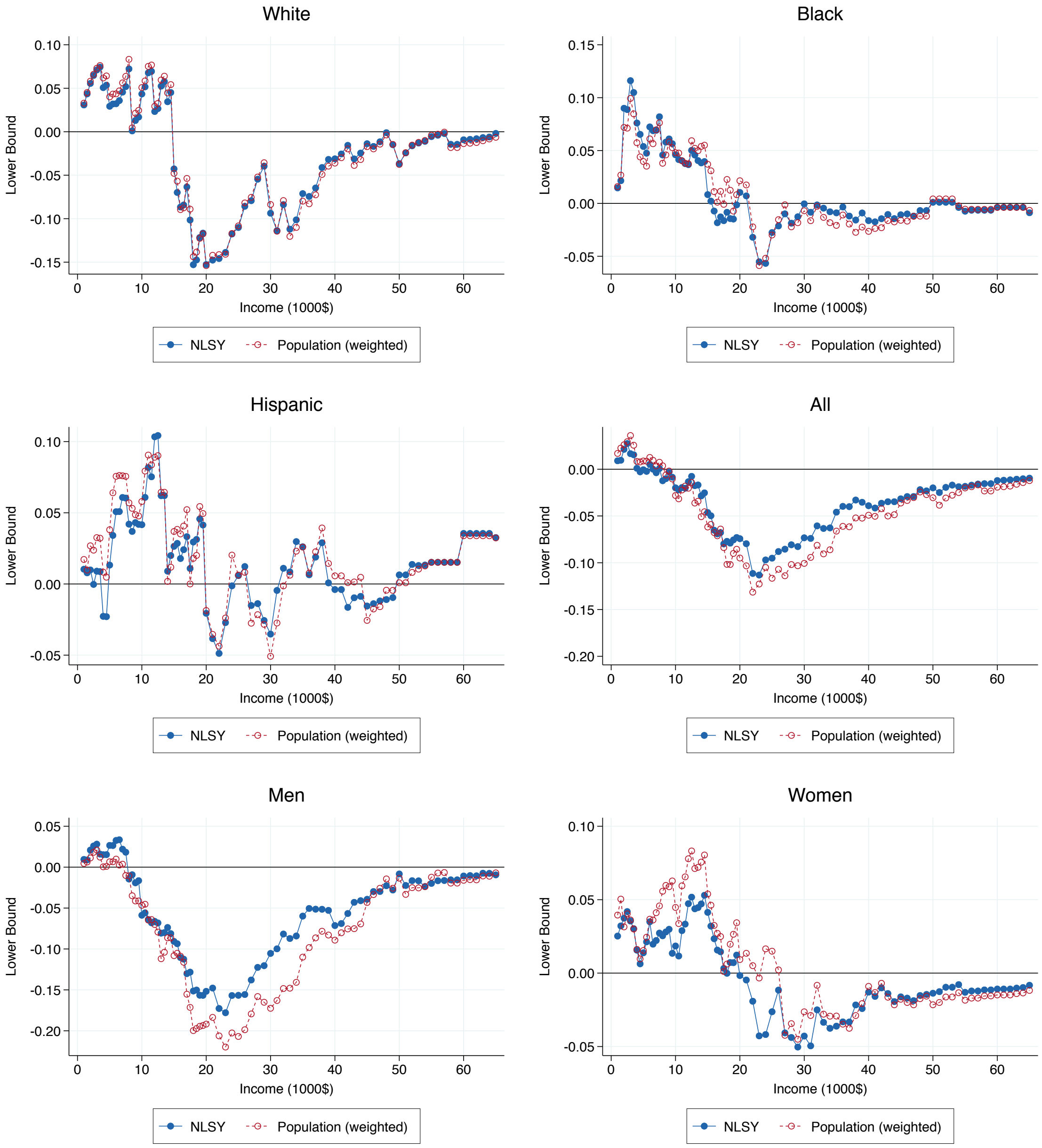


Figure A10. MTS-MIV bounds on the effect of Head Start on Income – Baseline and weighted estimates

A.2 The consequences of treatment misreporting for non-parametric bounds estimation

Misreporting In the following we assume that participants ($D^* = 1$) might falsely report that they did not participate in Head Start but that non-participants ($D^* = 0$) will never falsely report that they attended Head Start:

$$\Pr(\text{false positive}) = \Pr(D = 1, D^* = 0) = 0$$

We use the following notation

$$\begin{aligned}
\phi &\equiv \Pr(\text{false negative}) &&= \Pr(D = 0, D^* = 1) \\
p^* &\equiv \Pr(D^* = 1) && \\
p &\equiv \Pr(D = 1) &&= \Pr(D = 1, D^* = 1) + \Pr(D = 1, D^* = 0) \\
& &&= \Pr(D = 1, D^* = 1) \\
& &&= \Pr(D^* = 1) - \Pr(D = 0, D^* = 1) \\
& &&= p^* - \phi
\end{aligned}$$

where p^* is the true share of Head Start participants in the population and p is the reported share of participants.

When we allow for misreporting the cumulative potential outcome distributions can be decomposed as follows:

$$F_{Y(1)}(\gamma) = F_{Y(1)}(\gamma|D^* = 1) \cdot p^* + F_{Y(1)}(\gamma|D^* = 0) \cdot (1 - p^*) \tag{17}$$

$$F_{Y(0)}(\gamma) = F_{Y(0)}(\gamma|D^* = 1) \cdot p^* + F_{Y(0)}(\gamma|D^* = 0) \cdot (1 - p^*) \tag{18}$$

We start with rewriting equation (17) such that we can show how to obtain bounds around this cumulative potential outcome distribution when we allow for misreporting of Head

Start participation. Note that we can decompose $F_{Y(1)}(\gamma|D^*=1)$ as follows

$$\begin{aligned}
F_{Y(1)}(\gamma|D^*=1) &= F_{Y(1)}(\gamma|D^*=1,D=1)\cdot\Pr(D=1|D^*=1) \\
&\quad +F_{Y(1)}(\gamma|D^*=1,D=0)\cdot\Pr(D=0|D^*=1) \\
&= F_{Y(1)}(\gamma|D^*=1,D=1)\cdot\tfrac{\Pr(D=1,D^*=1)}{\Pr(D^*=1)} \\
&\quad +F_{Y(1)}(\gamma|D^*=1,D=0)\cdot\tfrac{\Pr(D=0,D^*=1)}{\Pr(D^*=1)} \\
&= F_{Y(1)}(\gamma|D^*=1,D=1)\cdot\tfrac{p}{p^*}+F_{Y(1)}(\gamma|D^*=1,D=0)\cdot\tfrac{\phi}{p^*}
\end{aligned}$$

Since we assume that there are no false positives, we have that

$$F_{Y(1)}(\gamma|D^*=1,D=1)=F_{Y(1)}(\gamma|D=1)=F_Y(\gamma|D=1)$$

Which implies that

$$F_{Y(1)}(\gamma|D^*=1) \quad =F_Y(\gamma|D=1)\cdot\tfrac{p}{p^*}+F_{Y(1)}(\gamma|D^*=1,D=0)\cdot\tfrac{\phi}{p^*} \tag{19}$$

Substituting for $F_{Y(1)}(\gamma|D^*=1)$ in equation (17) gives

$$F_{Y(1)}(\gamma)=F_Y(\gamma|D=1)\cdot p+F_{Y(1)}(\gamma|D^*=1,D=0)\cdot\phi+F_{Y(1)}(\gamma|D^*=0)\cdot(1-p^*) \tag{20}$$

Next, we consider the cumulative potential outcome distribution with no Head Start as potential treatment in Equation (18). Note that we can decompose $F_{Y(0)}(\gamma|D^*=0)$ as follows

$$\begin{aligned}
F_{Y(0)}(\gamma|D^*=0) = &\quad F_{Y(0)}(\gamma|D^*=0,D=1)\cdot\Pr(D=1|D^*=0) \\
&\quad +F_{Y(0)}(\gamma|D^*=0,D=0)\cdot\Pr(D=0|D^*=0) \\
= &\quad F_{Y(0)}(\gamma|D^*=0,D=1)
\end{aligned}$$

where the last lines follows from the assumption of no false positives ($\Pr(D=1|D^*=0)=$

0 and $\Pr(D=0|D^*=0)=1$). We further have that

$$\begin{aligned}F_{Y(0)}(\gamma|D=0) = & \ F_{Y(0)}(\gamma|D^*=0,D=0)\cdot\Pr(D^*=0|D=0)\\ & +F_{Y(0)}(\gamma|D^*=1,D=0)\cdot\Pr(D^*=1|D=0)\\ = & \ F_{Y(0)}(\gamma|D^*=0,D=0)\cdot\tfrac{\Pr(D^*=0,D=0)}{\Pr(D=0)}\\ & +F_{Y(0)}(\gamma|D^*=1,D=0)\cdot\tfrac{\Pr(D^*=1,D=0)}{\Pr(D=0)}\\ = & \ F_{Y(0)}(\gamma|D^*=0,D=0)\cdot\tfrac{1-p^*}{1-p}\\ & +F_{Y(0)}(\gamma|D^*=1,D=0)\cdot\tfrac{\phi}{1-p}\end{aligned}$$

Rewriting to get $F_{Y(0)}(\gamma|D^*=0,D=0)$ to the left-hand side of the equation gives

$$F_{Y(0)}(\gamma|D^*=0,D=0)=\frac{1-p}{1-p^*}F_{Y(0)}(\gamma|D=0)-\frac{\phi}{1-p^*}F_{Y(0)}(\gamma|D^*=1,D=0)$$

Since we have that $F_{Y(0)}(\gamma|D^*=0)=F_{Y(0)}(\gamma|D^*=0,D=0)$ we get

$$F_{Y(0)}(\gamma|D^*=0)=\frac{1-p}{1-p^*}F_{Y(0)}(\gamma|D=0)-\frac{\phi}{1-p^*}F_{Y(0)}(\gamma|D^*=1,D=0) \tag{21}$$

Finally, substituting for $F_{Y(0)}(\gamma|D^*=0)$ in equation (18) gives

$$F_{Y(0)}(\gamma)=F_{Y(0)}(\gamma|D^*=1)\cdot p^*+(1-p)\cdot F_{Y(0)}(\gamma|D=0)-\phi\cdot F_{Y(0)}(\gamma|D^*=1,D=0) \tag{22}$$

No-Assumption bounds In order to obtain No-Assumption bounds around $F_{Y(1)}(\gamma)$ we replace the unobserved cumulative potential outcome distributions, $F_{Y(1)}(\gamma|D^*=0)$ and $F_{Y(1)}(\gamma|D^*=1,D=0)$ in equation (20) by 0 and 1.

$$F_Y(\gamma|D=1)\cdot p\leq F_{Y(1)}(\gamma)\leq F_Y(\gamma|D=1)\cdot p+\phi+(1-p^*) \tag{23}$$

since $\phi+(1-p^*)=1-p$, allowing for false negatives does not change the No-Assumption bounds around $F_{Y(1)}(\gamma)$.

In order to obtain the No-Assumption lower bound around $F_{Y(0)}(\gamma)$ we replace the unobserved cumulative potential outcome distributions in Equation (22), $F_{Y(0)}(\gamma|D^*=1)$ by 0 and $F_{Y(0)}(\gamma|D^*=1,D=0)$ by 1.[19] To obtain the No-Assumption upper bound we replace $F_{Y(0)}(\gamma|D^*=1)$ by 1 and $F_{Y(0)}(\gamma|D^*=1,D=0)$ by 0.[20]

$$(1-p)\cdot F_Y(\gamma|D=0)-\phi \leq F_{Y(0)}(\gamma) \leq p^*+(1-p)\cdot F_Y(\gamma|D=0) \tag{24}$$

MTS-bounds Under the MTS assumption we have the following

$$F_{Y(0)}(\gamma|D^*=1) \geq F_{Y(0)}(\gamma|D^*=0)$$
$$F_{Y(1)}(\gamma|D^*=1) \geq F_{Y(1)}(\gamma|D^*=0)$$

We can use the MTS assumption to tighten the upper bound on $F_{Y(1)}(\gamma)$, because we can use $F_{Y(1)}(\gamma|D^*=1)$ as an upper bound for $F_{Y(1)}(\gamma|D^*=0)$ (as opposed to replacing it by 1 as in the No-Assumption bounds). This implies that we can use Equation (19) as an MTS upper bound on $F_{Y(1)}(\gamma)$.

$$F_Y(\gamma|D=1)\cdot p \quad \leq F_{Y(1)}(\gamma) \quad \leq F_Y(\gamma|D=1)\cdot \tfrac{p}{p^*}+F_{Y(1)}(\gamma|D^*=1,D=0)\,\tfrac{\phi}{p^*}$$

Finally we can replace $F_{Y(1)}(\gamma|D^*=1,D=0)$ by 1 to get the MTS bounds

$$F_Y(\gamma|D=1)\cdot p \quad \leq F_{Y(1)}(\gamma) \quad \leq F_Y(\gamma|D=1)\cdot \tfrac{p}{p^*}+\tfrac{\phi}{p^*} \tag{25}$$

Similarly we can use the MTS assumption to tighten the lower bound on $F_{Y(0)}(\gamma)$, because we can use $F_{Y(0)}(\gamma|D^*=0)$ as a lower bound for $F_{Y(0)}(\gamma|D^*=1)$ (as opposed to replacing it by 0 as in the No-Assumption bounds). This implies that we can use equation

[19] Due to the minus sign we get a lower bound by replacing $F_{Y(0)}(\gamma|D^*=1,D=0)$ by 1 (instead of 0).

[20] Due to the minus sign we get an upper bound by replacing $F_{Y(0)}(\gamma|D^*=1,D=0)$ by 0 (instead of 1).

(21) as an MTS lower bound on $F_{Y(0)}(\gamma)$.

$$\frac{1-p}{1-p^*}F_{Y(0)}(\gamma|D=0)-F_{Y(0)}(\gamma|D^*=1,D=0)\cdot\frac{\phi}{1-p^*}$$
$$\leq F_{Y(0)}(\gamma)\leq p^*+(1-p)\cdot F_Y(\gamma|D=0)$$

Finally we can replace $F_{Y(0)}(\gamma|D^*=1,D=0)$ by 1 to get the MTS bounds

$$\tfrac{1-p}{1-p^*}F_{Y(0)}(\gamma|D=0)-\tfrac{\phi}{1-p^*}\quad\leq F_{Y(0)}(\gamma)\quad\leq p^*+(1-p)\cdot F_Y(\gamma|D=0) \tag{26}$$

Monotone reporting selection Under the monotone reporting selection assumption we assume that among the treated ($D^*=1$) the mis-reporters are weakly positively selected

$$F_{Y(h)}(\gamma|D^*=1,D=0)\leq F_{Y(h)}(\gamma|D^*=1)\leq F_{Y(h)}(\gamma|D^*=1,D=1)$$

Using this assumption allows us to tighten the MTS upper bound on $F_{Y(1)}(\gamma)$ because we can replace $F_{Y(1)}(\gamma|D^*=1,D=0)$ by $F_Y(\gamma|D=1)$ since

$$F_{Y(1)}(\gamma|D^*=1,D=0)\leq F_{Y(1)}(\gamma|D^*=1,D=1)=F_Y(\gamma|D=1)$$

where the last equality follows from the assumption of no false positives. This give the following MTS-MRS bounds on $F_{Y(1)}(\gamma)$

$$F_Y(\gamma|D=1)\cdot p\leq F_{Y(1)}(\gamma)\leq F_Y(\gamma|D=1) \tag{27}$$

We can also use the MRS assumption to tighten the bounds around $F_{Y(0)}(\gamma)$. Note that

$$\begin{aligned}
F_Y(\gamma|D=0) &= F_Y(\gamma|D=0,D^*=0)\cdot\Pr(D^*=0|D=0) \\
&\quad +F_Y(\gamma|D=0,D^*=1)\cdot\Pr(D^*=1|D=0) \\
&= F_Y(\gamma|D=0,D^*=0)\cdot\tfrac{\Pr(D^*=0,D=0)}{P(D=0)} \\
&\quad +F_Y(\gamma|D=0,D^*=1)\cdot\tfrac{\Pr(D^*=1,D=0)}{\Pr(D=0)} \\
&= F_Y(\gamma|D=0,D^*=0)\cdot\tfrac{1-p^*}{1-p} \\
&\quad +F_Y(\gamma|D=0,D^*=1)\cdot\tfrac{\phi}{1-p} \\
&= F_{Y(0)}(\gamma|D=0,D^*=0)\cdot\tfrac{1-p^*}{1-p} \\
&\quad +F_{Y(1)}(\gamma|D=0,D^*=1)\cdot\tfrac{\phi}{1-p} \\
&= F_{Y(0)}(\gamma|D^*=0)\cdot\tfrac{1-p^*}{1-p} \\
&\quad +F_{Y(1)}(\gamma|D=0,D^*=1)\cdot\tfrac{\phi}{1-p}
\end{aligned}$$

Where the last line follows from the assumption of no false positives. Rewriting to get $F_{Y(0)}(\gamma|D^*=0)$ to the left-hand side gives

$$\begin{aligned}
F_{Y(0)}(\gamma|D^*=0) &= F_Y(\gamma|D=0)\cdot\frac{1-p}{1-p^*}-\frac{\phi}{(1-p^*)}\cdot F_{Y(1)}(\gamma|D=0,D^*=1) \\
&= F_Y(\gamma|D=0)-\frac{\phi}{1-p^*}\cdot\left(F_{Y(1)}(\gamma|D=0,D^*=1)-F_Y(\gamma|D=0)\right)
\end{aligned}$$

Under the MTS assumption we have that $F_{Y(0)}(\gamma)\geq F_{Y(0)}(\gamma|D^*=0)$. Substituting the above implies that

$$\begin{aligned}
F_{Y(0)}(\gamma) &\geq F_Y(\gamma|D=0)-\frac{\phi}{1-p^*}\cdot\left(F_{Y(1)}(\gamma|D=0,D^*=1)-F_Y(\gamma|D=0)\right) \\
&\geq F_Y(\gamma|D=0)-\frac{\phi}{1-p^*}\cdot\left(F_Y(\gamma|D=1)-F_Y(\gamma|D=0)\right)
\end{aligned}$$

where the last line follows from the MRS assumption. We thus have the following MTS-MRS bounds on $F_{Y(0)}(\gamma)$

$$
\begin{aligned}
&F_Y(\gamma|D=0) - \frac{\phi}{1-p^*}\cdot\left(F_Y(\gamma|D=1) - F_Y(\gamma|D=0)\right)\\
&\qquad\qquad\qquad\qquad \leq F_{Y(0)}(\gamma) \leq p^* + (1-p)\cdot F_Y(\gamma|D=0)
\end{aligned}
$$

MTS-MRS lower bound on the effect of Head Start The causal effect is defined as follows

$$
\triangle(\gamma) = \Pr(Y(1) > \gamma) - \Pr(Y(0) > \gamma) = F_{Y(0)}(\gamma) - F_{Y(1)}(\gamma)
$$

and the lower bound on the causal effect is obtained in the following way

$$
\begin{aligned}
LB_{\triangle(\gamma)} &= LB_{F_{Y(0)}(\gamma)} - UB_{F_{Y(1)}(\gamma)}\\
&= \left(F_Y(\gamma|D=0) - \frac{\phi}{1-p^*}\cdot\left(F_Y(\gamma|D=1) - F_Y(\gamma|D=0)\right)\right) - \left(F_Y(\gamma|D=1)\right)\\
&= \left(1 + \frac{\phi}{1-p^*}\right)\left(F_Y(\gamma|D=0) - F_Y(\gamma|D=1)\right)
\end{aligned}
$$