DISCUSSION PAPER SERIES

IZA

IZA DP No. 9915

Head Start and the Distribution of Long Term Education and Labor Market Outcomes

Monique De Haan
Edwin Leuven

April 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Head Start and the Distribution of Long Term Education and Labor Market Outcomes

Monique De Haan
University of Oslo,
CESifo, ESOP and Statistics Norway

Edwin Leuven
University of Oslo,
IZA, CESifo, CEPR and Statistics Norway

Discussion Paper No. 9915
April 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9915
April 2016

ABSTRACT

Head Start and the Distribution of Long Term Education and Labor Market Outcomes*

In this paper we investigate the effect of Head Start on long term education and labor market outcomes using data from the NLSY79. The contributions to the existing literature on the effectiveness of Head Start are threefold: (1) we are the first to examine distributional effects of Head Start on long term outcomes (2) we do not rely on quasi-experimental variation in Head Start participation but instead perform a nonparametric bounds analysis that relies on weak stochastic dominance assumptions and (3) we consider education and labor market outcomes observed for individuals in their early 30s. The results show that Head Start has a statistically significant positive effect on years of education, in particular for women, 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.

JEL Classification: H52, I21, J13, J24, J31

Keywords: Head Start, early intervention, long term outcomes, partial identification

Corresponding author:

Edwin Leuven
Department of Economics
University of Oslo
P.O. Box 1095 Blindern
0317 Oslo
Norway
E-mail: edwin.leuven@econ.uio.no

* We thank David Deming, Jim Heckman and seminar participants for valuable feedback and suggestions.

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, many studies that have evaluated Head Start using quasi-experimental designs find positive effects on medium and longer term outcomes such as high school completion, crime and health outcomes (Currie and Thomas, 1995, 2000; Deming, 2009; Ludwig and Miller, 2007; Carneiro and Ginja, 2014).

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, most studies observe individuals in their teens or early 20s. For certain outcomes, such as crime, these may be appropriate ages to measure the outcome variable. Measuring education in people's early 20s could however lead to truncation because individuals might not have finished their education. Similarly, labor market outcomes are better measured when individuals are in their early 30s if one wants to reduce life-cycle bias (Bhuller et al., 2016; Böhlmark and Lindquist, 2006; Haider and Solon, 2006).

In this study we 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 use a partial identification approach that does not require exogenous variation in Head Start participation but instead relies on weak stochastic dominance assumptions. Second, we investigate the impact of Head Start on education and labor market outcomes observed for individuals in their 30s. Third, instead of estimating a (local) average treatment effect, we estimate upper and lower bounds around 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 distributional impact of Head Start on 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.

To obtain informative bounds on the causal effect of Head Start we rely on two weak stochastic dominance assumptions. Since Head Start is targeted at disadvantaged children we first assume that the potential outcome distributions of Head Start participants are weakly stochastically dominated by those of nonparticipants. 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).

Combining these 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.

The remainder of the paper is organized as follows. In Section 2 we describe the Head Start program, the previous literature and how our paper contributes to this previous literature. In Section 4 we explain the partial identification approach and the identifying assumptions. Section 3 describes our data set, the National Longitudinal Study of Youth 1979, the sample restrictions and the construction of the outcome variables. The results are shown in Section 5 and finally Section 6 summarizes and concludes.

2 Background and literature

In 1965, Head start was launched by the Office of Economic Opportunity (OEO), with the goal to prepare children from disadvantaged backgrounds for compulsory schooling by providing these children and their parents with schooling, health, nutrition, and social welfare services. 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, while there are only a handful of studies that consider long term outcomes. The most recent findings on short term effects come from the Head Start Impact Study (HSIS). In the Head Start Impact Study eligible children were randomly assigned to a treatment group that could enroll in one of the participating Head Start centers or to a control group. This control group could not enroll in one of the participating Head Start centers, but these children could enroll in another preschool program including non-participating Head Start centers. Puma et al. (2010) compare the outcomes of the children in the treatment and control groups and find positive average effects on cognitive outcomes in preschool, but the effects disappear in kindergarten through 3rd grade. Bitler et al. (2014) use data from the Head Start impact Study to investigate the distributional effects of Head Start. They estimate instrumental variable quantile treatment effects and find substantial effects of Head Start on cognitive outcomes in preschool with the largest effects at the bottom end of the distribution and for Hispanics. The effects fade out in elementary school for the full sample, but the cognitive gains persist for some Spanish speakers.

Two recent studies address the issue that the control group could enroll in another preschool program. Feller et al. (2016) use a principal stratification framework and find strong positive short-term effects of Head Start for children whose counter-factual treatment would be home-based care, while they find no meaningful impact of Head Start for children whose counter-factual treatment would be other center-based care. Kline and Walters (2016) use a semi-parametric selection model and obtain similar findings as Feller et al. (2016); positive effects on test scores of Head Start compared to home-based care but Head Start is about as effective at raising test scores as competing preschools.

All the studies on long term effects of Head Start are based on quasi-experimental evidence. Garces et al. (2002) and Deming (2009) both estimate family-fixed effect models and assume that variation in Head Start participation between siblings is exogenous. Garces et al. (2002) use the 1995 wave of the PSID and find no statistically significant average effects on outcomes for individuals in their early 20s. For whites, however, they find a higher likelihood of high school completion (20 percentage points), higher college attendance (28 percentage points) and higher earnings from work (76 percent).[1] They do not find effects on the education and income of African-Americans.[2] Deming (2009) uses the CNLSY (the children of the women in the NLSY79) and finds strong positive effects on a summary index of adult outcomes. In addition he finds that Head Start participation increases high school graduation for blacks and boys (by about 10 percentage points), but not for non-blacks and girls. Effects on college enrollment are also found, in particular for blacks (14 percentage points) and girls (9 percentage points).

Carneiro and Ginja (2014) and Ludwig and Miller (2007) both use (fuzzy) regression discontinuity designs to identify the causal effect of Head Start. Carneiro and Ginja (2014) exploit fuzzy discontinuities based on income eligibility rules. They are unable to estimate effects for girls because their first-stages are insignificant, but they find effects of Head Start on crime and health outcomes for boys. They do not find significant effects on high school completion or college attendance. Ludwig and Miller (2007) exploit a discontinuity in Head Start funding rates at the OEO cutoff for grant-writing assistance. They find a drop in child mortality rates around the cutoff and they report evidence that suggest positive effects on high school completion and college attendance for eligible cohorts.

Although these studies on long-term outcomes tend to find positive effects of Head Start, they differ in the specific long-term outcomes that are affected as well as the subgroups that are found to benefit from Head Start. As pointed out by Elango et al. (2016), it is unclear wether

[1]The coefficient on earnings is not statistically significant.
[2]The do find effects on crime for African-Americans but not for whites.

the lack of consistency between these studies is due to differences in (counterfactual) treatment, differences in population, or problems related to the empirical approach.

We contribute to the existing literature on long term effects by estimating the effect of Head Start on the entire cumulative potential outcome distributions of education and wage income, using two nonparametric weak stochastic dominance assumptions that are explained and checked in Section 4. We will report results separately by gender and by race, as well as results where the counterfactual treatment is home-based care or mixture of home-based care and other preschool.

3 Data

We use 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.

As outcomes in our analysis, we use individuals' highest observed years of education as well as yearly wage income both reported in 1994, when the individuals were in their early 30s.[3,4] 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.[5] We restrict the main sample to Head Start participants and individuals who did not participate in Head Start nor any type of preschool. This means that in the main analysis we will estimate effects of Head Start relative to informal care and not relative to other types of preschool. We will also show results where we include individuals who attended another type of preschool in the estimation sample, this will create a counter-factual which is a mixture of preschool and informal care.

Basic background information such as age (birth year), gender and race is available in the

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

[4]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?".

[5]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?".

Table 1. Descriptive statistics

		Head Start		Race		
	All	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
Highest completed grade	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.

data. The respondents also provided information on parental education. For each parent the highest completed grade was reported. 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 the highest attained grade 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.[6]

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:[7]

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

For example, if Y is years of education then $\Delta(11)$ would be the effect of Head Start on obtaining more than 11 years of education i.e. a high school degree or more (and equivalently $-\Delta(11)$ would be the effect of obtaining no more than 11 years of education). We estimate (1) for values of γ over the whole support of $Y(h)$.

As shown in equation (1), the causal effect is the difference between two cumulative potential outcome distribution 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

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

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 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 that we will impose 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:*

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

Suppose we take the probability of obtaining at most 11 years of education as the potential outcome of interest, $F_{Y(h)}(11|X=x)$. The MIV assumption states that if everyone (or a random person) would receive the same treatment – either Head Start ($h=1$) or no Head Start ($h=0$) – then the probability of obtaining at most 11 years of education will, 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 $x \leq x^*$.

Figure 1 also illustrates how the MIV assumption can be exploited to tighten the bounds around $F_{Y(h)}(\gamma|X)$. It shows fictive upper and lower bounds on $F_{Y(h)}(\gamma|X)$ for three levels of parental education. For individuals whose parents have a middle education level, the potential probability of obtaining an education or labor market outcome of at most γ is assumed to be weakly higher, on average, than for individuals with high educated parents and weakly lower than for individuals with low educated parents. We can therefore use the upper (lower) bound from the sample of those with middle educated parents as an MIV upper (lower) bound for those with high (low) educated parents.

Finally, by taking the weighted average of the MIV bounds over all $x^* \in X$, we obtain the

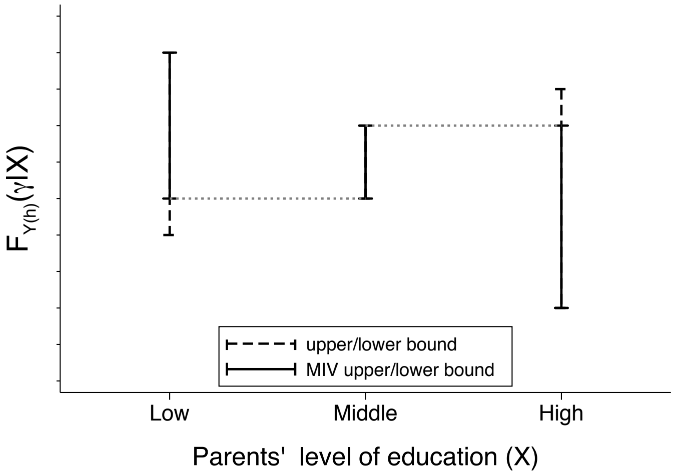


Figure 1. Example of how an MIV can tighten the bounds

following aggregate MIV-bounds on $F_{Y(h)}(\gamma)$.

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

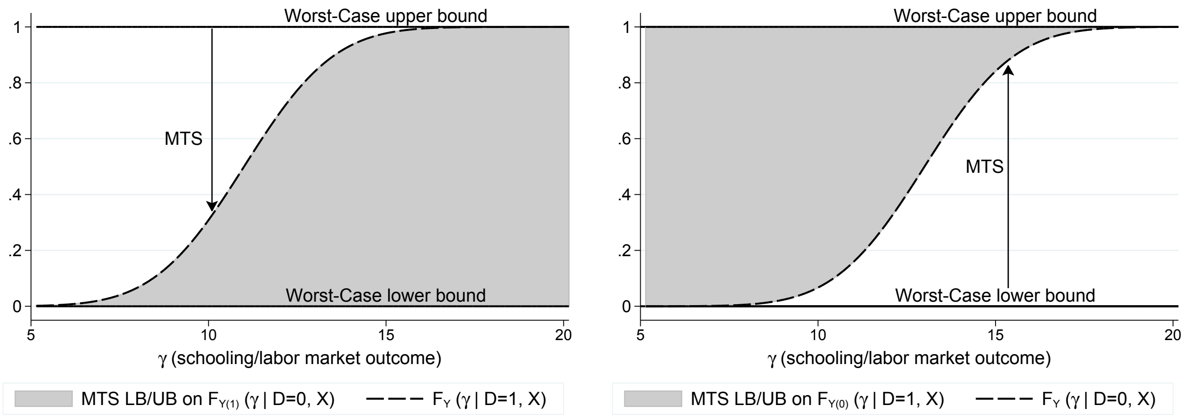
The second weak stochastic dominance assumption that we use to tighten the bounds is the Monotone Treatment Selection assumption, which is motivated by the eligibility criteria of Head Start as described in Section 2. Equation (8) shows the MTS assumption.

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

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

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 nonparticipants ($D = 0$).

Figure 2 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 of the



(a) Bounds around $F_{Y(1)}(\gamma|X)$ for nonparticipants　　**(b)** Bounds around $F_{Y(0)}(\gamma|X)$ for participants

Figure 2. Illustration of the MTS assumption

education or labor market outcome in case of Head Start as potential treatment; $F_{Y(1)}(\gamma|X)$. This cumulative potential outcome distribution is observed for participants and equals $F_Y(\gamma|D=1,X)$, but without imposing additional assumption all we know for the nonparticipants is that it will be between 0 and 1. Under the MTS assumption the potential outcome distribution of nonparticipants weakly stochastically dominates the potential outcome distribution of the participants, which 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 nonparticipants, $F_{Y(1)}(\gamma|D=0,X)$. Panel (b) shows that under a similar reasoning we can use the observed cumulative distribution of the nonparticipants, $F_Y\,(\gamma|D=0,X)$ as a lower bound on the unobserved cumulative potential outcome distribution for the participants, $F_{Y(0)}(\gamma|D=1,X)$. Equation (9) show the MTS bounds.

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

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 $F_{Y(0)}(\gamma)$.[8]

All bounds are consistent under the maintained assumptions but they may have finite-sample biases because bounds shown in Section 5 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 following Kreider and Pepper (2007), who propose a bias-correction method that uses the bootstrap distribution to estimate the finite-sample bias.[9] 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.[10]

4.2 Assumption check

The MIV assumption

The MTS and MIV assumptions are untestable since they involve counter-factual 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 counter-factual 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 3 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

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

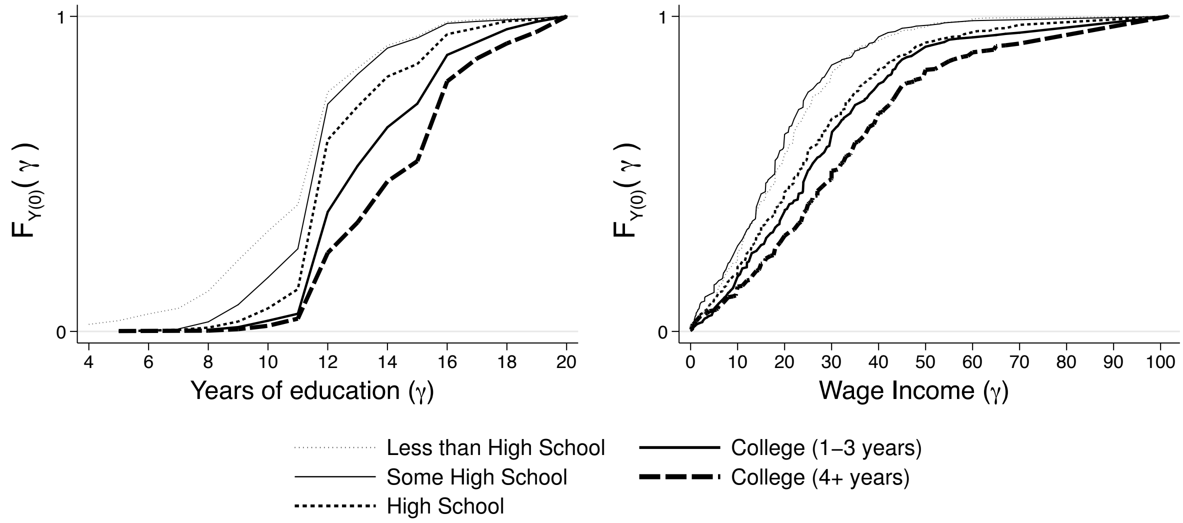
[9] Kreider and Pepper (2007) suggest to estimate the finite sample bias as $\hat{bias} = \left(\frac{1}{K}\sum_{k=1}^{K}\theta_k\right) - \widehat{\theta}$, where $\hat{\theta}$ is the initial estimate of the upper or lower bound and θ_k is the estimate of the k^{th} bootstrap replication. The bias-corrected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

[10] Equation 10 gives the formula for a 95-percent confidence interval:

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

where $\hat{lb}$ and $\hat{ub}$ 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.

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



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

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 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.

In Sections 5.3 and 5.4 we estimate bounds on the effect of Head Start separately by gender and by race and the MIV assumption therefore 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.

Table 2. Test of 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

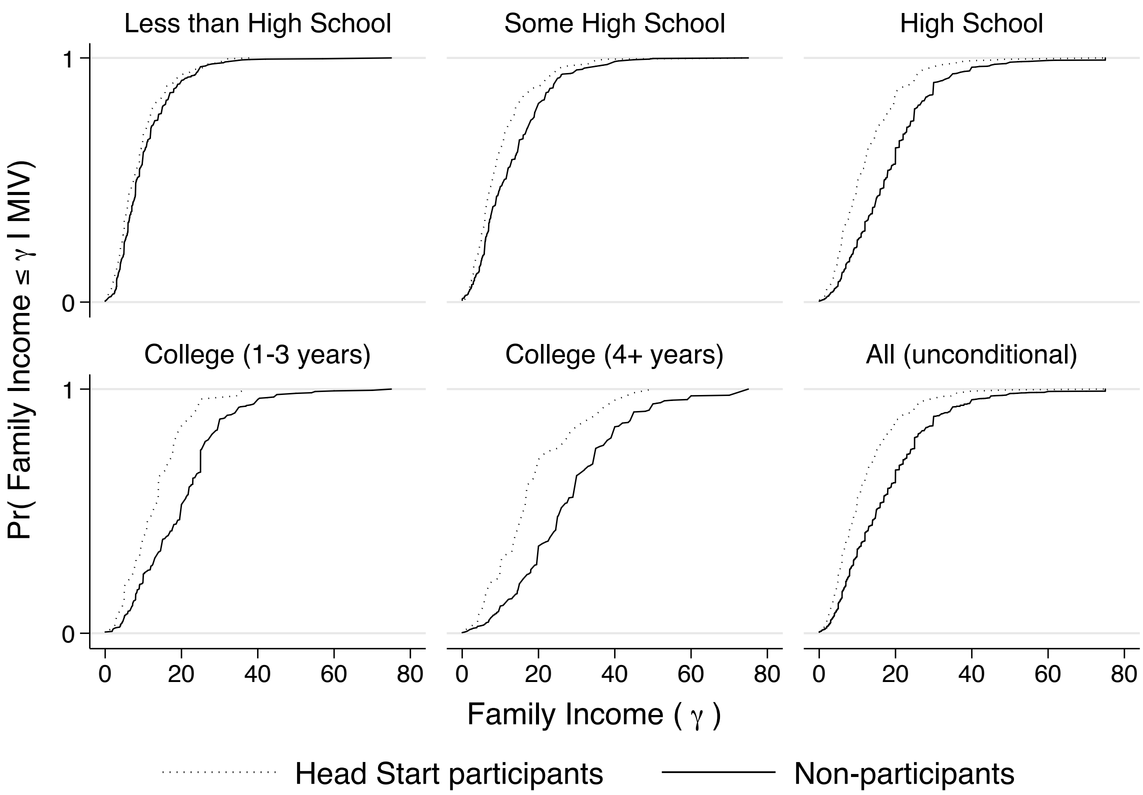
Note: Reported p-values are from one sided Kolmogorov-Smirnov tests, using data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations for education equal 4,873 (all), 2,425 (men), 2,448 (women), 3,172 (white), 1,044 (black), 657 (Hispanic).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

The MTS assumption

The motivation for using the MTS assumption comes from the eligibility rules which result in Head Start participants coming disproportionally from disadvantaged backgrounds, and because of that having potential outcome distributions that are weakly stochastically dominated by those of individuals that did not participate in Head Start. In Section 5 we will show results that are based on both the MTS and the MIV assumption which implies that the MTS should hold for each level of parental education. Although it is not possible to test this identifying assumption we can investigate whether background characteristics of the Head Start participants are indeed weakly stochastically dominated by those of nonparticipants within each sub-sample defined by the values of the MIV.

Figure 4 shows cumulative distributions of family income measured in 1978 when the individuals were between 14 and 18 years old.[11] 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 nonparticipants, which is in line with the MTS assumption. The first column in

[11]Family income could potentially be used as a MIV, but we do not do this for the following reasons. 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 4. MTS Check – Conditional (on MIV) CDFs of Family income at age 14-18 for Head Start participants and non-participants

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 nonparticipants is not rejected at all conventional significant levels. Figures A3 and A4 in the appendix report the cumulative distributions of family income for the participants and nonparticipants, 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 checks in Tables 2 and 3 strongly support our identifying assumptions.

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 4028 (all), 2018 (men), 2010 (women), 1957 (white), 1268 (black), 803 (Hispanic)

4.3 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 will 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 will use the following semi-monotone instrumental variable assumption

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

Suppose we take the probability of obtaining at most 11 years of education as the potential outcome of interest, $F_{Y(h)}(11)$. The MIV assumption states that if everyone (or a random person) would receive the same treatment – either Head Start ($h = 1$) or no Head Start ($h = 0$) – then the probability of obtaining at most 11 years of education will, 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.

Figures A6 and A7 and Table A1 in the appendix show the MIV-assumption check described in Section 4.2 for the case of two MIV's. For years of education as outcome we 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.

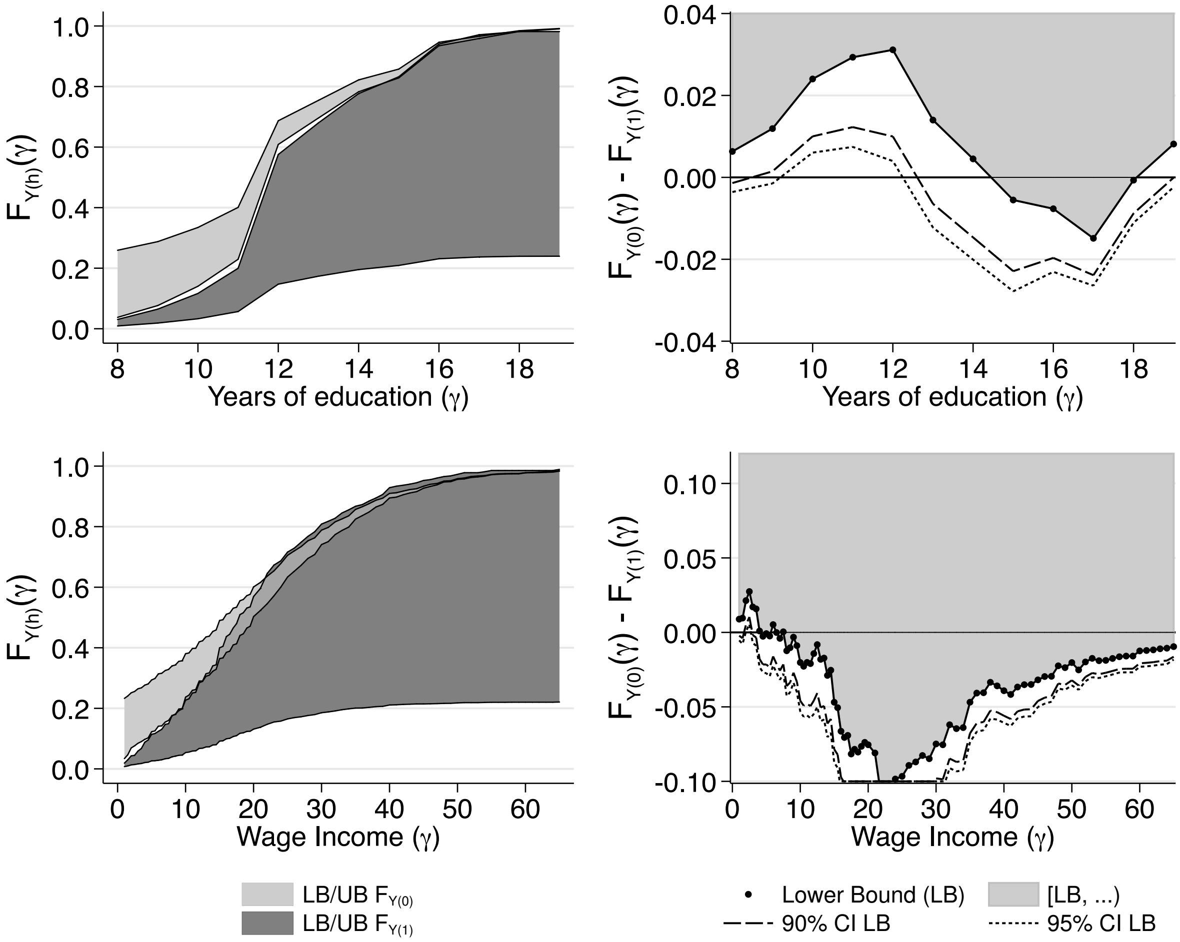
5 The effects of Head Start on long term outcomes

We now present the results from the bounds analysis. We start out by presenting bounds on the cumulative potential outcome distributions as well as lower bounds on the causal effects of Head Start across the distribution. We will first show these for the whole sample, for both education and earnings, after which we will proceed to our subsample analysis where we consider results separately by gender and by race.

5.1 Overall effects

The top left panel of Figure 5 shows the bounds on the cumulative potential outcome distribution of education for the main sample. The light grey area bounds the cumulative potential outcome distribution without Head Start ($F_{Y(0)}(\gamma)$), while the dark grey 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 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 year 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 5 where we bound the cumulative potential outcome distributions. The top right panel in Figure 5 shows the lower

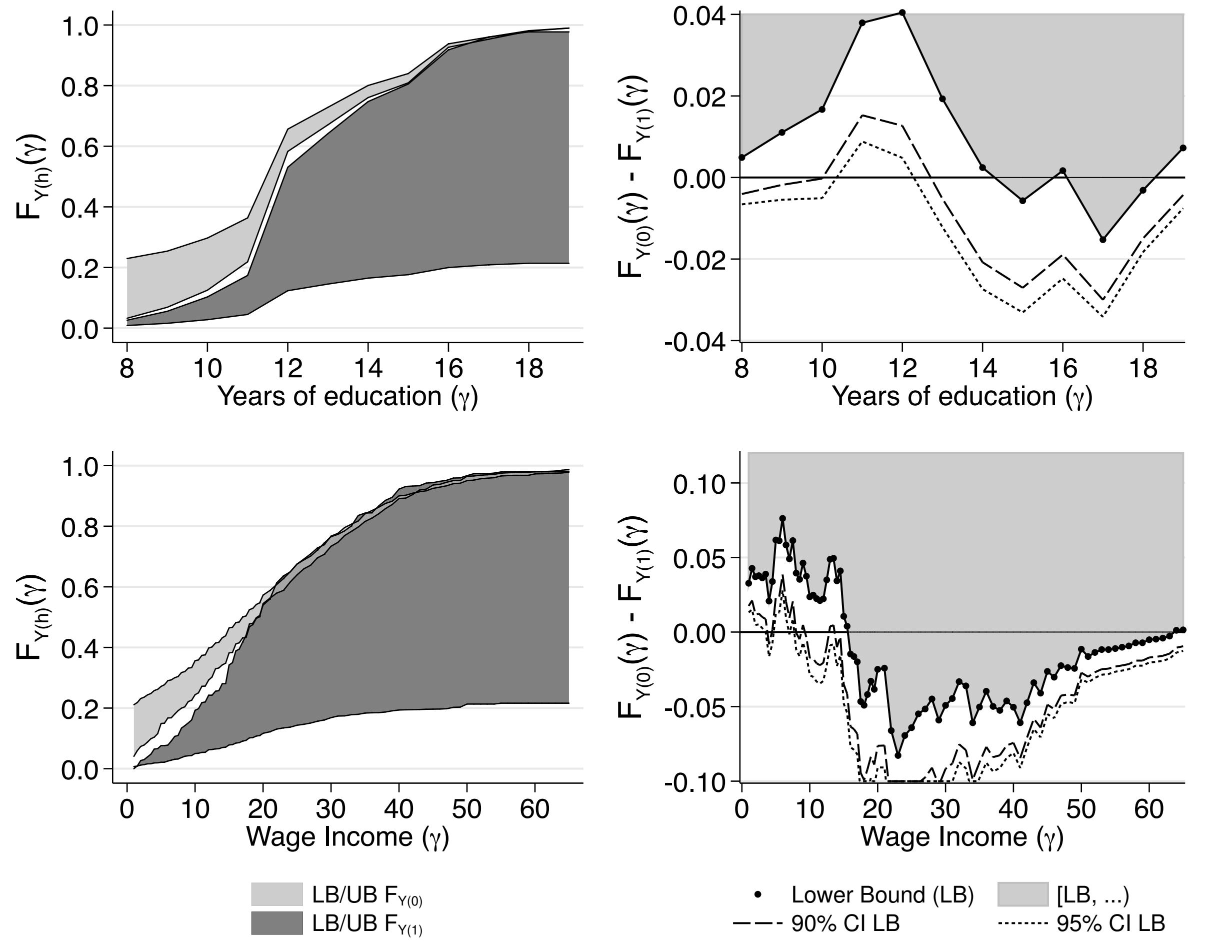


Note: Number of observations equals 4876 (years of education) and 3787 (wage income). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with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

Figure 5. MTS-MIV bounds on the effect of Head Start on education and earnings

bound on this causal effect at the different margins of education. 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. For example, the lower bound at $\gamma = 11$ is 0.03 which means that Head Start increases high school graduation (more than 11 years of education) rates by at least 3 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 top right panel in Figure 5 also shows the (lower bound of the) 90 and 95 percent confidence intervals. We see that 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 5 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



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

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 5 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.

5.2 Combining two monotone instruments

As described in Section 4.3, 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.

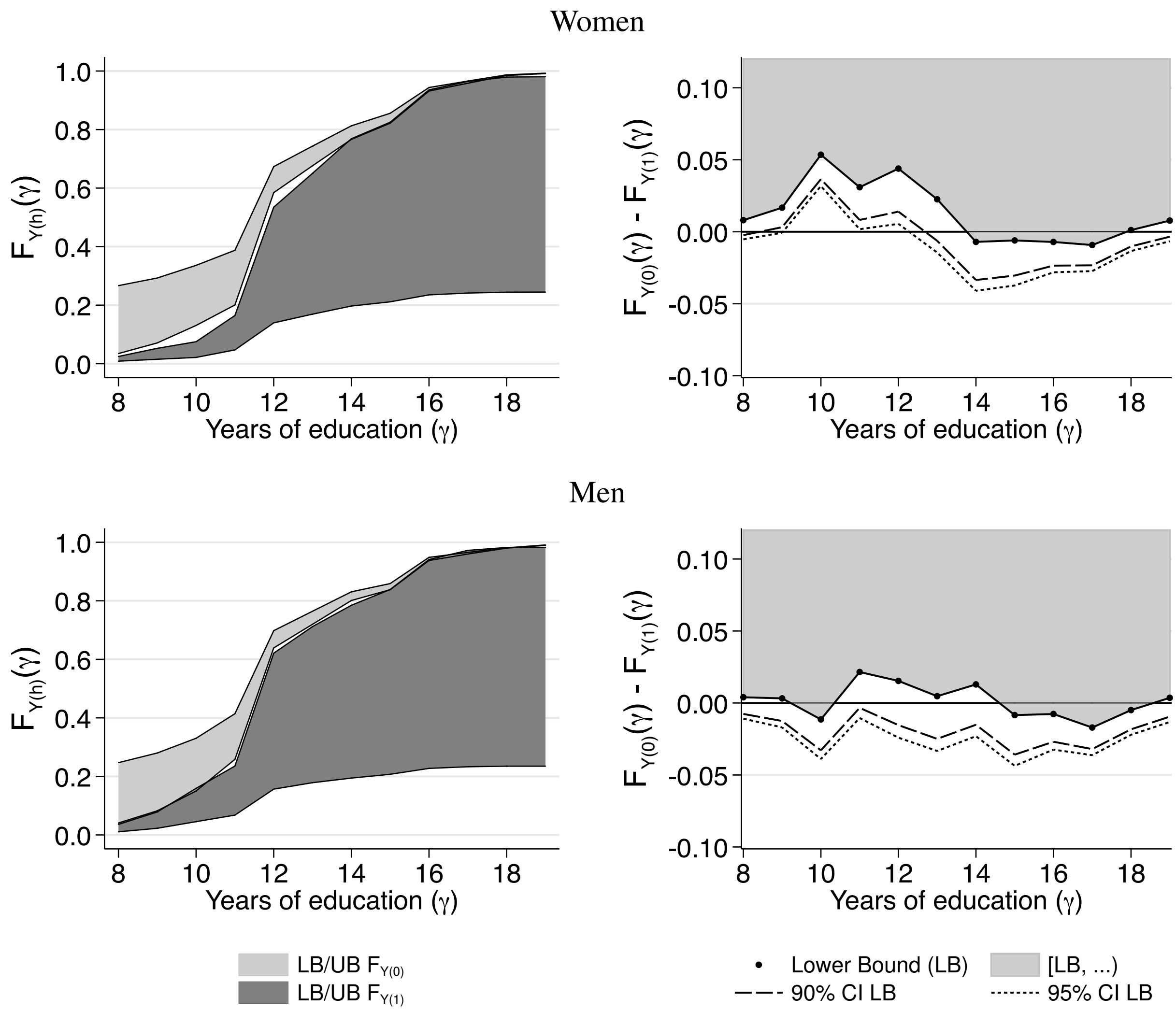
The top two panels of Figure 6 show the results for years of education when we use mother's and father's education as two separate MIV's. The top left panel looks very similar to the top left panel of Figure 5, but from the top right panel it becomes clear that the bounds using two monotone instruments 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.

The bottom two panels of Figure 6 report the bounds for wage income as outcome variable. These bounds are clearly tighter than in Figure 5 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.

5.3 Effects by gender

Many studies have documented that early childhood interventions affect boys and girls 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. In this subsection we estimate bounds on the effect of Head Start separately by gender, after which we will consider outcomes by race in Section 5.4. We will report results where we use the maximum of mother's and father's education as one MIV as we did in Section 5.1, because there are too few observations in (some of) the MIV-categories using two MIV's when we estimate bounds separately by gender and race.

The top right panel of Figure 7 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 7 reports the lower bounds on the effect on education

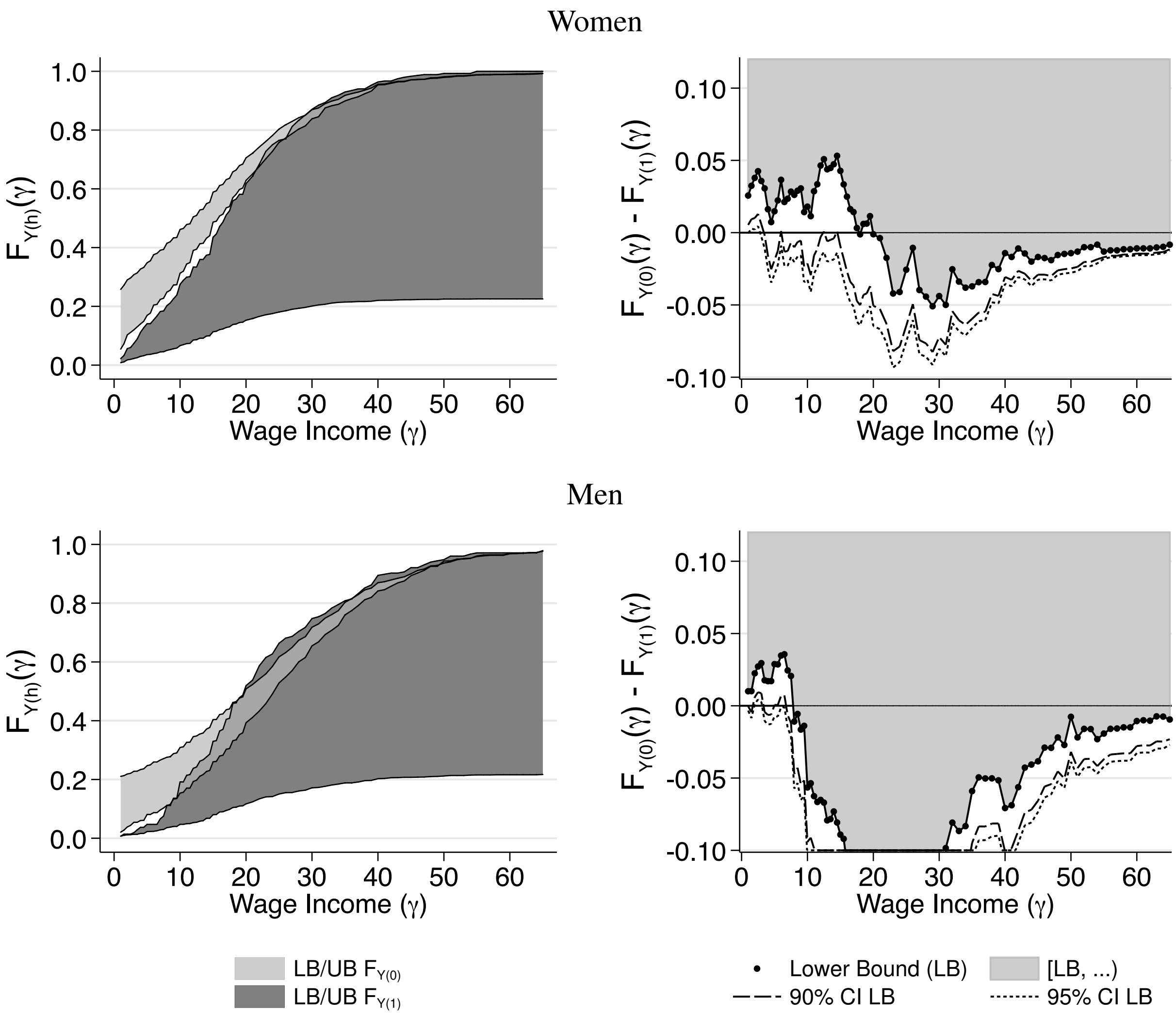


Note: Number of observations equals 2452 (women) and 2424 (men). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 7. The effect of Head Start on years of education – By gender

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 8 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



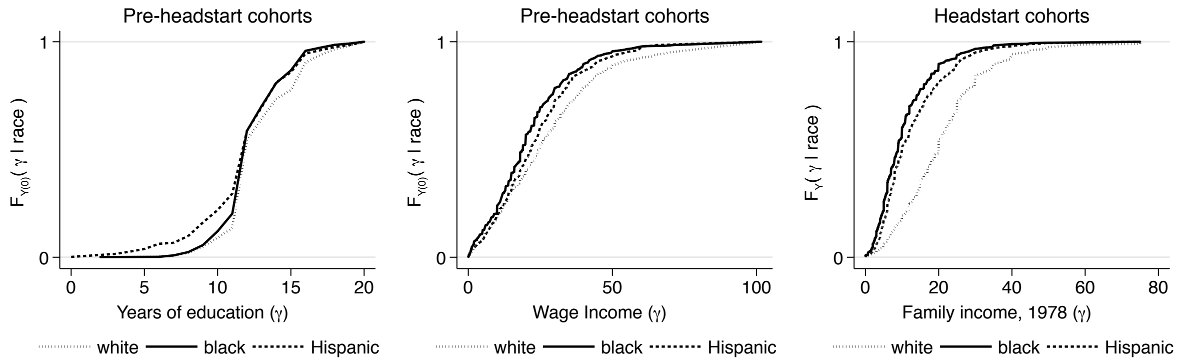
Note: Number of observations equals 1780 (women) and 2007 (men). Estimated bounds are bias-corrected using the bootstrap bias-correcting method proposed by Kreider and Pepper (2007). 90 and 95% confidence intervals are obtained using the method from Imbens and Manski (2004) with 999 bootstrap replications.

Figure 8. The effect of Head Start on wage income in 1993– By gender

of γ. For men we see in the bottom panel of Figure 8 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.

5.4 *Effects by race*

Following the literature we report results by race in Figures 10 and 11. Investigating the impact separately by race is in particular relevant in the context of Head Start, since its eligibility criteria



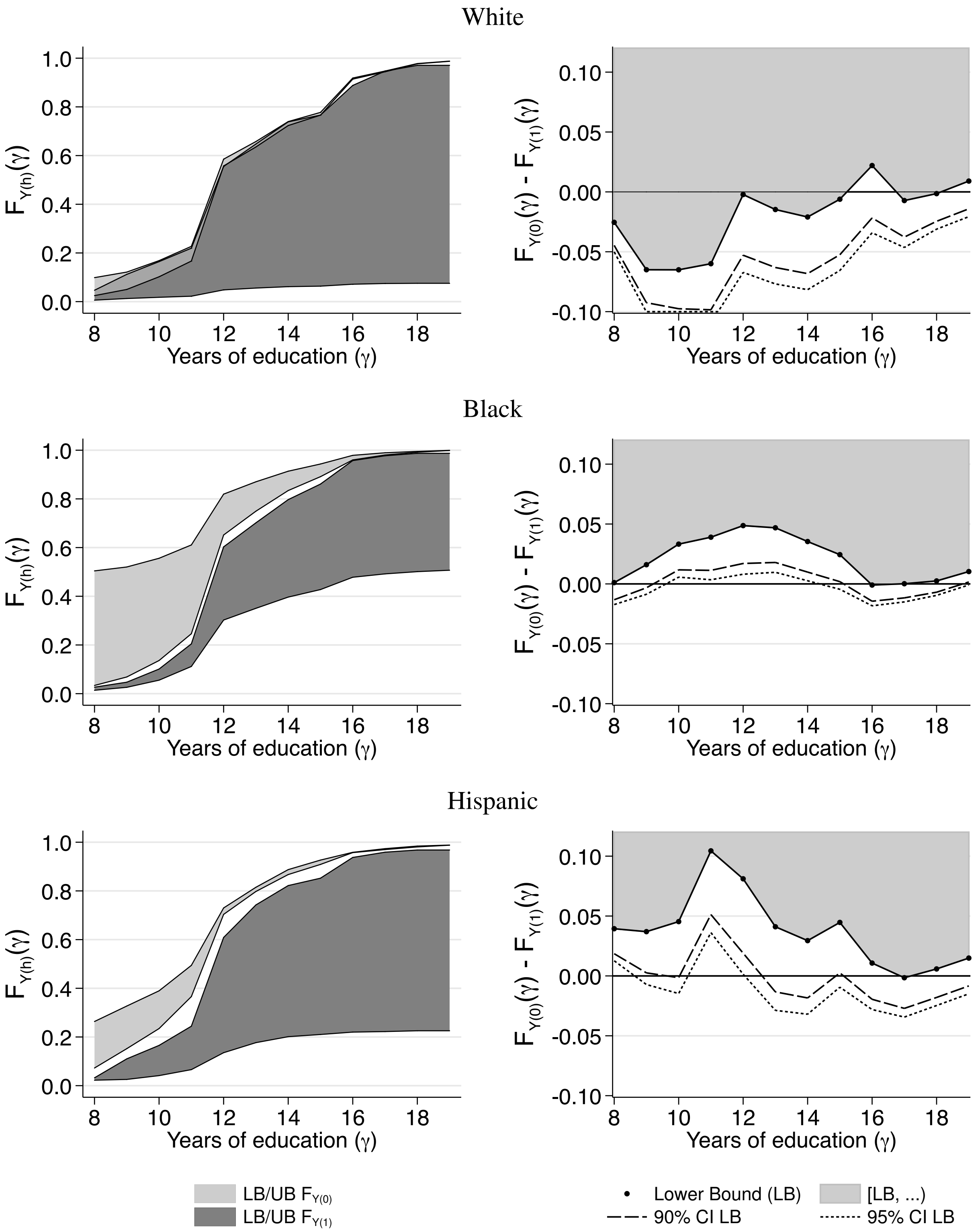
Note: Number of observations equals 4,873 (years of education), 2,153 (wage income, 1993), and 4,028 (family income, 1978)

Figure 9. CDF's of Y(0) in the pre-Head start cohorts and family income by race

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).

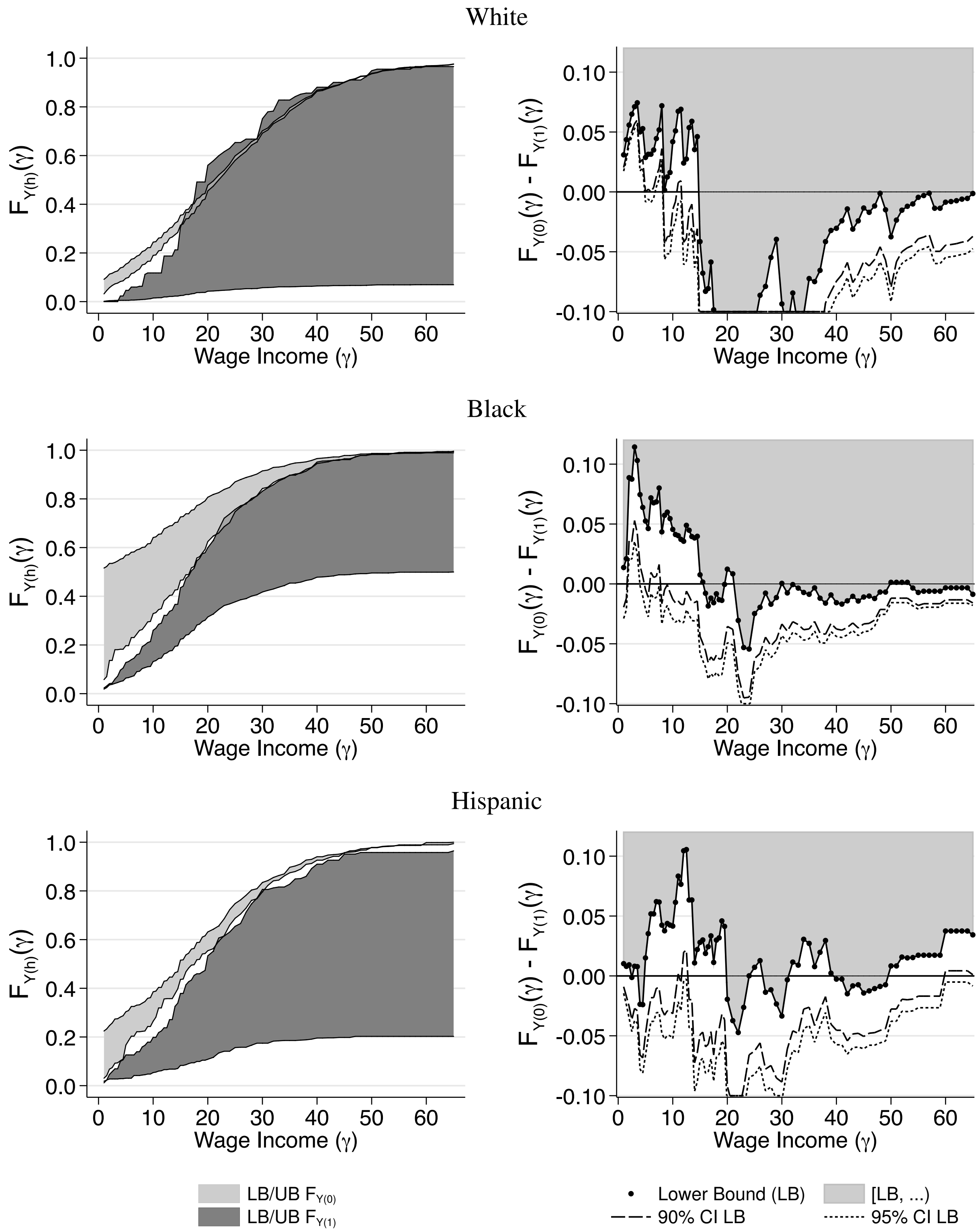
There are also reasons to expect heterogenous effect by race because we find the largest lower bounds at the bottom end of the distribution, which indicates that those with low ability and/or low background characteristics tend to benefit the most from participating in Head Start. That blacks and Hispanics are overrepresented at the lower end of the distribution is illustrated by Figure 9, 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.

We will first discuss the results for education after which we will turn to those for wage income. First consider the top panel in Figure 10 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



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 10. 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 11. The effect of Head Start on wage income in 1993– By race

high school graduation these lower bounds are around 5 percentage points and statistically significant at the 5 percent level.

The bottom panel in Figure 10 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.

Figure 11 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.

5.5 *The importance of the counterfactual*

As indicated by Elango et al. (2016) the evaluations of Head Start report contradictory evidence, in part because these studies compare Head Start with different counterfactual childcare arrangements. In addition, two recent papers, Kline and Walters (2016) and Feller et al. (2016), show that the finding that the effect of Head Start on cognitive test scores fades out is sensitive to the choice of the counter-factual treatment.

So far we have shown results where we compare the effectiveness of Head Start with informal care. To see whether the results are sensitive to the choice of the counter-factual, Figure 12 shows

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

results where we include individuals that attended another non-Head-Start preschool program in the group of nonparticipants. This means that we compare Head Start with a counter-factual 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 nonparticipants. 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 nonparticipants and those that attended another preschool program. This is in line with the MTS assumption.

Figure 12 shows that the results are qualitatively very similar to the results in Figure 5. The lower bounds on the effect of Head Start are however lower in Figure 12, for example, Head Start increases the probability of high school graduation by at least 3 percentage points when the counter-factual is informal care compared to 2 percentage points when the counter-factual is a mixture of informal care and center-based preschool. These results confirm that it is important to be explicit about the counter-factual and that the effects of Head Start seem strongest when informal (home-based) care is the alternative treatment.

5.6 Distributional 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 informative to know how the effects on the treated 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:[12]

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

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 distributional 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).

[12] Although it is straightforward to use the Monotone Treatment Selection assumption to bound the distributional effects of Head Start on the treated, this does not hold for the Monotone Instrumental Variable assumption.

6 Conclusion and discussion

There is currently no consensus in the literature about the effectiveness of Head Start. Many studies document positive effects of Head Start on short term outcomes, yet effects on cognitive outcomes appear to fadeout by 1st grade. It is unclear though whether the observed fadeout is real, or whether it is an artifact of the scaling of test scores (Bond and Lang, 2013; Cascio and Staiger, 2012). In addition as is shown by Kline and Walters (2016) and Feller et al. (2016), the finding of fadeout appears to depend on the counterfactual treatment. Regardless of the existence of any fade out in cognitive test scores, Head Start may also have important effects on non cognitive skills or the home environment, and ultimately one would therefore like to know whether Head Start affects long run life outcomes.

Assessing the effect of Head Start on the long run has however 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 distributional effects of Head Start on long term outcomes. It estimates the 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 and on the distributional effect of Head Start.

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 paint a consistent picture of the distributional long term effects of Head Start, and suggest that

Head Start benefits those that need it the most.

References

Bhuller, M., Mogstad, M., and Salvanes, K. G. (Forthcoming, 2016). Life cycle earnings, education premiums and internal rates of return. *Journal of Labor Economics*.

Bitler, M. P., Hoynes, H. W., and Domina, T. (2014). Experimental evidence on distributional effects of Head Start. Working Paper 20434, National Bureau of Economic Research, Cambridge, MA.

Böhlmark, A. and Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for sweden. *Journal of Labor Economics*, 24(4):879–896.

Bond, T. N. and Lang, K. (2013). The evolution of the black-white test score gap in grades K–3: The fragility of results. *Review of Economics and Statistics*, 95(5):1468–1479.

Carneiro, P. and Ginja, R. (2014). Preventing behavior problems in childhood and adolescence: evidence from Head Start. *American Economic Journal: Economic Policy*, 6(4):135–73.

Cascio, E. U. and Staiger, D. O. (2012). Knowledge, tests, and fadeout in educational interventions. Working Paper 18038, National Bureau of Economic Research, Cambridge, MA.

Currie, J. and Thomas, D. (1995). Does Head Start make a difference? *The American Economic Review*, 85(3):341–364.

Currie, J. and Thomas, D. (2000). School quality and the longer-term effects of Head Start. *Journal of Human Resources*, 35(4):755–74.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Elango, S., Heckman, J. J., Hojman, A., and Garcia, J. L. (2016). Early childhood education. In Moffit, R., editor, *Means-tested Transfer Programs in the United States II*. University of Chicago Press.

Feller, A., Grindal, T., Miratrix, L. W., and Page, L. (Forthcoming, 2016). Compared to what? variation in the impact of early childhood education by alternative care-type settings. *Annals of Applied Statistics*.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-term effects of Head Start. *American Economic Review*, 92(4):999–1012.

Gibbs, C., Ludwig, J., and Miller, D. L. (2011). Does Head Start do any lasting good? In Bailey, M. and Danziger, S., editors, *Legacies of the War on Poverty*, pages 36–65. Russell Sage Foundation, New York.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review*, 96(4):1308–1320.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

Kline, P. and Walters, C. (Forthcoming, 2016). Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics*.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., and Shonko, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, 103(27):10155–10162.

Kreider, B. and Pepper, J. V. (2007). Disability and employment: reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, 102(478):432–441.

Ludwig, J. and Miller, D. L. (2007). Does Head Start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics*, 122(1):159–208.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, 24(3):343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review, Papers & Proceedings*, 80(2):319–323.

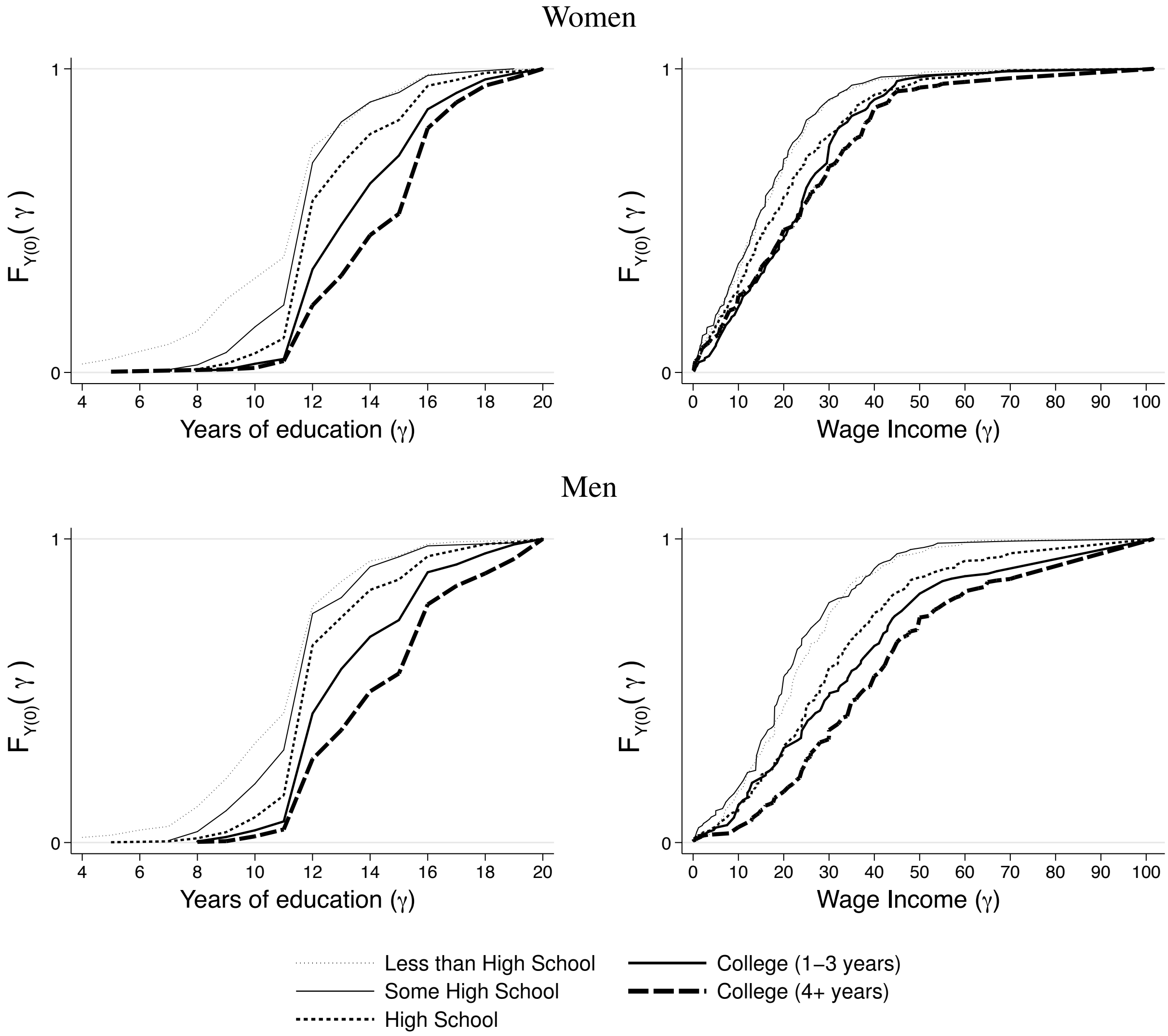
Manski, C. F. (1997). Monotone treatment response. *Econometrica*, 65(6):1311–1334.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68(4):997–1010.

McFadden, D. (1989). Testing for stochastic dominance. In Fomby, T. and Seo, T., editors, *Studies in the Economics of Uncertainty*, pages 113–134. Springer.

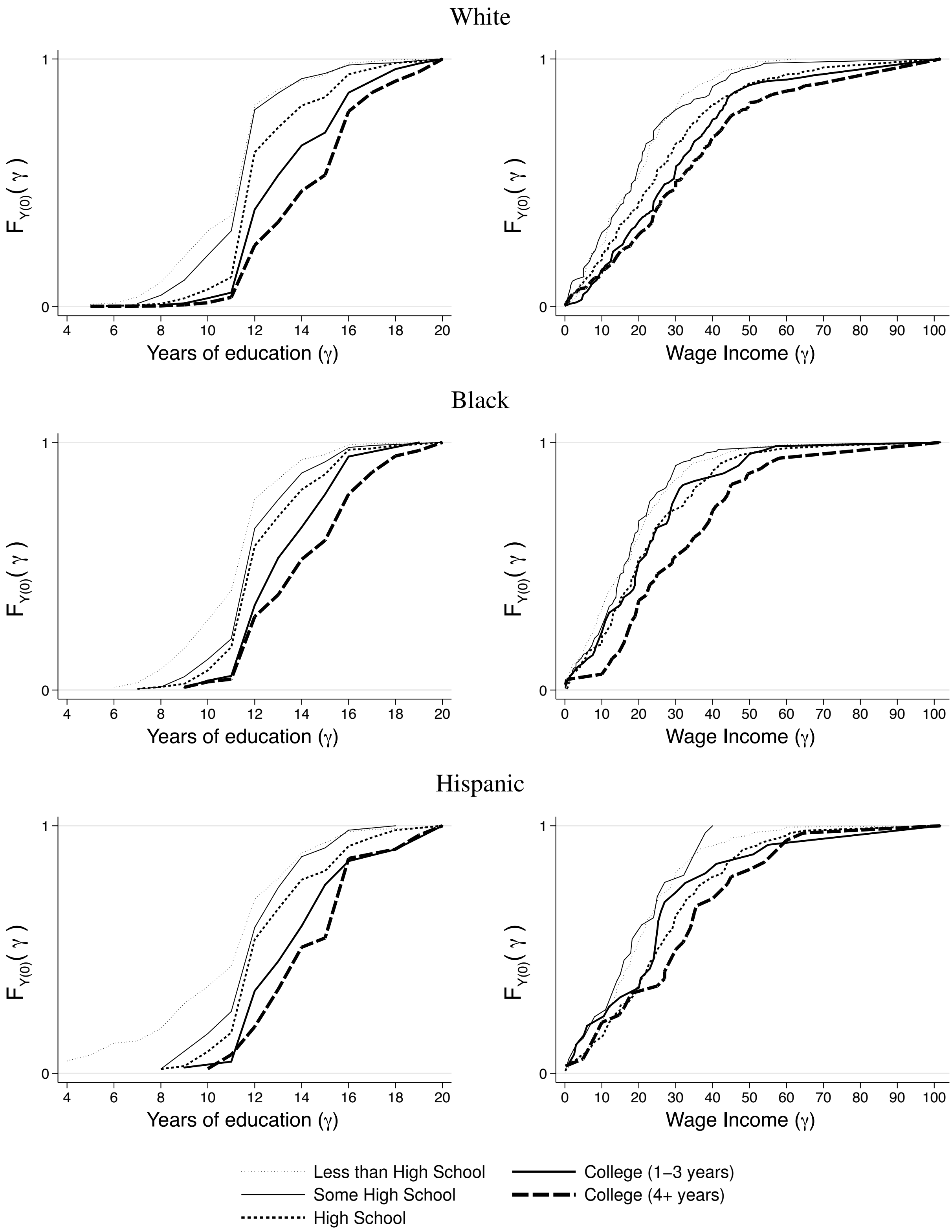
Puma, M., Bell, S., Cook, R., Heid, C., et al. (2010). *Head Start Impact Study. Final Report*. U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC.

A Appendix



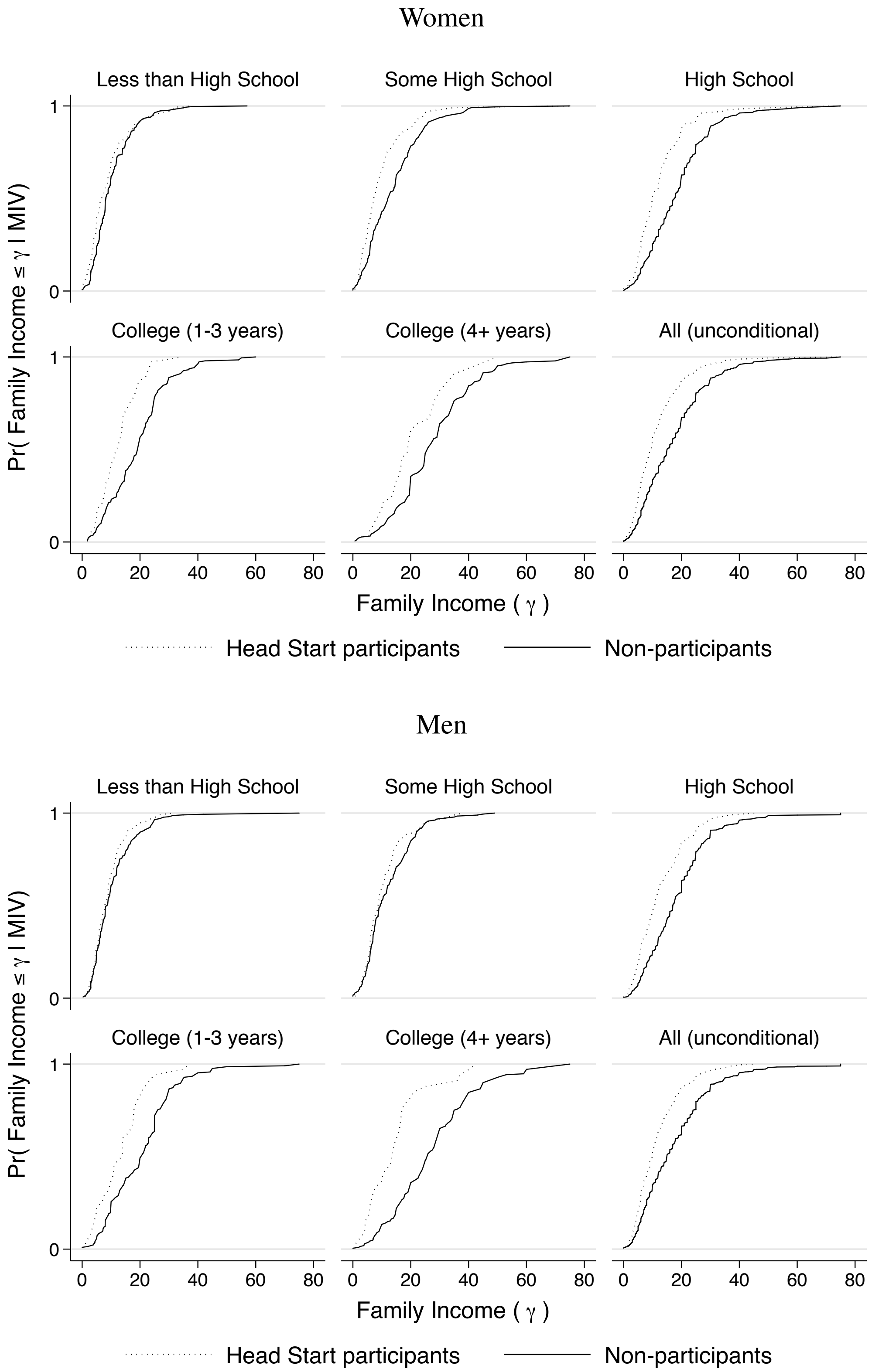
Note: Figures based on data on years of education and wage income for the pre-Head Start cohorts (born between 1957–1959). Number of observations for education equal 2425 (men) and 2448 (women). Number of observations for wage income equal 1099 (men) and 1054 (women).

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).

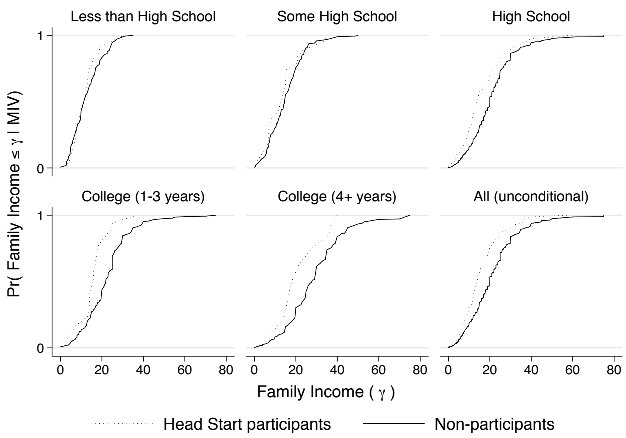
Figure A2. MIV Assumption Check – by race



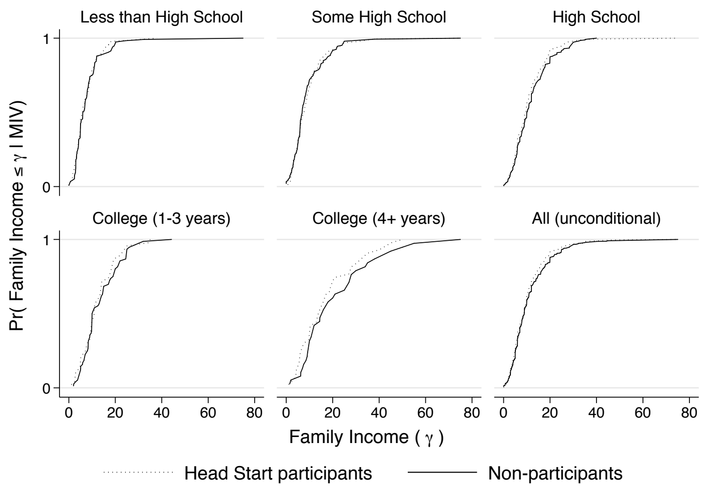
Note: Number of observations for women equal 439 (less than high school), 334 (some high school), 791 (high school), 227 (college 1-3 years), 219 (college 4+ years) and 2010 (all). Number of observations for men equal 422 (less than high school), 280 (some high school), 828 (high school), 246 (college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

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

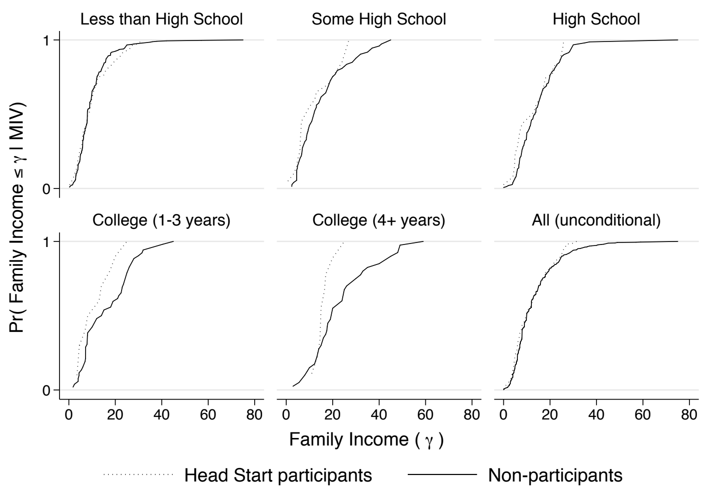
White



Black

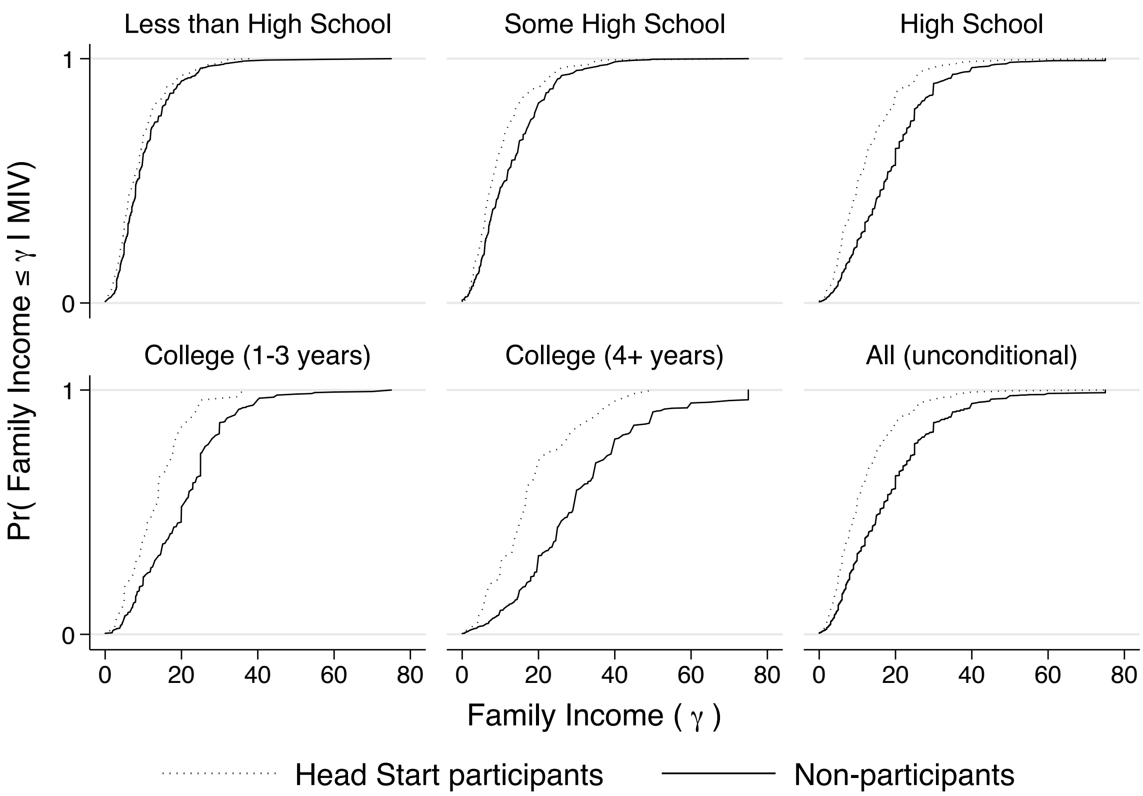


Hispanic



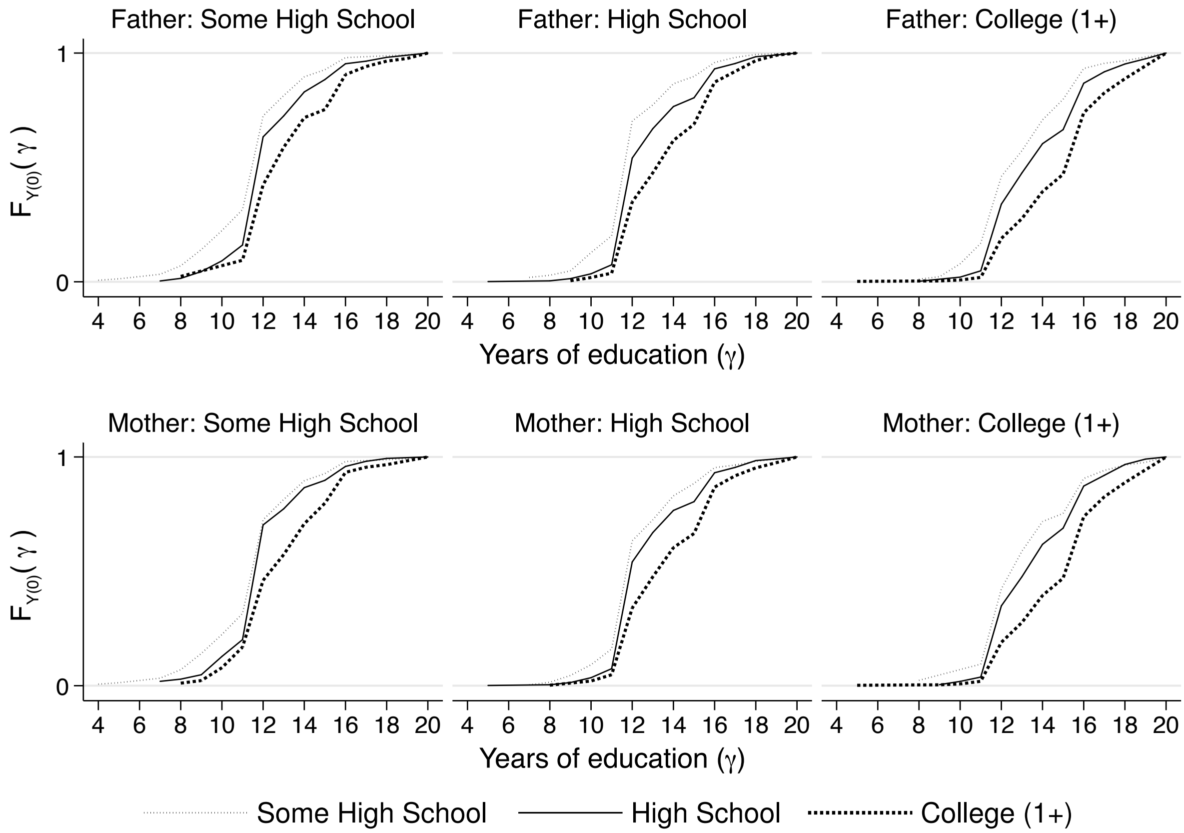
Note: Number of observations for white equal 201 (less than high school), 208 (some high school), 928 (high school), 289 (college 1-3 years), 331 (college 4+ years) and 1957 (all). Number of observations for black equal 252 (less than high school), 313 (some high school), 500 (high school), 122 (college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



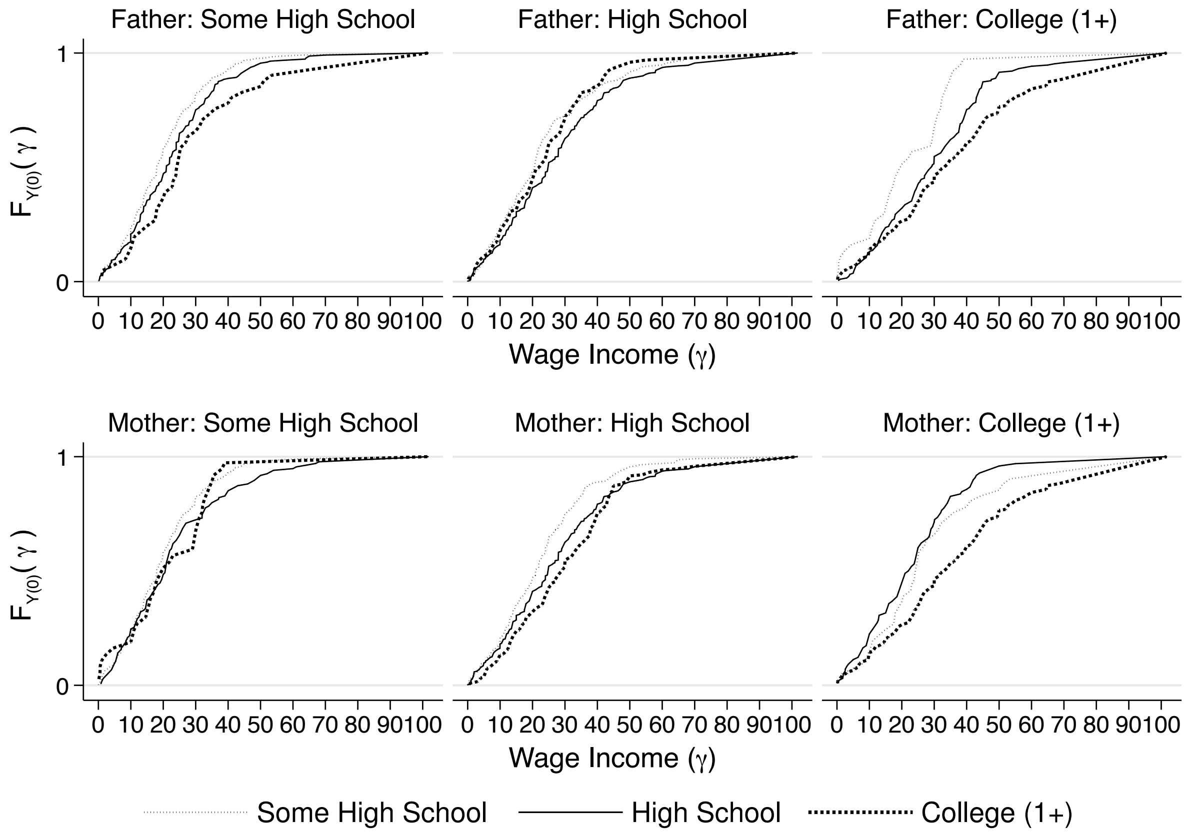
Note: Number of observations equal 947 (less than high school), 663 (some high school), 1825 (high school), 554 (college 1-3 years), 669 (college 4+ years) and 4658 (all).

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).