Appendix to:

Selection into Identification in Fixed Effects Models, with Application to Head Start

Contents

A Derivations, Proofs, and Extensions

A.1 Proof of Proposition 1

The proof of Proposition 1 closely follows the proofs of Theorem 2 in Angrist and Fernandez-Val (2013) and Theorem 1 in Aronow and Carnegie (2013). There are two key differences. First, we rely on Group ID Conditional Independence (Assumption 1), instead of the IV exclusion restriction. Second, we condition on two propensity scores, unlike Aronow and Carnegie (2013), who condition on $Pr(D_i = 1)$, and Angrist and Fernandez-Val (2013), who condition on discrete covariates.

Recall that we define $\hat{\delta}^t := \frac{1}{\sum_i \mathbf{1}(S_{g(i)}=1)} \sum_{i|S_{g(i)}=1} \widehat{w^t_{g(i)}} \cdot \hat{\delta}_{g,FE}$ and $w^t_{g(i)} := \frac{Q_x \cdot Pr[S=1]}{P_x \cdot Pr[T=1]}$. $w^t_{g(i)}$is well-defined, due to Assumption 4.

By Assumptions 1 and 3,

$$\hat{\delta}^t \rightarrow_p \mathbb{E}\left[w^t_{g(i)} \cdot \hat{\delta}_g | S_{g(i)} = 1\right] \tag{16}$$

From here, we simplify the g(i) notation to g.

By Assumption 1 and the law of iterated expectations,

$$\mathbb{E}\left[w^t_g \cdot \hat{\delta} | S_g = 1\right] = \mathbb{E}\left[w^t_g \left(Y_i\left(1\right) - Y_i\left(0\right)\right) | S_g = 1\right] \tag{17}$$

By the law of iterated expectations,

$$\begin{aligned}
\mathbb{E}[w^t_g \cdot (Y_i(1) - Y_i(0))|S = 1] &= \mathbb{E}[\mathbb{E}[w^t_g(Y(1) - Y(0))|S = 1, P_x, Q_x]|S_g = 1] && (18)\\
&= \mathbb{E}[\mathbb{E}[w^t_g(Y(1) - Y(0))|P_x, Q_x]|S_g = 1] && (19)\\
&= \mathbb{E}[w^t_g \mathbb{E}[(Y(1) - Y(0))|P_x, Q_x])|S_g = 1] && (20)\\
&= \mathbb{E}[w^t_g \Delta(P_x, Q_x)|S_g = 1] && (21)
\end{aligned}$$

Line (19) follows from (18) from the CFEI assumption that $\mathbb{E}[Y(1) - Y(0)|S, P_x, Q_x] = \mathbb{E}[Y(1) - Y(0)|P_x, Q_x]$. Line (20) follows from line (19) because w^t_g is a function of P_x and Q_x only, by definition. We define $\Delta(P_x, Q_x)$ to be the treatment effect conditional on P_x, Q_x: $\Delta(P_x, Q_x) \equiv \mathbb{E}[(Y(1) - Y(0))|P_x, Q_x]$ in line 20, so that line 21 follows from line 20 by definition.

Now let F be the distribution of (P_x, Q_x), and let $F(\cdot|S = 1)$ be the distribution conditional on $S = 1$. By Bayes rule,

$$
\begin{aligned}
\int w_g^t \Delta(P_x,Q_x)dF(P_x,Q_x|S=1) &= \int w_g^t \cdot \Delta(P_x,Q_x)\frac{Pr(S=1|P_x,Q_x)}{Pr(S=1)}dF(P_x,Q_x)\\
&= \int w_g^t \cdot \Delta(P_x,Q_x)\frac{P_x}{Pr(S=1)}dF(P_x,Q_x)\\
&= \int \frac{Q_x}{P_x}\frac{P(S=1)}{P(T=1)}\Delta(P_x,Q_x)\frac{P_x}{Pr(S=1)}dF(P_x,Q_x)\\
&= \int \Delta(P_x,Q_x)\frac{Q_x}{Pr(T=1)}dF(P_x,Q_x)\\
&= \int \Delta(P_x,Q_x)\frac{Pr(T=1|P_x,Q_x)}{Pr(T=1)}dF(P_x,Q_x)\\
&= \int \Delta(P_x,Q_x)dF(P_x,Q_x|T=1)
\end{aligned}
$$

By CFEI, $\int \Delta(P_x,Q_x)dF(P_x,Q_x|T=1) = \int \mathbb{E}[Y(1)-Y(0)|T=1,P_x,Q_x]dF(P_x,Q_x|T=1) = \mathbb{E}[Y(1)-Y(0)|T=1]$.

A.1.1 Extrapolating treatment effects to never-switchers

Note that the ATE for the target population can be written as a weighted average of the ATE for switchers, $\delta^{t,P_x>0}$, and the ATE for never-switchers, $\delta^{t,P_x=0}$:

$$\delta^t := Pr(P_x>0|T=1)\cdot\delta^{t,P_x>0} + (1-Pr(P_x>0|T=1))\cdot\delta^{t,P_x=0} \tag{22}$$

where $\delta^{t,P_x>0} = \mathbb{E}[Y(1)-Y(0)|T=1,P_x>0]$ and $\delta^{t,P_x=0} = \mathbb{E}[Y(1)-Y(0)|T=1,P_x=0]$.

Since we can not identify treatment effects for $P_x=0$ from the FFE design, we must impose an additional assumption that allows extrapolation of treatment effects for this group.

Assumption 5 (Treatment Effect Functional Form):

$$\mathbb{E}[Y(1)-Y(0)|\mathbf{X_g}] = H(\mathbf{\Phi};\mathbf{X_g})$$

with $H(\cdot)$ known, and $\mathbf{\Phi}$ parameters that can be consistently estimated. Under Assumption 5, $\hat{\mathbf{\Phi}}$ can be estimated using e.g. the regression $\hat{\delta}_g = \mathbf{\Phi'X_g}+\mathrm{u_g}$.

Proposition 2. *Under Assumptions 1, 2, 3, and 5, and assuming $Pr(P_x>0|T=1)$ can be consistently estimated, the ATE for the target t can be consistently estimated by*

$$\hat{\delta}^t = Pr(P_x>0|T=1)\cdot\widehat{\delta^{t,P_x>0}} + (1-Pr(P_x>0|T=1))\cdot\widehat{\delta^{t,P_x=0}}$$

where $\delta^{t,P_x>0}$comes from Equation 12 and $\delta^{t,P_x=0}$ is estimated from a projection of $\hat{\delta}_g$on X_g.

Proof:

Define $\widehat{\delta^{t,P_x=0}} := \frac{1}{\sum \mathbf{1}(T=1,P_x=0)}\sum_{T=1,P_x=0} H\left(\hat{\mathbf{\Phi}};\mathbf{X_g}\right)$.
Assumptions 1, 3, and 5 imply that

$$plim\ \widehat{\delta^{t,P_x=0}} = \mathbb{E}[Y(1)-Y(0)|T=1,P_x=0] \tag{23}$$

From the proof of Proposition 1, we have that

$$plim\ \widehat{\delta^{t,P_x>0}} = \mathbb{E}[Y(1) - Y(0)|T = 1, P_x > 0] \tag{24}$$

Then,

$$\begin{aligned} plim\ \widehat{\delta^t} &= Pr\left(P_x > 0|T = 1\right) \cdot \mathbb{E}[Y(1) - Y(0)|T = 1, P_x > 0] \quad (25)\\ &+ \left(1 - Pr\left(P_x > 0|T = 1\right)\right) \cdot \mathbb{E}[Y(1) - Y(0)|T = 1, P_x = 0] \\ &= \mathbb{E}[Y(1) - Y(0)|T = 1] \quad (26) \end{aligned}$$

A speculative alternative approach would be to take a "double robust" estimation approach. This is modeled after the double robust approach for estimating causal effects, as in Robins, Rotnitzky and Zhao (1995). The discussion in Chapter 17 of Imbens and Rubin (2015, pp. 399-400) notes that in the traditional setting, consistency of the estimated treatment effect requires either correct specification of the propensity score, or of the regression model. In our setting, instead of trying to model potential outcomes, we are estimating average treatment effects. Implementation would proceed by (i) defining weights $w^t_{g(i)}$ as in equation (13); (ii) estimating $\hat{\delta}_g = H\left(\hat{\mathbf{\Phi}}; \mathbf{X_g}\right)$ using these weights, and then predicting $\frac{1}{\sum \mathbf{1}(T=1)} \sum_{T_g=1} H\left(\hat{\mathbf{\Phi}}; \mathbf{X_g}\right)$ over the full target population.

A.2 Intuition for Propensity Score Weighting

In this section, we provide a simple derivation of the weighting scheme that we propose to obtain the ATE from the switchers sample by introducing a concrete example in which the treatment effect is determined by one discrete covariate, x, and in which there are only few groups in the switcher sample. For ease of exposition, we refer to groups as families and units within groups as kids.

A.2.1 Thought Experiment

Suppose that the target population is comprised of 75% black individuals and 25% white individuals. The switchers sample has 1 white family with 3 kids and 2 black families with 3 and 5 kids, respectively. Thus, to be representative of the target population, the white family should be given a weight of 25%. The share for each black family is proportional to the number of individuals in the family, normalized so that the total share across the two families is 75%. Thus, the first family should be given a weight of $0.75 \cdot \frac{3}{8}$, and the second family should be given a weight of $0.75 \cdot \frac{5}{8}$.

A.2.2 Notation

Under the setup above, the weight that should be given to a switcher family g where all individuals have race $x_i = x$, can be written as:

$$s_{gx} = \frac{\sum(\mathbf{1}(x_i = x)|T_g = 1)}{\sum T_g} \cdot \frac{\sum \mathbf{1}(g(i) = g)}{\sum(S_g|x_i = x)} \tag{27}$$

The first term, $\frac{\sum(\mathbf{1}(x_i=x)|T_g=1)}{\sum T_g}$, gives the share of individuals in the target population with race x. The second term, $\frac{\sum \mathbf{1}(g(i)=g)}{\sum(S_g|x_i=x)}$ gives the size of family g as a proportion of the switcher sample with race x.

Equivalently,

$$s_{gx} = Pr(x_i = x|T_g = 1) \cdot \frac{\sum \mathbf{1}(g(i) = g)}{Pr(x_i = x|S_g = 1) \cdot \sum S_g} \tag{28}$$

$$= \frac{Pr(x_i = x|T_g = 1)}{Pr(x_i = x|S_g = 1)} \cdot Pr(g(i) = g|S_g = 1) \tag{29}$$

A.2.3 Estimation

1. We obtain an estimate of $\hat{Q}_x = Pr(T_g = 1|x_i = x)$ as fitted values from a regression of T on indicator variables for each value of X.

   This is equal to $\frac{Pr(x_i=x|T=1)\cdot Pr(T=1)}{Pr(x_i=x)}$ by Bayes rule.

2. We obtain an estimate of $\hat{P}_x = Pr(S_g = 1|x_i = x)$ as fitted values from a regression of S on indicator variables for each value of X.

   This is equal to $\frac{Pr(x_i=x|S_g=1)\cdot Pr(S=1)}{Pr(x_i=x)}$ by Bayes rule. The ratio of (1) and (2) is $\frac{Pr(x_i=x|T_g=1)}{Pr(x_i=x|S_g=1)} \cdot \frac{Pr(T=1)}{Pr(S=1)}$.

3. To get s_{gx}, we need to multiply this ratio by $Pr(g(i) = g|S_g = 1)$ and divide by $\frac{Pr(T=1)}{Pr(S=1)}$. We then normalize the weights, which gives $s_{gx} = \frac{\frac{Q_x}{P_x}\cdot n_g}{\sum_{g\in G_S}\frac{Q_x}{P_x}\cdot n_g}$

A.3 Extension to Unit i Covariates

In this section, we discuss how to quantify the contribution of residual switchers to estimates and other considerations related to residual switchers, and how our key assumptions and proposition can be extended to accomodate C_i.

Defining residual switchers Consider the "deviations from group means" projection matrix, $M = I - H\left(H'H\right)^{-1} H$, with H a matrix of dummy variables for group membership: $H[i,j] = 1$ if unit i is a member of group j, and 0 otherwise. Let $\widetilde{D} = M \cdot D$ be deviations in treatment from group means. Basic switcher groups ("true switchers") are defined by having within-group variation in treatment: $V_g := Var\left(\widetilde{D}_i|g(i) = g\right) > 0$. Next consider the residual-maker matrix projecting on covariates C after taking deviations from group means, $L = I - (M\cdot C)\left(C'\cdot M\cdot C\right)^{-1}\left(C'\cdot M\right)$, and let $\ddot{D} = L\cdot M\cdot D$. With unit-varying covariates C_i, the variation that identifies the treatment effects is $V_{g,C} := Var\left(\ddot{D}|g(i) = g\right)$. For some groups g, it could be the case that $V_g = 0$, and also $V_{g,C} > 0$. We call these groups residual switchers.

How much do residual switchers matter for estimates? We can quantify this by calculating the share of variation in D_i coming from residual switchers, using a formula similar to the calculation of the effective number of observations.[63] In our PSID application, residual switchers provide 3% of

[63] In particular, the share of identification from residual switchers is equal to $1 - \frac{\sum_g Var(D_i|C_i,g(i)=g)\cdot\left(n_g-1\right)\cdot\mathbf{1}\left(S_g=1\right)}{\sum_g Var(D_i|C_i,g(i)=g)\cdot\left(n_g-1\right)}$. Alternatively, calculating the effective number of observations using Equation 3 (altering the variance to condition on C_i) would produce similar results

the variation used for identification of the Head Start FFE coefficient. Therefore, this contributes minimally to the FE estimate.

Unit i covariates in the one- and two-step reweighting approaches Introducing i-level covariates may also cause the two-step reweighting approach and the one-step in-regression reweighting approach to produce different results, for two reasons. The first reason is that the slope parameters on the C_i will be estimated with different weights for the two approaches. The second reason is that the two-step approach uses $G + G_S$ degrees of freedom (total number of groups + number of switching groups) to estimate the group fixed effects and group-specific treatment effects, while the one-step approach uses just G+1 degrees of freedom to estimate the group fixed effect and the (single) treatment effect. In both cases, the coefficients on the C_i are estimated from the remaining degrees of freedom; which could come from larger family sizes, or families that are non-switchers in D_i but which do have variation in C_i. This implies that in some cases it may be possible to include C_i in the one-step approach, but not in the two-step approach (e.g., if all groups only have two units).

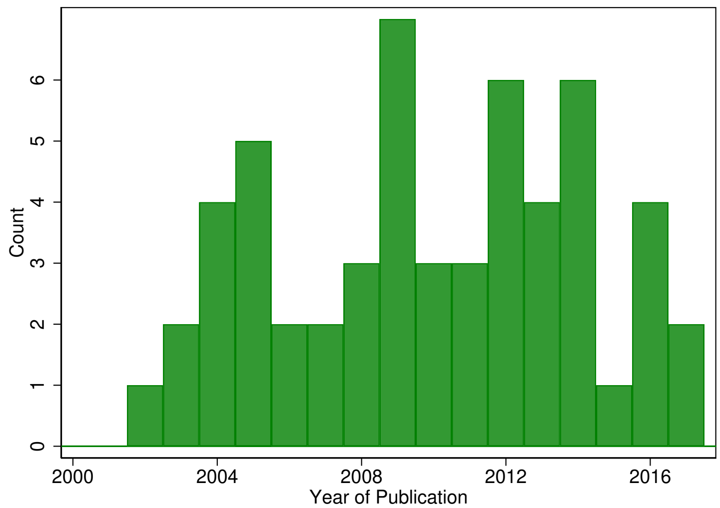
Finally, we discuss the modifications needed to the proof in Section A.1 to allow for residual switchers. We begin with a simplification of the model, in which outcomes are a linear function of treatment, individual covariates, and additively separable individual error terms: $Y_{ig} = \delta_g \cdot D_i + \beta \cdot C_i + \alpha_g + (u_{ig} \cdot D_i + \epsilon_{ig})$. We assume constant coefficients on C_i, and require the systematic part of the treatment effect to be constant within a group. There can also be an idiosyncratic component of the treatment effect, denoted by u_{ig}. We also now allow for individual covariates to enter into the propensity to be in the switching or target populations, respectively, as: $P_{X,C} = Pr\left[S_{g(i)} = 1 | X_g, C_i\right]$ and $Q_{X,C} = Pr\left[T_i = 1 | X_g, C_i\right]$. The IPW weights therefore vary at the individual level: $w_i^t = \frac{Q_{X,C}}{P_{X,C}} \cdot \frac{Pr(S=1)}{Pr(T=1)}$.

The assumptions from earlier now must now be modified slightly to recover the ATE. Assumption 1, the conditional fixed effects assumption now requires $\epsilon_{ig}, u_{ig} \perp D_i | C_i, \alpha_g$. This gives that $E\left[\hat{\delta}_g\right] = \delta_g$.Assumptions 2--4 and Proposition 1 will carry forward with these redefined terms.

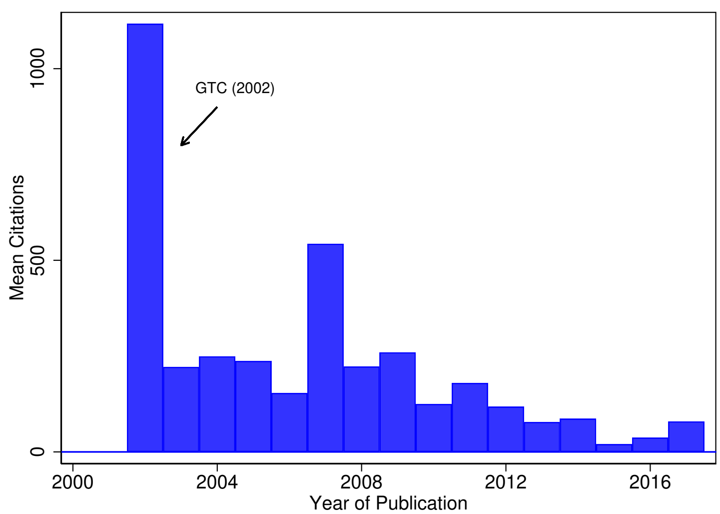
B Supplementary Figures and Tables

Figure B.1: Popularity of Family Fixed Effects Articles

(a) Publications by Year



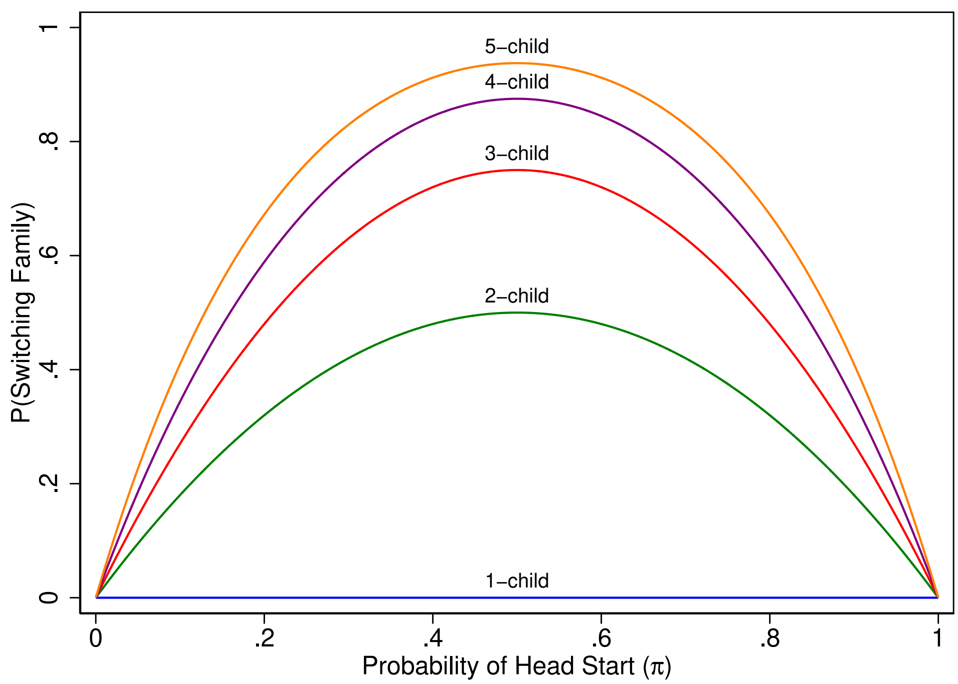
(b) Average Citations by Year of Publication



Notes: These figures display the data from our survey of FFE papers published from January 2000 to May 2017 in 11 leading journals that publish applied microeconomics articles. Figure (a) plots the number of FFE articles published in each year, and Figure (b) plots the average number of Google Scholar citations, as of May 2019, among the articles published in a given year.

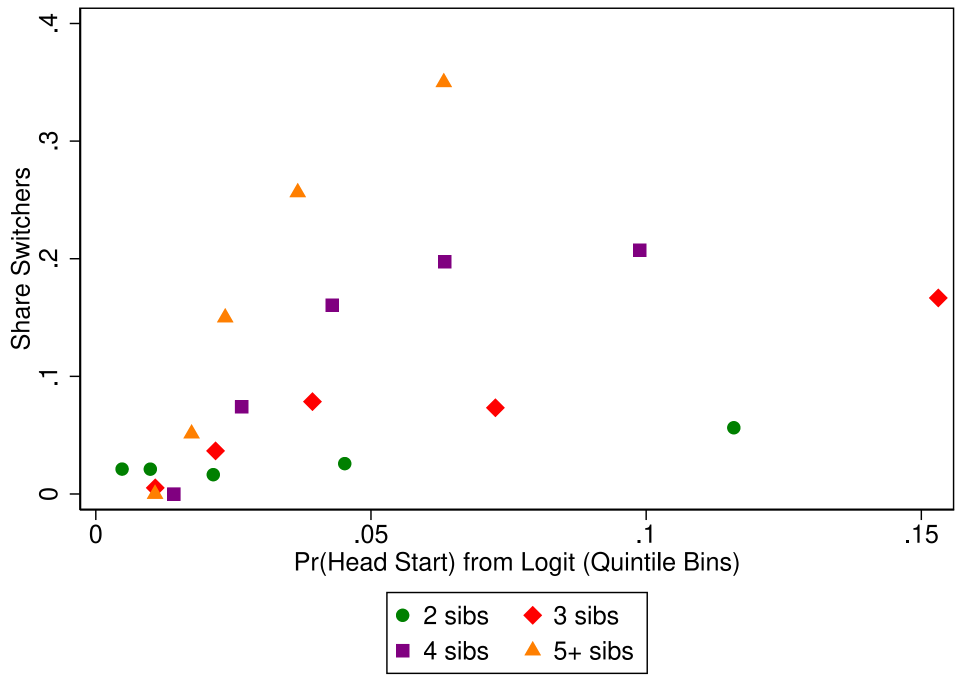
Figure B.2: Illustrative Model of the Role of Family Size in Switching

$$Pr(S_g = 1) = 1 - (1 - \pi)^{n_g} - \pi^{n_g}$$



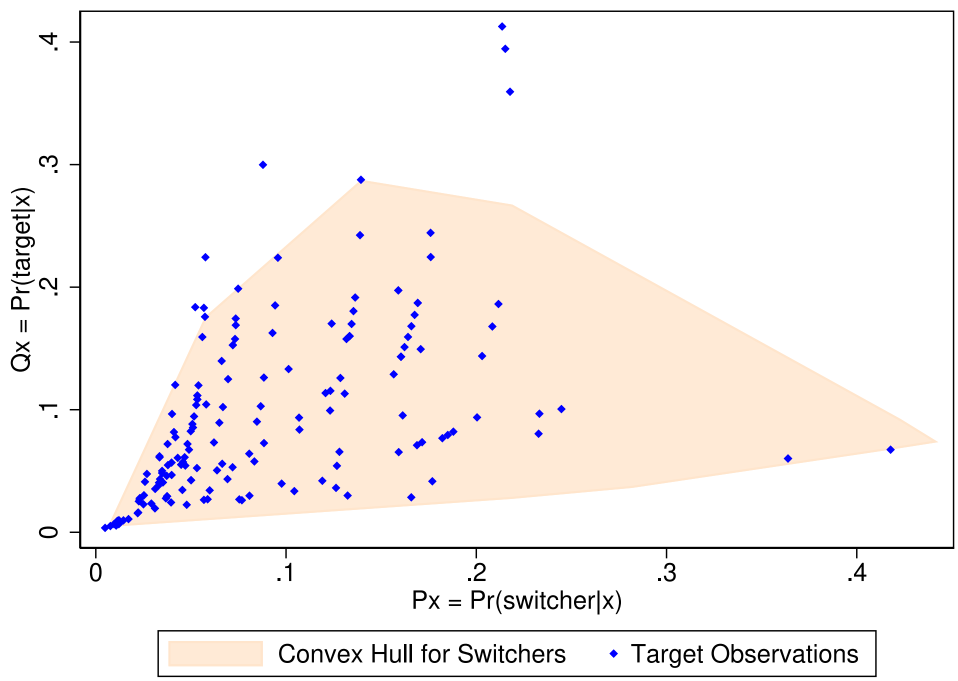
Notes: This figure plots the theoretical function: $P(HSSwitchingFamily) = 1 - (1 - \pi)^{n_g} - \pi^{n_g}$, where n_g is the number of children in a family and π is the probability of attending Head Start, for 2-, 3-, 4-, and 5 (plus)- child families.

Figure B.3: Predicted Pr(Head Start) and Switching by Family Size



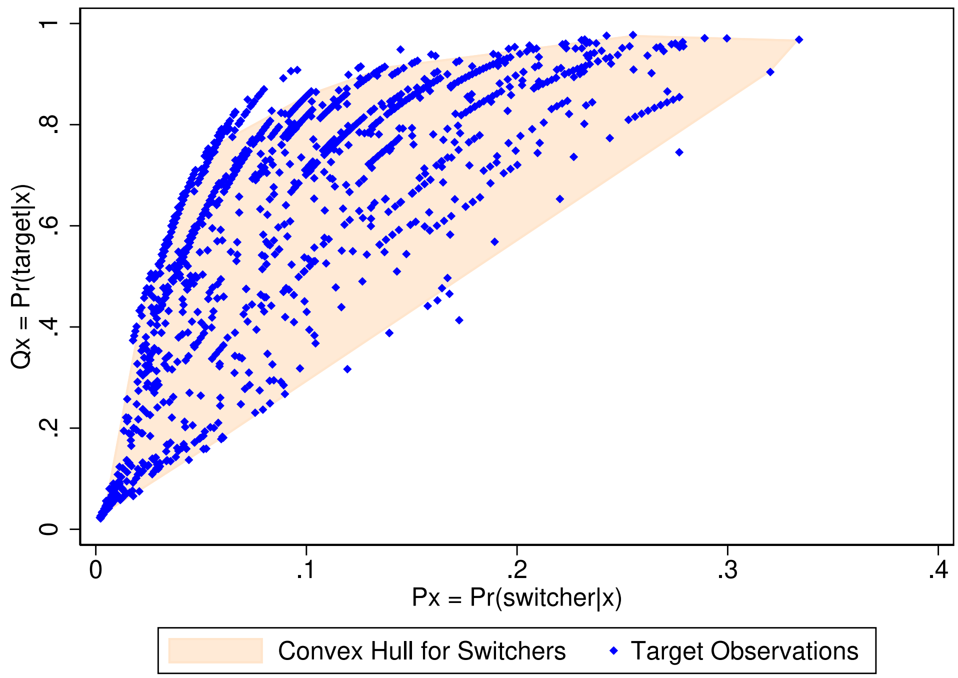
Notes: This figure shows the share of switching families by quintiles of the predicted probability of Head Start, which we estimate from a logit model using the covariates listed in Section 6.2. We define the quintiles and plot the share of switchings separately by family size (2, 3, 4, or 5+ children), where the color and shape of the markers indicate the number of children in the family.

Figure B.4: Examining Overlap in P_x and Q_x Between Switchers and Head Start Participants



Notes: This figure shows the convex hull for switchers of P_x and Q_x (Stata command cvxhull) in light orange, and all of the combinations of P_x and Q_x for individuals who are Head Start participants (the target). In total, 95% of Head Start participants lie within the convex hull.

Figure B.5: Examining Overlap in P_x and Q_x Between Switchers and Head-Start-Eligible



Notes: This figure shows the convex hull for switchers of P_x and Q_x (Stata command cvxhull) in light orange, and all of the combinations of P_x and Q_x for individuals who are Head Start eligible (the target). In total, 90% of Head Start eligibles lie within the convex hull.

Table B.1: Head Start Attendance and Within-Family Variation in Attendance by Family Size (PSID)

	Number of Children in Family:				
	2	3	4	5+	Total
Share of Family in Head Start (π)	0.157	0.222	0.195	0.206	0.182
Share with Switching	0.121	0.202	0.242	0.471	0.174
All Participants in HS in Family	0.096	0.125	0.093	0.049	0.102
No Participants in HS in Family	0.783	0.672	0.665	0.480	0.724

Notes: This table shows the sources of switching by family size. The first two rows show the likelihood of attending Head Start by family size and the likelihood of having variation in Head Start within a family (switching). The final two rows examines whether differences in rates of switching across family sizes are attributable to variation across family sizes in having all children attend Head Start (row 3) or variation in having no children attend Head Start (row 4).

Table B.2: Switchers are Different than Head-Start-Participating Non-Switchers

	(1) Switch	(2) Non-Switch, HS=1
Fraction female	0.562	0.530
Fraction African-American	0.516	0.714***
Mother's yrs education	9.283	10.304***
Father's yrs education	9.190	10.445***
Had a single mother at age 4	0.252	0.332***
Family income (age 3-6) (CPI adjusted)	31809	25915***
Mother employed, age 0	0.508	0.542
Mother employed, age 1	0.517	0.541
Mother employed, age 2	0.536	0.596**
Household size at age 4	5.487	4.353***
Fraction low birth weight	0.077	0.135***

Notes: This table presents summary statistics for Head-Start-switching families (column 1) and non-switching families that attended Head Start (column 2) together with stars indicating the p-value for the test of equality of the clums. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.3: Demographic Characteristics of Head Start Sample (PSID)

	All	Head Start	No Head Start	Sibling Sample
Head Start	0.076	1.000	0.000	0.073
	(0.27)	(0.00)	(0.00)	(0.26)
Other preschool	0.282	0.000	0.305	0.259
	(0.45)	(0.00)	(0.46)	(0.44)
Fraction African-American	0.150	0.618	0.111	0.154
	(0.36)	(0.49)	(0.31)	(0.36)
Fraction female	0.504	0.548	0.501	0.501
	(0.50)	(0.50)	(0.50)	(0.50)
Fraction low birth weight	0.060	0.114	0.056	0.061
	(0.24)	(0.32)	(0.23)	(0.24)
Had a single mother at age 4	0.112	0.296	0.091	0.103
	(0.32)	(0.46)	(0.29)	(0.30)
Fraction whose mother completed hs	0.717	0.632	0.724	0.689
	(0.45)	(0.48)	(0.45)	(0.46)
Fraction whose father completed hs	0.683	0.557	0.692	0.654
	(0.47)	(0.50)	(0.46)	(0.48)
Fraction eldest child in family	0.368	0.340	0.371	0.339
	(0.48)	(0.47)	(0.48)	(0.47)
Age in 1995	23.830	18.605	24.262	25.063
	(9.84)	(7.76)	(9.87)	(10.06)
Mother's yrs education	11.116	10.208	11.190	10.942
	(2.76)	(2.32)	(2.78)	(2.81)
Father's yrs education	11.238	10.159	11.314	11.076
	(3.23)	(2.70)	(3.25)	(3.35)
Family income (age 3-6) (CPI adjusted)	50339.121	28552.548	52718.519	50972.698
	(35814.01)	(17212.32)	(36509.36)	(37315.99)
Household size at age 4	4.535	4.814	4.504	4.778
	(1.68)	(2.06)	(1.63)	(1.64)
Observations	7372	1354	6018	5361

Notes: This table shows the mean demographic characteristics of the sample, weighted to be representative of 1995 population; see text for details. Standard deviations, shown in parentheses, are omitted for binary variables. CPI-adjusted income reported in 1999 dollars. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.4: Outcomes of Interest for Head Start Sample (PSID)

	All	Head Start	No Head Start	Sibling Sample
Fraction completed hs	0.913	0.878	0.916	0.912
Fraction attended some college	0.531	0.428	0.539	0.532
Fraction not booked/charged with crime	0.899	0.889	0.900	0.898
Avg. Earnings age 23-25 (CPI adjusted)	20410.159	14391.591	20817.636	20633.494
	(24927.76)	(12000.13)	(25517.27)	(26547.15)
Economic Sufficiency Index at 30	0.094	-0.601	0.151	0.096
	(1.03)	(1.05)	(1.01)	(1.03)
Economic Sufficiency Index at 40	0.020	-0.532	0.053	0.025
	(1.01)	(0.95)	(1.01)	(1.04)
Good Health Index at 30	0.004	-0.558	0.050	0.017
	(1.03)	(1.26)	(0.99)	(0.99)
Good Health Index at 40	0.011	-0.486	0.033	0.015
	(1.01)	(1.25)	(1.00)	(0.96)
Observations	7372	1354	6018	5361

Notes: This table shows the means for the main outcomes of interest, weighted to be representative of 1995 population; see text for details. Note that the fraction not booked/charged with a crime restricted to individuals that responded to the PSID in 1995 who were between the ages of 16 and 50 in that year. CPI-adjusted income reported in 1999 dollars. Standard deviations, shown in parentheses, are omitted for binary variables. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.5: Summary Statistics for Inputs to Summary Indices (PSID)

	All	Head Start	No Head Start	Sibling Sample
Inputs to Economic Sufficiency Index, Age 30				
Ever on AFDC/TANF by age 30	0.062	0.221	0.049	0.060
Fraction of last 5 yrs on Food Stamps/SNAP, age 30	0.064	0.151	0.056	0.071
	(0.20)	(0.30)	(0.19)	(0.22)
ln(mean earnings in last 5 years), age 30	9.661	9.415	9.676	9.659
	(1.06)	(0.91)	(1.07)	(1.07)
Fraction of last 5 yrs with positive earnings, age 30	0.895	0.887	0.896	0.898
	(0.25)	(0.26)	(0.25)	(0.25)
Fraction of last 5 yrs ever unemployed, age 30	0.146	0.173	0.144	0.150
	(0.24)	(0.27)	(0.23)	(0.24)
Mean Inc. Rel. Pov. in last 5 years, age 30	385.831	233.796	396.729	385.933
	(305.98)	(155.44)	(311.18)	(291.36)
Fraction completed college	0.209	0.073	0.220	0.220
	(0.41)	(0.26)	(0.41)	(0.41)
Inputs to Economic Sufficiency Index, Age 40				
Ever on AFDC/TANF by age 40	0.068	0.163	0.062	0.067
Fraction of last 5 yrs on Food Stamps/SNAP, age 40	0.043	0.098	0.040	0.043
	(0.16)	(0.25)	(0.16)	(0.16)
ln(mean earnings in last 5 years), age 40	9.962	9.779	9.968	9.957
	(1.15)	(0.90)	(1.16)	(1.15)
Fraction of last 5 yrs with positive earnings, age 40	0.850	0.867	0.849	0.849
	(0.31)	(0.29)	(0.31)	(0.31)
Fraction of last 5 yrs ever unemployed, age 40	0.094	0.122	0.093	0.098
	(0.20)	(0.24)	(0.19)	(0.20)
Mean Inc. Rel. Pov. in last 5 years, age 40	436.769	281.489	443.338	434.280
	(366.03)	(183.89)	(370.36)	(361.58)
Fraction of last 5 yrs owned home, age 40	0.500	0.287	0.510	0.522
	(0.44)	(0.42)	(0.44)	(0.44)
Inputs to Good Health Index, Age 30				
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 30	0.745	0.668	0.753	0.755
	(0.41)	(0.45)	(0.41)	(0.40)
Fraction of last 5 yrs reported good or better health, age 30	0.948	0.903	0.951	0.950
	(0.17)	(0.24)	(0.17)	(0.17)
Mean BMI in last 5 years, age 30	26.569	28.766	26.333	26.615
	(6.68)	(6.74)	(6.63)	(6.85)
Inputs to Good Health Index, Age 40				
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 40	0.738	0.714	0.739	0.728
	(0.42)	(0.44)	(0.42)	(0.42)
Fraction of last 5 yrs reported good or better health, age 40	0.919	0.871	0.921	0.922
	(0.22)	(0.29)	(0.22)	(0.22)
Mean BMI in last 5 years, age 40	27.504	30.191	27.327	27.433
	(5.92)	(7.42)	(5.77)	(5.85)
Observations	7372	1354	6018	5361

Notes: Weighted to be representative of 1995 population; see text for details. SD, in parentheses, are omitted for binary variables. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.6: N's for Control Covariates (PSID)

	All	Head Start	No Head Start	Sibling Sample
Head Start	7372	1354	6018	5361
Other preschool	7372	1354	6018	5361
Fraction African-American	7372	1354	6018	5361
Fraction female	7372	1354	6018	5361
Fraction low birth weight	5366	970	4396	4555
Had a single mother at age 4	6678	1285	5393	4672
Fraction whose mother completed hs	7231	1332	5899	5360
Fraction whose father completed hs	6596	1034	5562	4875
Fraction eldest child in family	7372	1354	6018	5361
Age in 1995	7372	1354	6018	5361
Mother's yrs education	7223	1331	5892	5356
Father's yrs education	6596	1034	5562	4875
Family income (age 3-6) (CPI adjusted)	6086	1145	4941	4338
Household size at age 4	6251	1187	5064	4420
Observations	7372	1354	6018	5361

Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.7: N's for Main Outcomes (PSID)

	All	Head Start	No Head Start	Sibling Sample
Fraction completed hs	7372	1354	6018	5361
Fraction attended some college	7372	1354	6018	5361
Fraction not booked/charged with crime	5005	802	4203	3591
Avg. Earnings age 23-25 (CPI adjusted)	4866	783	4083	3675
Economic Sufficiency Index at 30	7372	1354	6018	5361
Economic Sufficiency Index at 40	4085	613	3472	2845
Good Health Index at 30	4749	791	3958	3600
Good Health Index at 40	2228	312	1916	1673
Observations	7372	1354	6018	5361

Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.8: N's for Auxiliary Outcomes (PSID)

	All	Head Start	No Head Start	Sibling Sample
Inputs to Economic Sufficiency Index, 30				
Ever on AFDC/TANF by age 30	7372	1354	6018	5361
Fraction of last 5 yrs on Food Stamps/SNAP, age 30	4186	713	3473	2805
ln(mean earnings in last 5 years), age 30	4202	620	3582	3159
Fraction of last 5 yrs with positive earnings, age 30	4378	656	3722	3295
Fraction of last 5 yrs ever unemployed, age 30	4259	634	3625	3184
Mean Inc. Rel. Pov. in last 5 years, age 30	5293	891	4402	4068
Fraction completed college	7372	1354	6018	5361
Inputs to Economic Sufficiency Index, 40				
Ever on AFDC/TANF by age 40	4085	613	3472	2845
Fraction of last 5 yrs on Food Stamps/SNAP, age 40	1972	250	1722	1423
ln(mean earnings in last 5 years), age 40	1695	221	1474	1266
Fraction of last 5 yrs with positive earnings, age 40	1829	236	1593	1369
Fraction of last 5 yrs ever unemployed, age 40	1825	236	1589	1365
Mean Inc. Rel. Pov. in last 5 years, age 40	2152	296	1856	1613
Fraction of last 5 yrs owned home, age 40	2292	290	2002	1625
Inputs to Good Health Index, 30				
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 30	2267	385	1882	1742
Fraction of last 5 yrs reported good or better health, age 30	3763	579	3184	2806
Mean BMI in last 5 years, age 30	3248	587	2661	2528
Inputs to Good Health Index, 40				
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 40	1280	182	1098	930
Fraction of last 5 yrs reported good or better health, age 40	1463	182	1281	1116
Mean BMI in last 5 years, age 40	2037	307	1730	1486
Observations	7372	1354	6018	5361

Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.9: Effect of Head Start on Pre-Head-Start Outcomes (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Low birth weight					
Head Start	0.040*	0.045*	-0.016	-0.018	-0.029
	(0.021)	(0.023)	(0.026)	(0.033)	(0.042)
Other preschool	0.003	0.003	-0.012	-0.056**	-0.003
	(0.012)	(0.013)	(0.023)	(0.027)	(0.027)
Observations	5366	4555	4500	1872	2622
Disabled					
Head Start	-0.006	-0.017	-0.010	-0.016	-0.006
	(0.027)	(0.030)	(0.030)	(0.036)	(0.051)
Other preschool	0.018	0.018	0.021	0.032	0.017
	(0.019)	(0.022)	(0.028)	(0.049)	(0.032)
Observations	3516	2955	2661	1102	1555
Single mom at age 4					
Head Start	0.020	0.025	0.027	-0.007	0.051
	(0.015)	(0.020)	(0.024)	(0.022)	(0.040)
Other preschool	0.022**	0.020*	0.008	0.006	0.011
	(0.009)	(0.011)	(0.017)	(0.031)	(0.018)
Observations	6678	4672	4467	1939	2522
Family income (age 1) (CPI-adjusted)					
Head Start	250.552	151.038	1312.474	2194.434**	1791.555
	(643.230)	(648.685)	(1103.298)	(971.524)	(2107.734)
Other preschool	763.418	815.576	1484.736	2584.517**	1495.669
	(540.656)	(559.763)	(1266.153)	(1214.359)	(1478.018)
Observations	6219	4313	4023	1719	2298
Family income (age 2) (CPI-adjusted)					
Head Start	-1160.169*	-902.259	-1394.166	-1551.308	-562.303
	(650.029)	(670.967)	(1166.143)	(1061.856)	(2071.297)
Other preschool	-365.946	589.414	862.558	3726.183***	403.253
	(737.837)	(617.550)	(1344.132)	(1425.988)	(1532.799)
Observations	6274	4391	4151	1757	2388
Mom working at age 1					
Head Start	0.001	0.011	0.049	0.002	0.080
	(0.018)	(0.022)	(0.039)	(0.033)	(0.073)
Other preschool	-0.001	-0.002	-0.017	-0.078*	-0.014
	(0.013)	(0.016)	(0.030)	(0.043)	(0.034)
Observations	6219	4313	4023	1719	2298
Mom working at age 2					
Head Start	0.025	0.028	-0.041	-0.008	-0.077
	(0.021)	(0.023)	(0.040)	(0.036)	(0.073)
Other preschool	0.026*	0.032*	0.015	-0.013	0.017
	(0.015)	(0.018)	(0.031)	(0.044)	(0.036)
Observations	6274	4391	4151	1757	2388

Notes: Weighted to be representative of 1995 population; see text for details. SE clustered at 1968 family id in columns 1 and 2 and at mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.10: Test of Conditional Ignorability Assumption:

Do Individuals in the Target Population Have Differential Treatment Effects?

	Target= Eligible	Target= Participants
Some College (Whites, PSID)		
In Target	–	-0.100
	–	(0.065)
Observations	–	213
Economic Sufficiency Index (Whites, PSID)		
In Target	–	-0.032
	–	(0.100)
Observations	–	213
High School Graduation (NLSY)		
In Target	-0.017	-0.037
	(0.060)	(0.033)
Observations	537	668
Idle (NLSY)		
In Target	0.047	0.012
	(0.075)	(0.041)
Observations	537	668
Learning Disability (NLSY)		
In Target	-0.050	-0.000
	(0.040)	(0.021)
Observations	537	668
Poor Health (NLSY)		
In Target	-0.006	0.004
	(0.055)	(0.030)
Observations	537	668

Notes: Each cell of this table shows an estimate from a regression of the family-specific impact of Head Start on an indicator for whether an individual is in the target population. Regressions are weighted for balance on observables: target individuals are assigned a weight of $Pr(T_g = 1, S_g = 1|X_{ig})$ and non-target individuals are assigned a weight of $Pr(T_g = 0, S_g = 1|X_{ig})$. The first two panels use data from the PSID white sample, and the final four panels use data from the CNLSY. There are no individuals in the PSID white switcher sample that are ineligible for Head Start, which causes us to have missing estimates ("–").

Table B.11: Test of Conditional Ignorability Assumption:

Do Gains depend on the Fraction of Children attending Head Start? (PSID, College)

Head Start x 2 child family	-0.126	0.103
	(0.099)	(0.191)
Head Start x 3 child family	0.395*	0.368**
	(0.221)	(0.174)
Head Start x 4 child family	0.377	0.441**
	(0.271)	(0.186)
Head Start x 5plus child family	0.522**	0.429***
	(0.260)	(0.139)
Head Start x 3 child family x Fraction	-0.515	
	(0.431)	
Head Start x 4 child family x Fraction	-0.304	
	(0.524)	
Head Start x 5plus child family x Fraction	-0.979	
	(1.010)	
Head Start x Fraction attend		-0.457
		(0.328)
Observations	2986	2986

Notes: This table shows estimates from regressions of some college on on an indicator for Head Start, an interaction of Head Start with indicators for family size, and an interaction of Head Start with the fraction of children in the family that attended (column 2) or with the fraction of children in the family that attended and indicators for family size (column 1). The data come from the PSID white sample. Standard errors are clustered on mother id.

Table B.12: Test of Conditional Ignorability Assumption:

Do Gains depend on the Fraction of Children attending Head Start? (CNLSY)

	HS	Idle	LD	Poor Hlth	HS	Idle	LD	Poor Hlth
Head Start x 2 child family	0.059	-0.075	-0.019	-0.098**	0.201*	-0.078	0.033	0.008
	(0.050)	(0.060)	(0.025)	(0.048)	(0.117)	(0.145)	(0.090)	(0.116)
Head Start x 3 child family	0.040	-0.050	-0.049	0.106	0.186	-0.004	-0.021	0.048
	(0.191)	(0.208)	(0.137)	(0.183)	(0.127)	(0.151)	(0.102)	(0.122)
Head Start x 4 child family	0.394**	-0.059	-0.215*	-0.147	0.280**	-0.066	0.010	0.111
	(0.182)	(0.268)	(0.112)	(0.164)	(0.124)	(0.169)	(0.094)	(0.122)
Head Start x 5plus child family	0.666***	-0.332	0.266**	0.116	0.409***	-0.319**	-0.124	0.037
	(0.171)	(0.312)	(0.114)	(0.236)	(0.127)	(0.154)	(0.083)	(0.129)
Head Start x 3 child family x Fraction	0.000	0.097	-0.049	-0.329				
	(0.340)	(0.387)	(0.243)	(0.345)				
Head Start x 4 child family x Fraction	-0.516	-0.010	0.344	0.301				
	(0.396)	(0.516)	(0.241)	(0.284)				
Head Start x 5plus child family x Fraction	-0.995***	0.043	-1.196***	-0.444				
	(0.369)	(0.839)	(0.411)	(0.564)				
Head Start x Fraction					-0.286	0.006	-0.103	-0.214
					(0.209)	(0.264)	(0.170)	(0.209)
Observations	1251	1251	1247	1251	1251	1251	1247	1251

Notes: This table shows estimates from regressions of outcomes (shown in the headers) on on an indicator for Head Start, an interaction of Head Start with indicators for family size, and an interaction of Head Start with the fraction of children in the family that attended (columns 5-8) or with the fraction of children in the family that attended and indicators for family size (columns 1-4). The data come from the CNLSY sample. Standard errors are clustered on mother id.

Table B.13: Additional Estimates for Representative White Populations (PSID)

Using Post-Regression Reweighting Method

	FFE		Reweighted, Target =		
	GTC	Expand Sample	HS Eligible	Participants	Siblings
A. High School Graduation					
Head Start	0.203**	-0.015	-0.060	-0.043	-0.033
	(0.098)	(0.045)	(0.071)	(0.071)	(0.051)
Y Mean in Target	–	0.921	0.852	0.848	0.921
B. Good Health Index, Age 30					
Head Start	–	-0.265	-0.165	-0.341	-0.130
	–	(0.249)	(0.184)	(0.211)	(0.186)
Y Mean in Target	–	0.074	-0.061	-0.583	0.074

Notes: Columns 1 and 2 of this table show the FFE estimated impacts of Head Start from GTC (2002) and using our expanded sample for completion of high school (panel A) and the Good Health Index at age 30 (panel B). The remaining columns present reweighted estimates of the effect of Head Start for three target populations (shown in the column header) using the post-regression reweighting procedure described in the text. ”–” is used to indicate that the information is not available. Sample size is N=2,986 for the expanded sample in panel A, and 1,959 for the expanded sample in panel B, and 1,036 for GTC. Standard errors obtained by bootstrapping. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.14: Head Start Impact for Representative Black Eligible, Participants, and Siblings (PSID) Using Post-Regression Reweighting Method

	FFE		Reweighted, Target =		
	GTC	Expand Sample	HS Eligible	Participants	Siblings
A. High School Graduation					
Head Start	-0.025	-0.024	-0.020	-0.017	-0.017
	(0.065)	(0.031)	(0.033)	(0.031)	(0.033)
Y Mean in Target	–	0.862	0.854	0.896	0.862
B. Some College					
Head Start	0.023	-0.016	-0.022	-0.025	-0.020
	(0.066)	(0.036)	(0.038)	(0.039)	(0.037)
Y Mean in Target	–	0.396	0.376	0.423	0.396
C. Economic Sufficiency Index, Age 30					
Head Start	–	-0.023	-0.168**	-0.201***	-0.145*
	–	(0.102)	(0.076)	(0.076)	(0.076)
Y Mean in Target	–	-0.552	-0.626	-0.674	-0.552
D. Good Health Index, Age 30					
Head Start	–	0.024	-0.020	0.015	-0.030
	–	(0.249)	(0.129)	(0.127)	(0.125)
Y Mean in Target	–	-0.357	-0.381	-0.539	-0.357

Notes: Columns 1 and 2 of this table show the FFE estimated impacts of Head Start from GTC (2002) and using our expanded sample for completion of high school (panel A) and the Good Health Index at age 30 (panel B). The remaining columns present reweighted estimates of the effect of Head Start for three target populations (shown in the column header) using the post-regression reweighting procedure described in the text. ”–” is used to indicate that the information is not available. Sample size is N=2,369 for the expanded sample in panels A, B, and C, and 1,150 for the expanded sample in Panel D, and 762 for GTC. Standard errors obtained by bootstrapping. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.15: Head Start Impact for Representative Eligible Children, Participants, and Siblings Using Reweighting with Regression Extrapolation to Singletons

	Reweighted ATE, Target =	
	HS Eligible	Participants
A. Some College (PSID)		
Head Start	0.081	0.057
	(0.054)	(0.049)
Y Mean in Target	0.388	0.419
B. Economic Sufficiency Index, Age 30 (PSID)		
Head Start	-0.075	-0.054
	(0.094)	(0.092)
Y Mean in Target	-0.192	-0.455
C. High School Graduation (CNLSY)		
Head Start	0.041	0.047
	(0.031)	(0.031)
Y Mean in Target	0.734	0.766
D. Idle (CNLSY)		
Head Start	-0.059	-0.057
	(0.037)	(0.039)
Y Mean in Target	0.221	0.201
E. Learning Disability (CNLSY)		
Head Start	-0.034*	-0.041**
	(0.018)	(0.018)
Y Mean in Target	0.055	0.041
F. Poor Health (CNLSY)		
Head Start	-0.059**	-0.065**
	(0.027)	(0.027)
Y Mean in Target	0.098	0.074

Notes: Columns 1 and 2 present reweighted estimates of the effect of Head Start for the Head-Start-eligible and Head-Start-participant target populations, where the treatment effects for singletons are extrapolated from switchers using OLS. Sample size is N=2,986 for the expanded sample, and 1,036 for GTC. Standard errors obtained using bootstrap. * p < .10, ** p < .05, *** p < .01.

Table B.16: FFE Estimates Reweighted using Gibbons, Suarez, Urbancic (2018) Method

	FFE	GSU (2018) Reweight
	Baseline	Switchers
A. Some College (PSID)		
Head Start	0.120**	0.134***
	(0.053)	(0.053)
B. Economic Sufficiency Index, Age 30 (PSID)		
Head Start	-0.023	-0.081
	(0.102)	(0.094)
C. High School Graduation (CNLSY)		
Head Start	0.085***	0.084***
	(0.030)	(0.027)
D. Idle (CNLSY)		
Head Start	-0.072*	-0.068**
	(0.037)	(0.034
E. Learning Disability (CNLSY)		
Head Start	-0.059***	-0.053***
	(0.020)	(0.019)
F. Poor Health (CNLSY)		
Head Start	-0.069***	-0.059**
	(0.026)	(0.025)

Notes: Column 1 reprints the FFE estimate using our expanded sample for PSID outcomes and using our replication sample for CNLSY outcomes. Column 2 presents the estimate weighting family-level estimates by the sample share, as suggested in Gibbons, Urbancic, Suarez Serrato (2018). This "undoes" the conditional variance weighting of FFE, and produces an estimate that is interpretable as the ATE for switchers. Sample size is N=2,986 for the expanded sample. Standard errors are clustered on mother id. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B.17: Horse Race between Family Size and Index of Non-Family-Size Covariates (PSID White Sample)

	x Fam Size	x Index	Horse Race
Index = Predicted Head Start			
Head Start	0.025	0.073	0.008
	(0.063)	(0.069)	(0.072)
Head Start x 4plus child family	0.281**		0.250**
	(0.112)		(0.105)
Head Start x Tercile 1 Predicted Head Start		-0.049	-0.116
		(0.094)	(0.101)
Head Start x Tercile 2 Predicted Head Start		0.212*	0.125
		(0.113)	(0.111)
Observations	2986	2986	2986
Index = Predicted Finish College			
Head Start	0.025	-0.088	-0.130
	(0.063)	(0.083)	(0.100)
Head Start x 4plus child family	0.281**		0.266**
	(0.112)		(0.112)
Head Start x Tercile 1 Predicted Finish College		0.237**	0.155
		(0.112)	(0.121)
Head Start x Tercile 2 Predicted Finish College		0.260**	0.207
		(0.131)	(0.142)
Observations	2986	2986	2986

This table shows estimates from a FFE regression of attainment of some college on an indicator for attendance of Head Start, and an indicator for having a family with 4 or more children (Column 1), dummies for terciles of an index of predicted Head Start attendance (Column 2, Panel A), dummies for terciles of an index of the predicted likelihood of finishing college (Column 2, Panel B), and the combination of family size indicator and terciles of the index (Column 3). The predicted Head Start (finish college) index is created by regressing Head Start attendance (finish college) on all of the control variables in the PSID analysis, except for the household size variable.

C Supplementary PSID FFE Results

In this section, we discuss additional FFE results obtained using our expanded PSID sample.

We present the FFE results for the economic and health indices measured at age 40, together with the indices at age 30 for comparison, in Table C.1. Overall, the results suggest little support for a positive long term effect of Head Start. We come to the same conclusions when we aggregate the inputs using principal components analysis (see Table C.2. Our overall conclusions are not changed importantly by looking at specific outcomes or subsamples. We have also estimated regressions for each of the inputs to the economic and health indices, which we include in Tables C.3, and C.4, C.5, C.6. Table C.7 shows the regression results for the additional outcomes analyzed in GTC, earnings between ages 23 to 25, and not having committed a crime. Across these tables, there is no systematic evidence that Head Start impacts long term outcomes.[64]

Motivated by the prior findings of differential effects by gender in Carneiro and Ginja (2014); Deming (2009), in Table C.8 we look to see whether our mean results are obscuring this form of heterogeneity in our setting. Curiously, we find some evidence of significant negative effects of Head Start among men, in particular for health and economic outcomes at age 40. On the other hand, we find a positive and significant effect of Head Start on the probability that men attain some college. The effects estimated for women are never individually significant, but also not statistically different from men for many outcomes as indicated by the p-value of the difference in the table. The one exception is for economic outcomes observed at age 40, where women are found to have significantly better returns to Head Start participation than observed for men.

Another source of heterogeneity which could generate a discrepancy between our results and GTC is the fact that our sample includes later (younger) cohorts, whose Head Start experience may differ from earlier participants. In Table C.9, we find some support for a decreasing impact of Head Start across cohorts for the age 40 indices, but also find a larger improvement in the health index at age 30 for more recent cohorts. Thus, this does not appear to reconcile our findings.[65]

[64]Moreover, while we find a significant increase in ataiment of some college, when we examine the outcome of college completion, we obtain insignificant negative point estimates for the pooled sample (beta = -0.033, se = 0.023), for black children (beta = -0.014, se = 0.018), and for white children (beta = -0.058, se = 0.043).

[65]Moreover, when we instead use a binary indicator for more recent cohorts, we do not find a statistically significant difference in the impacts of Head Start, indicating that these results are sensitive to functional form assumptions.

Table C.1: Impact of Head Start on Economic Sufficiency Index and Good Health Index (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Economic Sufficiency Index, age 30					
Head Start	-0.147***	-0.117**	-0.090	-0.117	-0.023
	(0.043)	(0.050)	(0.064)	(0.081)	(0.102)
Other preschool	0.184***	0.181***	0.091	0.050	0.099
	(0.035)	(0.040)	(0.062)	(0.109)	(0.072)
Mean Y	0.094	0.096	0.096	-0.552	0.213
Observations	7372	5361	5361	2369	2986
Economic Sufficiency Index, age 40					
Head Start	-0.080	-0.071	-0.059	-0.170	-0.081
	(0.066)	(0.077)	(0.100)	(0.134)	(0.125)
Other preschool	0.112*	0.085	0.043	-0.270	0.118
	(0.059)	(0.077)	(0.107)	(0.223)	(0.122)
Mean Y	0.020	0.025	0.025	-0.670	0.142
Observations	4085	2845	2503	1065	1435
Good Health Index, Age 30					
Head Start	-0.349***	-0.320***	-0.148	0.024	-0.265
	(0.058)	(0.064)	(0.143)	(0.149)	(0.249)
Other preschool	0.087**	0.096**	0.081	0.040	0.106
	(0.038)	(0.045)	(0.076)	(0.159)	(0.084)
Mean Y	0.004	0.017	0.017	-0.357	0.074
Observations	4749	3600	3114	1150	1959
Good Health Index, Age 40					
Head Start	-0.201*	-0.175	-0.147	0.031	-0.146
	(0.118)	(0.141)	(0.202)	(0.201)	(0.393)
Other preschool	0.117	0.095	0.119	0.382*	0.038
	(0.094)	(0.115)	(0.130)	(0.210)	(0.150)
Mean Y	0.011	0.015	0.015	-0.290	0.062
Observations	2228	1673	1306	511	795

Notes: This table shows the estimates from regressions of either the Economic Sufficiency Index at age 30 (panel A), the Economic Sufficiency Index at age 40 (panel B), the Good Health Index at age 30 (panel C), or the Good Health Index at age 40 (panel D) on an indicator for participation in Head Start and control variables described in the text. Regressions are run on the whole sample (column 1), siblings (columns 2 and 3), black siblings (column 4) and white siblings (column 5). All columns include control variables, and columns 3, 4, and 5 include mother fixed effects. The Good Health Index includes measures of not smoking cigarettes, good self reported health and BMI, averaged over the previous 5 years. The Economic Sufficiency Index includes measures of high school graduation, attendance of some college, no receipt of Food Stamps/SNAP, no receipt of AFDC/TANF, average earnings, employment, and unemployment, averaged over the previous 5 years. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and at mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.2: Effect of Head Start on Economic and Health Principal Components (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Economic Sufficiency Principal Component, age 30					
Head Start	-0.174***	-0.140**	-0.100	-0.138	-0.031
	(0.058)	(0.068)	(0.084)	(0.109)	(0.128)
Other preschool	0.295***	0.285***	0.150*	0.071	0.166
	(0.051)	(0.057)	(0.087)	(0.150)	(0.101)
Mean Y	0.154	0.160	0.160	-0.731	0.321
Observations	7372	5361	5361	2369	2986
Economic Sufficiency Principal Component, age 40					
Head Start	-0.113	-0.093	-0.082	-0.219	-0.127
	(0.090)	(0.106)	(0.131)	(0.180)	(0.155)
Other preschool	0.209**	0.173	0.091	-0.291	0.183
	(0.086)	(0.113)	(0.145)	(0.296)	(0.167)
Mean Y	0.026	0.032	0.032	-0.968	0.199
Observations	4085	2845	2503	1065	1435
Good Health Principal Component, Age 30					
Head Start	-0.248***	-0.228***	-0.073	0.057	-0.159
	(0.047)	(0.052)	(0.121)	(0.131)	(0.208)
Other preschool	0.070**	0.069*	0.063	0.033	0.083
	(0.031)	(0.037)	(0.063)	(0.137)	(0.069)
Mean Y	0.003	0.013	0.013	-0.309	0.062
Observations	4749	3600	3114	1150	1959
Good Health Principal Component, Age 40					
Head Start	-0.143	-0.126	-0.101	0.044	-0.174
	(0.107)	(0.128)	(0.200)	(0.200)	(0.400)
Other preschool	0.101	0.077	0.121	0.288	0.062
	(0.089)	(0.110)	(0.104)	(0.221)	(0.117)
Mean Y	0.009	0.015	0.015	-0.259	0.056
Observations	2228	1673	1306	511	795

Notes:

Table C.3: Effect of Head Start on Inputs to Economic Sufficiency Index at age 30 (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
High School Graduate					
Head Start	0.007	-0.002	-0.011	-0.024	-0.015
	(0.018)	(0.021)	(0.026)	(0.031)	(0.045)
Other preschool	-0.002	-0.008	0.036*	-0.012	0.046*
	(0.011)	(0.014)	(0.021)	(0.048)	(0.024)
Mean Y	0.913	0.912	0.912	0.862	0.921
Observations	7372	5361	5361	2369	2986
Attended Some College					
Head Start	0.038	0.039	0.046	-0.016	0.120**
	(0.024)	(0.029)	(0.033)	(0.036)	(0.053)
Other preschool	0.068***	0.069***	0.034	-0.011	0.043
	(0.019)	(0.023)	(0.039)	(0.046)	(0.047)
Mean Y	0.531	0.532	0.532	0.396	0.556
Observations	7372	5361	5361	2369	2986
Fraction of last 5 yrs not on Food Stamps/SNAP, age 30					
Head Start	-0.018	0.011	0.043	0.042	0.076
	(0.015)	(0.017)	(0.033)	(0.037)	(0.055)
Other preschool	-0.003	0.007	-0.019	-0.019	-0.015
	(0.007)	(0.009)	(0.018)	(0.047)	(0.019)
Mean Y	0.936	0.929	0.929	0.831	0.949
Observations	4186	2805	2175	887	1285
Never on AFDC/TANF by age 30					
Head Start	-0.028*	-0.015	-0.009	-0.001	0.001
	(0.016)	(0.018)	(0.020)	(0.023)	(0.034)
Other preschool	0.022***	0.026***	0.004	-0.010	0.005
	(0.008)	(0.009)	(0.011)	(0.025)	(0.012)
Mean Y	0.938	0.940	0.940	0.819	0.962
Observations	7372	5361	5361	2369	2986
Fraction of last 5 yrs with positive earnings, age 30					
Head Start	0.041***	0.035**	0.061	0.026	0.088
	(0.015)	(0.017)	(0.038)	(0.034)	(0.072)
Other preschool	0.013	0.008	0.015	-0.047	0.027
	(0.011)	(0.013)	(0.019)	(0.048)	(0.020)
Mean Y	0.895	0.898	0.898	0.845	0.907
Observations	4378	3295	2800	1054	1740
Mean Inc. Rel. Pov. in last 5 years, age 30					
Head Start	-29.579***	-27.953**	-16.953	5.860	-24.477
	(10.548)	(12.160)	(14.369)	(12.890)	(23.499)
Other preschool	42.704**	46.790***	-1.326	-4.147	0.923
	(18.606)	(17.411)	(16.118)	(17.769)	(18.924)
Mean Y	385.831	385.933	385.933	224.651	412.236
Observations	5293	4068	3694	1514	2175
Fraction of last 5 yrs no unemployment, age 30					
Head Start	-0.007	-0.001	0.005	-0.013	0.056
	(0.015)	(0.016)	(0.030)	(0.031)	(0.049)
Other preschool	-0.017	-0.013	-0.029	0.022	-0.040
	(0.012)	(0.014)	(0.027)	(0.029)	(0.032)
Mean Y	0.854	0.850	0.850	0.807	0.857
Observations	4259	3184	2670	981	1683

Notes: This table shows estimates from regressions of the inputs to the Economic Sufficiency Index at age 30 on an indicator for participation in Head Start together with control variables described in the text. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.4: Effect of Head Start on Inputs to Economic Sufficiency Index at age 40 (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Fraction of last 5 yrs not on Food Stamps/SNAP, age 40					
Head Start	0.001	0.009	0.045	0.054	0.051
	(0.019)	(0.020)	(0.033)	(0.044)	(0.049)
Other preschool	0.001	0.003	-0.010	-0.013	-0.008
	(0.010)	(0.013)	(0.023)	(0.062)	(0.023)
Mean Y	0.957	0.957	0.957	0.866	0.971
Observations	1972	1423	1213	564	647
Never on AFDC/TANF by age 40					
Head Start	0.008	0.022	-0.009	-0.010	0.002
	(0.020)	(0.023)	(0.030)	(0.039)	(0.048)
Other preschool	0.016	0.019	0.018	-0.034	0.025
	(0.010)	(0.012)	(0.021)	(0.062)	(0.021)
Mean Y	0.932	0.933	0.933	0.778	0.959
Observations	4085	2845	2503	1065	1435
Fraction of last 5 yrs with positive earnings, age 40					
Head Start	0.026	0.022	0.021	0.073	-0.180
	(0.031)	(0.038)	(0.062)	(0.053)	(0.130)
Other preschool	-0.004	-0.012	-0.026	-0.135***	0.003
	(0.027)	(0.033)	(0.051)	(0.052)	(0.060)
Mean Y	0.850	0.849	0.849	0.856	0.847
Observations	1829	1369	1078	445	633
Mean Inc. Rel. Pov. in last 5 years, age 40					
Head Start	1.769	3.447	32.738	27.251	-11.620
	(21.347)	(26.202)	(30.410)	(24.095)	(56.148)
Other preschool	97.953**	101.861**	24.513	17.035	26.140
	(38.986)	(47.085)	(40.157)	(22.343)	(50.412)
Mean Y	436.769	434.280	434.280	234.965	466.741
Observations	2152	1613	1272	540	732
Fraction of last 5 yrs no unemployment, age 40					
Head Start	-0.003	-0.022	-0.028	-0.033	-0.046
	(0.022)	(0.027)	(0.047)	(0.056)	(0.083)
Other preschool	-0.011	-0.011	-0.026	-0.053	-0.016
	(0.017)	(0.021)	(0.037)	(0.060)	(0.044)
Mean Y	0.906	0.902	0.902	0.841	0.911
Observations	1825	1365	1073	440	633
Fraction of last 5 yrs owned home, age 40					
Head Start	-0.022	-0.024	0.045	-0.058	0.070
	(0.049)	(0.056)	(0.056)	(0.054)	(0.121)
Other preschool	-0.041	-0.053	-0.057	-0.079	-0.058
	(0.037)	(0.044)	(0.058)	(0.079)	(0.074)
Mean Y	0.500	0.522	0.522	0.324	0.554
Observations	2292	1625	1391	642	747

Notes: This table shows estimates from regressions of the inputs to the Economic Sufficiency Index at age 40 on an indicator for participation in Head Start together with control variables described in the text. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.5: Effect of Head Start on Inputs to Good Health Index at age 30 (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 30					
Head Start	-0.064*	-0.031	0.021	-0.127*	0.049
	(0.035)	(0.039)	(0.080)	(0.072)	(0.110)
Other preschool	-0.017	0.017	-0.011	-0.181**	0.012
	(0.021)	(0.024)	(0.052)	(0.091)	(0.056)
Mean Y	0.745	0.755	0.755	0.785	0.750
Observations	2267	1742	1174	376	796
Fraction of last 5 yrs reported good or better health, age 30					
Head Start	-0.001	0.001	0.042	0.047	0.039
	(0.012)	(0.013)	(0.031)	(0.034)	(0.052)
Other preschool	0.008	0.004	0.005	-0.009	0.010
	(0.008)	(0.010)	(0.016)	(0.035)	(0.017)
Mean Y	0.948	0.950	0.950	0.890	0.959
Observations	3763	2806	2292	829	1459
Negative Mean BMI in last 5 years, age 30					
Head Start	-1.063**	-0.982*	-0.485	1.408	-1.514
	(0.436)	(0.506)	(0.765)	(0.984)	(1.128)
Other preschool	0.046	-0.096	-0.332	-0.357	-0.202
	(0.266)	(0.313)	(0.441)	(1.069)	(0.472)
Mean Y	-26.569	-26.615	-26.615	-28.826	-26.267
Observations	3248	2528	1978	689	1286

Notes: This table shows estimates from regressions of the inputs to the Good Health Index at age 30 on an indicator for participation in Head Start together with control variables described in the text. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.6: Effect of Head Start on Inputs to Good Health Index at age 40 (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Fraction of last 5 yrs smoked less than 1 cigarette/day, age 40					
Head Start	-0.022	0.013	0.002	0.074	0.099
	(0.047)	(0.050)	(0.075)	(0.077)	(0.148)
Other preschool	0.003	0.041	-0.033	0.218**	-0.104
	(0.039)	(0.047)	(0.126)	(0.097)	(0.150)
Mean Y	0.738	0.728	0.728	0.713	0.731
Observations	1280	930	698	300	398
Fraction of last 5 yrs reported good or better health, age 40					
Head Start	0.010	0.008	0.013	0.021	0.002
	(0.034)	(0.039)	(0.059)	(0.061)	(0.144)
Other preschool	0.016	0.010	0.026	0.026	0.025
	(0.029)	(0.035)	(0.023)	(0.065)	(0.023)
Mean Y	0.919	0.922	0.922	0.871	0.930
Observations	1463	1116	884	398	486
Negative Mean BMI in last 5 years, age 40					
Head Start	-1.218**	-1.297*	-0.976	-0.475	0.501
	(0.613)	(0.731)	(0.867)	(1.055)	(1.251)
Other preschool	-0.330	-0.741	-1.861***	1.271	-2.360***
	(0.424)	(0.518)	(0.647)	(1.503)	(0.693)
Mean Y	-27.504	-27.433	-27.433	-29.491	-27.095
Observations	2037	1486	1116	413	703

Notes: This table shows estimates from regressions of the inputs to the Good Health Index at age 40 on an indicator for participation in Head Start together with control variables described in the text. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.7: Impact of Head Start on High School, College, Earnings, and Criminal Behavior (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
A. Completed High School					
Head Start	0.007	-0.002	-0.011	-0.024	-0.015
	(0.018)	(0.021)	(0.026)	(0.031)	(0.045)
Other preschool	-0.002	-0.008	0.036*	-0.012	0.046*
	(0.011)	(0.014)	(0.021)	(0.048)	(0.024)
R-Squared	0.098	0.105	0.028	0.050	0.038
Observations	7372	5361	5361	2369	2986
B. Completed Some College					
Head Start	0.038	0.039	0.046	-0.016	0.120**
	(0.024)	(0.029)	(0.033)	(0.036)	(0.053)
Other preschool	0.068***	0.069***	0.034	-0.011	0.043
	(0.019)	(0.023)	(0.039)	(0.046)	(0.047)
R-Squared	0.213	0.233	0.050	0.056	0.057
Observations	7372	5361	5361	2369	2986
C. Ln Earnings 23-25					
Head Start	0.040	0.032	0.064	0.057	0.113
	(0.056)	(0.066)	(0.109)	(0.142)	(0.158)
Other preschool	0.064	0.035	0.084	0.174	0.070
	(0.045)	(0.052)	(0.098)	(0.173)	(0.110)
R-Squared	0.151	0.161	0.131	0.095	0.152
Observations	4351	3309	2726	986	1736
D. Not Booked/Charged with Crime					
Head Start	-0.007	-0.012	-0.008	0.028	-0.068
	(0.025)	(0.031)	(0.033)	(0.028)	(0.064)
Other preschool	-0.006	0.007	-0.002	-0.022	0.002
	(0.014)	(0.017)	(0.033)	(0.036)	(0.039)
R-Squared	0.055	0.062	0.089	0.074	0.106
Observations	5005	3591	3206	1366	1836

Notes: This table shows estimates from regressions of high school graduation (panel A), some college attainment (panel B), ln earnings between ages 23 and 25 (panel C) and not being charged with a crime (panel D) on an indicator for participation in Head Start together with control variables described in the text. Among the 7,372 individuals in the sample, 1098 individuals are in families that have variation in the Head Start variable (347 families), among those for whom we observe completed education; 887 black (277 black families), and 211 white individuals (70 white families). Crime sample limited to individuals age ≥ 16 at the time of interview in 1995. Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.8: Impact of Head Start on Main Outcomes by Sex

	Black		White	
	Female	Male	Female	Male
High School				
Head Start x Sex	0.008	-0.062	0.005	-0.043
	(0.033)	(0.042)	(0.059)	(0.054)
P-value of Difference	0.092		0.497	
Some College				
Head Start x Sex	-0.012	-0.021	0.102	0.145***
	(0.044)	(0.045)	(0.074)	(0.053)
P-value of Difference	0.873		0.582	
Ln Earnings 23-25				
Head Start x Sex	0.265	-0.238	0.133	0.078
	(0.171)	(0.202)	(0.217)	(0.174)
P-value of Difference	0.037		0.834	
No Crime				
Head Start x Sex	0.038	0.016	-0.036	-0.112
	(0.035)	(0.041)	(0.073)	(0.089)
P-value of Difference	0.661		0.448	
Economic Sufficiency Index, age 30				
Head Start x Sex	-0.052	-0.197**	-0.099	0.078
	(0.099)	(0.090)	(0.112)	(0.141)
P-value of Difference	0.148		0.252	
Economic Sufficiency Index, age 40				
Head Start x Sex	-0.021	-0.363**	0.058	-0.271
	(0.173)	(0.164)	(0.140)	(0.184)
P-value of Difference	0.098		0.099	
Good Health Index, Age 30				
Head Start x Sex	0.042	-0.004	-0.198	-0.361
	(0.159)	(0.218)	(0.278)	(0.378)
P-value of Difference	0.838		0.690	
Good Health Index, Age 40				
Head Start x Sex	0.349	-0.672**	0.605	-1.099**
	(0.273)	(0.271)	(0.378)	(0.480)
P-value of Difference	0.014		0.004	

This table shows estimates from regressions of our main outcomes on an indicator for participation in Head Start interacted with an indicator for being female or male. The estimated interactions between Head Start and female (male) are shown in columns 1 and 3 (2 and 4). Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table C.9: Regression: Interaction with Cohort (Linear) (PSID)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Economic Sufficiency Index, age 30					
Head Start	-0.054	-0.033	-0.038	-0.081	0.094
	(0.066)	(0.073)	(0.086)	(0.104)	(0.153)
Head Start x trend	-0.010**	-0.010*	-0.009	-0.007	-0.017
	(0.005)	(0.006)	(0.007)	(0.008)	(0.013)
Mean Y	0.094	0.096	0.096	-0.552	0.213
Observations	7372	5361	5361	2369	2986
Economic Sufficiency Index, age 40					
Head Start	-0.042	-0.038	-0.030	-0.155	-0.026
	(0.084)	(0.093)	(0.104)	(0.136)	(0.118)
Head Start x trend	-0.014	-0.015	-0.031*	-0.050**	-0.029
	(0.012)	(0.013)	(0.017)	(0.025)	(0.019)
Mean Y	0.020	0.025	0.025	-0.670	0.142
Observations	4085	2845	2503	1065	1435
Good Health Index, Age 30					
Head Start	-0.318***	-0.291***	-0.113	-0.087	0.018
	(0.064)	(0.065)	(0.161)	(0.167)	(0.293)
Head Start x trend	-0.004	-0.004	-0.007	0.034**	-0.044
	(0.007)	(0.007)	(0.019)	(0.017)	(0.034)
Mean Y	0.004	0.017	0.017	-0.357	0.074
Observations	4749	3600	3114	1150	1959
Good Health Index, Age 40					
Head Start	-0.135	-0.110	-0.129	0.066	0.422
	(0.149)	(0.167)	(0.210)	(0.188)	(0.513)
Head Start x trend	-0.028	-0.034	-0.026	0.067	-0.186**
	(0.024)	(0.026)	(0.037)	(0.044)	(0.083)
Mean Y	0.011	0.015	0.015	-0.290	0.062
Observations	2228	1673	1306	511	795

Notes: This table shows estimates from regressions of the Economic Sufficiency and Good Health Indices on an indicator participation in Head Start interacted with a normed linear trend in year of birth (year of birth minus 1966, where 1966 represents the first year that Head Start was available). Estimates are weighted to be representative of 1995 population; see text for details. Standard errors are clustered at 1968 family id in column 1 and on mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

D Replication of GTC (2002)

D.1 Summary

In this appendix we describe the results of our replication of Garces, Thomas and Currie (2002) (GTC). We describe our replication methods is in the Section D.2.

Table D.1 below shows the summary statistics corresponding to Table 1 of GTC for our sample. We include GTC Table 1 for comparison as Table D.2. In general, the results across the two tables are similar, albeit not identical. The most notable difference is that we find a lower share of respondents participate in Head Start, although the difference is smaller for the sibling sample. The shares of respondents who graduate high school and college are higher in our sample than in GTC. We report average earnings from age 23-25 in nominal terms as well as adjusted to 1999 dollars. Our adjusted earnings are consistently higher than GTC's reported adjusted earnings, but our unadjusted earnings are quite close to their mean adjusted earnings. We suspect that GTC may have reported unadjusted earnings, although it is also possible that the discrepancy is due to a slightly larger sample of individuals with earnings in GTC's sample. Again, the number of observations we report in the final row of the table is based on the number of individuals responding to the Head Start participation question.

Our replication of the main regression results in GTC are shown in Table D.3. We include GTC's Table 2 as Table D.4 for comparison. Our regression results are qualitatively similar, especially for the larger samples (panels A, B, and C). GTC found few statistically significant results, one of which was a negative effect of Head Start on high school completion before including controls. We, too, find this negative and significant result, though ours is slightly smaller. The result in Column (6), which GTC find to be positive and significant, we do not find to be significant. Our results for the college outcomes are aligned with the findings in GTC. The magnitudes that we report are not statistically different from GTC and in particular we replicate the key finding that Head Start influences college going for white children and not for black children. Our replication of Panel C is qualitatively similar to GTC. We do not find a statistically significant decrease in black crime rates as GTC do, although our point estimates are consistently negative for blacks. Otherwise, our estimates are quite imprecise and not statistically different from GTC's.

Our earnings results (panel C replication) are quite different from GTC, but this may be due to differences in how we defined earnings rather than differences in our samples. This is apparent in the fact that we have many fewer observations than GTC beginning from column 2 onward, about 24% smaller in column 2 and 48% smaller in column 8.

D.2 Replication Methodology

This section documents the process of replicating Garces, Thomas and Currie (2002) (GTC) for future scholars wishing to repeat our steps. We describe three stages of the replication: construction of the dataset, iterations to identify the likely variable definitions, and our final decisions based on these iterations. We also include information about the mechanics of downloading the data and the variables we use.

D.2.1 Construction of Dataset

We begin by assembling data from the Panel Study of Income Dynamics (PSID), a nationally representative longitudinal dataset that forms the basis for the analysis in GTC. The PSID consists of the survey responses of household heads and their wives, which compose the annual household-level datasets ("family files"), as well as a smaller database of responses of all individuals in the household to a small set of questions ("cross-year individual files"). We merge the family files to the cross-year individual files using the "case id" number, which is present both on the individual and family files. We also merge responses of an individual's mother and father from the crossyear file for those individuals whose mother or father have been identified in the PSID crossyear file.

The result is a dataset with 71,285 individual observations, each of which contains the personal responses of an individual over time, the responses (usually given by the head of household) to the family interview questions for each year, and the responses of an individual's parents to the cross-year survey. The base dataset includes the Survey of Economic Opportunity "poverty oversample" and the Latino oversample, two populations specifically targeted by the PSID in order to improve the representativeness of the survey. We proceed by excluding the Latino oversample in accordance with GTC's footnote 4.

Next, we construct the variables needed to define our sample. GTC delineate the specifications for their sample throughout the paper, and in particular we rely on their descriptions in Section II and footnote 7. A key stratifying variable in GTC is race, which is also a limiting factor for the sample size since the GTC sample is restricted to only black and white individuals (see footnote 4 of GTC). Unfortunately, the PSID does not assign a race to each individual, so race must be imputed from the annual family responses about race. Specifically, the PSID surveys families about the race of the head and wife of the head of household, so an individual's race can only be identified if that individual becomes a head of household or his wife. Otherwise we must infer the race of the individual through their relation to the head of household or his wife.

The process of identifying race from the responses of other family members can be done at any age and from a variety of different family members, so we have experimented with using more and less restrictive definitions. We establish five definitions of race based on the relations through which we allow inference and the survey years over which we make the inference. These definitions are summarized over those two dimensions below in Table D.5.

The second limiting criterion is the age of individuals. GTC include respondents aged 18 and over in 1995, which results in a sample of respondents born between 1965 and 1977. They exclude the 1964 and 1965 cohorts. Since this sample restriction can be defined and replicated in a few different ways with PSID variables, we develop three candidate limitations on age and year of birth for individuals in our sample. We describe the criteria which define these alternative candidates in Table D.6.

The third criterion is to identify sets of siblings within the remaining sample that comprise the "siblings subsample." Since the identification strategy relies on the inclusion of a mother fixed effect, we define siblings as any two individuals who satisfy the race and age criteria for the sample and have the same unique mother identification number. The mother identification number is a combination of a family identifier and a personal identifying number which is assigned by the PSID. Individuals that do not have a mother identification number are excluded from the sibling subsample.

Next, we flag observations from the SEO poverty oversample with the intention of excluding them as GTC do. We ultimately do not exclude these observations because comparisons of the sample statistics with and without the SEO sample make us speculate that the results in GTC were generated from a sample that included the SEO sample.

We construct sample weights using CPS weights to make the sample representative of the 1995 white and African-American populations. Specifically, we collapsed the 1995 CPS weights to age-race-sex cells (year of birth is not available) and merge the cell weight onto each observation of our sample. Then, we divide the cell weight by the number of individuals in that age-race-sex cell who are in our sample and the resulting individual weight is what we use for our analysis.

D.2.2 Search for identical dataset construction

As mentioned previously, the sample construction criteria are clearly documented in GTC. For some dimensions, we could think of a few ways to define variables and samples in accordance with their descriptions. Therefore, we conducted tests to determine the procedures that would yield a dataset consistent with GTC, as well as to assess the stability of the results.

Our search iterations hinge on four parameters: inclusion or exclusion of the SEO oversample; the algorithm for identifying an individual's race; the criteria for age; and the order in which we dropped observations and weighted the sample. For this last parameter, we weighted the sample before dropping the Latino oversample as well as after. We do not present the results for the variations on this final parameter because the exercise clearly indicated that dropping the Latino oversample best matched GTC's results regardless of how the first three parameters were defined.

Table D.7 below shows the results of our iteration of the summary statistics results for a select set of variables. Our goal was to match the results to Table 1 in GTC, reproduced on the first row of the table. The number of observations we report is for the variable for Head Start participation, although some variables have fewer observations. For example, over half the observations for the income variable are missing. GTC also report one N for each column, although they also likely had fewer observations for variables like income.

Our sample is weighted based on race, gender, and age variables from the CPS, so we expect that the mean values for the weighted PSID sample should be similar to the CPS means. We include the CPS means for the three variables as a comparison. The definitions for age and race are as described in the previous section. There are a number of conclusions we draw from this table. First, we speculate that the 25.17 percent black reported in GTC is, in fact, 15.17 percent, which is much closer to the CPS means. Second, inclusion of the SEO oversample adds approximately 1,500 observations to our sample and brings us quite close to the size of the sample and sample means reported in GTC.

As we had hoped, moving from iteration to iteration substantially changes the number of observations, which suggest which decisions produced the sample of GTC. For example, holding SEO and age definitions constant, moving from our conservative definition of race (2) to the liberal definition (4) adds approximately 30 to 50 observations, an approximately 1.5 percent increase in sample size. The specification of age is also important for defining the sample size. For example, the movement from row 1, 1, 2 (N=3,286) to 1, 2, 2 (N=3,548) is an eight percent increase, and the subsequent movement to row 1, 3, 2 (N=4,187) is an 18 percent increase.

Despite the variability in sample size, our sample characteristics are not sensitive to the decisions along each of these dimensions. Additionally, while our results for these select variables are at times statistically different from those of GTC, we remain close to the magnitudes that they report. The race, gender, and age means are very similar across the specifications, likely on account of the weighting. The preschool participation and high school graduation rates are nearly identical throughout, especially when we include the SEO oversample. The exception to this pattern is Head Start participation. The SEO oversample increases the share of respondents who were in Head Start to close to nine percent, which is still lower than the 10.57 percent reported in GTC. We were unable to replicate this high incidence of Head Start participation throughout the iteration process, including in iterations not reported here.

We also performed iterations on the regression models from GTC's Table 2. GTC conduct a similar regression for each of four outcome variables: high school graduation, college graduation, crime, and later earnings. The first of these three are fairly similar: they are defined by one variable in the PSID. In this comparison table we only show results for high school graduation. On the other hand, compiling a consistent variable for earnings is trickier. Here we present results for one of our regressions, but in general we were not able to replicate the findings for this outcome variable.

There are eight different models in GTC. The first three are on the full sample, the sibling sample, and the sibling sample with controls. The next five models use mother fixed effects: first on the full sample, then the full sample split by whether the mother was white or black, and finally for the subset of mothers with less than a high school education, also split by race.

Table D.8 shows a comparison of the results. We show iterations on the same three age restrictions as above, as well as race definitions for definitions 4 and 5 as defined in the previous section. For each regression the corresponding result from GTC is shown on the first row.

Our regression results are qualitatively similar, especially for the larger samples (panel A). GTC found few statistically significant results, one of which was a negative effect of Head Start on high school completion (result A.1). We, too, replicate this negative and significant result, though ours are smaller. As can be noted in result A.4, our models using later earnings were similar to those in the paper. The result in B.4, which GTC find to be positive and significant, we do not find significant. However, all of our replications of this result fall within the confidence interval they use.

Among our various iterations, the results are stable. Only the result in A.1 has a difference of one standard error between estimates, with the rest of these results never straying more than half a standard error from each other.

D.3 Final dataset restrictions

Given our iteration exercises, our preferred sample definition includes the SEO poverty over-sample, uses age definition 1 and uses race definition 5 as explained in the first section of this appendix. Our choice of age and race definitions is appropriate for three reasons. First, they replicate the GTC adequately. Second, they are a reasonable method for a researcher not attempting to replicate findings. Third, they result in large samples, which is important for additional analyses.

D.4 More on the data

We downloaded the data files from http://simba.isr.umich.edu/Zips/ZipMain.aspx. Table D.9 shows the variables we downloaded.

Table D.1: Replication of Garces, Thomas, Currie (2002) Summary Statistics

	All	Head Start	No Head Start	Sibling Sample
Head Start	0.0873	1	0	0.103
	(.282)	(0)	(0)	(.304)
Other preschool	0.266	0	0.291	0.281
	(.442)	(0)	(.454)	(.45)
Fraction completed hs	0.851	0.752	0.860	0.854
	(.356)	(.432)	(.347)	(.353)
Fraction attended some college	0.468	0.339	0.481	0.482
	(.499)	(.474)	(.5)	(.5)
Avg. Earnings age 23-25	18543.5	13361.3	18962.7	20116.3
	(14929)	(12057)	(15062)	(17141)
Avg. Earnings age 23-25 (CPI adjusted)	20367.9	14730.7	20823.9	21734.8
	(15646)	(12950)	(15758)	(17521)
Fraction booked/charged with crime	0.0998	0.124	0.0975	0.106
	(.3)	(.33)	(.297)	(.308)
Fraction African-American	0.150	0.619	0.105	0.162
	(.357)	(.486)	(.307)	(.369)
Fraction female	0.502	0.533	0.499	0.475
	(.5)	(.499)	(.5)	(.5)
Age in 1995	23.67	23.14	23.72	23.14
	(3.44)	(3.5)	(3.43)	(3.28)
Fraction eldest child in family	0.345	0.335	0.346	0.364
	(.475)	(.472)	(.476)	(.481)
Fraction low birth weight	0.0608	0.110	0.0553	0.0560
	(.239)	(.314)	(.229)	(.23)
Mother's yrs education	11.36	10.00	11.49	11.17
	(2.58)	(2.44)	(2.56)	(2.54)
Fraction whose mother completed hs	0.772	0.585	0.790	0.770
	(.419)	(.493)	(.407)	(.421)
Father's yrs education	11.46	9.806	11.60	11.37
	(3.01)	(2.78)	(2.98)	(3)
Fraction whose father completed hs	0.725	0.475	0.747	0.717
	(.446)	(.5)	(.435)	(.451)
Family income (age 3-6) (CPI adjusted)	48040.3	30253.9	49699.4	48580.8
	(27470)	(15498)	(27756)	(29193)
Had a single mother at age 4	0.119	0.320	0.0998	0.108
	(.324)	(.467)	(.3)	(.31)
Household size at age 4	4.659	5.109	4.616	4.831
	(1.81)	(2.18)	(1.76)	(1.71)
Observations	3399	552	2847	1541

Notes: Weighted to be representative of 1995 population; see text for details.

Table D.2: GTC Table 1: Summary Statistics

	All sample	Head Start	Not in Head Start	Sibling Sample
Head Start	0.1057	1	0	0.1089
	(.0053)	(0)	(0)	(.0073)
Other preschool	0.2834	0.1333	0.3011	0.2771
	(.0077)	(.0151)	(.0085)	(.0105)
Pct. completed hs	0.7660	0.6465	0.7803	0.7721
	(.0074)	(.0216)	(.0079)	(.0101)
Pct. attended some college	0.3714	0.2508	0.3859	0.3880
	(.0085)	(.0196)	(.0093)	(.0117)
Average earnings between age 23-25	-	-	-	-
	-	-	-	-
Average earnings between age 23-25 - CPI adjusted	17290	12100	17810	17310
	(690)	(670)	(760)	(1000)
Pct. booked/charged with crime	0.0969	0.1104	0.0953	0.1004
	(.0051)	(.00139)	(.0054)	(.0070)
Pct. African-American	0.2517	0.7532	0.1924	0.2285
	(.0074)	(.00192)	(.0078)	(.0098)
Pct. female	0.5149	0.5641	0.5091	0.5075
	(.0085)	(.0220)	(.0093)	(.0117)
Age in 1995	23.66	23.35	23.70	23.65
	(.06)	(.15)	(.06)	(.08)
Pct. eldest child in family	0.5311	0.5089	0.5337	0.5057
	(.0056)	(.0141)	(.0061)	(.0076)
Pct. low birth weight	0.0699	0.1040	0.0659	0.0669
	(.0037)	(.0124)	(.0038)	(.0056)
Mother's yrs education	12.14	11.33	12.24	12.30
	(.04)	(.09)	(.04)	(.05)
Pct. whose mother completed hs	0.7037	0.5552	0.7212	0.7815
	(.0078)	(.0221)	(.0083)	(.0097)
Father's yrs education	11.60	10.19	11.76	12.23
	(.06)	(.14)	(.06)	(.07)
Pct. whose father completed hs	0.5612	0.2638	0.5964	0.6330
	(.0085)	(.0196)	(.0091)	(.0113)
Family income (age 3-6) - CPI adjusted	46230	26620	48540	47330
	(460)	(580)	(500)	(670)
Had a single mother at age 4	0.1642	0.4035	0.1359	0.1306
	(.0061)	(.0216)	(.0061)	(.0079)
Household size at age 4	4.59	4.97	4.55	4.84
	(.03)	(.09)	(.03)	(.04)
Observations	3255	489	2766	1742

Table D.3: Replication of Garces, Thomas, Currie (2002) Regressions

	All	Sibs	Controls	Mom FE	Blk, FE	Wht, FE
Panel A. High School						
Head Start	-0.064*	-0.017	0.009	0.031	-0.017	0.093
	(0.034)	(0.043)	(0.040)	(0.057)	(0.063)	(0.092)
Other Preschool	0.082***	0.076***	0.014	0.028	0.068	0.021
	(0.013)	(0.022)	(0.021)	(0.035)	(0.072)	(0.038)
Observations	3399	1541	1541	1541	615	923
Panel B. College						
Head Start	-0.027	-0.021	0.033	0.100*	-0.039	0.232**
	(0.035)	(0.053)	(0.045)	(0.059)	(0.059)	(0.094)
Other Preschool	0.200***	0.219***	0.098***	0.047	-0.062	0.059
	(0.025)	(0.034)	(0.033)	(0.044)	(0.101)	(0.049)
Observations	3399	1541	1541	1541	615	923
Panel C. Earnings						
Head Start	-0.139*	-0.142	-0.056	-0.041	0.427*	-0.322
	(0.074)	(0.108)	(0.113)	(0.191)	(0.245)	(0.261)
Other Preschool	0.067	-0.023	-0.125*	-0.013	0.286	-0.017
	(0.062)	(0.072)	(0.074)	(0.116)	(0.448)	(0.118)
Observations	2118	972	972	779	236	541
Panel D. No Crime						
Head Start	-0.028	0.069	-0.055	-0.086	0.065	-0.222*
	(0.028)	(0.050)	(0.049)	(0.070)	(0.044)	(0.125)
Other Preschool	-0.000	-0.020	0.004	-0.046	0.059	-0.059
	(0.015)	(0.019)	(0.020)	(0.038)	(0.052)	(0.043)
Observations	3387	1537	1537	1535	614	918

Notes: * p < .10, ** p < .05, *** p < .01. Weighted to be representative of 1995 population; see text for details. SE clustered at 1968 family id in column 1 and at mother id level otherwise.

Table D.4: GTC Table 2: Regressions

	(1) All	(2) Sibs	(3) Controls	(4) Mom FE	(5) Blk, FE	(6) Wht, FE	(7) Blk, l.e. HS	(8) Wht, l.e. HS
Completed high School								
Head Start	-0.089	-0.075	0.006	0.037	-0.025	0.203	0.000	0.283
	(0.026)	(0.035)	(0.034)	(0.053)	(.065)	(0.098)	(0.071)	(0.119)
Other Preschool	0.085	0.073	0.003	-0.032	-0.056	-0.014	-0.080	-0.019
	(0.016)	(0.022)	(0.021)	(0.038)	(0.064)	(0.048)	(0.077)	(0.067)
Difference	-0.174	-0.148	0.003	0.069	0.031	0.217	0.081	0.302
S.E Difference	0.028	0.037	0.039	0.062	0.085	0.105	0.097	0.126
N	3255	1742	1742	1742	706	1036	554	677
Attended Some College								
Head Start	-0.038	-0.016	0.075	0.092	0.023	0.281	0.031	0.276
	(0.023)	(0.033)	(0.033)	(0.056)	(.066)	(0.108)	(0.067)	(0.120)
Other Preschool	0.142	0.149	0.023	0.050	-0.007	0.095	0.022	0.0103
	(0.019)	(0.027)	(0.026)	(0.040)	(0.064)	(0.052)	(0.072)	(0.068)
Difference	-0.180	-0.165	0.052	0.042	0.030	0.186	0.009	0.173
S.E Difference	0.028	0.040	0.041	0.065	0.085	0.115	0.092	0.127
N	3255	1742	1742	1742	706	1036	554	677
ln(earnings 23-25)								
Head Start	-0.034	0.053	0.170	0.194	0.073	0.566	0.051	1.004
	(0.090)	(0.116)	(0.117)	(0.257)	(0.321)	(0.459)	(0.357)	(0.516)
Other Preschool	0.173	0.174	0.002	0.079	-0.087	0.146	0.124	0.136
	(0.063)	(0.086)	(0.082)	(0.171)	(0.287)	(0.219)	(0.341)	(0.306)
Difference	-0.207	-0.122	0.167	0.115	0.160	0.420	-0.073	0.868
S.E Difference	0.104	0.138	0.144	0.302	0.420	0.504	0.482	0.548
N	1383	728	728	728	272	456	216	320
Booked or charged with crime								
Head Start	0.023	0.041	0.012	-0.053	-0.116	0.122	-0.126	0.058
	(0.018)	(0.026)	(0.026)	(0.039)	(0.045)	(0.077)	(0.050)	(0.095)
Other Preschool	-0.017	- 0.022	-0.001	0.032	0.000	0.063	-0.023	0.147
	(0.011)	(0.016)	(0.017)	(0.028)	(0.045)	(0.036)	(0.056)	(0.054)
Difference	0.040	0.063	0.013	-0.085	-0.117	0.059	-0.103	-0.089
S.E Difference	0.020	0.028	0.030	0.045	0.059	0.082	0.070	0.100
N	3255	1742	1742	1742	706	1036	554	677

SE in parentheses.

Table D.5: Alternative Definitions of Race

Defn.	Survey Years		Relation to Head (or Wife)				
	1995	1985-1996	Head	Wife	Child	Parent	Sibling
1	X		X	X	X		
2	X		X	X	X	X	X
3		X	X	X	X		
4		X	X	X	X	X	X
5		X	X	X	X	X	X

Table D.6: Candidate limitations on birth year and age

Defn.	BirthYears			Age in 1995		
	1966-1977	Not 1965, 1978	No Restriction	>18	17-29	17-30
1	X			X		
2		X			X	
3			X			X

Table D.7: Iterations for Summary Statistics Table

			Black	Female	Age	Head Start	Preschool	High School	N
GTC(2002)			0.252	0.515	23.660	0.106	0.283	0.766	3255
CPS 1995			0.150	0.505	23.686				
Sample Iterations									
SEO	Age	Race							
0	1	2	0.149	0.497	22.952	0.078	0.302	0.822	1708
0	1	4	0.149	0.497	22.950	0.079	0.299	0.820	1735
0	2	2	0.154	0.499	22.859	0.079	0.309	0.811	1855
0	2	4	0.154	0.499	22.857	0.080	0.306	0.809	1883
0	3	2	0.150	0.503	23.713	0.076	0.286	0.820	2173
0	3	4	0.150	0.503	23.712	0.076	0.284	0.818	2204
1	1	2	0.153	0.498	22.959	0.089	0.290	0.788	3286
1	1	4	0.153	0.498	22.958	0.089	0.288	0.787	3333
1	2	2	0.157	0.500	22.926	0.087	0.292	0.782	3548
1	2	4	0.157	0.500	22.925	0.087	0.290	0.781	3597
1	3	2	0.150	0.503	23.710	0.082	0.276	0.788	4187
1	3	4	0.120	0.503	23.710	0.082	0.274	0.787	4244

Notes: First row corresponds to selections from Garces, Thomas and Currie (2002) table 1. Second row corresponds to 1995 CPS means, as described in the text of the appendix. The next 12 columns correspond to sample iterations on three criteria. The first is the inclusion (SEO=1) or exclusion (SEO=0) of the Survey of Economic Opportunity sample. The three age criteria and two race criteria are explained in detail in the previous table. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table D.8: Iterations for Regressions Table

Panel A.		HS, All			HS, Sib			HS, Mom FE			Log Earnings, All		
		b	se	N	b	se	N	b	se	N	b	se	N
GTC (2002)		-0.089	(0.026)	3255	-0.075	(0.035)	1742	0.037	(0.053)	1742	-0.034	(0.090)	1383
Sample Iterations													
Age	Race												
1	4	-0.075	(0.030)	3315	-0.035	(0.043)	1543	0.047	(0.075)	1543	-0.064	(0.106)	894
1	5	-0.071	(0.030)	3344	-0.025	(0.042)	1565	0.047	(0.075)	1565	-0.067	(0.105)	898
2	4	-0.073	(0.030)	3585	-0.034	(0.039)	1731	0.072	(0.077)	1731	-0.064	(0.104)	894
2	5	-0.067	(0.031)	3616	-0.024	(0.039)	1753	0.072	(0.076)	1753	-0.067	(0.104)	898
3	4	-0.052	(0.026)	4233	-0.046	(0.035)	2125	0.037	(0.063)	2125	-0.043	(0.092)	1132
3	5	-0.046	(0.027)	4264	-0.036	(0.035)	2147	0.036	(0.062)	2147	-0.046	(0.092)	1136

Panel B.		HS, Mom FE, Black			HS, Mom FE, White			HS, Mom<HS, Black			HS, Mom<HS, White		
		b	se	N	b	se	N	b	se	N	b	se	N
GTC (2002)		-0.025	(0.065)	706	0.203	(0.098)	1036	0	(0.071)	554	0.283	(0.119)	677
Sample Iterations													
Age	Race												
1	4	-0.030	(0.058)	625	0.133	(0.089)	898	-0.026	(0.058)	586	0.152	(0.099)	672
1	5	-0.030	(0.058)	625	0.133	(0.088)	920	-0.026	(0.058)	586	0.152	(0.098)	692
2	4	-0.028	(0.056)	702	0.181	(0.094)	1008	-0.025	(0.056)	649	0.203	(0.105)	759
2	5	-0.028	(0.056)	702	0.181	(0.092)	1030	-0.025	(0.056)	649	0.202	(0.104)	779
3	4	-0.043	(0.044)	858	0.120	(0.081)	1241	-0.045	(0.044)	797	0.136	(0.092)	961
3	5	-0.043	(0.044)	858	0.114	(0.079)	1263	-0.045	(0.044)	797	0.130	(0.088)	981

Notes: First row of each panel corresponds to selections from Garces, Thomas and Currie (2002) table 2. The three age criteria and two race criteria are explained in detail in the previous table. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table D.9: PSID Variables used in the analysis

Our Variable	PSID Original Variable	Description (derived variable)	Source
id1968	ER30001	Family identifier	Indiv. Cross year
pernum	ER30002	Personal identifier	Indiv. Cross year
relation1968-relation2001	ER30003, ER30022, ER30045, ER30069, ER30093, ER30119, ER30140, ER30162 ER30190, ER30219, ER30248 ER30285, ER30315, ER30345, ER30375, ER30401, ER30431, ER30465, ER30500, ER30537, ER30572, ER30608, ER30644 ER30691, ER30735, ER30808, ER33103, ER33203, ER33303, ER33403, ER33503, ER33603	Relation to head	Indiv. Cross year
caseid1968-caseid2011	ER30020 ER30043 ER30067 ER30091 ER30117 ER30138 ER30160 ER30188 ER30217 ER30246 ER30283 ER30313 ER30343 ER30373 ER30399 ER30429 ER30463 ER30498 ER30535 ER30570 ER30606 ER30642 ER30689 ER30733 ER30806 ER33101 ER33201 ER33301 ER33401 ER33501 ER33601 ER33701 ER33801 ER33901 ER34001 ER34101 ER33601	Fam. Intervew Number	Indiv. Cross year
edu1968-edu2011	ER30010 ER30052 ER30076 ER30100 ER30126 ER30147 ER30169 ER30197 ER30226 ER30255 ER30296 ER30326 ER30356 ER30384 ER30413 ER30443 ER30478 ER30513 ER30549 ER30584 ER30620 ER30657 ER30703 ER30748 ER30820 ER33115 ER33215 ER33315 ER33415 ER33516 ER33616 ER33716 ER33817 ER33917 ER34020 ER34119	Yrs. Education	Indiv. Cross year
age1995	ER33204	Age in 1995	Indiv. Cross year
birthyr1995	ER33206	Birthyear in 1995	Indiv. Cross year
headstart1995	ER33261	Head Start Response in 1995	Indiv. Cross year
preschool1995	ER33264	Preschool Response in 1995	Indiv. Cross year
preschool1995	ER33266	Crime Response in 1995	Indiv. Cross year
sex	ER32000	Sex	Indiv. Cross year
momid1968	ER32009	Mother's Family ID	Indiv. Cross year
mompernum	ER32010	Mother's Personal ID	Indiv. Cross year
dadid1968	ER32016	Father's Family ID	Indiv. Cross year
dadpernum	ER32017	Father's Personal ID	Indiv. Cross year
birthweight	ER32014	Birth weight	Indiv. Cross year

Our Variable	PSID Original Variable	Description (derived variable)	Source
crime1995	ER33266	Committed/Charged with Crime	Indiv. Cross year
parityofmom	ER32013	Parity of mom (Eldest)	Indiv. Cross year
h_edu1968-h_edu2011	V313 V794 V1485 V2197 V2823 V3241 V3663 V4198 V5074 V5647 V6194 V6787 V7433 V8085 V8709 V9395 V11042 V12400 V13640 V14687 V16161 V17545 V18898 V20198 V21504 V23333 ER4158 ER6998 ER9249 ER12222 ER16516 ER20457 ER24148 ER28047 ER41037 ER46981 ER52405	Education of Head (Mom, Dad Education)	Family Interviews
w_edu1968, w_edu1972-w_edu2011	V246 V2687 V3216 V3638 V4199 V5075 V5648 V6195 V6788 V7434 V8086 V8710 V9396 V11043 V12401 V13641 V14688 V16162 V17546 V18899 V20199 V21505 V23334 ER4159 ER6999 ER9250 ER12223 ER16517 ER20458 ER24149 ER28048 ER41038 ER46982 ER52406	Education of Wife of Head (Mom Education)	Family Interviews
h_sex1968-h_sex2011	V119 V1010 V1240 V1943 V2543 V3096 V3509 V3922 V4437 V5351 V5851 V6463 V7068 V7659 V8353 V8962 V10420 V11607 V13012 V14115 V15131 V16632 V18050 V19350 V20652 V22407 ER2008 ER5007 ER7007 ER10010 ER13011 ER17014 ER21018 ER25018 ER36018 ER42018 ER47318	Sex of Head (Single mom)	Family Interviews
f_tanf1994-f_tanf2011	ER3262 ER6262 ER8379 ER11272 ER14538 ER18697 ER22069 ER26050 ER37068 ER43059 ER48381	Family Received AFDC/TANF last year	Family Interviews
f_fs1994-f_fs2011	ER3059 ER6058 ER8155 ER11049 ER14255 ER18386 ER21652 ER25654 ER36672 ER42691 ER48007	Family Received Food Stamps last year	Family Interviews
h_cigs1986, h_cigs1999-h_cigs2011	V13442 ER15544 ER19709 ER23124 ER27099 ER38310 ER44283 ER49621	Cigarettes Per Day of Head	Family Interviews
w_cigs1986, w_cigs1999-w_cigs2011	V13477 ER15652 ER19817 ER23251 ER27222 ER39407 ER45380 ER50739	Cigarettes Per Day of Wife of Head	Family Interviews
h_wlbs1999-h_wlbs2011	ER15552 ER19717 ER23132 ER38320 ER44293 ER49631	Weight of Head (BMI)	Family Interviews
w_wlbs1999-w_wlbs2011	ER15660 ER19825 ER23259 ER27232 ER39417 ER45390 ER50749	Weight of Wife of Head (BMI)	Family Interviews

Our Variable	PSID Original Variable	Description (derived variable)	Source
h_srhealth1984-h_srhealth2011	V10877 V11991 V13417 V14513 V15993 V17390 V18721 V20021 V21321 V23180 ER3853 ER6723 ER8969 ER11723 ER15447 ER19612 ER23009 ER26990 ER38202 ER44175 ER49494	Self-Reported Health of Head	Family Interviews
w_srhealth1984-w_srhealth2011	V10884 V12344 V13452 V14524 V15999 V17396 V18727 V20027 V21328 V23187 ER3858 ER6728 ER8974 ER11727 ER15555 ER19720 ER23136 ER27113 ER39299 ER45272 ER50612	Self Reported Health of Head of Wife	Family Interviews
f_rentown1968-f_rentown2011	V103 V593 V1264 V1967 V2566 V3108 V3522 V3939 V4450 V5364 V5864 V6479 V7084 V7675 V8364 V8974 V10437 V11618 V13023 V14126 V15140 V16641 V18072 V19372 V20672 V22427 ER2032 ER5031 ER7031 ER10035 ER13040 ER17043 ER21042 ER25028 ER36028 ER42029 ER47329	Family Rents/Owns Home	Family Interviews
h_wages1968-h_wages2011	V251 V699 V1191 V1892 V2493 V3046 V3458 V3858 V4373 V5283 V5782 V6391 V6981 V7573 V8265 V8873 V10256 V11397 V12796 V13898 V14913 V16413 V17829 V20178 V21484 V23323 ER4140 ER6980 ER9231 ER12080 ER16463 ER20443 ER24116 ER27931 ER40921 ER46829 ER52237	Earnings of Head	Family Interviews
w_wages1968-w_wages2011	V76 V516 V1198 V1899 V2500 V3053 V3465 V3865 V4379 V5289 V5788 V6398 V6988 V7580 V8273 V8881 V10263 V11404 V12803 V13905 V14920 V16420 V17836 V19136 V20436 V23324 ER4144 ER6984 ER9235 ER12082 ER16465 ER20447 ER24135 ER27943 ER40933 ER46841 ER52249	Earnings of Wife of Head	Family Interviews

E Functional form choices with Binary Treatment and Binary Outcome

We now consider potential sensitivity to functional form modeling assumptions. For binary outcomes the usual choice of specifications include linear probability model (LPM), logit, and probit. In the cross-sectional setting, the conventional wisdom is that the choice among these options is fairly innocuous, especially when the objective is to recover the ATE.[66] We are not aware of any previous systematic exploration of these properties in extremely short-panel settings such as found in the FFE design. We demonstrate some complications that arise in such settings, and compare the performance of these estimators.

E.1 Specification choices

Empiricists commonly use LPM specification to estimate FE models. In our sample of papers, this is almost universally used as the primary, if not only, specification. We speculate that this is motivated by (1) the intuition carried over from the cross-sectional case that LPM models usually recover the ATE; (2) the benefit that the incidental parameters problem does not pollute the main parameters of interest (Chamberlain, 1980);[67] (3) computational ease, especially when paired with other complications to the research design such as many fixed effects, instrumental variables, etc.); and (4) the fact that the estimated coefficient β_{LPM} directly gives the estimate of the ATE.

Obtaining ATE from a nonlinear specification is not only less common, but also sometimes less straightforward. The conditional logit model, sometimes referred to as logit FE, consistently estimates β_{Logit} by conditioning on the number of successes in a family, but does not have a paired method for obtaining treatment effects. To obtain ATE, Wooldridge (2010, section 15.8) recommends employing a regular logit model and including family-level-means of control variables, i.e. "Chamberlain-Mundlak controls," (hereafter, Mundlak controls) rather than directly controlling for fixed effects (Mundlak, 1978; Chamberlain, 1980).[68]

Fernandez-Val (2009) examines the probit FE model. He proposes a bias-correction approach, which is based on the large-T asymptotic bias resulting from the incidental parameters problem. He also derives a "small bias" property for uncorrected/naive estimates of marginal effects for the probit FE model, and demonstrates this for panels of length as short as T=4. However it is not clear that the results in Fernandez-Val (2009) should apply in the family FE setting. This is because: (1)

[66] See Angrist and Pischke (2009, pg. 107) and Wooldridge (2010, section 15.6). In contrast, Cameron and Trivedi (2005, pg. 471) recommend limiting LPM's to exploratory analysis, and note that it does not do a good job making predicted probabilities for individual observations. In panel contexts, textbook treatments generally state that estimates should be fine using LPM (Wooldridge, 2010, pg. 608).

[67] Because this inconsistency is based on the panel length being fixed, the problem may be especially acute for short panels.

[68] The traditional implementation is to model the residual variance as having an i-level random effect, hence the terminology Correlated Random Effects given to this method. However, it is also possible to include family means of control variables and then estimate regular pooled logit or probit, as we will do.

we face extremely short panel lengths due to families commonly having only 2 children[69]; (2) his results apply only to the leading (order 1/T) bias term, but with very short panels the subsequent bias terms could still be relevant; and (3) there is an unresolved challenge of how to address the extrapolation from the estimation sample to singletons (when they are in the target population).[70]

Mundlak controls and naive-fixed-effects methods have the attractive properties of: (1) being easy to implement; (2) respecting the binary functional form of the left-hand-side (LHS) variable; and (3) straightforwardly obtaining ATEs. Nonetheless, empiricists' use of either of these options is uncommon; in our sample of 58 papers discussed in section 2 these methods are not used.

An additional complication with conditional or fixed effect logit and probit models is that they use less variation relative to LPM. With these models, for any families that have no variation in outcomes, i.e "all successes" or "all failures", the fixed effect parameters will be driven to +/- infinity, and these families will be dropped from estimation. This leaves only "double switchers": families with variation in both the outcome variable and the treatment variable. This means that moving from LPM to nonlinear specification is automatically tied to a change in estimation sample, which can reduce the effective sample size and may exacerbate the issues discussed in Section 3. In our application for example we see a reduction from 2986 individuals in the overall white "siblings sample" to 211 individuals in the "RHS switchers" sample to 98 individuals (from only 27 families) in the double switchers sample. A related issue is that the LPM results will depend on the fraction of observations in families that are not LHS switchers, whereas the logit model estimates will be invariant to the number of these non-switchers.

E.2 Obtaining Marginal Effects from Conditional Logit

In order to address the challenge of translating the conditional logit coefficient, β_{Logit}, into ATE units that can be compared with LPM results, we introduce a "two-step logit" model. The first step is the usual conditional logit estimator, used to obtain a consistent coefficient $\hat{\beta}$ for variables that change within-family. The second step estimates a random effects logit model (over the full sample, including non-switchers), while imposing the coefficient on the treatment variable (and on other individual-level variables) from the first step model. The purposes of the second step are (1) to estimate coefficients on family-level variables, so as (2) to assign an estimated "logit index" value to each observation, and (3) to estimate the variance of the family-level random effect σ_u^2. After the second step model is estimated, we then estimate the ATE using:

$$ATE_{2StepLogit} = \frac{1}{N}\sum_{i=1}^{N}\int_{u}(\hat{\beta}_{HeadStart} \cdot \Lambda(\hat{\beta}X_{if} + \hat{\gamma}Z_f + u) \cdot 1 - \Lambda(\hat{\beta}X_{if} + \hat{\gamma}Z_f + u)\phi(u)du \quad (30)$$

[69]We have reproduced the results for mean bias from his Table 4 for Probit and LPM-FS. We then reduced the panel size to T=2, and we find a detectible bias of -6.4% of the true ATE for the Probit, and no bias for the LPM-FS

[70]For singletons there is no ability to separately identify the value of the fixed effect from the idosyncratic error term. This is not a problem when the target population is either RHS switcher or all siblings. For these target populations, the naive logit FE or probit FE model could be used following the reweighting ideas presented above.

With $\hat{\beta}_{HeadStart}$ the coefficient on Head Start from the conditional logit first step; $\hat{\beta}$ the coefficient on i-level variables X_{if} from the conditional logit first step; $\hat{\gamma}$ the coefficients on family-level variables from the second step; and $\phi(u)$ the PDF from a normal distribution, with variance σ_u^2 estimated from the second step family-level random effects model. We have not yet found a prior implementation of this estimator in the literature; but it is similar in spirit to the two step fixed-effects logit proposed by Beck (2015).[71]

E.3 Selection in Nonlinear Models

A desirable feature of the two-step logit and the Mundlak controls models is that both allow the marginal effect of treatment, $\frac{\partial Pr(Y=1)}{\partial HeadStart}$, to vary across individuals. In both cases, the treatment effect depends on the "index value" for each individual. However, the models maintain an assumption of constant treatment effects in logit units (β_{Logit}). If the model is misspecified, and instead there are variable treatment effects for different individuals, and that a reweighted estimation sample might produce more reliable results, especially when trying to measure the ATE for a pre-specified target group. This consideration is analogous to the treatment effect heterogeneity discussed in section 4.2.

We propose employing the in-regression weights $\widetilde{s_f^{sw \to tg}} \cdot v_f$ as discussed above in section 4.2. That is, the weights are a combination of (1) propensity score weights derived from a multinomial logit model predicting "RHS switcher" status and "in target population" status, and (2) inverse within-family conditional variance of the treatement variable of interest. For expediency, we continue to estimate this conditional variance from a linear model, and to apply it directly to the second stage logit estimation step.

We explore some of these models in the context of our empirical example, and find some differences in the point estimates and precision across linear and nonlinear specifications. Compared with LPM, we find somewhat smaller and less precise impacts of Head Start on some college when we use the 2-step approach (point estimate: 0.086 (se: 0.059). We note that the slight decrease in precision here accompanies many fewer observations, which has fallen to 1200 for estimation of the logit beta instead of 2987 in the LPM.[72] The point estimate for the Mundlak controls is very similar to LPM, 0.126, but the standard errors are 20% larger (se: 0.053), so that the estimate is significant only at the 10% level.

E.4 Monte Carlo for Nonlinear Specifications

We next consider the bias of the different specifications in the context of a specific data generating process (DGP).

[71] Beck's second step is a logit FE (with dummies) estimator, with the β imposed from the conditional logit first stage. Then the estimated fixed effects are used to obtain the ATE.

[72] Note that in the second step, the ATE is calculated over the full population. Another difference is that we weight the conditional logit regressions using family averages of individual weights, since conditional logit does not accomodate individual weights.

For our simulations, we continue with the PSID data setup presented above in Section 4.4. We take the original data, and estimate a logit model predicting some college attainment, using as regressors family level variables and family-level averages of individual variables. From this model we construct a family level logit index variable, x_f. For each simulation, the underlying logit index for each individual is equal to x_f, plus the Head Start dummy multiplied by the Head Start (logit) treatment effect. We then turn this index into $Pr(y = 1)$ using the logistic CDF, and then randomly draw outcomes y. We consider three DGPs. The first of these has a constant treatment effect (in logit units). The second has a treatment effect that is zero for small families (with 2 or 3 children), and a larger treatment effect for families with 4 or more children. The third DGP has a variable treatment effect which is decreasing linearly in x_f. For all of the DGPs, the treatment effect in terms of $Pr\,(y = 1)$ will vary across target populations because different children have different logit indices. For DGPs 2 and 3, there is additional variability stemming from family characteristics.

We run 2,500 Monte Carlo replications. In each replication we estimate a basic LPM, and LPM reweighted for the target population. We also estimate our two-step logit model and a logit model with Mundlak controls. For each of these we estimate both an unweighted version and a version that is reweighted for the target population. We consider the same four target populations, and present the results in Table E.1. The first panel shows results for DGP 1, with constant (in logit units) treatment effects. For this DGP, all models perform well for target groups of switchers, with biases that are small and usually not distinguishable from zero. When we target siblings, all children, or Head Start participants, the LPM model exhibits a detectable bias, which is slightly reduced by reweighting. The proposed 2-step logit model and Mundlak model do better, with small bias. However when they are reweighted with an aim to be representative of the target population, they too have a detectable bias.

In DGP 2 we now have treatment effects that vary with family size. Here all of the basic models perform poorly, both LPM and our two logit variations. Reweighting helps dramatically here, for all three models.

For DGP 3 all models give biased results when we target all children or all siblings. The three reweighted models perform roughly equally well. Each of the specifications does well for estimating treatment effects for switchers, Head Start participants, and Head Start siblings, with small biases.

In results not reported, we also explored a naive logit fixed effects specification for target groups of RHS switchers and sbilings. For these groups, this method performs similarly to the LPM, 2-step Logit, and Mundlak logit discussed above.

E.5 Discussion of Specification Choices

In our literature sample, use of OLS/LPM methods is ubiquitous. Based on the results of this section, we recommend continued use of this method. For researchers who want to pursue a logit type specification, we believe that either the two-step logit model (based off of a conditional logit estimation first step) or a logit with Mundlak controls can perform well.

Table E.1: Monte Carlo Experiments: Bias of Linear and Nonlinear Models Relative to Target ATE, and Effectiveness of Reweighting

		LPM		Logit		Logit Reweight	
	True ATE	FE Baseline	Reweight	2-Step	Mundlak	2-Step	Mundlak
1. Constant TE							
Switchers	86.4	-0.5	-0.4	-0.3	-1.6*	-1.5	-1.8*
Siblings	78.8	7.0*	5.6*	2.0*	0.6	5.5*	4.3*
All	78.8	7.1*	5.8*	2.0*	0.7	5.7*	4.3*
HS Participants	88.1	-2.3*	-2.1*	1.0	-0.2	-2.8*	-1.0
2. Large Family TE							
Switchers	79.6	-10.2*	0.0	-11.5*	-10.8*	-2.5*	-0.9
Siblings	44.5	24.9*	2.6*	-9.1*	20.0*	1.1	2.0*
All	36.1	33.2*	1.6	0.5	28.3*	0.5	1.1
HS Participants	40.1	29.2*	-0.6	40.7*	30.9*	-1.7*	-0.1
3. TE linear in X_f							
Switchers	102.2	0.1	0.8	-1.5	-1.3	-2.3*	-1.1
Siblings	84.3	18.1*	9.3*	3.8*	10.4*	7.3*	7.9*
All	84.2	18.2*	9.5*	9.6*	10.5*	7.6*	8.0*
HS Participants	101.9	0.4	-0.2	2.6*	2.5*	-2.8*	0.8

Notes: This table shows the results from 2,500 Monte Carlo simulations for three different DGPs of some college attainment, presented separately in each panel of the table, and four different target populations, shown in each row of the panel. The true DGP is a logit model, and is discussed in Section E.4. The first panel shows results where Head Start has a constant treatment effect (TE) (on the logit index) for all individuals; the second shows results where Head Start (HS) has no effect on individuals from small families (3 or fewer children) and a large effect for families with many children (4 or more children); and the third panel shows results where effects are linear in X_f. Column 1, "True Beta," presents the true average increase in the probability of completing some college for participants in Head Start in the sample, which is a function of the DGP and sample composition. The remaining columns present the bias of various estimation strategies, defined as the difference between the estimated effects of Head Start and the true beta. Columns 2 and 3, LPM and LPM reweight, are defined as in Table 4. Columns 4 to 7 show the results from using the two step random effects estimator and Mundlak logit without and with propensity score weights, respectively. Reweighted estimates obtained using in-regression weighting, which accounts for the representativeness of switchers and the conditional variance of Head Start within families. All betas are multiplied by 1,000. * $p < .01$.

References

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics.* Princeton University Press.

Angrist, Joshua D., and Ivan Fernandez-Val. 2013. "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework." *Advances in Economics and Econometrics: Tenth World Congress*, , ed. Daron Acemoglu, Manuel Arellano and Eddie Dekel Vol. 3 of *Econometric Society Monographs*, 401 – 434. Cambridge University Press.

Aronow, Peter M., and Allison Carnegie. 2013. "Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable." *Political Analysis*, 21(4): 492–506.

Beck, Nathaniel. 2015. "Estimating Grouped Data Models with a Binary Dependent Variable and Fixed Effects: What are the Issues? Comments Prepared for Delivery at the Annual Meeting of the Society for Political Methodology." *mimeo.*

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

Carneiro, Pedro, and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy*, 6(4): 135–173.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies*, 47(1): 225–238.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics*, 1(3): 111–134.

Fernandez-Val, Ivan. 2009. "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models." *Journal of Econometrics*, 150(1): 71 – 85.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46(1): 69–85.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1995. "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data." *Journal of the American Statistical Association*, 90(429): 106–121.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT Press.