All In the Family: Assessing the External Validity of Family Fixed Effects Estimates and the Long Term Impact of Head Start

Douglas L. Miller[‡], Na'ama Shenhav[†], Michel Z. Grosz[*],

July 23, 2017

Abstract

Many papers use family fixed effects (FFE) to identify causal impacts of interventions in non-experimental settings. Empiricists in this literature are frequently concerned with the internal validity of estimates; in this paper we provide a detailed examination of aspects of the external validity of FFE with a binary treatment. The FFE design induces a selection into the identifying sample that varies systematically across families. Estimates frequently rely on identification from a small subset of the sibling population. We show that FFE is likely to produce a LATE that overweights larger families, departing from the overall average marginal treatment effect (AMTE) among siblings. Finally, we examine sensitivity to specification for binary outcomes. We show that OLS is unbiased, and provide a novel method for recovering AMTEs from Conditional Logit specifications. We apply these insights to examine the long-term effects of Head Start in the PSID, utilizing an expanded dataset of outcomes up to age 40 to update Garces, Thomas, and Currie (2002). Using FFE, we find that participation in Head Start increases the likelihood of completing some college by 12 ppts. We estimate this LATE to be 50% larger than the AMTE for siblings. We find no evidence of strong positive effects of Head Start on overall economic or physical well-being. We conclude that alternative research design strategies should be pursued to gain representative evidence about the long term impact of Head Start.

[‡] Department of Policy Analysis and Management, Cornell University and NBER
[†] Department of Economics, Dartmouth College.
[*] Department of Economics, University of California, Davis.

1. Introduction

Family Fixed Effects (FFE) are frequently used to obtain causal identification of an attribute, intervention, or policy, the ``treatment" of interest. This strategy identifies causal impacts by comparing siblings with different exposure to the treatment. The main advantage of this approach is that it allows researchers to eliminate family-level correlates of treatment, which may be a concern for estimation. However, since within-family decisions may still be endogenous, FFE are typically employed in estimation when stronger experimental or quasi-experimental research designs are not available[1].

This research design has been a key source of identification of the impact of Head Start, or federally funded preschool (Currie & Thomas, 1995; Garces, Thomas and Currie, 2002 (hereafter GTC); Deming 2009). Head Start has a budget of $8.6 billion dollars and enrolls over 900,000 children annually, the equivalent of about 60% of the number of 3 and 4 year old children in poverty (Carneiro & Ginja 2014).[2] A growing body of research credits Head Start with having positive impacts on economic and non-cognitive outcomes of participants measured in adulthood, giving credence to continued support of the program (see Gibbs et al., 2013 for a review). Nonetheless, there remains an ongoing debate regarding the effectiveness of the program, which raises the stakes for estimation of the program's impact.

The internal validity of FFE estimates form a focus of this empirical literature. Potential threats to this claim include measurement error, within-family selection into treatment, and sibling treatment

---

[1] This empirical strategy is pervasive in the Head Start literature, (Currie & Thomas 1995, 1999, Garces, Thomas, and Currie 2002, Deming 2009, Currie & Neidell 2007, Bauer and Schanzenbach (2016)). It has also used to estimate the impact of other public policies, including public housing assistance (Anderson et. al. 2016), the WIC program (Rossin-Slater 2013; Currie and Rajani 2014), as well as the causes and consequences of early-life health (Abrevaya 2006; Almond et. al. 2005; Black et. al. 2007; Figlio et. al. 2014; and many others).

[2] See Gibbs et. al. (2013) for an overview of the program, including programmatic details. Statistics on Head Start enrollment available at http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2014-hs-program-factsheet.html

spillovers[3]. Our focus in this paper is across-family selection into "identifying variation" when the treatment is binary, which has implications for the external validity of FFE.[4] Because this method requires within-family variation in treatment, the FFE design induces a selection (at the level of the family) into the identifying sample. We raise two implications of this for the number of identifying observations and the average treatment effect in the discussion that follows. The empirical literature to date also focuses on the OLS/Linear Probability Model. We examine OLS performance in comparison to Logit specifications.

The first implication of the FFE selection into identifying sample is a reduction in identifying variation. It is well known that in such a design the impact of the binary Head Start variable is identified by "switchers:" families in which some children attend Head Start while others do not. This necessarily reduces the amount of variation available, but how substantially this variation is reduced may be underappreciated. For example, in our sample of 5,355 children with siblings, only 1,098 children reside in "Head Start switching" households.[5] Further, we show that the loss of sample variation is systematically related to observables – it is lowest for families whose probability of treatment is closest to 0.5, and grows as treatment probabilities approach 0 or 1. There is also greater loss for smaller family sizes. This can have important differential selection effects within an overall sample. For example, in the PSID data from our application, we lose 93% of the sibling sample for white children, but only 62% of the sample for black children.

Second, we show that the family fixed effects design has a meaningful change in the types of families that identify treatment effects. The LATE obtained by this approach is based more heavily on

[3] See e.g. Bound and Solon (1999).
[4] While we focus on Family FE research designs, the lessons will also apply to other short-panel designs
[5] We are by no means the first to raise this point. Indeed, Currie and Thomas (1999) present a similar breakdown in their analysis of Head Start's impact on Hispanic Children. Footnote 7 in that paper suggests a reduction in sample size from 750 children to 222 in "Head Start switching families". Anderson et. al. (2010) also include the counts of children in switching families in their analysis.

larger families, due to the fact that "switchers" are much more likely to occur in families with several children. For example, in our setting individuals with 3 siblings make up just 11% of the overall sample, but they comprise nearly 18% of the "switchers" sample. We also find suggestive evidence that the effect of Head Start varies by the number of siblings in the family. Thus, the FFE treatment effect which is a weighted average of these LATEs is no longer representative of siblings, let alone the population of Head Start participants.

The intuition for the change in identifying variation is not specific to FFE; any estimator can shift relative weights across this heterogeneity (in addition to throwing out bad variation). To our knowledge, there is no discussion of the change in complier, ``switcher", characteristics for the FFE LATE, although a growing literature develops this idea in the IV/LATE context (Angrist and Fernandez Val 2014).[6] We suggest a combination of methods for addressing this. For transparency, the difference between FFE estimates and OLS should be decomposed to distinguish between the change resulting from deviations in the identifying complier population and the change in the identifying variation. Further, FFE treatment effects should be reweighted of estimates to account for differences in complier characteristics in FFE, similar to Angrist and Fernandez-Val (2014). We find that half of the difference between the ordinary least squares (OLS) and fixed effects regression may be attributable to this LATE among "switcher" families.

Third, we consider the sensitivity of FFE estimates to model choice when the outcome is binary (such as high school completion). Our analysis is motivated by the fact that the OLS linear probability model (LPM) is ubiquitous in the FFE empirical literature, despite the fact that this model uses more observations compared with nonlinear fixed effects models (such as Logit and Probit). This discrepancy arises because non-linear models only use observations with within-family variation in the outcome. As

[6] Loken et al (2012) discuss reweighting of in context of nonlinear treatment-response relationship

a starting point for the sensitivity analysis, we present a new method to recover AMTEs following a Conditional Logit estimation, using a second-step Random Effects Logit estimation. Monte Carlo simulations indicate that LPM models perform well in this setting. Simulations also show that logit models produce comparable AMTE results when either family-level covariates are a sufficient proxy for unobserved family characteristics within the logit model (Mundlak approach) or when our two-step estimation is used to obtain AMTE from the conditional logit model.

Based on these findings we propose new standards for practice when presenting results using similar research designs. We argue that best practice should employ the following steps: (1) clearly show not only total sample size, but additionally sample size when limited to "switching" families (and also for relevant subsamples within the data); (2) show how the family-size distribution compares between the "switching sample" and the overall sample; (3) if using a Logit or Probit model for binary outcome variables, additionally show the sample size based on "switching" families in both the dependent, left-hand-side (LHS) and independent, right-hand-side (RHS) variable; (4) consider showing a diagnostic graph along the lines of our Figure 2; (5) when facing binary outcome variables, show sensitivity to parametrization of the LHS model, such as LPM vs. our two-step procedure; (6) consider a "reweighting" of family-size OLS specific estimates to show the impact of the change in LATE when moving from OLS to FE models, along the lines of our Table 12.

In the second part of our paper, we present results on the long run impacts of Head Start using data from the PSID. Relative to prior evaluations of Head Start using FFE and the PSID, we use a sample three times as large in size, include longer run (up to age 40) outcomes, and expand the set of outcomes under consideration. Using the expanded sample, one of our principal findings echoes the results of prior studies: Head Start enrollment is associated with long-run improvements in college-going (GTC, Deming 2009). We estimate that Head Start leads to an 12 percentage point increase in the likelihood of attending some college for white children.[7] This effect is remarkably stable across cohorts, indicating that the estimate does not reflect the impact of a particular vintage of the program, and we find similar effects for both males and females. Using our reweighting approach, we estimate that the FFE is 50% larger than the AMTE for the representative sibling sample.

However, in contrast to previous findings, we find no significant improvements in high school completion or reductions in criminal activity. We also find no significant impacts on several summary indices of long run economic and health outcomes (results suggest negative point estimates), or on college completion. The effects are imprecisely measured, and we cannot rule out Head Start participation leading to moderate improvements in outcomes – up to 0.2 SD increase in our summary measure of economic sufficiency for whites. Nonetheless, the totality of the null results stand in contrast to the sizable effects on college attendance. We interpret our findings as pointing toward "increased uncertainty" rather than "zero effects" of the program: our findings should widen researchers' and policy makers' confidence intervals regarding Head Start's effectiveness, but not greatly change the location of center of their distribution of those beliefs.

Our study builds on a growing body of evidence that finds that the effects of Head Start are present beyond pre-school, despite the apparent impermanence of children's test score gains (Gibbs et. al. 2011)[8]. We also contribute to a broader literature using quasi-experimental methods to identify the long term effects of Head Start. Other research designs have exploited a supply shock during the early years of the program (Ludwig & Miller 2007, Thompson 2017), and regression discontinuity designs using income eligibility thresholds (Carneiro & Ginja 2014). These studies find improvements in childhood health and increases in educational attainment among earlier cohorts of participants, and

[7] Deming (2009) also finds increases in college attendance (6 ppts) among black participants.
[8] Test score "fade out" has been most convincingly demonstrated in the 2002 National Head Start Impact Study (US DHHS 2010), a large-scale randomized control trial of Head Start access (N=4,667 children).

reductions in behavioral problems, health problems, and obesity in later childhood and early adolescents for later cohorts of participants. We evaluate the effect of Head Start on longer-run outcomes relative to both of these studies through longitudinal tracking of individuals, which is an advantage over Ludwig & Miller (2007).

2. Survey of FFE Literature

As we have discussed, family fixed effects have been employed in several foundational papers in the Head Start evaluation literature. To analyze the prevalence of this strategy more broadly, we conduct a survey of publications from January 2000 to May 2017 in six leading journals that publish applied microeconomics articles.[9] We include all studies that use family fixed effects as a primary or secondary strategy.[10]

Our literature review yields 33 papers. We provide descriptive statistics of these articles in Table 1. The first panel tabulates the frequency of binary treatments and binary outcomes across the sample of papers, the focus of our methodological insights. These forms of variables appear frequently. Nearly two-thirds (20) of the papers have a binary treatment of interest and 13 additionally have a binary outcome. The second and third panels show the varied topics that appear in the sample, spanning health, public, education, and labor fields.

The final panel of the table summarizes the distribution of sample sizes used with FFE. The samples are frequently not limited to families with variation in the treatment variable; therefore, the sample size in the table is an upper bound on the number of observations used for identification. The median number of sibling observations is 3,990, or roughly 75% of the PSID sample in our analysis. It is

---

[9] We include the Quarterly Journal of Economics, American Economic Review (including Papers and Proceedings), AEJ: Applied, AEJ: Economic Policy, Journal of Labor Economics, Journal of Public Economics

[10] We use the following search terms on each journal's website to identify relevant papers: family, sibling, twin, mother, father, brother, sister, fixed effect, fixed effect, birthweight.

important to note that there is a high variance in sample size across samples, indicating that there is not a threshold for FFE analyses. The bottom 25% of papers have fewer than 600 observations, while the top 25% have over 200,000 sibling observations.

Figure 1 illustrates the salience of this estimation strategy over time. The top graph shows the number of articles published in each year. It shows increasing prevalence of studies using FFE from 1 to 3 per year during the early 2000's, a "boom" of 6 articles in 2009, and roughly 3 per year thereafter. The bottom graph shows the mean number of citations per article over time using data from Google Scholar.[11] On average a FFE article obtains 187 citations. GTC (2002), our only observation in that year, has 943 citations.

3. Family Fixed Effects: effective sample size, sample composition, and related issues

We now turn to the examination of three methodological issues that arise within the Family Fixed Effects research design. These include: (1) the dramatic reduction in identifying variation that results from the family fixed effects design; (2) a meaningful change in the types of families that are used to identify the effects of Head Start, shifting the focus toward larger families; and (3) sensitivity to model choice when family fixed effects are used together with a binary outcome variable. We also illustrate the issue in the context of a particular example based on micro data. This example is drawn from our analysis of the impacts of Head Start in the second part of the paper. In this section be provide an outline of the data for the example, with greater sample and estimation details presented in sections 3 and 4 below.

Our empirical example is based on a sample from the Panel Study of Income Dynamics. We use 2986 white children born in the years 1954-1987. The key treatment variable is a dummy variable for

[11] Google scholar citations are as of May 1, 2017.

ever having attended the Head Start programs; and the outcome variable of interest is a dummy variable for ever having attended some college. Important control variables include a dummy for other preschool attendance, and parental and early-childhood socioeconomics controls. The coefficient on Head Start in a cross-section regression is 0.049 (s.e. = 0.044). When mother fixed effects are added, the coefficient becomes 0.120 (s.e. = 0.053). This result indicates that the impact of Head Start participation on College attendance is meaningful in magnitude, and statistically significantly different from zero[12].

3.1 First issue: Fewer Observations in "Switching" Families

It is well-known that FFE identifies treatment effects from "switcher" families, the "complier" population. This implies that the ex-post effective number of observations—that is, those that contribute to identifying the treatment effect—may be quite small.

We demonstrate the reduction in identifying variation by creating a visualization of the within-family variation in Head Start in the PSID sample. In Figure 2, we show a scatterplot of the within-family deviation in Head Start, i.e $HeadStart_i - \overline{HeadStart_m}$, against the within-family deviation in attainment of some college for the white sample.[13] The size of each symbol is weighted by the number of students. A striking feature of the graph is the large mass at (0,0), indicating that many families have no variation in Head Start participation and no variation in the high school graduation of their children. Moreover, there are many additional families with no within-family deviation in Head Start, as illustrated by the vertical alignment of large bubbles. In Figure 3 we remove observations for families with no variation in

[12] Further details for these models can be found in section 5 and tables 3 and 10. Here we limit presentation of details, for the purpose of keeping the illustrative example as simple as possible.

[13] A value of 0.5 along the horizontal axis, for example, means that a person went to Head Start in a family where half the children attended Head Start. Values other than 0.5 and -0.5 are possible because not all families have just two children. Values of -0.5 and 0.5 are also possible in families with more than 2 children, if equal numbers of children participated as did not. A value of -0.75 means that a person did not go to Head Start in a family where three quarters of the children did.

Head Start to mimic the identifying variation used to estimate the effects of Head Start. The substantial reduction in the number of observations, from 4761 to 211, is visually apparent.

Importantly, the reduction in identifying observations may be unequal across subsamples of the data. To gain intuition about which characteristics might determine switching – and hence, influence the reduction in observations – we build a simple model of the Head Start participation decision within families. For now, we assume that the probability of attending Head Start is a constant, π, such that the likelihood of attending is a lottery within families. The probability of switching, is then a function of π and family size, X_f:

$$P(HS_{i,switcher} = 1) = 1 - P((\sum_f HS_i) = 0) - P((\sum_f HS_i) = X_f)$$

$$= 1 - (1-\pi)^{X_f} - \pi^{X_f}$$

Figure 4 graphs the relationship between $P(HS_{i,switcher} = 1)$ and π for families with 1, 2, 3, 4, or 5 children. It shows that the probability of switching has an inverse-U-shaped relationship with π, and that for a given level of π, the likelihood of being in a switching family is increasing with X. The markers on the graph show the observed probability of Head Start and likelihood of being in a switching family in the PSID. As in the stylized model, the observed likelihood of switching is increasing with X, although the difference in the likelihood of switching is much smaller than the model due to the fact that the probability of Head Start is not independent across children in the data. The observed likelihood of switching is also increasing with π, following the inverse-U.

The inverse-U relationship between π and switching implies that the reduction in observations will be larger for populations with very high and very low π, and smaller when π is close to 0.5. Figure 4 shows the observed probability of Head Start attendance is much higher and closer to 0.5 for black families relative to white families. Consistent with this, the switching probability is larger for black families. Therefore, FFE only uses 7% of the sibling sample for white children, but 38% of the sibling sample for black children.[14]

To address this issue, we recommend that researchers should take the following steps: clearly show not only total sample size, but also the sample size when limited to "switching" families (and also to do so for relevant subsamples within the data); and consider showing a diagnostic graph along the lines of our Figure 1.

3.2 Second issue: "LATE" Among Switchers

In addition to the smaller N, the FFE strategy induces a change in the population that contributes identifying variation, the "complier" characteristics. In the previous section we showed that switching is an increasing function of family size. Here, we examine the consequences of the family-size composition of compliers for estimates. Family-size-specific treatment effects are given by the coefficient on the interaction between Head Start and an indicator for family size, $\delta_{x,}$ which we assume is heterogeneous across X.

We further assume that δ_x is monotonically increasing in X. While this assumption is made in part for convenience, we argue that it is a reasonable extension of existing models of child investment and empirical evidence. In particular, the classic quantity-quality tradeoff of children implies that parental resources (time, money) per child is decreasing in X. If Head Start serves as a substitute for parental resources, then its effects should be largest for children from larger families. Second, empirically Head Start has been shown to have larger effects on children with higher endowments, who also receive greater parental investments in larger families (Aizer and Cunha, 2012). Therefore, if

[14] Note that we are focusing on race, but that this notion can be generalized to other family characteristic, such as SES, that determine π.

parental investments are complements to Head Start, then its effects may also be greater for children from larger families.

We now consider the implications of family-size-based switching for the coefficient obtained in the cross-section and FFE. In the cross-section, the OLS estimate for the population of siblings can be written as a weighted average of δ_x's (Angrist and Pischke, 2009)[15]:

$$\delta_{HS,OLS} = \frac{\sum_x \delta_x [P(HS_i = 1 | X_f = x)(1 - P(HS_i = 1 | X_f = x))] P(X_f = x)}{\sum_x [P(HS_i = 1 | X_f = x)(1 - P(HS_i = 1 | X_f = x))] P(X_f = x)} \qquad (1)$$

Now if we restrict OLS to families that will be included in the FFE model, then we obtain $\delta_{HS,OLS,FESample}$ as:

$$\delta_{HS,OLS,FESample} = \frac{\sum_x \delta_x (\sigma^2_{HS_i,switcher}) P(X_f = x | HS_{i,switcher} = 1)}{\sum_x (\sigma^2_{HS_i,switcher}) P(X_f = x | HS_{i,switcher} = 1)} \qquad (2)$$

where $\sigma^2_{HS_i,switcher} = P(HS_i = 1 | X_f = x, HS_{i,switcher} = 1)(1 - P(HS_i = 1 | X_f = x, HS_{i,switcher} = 1))$

Moving to the FFE estimate, we substitute $\delta_{x,FE}$ for δ_x:

$$\delta_{HS,FE} = \frac{\sum_x \delta_{x,FE} (\sigma^2_{HS_i,switcher}) P(X_f = x | HS_{i,switcher} = 1)}{\sum_x (\sigma^2_{HS_i,switcher}) P(X_f = x | HS_{i,switcher} = 1)}$$

The correlation between switching and family size can create a wedge between the estimated average effect of Head Start for the whole sample and the switching sample (without altering the family-

[15] See equation 3.3.7 on page 75 of Angrist & Pischke (2009).

specific estimate) in two ways. First, the share of families in families with more children is likely to be larger in the switching sample (the last term in the numerator), which would tend to upweight the estimated effect for those families. Second, and more ambiguous, it may affect the variance of Head Start (the first two terms of the numerator and the denominator), which could exacerbate or counteract this upweighting.

We now look to empirically assess the relevance of the change in sample composition from the family fixed effects design for our setting. The first two rows of Table 2 are akin to the markers in Figure 4, showing a monotonic relationship between the likelihood of switching and family size. In the next two rows of the table, we decompose the cause of this relationship into two factors: differences in the likelihood of having all children attend Head Start, and differences in the likelihood of having no children attend Head Start. They indicate that increased switching among families with three or more children is due in large part to being less likely to have no children attend Head Start.[16]

Since it is now apparent that the probability of switching correlated with family size, we turn to consider the implications for our regression coefficients. Columns 1-3 of Table 3, Panel A show how the probability of each of these family sizes changes going from the whole sample to the switching sample. Consistent with the above results, the proportion of 5+ child families in the switching sample is roughly twice the proportion in the overall sample, while the share of 3 and 4-child families is roughly constant. This will tend to upweight the coefficients of 5+-child families in the regression. Panel B of the table shows that that variance in Head Start is higher for every family size in the switching sample. The

[16] Appendix Figures B1 and B2 confirm and elaborate upon the relationship between switching and family size by separately looking at switching within families from Head Start to no Head Start and vice versa. Figure B1 shows that conditional on having the first child attend Head Start, subsequent Head Start decisions across children are quite similar across family sizes. Figure B2 instead looks at the likelihood of switching from no Head Start to Head Start across children. Conditional on not having a first or second child attend Head Start, families with 5 or more children are significantly more likely to have a third, fourth, or fifth child attend.

increase is relatively similar across family sizes, however, so the average effect is unlikely to be affected by this. In Panel C we calculate the ``A+P weights'' from Equations 1 and 2, which combine the inputs from Panels A and B. Going from the sibling sample to the switchers sample, the A+P weight for 2-child families declines by over 25% and the weight for 3-child families declines by 15%. On the other hand, the weight for 5+ families nearly doubles from .134 to .243, and the weight for 4-child families increases by over 25%.

We quantify the importance of the change in complier characteristics in two ways. Our first approach decomposes the change in the coefficient estimated from OLS and FFE into a change in identification and the change in identification.[17]

$$\delta_{FE} - \delta_{OLS} = \underbrace{(\delta_{FE} - \delta_{OLS,switchers})}_{\Delta Identification} + \underbrace{(\delta_{OLS,switchers} - \delta_{OLS})}_{\Delta LATE}$$

The second option is to reweight the FFE estimates to match the representative sibling population. This approach is similar to the "Late-Reweight" concept in Angrist and Fernandez-Val (2014).[18] The reweighted coefficient gives the estimated impact of Head Start for a broader population of Head Start participants under the assumption that the Head Start treatment effect does not vary within family size across switching and non-switching families. We implement this as:

$$\delta_{FE,sibs} = \frac{\sum_x \delta_{x,FE}(\sigma^2_{HS_i})P(X_f = x \mid HS_i = 1)}{\sum_x (\sigma^2_{HS_i})P(X_f = x \mid HS_i = 1)}$$

[17] Our decomposition is a special case of Equation 13 in Loken et al (2012), which provides a general formula for the comparison of OLS and FE estimators.
[18] See equation 9 of Angrist and Fernandez-Val (2014).

This issue has a meaningful impact on the coefficients we estimate our data example. For this sample, the OLS coefficient on Head Start is 0.049 (se=0.044), and the fixed effects coefficient is 0.120 (se=0.053). When we examine how the coefficient on Head Start varies by family size, we find different effects, with larger families showing greater returns. Taking the cross-section family-size-specific coefficients and re-weighting by the fixed-effects regression weights, we obtain a weighted coefficient of 0.069. So approximately 1/3 of the change from OLS to FE is driven by the change in LATE weights; with the other 2/3 driven by change in identifying variation. Reweighting the FE coefficients by the sibling weights produces a coefficient of 0.08. This implies that the LATE is 50% larger than the AMTE for the sibling population. We present this example in greater detail in Section 5 and Table 10 below.

Our recommendation to practitioners using family FE models is to show how the switcher sample compares to the overall sample in terms of the distribution of family size, and in terms of the distribution in terms of the regression weights. We also recommend that researchers also explore a re-weighting exercise along the lines presented above.

3.3 Third issue: Functional form specification for FFE with Binary Treatment and Binary Outcome Variable

The third issue we consider is potential sensitivity to functional form modeling assumptions. For binary LHS variables the usual choice of specifications includes Linear Probability Model, Logit, and Probit. In the cross-section setting, the conventional wisdom is that the choice among these options is fairly innocuous, especially when the objective is to recover the AMTE[19]. There are some complications in panel settings, especially when facing a short-panel such as found in the family fixed effects design.

3.3.1 Specification choices

Empiricists commonly employ an LPM specification to estimate family-FE models. We speculate that this is motivated by (1) the intuition (carried over from the cross-sectional case) that LPM models usually recover the AMTE[20]; (2) the benefit that the incidental parameters problem does not pollute the main parameters of interest[21]; (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 AMTE.

It is important to note that obtaining AMTE from a non-linear specification is not only less common, but also less straightforward[22]. In particular, the conditional logit model[23], which consistently estimates β by conditioning on the number of successes in a family, does not have a paired method for obtaining treatment effects. To obtain AMTE, Wooldridge (2010, section 15.8) recommends the inclusion of the Chamberlain-Mundlak controls, family-means of control variables, rather than directly controlling for fixed effects (Mundlak, 1978; Chamberlain, 1980).[24] We believe that empiricists' use of these options is uncommon; in our sample of 19 papers discussed in section 2.1 these methods are not

---

[19] Angrist and Pishke (2009, pg. 107), Wooldridge (2010, section 15.6). In contrast, Cameron and Trivedi (2005, pg. 471) recommend limiting LPMs to exploratory analysis.

[20] We are not aware of systematic exploration of whether the good properties in (1) easily transfer to the short-panel case. Textbook treatments generally state that things should be fine (Wooldridge, 2010, pg. 608; Cameron & Trivedi, 2005, pg. 471), but sometimes (as in Cameron & Trivedi) caution against LPM if the goal is to make predicted probabilities for individual observations.

[21] See Chamberlain (1980). Because this inconsistency is based on the panel length being fixed, the problem may be especially acute for short panels such as found in family fixed effects models.

[22] See Wooldridge (2010, section 15.8) for a discussion of non-linear specifications.

[23] Conditional logit models are also referred to as logit fixed effects. See Chamberlain (1980) for more detail.

[24] 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 the controls $\bar{X}_i$ and then estimate regular pooled Logit or Probit.

used. However, these methods have the attractive properties of: (1) being easy to implement; (2) respecting the LHS binary functional form; and (3) directly obtaining AMTEs.

An additional complication is that conditional logit models also use less variation relative to OLS LPM. With these models, for any families that are "all successes" or "all failures", the fixed effect parameters will be driven to +/- infinity, and these families will be dropped from estimation. We refer to these as “double switchers”: families with variation in both the outcome variable and the treatment variable. This means: (1) the estimation sample can become quite a bit smaller (exacerbating the issues discussed in section 3.1); and (2) a change in specification (e.g. from LPM to Logit) is automatically tied to a change in estimation sample (from "RHS switchers" to "RHS and LHS switchers").

In our application for example we see a reduction from 2986 individuals in the overall "siblings sample" to 211 individuals in the "RHS switchers" sample to 98 individuals (from only 27 families) in the "LHS and RHS switchers sample". This can be seen in the change from Figure 3 to Figure 5. The horizontal line of observations is lost when Logit FE models are used. Third, the OLS/LPM results will depend on the fraction of “LHS not switcher” observations, whereas the Logit model estimates will be invariant to the number of these non-switchers. Finally, the selection into “double switchers” may exacerbate the skewing of the variation toward larger families, exacerbating the LATE changes discussed in section 3.2.

For our simulations, we will consider estimation of LPM on the “double switcher” sample as well as 3 ad-hoc approaches for obtaining AMTE from the conditional logit model, and introduce a 2-step model as a new approach for obtaining AMTE’s based on conditional Logit estimation.

With a Logit β in hand, the MTE for an individual with logit-index z* is given by $MTE(z^*) = \beta \cdot \Lambda(z^*)\big(1 - \Lambda(z^*)\big)$, with $\Lambda(z^*)$ the Logistic function. Translation of this coefficient to AMTE requires “representative” z*, or alternatively a distribution of values of z* to integrate over. The first ad-hoc approach we consider is to evaluate the MTE at the average outcome in the Logit sample: $z^* = (\bar{y}_i; i \in Logit\ Sample)$[25]. The second ad-hoc approach recognizes that this ignores the non-LHS-switchers. Presumably these non-LHS switchers have relatively extreme values of the Logit index. The second approach assumes the MTE for these observations is 0, and “scales down” the estimate from the first approach. The third ad-hoc approach evaluates the MTE at the average outcome of the OLS/LPM sample: $z^* = (\bar{y}_i; i \in OLS/LPM\ Sample)$.

The final estimator we consider is one we describe as 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, imposing the coefficient on the treatment variable (and 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, then we estimate the AMTE using:

$$AMTE_{2\,step\,RE\,Logit} = \frac{1}{N}\sum_{i=1}^{N}\int_{u}\left(\hat{\beta}_{Head\,Start}\cdot\Lambda\big(\hat{\beta}X_{if}+\hat{\gamma}Z_f+u\big)\cdot\Big(1-\Lambda\big(\hat{\beta}X_{if}+\hat{\gamma}Z_f+u\big)\Big)\right)\phi(u)du$$

With $\hat{\beta}$ being the coefficients on within-family variables 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

[25] We compute standard errors for the marginal effects to force preservation of the t-statistics from the Conditional logit coefficients.

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)[26].

We explore some of these models in the context of our empirical example in Table 11. In Column (1), we show our baseline specification from the linear probability model for the white sample[27]. Then, in Column (3) we show the results from a logit regression, with calculated marginal effects at the mean. The logit marginal effect suggests that Head Start increases the likelihood of completing some college by 18 percentage points, although the standard errors have more than doubled relative to the LPM, such that the estimate is not statistically significantly different from 0. The imprecision here is possibly driven by the few observations, which is now 1200 instead of 2987 in the LPM. Of the 1200, only 98 individuals are in families that have variation in both Head Start and in the attainment of some college.

In order to separate the effects of the sample selection (into the "double switchers" sample) from the functional form assumptions, in Column (2) we re-estimate the LPM using the same set of observations from the logit model. Here, Head Start is estimated to increase the likelihood of completing some college by a marginally significant 17.1 percentage points. In this setting we find a curious result: adding the horizontal line of "0" data (column 1) leads the OLS/LPM model to be more likely to reject the hypothesis of no effect!

Column 4 presents results from "scaling down" the Column 3 estimate, assuming that the treatment effect for non-LHS switchers was 0. The 5th column estimates the treatment effect at the outcome average for the LPM sample. This does not differ much from Column 3. In the 6th column we

---

[26] Beck's second step is a Logit-Fixed-Effects (with dummies) estimator, with the β imposed from the conditional Logit first stage. Then the estimated fixed effects are used to obtain AMTEs.
[27] In this table we employ slightly different weights than in earlier results. We use family weights (which we construct as the within-family average of the individual weights) for all specifications here. This is because the conditional logit model requires constant weights within a family, and we want to keep the weighting scheme fixed across specifications.

present results from our Two-step random effects Logit model. The results here (0.087) are fairly similar to those of LPM (0.095), although the estimated standard error is somewhat larger and this model does not reject the hypothesis of no effect. Finally, we present results from a Logit with Mundlak-type controls. The point estimate (0.089) is again similar, and is significant at the 10% level.

3.3.2 Illustrative model: analytic results

The results in Table 11 do not directly give firm guidance as to which result is most believable. To explore this issue further, we first consider a simplified setting in which closed form solutions for some of these models are readily derived. We examine the case where all families (indexed by f=1,...,F) contain exactly two siblings, and one of the siblings is treated while the other is not[28]. We assume no covariates enter the model beyond treatment status. Then each family can be fully characterized by the pattern of outcomes among the siblings. We label families as "type 00" when $(y_{f,1}, y_{f,2}) = (0,0)$; "type 01" when $(y_{f,1}, y_{f,2}) = (0,1)$; and so on. Within this framework any dataset can be fully characterized by the number of families of each type: (#00, #01, #10, #11). Equivalently, a dataset can be characterized by the total number of families F and the share of families of each type: $(s_{00}, s_{01}, s_{10}, s_{11})$; with s_{00}=#00/F (and similarly for the other shares).

Appendix Table B1 presents formulas for the LPM and Conditional Logit coefficients, their corresponding estimated standard errors, estimated MTEs, and the test statistic for the null hypothesis of zero effect of treatment. We also include the formulas for the three ad-hoc AMTE for the conditional logit model.[29]

[28] We assume away considerations of birth order, and simplify further by labeling the first sibling as the untreated one, and the second sibling as the treated one: $(x_{f,1}, x_{f,2}) = (0,1)$.

[29] Note that the average $\overline{y_i}$ in the Logit estimation sample. In this case $\overline{y_{Logit}} = 0.5$ (because only LHS switchers are used) and so $\widehat{ATE}_{Logit,LHS-switch} = 0.25 \cdot \beta_{Logit}$.

The formulas in Table B1 are derived in the Appendix.[30] We use notation s_{01} to represent the share of the sample that is "type 01" (and similarly for the other shares), and $cs_{01} = \frac{s_{01}}{s_{01}+s_{10}}$ for the "conditional share" of "type 01" families among right-hand-side switching (type 01 and type 10) families. The table shows that for a given sample, the estimated AMTEs can be sensitive to specification choice. This is also the case for hypothesis testing.

Figure 6 further illustrates the differences in estimated treatment effects across specifications. The estimated MTEs for the LPM estimator are shown in bold red and the Conditional Logit "scaled down" estimator are in bold blue. This is based on data where s_{01}/s_{10}=2 for all cases, and with F=500 families. The x-axis varies the share of "non LHS switcher" families from 0 to 1. Both estimation models always have the same sign of estimated MTE, but the "scaled down" Conditional Logit always estimates a lower MTE than the LPM.

We also show the MTE's for alternative specifications. The top horizontal blue line shows the results from estimating the LPM model on the "LHS and RHS switcher sample", which will always give the estimate of $\beta_{LPM,LHS-switch} = \frac{2}{3} - \frac{1}{3} = 0.33$. The middle horizontal red line show the results from scaling the conditional logit by the y-bar in the LHS switcher sample: $\widehat{AMTE}_{Logit,LHS-switch} = 0.25 \cdot \beta_{Logit} = 0.25 \cdot ln\left(\frac{2}{3}/\frac{1}{3}\right) = 0.173$. Finally, the curved blue line shows the results from evaluating the conditional logit at the $\bar{y}_{LPM}$ average outcome of the LPM sample. This graph illustrates the fact that for any given sample, the estimated MTE can depend significantly on the estimation model used.

[30] Appendix construction is in progress.

3.3.3 Illustrative model: Monte Carlo simulation results

The previous subsection showed that for a particular sample, the estimated results can depend importantly on specification choice. We next consider the properties (bias, sampling variability, and accuracy of hypothesis testing) of the different specifications, in the context of a specific data generating process (DGP).

For our simulations, the DGPs we continue with the two-sibling family setup. We assign to each family an "index value" $x_f \sim N(\mu_x, \sigma_x^2)$. We also allow for an observable family variable, z_f, generated to have variance 1 and which may be correlated with x_f, $corr(x_f, z_f) = \rho$. For each sibling, we assign their individual index value to be $x_{i,f} = x_f - (\beta_{Head\,Start}^{true}/2) + \beta_{Head\,Start}^{true} \cdot 1(Head\ Start = 1)$. We set $\beta_{Head\,Start}^{true} = 0.2$. We then assign to each individual a probability for the outcome variable, $P(y = 1) = \Lambda(x_{if})$, with $\Lambda(\cdot)$ the Logit CDF function, and then assigning a random outcome $y_i \in \{0,1\}$ to that individual.

We conduct 24 different simulations, covering all combinations of the following parameters: $\mu_x \in \{-1.5, -1, -0.5, 0\}$; $\sigma_x \in \{0.5, 1\}$; and $\rho \in \{0, 0.75, 1\}$. In each simulation we have 400 families, and we run 2000 MC reps per model. Results are fairly similar across simulations, and so for brevity we present results for only one.

Table 4 presents the results of our simulation for one of the models, with $\mu_x = 0$, $\sigma_x = 1$, and $\rho = 0.75$. For each of the estimators considered, we compute the bias of the estimated AMTE, as well as the RMSE. The LPM estimator, applied to the overall sample, is one of the best performing estimators, with a zero bias and a low RMSE. However, estimating the LPM on the sample of "LHS switchers" produces upward-biased results (and a corresponding large RMSE).

The treatment effects derived from the conditional logit perform unevenly, and are inferior to the Mundlak logit model. Column 3 and 4 show that calculating the AMTE at the Logit sample mean outcome produces significant bias; and scaling down toward zero for the non-switchers results in an overcompensation and a negative bias. In this model, computing the AMTE based on the $\bar{y}$ of the full/LPM sample produces a modest bias (but still much larger than that of the LPM model in column 1). The two-step Random Effects model does the best of all of the models based on the Conditional Logit; but it still has a slightly larger bias than that of LPM. Finally, column 7 shows that the Mundlak-type logit model performs very well. Overall, this table indicates that the models that produce the least bias are the LPM and Mundlak-type Logit model. However, the two-step Random Effects model is a reasonable non-linear alternative to the Mundlak logit, and may be particularly useful if family-level covariates are limited.

3.3.4 Discussion of Specification Choices

In this section we have examined the role of functional form modeling assumptions in family fixed effects models. The most striking fact is that the specification choice (OLS/LPM versus Logit, for example) can dramatically change the estimation sample used in estimation across the two methods. Despite this difference, our Monte Carlo simulations indicate that these different specifications (and samples) can recover similar AMTEs. However, to obtain the AMTEs using the Conditional Logit model requires significant care in going from "Logit βs" to AMTEs. We present a new method to do so - a 2-step Conditional-Logit Random Effects-Logit model. This model performs adequately in our simulations, although not quite as well as OLS/LPM, or a Mundlak-type Logit.

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.

4. Effects of Head Start: Data

We now turn to examining the impact of Head Start on long run outcomes, using a sample of individuals surveyed in the Panel Study of Income Dynamics (PSID)[31]. The Panel Study of Income Dynamics began in 1968 as a survey of roughly 5,000 households and has followed the members of these founding households and their children longitudinally. The longitudinal nature of the study allows sibling comparisons during adolescence as well as later in life.

We begin our analysis with a sample constructed as in GTC. This sample includes all black or white individuals born between 1966 and 1977, and excludes Hispanic individuals. This sample is intended to be representative of the Head Start population during the early years of the program.[32] We provide a detailed description of our replication of GTC in Appendix A. Despite some minor differences, the two samples are qualitatively similar. The summary statistics are often within a third of a standard deviation of each other. Moreover, the estimated effects of Head Start in this sample are similar to those estimated in GTC. Consistent with that study, we find large and significant effects of Head Start on the probability that whites attain some college, and large point estimates for high school graduation, though in our case these are not statistically significant. However, we do not find a meaningful reduction in the probability of committing a crime resulting from participation in Head Start, and in fact in some subsamples find an effect in the opposite direction.[33] On the whole the replication corresponds well with the original study.

[31] In particular, we make use of the 2011 cross-year files, and the 1968 to 2011 waves of the PSID family interview files.
[32] As pointed out in Garces et al. (2002), the number of immigrants was much smaller between the years 1960-1980, such that it is unlikely that many Hispanic immigrants would have benefited from Head Start.
[33] However we believe these cases are driven by situations where there are rather few observations identifying the coefficients. As such we believe that this lack of correspondence may be driven by very minor (and un-diagnosable) differences in specification and/or dataset construction. We return to this topic in more detail in section 5.

For most of the analyses in this paper, we use a sample that substantially expands and modifies the GTC sample. The first modification is to include survey participants from the Survey of Economic Opportunity (SEO), which added an oversample of poor households. We discuss this decision in more detail in Appendix A. Second, we expand the sample to include individuals born between 1978 and 1987. The individuals in these cohorts were too young when the analysis in GTC was performed to observe their education and early career outcomes. Third, we include older siblings of all individuals, including those born prior to 1966. These early cohorts were too old to benefit from the introduction of Head Start, and serve as a plausible control group for the early cohorts.

In addition to modifications of the sample, we also expand the number of outcomes under analysis in order to gain a more extensive understanding of the channels by which Head Start affects children's lives. Recognizing that multiple testing of individual outcomes creates violations of standard inference techniques, we follow the established practice of distilling the measures to summary indices (see, e.g., Anderson 2008, Kling et al. 2007, Hoynes et al. 2012). This approach reduces the number of hypotheses tested, limiting concerns of multiple-hypothesis testing, and may increase the power of the analysis (Kling et al. 2007).

We create four indices to capture economic and health outcomes observed for individuals at age 30 and 40. The "economic sufficiency index" includes measures of educational attainment, receipt of AFDC/TANF, food stamps, mean earnings, mean family income relative to the poverty threshold, the fraction of years with positive earnings, the fraction of years that the individual did not report an unemployment spell, and homeownership. The "good health index" summarizes the following component measures: non-smoking, report of good health, and negative of mean BMI. See Table B.3 for descriptive statistics of the inputs to the indices.[34] Note that we intentionally include both positive and negative measures of well-being in each index in order to be able to capture both reductions in undesirable outcomes as well as absolute improvements in well-being.

The process of creating each index follows the procedure described in Kling et al. (2007). In particular, we standardize each component of the index by subtracting the mean outcome for non-treated children, defined as children that did not attend any form of pre-school, and then dividing the result by the standard deviation of the outcome for non-treated children.[35] The summary index takes a mean of these standardized measures. We also extract the first principal component of the standardized variables for "economic sufficiency" and for "good health". Later we use these as alternative outcome variables.[36]

Table 5 reports sample descriptive statistics for the expanded sample we construct. For ease of comparison with our earlier replication, we include means for the entire sample, the subsamples of Head Start participants/non-participants, and for the sample of individuals with siblings. We present the means of the analyzed outcomes in Table 6. In each of these tables, the number of observations varies for each of the reported means; for parsimony, in the table we only report the number of individuals in the sample for whom we have information on their attendance of Head Start[37]. A full accounting of the number of observations for each characteristic or outcome is available in Tables B.3, B.4, and B.5.

---

[34] The number of observations included in the table is the number of individuals with a response to the Head Start attendance question; the number of observations for each reported mean is available in Table B.4.

[35] Consistent with Kling et al. (2007), we generate a summary index for any individual for whom we observe a response for one component of the index. Missing components of the index are imputed as the mean of the outcome conditional on treatment status. For example, if a former Head Start participant is missing an outcome, it is imputed as the mean outcome of other Head Start participants. Likewise for other pre-school, or non-preschool participants.

[36] As alternative measures of well-being, we generate four additional indices using the same four variable partitions (economic outcomes at 30,40; health outcomes at 30,40) analyzed in a principal component analysis. The extra indices are created as a weighted average of the variables using the weights of the first principal component.

[37] The variable "ever booked or charged with a crime" was only collected in the 1995 wave, and so is only relevant for cohorts old enough to be at risk for that outcome by 1995.

5. Effects of Head Start: Empirical Strategy

The empirical strategy takes advantage of within-family variation in participation in Head Start to identify the long term impact of the program. Following GTC, we estimate:

(1) $$Y_{im} = \alpha + \beta_1 HeadStart_{im} + \beta_2 OtherPre\text{-}School_{im} + X_{im}\gamma + \delta_m + \varepsilon_{im}$$

where Y_{im} represents a long-term outcome for individual *i* with mother *m*. HeadStart indicates whether a child reports participation in the program, and OtherPre-School indicates participation in other Pre-school (and no participation in Head Start)[38]. The vector X_{im} includes a large number of controls for individual and family characteristics to absorb differences in personal and household characteristics which may be correlated with one's participation in Head Start and long term outcomes. These controls fall into three broad categories: demographics, including an individual's year of birth, sex, race, and an indicator for being low birth weight; family background, such as mother and father's years of education; and family economic circumstances during early childhood, including an indicator for having a single mother at age 4, 4-knot splines in annual family income for each age 0, 1, and 2, a fourth spline based on average family income between ages 3 and 6, indicators for mother's employment status at ages 0, 1, and 2, and household size at age 4.[39] δ_m is a mother fixed effect which enables comparisons across siblings with a shared mother. We estimate equation (1) using weights to make the sample representative of the national population,[40] and cluster standard errors on mother. When Y_{im} is a binary

[38] These two variables are in this way defined so as to be mutually exclusive, with "neither Head Start nor other pre-school" as the omitted category. Since Head Start only became available in 1965, we recode Head Start attendance to be "other preschool" for the 1961 and older cohorts.
[39] Missing control variables are imputed at the mean. We include an indicator variable for these imputed observations. This is a more expansive set of covariates relative to GTC, which did not include controls for maternal employment or family income prior to age 3.
[40] Following Garces et al. (2002), we generate representative population weights from the 1995 March CPS.

variable, we estimate linear probability models as a main specification and check the sensitivity of our results to alternative models.

The coefficient of interest is β_1, the impact of Head Start on long term outcomes compared to no preschool.[41] These coefficients take on a causal interpretation under the assumption that within-families, and conditional on other covariates, the child care decision across siblings is as good as random, and that the treatment effect does not spill over to siblings.[42]

The standard test of the identifying assumption is to look for balance in observables across siblings within families. Deming (2009) finds little evidence that Head Start attendance is correlated with observable differences across siblings, which suggests that the magnitude of selection may be small. In Table B.6, we examine the plausibility of the identifying assumption by testing the correlation between participation in Head Start and observable pre-Head Start individual and family characteristics. For the white sample, there are few statistically significant correlations, which suggest that the assumption may be reasonable. For the black sample, participation in Head Start is correlated with a greater likelihood of having higher income at age 1, and lower income at age 2. These correlations may raise concerns that black families may tend to send their children to Head Start after a rupture in the family or after an income shock, which may bias the estimated effects downward.[43] However, given the many hypotheses being tested in this table, we acknowledge the possibility that these significant findings might be spurious. Moreover, these results are somewhat sensitive, becoming insignificant when we drop observations with imputed controls. We are therefore uncertain how worrying these estimates are.

[41] For context we also present β_2 in our tables, but we do not discuss those results.
[42] See Bound and Solon (1999) for a more extensive discussion of this issue.
[43] We find a similar pattern of results when we restrict the sample to the early cohorts observed in Garces et al. (2002). See Table B1 for details.

6. Effects of Head Start: Results

We present the results for high school attainment, attending some college, log earnings, and crime for the full, expanded sample in Table 7. Column (1) presents results for the whole sample, column (2) for the subset of individuals for whom we observe siblings, column (3) adds mother fixed effects, and columns (4) and (5) maintain the same specification and stratify by race.[44] We focus on results in columns (3)-(5), which employ the mother fixed effects.

Overall, the results indicate that Head Start does not have a statistically significant impact on many of these long term variables. One notable exception is that we do find that attending Head Start leads to a 12 percentage point increase in the probability of attaining some college for white children. Though the effect size is half of that estimated with the earlier sample, this effect is sizeable and economically important. Participation in Head Start does not have a statistically significant impact on high school attainment, earnings between ages 23-25, or on the probability of not having committed a crime. Nonetheless, these results imply that Head Start shows significant long term effects on this measure of education. 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).

One limitation of these results is that many of the estimates are relatively noisy, such that the 95% confidence intervals allow for a sizeable positive impact of Head Start in spite of the small or negative point estimate. Therefore, we seek to increase the precision of estimation by pooling together a larger set of outcomes in a summary index. Table 8 presents the results for economic and health

[44] Control variables play an important role in absorbing observed heterogeneity, and therefore we include them in each of these specifications. This is a slightly different table set up than GTC or our replication of that paper, in which we omitted control variables in the first two columns.

indices measured at age 30 and at age 40.[45] The results suggest little support for a positive long term effect of Head Start.[46] In fact, the only statistically significant relationship is of non-Head Start pre-school on economic sufficiency at age 30. It bears emphasizing, though, that the results are not precisely estimated, which limits our ability to place much confidence in the point estimates. In Section 5, we explore potential explanations for the lack of precision in our estimation, and other limitations of our findings.

Motivated by the prior findings of differential effects by gender[47], in Tables 9 and 10 we look to see whether our mean results are obscuring this form of heterogeneity in our setting. Curiously, in Table 9, we find some evidence of negative effects of Head Start among men, in particular for health and economic outcomes at age 40. On the other hand, in Table 10, 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 outcomes than observed for men.

6.1 Treatment Effect variation, Family Size, and LATE

[45] We show a parallel table of results using the principal component outcomes in Appendix Table B.7. The findings are qualitatively similar for Head Start, and more positive for preschool.
[46] Appendix Tables B.8-B.11 include regressions for the components of the economic sufficiency index at ages 30 and 40 and of the good health index at ages 30 and 40. Consistent with the index, we find few statistically significant effects of Head Start for these components, with the exception of the attainment of some college for whites, and the fraction of the last 5 years with positive earnings at age 40 for blacks. It is worth highlighting that we also examine the effect of Head Start on completing college, and find no statistically significant effect.
[47] An additional source of potential heterogeneity in the results is the vintage of Head Start which a child attended. In Appendix Tables B.12 and B.13, we examine differential impacts of Head Start for “later cohorts,” defined as children born after 1977 (the median in the sample), and Appendix Table B.14 shows the results from an interaction with a linear trend in cohort, where the trend is normed to take on a 0 value for the 1966 cohort. We do not find any systematic differences in outcomes across cohorts.

In Table 12, we show that the effect of Head Start on completing some college is significantly higher among white children in families with 5 or more children.[48][49] One possible explanation for this heterogeneity is that children with higher initial endowments receive greater parental investments in larger families, and also benefit more from Head Start (Aizer and Cunha 2012). This implies that even without changing the specification from OLS to Fixed Effects, we would expect to have a larger estimated effect of Head Start in the within-family model due to the change in weighting.[50]

For this sample, the OLS coefficient on Head Start is 0.049 (se=0.044), and the fixed effects coefficient is 0.120 (se=0.053). The discussion above raises the question: how much of this change is due to “moving from bad variation (between families) to good variation (within families)”, and how much is due to changing the weighting across different effect sizes? To explore this question, we estimate both the OLS and the FE models allowing the treatment effect to vary by family size. We also compute the regression “implied weights” that link the family-size-specific coefficients to the basic overall coefficient. This allows us to perform counterfactual analyses in which we hold the family-size-specific coefficients fixed, but allow the composition of the sample to change. We do this for both the OLS specification with “All Weights”, the OLS weights for the whole sample, and the FE specification with “Switch Weights”, the weights for the FE sample. As a bridge we also compute this for the OLS specification on the sample of siblings excluding singletons using “Sib.Weights”.

In the bottom Panel of Table 12, we illustrate how the “average effect” of Head Start is affected by this LATE, using two weighting schemes. First, we calculate the implied regression weights for each

[48] This specification also includes indicators for having 1, 2, 3, 4, 5+, or unknown number of children.
[49] Table B.15 contains the equivalent results for the whole sample.
[50] To investigate whether the greater returns to Head Start among larger families were due to returns to other observables, we estimated an auxiliary model with interactions between Head Start treatment and family size, and additionally interactions between Head Start and a set of dummies for terciles of an index of SES. The results indicate that treatment effect heterogeneity is driven by family size instead of the other observables.

specification. These weights are shown in Panel columns (4)-(6) of Table B.16[51]. We verify that these weights work as intended, weighting the fixed effect coefficients by the weights for the switcher sample, obtaining 0.123 (very close to the FE estimate of 0.12). Using these weighted averages, we can quantify the extent to which the change in coefficients from the cross-section to the fixed effects specification is due to a change in the identifying variation or a change in the LATE. Specifically, in column 1 the coefficient increases from 0.046[52] to 0.069 simply by altering the weighting scheme to match the implied regression weights of the fixed effects sample. Then, moving from column 1 to column 2, the change in the identifying variation increases the coefficient further from 0.069 to 0.123. Thus, roughly one-third of the change in the coefficient is attributable to the change in the composition of the sample (LATE), while the other half results from using a cleaner source of identifying variation.

Table 12 also provides the FE estimate reweighted to be representative of the sibling population. It indicates that the Head Start leads to an 8.3 p.p. increase in some college attendance for children with siblings. This implies that the FFE estimate is roughly 50% too large.

Next, we examine whether switcher families differ from non-switcher families on other dimensions of observable characteristics, focusing on the sample of white families. Table B.17 indicates that in addition to having a larger family size, children in switcher families tend to have parents with significantly less education than children in non-switcher families (columns 4 and 5). These differences in parental education are significant even once we have accounted for family size (columns 6 and 7). We also see that the family income during pre-school of children in switcher families is significantly lower than non-switcher families overall (some of which may have too high of income to ever qualify for Head

---

[51] These vary slightly from Table 3 because they are only for the white sample.

[52] This weighted average of the cross-sectional coefficients (.041) differs slightly from the OLS result for this sample (.044) due to the inclusion of additional controls for family size in the former regressions.

Start), but significantly higher than children that attended Head Start in non-switcher families. Similar patterns are also present when we repeat this analysis for the whole sample in Table B.18.

In Table B.19 we hone our comparisons to ask whether *conditional on attending Head Start*, individuals from switcher families are observably different from non-switcher families. Now, the parental education levels of Head Start attendees from switching and non-switching families are statistically indistinguishable. However, switcher families still tend to have a significantly higher family income during pre-school ages and to have larger families (column 5). One possible interpretation of these results is that more disadvantaged families are less likely to be switchers because they are more likely to be consistently eligible for Head Start. This may raise questions about the external validity of estimates from the switcher sample.

7. Conclusion

Family Fixed Effects (FFE) are frequently used to obtain causal identification of an attribute, intervention, or policy, the ``treatment" of interest. In this paper, we present new results regarding across-family selection into identifying variation when the treatment is binary, which has implications for the external validity of FFE. Because this method requires within-family variation in treatment, the FFE design induces a selection (at the level of the family) into the identifying sample.

We show that the selection into identifying variation causes has a meaningful change in the types of families that identify treatment effects. First, the number of observations used for identification may be significantly smaller than the sample size typically reported. Further, the loss of sample variation is systematically related to observables – it is lowest for families whose probability of treatment is closest to 0.5, and grows as treatment probabilities approach 0 or 1. Second, we show that the LATE obtained by this approach is based more heavily on larger families, due to the fact that "switchers" are much more likely to occur in families with several children.

We also provide new results on the (in)sensitivity of FFE estimates to model choice when the outcome is binary (such as high school completion). As a starting point for the sensitivity analysis, we present a new method to recover AMTEs following a Conditional Logit estimation, using a second-step Random Effects Logit estimation. We compare the performance of alternative specifications in the context of our empirical example as well as Monte Carlo simulations. The resulting preferred estimators are OLS, Logit with Mundlak-type controls, and our new two-step estimator.

The data structure and research design (family fixed effects) are not limited to the PSID and Head Start. When researchers are employing a family fixed effects design, we argue that best practice should employ the following steps: (1) clearly show not only total sample size, but additionally sample size when limited to "switching" families (and also for relevant subsamples within the data); (2) show how the family-size distribution compares between the "switching sample" and the overall sample; (3) if using a Logit or Probit model for binary outcome variables, additionally show the sample size based on "switching" families in both the dependent, left-hand-side (LHS) and independent, right-hand-side (RHS) variable; (4) consider showing a diagnostic graph along the lines of our Figure 2; (5) when facing binary outcome variables, show sensitivity to parametrization of the LHS model, such as LPM vs. our two-step procedure; (6) consider a "reweighting" of family-size OLS specific estimates to show the impact of the change in LATE when moving from OLS to FE models, along the lines of our Table 12.

In our application, we present results on the long run impacts of Head Start using data from the PSID. Relative to prior evaluations of Head Start using FFE and the PSID, we use a sample three times as large in size, include longer run (up to age 40) outcomes, and expand the set of outcomes under consideration. Using the expanded sample, we estimate that Head Start leads to a 12 percentage point increase in the likelihood of attending some college for white children. Using our reweighting approach, we estimate that the FFE is 50% larger than the AMTE for the representative sibling sample. However, in contrast to previous findings, we find no significant improvements in high school completion or reductions in criminal activity. We also find no significant impacts on several summary indices of long run economic and health outcomes (results suggest negative point estimates), or on college completion. We interpret our findings as pointing toward "increased uncertainty" rather than "zero effects" of the program: our findings should widen researchers' and policy makers' confidence intervals regarding Head Start's effectiveness, but not greatly change the location of center of their distribution of those beliefs.

References

Abrevaya, J. (2006): "Estimating the Effect of Smoking on Birth Outcomes using a Matched Panel Data Approach*," Journal of Applied Econometrics*, 21(4): 489-519.

Aizer, Anna, and Flávio Cunha (2012): "The Production of Human Capital: Endowments, Investments and Fertility." Working Paper. National Bureau of Economic Research. http://www.nber.org/papers/w18429.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee (2005): "The Costs of Low Birth Weight," *The Quarterly Journal of Economics,* 120(3): 1031-1083.

Andersson, Fredrik, John C. Haltiwanger, Mark J. Kutzbach, Giordano E. Palloni, Henry O. Pollakowski, and Daniel H. Weinberg (2016): "Childhood Housing and Adult Earnings: A Between-Siblings Analysis of Housing Vouchers and Public Housing." Working Paper. National Bureau of Economic Research. http://www.nber.org/papers/w22721.

Anderson, Kathryn H., James E. Foster, and David E. Frisvold (2010): "Investing in Health: The Long Term Impact of Head Start on Smoking," *Economic Inquiry*, 48(3): 587-602.

Anderson, Michael L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481-1495.

Angrist, Joshua D., and Iván Fernández-Val. "ExtrapoLATE-Ing: External Validity and Overidentification in the LATE Framework." Chapter. In *Advances in Economics and Econometrics: Tenth World Congress*, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 3:401–34. Econometric Society Monographs. Cambridge: Cambridge University Press, 2013. doi:10.1017/CBO9781139060035.012.

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

Bauer, Lauren, Diane W. Schanzenbach (2016): "The Long-Term Impact of the Head Start Program" Hamilton Project." Accessed October 26, 2016. http://www.hamiltonproject.org/papers/the_long_term_impacts_of_head_start.

Nathanial Beck, 2015, mimeo. "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, University of Rochester, July, 2015. Version of July 17, 2015.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2007): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *The Quarterly Journal of Economics*, 122(1): 409-439.

Bound, John, and Gary Solon (1999): "Double trouble: on the value of twins-based estimation of the return to schooling," *Economics of Education Review*, Vol. 18, Issue 2, pp. 169-182.

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

Chamberlain, Gary, 1980. "Analysis of Covariance with Qualitative Data," The Review of Economic Studies, Vol. 47, No. 1, pp. 225-238.

Currie, Janet and Matthew Neidell (2007) "Getting Inside the `Black Box' of Head Start Quality: What Matters and What Doesn't," *Economics of Education Review*, 26(1): 83-99.

Currie, Janet, and Ishita Rajani (2014). "Within-Mother Estimates of the Effects of WIC on Birth Outcomes in New York City." Working Paper. National Bureau of Economic Research.

Currie, Janet and Duncan Thomas (1995) "Does Head Start Make a Difference?" *American Economic Review*. 85(3): 341-364.

Currie J, Thomas D. (1999) "Does Head Start help Hispanic Children?" *Journal of Public Economics*, 74(2):235-262.

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

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth (2014) "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review*. 104(12): 3921-3955.

Garces, Eliana, Duncan Thomas, and Janet Currie (2002): "Longer-Term Effects of Head Start," *American Economic Review*, 92, 999-1012.

Gibbs, Chloe, Jens Ludwig, and Douglas L. Miller (2013): "Head Start Origins and Impacts," in *Legacies of the War of Poverty*, ed. by M. J. Bailey and S. Danziger, Russell Sage Foundation.

Hoynes, Hilary W., Diane W. Schanzenbach, and Doug Almond (2012): "Long Run Impacts of Childhood Access to the Safety Net," Working Paper 18535, National Bureau of Economic Research.

Johnson, Rucker and C. Kirabo Jackson (2017): "Reducing Inequality Through Dynamic Complementarity: Evidence from Head Start and Public School Spending," NBER Working Paper 23489. June 2017.

Klein, Joe (2011) "Time to ax public programs that don't yield results." *Time Magazine*. July 7, 2011.

Kline, Patrick and Walters, Christopher (forthcoming): "Evaluating Public Programs with Close Substitutes: The Case of Head Start," Quarterly Journal of Economics.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," Econometrica, 75, 83-119.

Løken, Katrine V., Magne Mogstad, and Matthew Wiswall. "What Linear Estimators Miss: The Effects of Family Income on Child Outcomes." *American Economic Journal: Applied Economics* 4, no. 2 (April 2012): 1–35. doi:10.1257/app.4.2.1.

Ludwig, Jens and Douglas L. Miller (2007) "Does Head Start Improve Children's Life Chances? Evidence from a Regression-Discontinuity Design." *Quarterly Journal of Economics*. 122(1): 159-208.

Mundlak, Yair (1978) "On the Pooling of Time Series and Cross Section Data," *Econometrica*, Vol. 46, No. 1, pp. 69-85.

Rossin-Slater, Maya (2013). "WIC in Your Neighborhood: New Evidence on the Impacts of Geographic Access to Clinics." *Journal of Public Economics* 102: 51–69.

Thompson, Owen (2017). "Head Start's Impact in the Very Long Run," forthcoming, *Journal of Human Resources*.

U.S. Department of Health and Human Services, Administration for Children and Families. 2010. "Head Start Impact Study: Final Report." Washington, DC. https://www.acf.hhs.gov/opre/resource/head-start-impact-study-final-report

Wooldridge, Jeffrey M. (2010) *Econometric Analysis of Cross Section and Panel Data*, Second edition, The MIT Press.

Appendix A: Replication of Garces, Thomas and Currie (2002).

A.1. Summary

In this appendix we describe the results of our replication of Garces, Thomas and Currie (2002) (GTC), which comprises the starting point for this project. An explanation of our replication strategy is in the next appendix.

Table A.1 below shows the summary statistics for our preferred specification, which can be compared to Table 1 of GTC, included as Table A.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 use 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 main results are our replication of GTC's Table 2. Note that in each of the regressions we cluster along the family identifier in column 1, as opposed to by common mother, because not all observations have a mother identified. We cluster on mother for the rest of the columns. Table A.3 below shows the regression results.

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 significant. Our results for the college outcomes are aligned with the findings in GTC. The magnitudes that we report are not statistically different than 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[53], although our point estimates are consistently negative for blacks. Otherwise, our estimates are quite imprecise and not statistically different than GTC's.

Our earnings results (panel C replication) are quite different than 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.

A.2. Replication Methodology

Note: things to check are highlighted. There are some smaller tables here, too, so we've just included snapshots of them.

---

[53] To be consistent with the later analyses, we analyze effects on the likelihood of *not* having committed a crime, and therefore should be compared with the coefficients in Panel D of GTC multiplied by negative one.

The purpose of this appendix is to document the process of replicating Garces, Thomas and Currie (2002) (GTC) for future scholars wishing to repeat our steps and as a jumping off point for this work exploring the long term effects of Head Start. This appendix describes 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 further information about the mechanics of downloading the data and the variables we use.

A.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 (``crossyear 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 crossyear 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.

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 footnotes 4 and 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 A.5.

We also exclude the Latino oversample in accordance with GTC's footnote 4.[54]

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 A.6.

[54] GTC footnote 4 states that ``we have excluded Hispanics from this study".

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.[55] 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.[56] 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.

A.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 A.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.[57]

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.

[55] See GTC footnote 4.
[56] GTC describe in their footnote 4 that the weights are ``constructed so that the joint distribution of race, sex, and year of birth in our sample matches the joint distribution in the 1995 CPS."
[57] See GTC footnote 15.

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 than 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 A.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.

Final dataset restrictions

Given our iteration exercises, our preferred sample definition includes the SEO poverty oversample, 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.

More on the data

We downloaded the data files from http://simba.isr.umich.edu/Zips/ZipMain.aspx. Table A.7. shows the variables we downloaded:

I Tables

Table 1: Summary of Family FE Articles

	Binary Indep.	Binary Dep.	Both Binary	Total
AEJ: Applied	7	5	4	9
AEJ: Economic Policy	1	1	1	1
AER	3	1	1	5
AER Papers and Proceedings	2	2	1	3
Journal of Labor Economics	2	1	1	5
Journal of Public Economics	4	5	4	6
QJE	1	4	1	4
Total	20	19	13	33
Common Dependent Variables				
Schooling/Attainment	12			
Earnings	10			
Test Score	8			
Birth Weight	6			
Behavioral Issues/Crime	4			
Height/BMI	3			
Common Independent Variables				
Birth Weight	4			
Pre-School	3			
Means-Tested Public Program	2			
Death of Family Member	2			
Bombing/Radiation	2			
Employment/ Employment conditions	2			
Observations by Sample				
	Siblings N	Total N		
p10	428	1,212		
p25	619	3,255		
p50	3,990	17,501		
p75	217,412	405,802		
p90	1,095,863	1,582,142		
Year Publication Min/Max	2002	2017		

Table 2: Probability of Head Start (Any, All, None) by Family Size

	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
Observations	2003				

Table 3: Composition of Sample and Weights in Regression Estimates Across Sibling, Switcher Samples

	1	2	3	4	5 +
A. Shares					
All Sample	0.123	0.273	0.238	0.147	0.134
Siblings Sample	0.000	0.345	0.300	0.186	0.169
Switchers Sample	0.000	0.210	0.271	0.197	0.322
B. Variance					
All Sample	0.089	0.104	0.121	0.127	0.132
Siblings Sample	0.000	0.024	0.050	0.059	0.068
Switchers Sample	0.000	0.085	0.166	0.197	0.223
C. A+P weights					
All Sample	0.171	0.257	0.284	0.117	0.101
Siblings Sample	0.000	0.338	0.374	0.154	0.134
Switchers Sample	0.000	0.246	0.315	0.197	0.243
Observations	7372	7372	7372	7372	7372

Table 4: Monte Carlo Results

	LPM		Condl. Logit AMTE				Mundlak-Type
	LPM Sample	Logit Sample	At Logit Y-bar	Scaled Down	At LPM Y-bar	Two-Step RE	Margins
Bias	0.0001	0.059	0.093	-0.020	0.009	0.002	0.00003
RMSE	0.032	0.097	0.040	0.036	0.040	0.033	0.032

Table 5: Summary Statistics, New Sample Characteristics

	All	Head Start	No Head Start	Sibling Sample
Head Start	0.076	1.000	0.000	0.073
Other preschool	0.282	0.000	0.305	0.259
Fraction African-American	0.150	0.618	0.111	0.154
Fraction female	0.504	0.548	0.501	0.501
Fraction low birth weight	0.060	0.114	0.056	0.061
Had a single mother at age 4	0.112	0.296	0.091	0.103
Fraction whose mother completed hs	0.717	0.632	0.724	0.689
Fraction whose father completed hs	0.683	0.557	0.692	0.654
Fraction eldest child in family	0.368	0.341	0.371	0.339
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	7363	1345	6018	5355

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 6: Summary Statistics, New Sample Outcomes

	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	14391	20817.636	20633
	(24927)	(12000)	(25517)	(26547)
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	7363	1345	6018	5355

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 7: Base Regressions, New Sample

	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: 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. Regressions incorporate full set of controls from the preferred specification, which are more extensive than those included in Garces et al. (2002).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. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table 8: Regression: Good Health Index, Economic Sufficiency Index

	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: 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. 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. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table 9: Regression: Interaction with Female (Adult Outcomes)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Economic Sufficiency Index, age 30					
Males in Headstart	-0.158***	-0.124**	-0.069	-0.197**	0.078
	(0.055)	(0.063)	(0.079)	(0.090)	(0.141)
Females in Headstart	-0.138**	-0.112*	-0.106	-0.052	-0.099
	(0.057)	(0.068)	(0.074)	(0.099)	(0.112)
P-Value of Difference	0.779	0.892	0.662	0.148	0.252
Mean Y	0.094	0.096	0.096	-0.552	0.213
Observations	7372	5361	5361	2369	2986
Economic Sufficiency Index, age 40					
Males in Headstart	-0.190**	-0.179*	-0.142	-0.363**	-0.271
	(0.084)	(0.097)	(0.127)	(0.164)	(0.184)
Females in Headstart	0.008	0.018	-0.001	-0.021	0.058
	(0.082)	(0.098)	(0.117)	(0.173)	(0.140)
P-Value of Difference	0.058	0.113	0.317	0.098	0.099
Mean Y	0.020	0.025	0.025	-0.670	0.142
Observations	4085	2845	2503	1065	1435
Good Health Index, Age 30					
Males in Headstart	-0.386***	-0.310***	-0.204	-0.004	-0.361
	(0.102)	(0.091)	(0.209)	(0.218)	(0.378)
Females in Headstart	-0.324***	-0.327***	-0.114	0.042	-0.198
	(0.066)	(0.082)	(0.151)	(0.159)	(0.278)
P-Value of Difference	0.599	0.878	0.665	0.838	0.690
Mean Y	0.004	0.017	0.017	-0.357	0.074
Observations	4749	3600	3114	1150	1959
Good Health Index, Age 40					
Males in Headstart	-0.324*	-0.406*	-0.513	-0.672**	-1.099**
	(0.188)	(0.223)	(0.318)	(0.271)	(0.480)
Females in Headstart	-0.140	-0.050	0.048	0.349	0.605
	(0.134)	(0.159)	(0.225)	(0.273)	(0.378)
P-Value of Difference	0.397	0.171	0.130	0.014	0.004
Mean Y	0.011	0.015	0.015	-0.290	0.062
Observations	2228	1673	1306	511	795

Notes: 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. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table 10: Regression: Interaction with Female (GTC Outcomes)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
High School					
Males in Headstart	-0.010	-0.016	-0.039	-0.062	-0.043
	(0.027)	(0.033)	(0.032)	(0.042)	(0.054)
Females in Headstart	0.021	0.010	0.010	0.008	0.005
	(0.018)	(0.022)	(0.030)	(0.033)	(0.059)
P-Value of Difference	0.279	0.448	0.180	0.092	0.497
Mean Y	0.913	0.912	0.912	0.862	0.921
Observations	7372	5361	5361	2369	2986
Some College					
Males in Headstart	0.024	0.010	0.053	-0.021	0.145***
	(0.031)	(0.038)	(0.037)	(0.045)	(0.053)
Females in Headstart	0.050	0.062*	0.042	-0.012	0.102
	(0.031)	(0.037)	(0.042)	(0.044)	(0.074)
P-Value of Difference	0.503	0.269	0.801	0.873	0.582
Mean Y	0.531	0.532	0.532	0.396	0.556
Observations	7372	5361	5361	2369	2986
Ln Earnings 23-25					
Males in Headstart	-0.148*	-0.128	-0.180	-0.238	0.078
	(0.080)	(0.091)	(0.144)	(0.202)	(0.174)
Females in Headstart	0.161**	0.136	0.207	0.265	0.133
	(0.075)	(0.088)	(0.142)	(0.171)	(0.217)
P-Value of Difference	0.004	0.031	0.039	0.037	0.834
Mean Y	9.588	9.578	9.578	9.207	9.630
Observations	4351	3309	2726	986	1736
No Crime					
Males in Headstart	-0.028	-0.046	-0.025	0.016	-0.112
	(0.042)	(0.051)	(0.048)	(0.041)	(0.089)
Females in Headstart	0.012	0.017	0.003	0.038	-0.036
	(0.019)	(0.022)	(0.037)	(0.035)	(0.073)
P-Value of Difference	0.328	0.190	0.587	0.661	0.448
Mean Y	0.899	0.898	0.898	0.897	0.898
Observations	5005	3591	3206	1366	1836

Notes: 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. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table 11: Regression: Logit Some College, GMS White Sample

	LPM		Condl. Logit AMTE				Mundlak-Type
	LPM Sample	Logit Sample	At Logit Y-bar	Scaled Down	At LPM Y-bar	Two-Step RE	Margins
Head Start	0.095**	0.171	0.180	0.083	0.172	0.086	0.089*
	(0.048)	(0.111)	(0.117)	(0.054)	(0.112)	(0.059)	(0.046)
Observations	2986	1200	1200	1200	1200	1200	2989

Notes: 211 white individuals are in families that have variation in the Head Start variable, among those for whom we observe completed education; Of those, 98 also come from families that have variation in the attainment of some college. White individuals only. See text for calculation of marginal effects. Weighted using family mean of individual weights. SE clustered at 1968 family id for non-family fixed effect specifications and at mother id level otherwise. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

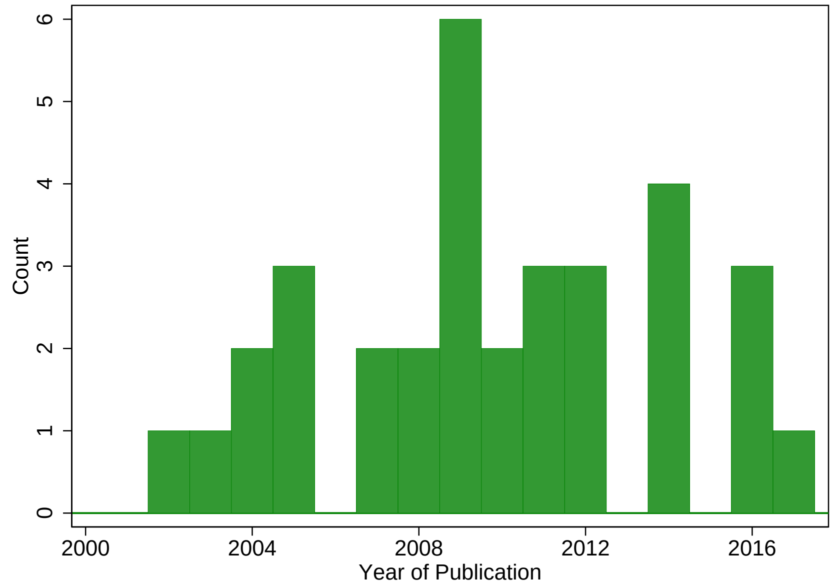
Table 12: Cross-Sectional Regression Coefficients (Some College, White) - Interaction with Number of Children in Family

	CX	FE
Head Start x 1 child family	0.169*	
	(0.091)	
Head Start x 2 child family	0.038	-0.126
	(0.079)	(0.099)
Head Start x 3 child family	-0.030	0.152**
	(0.087)	(0.075)
Head Start x 4 child family	-0.053	0.251***
	(0.100)	(0.091)
Head Start x 5+ child family	0.572***	0.348***
	(0.119)	(0.126)
Head Start x Unknown child family	-0.099	
	(0.108)	
Observations	4258	2986
N Non-Switch/Switch		
A+P All Weights	0.046	
A+P Sib. Weights	0.037	0.083
A+P Switch Weights	0.069	0.123

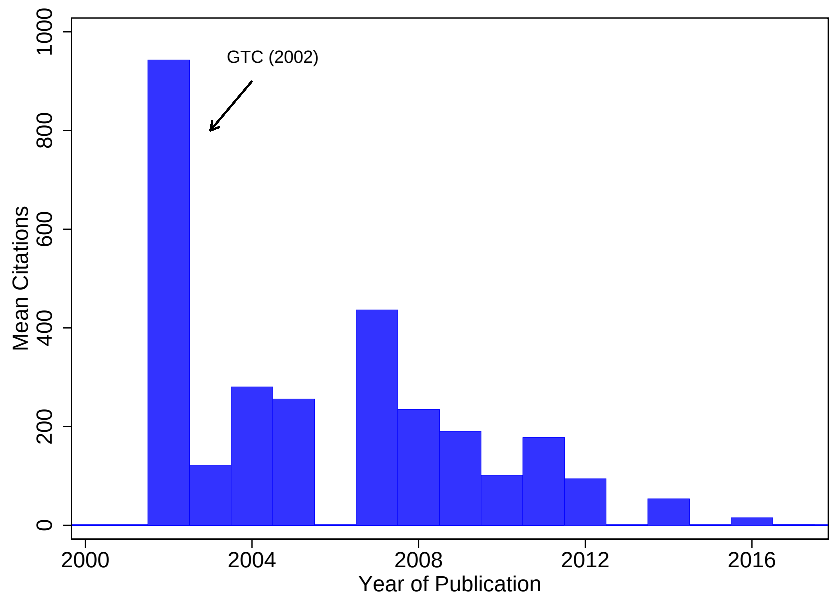
Notes: Columns 3 and 4 show the coefficients from one regression that interacts an indicator for Head Start with the number of children in the family and whether the family have variation in Head Start attendance. Columns 1, 3, and 4 include controls, but not mother f.e., and SE are clustered at 1968 family id. Column 2 includes mother fixed effects, and SE clustered by mother id. The bottom rows of columns 1 and 2 show the weighted average of the coefficients and the respective standard errors when using weights determined by the overall distribution of families, the distribution of 2+ child families, and the distribution of 2+ child families that have variation in Head Start attendance. The coefficient on Head Start from a cross-sectional regression restricting to the white sample is 0.049 (se: 0.044) (0.045 (se: .044) when dummies for child size are added). When the coefficients on child-interacted Head Start obtained in column 1 are weighted by the implied regression weights for all/sibling/switcher families, the weighted average coefficient changes from 0.046 to 0.037 to 0.069. By comparison, the weighted average of the column 2 coefficients using the implied regression weights for switcher families is 0.123; and the coefficient on Head Start from the family f.e. regression is 0.120 (se: 0.053). See text for details. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

II Figures

Figure 1: Number of Family Fixed Effects Articles, Citations by Year of Publication

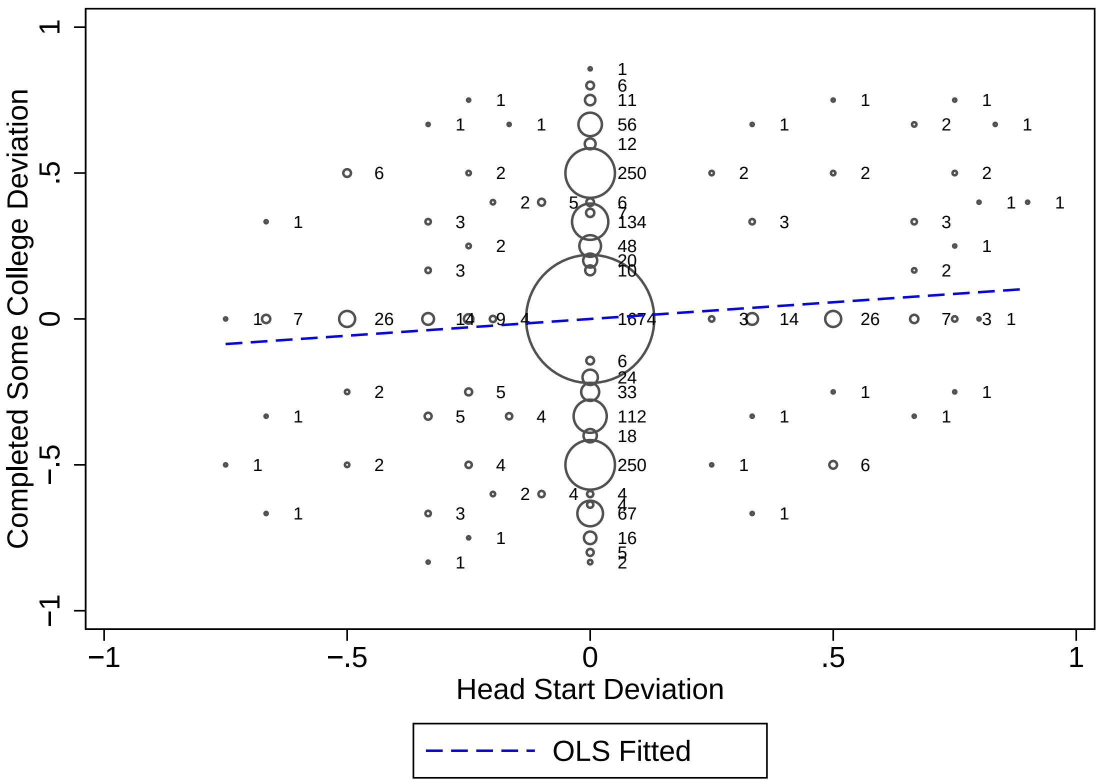


(a) Publications by Year



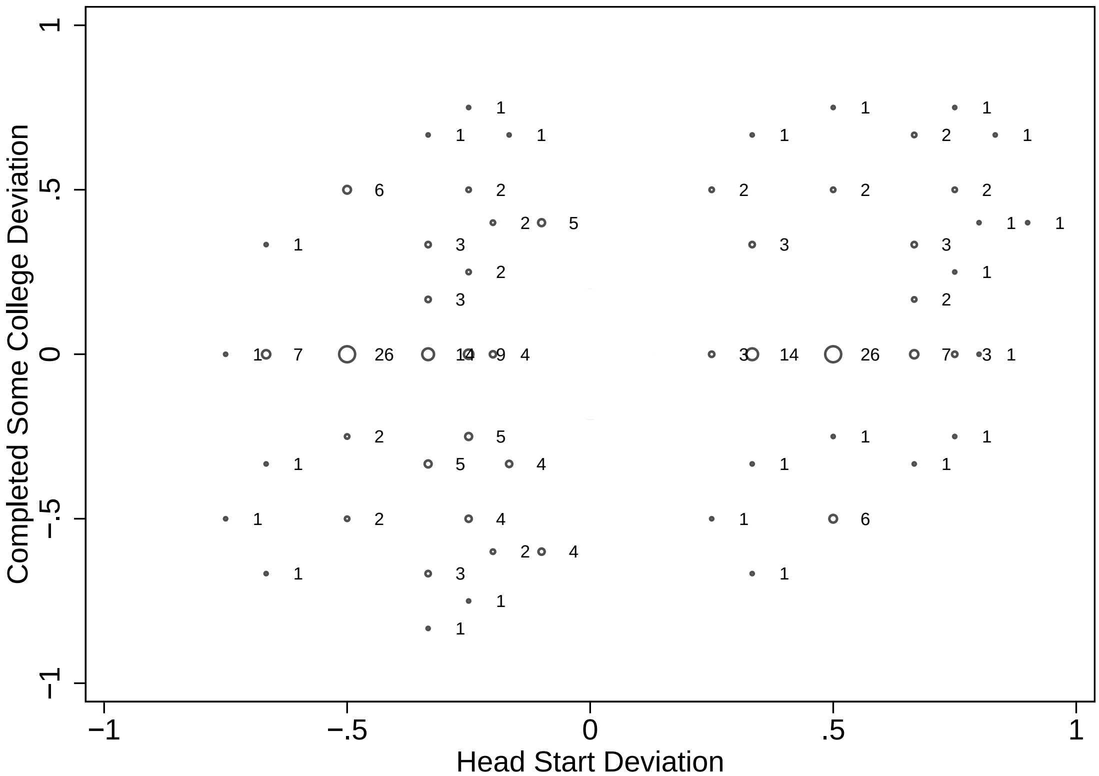
(b) Average Citations by Year of Publication

Figure 2: Raw variation within Families in Head Start, Some College (GMS Sample, whites)



Notes: Size of bubbles represents number of individuals (unweighted). Source: Panel Study of Income Dynamics, 1968-2011 waves.

Figure 3: Variation within Families in Head Start, Some College, Remove Zero Variation in Head Start (GMS Sample, whites)



Notes: Size of bubbles represents number of individuals (unweighted). N = 211 individuals across 70 families. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Figure 4: Probability Family is in Switcher Sample as Function of π

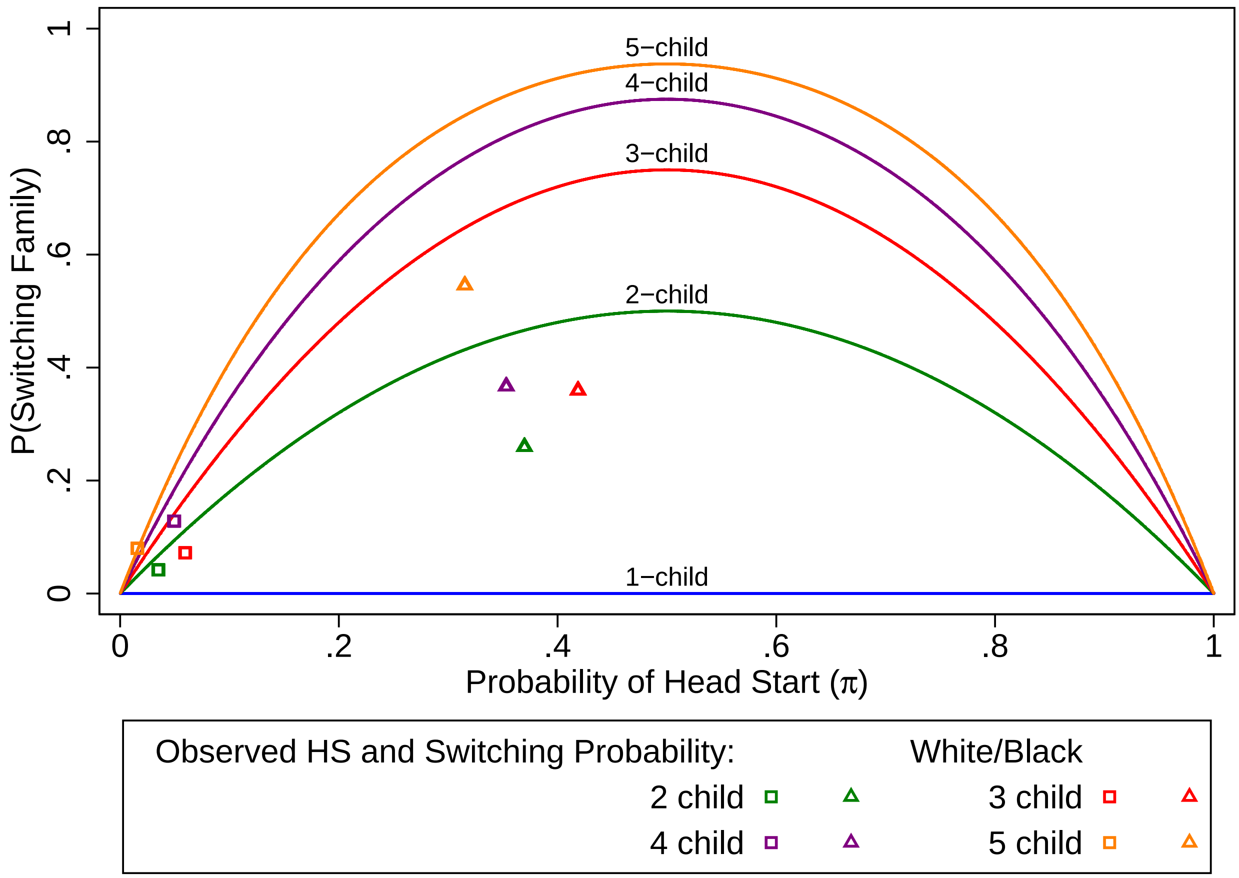
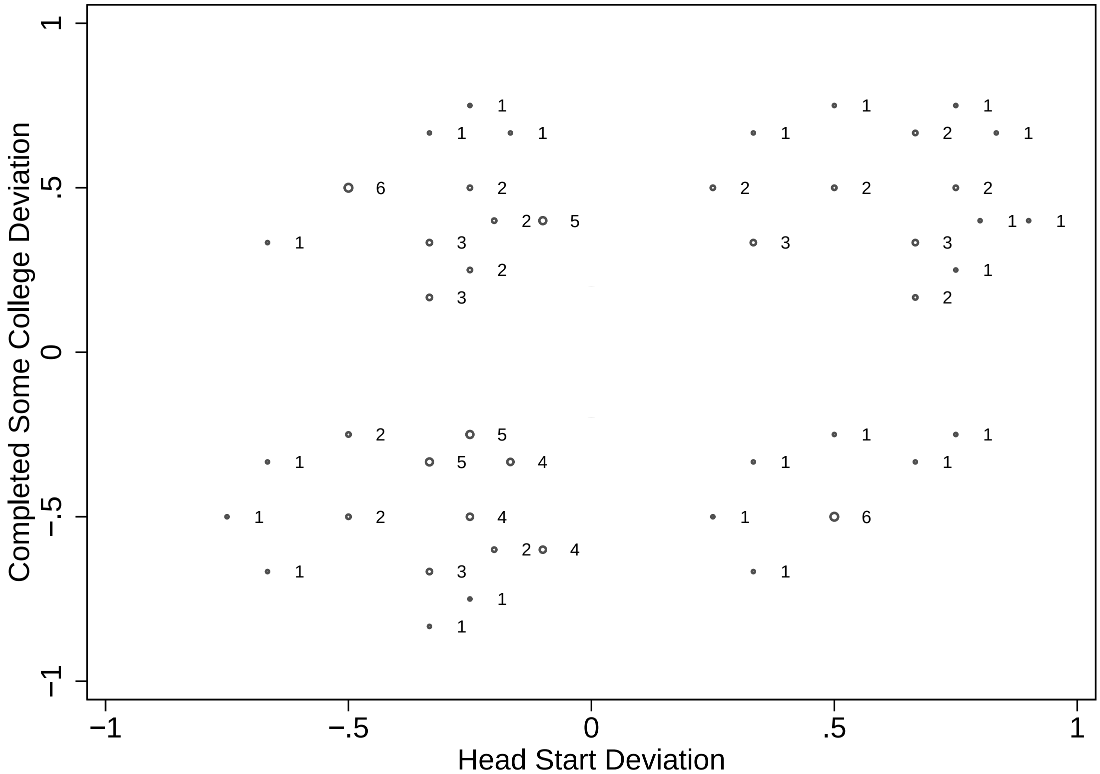
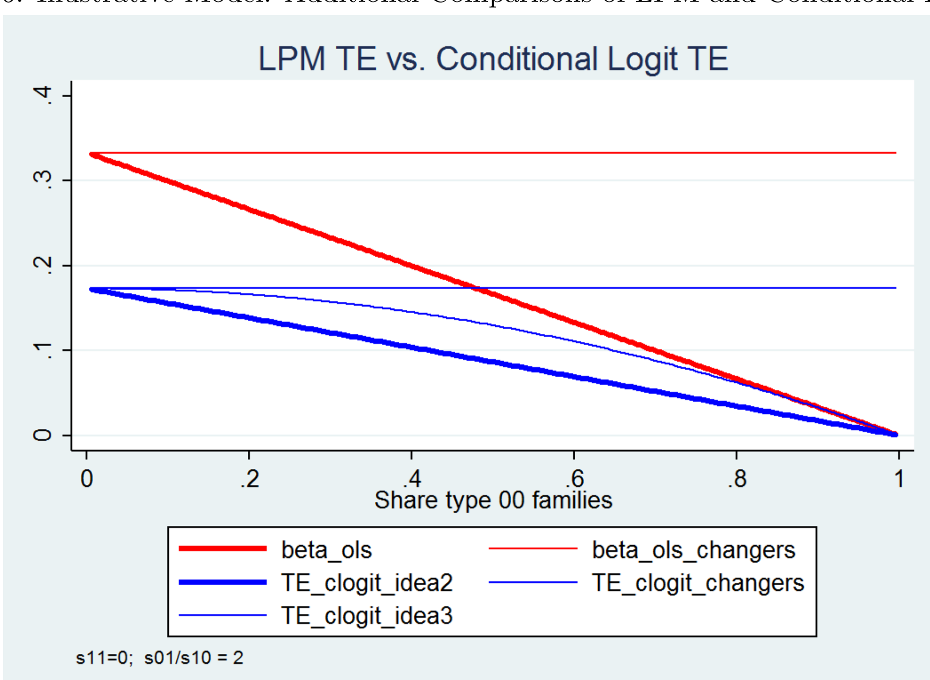


Figure 5: Variation within Families in Head Start, Some College, Remove Zero Variation in Head Start and College (GMS Sample, whites)



Notes: Size of bubbles represents number of individuals (unweighted). N= 98 individuals across 27 families. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Figure 6: Illustrative Model: Additional Comparisons of LPM and Conditional Logit TE



III Replication of GTC

Table A.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 A.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 A.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 A.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
	(0.0260)	(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 A.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[1]		X	X	X	X	X	X

Table A.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 A.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 A.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 A.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

IV Appendix

Figure B.1: Correlation in Head Start Attendance: Probability Child Attended Conditional on Previous Siblings Attending

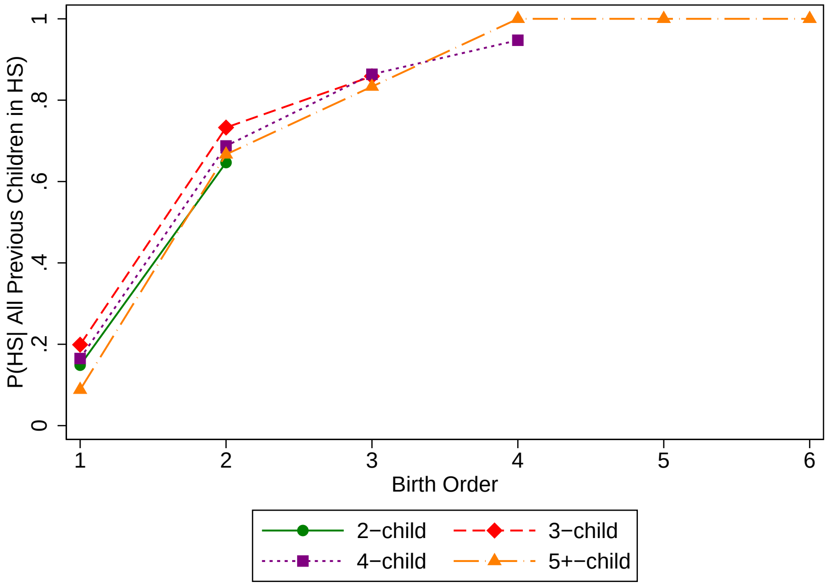


Figure B.2: Correlation in Not Attending Head Start: Probability Child Did Not Attend Conditional on Previous Siblings Not Attending

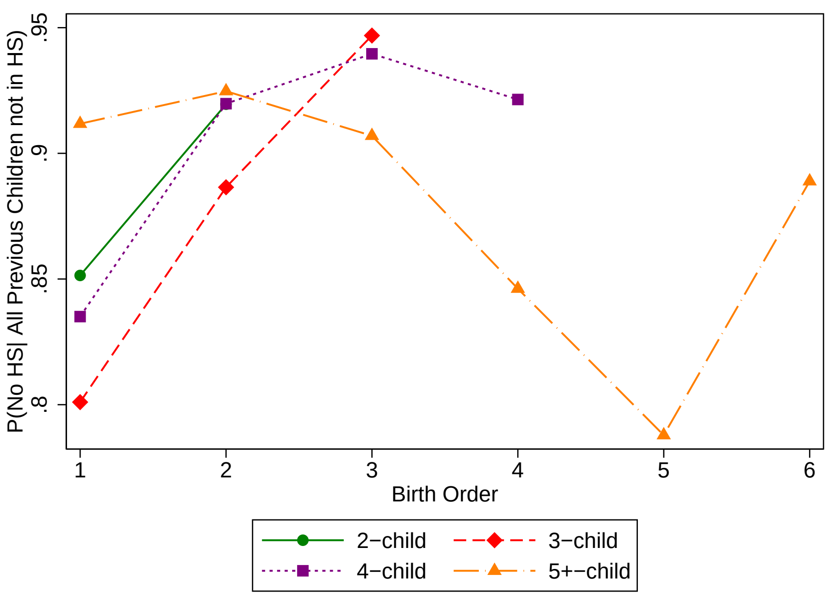


Table B.1: Analytic Formulas for Illustrative Model

	OLS/LPM	Conditional Logit
β	$s_{01} - s_{10}$	$ln(\frac{cs_{01}}{1-cs_{01}})$
se^2	$\frac{1}{F-1}[(s_{01}+s_{10}) - \frac{1}{F}(s_{01}-s_{10})^2]$ $= \frac{1}{F-1}(s_{01}+s_{10}) + \mathcal{O}(F^{-2})$	$\frac{s_{01}+s_{10}}{F \cdot s_{01} \cdot s_{10}}$
Approx. t^* for H_0: TE=0	$\sqrt{F}\frac{s_{01}-s_{10}}{\sqrt{s_{01}+s_{10}}}$	$\sqrt{\frac{F \cdot s_{01} \cdot s_{10}}{s_{01}+s_{10}}} \cdot ln(\frac{cs_{01}}{1-cs_{01}})$
TE at $\overline{y}$	$s_{01} - s_{10}$	$ln(\frac{cs_{01}}{1-cs_{01}}) \cdot \overline{y}(1-\overline{y})$
β, LPM for "LHS switchers"	$2cs_{01} - 1 = \frac{s_{01}-s_{10}}{s_{01}+s_{10}}$	

Table B.2: Summary Statistics, Auxiliary New Sample Outcomes

	All	Head Start	No Head Start	Sibling Sample
Inputs to Economic Sufficiency Index, 30				
Ever on AFDC/TANF by age 30	0.062	0.220	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
Inputs to Economic Sufficiency Index, 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, 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, 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	7363	1345	6018	5355

Notes: Weighted to be representative of 1995 population; see text for details. SD, in parentheses, are omitted for binary variables.

Table B.3: N's, New Sample Characteristics

	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.4: N's, Auxiliary New Sample Outcomes

	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.5: N's, New Sample Outcomes

	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.6: Regression: Pre-Head Start Outcomes, New Sample

	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	0.000**	-0.000***	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Other preschool	-0.000***	-0.000***	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6219	4313	4023	1719	2298
Family income (age 2) (CPI adjusted)					
Head Start	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Other preschool	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
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.7: Regression: Principal Component

	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: Outcomes are indices created using weights from principal component analysis. See text for details. * p < .10, ** p < .05, *** p < .01. 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.

Table B.8: Regression: Inputs to Economic Sufficiency Index at age 30

	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: 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. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.9: Regression: Inputs to Economic Sufficiency Index at age 40

	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: . 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. High school graduation and some college attendance are also inputs to the index but are not shown here. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.10: Regression: Inputs to Good Health Index at age 30

	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: 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. * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.11: Regression: Inputs to Good Health Index at age 40

	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: 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. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.12: Regression: Interaction with Cohort (Binary, Adult Outcomes)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
Economic Sufficiency Index, age 30					
Head Start	-0.096*	-0.069	-0.064	-0.106	0.038
	(0.056)	(0.063)	(0.076)	(0.097)	(0.126)
Later Head Start Cohorts	-0.121*	-0.122	-0.100	-0.051	-0.196
	(0.071)	(0.082)	(0.099)	(0.122)	(0.175)
Mean Y	0.094	0.096	0.096	-0.552	0.213
Observations	7372	5361	5361	2369	2986
Good Health Index, Age 30					
Head Start	-0.289***	-0.259***	-0.042	-0.052	-0.006
	(0.066)	(0.069)	(0.145)	(0.162)	(0.259)
Later Head Start Cohorts	-0.184*	-0.194	-0.445*	0.281	-0.780*
	(0.111)	(0.122)	(0.253)	(0.276)	(0.409)
Mean Y	0.004	0.017	0.017	-0.357	0.074
Observations	4749	3600	3114	1150	1959

Notes: 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. Later Head Start cohorts defined as individuals born after the median birth year among individuals born after Head Start became available (1966) - in the sample, roughly 1977. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.13: Regression: Interaction with Cohort (Binary, GTC Outcomes)

	All	Sibs	Mom FE	Blk, FE	Wht, FE
High School					
Head Start	-0.005	-0.010	-0.015	-0.015	-0.036
	(0.025)	(0.029)	(0.030)	(0.035)	(0.055)
Later Head Start Cohorts	0.028	0.019	0.009	-0.045	0.061
	(0.026)	(0.031)	(0.039)	(0.041)	(0.075)
Mean Y	0.913	0.912	0.912	0.862	0.921
Observations	7372	5361	5361	2369	2986
Some College					
Head Start	0.059*	0.052	0.055	0.003	0.120**
	(0.031)	(0.038)	(0.037)	(0.042)	(0.059)
Later Head Start Cohorts	-0.046	-0.032	-0.034	-0.071	-0.000
	(0.039)	(0.045)	(0.049)	(0.053)	(0.087)
Mean Y	0.531	0.532	0.532	0.396	0.556
Observations	7372	5361	5361	2369	2986
Ln Earnings 23-25					
Head Start	-0.002	-0.039	0.017	0.055	0.035
	(0.068)	(0.081)	(0.120)	(0.151)	(0.182)
Later Head Start Cohorts	0.090	0.168*	0.174	0.019	0.248
	(0.089)	(0.102)	(0.183)	(0.212)	(0.262)
Mean Y	9.588	9.578	9.578	9.207	9.630
Observations	4351	3309	2726	986	1736
No Crime					
Head Start	-0.011	-0.018	-0.007	0.031	-0.067
	(0.027)	(0.033)	(0.033)	(0.028)	(0.064)
Later Head Start Cohorts	0.063**	0.074**	0.081	0.003	0.172
	(0.030)	(0.036)	(0.059)	(0.062)	(0.129)
Mean Y	0.899	0.898	0.898	0.897	0.898
Observations	5005	3591	3206	1366	1836

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. Later Head Start cohorts defined as individuals born after the median birth year among individuals born after Head Start became available (1966) - in the sample, roughly 1977. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.14: Regression: Interaction with Cohort (Linear)

	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: 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. Trend has been normed so that 0 is the minimum birth year among individuals born while Head Start was available (1966). * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.15: Cross-Sectional Regression Coefficients (Some College, All) - Interaction with Number of Children in Family

	CX	FE
Head Start x 1 child family	0.091*	
	(0.048)	
Head Start x 2 child family	0.052	-0.082
	(0.044)	(0.060)
Head Start x 3 child family	-0.000	0.029
	(0.049)	(0.047)
Head Start x 4 child family	-0.061	0.148**
	(0.061)	(0.067)
Head Start x 5+ child family	0.156**	0.111
	(0.075)	(0.075)
Head Start x Unknown child family	-0.027	
	(0.069)	
Observations	7372	5361
N Non-Switch/Switch		
A+P All Weights	0.036	
A+P Sib. Weights	0.029	0.021
A+P Switch Weights	0.039	0.045

Notes: Columns 3 and 4 show the coefficients from one regression that interacts and indicator for Head Start with the number of children in the family and whether the family have variation in Head Start attendance. Columns 1, 3, and 4 include controls, but not mother f.e., and SE are clustered at 1968 family id. Column 2 includes mother fixed effects, and SE clustered by mother id. The bottom rows of columns 1 and 2 show the weighted average of the coefficients and the respective standard errors when using weights determined by the overall distribution of families, the distribution of 2+ child families, and the distribution of 2+ child families that have variation in Head Start attendance. * p < .10, ** p < .05, *** p < .01. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.16: Share of Sample By Number of Children in Family, White Sample

	Shares			A+P Weights		
	All	Sibs	Switchers	All	Sibs	Switchers
1 child family	0.120	0.000	0.000	0.171	0.000	0.000
2 child family	0.283	0.358	0.246	0.261	0.358	0.294
3 child family	0.236	0.299	0.260	0.297	0.407	0.317
4 child family	0.147	0.186	0.254	0.115	0.158	0.247
5+ child family	0.123	0.156	0.240	0.056	0.077	0.141
Unknown child family	0.091	0.000	0.000	0.099	0.000	0.000
Observations	4258	2989	213	4258	2989	213

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

Table B.17: Comparison of Switcher and Non-Switcher families, White Sample

	(1) Switch	(2) Non-Switch	(3) Non-Switch, HS=1	(4) T-Stat. (1)=(2)	(5) T-Stat. (1)=(3)	(6) Beta Switch	(7) T-Stat (6)
Fraction female	0.586	0.495	0.473	2.592	1.705	0.026	0.502
Fraction African-American	0.000	0.000	0.000	.	.	0.000	0.000
Mother's yrs education	9.395	11.363	9.823	-9.449	-1.323	-0.227	11.334
Father's yrs education	9.599	11.531	10.281	-8.890	-2.124	-0.337	11.494
Had a single mother at age 4	0.137	0.065	0.179	2.751	-0.846	0.026	0.072
Family income (age 3-6) (CPI adjusted)	35930.879	55474.857	29116.215	-12.542	2.812	-4432.725	53824.734
Mother employed, age 0	0.508	0.567	0.450	-1.402	0.808	0.091	0.556
Mother employed, age 1	0.506	0.538	0.485	-0.772	0.294	0.064	0.531
Mother employed, age 2	0.513	0.548	0.575	-0.881	-0.871	0.070	0.536
Household size at age 4	5.313	4.419	4.246	5.521	4.746	0.721	4.450
Fraction low birth weight	0.059	0.054	0.173	0.270	-2.133	-0.007	0.054
Observations	213	3474	78	3687	291	4258	4258

Notes: White sample only. Weighted to be representative of 1995 population; see text for details. Columns 1, 2, and 3 present the means of switcher families (have variation in Head Start), non-switcher families (no variation in Head Start), and non-switcher families that attend Head Start, respectively. Column 4 shows the coefficient from a regression of the variable shown in the row heading on an indicator for being in a switcher family, and Column 5 shows the corresponding standard errors, clustered by id1968. All controls from the main specification are included, excluding the variable shown in the row heading. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.18: Comparison of Switcher and Non-Switcher families, All Sample

	(1) Switch	(2) Non-Switch	(3) Non-Switch, HS=1	(4) T-Stat. (1)=(2)	(5) T-Stat. (1)=(3)	(6) Beta Switch	(7) T-Stat (6)
Fraction female	0.562	0.495	0.530	4.067	1.362	0.024	0.504
Fraction African-American	0.516	0.111	0.714	25.877	-9.017	0.249	0.150
Mother's yrs education	9.283	11.230	10.304	-21.590	-9.174	-0.140	11.116
Father's yrs education	9.190	11.371	10.445	-19.594	-8.976	-0.389	11.238
Had a single mother at age 4	0.252	0.099	0.332	10.049	-3.581	0.055	0.112
Family income (age 3-6) (CPI adjusted)	31808.717	52574.453	25914.881	-24.735	6.725	-4758.501	50339.121
Mother employed, age 0	0.508	0.570	0.542	-3.099	-1.292	0.055	0.557
Mother employed, age 1	0.517	0.543	0.541	-1.342	-0.925	0.058	0.536
Mother employed, age 2	0.536	0.554	0.596	-0.951	-2.353	0.118	0.544
Household size at age 4	5.487	4.451	4.353	12.343	10.883	0.755	4.535
Fraction low birth weight	0.077	0.058	0.135	1.971	-3.464	0.010	0.060
Observations	1103	5500	795	6603	1898	7372	7372

Notes: Weighted to be representative of 1995 population; see text for details. Columns 1, 2, and 3 present the means of switcher families (have variation in Head Start), non-switcher families (no variation in Head Start), and non-switcher families that attend Head Start, respectively. Column 4 shows the coefficient from a regression of the variable shown in the row heading on an indicator for being in a switcher family, and Column 5 shows the corresponding standard errors, clustered by id1968. All controls from the main specification are included, excluding the variable shown in the row heading. Source: Panel Study of Income Dynamics, 1968-2011 waves.

Table B.19: Comparison of Switcher and Non-Switcher families Conditional on Head Start, White Sample

	(1) Switch	(2) Switch, HS=1	(3) Non-Switch, HS=1	(4) T-Stat. (1)=(3)	(5) T-Stat. (2)=(3)
Fraction female	0.586	0.559	0.473	1.705	1.109
Fraction African-American	0.000	0.000	0.000	.	.
Mother's yrs education	9.395	10.357	9.823	-1.323	1.434
Father's yrs education	9.599	10.050	10.281	-2.124	-0.580
Had a single mother at age 4	0.137	0.176	0.179	-0.846	-0.048
Family income (age 3-6) (CPI adjusted)	35930.879	36239.011	29116.215	2.812	2.420
Mother employed, age 0	0.508	0.492	0.450	0.808	0.492
Mother employed, age 1	0.506	0.565	0.485	0.294	0.947
Mother employed, age 2	0.513	0.489	0.575	-0.871	-1.035
Household size at age 4	5.313	5.060	4.246	4.746	3.060
Fraction low birth weight	0.059	0.076	0.173	-2.133	-1.618
Observations	213	89	78	291	167

Notes: White sample only. Weighted to be representative of 1995 population; see text for details. Columns 1, 2, and 3 present the means of switcher families (have variation in Head Start), individuals that attended Head Start in switcher families, and non-switcher families that attend Head Start, respectively. Column 4 shows the coefficient from a regression of the variable shown in the row heading on an indicator for being in a switcher family, and Column 5 shows the corresponding standard errors, clustered by id1968. All controls from the main specification are included, excluding the variable shown in the row heading. Source: Panel Study of Income Dynamics, 1968-2011 waves.