Assessing the Performance of Nonexperimental Estimators for Evaluating Head Start

Journal of Labor Economics, forthcoming

Andrew Griffen (U. of Tokyo)
and
Petra E. Todd (U. of Pennsylvania)

January 21, 2017

Introduction

- There is a long-standing debate in the literature over whether the impacts of program interventions can be reliably evaluated without data from a randomized experiment.
- Randomization
  - generates a control group that has the same distribution of observed and unobserved characteristics as the treatment group.
  - also has drawbacks (high cost, disruption, selective attrition)
- Nonexperimental estimators
  - tend to be less costly and disruptive
  - accumulated evidence indicates that impact estimates can be sensitive to the estimator used (Ashenfelter (1978), Bassi (1984), Ashenfelter and Card (1985), Lalonde (1986) and Fraker and Maynard (1987))

Literature

- Builds on an earlier literature that evaluates the performance of nonexperimental estimators. (LaLonde (1986), Heckman and Hotz (1989), Dehejia and Wahba (1999,2001), Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998).)
- Impact estimates based on experimental data provide a benchmark against which to gauge the performance of alternative nonexperimental estimators.

Literature

- LaLonde (1986) used data from the National Supported Work (NSW) Demonstration experiment combined with non experimental data from the CPS and PSID.
  - Applied a number of standard evaluation estimators, including simple regression adjustment, difference-in-differences, and the two-step estimator of Heckman (1979).
  - Found alternative estimators produce different estimates and that non experimental estimators are generally not reliable.
- Heckman and Hotz (1989) also used the NSW data and applied pre-program exogeneity tests that eliminated estimators with the largest bias. broad range of specification tests to guide the choice among nonexperimental estimators.

Literature

- More recent debate over the performance of nonexperimental estimators also considers other types of program interventions and outcomes.
- Some papers find that propensity score matching performs well against an experimental comparison (Diaz and Handa, 2006; Handa and Maluccio, 2010; Bifulco, 2012, McKenzie, Stillman and Gibson, 2010).

Literature

- Other work argues that even under ideal conditions, propensity score matching does not mimic experimental results (Agodini and Dynarski, 2004; Arceneaux, Gerber, and Green, 2006; McKenzie, Stillman, and Gibson, 2010; Peikes, Moreno and Orzol, 2008; Wilde and Hollister, 2007).
- Similar to HIT and HIST, Cook, Shadish and Wong (2008) argue that nonexperimental estimators perform better when there exists a rich set of control variables including baseline outcomes, outcomes are measured across data sets and where the variables affecting selection into treatment are known and measured.

Goals of this paper

- This paper evaluates performance of non experimental estimators in assessing the effect of a early education program targeted at children from disadvantaged backgrounds aged 3-5.
- Experimental data come from the Head Start Impact Study (HSIS) and the nonexperimental data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B).
- Compare experimental and nonexperimental impact estimates on a variety of outcome measures related to children's achievement, children's health outcomes, parental behaviors, and parental labor supply.
- Nonexperimental estimators include cross-section and difference-in-difference regression estimators and cross-section and difference-in-difference matching estimators.
- Relative to previous literature consider a larger number of outcome variables

Previous Literature on Head Start

- Widely studied program (in economics, education and psychology) with hundreds of quasi- and non-experimental evaluations (McGroder, 1990; Zigler and Styfco, 2004; Shager et al., 2012).
- Most recent experimental study is Puma et. al. (2005)
  - Found some positive impacts at the end of the first Head Start year, especially in reading assessments and some parental behaviors.
  - Found complete fade-out and negligible impacts in most domains and outcomes by the end of 1st grade and continuing into 3rd grade.
  - Mechanism for fade-out is unclear.
  - HSIS data reanalysis estimate LATE parameters for different subgroups and find larger gains associated with switching from the home into Head Start than from other formal child care into Head Start (Feller et al. 2014, Kline and Walters, 2014).

Previous Literature on Head Start

- A fairly common finding in the earlier non-experimental literature is that Head Start impacts on child outcomes “fade-out" after the child leaves the program (Westinghouse Study, 1969; McKey et al., 1985; Currie and Thomas, (1995,1999)).
- Phillips and White (2004) describe how these findings plus increasing pressure for measuring effectiveness of government programs in the 1990s led to a push for a national Head Start evaluation.

Previous Literature on Head Start

- Papers also find impacts of Head Start on longer term outcomes (Garces et al. 2002; Deming, 2009; Ludwig and Miller, 2007; Carneiro and Ginja, 2014).
- Short-run fade-out plus long term impacts would mimic the findings of the Perry Preschool impact.
- HSIS experimental results did not end the production of papers using nonexperimental impact evaluation of Head Start (Pigott and Israel, 2005; Zhai et al., 2011; Zhai et al., 2013; Lee et al., 2014a; Lee et al., 2014b)

The Evaluation Problem

Y_1 denote the outcome conditional on participation
Y_0 denotes the outcome conditional on non-participation

Program impact is

$$\Delta = Y_1 - Y_0.$$

For each person, only Y_1 or Y_0 is observed, so Δ is not directly observed.

$D = 1$ for program participants for whom Y_1 is observed.
$D = 0$ for nonparticipants for whom Y_0 is observed.

The Evaluation Problem

- X denotes a vector of observed covariates whose distribution is assumed not to be affected by the program (e.g. gender, race/ethnicity, geographic location).
- Experiments are designed to provide evidence on the so-called *average impact of treatment on the treated*

$$\begin{aligned} ATT(X) &= E(\Delta|X, D=1) \\ &= E(Y_1 - Y_0|X, D=1) \\ &= E(Y_1|X, D=1) - E(Y_0|X, D=1) \end{aligned}$$

- For other parameters of interest, see Heckman, Lalonde and Smith (1999)

Nonexperimental Estimators

- Use two types of data to impute counterfactual outcomes for program participants: data on participants prior to entering the program and data on nonparticipants.
- We consider regression-based methods and matching methods

Cross-section regression estimators

Let i denote the individual and t the time period

$$\begin{aligned} Y_{1it} &= \varphi_1(X_{it}) + U_{1it} \\ Y_{0it} &= \varphi_0(X_{it}) + U_{0it}, \end{aligned}$$

where U_{1it} and U_{0it} are distributed independently across persons and satisfy $E(U_{1it}|X_{it}) = 0$ and $E(U_{0it}|X_{it}) = 0$.

Cross-section regression estimators

The $ATT(X_{it})$ parameter is

$$ATT(X_{it}) = \varphi_1(X_{it}) - \varphi_0(X_{it}) + E(U_{1it} - U_{0it}|X_{it}, D_i = 1)$$

Add and subtract $D_i E(U_{1it} - U_{0it}|X_{it}, D_i = 1)$ to get:

$$Y_{it} = \varphi_0(X_{it}) + D_i ATT(X_{it}) + U_{0it} + D_i[U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|X_{it}, D_i = 1)] \quad (1)$$

Defining the error term as

$$\varepsilon_{it} = U_{0it} + D_i[U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|X_{it}, D_i = 1)]$$

consistency requires

$$E(U_{0it}|X_{it}, D_i = 1) = 0,$$

Difference-in-Difference regression estimators

Uses pre- and post-program data (t' and t data) on $D = 1$ and $D = 0$ persons.
The outcomes in the post program and pre-program time periods can be written as

$$\begin{aligned} Y_{it} &= \varphi_0(X_{it}) + D_i ATT(X_{it}) + U_{0it} + D_i(U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_i = 1 \\ Y_{it'} &= \varphi_0(X_{it'}) + U_{0it'} \end{aligned}$$

Taking differences,

$$\begin{aligned} Y_{it} - Y_{it'} = \varphi_0(X_{it}) - \varphi_0(X_{it'}) + D_i ATT(X_{it}) + (U_{0it} - U_{0it'}) + \\ D_i(U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_i = 1, X_{it})) \end{aligned}$$

Defining the error term as

$$\tilde{\varepsilon} = (U_{0it} - U_{0it'}) + D_i(U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_i = 1, X_{it}))$$

consistency requires

$$E(\tilde{\varepsilon}|X_{it}, D_i) = 0$$

.

- A special case where this assumption would be satisfied is if $U_{0it} = f_i + v_{it}$ where f_i depends on i but does not vary over time and v_{it} is a random error term (i.e. U_{0it} satisfies a fixed effect assumption)
- Lalonde (1986) implements both the standard estimator just described and an "unrestricted" version that includes $Y_{it'}$ as a right-hand-side variable.

Cross-section matching Methods

- Traditional matching estimators pair each program participant with an observably similar nonparticipant (Rosenbaum and Rubin, 1983).
- Researchers usually assume that there exists a set of observables Z for which the non-participation outcome Y_0 is independent of participation status D conditional on Z

$$Y_0 \perp\!\!\!\perp D \mid Z \, .$$

It is also assumed that for all Z there is a positive probability of either participating ($D = 1$) or not participating ($D = 0$), i.e.

$$0 < \Pr(D = 1|Z) < 1.$$

Under these assumptions, the mean impact of the program can be written as

$$\begin{aligned}\Delta &= E(Y_1 - Y_0|D=1) \\ &= E(Y_1|D=1) - E_{Z|D=1}\{E_Y(Y_0|D=1,Z)\} \\ &= E(Y_1|D=1) - E_{Z|D=1}\{E_Y(Y_0|D=0,Z)\},\end{aligned}$$

where the first term can be estimated from the treatment group and the second term from the mean outcomes of the matched (on Z) comparison group.

Reducing the Dimensionality

- Matching may be difficult to implement when the set of matching variables Z is large.
- We use Rosenbaum and Rubin (1983) result that when Y_0 outcomes are independent of program participation conditional on Z they are also independent of participation conditional on the *propensity score*, $\Pr(D = 1|Z)$
- Provided that the propensity score can be estimated parametrically (or semiparametrically at a rate faster than the nonparametric rate), the dimensionality of the matching problem is reduced.

Matching Estimators

Let $P = \Pr(D = 1|Z)$.

$$\hat{\alpha}_M = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} [Y_{1i} - \hat{E}(Y_{0i}|D = 1, P_i)]$$

where

$$\hat{E}(Y_{0i}|D = 1, P_i) = \sum_{j \in I_0} W(i,j) Y_{0j},$$

and where I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support , and n_1 the number of persons in the set $I_1 \cap S_P$.
- The match for each participant $i \in I_1 \cap S_P$ is constructed as a weighted average over the outcomes of non-participants, where the weights $W(i,j)$ depend on the distance between P_i and P_j.

Difference-in-difference matching

- There may be systematic differences between participant and nonparticipant outcomes, even after conditioning on observables.
    - because of program selectivity on unmeasured characteristics
    - because of levels differences in outcomes across different markets in which the participants and nonparticipants reside,
    - because outcomes outcomes for participants and nonparticipants are measured in different ways (as when data are collected using different survey instruments).

Difference-in-difference matching

- As described in HIT (1997) and HIST (1998), allows for temporally invariant differences in outcomes between participants and nonparticipants.
- Analogous to the regression estimator but does not impose the linear functional form restriction and reweights observations according to the weighting functions used by the matching estimators.

Assumes that

$$E(Y_{0t} - Y_{0t'}|P, D = 1) = E(Y_{0t} - Y_{0t'}|P, D = 0),$$

where t and t' are time periods after and before the program enrollment date.

- Also requires support condition in both time periods.

$$\hat{\alpha}_{KDM} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left\{ (Y_{1ti} - Y_{0t'i}) - \sum_{j \in I_0 \cap S_P} W(i,j)(Y_{0tj} - Y_{0t'j}) \right\},$$

Head Start Impact Study Data

- Head Start serves nearly 1 million children between the ages of 3 and 5 from low income families in the US and has the goal to "[p]romote school readiness by enhancing the social and cognitive development of children."
- Despite long-term popular and political support for Head Start since its inception in 1965, the 105th U.S. Congress in 1998 took the unusual step to mandate a randomized evaluation of Head Start (P.L. 105-285).

Head Start Impact Study Data

- Head Start centers were selected for inclusion in the experiment according to a stratified randomized sampling design that was intended to be largely representative of the entire population of Head Start centers, except it was required that the centers be oversubscribed. (85% of the centers satisfied this criteria)
- 2,783 children randomized to received an offer of Head Start and 1,884 children randomized out to serve as a control group.

Head Start Impact Study Data

- A group of 4 year-olds who were randomized to receive an offer of Head Start or not (HSIS 4 year-olds) and a group of 3 year-olds who were randomized to receive an offer of Head Start or to receive a delayed option to reapply for Head Start the subsequent year (HSIS 3 year-olds).
- The study followed children longitudinally and collected follow-up data at the end of the first Head Start year, the second Head Start year (for the HSIS 3 year olds) and at the end of kindergarten, 1st grade and 3rd grade.
- We focus on the first year impacts in the current study for the two groups of children

Head Start Impact Study Data

- The HSIS reported first year impact estimates on 33 outcomes measures across four domains: cognitive, socio-emotional, health and parenting.
- We can only find comparable outcomes in the ECLS-B data for a subset of the measures.
- We also examine parental labor supply and family income, although they were not among the original outcomes considered by the HSIS

Head Start Impact Study Data

- In general, the pattern of impact estimates obtained from the HSIS can be described as small and insignificant.
- Of the 33 outcomes considered in the first year findings, 9 impact estimates were statistically significant for the 3 year olds and 7 impact estimates were statistically significant for the 4 year olds.
- Most of these significant outcomes were early reading assessments in the cognitive domain and parenting behaviors.
- Impact estimates at kindergarten entry, the end of 1st grade and the end of 3rd grade were not significant.

Early Childhood Longitudinal Study - Birth Cohort

- Nationally representative panel of 14,000 children born in 2001.
- Followed children from birth until kindergarten entry and collected detailed information about family background, home environment, maternal work decisions, family income, child care usage and cognitive achievement outcomes.
- Covers the 2001 cohort of children born in the US compared HSIS children who were born from 1997 to 1999.
- Drew some sample questions directly from the HSIS, which increases the likelihood of finding similar questions across data sets.
- Also sent questionnaires directly to child care providers. Parental reports of Head Start use in the ECLS-B overstate participation relative to national statistics

Early Childhood Longitudinal Study - Birth Cohort

- Combine information from the provider Head Start reports with information from the parental survey about when the child first began to participate in Head Start.
- To mimic the design of the HSIS baseline data collection, we assign children to the 4 year old Head Start group (ECLS-B 4 year olds) if they both reported being in Head Start and they had begun Head Start with two months of the ECLS-B 3rd round.
- Other children who began Head Start earlier are assigned to the 3 year old cohort (ECLS-B 3 year olds).
- Drop families with total income greater than $\$50,000$: unlikely to Head Start eligible

Outcome variables

- Cognitive:
  - Reading percentile score
  - Math percentile scores
- Health
  - Child health good or excellent? 1 Yes, 0 No
  - Child has health insurance? 1 Yes, 0 No
- Parenting
  - Parent read to the child in the last week? 1 Yes, 0 No
  - Parental safety practices scale consisting of frequency of car seat and smoke detector use
  - Spanked the child in the last week? 1 Yes, 0 No
  - Used a time out in the last week? 1 Yes, 0 No
- Labor
  - Household income ($ / year)
  - Maternal employment: 1 Full/Part time, 0 otherwise
  - Paternal employment: 1 Full/Part time, 0 otherwise

Table 1: Descriptive statistics HSIS and ECLS-B at baseline

		ECLS-B 3-year-olds			ECLS-B 4-year-olds	
	HSIS 3 year olds	Head Start	Non-Head Start	HSIS 4 year olds	Head Start	Non-Head Start
% children female	0.52	0.47	0.50	0.49	0.44	0.48
	(0.69)	(0.71)	(0.70)	(0.68)	(0.68)	(0.69)
% children black	0.29	0.35	0.18	0.21	0.16	0.17
	(0.61)	(0.63)	(0.47)	(0.54)	(0.42)	(0.44)
% children Hispanic	0.28	0.32	0.31	0.34	0.45	0.33
	(0.60)	(0.71)	(0.65)	(0.63)	(0.70)	(0.65)
% children white	0.23	0.26	0.45	0.23	0.34	0.45
	(0.55)	(0.69)	(0.71)	(0.53)	(0.70)	(0.70)
Age child baseline (months)	45.13*	24.41	24.44	55.81*	51.04	52.36
	(5.38)	(1.71)	(1.65)	(6.15)	(4.38)	(5.54)
Age child at assessment (months)	50.45*	53.07	52.54	61.25*	63.73	64.65
	(5.29)	(5.68)	(5.70)	(6.12)	(4.18)	(4.99)
Years of maternal education	11.87	11.96	12.50	11.61	11.91	12.32
	(2.55)	(2.59)	(3.09)	(2.75)	(2.36)	(3.02)
Age mothers	28.27	28.76	29.44	29.21	28.06	29.38
	(7.86)	(8.81)	(8.09)	(8.08)	(8.27)	(8.22)

Table 1: Descriptive statistics HSIS and ECLS-B at baseline **cont'd**

		ECLS-B 3 year olds			ECLS-B 4 year olds	
	HSIS 3-year-olds	Head Start	Non-Head Start	HSIS 4 year olds	Head Start	Non-Head Start
% mothers married	0.38	0.41	0.56	0.38*	0.57	0.56
	(0.66)	(0.72)	(0.69)	(0.65)	(0.69)	(0.69)
% mothers separated	0.05	0.05	0.03	0.07	0.05	0.04
	(0.27)	(0.26)	(0.23)	(0.33)	(0.36)	(0.28)
% mothers divorced	0.05	0.06	0.06	0.07*	0.03	0.07
	(0.27)	(0.35)	(0.36)	(0.34)	(0.20)	(0.36)
% mothers never married	0.34*	0.45	0.32	0.29	0.31	0.30
	(0.63)	(0.72)	(0.65)	(0.60)	(0.64)	(0.62)
% teenage mothers	0.13	0.14	0.08	0.16	0.17	0.09
	(0.44)	(0.56)	(0.39)	(0.48)	(0.54)	(0.40)
Number children under age 6	1.80	1.86	1.74	1.72	1.78	1.73
	(1.12)	(1.12)	(1.06)	(0.98)	(0.90)	(1.04)
% live in urban area	0.79	0.83	0.82	0.84	0.75	0.82
	(0.56)	(0.51)	(0.56)	(0.49)	(0.60)	(0.55)
% speak English at home	0.77	0.77	0.77	0.70	0.67	0.77
	(0.57)	(0.60)	(0.58)	(0.63)	(0.66)	(0.58)
% own house	0.28	0.24	0.37	0.26	0.33	0.41
	(0.62)	(0.61)	(0.68)	(0.59)	(0.64)	(0.69)
N	2449	550	3500	1993	200	3300
F-test		0.00			0.00	

Table 2: Mean outcomes HSIS and ECLS-B at baseline

		ECLS-B 3 year olds			ECLS-B 4 year olds	
	HSIS 3 year olds	Head Start	Non-Head Start	HSIS 4 year olds	Head Start	Non-Head Start
Cognitive						
Reading percentile	-	-	-	31.24	32.46	40.55
				(24.14)	(32.18)	(37.84)
Mathematics percentile	-	-	-	33.78	34.27	41.96
				(30.26)	(36.41)	(37.39)
Health						
Health good/excellent	0.79	0.82	0.86	0.81	0.79	0.84
	(0.57)	(0.55)	(0.46)	(0.51)	(0.57)	(0.48)
Health insurance	0.88*	0.97	0.95	0.87*	0.96	0.92
	(0.49)	(0.28)	(0.31)	(0.44)	(0.24)	(0.38)

Table 2: Mean outcomes HSIS and ECLS-B at baseline **cont'd**

	HSIS 3 year olds	ECLS-B 3 year olds		HSIS 4 year olds	ECLS-B 4 year olds	
		Head Start	Non-Head Start		Head Start	Non-Head Start
Parenting						
Read to child	0.34	0.31	0.37	0.35	0.31	0.29
	(0.66)	(0.71)	(0.68)	(0.65)	(0.66)	(0.64)
Parental safety practices	3.76	3.79	3.83	3.79	3.74	3.66
	(0.77)	(0.68)	(0.59)	(0.59)	(0.65)	(0.83)
Used time out	0.63	0.56	0.56	0.61	0.70	0.69
	(0.66)	(0.72)	(0.70)	(0.67)	(0.64)	(0.63)
Spanked child	0.48	0.53	0.51	0.43	0.38	0.38
	(0.69)	(0.73)	(0.70)	(0.68)	(0.68)	(0.67)
Labor						
Household income	17286*	22375	24993	18672*	25573	24569
	(15133)	(25398)	(17294)	(19251)	(37359)	(16326)
Mother works	0.52*	0.40	0.51	0.50*	0.36	0.54
	(0.68)	(0.70)	(0.70)	(0.67)	(0.67)	(0.69)
Father works (if present)	0.82*	0.94	0.92	0.86	0.86	0.92
	(0.52)	(0.27)	(0.38)	(0.44)	(0.51)	(0.37)
F-test		0.00			0.00	

Table 3: Replicating HSIS experimental impacts

	Intent to Treat		IV (LATE)		Treatment on the Treated	
	3-year-olds	4-year-olds	3-year-olds	4-year-olds	3-year-olds	4-year-olds
Cognitive						
Reading percentile	4.28*	4.56*	6.14*	6.88*	5.31*	7.23*
	(1.10)	(1.31)	(1.56)	(2.02)	(1.86)	(2.18)
Mathematics percentile	2.35	2.10	3.39	3.19	0.86	4.32
	(1.80)	(1.51)	(2.60)	(2.31)	(3.18)	(2.54)
Health						
Health good/excellent	0.05	-0.03	0.07	-0.04	0.00	0.00
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Health insurance	0.01	0.01	0.01	0.02	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)

Table 3: Replicating HSIS experimental impacts (cont'd)

	Intent to Treat		IV (LATE)		Treatment on the Treated	
	3-year-olds	4-year-olds	3-year-olds	4-year-olds	3-year-olds	4-year-olds
Parenting						
Read to child	0.07*	0.03	0.10*	0.04	0.11*	0.07
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)
Parental safety practices	0.03	0.02	0.04	0.03	0.09	0.01
	(0.03)	(0.03)	(0.05)	(0.04)	(0.07)	(0.05)
Used time out	-0.04	-0.07*	-0.06	-0.11*	-0.01	-0.10
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
Spanked child	-0.07*	0.00	-0.10*	0.00	-0.10	-0.08
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
Labor						
Household income	-104	196	-151	294	-485	501
	(734)	(769)	(1063)	(1156)	(1369)	(1329)
Mother works	-0.06	0.01	-0.08	0.01	-0.03	0.00
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
Father works (if present)	0.05	0.03	0.07	0.04	0.04	0.08
	(0.03)	(0.03)	(0.05)	(0.04)	(0.06)	(0.05)

Table 5: Distribution of child care choices

	Head Start	Center	Home	Other
HSIS treatment (program participants)				
3-year-olds	100.0	0.0	0.0	0.0
4-year-olds	100.0	0.0	0.0	0.0
HSIS control (nonparticipants)				
3-year-olds	0.0	31.1	44.2	24.7
4-year-olds	0.0	41.0	41.6	17.4
ECLS-B treatment (program participants)				
3rd round	100.0	0.0	0.0	0.0
ECLS-B control (nonparticipants)				
3rd round	0.0	35.3	41.8	22.9

Table 6a: Regression based estimators for **3 year olds**

		ECLS-B Non-Experimental Head Start Impact Estimates					
	HSIS Impact	Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Cognitive							
Reading percentile	5.31*	0.72	4.43				
	(1.86)	(2.13)	(1.96)				
% Δ from HSIS Impact		-87	-17				
Δ from HSIS Impact (σ)		-2.47	-0.48				
Mathematics percentile	0.86	0.51	2.41				
	(3.18)	(2.07)	(1.94)				
% Δ from HSIS Impact		-41	181				
Δ from HSIS Impact (σ)		-0.11	0.49				
F-test		0.63	1.00				
Health							
Child health good/excellent	0.00	-0.02	0.03	0.03	-0.01	-0.01	0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
% Δ from HSIS Impact		-469	504	523	-282	-244	85
Δ from HSIS Impact (σ)		-0.49	0.52	0.54	-0.29	-0.25	0.09
Health insurance	0.02	0.04	0.03	0.01	-0.01	0.03	0.01
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
% Δ from HSIS Impact		51	2	-46	-136	10	-63
Δ from HSIS Impact (σ)		0.52	0.02	-0.47	-1.37	0.10	-0.64
F-test		1.00	1.00	1.00	0.97	1.00	1.00

Table 6a: Regression based estimators for **3 year olds cont'd**

	HSIS Impact	ECLS-B Non-Experimental Head Start Impact Estimates					
		Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Parenting							
Read to child	0.11*	-0.06†	-0.03†	0.01	-0.05†	-0.03†	-0.05†
	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
% Δ from HSIS Impact		-150	-125	-93	-143	-130	-149
Δ from HSIS Impact (σ)		-3.40	-2.84	-2.11	-3.25	-2.95	-3.38
Parental safety practices	0.09	-0.19†	-0.05	-0.14†	-0.08	-0.17†	-0.07
	(0.07)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
% Δ from HSIS Impact		-319	-164	-269	-189	-301	-186
Δ from HSIS Impact (σ)		-4.17	-2.14	-3.51	-2.47	-3.94	-2.43
Used time out	-0.01	-0.02	0.04	-0.02	0.01	-0.02	0.03
	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)
% Δ from HSIS Impact		-130	699	-209	182	-141	494
Δ from HSIS Impact (σ)		-0.18	0.95	-0.28	0.25	-0.19	0.67
Spanked child	-0.10	0.03†	0.01	0.00	0.01	0.02	0.00
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
% Δ from HSIS Impact		130	112	103	110	122	105
Δ from HSIS Impact (σ)		2.43	2.09	1.92	2.05	2.29	1.96
F-test		0.15	0.43	0.36	0.37	0.18	0.29

Table 6a: Regression based estimators for **3 year olds cont'd**

	HSIS Impact	ECLS-B Non-Experimental Head Start Impact Estimates					
		Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Labor							
Household income	-485	-9589†	-4859†	-6602†	-4866†	-7123†	-4881†
	(1369)	(1254)	(1041)	(1303)	(1120)	(1193)	(1121)
% Δ from HSIS Impact		-1877	-902	-1261	-903	-1369	-906
Δ from HSIS Impact (σ)		-6.65	-3.19	-4.47	-3.20	-4.85	-3.21
Mother works	-0.03	-0.02	0.02	0.09	0.07	0.02	0.04
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
% Δ from HSIS Impact		7	163	429	372	183	237
Δ from HSIS Impact (σ)		0.03	0.79	2.09	1.81	0.89	1.15
Father works (if present)	0.04	-0.07	-0.07	-0.10	-0.14†	-0.08	-0.10
	(0.06)	(0.04)	(0.05)	(0.05)	(0.07)	(0.05)	(0.06)
% Δ from HSIS Impact		-251	-269	-336	-415	-282	-325
Δ from HSIS Impact (σ)		-1.81	-1.94	-2.43	-3.00	-2.04	-2.35
F-test		0.05	0.36	0.12	0.24	0.12	0.33
F-test overall		0.02	0.58	0.09	0.18	0.05	0.23

Table 6b: Regression based estimators for **4 year olds**

		ECLS-B Non-Experimental Head Start Impact Estimates					
	HSIS Impact	Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Cognitive							
Reading percentile	7.23*	-2.08†	4.28	5.75	6.26	2.57	5.80
	(2.18)	(2.93)	(3.18)	(2.78)	(2.77)	(2.58)	(2.81)
% Δ from HSIS Impact		-129	-41	-20	-13	-64	-20
Δ from HSIS Impact (σ)		-4.26	-1.35	-0.68	-0.44	-2.13	-0.66
Mathematics percentile	4.32	-3.62†	1.83	4.12	4.53	1.72	3.91
	(2.54)	(2.52)	(2.86)	(2.33)	(2.59)	(1.95)	(2.31)
% Δ from HSIS Impact		-184	-58	-5	5	-60	-10
Δ from HSIS Impact (σ)		-3.12	-0.98	-0.08	0.08	-1.02	-0.16
F-test		0.41	1.00	1.00	1.00	0.90	1.00
Health							
Child health good/excellent	0.00	-0.00	0.01	0.05	0.04	0.01	0.02
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
% Δ from HSIS Impact		-172	270	2209	2041	549	957
Δ from HSIS Impact (σ)		-0.07	0.12	0.96	0.89	0.24	0.42
Health insurance	0.01	0.03	0.07	-0.01	0.04	0.02	0.06
	(0.03)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)
% Δ from HSIS Impact		130	450	-176	210	33	334
Δ from HSIS Impact (σ)		0.50	1.72	-0.67	0.80	0.12	1.27
F-test		1.00	0.81	1.00	1.00	1.00	0.94

Table 6b: Regression based estimators for **4 year olds cont'd**

	HSIS Impact	ECLS-B Non-Experimental Head Start Impact Estimates					
		Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Parenting							
Read to Child	0.07	0.01	0.05	-0.01	-0.06	0.00	-0.01
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
% Δ from HSIS Impact		-81	-26	-115	-185	-95	-116
Δ from HSIS Impact (σ)		-1.03	-0.33	-1.46	-2.35	-1.21	-1.48
Parental safety practices	0.01	-0.01	-0.02	-0.09	-0.12	-0.05	-0.10
	(0.05)	(0.07)	(0.08)	(0.07)	(0.10)	(0.06)	(0.08)
% Δ from HSIS Impact		-184	-252	-812	-1051	-494	-868
Δ from HSIS Impact (σ)		-0.53	-0.72	-2.32	-3.00	-1.41	-2.48
Used time out	-0.10	0.07†	0.07†	0.07†	-0.00	0.07†	0.03
	(0.06)	(0.05)	(0.05)	(0.06)	(0.07)	(0.04)	(0.05)
% Δ from HSIS Impact		164	164	165	99	164	128
Δ from HSIS Impact (σ)		3.05	3.04	3.06	1.85	3.05	2.38
Spanked child	-0.08	0.04	0.03	0.05	0.08	0.05	0.06
	(0.06)	(0.05)	(0.06)	(0.04)	(0.06)	(0.04)	(0.05)
% Δ from HSIS Impact		157	135	172	201	162	176
Δ from HSIS Impact (σ)		2.01	1.73	2.20	2.57	2.08	2.25
F-test		0.70	0.85	0.59	0.58	0.58	0.70

Table 6b: Regression based estimators for **4 year olds cont'd**

		ECLS-B Non-Experimental Head Start Impact Estimates					
	HSIS Impact	Mean difference	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Unrestricted difference-in-differences	Unrestricted difference-in-differences with controls
Labor							
Household income	501	-6519†	-1987	-5901†	-2502	-5831†	-2545
	(1329)	(2095)	(1576)	(2376)	(1673)	(2008)	(1652)
% Δ from HSIS Impact		-1401	-497	-1277	-599	-1263	-608
Δ from HSIS Impact (σ)		-5.28	-1.87	-4.82	-2.26	-4.77	-2.29
Mother works	0.00	-0.06	-0.01	0.11	0.05	0.03	-0.00
	(0.06)	(0.06)	(0.07)	(0.06)	(0.08)	(0.05)	(0.07)
% Δ from HSIS Impact		-1589	-312	2541	1144	571	-109
Δ from HSIS Impact (σ)		-1.13	-0.22	1.80	0.81	0.40	-0.08
Father works (if present)	0.08	-0.03	-0.02	0.08	-0.00	0.02	-0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.03)	(0.04)
% Δ from HSIS Impact		-135	-130	7	-104	-73	-142
Δ from HSIS Impact (σ)		-1.98	-1.91	0.11	-1.53	-1.07	-2.08
F-test		0.30	0.95	0.46	0.87	0.43	0.80
F-test overall		0.51	0.99	0.90	0.97	0.85	0.97

Table 7a: Probit propensity score model
Dependent variable: Head Start participation

	Coefficients	Marginal effects
Child female	-0.0856	-0.0185
	(0.080)	(0.017)
Child black	0.1900	0.0410
	(0.160)	(0.034)
Child Hispanic	-0.0078	-0.0017
	(0.161)	(0.035)
Child white	-0.3277*	-0.0707*
	(0.164)	(0.035)
Child age baseline	-0.0065	-0.0014
	(0.020)	(0.004)
Child age at assessment	0.0065	0.0014
	(0.021)	(0.005)
Mother years education	0.0023	0.0005
	(0.018)	(0.004)
Age mother	0.0054	0.0012
	(0.008)	(0.002)
Married	-0.4222*	-0.0911*
	(0.250)	(0.054)
Separated	-0.3842	-0.0829
	(0.318)	(0.069)
Divorced	-0.4472	-0.0965
	(0.290)	(0.063)
Never Married	-0.4432*	-0.0956*
	(0.251)	(0.054)
Teenage mother	-0.0516	-0.0111
	(0.253)	(0.055)

Table 7a: Probit propensity score model **cont'd**
Dependent variable: Head Start participation

Number children under age 6	0.0487	0.0105
	(0.048)	(0.010)
Urban area	-0.2798*	-0.0604*
	(0.117)	(0.025)
English primary language	-0.0441	-0.0095
	(0.125)	(0.027)
Own house	-0.7092*	-0.1530*
	(0.419)	(0.090)
Constant	-0.3770	
	(0.866)	
Additional controls	Yes	
N	2850	
N Head Start	450	
N non-Head Start	2400	
Classification Rate	64%	
Balancing tests (# Passed/ # Variables)	18/18	
Percent in Common Support	100%	

Table 7b: Propensity score balancing tests

	p-value
Child female	0.44
Child black	0.11
Child Hispanic	0.96
Child white	0.99
Child age baseline	0.82
Child age at assessment	0.72
Mother years education	0.86
Age mother	0.36
Married	0.73
Separated	0.29
Divorced	0.82
Never married	0.47
Teenage mother	0.37
Number children under age 6	0.61
Urban area	0.10
English primary language	0.94
Own house	0.08
Household income baseline	0.92

Notes: For each variable Z_k, we estimated $Z_k = \beta_0 + \beta_1 \hat{P}(Z) + \beta_2 \hat{P}(Z)^2 + \beta_3 \hat{P}(Z)^3 + \beta_4 \hat{P}(Z)^4 + \beta_5 D\hat{P}(Z) + \beta_6 D\hat{P}(Z)^2 + \beta_7 D\hat{P}(Z)^3 + \beta_8 D\hat{P}(Z)^4 + \eta$, where D is a dummy that equals 1 if the child attended Head Start and 0 otherwise. The p-values are from an F-test on the joint null that the coefficients on the terms interacted with a Head Start dummy are zero.

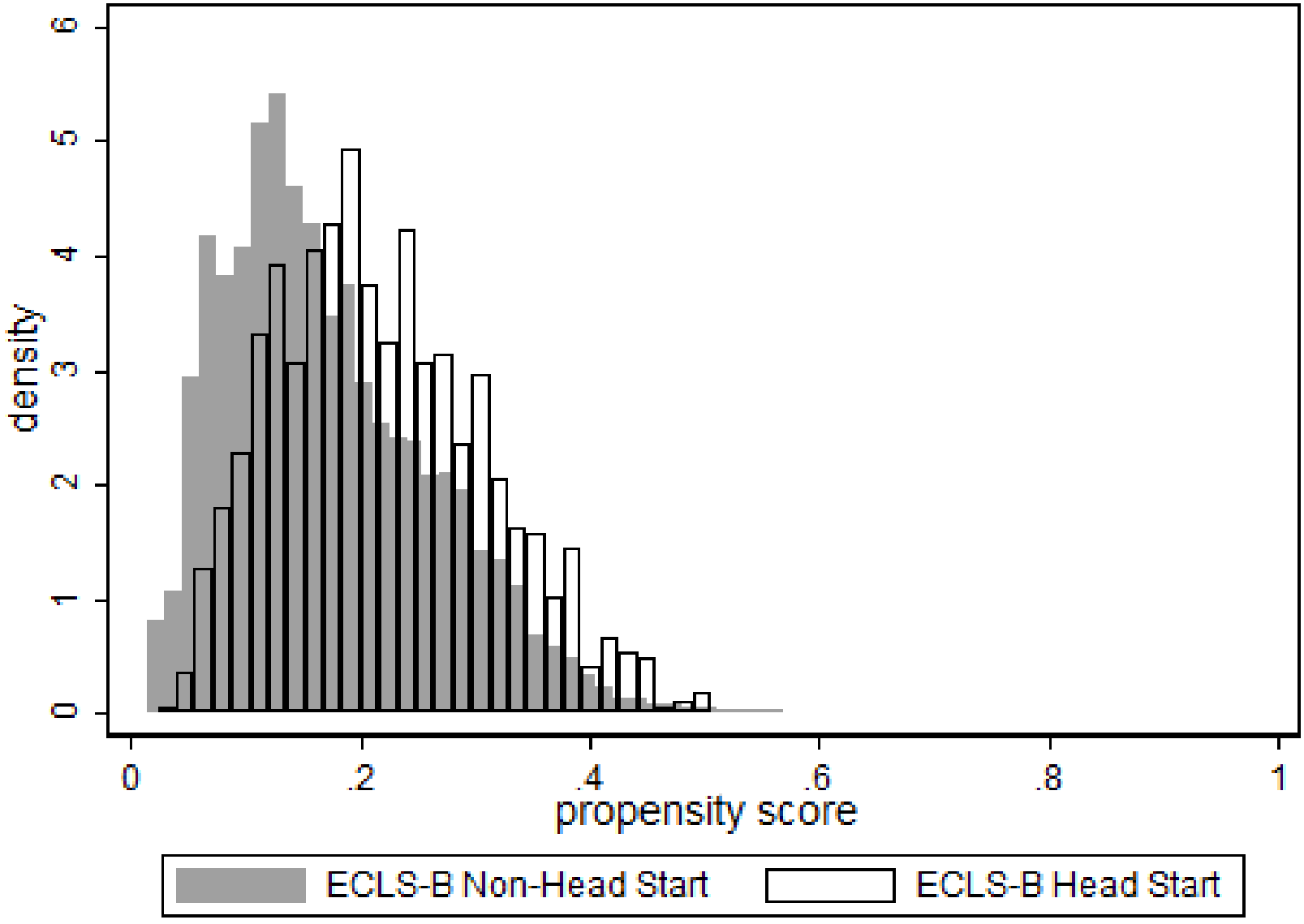

Figure 1: Propensity score histograms for common covariates in HSIS and ECLS-B

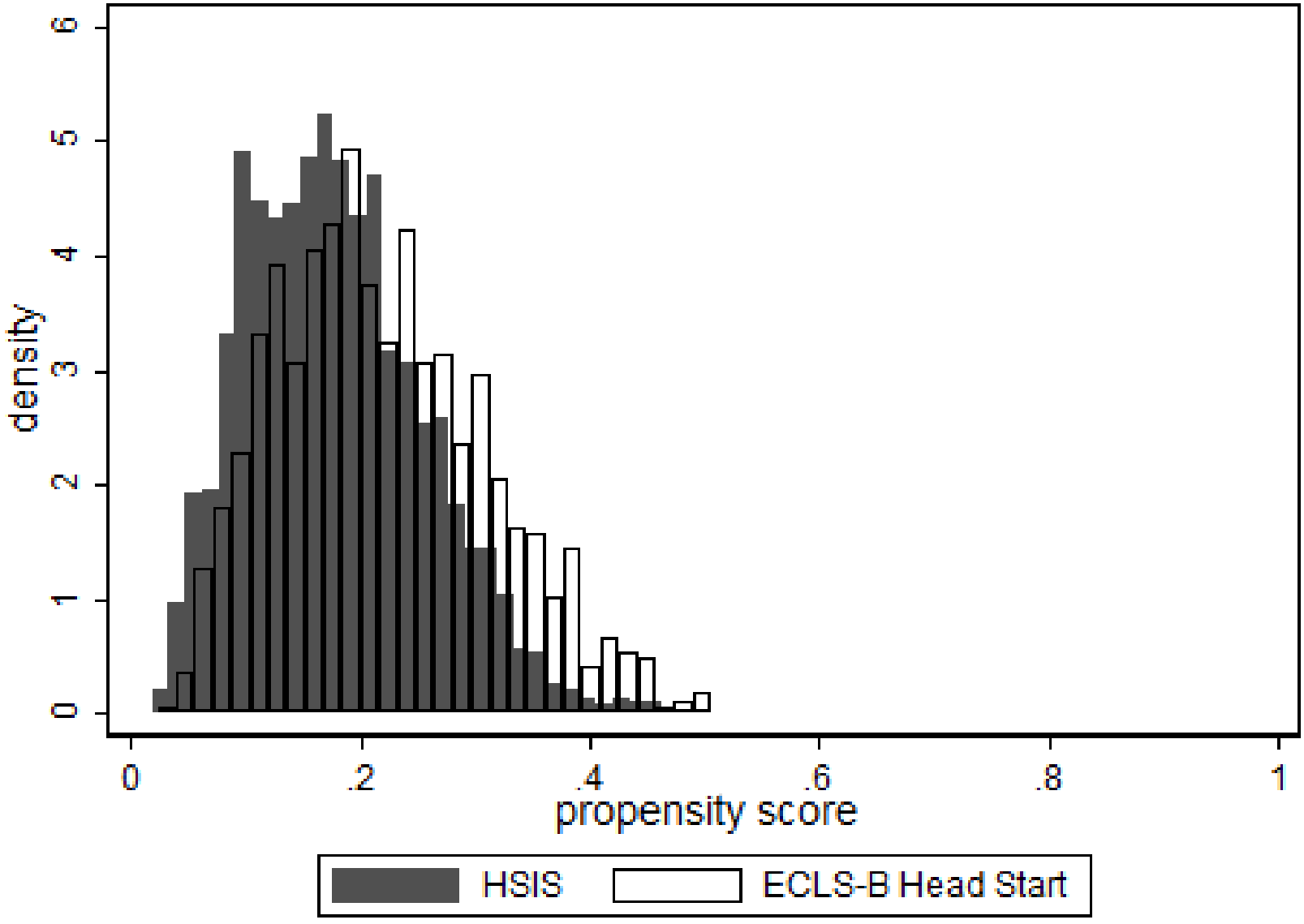

Figure 1: Propensity score histograms for common covariates in HSIS and ECLS-B

Table 8: Cross sectional local linear matching estimators

	3-year-olds Head Start Impact		% Δ from	Δ from	4-year-olds Head Start Impact		% Δ from	Δ from
	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)
Cognitive								
Reading percentile	5.31*	2.92	-45	-1.29	7.23*	-0.14	-102	-3.37
	(1.86)	(2.37)			(2.18)	(3.81)		
Mathematics percentile	0.86	2.07	141	0.38	4.32	-1.73	-140	-2.38
	(3.18)	(2.19)			(2.54)	(3.18)		
F-test	0.93				0.76			
Health								
Child health good/excellent	0.00	0.03	639	0.66	0.00	0.01	521	0.23
	(0.04)	(0.03)			(0.05)	(0.05)		
Health insurance	0.02	0.03	5	0.05	0.01	0.05	279	1.07
	(0.02)	(0.01)			(0.03)	(0.01)		
F-test	0.99				0.99			

Table 8: Cross sectional local linear matching estimators **(cont'd)**

	3-year-olds				4-year-olds			
	Head Start Impact		% Δ from	Δ from	Head Start Impact		% Δ from	Δ from
	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)
Parenting								
Read to child	0.11*	-0.05†	-141	-3.20	0.07	0.05	-22	-0.28
	(0.05)	(0.04)			(0.05)	(0.06)		
Parental safety practices	0.09	-0.06	-175	-2.29	0.01	0.06	356	1.02
	(0.07)	(0.06)			(0.05)	(0.08)		
Used time out	-0.01	0.02	434	0.59	-0.10	0.06†	153	2.85
	(0.05)	(0.03)			(0.06)	(0.05)		
Spanked child	-0.10	0.01	110	2.05	-0.08	0.03	140	1.79
	(0.05)	(0.04)			(0.06)	(0.07)		
F-test	0.36				0.84			
Labor								
Household income	-485	-8558†	-1665	-5.90	501	-3597	-818	-3.08
	(1369)	(1577)			(1329)	(1637)		
Mother works	-0.03	-0.01	71	0.35	0.00	-0.08	-1871	-1.33
	(0.05)	(0.04)			(0.06)	(0.07)		
Father works (if present)	0.04	-0.08	-275	-1.99	0.08	0.02	-76	-1.12
	(0.06)	(0.05)			(0.05)	(0.06)		
F-test	0.12				0.74			
F-test overall	0.19				0.96			

Table 9: Difference-in-difference local linear matching estimators

	3-year-olds Head Start Impact		% Δ from	Δ from	4-year-olds Head Start Impact		% Δ from	Δ from
	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)
Cognitive								
Reading percentile					7.23*	7.35	2	0.05
					(2.18)	(3.01)		
Mathematics percentile					4.32	4.90	13	0.23
					(2.54)	(2.53)		
F-test						0.99		
Health								
Child health good/excellent	0.00	0.00	-45	-0.05	0.00	0.02	754	0.33
	(0.04)	(0.04)			(0.05)	(0.05)		
Health insurance	0.02	-0.01	-127	-1.29	0.01	0.01	0	0.00
	(0.02)	(0.02)			(0.03)	(0.01)		
F-test		0.97				0.99		

Table 9: Difference-in-difference local linear matching estimators **(cont'd)**

	3-year-olds				4-year-olds			
	Head Start Impact		% Δ from	Δ from	Head Start Impact		% Δ from	Δ from
	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)	HSIS	ECLS-B	HSIS Impact	HSIS Impact (σ)
Parenting								
Read to child	0.11*	-0.03†	-127	-2.88	0.07	-0.03	-139	-1.77
	(0.05)	(0.03)			(0.05)	(0.05)		
Parental safety practices	0.09	-0.07	-184	-2.40	0.01	-0.05	-523	-1.49
	(0.07)	(0.05)			(0.05)	(0.08)		
Used time out	-0.01	0.00	138	0.19	-0.10	0.04	136	2.52
	(0.05)	(0.05)			(0.06)	(0.08)		
Spanked child	-0.10	0.01	115	2.14	-0.08	0.08	210	2.69
	(0.05)	(0.05)			(0.06)	(0.06)		
F-test	0.36				0.72			
Labor								
Household income	-485	-4830†	-896	-3.17	501	-2959	-690	-2.60
	(1369)	(1081)			(1329)	(1407)		
Mother works	-0.03	0.06	317	1.54	0.00	0.10	2159	1.53
	(0.05)	(0.05)			(0.06)	(0.08)		
Father works (if present)	0.04	-0.12	-383	-2.77	0.08	0.07	-14	-0.20
	(0.06)	(0.07)			(0.05)	(0.05)		
F-test	0.24				0.79			
F-test overall	0.22				0.99			

Table 10a: Heckman-Hotz tests for regression based estimators

	ECLS-B 3-year-olds	ECLS-B 4-year-olds				
	Regression with controls	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Lagged value-added	Lagged value-added with controls
Cognitive						
Reading percentile		-0.51				
		(1.77)				
Mathematics percentile		-1.18				
		(2.07)				
Health						
Child health good/excellent	0.02	-0.03	-0.06	-0.07	-0.05	-0.04
	(0.02)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)
Health insurance	0.02	0.02	0.05	0.04	0.03	0.02
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)

Table 10a: Heckman-Hotz tests for regression based estimators **(cont'd)**

	ECLS-B 3-year-olds	ECLS-B 4-year-olds				
	Regression with controls	Regression with controls	Difference-in-differences	Differences-in-differences with controls	Lagged value-added	Lagged value-added with controls
Parenting						
Read to child	0.01	0.07	0.03	-0.01	0.03	0.05
	(0.04)	(0.06)	(0.07)	(0.07)	(0.05)	(0.06)
Parental safety practices	0.02	0.04	0.08	0.04	0.08	0.04
	(0.04)	(0.07)	(0.06)	(0.08)	(0.05)	(0.07)
Used time out	0.03	0.04	-0.00	0.03	0.00	0.04
	(0.04)	(0.06)	(0.06)	(0.07)	(0.05)	(0.05)
Spanked child	-0.00	-0.03	0.04	0.04	0.01	-0.01
	(0.04)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)
Labor						
Household income	35	20	4318	1413	2423	47
	(45)	(90)	(2443)	(2119)	(2687)	(91)
Mother works	-0.04	-0.10	-0.14*	-0.12	-0.16*	-0.11
	(0.04)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)
Father works (if present)	0.03	-0.03	-0.02	0.08	-0.04	0.06
	(0.03)	(0.07)	(0.06)	(0.06)	(0.05)	(0.06)

Table 12a: Bias summary statistics by model and domain

Domain	Regression without lagged variables	Regression with lagged variables	Cross section matching	Difference-in-differences matching	Overall
Cognitive	74	25	107	7	52
	0.82	*0.66*	*1.86*	*0.14*	*0.90*
Health	306	494	361	232	411
	0.59	*0.57*	*0.50*	*0.42*	*0.54*
Parenting	210	248	191	196	227
	1.73	*2.15*	*1.76*	*2.01*	*2.01*
Labor (excluding income)	219	454	573	718	475
	1.22	*1.48*	*1.20*	*1.51*	*1.41*
Income	699	1023	1241	793	975
	2.53	*3.73*	*4.49*	*2.89*	*3.55*
Overall	249	394	372	349	362
	1.34	*1.71*	*1.69*	*1.49*	*1.62*
Only models passed Heckman-Hotz					
Cognitive	49				49
	1.16				*1.16*
Health	306	814	580	377	603
	0.59	*0.67*	*0.44*	*0.16*	*0.56*
Parenting	210	313	191	252	256
	1.73	*2.20*	*1.76*	*2.12*	*1.99*
Labor (excluding income)	219	263	141	1086	337
	1.22	*0.95*	*1.15*	*0.87*	*1.05*
Labor (excluding income)	699	937	818	690	832
	2.53	*3.53*	*3.08*	*2.60*	*3.11*
Overall	264	495	281	514	398
	1.42	*1.78*	*1.54*	*1.46*	*1.60*

Conclusion

- Some of the methods for some of the outcomes reproduce the experimental results fairly closely, but a priori it would be difficult to know whether the estimator would work well for any particular outcome.
- Pre-program exogeneity tests (Heckman-Hotz tests) were not found to be very discriminating in isolating the best performing estimators/outcomes.
- Estimated bias varies substantially across outcome measures, even more so than across methods. Outcomes, such as child test scores, tend to have smaller biases, regardless of the estimation method, but other outcomes, such as household income, exhibit consistently large biases.

Conclusion (cont'd)

- The difference-in-difference regression and matching estimators tend to exhibit lower biases than the cross-section estimators.
- The literature thus far has focused on the question of which estimation methods are more reliable, but the context in which the program operates and the specific outcome measures of interest are also important determinants of estimator performance.