Estimating Dynamic Treatment Effects from Project STAR

Weili Ding
Queen's University

Steven F. Lehrer
Queen's University and NBER

October 2006

First Version: November 2003

MOTIVATION

Program Evaluation: Non-Compliance of Human Subjects
Multi-Period Nature of the Interventions
Past Studies treat non-compliance as random and ignore multi-period nature of the experimental design

Economics: Structural Parameters of the Education Production Functions, Flexible Estimation Approaches
Relationship between Economic Estimates and Dynamic Treatment Parameters in Policy Evaluation

Public Policy: Multi-Billion Dollar Class Size Reduction Initiatives Throughout North America Over the Past Decade
Incredible Attention Paid to Project STAR in Policy Debates

POLICY RESPONSE TO PROJECT STAR

United States Department of Education (1998)
“In sum, due to the magnitude of the Project STAR longitudinal experiment, the design, and the care with which it was executed, the results are clear: This research leaves no doubt that small classes have an advantage over larger classes in student performance in the early primary grades”

The Manitoba Teachers' Society: Written Submission to Class Size and Composition Commission
“Tennessee's Project STAR the most reputable and frequently cited study of class size”; “Nevertheless, the "tipping point" does seem to be between 19 and 20 students”

Frederick Mosteller (1995), the STAR Project is considered "*one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools*."

Highly Politicized Area of Debate

PAST FINDINGS

Jeremy Finn and C.M. Achilles (Fall 1990) state "*This research leaves no doubt that small classes have an advantage over larger classes in reading and math in early primary grades*." "*The effects of reduced-size classes were found on every achievement measure administered in Project STAR* ".

Krueger (1999) finds ("*smaller overall effects*") "*test score ad-vantage of students in small classes expands by 1% per year in subsequent years*"

Hanushek (1999) "*one would expect the differences (between class types) to get wider through the grades as they continue to get more resources (treatment)*"

Finn and Achilles (1999) recast STAR results using grade equivalent scores as achievement outcomes to show effects increase with each grade.

What's the nature and magnitude of small class effect?
Is the effect persistent?
Is non-compliance a problem with STAR data?

PROJECT STAR

The Student Teacher Achievement Ratio (STAR) project was a large-scale, four-year, experimental study of reduced class size

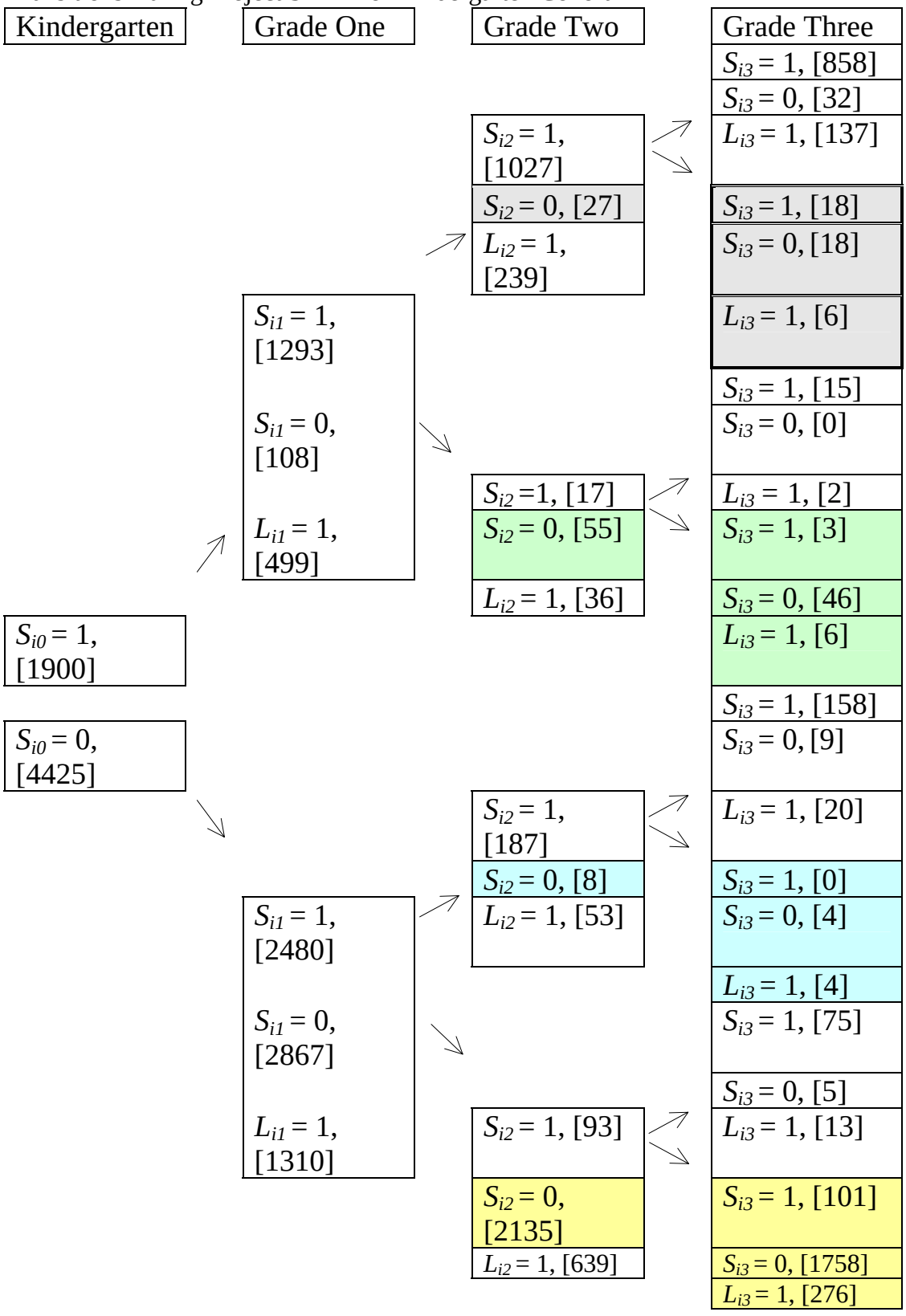
Over 6,000 students in 79 schools upon entering kindergarten were randomly assigned into one of three interventions: small class, regular class and regular-with-aide class.
Classroom teachers were also randomly assigned to classes.

The interventions were initiated as the students entered school in kindergarten and continued through the third grade.

Random assignment is designed to circumvent selection in treatment.

A variety of implementation problems came up in those four years.

Figure 1: Transitions During Project STAR for Kindergarten Cohort



STAR IMPLEMENTATION PROBLEMS

Only 45.16% of the kindergarten sample initially assigned to small classes conformed to their assignments throughout all four years of the STAR experiment.

Only 39.73% of those initially assigned to regular classes conformed.

Refreshment samples *NOT* randomly assigned w.r.t. class type.

Subjects exit the sample differently with regard to initial random assignment.

Coefficient 0.065 (0.030)

Simple test: Are individuals who subsequently leave the STAR experiment systematically different from those who remain in terms of initial behavioral relationships?

$$A_{ijK} = \alpha_x' X_{ijK} + \alpha_l' L_{ij} X_{ijK} + v_j + \varepsilon_{ijK}$$

Results suggest attrition is non-random. (Table 1)

Table 1: Are Attritors Different from Non-attritors

Subject Area	Mathematics	Reading	Word Recognition
Kindergarten Class Type	10.434	6.513	7.370
	(2.332)	(1.440)	(1.628)
White or Asian Student	20.499	8.608	8.505
	(2.760)	(2.005)	(2.524)
Female Student	2.587	3.349	2.488
	(1.363)	(1.074)	(1.296)
Student on Free lunch	-13.729	-12.239	-13.916
	(1.679)	(1.187)	(1.480)
Years of Teaching Experience	0.323	0.255	0.329
	(0.220)	(0.123)	(0.135)
White Teacher	-.926	-1.577	-1.578
	(4.366)	(3.068)	(3.506)
Teacher has Master Degree	-1.482	-1.211	-0.491
	(2.396)	(1.423)	(1.729)
Attrition Indicator	-17.305	-13.674	-13.198
	(3.838)	(2.537)	(3.251)
Attrition Indicator Interacted with	-5.383	-2.069	-3.004
Kindergarten Class Type	(2.616)	(1.686)	(2.045)
Attrition Indicator Interacted with	-3.949	-.259	-1.177
White or Asian Student	(2.732)	(1.824)	(2.368)
Attrition Indicator Interacted with	5.597	2.943	3.750
Female Student	(2.078)	(1.454)	(1.739)
Attrition Indicator Interacted with	-5.186	-0.496	0.549
Student on Free lunch	(2.384)	(1.554)	(1.891)
Attrition Indicator Interacted with	0.188	0.075	-0.060
Years of Teaching Experience	(0.210)	(0.131)	(0.164)
Attrition Indicator Interacted with	1.263	2.269	0.642
White Teacher	(3.490)	(2.133)	(2.678)
Attrition Indicator Interacted with	-1.370	0.939	1.552
Teacher has Master Degree	(2.490)	(1.586)	(1.876)
Number of Observations (R-Squared)	5810 (0.305)	5729 (0.295)	5789 (0.259)
Joint Effect of Attrition on Constant	42.39	32.68	25.76
and Coefficient Estimates	[0.000]	[0.000]	[0.000]
Joint Effect of Attrition on all	3.14	1.23	1.45
Coefficient Estimates but not constant	[0.003]	[0.280]	[0.181]
Effect of Attrition	20.33	29.06	16.48
on Constant Alone	[0.000]	[0.000]	[0.000]

Note:Regressions include school indicators. Standard errors corrected at the classroom level are in ( ) parentheses. Probability > F are in [ ] parentheses.

THE BASIC CASE

Define M = 1 Initially assigned to *small class*
M = 0 Initially assigned to regular class

$S_t = 1$ Attending *small class* in period *t*
$S_t = 0$ Attending regular class in period *t*

A_t is Achievement at the end of period *t*.

$$A_t = S_t A_{1t} + (1 - S_t) A_{0t}$$

GOAL: Estimate $ATE = E(A_{1t} - A_{0t})$: Causal Effect of Treatment.

Sometimes we are interested in $ITT = \overline{A}_{M=1} - \overline{A}_{M=0}$: Causal Effect of Treatment Assignment (not proper for mandatory policy like class size).

NON-COMPLIANCE

Dropout Bias M=1, S=0

Substitution Bias M=0, S=1

Under dropout and substitution, we can use initial random assignments as instrument to estimate Local Average Treatment Effects (LATE or CACE; Angrist, Imbens and Rubin (1996))

$$ITT^{IV} = \overline{A}^{C}_{M=1} - \overline{A}^{C}_{M=0} = \frac{\overline{A}_{M=1} - \overline{A}_{M=0}}{\Pr(S_t = 1 \mid M = 1) - \Pr(S_t = 1 \mid M = 0)}$$

Note that in general LATE ≠ ATE

ATTRITION

Define L_{t+1}=1 Leave the STAR sample at the end of year *t* L_{t+1}=0
Remain in the STAR sample at end of year *t*

Frangakis and Rubin (1999) demonstrate that neither ITT or LATE (CACE) are robust to attrition bias.

Attrition Bias

- -Selection Due to Observables
- -Selection Due to Unobservables

Balke and Pearl (97) demonstrate that in the face of imperfect compliance traditional estimates are potentially misleading as they may lie entirely outside the bounds for an average causal effect of the intervention.

Construct bounds using 2 approaches Horowitz and Manski (2000) and Lee (2005)

Table 3: Traditional Single Period Causal Estimates of The Impacts of Reduced Class Size

Method	Mathematics			Reading			Word Comprehension		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
ITT Analysis	9.023 (1.946)	5.389 (2.48)	4.329 (2.048)	10.409 (2.288)	4.698 (2.316)	7.189 (1.917)	9.036 (2.377)	4.622 (2.314)	8.127 (2.45)
IV Analysis (LATE)	10.353 (2.231)	6.417 (2.953)	5.781 (2.035)	11.948 (2.625)	5.602 (2.536)	9.584 (2.633)	10.296 (2.696)	5.513 (2.748)	10.862 (3.351)
Attrition Test for ITT	61.60 [0.00]	79.89 [0.00]	86.12 [0.00]	62.21 [0.00]	89.73 [0.00]	78.22 [0.00]	36.55 [0.00]	73.2 1 [0.00]	63.19 [0.00]
Attrition Test for IV	26.93 [0.00]	22.43 [0.00]	22.11 [0.00]	22.21 [0.00]	25.69 [0.00]	18.90 [0.00]	14.22 [0.00]	18.93 [0.00]	13.41 [0.00]
Horowitz-Manski Bounds	{-73.174, 86.647}	{-111.487, 117.794	{-144.049, 149.045}	{-67.039, 81.720}	{-114.397, 120.315}	{-130.950, 135.940}	{-102.133, 119.564}	{-97.005, 105.765}	{-112.028, 118.986}
Horowitz-Manski Bounds with school covariates	{-93.935, 94.805}	{-87.347 87.861}	{-97.246, 98.414}	{-82.452, 83.447}	{-89.447, 90.532}	{-87.716, 88.713}	{-98.434, 98.632}	{-78.441, 78.359}	{-77.924, 79.558}
Lee Bounds with no covariates	{5.844, 11.126	{-3.010 6.502}	{-3.296, 6.005}	{5.443, 12.152}	{-2.575, 6.387}	{-0.834, 7.627}	{5.471, 11.407}	{-1.097, 7.110}	{-2.523, 8.547}
Lee Bounds with school covariates	{0.487, 20.992}	{-6.782 20.011}	{-8.055, 17.619}	{-2.024, 24.360}	{-8.177 18.779}	{-6.286, 19.227}	{-4.357, 26.532}	{-9.301, 21.683}	{-8.796, 22.948}
ITT Ignoring Selective Attrition	9.297 (1.894)	5.554 (2.070)	4.034 (1.637)	10.659 (2.141)	5.488 (1.86)	5.437 (1.545)	9.737 (2.283)	5.649 (2.004)	6.436 (1.922)
IV Ignoring Selective Attrition	10.852 (1.449)	6.707 (1.854)	5.642 (2.035)	12.45 (1.87)	6.783 (1.837)	7.56 (1.959)	11.323 (1.951)	6.918 (2.081)	9.004 (2.352)

Note: All of the ITT and IV estimates are statistically significant at the 5% level. The IV and ITT analyses include the full history of teacher inputs, free lunch status, race, gender and school indicators. Standard errors corrected at the classroom level are in parentheses. For the specification tests the Probability that the Null is rejected is contained in [] brackets. For all the bounds on the ATE analysis {lower bound, upper bound}.

OVERVIEW OF ESTIMATION STRATEGY

More generally,

$$A_{ij2} = S_{i1}S_{i2}A_i^{11} + (1-S_{i1})S_{i2}A_i^{01} + S_{i1}(1-S_{i2})A_i^{10} + (1-S_{i1})(1-S_{i2})A_i^{00}$$

Linearized version of Education Production Function

$$A_{i2} = \beta_{x2}'X_{i2} + \beta_{x1}'X_{i1} + \beta_{S2}'S_{i2} + \beta_{S1}'S_{i1} + \beta_{S12}'S_{i2}S_{i1} + v_i + t_2 + \varepsilon_{i2}$$

$$A_{i1} = \alpha_{x1}'X_{i1} + \alpha_{S1}'S_{i1} + v_i + \varepsilon_{i1}$$

Remove effect of v_i by taking first differences

$$A_{i2} - A_{i1} = \beta_{x2}'X_{i2} + (\beta_{x1} - \alpha_{x1})'X_{i1} + \beta_{S2}'S_{i2} + (\beta_{S1} - \alpha_{S1})'S_{i1} + \beta_{S12}'S_{i2}S_{i1} + \varepsilon_{i2}^*$$

$$A_{i1} = \alpha_{x1}'X_{i1} + \alpha_{S1}'S_{i1} + \varepsilon_{i1}^*$$

Estimate System of Equations.

INVERSE PROBABILITY WEIGHTING

1) Estimate Attrition Logit

$$\Pr(L_{i2}=0 \mid A_{i1}, X_{i1}\} = 1\{\alpha' Z_{i1} + \omega_{i1} \geq 0\}$$

$$\widehat{p}_{i1} = F(\widehat{\alpha}' Z_{i1})$$

2) Use $\widehat{p}_{i1}$ to reweight data

$$\frac{A_{i2} - A_{i1}}{\widehat{p}_{i1}} = \frac{\beta_{x2}' X_{i2} + (\beta_{x1} - \alpha_{x1})' X_{i1} + \beta_{S2}' S_{i2} + (\beta_{S1} - \alpha_{S1})' S_{i1} + \beta_{S12}' S_{i2} S_{i1} + \varepsilon^*_{i2}}{\widehat{p}_{i1}}$$

$$A_{i1} = \alpha_{x1}' X_{i1} + \alpha_{S1}' S_{i1} + \varepsilon_{i1}^*$$

Use structural parameters to calculate dynamic treatment effects

$$T^{(1,1)(0,0)}(1,1) = \beta_{S2} + \beta_{S1} + \beta_{S12}$$

$$T^{(1,1)(1,0)}(1,1) = \beta_{S2} + \beta_{S12}$$

$$T^{(0,1)(0,0)}(0,1) = \beta_{S2}$$

DETAILS

Miquel (2003) proves that the full sequence of causal effects are estimated under the standard assumptions of

- common trend
- no pretreatment effects
- common support condition

Easy to extend to T periods.

-Assume attrition is an absorbing state

Note we also consider several methods that place bounds on the ATE.

DATA

Use scaled test scores for norm referenced test as outcome measures
Three subjects separately
Only the kindergarten cohort

Table 2: Testing Randomization of Student Characteristics across Class Types

	Kindergarten	Grade One	Grade Two	Grade Three
INCOMING STUDENTS				
White or Asian Student	2.35*10E-4	-0.275	-0.061	7.63*10E-4
	(0.012)	(0.193)	(0.041)	(0.063)
Female Student	0.012	0.199	-0.020	-0.017
	(0.019)	(0.126)	(0.021)	(0.028)
Student on Free lunch	-8.74*10E-3	-0.262	0.013	-0.057
	(0.017)	(0.167)	(0.022)	(0.037)
Joint Test of Student	0.29	1.83	1.24	1.01
Characteristics	[0.831]	[0.150]	[0.301]	[0.392]
Number of Observations	6300	2211	1511	1181
R Squared	0.318	0.360	0.248	0.411
FULL SAMPLE				
White or Asian Student	2.35*10E-4	-0.003	-0.008	-0.021
	(0.012)	(0.021)	(0.025)	(0.027)
Female Student	0.012	0.007	0.004	0.008
	(0.019)	(0.009)	(0.009)	(0.009)
Student on Free lunch	-8.74*10E-3	-0.038	-0.030	-0.044
	(0.017)	(0.016)	(0.016)	(0.016)
Joint Test of Student	0.29	2.05	1.38	2.98
Characteristics	[0.831]	[0.114]	[0.255]	[0.037]
Number of Observations	6300	6623	6415	6500
R Squared	0.318	0.305	0.328	0359

Note:Regressions include school indicators. Standard errors corrected at the school level are in ( ) parentheses. Probability > F are in [ ] parentheses.

Table 1: Summary Statistics

	Kindergarten	Grade One	Grade Two	Grade 3
Class Size	19.914 (3.827)	20.334 (4.017)	20.217 (4.118)	20.400 (4.441)
Receiving Small Class Treatment	0.314 (0.464)	0.350 (0.477)	0.373 (0.484)	0.396 (0.489)
Math Test Score	500.038 (44.979)	545.939 (40.405)	594.427 (43.499)	627.977 (40.181)
Reading Test Score	445.673 (31.438)	541.754 (52.412)	599.326 (43.390)	625.634 (37.125)
Word Recognition Test Score	444.702 (37.295)	532.811 (46.788)	600.021 (47.118)	622.771 (43.932)
Free Lunch Status	0.359 (0.480)	0.371 (0.483)	0.354 (0.478)	0.353 (0.478)
Student is White of Asian	0.753 (0.432)	0.753 (0.432)	0.753 (0.432)	0.753 (0.432)
Student is Female	0.518 (0.500)	0.518 (0.500)	0.518 (0.500)	0.518 (0.500)
Teacher Race is Non-White	0.129 (0.335)	0.140 (0.347)	0.178 (0.383)	0.165 (0.372)
Teacher has a Masters Degree	0.377 (0.485)	0.343 (0.475)	0.363 (0.481)	0.443 (0.497)
Teacher Years of Experience	9.447 (5.497)	11.713 (8.625)	13.076 (8.567)	13.547 (8.471)

Note: Each cell reports the mean and standard deviations are presented in parentheses. There are 2239 students who participated and completed all three exams in each year of the experiment.

RESULTS I

We find that small class attendance is effective in kindergarten.

GRADE ONE (Table 4)

- Small class attendance in kindergarten and grade one are both positive.
- No additional non-linear benefit.
- However the effect of attending small classes in both years is not significantly different from attending in either year only.
- Economic significance of grade one benefit is larger.

RESULTS II

GRADE TWO AND THREE

- No lasting beneficial small class impacts in mathematics, reading and word recognition.
- Grade Two: Positive in Math Only.
- Grade Three: Negative in all Subject Areas

What is going on?

We investigate who gains the most in the classroom.

Weakest Students in Mathematics

Environmental Stability

Increased Variation in Small Classes Relative to Regular Classes

Small Classes do not seem to work unconditionally

Table 4: Structural Estimates of the Treatment Parameters in Education Production Functions

Subject Area	Mathematics	Reading	Word Recognition
Kindergarten			
S_{iK}	8.595 (1.120)***	5.950 (0.802)***	6.342 (0.945)***
Grade One			
S_{iK}	7.909 (4.625)**	8.785 (5.284)**	11.868 (6.722)**
S_{i1}	9.512 (3.307)***	9.315 (4.350)***	15.394 (5.730)***
$S_{iK}S_{i1}$	-6.592 (5.648)	-2.229 (6.992)	-11.060 (8.965)
Grade Two			
S_{iK}	-2.078 (7.276)	11.320 (7.240)	9.959 (8.438)
S_{i1}	-4.010 (3.855)	-20.036 (19.189)	4.298 (7.763)
S_{i2}	15.150 (5.430)***	3.040 (4.428)	0.526 (5.814)
$S_{iK}S_{i1}$	3.851 (11.678)	1.148 (24.059)	-12.074 (17.673)
$S_{iK}S_{i2}$	-4.049 (13.112)	-31.513 (17.366)**	-23.084 (13.237)**
$S_{i1}S_{i2}$	-4.944 (6.617)	25.122 (19.480)	7.868 (8.537)
$S_{iK}S_{i1}S_{i2}$	6.653 (16.067)	23.634 (28.632)	30.111 (19.851)
Grade Three			
S_{iK}	-7.298 (10.901)	1.215 (10.372)	13.071 (12.202)
S_{i1}	43.514 (32.898)	22.083 (30.097)	-6.920 (37.200)
S_{i2}	25.263 (42.080)	-22.085 (26.069)	-25.024 (22.031)
S_{i3}	-6.835 (3.932)**	-10.590 (4.179)***	-12.738 (5.952)***
$S_{iK}S_{i1}$	-38.612 (30.944)	7.978 (39.071)	-18.002 (32.872)
$S_{iK}S_{i2}$	37.355 (28.625)	-42.740 (25.731)**	-2.932 (22.527)
$S_{iK}S_{i3}$	-39.819 (19.922)	17.870 (18.147)	7.328 (14.855)
$S_{i1}S_{i2}$	-61.947 (52.749)	25.388 (35.964)	-7.586 (36.814)
$S_{i1}S_{i3}$	17.163 (43.057)	-6.613 (32.183)	-7.954 (29.718)
$S_{i2}S_{i3}$	-14.366 (42.280)	35.547 (22.836)	29.203 (26.267)
$S_{iK}S_{i1}S_{i3}$	-4.651 (52.881)	-41.180 (43.335)	-14.706 (35.985)
$S_{iK}S_{i1}S_{i2}S_{i3}$	48.084 (48.704)	6.834 (30.521)	14.377 (33.920)

Note: Corrected standard errors in parentheses. The sequences $S_{iK}S_{i1}S_{i2}$, $S_{iK}S_{i2}S_{i3}$ and $S_{i1}S_{i2}S_{i3}$ lack unique support to permit identification in grade 3.

Table 5: Dynamic Average Treatment Effect for the Treated Estimates

Subject Area	Mathematics	Reading	Word Recognition
Kindergarten			
$\tau^{(1)(0)}(1)$	8.595 (1.120)***	5.950 (0.802)***	6.342 (0.945)***
Grade One			
$\tau^{(0,1)(0,0)}(0,1)$	9.512 (3.307)***	9.315 (4.350)***	15.394 (5.730)***
$\tau^{(1,0)(0,0)}(1,0)$	7.909 (4.625)**	8.785 (5.284)**	11.868 (6.722)**
$\tau^{(1,1)(0,0)}(1,1)$	10.829 (8.021)*	15.872 (9.787)*	16.203 (12.587)*
$\tau^{(1,1)(1,0)}(1,1)$	2.920 (6.544)	7.086 (8.235)	4.334 (10.640)
$\tau^{(1,1)(0,1)}(1,1)$	1.317 (7.300)	6.556 (8.764)	0.808 (11.205)
$\tau^{(0,1)(1,0)}(0,1)$	1.603 (5.686)	0.530 (6.844)	4.066 (8.833)
Grade Two			
$\tau^{(0,0,1)(0,0,0)}(0,0,1)$	15.150 (5.430)***	3.040 (4.428)	0.526 (5.814)
$\tau^{(1,0,0)(0,0,0)}(1,0,0)$	-2.078 (7.276)	11.320 (7.240)*	9.959 (8.438)
$\tau^{(1,1,1)(0,0,0)}(1,1,1)$	10.574 (26.606)	12.714 (50.199)	17.603 (33.463)
$\tau^{(1,1,1)(1,0,0)}(1,1,1)$	12.651 (25.589)	1.394 (49.674)	7.644 (32.381)
$\tau^{(1,1,1)(1,1,0)}(1,1,1)$	12.810 (22.436)	20.282 (38.993)	15.421 (25.999)
$\tau^{(0,1,1)(0,0,0)}(0,1,1)$	6.196 (9.400)	8.125 (27.700)	12.691 (12.920)
$\tau^{(0,0,1)(1,0,0)}(0,0,1)$	17.228 (9.084)**	-8.208 (8.490)	-9.433 (10.249)
Grade Three			
$\tau^{(0,0,0,1)(0,0,0,0)}(0,0,0,1)$	-6.835 (3.932)**	-10.590 (4.179)***	-12.738 (5.952)***
$\tau^{(1,1,1,1)(0,0,0,0)}(1,1,1,1)$	-2.148 (129.436)	-17.192 (93.135)	-20.985 (102.228)
$\tau^{(1,1,1,1)(1,1,0,0)}(1,1,1,1)$	0.247 (120.810)	-22.487 (81.117)	-35.114 (85.973)
$\tau^{(1,1,1,1)(1,1,1,0)}(1,1,1,1)$	-0.424 (96.033)	10.115 (63.543)	7.262 (70.360)
$\tau^{(1,1,1,1)(0,1,1,1)}(1,1,1,1)$	-4.940 (86.378)	-20.263 (64.365)	-30.626 (75.468)
$\tau^{(0,1,1,1)(0,0,0,0)}(0,1,1,1)$	2.792 (96.397)	3.071 (67.314)	9.641 (68.958)
$\tau^{(0,0,1,1)(0,0,0,0)}(0,0,1,1)$	4.062 (59.781)	-3.472 (37.243)	-2.215 (32.284)
$\tau^{(0,0,1,1)(1,1,0,0)}(0,0,1,1)$	6.458 (75.714)	-8.767 (59.001)	-16.344 (64.043)

Note: Standard Errors in parentheses.
***,** indicate statistical significance at the 5%, and 10% level respectively

RESULTS III

DuMouchel and Duncan Test confirm accounting for attrition due to observables is crucial.

Likelihood ratio tests reject Null of no selection on unobservables.

ROBUSTNESS CHECK

1) No non-linear effects of treatment
2) Simpler attrition model (less lags, more data)
3) Always versus never
    - No major differences in structural parameters
    - Larger Sample supports limited lasting effect of kindergarten small class attendance to reading or word recognition later on (not strong).
        - Treatment benefits disappear between two prime arms

Table 6: Tests of Weighted versus Unweighted Estimates

Subject Area	Mathematics	Reading	Word Recognition
Grade One	8.74 [0.000]	3.39 [0.000]	1.35 [0.169]
Grade Two	1.48 [0.071]	3.86 [0.000]	2.08 [0.002]
Grade Three	1.72 [0.008]	1.91 [0.002]	1.03 [0.424]

Note: Probability > F are in [ ] parentheses.

Table 7: Likelihood Ratio Tests for the Presence of Selection on Unobservables

Subject Area	Mathematics	Reading	Word Recognition
Grade One	2661.91 [0.000]	4468.98 [0.000]	3293.98 [0.000]
Grade Two	1648.11 [0.000]	1478.86 [0.000]	5480.28 [0.000]
Grade Three	1606.95 [0.000]	1421.94 [0.000]	839.84 [0.000]

Note: Probability > χ^2 are in [ ] parentheses.

Table 8: Structural Estimates of the Treatment Parameters in Education Production Functions using Simpler Attrition Model to Account for Test Completion

Subject Area	Mathematics	Reading	Word Recognition
Kindergarten			
S_{iK}	8.595 (1.120)***	5.950 (0.802)***	6.342 (0.945)***
Grade One			
S_{iK}	12.794 (4.742)***	11.221 (5.088)***	12.580 (5.433)***
S_{i1}	10.322 (2.798)***	4.032 (2.962)	9.282 (3.568)***
$S_{iK}S_{i1}$	-12.748 (5.461)***	-3.164 (5.914)	-10.514 (6.603)
Grade Two			
S_{iK}	8.993 (7.063)	17.40 (8.054)***	-1.690 (4.068)
S_{i1}	-15.755 (11.672)	-37.592 (16.710)***	-23.035 (16.522)
S_{i2}	9.001 (4.839)**	-2.471 (4.4149)	7.278 (8.297)
$S_{iK}S_{i1}$	0.437 (15.122)	-0.044 (22.636)	0.061 (21.173)
$S_{iK}S_{i2}$	-0.933 (8.931)	-19.001 (11.704)	-10.165 (21.262)
$S_{i1}S_{i2}$	14.477 (12.686)	43.044 (17.248)***	29.128 (17.002)**
$S_{iK}S_{i1}S_{i2}$	-7.712 (16.250)	8.050 (24.184)	9.189 (28.858)
Grade Three			
S_{iK}	2.512 (11.252)	12.487 (9.726)	20.241 (11.072)**
S_{i1}	7.347 (11.921)	3.743 (19.584)	3.533 (27.390)
S_{i2}	32.700 (25.589)	-14.059 (11.435)	-16.140 (8.272)**
S_{i3}	-2.991 (3.932)	-3.547 (3.411)	-5.491 (4.815)
$S_{iK}S_{i1}$	-2.424 (19.982)	-14.738 (27.662)	-18.626 (33.645)
$S_{iK}S_{i2}$	42.515 (28.165)	-19.929 (26.944)	-49.423 (35.623)
$S_{iK}S_{i3}$	-9.926 (26.641)	20.363 (23.145)	29.862 (26.369)
$S_{i1}S_{i2}$	-30.957 (29.537)	6.710 (27.010)	-3.718 (36.282)
$S_{i1}S_{i3}$	-34.354 (28.549)	-45.065 (25.648)**	-65.591 (29.914)***
$S_{i2}S_{i3}$	-27.291 (25.802)	13.957 (11.755)	25.368 (9.699)***
$S_{iK}S_{i1}S_{i2}$	-43.321 (34.722)	38.333 (40.920)	94.618 (53.809)**
$S_{i1}S_{i2}S_{i3}$	66.369 (39.566)**	46.807 (31.803)	69.728 (38.514)**
$S_{iK}S_{i1}S_{i2}S_{i3}$	8.646 (28.371)	-34.171 (28.758)	-72.552 (36.493)***

Note: Corrected standard errors in parentheses. The sequences $S_{iK}S_{i1}S_{i3}$ and $S_{iK}S_{i2}S_{i3}$ lack unique support to permit identification in grade 3.

Table 3: Traditional Single Period Causal Estimates of The Impacts of Reduced Class Size

Method	Mathematics			Reading			Word Comprehension		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
ITT Analysis	9.023 (1.946)	5.389 (2.48)	4.329 (2.048)	10.409 (2.288)	4.698 (2.316)	7.189 (1.917)	9.036 (2.377)	4.622 (2.314)	8.127 (2.45)
IV Analysis (LATE)	10.353 (2.231)	6.417 (2.953)	5.781 (2.035)	11.948 (2.625)	5.602 (2.536)	9.584 (2.633)	10.296 (2.696)	5.513 (2.748)	10.862 (3.351)
Attrition Test for ITT	61.60 [0.00]	79.89 [0.00]	86.12 [0.00]	62.21 [0.00]	89.73 [0.00]	78.22 [0.00]	36.55 [0.00]	73.2 1 [0.00]	63.19 [0.00]
Attrition Test for IV	26.93 [0.00]	22.43 [0.00]	22.11 [0.00]	22.21 [0.00]	25.69 [0.00]	18.90 [0.00]	14.22 [0.00]	18.93 [0.00]	13.41 [0.00]
Horowitz-Manski Bounds	{-73.174, 86.647}	{-111.487, 117.794	{-144.049, 149.045}	{-67.039, 81.720}	{-114.397, 120.315}	{-130.950, 135.940}	{-102.133, 119.564}	{-97.005, 105.765}	{-112.028, 118.986}
Horowitz-Manski Bounds with school covariates	{-93.935, 94.805}	{-87.347 87.861}	{-97.246, 98.414}	{-82.452, 83.447}	{-89.447, 90.532}	{-87.716, 88.713}	{-98.434, 98.632}	{-78.441, 78.359}	{-77.924, 79.558}
Lee Bounds with no covariates	{5.844, 11.126	{-3.010 6.502}	{-3.296, 6.005}	{5.443, 12.152}	{-2.575, 6.387}	{-0.834, 7.627}	{5.471, 11.407}	{-1.097, 7.110}	{-2.523, 8.547}
Lee Bounds with school covariates	{0.487, 20.992}	{-6.782 20.011}	{-8.055, 17.619}	{-2.024, 24.360}	{-8.177 18.779}	{-6.286, 19.227}	{-4.357, 26.532}	{-9.301, 21.683}	{-8.796, 22.948}
ITT Ignoring Selective Attrition	9.297 (1.894)	5.554 (2.070)	4.034 (1.637)	10.659 (2.141)	5.488 (1.86)	5.437 (1.545)	9.737 (2.283)	5.649 (2.004)	6.436 (1.922)
IV Ignoring Selective Attrition	10.852 (1.449)	6.707 (1.854)	5.642 (2.035)	12.45 (1.87)	6.783 (1.837)	7.56 (1.959)	11.323 (1.951)	6.918 (2.081)	9.004 (2.352)

Note: The IV and ITT analyses include the full history of teacher inputs, free lunch status, race, gender and school indicators. Standard errors corrected at the classroom level are in parentheses. For the specification tests the Probability that the Null is rejected is contained in [] brackets. For all the bounds on the ATE analysis {lower bound, upper bound}.

Table 8:

Subject Area	Mathematics	Reading	Word Recognition
Grade One			
S_{iK}	4.174 (2.829)	9.351 (2.805)	5.434 (3.250)
S_{i1}	6.608 (2.488)	2.779 (2.582)	6.415 (3.016)
Grade Two			
S_{iK}	6.191 (4.034)	10.479 (4.340)	6.035 (4.659)
S_{i1}	-8.916 (5.191)	-6.529 (5.949)	0.742 (5.784)
S_{i2}	12.805 (4.152)	5.730 (4.659)	4.114 (4.138)
Grade Three			
S_{iK}	0.131 (5.286)	8.885 (5.088)	12.057 (5.940)
S_{i1}	-1.168 (7.588)	-0.057 (7.500)	-5.097 (8.118)
S_{i2}	11.747 (7.162)	3.152 (6.784)	11.079 (7.655)
S_{i3}	-2.596 (3.717)	-1.370 (3.244)	-6.679 (4.691)
DYNAMIC TREATMENT EFFECTS			
Grade One			
$\tau^{(1,1)(0,0)}(1,1)$	10.782 (3.767)	11.933 (3.813)	11.849 (4.434)
$\tau^{(1,1)(1,0)}(1,1)$	6.608 (2.488)	2.779 (2.582)	6.415 (3.016)
$\tau^{(1,1)(0,1)}(1,1)$	4.174 (2.829)	9.351 (2.805)	5.434 (3.250)
$\tau^{(0,1)(1,0)}(0,1)$	2.434 (3.767)	-6.572 (3.813)	0.981 (4.434)
Grade Two			
$\tau^{(0,0,1)(0,0,0)}(0,0,1)$	12.805 (4.152)*	5.730 (4.659)	4.114 (4.138)
$\tau^{(1,0,0)(0,0,0)}(1,0,0)$	6.191 (4.034)	10.479 (4.340)*	6.035 (4.659)
$\tau^{(1,0,0)(0,1,0)}(1,0,0)$	15.107 (6.574)*	8.942 (7.364)	17.154 (7.427)*
$\tau^{(1,1,1)(0,0,0)}(1,1,1)$	10.080 (7.776)	9.680 (8.714)	10.891 (8.502)
$\tau^{(11,1)(1,0,0)}(1,1,1)$	3.889 (6.647)	-0.799 (7.556)	3.372 (7.112)
$\tau^{(0,0,1)(0,1,0)}(0,0,1)$	21.721 (6.647)*	12.259 (7.556)***	4.856 (7.112)
$\tau^{(0,0,1)(1,0,0)}(0,0,1)$	6.614 (5.789)	-4.749 (6.367)	-1.921 (6.231)
Grade Three			
$\tau^{(0,0,0,1)(0,0,0,0)}(0,0,0,1)$	-2.596 (3.717)	-1.370 (3.244)	-6.679 (4.691)
$\tau^{(1,0,0,0)(0,0,0,0)}(1,0,0,0)$	0.131 (5.286)	8.885 (5.088)***	12.057 (5.940)**
$\tau^{(1,1,1,1)(0,0,0,0)}(1,1,1,1)$	8.114 (12.273)	10.580 (11.776)	11.360 (13.483)
$\tau^{(1,1,1,1)(1,1,0,0)}(1,1,1,1)$	9.151 (8.069)	1.782 (7.520)	4.400 (8.798)
$\tau^{(1,1,1,1)(1,1,1,0)}(1,1,1,1)$	-2.596 (3.717)	-1.370 (3.244)	-6.679 (4.691)
$\tau^{(1,1,1,1)(0,1,1,1)}(1,1,1,1)$	0.131 (5.286)	8.885 (5.088)***	12.057 (5.940)**
$\tau^{(0,1,1,1)(0,0,0,0)}(0,1,1,1)$	7.983 (11.076)	1.695 (10.621)	-0.697 (12.104)
$\tau^{(1,1,1,1)(1,0,0,0)}(1,1,1,1)$	7.983 (11.076)	1.695 (10.621)	-0.697 (12.104)
$\tau^{(0,0,1,0)(,0,0,0,0)}(0,0,1,0)$	11.747 (7.162)	3.152 (6.784)	11.079 (7.655)
$\tau^{(1,0,1,0)(0,0,0,0)}(1,0,1,0)$	11.878 (6.426)***	12.037 (6.034)**	23.129 (7.570)*
$\tau^{(1,0,0,0)(0,0,1,0)}(1,0,0,0)$	-11.616 (6.426)***	5.733 (6.034)	0.971 (7.570)

Note: Standard Errors in parentheses.

***,** indicate statistical significance at the 5%, and 10% level respectively

Table 9

	Reading	Math	Word
Weighted Estimates			
Kindergarten	5.95 (1.28)	8.595 (2.025)	6.342 (1.415)
Grade One	-11.642 (14.768)	-4.655 (11.692)	-16.51 (16.263)
Grade Two	2.131 (6.533)	10.2 (6.844)	-1.095 (7.552)
Grade Three	15.231 (7.322)	13.134 (7.882)	5.714 (8.913)
Unweighted Estimates			
Grade One	-0.044 (2.97)	-3.576 (2.369)	-1.770 (3.426)
Grade Two	6.032 (2.258)	7.498 (2.465)	5.969 (2.497)
Grade Three	4.626 (2.624)	2.438 (2.902)	5.868 (3.366)

Note: Each coefficient. Standard errors corrected at the classroom level are in parentheses.

CONCLUSIONS

Multi-period trials with non-compliance are common in both economics and clinical medicine

-Traditional Estimators may not recover parameters of interest

We adopt a simple estimation approach for this setting and apply it to data from Project STAR.

-Small classes does not seem to have Non-Linear Benefit.

-One dose in kindergarten may be all that's needed.

-No significant dynamic benefits in the higher grade.

-Achievement gap closing or reversed in higher grades may be due to a tradeoff between variation in past performance and class size as well as teachers teaching to the bottom, particularly in mathematics.

- More understanding of the trade-off between increased student variability, class size and teaching methods is needed

THE FULL DATA SET SHOULD BE MADE AVAILABLE TO THE GENERAL RESEARCH COMMUNITY (attendance, teaching strategy, group size, teaching time to each group, principal evaluation, teacher evaluation of student classroom behavior, which district 42, teaching attitude etc. observational data to get to the nature of how class size operates)