Precept 11: Unmeasured Confounding

Soc 500: Applied Social Statistics

Simone Zhang

Princeton University

December 2016

Acknowledgments: Many thanks to Ryan Parsons, Xinyi Duan, Ian Lundberg, and Jeremy Cohen for slides on the papers presented here. Material also comes from Matt Blackwell, Adam Glynn, Jens Hainmueller and the lectures.

Today's Agenda

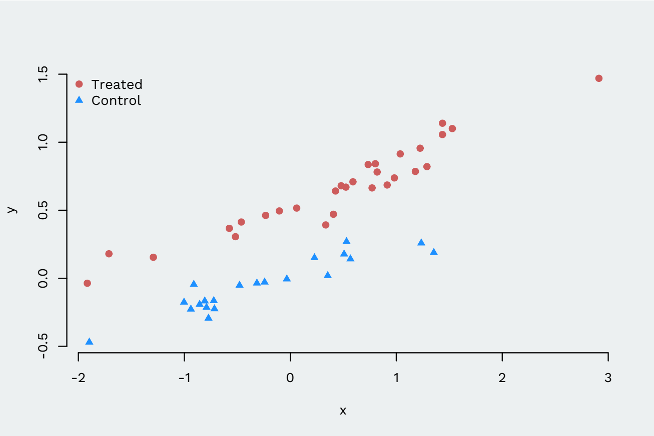
- Imputation Estimator
- Instrumental Variables (+ 3 examples)
- Regression discontinuity

Imputation estimators

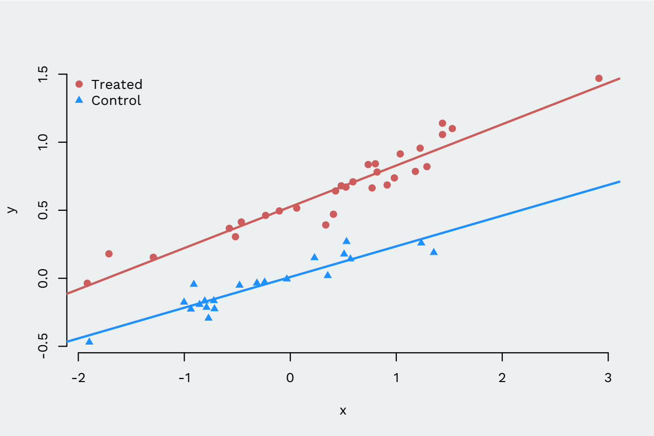
- Allowing for heterogeneous treatment effects
- Impute the treated potential outcomes with $\widehat{Y}_i(1) = \hat{\mu}_1(X_i)$
- Impute the control potential outcomes with $\widehat{Y}_i(0) = \hat{\mu}_0(X_i)$
- Procedure:
  - Regress Y_i on X_i in the treated group and get predicted values for all units (treated or control).
  - Regress Y_i on X_i in the control group and get predicted values for all units (treated or control).
  - Take the average difference between these predicted values.
- Mathematically, it looks like this:

$$\tau_{imp} = \frac{1}{N} \sum_i \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)$$

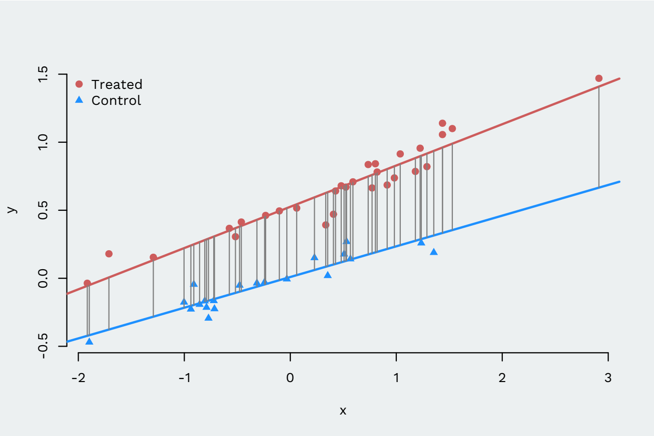
Imputation estimator visualization



Imputation estimator visualization



Imputation estimator visualization



Simple imputation estimator

- Use `predict()` from the within-group models on the data from the entire sample.

R Code

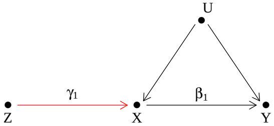
```
## Model on the untreated
model0 <- lm(outcome ~ explan,
 data = subset(my_data, treated == 0))

## Model on the treated
model1 <- lm(outcome ~ explan,
  data = subset(my_data, treated == 1))

## Take the average difference
p1 <- predict(model1, newdata = my_data)
p0 <- predict(model0, newdata = my_data)
mean(p1-p0)
```

Instrumental Variables (IV)

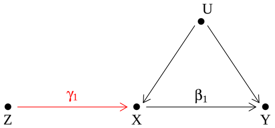
$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$E[U_i|X_i] \neq 0$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$
$$E[U_i|Z_i] = 0$$
$$Cov[X_i, Z_i] \neq 0$$



The IV Estimator

With our assumed model,

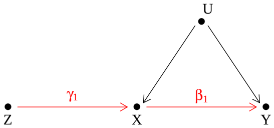
- regressing X on Z identifies γ_1



The IV Estimator

With our assumed model,

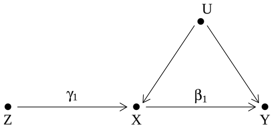
- regressing X on Z identifies γ_1
- regressing Y on Z identifies $\gamma_1 \cdot \beta_1 =$



The IV Estimator

With our assumed model,

- regressing X on Z identifies γ_1
- regressing Y on Z identifies $\gamma_1 \cdot \beta_1 =$
- $\frac{\widehat{\gamma_1 \cdot \beta_1}}{\widehat{\gamma_1}}$ identifies $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$



Review of Key Assumptions

1. Exogeneity of the instrument
2. Exclusion restriction
3. First-stage relationship
4. Monotonicity

Assumption 1: Exogeneity of the Instrument

- Essentially we need the instrument to be randomized:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability. But why believe conditional ignorability for the instrument but not the treatment?

Assumption 2: Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d,1) = Y_i(d,0) \quad \text{for } d = 0,1$$

- Given this exclusion restriction, we know that the potential outcomes for each treatment status only depend on the treatment, not the instrument:

$$Y_i(1) \equiv Y_i(1,1) = Y_i(1,0)$$
$$Y_i(0) \equiv Y_i(0,1) = Y_i(0,0)$$

- Random assignment of the instrument is not sufficient for exclusion
- **NOT A TESTABLE ASSUMPTION**

Assumption 3: First Stage Relationship

- The instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

- Implies that
  - $\text{Cov}(D_i, Z_i) \neq 0$,
  - $0 < P(Z = 1) < 1$
  - $P(D_1 = 1) \neq P(D_0 = 1)$
- This is testable by regressing D on Z
- Note that even a weak instrument can induce a lot of bias. Thus, for practical sample sizes you need a **strong** first stage effect.

Assumption 4: Monotonicity

- To allow for heterogenous effects we need to make a new assumption about the relationship between the instrument and the treatment.
- **Monotonicity** says that the presence of the instrument **never dissuades** someone from taking the treatment:

$$D_i(1) - D_i(0) \geq 0$$

- Note if this holds in the opposite direction $D_i(1) - D_i(0) \leq 0$, we can always rescale D_i to make the assumption hold.

Monotonicity means no defiers

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- We sometimes call assumption 4 **no defiers** because the monotonicity assumption rules out the existence of defiers.
- This means we can now sometimes identify the subgroup
- Anyone with $D_i = 1$ when $Z_i = 0$ must be an **always-taker** and anyone with $D_i = 0$ when $Z_i = 1$ must be a **never-taker**.

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Based on these IV assumptions we can identify three effects:

1. The **first stage effect**: Effect of Z **on** D.
2. **Reduced form** or **intent-to-treat** effect: Effect of Z **on** Y.
3. The **instrumental variable** treatment effect: Effect of D **on** Y, using only the exogenous variation in D that is induced by Z.

First Stage Effect in JTPA

First stage effect: Z on D: $\hat{\pi}_1 = \frac{Cov[D,Z]}{V[Z]}$

R Code
```
> cov(d[,c("earnings","training","assignmt")])
             earnings    training    assignmt
earnings 2.811338e+08 685.5254685 257.0625061
training 6.855255e+02   0.2456123   0.1390407
assignmt 2.570625e+02   0.1390407   0.221713
```

R Code
```
> 0.1390407/0.2217139
[1] 0.6271177
```

First Stage Effect in JTPA

R Code

```
> summary(lm(training~assignmt,data=d))

Call:
lm(formula = training ~ assignmt, data = d)

Residuals:
     Min       1Q   Median       3Q      Max
-0.64165 -0.01453 -0.01453  0.35835  0.98547

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.014528   0.006529   2.225   0.0261 *
assignmt    0.627118   0.007987  78.522   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 0.398 on 11202 degrees of freedom
Multiple R-squared: 0.355,        Adjusted R-squared: 0.355
F-statistic:  6166 on 1 and 11202 DF,  p-value: < 2.2e-1
```

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$
\begin{aligned}
Y &= \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2 \\
Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\
Y &= \gamma_0 + \gamma_1 Z + u_3
\end{aligned}
$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$.

Reduced Form/Intent-to-treat Effect

R Code

```
> summary(lm(earnings~assignmt,data=d))

Call:
lm(formula = earnings ~ assignmt, data = d)

Residuals:
   Min     1Q Median     3Q    Max
-16200 -13803  -4817   8950 139560

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15040.5      274.9  54.716  < 2e-16 ***
assignmt      1159.4      336.3   3.448 0.000567 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 16760 on 11202 degrees of freedom
Multiple R-squared: 0.00106,        Adjusted R-squared: 0.000971
F-statistic: 11.89 on 1 and 11202 DF,  p-value: 0.000566
```

Instrumental Variable Effect: Wald Estimator

Instrumental Variable Effect: $\alpha_1 = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]}{Cov[D,Z]}$

R Code

```
> cov(d[,c("earnings","training","assignmt")])
              earnings    training    assignmt
earnings 2.811338e+08 685.5254685 257.0625061
training 6.855255e+02   0.2456123   0.1390407
assignmt 2.570625e+02   0.1390407   0.221713
```

R Code

```
> 257.0625061/0.1390407
[1] 1848.829
```

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\ Y &= \alpha_0 + \alpha_1 (\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \\ Y &= (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1 (\hat{\pi}_1 Z) + u_2 \end{aligned}$$

- Intuition: Retain only variation in D that is induced by Z, "purged" of endogeneity

Instrumental Variable Effect: Two Stage Least Squares

Point estimates from 2nd regression are equivalent to IV estimator, but the SEs are not quite correct since they ignore the estimation uncertainty in $\hat{\pi}_0$ and $\hat{\pi}_1$. The following function corrects for that:

R Code

```
> training_hat <- lm(training~assignmt,data=d)$fitted
> summary(lm(earnings~training_hat,data=d))

Call:
lm(formula = earnings ~ training_hat, data = d)

Residuals:
   Min     1Q Median     3Q    Max
-16200 -13803  -4817   8950 139560

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   15013.6      281.3  53.375  < 2e-16 ***
training_hat   1848.8      536.2   3.448 0.000567 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 16760 on 11202 degrees of freedom
Multiple R-squared: 0.00106,       Adjusted R-squared: 0.000971
F-statistic: 11.89 on 1 and 11202 DF,  p-value: 0.0005669
```

Instrumental Variable Effect: Two Stage Least Squares

R Code

```
> library(AER)
> summary(ivreg(earnings ~ training | assignmt,data = d))
Call:
ivreg(formula = earnings ~ training | assignmt, data = d)
Residuals:
   Min     1Q Median     3Q    Max
-16862 -13716  -4943   8834 140746
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15013.6      280.6  53.508  < 2e-16 ***
training      1848.8      534.9   3.457 0.000549 ***
---
Residual standard error: 16720 on 11202 degrees of freedom
Multiple R-Squared: 0.00603,        Adjusted R-squared: 0.005941
Wald test: 11.95 on 1 and 11202 DF,  p-value: 0.0005491
```

Vietnam Draft Lottery

Angrist, Joshua. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery." *American Economic Review* 80: 313-336.

Background

Key question:

Does military service cause changes in earnings?

Problem with prior research

Unobserved factors might affect both military service and earnings.

$$\begin{tikzcd} & S \arrow[r] & Y \\ U \arrow[ur] \arrow[urr, bend right] & & \end{tikzcd}$$

Instrumental variables approach

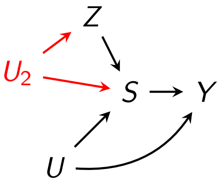
Find Z which affects earnings Y *only* through its affect on military service T.

$$
\begin{tikzcd}
Z \arrow[dr] & & \\
& S \arrow[r] & Y \\
U \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

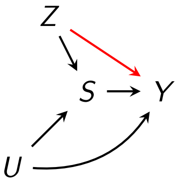
Proposed instrument Z: Vietnam draft lottery numbers

Assumptions

Exogeneity ✓



Exclusion restriction



Assumptions

- **First-stage relationship** - Does the draft actually induce people to serve in the military?
- **Monotonicity** - Who would the defiers be?

The Vietnam draft

- Televised drawing of Random Sequence Numbers (1-365) which assigned draft priorities to birth dates
- Ceiling set so only those below the ceiling were drafted
- Men were drafted in 1970-1972
- Ceilings were 195 in 1970, 125 in 1971, and 95 in 1972

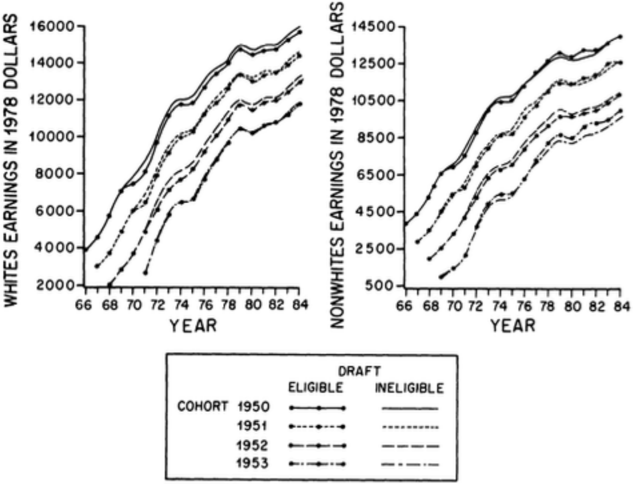
What it looked like

Rep. Alexander Pirnie, R-NY, draws the first capsule in the lottery drawing held on December 1, 1969.

Variables

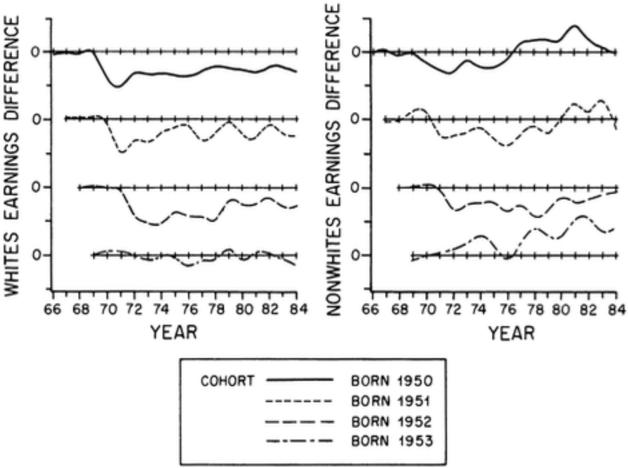
- Outcome 1: Social Security Continuous Work History Sample (CWHS)
  - 1% sample of population
  - 1964-1984
  - Only Social Security earnings, up to taxable maximum
- Outcome 2: IRS total compensation
  - Aggregated within cells defined by year of earnings, year of birth, race, and five consecutive lottery numbers
  - 1978 on

Descriptive evidence



Peeling apart following the draft

Descriptive evidence

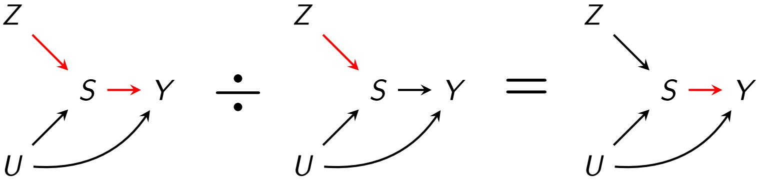


Notes: The figure plots the difference in FICA taxable earnings by draft-eligibility status for the four cohorts born 1950–53. Each tick on the vertical axis represents $500 real (1978) dollars.

FIGURE 2. THE DIFFERENCE IN EARNINGS BY DRAFT-ELIGIBILITY STATUS

Instrumental variables estimate

$$\hat{\alpha} = \frac{\bar{y}^e - \bar{y}^n}{\hat{p}^e - \hat{p}^n}$$



The effect of military service is the difference in mean earnings between those eligible $\bar{y}^e$ and not eligible $\bar{y}^n$, divided by the difference in rates of military service rates between them.

Potential issue: Draft eligibility only changed the probability of veteran status by 0.10 to 0.16.

Single year estimates

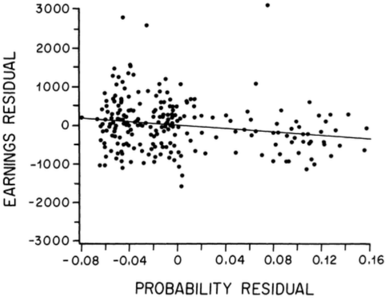
TABLE 3—WALD ESTIMATES

Cohort	Year	Draft-Eligibility Effects in Current $: FICA Earnings (1)	Adjusted FICA Earnings (2)	Total W-2 Earnings (3)	$\hat{p}^e - \hat{p}^n$ (4)	Service Effect in 1978 $ (5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8
		(281.9)	(330.1)	(376.0)		(1,959.3)

Efficient estimator: Pool years

$$\bar{y}_{ctj} = \beta_c + \delta_t + \hat{p}_{cj}\alpha + \bar{u}_{ctj}$$

$\hat{p}_{cj}$ estimated from the Defense Manpower Data Center administrative records and CWHS data on cohort size. SIPP used for 1950 cohort.



Earnings loss of about $2,000.

Possible mechanism: Lost work experience

Can't identify the causal effect of work experience on earnings, but assuming a parametric model, the results agree with the claim that military service reduces earnings through lost work experience.

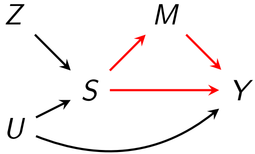
TABLE 5—EARNINGS-FUNCTION MODELS FOR THE VETERAN EFFECT, WHITES BORN 1950–52

Parameter	Model (5): Loss of Experience (1)	Model (6): Loss of Experience, Reduced Growth Rate (2)	Model (7): Unrestricted Reduced Form (3)
Experience Slope, β_0	0.1022	0.1016	0.1016
	(0.007)	(0.007)	(0.007)
Experience Squared, γ	−0.0027	−0.0025	−0.0025
	(0.0003)	(0.0003)	(0.0003)
Veteran Effect on Slope, β_1		−0.0035	
		(0.0023)	
Veteran Loss of Experience, l	2.08	1.84	
	(0.38)	(0.43)	
$\pi_1 = -[\beta_0 l - \gamma l^2 + \beta_1 l]$			−0.189
			(0.052)
$\pi_2 = -[2\gamma l - \beta_1]$			0.006
			(0.004)
Age at Which Reduced Form Veteran Effect $(\pi_1 + \pi_2 x_{ct}) = 0$			50.1
			(15.9)
χ^2(dof)	1.41(1)		813.57(1247)

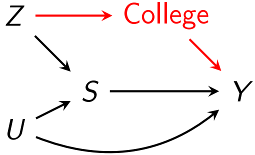
Notes: Standard errors in parentheses.
The table reports estimates of experience-earnings profiles that include parameters for the effect of veteran status. Estimates are of equations (5), (6) and (7) in the text. The estimating sample includes FICA taxable earnings from 1975–84 for men born 1950, 1976–84 earnings for men born in 1951, and 1977–84 earnings for men born 1952. The estimation method is optimally weighted Two-Sample Instrumental Variables for a nonlinear model in columns (1) and (2), and for a linear model in column (3).

Limitations

- Treatment effect heterogeneity: Model estimates the local average treatment effect (LATE) for compliers. May not reflect the effect on those who volunteer.
- Estimates the total effect



- Exclusion restriction



Amusing conclusion

Angrist disproves his prior work:

> Finally, there remains the question of reconciling the loss of earnings to Vietnam era veterans with the apparent benefits of military service to veterans of World War II and other eras (Rosen and Taubman, 1982; Berger and Hirsch, 1983). Elsewhere, Alan Krueger and I have argued that the need for reconciliation is, at least in part, illusory (Angrist and Krueger, 1989). Although OLS regressions usually show that the effect of World War II veteran status is large, positive, and significant, these results may actually be a consequence of selection bias. By exploiting the fact that World War II veteran status is also correlated with exact date of birth, we have implemented an instrumental variables estimation strategy similar in spirit to the one used here. The results of this procedure indicate that the true impact of World War II veteran status on earnings is no larger than zero and may well be negative.

Figlio, David N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy* 2(4): 376-94.

Research Question

- Question: What is the causal effect of having a disruptive student on the academic and behavioural outcomes of other students in the class?
- What are possible confounders? In what direction might those confounders bias our results?

$$
\begin{tikzcd}
 & T \arrow[r] & Y \\
U \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

The Natural Experiment

"I propose an unusual identification strategy to estimate the effects of disruptive students on peer behavior and academic outcomes. I suggest that boys with names most commonly given to girls may be more prone to misbehavior as they get older. The argument goes as follows: Up until a certain point in childhood, boys with names associated with girls are unaffected by their names, either positively or negatively. But as they enter middle school and (1) become more aware of their own sexuality and (2) are mixed with a new group of children (including those older than they are) who did not attend their elementary school, boys with names associated with girls may begin to misbehave in school at a disproportionate rate."

- Data: Names, classroom assignment, behavioural problems, and student test scores from a large Florida school district in SY 1996-97 through SY 1999-2000.

What are the key variables?

- **What's the instrument (Z)?**

What are the key variables?

- **What's the instrument (Z)?**
  Having a male student with a female name in a class

What are the key variables?

- **What's the instrument (Z)?**
  Having a male student with a female name in a class
- **What's the treatment (T)?**

What are the key variables?

- **What's the instrument (Z)?**
  Having a male student with a female name in a class
- **What's the treatment (T)?**
  Having one or more disruptive students in the class

What are the key variables?

- **What's the instrument (Z)?**
  Having a male student with a female name in a class
- **What's the treatment (T)?**
  Having one or more disruptive students in the class
- **What are the outcomes (Y)?**

What are the key variables?

- **What's the instrument (Z)?**
  Having a male student with a female name in a class
- **What's the treatment (T)?**
  Having one or more disruptive students in the class
- **What are the outcomes (Y)?**
  Academic performance and getting suspended (others in the classroom)

Boys with female names

- Boys' names given overwhelmingly to girls (most common in FL between 1989 and 1994): Alexis (given 90 percent of the time to girls), Courtney (94 percent), Shannon (92 percent), Kelly (93 percent), Shelby (95 percent), and Ashley (99 percent).
- Among the broader set of names given more frequently to girls than to boys, the most common names, in addition to Alexis and Courtney, are Taylor (71 percent female), Dominique (66 percent), Jamie (81 percent), and Ariel (80 percent).

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics? (Maybe, conditional on immigrant status, race/ethnicity, family income)
2. **Exclusion Restriction**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics? (Maybe, conditional on immigrant status, race/ethnicity, family income)
2. **Exclusion Restriction**
   Can having a female name have an effect on peer outcomes other than through the student's disruptive behaviour?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics? (Maybe, conditional on immigrant status, race/ethnicity, family income)
2. **Exclusion Restriction**
   Can having a female name have an effect on peer outcomes other than through the student's disruptive behaviour?
3. **First-stage relationship**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics? (Maybe, conditional on immigrant status, race/ethnicity, family income)
2. **Exclusion Restriction**
   Can having a female name have an effect on peer outcomes other than through the student's disruptive behaviour?
3. **First-stage relationship**
   Does a boy having a female name actually induce him to act out disruptively in class?
4. **Monotonicity**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Are names randomly assigned to kids? Are they as-if randomly assigned conditional on observed characteristics? (Maybe, conditional on immigrant status, race/ethnicity, family income)
2. **Exclusion Restriction**
   Can having a female name have an effect on peer outcomes other than through the student's disruptive behaviour?
3. **First-stage relationship**
   Does a boy having a female name actually induce him to act out disruptively in class?
4. **Monotonicity**
   Are there defiers?

IV Assumption Check - First Stage Relationship

Table 1. Rates of 5+ Day Suspensions before and after "Natural" School Transitions, All Boys

	ELEMENTARY GRADES			FIRST YEAR OF MIDDLE SCHOOL			
	Boys with "Feminine" Names	Boys with "Masculine" Names	Difference	Boys with "Feminine" Names	Boys with "Masculine" Names	Difference	Difference in Difference
All boys	0.017	0.015	0.002	0.091	0.067	0.024*	0.022*
Boys from poor families	0.026	0.027	−0.001	0.167	0.111	0.056*	0.057*
Boys from other families	0.007	0.005	0.002	0.028	0.029	0.001	−0.001
Black boys from poor families	0.040	0.041	−0.001	0.217	0.156	0.061*	0.062*
White boys from poor families	0.005	0.011	−0.006	0.103	0.065	0.038*	0.044*
Hispanic boys from poor families	0.012	0.016	−0.004	0.087	0.074	0.013*	0.017*

*statistically significant at 5%.

Starting in middle school, boys in female names begin to be more disruptive.

Reduced Form

Table 1. Rates of 5+ Day Suspensions before and after "Natural" School Transitions, All Boys

	ELEMENTARY GRADES			FIRST YEAR OF MIDDLE SCHOOL			
	Boys with "Feminine" Names	**Boys with "Masculine" Names**	**Difference**	**Boys with "Feminine" Names**	**Boys with "Masculine" Names**	**Difference**	**Difference in Difference**
All boys	0.017	0.015	0.002	0.091	0.067	0.024*	0.022*
Boys from poor families	0.026	0.027	−0.001	0.167	0.111	0.056*	0.057*
Boys from other families	0.007	0.005	0.002	0.028	0.029	0.001	−0.001
Black boys from poor families	0.040	0.041	−0.001	0.217	0.156	0.061*	0.062*
White boys from poor families	0.005	0.011	−0.006	0.103	0.065	0.038*	0.044*
Hispanic boys from poor families	0.012	0.016	−0.004	0.087	0.074	0.013*	0.017*

*statistically significant at 5%.

Boys with female names tend to get suspended at higher rates than boys with masculine names.

Results

Table 4. Instrumental Variables Results of Effects of Disruptive Peers

Specification of Disruptive Classmate Measure	DEPENDENT VARIABLE: Mathematics Test Score (National Percentile Ranking)	DEPENDENT VARIABLE: Student Suspended at Least Once for Five or More Days	Partial R^2 of Instrumental Variables in First-Stage Regression
Fraction of classmates suspended at least once for five or more days	–100.9* (22.5)	0.932* (0.144)	0.027
Fraction of classmates suspended multiple times during the year	–96.4* (23.1)	0.778* (0.151)	0.025
At least 5% of classmates suspended for five or more days	–6.5* (1.5)	0.061* (0.011)	0.028
At least 10% of classmates suspended for five or more days	–17.7* (4.8)	0.163* (0.035)	0.025

- Each cell represents a different regression
- Note that these effects represent the effects of moving from 0% to 100% disruptive peers

Results

"To put these estimates in perspective, in a typical classroom of thirty students, the estimates suggest that adding one additional disruptive child to the classroom results in **reduced peer mathematics test scores of about four national percentiles** and about a **three percentage point increased likelihood that a peer will get into serious trouble at school**, as measured by being suspended at least once for five or more days."

Quarter of Birth Example

Angrist, Joshua and Alan Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4).

Research Question

- Question: What is the causal effect of education on earnings?
- What are possible confounders? In what direction might those confounders bias our results?

$$\begin{tikzcd} & T \arrow[r] & Y \\ U \arrow[ur] \arrow[urr, bend right] & & \end{tikzcd}$$

The Natural Experiment

"The experiment stems from the fact that children born in different months of the year start school at different ages, while compulsory schooling laws generally require students to remain in schools until their sixteenth or seventeenth birthday. In effect, the interaction of school-entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months."

- Data: Men from the 1980 Census Public Use Sample

What are the key variables?

- **What's the instrument (Z)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education
- **What's the outcome (Y)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education
- **What's the outcome (Y)?**
  Earnings

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**
   Does birth quarter induce variation in time spent in school?
4. **Monotonicity**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**
   Does birth quarter induce variation in time spent in school?
4. **Monotonicity**
   Are there defiers?

IV Assumption Check - First Stage Relationship

We can check by regressing treatment on the instrument. We can also gain more confidence by examining plots of the relationship:

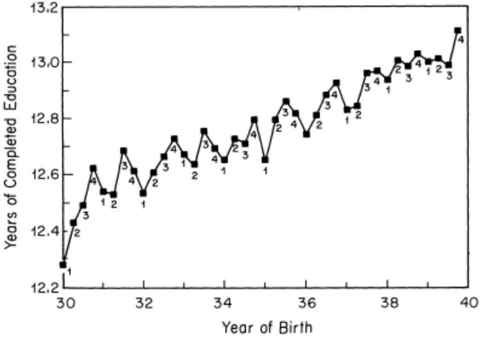
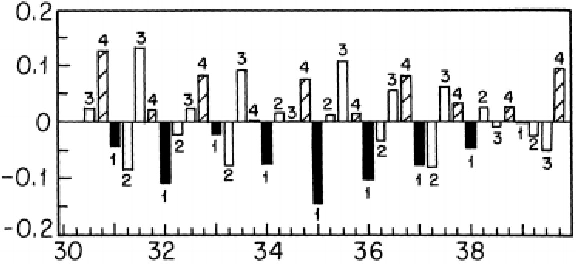


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Men born earlier in the year have less schooling

IV Assumption - First Stage Relationship



Also: differences across states suggest that compulsory school laws do keep students enrolled longer than they might want.

Reduced Form

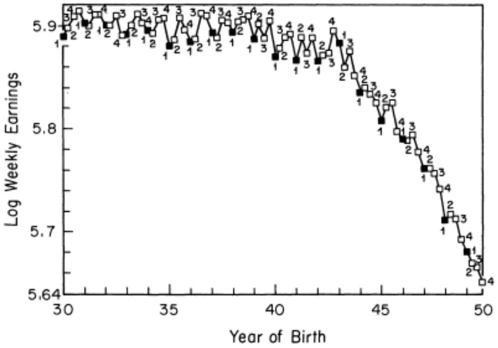


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Differences in schooling due to quarter of birth appear to translate into different earnings.

2SLS Results: White Men, 1930s Cohorts

TABLE VII
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673	0.0928	0.0673	0.0907	0.0628	0.0831	0.0628	0.0811
	(0.0003)	(0.0093)	(0.0003)	(0.0107)	(0.0003)	(0.0095)	(0.0003)	(0.0109)
Race (1 = black)	—	—	—	—	−0.2547	−0.2333	−0.2547	−0.2354
					(0.0043)	(0.0109)	(0.0043)	(0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705	0.1511	0.1705	0.1531
					(0.0029)	(0.0095)	(0.0029)	(0.0107)
Married (1 = married)	—	—	—	—	0.2487	0.2435	0.2487	0.2441
					(0.0032)	(0.0040)	(0.0032)	(0.0042)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0757	−0.0880	—	—	−0.0778	−0.0876
			(0.0617)	(0.0624)			(0.0603)	(0.0609)
Age-squared	—	—	0.0008	0.0009	—	—	0.0008	0.0009
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	163 [179]	—	161 [177]	—	164 [179]	—	162 [177]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 150 quarter-of-birth times state-of-birth interactions. Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is the same as in Table VI. Sample size is 329,509.

2SLS Results: Black Men, 1930s Cohorts

TABLE VIII

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR BLACK MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0672	0.0635	0.0671	0.0555	0.0576	0.0461	0.0576	0.0391
	(0.0013)	(0.0185)	(0.0003)	(0.0199)	(0.0013)	(0.0187)	(0.0013)	(0.0199)
SMSA (1 = center city)	—	—	—	—	0.1885	0.2053	0.1884	0.2155
					(0.0142)	(0.0308)	(0.0142)	(0.0324)
Married (1 = married)	—	—	—	—	0.2216	0.2272	0.2216	0.2307
					(0.0193)	(0.0136)	(0.0100)	(0.0140)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
49 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0309	−0.3274	—	—	−0.2978	−0.3237
			(0.2538)	(0.2560)			(0.0032)	(0.2497)
Age-squared	—	—	0.0033	0.0035	—	—	0.0032	0.0035
			(0.0028)	(0.0028)			(0.0027)	(0.0028)
χ^2 [dof]	—	184 [176]	—	181 [173]	—	178 [176]	—	175 [173]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 147 quarter-of-birth times state-of-birth interactions. (There are no black men in the sample born in Hawaii.) Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is drawn from the 1980 Census. Sample size is 26,913.

Note the returns for black men appear to be smaller

Regression Discontinuity

- Key idea is to exploit an arbitrary assignment rule to identify a causal quantity.
- Remember that we are only identifying an effect at the boundary.

Setup

- The basic idea behind RDDs:
  - X_i is a **forcing variable.**
  - Treatment assignment is determined by a cutoff in X_i.

$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

- X_i can be related to the potential outcomes and so comparing treated and untreated units does not provide causal estimates
- Assume relationship between X and the potential outcomes Y_1 and Y_0 is **smooth** around the threshold $\rightsquigarrow$ discontinuity created by the treatment to estimate the effect of D on Y at the threshold

Design

- **Sharp RD**: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(D_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(D_i = 1 | X_i = c - \varepsilon) = 0$$

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We want to investigate the behavior of the outcome around the threshold: $\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]$
- Under certain assumptions, this quantity identifies the ATE at the threshold: $\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c]$

Identification

Identification Assumption

1. $Y_1, Y_0 \perp\!\!\!\perp D|X$ *(trivially met)*
2. $0 < P(D=1|X=x) < 1$ *(always violated in Sharp RDD)*
3. $E[Y_1|X,D]$ *and* $E[Y_0|X,D]$ *are continuous in* X *around the threshold* $X=c$ *(individuals have imprecise control over* X *around the threshold)*

Identification Result

The treatment effect is identified at the threshold as:

$$\begin{aligned} \alpha_{SRDD} &= E[Y_1 - Y_0|X=c] \\ &= E[Y_1|X=c] - E[Y_0|X=c] \\ &= \lim_{x \downarrow c} E[Y_1|X=x] - \lim_{x \uparrow c} E[Y_0|X=x] \end{aligned}$$

Without further assumptions α_{SRDD} *is only identified at the threshold.*

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.
- For instance, if people sort around threshold, then you might get jumps other than the one you care about.
- If things other than the treatment change at the threshold, then that might cause discontinuities in the potential outcomes.

Fuzzy RD

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

$$\lim_{x \downarrow c} \Pr[D_i = 1 | X_i = x] \neq \lim_{x \uparrow c} \Pr[D_i = 1 | X_i = x]$$

- In the sharp RD, this is also true, but it further required the jump in probability to be from 0 to 1.
- Fuzzy RD is often useful when the a threshold encourages participation in program, but does not actually force units to participate.
- Sound familiar? Fuzzy RD is just IV!

Fuzzy RD is IV

- Forcing variable is an **instrument**: affects Y_i, but only through D_i (at the threshold)
- Let $D_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.
- $D_i(x) = 1$ if unit i would take treatment when X_i was x
- $D_i(x) = 0$ if unit i would take control when X_i was x

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $D_i(c+e) \geq D_i(c-e)$ for all $0 < e < \varepsilon$

No one is discouraged from taking the treatment by crossing the threshold.

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, D_i(x)\} \perp\!\!\!\perp X_i$$

Basically, in an ε-ball around c, the forcing variable is randomly assigned.

The Elite Illusion

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82(1): 137-196.

Background

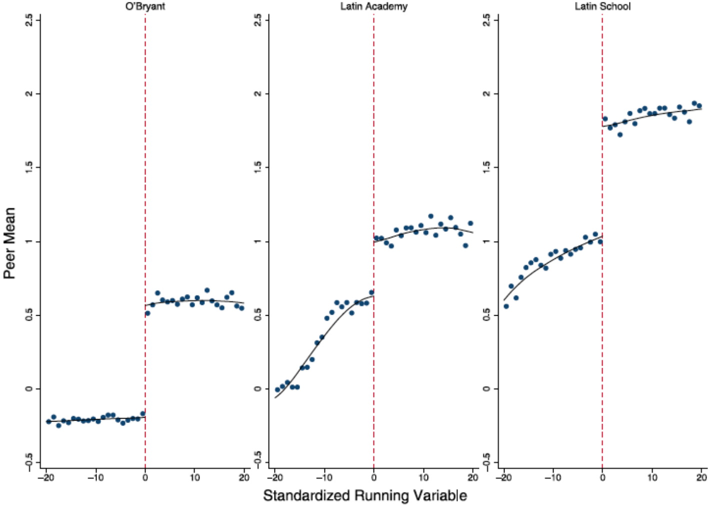
Key question:
Do peer effects influence the educational returns to attending an exam school?
Problem with prior research:
Lots of selection issues! Since exam schools seek to admit the highest achievers, those who go to exam schools might look systematically different from people who do not go to exam schools.

Identification strategy

Fuzzy RD design!
Intuition: students who just barely missed the cut-off for an offer of admission should be comparable to students who just barely passed the cut-off.

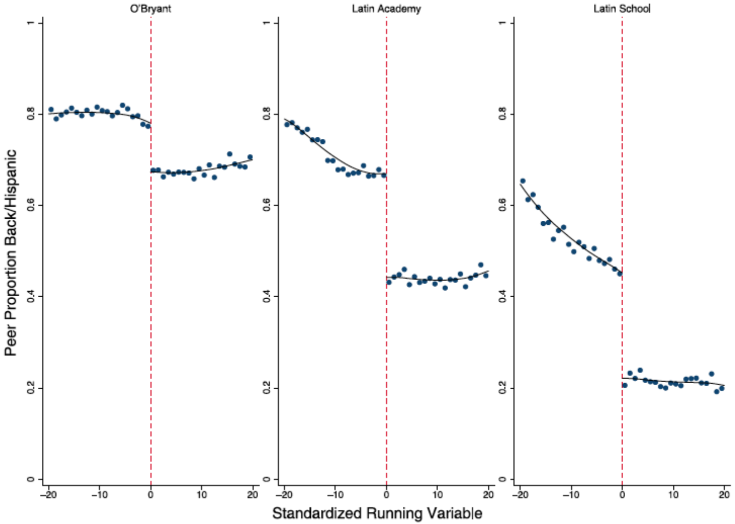
- **What's the running variable?**
  A composite academic score constructed as a weighted average of applicants' standardized math and English GPA, along with standardized scores on four parts of an exam (verbal, quantitative, reading, and math).
- **What's the instrument?**
  Offer of admission to an exam school
- **What's the treatment (T)?**
  Attending an exam school with different peer characteristics
- **What are the outcomes (Y)?**
  Academic performance

Academic Achievement of Peers



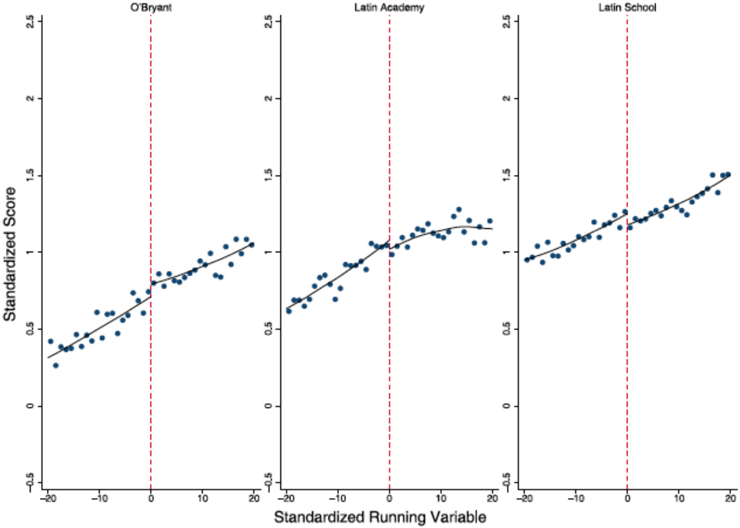
(a) Baseline peer math score at Boston exam schools for 7th and 9th grade applicants

Demographic Composition of Peers



(b) Proportion black or Hispanic at Boston exam schools

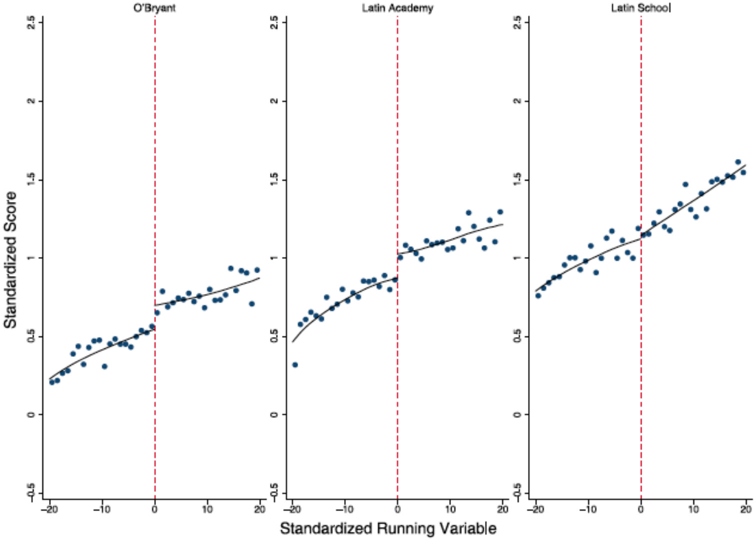
10th Grade Math Scores



(a) 10th grade math at Boston exam schools for 7th and 9th grade applicants

FIGURE 4.—This figure shows the average 10th grade math (a) and English (b) MCAS scores

10th Grade English Scores



(b) 10th grade English at Boston exam schools for 7th and 9th grade applicants

Elite Illusion?

BOSTON REDUCED-FORM ESTIMATES: MCAS MATH AND ENGLISH[a]

		Parametric Estimates				Nonparametric (DM) Estimates			
Application Grade	Test Grade	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
				Panel A. Math					
7th	7th and 8th	−0.125	−0.105	0.002	−0.079	−0.093	−0.144*	0.012	−0.086**
		(0.100)	(0.093)	(0.099)	(0.054)	(0.071)	(0.074)	(0.060)	(0.034)
		4,047	4,208	3,786	12,041	3,637	4,000	3,067	10,704
7th and 9th	10th	0.066	−0.097	−0.056	−0.018	0.067	−0.047	−0.064**	0.000
		(0.066)	(0.085)	(0.051)	(0.036)	(0.045)	(0.047)	(0.028)	(0.026)
		3,389	2,709	2,459	8,557	3,083	2,027	1,827	6,937
7th and 9th	7th, 8th, and 10th	−0.038	−0.102	−0.020	−0.054	−0.020	−0.115**	−0.016	−0.053**
		(0.068)	(0.067)	(0.072)	(0.039)	(0.049)	(0.049)	(0.043)	(0.024)
		7,436	6,917	6,245	20,598	6,720	6,027	4,894	17,641
				Panel B. English					
7th	7th and 8th	−0.061	−0.092	−0.187***	−0.110**	−0.062	0.012	−0.128***	−0.063**
		(0.078)	(0.067)	(0.065)	(0.043)	(0.041)	(0.042)	(0.037)	(0.025)
		4,151	4,316	3,800	12,267	3,931	3,762	3,533	11,226
7th and 9th	10th	0.108	0.136	0.028	0.095*	0.140***	0.182***	−0.002	0.113***
		(0.079)	(0.096)	(0.085)	(0.053)	(0.048)	(0.057)	(0.065)	(0.036)
		3,398	2,715	2,463	8,576	3,308	1,786	1,916	7,010
7 th and 9th	7th, 8th, and 10th	0.014	−0.001	−0.106*	−0.026	0.029	0.067	−0.089***	0.002
		(0.055)	(0.070)	(0.061)	(0.039)	(0.034)	(0.042)	(0.032)	(0.023)
		7,549	7,031	6,263	20,843	7,239	5,548	5,449	18,236

[a]This table reports estimates of the effects of exam school offers on MCAS scores. The sample covers students within 20 standardized units of offer cutoffs. Parametric models include a cubic function of the running variable, allowed to differ on either side of offer cutoffs. Nonparametric estimates use the edge kernel, with bandwidth computed following DesJardins and McCall (2008) and Imbens and Kalyanaram (2012), as described in the text. Optimal bandwidths were computed separately for each school. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates and for estimates pooling outcomes also cluster on student. Sample sizes are shown below standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.

Questions?