

## Mastering Modern IV

Master Joshway

ASSA Continuing Ed: January 2020

### Sometimes You Get What You Need

- Modern IV distinguishes *internal* from *external* validity
- A good instrument – by definition – captures an internally valid causal effect: treatment effects on subjects for whom the instrument changes treatment
- External validity is the predictive value of internally valid estimates in a new context
- Examples
  - Draft-lottery estimates of effects of Vietnam-era military service
  - Quarter-of-birth estimates of the economic returns to schooling
  - Fertility experiments TBD
- The theory of a heterogeneous world
  - Quasi-experimental designs capture causal effects for a well-defined subpopulation, usually a proper subset of the treated
  - In models with variable treatment intensity, we get effects over a limited but knowable range

## Children and Their Parents Labor Supply

- A causal model for the impact of a third child on mothers with at least two:

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \alpha + \rho D_i + \eta_i$$

Constant FX? Parameter  $\rho$  is *the thing that must be named*

- Dependent variables = employment, hours worked, weeks worked, earnings
  - $D_i = 1[kids > 2]$  for samples of mothers with at least two children
  - $Z_i$  indicates twins or same-sex sibships at second birth
- With a single Bernoulli instrument and no covariates, the IV estimand is the Wald formula

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

- Instruments ready?

## Effects for Whom?

## The LATE Framework

- $Y_i(d, z)$  denotes the potential outcome for  $i$  when treatment status  $D_i = d$  and instrument  $Z_i = z$
- Double-indexed potentials mean instrumental variables might change outcomes directly
- We assume, however, that IV initiates a causal chain: the instrument,  $Z_i$ , affects  $D_i$ , which in turn affects  $Y_i$
- To build these links, define *potential treatment status*, indexed by values of  $Z_i$ :
  - $D_{1i}$  is  $i$ 's treatment status when  $Z_i = 1$
  - $D_{0i}$  is  $i$ 's treatment status when  $Z_i = 0$
- Observed treatment status is therefore

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$$

- The causal effect of  $Z_i$  on  $D_i$  is  $D_{1i} - D_{0i}$

## Independence and First Stage

**Independence.** The instrument is as good as randomly assigned:

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

- Sibling sex mix and multiple births are independent of potential outcomes and potential treatments
- Independence implies that the **first-stage** is the average causal effect of  $Z_i$  on  $D_i$ :

$$\begin{aligned} E[D_i | Z_i = 1] - E[D_i | Z_i = 0] &= E[D_{1i} | Z_i = 1] - E[D_{0i} | Z_i = 0] \\ &= E[D_{1i} - D_{0i}] \end{aligned}$$

- Independence is likewise sufficient for a causal interpretation of the **reduced form**:

$$E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0] = E[Y_i(D_{1i}, 1) - Y_i(D_{0i}, 0)]$$

## Exclusion

Our journey from causal RF to treatment effect starts with:

**Exclusion.** The instrument affects  $Y_i$  only through  $D_i$ , that is,

$$\begin{aligned}Y_i(1, 1) &= Y_i(1, 0) \equiv Y_{1i} \\ Y_i(0, 1) &= Y_i(0, 0) \equiv Y_{0i}\end{aligned}$$

- The exclusion restriction means  $Y_i$  can be written

$$\begin{aligned}Y_i &= Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i,\end{aligned}$$

for single-index potentials  $Y_{1i}$  and  $Y_{0i}$  *that satisfy independence*

- Exclusion means quarter of birth affects earnings only through schooling; sex mix affects labor supply only by changing family size

## Monotonicity

A useful technical assumption:

**Monotonicity.**  $D_{1i} \geq D_{0i}$  for everyone (or vice versa).

- By virtue of monotonicity,  $E[D_{1i} - D_{0i}] = P[D_{1i} > D_{0i}]$
- Consider a latent-index model

$$D_i = \begin{cases} 1 & \text{if } \gamma_0 + \gamma_1 Z_i > v_i \\ 0 & \text{otherwise} \end{cases}$$

where  $v_i$  is "random utility"

- This model characterizes potential treatment assignments as

$$\begin{aligned}D_{0i} &= 1[\gamma_0 > v_i] \\ D_{1i} &= 1[\gamma_0 + \gamma_1 > v_i],\end{aligned}$$

clearly satisfying monotonicity

## Better LATE . . .

- The *independence assumption* says the instrument is as good as randomly assigned
- The *exclusion restriction* means that causal effects of the instrument on outcomes are due solely to effects of the instrument on  $D_i$ 
  - Exclusion is usually more controversial than independence
- We also assume there's a *first-stage*; by virtue of *monotonicity*, this is the share of the population for which  $D_i$  is changed by  $Z_i$
- Given these assumptions, we have:

### THE LATE THEOREM

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

- Proof - See MHE 4.4.1

## The Compliant Subpopulation

*LATE compliers have  $D_{1i} > D_{0i}$*

- This language comes from randomized trials where  $Z_i$  is treatment assigned and  $D_i$  is treatment received (an apt analogy)
- LATE assumptions partition the world:
  - Compliers  $D_{1i} > D_{0i}$
  - Always-takers  $D_{1i} = D_{0i} = 1$
  - Never-takers  $D_{1i} = D_{0i} = 0$
- IV says nothing about always-takers and never-takers because treatment status for these types is unchanged by the instrument
  - An analogy: panel models with fixed effects identify effects only for "changers"
- Assuming effects are the same for all three groups returns us to the constant-effects model

## The Compliant Subpopulation (cont.)

- From

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i,$$

we see that  $\{D_i = 1\}$

$$= \{D_{0i} = D_{1i} = 1\} \cup \{\{D_{1i} - D_{0i} = 1\} \cap \{Z_i = 1\}\}$$

- In other words . . .

$$\{\text{treated}\} = \{\text{always-takers}\} + \{\text{compliers assigned } Z_i = 1\}$$

- Effects on the treated average those for always-takers and compliers
  - $Z_i = 1$  compliers are representative of all
- Characterizing compliers
  - How many? The first stage!
  - What are their X's? See MHE 4.4.4

## IV in Randomized Trials (An Analogy Realized)

*RCTs are beset by noncompliance: Some randomly assigned to the treatment group are untreated*

- *Intention-to-treat* analysis (contrasts by treatment assigned) preserve independence but is diluted by non-compliance
- *Per-protocol* analysis (contrasts by treatment received) are contaminated by selection bias
- IV solves this problem:  $Z_i$  indicates random assignment to the treatment group;  $D_i$  indicates treatment received
- No always-takers! (no controls are treated), so LATE = TOT:

$$\begin{aligned} \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1]} &= \frac{\text{ITT effect}}{\text{compliance rate}} \\ &= E[Y_{1i} - Y_{0i}|D_i = 1] \end{aligned}$$

- Direct proof (Bloom, 1984; See MHE 4.4.3)

## Are we there yet?

**#TuesdayThoughts**  
86.9K Tweets

**Roy Halladay**  
6,918 Tweets

**#ParisAgreement**  
8,757 Tweets

**#DefiningJustice**

**#BeyondSilos**

**John Mandrola, MD** @drjohnm · Nov 5  
Can u elebotate? I'm not sure what you mean.

2 1 2

**Robert W. Yeh MD MBA** @rwyeh · Nov 5  
We don't effectively deal with non-compliance/crossover in RCT analyses. Neither ITT nor per protocol estimate true effect in the treated.

4 6

**Seth Trueger** @MDaware · Nov 5  
isn't that a feature, not a bug of ITT? in real practice when we try a treatment, not everyone can complete it

1 1


**Robert W. Yeh MD MBA** @rwyeh · Nov 5  
It just answers a different question. If question is "does this Rx actually work when received vs not received" it gives a biased answer.

1 1 2

**Jeremy Sussman** @JeremySussman  
Follow

Replied to @rwyeh @metrics52 and 3 others

**Fwiw, we tried to explain this problem and the solution for docs a few years ago.**



**An IV for the RCT: using instrumental variables to adjust for...**  
Although the randomised controlled trial is the "gold standard" for studying the efficacy and safety of medical treatments, it is not necessarily free from bias. When patients do not follow the...  
ncbi.nlm.nih.gov

8:33 AM - 5 Nov 2017

**Find people you know**  
Import your contacts from G

[Connect other address books](#)

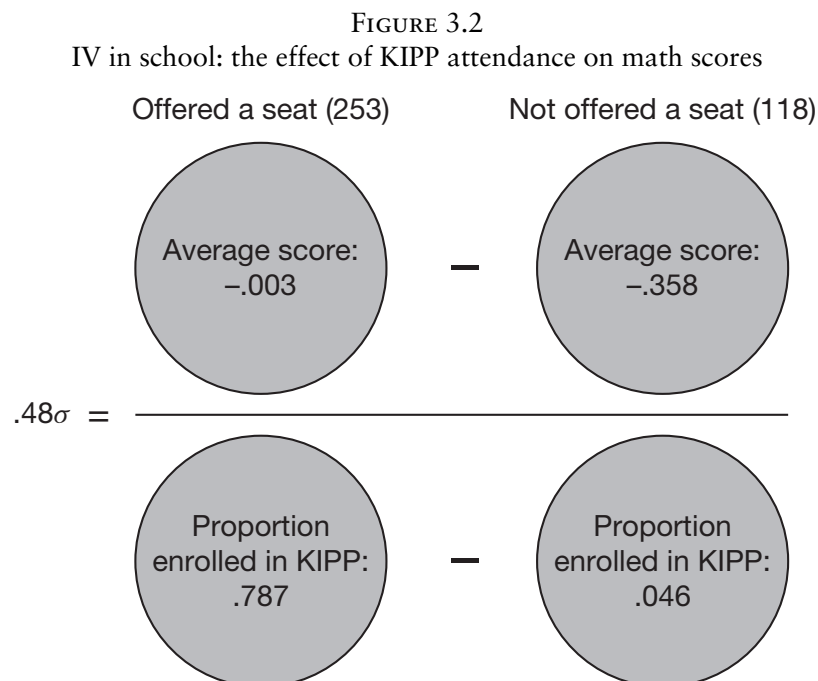
© 2017 Twitter About Help Center  
Privacy policy Cookies Ads info Br  
Blog Status Apps Jobs Advertise  
Marketing Businesses Developers

# Bloom Waits for Superman

## The Charter Conundrum

- Charter schools (featured in [Waiting for Superman](#)) are publicly-funded private schools with a time-limited warrant to operate in public school districts
  - Host districts pay their PPE to charters for each pupil enrolled
  - Charters are granted conditional on good governance and good performance; many are lost or revoked
  - Unlike public sector teachers, charter teachers typically aren't unionized; many are inexperienced and uncredentialed
- Urban charter students do better than traditional public school peers: causal effect or selection bias?
  - Charter applicants often have better baseline (pre-enrollment scores)
- MIT's SEII researchers answer the charter causal challenge by playing the lottery
- Over-subscribed Massachusetts charters admit by random assignment

## The KIPP Lottery Does the Heavy Lifting (MM Chpt 3)



*Note:* The effect of Knowledge Is Power Program (KIPP) enrollment described by this figure is  $.48\sigma = .355\sigma / .741$ .



## The Four Types of Children

TABLE 3.2  
The four types of children

		Lottery losers $Z_i = 0$	
		Doesn't attend KIPP $D_i = 0$	Attends KIPP $D_i = 1$
Lottery winners $Z_i = 1$	Doesn't attend KIPP $D_i = 0$	Never-takers ( <i>Normando</i> )	Defiers
	Attends KIPP $D_i = 1$	Compliers ( <i>Camila</i> )	Always-takers ( <i>Alvaro</i> )

Note: KIPP = Knowledge Is Power Program.

- With few like Alvaro, **LATE=TOT**:

$$\begin{aligned}
 & E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}] \\
 &= E[Y_{1i} - Y_{0i} | D_{1i} = 1] = E[Y_{1i} - Y_{0i} | D_{1i} = 1, Z_i = 1] \\
 &= E[Y_{1i} - Y_{0i} | D_i = 1]
 \end{aligned}$$

Remember O.J.?

## Abuse Busters

The police were called to O.J.'s Rockingham mansion 9 times; he was arrested only once. The Minneapolis Domestic Violence Experiment (MDVE; Sherman and Berk, 1984) boldly evaluated the police response to domestic violence . . .

- Police were randomly assigned to advise, separate, or arrest
- Substantial compliance problems as officers reacted in the field:

Table 1: Assigned and Delivered Treatments in Spousal Assault Cases

Assigned Treatment	Delivered Treatment			Total
	Arrest	Coddled		
		Advise	Separate	
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0(314)

## MDVE First-Stage and Reduced Forms

- IV analysis in Angrist (2006)

Table 2: First Stage and Reduced Forms for Model 1

	Endogenous Variable is Coddled			
	First-Stage		Reduced Form (ITT)	
	(1)	(2)*	(3)	(4)*
Coddled-assigned	0.786 (0.043)	0.773 (0.043)	0.114 (0.047)	0.108 (0.041)
Weapon		-0.064 (0.045)		-0.004 (0.042)
Chem. Influence		-0.088 (0.040)		0.052 (0.038)
Dep. Var. mean		0.567 (coddled-delivered)		0.178 (failed)

## MDVE OLS and 2SLS

Table 3: OLS and 2SLS Estimates for Model 1

Endogenous Variable is Coddled				
	OLS		IV/2SLS	
	(1)	(2)*	(3)	(4)*
Coddled-delivered	0.087 (0.044)	0.070 (0.038)	0.145 (0.060)	0.140 (0.053)
Weapon		0.010 (0.043)		0.005 (0.043)
Chem. Influence		0.057 (0.039)		0.064 (0.039)

- Columns 3 and 4 estimate *the effect of coddling on the coddled* (those assigned to be arrested are arrested: there are no "coddling always-takers")
- Selective compliance attenuates OLS, but IV (2SLS) fixes this

# Superman Returns!

## Distribution Treatment Effects (ACR)

Abadie (2002) shows that for any function,  $g(Y_i)$

$$\frac{E[D_i g(Y_i) | Z_i = 1] - E[D_i g(Y_i) | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]} = E[g(Y_{1i}) | D_{1i} > D_{0i}]$$

$$\frac{E[(1 - D_i) g(Y_i) | Z_i = 1] - E[(1 - D_i) g(Y_i) | Z_i = 0]}{E[1 - D_i | Z_i = 1] - E[1 - D_i | Z_i = 0]} = E[g(Y_{0i}) | D_{1i} > D_{0i}]$$

- Set  $g(Y_i) = Y_i$  to estimate marginal potential outcome means
- Set  $g(Y_i) = 1[Y_i < c]$  to capture

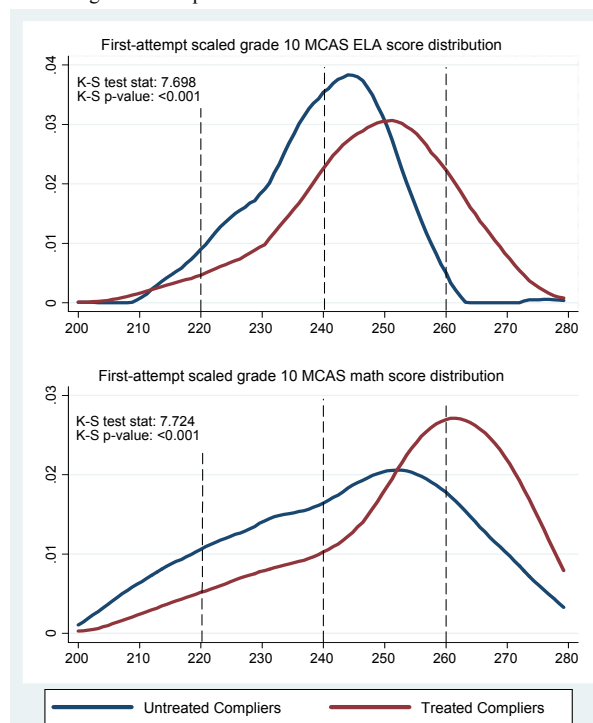
$$E\{1[Y_{ji} < c] | D_{1i} > D_{0i}\} = P[Y_{ji} < c | D_{1i} > D_{0i}],$$

the *distributions* of  $Y_{1i}$  and  $Y_{0i}$

- Angrist *et al.* (JOLE 2016) used this to study charter school effects on achievement *distributions*

## Superman Distributes Achievement Gains at Boston Charter High Schools

Figure 1: Complier Distributions for MCAS Scaled Scores



# Schooling IV

## Questions of Variable Intensity (summary)

Variable  $s_i$  takes on values in the set  $\{0, 1, \dots, \bar{s}\}$ , generating  $\bar{s}$  unit causal effects,  $Y_{s_i} - Y_{s-1,i}$

- A linear model assumes these are the same for all  $s$  and for all  $i$ , obviously unrealistic
- Fear not! 2SLS generates a weighted average of unit causal effects
  - Suppose dummy instrument,  $z_i$  (indicating late quarter births) is used to estimate the returns to schooling
  - Let  $s_{1i}$  denote the schooling  $i$  gets if  $z_i = 1$ ; let  $s_{0i}$  denote the schooling  $i$  gets if  $z_i = 0$
  - We observe  $s_i = s_{0i}(1 - z_i) + z_i s_{1i}$
- Assumptions:
  - Independence and Exclusion  $\{Y_{0i}, Y_{1i}, \dots, Y_{\bar{s}i}; s_{0i}, s_{1i}\} \perp\!\!\!\perp z_i$
  - First Stage  $E[s_{1i} - s_{0i}] \neq 0$
  - Monotonicity  $s_{1i} - s_{0i} \geq 0 \quad \forall i$  (or vice versa)

## Average Causal Response (ACR)

Angrist and Imbens (1995) show

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]} = \sum_{s=1}^{\bar{s}} \omega_s E[Y_{si} - Y_{s-1,i} | S_{1i} \geq s > S_{0i}]$$

where

$$\omega_s = \frac{P[S_{1i} \geq s > S_{0i}]}{\sum_{j=1}^{\bar{s}} P[S_{1i} \geq j > S_{0i}]}$$

Weights  $\omega_s$  are non-negative and sum to 1.

- The ACR is a weighted average of the *unit causal response* along the length of a potentially nonlinear causal relation
- $E[Y_{si} - Y_{s-1,i} | S_{1i} \geq s > S_{0i}]$ , is the average difference in potential outcomes for *compliers at point s*
- Here, compliers are those the instrument moves from treatment intensity less than  $s$  to at least  $s$

## The ACR Weighting Function

- By Monotonicity, the group of compliers at point  $s$  has size:

$$\begin{aligned} P[S_{1i} \geq s > S_{0i}] &= P[S_{1i} \geq s] - P[S_{0i} \geq s] \\ &= P[S_{0i} < s] - P[S_{1i} < s] \end{aligned}$$

- By Independence, this is a difference in treatment CDFs given  $Z_i$ :

$$P[S_{0i} < s] - P[S_{1i} < s] = P[S_i < s | Z_i = 0] - P[S_i < s | Z_i = 1]$$

- The mean of a non-negative random variable is one minus the CDF:

$$\begin{aligned} &E[S_i | Z_i = 1] - E[S_i | Z_i = 0] \\ &= \sum_{j=1}^{\bar{s}} (P[S_i < j | Z_i = 0] - P[S_i < j | Z_i = 1]) = \sum_{j=1}^{\bar{s}} P[S_{1i} \geq j > S_{0i}] \end{aligned}$$

ACR weights are normalized by the first-stage

## QOB IV Reprise

The ACR weighting function shows us where the action is . . .

- AI-95 version of AK-91 Wald
- $S_i$  is years of schooling
- $Z_i$  compares men born in 1st and 4th quarters
- Diffs in CDFs by QOB (first vs. fourth quarter births)  $\Rightarrow$

Table 1. Compulsory School Attendance

	(1) Born in 1st quarter of year	(2) Born in 4th quarter of year	(3) Difference (std. error) (1) – (2)
<i>Panel A: Wald Estimates for 1970 Census—Men Born 1920–1929*</i>			
ln (weekly wage)	5.1485	5.1578	–.00935 (.00374)
Education	11.3996	11.5754	–.1758 (.0192)
Wald est. of return to education			.0531 (.0196)
OLS est. of return to education <sup>b</sup>			.0797 (.0005)
<i>Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939</i>			
ln (weekly wage)	5.8916	5.9051	–.01349 (.00337)
Education	12.6881	12.8394	–.1514 (.0162)
Wald est. of return to education			.0891 (.0210)
OLS est. of return to education			.0703 (.0005)

## Empirical Weighting Function

- For men born 1920-29 in the 1970 Census

Angrist and Imbens: Estimation of Average Causal Effects

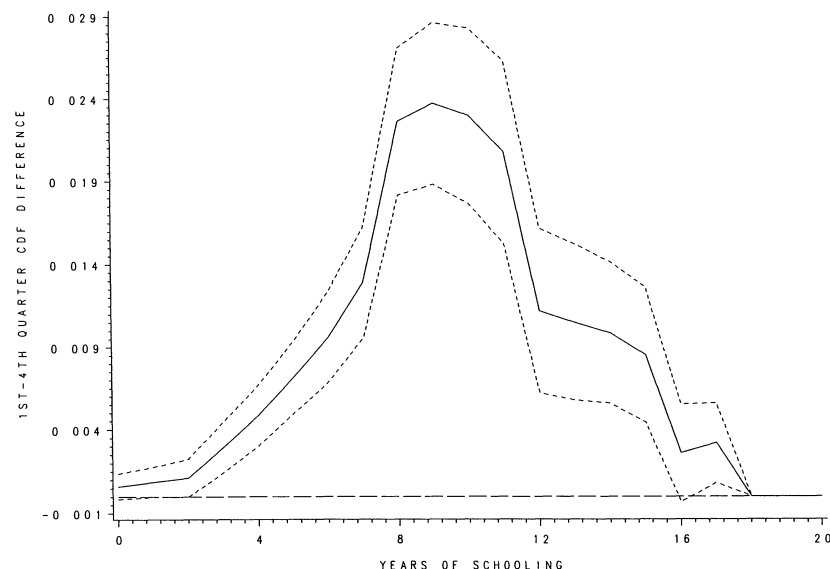


Figure 3. First-Fourth Quarter Difference in Schooling CDF (Men Born 1920–1929, Data From the 1970 Census). Dotted lines are 95% confidence intervals.

## More Variable Treatment Intensities

- Returns to schooling identified using compulsory attendance and child labor laws (Acemoglu and Angrist, 2000)
- Class size (Angrist and Lavy, 1999; Krueger, 1999)
  - $Y_i$  is a test score;  $S_i$  is class size
  - $Z_i$  is Maimonides Rule or random assignment
- GRE test preparation (Powers and Swinton, 1984)
  - $Y_i$  is GRE analytical score;  $S_i$  is hours of study
  - $Z_i$  is randomly assigned letter of encouragement
- Maternal smoking (Permutt and Hebel, 1989)
  - $Y_i$  is birth weight;  $S_i$  is mother's pre-natal smoking
  - $Z_i$  is randomly assigned offer of anti-smoking counseling
- **Quantity-quality trade-offs** (Angrist, Lavy, and Schlosser, 2010)
  - $Y_i$  is schooling, earnings, etc.;  $S_i$  is sibship size
  - $Z_i$  is derived from twins and sibling-sex composition

QQ



## Validating External Validity (summary)

- MM Chpt 3 (ALS 2010) compares 2SLS estimates of the quantity-quality trade-off using twins and sex-mix instruments

- Twins take no never-takers! Twins LATE is therefore

$$E[Y_{1i} - Y_{0i} | D_i = 0]; \text{ where } D_i \text{ indicates more than two}$$

Twins compliers want to stop at two; they're highly educated

- Angrist (2004) shows same-sex LATE is close to ATE by virtue of a symmetric first stage
- Twinning mostly causes a one-child shift; while sex-composition increases childbearing at high parities:
  - QQ twins 1st stage
  - QQ same-sex 1st stage
- Yet the answer always comes out: **no (or positive) effects**. That's one kinda external validity!
  - Angrist and Fernandez-Val (2013) propose another

## Summary

- IV provides a powerful and flexible framework for causal inference
  - An alternative to random assignment with a strong claim on internal validity when the instruments are good
  - A solution to the compliance problem in randomized trials
  - A strategy for the analysis of many observational designs
- Distribution treatment effects? Identified!
  - kappa-weighting (Abadie 2003) extends LATE to nonlinear and quantile models
- IV produces weighted averages of ordered and continuous treatment effects, a generalized LATE
  - The weighting function describes the range of variation covered
- LATE spec tests: No first stage? No reduced form! (Kitagawa 2015)

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

Variable	1980 PUMS			1990 PUMS			1980 PUMS		
	Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Twins-2	Wald estimate using as covariate:	
		More than 2 children	Number of children		More than 2 children	Number of children		More than 2 children	Number of children
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	-0.0080 (0.0016)	-0.133 (0.026)	-0.104 (0.021)	-0.0053 (0.0015)	-0.084 (0.024)	-0.063 (0.018)	-0.0459 (0.0086)	-0.076 (0.014)	-0.057 (0.011)
Weeks worked	-0.3826 (0.0709)	-6.38 (1.17)	-5.00 (0.92)	-0.3233 (0.0743)	-5.15 (1.17)	-3.87 (0.88)	-1.982 (0.386)	-3.28 (0.63)	-2.45 (0.47)
Hours/week	-0.3110 (0.0602)	-5.18 (1.00)	-4.07 (0.78)	-0.2363 (0.0620)	-3.76 (0.98)	-2.83 (0.73)	-1.979 (0.327)	-3.28 (0.54)	-2.44 (0.40)
Labor income	-132.5 (34.4)	-2208.8 (569.2)	-1732.4 (446.3)	-119.4 (42.4)	-1901.4 (670.3)	-1428.0 (502.6)	-570.8 (186.9)	-946.4 (308.6)	-705.2 (229.8)
ln(Family income)	-0.0018 (0.0041)	-0.029 (0.068)	-0.023 (0.054)	-0.0085 (0.0047)	-0.136 (0.074)	-0.102 (0.056)	-0.0341 (0.0223)	-0.057 (0.037)	-0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

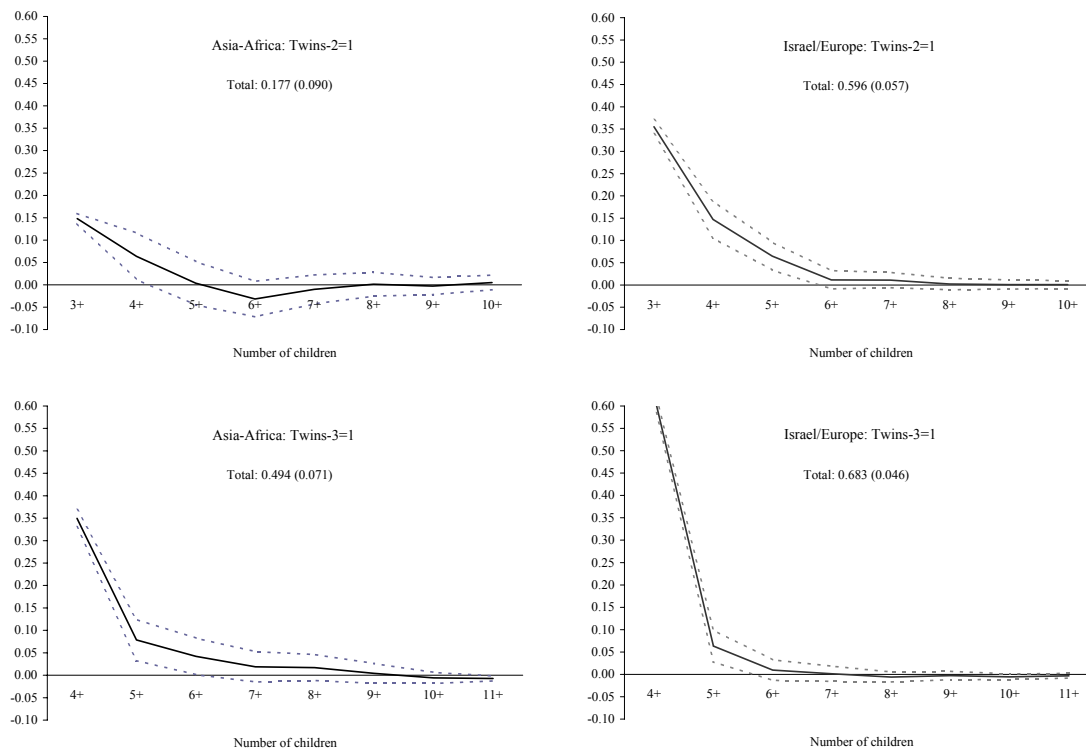


Figure 1: First borns in the 2+ sample, first stage effects of twins-2 (top panel). First and second borns in the 3+ sample, first stage effects of twins-3 (bottom panel).

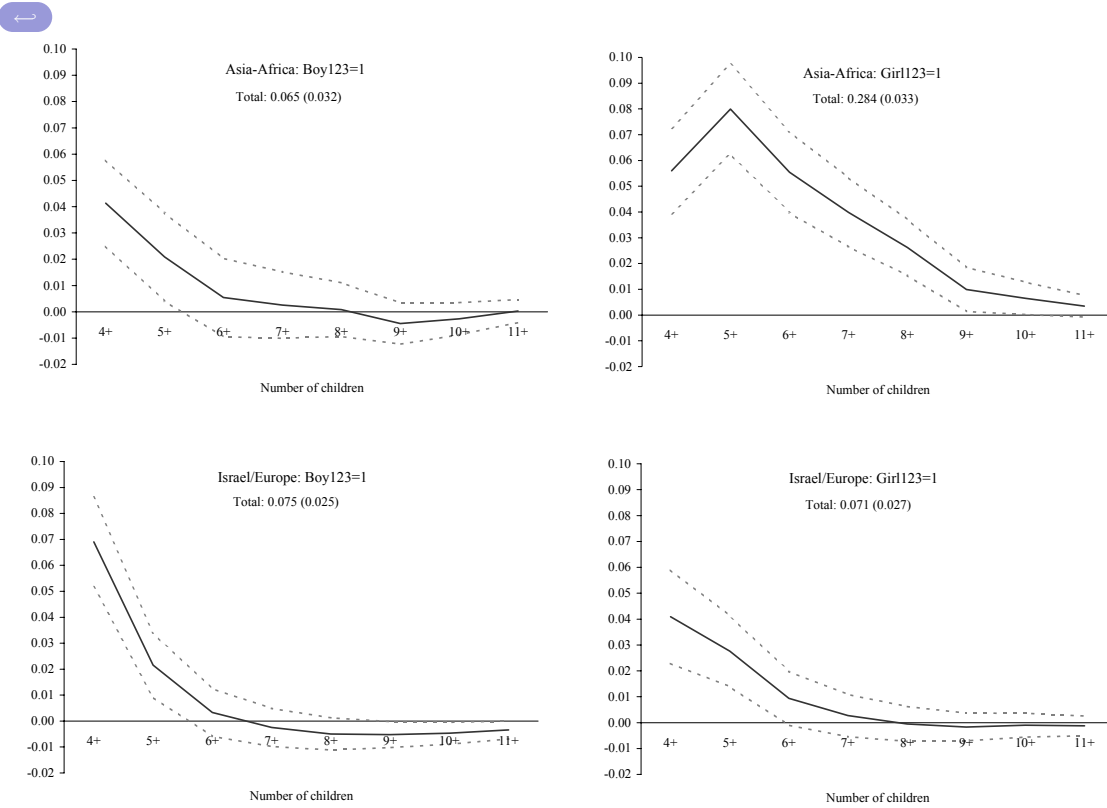


Figure 3: First and second borns 3+ sample. First stage effects by ethnicity and type of sex-mix.

Table 3.3: Estimates of the Quantity-Quality Trade-off								
Outcome	OLS		2SLS Instrument list					
	Basic controls (1)	All controls (2)	Twins (3)	Twins, TwinsAA (4)	Samesex (5)	Samesex, SamesexAA (6)	Twins, Samesex (7)	Twins, Samesex, SamesexAA (8)
Highest grade completed	-0.252 (0.005)	-0.145 (0.005)	0.174 (0.166)	0.105 (0.131)	0.318 (0.210)	0.315 (0.210)	0.237 (0.128)	0.186 (0.112)
Years of schooling $\geq 12$	-0.037 (0.001)	-0.029 (0.001)	0.030 (0.028)	0.024 (0.021)	0.001 (0.033)	0.002 (0.033)	0.017 (0.021)	0.016 (0.018)
Some College (age $\geq 24$ )	-0.049 (0.001)	-0.023 (0.001)	0.017 (0.052)	0.026 (0.046)	0.078 (0.054)	0.080 (0.055)	0.048 (0.037)	0.049 (0.035)
College graduate (age $\geq 24$ )	-0.036 (0.001)	-0.015 (0.001)	-0.021 (0.045)	-0.006 (0.041)	0.125 (0.053)	0.127 (0.053)	0.052 (0.032)	0.049 (0.031)

Notes: This table reports OLS estimates of the coefficient on sibship size in columns 1-2. 2SLS estimates appear in columns 3-8. Instruments with an 'AA' suffix are interaction terms with an AA dummy. The sample includes first borns from families with 2 or more births. OLS estimates for column 2 include indicators for age and sex. Estimates for columns 2-8 are from models that include the controls used for first stage models reported in the previous table. Robust standard errors are reported in parenthesis.