Instrumental Variables (Take 2): Causal Effects in a Heterogeneous World

Josh Angrist

MIT 14.387 (Fall 2014)

Our Constant-Effects Benchmark

- The traditional IV framework is the linear, constant-effects world discussed in Part 1. With Bernoulli treatment, D_i, we have

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

- η_i is not a regresson error (Y_{0i} is not independent of D_i), so OLS fails to capture causal effects
- Using an instrument, Z_i, that's independent of Y_{0i} but correlated with D_i, we have

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

- Constant FX focuses our attention on omitted variables bias, abstracting from more subtle concerns
- Time now to allow for the fact that $Y_{1i} - Y_{0i}$ need not be (in some cases, cannot be) the same for everyone

Sometimes You Get What You Need

- In a heterogeneous world, we distinguish between *internal validity* and *external validity*
- A good instrument – by definition – captures an internally valid causal effect. This is the causal effect on the group subject to (quasi-) experimental manipulation
- External validity is the predictive value of internally valid causal estimates in contexts beyond those generating the estimates
- Examples
  - Draft-lottery estimates of the effects of Vietnam-era military service
  - Quarter-of-birth estimates of the effects of schooling on earnings
  - Regression-discontinuity estimates of the effects of class size
- In 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 typically get effects over a limited but knowable range

Roadmap

1. An example: the effect of childbearing on labor supply
   - Two good instruments, two good answers
2. The theory of instrumental variables with heterogeneous potential outcomes
   - Notation and framework
   - The LATE Theorem
3. Implications for the design and analysis of field trials
   - The Bloom Result
   - Illustration: JTPA and MDVE
4. All about compliers: Kappa and QTE
5. Average causal response in models with variable treatment intensity
   - The ACR theorem and weighting function
   - A world of *continuous* activity
6. External Validity (first pass)

Children and Their Parents Labor Supply

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

$$\mathrm{Y}_i = \mathrm{Y}_{0i} + \mathrm{D}_i(\mathrm{Y}_{1i} - \mathrm{Y}_{0i}) = \alpha + \rho \mathrm{D}_i + \eta_i$$

  Constant FX? Here, ρ is *the thing that must be named*

- Dependent variables = employment, hours worked, weeks worked, earnings
  - $\mathrm{D}_i = 1[kids > 2]$ 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

$$\rho = \frac{Cov(\mathrm{Y}_i, \mathrm{Z}_i)}{Cov(\mathrm{D}_i, \mathrm{Z}_i)} = \frac{E[\mathrm{Y}_i|\mathrm{Z}_i = 1] - E[\mathrm{Y}_i|\mathrm{Z}_i = 0]}{E[\mathrm{D}_i|\mathrm{Z}_i = 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0]}$$

- **Results**

The LATE Framework

- Let $\text{Y}_i(d, z)$ denote the potential outcome of individual i were this person to have treatment status $\text{D}_i = d$ and instrument value $\text{Z}_i = z$
- Note the double-indexing: candidate instruments *might* have a direct effect on outcomes
- We assume, however, that IV initiates a causal chain: the instrument, Z_i, affects D_i, which in turn affects Y_i
- To flesh this out, we first define *potential treatment status,* indexed against Z_i :
  - D_{1i} is i's treatment status when $\text{Z}_i = 1$
  - D_{0i} is i's treatment status when $\text{Z}_i = 0$
- Observed treatment status is therefore

$$\text{D}_i = \text{D}_{0i} + (\text{D}_{1i} - \text{D}_{0i})\text{Z}_i$$

- The causal effect of Z_i on D_i is $\text{D}_{1i} - \text{D}_{0i}$

LATE Assumptions (Independence and First Stage)

Independence. The instrument is as good as randomly assigned:

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

- Independence means that draft lottery numbers 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 sufficient for a causal interpretation of the **reduced form**. Specifically,

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

- RF is the causal effect of the *instrument* on the dependent variable, but we have yet to link this to treatment

LATE Assumptions (Exclusion)

Our journey from RF to treatment effects starts here:

Exclusion. The instrument affects Y_i only through D_i:

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

- 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 Y_{1i} and Y_{0i} that satisfy the independence assumption
- Exclusion means draft lottery numbers affect earnings only via veteran status; quarter of birth affects earnings only through schooling; sex comp affects labor supply only by changing family size

LATE assumptions (Monotonicity)

A necessary technical assumption:

Monotonicity. $\mathrm{D}_{1i} \geq \mathrm{D}_{0i}$ for everyone (or vice versa).

- By virtue of monotonicity, $E\left[\mathrm{D}_{1i} - \mathrm{D}_{0i}\right] = P\left[\mathrm{D}_{1i} > \mathrm{D}_{0i}\right]$
- Interpreting monotonicity in latent-index models:

$$\mathrm{D}_i = \begin{array}{cc} 1 & \text{if } \gamma_0 + \gamma_1 \mathrm{Z}_i > v_i \\ 0 & \text{otherwise} \end{array}$$

where v_i is a random factor.

- This model characterizes potential treatment assignments as:

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

which clearly satisfy monotonicity

The LATE Theorem

Assumption recap:

- The independence assumption is sufficient for identification of the causal effects of the *instrument*
- The exclusion restriction means that the causal effect of the instrument on the dependent variable is due solely to the effect of the instrument on D_i
    - Exclusion is (or should be) more controversial than independence
- We also assume there *is* a first-stage; by virtue of monotonicity, this is the proportion 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 are subjects with $D_{1i} > D_{0i}$

- This language comes from randomized trials where Z_i is treatment assigned and D_i is treatment received (more on this soon)

- 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 is uninformative for always-takers and never-takers because treatment status for these types is unchanged by the instrument
  - An analogy: panel models with fixed effects identify treatment effects only for "changers"

- Of course, we can assume effects are the same for all three groups (a version of the constant-effects model)

The Compliant Subpopulation (cont.)

- From the fact that

$$\text{D}_i = \text{D}_{0i} + (\text{D}_{1i} - \text{D}_{0i})\text{Z}_i,$$

we see that:

$$\{\text{D}_i = 1\} =$$
$$\{\text{D}_{0i} = \text{D}_{1i} = 1\} \cup \{\{\text{D}_{1i} - \text{D}_{0i} = 1\} \cap \{\text{Z}_i = 1\}\}$$

- In other words . . .

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

- TOT is therefore a weighted average of effects on always-takers and compliers (compliers rolling $\text{Z}_i = 1$ are representative of all compliers)

IV in Randomized Trials

The compliance problem in RCTs: Some randomly assigned to the treatment group are untreated

- When compliance is voluntary, an *as-treated* analysis is contaminated by selection bias
- *Intention-to-treat* analyses preserve independence but are diluted by non-compliance
- IV solves this problem: Z_i, is a dummy variable indicating random assignment to the treatment group; D_i is a dummy indicating whether treatment received or taken
- No always-takers! (no controls are treated), so LATE = TOT

THE BLOOM RESULT.

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

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

Bloom Example 1: Training

The Job Training Partnership Act (JTPA) included a large randomized trial to evaluate the effect of training on earnings

- The JTPA *offered* treatment randomly; participation was voluntary
- Roughly 60 percent of those offered training received it
- IV setup
    - D_i indicates those who received JTPA services
    - Z_i indicates the random offer of treatment
    - Y_i is earnings in the 30 months since random assignment
- The first-stage here is approximately the compliance rate

$$E[\mathrm{D}_i|\mathrm{Z}_i = 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0] \cong P[\mathrm{D}_i = 1|\mathrm{Z}_i = 1] = .60$$

(.62 of Z_i =1 group trained; .02 of Z_i =0 group also trained)

- **Table 4.4.1** Selection bias in OLS (as delivered); ITT (as assigned) is diluted; IV (TOT) is . . . just right!

Bloom Example 2: Battered Wives

What's the best police response to domestic violence? The Minneapolis Domestic Violence Experiment (MDVE; Sherman and Berk, 1984) boldly tried to find out . . .

- Police were randomly assigned to advise, separate, or arrest
- Substantial compliance problems as officers made their own judgements in the field

Table 1: Assigned and Delivered Treatments in Spousal Assault Cases

Assigned Treatment	Delivered Treatment			
		Coddled		
	Arrest	Advise	Separate	Total
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

- 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)

How Many Compliers You Got?

- Given monotonicity, we have

$$\begin{aligned} P[\mathrm{D}_{1i} > \mathrm{D}_{0i}] &= E[\mathrm{D}_{1i} - \mathrm{D}_{0i}] = E[\mathrm{D}_{1i}] - E[\mathrm{D}_{0i}] \\ &= E[\mathrm{D}_i|\mathrm{Z}_i = 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0] \end{aligned}$$

  The first stage tells us how many!
- And among the treated?
  - Start with the definition of conditional probability:

$$\begin{aligned} P[\mathrm{D}_{1i} &> \mathrm{D}_{0i}|\mathrm{D}_i = 1] \\ &= \frac{P[\mathrm{D}_i = 1|\mathrm{D}_{1i} > \mathrm{D}_{0i}]P[\mathrm{D}_{1i} > \mathrm{D}_{0i}]}{P[\mathrm{D}_i = 1]} \\ &= \frac{P[\mathrm{Z}_i = 1](E[\mathrm{D}_i|\mathrm{Z}_i = 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0])}{P[\mathrm{D}_i = 1]} \end{aligned}$$

    An easy calculation, proportional to the first stage
- **Sample complier counts**

Counting ...

Table 4.4.2
Probabilities of compliance in instrumental variables studies

Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	Compliance Probabilities: $P[D_1 > D_0 \mid D=1]$ (8)	Compliance Probabilities: $P[D_1 > D_0 \mid D=0]$ (9)
Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
		Non-white men born in 1950	.163	.060	.534	.197	.033
More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
	First two children are same sex		.381	.060	.506	.080	.048
High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

Characterizing Compliers

- Are sex-comp compliers more or less likely to be college graduates (indicated by $x_{1i} = 1$) than other women?

$$
\begin{aligned}
& \frac{P[x_{1i} = 1 | \mathrm{D}_{1i} > \mathrm{D}_{0i}]}{P[x_{1i} = 1]} \\
= \quad & \frac{P[\mathrm{D}_{1i} > \mathrm{D}_{0i} | x_{1i} = 1]}{P[\mathrm{D}_{1i} > \mathrm{D}_{0i}]} \\
= \quad & \frac{E[\mathrm{D}_i | \mathrm{Z}_i = 1, x_{1i} = 1] - E[\mathrm{D}_i | \mathrm{Z}_i = 0, x_{1i} = 1]}{E[\mathrm{D}_i | \mathrm{Z}_i = 1] - E[\mathrm{D}_i | \mathrm{Z}_i = 0]}
\end{aligned}
$$

- The relative likelihood a complier is a college grad is given by the ratio of the first stage for college grads to the overall first stage

- **Sample complier characteristics**

Characterizing ...

TABLE 4.4.3
Complier characteristics ratios for twins and sex composition instruments

		Twins at Second Birth		First Two Children Are Same Sex	
Variable	$P[x_{1i} = 1]$ (1)	$P[x_{1i} = 1 \mid D_{1i} > D_{0i}]$ (2)	$P[x_{1i} = 1 \mid D_{1i} > D_{0i}] / P[x_{1i} = 1]$ (3)	$P[x_{1i} = 1 \mid D_{1i} > D_{0i}]$ (4)	$P[x_{1i} = 1 \mid D_{1i} > D_{0i}] / P[x_{1i} = 1]$ (5)
Age 30 or older at first birth	.0029	.004	1.39	.0023	.995
Black or hispanic	.125	.103	.822	.102	.814
High school graduate	.822	.861	1.048	.815	.998
College graduate	.132	.151	1.14	.0904	.704

Notes: The table reports an analysis of complier characteristics for twins and sex composition instruments. The ratios in columns 3 and 5 give the relative likelihood that compliers have the characteristic indicated at left. Data are from the 1980 census 5 percent sample, including married mothers aged 21–35 with at least two children, as in Angrist and Evans (1998). The sample size is 254,654 for all columns.

Distribution Treatment Effects

- LATE, $E[\text{Y}_{1i}-\text{Y}_{0i}|\text{D}_{1i}>\text{D}_{0i}]$, is an average causal effect. We turn now to the *distribution* of potential outcomes for compliers.
- Abadie (2002) showed that, for any measurable function, $g(\text{Y}_{ji})$,

$$\frac{E[\text{D}_i g(\text{Y}_i)|\text{Z}_i=1]-E[\text{D}_i g(\text{Y}_i)|\text{Z}_i=0]}{E[\text{D}_i|\text{Z}_i=1]-E[\text{D}_i|\text{Z}_i=0]} = E[g(\text{Y}_{1i})|\text{D}_{1i}>\text{D}_{0i}]$$

$$\frac{E[(1-\text{D}_i)g(\text{Y}_i)|\text{Z}_i=1]-E[(1-\text{D}_i)g(\text{Y}_i)|\text{Z}_i=0]}{E[1-\text{D}_i|\text{Z}_i=1]-E[1-\text{D}_i|\text{Z}_i=0]}$$

$$= \quad E[g(\text{Y}_{0i})|\text{D}_{1i}>\text{D}_{0i}]$$

- Set $g(\text{Y}_{ji})=\text{Y}_{ji}$ to capture marginal mean outcomes; set $g(\text{Y}_{ji})=1[\text{Y}_{ji}<c]$ to capture distributions:

$$E\{1[\text{Y}_{ji}<c]|\text{D}_{1i}>\text{D}_{0i}\}=P[\text{Y}_{ji}<c|\text{D}_{1i}>\text{D}_{0i}$$

- **Charter school IV and the distribution of test scores**

Charter School 2SLS

Table 3: Lottery Estimates of Effects on 10th-Grade MCAS Scores by Projected Senior Year

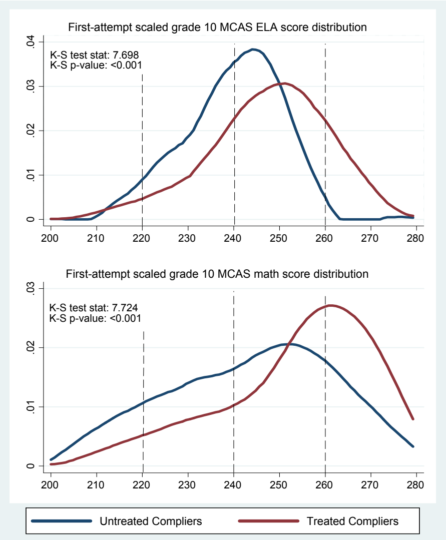
	Charter Enrollment			MCAS Scores			
	Non-offered mean	Immediate Offer	Waitlist Offer	Non-charter Mean	Immediate Offer	Waitlist Offer	Charter Effect
Subject	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 2006-2013 (MCAS outcome sample)							
Standardized ELA	0.104	0.373***	0.239***	-0.285	0.148***	0.136***	0.408***
	[0.306]	(0.047)	(0.042)	[0.833]	(0.046)	(0.044)	(0.102)
N							3685
Standardized Math	0.106	0.374***	0.241***	-0.233	0.221***	0.152***	0.592***
	[0.307]	(0.048)	(0.041)	[0.911]	(0.058)	(0.054)	(0.117)
N							3629
Panel B: 2006-2012 (NSC outcome sample)							
Standardized ELA	0.103	0.369***	0.231***	-0.296	0.111**	0.074	0.301**
	[0.305]	(0.055)	(0.049)	[0.830]	(0.054)	(0.050)	(0.123)
N							3009
Standardized Math	0.104	0.371***	0.234***	-0.241	0.187***	0.126**	0.504***
	[0.306]	(0.055)	(0.049)	[0.893]	(0.068)	(0.064)	(0.141)
N							2965

Notes: This table reports 2SLS estimates of the effects of Boston charter attendance on 10th-grade MCAS test scores. The sample includes students projected to graduate between 2006 and 2013. The endogenous variable is an indicator for charter attendance in 9th or 10th grade. The instruments are immediate and waitlist offer dummies. Immediate offer is equal to one when a student is offered a seat in any charter school immediately following the lottery, while waitlist offer is equal to one for students offered seats later. All models control for risk sets, 10th grade calendar year dummies, race, sex, special education, limited English proficiency, subsidized lunch status, and a female by minority dummy. Standard errors are clustered at the school-year level in 10th grade. Means are for non-charter students.
*significant at 10%; **significant at 5%; ***significant at 1%

From Angrist, et al. (July 2013) LTO

Score Distributions

Figure 1: Complier Distributions for MCAS Scaled Scores



All About Kompliers

Theorem

ABADIE KAPPA. Suppose the assumptions of the LATE theorem hold conditional on covariates, X_i. *Let* $g(Y_i, D_i, X_i)$ *be any measurable function of* (Y_i, D_i, X_i) *with finite expectation. Define*

$$\kappa_i = 1 - \frac{D_i(1 - Z_i)}{1 - P(Z_i = 1|X_i)} - \frac{(1 - D_i)Z_i}{P(Z_i = 1|X_i)}.$$

Then

$$E[g(Y_i, D_i, X_i)|D_{1i} > D_{0i}] = \frac{E[\kappa_i g(Y_i, D_i, X_i)]}{E[\kappa_i]}. \quad (1)$$

Proof.

By monotonicity, those with $D_i(1-Z_i) = 1$ are always-takers because they have $D_{0i} = 1$, while those with $(1-D_i)Z_i = 1$ are never-takers because they have $D_{1i} = 0$. Kappa removes means for always-takers and never-takers from marginal means, leaving the average for compliers. □

Using Kappa

- Sketch of proof: Kappa uses this relation, true by monotonicity:

$$E|Y|c] = \frac{1}{P(c)} \{E[Y] - E[Y|AT]P(AT) - E[Y|NT]P(NT)\}$$

- Who cares? *Conditional on compliance, treatment is ignorable*:

$$\{Y_{1i}, Y_{0i}\} \amalg D_i | D_{1i} > D_{0i},$$

so we can use κ to approximate a causal CEF, by solving:

$$(\alpha, \beta) = \arg\min_{a,b} E\{\kappa_i (Y_i - h\left[aD_i + X_i'b\right])^2\} \tag{2}$$

for any linear *or nonlinear* approx function, $h[\alpha D_i + X_i'\beta]$

- Suppose, for example,

$$h\left[\alpha D_i + X_i'\beta\right] = \Phi\left[\alpha D_i + X_i'\beta\right]$$

This gives "best Probit approximation" to a causal CEF with endogenous treatment

Quantile Treatment Effects

- QR models conditional distributions. Assume:

$$Q_\tau(\mathrm{Y}_i|\mathrm{X}_i) = \gamma_\tau'\mathrm{X}_i$$

Then γ_τ solves

$$\gamma_\tau = \arg\min_c E\{\rho_\tau(\mathrm{Y}_i - \mathrm{X}_i'c)\}$$

where $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ is the "check function."

  - If the CQF is nonlinear, QR provides a regression-like minimum weighted MSE approx to it; see Angrist, Chernozhukov and Fernandez-Val, 2006)

- Kappa captures a causal *quantile treatment effect*, α_τ, in

$$Q_\tau(\mathrm{Y}_i|\mathrm{X}_i, \mathrm{D}_i, \mathrm{D}_{1i}{>}\mathrm{D}_{0i}) = Q_\tau(\mathrm{Y}_{Di}|\mathrm{X}_i, \mathrm{D}_{1i}{>}\mathrm{D}_{0i}) = \alpha_\tau\mathrm{D}_i + \mathrm{X}_i'\beta_\tau, \tag{3}$$

by solving:

$$(\alpha_\tau, \beta_\tau) = \arg\min_{a,b} E\{\kappa_i\rho_\tau(\mathrm{Y}_i - a\mathrm{D}_i - \mathrm{X}_i'b)\}$$

- **QR 'n QTE**

The JTPA Redux

Means and Standard Deviations

	Entire Sample	Assignment: Treatment	Assignment: Control	Assignment: Diff. (t-stat.)	Treatment: Trainees	Treatment: Non-trainees	Treatment: Diff. (t-stat.)
A. Men							
Number of observations	5,102	3,399	1,703		2,136	2,966	
Treatment							
Training	.42	.62	.01	.61			
	[.49]	[.48]	[.11]	(70.34)			
Outcome variable							
30 month earnings	19,147	19,520	18,404	1,116	21,455	17,485	3,970
	[19,540]	[19,912]	[18,760]	(1.96)	[19,864]	[19,135]	(7.15)
Baseline Characteristics							
Age	32.91	32.85	33.04	−.19	32.76	33.02	−.26
	[9.46]	[9.46]	[9.45]	(−.67)	[9.64]	[9.32]	(−.95)
High school or GED	.69	.69	.69	−.00	.71	.68	.03
	[.45]	[.45]	[.45]	(−.12)	[.44]	[.45]	(2.46)
Married	.35	.36	.34	.02	.37	.34	.03
	[.47]	[.47]	[.46]	(1.64)	[.47]	[.46]	(2.82)
Black	.25	.25	.25	.00	.26	.25	.01
	[.44]	[.44]	[.44]	(.04)	[.44]	[.43]	(.48)
Hispanic	.10	.10	.09	.01	.10	.09	.01
	[.30]	[.30]	[.29]	(.70)	[.31]	[.29]	(1.60)
Worked less than 13 weeks in past year	.40	.40	.40	.00	.40	.40	−.00
	[.47]	[.47]	[.47]	(.56)	[.47]	[.47]	(−.32)

TABLE II

QUANTILE REGRESSION AND OLS ESTIMATES

Dependent Variable: 30-month Earnings

		Quantile				
	OLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of Training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	−2,354	−134	−500	−2,084	−3,576	−3,609
	(626)	(194)	(324)	(684)	(1,087)	(1,331)
Hispanic	251	91	278	925	−877	−85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked less than 13	−6,582	−1,090	−3,097	−7,610	−9,834	−9,951
weeks in past year	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	−216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)

TABLE III

QUANTILE TREATMENT EFFECTS AND 2SLS ESTIMATES

Dependent Variable: 30-month Earnings

	2SLS	Quantile 0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of Training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	−2,349	−171	−377	−2,656	−4,182	−3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked less than 13 weeks in past year	−6,575	−1,932	−4,195	−7,009	−9,289	−9,078
	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	−134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)

Questions of Variable Intensity

Average Causal Response

Variable intensity s_i takes on values in the set $\{0, 1, ..., \bar{s}\}$. There are $\bar{s}$ unit causal effects, $Y_{si} - Y_{s-1,i}$.

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

The ACR Theorem

Angrist and Imbens (1995) show

$$\frac{E[\mathrm{Y}_i|\mathrm{Z}_i = 1] - E[\mathrm{Y}_i|\mathrm{Z}_i = 0]}{E[\mathrm{S}_i|\mathrm{Z}_i = 1] - E[\mathrm{S}_i|\mathrm{Z}_i = 0]} = \sum_{s=1}^{\bar{s}} \omega_s E[Y_{si} - Y_{s-1,i}|\mathrm{S}_{1i} \geq s > \mathrm{S}_{0i}]$$

where

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

The weights, ω_s, are non-negative and sum to 1.

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

The ACR Weighting Function

- The size of the group of compliers at point s is:

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

- By Independence, this is an observed CDF difference:

$$P[\mathrm{s}_{0i} < s] - P[\mathrm{s}_{1i} < s] = P[\mathrm{s}_i < s|\mathrm{z}_i = 0] - P[\mathrm{s}_i < s|\mathrm{z}_i = 1]$$

- Finally, because the mean of a non-negative random variable is one minus the CDF,

$$\begin{aligned} & E\left[\mathrm{s}_i|\mathrm{z}_i = 1\right] - E\left[\mathrm{s}_i|\mathrm{z}_i = 0\right] \\ = & \sum_{j=1}^{\bar{s}} \left(P\left[\mathrm{s}_i < j|\mathrm{z}_i = 0\right] - P\left[\mathrm{s}_i < j|\mathrm{z}_i = 1\right]\right) = \sum_{j=1}^{\bar{s}} P\left[\mathrm{s}_{1i} \geq j > \mathrm{s}_{0i}\right] \end{aligned}$$

- The ACR weighting function is given by the difference in the CDFs of treatment intensity with the instrument switched off and on, normalized by the first-stage

QOB Estimates of the Returns to Schooling

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

- Angrist and Krueger (1991) do Wald
- s_i is years of schooling
- z_i indicates men born in the 1st quarter (old at school entry)
- Diffs in CDFs by QOB (first vs. fourth quarter births)$\Longrightarrow$

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)
Panel A: Wald Estimates for 1970 Census—Men Born 1920–1929[a]			
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[b]			.0797 (.0005)
Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939			
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

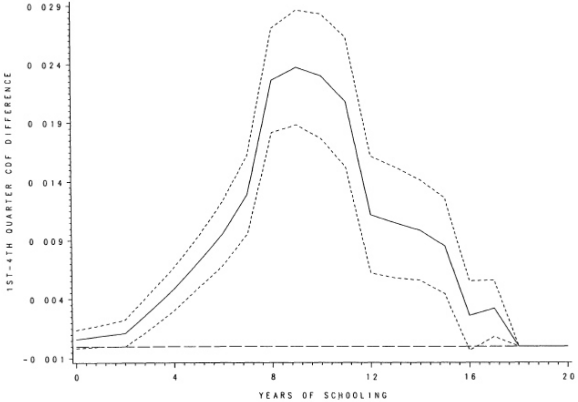


Figure 3. First–Fourth Quarter Difference in Schooling CDF (Men Born 1920–1929, Data From the 1970 Census). Dotted lines are 95% confidence intervals.

More Variable Treatment Intensities

- Returns to schooling again, identified using compulsory attendance and child labor laws (Acemoglu and Angrist, 2000)
- Class size (Angrist and Lavy, 1999; Krueger, 1999)
  - Y_i is test score; S_i is class size
  - Z_i is Maimonides Rule (regression-discontinuity) 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 birthweight; 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

So Long and Thanks for All the Fish

- Let $q_i(p)$ denote quantity demanded in market i at hypothetical price p, a continuous function
- The slope of this demand curve is $q_i'(p)$; with quantity and price measured in logs, this is an elasticity
- The Wald estimator using a $stormy_i$ instrument is

$$\frac{E[q_i|stormy_i = 1] - E[q_i|stormy_i = 0]}{E[p_i|stormy_i = 1] - E[p_i|stormy_i = 0]} = \frac{\int E[q_i'(t)|\ p_{1i} \geq t > p_{0i}]P[p_{1i} \geq t > p_{0i}]dt}{\int P[p_{1i} \geq t > p_{0i}]dt},$$

where p_{1i} and p_{0i} are potential prices indexed by $stormy_i$

- This is a weighted average derivative with weighting function $P[p_{1i} \geq t > p_{0i}] = P[p_i \leq t|z_i = 0] - P[p_i \leq t|z_i = 1]$ at price t

Continuous Special Cases

1. *Linear*: $q_i(p) = \alpha_{0i} + \alpha_{1i}p$, for random coefficients, α_{0i} and α_{1i}. Then, we have,

$$\frac{E[q_i|stormy_i = 1] - E[q_i|stormy_i = 0]}{E[p_i|stormy_i = 1] - E[p_i|stormy_i = 0]} = \frac{E[\alpha_{1i}(p_{1i} - p_{0i})]}{E[p_{1i} - p_{0i}]}, \tag{4}$$

   a weighted average of α_{1i}, with weights proportional to the price change induced by the weather in market i.

2. *Additive nonlinear*:

$$q_i(p) = Q(p) + \eta_i \tag{5}$$

   By this we mean $q_i'(p) = Q'(p)$ every day or in every market. ACR becomes,

$$\int Q'(t)\omega(t)dt, \text{where } \omega(t) = \frac{P[p_{1i} \geq t > p_{0i}]}{\int P[p_{1i} \geq r > p_{0i}]dr}$$

- Case 1 emphasizes heterogeneity; Case 2 focuses on nonlinearity
- Y'allah, let's fish!

The Elasticity of Demand for Fish

TABLE 5

Two-stage-least-squares estimates of demand function with stormy and mixed as instruments

Variable	est.	(s.e.)	est.	(s.e.)
Av. price effect	–1·01	(0·42)	–0·947	(0·46)
Monday			–0·013	(0·18)
Tuesday			–0·51	(0·18)
Wednesday			–0·56	(0·17)
Thursday			0·10	(0·17)
Weather on shore			0·02	(0·16)
Rain on shore			0·07	(0·16)

Note: Standard errors are reported in parentheses.

book versions of instrumental variables with the demand function equal to

$$\ln q_t^d(p) = \beta_0 + \beta_1 \cdot \ln p + \beta_2 \cdot x + \varepsilon_t^d,$$

again using observations with $z_t = z$ and $z_t = z'$ and using the indicator for $z_t = z'$ as the instrument. Note that under the additivity assumption (Assumption 5) the value of $\tilde{\beta}(z, z', x)$ does not vary with x.

Table 5 presents two-stage-least-squares estimates using both the stormy and mixed instruments, with and without the additional regressors. These estimates are close to those based on the single instruments, with slightly lower standard errors.

Shelter from the Storm (or Mixed Conditions)

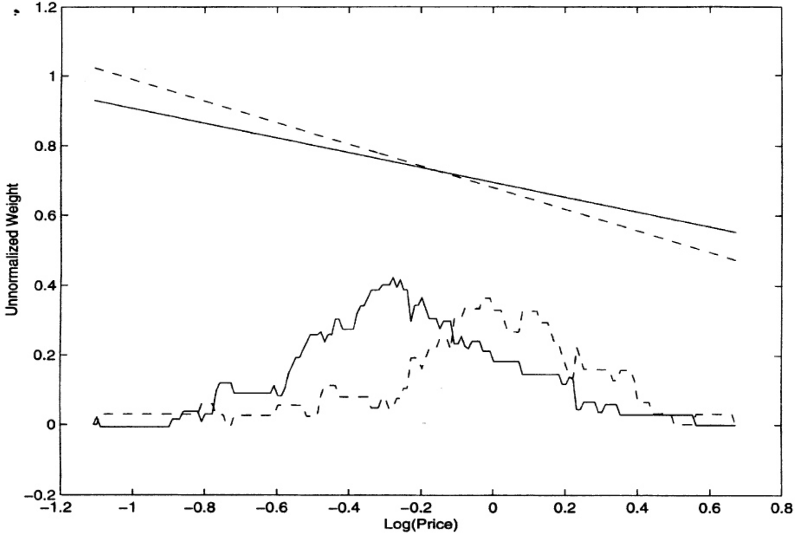


FIGURE 4

Weight and regression functions for different binary instruments

External Validity (Prequel)

How to assess the predictive value of IV estimates?

- Angrist, Lavy, and Schlosser (2010) compare IV estimates of the QQ trade-off using twins and sex-composition instruments
  - Twins and sex composition affect very different types of people: with twins, there are no never-takers, so LATE is

    $$E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{D}_i = 0]; \text{ where } \mathrm{D}_i \text{ indicates more than two}$$

    Twins compliers want to stop at two, hence they're more educated than samesex compliers (and others)
  - Twinning mostly causes a one-child shift; while sex-composition increases childbearing at high parities: **twins 1st stage**; **QQ samesex 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

- The IV paradigm 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 (the biomed RCT world has been slow to absorb this; e.g., AIDS vaccine trials)
    - A flexible strategy for the analysis of observational designs
- Kappa weighting extends the LATE framework to nonlinear and quantile models
- IV produces weighted averages of ordered and contnuous treatment effects; the weighting function describes the range contributing to a particular estimate
- Up next: DD and RD ... these too are often IV!

Tables and Figures

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*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.

TABLE 4.4.1
Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status (OLS)		Comparisons by Assignment Status (ITT)		Instrumental Variable Estimates (IV)	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Notes: Authors' tabulation of JTPA study data. The table reports OLS, ITT, and IV estimates of the effect of subsidized training on earnings in the JTPA experiment. Columns 1 and 2 show differences in earnings by training status; columns 3 and 4 show differences by random-assignment status. Columns 5 and 6 report the result of using random-assignment status as an instrument for training. The covariates used for columns 2, 4, and 6 are high school or GED, black, Hispanic, married, worked less than 13 weeks in past year, AFDC (for women), plus indicators for the JTPA service strategy recommended, age group, and second follow-up survey. Robust standard errors are shown in parentheses. There are 5,102

Table 4.4.2
Probabilities of compliance in instrumental variables studies

Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D = 1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z = 1]$ (7)	Compliance Probabilities $P[D_1 > D_0 \mid D = 1]$ (8)	$P[D_1 > D_0 \mid D = 0]$ (9)
Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
		Non-white men born in 1950	.163	.060	.534	.197	.033
More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
	First two children are same sex		.381	.060	.506	.080	.048
High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

Table 4.4.3
Complier characteristics ratios for twins and sex composition instruments

		Twins at Second Birth		First Two Children Are Same Sex	
Variable	$P[x_{1i}=1]$ (1)	$P[x_{1i}=1 \mid \text{D}_{1i} > \text{D}_{0i}]$ (2)	$P[x_{1i}=1 \mid \text{D}_{1i} > \text{D}_{0i}]/P[x_{1i}=1]$ (3)	$P[x_{1i}=1 \mid \text{D}_{1i} > \text{D}_{0i}]$ (4)	$P[x_{1i}=1 \mid \text{D}_{1i} > \text{D}_{0i}]/P[x_{1i}=1]$ (5)
Age 30 or older at first birth	.0029	.004	1.39	.0023	.995
Black or hispanic	.125	.103	.822	.102	.814
High school graduate	.822	.861	1.048	.815	.998
College graduate	.132	.151	1.14	.0904	.704

Notes: The table reports an analysis of complier characteristics for twins and sex composition instruments. The ratios in columns 3 and 5 give the relative likelihood that compliers have the characteristic indicated at left. Data are from the 1980 census 5 percent sample, including married mothers aged 21–35 with at least two children, as in Angrist and Evans (1998). The sample size is 254,654 for all columns.

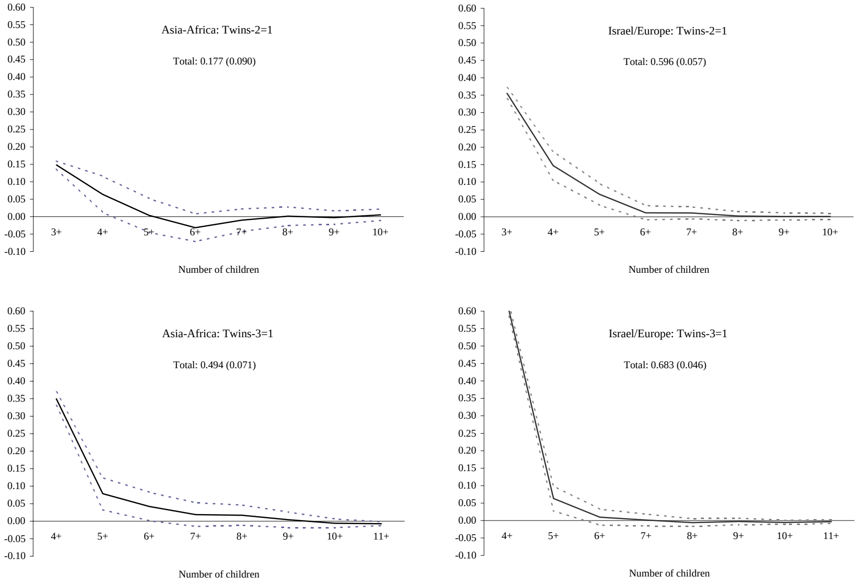


Figure 1: First borns in the 2+ sample, first stage effects of twins-2 (top panel). First and second borns in the 3+ sample, first stage effects of twins-3 (bottom panel).

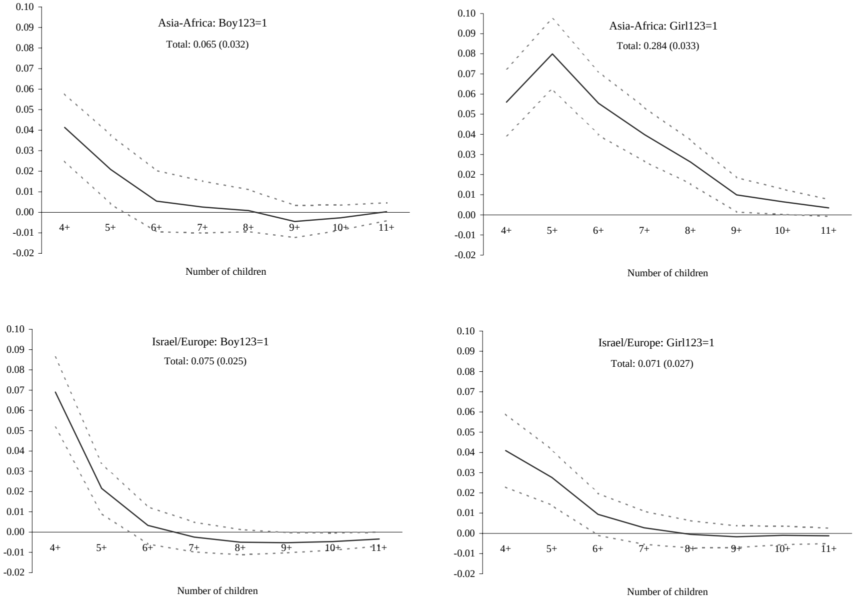


Figure 3: First and second borns 3+ sample. First stage effects by ethnicity and type of sex-mix.

Table 3.3: Estimates of the Quantity-Quality Trade-off

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