MIT 14.32/14.320 **Spring 2023**
J. Angrist

Lecture Note 11

Instrumental Variables and 2SLS for OVB

1 Recap: Regression and the CIA

- Recall the potential outcomes model for effects of private university attendance (P_i) on wages:
  - $Y_{0i} = \alpha + \eta_i$, where $E[Y_{0i}] = \alpha$; assume $Y_{1i} - Y_{0i} = \delta$. This means:
  $$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})P_i = \alpha + \delta P_i + \eta_i \tag{1}$$
  where δ is a new Greek name for the causal effect of private college attendance
- The CEF of Y_i given P_i is linear, so the regression of Y_i on P_i generates a difference in means:
  $$\frac{C(Y_i, P_i)}{V(P_i)} = E[Y_i|P_i = 1] - E[Y_i|P_i = 0] = \delta + \{E[\eta_i|P_i = 1] - E[\eta_i|P_i = 0]\}$$
  - Uncontrolled comparisons equal the causal effect of interest plus selection bias
- Regression captures causal effects by invoking a *conditional independence assumption* (CIA):
  $$E[\eta_i|P_i, X_i] = E[\eta_i|X_i] = \gamma' X_i, \tag{2}$$
  for a set of observed control variables, X_i. Equivalently,
  $$\eta_i = \gamma' X_i + u_i$$
  where u_i and X_i are uncorrelated by construction and u_i and P_i are uncorrelated by the CIA (look back at Sec. 4 of LN6)
  - In LN6 and MM Chpt 2, the variables in X_i are dummies for Barrons selectivity groups or simply the average SAT scores of colleges applied to and dummies for number of apps submitted
- The CIA yields a causal regression model:
  $$Y_i = \alpha + \delta P_i + \gamma' X_i + u_i, \tag{3}$$
  that's free of OVB
- Nice work if you can get it! Alas, you won't always be so lucky in the controls department. What's a young 'metrics master to do?

2 Instrumental Variables Eliminate Selection Bias

2.1 Waiting for Superman

- Many children in large urban school districts leave school with poor reading and math skills. Economists have shown that lack of basic skills perpetuates poverty.
- Charter schools—privately managed public schools—offer a possible solution
  - Charters are funded by the host district, but free to deviate from local district requirements (such as restrictions on who can be hired to teach) and to opt out of the restrictive teachers' union contracts that determine pay and job security at traditional public schools
  - The Knowledge is Power Program (KIPP) is iconic in the charter universe, serving mostly urban minority students. KIPP's "No Excuses/High Expectations" charter recipe features a long school day and year, data-driven instruction, Teach for American (TFA) interns and extensive tutoring, and an emphasis on classroom comportment
  - KIPP students tend to do better than other students in the district they hail from. But this may reflect selection bias. The KIPP parent, after all, knows enough to find their child a coveted seat at KIPP

  Here's what a leading critic says:

  > KIPP students, as a group, enter KIPP with substantially higher achievement than the typical achievement of schools from which they came. . . . [T]eachers told us either that they referred students who were more able than their peers, or that the most motivated and educationally sophisticated parents were those likely to take the initiative . . . and enroll in KIPP.

- The charter selection bias story
  - Let D_i denote attendance at KIPP and Y_i be an achievement test outcome
  - Under constant causal effects,

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = \alpha + \lambda D_i + \eta_i, \tag{4}$$

  where λ ("lambda") is a Greek name for the causal effect of interest. Once again, simple comparisons are confounded by selection bias:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \lambda + \{E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]\}$$

2.2 Defining Instruments (on the path to causal KIPP effects)

- An <u>instrumental variable</u> (Z_i) for KIPP attendance is correlated with D_i but uncorrelated with Y_{0i}. **In the context of equation (4), instrument Z_i is assumed to satisfy two conditions:**

$$C(Z_i,\, D_i) \neq 0 \tag{5}$$
$$C(Z_i,\, Y_{0i}) = C(Z_i,\, \eta_i) = 0 \tag{6}$$

  The second of these is often called an *independence assumption* or an *exclusion restriction* (or both). These two conditions imply:

$$C(Z_i,\, Y_i) = C(Z_i,\, D_i)\lambda \tag{7}$$
$$\lambda = \frac{C(Z_i,\, Y_i)}{C(Z_i,\, D_i)} = \frac{C(Z_i,\, Y_i)/V(Z_i)}{C(Z_i,\, D_i)/V(Z_i)} \tag{8}$$

- IV logic: the only reason Z_i is correlated with Y_i is because it's correlated with D_i
- Because we can solve for λ given information on the joint distribution of $\{Y_i, D_i, Z_i\}$, parameter λ is said to be *identified* by instrument Z_i.
  - *Identification* problems are distinct from *estimation* problems
- The top and bottom of the IV ratio, (8), are central to the IV story, so we christen them:

$$\frac{Reduced\ \ Form}{First\ \ Stage} = \frac{C(Z_i,\, Y_i)/V(Z_i)}{C(Z_i,\, D_i)/V(Z_i)} = \frac{\rho}{\phi} = \lambda$$

- The *IV estimator* can be written as a ratio of estimated regression coefficients:

$$\hat{\lambda}_{IV} = \frac{\hat{\rho}}{\hat{\phi}}, \tag{9}$$

  where $\hat{\rho}$ and $\hat{\phi}$, are the *estimated reduced form* and *estimated first stage* coefficients, computed by OLS regressions of Y_i and D_i on Z_i
  - Given assumptions (5) and (6), $\hat{\lambda}_{IV}$ is a consistent (though not unbiased) estimator of λ, with an asymptotically Normal sampling distribution and standard error formulas that we will derive later

An alternate path to IV: Long Regression w/o Controls

- We'd like to control for factors like ability and family background when estimating KIPP effects on learning
  - Suppose this is the "long regression" we'd like to run:

$$Y_i = \alpha_l + \lambda_l D_i + \gamma' A_i + u_i, \tag{10}$$

    where A_i is a vector of (awesome) ability and family background controls, and λ_l is the KIPP effect of interest
- Alas, important control variables are unobserved. For example, ability is hard to measure.
  - Instrumental Variables (IV) methods allow us to recover the coefficient of interest in a long regression when long-regression controls are unavailable. In addition to the first stage requirement, condition (5), this justification for IV requires that Z_i be uncorrelated with omitted variables and the residual thats left over. That is, we replace (6) with $C(Z_i, A_i) = C(Z_i, u_i) = 0$ in (10). This yields:

$$\lambda_l = \frac{C(Z_i,\, Y_i)}{C(Z_i,\, D_i)} \tag{11}$$

2.3 Playing the KIPP Lottery

- So what's the instrument for KIPP? Like all Massachusetts charter schools, KIPP Lynn assigns seats by lottery when over-subscribed
  - A research jackpot!
- In this case, instrument Z_i is a dummy variable indicating the set of KIPP applicants randomly offered a KIPP seat

- Because the lottery is how most KIPP applicants get seated there, (5) is satisfied
- Because lottery offers are randomly assigned, they're likely to be independent of potential outcomes, satisfying (6)

- Bernoulli (dummy) instruments generate a useful simplification of (8):

$$\lambda = \frac{C(Z_i, Y_i)/V(Z_i)}{C(Z_i, D_i)/V(Z_i)} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \tag{12}$$

  - We can therefore construct IV estimates using a ratio of differences in means:

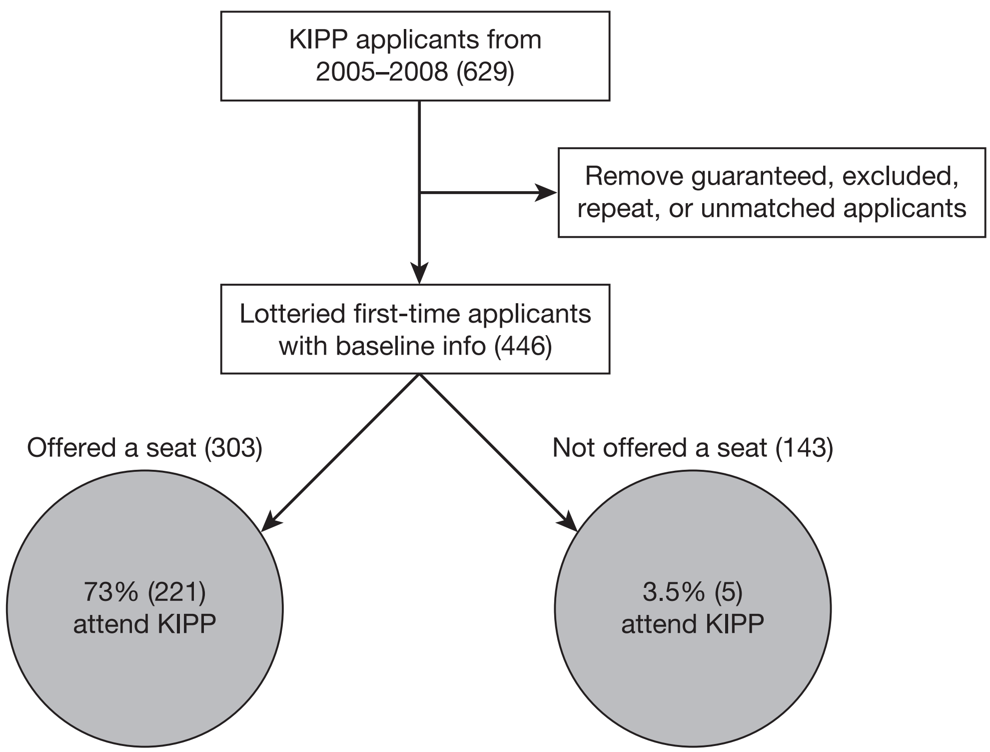
$$\hat{\lambda}_{IV} = (\bar{Y}_1 - \bar{Y}_0)/(\bar{D}_1 - \bar{D}_0), \tag{13}$$

    where $\bar{Y}_j$ and $\bar{D}_j$ are sample means of Y_i and D_i conditional on $Z_i = j$
  - The formula in (13) is called a *Wald estimator*, after Wald (1940)
  - Since D_i is also Bernoulli, $\bar{D}_j$ is equal to the KIPP enrollment rate conditional on $Z_i = j$

The KIPP First Stage

- This figure diagrams the lottery first stage for applicants to KIPP Lynn, applying for 5th and 6th grade seats in the years 2005-2008

FIGURE 3.1
Application and enrollment data from KIPP Lynn lotteries



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

- Here, $\bar{D}_1 = .73$ and $\bar{D}_0 = .035$, generating a large first-stage
- Lotteries ensure *ceteris paribus* comparisons between applicants with $Z_i = 0$ and $Z_i = 1$
- The table below describes KIPP's 2005-8 applicants for 5th and 6th grade seats; outcomes are from the end of these grades (for the 371 tested lottery applicants; baseline scores are from 4th grade). Test scores are standardized to the state mean and variance.

TABLE 3.1
Analysis of KIPP lotteries

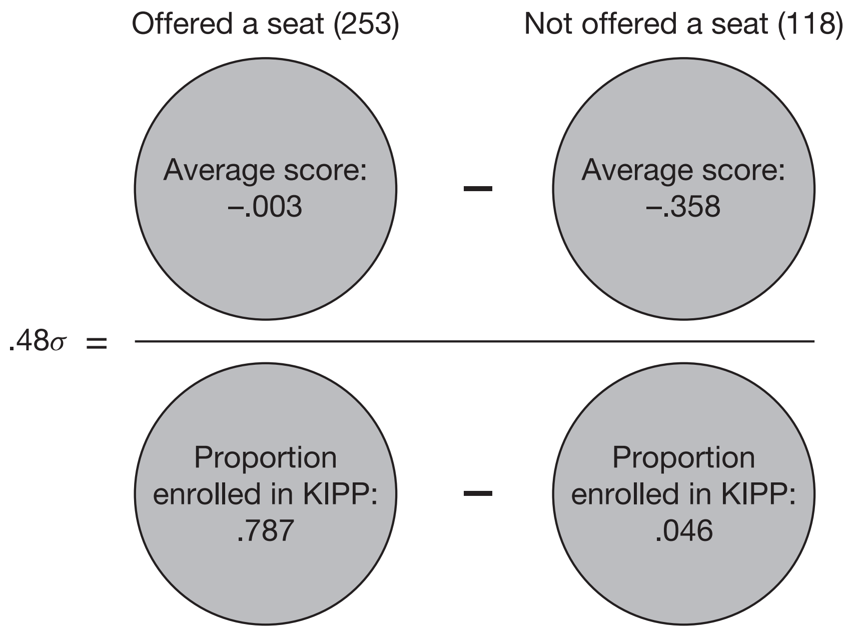
		KIPP applicants			
	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
		Panel A. Baseline characteristics			
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
		Panel B. Outcomes			
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows differences between KIPP students and applicants who did not attend KIPP. Standard errors are reported in parentheses.

Superman Arrives

This graphic fills in the numbers for KIPP lottery IV

FIGURE 3.2
IV in school: the effect of KIPP attendance on math scores



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

3 IV is always LATE

3.1 The Four Types of Children

- KIPP lottery offers affect KIPP enrollment for many applicants . . . but not all
  - Some offered a seat at KIPP nevertheless go elsewhere
    (roughly 3/4 of those offered a seat enroll)
  - A few not offered a seat in the lottery manage to find one anyway
    (about 5% of lottery losers get in anyway)
- How should we interpret IV estimates in light of this fact?

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 (*Normando*)	Defiers
	Attends KIPP $D_i = 1$	Compliers (*Camila*)	Always-takers (*Alvaro*)

Note: KIPP = Knowledge Is Power Program.

(*Actually there are are only three types: no defiers allowed!*)

- In a world of heterogeneous potential outcomes,

$$\lambda = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|C_i = 1],$$

where C_i indicates *compliers*, like Camila

- Parameter $E[Y_{1i} - Y_{0i}|C_i = 1]$ is called a *local average treatment effect* (LATE); Compliers are a subset of the treated.
- In general, LATE differs from the effect of treatment on the treated, $E[Y_{1i} - Y_{0i}|D_i = 1]$, because some treated applicants, like Alvaro, are *always-takers*
  - As detailed in MHE, the proportion of always-takers is given by $E[D_i|Z_i = 0]$
  - With few always-takers (as in the KIPP lottery), we expect:

$$E[Y_{1i} - Y_{0i}|C_i = 1] \approx E[Y_{1i} - Y_{0i}|D_i = 1]$$

3.2 LATE Again: Effects of Vietnam-Era Military Service (Angrist 1990)

- From 1970-73, Uncle Sam selected soldiers in a *draft lottery*: Men born 1950-53 were called up by random sequence numbers (RSN), assigned to their DOB
- Men born in 1950 with RSN<195 were draft-eligible; code instrument Z_i from this
- This MHE table summarizes the draft lottery IV story, where Vietnam veteran status is the variable instrumented:

Table 4.1.3

IV Estimates of the Effects of Military Service on the Earnings of White Men born in 1950

Earnings year	Earnings: Mean	Earnings: Eligibility Effect	Veteran Status: Inelig. Mean	Veteran Status: Eligibility Effect	Wald Estimate of Veteran Effect
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8 (210.5)	.182	.159 (.040)	-2,741 (1,324)
1971	3,338	-325.9 (46.6)			-2050 (293)
1969	2,299	-2.0 (34.5)			

Note: Adapted from Table 5 in Angrist and Krueger (1999) and author tabulations. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Vete status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.

- What's the LATE interpretation here? What do the key IV assumptions require?
  - Independence vs. exclusion [MHE 4.4.1]: define $Y_i(z, d)$ as potential outcomes indexed against *both* instrument and treatment
  - Exclusion
$$Y_i(0, d) = Y_i(1, d);\ d = 0, 1$$
  - Independence
$$E[Y_i(z, d)|Z_i] = E[Y_i(z, d)]$$
- A Vietnam-era update:

Angrist, Chen, and Song (2011)

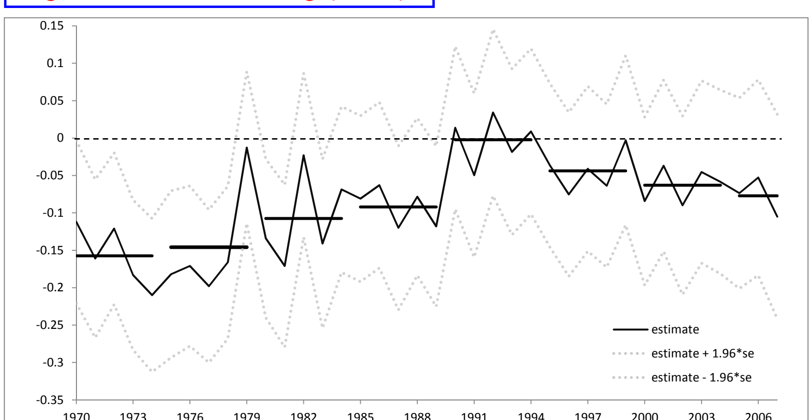


Figure 1. Draft-lottery Estimates of Vietnam-era Service Effects on ln(Earnings) for White Men Born 1950-52

- IV is everywhere! Reconsider, for example, the Carter, Greenberg and Walker (2017) class computer RCT and the OHP Medicaid effects in Taubman, et al. (2014). In fact, the OHP is done as IV in both Finkelstein, et al. (2012) and Taubman, et al. (2014). Ponder the first stage in these imperfectly-randomized randomized trials.
- Here's MM OHP Table 1.5 again:

TABLE 1.5
OHP effects on insurance coverage and health-care use

	Oregon		Portland area	
Outcome	Control mean (1)	Treatment effect (2)	Control mean (3)	Treatment effect (4)
A. Administrative data				
Ever on Medicaid	.141	.256 (.004)	.151	.247 (.006)
Any hospital admissions	.067	.005 (.002)		
Any emergency department visit			.345	.017 (.006)
Number of emergency department visits			1.02	.101 (.029)
Sample size	74,922		24,646	
B. Survey data				
Outpatient visits (in the past 6 months)	1.91	.314 (.054)		
Any prescriptions?	.637	.025 (.008)		
Sample size	23,741			

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on insurance coverage and use of health care. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

4 Two-Stage Least Squares

In practice, we do IV by doing two-stage least squares (2SLS). This allows us to add covariates (controls) and to use multiple instruments to generate more precise IV estimates.

4.1 2SLS Derived

- 2SLS starts with the first-stage equation:

$$D_i = \alpha_1 + \phi Z_i + e_{1i}.$$

Save the first-stage fitted values this generates:

$$\hat{D}_i = \alpha_1 + \phi Z_i,$$

then regress Y_i on fits:

$$Y_i = \alpha_2 + \lambda_{2SLS}\hat{D}_i + e_{2i}. \tag{14}$$

It's easy to show that λ_{2SLS} in (14) equals the IV ratio (8) in both population and sample:

$$\lambda_{2SLS} = \frac{C(Y_i, \hat{D}_i)}{V(\hat{D}_i)} = \frac{\phi C(Y_i, Z_i)}{\phi^2 V(Z_i)} = ...$$

Covs in the mix

- Suppose the causal model of interest includes covariates, X_i:

  $$Y_i = \alpha_2' X_i + \lambda_{2SLS} D_i + \eta_i \tag{15}$$

  In the Superman story, for example, X_i includes dummies for application year (KIPP offers are randomized conditional on this)

- Write the first stage with covariates as the sum of first-stage fitted values plus first-stage residuals:

  $$D_i = X_i^{'}\alpha_1 + \phi Z_i + e_{1i} = \hat{D}_i + e_{1i} \tag{16}$$

  The 2SLS second stage is OLS on:

  $$Y_i = \alpha_2' X_i + \lambda_{2SLS}\hat{D}_i + e_{2i} \tag{17}$$

- Why does this work? The key is that the second-stage residual is:

  $$e_{2i} = \lambda_{2SLS} e_{1i} + \eta_i,$$

  with $E[\hat{D}_i e_{1i}] = 0$ by construction and $E[\hat{D}_i \eta_i] = 0$ by assumption (6), so $E[\hat{D}_i e_{2i}] = 0$ too.

  - Unlike in LN8, we're not bothering here to distinguish between estimated and population first-fitted values. The fact that in practice we must estimate first-stage fitted values turns out not to matter for 2SLS to work in large samples

- The reduced form regression for the model with covariates is:

  $$Y_i = X_i^{'}\alpha_0 + \rho Z_i + e_{0i} \tag{18}$$

  Equation (18) is obtained by substituting (16) into (15)

  - The reduced form residual, e_{0i}, is the same as the second-stage residual, e_{2i}
  - λ_{2SLS} is still the ratio of *reduced form* to *first stage* coefficients on the instrument:

    $$\lambda_{2SLS} = \frac{\rho}{\phi}$$

    (show this)

Mightier with more instruments

- Blessed with more than one instrument?
  - In the Superman story, we might code one dummy for lottery offers made immediately (on lottery night) and one for offers made later (to applicants on a waiting list)
- Call these Z_{1i} and Z_{2i} and add 'em both to the first stage when baking the fits:

$$D_i = X_i^{'}\alpha_1 + \phi_1 Z_{1i} + \phi_2 Z_{2i} + e_{1i}$$

  The second stage, equation (17), stays the same
- Models with more instruments than necessary are said to be *over-identified*; the resulting 2SLS estimates are more precise than *just-identified* estimates using one instrument alone

5 Where Do Babies Come From?

5.1 Family Size Effects on Female Labor Supply (Angrist and Evans, 1998)

- Lotteries are awesome! Other instruments come from deep institutional knowledge or an understanding of the mechanisms and forces that determine treatment.
- Consider effects of childbearing on female labor supply and earnings. From 1960-80, married women's LFP approached that of men. At the same time, marital fertility fell from 2 to 3. (My family illustrates!)
  - Did the sharp decline in family size cause women to work more? Or, is the negative correlation between childbearing and labor supply an artifact of selection bias?
- AE98 uses two instruments to identify effects of having more than two kids on parents' work and earnings
  - A *twins instrument*, Z_{1i} indicates multiple second births (buy one, get one free)
  - A *samesex instrument*, Z_{2i} indicates mothers of two boys and two girls at parities 1 and 2 (diversify your sibling-sex portfolio)

(Stata for AE98 follows)

5.2 The Quantity-Quality Trade-Off (Angrist, Lavy, and Schlosser, 2010)

- In the 1970s and 1980s, governments around the world discouraged childbearing in the belief that large families decrease living standards
  - China's One Child Policy is the most (in)famous of these anti-natalist policies
  - Economists refer to the relationship between family size and living standards as a *quantity-quality tradeoff*
- Are larger families really impoverished by their size? If only we could randomize the number of children and find out!
  - Angrist, Lavy, and Schlosser (2010) exploit AE98-style natural experiments for family size in samples of women with 2 or more children
    - *twins instruments*, Z_{1i} now indicate multiples at various births (e.g., twins@2, twins@3)
    - *samesex instruments*, Z_{2i} now indicate mothers of various samesex sibships (e.g., 3 girls@3)
  - Z_{1i} and Z_{2i} are both highly predictive of the number of children born in family i
  - And are arguably independent of the *potential* human capital of the *first-borns* in these families (the sample used to construct the tables below consists of first-born non-twin Israeli Jews aged 18-60 in the Census, whose mothers were born after 1930 and had their first birth between the ages of 15-45; instruments here use info on first and second births only)

TABLE 3.4
Quantity-quality first stages

	Twins instruments		Same-sex instruments		Twins and same-sex instruments
	(1)	(2)	(3)	(4)	(5)
Second-born twins	.320	.437			.449
	(.052)	(.050)			(.050)
Same-sex sibships			.079	.073	.076
			(.012)	(.010)	(.010)
Male		−.018		−.020	−.020
		(.010)		(.010)	(.010)
Controls	No	Yes	No	Yes	Yes

Notes: This table reports coefficients from a regression of the number of children on instruments and covariates. The sample size is 89,445. Standard errors are reported in parentheses.

Table 3.5
OLS and 2SLS estimates of the quantity-quality trade-off

Dependent variable	OLS estimates (1)	2SLS estimates: Twins instruments (2)	Same-sex instruments (3)	Twins and same-sex instruments (4)
Years of schooling	−.145	.174	.318	.237
	(.005)	(.166)	(.210)	(.128)
High school graduate	−.029	.030	.001	.017
	(.001)	(.028)	(.033)	(.021)
Some college (for age $\geq$ 24)	−.023	.017	.078	.048
	(.001)	(.052)	(.054)	(.037)
College graduate (for age $\geq$ 24)	−.015	−.021	.125	.052
	(.001)	(.045)	(.053)	(.032)

Notes: This table reports OLS and 2SLS estimates of the effect of family size on schooling. OLS estimates appear in column (1). Columns (2), (3), and (4) show 2SLS estimates constructed using the instruments indicated in column headings. Sample sizes are 89,445 for rows (1) and (2); 50,561 for row (3); and 50,535 for row (4). Standard errors are reported in parentheses.

6 Sampling Variance of 2SLS Estimates

- Here's equation (17) without covariates and with the second-stage residual written out:

$$Y_i = \alpha + \lambda_{2SLS}\hat{D}_i + [\eta_i + \lambda_{2SLS}(D_i - \hat{D}_i)], \tag{19}$$

2SLS is OLS on this second-stage equation:

$$\hat{\lambda}_{2SLS} = \frac{\sum Y_i(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2},$$

Substituting for Y_i, we have:

$$\begin{aligned} \hat{\lambda}_{2SLS} &= \alpha\frac{\sum(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2} + \lambda_{2SLS}\left[\frac{\sum \hat{D}_i(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2}\right] + \frac{\sum \eta_i(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2} + \lambda_{2SLS}\frac{\sum(D_i - \hat{D}_i)(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2} \\ &= \lambda_{2SLS} + \frac{\sum \eta_i(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2} \end{aligned} \tag{20}$$

(why do the first and last terms in the first line equal 0? why does the term in square brackets in the first line equal 1?)

- Assuming η_i is homoscedastic with variance σ^2_η, the asymptotic standard error of $\hat{\lambda}_{2SLS}$ is:

$$SE(\hat{\lambda}_{2SLS}) = \frac{1}{\sqrt{n}} \frac{\sigma_\eta}{\sigma_{\hat{D}}}$$

  where σ_η is the standard deviation of η_i and $\sigma_{\hat{D}}$ is the standard deviation of first-stage fits, $\hat{D}_i$

Notes

- The standard errors generated by OLS estimation of (19) (i.e., "manual 2SLS") are wrong (why?)
  - Stata `ivregress` gets 'em right
- $SE(\hat{\lambda}_{2SLS})$ is an asymptotic formula, derived under something like classical assumptions, but even given these assumptions, valid only in large samples
  - Robust, clustered, and Newey-West standard errors for 2SLS are known to Stata (though also valid only in large samples)
  - For more on 2SLS inference, see the MM Chapter 3 appendix and MHE Chapter 4
- We can claim only that $\hat{\lambda}_{2SLS}$ is consistent and asymptotically Normally distributed; as a rule 2SLS estimates are biased
  - The bias of 2SLS is proportional to the number of instruments in an over-identified model and inversely proportional to the F statistic that tests instrument relevance in the first stage
    * With many weak instruments, 2SLS estimates are likely to be misleadingly close to the corresponding OLS estimates
    * Given a reasonably strong first stage, *just-identified* 2SLS estimates (one instrument for one endogenous regressor) are approximately unbiased