Instrumental Variables (IV) and 2SLS for OVB

1 Recap: Regression and the CIA

- Recall the causal regression model for effects of private university attendance (P_i) on wages from previous notes:
  - $Y_{0i} = \alpha + \eta_i$, 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 produces a difference in means:
  $$E[Y_i|P_i = 1] - E[Y_i|P_i = 0] = \delta + \{E[\eta_i|P_i = 1] - E[\eta_i|P_i = 0]\}$$
  - 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 controls, X_i. Equivalently,
  $$\eta_i = \gamma' X_i + u_i$$
  where u_i and X_i are uncorrelated
  - In MM Chpt 2, the variables in X_i are dummies for Barrons selectivity groups or the average selectivity of colleges applied to
- 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
- *Often, however, we're not so lucky in the controls department. Even so, masters know that ...*

2 Instrumental Variables Eliminate Selection Bias

2.1 Waiting for Superman

- Many children in large urban districts leave school with poor reading and math skills. This perpetuates poverty and increases inequality.
- 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 and to opt out of collective bargaining agreements that cover traditional public school teachers

- The Knowledge is Power Program (KIPP) is iconic in the charter universe, serving mostly urban minority students. KIPP's "No Excuses" charter recipe includes a long school day and year, data-driven instruction, TFA and tutoring, and emphasizes discipline and comportment
- KIPP students tend to do better than other students in the district they hail from. But is this selection bias or a causal effect?

  Here's what the critics say:

  > 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 Superman selection 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 is yet another Greek name for the causal effect of interest. Again, we're 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

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

$$C(Z_i, D_i) \neq 0 \tag{5}$$

$$C(Z_i, \eta_i) = 0 \tag{6}$$

- These conditions imply

$$C(Z_i, Y_i) = C(Z_i, D_i)\lambda_{IV} \tag{7}$$

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

  Because we can solve for the causal coefficient of interest from observable moments (in this case, variances and covariances), λ is said to be *identified*

  - *Identification* problems are distinct from *estimation* problems
  - We put the subscript "IV" on λ_{IV} because, as we'll soon see, IV estimators identify a particular type of causal effect

- The *IV estimator* is a ratio of regression estimates:

$$\hat{\lambda}_{IV} = \frac{s_{ZY}/s_Z^2}{s_{ZD}/s_Z^2} \tag{9}$$

  where s_{ZY} etc. are sample covariances and variances

- Given assumptions (5) and (6), $\hat{\lambda}_{IV}$ is a consistent (though not unbiased) estimator of λ, with an asymptotically Normal sampling distribution that we derive later

- The top and bottom of the IV ratio, (8), are central to the IV story, so we christen them:

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

Sample analogs, denoted $\hat{\rho}$ and $\hat{\phi}$, are called "reduced form estimates" and "first stage estimates"

  - Nice work if you can get it!
  - Do such miraculous instruments exist?

Pedestrian IV: Long Regression w/o Controls

- Another way to motivate IV: when estimating KIPP effects, we'd like to control for factors like ability and family background

  - Denote these by A_i. Suppose this is the "long regression" we'd like to run:

$$Y_i = \alpha_l + \rho_l D_i + \gamma' A_i + e_i \tag{10}$$

- 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 even when long-regression controls are unavailable. In addition to the first stage requirement, condition (5), this formulation requires that Z_i be uncorrelated with omitted variables and the residual thats left over. That is, we replace (6) with $C(Z_i, A_i) = C(Z_i, e_i) = 0$ in (10).

2.3 Playing the KIPP Lottery

- Like all Massachusetts charter schools, KIPP Lynn assigns seats by lottery (i.e., at random) 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):

$$\lambda_{IV} = \frac{C(Z_i,\, Y_i)/V(Z_i)}{C(Z_i,\, D_i)/V(Z_i)} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \tag{11}$$

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

$$\hat{\lambda}_{IV} = (\bar{Y}_1 - \bar{Y}_0)/(\bar{D}_1 - \bar{D}_0), \tag{12}$$

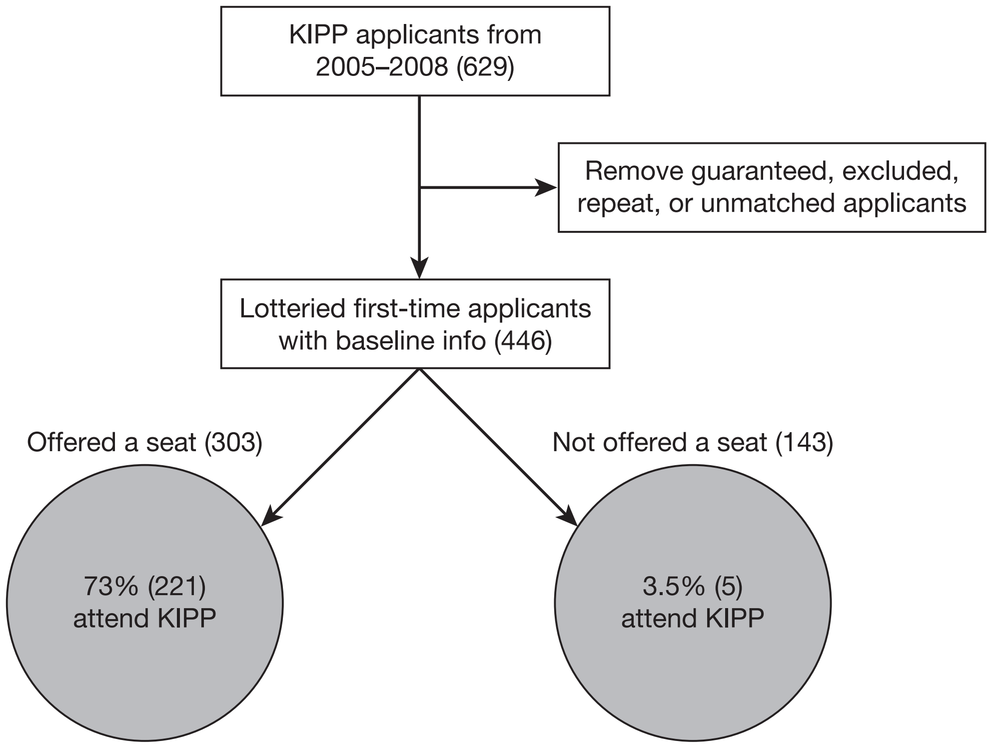
    where $\bar{Y}_j$ and $\bar{S}_j$ are sample means of Y_i and S_i conditional on $Z_i = j$

  - The formulation in (12) is called a *Wald estimator* after Wald (1940)

The KIPP 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.

- Lotteries at KIPP Lynn make *ceteris* into *paribus*

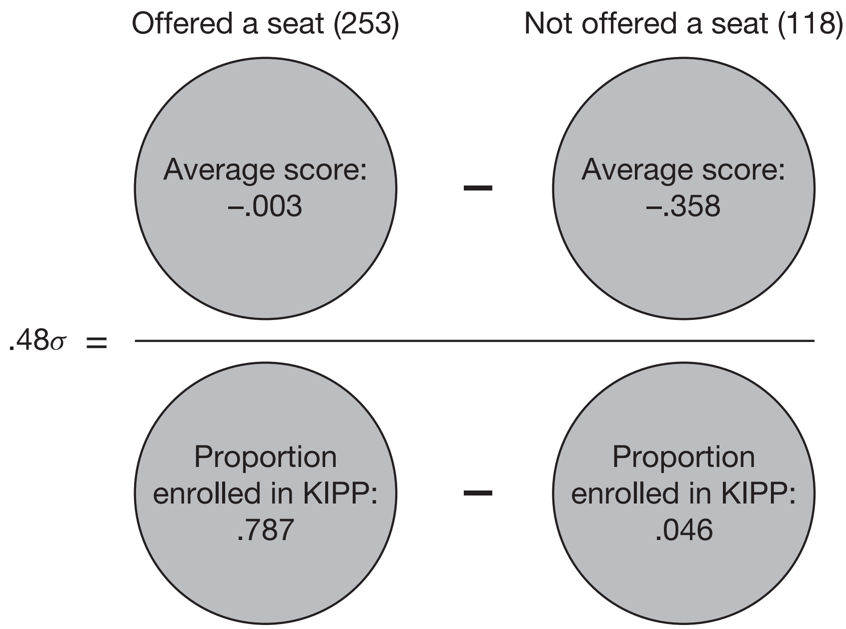
- The table below describes KIPP's 2005-8 applicants for 5th and 6th grade seats; outcomes are from the end of these grades (for the 371 tested lottery applicants; baseline scores are from 4th grade)

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.

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 With Heterogeneous Potential Outcomes

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
  - A few not offered a seat in the lottery manage to sneak 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: no defiers allowed!*)

- In a world of heterogeneous potential outcomes,

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

  where C_i indicates compliers, like Camila

- Parameter $E[Y_{1i} - Y_{0i}|C_i = 1]$ is called a *local average treatment effect* (LATE)
- In general, LATE differs from the effect of treatment on the treated, $E[Y_{1i} - Y_{0i}|D_i = 1]$, because some treated are *always-takers*, like Alvaro
  - 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
- This table is from MHE:

Table 4.1.3

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

	Earnings		Veteran Status		Wald Estimate of Veteran Effect
Earnings year	Mean	Eligibility Effect	Inelig. Mean	Eligibility 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?
- A Vietnam update:

Angrist, Chen, and Song (2011)

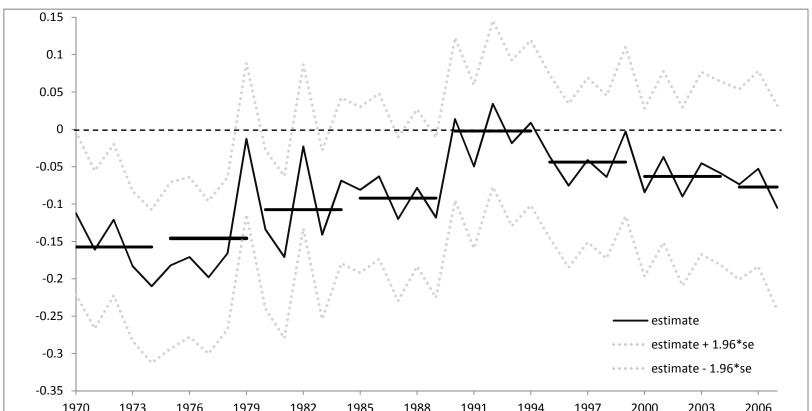


Figure 1. Draft-lottery Estimates of Vietnam-era Service Effects on ln(Earnings) for White Men Born 1950-52

- 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)

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 a more efficient (precise) IV estimate.

4.1 2SLS Derived

- Here's a nifty way to compute IV estimates: First, regress D_i on Z_i

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

  and save the first-stage fitted values:

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

  Then regress Y_i on these

$$Y_i = \alpha_2 + \lambda_{2SLS}\hat{D}_i + e_{2i}, \tag{13}$$

  It's easy to show (be sure you can) that λ_{2SLS} in (13) equals λ_{IV} in (8) in both population and sample

Covs in the mix

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

$$Y_i = \alpha_2' X_i + \lambda_{2SLS} D_i + \eta_i \tag{14}$$

  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:

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

  2SLS in this case is OLS on the second-stage equation with covariates:

$$Y_i = \alpha_2' X_i + \lambda_{2SLS}\hat{D}_i + e_{2i} \tag{15}$$

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

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

  and both pieces on the RHS are orthogonal to $\hat{D}_i$, that is, $E[\hat{D}_i e_{2i}] = 0$.

  – The first stage and reduced form regressions for this model are,

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

$$Y_i = X_i^{'}\alpha_0 + \rho Z_i + e_{0i} \tag{17}$$

    Equation (17) is obtained by substituting (16) into (14).

  – λ_{2SLS} is still the ratio of *RF* to *1st Stage* coefficients:

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

    (show this)

  – In practice, we plug the *estimated* first stage fits into the second stage. These are:

$$D_i = X_i^{'}\hat{\alpha}_1 + \hat{\phi} Z_i + \hat{e}_{1i} = \hat{D}_i^* + \hat{e}_{1i}$$

    We rely, therefore, on the fact that *plim* $\frac{1}{N}\sum \hat{D}_i^* \hat{e}_{2i} = 0$. This allows us to say that $\hat{\lambda}_{2SLS}$ is a *consistent* estimator of the causal effect of D_i on Y_i but it is not unbiased. Similarly, we claim only consistency for the sample analog of $\frac{\hat{\rho}}{\hat{\phi}}$.

Mightier with more instruments

- Blessed with more than one instrument?
  - In the Superman story, we might use dummies for lottery offers made immediately (on lottery night) and later (to applicants on a waiting list)
- Add 'em 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 (15), stays the same

- Models with more instruments than necessary are said to be *over-identified*

5 So Where *Do* Babies Come From?

- Lotteries are awesome! Other instruments come from deep institutional knowledge, revealing, for example, the effect of children on their parents' labor supply (Angrist and Evans, 1998)

5.1 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 small families increase living standards

```
1 .
2 . *All women sample:
3 . keep if ((agem1>=21 & agem1<=35) & (kidcount>=2) & (ageq2nd1>4) & (agefstm>=15) & (asex==0) & (aage==0) & (aqtrbrth==0) & (asex2nd==0) & (aage2nd==0))
  (532,427 observations deleted)

4 .   /*& (agefstd>=15 | agefstd==.)*/
5 .
6 . *keep if (msample==1)
7 .
8 . sum agem1 kidcount ageq2nd1 agefstm weeksm1 workedm morekids agem1 boy1st boy2nd blackm hispm othracem multi2nd samesex msample
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agem1	394,840	30.1248	3.509685	21	35
kidcount	394,840	2.552069	.8083876	2	12
ageq2nd1	394,840	26.36489	14.61527	5	70
agefstm	394,840	20.13956	2.949069	15	33
weeksm1	394,840	20.83419	22.28601	0	52
workedm	394,840	.5654873	.4956935	0	1
morekids	394,840	.4020641	.4903154	0	1
agem1	394,840	30.1248	3.509685	21	35
boy1st	394,840	.511088	.4998777	0	1
boy2nd	394,840	.5109614	.4998805	0	1
blackm	394,840	.1189343	.3237115	0	1
hispm	394,840	.03004	.1706976	0	1
othracem	394,840	.028685	.16692	0	1
multi2nd	394,840	.0085604	.0921258	0	1
samesex	394,840	.5053895	.4999716	0	1
msample	394,840	.6449499	.4785291	0	1

```
9 .
10 . *OLS:
11 . reg weeksm1 morekids agem1 agefstm boy1st boy2nd blackm hispm othracem, r
```

Linear regression

Number of obs	=	394,840
F(8, 394831)	=	4589.07
Prob > F	=	0.0000
R-squared	=	0.0778
Root MSE	=	21.402

weeksm1	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
morekids	-8.978191	.0705666	-127.23	0.000	-9.1165	-8.839883
agem1	1.466036	.0105266	139.27	0.000	1.445404	1.486668
agefstm	-1.423913	.0131709	-108.11	0.000	-1.449728	-1.398099
boy1st	-.1153498	.0681462	-1.69	0.091	-.2489143	.0182147
boy2nd	-.1773649	.0681483	-2.60	0.009	-.3109335	-.0437963
blackm	6.451669	.1103587	58.46	0.000	6.235369	6.667968
hispm	-.7810209	.1956389	-3.99	0.000	-1.164467	-.3975744
othracem	2.860371	.2109436	13.56	0.000	2.446928	3.273814
_cons	8.280615	.3199806	25.88	0.000	7.653463	8.907767

```
12 . reg workedm morekids agem1 agefstm boy1st boy2nd blackm hispm othracem, r
```

Linear regression

Number of obs	=	394,840
F(8, 394831)	=	3032.83
Prob > F	=	0.0000
R-squared	=	0.0537
Root MSE	=	.48222

workedm	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
morekids	-.1764489	.0016171	-109.11	0.000	-.1796184	-.1732793
agem1	.0241995	.0002424	99.84	0.000	.0237244	.0246745
agefstm	-.0291002	.0002967	-98.07	0.000	-.0296818	-.0285187
boy1st	-.0005312	.0015353	-0.35	0.729	-.0035404	.002478
boy2nd	-.0040863	.0015353	-2.66	0.008	-.0070955	-.0010771
blackm	.1060263	.0023474	45.17	0.000	.1014255	.110627
hispm	-.0309759	.0046057	-6.73	0.000	-.0400029	-.0219488
othracem	.0420805	.0046453	9.06	0.000	.0329759	.0511852
_cons	.4829654	.0075603	63.88	0.000	.4681474	.4977834

```
13 .
14 . *first stage and weeks reduced form: twins
15 . reg morekids multi2nd, r
```

```
Linear regression                               Number of obs     =    394,840
                                                F(0, 394838)      =          .
                                                Prob > F          =          .
                                                R-squared         =     0.0128
                                                Root MSE          =     .48716

------------------------------------------------------------------------------
             |               Robust
    morekids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    multi2nd |   .6030987    .000782   771.25   0.000      .601566    .6046313
       _cons |   .3969013    .000782   507.56   0.000     .3953687     .398434
------------------------------------------------------------------------------

16 . reg weeksm1 multi2nd, r

Linear regression                               Number of obs     =    394,840
                                                F(1, 394838)      =      27.19
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0001
                                                Root MSE          =     22.285

------------------------------------------------------------------------------
             |               Robust
     weeksm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    multi2nd |  -1.975956   .3789719    -5.21   0.000    -2.718729   -1.233182
       _cons |    20.8511   .0356232   585.32   0.000     20.78128    20.92092
------------------------------------------------------------------------------

17 . *Wald for twins
18 . ivregress 2sls weeksm1 (morekids = multi2nd)

Instrumental variables (2SLS) regression          Number of obs   =    394,840
                                                  Wald chi2(1)    =      26.71
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0138
                                                  Root MSE        =     22.132

------------------------------------------------------------------------------
     weeksm1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    morekids |  -3.276339   .6339241    -5.17   0.000    -4.518807   -2.033871
       _cons |   22.15149   .2573002    86.09   0.000     21.64719    22.65579
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   multi2nd

19 .
20 . *first stage and weeks reduced form: samesex
21 . reg morekids samesex, r

Linear regression                               Number of obs     =    394,840
                                                F(1, 394838)      =    1461.73
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0037
                                                Root MSE          =     .48941

------------------------------------------------------------------------------
             |               Robust
    morekids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |    .059544   .0015574    38.23   0.000     .0564915    .0625965
       _cons |   .3719712   .0010937   340.10   0.000     .3698276    .3741148
------------------------------------------------------------------------------

22 . reg weeksm1 samesex, r

Linear regression                               Number of obs     =    394,840
                                                F(1, 394838)      =      28.50
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0001
                                                Root MSE          =     22.285

------------------------------------------------------------------------------
             |               Robust
     weeksm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |  -.3786749   .0709378    -5.34   0.000    -.5177109   -.2396389
       _cons |   21.02557   .0505151   416.22   0.000     20.92656    21.12457
------------------------------------------------------------------------------

23 . *Wald for samesex
```

```
24 . ivregress 2sls weeksm1 (morekids = samesex)

   Instrumental variables (2SLS) regression         Number of obs   =    394,840
                                                    Wald chi2(1)    =      29.00
                                                    Prob > chi2     =     0.0000
                                                    R-squared       =     0.0173
                                                    Root MSE        =     22.093

   ------------------------------------------------------------------------------
        weeksm1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
       morekids |  -6.359578   1.181014    -5.38   0.000    -8.674324   -4.044833
          _cons |   23.39115   .4761434    49.13   0.000     22.45792    24.32437
   ------------------------------------------------------------------------------
   Instrumented:  morekids
   Instruments:   samesex

25 .
26 . *check for balance
27 . reg agefstm multi2nd agem1 boy1st boy2nd blackm hispm othracem, r

   Linear regression                              Number of obs     =    394,840
                                                  F(7, 394832)      =   16568.12
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.1941
                                                  Root MSE          =     2.6475

   ------------------------------------------------------------------------------
                |               Robust
        agefstm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
       multi2nd |   .1752039   .0457188     3.83   0.000     .0855964    .2648115
          agem1 |   .3290559   .0010836   303.66   0.000      .326932    .3311798
         boy1st |   .0094222   .0084286     1.12   0.264    -.0070976     .025942
         boy2nd |   .0193927   .0084301     2.30   0.021       .00287    .0359155
         blackm |  -1.427554   .0120888  -118.09   0.000    -1.451248    -1.40386
          hispm |  -.5792559   .0234584   -24.69   0.000    -.6252337   -.5332781
       othracem |   .6226536   .0280525    22.20   0.000     .5676715    .6776357
          _cons |   10.37992   .0320333   324.04   0.000     10.31713     10.4427
   ------------------------------------------------------------------------------

28 . reg educm multi2nd agem1 agefstm boy1st boy2nd blackm hispm othracem, r

   Linear regression                              Number of obs     =    394,840
                                                  F(8, 394831)      =    9264.86
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.2109
                                                  Root MSE          =     2.1325

   ------------------------------------------------------------------------------
                |               Robust
          educm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
       multi2nd |  -.0074761   .0373593    -0.20   0.841    -.0806992    .0657471
          agem1 |   .0221166   .0010089    21.92   0.000     .0201392     .024094
        agefstm |   .3330336   .0014574   228.51   0.000     .3301771    .3358902
         boy1st |   .0036823   .0067884     0.54   0.588    -.0096229    .0169874
         boy2nd |   .0075476   .0067899     1.11   0.266    -.0057603    .0208555
         blackm |   .2191673   .0101611    21.57   0.000     .1992517    .2390828
          hispm |  -2.374502   .0326049   -72.83   0.000    -2.438406   -2.310597
       othracem |   -.531052   .0321654   -16.51   0.000    -.5940952   -.4680088
          _cons |    4.80712   .0341787   140.65   0.000     4.740131    4.874109
   ------------------------------------------------------------------------------

29 . reg agefstm samesex agem1 boy1st boy2nd blackm hispm othracem, r

   Linear regression                              Number of obs     =    394,840
                                                  F(7, 394832)      =   16567.41
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.1941
                                                  Root MSE          =     2.6475

   ------------------------------------------------------------------------------
                |               Robust
        agefstm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
        samesex |   .0217124   .0084313     2.58   0.010     .0051874    .0382375
          agem1 |   .3290979   .0010836   303.71   0.000     .3269741    .3312217
         boy1st |   .0089868   .0084313     1.07   0.286    -.0075383     .025512
         boy2nd |   .0188476   .0084318     2.24   0.025     .0023216    .0353737
         blackm |  -1.427105    .012089  -118.05   0.000    -1.450799   -1.403411
          hispm |  -.5793764   .0234604   -24.70   0.000     -.625358   -.5333947
       othracem |   .6225523   .0280511    22.19   0.000      .567573    .6775316
          _cons |   10.36963   .0323018   321.02   0.000     10.30632    10.43294
```

```
30 . reg educm samesex ageml agefstm boy1st boy2nd blackm hispm othracem, r

   Linear regression                               Number of obs     =    394,840
                                                   F(8, 394831)      =    9264.95
                                                   Prob > F          =     0.0000
                                                   R-squared         =     0.2109
                                                   Root MSE          =     2.1325

   ------------------------------------------------------------------------------
                |               Robust
          educm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
        samesex |  -.0087736   .0067908    -1.29   0.196    -.0220833    .0045362
          ageml |   .0221105   .0010089    21.92   0.000     .0201332    .0240878
        agefstm |   .3330388   .0014574   228.51   0.000     .3301823    .3358953
         boy1st |   .0038694   .0067906     0.57   0.569    -.0094399    .0171787
         boy2nd |   .0077413   .0067907     1.14   0.254    -.0055683    .0210509
         blackm |   .2191597   .0101598    21.57   0.000     .1992469    .2390726
          hispm |  -2.374498   .0326047   -72.83   0.000    -2.438402   -2.310594
       othracem |  -.5310898   .0321659   -16.51   0.000     -.594134   -.4680457
          _cons |   4.811376   .0343194   140.19   0.000     4.744112    4.878641
   ------------------------------------------------------------------------------

31 .
32 . *2sls: weeks (twins, w/covs)
33 . ivregress 2sls weeksm1 (morekids = multi2nd) ageml agefstm boy1st boy2nd blackm hispm othracem, r

   Instrumental variables (2SLS) regression          Number of obs   =    394,840
                                                     Wald chi2(8)    =   18168.92
                                                     Prob > chi2     =     0.0000
                                                     R-squared       =     0.0654
                                                     Root MSE        =     21.545

   ------------------------------------------------------------------------------
                |               Robust
        weeksm1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
       morekids |  -3.712292   .6036268    -6.15   0.000    -4.895379   -2.529205
          ageml |   1.307164   .0209401    62.42   0.000     1.266122    1.348205
        agefstm |  -1.186511   .0300822   -39.44   0.000    -1.245471   -1.127551
         boy1st |  -.0804947   .0687157    -1.17   0.241    -.2151751    .0541857
         boy2nd |  -.1385996   .0687455    -2.02   0.044    -.2733383   -.0038609
         blackm |   6.075055   .1192318    50.95   0.000     5.841365    6.308745
          hispm |  -1.603621   .2187741    -7.33   0.000    -2.032411   -1.174832
       othracem |   2.482386    .216284    11.48   0.000     2.058477    2.906295
          _cons |   6.210914    .398797    15.57   0.000     5.429286    6.992542
   ------------------------------------------------------------------------------
   Instrumented:  morekids
   Instruments:   ageml agefstm boy1st boy2nd blackm hispm othracem multi2nd

34 .
35 . *2sls: weeks (samesex, w/covs)
36 . ivregress 2sls weeksm1 (morekids = samesex) ageml agefstm boy1st boy2nd blackm hispm othracem, r

   Instrumental variables (2SLS) regression          Number of obs   =    394,840
                                                     Wald chi2(8)    =   18252.28
                                                     Prob > chi2     =     0.0000
                                                     R-squared       =     0.0726
                                                     Root MSE        =     21.462

   ------------------------------------------------------------------------------
                |               Robust
        weeksm1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
       morekids |   -5.55877   1.117829    -4.97   0.000    -7.749673   -3.367866
          ageml |   1.362872   .0352894    38.62   0.000     1.293706    1.432038
        agefstm |  -1.269755   .0520245   -24.41   0.000    -1.371722   -1.167789
         boy1st |  -.0927166   .0687273    -1.35   0.177    -.2274196    .0419864
         boy2nd |  -.1521926   .0688292    -2.21   0.027    -.2870953   -.0172899
         blackm |   6.207114   .1364545    45.49   0.000     5.939668     6.47456
          hispm |  -1.315178   .2625227    -5.01   0.000    -1.829713   -.8006428
       othracem |   2.614926   .2260306    11.57   0.000     2.171914    3.057938
          _cons |   6.936651   .5431087    12.77   0.000     5.872177    8.001124
   ------------------------------------------------------------------------------
   Instrumented:  morekids
   Instruments:   ageml agefstm boy1st boy2nd blackm hispm othracem samesex

37 .
38 . *2sls: weeks (overid, w/covs)
39 . ivregress 2sls weeksm1 (morekids = multi2nd samesex) ageml agefstm boy1st boy2nd blackm hispm othracem, r

   Instrumental variables (2SLS) regression          Number of obs   =    394,840
```

```
                                                  Wald chi2(8)     =   18224.63
                                                  Prob > chi2      =     0.0000
                                                  R-squared        =     0.0674
                                                  Root MSE         =     21.522

------------------------------------------------------------------------------
             |               Robust
    weeksm1  |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    morekids |  -4.141475   .5311981    -7.80   0.000    -5.182605   -3.100346
       agem1 |   1.320112   .0190752    69.21   0.000     1.282725    1.357499
     agefstm |   -1.20586   .0271733   -44.38   0.000    -1.259118   -1.152601
      boy1st |  -.0833355   .0686168    -1.21   0.225    -.2178218    .0511509
      boy2nd |  -.1417591   .0686413    -2.07   0.039    -.2762936   -.0072245
      blackm |    6.10575   .1173185    52.04   0.000      5.87581     6.33569
       hispm |  -1.536577   .2138234    -7.19   0.000    -1.955663   -1.117491
    othracem |   2.513193   .2151399    11.68   0.000     2.091526    2.934859
       _cons |   6.379599   .3822657    16.69   0.000     5.630372    7.128826
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   agem1 agefstm boy1st boy2nd blackm hispm othracem multi2nd samesex

40 .
41 . *manual 2SLS
42 . reg morekids multi2nd samesex agem1 agefstm boy1st boy2nd blackm hispm othracem

      Source |       SS           df       MS      Number of obs   =   394,840
-------------+----------------------------------   F(9, 394830)    =   4708.57
       Model |  9200.59068         9  1022.28785   Prob > F        =    0.0000
    Residual |  85722.3271   394,830   .21711199   R-squared       =    0.0969
-------------+----------------------------------   Adj R-squared   =    0.0969
       Total |  94922.9177   394,839  .240409174   Root MSE        =    .46595

------------------------------------------------------------------------------
    morekids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    multi2nd |   .6049071   .0080499    75.14   0.000     .5891295    .6206847
     samesex |   .0614735   .0014839    41.43   0.000     .0585652    .0643819
       agem1 |   .0301281   .0002318   129.97   0.000     .0296738    .0305825
     agefstm |  -.0452589   .0002801  -161.58   0.000    -.0458079   -.0447099
      boy1st |  -.0080449   .0014839    -5.42   0.000    -.0109533   -.0051366
      boy2nd |  -.0084413   .0014839    -5.69   0.000    -.0113497   -.0055329
      blackm |   .0696438   .0023467    29.68   0.000     .0650443    .0742433
       hispm |   .1565985    .004367    35.86   0.000     .1480392    .1651578
    othracem |   .0729161   .0044574    16.36   0.000     .0641797    .0816525
       _cons |   .3630539   .0072301    50.21   0.000     .3488831    .3772248
------------------------------------------------------------------------------

43 . predict more_hat if e(sample)
(option xb assumed; fitted values)

44 . reg weeksm1 more_hat agem1 agefstm boy1st boy2nd blackm hispm othracem, r

Linear regression                               Number of obs     =    394,840
                                                F(8, 394831)      =    2227.06
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0420
                                                Root MSE          =     21.813

------------------------------------------------------------------------------
             |               Robust
    weeksm1  |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    more_hat |  -4.141476   .5336062    -7.76   0.000    -5.187328   -3.095624
       agem1 |   1.320112   .0192077    68.73   0.000     1.282466    1.357759
     agefstm |   -1.20586   .0273376   -44.11   0.000     -1.25944   -1.152279
      boy1st |  -.0833355   .0695413    -1.20   0.231    -.2196343    .0529634
      boy2nd |  -.1417591   .0695644    -2.04   0.042    -.2781032   -.0054149
      blackm |    6.10575    .118865    51.37   0.000     5.872778    6.338722
       hispm |  -1.536577   .2171298    -7.08   0.000    -1.962145   -1.111009
    othracem |   2.513193   .2175725    11.55   0.000     2.086757    2.939628
       _cons |   6.379599   .3862117    16.52   0.000     5.622636    7.136563
------------------------------------------------------------------------------

45 .
46 . log close
      name:  <unnamed>
       log:  /Users/joshangrist/Documents/teaching/14.32/2020/notes/LN14/AE98/AE98for1432.smcl
  log type:  smcl
 closed on:  17 Apr 2020, 21:26:35
-------------------------------------------------------------------------------------------------------------------
```

  - China's One Child Policy is the most (in)famous of these anti-natalist policies
  - Economists call the relationship between family size and living standards the *quantity-quality tradeoff*

- Are larger families really impoverished by their size? If only we could randomize the number of children and find out!
  - Angrist and Evans (1998) and Angrist, Lavy, and Schlosser (2010) run natural experiments on family size in samples of women with 2 or more children
    - The *twins instrument*, Z_{1i} indicates multiple second births (buy one, get one free!)
    - The *samesex instrument*, Z_{2i} indicates mothers of two boys and two girls at parities 1 and 2 (diversify your sibling-sex portfolio!)
  - Z_{1i} and Z_{2i} are both highly predictive of the number of children born in family i
  - They're arguably independent of the potential human capital of the first-borns in these families (samples 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)

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 (.052)	.437 (.050)			.449 (.050)
Same-sex sibships			.079 (.012)	.073 (.010)	.076 (.010)
Male		−.018 (.010)		−.020 (.010)	−.020 (.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

		2SLS estimates		
Dependent variable	OLS estimates (1)	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 ≥ 24)	−.023	.017	.078	.048
	(.001)	(.052)	(.054)	(.037)
College graduate (for age ≥ 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 (15) without controls and with the second-stage residual written out:

$$Y_i = \alpha + \lambda_{2SLS}\hat{D}_i + [\eta_i + \lambda(D_i - \hat{D}_i)], \tag{18}$$

- 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:

$$\begin{aligned}\hat{\lambda}_{2SLS} &= \lambda_{2SLS}\frac{\sum \hat{D}_i(\hat{D}_i - \bar{D})}{\sum(\hat{D}_i - \bar{D})^2} + \frac{\sum \hat{D}_i\eta_i}{\sum(\hat{D}_i - \bar{D})^2} + \lambda_{2SLS}\frac{\sum \hat{D}_i(D_i - \hat{D}_i)}{\sum(\hat{D}_i - \bar{D})^2} \\ &= \lambda_{2SLS} + \frac{\sum \hat{D}_i\eta_i}{\sum(\hat{D}_i - \bar{D})^2}\end{aligned} \tag{19}$$

- The last term in the first line above is zero (why?)
- 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 std dev of residual η_i and $\sigma_{\hat{D}}$ is the std dev of first-stage fitted values, $\hat{D}_i$

Notes

- The standard errors generated by OLS estimation of (18) 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
- Likewise, we can say only that $\hat{\lambda}_{2SLS}$ is consistent; 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 first-stage F statistic for the instruments
    * 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
  - Robust, clustered, and Newey-West standard errors for 2SLS are known to Stata (again, valid only in large samples)
    * For more on 2SLS inference, see the MM Chapter 3 appendix and MHE chapter 8