LECTURE 12: INSTRUMENTAL VARIABLES (IV) REGRESSION: AN INTRODUCTION

March 16, 2010

Plan for Today

1. Instrumental Variables
2. Example: Vouchers for private school
3. Preview of reading for Thursday

Exogenous Assignment

Definition: Beyond manipulation by the participants themselves. Membership in the treatment or control group is totally independent of participants' own motivations and decisions.

Methods for Ensuring Treatment Exogeneity

- ✓ *True Experiment*- The experimenter explicitly and randomly assigns treatment conditions exogenously to groups
- ✓ *Natural Experiment*- An external agency, other than the experimenter, assigns treatment conditions exogenously to groups.
- ✓ *Instrumental Variables Estimation (IVE)*- An analytic strategy that is used to tease out any treatment exogeneity that is present in the question predictor so that it can be used directly in the estimation process.

Sources of Exogeneity

1. Natural disaster or abrupt change in policy
   - Dynarski (2003) SSB policy
2. Differences in policies or practices that occur across geographical boundaries
   - Tyler (2000) differences in state policies about the minimum score required to pass the GED
3. Forcing variables caused by sharp cutoffs in test score, class size, etc.

Potential Problems with OLS Estimates of Causal Effects

$$Y_i = \alpha + \beta_{OLS} X_i + \varepsilon_i$$

1. Omitted variables bias
2. Selection bias
3. Endogeneity (simultaneous causation) bias

All result in: Corr(X,ε)≠0.

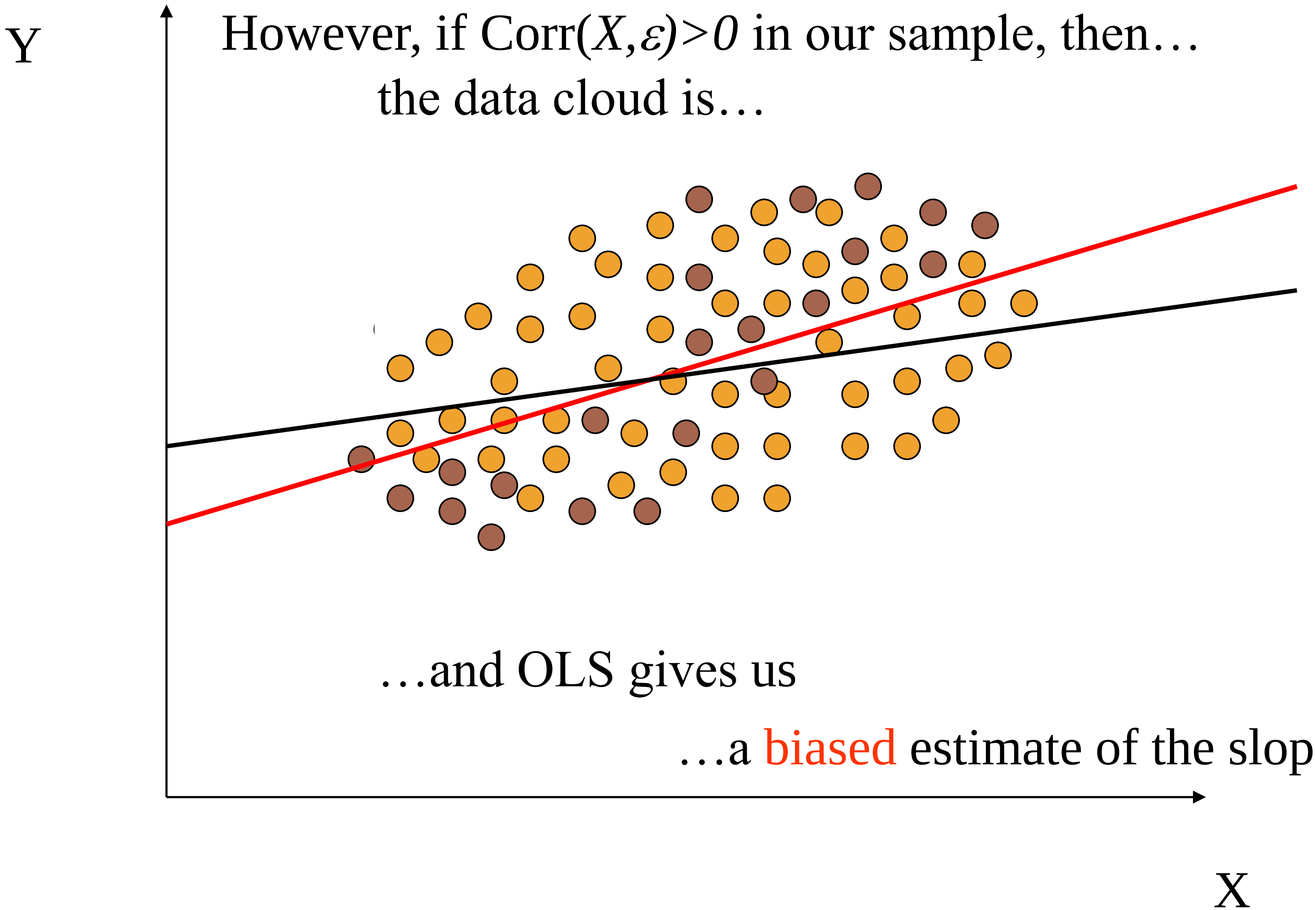
General definitions:

- Exogenous variable → not correlated with ε
- Endogenous variable → correlated with ε

We assume Corr(X,ε) = 0, but what if there is a systematic relationship between them in the data? Say, Corr(X,ε) > 0….

However, if Corr(X,ε)>*0* in our sample, then…

the data cloud is…



…and OLS gives us

…a biased estimate of the slope!

OLS Model

Outcome: Student Achievement

Dummy predictor (0 = no, 1=yes)

Covariates (or controls)

$$\text{Model}: \boxed{Y_i} = \beta_0 + \boxed{\beta_1}\boxed{V_i} + \gamma \boxed{X_i} + \boxed{\varepsilon_i}$$

Residual

- β_1 represents the *causal effect* of *lottery assignment* ("intent to treat") on *student achievement*
- With random assignment, we can obtain an *unbiased estimate* using *OLS regression analysis*

But, say we want to know…

- whether the i^{th} student actually attends private school
  - But we know that some parents will send their kids to private school without the voucher because
    - they want their kids taught in a religious setting
    - they want their kids out of the public schools
    - they have greater financial resources, etc.
  - And some parents who could send their kids to private school with the voucher, will not use it due to
    - Transportation issues
    - Cost issues, etc.

Replace the offer of treatment with a new dummy predictor which indicates ***whether the i^th^ student attends private school*** (0 = no, 1=yes)

$$\text{Model}: Y_i = \beta_0 + \beta_1 \boxed{V_i} + \gamma X_i + \boxed{\varepsilon_i}$$

What is the problem?

- Values of V now depend on *unobserved personal and family characteristics*
- The same *unobserved characteristics* that made them choose, or not choose, a private school (such as motivation or resources) *may also predict the student achievement outcome* directly.

➢ But, these *unobserved characteristics* are not explicitly included as predictors in the model, and *must form part of the residual, ε.*

➢ Therefore the *residuals* are now *correlated with V.*

The Problem

- We're concerned that *unobserved* effects (like family income, motivation, resources, etc) impact the outcome, say test scores, but are omitted as predictors and are potentially correlated with private school attendance, leading to a *biased OLS* estimate of the effect of *attendance at private school* on *test scores*, β_1.

The Instrumental Variable (IV) Solution

$$Y_i = \alpha + \beta_{OLS} X_i + \varepsilon_i, \text{corr}(X, \varepsilon \neq 0)$$

- IV Regression breaks the variation in *X* into two parts:
  1. Part that is correlated with ε (bad)
  2. Part that is uncorrelated with ε (good)
- How? With an instrumental variable, Z_i, that is uncorrelated with ε_i
- A <u>valid</u> instrument lets us isolate variation in *X* that is unrelated to ε—it is "as if" randomly assigned
- We can then use this variation to estimate β and get an unbiased estimate of a true causal effect

Conditions for a valid instrument

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

For an instrumental variable (an "***instrument***") Z to be valid, it must satisfy two conditions:

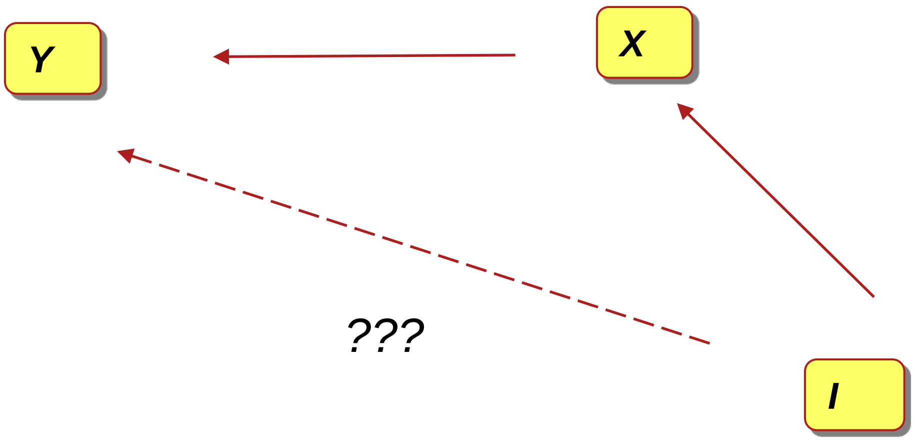
1. ***Instrument relevance***: $\text{corr}(Z_i, X_i) \neq 0$
2. ***Instrument exogeneity***: $\text{corr}(Z_i, \varepsilon_i) = 0$

In other words, the instrument must be correlated with X but must not be causally related to Y—except through its relationship with X.

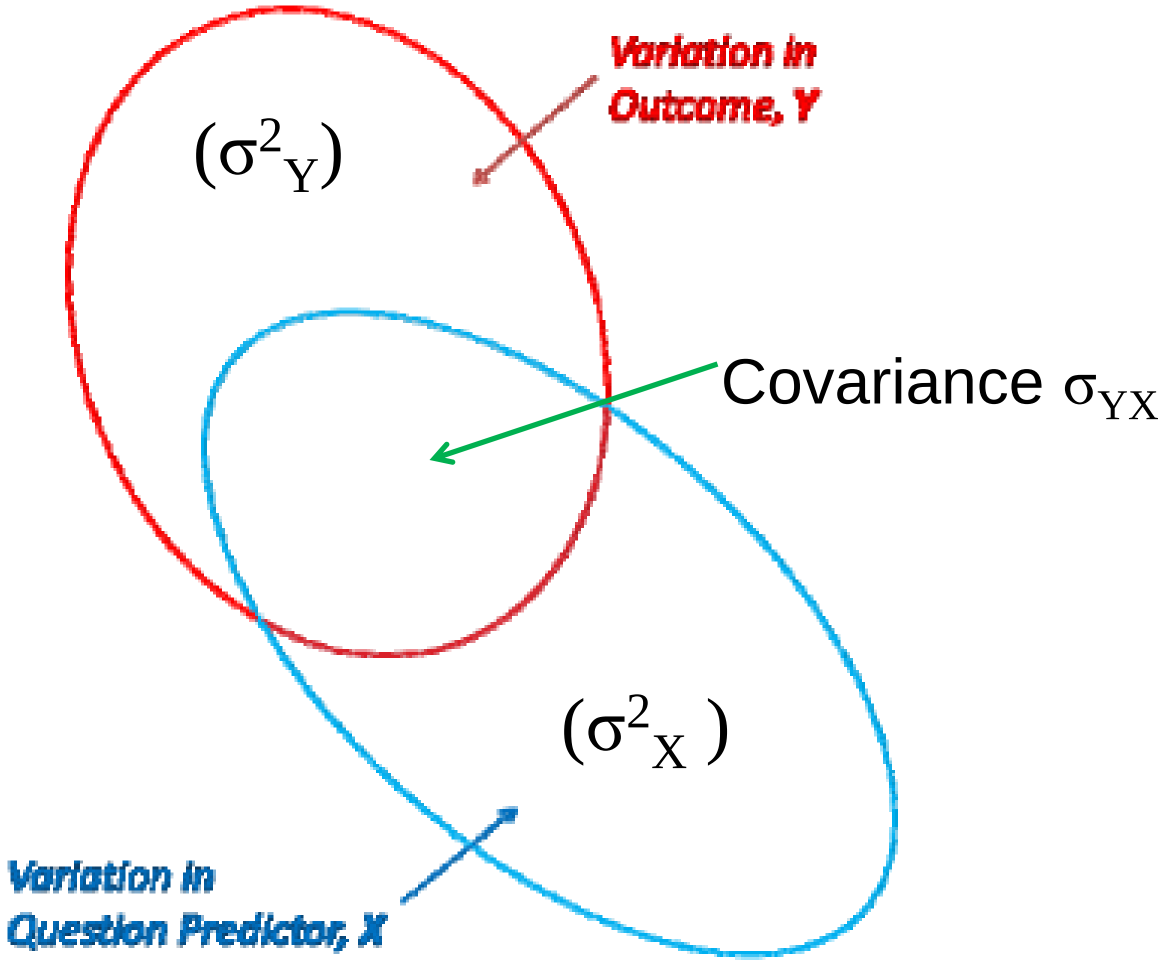
Suppose for now that you have such a Z_i → How can you use it to estimate β?

Instrumental Variables

- The instrument must act on the outcome only through the question predictor



OLS Approach



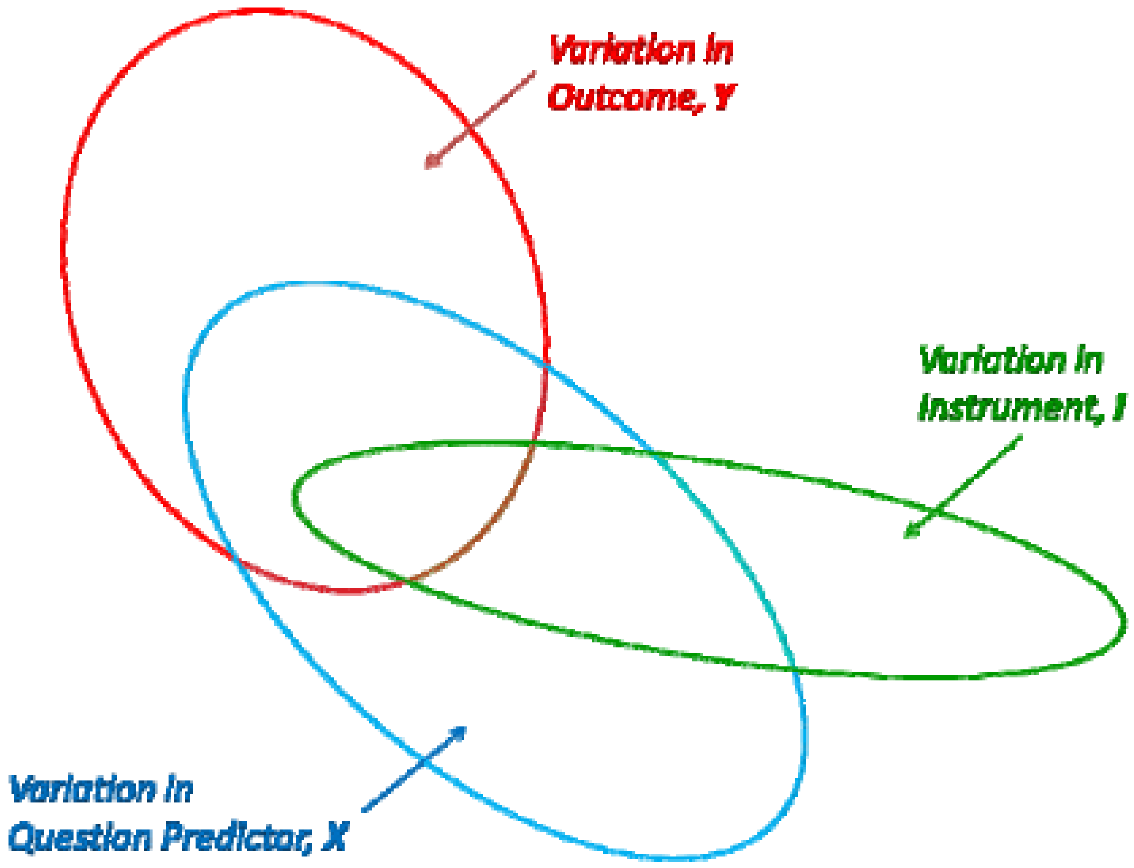
OLS design (ignoring endogeneity)

Want to know the effects of attending private school on reading test scores in first grade, controlling for kindergarten reading scores

```
. reg   read1  read0  privt1

      Source |       SS       df       MS              Number of obs =    1449
-------------+------------------------------           F(  2,  1446) =  389.32
       Model |  232982.998     2  116491.499           Prob > F      =  0.0000
    Residual |  432668.989  1446  299.217835           R-squared     =  0.3500
-------------+------------------------------           Adj R-squared =  0.3491
       Total |  665651.988  1448  459.704411           Root MSE      =  17.298

------------------------------------------------------------------------------
       read1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       read0 |   .5876687   .0210745    27.89   0.000     .5463288    .6290085
      privt1 |   1.298631   .9107364     1.43   0.154    -.4878752    3.085137
       _cons |   10.77581   .7874507    13.68   0.000      9.23114    12.32048
------------------------------------------------------------------------------
```

But, is this biased?

IV Approach



Variation in
Outcome, Y
Variation in
Instrument, I
Variation in
Question Predictor, X

First Stage: Use offer of treatment to tease out the exogenous part of attending private school

1^{st} Stage Model: $privt1_i = \alpha_0 + \alpha_1 treat_i + \alpha_2 read0_i + \delta_i$

1^{st} stage outcome

Instrument

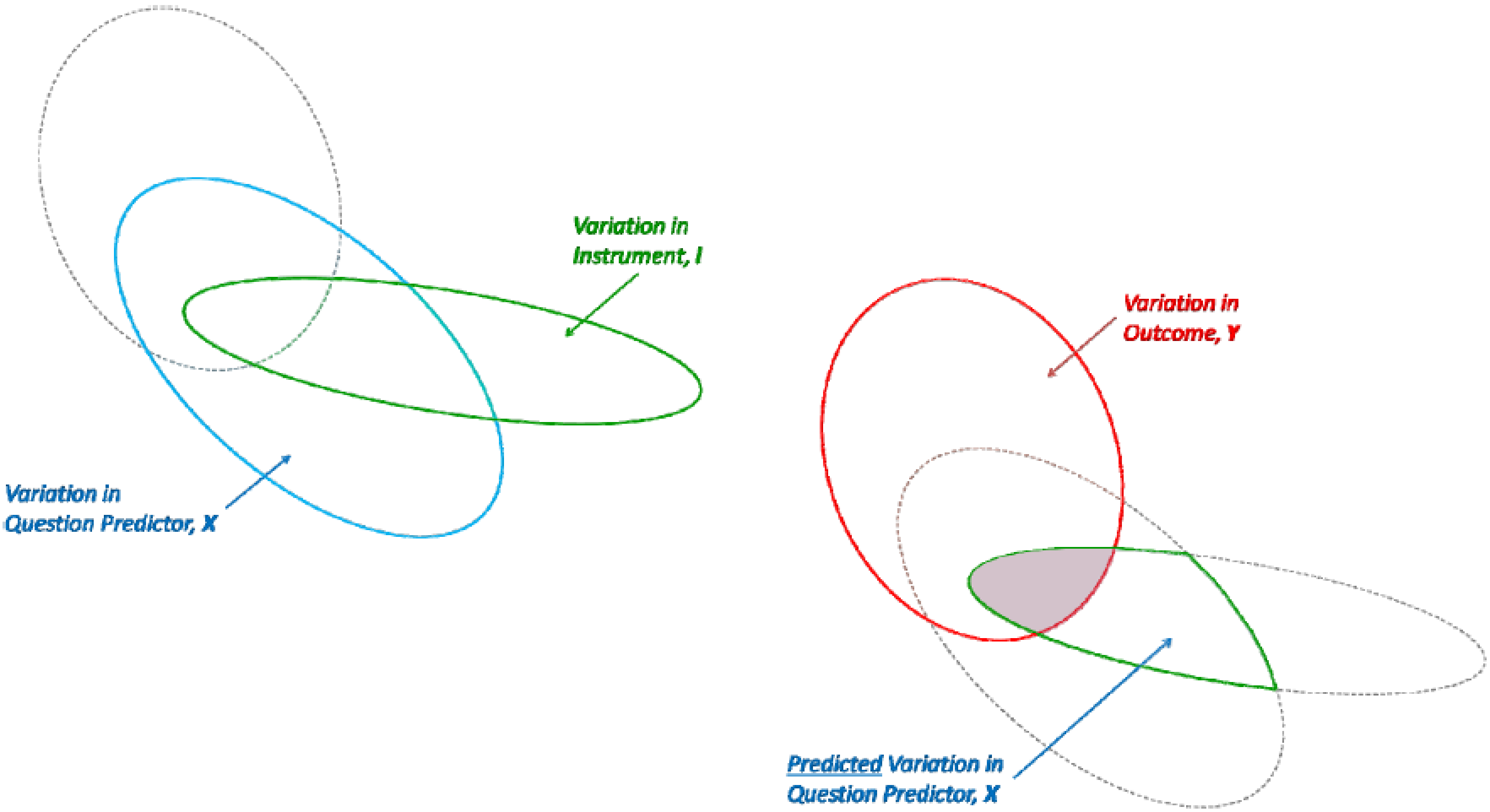
All control predictors that will be included in the *2^{nd} stage* model should *also* be included in the *1^{st} stage* model.

```
. reg  privt1  read0  treat

      Source |       SS       df       MS              Number of obs =    1449
-------------+------------------------------           F(  2,  1446) = 1044.75
       Model |  213.247597     2  106.623799           Prob > F      =  0.0000
    Residual |  147.573659  1446  .102056472           R-squared     =  0.5910
-------------+------------------------------           Adj R-squared =  0.5904
                                                              SE     =  .31946
                                                       ----------------
      privt1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------
       read0 |   .0003896   .0003895     1.00   0.31
       treat |   .7700955     .01685    45.70   0.00
       _cons |   .0447431   .0153769     2.91   0.00
------------------------------------------------------

. predict privthat
```

Fitted 1^{st} Stage Model: $\widehat{privt1}_i = \hat{\alpha}_0 + \hat{\alpha}_1 treat_i + \hat{\alpha}_2 read0_i$

The predicted values estimate *that part of each person's value of P that is* ***exogenous***.

1 st Stage vs. 2nd Stage



2^{nd} stage model

$$2^{nd}\text{ Stage Model}: \boxed{read1_i} = \beta_0 + \beta_1 \boxed{pri\hat{v}t1_i} + \gamma \boxed{read0_i} + \varepsilon_i$$

2^{nd} stage outcome, *reading achievement*

Predicted (exogenous) part of *attendance at private school*

Control

```
. reg    read1  read0  privthat

      Source |       SS       df       MS              Number of obs =    1449
-------------+------------------------------           F(  2,  1446) =  388.21
       Model |  232549.631     2  116274.816           Prob > F      =  0.0000
    Residual |  433102.356  1446  299.517536
-------------+------------------------------
       Total |  665651.988  1448  459.704411

------------------------------------------------------------
       read1 |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------------
       read0 |   .5875379   .0210866    27.86   0.000
    privthat |   .9060803   1.185346     0.76   0.445
       _cons |    10.9627   .8665962    12.65   0.000
------------------------------------------------------------
```

Causal effect of attendance at private school is smaller than in the earlier *OLS* analysis (0.91 vs. 1.30), suggesting that the earlier *OLS* estimate was *biased*.

IV Estimation in Stata: ivreg

STATA command

2nd stage outcome, *reading achievement*

Control predictor

Hypothesized 1st stage equation specifying the relationship between the potentially endogenous second-stage predictor and the instrument

```
. ivreg    read1  read0 (privt1=  treat)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    1449
-------------+------------------------------           F(  2,  1446) =  388.55
       Model |  232927.409     2  116463.704           Prob > F      =  0.0000
    Residual |  432724.579  1446  299.256279           R-squared     =  0.3499
-------------+------------------------------           Adj R-squared =  0.3490
          al |  665651.988  1448  459.704411           Root MSE      =  17.299

------------------------------------------------------------------------------
       read1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      privt1 |   .9060804   1.184829     0.76   0.445    -1.418087    3.230248
       read0 |   .5875379   .0210774    27.88   0.000     .5461924    .6288834
       _cons |    10.9627   .8662181    12.66   0.000     9.263526    12.66188
------------------------------------------------
Instrumented:   privt1
Instruments:    read0 treat
------------------------------------------------------------------------------
```

Estimate identical to the *2SLS* estimate

Specification of the instrumentation (notice the automatic inclusion of the control predictor)

IV Estimation in Stata: ivreg

```
. help ivreg

help for ivreg                                          manual:  [R] ivreg
                                                        dialogs:  ivreg  predict

Instrumental variables and two-stage least squares regression

        ivreg depvar [varlist1] (varlist2=varlist_iv) [weight] [if exp] [in range] [, level(#)
              beta hascons noconstant robust cluster(varname) first noheader eform(string)
              depname(varname) mse1 ]
```

```
Examples

    . ivreg y1 (y2 = z1 z2 z3) x1 x2 x3
    . ivreg y1 x1 x2 x3 (y2 = z1 z2 z3)
```

An example

- Howell et al, “School Vouchers and Academic Performance: Results from Three Randomized Field Trials” (2002)

The Voucher Debate

- Empirical question #1: impact of attending private school
- Empirical question #2: impact of choice on public schools
- Previous literature on private school effects:
  - Attainment higher in private schools
  - Achievement higher for urban minorities
  - Based entirely on observational data → do differences reflect causal effects?

Selection Bias

- Experimental design used to address the fundamental problem of selection bias
  - More eligible applicants than voucher slots
  - Applicants randomly assigned to receive or not receive a school voucher → equally motivated
  - Baseline data used to assess randomization
  - Any differences should be due to voucher receipt

The Programs Under Evaluation

Table 1. Description of the voucher programs.

	New York, NY	Dayton and Montgomery County, OH	Washington, DC
Name of program	School Choice Scholarships Foundation	Parents Advancing Choice in Education	Washington Scholarship Fund
First year of program	1997–1998	1998–1999	1998–1999
Max. amount of scholarship	$1400	$1200	$1700
Eligible grades in first year	1–4	K–12	K–8
Income eligibility	Eligible for federal free lunch program	Up to 2× federal poverty line	Up to 2.5× federal poverty line
Num. students from public schools that were tested at baseline	1,960	803	1,582
Response rate in 1st year	82%	56%	63%
Response rate in 2nd year	66%	49%	50%

Results: Effects of a Voucher Offer (Intent to Treat)

$$Y_t = \alpha + \beta_1 \mathrm{V} + \beta_2 \mathrm{Y}_{0\mathrm{R}} + \beta_3 \mathrm{Y}_{0\mathrm{M}} + u$$

Table 5. Effect of a voucher offer on the test scores of African Americans and other ethnic groups in three cities after 1 and 2 years.

	New York, NY		Dayton, OH		Washington, DC	
	Af. Am. (1)	Oth. Ethn.[1] (2)	Af. Am. (3)	Oth. Ethn.[2] (4)	Af. Am. (5)	Oth. Ethn.[3] (6)
Second Year						
Offered Voucher	3.27**	–1.04	3.46*	–0.08	3.80***	–0.08
	(1.50)	(1.50)	(1.98)	(3.96)	(1.16)	(0.42)
Baseline Scores						
Math	0.37***	0.37***	0.22***	0.39***	0.40***	0.42***
	(0.04)	(0.03)	(0.05)	(0.07)	(0.03)	(0.18)
Reading	0.29***	0.40***	0.37***	0.36***	0.14***	0.24
	(0.03)	(0.03)	(0.04)	(0.07)	(0.02)	(0.15)
Constant	0.79	10.94	11.52***	15.47***	6.49***	11.77**
Adjusted R^2	.43	.47	.34	0.50	.34	.45
(N)	497	699	273	96	668	42

Practical Issues: Non-compliance

Table 6. Attendance patterns among treatment and control groups.

	New York	Dayton	Washington
All Students	%	%	%
Individuals offered a voucher who attended a private school in 1st year	82	78	68
Individuals not offered a voucher who attended a private school in 1st year	5	18	11
Individuals offered a voucher who attended a private school both years	79	60	47
Individuals not offered a voucher who attended a private school both years	3	10	8

Two issues:

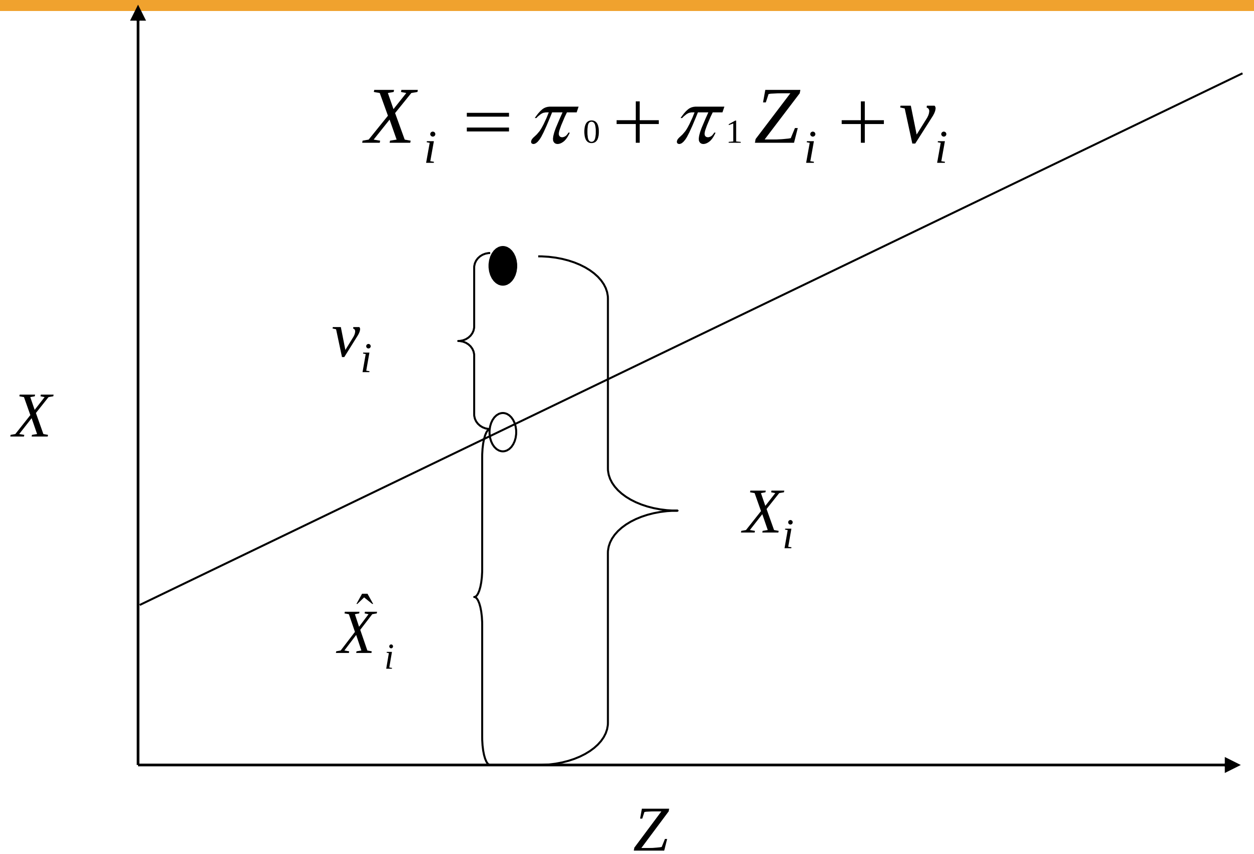
1. Some treatment group members don't attend private schools
2. Some control group members did!

Two Stage Least Squares (2SLS)

2SLS involves two stages/regressions:

1. First regress X on Z using OLS to isolate the part of X that is uncorrelated with u:
   - (1) $X_i = \pi_0 + \pi_1 Z_i + v_i$
   - Compute the predicted values of X_i, where $\hat{X}_i = \pi_0 + \pi_1 Z_i$

First Stage Regression



Because z is by assumption uncorrelated with ε, $\hat{X}$ must be uncorrelated with ε.

Two Stage Least Squares (2SLS)

2SLS involves two stages/regressions:

1. First regress X on Z using OLS to isolate the part of X that is uncorrelated with u:
   - (1) $X_i = \pi_0 + \pi_1 Z_i + v_i$
   - Compute the predicted values of X_i, where $\hat{X}_i = \pi_0 + \pi_1 Z_i$

2. Then regress Y on $\hat{X}_i$ using OLS:
   - (2) $Y_i = \alpha + \beta_{IV} \hat{X}_i + u_i$
   - Exclude the instrument from the 2nd stage regression

2SLS

Stage 1 Model: Predict "voucher use" using assignment status (randomized lottery)

$$(1)\ \widehat{VoucherUse}_i = \beta_0 + \beta_1 Lottery_i + \beta_2 X_i + \mu_i$$

Stage 2 Model: Use exogenous variation in voucher use to predict outcome

$$(2)\ Y_i = \delta_0 + \delta_1 \widehat{VoucherUse}_i + \delta_2 X_i + \varepsilon_i$$

IV Estimation

- The advantage of IV estimation:
  - β_{IV} offers **unbiased** estimates of β, as opposed to β_{OLS}

- Disadvantage of IV estimation:
  - Because we are using an estimate of X_i we have lost some information
  - Loss of information is always reflected in **larger standard errors.**

Another Disadvantage

- Because we are using only part of the variation in T_i to estimate β, we are no longer estimating the average treatment effect (ATE)
- Instead, we can only estimate **local average treatment effects (LATE)** → effect for "compliers" only (those whose behavior is associated with the instrument)

Local Average Treatment Effect

- It is only the variation in X that is affected by the instrument that has been capitalized upon in estimating the outcome. The estimate does not provide any information about the impact of X on the outcome for individuals whose decision about X was not influenced by the instrument.
- If the effect of X on the outcome is "homogeneous" across all sectors of the population, then the "average" and the "local average" treatment effects will be identical and both represented by the same population slope, β_1.

Rank Condition

- *For every endogenous predictor included in the second stage, there must be at least one instrument included in the first stage.*
- If we include one potentially endogenous main effect and three potentially endogenous interactions in the second-stage model, then we must include at least four instruments in the first-stage model.
- Just use the corresponding interactions between the original instrument for the main effect of *X* and its interactions with the same exogenous covariates

3 Key Assumptions for IV

1. ***The instrument must be correlated with the predictor***

- If *Assumption #1 is violated* and instrument, *I*, is *uncorrelated* with *X*:
  - All fitted values from the 1st stage equation will be constant and equal to the sample mean of *X*.
  - There will be no variability in the values of the “instrumented predictor” in the 2nd stage equation.
  - The estimated value of β_1 will be zero, regardless of its value in the population.

3 Key Assumptions for IV

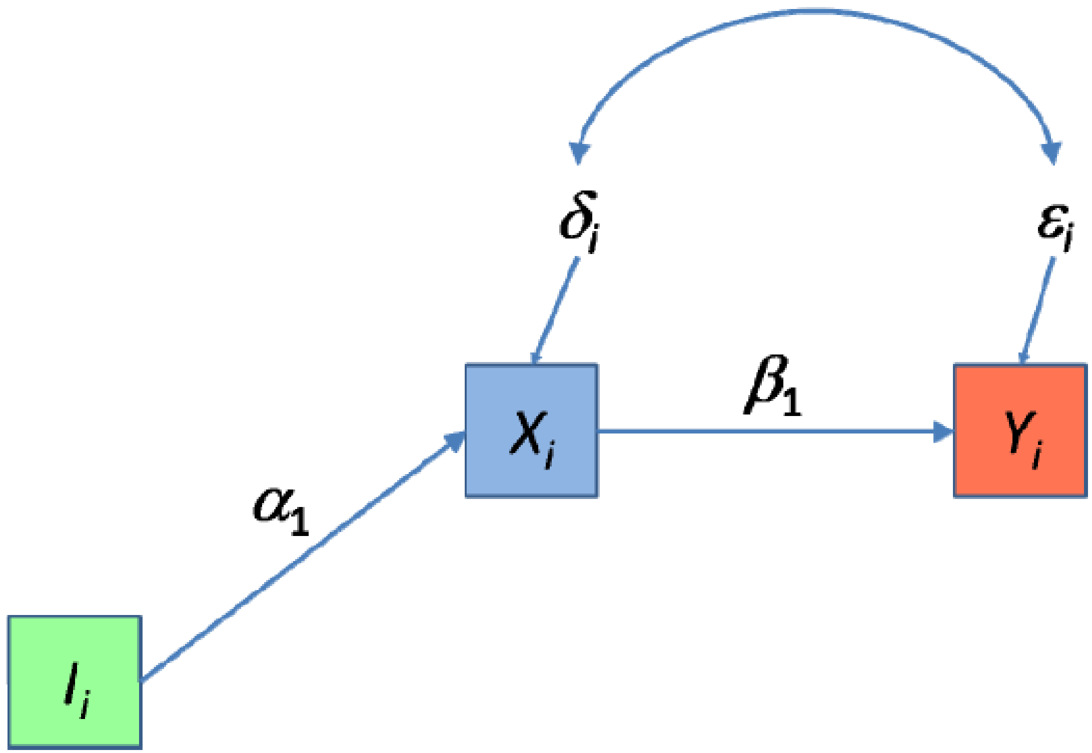
2. ***The instrument must be uncorrelated with the unobserved effects that have made the question predictor endogenous in the first place (i.e., it must be uncorrelated with the residuals in the 2^{nd} stage model).***

If *Assumption #2 is violated*, and *I* and ε are *correlated*:

- Then and ε will be correlated too, because is just a function of *I* (see the fitted 1^{st} stage equation).
- This means that the "instrumented" predictor will still be correlated with the residuals in the 2^{nd} stage model, and your OLS estimate of β_1 will still be biased.

3 Key Assumptions for IV

3. The instrument must act on the outcome only through the question predictor



Using IV Regression to Estimate Treatment-on-Treated (TOT) Effects

- Solution: use the offer of a voucher as an instrument for private school attendance
  - Relevant?
    - Correlated with private school attendance
  - Exogenous?
    - It was random (and no reason to think losing lottery affects outcomes directly)
- Two stage least squares:

$$P = \alpha_1 + \beta_1 \mathrm{V} + \beta_2 \mathrm{Y}_{0\mathrm{R}} + \beta_3 \mathrm{Y}_{0\mathrm{M}} + u_1$$

$$Y_t = \alpha_1 + \beta_4 \hat{\mathrm{P}} + \beta_5 \mathrm{Y}_{0\mathrm{R}} + \beta_6 \mathrm{Y}_{0\mathrm{M}} + u_2$$

Results: Private school effects

Table 7. Effect of switching from a public to a private school on the test scores of African Americans and other ethnic groups in three cities after 1 and 2 years.

	New York, NY		Dayton, OH		Washington, DC	
	Af. Am. (1)	Oth. Ethn. (2)	Af. Am. (3)	Oth. Ethn. (4)	Af. Am. (5)	Oth. Ethn. (6)
Second Year						
Private School	4.41**	–1.54	6.45*	–0.19	9.22***	–0.14
	(2.03)	(2.23)	(3.66)	(8.96)	(2.86)	(9.77)
Baseline Scores						
Math	0.37***	0.37***	0.23***	0.39***	0.39***	0.42**
	(0.04)	(0.03)	(0.05)	(0.08)	(0.03)	(0.19)
Reading	0.29***	0.40***	0.37***	0.36***	0.13***	0.24
	(0.03)	(0.03)	(0.04)	(0.08)	(0.02)	(0.15)
Constant	0.44	11.11	10.77***	15.52***	6.49***	11.76*
Adjusted R^2	0.42	0.47	0.35	0.50	0.33	0.45
(N)	497	699	273	96	668	42

Conclusions

- Voucher use benefits African Americans only
  - Explanations?
  - Implications?

IV Estimation: Summary

- A valid instrument isolates variation in a potentially endogenous variable that is "as if" randomly assigned
  - Relevance evaluated with 1st stage regression results
  - Exogeneity can not be evaluated directly

- Drawbacks include larger standard errors and limited external validity, but estimation is straightforward

The Real Challenge in IV Estimation…

Finding valid, defensible instruments

Other kinds of instruments

- When investigating the impact of *educational attainment* on *labor market outcomes,* researchers used several different instruments for educational attainment:
  - ✓ Card (1993) used the *presence of a nearby college* to instrument for amount of schooling.
  - ✓ Butcher & Case (1994) used *family sibling composition* to instrument for educational attainment.
  - ✓ Angrist & Krueger (1991a) use *quarter of an individual's birth* to instrument for completed schooling.
  - ✓ Angrist & Krueger (1991b) use *Vietnam-era draft lottery numbers* to instrument for number of years of education.

Currie & Moretti (2003)

- Currie & Moretti (2003) use the *availability of colleges in the woman's county in her 17th year* as an instrument for *mother's educational attainment* when investigating the effect of mother's educational attainment on birth outcomes (birth weight, gestational age).