Instrumental Variables Estimation: Why, What and How?

Elaine Liu, Ph.D.
Department of Economics
University of Houston

CERCIT
UTMB Galveston
September 16, 2011

Topics That Will Be Covered

- **Why** use IV?
  - Discussion of endogeneity bias
  - Statistical motivation for IV
- **What** is an IV?
  - Identification issues
  - Statistical properties of IV estimators
- **How** is an IV model estimated?
  - Software and data examples
  - Diagnostics: IV relevance, IV exogeneity, Hausman
- **Keywords: Endogeneity, Exogeneity, L.A.T.E., overidentified**

2

Some example

- Link

"Hurricane Ike's Impact On Whether Teenagers Drink, Use Drugs"

Dr. Jeff Temple

"Kids who did not evacuate and were thus exposed to the hurricane itself, we found that those kids were more likely to report substance use and teen dating violence than the kids that did evacuate."

3

Some reasons

- Those who stayed are a selected group
  - Their parents may be more risk loving
  - Their parents may lack the resources
- Post-Traumatic Stress Disorder (PTSD)
- Correlation ≠ Causation
- Any solution?
  - RCT, IV, propensity score matching

4

Review of the Linear Model

- Population model: $\mathbf{Y = \alpha + \beta X + \varepsilon}$
- Sample model: $\mathbf{Y = a + bX + e}$
  - Least squares (LS) estimator of β:
- $\mathbf{b_{LS} = (X'X)^{-1}X'Y = Cov(X,Y) / Var(X)}$
- Key assumption of the linear model:

  $\mathbf{E(X'e) = Cov(X,e) = E(e \mid X) = 0}$
  - Exogeneity assumption = X is uncorrelated with the unobserved determinants of Y

5

Review of the Linear Model

- Important statistical property of the LS estimator **under exogeneity**:

  $\mathbf{E(b_{LS}) = \beta + Cov(X,e) / Var(X)}$

X is called **"endogenous"** if x correlated with the error) That is, whenever Cov(x,e) ≠ 0

6

Endogeneity and the Evaluation Problem

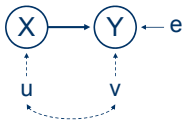
- When is the exogeneity assumption violated?
  - Measurement error
  - Reverse causation
  - Omitted Variables Bias

7

When Is the Exogeneity Assumption Violated?

(1) Measurement error in X (u) that is correlated with M.E. in Y (v) or with the model error (e)

- Classical M.E. leads to **attenuation**, $0 < E(b_{LS}) < \beta$, but non-random M.E. (or correlation between M.E. and X, Y, V, and/or e) introduces unknown biases



8

When Is the Exogeneity Assumption Violated?

(2) Instantaneous causation of Y on X

- Direction of the bias depends on what the sign is for the feedback effect, Y → X
  - If positive, $E(b_{LS}) > \beta$, so overestimate true effect
  - If negative, $E(b_{LS}) < \beta$, so underestimate true effect and in severe cases can even flip the sign so that $E(b_{LS}) < 0$ even though $\beta > 0$

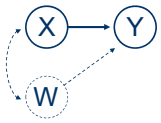
This non-recursivity complicates the relationship between price and quantity in economics

9

When Is the Exogeneity Assumption Violated?

(3) Omitted variable (W) that is correlated with both X and Y

- Classic problem of omitted variables bias
  - Coefficient on X will absorb the indirect path through W, whose sign depends on Cov(X,W) and Cov(W,Y)



10

Example: Effectiveness of Chemotherapy for Advanced Lung Cancer in the Elderly: Instrumental Variable and Propensity Analysis

- *Journal of Clinical Oncology*, Vol 19, Issue 4 (February), 2001: 1064-1070
- Use instrumental variable analysis to simulate the conditions of a randomized trial in a retrospective cohort of patients over age 65 from the Survival, Epidemiology, and End Results (SEER) tumor registry
- "randomized clinical trials,…,are usually carried out on highly selected patients, in centers of excellence, and under optimal protocol-driven conditions. While a trial may have been well conducted, and thus be internally valid, the results may not be generalizable, or externally valid"

11

Not a Problem With Endogeneity

- Population model: $\mathbf{Y = \alpha + \beta X + \varepsilon}$
- Sample model: $\mathbf{Y = a + bX + e}$
- Not a representative sample of population
  - Y= patient outcome
  - X= chemotherapy (can be quantity or an indicator variable)

12

Example #3: Does Maternal Smoking Harm Infant Health?

- True Model: $Y_i = \alpha + \beta_1 S_i + \beta_2 A_i + v_i$
  - Y= infant health outcome (birth weight)
  - S= maternal smoking (can be quantity or an indicator variable)
  - A = other factors (observables and unobservables) such as education, maternal drinking, prenatal care, "good mother" indicator

13

Example : Maternal Smoking

- Measurement error
  - Measures of intensity of smoking can be noisy (M.E. in X) -> resulting in attenuation at best
  - Mothers may not be truthful about whether they smoke during pregnancy (not a classical measurement error)
- Omitted variables
  - Unobserved characteristics such as how *caring* the mother is, or how *responsible* the mother is

14

Regression Estimation Ignoring Omitted Variables

Thus, estimation of the following would be biased:

$$Y_i = a + b_1 S_i + e_i$$

because S_i and e_i are correlated as

$$E[b_1]=E[Y'S]/E[S'S]$$
$$=E[(\beta_1 S_i+\beta_2 A_i+v_i)'S] / E[S'S]$$
$$= \beta_1 + E[(\beta_2 A_i+v_i)'S] / E[S'S]$$
$$= \beta_1 + \beta_2 E[A_i'S] / E[S'S]$$
$$\neq \beta_1$$

15

Regression Estimation Ignoring Omitted Variables

- What does LS estimate when A is omitted?

$$b_{LS=} \beta_1 + \beta_2 \; E[A_i'S] / E[S'S]$$

True impact of smoking on birth weight (–)

Impact of responsibility on birth (–)

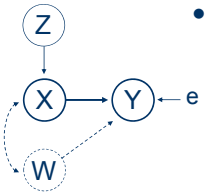
Difference in "responsibility" between smoking and non-smoking mother (+)

- The effect of smoking on birth weight will be overestimated

16

Instrumental Variables Estimation Is a Viable Approach

- An "instrumental variable" for X is one solution to the problem



- Requirements for Z to be a valid instrument for X
  - Relevant = Correlated with X
  - Exogenous = Not correlated with Y **but through** its correlation with X

17

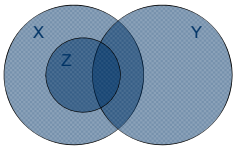
Important Point about Instrumental Variables Models

- I often hear..."A good instrument should not be correlated with the dependent variable"
  - WRONG!!!
- Z has to be correlated with Y, otherwise it is useless as an instrument
  - It can only be correlated with Y through X
- **A good instrument must not be correlated with the error term (exclusion restriction)**

18

Important Point about Instrumental Variables Models

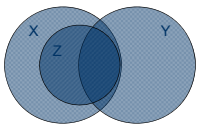
- Not all of the available variation in X is used
  - **Only that portion of X which is "explained" by Z is used to explain Y**



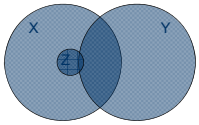
X = Endogenous variable
Y = Response variable
Z = Instrumental variable

19

Important Point about Instrumental Variables Models



Best-case scenario: A lot of X is explained by Z, and most of the overlap between X and Y is accounted for



Realistic scenario: Very little of X is explained by Z, or what is explained does not overlap much with Y

20

Any Suggestion for Instruments?

- #1 Study: Whether one stays in Galveston or Not
- #2 Study: Chemotherapy on Elderly Cancer Patient
  - significant unexplained geographic variation, we divided Health Care Service Areas (HCSA) into quintiles of chemotherapy utilization
- #3 Study: Maternal Smoking

21

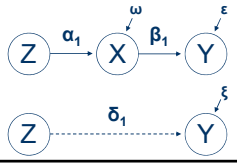
Instrumental Variables Terminology

- Three different models to be familiar with
  - First stage: $\mathbf{X = \alpha_0 + \alpha_1 Z + \omega}$
  - Structural model: $\mathbf{Y = \beta_0 + \beta_1 X + \varepsilon}$
  - Reduced form: $\mathbf{Y = \delta_0 + \delta_1 Z + \xi}$
- An interesting equality:

  $\mathbf{\delta_1 = \alpha_1 \times \beta_1}$

  so…

  $\mathbf{\beta_1 = \delta_1 / \alpha_1}$



22

More on the Method of Two-Stage Least Squares (2SLS)

- Step 1: $\mathbf{X = a_0 + a_1 Z_1 + a_2 Z_2 + \cdots + a_k Z_k + u}$
  - Obtain fitted values ($\hat{X}$) from the first-stage model
- Step 2: $\mathbf{Y = b_0 + b_1 \hat{X} + e}$
  - Substitute the fitted $\hat{X}$ in place of the original X
  - Note: If done manually in two stages, the standard errors are based on the wrong residual

    $\mathbf{e = Y - b_0 - b_1 \hat{X}}$ when it should be $\mathbf{e = Y - b_0 - b_1 X}$
- Best to just let the software do it for you

23

Including Control Variables in an IV/2SLS Model

- Control variables (W's) should be entered into the model at both stages
  - First stage: $\mathbf{X = a_0 + a_1 Z + a_2 W + u}$
  - Second stage: $\mathbf{Y = b_0 + b_1 \hat{X} + b_2 W + e}$
- For Study #2:

$$\text{IV estimate} = \frac{\text{Adjusted Survival}_{\text{Hi}} - \text{Adjusted Survival}_{\text{Lo}}}{\text{P(chemotherapy}|\text{Hi)} - \text{P(chemotherapy}|\text{Lo)}}$$

24

Instrumental Variables and Randomized Experiments

- Imperfect compliance in randomized trials
  - Some individuals assigned to treatment group will not receive T_x, and some assigned to control group will receive T_x
    - Assignment error; subject refusal; investigator discretion
  - Some individuals who receive T_x will not change their behavior, and some who do not receive T_x will change their behavior

25

Instrumental Variables and Randomized Experiments

- Two different measures of treatment (X)
  - Treatment assigned = Exogenous
    - Intention-to-treat (ITT) analysis
      - Reduced-form model: $\mathbf{Y = \delta_0 + \delta_1 Z + \xi}$
    - Often leads to underestimation of treatment effect
  - Treatment delivered = Endogenous
    - Treatment on the Treated (TOT) : Individuals who do not comply probably differ in ways that can undermine the study
    - Self-selection ∴ bias and inconsistency

26

Sexton and Hebel (1984), *J.A.M.A.*

- Maternal smoking and birth weight
  - Sexton and Hebel (1984)
    - Sample of pregnant women who were confirmed smokers, recruited from prenatal care registrants
      - At least 10 cigarettes per day and not past 18th week
  - Random assignment of staff assistance in a smoking cessation program
    - Personal visits; telephone and mail contacts
  - But...some smokers in treatment group did not quit and some smokers in control group did quit

27

Sexton and Hebel (1984), *J.A.M.A.*

Table 2.—Smoking Characteristics of Control and Treatment Groups at Eighth Month of Pregnancy*

	Control Group	Treatment Group
% reporting smoking, cigarettes/day†		
N	395‡	388§
0	20.0	43.0
1-5	12.7	19.1
6-10	22.0	16.2
11-20	31.4	17.8
≥20	13.9	3.9
➔ Mean ± SD‖	12.8 ± 11.5	6.4 ± 6.7
Thiocyanate, µmole/L		
N	389	380
Mean ± SD¶	2,452 ± 1,228	2,094 ± 1,209

Table 4.—Measurement of Status of Newborn*

	Control Group (N=438)		Treatment Group (N=429)		
	N	Mean ± SD	N	Mean ± SD	t
Primary factor					
➔ Birth weight, g	438	3,186 ± 656	429	3,278 ± 627	2.28†
<2,500, %	438	8.9	429	6.8	. . .
<1,500, %	438	1.1	429	1.9	. . .
Other factors					
Birth length, cm	410	49.70 ± 4.16	406	50.30 ± 3.67	2.16†
Head circumference, cm	421	34.06 ± 2.44	413	34.14 ± 1.79	0.54
Apgar score (1 min)	430	8.01 ± 1.56	425	7.96 ± 1.66	0.49
<7, %	430	10.5	425	10.6	. . .
Apgar score (5 min)	431	9.00 ± 0.92	425	8.97 ± 1.09	0.45
<7, %	431	1.9	425	2.8	. . .
Gestational age, wk	438	39.60 ± 2.77	429	39.65 ± 2.77	0.22

(1) First-stage model

Mean cigarettes smoked:

Treatment = 6.4

Control = 12.8

First-stage effect: $\mathbf{b_{FS} = -6.4}$

(2) Reduced-form model

Mean birth weight:

Treatment = 3,278g

Control = 3,186g

Reduced-form effect: $\mathbf{b_{RF} = 92}$

(3) Structural model

Effect of smoking frequency on mean birth weight:

$\mathbf{b_{IV} = 92 / -6.4 = -14.4g}$

Each cigarette (per day) reduces birth weight by 14.4 grams

28

Sexton and Hebel (1984), *J.A.M.A.*

- As an interesting aside, it's also possible to estimate the effect of continuing smoking (vs. quitting) from the data
  - First stage: $b_{FS} = -0.23$ (57% vs. 80% smokers)
  - Reduced form: $b_{RF} = 92$g
  - Structural: $b_{IV} = 92 / -0.23 = -400$g
- Women who kept smoking by the 8th month of pregnancy bore children who were 400 grams lighter, on average

29

Technical Conditions Required for Model Identification

- **Order condition** = At least the same # of IV's as endogenous X's
  - Just-identified model: # IV's = # X's
  - Overidentified model: # IV's > # X's
- **Rank condition** = At least one IV must be significant in the first-stage model
  - Number of linearly independent columns in a matrix
    - E(X | Z,W) cannot be perfectly correlated with E(X | W)

30

Instrumental Variables and Local Average Treatment Effects

- Definition of a L.A.T.E.
  - The average treatment effect for individuals "who can be induced to change [treatment] status by a change in the instrument"
    - Imbens and Angrist (1994, p. 470)
  - The average causal effect of X on Y for "compliers," as opposed to "always takers" or "never takers"
    - Not a particularly well-defined (sub)population
- L.A.T.E. is instrument-dependent, in contrast to the population A.T.E.

31

L.A.T.E. in the Previous Two Examples

- In the maternal smoking study...
  - For women who reduced their smoking frequency because they were assigned to the intervention, each one-cigarette reduction resulted in a 14-gram increase in birth weight (from mean 11 cigarettes)
- In the Galveston study..
  - For children whose parents may not evacuate because of…. (the IV)….
- In the Cancer study…

32

Bound, Jaeger, and Baker (1995), *J.A.S.A.*

- Potential problems with
  - Small Cov(X,Z) introduces finite-sample bias, which will be exacerbated with the inclusion of many IV's
  - Instrument may not satisfy "exclusion restriction"
    - Even small Cov(Z,e) will cause inconsistency, and this will be exacerbated when Cov(X,Z) is small
    - Are the ones we proposed before good instrument?

33

Bound, Jaeger, and Baker (1995), *J.A.S.A.*

- Even if the instrument is "good," matters can be made far worse with IV as opposed to LS
  - Weak correlation between IV and endogenous regressor can pose severe finite-sample bias
    - And…really large samples won't help, especially if there is even weak endogeneity between IV and error
- First-stage diagnostics provide a sense of how good an IV is in a given setting
  - F-test and partial-R^2 on IV's

34

Useful Diagnostic Tools for IV Models

- Tests of instrument relevance
  - Weak IV's → Large variance of b_{IV} as well as potentially severe finite-sample bias
- Tests of instrument exogeneity
  - Endogenous IV's → Inconsistency of b_{IV} that makes it no better (and probably worse) than b_{LS}
- Durbin-Wu-Hausman test
  - Endogeneity of the problem regressor(s)

35

Tests of Instrument Relevance

- Diagnostics based on the F-test for the joint significance of the IV's
  - Nelson and Startz (1990); Staiger and Stock (1997)
  - Bound, Jaeger, and Baker (1995)
- There is a growing econometric literature on the "weak instrument" problem
- **Rule of Thumb:** F-statistic against the null that the instruments are irrelevant in the first-stage regression should be larger than 10.

36

Tests of Instrument Exogeneity

- Model must be overidentified, i.e., more IV's than endogenous X's
  - H_0: All IV's uncorrelated with structural error
- Overidentification test:
  1. Estimate structural model
  2. Regress IV residuals on all exogenous variables
  3. Compute $N \cdot R^2$ and compare to chi-square
     - df = # IV's – # endogenous X's

37

Durbin-Wu-Hausman (DWH) Test

- If the so-called endogenous regressor is in fact exogenous, then the IV is still consistent
- Balances the consistency of IV against the efficiency of LS
  - H_0: IV and LS both consistent, but LS is efficient
  - H_1: Only IV is consistent
- DWH test for a single endogenous regressor:

  $\mathbf{DWH = (b_{IV} - b_{LS})^2 / v(b_{IV} - b_{LS})}$

  $\mathbf{With\ some\ assumption\ (b_{IV} - b_{LS})^2 / v(b_{IV}) - v(b_{LS})}$
  - If |DWH| > 1.96, then X is endogenous and IV is the preferred estimator despite its inefficiency

38

Durbin-Wu-Hausman (DWH) Test

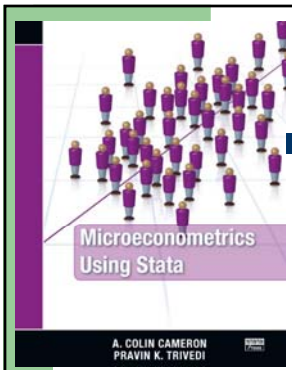
- A roughly equivalent procedure for DWH:
  1. Estimate the first-stage model
  2. Include the first-stage residual in the structural model along with the endogenous X
  3. Test for significance of the coefficient on residual

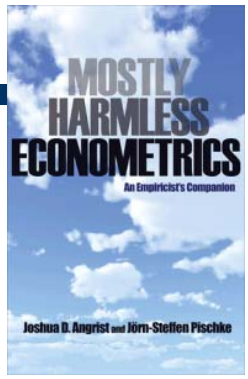
39

Software Considerations

- I have a strong preference for Stata
- For version 10 or above
  - Classic routine (*-ivregress-*)
  - Non-linear models: *-ivprobit-* and *-ivtobit-*
- Useful post-estimation routines
  - Overidentification: *- estat overid –*
  - Reporting first stage: *estat firststage*
  - Endogeneity of X in LS model: *-estat endogenous-*

40





41

Software Considerations

- Basic model specification in Stata

  ivregress 2sls y (x = z) w , *options*

  y = dependent variable
  x = endogenous variables (can have more than one here)
  z = instrumental variable (can have more than one)
  w = other control variable(s)
  - Useful options: *first*, *robust*, *cluster(varname)*
  - Overidentification: *-overid-*
  - Endogeneity of X in LS model: *-ivendog-*

42

Software Considerations

- For SAS users: Proc Syslin (SAS/ETS)
  - Basic command:

```
proc syslin data=dataset 2sls options1;
endogenous x;
instruments z w;
model y = x w / options2;
weight wtvar;
run;
```

  - Useful "options1": *first*
  - Useful "options2": *overid*

43

Software Considerations

- For SPSS users: 2SLS
  - Basic command:

```
2sls y with x w
/ instruments z w
/ constant.
```

  - For point-and-click aficionados
    - Analyze → Regression → Two-Stage Least Squares
    - DEPENDENT, EXPLANATORY, and INSTRUMENTAL

44

Software Considerations

- For Limdep users: 2SLS
  - Basic command:

```
2SLS ; Lhs = y
     ; Rhs = one, x, w
     ; Inst = one, z, w
     ; Wts = wtvar
     ; Dfc $
```

45

Stata Commands for the Foregoing Example

- Regression model ignoring endogeneity:

```
reg y x w
```

- First-stage regression model:

```
reg x z1 z2 w
```

  - With controls and multiple IV's, test relevance:

```
test z1 z2
```

- 2SLS regression model:

```
ivregress y (x = z1 z2) w
```

46

Stata Commands for the Foregoing Example

- Manual post hoc commands
  - Get residuals for regression-based overid. test:
    - After 2SLS model: **predict IVresid if e(sample), resid**
    - Then: **reg IVresid z1 z2**
  - Get residuals for regression-based DWH test:
    - After first-stage model: **predict FSresid if e(sample), resid**
    - Then: **reg y x w FSresid**
- "Canned" post hoc commands
  - After 2SLS model: **estat overid** and **estat endogenous**

47

Rules for Good Practice with Instrumental Variables Models

- IV models can be very informative, but it's your job to convince your audience
  - Show the first-stage model diagnostics
    - Even the most clever IV might not be sufficiently strongly related to X to be a useful source of identification
  - Report test(s) of overidentifying restrictions
    - An invalid IV is often worse than no IV at all
  - Report LS endogeneity (DWH) test

48

Rules for Good Practice with Instrumental Variables Models

- Most importantly, TELL A STORY about why a particular IV is a "good instrument"
- Something to consider when thinking about whether a particular IV is "good"

49

References

- Bound, Jaeger & Baker. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association, 90,* 443-450.
- Imbens & Angrist. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62,* 467-475.
- Kelejian. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variable. *Journal of the American Statistical Association, 66,* 373-374.
- Sexton & Hebel. (1984). A clinical trial of change in maternal smoking and its effect on birth weight. *Journal of the American Medical Association, 251,* 911-915.
- Staiger and Stock. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65,* 557-586.
- http://www.stata.com/support/faqs/stat/bias.html
- Craig C. Earle, Jerry S. Tsai, Richard D. Gelber, Milton C. Weinstein, Peter J. Neumann, Jane C. Weeks (2001). *Journal of Clinical Oncology*, Vol 19, Issue 4 (February), 2001: 1064-1070
- Temple, Jeff; Patricia van den Berg, Fred Thomas, James Northcutt, Dr. Christopher Thomas and Daniel Freeman Jr.July/August 2011 *American Journal of Disaster Medicine.*

50

Note: Instrumental Variables Models

- Finite-sample bias
  - In other words, $\mathbf{E(b_{IV}) \neq \beta}$ in small sample
  - The appeal of IV derives from its consistency
    - "Consistency" is a way of saying that $\mathbf{E(b) \rightarrow \beta}$ as $\mathbf{N \rightarrow \infty}$
    - So…IV studies often have very large samples
  - But with endogeneity, $\mathbf{E(b_{LS}) \neq \beta}$ and $\mathbf{plim(b_{LS}) \neq \beta}$ anyway
- Asymptotic behavior of IV

  $\mathbf{plim(b_{IV}) = \beta + Cov(Z,e) / Cov(Z,X)}$
  - If Z is truly exogenous, then $\mathbf{Cov(Z,e) = 0}$

51