CYBER SEMINAR HANDOUT

INSTRUMENTAL VARIABLES

Douglas O. Staiger, Ph.D.
Dartmouth College and the NBER

Thursday, March 28th, 2002
1:00 PM - 3:00 PM

Academy for Health Services Research and Health Policy

OUTLINE AND REFERENCES

(a `*' indicates useful background reading)

I. Introduction to Instrumental Variable (IV) methods

* Newhouse, J. and M. McClellan, "Econometrics in Outcomes Research: The Use of Instrumental Variables," *Annual Review of Public Health*, 19:17-34, 1998.

* Angrist, J. and A. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4):69-85, 2001.

II. Successful application of IV: estimating treatment effects

* McClellan, M., B. McNeil and J. Newhouse, "Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?" *JAMA*, 272(11):859-866, Sept. 1994.

McClellan, M. & H. Noguchi, "Treatment effect estimation with observational data: Validity and interpretation," manuscript, Stanford University, 1998.

III. Interpretation of IV: What does IV estimate?

Angrist, J., G. Imbens and D. Rubin, "Identification of Causal Effects Using Instrumental Variables," *JASA*, 91:444-455, 1996.

Gruber, J., P. Levine and D. Staiger, "Abortion Legalization and Child Living Circumstances: Who is the `Marginal Child'," *The Quarterly Journal of Economics*, 114(1):263-291, 1999.

IV. Problems with IV: Weak instruments and reporting error

Staiger, D. and J. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3):557-586, 1997.

Kane, T., C. Rouse and D. Staiger, "Estimating Returns to Schooling When Schooling is Misreported," National Bureau of Economic Research Working Paper #7235, 1999.

Geppert, McClellan and Staiger, "How Do Good Hospitals Do It? Estimating the Effects of Medical Practice," manuscript, Stanford University, 2000.

Problem: Estimate effect of treatment (T) on observed outcome (Y).
i.e., estimate β_1 in:

(1) $Y_i = \beta_0 + T_i\beta_1 + \varepsilon_i == X_i\beta + \varepsilon_i$ (X=[1 T])

For simplicity, suppose:

- dichotomous treatment variable: T=1 if treated, 0 otherwise
- homogeneous treatment effect (β)
- linear
- no covariates

Least-squares estimate of (1) yields standard "experimental" estimator:

$$\hat{\beta}_1^{OLS} = \overline{Y}_{T=1} - \overline{Y}_{T=0}$$

Key assumption of least-squares:

$$E(X'\varepsilon) = 0$$
$$\Rightarrow X'\left(Y - X\hat{\beta}_{OLS}\right) = 0$$
$$\Rightarrow \hat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

Unlikely to hold due to standard omitted variable problem:
Treatment related to relevant, but omitted variables (W).

i.e. suppose we partition ε such that $\varepsilon = W\delta + \nu$, where $E(X'\nu)=0$.
Then validity of OLS estimates of (1) require treatment to be uncorrelated with omitted variables:

E(T'W) = 0 ==> E(W|T=1) = E(W|T=0)

Four solutions to this problem:

1. Randomized Controlled Trial

   RCT is designed to ensure key OLS assumption: $E(T'\varepsilon)=E(T'W)=0$.

2. "Natural" Experiments

   Find similar observations with different treatment for "arbitrary" reasons (e.g. regulatory rules, law changes)

   - "Difference-in-Difference" estimates
   - Discontinuity design

3. Adjustment for Observable Differences

   Attempt to condition on sufficient W's s.t. $E(T'\nu)=0$
   → treatment is random/ignorable conditional on W

   Then estimate directly by least squares:

   (1') $Y = \beta_0 + T\beta_1 + W\delta + \nu$

   Variants on this approach include:
   - Matching, Case-Control
   - Regression
   - Fixed effects (sibling/person as own control)
   - propensity score

4. Instrumental Variables
(Includes 2SLS, LIML, many GMM, and Heckman selection)

Suppose exists variable (Z) that is:

- correlated with treatment: $E(Z'T) \neq 0$
- Uncorrelated with outcome, conditional on treatment: $E(Z'\varepsilon)=0$

Basis for IV estimator:

$$E(Z'\varepsilon) = 0$$
$$\Rightarrow Z'\left(Y - X\hat{\beta}_{IV}\right) = 0$$
$$\Rightarrow \hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$

If Z is dichotomous with no covariates, simple interpretation:

$$\hat{\beta}_{IV} = \frac{\overline{Y}_{Z=1} - \overline{Y}_{Z=0}}{\overline{T}_{Z=1} - \overline{T}_{Z=0}}$$

Note:

- Wald estimator
- Analogous to randomization
- Key assumption: E(W|Z=1) = E(W|Z=0)
- Estimate can "balance on the head of a pin!"

SUCCESSFUL APPLICATION OF IV:
ESTIMATING TREATMENT EFFECTS IN AMI

McClellan, M., B. McNeil and J. Newhouse, "Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?" *JAMA*, 272(11):859-866, Sept. 1994.

- Medicare claims data, elderly with heart attack (AMI), 1987-91
- Treatment: Cardiac Catheterization (marker for aggressive care)
- Outcome: Survival to 1 day, 30 days, 90 days, etc.
- Instrument: Is nearest hospital a catheterization hospital?

  Differential Distance =
  (distance to nearest cath) - (distance to nearest non-cath)

  based on zipcode of residence, zip code of hospital

Is this a good instrument?

1. Correlated with treatment (Cath)?

   26.2% get Cath if nearest is Cath hospital, 19.5% if not

2. Uncorrelated with patient severity that is *observable* in claims?

   Differential distance unrelated to age, comorbid disease.

- Major Results:

1. Least squares dramatically overstates treatment effect, because Cath associated with fewer risk factors.

2. IV estimates suggest Cath associated with 5-10 percentage point reduction in mortality; nearly all in 1st day.

3. Can successfully use administrative claims data to estimate effect of more aggressive treatment.

McClellan & Noguchi, "Treatment effect estimation with observational data: Validity and interpretation," manuscript, Stanford University, 1998. (Tables 1-2 below)

Geppert, McClellan and Staiger, "How Do Good Hospitals Do It? Estimating the Effects of Medical Practice," manuscript, Stanford University, 2000.
(Table 4 below)

Among other things, replicate & validate earlier work with more comprehensive control variables, alternative instruments:

-- Data from Cooperative Cardiovascular Project (CCP)
Chart data for appx. 180,000 AMI patients from 1994-95
Linked Medicare claims data

-- Treatments and outcomes of AMI in elderly as in earlier work

-- Instruments:
(1) Differential distance
(2) Variation in hospital Cath rate (>4000 dummies)

Key questions:

1. Are severity measures *unobserved in claims data* uncorrelated with instrument (differential distance)?

2. Are OLS results closer to IV with more extensive controls?

3. Are IV results robust to more extensive controls?

4. Are IV results robust to alternative instruments?

TABLE 1: DESCRIPTIVE STATISTICS
1994-95 ELDERLY AMI PATIENTS (CCP PROJECT)

RATE (%)	FULL COHORT (N=180,178)	NO CATH WITHIN 90 DAYS (N=96,323)	CATH WITHIN 90 DAYS (N=83,855)
PATIENT CHARACTERISTICS			
DEMOGRAPHIC VARIABLES			
Female	47.8	53.8	41.0
Black	6.7	7.4	5.8
Age in Years	75.1	78.2	71.6
(Standard Deviation)	(9.2)	(9.2)	(7.8)
Urban	72.1	71.1	73.2
CORMORBIDITY VARIABLES			
Mobility: Walks Independently	78.0	68.0	89.5
Mobility: Unknown Mobility	3.0	4.3	1.5
Dementia/Alzheimer Disease	6.1	10.4	1.2
Diabetes	31.4	33.1	29.5
CVA/Stroke	14.2	17.9	9.8
Angina/Chest Pain	46.8	42.3	51.9
CHF or Pumonary Edema	21.6	30.5	11.2
SEVERITY VARIABLES ON ADMISSION			
MI Confirmed by LDH, CPK-MB, or EKG	93.4	92.0	95.0
Verbal: Oriented/Converses	91.2	85.5	97.8
Heart Rate>100 (with imputations)	26.2	32.5	18.9
40<=Mean Arterial Pressure<80 (with imputations)	14.0	17.0	10.6
Time since chest pain started<=6 hours	49.8	41.8	58.9
Blood Urea Nitrogen>40 (with imputations)	8.0	12.2	3.1
SUMMARY COMORBIDITY AND SEVERITY			
Killip Class = 1	63.6	55.7	72.7
Killip Class = 2	9.3	9.0	9.7
Killip Class = 3	26.0	34.0	16.9
Killip Class = 4	1.0	1.3	0.6
30-Day Predicted Mortality	18.6	25.5	10.6
1-Year Predicted Mortality	32.1	42.8	19.8

TABLE 1: DESCRIPTIVE STATISTICS
1994-95 ELDERLY AMI PATIENTS (CCP PROJECT)

RATE (%)	FULL COHORT (N=180,178)	NO CATH WITHIN 90 DAYS (N=96,323)	CATH WITHIN 90 DAYS (N=83,855)
TREATMENTS BEFORE ADMISSION			
Beta Blocker at Arrival	18.9	17.4	20.7
CA++ Blocker at Arrival	34.8	34.3	35.5
Admit to Catheterization Hosp.	67.5	59.9	76.2
Admit to Revascularizton Hosp.	41.7	32.8	51.8
Admit to High-Volume Hosp.	69.5	64.8	75.0
TREATMENTS AFTER ADMISSION			
1-Day Catherization	12.7	0.0	27.4
90-Day Catherization	46.5	0.0	100.0
1-Day CABG	1.2	0.0	2.7
90-Day CABG	14.3	0.0	30.7
1-Day PTCA	5.6	0.0	12.0
90-Day PTCA	17.7	0.0	38.0
Heparin > 4000 U.	66.2	53.1	81.2
Thrombolytics After Arrival	16.5	10.0	24.0
ASA after Arrival	76.5	66.5	87.9
IV NTG After Arrival	50.3	39.0	63.4
Beta Blocker after Arrival	45.2	36.1	55.6
Antidepressant after Arrival	6.4	7.6	5.2
TREATMENTS AT DISCHARGE			
ACE Inhibitor at Discharge	23.6	25.3	21.6
ASA at Discharge	46.4	39.6	54.2
Beta Blocker at Discharge	26.3	22.1	31.1
CA++ Blocker at Discharge	25.3	23.6	27.3
Antidepressant at Discharge	4.2	5.0	3.2
OUTCOMES			
1-Day Mortality	5.7	10.0	0.9
7-Day Mortality	12.8	21.0	3.3
30-Day Mortality	18.8	29.3	6.8
1-Year Mortality	32.5	47.8	14.9
1-Year Total Hopital Days (days)	16.4	14.8	18.1
1-Year Total Expenditures ($)	16,611	11,651	22,277

TABLE 2: DESCRIPTIVE STATISTICS BY DIFFERENTIAL DISTANCE TO HOSPITALS

RATE (%)	DIFFERENTIAL DISTANCE TO CATH HOSPITAL		DIFFERENTIAL DISTANCE TO HIGH-VOLUME HOSPITAL	
	< = -0.5 mi (N=91,229)	> -0.5 mi (N=88,949)	< = -0.5 mi (N=88,568)	> -0.5 mi (N=91,610)
PATIENT CHARACTERISTICS				
DEMOGRAPHIC VARIABLES				
Female	48.2	47.5	48.4	47.3
Black	7.4	5.9	6.2	7.1
Age in Years	75.1	75.1	75.3	74.9
(Standard Deviation)	(9.1)	(9.3)	(9.1)	(9.4)
Urban	87.7	56.0	85.4	59.1
CORMORBIDITY VARIABLES				
Mobility: Walks Independently	78.0	78.0	78.2	77.8
Mobility: Unknown Mobility	3.0	3.0	2.8	3.2
Dementia/Alzheimer Disease	6.4	5.8	6.3	5.9
Diabetes	31.3	31.5	31.5	31.4
CVA/Stroke	14.4	13.9	14.2	14.1
Angina/Chest Pain	47.4	46.1	47.6	46.0
CHF or Pumonary Edema	21.5	21.6	21.9	21.3
SEVERITY VARIABLES ON ADMISSION				
MI Confirmed by LDH, CPK-MB, or EKG	93.7	93.1	93.6	93.2
Verbal: Oriented/Converses	91.1	91.2	91.2	91.2
Heart Rate>100 (with imputations)	26.7	25.7	26.9	25.5
40<=Mean Arterial Pressure<80 (with imputations)	14.1	13.9	14.1	14.0
Time since chest pain started<=6 hours	50.8	48.6	50.2	49.3
Blood Urea Nitrogen>40 (with imputations)	8.1	7.9	8.3	7.7
SUMMARY COMORBIDITY AND SEVERITY				
Killip Class = 1	63.1	64.1	62.7	64.5
Killip Class = 2	9.4	9.2	9.2	9.5
Killip Class = 3	26.4	25.6	27.1	25.0
Killip Class = 4	1.1	1.0	1.0	1.0
30-Day Predicted Mortality	18.4	18.7	18.6	18.6
1-Year Predicted Mortality	32.0	32.2	32.3	31.9

TABLE 2: DESCRIPTIVE STATISTICS BY DIFFERENTIAL DISTANCE TO HOSPITALS

RATE (%)	DIFFERENTIAL DISTANCE TO CATH HOSPITAL		DIFFERENTIAL DISTANCE TO HIGH-VOLUME HOSPITAL	
	< = -0.5 mi (N=91,229)	> -0.5 mi (N=88,949)	< = -0.5 mi (N=88,568)	> -0.5 mi (N=91,610)
PATIENT TREATMENTS AND OUTCOMES				
TREATMENTS BEFORE ADMISSION				
Beta Blocker at Arrival	19.5	18.3	20.1	17.7
CA++ Blocker at Arrival	35.5	34.2	35.6	34.1
Admit to Catheterization Hosp.	91.4	43.0	79.1	56.2
Admit to Revascularizton Hosp.	54.0	29.0	46.7	36.8
Admit to High-Volum Hosp.	81.9	56.9	91.0	48.7
TREATMENTS AFTER ADMISSION				
1-Day Catherization	14.8	10.6	13.0	12.5
90-Day Catherization	49.7	43.3	47.6	45.5
1-Day CABG	1.5	1.0	1.2	1.3
90-Day CABG	14.7	13.8	14.5	14.0
1-Day PTCA	6.6	4.5	5.7	5.4
90-Day PTCA	19.2	16.1	18.1	17.2
Heparin > 4000 U.	68.6	63.7	67.6	64.7
Thrombolytics After Arrival	17.3	15.7	16.7	16.3
ASA after Arrival	78.3	74.6	77.5	75.5
IV NTG After Arrival	52.9	47.7	52.1	48.6
Beta Blocker after Arrival	47.1	43.2	47.7	42.7
Antidepressant after Arrival	6.6	6.3	6.5	6.4
TREATMENTS AT DISCHARGE				
ACE Inhibitor at Discharge	24.4	22.9	24.2	23.1
ASA at Discharge	48.6	44.1	47.4	45.4
Beta Blocker at Discharge	27.9	24.7	28.2	24.5
CA++ Blocker at Discharge	26.8	23.9	26.1	24.6
Antidepressant at Discharge	4.4	3.9	4.3	4.0
OUTCOMES				
1-Day Mortality	5.5	6.0	5.3	6.1
7-Day Mortality	12.5	13.1	12.1	13.4
30-Day Mortality	18.6	19.1	18.3	19.3
1-Year Mortality	32.2	32.8	32.3	32.7
1-Year Total Hopital Days (days)	16.4	16.3	17.0	15.7
1-Year Total Hopital Expenditures ($)	16,804	16,396	17,281	15,928

Table 4

Estimates of effect of catheterization on 1-day, 30-day and 1-year mortality
Comparison of estimates from GMM, IV and OLS methods with and without detailed patient covariates
Based on AMI admissions from the CCP Project, 1994-95
(Standard errors of estimates in parentheses)

90-day catheterization effects	1-day mortality	30-day mortality	1-year mortality	1-day mortality	30-day mortality	1-year mortality
	detailed patient covariates			demographic controls only		
GMM	-4.7 (0.6)	-8.0 (1.0)	-12.0 (1.1)	-4.9 (0.7)	-8.0 (1.0)	-14.9 (1.1)
IV	-7.7 (2.1)	-11.3 (3.4)	-12.2 (3.9)	-8.4 (2.1)	-10.7 (3.5)	-10.9 (4.1)
OLS	-5.9 (0.1)	-12.8 (0.2)	-15.8 (0.2)	-9.0 (0.1)	-20.8 (0.2)	-28.7 (0.2)

GMM estimates are derived as described in the text. IV estimates are from two stage least squares, using differential distance to a catheterization hospital as the instruments. Sample includes 180,225 patients and 4005 hospitals.

Conclusions

1. Observed individual covariates can be used to assess bias of alternative methods for estimating treatment effects with observational data.

2. Methods that attempt to adjust for observable differences are quite sensitive to the use of more detailed chart data, and yield biased estimates of treatment effects in commonly available datasets.

3. IV methods for evaluating AMI treatment are not sensitive to the use of more detailed chart data, and appear to have minimal bias.

Many other applications of IV, using variety of instruments:

- Geography as an instrument
  (distance, rivers, small area variation)

- Legal/political institutions as an instrument
  (laws, election dynamics)

- Administrative rules as an instrument
  (wage/staffing rules, reimbursement rules, eligibility rules)

- Naturally occurring randomization
  (draft, birth date, lottery, roommate assignment, weather)

INTERPRETATION OF IV:
WHAT DOES IV ESTIMATE?

- So far, we have assumed that the treatment effect is homogeneous.

- What if the treatment effect is heterogeneous?
  - e.g. depends on patient severity
  - Then what does IV estimate? Effect for what population?
  - Angrist, Imbens and Ruben (1996) give careful answer.

- Think of RCT's
  - Criterion for inclusion in trial
  - Estimate treatment effect in well-defined population
  - Always issues of external validity (to general population)

- Analogous issues arise with IV estimates
  - Who are the "marginal" patients (or "compliers"), whose treatment is effected by the instrument?
  - IV estimates treatment effect among these "marginal" patients.
  - Often not estimating treatment effect in general population.

Example: McClellan, McNeil, Newhouse (1994)

Who are being treated?
Conceptually, the shaded areas below.

	Nearest to non-Cath Hospital	**Nearest to Cath Hospital**
	Less Effect of Treatment	
Never Treated →		
Compliers →		
Always Treated →		
	More Effect of Treatment	

So we would *expect*:

1. Patients with most to gain are always treated.

2. "Marginal" patients tend to have smaller effects of the treatment.

3. IV like RCT on patients thought to be least appropriate for treatment.

4. Average treatment effect among all those being treated may be higher.

Identifying "compliers"

Key issue in interpreting IV is identifying the "compliers", i.e. the margin on which your instrument is working.

In practice, three types of evidence are used for this purpose:

1. What range of variation in the endogenous variable is being generated?

    - Does instrument identify groups with wide or narrow range of treatment?

    - For continuous treatment (e.g. dose), does instrument have effects throughout distribution or in narrow range?

2. Does instrument have larger effect on odds-ratio of treatment among some sub-populations in the sample?

3. When instrument increases treatment rate, does it also affect the average characteristics among the treated?

    - If so, "compliers" were different from the "always treated."

    - Use this information to infer characteristics of marginal patient.

    - See Gruber, Levine and Staiger, "Abortion Legalization and Child Living Circumstances: Who is the `Marginal Child'," *The Quarterly Journal of Economics*, 114(1):263-291, 1999.

PROBLEMS WITH IV ESTIMATION:
WEAK INSTRUMENTS AND REPORTING ERROR

Outline

I. Testing key assumptions of IV:
   A1. Instruments are correlated with treatment
   A2. Instruments are uncorrelated with error

II. IV estimates when A1 fails: Weak instruments.

   See Staiger, D. and J. Stock, "Instrumental Variables Estimation with Weak Instruments," Econometrica, 1997.

   Application: Geppert, McClellan and Staiger, "How Do Good Hospitals Do It? Estimating the Effects of Medical Practice," manuscript, Stanford University, 2000.

III. IV estimates when A2 fails: Reporting error in treatment.

   See Kane, Rouse and Staiger, "Estimating Returns to Schooling when Schooling is Misreported," NBER wp #7235, 1999.

I. Testing Key Assumptions of IV

Suppose we have instrument(s) Z, and estimate by IV:

(1') $Y = \beta_0 + T\beta_1 + X\theta + \varepsilon$

The two key assumptions of IV are testable:

Assumption 1: The instruments are correlated with the treatment

Can be tested with "First-stage F-statistic" testing $\Pi=0$ from first stage regression:

$$T = X\alpha + Z\Pi + u$$

where small values of first-stage F imply failure of assumption 1.

Assumption 2: The instruments are uncorrelated with the error

Can be tested if over-identified (more instruments than treatments) using an auxiliary regression of $\hat{\varepsilon}_{IV}$ on X and Z: large values of N. R^2 imply failure of assumption 2 (under null A2, N. $R^2 \sim \chi^2(k)$ with k = #over-id restrictions).

WEAK INSTRUMENTS

In many applications of IV the instruments are weak, i.e. the first-stage F is small or not significant.

Usual asymptotic properties of IV (consistency, normality) assume that first-stage F is infinite and are poor guides when F is small.

Range of well-known examples from economics:

Angrist & Krueger, Quarterly Journal of Economics, 1991.
Use Census data: over 329,000 observations, up to 180 instruments.
First-stage F-statistic often under 10 and borderline significance.
Bound, Jaeger and Baker (JASA 1995), replicate results using random numbers as instruments!

Campbell and Mankiw, NBER Macroeconomics Annual , 1989.
Use annual US data, under 100 observations, 1-5 instruments.
First-stage F-statistic is always under 10, often insignificant.
Nelson and Startz, (Econometrica, 1990) monte carlos show that their estimates and standard errors are biased, with small sample distributions quite different from asymptotic approximations.

Staiger & Stock (1997) provide alternative asymptotic representations for IV estimator and test statistics which, loosely speaking, hold the first-stage F fixed asymptotically.

Key findings from these weak-instrument asymptotics include:

1. Can explain anomalies (bias, non-normal distributions) in well-known examples, as well as wide range of monte carlo evidence.

2. Properties of IV depend on only three parameters: the first-stage F-statistic (as defined above), the number of instruments, and the amount of bias in the OLS estimates.

3. With weak instruments:

- Two-Stage Least-Squares (2SLS) estimates biased toward OLS, with bias relative to OLS generally well approximated by 1/(first-stage F).

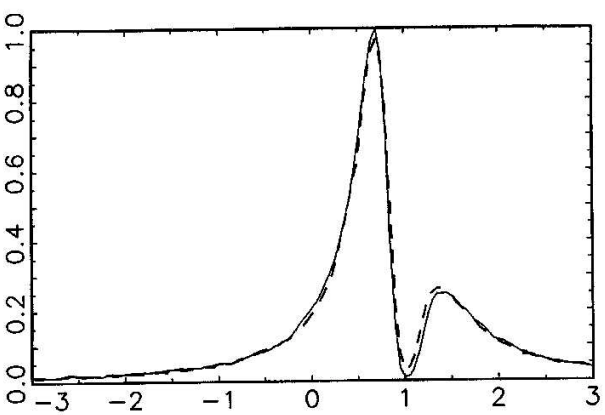
- 2SLS confidence intervals are too short, particularly with many instruments and/or a first-stage F under 10.

- Limited Information Maximum Likelihood (LIML) estimates are median unbiased, and confidence intervals are fairly accurate particularly when first-stage F can reject at 1% or greater level.

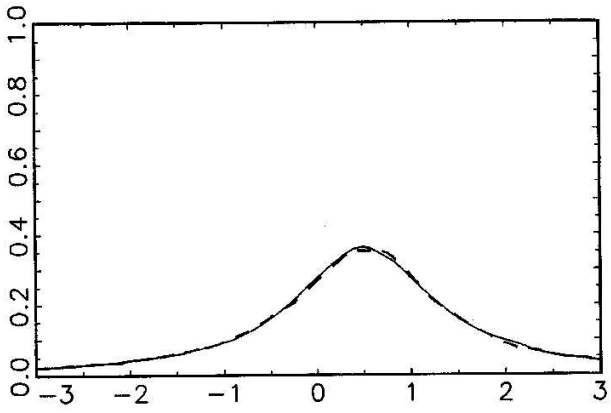
- Valid methods for constructing confidence intervals are discussed in Staiger & Stock.

- Other test-statistics are affected:
  1. Hausman-Wu tests tend to under-reject, Durbin form is valid.
  2. Over-id tests for 2SLS (LIML) tend to over-reject (under-reject).

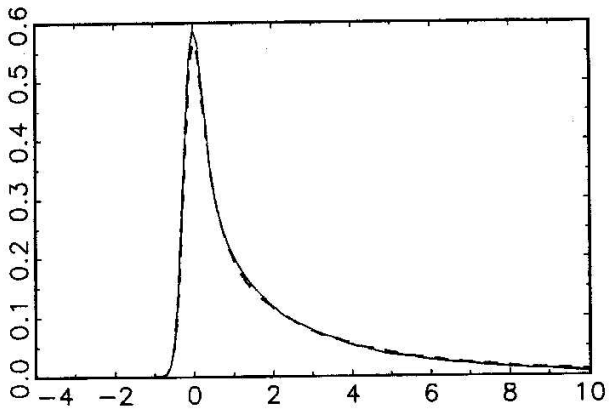
Figures from Staiger & Stock (1997)



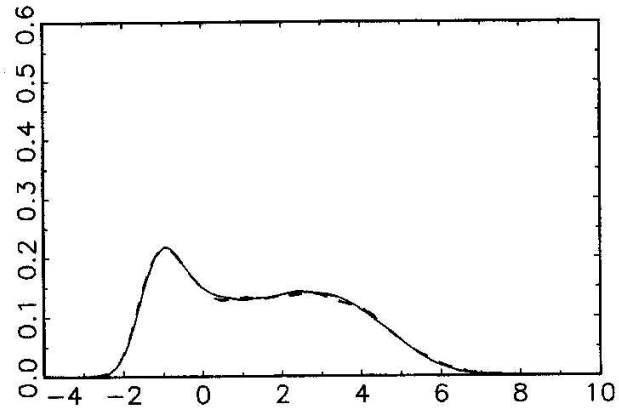
(a) $\hat{\beta}_{\text{LIML}}$ $(=\hat{\beta}_{\text{TSLS}})$: Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.



(b) $\hat{\beta}_{\text{LIML}}$: Design II, ρ=.5, K_2=100, $\lambda'\lambda/K_2$=0.



(c) t_{LIML} $(=t_{\text{TSLS}})$: Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.



(d) t_{LIML}: Design II, ρ=.5, K_2=100, $\lambda'\lambda/K_2$=0.

FIGURE 1—Asymptotic (solid line) and finite-sample (dashed line) pdf's of the LIML estimator and t statistic. True $\beta_0 = 0$; plim$(\hat{\beta}_{\text{OLS}}) = \rho$.

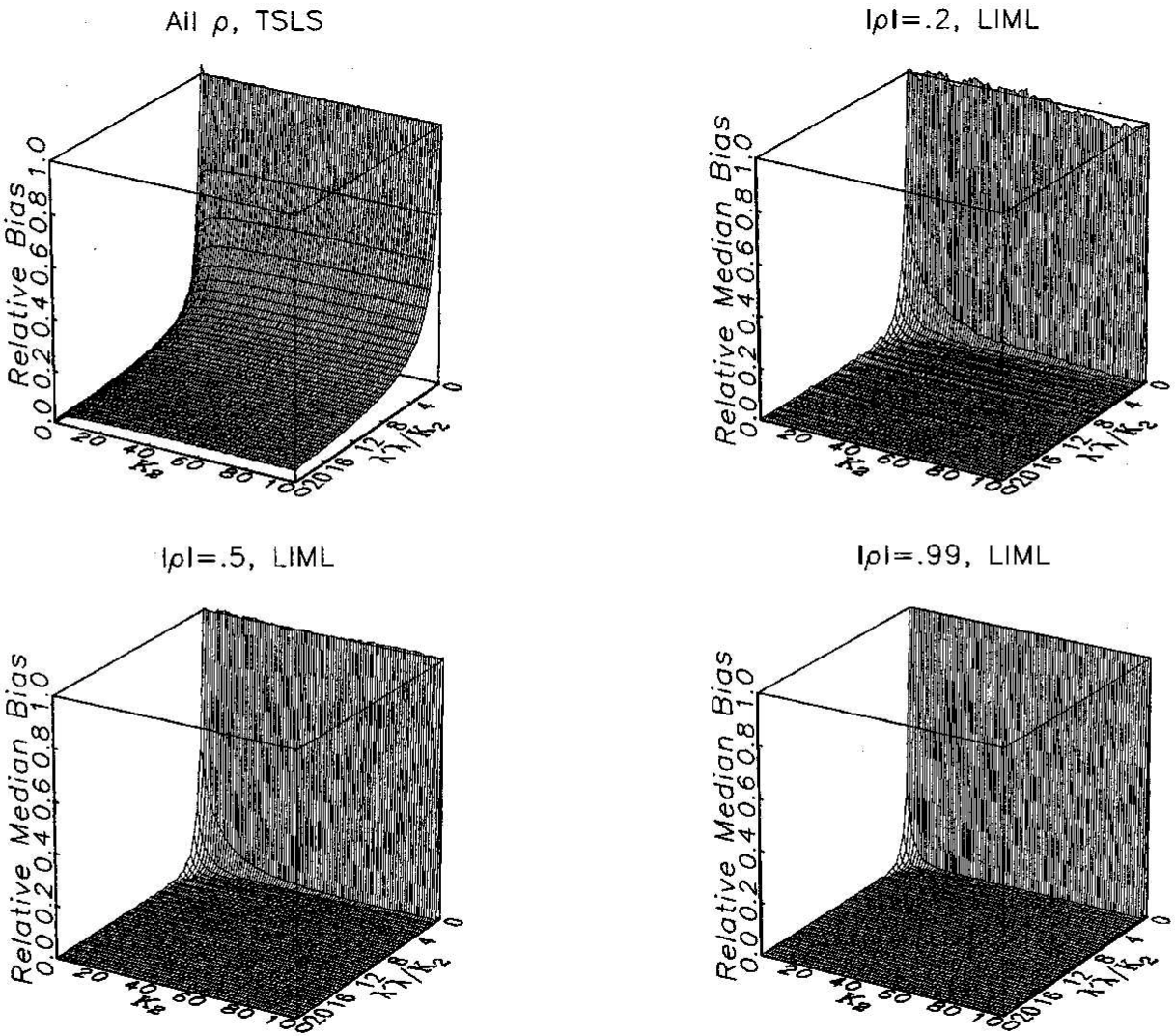


FIGURE 2.—Asymptotic bias of $\hat{\beta}_{\text{TSLS}}$ and asymptotic median bias of $\hat{\beta}_{\text{LIML}}$, as a fraction of the bias of $\hat{\beta}_{\text{OLS}}$.

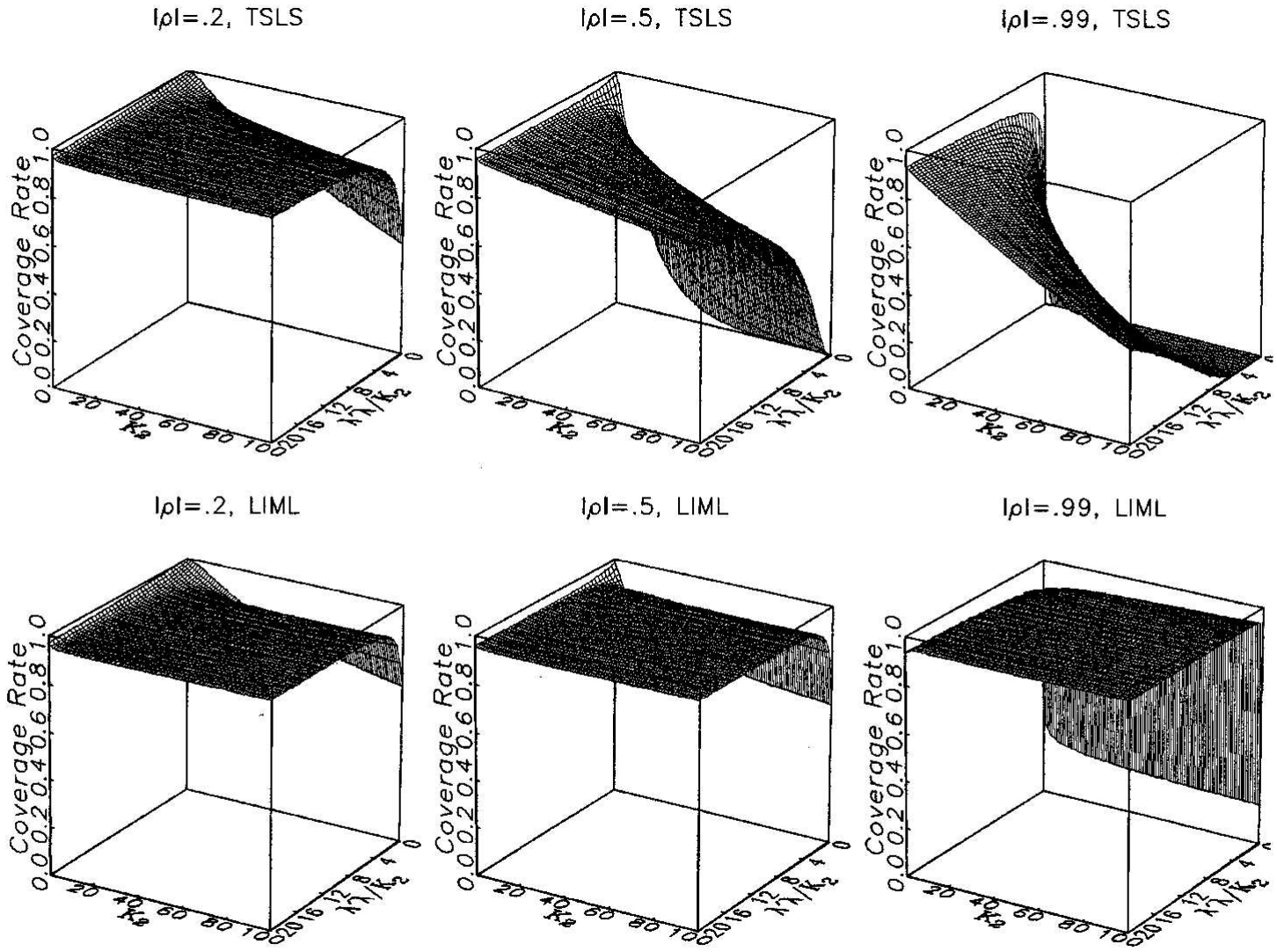


FIGURE 3.—Asymptotic coverage rates of conventional 95% TSLS and LIML confidence intervals.

Angrist/Krueger example

Data: US Census 5% PUMS

Sample: 329,509 men born 1930-39

$\ln(\text{earnings}) = \text{Education}*\beta + X\theta + e$

Instruments: (Quarter of Birth)*(Year of birth)
(Quarter of Birth)*(State of birth)

Instruments: 178

First-stage F: 1.869
(p-value) (0.000)

	Estimate of β	*95% Confidence interval*
OLS	0.063	(0.062,0.063)
2SLS	0.081	(0.060,0.102)
2SLS with random instruments	0.060	(0.031,0.089)
LIML	0.098	(0.068, 0.128)
Valid Confidence interval (Anderson-Rubin)		(-0.015,0.240)

Geppert, McClellan and Staiger (2000)

1. Use between-hospital variation in treatment intensity (e.g. cath rate) as instrument to estimate treatment effects

   Equivalent to using >4000 hospital dummies as instruments

2. But instruments are weak: 1st Stage F-statistic is 10-25

   → 2SLS estimates have small bias (1/F) towards OLS

   → 2SLS SEs are too small (many instruments, modest F)

3. Using hierarchical structure, we develop alternative GMM estimation procedure to correct estimates & SEs.

REPORTING ERROR IN TREATMENT

IV estimates are consistent in the presence of classical measurement error in the treatment variable (e.g. mean zero, independent error).

However, measurement error cannot be classical in a dichotomous treatment variable → error must be negatively correlated with treatment.

As a result, if there is reporting error in treatment:

- IV estimates of treatment effects are biased
- IV estimates tend to overstate magnitude of the treatment effect

Why does reporting error bias IV estimates?

Consider the problem of estimating a treatment effect:

(1) $Y = \beta_0 + T\beta_1 + e$

with a valid instrument Z: $E(Z'T)\neq 0$, $E(Z'\varepsilon)=0$
(e.g. Z could be assignment in a RCT)

What if we observe the treatment with error? i.e.:

T^0 = observed treatment variable
$\alpha_1 = pr(T^0=1|T=0)$, $\alpha_2 = pr(T^0=0|T=1)$
So that α_1 and α_2 are the error rates in the observed variable.

More compactly: $T^0 = \alpha_1 + (1- \alpha_1 - \alpha_2)T + \nu$

We can rewrite (1) as:

(1') $Y = \beta_0 + T^0\beta_1 + u,$

where $u = \varepsilon + \beta_1 (T-T^0)$
$= \varepsilon + \beta_1 (\alpha_1 + \alpha_2)T$ + other terms

IV estimates based on the observed treatment are now invalid because $E(Z'u)\neq 0$ (since T is in u, and $E(Z' T)\neq 0$)

In this case, 2SLS overstates treatment effect in proportion to the sum of the error rates:

$$\hat{\beta}_1^{2SLS} \Rightarrow {E(Z'Y)}\Big/{E\left(Z'T^0\right)}$$

$$\Rightarrow {E(Z'Y)}\Big/{\left[\left(1-\alpha_1-\alpha_2\right)E(Z'T)\right]}$$

$$\Rightarrow {\beta_1}\Big/{\left[1-\alpha_1-\alpha_2\right]}$$

Intuition from Wald estimator (when Z dichotomous):

$$\hat{\beta}_1^{IV} = \frac{\overline{Y}_{Z=1} - \overline{Y}_{Z=0}}{\overline{T}_{Z=1}^0 - \overline{T}_{Z=0}^0}$$

We correctly estimate the difference in the outcomes in the numerator, but understate the difference in the treatments in the denominator.

What can be done?

1. Knowledge of error rates (e.g. validation study) useful in evaluating magnitude of bias and correcting bias.

2. Standard IV specification tests (e.g. over-id test) have no power to detect this problem.

3. Can construct consistent estimates of treatment effect and error rates if have two reports of treatment with independent reporting errors (e.g. report from patient and from health care provider). For details, see Kane, Rouse and Staiger (1999).

Conclusions

1. IV is useful, practical alternative to Randomized Controlled Trials.
    - Many successful examples
    - Increasingly wide range of applications

2. Like RCT's, external validity depends on the population being studied.
    - What population's treatment is being affected by instrument?
    - Effect in marginal population may differ from average effect.

3. Key to success is finding *good instruments*!
    - Must carefully evaluate key assumptions of IV in any application:
        1. Instrument is correlated with treatment
        2. Instrument is not correlated with error term
    - Best instruments work like RCT's: Run a "natural" experiment.