# Instrumental Variables: A simpleminded introduction

Bernard Black
Univ of Texas, Law School &
McCombs School of Business
CELS 2007

# Overview

- Why Use Instruments
- How to Find Them
- How to Use Them
- How to Misuse Them (With Examples)
- Warning:  I'm an improperly trained cowboy econometrician.

# 1. Standard Regression Setup

- Assume familiarity with basic multivariable regression:

$$y = a + \mathbf{b} * \mathbf{x} + \varepsilon \qquad (1)$$

- y = dependent variable
- x = independent variable(s)
- (**bold** → could be many x's)
- ε = "error" (part of y that is not predicted by the regression equation)

# Regression Coefficients: Association, not Causation

- Standard understanding:
  - Equation (1) shows the unobserved "true" relationship in an overall “population”
  - We "estimate" the coefficients a, and **b** by running a regression on an observed "sample", assumed to be drawn at random from this population
- Regression coefficients **b** show the correlation of
  - Partial correlation, holding constant other x’s
  - **Not** causation
  - We say x "predicts" y, or x "is associated with" y
- Often want to know: Is the effect causal?

# Example 1: Board composition and firm value

From Bhagat & Black (2002)

| | Measured in early 1991 | | |
|---|---|---|---|
| **Dependent variable** | Board independence | Board size | No of 5% outside holders |
| *mean Tobin's q, year ends* 1991-1993 | **-.22 (t=2.09)**** | -.018 (t=1.81)* | **-.067 (t=2.92)**** |

Higher independence predicts lower Tobin's q over next 3 years.

Can we tell a noncausal story?

# Example 2: Defense cost and litigation payout

From Black, Hyman et al. (2007)

| **Dependent variable** | year | ln(payout) | dummy (trial started) |
|---|---|---|---|
| *ln(defense cost)* | **0.051 (t=31.52)*** ** | **0.341 (t=44.35)*** ** | **0.578 (t=14.00)*** ** |

Higher payout predicts higher defense cost.
Can we tell a noncausal story?
What if we flipped dependent and independent variables?

# Example 3: Death penalty and murder rates

From Dehzbakhsh & Shepherd (2004), as discussed in Donohue & Wolfers (2005)

| | death penalty dummy | |
|---|---|---|
| **Dependent variable** | original paper | Donohue-Wolters replication |
| *murder rate* | **-0.87 (σ=0.21)*** ** | -0.47 (σ=0.74) |

Suppose death penalty did predict lower (or higher) murder rate? Can we tell a noncausal story?

# Why Use Instruments: Helps to Get at Causation

- Refine basic multivariable regression:

$$y = a + b * x + \mathbf{c} * \mathbf{z} + \varepsilon \qquad (2)$$

- y = dependent variable
- x – possibly "endogenous" "independent" variable
- **z** = other independent variable(s), assumed *exogenous*
- ε = error

# Scope of Talk

- Assume:
  - one possibly endogenous variable x
  - (mostly) one possible instrument w
- econometricians call this "just identified"
  - this means number of w's = number of x's
  - "overidentified" ➔ more w's than x's
  - "identification" = "we can infer something about causation"
- BUT having instrument(s) does **not** mean you have good identification
  - that takes a **valid** instrument . . .

# Walk before you run: Common errors in empirical papers

- I'll also assume:
  - you know about heteroskedasticity consistent standard errors ("robust" option in Stata)
  - you know how (and maybe even when) to use clusters ("cluster" option in Stata)
  - if you have "panel" data, you know about fixed and random effects ("xtreg" in Stata, with fe or re option), and time dummies
- You know your data really well
  - checked for outliers
  - checked for high correlation among indep. variables
  - summary statistics make sense

# What does "endogeneity" mean

- Could mean y causes x (reverse causation)
  - low Tobin's q ➔ high board independence
  - murder rate ➔ death penalty
- Could mean an unobserved, "omitted variable" predicts both x and y
  - firm characteristics ➔ both Tobin's q and board structure
  - expected damages ➔ payout, defense cost
  - state culture ➔ murder rate, death penalty

# How Does An Instrument Help?

- Core idea: replace possibly endogenous x

- If (i) we have a valid instrument and (ii) instrumented-x predicts y, then
  - this is evidence that x causes y
  - one calls this "identification"

# Required for Valid Instrument

- A valid instrument w should be:
- **exogenous** (comes from "outside" the regression system): y can't cause w
- **correlated** with x (hopefully strongly)
- predicts y **only** through x
  - not directly -- not (partially) correlated with y
  - not indirectly through another variable:
    - whether included in the z's or not!
  - formally: w must be uncorrelated with ε in the population

# Bad instrument: different proxy for board independence

- Bhagat & Black (2002) use INDEP =

  inside directors as their main measure
- in a "three stage least squares" analysis, then use (fraction of indep directors) as an instrument for INDEP
- What's wrong with this?

# Better instrument: Large firm dummy in Korea

- Black, Jang & Kim (2006) have a better instrument:

  Korea, in 1999, requires 50% outside directors & audit committee for large firms (> 2 trillion won), not for smaller firms

  – Idea: use large firm dummy to instrument for combined effect of both rules

  – Does this work?

# Two Stage Least Squares (2SLS)

| | 1st stage | 2nd stage |
|---|---|---|
| | Board Structure Subindex | Tobin's *q* |
| Instrumented Board Structure Index | | **0.0112*** ** |
| | | **(3.44)** |
| Large firm dummy | **11.86*** ** | |
| | **(14.01)** | |
| *Ln*(assets) | 0.157 | **-0.041*** ** |
| | (0.78) | **(3.45)** |
| KCGI - Board Structure Index | **0.043** ** | **0.006*** ** |
| | **(2.41)** | **(4.66)** |
| Other Control Variables | yes | yes |
| Adjusted $R^2$ | 0.771 | 0.324 |

# What if we omit the rest of KCGI

| | 1st stage | 2nd stage |
|---|---|---|
| | Board Structure Subindex | Tobin's *q* |
| Instrumented Board Structure Index | | **0.0133** |
| | | **(3.86)*** ** |
| Large firm dummy | **12.05** | |
| | **(14.44)*** ** | |
| *Ln*(assets) | | **-0.035** |
| | | **(2.85)*** ** |
| KCGI - Board Structure Iindex | omitted | omitted |
| Other Control Variables | yes | yes |

1. Effect of moving firm from small to large: 12.05 * .0133 = 0.160
2. Higher 2d stage coeff: .0133 vs. .0112in prior slide. Why?

# Instrument instead for Board Procedure?

| | 1st stage | 2nd stage |
|---|---|---|
| | Board *Procedure* Subindex | Tobin's *q* |
| Instrumented Board *Procedure* Index | | **0.0756** |
| | | **(3.86)*** ** |
| Large firm dummy | **2.12** | |
| | **(5.60)*** ** | |
| *Ln*(assets) | | **-0.035** |
| | | **(2.85)*** ** |
| KCGI - Board Structure Index | omitted | omitted |
| Other Control Variables | yes | yes |

1. Effect of move from small to large: 2.12 * .0756 = 0.160
2. Higher 2d stage coeff: .0756 vs. .0133. Why?
3. Identical t-stat in second stage. Why?

# What's wrong here?

- Large firm dummy is still exogenous and correlated with instrumented variable
- It predicts Tobin's q *partly* through Board Procedure, but not **only**
  - Predicts Tobin's q mostly through an omitted variable, Board Structure
  - 2SLS procedure forces instrument to affect Tobin's q only through Board Procedure ➔ coeff. too large
  - *t*-stat is the same, because it measures the power of the ***instrument*** to predict Tobin's *q*
    - **not** the power of the instrumented variable

# Can we test for a valid instrument?

- Sadly, no.
- Can test for some obvious problems
  - does instrument directly predict y, controlling for instrumented variable
- Sometimes can conduct out of sample test
  - Korea study, large firm dummy should ***not*** predict Tobin's q in 1998
- But for the subtler ones, no good test
  - **If** valid instrument, can test for reverse causation (Hausman test)
  - If have two instruments for one variable, can test each for endogeneity by assuming the other is a good instrument (Hansen test)
    - If multiple valid instruments, coefficients should not depend (much) on which subset we use
- In the end, we're often stuck with logic on what will probably be a good instrument

# Instrumenting for payout

- Return to defense costs paper.
- Payout and defense cost are endogenous.
- What can we do?
- Possible IV: Basic claimant characteristics (age, employed).
  - *Could* predict defense cost only through effect on expected damages, not directly

| **Dep. variable** | ln(payout) |
|---|---|
| *ln(def. cost)* | **0.341 (t=44.35)** |

# IVs: *ln*(age+1), baby dummy

| | OLS | 1st Stage | 2nd Stage |
|---|---|---|---|
| **Dependent Variable** | **ln(def. cost)** | **ln(payout)** | **ln(def. cost)** |
| **ln(payout)** | **0.341** | | **0.712** |
| | **(44.35)*** ** | | **(16.02)*** ** |
| **ln (policy limits)** | **0.043** | **0.195** | **-0.025** |
| | **(5.50)*** ** | **(21.92)*** ** | **(2.11)** ** |
| **Dummy trial started** | **0.578** | **0.326** | **0.460** |
| | **(14.00)*** ** | **(5.71)*** ** | **(9.51)*** ** |
| **multi-defendant dummy** | **0.105** | **0.283** | -0.018 |
| | **(5.04)*** ** | **(11.81)*** ** | (0.66) |
| **ln(time open)** | **0.705** | **-0.084** | **0.725** |
| | **(39.49)*** ** | **(5.21)*** ** | **(39.06)*** ** |
| **IV: ln(age+1)** | | **-0.094** | |
| | | **(6.86)*** ** | |
| **IV: baby dummy** | | **0.228** | |
| | | **(3.82)*** ** | |
| **Adj. $R^2$** | 0.525 | 0.160 | 0.427 |
| **Hansen Overident. Test** | | chi-2=0.248 (p = 0.62**)** | |

# IVs for death penalty?

- Dezhbakhsh, Rubin & Shepherd (ALER, 2003) use the following instruments (borrowed from John Lott's work on concealed handgun laws):
  - police spending
  - judicial spending
  - prison admission rate
  - Republican vote share in Presidential elections (1976-1996)
- What might be wrong with these?
- Is an out of sample test available?

# Have instrument, will travel

- One set of instruments, many papers
  - Shepherd uses same instruments for 3-strike laws (J Leg Stud, 2002), truth-in-sentencing legislation(J L Econ, 2002), and sentencing guidelines (work pap 2004)
  - Rubin & Dezhbakhsh (Int Rev L Econ, 2003) use these instruments to study concealed gun laws (following Lott).
- What ***must*** be wrong with that?

# Local Average Treatment Effect

- IV result applies only to the "treatment group"
  - Korea paper, large and near-large firms
- Earnings = a + b*(yrs of college) + **c*z** + ε
  - Observe:  b positive
    - Can we tell a noncausal story?
  - IV: miles to nearest college
    - Is this a valid instrument?
  - If so, for whom?

# Lots of papers have bad instruments
# Some warning signs

- Glides over why the instrument is valid
- First-stage results not shown
- Instrument might predict y directly
- Instrument might predict y through omitted variable
  - one instrument, many uses is a clue
- Second stage coefficients too large
  - Much larger than OLS

# Some thought experiments

Waldman and autism

- Theory:  TV watching causes autism in sensitive subpopulation
  - basis – observed own child
- To test, need exogenous variable, which predicts TV watching, but is unrelated to other factors which might predict autism
- IV:  rain days in Pacific Northwest
  - more TV watching in rainier years
  - find more autism among 2-year-olds in rainier years
- Is rain a valid instrument for TV watching?

# Class size

- test score = a + b*(class size) + d*(school

- Observe, b negative
  - Can we tell a noncausal story?
- IV (Angrist & Lavy): Maimonides rule in Israel: class size ≤ 40
  - "regression discontinuity" approach