Interpreting Did and IV Estimates: ATE, ATET, LATE, and All That

Bernard Black
Univ of Texas (through 2010)
Northwestern (from 09.01.2010)
Law School & Kellogg School of Management

CELS, 2009
Main goal

• Assume: You have a (reasonably) **valid** instrumental variable (IV) or difference-in-differences (DiD) study design
  – How to achieve that: different talk(s)

• This provides evidence of causation for a **subpopulation**

• Goal here: What can you learn from the study
  – Hint: Less than you might think
Proceed by Example

• See what we learn from:
  – One of my wife’s studies
  – One of my own studies
    • Black, Jang & Kim, *Does Corporate Governance Affect Firms' Market Values? Evidence from Korea* (JLEO 2006)
  – Two classic IV studies
    • Angrist & Kreuger, *Does Compulsory School Attendance Affect Schooling and Earnings?* (QJE 1991)
Simple setup (0-1 “treatment”)

• Treatment is “on” \( (D_i = 1) \) or “off” \( (D_i = 0) \)
  – \( i \) indexes the subjects of the treatment: people or firms or states

• We are interested in how treatment affects outcome \( y_i \) for subject \( i \)
  \[ [y_i | D_i = 1] - [y_i | D_i = 0] \]
  Or, more compactly: \( y_{1i} - y_{0i} \)

• Any subject is treated or not

• So we observe one of these, but not the other
  – Regression analysis lets us estimate the unobserved outcome
Regression design

• OLS: \[ y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon \]
  \( X = \) set of control variables
  \( \beta \) is estimate of \( E[y_{1i} - y_{0i}] \)

• This is association, not causation. Could have:
  - reverse causation: \( y \) causes \( D \)
  - omitted variable: both \( y \) and \( D \) are associated with an unobserved variable \( w \)
    • if so, regression estimate \( \beta \) is biased

• IV and DiD provide ways to get unbiased estimate
Litvak (2007) research design

- “natural” experiment: SOX applies to all US firms (so no within-US control group), and to some foreign firms
  - SOX applies, if firm is cross-listed in US on level 2 or 3
  - Does not apply if firm is cross-listed on level 1 and 4
  - (Obviously) does not apply if firm is not cross-listed

- DiDiD design:
  - stock price return during SOX adoption events in 2002 (first difference)
  - “pair return”: return to level-23 firm relative to propensity matched non-cross-listed firm from same country (second “pair” difference)
    - n = 431 (this will become important later)
  - to level-23 pairs relative to level-14 pairs (third difference)
Litvak main result

• Negative return to level-23 pairs versus level-14 pairs: $\Sigma$(SOX events) = -5.8% ($t = 5.48$)

• Assume investors are right in predicting effect of SOX on firm value
  – apparently valid through 2005 (see Litvak 2008).

• What can we learn (and not learn)?
Logic of research design

• Narrow event study windows ➔ stock price return more likely due to SOX
• **Pair** return controls for unobserved developments in home country
• Difference in level-23 and level-14 pair returns controls for unobserved US developments that affect all cross-listed firms
• But:
  – limited control variables
    • limited data availability, especially for non-cross-listed firms
  – must assume: no relevant unobserved difference between level-23 and level-14 pairs
  – usual event study assumptions (smart investors, efficient markets)
What inferences can we draw?

• All level-23 firms were “treated” by SOX
• Mean -5.8% difference in pair returns is estimate of **Average Treatment Effect (ATE)**
• In symbols: \( \text{ATE} = E[y_{1i} - y_{0i}] \)
  
  - \( E \) = expectation operator
  - \( y_{1i} \) = value of firm \( i \) if treated by SOX (\( D_i = 1 \))
  - \( y_{0i} \) = value of firm \( i \) if **not** treated by SOX (\( D_i = 0 \))

  **not** observed; estimated using the DiDiD procedure

• US firms were treated too . . .
  – Can we estimate a -5.8% reaction to SOX for these firms?
ATE for whom?

• On the **treated population**
  – Litvak studies only level-23 foreign firms
  – no assurance that **observed** ATE on level-23 firms
    = **unobserved** treatment effect on US firms
  – Compare Zhang (2008) (event study of US firm reactions, using non-US firms as control)
    • advantage: measures ATE for US firms
    • concern: weaker control group ➔ unobserved differences between treatment and control groups
ATET (Average Treatment Effect on the Treated)

• ATET = ATE on the treated
• Formally: \( \text{ATET} = E[(y_{1i} - y_{0i}) | D_i = 1] \)
• If whole population is treated, then ATE = ATET
  
  — For Litvak’s study, ATE = ATET
  — But ATET\(\text{T}\) terminology is still useful: suggests need to focus on who is treated
How about other foreign firms?

• Level-14 firms were not treated
  – Technical answer: Strictly speaking, we don’t know ATE
  – Practical question: Are they similar enough to level-23 firms to infer SOX impact, *if they had been treated*?
    • No clear answer, must study how they differ.

• Non-cross-listed matched firms?
  – Matching procedure ensures they are similar to level-23 firms in propensity to cross-list
  – Does that imply similar reaction to SOX?

• Other non-cross listed firms?
  – Less similar to level-23 firms ➔ less likely that one can infer hypothetical effect if treated
Black, Jang and Kim

• IV study (2006); DiD study (2009, with Park)
• Large Korean firms (assets > 2 trillion won) are subject to legal shock in 1999. Must have:
  – 50% outside directors
  – audit committee, majority of outside directors
  – outside director nomination committee
• Small firms are not treated, form control group
• Core result: Large firm Tobin’s $q$ rises by 0.16 ($t = 3.86$) relative to small firms
Korea Study: ATE and ATET

• All large firms are treated:
  • $ATE = ATET = 0.16$ increase in Tobin’s $q$

• What can we infer about value of these governance reforms for small firms?
  – Technically, nothing (they are not treated)
  – But if similar to large firms, could see similar effect
    • mid-sized (say 0.5-2T won): likely similar effects
    • smaller firms: less clear
  – BJK report similar coefficients for large firms and for small firms who are voluntary adopters
Angrist & Krueger on school dropouts

- Core idea: dropouts allowed at age 16 (say)
- Some students are closer to graduation when they turn 16 (depends on birth date)
  - Assume start K if age 5 by Jan. 1 of next year
  - Each K class includes ages 4.08-5.07 (year.months)
    - 4.08 kids will be almost halfway through grade 11 when they turn 16
    - 5.07 kids will turn 16 about halfway through grade 10
- Predict (and observe): more schooling/fewer dropouts if younger when start K
Birthdate as instrument for schooling

• Classic problem in labor economics: impact of schooling on future income

• Want to estimate:

\[
\text{income} = \alpha + \beta \cdot (\text{years of school}) + \gamma \cdot X + \delta \cdot \text{ability} + \epsilon
\]

• But can’t observe ability, so actually estimate:

\[
\text{income} = \alpha + \beta \cdot (\text{years of school}) + \gamma \cdot X + \epsilon
\]

  – omitted variable problem
  – estimate of \( \beta \) biased \text{upward} if ability \( \rightarrow \) schooling

• need instrument for schooling, uncorrelated with ability
  – birthdate, relative to K cutoff date, is plausible instrument
More complex setup (still 0-1 variables)

- A treatment is “on” ($D_i = 1$) or “off” ($D_i = 0$)
  - *Kids enter K younger* ($D_i = 1$) or older ($D_i = 0$)
- *Treatment affects outcomes:* *Younger kids stay in school longer, on average*
  - $E[\text{school years}_{1i}] > E[\text{school years}_{0i}]$
- Treatment never causes reverse response
  - $\text{school years}_{1i} \geq \text{school years}_{0i}$ for all $i$
  - *No one gets less school because entered K younger*
Instrument for whom?

• Put aside “weak instrument” problem with their study (not known at the time):
  – Estimated 4-6% higher earnings per year of school
    • Not far from OLS estimates with good controls

• Who is affected by the “treatment”?
  – not your kids or mine, who will graduate HS anyway
    – Let $G_i = 1$ if graduate; 0 otherwise
    – Observe: [school years$_{i} | G_i = 1$] is independent of $D_i$
  – instead marginal kids, who would have dropped out if they were older at entry ($G_i = 0 | D_i = 0$) $\approx G_{0i} = 0$
    • some graduate if $D_i = 1$; some still don’t
Don’t know ATE or ATET

• No reason to believe value of schooling is same for both groups
• → don’t know ATE = average effect on all kids of being “treated” (being young when enter K)
  – don’t know ATET either (average treatment effect on kids who were young when enter K)
• What do we know?
LATE

• We know the **local** average treatment effect (LATE) – ATE for the marginal kids
• Formally: We observe, through the 2SLS procedure:

\[
\text{LATE} = \mathbb{E}[(y_{1i} - y_{0i}) | G_{0i} = 0]
\]

  \[G_{Di} = 1 \text{ if graduate; } 0 \text{ if drop out}\]

  \[y_{1i} = \text{earnings if treated (if young when enter K)}\]

  \[y_{0i} = \text{earnings if not treated (older when enter K)}\]

• Can think of this as ATE on the **effectively treated** (subpopulation of treated for whom treatment matters)
Is This Useful to Know?

• Not if your policy concern is the average return to extra education for all kids.
• But suppose you are a state legislator, considering changing the minimum school leaving age, or the age at school entry?
  – You might care about effect on the marginal kids
McClellan et al. on AMI treatment

• Observe:
  – **huge** difference in AMI (heart attack) survival for who receive cardiac catheterization ($D_i = 1$), and those who don’t
    • 37% higher 4-year survival (67% vs. 30%)

• But also observe:
  – patients who receive catheterization are younger and healthier than those who don’t ➔ better outcomes in any case

• Survival difference shrinks if control for observable health markers, still large
  – could reflect unobserved health differences
  – could reflect unobserved treatment differences
    • cath-equipped hospitals might be better at non-cath treatment
• Instrument: Distance to cath hospital

• Observe: only major hospitals perform cath
• AMI victims usually taken to nearest hospital
  – some later transfer to cath hospital, some don’t
  – let $b_i = 1$ if extra distance from home to cath hospital, versus nearest hospital < 2.5 miles
  – $E[D_i|b_i = 1] > E[D_i|b_i = 0]$
  – 26.2% vs. 19.5% (within 90 days)

• Observable health markers for AMI victims are independent of $b_i$

• Use $b_i$ as instrument for $D_i$
McClellan et al. results

• 4-year survival:
  \[ E[y_i|b_i=1] - E[y_i|b_i=0] = 58.5\% - 58.1\% = 0.4\% \]

• Implies that \( D_i = 1 \) improves survival by:

\[
\text{effect of cath on survival} = \frac{\Delta\text{(survival)}}{\Delta\text{(cath rate)}} = \frac{0.4\%}{0.067} = 6.0\%
\]

\[
E[y_i|D_i=1] - E[y_i|D_i=0] = \frac{E[y_i|b_i=1] - E[y_i|b_i=0]}{E[D_i|b_i=1] - E[D_i|b_i=0]}
\]

Sometimes called a “Wald” estimate
This is LATE -- on whom?

• Who is affected by background condition ($b_i = 0$ or 1)?

• Some patients will get cath anyway
  – Those who benefit the most, one hopes

• Some won’t get cath anyway
  – Hopefully those who won’t benefit (too sick already, otherwise not indicated)

• Distance to cath hospital affects marginal patient
  – gets cath only if close to cath hospital
Terminology: always takers; never takers, compliers

- **Always takers**: get treatment \((D_i = 1)\) no matter what
- **Never takers**: no treatment \((D_i = 0)\) no matter what
  - For both: \(D_i \text{ indep. of } b_i\)
- **Compliers**: Treatment depends on background
  - \(D_i = 1\) if and only if \(b_i = 1\)
- **LATE** (6.0% higher survival) is average effect on compliers (on the marginal patients)
- Terms developed for partly voluntary action
  - e.g., enroll in training program if eligible; complete diet or exercise program; serve in Army if drafted
  - but apply here too
• Clinical medical trial (low-fat diet, say):
  – Sign up 2,000 people; half get “treated”; half are controls
  – Only some complete the program
    • Dropouts are not random
  – So ATE for those who complete treatment is endogenous and biased.

• Use intention to treat ($b_i = 1$) as instrument for actual treatment
LATE with intent to treat

- Wald estimate of LATE

\[ LATE = \frac{\Delta(\text{expected outcome based on intent-to-treat})}{\Delta(\text{likelihood of treatment based on intent-to-treat})} \]

- Compare McClellan AMI study:
  - “intent”: treat those close to cath hospital with cath
  - But some who lived close \((b_i=1)\) got no cath
  - And some who lived farther away \((b_i = 0)\) got cath
  - Denominator is difference in cath rate between these two groups.
• ATET is average effect on those who are treated: always takers and compliers
• ATET = weighted average of LATE and (effect on always takers)
Instrument Validity

• Standard, and large, problem with instrument validity: There is no test!
• Valid instrument must be (i) exogenous; (ii) correlated, preferably strongly, with instrumented variable, and (iii) predict dependent variable only indirectly, through the instrumented variable
  – (i) is sometimes easy
  – (ii) can be estimated
  – (iii) [called the “exclusion restriction”] is the hardest to satisfy, and there is no test, only logic.
• There is, however, a Hausman test for endogeneity, assuming instrument is valid
  – Is IV coefficient estimate significantly different than OLS estimate
  – Caveat: weak instrument ➔ weak test

• And a Hansen “test of overidentifying restrictions”: validity of two or more instruments assuming one is valid
  – does IV coefficient estimate change significantly when add second instrument
  – and “difference in Hansen” test when add third instrument
What can these tests really tell you?

• Hausman: Is there evidence of endogeneity for the compliers?

• Hansen: Can fail this test if:
  
  (i) second instrument is not valid, or
  
  (ii) different instruments operate on different subpopulations with different LATEs
  
  – Test can’t distinguish between these
    • two effects could offset, leading to false comfort with multiple instruments
Example of different subpopulations

• Effect of additional child on mother’s earnings
  – focus on 3+ kids versus 2 (Angrist & Evans, 1998)
    • instrument 1: first two kids of same sex
      – weakly predicts greater probability of third child
    • instrument 2: second birth is twins
      – strongly predicts greater probability of third child

• What’s your prediction:
  – Will treatment effects be similar?
  – If different, which should be larger?
Same sex vs. twins estimates

• Married women, age 21-35, lots of controls
• Same sex estimate on probability of working:
  – 13.5% [95% CI = 7.7-19.3%]
• Twins estimate:
  – 8.3% [95% CI = 4.9-11.7%]
  – (Probably) different, with expected sign
  – If twins are random event,
    • Twins estimate is close to ATET = ATE
    • But on whom?: Effect on whom?
      – women with one child, and desire or willingness to have 2+
  – Same sex estimate is LATE
Wrapup

• Limitations of LATE are frustrating
  – Example: Benmelech (2009): railroad gauge in 1800s as instrument for asset specificity of railroad care
    • test effect of asset specificity on capital structure
      – generalization to other assets?
      – generalization to today’s financial markets??
    – Maybe still “Better LATE than Nothing” (Imbens, 2009)

• Lots of complexity left untouched:
  – variable intensity treatments (not just 0-1)
  – adding control variables