Microeconometric Policy Evaluation
Instrumental Variables

Monica Costa Dias
Institute for Fiscal Studies
Research Methods Festival
July 4th, 2012
A simple model of potential outcomes

• Simple binary treatment 0/1 for untreated (or treatment 0) and treated (or treatment 1), respectively
• $d_i$ represents the “treatment status” of individual $i$
• Each individual has two counterfactual outcomes, $y_i^0/y_i^1$, depending on treatment status
• We define
  $$y_i^0 = \beta + u_i$$
  $$y_i^1 = \beta + \alpha_i + u_i$$
• The observed outcome of individual $i$ is $y_i$
  $$y_i = y_i^0 + d_i (y_i^1 - y_i^0)$$
  $$= \beta + d_i \alpha_i + u_i$$
The treatment effect

- Wish to assess impact of treatment relative to no treatment on the outcome $y$
- For individual $i$ this is $\alpha_i = y_i^1 - y_i^0$: individual level causal effect
- *Missing data problem*: the treatment effect ($\alpha_i$) or the two potential outcomes $(y_i^0, y_i^1)$ cannot be directly measured for any individual
- We can hope to identify some features of the distribution of treatment effects, but not the individual treatment effect
Two main difficulties faced by evaluation studies

1. The treatment effect, $\alpha_i$, is heterogeneous
2. Selection into treatment may depend on both counterfactual outcomes, $(y_i^0, y_i^1)$, and thus on the gain from treatment, $\alpha_i$

Evaluation methods tend to be designed to identify some feature of the distribution of $\alpha_i$

We will start by focusing on the ATT but will then move to other moments of the distribution of the treatment effect
Consider an iid sample \( \{(y_i, d_i)\}_{i=1,...,N} \) and the linear regression \( y_i = \beta + \alpha d_i + e_i \). The OLS estimator of \( \alpha \) is

\[
\hat{\alpha}^{OLS} = \frac{1}{N} \sum_i y_i d_i - \frac{1}{N^2} \sum_i y_i \sum d_i \\
\frac{1}{N} \sum_i d_i^2 - \left( \frac{1}{N} \sum d_i \right)^2
\]

which identifies the parameter

\[
\alpha^{OLS} = E[\alpha_i | d_i = 1] + E[y_i^0 | d_i = 1] - E[y_i^0 | d_i = 0]
\]

- Heterogeneity: the first term is \( ATT = E[y_i^1 | d_i = 1] - E[y_i^0 | d_i = 1] \)
- Selection bias: the second term equals \( E[u_i | d_i = 1] - E[u_i | d_i = 0] \) and suggests treated and untreated are different
- **Selection on the unobservables**: conditioning on observables \( X \) may not change this result
IV directly addresses the problem of selection on the unobservables

- Selection creates compositional differences between treated and untreated

IV solution: find variable(s) $Z$ affecting selection but not outcomes

- Changes in $Z$ induce changes in treatment status without affecting outcomes
- Under certain conditions, variation in $Z$ can be used to compare otherwise identical individuals and identify the treatment effect
- $Z$ are the \textit{exogenous instruments}
- Similar to a “natural experiment”: find an event ($z = 0, 1$) that assigns individuals to treatment randomly
The model

- Omit observed variables: assume alignment of observed covariates
- Consider single instrument $z$ for simplicity
- The selection model of outcomes is

$$y_i = \beta + \alpha_i d_i + u_i$$
$$= \beta + \alpha_i d_i + [u_i + d_i(\alpha_i - \alpha)]_{=e_i}$$
$$d_i = 1 [g(z_i, v_i) \geq 0]$$

- Selection on the unobservables: $(e, v)$ are related - $(\alpha, v)$ and/or $(u, v)$ not independent
The outcome equation simplifies to

$$y_i = \beta + \alpha d_i + u_i$$

If $z$ unrelated to $y$ other then through $d$

$$E(y_i \mid z_i = z) = \beta + \alpha P(d_i = 1 \mid z) + E(u_i \mid z)$$

$$= \beta + \alpha P(z)$$

Choose $z^*$ and $z^{**}$ such that $P(d_i = 1 \mid z^*) \neq P(d_i = 1 \mid z^{**})$ and contrast the 2 groups

$$E(y_i \mid z^*) - E(y_i \mid z^{**}) = \alpha [P(z^*) - P(z^{**})] \quad \text{implying} \quad \alpha_{IV} = \frac{E(y_i \mid z^*) - E(y_i \mid z^{**})}{P(z^*) - P(z^{**})} = \alpha$$

If $z$ continuous it is more efficient to use all its variation

$$\text{cov}(y, z) = \alpha \text{cov}(d, z) + \text{cov}(u, z) \quad \text{implying} \quad \alpha_{IV} = \frac{\text{cov}(y, z)}{\text{cov}(d, z)}$$
Identification hinges on 3 assumptions

1. Homogeneity: $\alpha_i = \alpha$ for all $i$
2. $z$ determines participation: $P(d_i = 1 \mid z^*) \neq P(d_i = 1 \mid z^{**})$ (or $g$ is a non-trivial function of $z$)
3. Exclusion: $E(u \mid z) = E(u)$

When are these assumptions violated?

- returns from treatment unlikely to be homogeneous
- weak instruments - if $z$ has insufficient variation or is weakly related to $d$ $\rightarrow$ imprecise estimates of $\alpha$
- may be difficult to find data on a variable that does not affect simultaneously $d$ and $y$
The general model of outcomes is

\[ y_i = \beta + \alpha d_i + [u_i + d_i(\alpha_i - \alpha)] = e_i \]

Classical IV now identifies

\[ \alpha^{IV} = \alpha + \frac{E(e_i | z^*) - E(e_i | z^{**})}{P(z^*) - P(z^{**})} \]

unless the IV condition \( E(y_i | z_i = z) = \beta + \alpha p(z) \) still holds, meaning

\[ E(e_i | z_i = z) = E(u_i | z_i = z) + P(d_i = 1 | z)E(\alpha_i - \alpha | d_i = 1, z_i = z) = 0 \]

In particular, the IV condition requires individuals not to have, or not to act upon, information about their own idiosyncratic gains.

Violation of the classical IV condition means \( z \) affects outcomes through ways other than \( d \).
Homogeneity (or ignorance) is not compelling: individuals expected to use more and better information about their own potential outcomes then can be observed.

Under an additional assumption, Imbens and Angrist (1994, Econometrica) offer an interpretation to the IV estimator: LATE

- Suppose there exists a variable $z$ capable of inducing individuals to change treatment status for reasons unrelated to potential outcomes.
- Imagine having data on 2 groups with different realisations of $z$ but otherwise similar.
- Observed differences in mean outcomes can then be attributed to differences in participation rates due to $z$ only.
- In special cases, such differences can be use to identify the impact of treatment on the subpopulation of compliers.
LATE: assumptions

- Remember the model of outcomes
  \[ y_i = \beta + \alpha d_i + [u_i + d_i(\alpha_i - \alpha)] = e_i \]

- Consider a binary instrument \((z = 0/1)\) such as an exogenous policy reform.

- Define the function \(d_{iz}\) as the treatment status of individual \(i\) under policy \(z\):
  \[ d_{iz} = 1(g(z, v_i) > 0) \]

- LATE requires stronger assumptions than classical IV to compensate for the lack of homogeneity.

  1. \(z\) determines participation (\(g\) is a non-trivial function of \(z\) - IV2)
  2. Exclusion: \(E(u_i|z) = E(u_i)\) (IV3)
  3. \((\alpha, v)\) are jointly independent of \(z\)
Assumptions 2 and 3 impose

- potential outcomes \((y^0, y^1)\) are not affected by the policy regime
- \(z\) is exogenous in the participation equation

\[
p(d_i = 1|z_i = z) = P(g(z, v_i) > 0) = P(d_{iz} = 1) = P(z)
\]

And can be used to derive

\[
E(y_i|z_i = z) = \beta + P(d_i = 1|z) E(\alpha_i|d_i = 1, z)
= \beta + P(d_{iz} = 1) E(\alpha_i|d_{iz} = 1)
\]

Contrasting the policy regimes under additional assumption 1:

\[
E(y_i|z_i = 1) - E(y_i|z_i = 0)
= P[d_{i1} - d_{i0} = 1] E[\alpha_i | d_{i1} - d_{i0} = 1] - P[d_{i1} - d_{i0} = -1] E[\alpha_i | d_{i1} - d_{i0} = -1]
\]
Contrasting the policy regimes under additional assumption 1:

$$E(y_i | z_i = 1) - E(y_i | z_i = 0)$$

$$= P [d_{i1} - d_{i0} = 1] E [\alpha_i | d_{i1} - d_{i0} = 1] - P [d_{i1} - d_{i0} = -1] E [\alpha_i | d_{i1} - d_{i0} = -1]$$

The above expression is useless unless

- homogeneous effects:  
  $$E(y_i | z_i = 1) - E(y_i | z_i = 0) = P [d_{i1} \neq d_{i0}] E [\alpha_i]$$
  - impose additional monotonicity assumption

- **Monotonicity**:  
  $$d_{i0} \geq (\leq) d_{i1}$$ for all $$i$$ (with strict inequality for some $$i$$)

This is to say that either $$P [d_{i1} - d_{i0} = 1] = 0$$ or $$P [d_{i1} - d_{i0} = -1] = 0$$, but not both

Notice that an index restriction in the participation rule (meaning $$\nu$$ is additively separable) implies the monotonicity assumption
LATE: identification

Suppose \( p[d_{i1} - d_{i0} = -1] = 0 \)

- any \((z = 0)\)-participant is also a \((z = 1)\)-participant
- Then

\[
\alpha^{\text{LATE}} = E[\alpha_i \mid d_{i1} - d_{i0} = 1] = \frac{E(y_i \mid z_i = 1) - E(y_i \mid z_i = 0)}{P(z_i = 1) - P(z_i = 0)}
\]
Local assumptions and local parameters

Controversy surrounding LATE

- shows IV can be meaningless when effects are heterogeneous
- if monotonicity assumption justified, LATE can be an interesting approach to compare two policy regimes
- but in generally results are instrument-dependent and LATE measures effects on a not clearly defined population
- interpretation particularly cumbersome when \( z \) continuous
We have studied two different parameters - ATT and LATE
- both average over parts of the distribution of treatment effects
- makes it difficult to interpret and synthetise results

How they relate to each other is formalised by the Marginal Treatment Effect (MTE)
- First introduced by Bjorklund and Moffit (1987) to quantify the impact of treatment on individuals just indifferent about participation
- Heckman and Vytlacil (1999, 2001, 2006) use the MTE as a unifying parameter in the treatment effect literature
  - basis for definition of all other average treatment effect parameters
  - and for their interpretation
- They notice LATE can be measured for infinitesimal changes in the instrument $z$ to form the MTE
MTE: definition

- Consider a continuous instrument, $z$
- And the selection model of outcomes after imposing an index restriction on the selection rule

$$y_i = \beta + \alpha d_i + [u_i + d_i(\alpha_i - \alpha)] = e_i$$

$$d_i = 1[v_i < g(z_i)]$$

- For a given value $z$
  - participants are those drawing $v_i < g(z)$
  - the marginal (indifferent) participant draws $v_i = g(z)$
- MTE: effect on individuals drawing a specific value of $v$, say $g(z)$

$$E(y_i^1 - y_i^0 | v_i = g(z)) = E(\alpha_i | v_i = g(z)) = \alpha^{MTE}(g(z))$$
Assume we are under the LATE assumptions 1 to 3 together with the index restriction.

Let $F_v$ be cdf of $v$ and write, for $z_i = z$

$$P(z) = P(v_i < g(z))$$
$$= F_v(g(z))$$

Under the index restriction

$$v_i < g(z) \iff F_v(v_i) < F_v(g(z)) \iff \tilde{v}_i < P(z)$$

where $\tilde{v} = F_v(v)$ follows a uniform $[0,1]$ distribution.

Now, for a given $z$ and $p = P(z)$:

- a participant is someone drawing $\tilde{v}_i$ below $p = P(z)$
- indifference regarding participation occurs at $\tilde{v}_i = p$
- MTE redefined as the impact of treatment at a point $p$ in the distribution of $\tilde{v}$

$$\alpha^{MTE}(p) = E(\alpha_i | \tilde{v}_i = p)$$
MTE independent of $z$ since $z$ contains no information on expected gains after conditioning on $\tilde{v}$ (LATE assumptions)

$$\alpha^{MTE}(p) = E(\alpha_i|\tilde{v}_i = p, z_i) \text{ for any value } z_i$$

Thus MTE is the average impact of treatment on individuals drawing a specific value of $\tilde{v}$, irrespective of $z$.

But for those indifferent at $p$ - meaning $z_i : \tilde{v}_i = p = P(z_i)$

$$\alpha^{MTE}(p) = E(\alpha_i|\tilde{v}_i = p, P(z_i) = p)$$

This expression justifies the interpretation of MTE as the impact of treatment on individuals at the margin of participation.

It also supports the identification strategy using LIV.
MTE: Local IV

- Under LATE assumptions 2 and 3 together with additive separability of $v$

  \[
  E(y_i | z_i = z) = \beta + P(z) E(\alpha_i | z, d_i = 1) \\
  = \beta + P(z) E(\alpha_i | \tilde{v}_i < P(z)) \\
  = E(y_i | P(z))
  \]

- Further imposing the first LATE assumption and contrasting two points in the domain of $z$, say $(z^*, z^{**})$

  \[
  \alpha^{LATE}(z^*, z^{**}) = \frac{E(y | z^*) - E(y | z^{**})}{P(z^*) - P(z^{**})} \\
  = \frac{E(y | P(z^*)) - E(y | P(z^{**}))}{P(z^*) - P(z^{**})}
  \]

- Taking the limits as $z^*$ and $z^{**}$ become arbitrarily close

  \[
  \alpha^{LIV}(P(z)) = \frac{\partial E(y | P(z))}{\partial P(z)}
  \]

- LIV stands for Local IV - a formulation of the MTE parameter using individuals at the margin of participation at $P(z)$
The derivation of LIV suggests an estimation procedure for the local MTE

1. estimate \( P(z) \) and compute the predicted values \( \hat{p} \)
2. regress \( y \) on \( P(z) \) non-parametrically - say using local polynomials
3. differentiate with respect to \( P(z) \)

If \( z \) can induce variation in \( P(z) \) over the full support \((0, 1)\), it is possible to estimate the whole distribution of MTEs

In which case all population parameters can be derived from the MTE
MTE: recovering ATT

Recovering the ATT requires a little more work.

- At each point \( p \), the ATT is the impact of treatment on participants at such propensity score:

\[
\alpha^{ATT}(p) = \int_0^p \alpha^{MTE}(\tilde{v}) \ dF_{\tilde{v}}(\tilde{v} | \tilde{v} < p)
\]

\[
= \int_0^p \alpha^{MTE}(\tilde{v}) \frac{1}{p} \ d\tilde{v}
\]

- and the overall ATT is

\[
\alpha^{ATT} = \int_0^1 \alpha^{ATT}(\tilde{v}) \ f_p(p | d = 1) \ dp
\]

- An estimator of the ATT is the empirical counterpart of the above parameter