Identification of Causal Effects Using Instrumental Variables

Patrick Lam

November 27, 2007

How does the potential outcomes framework used by statisticians relate to the instrumental variables (IV) approach used by econometricians? Under what conditions can you make causal interpretations of the IV estimand, and what are the consequences of violations of the assumptions?

Example: What is the causal effect of serving in the military during the Vietnam War on post-war civilian mortality?

The Econometric Approach

Economists typically estimate causal effects through structural equation models, which rely on specification of systems of equations with parameters and variables that attempt to capture behavioral relationships and specify the causal links between variables. They often exploit the presence of instrumental variables, which are included in some equations and excluded in others, to deal with issues of omitted variable bias or endogeneity.

Notation

- D: binary treatment variable (serving in the military = 1) - explanatory variable
- Z: binary treatment assignment (drafted = 1) - instrument
- Y: binary outcome variable (death = 1)

Structural Equation Models

Structural equation models are stochastic models in which each equation represents a causal link, rather than a mere empirical association. For person i, Y_i is the observed health outcome, D_i is the observed treatment, and Z_i is the observed draft status. This dummy endogenous variables model would have the form

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \epsilon_i \tag{1}$$

$$D_i^* = \alpha_0 + \alpha_1 \cdot Z_i + \nu_i \tag{2}$$

and

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases}$$

β_1 represents the causal effect of D on Y. Note the features of the model, including an underlying linear structure, constant coefficients, and a reliance on error terms to characterize omitted variables.

Assumptions

In order to identify β_1, three assumptions are typically made:

A1. $E[Z_i \cdot \epsilon_i] = 0$: Z_i is uncorrelated with the error in Equation (1), which combined with the absence of Z in Equation (1) captures the notion that any effect of Z on Y must be through an effect of Z on D.

A2. $E[Z_i \cdot \nu_i] = 0$: Z_i is uncorrelated with the error in Equation (2), which captures the notion that the assignment of the instrument is ignorable or exogenous to the treatment.

A3. $\text{cov}(D_i, Z_i) \neq 0$: An equivalent interpretation is requiring that $\alpha_1 \neq 0$. This assumption and Equation (2) combined implies that Z has a causal effect on D.

If Z_i satisfies these assumptions, then it is considered an instrumental variable in this model. The IV estimator, $\hat{\beta}_1^{\text{IV}}$, is simply the ratio of the effect of Z on Y divided by the effect of Z on D:

$$\hat{\beta}_1^{\text{IV}} = \frac{\widehat{\text{cov}}(Y_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} \tag{3}$$

The Potential Outcomes Approach

Statisticians often prefer an alternative framework for a causal interpretation of the IV estimand based on potential outcomes and the Rubin Causal Model (RCM).

Notation

- Z, D, and Y: the treatment assignment (instrument), treatment variable, and outcome variable as before.
- $\mathbf{Z}$: an N-dimensional vector of assignments with ith element Z_i.
- $\mathbf{D}$: an N-dimensional vector of treatments with ith element D_i.
- $D_i(\mathbf{Z})$: an indicator for whether person i would serve given the randomly allocated vector of assignments $\mathbf{Z}$.
- $Y_i(\mathbf{Z}, \mathbf{D})$: the outcome for person i given the vector of treatment indicators $\mathbf{D}$ and vector of assignments $\mathbf{Z}$.

$Y_i(\mathbf{Z}, \mathbf{D})$ and $D_i(\mathbf{Z})$ are potential outcomes, which are only partially observed.

Definitions

- **Never-taker** ($D_i(1) = D_i(0) = 0$): a person who does not take the treatment regardless of treatment assignment.
- **Always-taker** ($D_i(1) = D_i(0) = 1$): a person who takes the treatment regardless of treatment assignment.

- **Complier** ($D_i(1) = 1, D_i(0) = 0$): a person who takes the treatment when assigned and does not take the treatment when not assigned.
- **Defier** ($D_i(1) = 0, D_i(0) = 1$): a person who takes the treatment when not assigned and does not take the treatment when assigned.

Table 1: Table of Notation and Definitions

			$D(\mathbf{Z})$		$Y(\mathbf{Z},\mathbf{D})$				
i	$\mathbf{Z}$	$\mathbf{D}$	$D(1)$	$D(0)$	$Y(1,1)$	$Y(0,0)$	$Y(0,1)$	$Y(1,0)$	Type
1	1	1	1	(1)	y	?	?	?	Always-taker
2	0	0	(0)	0	?	y	?	?	Never-taker
3	1	1	1	(0)	y	?	?	?	Complier
4	0	1	(0)	1	?	?	y	?	Defier

Numbers in parentheses or ? represent unobserved outcomes.

Assumptions

B1. **Stable Unit Treatment Value Assumption** (SUTVA):

(a) If $Z_i = Z_i'$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$.

(b) If $Z_i = Z_i'$ and $D_i = D_i'$, then $Y_i(\mathbf{Z},\mathbf{D}) = Y_i(\mathbf{Z}',\mathbf{D}')$.

SUTVA implies that the potential outcomes for each person are unrelated to the treatment status of other individuals. If Z_i does not change, then $D_i(Z_i)$ does not change regardless of how Z_{-i} changes. Similarly, if Z_i and D_i do not change, then $Y_i(Z_i, D_i)$ does not change regardless of how Z_{-i} and D_{-i} change. This assumption allows us to write $Y_i(\mathbf{Z},\mathbf{D})$ and $D_i(\mathbf{Z})$ as $Y_i(Z_i, D_i)$ and $D_i(Z_i)$.

Figure 1: SUTVA Assumption implies absence of dotted arrows.

$$\begin{tikzcd}
Z_i \arrow[r] \arrow[dr, dotted] \arrow[drr, dotted] & D_i \arrow[r] \arrow[dr, dotted] & Y_i \\
Z_j \arrow[r] & D_j \arrow[r] & Y_j
\end{tikzcd}$$

Example: The veteran status of any man at risk of being drafted in the lottery was not affected by the draft status of others at risk of being drafted, and, similarly, that the civilian mortality of any such man was not affected by the draft status of others.

B2. **Random (Ignorable) Assignment of** Z_i (instrument):

$$\Pr(\mathbf{Z} = \mathbf{c}) = \Pr(\mathbf{Z} = \mathbf{c}')$$

The treatment assignment is randomly assigned.

Example: Assignment of draft status was random.

B3. **Exclusion Restriction**

$$\mathbf{Y}(\mathbf{Z}, \mathbf{D}) = \mathbf{Y}(\mathbf{Z}', \mathbf{D}) \text{ for all } \mathbf{Z}, \mathbf{Z}' \text{ and for all } \mathbf{D}$$

The exclusion restriction implies that any effect of Z on Y must be via an effect of Z on D.

Figure 2: Exclusion assumption implies absence of dotted arrow.

$$\begin{tikzcd} Z_i \arrow[r] \arrow[rr, bend right, dotted] & D_i \arrow[r] & Y_i \end{tikzcd}$$

Example: Civilian mortality risk was not affect by draft status once veteran status is taken into account.

B4. **Nonzero Average Causal Effect (ACE) of** Z **on** D

$$E[D_i(1) - D_i(0)] \neq 0$$

There is an effect of Z on D.

Example: Having a low lottery number increases the average probability of service.

B5. **Monotonicity**

$$D_i(1) \geq D_i(0) \text{ for all } i = 1, \ldots, N$$

Monotonicity implies that there are no defiers in the population.

Example: There is no one who would have served if given a high lottery number, but not if given a low lottery number.

Identifying Causal Effects

The goal is to estimate the unbiased causal effect of treatment D on outcome Y. However, assignment of the treatment D is not ignorable, so we must use an IV-like approach. We begin by finding the effect of Z on Y. At the individual unit level, with SUTVA (B1) and the exclusion restriction (B3), the effect is given by

$$\begin{aligned} Y_i(1, D_i(1)) - Y_i(0, D_i(0)) &= Y_i(D_i(1)) - Y_i(D_i(0)) \text{ by exclusion restriction} \\ &= [Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))] - [Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))] \\ &= (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0)) \end{aligned}$$

which is the product of the causal effect of D on Y and the causal effect of Z on D. For always-takers and never-takers, $D_i(1) - D_i(0) = 0$. The monotonicity assumption (B5) says that there are no defiers, so $D_i(1) - D_i(0) \geq 0$. Thus, Z only has a non-zero effect on Y for compliers. The expected average causal effect of Z on Y is just the average causal effect of D on Y for compliers times the probability an individual is a complier

$$E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] = E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \cdot P[D_i(1) - D_i(0) = 1]$$

Since $P[D_i(1)-D_i(0)=1]=E[D_i(1)-D_i(0)]$ by the monotonicity assumption (no defiers) and $E[D_i(1)-D_i(0)]\neq 0$ by the non-zero ACE of Z on D assumption (B4), we can rewrite the equation as

$$E[(Y_i(1)-Y_i(0))|D_i(1)-D_i(0)=1]=\frac{E[Y_i(D_i(1),1)-Y_i(D_i(0),0)]}{E[D_i(1)-D_i(0)]}$$

This is known as the Local Average Treatment Effect (LATE) and is equivalent to the IV estimand. The numerator is the effect of Z on Y and the denominator is the effect of Z on D. This effect is a local average treatment effect because it is only the unbiased effect of D on Y on the subpopulation of compliers, and not on the whole population.

With the assumption of random assignment of Z (B2) and SUTVA, we can estimate the LATE. The LATE is simply the ratio of the *average intention-to-treat effects*, Z on Y and Z on D. The average intention-to-treat effects are given by

- Z on Y: $\frac{\sum_i Y_i Z_i}{\sum_i Z_i}-\frac{\sum_i Y_i(1-Z_i)}{\sum_i(1-Z_i)}$
- Z on D: $\frac{\sum_i D_i Z_i}{\sum_i Z_i}-\frac{\sum_i D_i(1-Z_i)}{\sum_i(1-Z_i)}$

which are simply mean differences between the two treatment assignments. We can then calculate the LATE.

$$\text{LATE}=\frac{E[Y_i(D_i(1),1)-Y_i(D_i(0),0)]}{E[D_i(1)-D_i(0)]}=\frac{\frac{\sum_i Y_i Z_i}{\sum_i Z_i}-\frac{\sum_i Y_i(1-Z_i)}{\sum_i(1-Z_i)}}{\frac{\sum_i D_i Z_i}{\sum_i Z_i}-\frac{\sum_i D_i(1-Z_i)}{\sum_i(1-Z_i)}}$$

Comparing the Two Approaches

Table 2 shows the equivalent assumptions and results from both approaches.

Table 2: Comparing the Econometric and Potential Outcomes Approaches

Econometric Approach	Potential Outcomes Approach
A1. $E[Z_i\cdot\epsilon_i]=0$	B3. Exclusion Restriction
A2. $E[Z_i\cdot\nu_i]=0$	B2. Random (Ignorable) Assignment of Z
A3. $\text{cov}(D_i,Z_i)\neq 0$	B4. Nonzero Average Causal Effect of Z on D
Implied by constant α_0 and α_1	B5. Monotonicity
Implied by independence of observations	B1. SUTVA
IV estimand: β_1^{IV}	Local Average Treatment Effect (LATE)

Although it is clear from the potential outcomes framework that the IV estimand is only a local average treatment effect on the compliers, the econometric approach does not explicitly state this clearly. Therefore, it is important to be clear that the IV estimand is only over the subpopulation of compliers. Many interpretations of the IV in the econometric approach make the assumption that the LATE is the same as the population effect, which in essence assumes that the effect of the treatment is the same for compliers and noncompliers. This is a stronger assumption than the model allows.

References

Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91(434):444–455.