Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations

James Heckman

1995

Treatment Model

$$Y = \begin{cases} Y_0 & \textit{if} \quad D = 0 \\ Y_1 & \textit{if} \quad D = 1 \end{cases}$$

or

$$Y = (1 - D)Y_0 + DY_1$$

D: enrollment in military service.

Y_j: earnings.

The Evaluation Problem

Estimate the gain into the program

$$\Delta = (Y_1 - Y_0)$$

The Mean Effect of the Treatment of the Treated:

$$E(\Delta | D = 1, X) = E(Y_1 - Y_0 | D = 1, X)$$

$E(Y_1 | D = 1, X)$ is known but $E(Y_0 | D = 1, X)$ is not.

Random Assignment Effect

$$E(\Delta | X) = E(Y_1 - Y_0 | X)$$

This can only be estimated if the program participation or nonparticipation is universal. That is, the group that is randomly assigned or not assigned to the program has to be the representative of the entire population. This is unlikely to be the case.

Local Average Treatment Effect

The treatment assignment is a function of the instrument z and some other unobservable V, which is assumed to be random.

$$E\left[Y_1(z) - Y_0(z')|V : D(z, V) = 1, D(z', V) = 0\right]$$

Econometric Specification

$$E[Y_0|X] = \mu_0(X)$$
$$E[Y_1|X] = \mu_1(X)$$

Hence,

$$Y_0 = \mu_0(X) + U_0$$
$$Y_1 = \mu_1(X) + U_1$$

$E[U_0|X] = 0,\ E[U_1|X] = 0.$

Then,

$$Y = \mu_0(X) + D\left(\mu_1(X) - \mu_0(X) + U_1 - U_0\right) + U_0$$

$\mu_1(X) - \mu_0(X)$: average gain from program participation of people with characteristics X.
$U_1 - U_0$: idiosyncratic gain from program participation.

Random Assignment Effect:Mean Effect of Treatment

$$E\left(Y_1 - Y_0 | X\right) = \mu_1(X) - \mu_0(X)$$

The Mean Effect of the Treatment of the Treated:

$$E\left(Y_1 - Y_0 | D = 1, X\right) = \mu_1(X) - \mu_0(X) + E\left[U_1 - U_0 | D = 1, X\right]$$

The mean effect of treatment given X and the treatment of the treated are different by $E\left[U_1 - U_0 | D = 1, X\right]$.

Mean Selection Bias

$$Y = \mu_0(X) + D\left(\mu_1(X) - \mu_0(X) + U_1 - U_0\right) + U_0$$

If we look at the outcome difference between program participants and nonparticipants, then

$$\begin{aligned} & E\left(Y_1 | D = 1, X\right) - E\left(Y_0 | D = 0, X\right) \\ = \; & \mu_1(X) + E\left[U_1 | D = 1, X\right] - \mu_0(X) - E\left[U_0 | D = 0, X\right] \\ = \; & \mu_1(X) - \mu_0(X) + E\left[U_1 - U_0 | D = 1, X\right] \\ & + E\left[U_0 | D = 1, X\right] - E\left[U_0 | D = 0, X\right] \end{aligned}$$

Mean bias from the mean treatment effect:

$$E\left[U_1 - U_0 | D = 1, X\right] + E\left[U_0 | D = 1, X\right] - E\left[U_0 | D = 0, X\right]$$

Mean Selection bias from the mean treatment effect of the treated:

$$E\left[U_0|D=1,X\right]-E\left[U_0|D=0,X\right]$$

Treatment effect of the treated does not face selection bias if selection does not depend on the error term in non-treatment outcome.
Mean treatment effect and the mean treatment effect of the treated are equal if

$$E\left[U_1-U_0|D=1,X\right]=0$$

The Instrumental Variable Estimation

Assumption 1: The error term should be mean independent to the instrument.

$$Y = \mu_0(X) + D\left(\mu_1(X) - \mu_0(X)\right) + D\left(U_1 - U_0\right) + U_0$$

Assumption 1 (IV for mean treatment effect estimation)

$$E\left[D\left(U_1 - U_0\right) + U_0 | X, Z\right] = 0$$

$$\begin{aligned} Y &= \mu_0(X) + D\left[\mu_1(X) - \mu_0(X) + E\left(U_1 - U_0 | D = 1, X\right)\right] \\ &\quad + D\left(U_1 - U_0\right) + U_0 - DE\left(U_1 - U_0 | D = 1, X\right) \end{aligned}$$

Assumption 1 a (IV for mean treatment effect of the treated estimation)

$$E\left[D\left(U_1 - U_0\right) + U_0 - DE\left(U_1 - U_0 | D = 1, X\right) | X, Z\right] = 0$$

Assumption 2: The IV should be correlated with D.

$$E\left[D|X,Z\right] = Pr(D=1|X,Z) \neq constant$$

Assumptions 1 and 2 imply

$$E\left[Y|X,Z\right] = \mu_0(X) + Pr(D=1|X,Z)\left(\mu_1(X) - \mu_0(X)\right)$$

Assumptions 1a and 2 imply

$$E\left[Y|X,Z\right] = \mu_0(X) + Pr(D=1|X,Z)E\left[\Delta|D=1,X,Z\right]$$

where

$$E\left[\Delta|D=1,X,Z\right] = \mu_1(X) - \mu_0(X) + E\left(U_1 - U_0|D=1,X\right)$$

the mean treatment of the treated.

IV Estimator

Mean Treatment Effect

$$E[\Delta|X] = \frac{E[Y|X,z] - E[Y|X,z']}{Pr(D=1|X,z) - Pr(D=1|X,z')}$$

Mean Treatment Effect of the Treated

$$E[\Delta|D=1,X] = \frac{E[Y|X,z] - E[Y|X,z']}{Pr(D=1|X,z) - Pr(D=1|X,z')}$$

Military draft IV by Angrist is not a valid IV for the Mean Treatment of the Treated Effect.
Only persons with high returns to military service $(U_1 - U_0)$ are voluntarily going to serve in the military $(Z = 0, D = 1)$. On the other hand, even individuals with low returns to service $(U_1 - U_0)$ are serving in the military if drafted $(Z = 1, D = 1)$. Then, the mean treatment of the treated estimate measures

$$\mu_1(X) - \mu_0(X) + E\left(U_1 - U_0 | D = 1, X, Z\right)$$

which depends on Z.

Military draft IV by Angrist is in general not a valid IV for the Mean Treatment Effect. The conditional mean of the error term is a function of Z.

$$
\begin{aligned}
& E\left[D\left(U_1 - U_0\right) + U_0 | X, Z\right] \\
= \quad & E\left[U_1 - U_0 | D = 1, X, Z\right] + E\left[U_0 | X, Z\right]
\end{aligned}
$$

$$E\left[\left(U_1 - U_0\right) | D = 1, X, Z = 0\right] \neq E\left[D\left(U_1 - U_0\right) | D = 1, X, Z = 1\right]$$

Military draft IV is only valid for the estimation of mean treatment of the treated if the treatment decision of individuals do not depend on the individual specific returns to outcome, which implies.

$$E\left[U_1 - U_0 | D = 1, X, Z\right] = E\left[U_1 - U_0 | D = 1, X\right]$$

which is an unreasonable assumption.

Local Average Treatment Effect

$$
\begin{aligned}
& E\left[Y_1 - Y_0 | D(z) = 1, D(z') = 0\right] \\
= \quad & \frac{E\left[Y|Z = z\right] - E\left[Y|Z = z'\right]}{Pr(D = 1|Z = z) - Pr(D = 1|Z = z')}
\end{aligned}
$$

- LATE is defined by the instrument, which is external to the outcome equation.
- How should one make structural interpretation of the estimation result? If we interpret LATE as the estimate of the structural parameter of the outcome equation, then the structural parameter changes with the value of the instrument. This is against the idea that structural parameters should be policy invariant.

- Instrument is a policy variable: The effect of individuals who change treatment status by a specific policy change.
- Instrument is a personal characteristic: distance to school. Returns to education that is estimated based on individuals who will change schooling under certain change in distance to school.

- LATE:

$$E\left[Y_1 - Y_0 | V : D(z, V) = 1, D(z', V) = 0\right]$$

- Mean Treatment Effect of the Treated given z

$$\begin{aligned} & E\left[Y_1 - Y_0 | V : D(z, V) = 1, z, X\right] \\ = \;& E\left[Y_1 - Y_0 | V : D(z, V) = 1, D(z', V) = 0\right] \\ & Pr(V : D(z, V) = 1, D(z', V) = 0) \\ & + E\left[Y_1 - Y_0 | V : D(z, V) = 1, D(z', V) = 1\right] \\ & Pr(V : D(z, V) = 1, D(z', V) = 1) \\ = \;& LATE(z, z') Pr(V : D(z, V) = 1, D(z', V) = 0) \\ & + E\left[Y_1 - Y_0 | V : D(z, V) = 1, D(z', V) = 1\right] \\ & Pr(V : D(z, V) = 1, D(z', V) = 1) \end{aligned}$$

That is, mean treatment effect of the treated is the probabilty sum of LATE and the mean treatment effect of those who are always treated (always takers).

- Mean treatment effect of the treated and LATE are equal if the policy change from z' to z has no effect on those who are always takers. But that won't be true if the policy change is costly and its cost is shared among all people.
- In small program with limited potential participants, cost that is shared among everybody may be tiny and can be ignored. But in that case, the potential participants of the program may not be the representative population.