Instrumental Variables Notes

Thomas Leavitt

2024-06-26

Setup

Under SUTVA, we can partition participants into 4 principal strata related to participants' compliance with treatment assignment. Let $d_i(0)$ denote whether participant i would receive the treatment if assigned to the control condition; also let $d_i(1)$ denote whether participant i would receive the treatment if assigned to the treatment condition. The 4 principal strata in terms of $d_i(0)$ and $d_i(1)$ are below.

$z_i = 0$	$z_i = 1$	Principal Stratum
$d_i(0) = 1$	$d_i(1) = 1$	*Always-Taker*
$d_i(0) = 0$	$d_i(1) = 1$	*Complier*
$d_i(0) = 1$	$d_i(1) = 0$	*Defier*
$d_i(0) = 0$	$d_i(1) = 0$	*Never-Taker*

Goal

We would like to estimate the average effect of Z on Y among Compliers. Because we do not know which units are Compliers, we cannot condition on them in estimating the average treatment effect of Z on Y. We can unbiasedly and consistently estimate two unconditional quantities, the average effect of Z on Y and the average effect of Z on D. Our goal is to show that, under specific assumptions, the ratio of these two unconditional average effects is equal to the conditional average effect of Z on Y among Compliers. This equality is helpful because it means that, under random assignment, we can consistently estimate the conditional average effect among Compliers despite our inability to directly condition on Compliers.

Proof

First, we will derive a new expression for the ITT_D. Then we will derive a new expression for the ITT_Y. Then we will show that the ratio of the new expression of ITT_Y to the new expression of ITT_D is equal to the conditional average effect of Z on Y among Compliers.

We can write the ITT_D, as

$$\text{ITT}_D = \left(\frac{1}{N}\right) \sum_{i=1}^{N} d_i(1) - d_i(0).$$

Note that, for all Always-Takers and Never-Takers, $d_i(1) - d_i(0) = 0$; by contrast, for Compliers, $d_i(1) - d_i(0) = 1$ and, for Defiers, $d_i(1) - d_i(0) = -1$. Hence, $\sum_{i=1}^{N} d_i(1) - d_i(0)$ is the number of Compliers minus the number of Defiers, which implies that ITT_D is

$$\text{ITT}_D = \left(\frac{1}{N}\right) \left(N^{\text{C}} - N^{\text{D}}\right) = \pi^{\text{C}} - \pi^{\text{D}}.$$

The assumption of No Defiers implies that $\pi^{\text{D}} = 0$; hence,

$$\text{ITT}_D = \pi^{\text{C}}. \tag{1}$$

Under the additional assumptions of (1) Exclusion Restriction (i.e., treatment assignment affects the outcome only through receipt of the treatment), (2) No Defiers and (3) Existence of Compliers, the ratio of ITT_Y to ITT_D is equal to the CACE. To see why, first note that, under SUTVA, ITT_Y is

$$\text{ITT}_Y = \left(\frac{1}{N}\right) \sum_{i=1}^{N} \tau_i.$$

We can decompose $\sum_{i=1}^{N} \tau_i$ into the sum of individual effects among the four compliance types:

$$\sum_{i=1}^{N} \tau_i = N^{\text{AT}}\tau^{\text{AT}} + N^{\text{C}}\tau^{\text{C}} + N^{\text{D}}\tau^{\text{D}} + N^{\text{NT}}\tau^{\text{NT}},$$

where the sum of individual effects for individuals in one compliance type is the number of individuals belonging to that type multiplied by the average effect among individuals belonging to that type, e.g., $N^{\text{AT}}\tau^{\text{AT}}$.

The Exclusion Restriction stipulates that the individual effects for all Always-Takers and Never-Takers are 0,

which implies that the average effects among *Always-Takers* and *Never-Takers* are 0. Hence, ITT_Y is

$$\text{ITT}_Y = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\tau_i = \left(\frac{1}{N}\right)\left[N^{\text{C}}\tau^{\text{C}} + N^{\text{D}}\tau^{\text{D}}\right] = \pi^{\text{C}}\tau^{\text{C}} + \pi^{\text{D}}\tau^{\text{D}},$$

which, as the assumption of No Defiers implies, is equal to

$$\text{ITT}_Y = \pi^{\text{C}}\tau^{\text{C}}. \tag{2}$$

Then, invoking the expression for the ITT_D in (1), it follows that the ratio of ITT_Y to ITT_D is

$$\frac{\text{ITT}_Y}{\pi^{\text{C}}}.$$

The Existence of Compliers implies that this ratio is well-defined (i.e., that it does not divide by 0). Then, plugging in the expression for ITT_Y in (2) implies that the ratio of ITT_Y to ITT_D is

$$\frac{\pi^{\text{C}}\tau^{\text{C}}}{\pi^{\text{C}}} = \tau^{\text{C}},$$

which is the CACE.

Under random assignment, we can unbiasedly estimate both ITT_Y and ITT_D, which, since $\text{CACE} = \dfrac{\text{ITT}_Y}{\text{ITT}_D}$, means that we can consistently estimate the CACE. That is, with a sufficiently large experiment, we would generate a CACE estimate close to the true CACE with probability close to 1.