

Sociology 750 - Spring 2004 - Freese - The Logic of Instrumental Variables

Really, truly, even though instrumental variables techniques are often talked about like they represent “advanced” or “technical” or “sophisticated” data analytic methods (at least by the modest standards of quantitative sociology), the underlying logic of what’s going on is this simple.

Inmates in Cell Block A are given the option of joining a voluntary education program. A study followed up 200 former Block A prisoners to see if they had been arrested for another offense within a year after their release. As it happens, 100 of these prisoners had participated in the education program and 100 had not. Say that:

- Of the 100 prisoners who did not participate in the program, 60 had been arrested
- Of the 100 who did participate in the program, only 20 had been arrested

What to make of this? We know that 20 participants were arrested. The key question for causal inference is how many participants would have been arrested had there been no reading program.

If participation in the reading program had been randomly assigned, we could infer that the group of non-participants provides the best guess as to what would have happened to the participants had they not participated. That is, we would infer that 60 of them would have been arrested, or 40 more than were actually arrested.

The problem with the reading program being voluntary is that we have every reason to believe that there could be important differences between prisoners motivated to sign up for the reading program and those who do not. Had the reading program been cancelled before it ever began, it could still be that only 20 of those who had signed up for the program would be re-arrested.

Under normal circumstances, we would be at an inferential impasse here, at which we could only lament that a randomized experiment was not done.

Lo, the same prison also has a Block B, in which a voluntary education program was not offered to anyone. Whether prisoners happened to be assigned to Block A or Block B was not in any way determined by characteristics of the prisoners, and so we can presume assignment into one cell block or the other was a random matter. We also followed up on 200 Block B prisoners to see if they had been re-arrested within a year of their release.

Say that x of these prisoners had been arrested.

The key move is to realize that Block B (as a whole) provides us with the best picture of what would have happened to Block A (as a whole) if the program had not been offered there as well. So, when the reading program was offered to Block A prisoners, 80 of 200 prisoners were re-arrested; had the reading program not been offered, we estimate that x of 200 prisoners would have been re-arrested.

What if x was 80? This would mean that we estimate that the same number of Block A prisoners would have been re-arrested regardless of whether the reading program was offered. In other words, the reading program did not make a difference. The observed difference in recidivism rates between participants and non-participants was entirely an artifact of selection.

What if x was 120? We estimate that 120 Block A prisoners total would have been re-arrested, and we already know that 60 people who did not participate in the program were rearrested. So the other 60 people would have to come from the group of participants. We infer that, were it not for the reading program, 60 participants would have been arrested, but instead, because of the reading program, only 20 were. The difference we observed between participants and non-participants in Block A was entirely the result of the program.

What if x was 110? We estimate that 110 Block A prisoners total would have been re-arrested, and we already know that 60 non-participants were re-arrested. So the remaining 50 would have to come from the group of participants. We infer that, were it not for the reading program, 50 participants would have been arrested, but instead, because of the reading program, only 20 were. The program can therefore be credited with preventing 30 re-arrests; the remaining difference between participants and non-participants in Block A is an artifact of selection.

What are the key assumptions behind this inference?

1. That the recidivism rate for Block B (as a whole) can be used as a proxy for what would be the recidivism rate for Block A (as a whole) if there had been no reading program.
2. That the same number of Block A non-participants would have been re-arrested had no reading program not been offered (in other words, that the reading program had no effect—positive or negative—on the probability of recidivism for those who could have participated in it but did not).

Other points of note:

1. What we have done here is estimate the “treatment effect for the treated”; that is, the effect of the reading program for those who actually participated in the program. We might be interested in knowing how many of the 100 non-participants in the reading program would have been re-arrested had all prisoners been required to participate in the reading program. Using the approach outlined above, we can only get at this question (the “treatment effect for the non-treated”) if there was a Cell Block C where all prisoners were required to participate in the reading program.
2. In a real statistical analysis, we would be providing estimates of the uncertainty of the various estimates of the effect of the reading program above.