Loss of precision with IV

Steve Pizer, Austin Frakt

14 May 2013

Claim: All other things being equal, relative to the precision of the treatment effect determined with an RCT, one based on an IV analysis has a precision that is $1/ R^2$ larger, where R^2 is the coefficient of determination from a mean-adjusted version of the first stage.

Proof: Let Y be the outcome, T the treatment indicator, Z the instrument. All are vectors of length N. Assume n observations are assigned to treatment. Define $X = T - \langle T \rangle = T - p$, where $p = n/N$. Assume that

1. $Y = X\beta + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ and
2. $X = Z\gamma + \upsilon$, where $E[Z'\varepsilon|X] = 0$ and $\upsilon \sim \mathcal{N}(0, \mu^2 I)$.

Then we have the standard OLS and 2SLS results:

3. $\gamma_{ols} \sim \mathcal{N}((Z'Z)^{-1}Z'X, \mu^2(Z'Z)^{-1})$,
4. Define $P_Z = Z (Z'Z)^{-1}Z'$,
5. $\beta_{ols} \sim \mathcal{N}((X'X)^{-1}X'Y, \sigma^2(X'X)^{-1})$, and
6. $\beta_{IV} \sim \mathcal{N}((X'P_ZX)^{-1}X'P_ZY, \sigma^2(X'P_ZX)^{-1})$.

Let R^2 be the coefficient of determination for the estimation of the first stage (equation 2). Using the definition of R^2,

7. $R^2 = 1 - \upsilon' \upsilon / X'X = X'P_ZX / X'X$.

Therefore,

8. $R^2X'X = X'P_ZX$.

Since $X = T - p$, $X'X = n - 2pn + np = n(1-p) = Np(1-p)$. As a result, the variance of β_{IV} can be simplified to

9. $Var(\beta_{IV}) = \sigma^2(X'P_ZX)^{-1} = \sigma^2 / Np(1-p)R^2$.

Whereas the variance of β_{ols} is

10. $Var(\beta_{ols}) = \sigma^2(X'X)^{-1} = \sigma^2 / Np(1-p)$.

So, the effect of the IV is to increase the variance of the estimate of the treatment effect by a factor of

11. $Var(\beta_{IV}) / Var(\beta_{ols}) = 1/R^2$.

The same result is obtained by Wooldridge.[1]

[1] See page 511 of Wooldridge, J. M. (2009). Introductory econometrics: a modern approach. 4th edition. South-Western Pub.

Application: Let's use the above to compute the sample size needed to obtain 80% power for the point estimate of the change in proportion of the population with elevated glycated hemoglobin (>6.5%) from the Oregon Medicaid study.[2] Here are the key numbers from that study:

- Number on Medicaid:
  - 2,723 (=6,387 lottery winners x 42.64% compliance rate) plus
  - 1,081 (=5,842 lottery losers x 18.5% noncompliance rate)
  - Total on Medicaid, n: 3,804
- Number not on Medicaid:
  - 4,761 (=5,842 lottery losers x 81.5% compliance rate) plus
  - 3,664 (=6,387 lottery winners x 57.36% noncompliance rate)
  - Total not on Medicaid: 8,425
- Total sample, N: 3,804 + 8,425 = 12,229
- Mean baseline glycated hemoglobin (GH) > 6.5%: 5.1% (estimated from the 5,842 randomized to not Medicaid). Note that one could correct this for the fact that that some randomized to not Medicaid did not comply, as given above. We ignore this detail here.
- Point estimate of effect of Medicaid on GH > 6.5%: -0.93 percentage points (estimated with IV). That is, the point estimate of the GH > 6.5% rate for the Medicaid group is 4.17%.

From these, we can infer that the probability of being on Medicaid is p = 0.311 (=3,804/12.229), the probability of winning the lottery is q = 0.522 (=6,387/12,229), and the probability of winning the lottery and being on Medicaid is r = 0.223 (=2,723/12,229). The relative size of the Medicaid vs. non-Medicaid group is 0.452 (=3,804/8,425).[3]

Using Stata's sampsi or an online calculator,[4] we can infer that, had this been an RCT, the number randomized to Medicaid and control to obtain the effect size reported above at 80% power would have had to have been about 5,900 and 13,100, respectively, or about 1.55 times larger than the study was.

However, there was selected enrollment into Medicaid due to noncompliance (aka, crossover or contamination), which could have biased the results had this been run as a straight RCT. Consequently, the investigators instrumented for Medicaid enrollment with a variable that had been used to randomize participants into eligibility to apply for Medicaid (the lottery).

The data above are sufficient to run a simulation of the first stage to determine R^2. In fact, the setup is simple enough that we can derive a closed form expression for R^2 using the facts that (a) X'X and Z'Z are scalars, (b) they are zero mean versions of the Medicaid and randomization (lottery) indicators, and (c) r < q. Omitting some algebra:

12. $R^2 = (X'Z)^2 / [\ (X'X)(Z'Z)\ ] = [\ N(r-qp)\ ]^2 / [\ (Np(1-p))\ (Nq(1-q))\ ] = (r-qp)^2 / (p(1-p))(q(1-q))$.

[2] Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., ... & Finkelstein, A. N. (2013). The Oregon Experiment—Effects of Medicaid on Clinical Outcomes. New England Journal of Medicine, 368(18), 1713-1722. All figures come from Tables 1 and 2 of the paper and S9 of the supplementary appendix.

[3] Why not use the treatment and control group sizes as given by the randomization according to the lottery? Our answer is that, according to our proof, what matters is the contents of the vector X, not the vector Z. The former includes indicators of who actually did and did not receive the intervention. The latter is the randomization or instrument.

[4] http://statpages.org/proppowr.html

Note that if only lottery winners are on Medicaid, though not necessarily all of them (no crossover from the lottery losers group), then r = p and this result simplifies to p(1-q) / q(1-p), which you have to admit, is a pretty nice result. You can do the math yourself for the case r = q, everybody who wins the lottery enrolls in Medicaid, i.e., no crossover from the lottery winners group. And, finally, when there is no crossover from either group, r = p = q and $R^2 = 1$. How much fun is this?

Plugging in r, p, and q from above, we get $R^2 = 0.0678$, a result we also confirmed with a simulation in Stata. Consequently, the sample sizes needed for 80% power for the example above would have had to have been $1/ R^2 = 14.8$ times larger than 5,900 and 13,100 for the Medicaid and non-Medicaid groups, respectively, i.e., about 87,300 and 193,900, respectively.

For the elevated GH result only, the Oregon Medicaid study was underpowered by a factor of 14.8 x 1.55 = 22.9.

This analysis does not account for the study's survey weighting, which has the effect of reducing sample (see top of page 8 of the supplemental appendix to the study). Therefore, for the GH analysis, the study was underpowered by a factor of *at least* 22.9.