FLS 6415: Replication 4 - Instrumental Variables

April 2020

First read the paper by Albertson and Lawrence (2009) on the course website.

The replication data is in the file *IV.csv*, and the important variables are described below. The rest of the variables in the data are just control variables for use in Q10. We are only going to replicate 'Study 2' of the paper (details from page 290), where the research question is whether watching a TV program on affirmative action increased knowledge of and support for a proposition to *eliminate* affirmative action policies.

Table 1: Important Variables

Variable	Description
Watched_Program	The respondent actually watched the TV program
Encouraged_to_Watch	The respondent was encouraged to watch the TV program during the phone call
Info_Proposition	The respondent's self-assessment of whether they are well-informed about the Proposition

1. What is the treatment? What is the control? What is the outcome? What is the instrument for treatment?

2. If we did not know about instrumental variables, the basic *observational* regression we might run is to examine directly how the treatment variable is related to the outcome variable. Run this basic observational regression and interpret the results. *NB:* (The outcome variable is an ordered categorical variable - feel free to use either an ordered multinomial logit model or a simple linear OLS regression for all the questions, it doesn't make much difference to the results).

3. Do you trust the treatment effect estimates from Q2? What are the major threats to causal inference here? Provide concrete examples of why the estimate in Q2 might be wrong.

4. To conduct an instrumental variables analysis, we first need to make sure we have a strong 'first stage', i.e. that our instrument predicts our treatment variable. Using a simple regression, what is the evidence about the strength of our first stage?

5. Now let's perform the 2-Stage Least Squares instrumental variables methodology. First, save the fitted values of the first stage regression from Q4 as another column in your data.

6. Next, run the second-stage regression of the outcome variable on those fitted values from Q5. Carefully interpret the Instrumental Variables regression result.

7. Conduct the equivalent all-in-one IV approach to the previous analysis using *ivreg* in the *AER* library in R or *ivreg2* in Stata. Compare the result with the result from Q6.

8. A crucial assumption for the instrumental variables regression is the *exclusion restriction*: that the instrument ONLY affects the outcome through the treatment, and not through any other mechanism. We have to support this assumption by theory and supportive qualitative evidence as it cannot be empirically verified. Make the argument that the encouragement to watch the program through the phone call ONLY affects participants' information about the proposition through its effect on watching the program.

9. Now pretend you are a reviewer/critic and make the argument that the exclusion restriction assumption is likely to be *false*.

10. To what group of people ('population') does our estimate of the causal effect of treatment apply? How generalizable would our results be?

11. The authors' analysis in Table 4 is more complicated than ours only because it includes control variables in an attempt to make sure the instrument satisfies the exclusion restriction. Add the control variables to *both* the first and second stage 2SLS methodology regressions and interpret the results (it may still be slightly different from the values in Table 4).