AMERICAN UNIVERSITY - CENTRAL ASIA 05.12.2022; 08.12.2022
Econometrics I (ECO 608)
Nurgul Tilenbaeva

Seminar 8

Solutions to Endogeneity: Instrumental Variables Estimation

1. Consider a simple model to estimate the effect of personal computer (PC) ownership on college grade point average for graduating seniors at a large public university:

   $GPA = \beta_0 + \beta_1 PC + u,$

   where

   `PC`=binary variable indicating PC ownership.

   (a) Why might `PC` ownership be correlated with u?

   (b) Explain why `PC` is likely to be related to parents' annual income. Does this mean parental income is a good IV for `PC`? Why or why not?

   (c) Suppose that, four years ago, the university gave grants to buy computers to roughly one-half of the incoming students, and the students who received grants were randomly chosen. Carefully explain how you would use this information to construct an instrumental variable for `PC`.

2. Suppose that you wish to estimate the effect of class attendance on student performance. A basic model is

   $stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u,$

   where

   `stndfnl`=standardized outcome on a final exam;
   `atndrte`=percentage of classes attended;
   `priGPA`=prior college grade point average;
   `ACT`=ACT score.

   (a) Let `dist` be the distance from the students' living place to the lecture hall. Do you think `dist` is uncorrelated with u?

   (b) Assuming that `dist` and u are uncorrelated, what other assumption must `dist` satisfy to be a valid IV for `atndrte`?

3. In an article, Evans and Schwab (1995) studied the effects of attending a Catholic high school on the probability of attending college. For concreteness, let `college` be a binary variable equal to unity if a student attends college, and zero otherwise. Let `CathHS` be a binary variable equal to one if the student attends a Catholic high school. A linear probability model is

   $college = \beta_0 + \beta_1 CathHS + otherfactors + u,$

   where the other factors include gender, race, family income, and parental education.

   (a) Why might `CathHS` be correlated with u?

   (b) Evans and Schwab have data on a standardized test score taken when each student was a sophomore. What can be done with this variable to improve the ceteris paribus estimate of attending a Catholic high school?

   (c) Let `CathRel` be a binary variable equal to one if the student is Catholic. Discuss the two requirements needed for this to be a valid IV for `CathHS` in the preceding equation. Which of these can be tested?

   (d) Not surprisingly, being Catholic had a significant positive effect on attending a Catholic high school. Do you think `CathRel` is a convincing instrument for `CathHS`?

4. Import the Stata data file `"wage2"` from the e-course platform. This data set contains information on monthly earnings, education, several demographic variables, and IQ scores for 935 men in 1980.

   (a) Estimate the regression model by OLS:

   $log(wage) = \beta_0 + \beta_1 educ + u$

   What is the estimated return for another year of education?

   (b) Why might `educ` be correlated with u? What is the potential source of endogeneity?

   (c) The variable `sibs` is the number of siblings. Explain why `educ` and `sibs` might be correlated, and what is the expected sign of the correlation? Regress `educ` on `sibs` to determine whether there is

a statistically significant correlation, and whether the sign of the correlation is consistent with your expectations.

(d) Use `sibs` as an IV for `educ`. Report and interpret the results.

(e) The variable `brthord` is birth order (`brthord` is one for a first-born child, two for a second-born child, and so on). Explain why `educ` and `brthord` might be negatively correlated. Regress `educ` on `brthord` to determine whether there is a statistically significant negative correlation.

(f) Use `brthord` as an IV for `educ`. Report and interpret the results.

(g) Now, suppose that we include number of siblings as an explanatory variable in the wage equation; this controls for family background, to some extent:

$log(wage) = \beta_0 + \beta_1 educ + \beta_2 sibs + u.$

Suppose that we want to use `brthord` as an IV for `educ`, assuming that `sibs` is exogenous. The reduced form for `educ` is

$educ = \pi_0 + \pi_1 sibs + \pi_2 brthord + v.$

State and test the identification assumption.

(h) Estimate the equation from part (g) using `brthord` as an IV for `educ`. Report and interpret the results.

5. Import the Stata data file `"fertil2"` from the e-course platform. These data include, for women in Botswana during 1988, information on number of children, years of education, age, and religious and economic status variables.

(a) Estimate the model

$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$

by OLS, and interpret the estimates. In particular, holding `age` fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

(b) The variable `frsthalf` is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that `frsthalf` is uncorrelated with the error term from part (a),

show that `frsthalf` is a reasonable IV candidate for `educ`. (*Hint:* You need to do a regression.)

(c) Estimate the model from part (a) by using `frsthalf` as an IV for `educ`. Compare the estimated effect of education with the OLS estimate from part (a).