R300 – Advanced Econometric Methods

PROBLEM SET 6 - QUESTIONS

Posted on Fri. November 15 Due on Tue. November 26, 2018 at noon

1. Take the simple instrumental-variable model

$$y_i = x_i\beta + v_i$$
$$x_i = z_i\pi + u_i$$

where

$$\begin{pmatrix} v_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \rho\sigma_v\sigma_u \\ \rho\sigma_v\sigma_u & \sigma_u^2 \end{pmatrix} \right)$$

is independent of z_i. Note that both x_i and z_i are scalars. Assume for simplicity that z_i is zero mean.

(i) Derive the conditional mean function $E(y_i|x_i, u_i)$.

(ii) Under what conditions on the joint distribution of (v_i, u_i) would a least-squares regression of y_i on x_i yield a consistent estimator of β?

(ii) Suppose that you would observe u_i. How would you estimate β based on your result under (i)?

2. Continue on with the setup from the previous question.

(i) Still supposing that you observe u_i, could you use your result from 1.(i) to construct a formal test for endogeneity? Explain how you would do so and present the testing procedure (including statement of the null hypothesis, the test statistic used, and its limit distribution under the null).

(ii) If you do not observe u_i you could replace it by the residual from a least-squares regression of x_i on z_i. Show that, when doing so, your estimator of β under (ii) equals 2SLS.

3. Consider a situation where

$$y_i = g(x_i, v_i);$$

the function g is unknown, x_i is binary (0,1) and endogenous, and we have instrument z_i that is also binary (0,1). Suppose that

$$x_i = \{h(z_i) \geq u_i\},$$

where the function h is strictly increasing (but otherwise unknown). You may assume that the unobservables (v_i, u_i) are independent of z_i.

We can think about x_i as participation to a job-training program and z_i as a variable that makes participation to that program easier, such as an exemption from a participation fee, for example.

(i) Show that, if x_i would be exogenous (that is, if u_i and v_i would be independent), then the ordinary least-squares estimator of the slope in a regression of y_i on x_i (and a constant) estimates

$$E(y_i|x_i = 1) - E(y_i|x_i = 0),$$

i.e., the average treatment effect.

(ii) Show that the instrumental-variable estimator of the slope parameter in a linear model of y_i on x_i (and a constant) estimates

$$\frac{E(y_i|z_i = 1) - E(y_i|z_i = 0)}{E(x_i|z_i = 1) - E(x_i|z_i = 0)}.$$

4. Continue on with the setup from the previous question.

(i) Show that the estimand in 3.(ii) can be written as

$$\frac{\int_{h(0)}^{h(1)} \left(E(y_i|x_i = 1, u_i = u) - E(y_i|x_i = 0, u_i = u)\right) f(u)\, du}{P(h(0) \leq u_i \leq h(1))},$$

where $f(u)$ is the density of u_i at u.

(ii) Give a precise interpretation of

$$E(y_i|x_i = 1, u_i = u) - E(y_i|x_i = 0, u_i = u).$$

(iii) Can you give an interpretation to the instrumental-variable estimand? To do so think about how individuals would change x_i as a function of z_i. Would everyone change his participation decision to the program if the participation fee is waived or imposed?

(iv) What happens to the estimand as $|h(1) - h(0)|$ becomes larger? Explain.