R300 – Advanced Econometric Methods

PROBLEM SET 6 - SOLUTIONS

Posted on Fri. November 15 Due on Mon. November 23, 2020

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)?

(i) By normality of the errors,

$$v_i|u_i \sim N\left(\rho\frac{\sigma_v}{\sigma_u}u_i, (1-\rho^2)\,\sigma_v^2\right),$$

and so

$$E(y_i|x_i, u_i) = x_i\beta + \rho\frac{\sigma_v}{\sigma_u}u_i$$

or, equivalently,

$$y_i = x_i\beta + \rho\frac{\sigma_v}{\sigma_u}u_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, (1-\rho^2)\,\sigma_v^2).$$

(ii) When $\rho = 0$.

(iii) The last displayed equation under (i) above is a (classical) linear regression model, as ε_i is independent of both regressors x_i and u_i. Assuming u_i to be observable in the data, the least-squares estimator of β and $\gamma := \rho\sigma_v/\sigma_u$ is simply

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \sum_i x_i^2 & \sum_i x_i u_i \\ \sum_i x_i u_i & \sum_i u_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i x_i y_i \\ \sum_i u_i y_i \end{pmatrix}$$
$$= \frac{1}{\sum_i x_i^2 \sum_i u_i^2 - (\sum_i x_i u_i)^2} \begin{pmatrix} \sum_i u_i^2 & -\sum_i x_i u_i \\ -\sum_i x_i u_i & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \sum_i x_i y_i \\ \sum_i u_i y_i \end{pmatrix}.$$

If we let $\hat{r}_i$ be the residual of a regression of x_i on u_i, then this yields

$$\hat{\beta} = \frac{\sum_i u_i^2 \sum_i x_i y_i - \sum_i x_i u_i \sum_i u_i y_i}{\sum_i x_i^2 \sum_i u_i^2 - (\sum_i x_i u_i)^2} = \frac{\sum_i \hat{r}_i y_i}{\sum_i \hat{r}_i x_i} = \frac{\sum_i \hat{r}_i y_i}{\sum_i \hat{r}_i^2}.$$

This is, of course, nothing else than standard multiple regression, and the Frish-Waugh theorem in particular. The interesting part of this application here is that conditioning on u_i, in fact, solves the endogeneity of x_i in the above equation. This is so as, conditional on u_i the remaining variation in x_i comes from the instrumental variable z_i, which is exogenous (as it is independent of v_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.

(i) In the same way as above we can write the least-squares estimator of γ as

$$\hat{\gamma} = \frac{\sum_i \hat{q}_i y_i}{\sum_i \hat{q}_i^2}$$

where $\hat{q}_i$ is the residual of a regression of u_i on x_i. As $\gamma = \rho\sigma_v/\sigma_u$, the null of exogeneity ($\rho = 0$) is equivalent to testing the null $H_0 : \gamma = 0$ against the alternative of endogeneity, which is $H_1 : \gamma \neq 0$. Here we can do this by a standard two-sided 't-test' (i.e., Likelihood-ratio test); we reject the null if $|\hat{\gamma}/\text{se}(\hat{\gamma})| > t_{\alpha/2;n-2}$.

(ii) A least-squares fit decomposes x_i as

$$x_i = \hat{x}_i + \hat{u}_i, \quad \text{(fit + residual)}$$

where the components in this decomposition are uncorrelated. Our two-step approach obtained by regressing y_i on x_i and $\hat{u}_i$ gives

$$\frac{\sum_i \tilde{r}_i y_i}{\sum_i \tilde{r}_i^2}$$

where $\tilde{r}_i$ are the residuals from a regression of x_i on $\hat{u}_i$. Now, $x_i = \hat{x}_i + \hat{u}_i$ and so

$$\tilde{r}_i = x_i - \frac{\sum_j x_j \hat{u}_j}{\sum_j \hat{u}_j^2}\,\hat{u}_i = \hat{x}_i + \hat{u}_i - \frac{\sum_j x_j \hat{u}_j}{\sum_j \hat{u}_j^2}\,\hat{u}_i = \hat{x}_i + \hat{u}_i\left(1 - \frac{\sum_j x_j \hat{u}_j}{\sum_j \hat{u}_j^2}\right) = \hat{x}_i,$$

where the last transition follows form the fact that

$$\frac{\sum_i x_i \hat{u}_i}{\sum_i \hat{u}_i^2} = \frac{\sum_i (\hat{x}_i + \hat{u}_i)\hat{u}_i}{\sum_i \hat{u}_i^2} = \frac{\sum_i \hat{u}_i^2}{\sum_i \hat{u}_i^2} = 1.$$

Thus,

$$\frac{\sum_i \tilde{r}_i y_i}{\sum_i \tilde{r}_i^2} = \frac{\sum_i \hat{x}_i y_i}{\sum_i \hat{x}_i^2},$$

which is the familiar expression for two-stage least-squares.

---

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

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

(i) When x_i is exogenous we can saturate our model and write

$$E(y_i|x_i) = \alpha + x_i\beta.$$

A least-squares regression will estimate β and, indeed,

$$\beta = E(y_i|x_i = 1) - E(y_i|x_i = 0)$$

as the right-hand side is simply $(\alpha + \beta) - \alpha$.

(ii) Straightforward calculations (using the binary nature of the variables) give

$$\text{cov}(z_i, y_i) = \{E(y_i|z_i = 1) - E(y_i|z_i = 0)\}\; P(z_i = 1)\, P(z_i = 0)$$
$$\text{cov}(z_i, x_i) = \{E(x_i|z_i = 1) - E(x_i|z_i = 0)\}\; P(z_i = 1)\, P(z_i = 0)$$

and so the instrumental-variable estimand is

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

as claimed.

---

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.

(i) The denominator equals

$$E(x_i|z_i = 1) - E(x_i|z_i = 0) = P(u_i \leq h(1)) - P(u_i \leq h(0)) = P(h(0) \leq u_i \leq h(1)).$$

For the numerator, note that

$$E(y_i|z_i = 1) = E(y_i|x_i = 1, u_i \leq h(1))\, P(u_i \leq h(1)) + E(y_i|x_i = 0, u_i > h(1))\, P(u_i > h(1)),$$

and, similarly,

$$E(y_i|z_i = 0) = E(y_i|x_i = 1, u_i \leq h(0))\, P(u_i \leq h(0)) + E(y_i|x_i = 0, u_i > h(0))\, P(u_i > h(0)).$$

Now,

$$\begin{aligned} E(y_i|x_i = 1, u_i \leq h(1)) &= \int_{-\infty}^{+\infty} E(y_i|x_i = 1, u_i)\, f(u_i|u_i \leq h(1))\, du_i \\ &= \frac{\int_{-\infty}^{h(1)} E(y_i|x_i = 1, u_i)\, f(u_i)\, du_i}{P(u_i \leq h(1))}, \\ E(y_i|x_i = 1, u_i \leq h(0)) &= \int_{-\infty}^{+\infty} E(y_i|x_i = 1, u_i)\, f(u_i|u_i \leq h(0))\, du_i \\ &= \frac{\int_{-\infty}^{h(0)} E(y_i|x_i = 1, u_i)\, f(u_i)\, du_i}{P(u_i \leq h(0))}, \end{aligned}$$

and

$$\begin{aligned} E(y_i|x_i = 0, u_i > h(1)) &= \int_{-\infty}^{+\infty} E(y_i|x_i = 0, u_i)\, f(u_i|u_i > h(1))\, du_i \\ &= \frac{\int_{h(1)}^{+\infty} E(y_i|x_i = 0, u_i)\, f(u_i)\, du_i}{P(u_i > h(1))}, \\ E(y_i|x_i = 0, u_i > h(0)) &= \int_{-\infty}^{+\infty} E(y_i|x_i = 0, u_i)\, f(u_i|u_i > h(0))\, du_i \\ &= \frac{\int_{h(0)}^{+\infty} E(y_i|x_i = 0, u_i)\, f(u_i)\, du_i}{P(u_i > h(0))}. \end{aligned}$$

Therefore, the numerator of the IV estimand becomes just

$$\int_{h(0)}^{h(1)} (E(y_i|x_i = 1, u_i) - E(y_i|x_i = 0, u_i))\, f(u_i)\, du_i,$$

and the estimand itself is

$$\frac{\int_{h(0)}^{h(1)} (E(y_i|x_i = 1, u_i) - E(y_i|x_i = 0, u_i))\, f(u_i)\, du_i}{P(h(0) \leq u_i \leq h(1))}.$$

Note that this equals

$$E(y_i|x_i = 1, h(0) < u_i \leq h(1)) - E(y_i|x_i = 0, h(0) < u_i \leq h(1)).$$

(ii) The difference

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

is the average treatment effect for an individual whose (first-stage) unobserved heterogeneity equals u. Call this the marginal treatment effect; $\gamma(u)$, say.

(iii) The instrumental-variable estimand is

$$E(\gamma(u_i)|u_i \in [h(0), h(1)]),$$

which is an average treatment effect for the subpopulation of individuals whose unobserved heterogeneity lies in the interval $[h(0), h(1)]$. Individuals for whom $u_i \in (\infty, h(0))$ would have $x_i = 1$ for both $z_i = 0$ and $z_i = 1$; they would participate to the program no matter the tuition waiver. Individuals for whom $u_i \in (h(1), +\infty)$ would have $x_i = 0$ for both $z_i = 0$ and $z_i = 1$; they would not participate to the program even if the tuition was waived. The last group of individuals, then, has $u_i \in [h(0), h(1)]$. These individuals participate if sponsored but not otherwise; i.e., the instrument affects their behavior. It is exactly for this subpopulation of units that we can identify the average treatment effect.

(iv) What happens to the estimand as $|h(1) - h(0)|$ becomes larger? Explain. All units will react to changes in the instrument and

$$E(\gamma(u_i)|u_i \in [h(0), h(1)]) \rightarrow E(\gamma(u_i)),$$

the average treatment effect for the entire population.