R300 – Advanced Econometric Methods

PROBLEM SET 7 - SOLUTIONS

Due on Mon. November 30

1. Consider the linear instrumental-variable model (in matrix notation)

$$\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{\varepsilon}$$

with more instruments then covariates. You may assume that $\boldsymbol{\varepsilon}$ is homoskedastic throughout.

You want to test the null hypothesis that $\theta = 0$ against the two-sided alternative that $\theta \neq 0$.

Set up the Wald, LR-type, and LM-type test statistics for this null and show that they are all numerically equivalent to each other here.

Let $\hat{\theta}$ be the unconstrainted optimal GMM estimator which equals

$$\hat{\theta} = (\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{y}), \qquad \boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z},'$$

that is, 2SLS. For the sequel, also define the matrix that projects on $\boldsymbol{P_Z X}$, i.e.,

$$\boldsymbol{P}_{\boldsymbol{P_Z X}} = \boldsymbol{P_Z X}(\boldsymbol{X}'\boldsymbol{P_Z X})^{-1}\boldsymbol{X}'\boldsymbol{P_Z},$$

and

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\theta}, \qquad \hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/n,$$

that is, the residuals and the residual variance.

(i) The Wald statistic is

$$\frac{\hat{\theta}'(\boldsymbol{X}'\boldsymbol{P_Z X})\,\hat{\theta}}{\hat{\sigma}^2} = \frac{\boldsymbol{y}'\boldsymbol{P_Z X}(\boldsymbol{X}'\boldsymbol{P_Z X})^{-1}\boldsymbol{X}'\boldsymbol{P_Z y}}{\hat{\sigma}^2} = \frac{\boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{P_Z X}}\boldsymbol{y}}{\hat{\sigma}^2}.$$

(ii) The LR statistic is

$$\frac{\boldsymbol{y}'\boldsymbol{P_Z}\boldsymbol{y}}{\hat{\sigma}^2} - \frac{\hat{\varepsilon}'\boldsymbol{P_Z}\hat{\varepsilon}}{\hat{\sigma}^2} = \frac{(\boldsymbol{X}\hat{\theta}+\hat{\varepsilon})'\boldsymbol{P_Z}(\boldsymbol{X}\hat{\theta}+\hat{\varepsilon})}{\hat{\sigma}^2} - \frac{\hat{\varepsilon}'\boldsymbol{P_Z}\hat{\varepsilon}}{\hat{\sigma}^2}$$
$$= \frac{\hat{\theta}'\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X}\hat{\theta} + 2\,\hat{\theta}'\boldsymbol{X}'\boldsymbol{P_Z}\hat{\varepsilon}}{\hat{\sigma}^2}$$
$$= \frac{\hat{\theta}'\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X}\hat{\theta} + 2\,\hat{\theta}'\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{y} - 2\,\hat{\theta}'\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X}\hat{\theta}}{\hat{\sigma}^2}$$
$$= \frac{\boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{P_Z X}}\boldsymbol{y}}{\hat{\sigma}^2}.$$

(iii) The LM statistic is

$$\frac{\boldsymbol{y}'\boldsymbol{P_Z}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{y}}{\hat{\sigma}^2} = \frac{\boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{P_Z X}}\boldsymbol{y}}{\hat{\sigma}^2}.$$

So, indeed, all three statistics are the same.

---

2. Consider the simple binary-choice model

$$y = \{x_i\beta \geq v_i\},$$

where x_i is a scalar continuous regressor. The complication is that x_i is not independent of u_i. Moreover, we have

$$x_i = z_i\gamma + u_i,$$

and

$$\begin{pmatrix} v_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_u \\ \rho\sigma_u & \sigma_u^2 \end{pmatrix} \right),$$

and these errors are independent of $z_i \sim N(0,1)$.

(i) Derive an expression for $E(y_i|x_i,u_i)$.

(ii) Suppose that you would observe u_i in the data. How could you use your answer to (i) to estimate the parameters of the model?

(iii) Derive an expression for the linear instrumental-variable estimand in this model. That is, compute $\text{cov}(y_i,z_i)/\text{cov}(x_i,z_i)$. Is this a meaningful quantity to estimate? In answering this you may find it useful to know that

$$\int_{-\infty}^{+\infty} x\,\Phi(a+bx)\,\phi(x)\,dx = \tfrac{b}{\sqrt{1+b^2}}\,\phi\left(\tfrac{a}{\sqrt{1+b^2}}\right), \qquad \int_{-\infty}^{+\infty}\Phi(a+bx)\,\phi(x)\,dx = \Phi\left(\tfrac{a}{\sqrt{1+b^2}}\right),$$

for constants a and b.

(i) By joint normality of the errors,

$$v_i|u_i \sim N(\rho/\sigma_u u_i, (1-\rho^2)).$$

Hence,

$$P(y_i = 1|x_i, u_i) = P\left(v_i \leq x_i\beta|\, x_i, u_i\right) = \Phi\left(\frac{x_i\beta - \rho/\sigma_u u_i}{\sqrt{1-\rho^2}}\right).$$

(ii) If u_i is observed the likelihood function would be

$$\prod_i \Phi\left(\frac{x_i\beta - \rho/\sigma_u u_i}{\sqrt{1-\rho^2}}\right)^{y_i} \left(1 - \Phi\left(\frac{x_i\beta - \rho/\sigma_u u_i}{\sqrt{1-\rho^2}}\right)\right)^{1-y_i},$$

which can be maximized with respect to the parameters to get an efficient estimator.

In practice, a feasible version leads to a two-step procedure where, first, we estimate u_i by the residual of a least-squares regression of x_i on z_i (say $\hat{u}_i$) and, second, proceed as before with $\hat{u}_i/\hat{\sigma}_u$ replacing u_i/σ_u. (Note that this replacement requires an adjustment to the usual standard errors in order to yield valid inference!) This estimator is frequently referred to as the Rivers-Vuong estimator. It is an example of what is known as a control-function estimator. Two other examples of such a procedure are Heckman's sample-selection estimator (Heckit or Heckman in Stata depending on your version) and, yes, two-stage least squares. In Stata the command for the Rivers-Vuong estimator is called (somewhat unfortunately) ivprobit.

(iii) We need to calculate

$$\text{cov}(z_i, y_i) = E\left(z_i\,\Phi\left(\frac{z_i(\gamma\beta)}{\sqrt{1-\rho^2+(\beta\sigma_u-\rho)^2}}\right)\right) = \frac{\gamma\beta}{\sqrt{1-\rho^2+(\beta\sigma_u-\rho)^2+(\gamma\beta)^2}}\,\phi(0)$$

and

$$\text{cov}(z_i, x_i) = \gamma.$$

The IV estimand is, therefore,

$$\beta_{IV} = \frac{\beta}{\sqrt{1-\rho^2+(\beta\sigma_u-\rho)^2+(\gamma\beta)^2}}\,\phi(0).$$

This quantity does not have an obvious meaningful interpretation. When x_i is independent of v_i (when $\rho = 0$) it would (at least in this design) co-incide with the average marginal effect.

The marginal effect we would be interested in would an expected change for an exogenous change in x_i. Because x_i is endogenous this is not $\partial E(y_i|x_i)/\partial x_i$. Conditional on u_i, however, variation in $x_i = x$ is exogenous and so we could consider recovering the average marginal effect for a given x_i as

$$\int \left. \frac{\partial E(y_i|x_i,u)}{\partial x_i} \right|_{x_i=x} \phi(u/\sigma_i)/\sigma_u \, du$$

and the average of these (over the regressor) as

$$E\left(\int \left. \frac{\partial E(y_i|x_i,u)}{\partial x_i} \right|_{x_i=x} \phi(u/\sigma_i)/\sigma_u \, du \right).$$

With some additional calculation efforts we can verify that this average marginal effect is

$$\frac{\beta}{\sqrt{1+\beta^2(\gamma^2+\sigma_u^2)}} \phi(0).$$

---

3. Now suppose that

$$y_i = \beta_0 + x_i\beta_1 + v_i, \qquad x_i = \begin{cases} 1 & \text{if } z_i\gamma \geq u_i \\ 0 & \text{if } z_i\gamma < u_i \end{cases},$$

where

$$\begin{pmatrix} v_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \rho\sigma_v \\ \rho\sigma_v & 1 \end{pmatrix} \right),$$

and these errors are independent of $z_i \sim N(0,1)$. We observe a random sample on (y_i, x_i, z_i).

(i) Show that x_i is endogenous unless $\rho = 0$ by computing $E(x_i v_i)$. In answering this you may again find it useful to know that

$$\int_{-\infty}^{+\infty} x\, \Phi(a+bx)\, \phi(x)\, dx = \tfrac{b}{\sqrt{1+b^2}}\, \phi\left(\tfrac{a}{\sqrt{1+b^2}}\right), \qquad \int_{-\infty}^{+\infty} \Phi(a+bx)\, \phi(x)\, dx = \Phi\left(\tfrac{a}{\sqrt{1+b^2}}\right),$$

for constants a and b.

(ii) Derive conditions on the parameters in the model for z_i to be a relevant instrument.

(iii) Discuss how the strength of the instrument varies as a function of the parameters of the model.

(i) We have

$$
\begin{aligned}
E(x_i v_i) &= E(\{z_i\gamma \geq u_i\}\, v_i) \\
&= E\left(\Phi\left(\frac{z_i\gamma - \rho/\sigma_v v_i}{\sqrt{1-\rho^2}}\right)\, v_i\right) \\
&= E\left(\int_{\infty}^{+\infty} (v/\sigma_v)\, \Phi\left(\frac{z_i\gamma}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}}\, v/\sigma_v\right)\, \phi(v/\sigma_v)\, dv\right) \\
&= -\rho\sigma_v \int_{-\infty}^{+\infty} \phi(z\gamma)\, \phi(z)\, dz \\
&= -\frac{\rho\sigma_v}{\sqrt{1+\gamma^2}}\, \phi(0).
\end{aligned}
$$

The severity of the endogeneity as a function of the parameters ρ, σ_v, and γ is intuitive. Stronger correlation between the errors and a smaller signal-to-noise ratio in y_i both increase $E(x_i v_i)$ (in magnitude). A stronger signal to noise ratio in x_i on the other hand reduces $E(x_i v_i)$ (in magnitude).

(ii) We look at the reduced form

$$x_i = \delta_0 + z_i\delta_1 + \varepsilon_i,$$

where

$$\delta_1 = \frac{\text{cov}(z_i, x_i)}{\text{var}(z_i)} = E(z_i x_i),$$

with the last transition being a consequences of $z_i \sim N(0,1)$. We calculate

$$\delta_1 = E(z_i\{z_i\gamma \geq u_i\}) = E\left(z_i\, \Phi(z_i\gamma)\right) = \textstyle\int_{-\infty}^{+\infty} z\Phi(z\gamma)\, \phi(z)\, dz = \dfrac{\gamma}{\sqrt{1+\gamma^2}}\, \phi(0).$$

So, $\delta_1 \neq 0$ is equivalent to $\gamma \neq 0$. This is intuitive. It states that z_i will be a relevant instrument if it affects x_i. This can be tested by comparing the t-statistic for the null that $\delta_1 = 0$ (against the alternative $H_1 : \delta_1 \neq 0$) against the quantiles of the standard normal distribution.

Validity of the instrument follows from the fact that it is excluded from the structural equation for y_i and it is independent of v_i. Because we only have one instrument here the parameters are just identified and we cannot formally test this exclusion restriction on z_i.

(iii) The function $\gamma/\sqrt{1+\gamma^2}$ is increasing in $|\gamma|$ and symmetric around zero. Its lower and upper limits are -1 and 1, respectively. Hence, δ_1 will reach its maximum magnitude as $|\gamma| \to \infty$. The relation between γ and δ_1 is again intuitive. Note also that the DGP restricts $\delta_1 \in [-1, 1]$. This makes sense, as larger γ increase the correlation between x_i and $\text{sign}(z_i)$.

The standard deviation of δ_1 when estimated is

$$n^{-1/2}\, \sigma_\varepsilon / \sigma_z = \sigma_\varepsilon / \sqrt{n}.$$

The concentration parameter is

$$\pi = \delta_1^2 \sum_i z_i^2 / \sigma_\varepsilon^2 \overset{a}{\sim} \delta_1^2 / (\sigma_\varepsilon^2 / n) = \delta_1^2 / \text{var}(\hat{\delta}_1).$$

So for a given δ_1 value the instrument will be stronger when the signal to noise ratio in x is larger or (equivalently) when δ_1 is estimated more precisely.

---

4. Continue with the setup from the previous question.

(i) Your colleague who has not taken this course says that you should not use 2SLS here. His argument is that, because the variable x_i is discrete, a linear probability model that decomposes $x_i = \hat{x}_i + \hat{\varepsilon}_i$ by least squares gives an incorrect specification of $E(x_i|z_i)$ and so the resulting 2SLS estimator,

$$\frac{\sum_i (\hat{x}_i - \overline{x})(y_i - \overline{y})}{\sum_i (\hat{x}_i - \overline{x})^2}$$

will be inconsistent. Do you agree? Explain.

(ii) Our model implies that

$$E(v_i|z_i) = 0.$$

Derive the optimal moment condition implied by this to estimate β.

(iii) In the previous question you provided a function $\varphi(z_i)$ for which

$$E(\varphi(z_i)\, v_i) = 0.$$

In practice you will have to estimate this function φ. How would you proceed here?

(i) It is correct that the conditional mean function $E(x_i|z_i)$ must be nonlinear in z_i in this case. However, 2SLS does not require you to estimate this conditional mean. Indeed, 2SLS requires you to decompose x_i into two blocks: one ($\hat{x}_i$) being a function of z_i that contains the exogenous variation in x_i, the other an orthogonal part ($\hat{\varepsilon}_i$) that contains the endogenous variation in x_i. This is exactly what least squares does for you independent of the nature of x_i.

(ii) The optimal unconditional moment condition (up to sign) here is

$$E(\varphi(z_i)\, v_i) = 0, \qquad \varphi(z_i) = \begin{pmatrix} 1 \\ E(x_i|z_i) \end{pmatrix} E(v_i^2|z_i)^{-1}.$$

Given that $E(x_i|z_i) = \Phi(z_i\gamma)$ and that v_i is independent of z_i with variance σ_v^2 the optimal moment condition becomes

$$E\left(\begin{pmatrix} 1 \\ \Phi(z_i\gamma) \end{pmatrix} \frac{(y_i - \beta_0 - x_i\beta_1)}{\sigma_v^2}\right) = 0$$

The covariance matrix of these moment conditions is

$$\begin{pmatrix} 1 & E(\Phi(z_i\gamma))/\sigma_v^2 \\ E(\Phi(z_i\gamma))/\sigma_v^2 & E(\Phi(z_i\gamma)^2)/\sigma_v^2 \end{pmatrix}$$

and the asymptotic variance of the optimal estimator equals the inverse of this matrix.

(iii) We proceed in multiple steps:

1. Estimate a probit model (by maximum likelihood) for x_i given z_i to get an estimate of γ, say $\hat{\gamma}$.

2. Run an instrumental-variable estimator with instruments 1 and $\Phi(z_i\hat{\gamma})$—or, indeed, any other valid instrument; any transformation of z_i would suffice here—to obtain a first-step consistent estimator of $\beta = (\beta_0, \beta_1)'$ and use these to construct the residuals $\hat{v}_i$.

3. Estimate σ_v^2 by the sample variance of the $\hat{v}_i$ and run a new instrumental-variable estimator, now using instruments $1/\hat{\sigma_v^2}$ and $\Phi(z_i\hat{\gamma})/\hat{\sigma}_v^2$.
