# R300 – Advanced Econometric Methods

**PROBLEM SET 7 - QUESTIONS**

**Due on Mon. November 30, 2020**

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.

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 $b$.

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

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

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.

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 $\beta$.

(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 $\varphi$. How would you proceed here?