Exogeneity cannot be determined from the joint distribution of observables

DA Freedman, Statistics, UCB 15 November 2004

The linear probability model

Lat ξ, ζ be independent random variables, with mean 0. (To match Schneider et al, ξ may be taken as two-valued.) Let

$$X = \alpha\xi + \beta\zeta, \quad U = \gamma\xi + \zeta.$$

The response variable Y in a linear probability model is 0 or 1 with

$$P(Y = 1|\xi, \zeta) = a + bX + U.$$

The observables are ξ, X, Y. The parameters are $a, b, \alpha, \beta, \gamma$ and the distribution of ζ. For the model to hold, the random variables ξ, ζ must be bounded, and the parameters must be restricted to suitably narrow intervals.

Since X will be endogenous if $\beta \neq 0$, the idea is to use ξ as an instrument. However, the exogeneity of ξ cannot be determinued from the joint distribution of observables. The joint distribution of ξ, X identifies α and the distribution of $\beta\zeta$. We must now consider

$$\begin{aligned} P(Y = 1|\xi, X) &= a + bX + E(U|\xi, X) \\ &= a + bX + \gamma\xi + E(\zeta|\xi, \zeta) \\ &= a + bX + \gamma\xi + \zeta \\ &= a + bX + \gamma\xi + \frac{X - \alpha\xi}{\beta} \\ &= a + \Big(\gamma - \frac{\alpha}{\beta}\Big)\xi + \Big(b + \frac{1}{\beta}\Big)X. \end{aligned}$$

The identifiable parameters are therefore

$$\alpha, \quad a, \quad \gamma - \frac{\alpha}{\beta}, \quad b + \frac{1}{\beta},$$

and the distribution of $\beta\zeta$. But ξ is exogenous iff $\gamma = 0$. The crucial parameter γ is not separately identifiable, nor is $\gamma = 0$.

The probit model

The same construction can be used for the probit. Suppose ξ, ζ are independent normal variables, with mean 0 and respective variances σ^2, τ^2. Let

$$X = \alpha\xi + \beta\zeta, \quad U = \gamma\xi + \zeta.$$

The response variable Y is

$$1 \text{ if } a + bX + U > 0, \text{ else } 0.$$

The observables are ξ, X, Y. The parameters are $a, b, \alpha, \beta, \gamma, \sigma^2, \tau^2$. Here, X is endogenous unless $\beta = 0$. We want to use ξ as an instrument for X, but ξ is endogenous unless $\gamma = 0$.

For simplicity, suppose $\beta \geq 0$. We get $\alpha, \sigma, \beta\tau$ from the joint distribution of ξ, X. We must now compute

$$P(Y = 1|\xi, X) = P(a + bX + U > 0|\xi, X). \tag{1}$$

By the previous argument, $P(Y = 1|\xi, X) = 1$ if

$$a + \Big(\gamma - \frac{\alpha}{\beta}\Big)\xi + \Big(b + \frac{1}{\beta}\Big)X > 0;$$

otherwise, the conditional probability is 0. For simplicity, suppose $a \neq 0$. Then (1) identifies

$$\text{sign}(a), \quad \frac{\gamma}{a} - \frac{\alpha}{a\beta}, \quad \frac{b}{a} + \frac{1}{a\beta}.$$

In both examples,

$$U = \frac{X - \alpha\xi}{\beta} + \gamma\xi = \Big(\gamma - \frac{\alpha}{\beta}\Big)\xi + \frac{X}{\beta}$$

is computable from ξ, X. However, the computation uses the parameters α, β, γ. Therefore, from a statistical perspective, U remains latent. In the probit model, it may be more attractive to introduce yet another independent normal variable η, with mean 0 and (say) known variance 1, setting $U = \gamma\xi + \zeta + \eta$. Then (1) is

$$\Phi\left(a + \Big(\gamma - \frac{\alpha}{\beta}\Big)\xi + \Big(b + \frac{1}{\beta}\Big)X\right),$$

where Φ is the standard normal. The identifiable parameters are then

$$\alpha, \quad \sigma, \quad \beta\tau, \quad a, \quad \gamma - \frac{\alpha}{\beta}, \quad b + \frac{1}{\beta}.$$

We need τ so that β itself is not identifiable: if α, β are identifiable, so is γ. For similar results in the regression model, see exercise 8E5 in *Statistical Models*.