Exam #2

Economics 435: Quantitative Methods

Fall 2008

1 Dynamic panel data models

Suppose that we have a panel data set are estimating the following model:

$$y_{i,t} = a_i + \beta_1 y_{i,t-1} + u_{i,t} \tag{1}$$

where a_i is an individual-specific fixed effect, $y_{i,t}$ is the outcome variable of interest, and $u_{i,t}$ is a period-specific unobserved factor. We assume that the individuals in the data set are a random sample and that $y_{i,t}$ has positive variance both within and across individuals. This is going to require:

$$var(u_{i,t}) > 0 \tag{2}$$

This model differs from the standard panel data model in that one of the explanatory variables (in this case, the only one) is the previous value of the outcome variable. A model of this type is called a *dynamic* panel data model. Usually such models contain additional explanatory variables and period fixed effects, but I will leave those out for simplicity.

a) In most panel data models we would assume strict exogeneity. In this case strict exogeneity means:

$$E(u_{i,t}|a_i, y_{i,1}, y_{i,2}, \ldots, y_{i,T}) = 0$$

Prove[1] that strict exogeneity cannot hold in this model, given conditions (1) and (2).

b) Although strict exogeneity is inconsistent with this model, it is possible to make a weaker assumption that is logically consistent with the model and allows for identification of β_1. Assume for the rest of this question that:

$$E(u_{i,t}|a_i, y_{i,1}, \ldots, y_{i,t-1}) = 0 \tag{3}$$

Prove that under assumptions (1), (2), and (3), the following estimator will consistently estimate β_1:

$$\hat{\beta}_1^{IV} = \frac{c\hat{o}v(\Delta y_{i,t}, y_{i,t-2})}{c\hat{o}v(\Delta y_{i,t-1}, y_{i,t-2})}$$

This estimator corresponds to the IV regression of $\Delta y_{i,t}$ on $\Delta y_{i,t-1}$ using $y_{i,t-2}$ as an instrument.

c) Prove that under assumptions (1), (2), and (3), the standard first differences regression

$$\hat{\beta}_1^{FD} = \frac{c\hat{o}v(\Delta y_{i,t}, \Delta y_{i,t-1})}{v\hat{a}r(\Delta y_{i,t-1})}$$

will not consistently estimate β_1.

[1] A few hints:

1. Strict exogeneity implies that $u_{i,t}$ is uncorrelated (has zero covariance) with a_i and all of the y's.
2. Use this result to prove that strict exogeneity implies both that $cov(u_{i,t}, y_{i,t}) = 0$ and $cov(u_{i,t}, y_{i,t}) \neq 0$.

2 Instrumental variables and local average treatment effects

Suppose we are using a scholarship experiment to measure the effect of university attendance on wages.

We have a random sample on (y, p, z) where: $z_i = I(\text{received scholarship})$, $p_i = I(\text{attended university})$, and y_i is subsequent wages.

Let $p_i(z)$ and $y_i(p)$ be the potential outcome functions for university attendance and wages[2] respectively.

The scholarship is assigned by a random mechanism:

$$\Pr(z_i = 1 | p_i(0), p_i(1), y_i(0), y_i(1)) = \Pr(z_i = 1) = q$$

The choice of whether to attend university is not random. Each individual falls into one of four categories:

1. "Always-takers" go to university whether or not they receive the scholarship, i.e., $p_i(0) = p_i(1) = 1$.
2. "Never-takers" never go to university, i.e., $p_i(0) = p_i(1) = 0$.
3. "Compliers" go to university only if they receive the scholarship, i.e., $p_i(0) = 0, p_i(1) = 1$.
4. "Defiers" go to university only if they don't receive the scholarship $p_i(0) = 1, p_i(1) = 0$.

We assume that no one is a defier, i.e., that $\Pr(defier) = 0$.

We are interested in the treatment effect of university attendance on wages, i.e.:

$$TE_i = y_i(1) - y_i(0)$$

a) Which of the following quantities can be identified in the data?

1. $E(p|z = 1)$ and $E(p|z = 0)$
2. $E(y|z = 1)$ and $E(y|z = 0)$.
3. $E(y|p = 1)$ and $E(y|p = 0)$.

b) Find $\Pr(\textit{complier})$ in terms of those quantities listed in part (a) that can be identified in the data.

c) Prove that

$$E(TE|\textit{complier}) = \frac{E(y|z = 1) - E(y|z = 0)}{E(p|z = 1) - E(p|z = 0)}$$

This quantity is known as the "local average treatment effect" or LATE.

d) Is $E(TE|\textit{complier})$ identified?

e) Are $E(TE|never - taker)$, $E(TE|always - taker)$, or $E(TE)$ identified?

f) Prove that

$$\frac{cov(z, y)}{cov(z, p)} = \frac{E(y|z = 1) - E(y|z = 0)}{E(p|z = 1) - E(p|z = 0)}$$

g) Given the results above, prove that the IV regression coefficient:

$$\hat{\beta}^{IV} \equiv \frac{\hat{cov}(y, z)}{\hat{cov}(p, z)}$$

(where $\hat{cov}$ is the usual consistent estimator of the covariance) is a consistent estimator of the local average treatment effect.

[2] Note that by writing the potential wage as $y_i(p)$ I am assuming that the scholarship's only effect on wages is through its effect on attendance. This rules out, for example, the possibility that scholarship-receiving students use their improved financial position to reduce work hours and spend more time studying; or to spend more time at the pub.