Exam #2

Economics 435: Quantitative Methods

Fall 2010

Please answer the question I ask - no more and no less - and remember that the correct answer is often short and simple.

1 Short answer questions

a) For each of these statements, indicate whether the statement is true or false.

1. If x and y are independent, then $cov(x, y) = 0$.
2. If x and y are independent, then $E(y|x) = E(y)$.
3. If $cov(x, y) = 0$ then x and y are independent.
4. If $cov(x, y) = 0$ then $E(y|x) = E(y)$.
5. If $E(y|x) = E(y)$ then $cov(x, y) = 0$.
6. If $E(y|x) = E(y)$, then x and y are independent.

No need to prove, just identify which statements are true and which are false.

b) Let:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Which of the following problems make the OLS estimator of β_1 inconsistent?

1. Substantial correlation between x_1 and x_2.
2. Nonzero correlation between x_1 and u.
3. Heteroskedasticity of u.
4. Homoskedasticity of u.
5. Classical measurement error in x_1.
6. Classical measurement error in y.

7. A sampling probability that depends on x_1.
8. A sampling probability that depends on y.

c) Suppose that a is independent of both b and c. Prove that:

$$cov(ab, c) = E(a)cov(b, c)$$

Note that this result will be useful for part (d) of question (3).

2 Plausibly exogenous instruments

The standard theory justifying the use of instrumental variables to estimate structural or causal models requires the assumption that we are certain that the instruments are exogenous.

In practice, researchers often use instruments that are only what some have called "plausibly exogenous." That is, we have no particular reason to believe the instruments are not exogenous, so we hope any deviation from exogeneity is probably small enough to ignore.

We will analyze this situation in a simple setting. Suppose we have a random sample of size n on a scalar outcome y_i, a scalar explanatory variable x_i, and a scalar instrument z_i. Suppose that our structural model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where β_1 has a causal interpretation and so x_i and u_i are potentially correlated. We have a candidate instrumental variable z_i that is relevant:

$$cov(x_i, z_i) \neq 0$$

but not necessarily exogenous.

Use the following notation in answering this question:

$$\begin{aligned}
\rho_{x,u} &= corr(x_i, u_i) \\
\rho_{z,u} &= corr(z_i, u_i) \\
\rho_{x,z} &= corr(x_i, z_i) \\
\sigma_u^2 &= var(u_i) \\
\sigma_x^2 &= var(x_i) \\
\sigma_z^2 &= var(z_i)
\end{aligned}$$

To keep the problem simple, assume $\rho_{x,u} \geq 0$, $\rho_{z,u} \geq 0$, and $0 \leq \rho_{x,z} < 1$.

a) Let

$$\hat{\beta}_1^{OLS} = \frac{\hat{cov}(x_i, y_i)}{\hat{var}(x_i)}$$

be the coefficient from the OLS regression of y_i on x_i. Find plim $(\hat{\beta}_1^{OLS} - \beta_1)$ in terms of $(\rho_{x,u}, \rho_{z,u}, \rho_{x,z}, \sigma_u, \sigma_x, \sigma_z)$.

b) Let

$$\hat{\beta}_1^{IV} = \frac{c\hat{o}v(z_i, y_i)}{c\hat{o}v(z_i, x_i)}$$

be the coefficient from the IV regression. Find plim $\hat{\beta}_1^{IV} - \beta_1$ in terms of $(\rho_{x,u}, \rho_{z,u}, \rho_{x,z}, \sigma_u, \sigma_x, \sigma_z)$.

c) Find a condition on $(\rho_{x,u}, \rho_{z,u}, \rho_{x,z})$ under which the inconsistency (i.e. the difference between the probability limit of an estimator and the quantity it is estimating) of IV is less than that of OLS.

d) Although $\rho_{x,u}$ and $\rho_{z,u}$ are not identified, we can estimate $\rho_{x,z}$. Suppose you estimate of $\rho_{x,z}$ is about 0.2, and you would like to put the following sentence into your paper:

> "Our IV regression will be inconsistent if our instrument fails to be exogenous, but it will be less inconsistent than the OLS regression as long as the instrument's correlation with unobservables is less than ________ times the explanatory variable's correlation with unobservables"

Fill in the blank.

3 Treatment effects in panel data

Suppose we have a panel data set constructed from a random sample of individuals indexed by $i = 1, 2, \ldots, n$ observed at a fixed set of points in time $t = 1, 2, \ldots, T$. We observe an outcome of interest y_{it} and a binary treatment $x_{it} \in \{0, 1\}$.

Suppose that y_{it} is determined by the potential outcome function $y_{it}^p(x)$. That is, $y_{it}^p(1)$ is the outcome person i would have experienced in time period t if he received the treatment (i.e., if x_{it} were equal to one), and $y_{it}^p(1)$ is the outcome he would have experienced if not (i.e., if x_{it} were equal to zero). The actual outcome is:

$$y_{it} = y_{it}^p(x_{it})$$

the treatment effect for person i in time period t is:

$$TE_{it} = y_{it}^p(1) - y_{it}^p(0)$$

and the average treatment effect is:

$$ATE = E(TE_{it})$$

a) The standard fixed effects model is:

$$y_{it} = a_i + \beta x_{it} + u_{it} \tag{1}$$

$$E(u_{it}|x_{i1}, \ldots, x_{iT}) = 0 \tag{2}$$

where β is interpreted as effect of x on y. Find $y_{it}^p(1)$, $y_{it}^p(0)$, TE_{it} and ATE for this model.

b) Which of the following implicit assumptions about potential outcomes are being made by the standard fixed effects model?

1. The treatment is unrelated to unobserved outcome-relevant factors.
2. Changes in treatment status are unrelated to changes in unobserved outcome-relevant factors.
3. The treatment effect is constant across individuals and time.

c) Now suppose that treatment effects are individual-specific. That is, instead of (1) and (2) assume:

$$y_{it}^p(x) = a_i + \beta x + u_{it}^p(x) \tag{3}$$
$$E(u_{it}^p(x)|a_i, x_{i1}, \ldots, x_{iT}) = 0 \tag{4}$$

Find TE_{it} and ATE for this model.

d) In addition to (3) and (4) above, assume that

$$u_{it}^p(x) \text{ is independent of } (a_i, x_{i1}, \ldots, x_{iT}) \tag{5}$$

Let:

$$\hat{\beta} = \frac{c\hat{o}v(\Delta x_{it}, \Delta y_{it})}{v\hat{a}r(\Delta x_{it})} \rightarrow^p \frac{cov(\Delta x_{it}, \Delta y_{it})}{var(\Delta x_{it})}$$

be the usual first-difference (FD) estimator of the effect of x_{it} on y_{it}. Prove that $\hat{\beta} \rightarrow^p ATE$.