Exam #2

Economics 435: Quantitative Methods

Fall 2009

1 The fixed effects IV estimator

Suppose that we have a panel data set (i.e., a random sample of individuals indexed with $i = 1, 2, \ldots, n$ observed for time periods $t = 1, 2, \ldots, T$) with data on (x_{it}, y_{it}, z_{it}) for each individual and time period. We are interested in estimating the fixed effects model:

$$y_{it} = a_i + \beta_1 x_{it} + u_{it} \tag{1}$$

where a_i is an individual-level fixed effect and β_1 is our parameter of interest. For simplicity, this specification does not include any time fixed effects.

Suppose we are unwilling in this case to believe the conventional strict exogeneity assumption on x_{it}. That is we think that changes in x_{it} over time may be correlated with changes in u_{it}. However, we have an instrumental variable z_{it} whose changes are correlated with changes in x_{it} but uncorrelated with changes in u_{it}. That is, it obeys the (strict) exogeneity condition:

$$E(u_{it}|z_{i1}, z_{i2}, \ldots, z_{iT}, a_i) = 0 \tag{2}$$

as well as the relevance condition:

$$cov(\Delta x_{it}, \Delta z_{i,t}) \neq 0 \tag{3}$$

where $\Delta x_{it} = x_{it} - x_{it-1}$, etc.

a) Let:

$$\hat{\beta}_1^{FDIV} \equiv \frac{c\hat{o}v(\Delta y_{it}, \Delta z_{i,t})}{c\hat{o}v(\Delta x_{it}, \Delta z_{i,t})}$$

Prove that under assumptions (1)-(3), $\hat{\beta}_1^{FDIV}$ is a consistent estimator of β_1.

b) Suppose we are also willing to assume that:

$$cov(z_{it}, a_i) = 0 \tag{4}$$

Let $\hat{\beta}_1^{IV}$ be the conventional IV estimator:

$$\hat{\beta}_1^{IV} \equiv \frac{c\hat{o}v(y_{it}, z_{i,t})}{c\hat{o}v(x_{it}, z_{i,t})}$$

Prove that under assumptions (1)-(4), $\hat{\beta}_1^{IV}$ is a consistent estimator of β_1.

c) Suppose that you are trying to estimate the elasticity of demand for gasoline, and that you have state-level panel data on log consumption (y_{it}), log prices (x_{it}), and log taxes (per liter) (z_{it}). *Using plain language*, describe what we need to assume in order to consistently estimate the elasticity of gasoline demand using $\hat{\beta}_1^{FDIV}$.

d) *Using plain language*, describe what we need to assume in order to consistently estimate the elasticity of demand using $\hat{\beta}_1^{IV}$.

2 Partial effects in the logit model

Suppose that we have a random sample of data on $(y, x_1, \ldots, x_k)$ where y is a binary outcome variable. We would like to estimate a logit model:

$$\Pr(y = 1|x_1, \ldots, x_k) = \Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

where

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

is the CDF of the logistic distribution.

However, instead of the logit coefficients themselves, we would like to estimate partial effects for x_1. We define the partial effect for a particular $(x_1, \ldots, x_k)$ as:

$$PE(x_1, \ldots, x_k) = \frac{\partial \Pr(y = 1|x_1, \ldots, x_k)}{\partial x_1}$$

and the partial effect at the average as the PE evaluated at the sample average of each explanatory variable:

$$PEA = PE(\bar{x}_1, \ldots, \bar{x}_k)$$

If this notation or setup is unclear to you please feel free to ask for clarification.

a) Let $\lambda(z)$ be the PDF of the logistic distribution. Find $\lambda(z)$.

b) Find the PE as a function of $(\beta_0, \beta_1, \ldots, \beta_k)$ and $(x_1, \ldots, x_k)$. You can leave $\lambda(.)$ as part of your answer.

c) Find the PEA as a function of $(\beta_0, \beta_1, \ldots, \beta_k)$ and $(\bar{x}_1, \ldots, \bar{x}_k)$. You can leave $\lambda(.)$ as part of your answer.

d) As your answer to the previous question suggests, it will not be possible to estimate the PEA directly from the regression output. Fortunately, there is a simple way to set up the regression so that the PEA is easily estimated from the regression output. Let:

$$\begin{aligned} \tilde{x}_1 &= x_1 - \bar{x}_1 \\ &\vdots \\ \tilde{x}_k &= x_k - \bar{x}_k \end{aligned}$$

Show that there exists a $\tilde{\beta}_0$ such that

$$\Pr(y = 1|x_1, \ldots, x_k) = \Lambda(\tilde{\beta}_0 + \beta_1 \tilde{x}_1 + \cdots + \beta_k \tilde{x}_k)$$

The easiest way to do this is to solve for $\tilde{\beta}_0$.

e) Find the PEA as a function of $(\tilde{\beta}_0, \beta_1)$ and $\lambda(.)$.

3 Combining samples with population data

Sometimes we can address the issue of nonrandom sampling by combining the information in our sample with information on the population from other sources.

Suppose we are interested in seeing how the probability of imprisonment varies with whether a person has completed high school. Our model is:

$$prison = \beta_0 + \beta_1 grad + u$$

where *prison* is a binary indicator of whether the person is currently in prison, and *grad* is a binary indicator of whether the person completed high school.

We have two data sources on our population of interest (adult residents of Canada). The first data source is from the Census, and provides the exact proportion of individuals in the population that are in prison

$$\mu_p = \Pr(prison = 1)$$

and the exact proportion that have completed high school

$$\mu_g = \Pr(grad = 1)$$

Note that the Census is a complete enumeration of the population of interest, so μ_p and μ_g should be treated as fixed quantities and not random variables.

The second data source is a random sample of n prisoners (i.e., $prison = 1$ for every person in the sample). Let $\overline{grad}$ be the proportion of prisoners in our data who have completed high school.

a) Briefly (two sentences at most) explain why we need both data sets to estimate β_1.

b) Let

$$\hat{\beta}_1 = \overline{grad}\frac{\mu_p}{\mu_g(1-\mu_g)} - \frac{\mu_p}{1-\mu_g}$$

Show that this is a consistent estimator of β_1.

c) Is $\hat{\beta}_1$ also unbiased? Prove it.

d) Find the standard deviation of $\hat{\beta}_1$.

e) Find a consistent estimator of the standard deviation of $\hat{\beta}_1$.