Exam #2 Answer Key

Economics 435: Quantitative Methods

Fall 2011

1 Short answers

a) Hopefully you've done this a few times before, but here it is again:

$$\begin{aligned}
E(u) &= E(E(u|x)) \qquad \text{(law of iterated expectations)} \\
&= E(0) \qquad \text{(given)} \\
&= 0 \\
E(xu) &= E(E(xu|x)) \qquad \text{(law of iterated expectations)} \\
&= E(xE(u|x)) \qquad \text{(conditioning)} \\
&= E(x0) \qquad \text{(given)} \\
&= 0 \\
cov(x,u) &= E(xu) - E(x)E(u) \qquad \text{(a result we use all the time)} \\
&= 0 - E(x)0 \\
&= 0
\end{aligned}$$

b)

$$\begin{aligned}
\text{plim } e^{\bar{x}} &= e^{\text{plim } \bar{x}} \qquad \text{(Slutsky theorem)} \\
&= e^{\mu} \qquad \text{(law of large numbers)}
\end{aligned}$$

c) If $g(\theta) = e^{\theta}$, then the rules of differentiation imply that $g'(\theta) = e^{\theta}$ as well. So:

$$\sqrt{n}(e^{\bar{x}} - e^{\mu}) \quad \rightarrow^{D} \quad N(0, e^{2\mu}\sigma^2)$$

d) The Central Limit Theorem tells us that the sample average of a large random sample is approximately normally distributed. This is useful for performing inference (hypothesis tests and confidence intervals) because inference requires us to know the probability distribution of our statistics - most of the time the finite-sample distribution is intractable, so having a convenient approximation is very helpful.

2 The 2011 Census

a) We now have a random sample of size $n = NSR$, so

$$var(\bar{y}) = \frac{\sigma_y^2}{NSR}$$

b)

$$\begin{aligned}
\sqrt{\frac{var(\bar{y}; S = 0.333333, R = 0.5)}{var(\bar{y}; S = 0.2, R = 0.95)}} &= \sqrt{\frac{\frac{\sigma_y^2}{N \times 0.333 \times 0.5}}{\frac{\sigma_y^2}{N \times 0.2 \times 0.95}}} \\
&= \sqrt{\frac{0.19}{0.1667}} \\
&= \sqrt{1.14} \\
&= 1.07
\end{aligned}$$

It raises the variance by about 14%, and the standard error by about 7%.

c) No.

d) Since $E(y_i|r_i = 0) \in [a, b]$, we have

$$E(y_i|r_i = 1)R + a(1 - R) \leq E(y_i) \leq E(y_i|r_i = 1)R + b(1 - R)$$

e) The lower bound would be:

$$\bar{y}R + a(1 - R)$$

and the upper bound would be:

$$\bar{y}R + b(1 - R)$$

f) The percentage lies between $0.10 * 0.95$ and $0.10 * 0.95 + 0.05$, or 9.5% and 14.5%.

g) The percentage lies between $0.10 * 0.5$ and $0.10 * 0.5 + 0.5$ or between 5% and 55%.

h)

$$\begin{aligned}
\theta &= E(y_i|x_i = 1) - E(y_i|x_i = 0) \\
&= (E(y_i|x_i = 1, r_i = 1) \Pr(r_i = 1) + E(y_i|x_i = 1, r_i = 0) \Pr(r_i = 0)) \\
&\quad - (E(y_i|x_i = 0, r_i = 1) \Pr(r_i = 1) + E(y_i|x_i = 0, r_i = 0) \Pr(r_i = 0)) \\
&= (E(y_i|x_i = 1, r_i = 1) - E(y_i|x_i = 1, r_i = 1)) \Pr(r_i = 1) \\
&\quad + (E(y_i|x_i = 1, r_i = 0) - E(y_i|x_i = 1, r_i = 0)) \Pr(r_i = 0) \\
&= \beta_1 R + (E(y_i|x_i = 1, r_i = 0) - E(y_i|x_i = 1, r_i = 0)) (1 - R)
\end{aligned}$$

We can determine that $(E(y_i|x_i = 1, r_i = 0) - E(y_i|x_i = 1, r_i = 0))$ lies between $(0 - 1) = -1$ and $(1 - 0) = 1$, so

$$\beta_1 R - (1 - R) \leq \theta \leq \beta_1 R + (1 - R)$$

i) The bounds would be:

$$\hat{\beta}_1 R - (1-R) \leq \theta \leq \hat{\beta}_1 R + (1-R)$$

j)

$$-0.05 = (0 \times 0.95) - (1 - 0.95) \leq \theta \leq (0 \times 0.95) + (1 - 0.95) = 0.05$$

So being in a female-headed household may raise or lower the probability of living in poverty by as much as 5%.

k)

$$-0.5 = (0 \times 0.5) - (1 - 0.5) \leq \theta \leq (0 \times 0.5) + (1 - 0.5) = 0.5$$

So being in a female-headed household may raise or lower the probability of living in poverty by as much as 50%.

3 Cluster data with fixed effects

a) First we show:

$$\begin{aligned}
\bar{y}_s &= \frac{1}{n_s} \sum_{i:s(i)=s} y_i \\
&= \frac{1}{n_s} \sum_{i:s(i)=s} a_{s(i)} + \beta_1 FDK_{s(i)t(i)} + \beta_2 x_i + u_i \\
&= \frac{1}{n_s} \sum_{i:s(i)=s} a_s + \beta_1 \frac{1}{n_s} \sum_{i:s(i)=s} FDK_{st(i)} + \frac{1}{n_s} \sum_{i:s(i)=s} \beta_2 x_i + \frac{1}{n_s} \sum_{i:s(i)=s} u_i \\
&= a_s + \beta_1 \overline{FDK}_s + \beta_2 \bar{x}_s + \bar{u}_s \\
y_i - \bar{y}_{s(i)} &= \left(a_{s(i)} + \beta_1 FDK_{s(i)t(i)} + \beta_2 x_i + u_i\right) - \left(a_{s(i)} + \beta_1 \overline{FDK}_{s(i)} + \beta_2 \bar{x}_{s(i)} + \bar{u}_{s(i)}\right) \\
&= \beta_1 (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}) + \beta_2 (x_i - \bar{x}_{s(i)}) + (u_i - \bar{u}_{s(i)})
\end{aligned}$$

Then we show:

$$
\begin{aligned}
& E\left(u_i - \bar{u}_{s(i)} | (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)})\right) \\
= \; & E\left(E\left(u_i - \bar{u}_{s(i)} \middle| a_{s(i)}, \left\{FDK_{s(i)t}\right\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s(i)} \right) \middle| (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)}) \right) \\
& \quad \text{(by the Law of Iterated Expectations)} \\
= \; & E\left(E\left(u_i - \frac{1}{n_{s(i)}} \sum_{j:s(j)=s(i)} u_j \middle| a_{s(i)}, \left\{FDK_{s(i)t}\right\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s(i)} \right) \middle| (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)}) \right) \\
& \quad \text{(by substitution)} \\
= \; & E\left(E\left(u_i \middle| a_{s(i)}, \left\{FDK_{s(i)t}\right\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s(i)} \right) \middle| (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)}) \right) \\
- \; & \frac{1}{n_{s(i)}} \sum_{j:s(j)=s(i)} E\left(E\left(u_j \middle| a_{s(i)}, \left\{FDK_{s(i)t}\right\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s(i)} \right) \middle| (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)}) \right) \\
& \quad \text{(by linearity of the expected value)} \\
= \; & E\left(0 | (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)})\right) - \frac{1}{n_{s(i)}} \sum_{j:s(j)=s(i)} E\left(0 | (FDK_{s(i)t(i)} - \overline{FDK}_{s(i)}), (x_i - \bar{x}_{s(i)})\right) \\
& \quad \text{(by strict exogeneity)} \\
= \; & 0 - \frac{1}{n_{s(i)}} \sum_{j:s(j)=s(i)} 0 \\
= \; & 0
\end{aligned}
$$

b) First we show:

$$
\begin{aligned}
\bar{y}_{st} &= \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} y_i \\
&= \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} a_{s(i)} + \beta_1 FDK_{s(i)t(i)} + \beta_2 x_i + u_i \\
&= \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} a_s + \beta_1 \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} FDK_{st} + \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} \beta_2 x_i + \frac{1}{n_{st}} \sum_{i:s(i)=s,t(i)=t} u_i \\
&= a_s + \beta_1 FDK_{st} + \beta_2 \bar{x}_{st} + \bar{u}_{st} \\
\bar{y}_{st} - \bar{y}_s &= (a_s + \beta_1 FDK_{st} + \beta_2 \bar{x}_{st} + \bar{u}_{st}) - \left(a_s + \beta_1 \overline{FDK}_s + \beta_2 \bar{x}_s + \bar{u}_s\right) \\
&= \beta_1 (FDK_{st} - \overline{FDK}_s) + \beta_2 (\bar{x}_{st} - \bar{x}_s) + (\bar{u}_{st} - \bar{u}_s)
\end{aligned}
$$

Then we show:

$$
\begin{aligned}
& E\left(\bar{u}_{st} - \bar{u}_s | (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s)\right) \\
= \; & E\left(E\left(\bar{u}_{st} - \bar{u}_s \middle| \, a_s, \{FDK_{st}\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s} \right) \middle| \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
& \qquad \text{(by the Law of Iterated Expectations)} \\
= \; & E\left(E\left(\frac{1}{n_{st}} \sum_{j:s(j)=s,t(j)=t} u_j - \frac{1}{n_s} \sum_{j:s(j)=s} u_j \middle| \, a_s, \{FDK_{st}\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s} \right) \middle| \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
& \qquad \text{(by substitution)} \\
= \; & \frac{1}{n_{st}} \sum_{j:s(j)=s,t(j)=t} E\left(E\left(u_j \middle| \, a_s, \{FDK_{st}\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s} \right) \middle| \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
- \; & \frac{1}{n_s} \sum_{j:s(j)=s} E\left(E\left(u_j \middle| \, a_s, \{FDK_{st}\}_{t=2009}^{2011}, \{x_j\}_{j:s(j)=s} \right) \middle| \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
& \qquad \text{(by linearity of the expected value)} \\
= \; & \frac{1}{n_{st}} \sum_{j:s(j)=s,t(j)=t} E\left(0 | \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
- \; & \frac{1}{n_s} \sum_{j:s(j)=s} E\left(0 | \, (FDK_{st} - \overline{FDK}_s), (\bar{x}_{st} - \bar{x}_s) \right) \\
& \qquad \text{(by strict exogeneity)} \\
= \; & \left(\frac{1}{n_{st}} \sum_{j:s(j)=s,t(j)=t} 0 \right) - \left(\frac{1}{n_s} \sum_{j:s(j)=s} 0 \right) \\
= \; & 0
\end{aligned}
$$

c) Estimate the OLS regression of $(y_i - \bar{y}_{s(i)})$ on $(FDK_{s(i)t(i)} - \overline{FDK}_{s(i)})$ and $(x_i - \bar{x}_{s(i)})$.

d) Estimate the OLS regression of $(y_{st} - \bar{y}_s)$ on $(FDK_{st} - \overline{FDK}_s)$ and $(\bar{x}_{st} - \bar{x}_s)$.