Econometrics 140B
Fall 2017 Midterm 2

1. What are the 2 requirements of a valid instrument?

   Answer: The 2 requirements of a valid instrument are that $Cov(x_i, z_i \neq 0$ and $Cov(z_i, \varepsilon_i) - 0$ where x_i is the endogenous regressor, z_i is the instrument, and ε_i is the error term.

2. For a set of data, you first estimate the model

$$earnings = \beta_0 + \beta_1 education + u,$$

   and obtain the OLS estimate of β_1. You then obtain a valid instrument for education and use it to obtain the 2SLS estimate of β_1. Carefully explain which of these estimates you think will be larger. How will the standard errors of the estimates differ?

   Answer: Assuming education is correlated with higher earnings and education is positively correlated with u because higher education is correlated with higher ability and therefore higher earnings. Then β_1^{OLS} will be larger because it is biased upwards and 2SLS removes that bias. The standard errors of 2SLS will be larger.

3. For the model

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

   assume that $\mathbb{E}(u_i|x_i) = 0$ and $\mathbb{E}(u_i) = 0$.

   (a) Derive the OLS estimator of β_1.

Answer:

$$
\begin{aligned}
\widehat{\beta}^{OLS} &= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (y - x\beta_1 - \beta_0)^2 \\
\frac{\partial}{\partial \beta_0} &= \sum (-2)(y - x\beta_1 - \beta_0) = 0 \\
&\implies \sum y - \sum x\beta_1 - \sum \beta_0 = 0 \\
&\implies \sum y - \sum x\beta_1 = \sum \beta_0 \\
&\implies \sum y - \sum x\beta_1 = n\beta_0 \\
&\implies \frac{1}{n}\sum y - \frac{1}{n}\sum x\beta_1 = \beta_0 \\
&\implies \overline{y} - \beta_1\overline{x} = \beta_0 \\
\frac{\partial}{\partial \beta_1} &= \sum (-2x)(y - x\beta_1 - \beta_0) = 0 \\
&\implies \sum yx - \sum x^2\beta_1 - \sum x\beta_0 = 0 \\
&\implies \sum yx = \beta_1 \sum x^2 - \beta_0 \sum x \\
&\implies \sum yx = \beta_1 \sum x^2 - (\overline{y} - \beta_1\overline{x}) \sum x \\
&\implies \sum yx - \overline{y}\sum x = \beta_1 \left(\sum x^2 - \overline{x} \sum x \right) \\
&\implies \widehat{\beta}_1^{OLS} = \frac{\sum yx - \overline{y}\sum x}{\sum x^2 - \overline{x}\sum x} \\
&\vdots \\
&\implies \widehat{\beta}_1^{OLS} = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sum (x - \overline{x})^2}
\end{aligned}
$$

(b) Under the assumption that $\mathbb{E}[u_i|x] = 0$, prove that the OLS estimator of β_1 is unbiased.

Answer:

$$
\begin{aligned}
\mathbb{E}[\widehat{\beta}_1] &= \mathbb{E}\left[\frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\frac{\sum(x-\overline{x})(\beta_0+\beta_1x_i+u_i-\frac{1}{n}\sum(\beta_0+\beta_1x_i+u_i))}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\frac{\sum(x-\overline{x})(\beta_0+\beta_1x_i+u_i-\frac{1}{n}\sum\beta_0-\frac{1}{n}\sum\beta_1x_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\frac{\sum(x-\overline{x})(\beta_0+\beta_1x_i+u_i-\beta_0-\beta_1\overline{x}-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\frac{\sum(x-\overline{x})(\beta_1(x_i-\overline{x})+u_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\frac{\beta_1\sum(x-\overline{x})^2}{\sum(x-\overline{x})^2}+\frac{\sum(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}\left[\beta_1+\frac{\sum(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \mathbb{E}[\beta_1]+\mathbb{E}\left[\frac{\sum(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\right] \\
&= \beta_1+\mathbb{E}\left[\frac{\sum(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)}{\sum(x-\overline{x})^2}\middle| x_i\right] \\
&= \beta_1+\frac{1}{\sum(x-\overline{x})^2}\mathbb{E}\left[\sum(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)\middle| x_i\right] \\
&= \beta_1+\frac{1}{\sum(x-\overline{x})^2}\sum\mathbb{E}\left[(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)\middle| x_i\right] \\
&= \beta_1+\frac{1}{\sum(x-\overline{x})^2}\sum(x-\overline{x})\cancelto{0\ \textbf{by assumption}}{\mathbb{E}\left[(u_i-\frac{1}{n}\sum u_i)\middle| x_i\right]} \\
&= \beta_1 \qquad \therefore \widehat{\beta}_1 \textbf{ is an unbiased estimator of } \beta_1.
\end{aligned}
$$

(c) Under the assumption that $\mathbb{E}[u_i|x] \neq 0$, derive the bias of the OLS estimator. What is the direction of the bias if $\beta_1 > 0$ and $Cov(x_i, u_i) < 0$?

Answer: From part (b):

$$
\begin{aligned}
\mathbb{E}[\widehat{\beta}_1] &= \beta_1+\frac{1}{\sum(x-\overline{x})^2}\sum\mathbb{E}\left[(x-\overline{x})(u_i-\frac{1}{n}\sum u_i)\middle| x_i\right] \\
&\approx \beta_1+\frac{Cov(x_i,u_i)}{Var(x_i)}
\end{aligned}
$$

Therefore, $\widehat{\beta}_1$ is biased downwards.