Midterm Solutions

ECON 210 Econometrics A

Evan Zuofu Liao*

October 29, 2015

Question 1

(a) Let's try to find the expected value of $\bar{X}_n \bar{Y}_n$

$$
\begin{aligned}
\mathbb{E}\left[\bar{X}_n \bar{Y}_n\right] &= \mathbb{E}\left[\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)\Big(\frac{1}{n}\sum_{i=1}^{n} Y_i\Big)\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\Big(\sum_{i=1}^{n} X_i\Big)\Big(\sum_{i=1}^{n} Y_i\Big)\right] \\
&= \frac{1}{n^2}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[X_i Y_j\right]\Big) \\
&= \frac{1}{n^2}\Big(\sum_{i=j}\mathbb{E}\left[X_i Y_j\right] + \sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left[X_i Y_j\right]\Big) \\
&= \frac{1}{n^2}\Big(n\mathbb{E}\left[XY\right] + \sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left[X_i\right]\mathbb{E}\left[Y_j\right]\Big) \\
&= \frac{1}{n^2}\Big(n\mathbb{E}\left[XY\right] + n(n-1)\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]\Big) \\
&= \frac{1}{n^2}\Big(n^2\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] + n\mathrm{Cov}\left(X,Y\right)\Big) \\
&= \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] + \frac{1}{n}\mathrm{Cov}\left(X,Y\right)
\end{aligned}
$$

Now the question reduces to check whether or not $\mathrm{Cov}(X,Y) = 0$. Therefore, if the covariance of X and Y is zero, $\bar{X}_n \bar{Y}_n$ is an unbiased estimator of $\mathbb{E}[X]\mathbb{E}[Y]$. Otherwise we cannot conclude that the estimator is unbiased.

*Comments and questions to evanliao@uchicago.edu.

(b) Using the Continuous Mapping Theorem, we know that

$$\left.\begin{array}{r}\bar{X}_n \xrightarrow{p} \mathbb{E}[X] \\ \bar{Y}_n \xrightarrow{p} \mathbb{E}[Y]\end{array}\right\} \xrightarrow{CMT} g(\bar{X}_n, \bar{Y}_n) \xrightarrow{p} g(\mathbb{E}[X], \mathbb{E}[Y])$$

where $g : \mathbb{R}^2 \to \mathbb{R}$ is defined such that $g(a, b) = ab$

Question 2

(a) We can take advantage of the hint and prove that $\frac{\mathrm{Cov}[Y,D]}{\mathrm{Var}[D]} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$. This is exactly the same property that we have proved in the first problem set. For details of the proof, see problem set 1 solutions.

Alternatively, we can show

$$\begin{aligned}\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] &= \mathbb{E}[\beta_0 + \beta_1 D + U|D=1] - \mathbb{E}[\beta_0 + \beta_1 D + U|D=0] \\ &= \beta_0 + \beta_1 + \mathbb{E}[U|D=1] - \beta_0 - \mathbb{E}[U|D=0] \\ &= \beta_1 + \mathbb{E}[U|D=1] - \mathbb{E}[U|D=0] \\ &= \beta_1\end{aligned}$$

where the last equality exploits the fact that $\mathbb{E}[U|D=1] = \mathbb{E}[U|D=0] = 0$. To see this, we use the two first order conditions from the OLS minimization problem

$$\begin{aligned}\mathbb{E}[DU] &= \mathbb{E}[DU|D=1]\Pr(D=1) + \mathbb{E}[DU|D=0]\Pr(D=0) \\ &= \mathbb{E}[U|D=1]\Pr(D=1)\end{aligned}$$

Thus, using the FOC and the assumption that $\Pr(D=1) > 0$, we know that $\mathbb{E}[U|D=1] = 0$. Similarly,

$$\mathbb{E}[U] = \mathbb{E}[U|D=1]\Pr(D=1) + \mathbb{E}[U|D=0]\Pr(D=0) = 0$$

Again, given that $\mathbb{E}[U|D=1] = 0$ and assume $\Pr(D=0) > 0$ we find that $\mathbb{E}[U|D=0] = 0$.

(b) OLS implies that the causal effect of college education on wages is simply the difference in the conditional expectations, that is, the difference in expected wages between graduating from college and not graduating from college.

(c) Given $\mathbb{E}[U|D=0]=0$ from part (a), we have

$$\begin{aligned}\mathbb{E}[Y|D=0] &= \mathbb{E}[\beta_0+\beta_1 D+U|D=0]\\ &= \beta_0+\mathbb{E}[U|D=0]\\ &= \beta_0\end{aligned}$$

(d) In this simple model, the OLS identifies the intercept β_0 such that it can be interpreted as the expected wage for not having a college degree.

(e) Based on the results in part (a), we can propose the following estimator

$$\begin{aligned}\hat{\beta}_1 &= \frac{1}{n_1}\sum_{i=1}^{n} Y_i D_i - \frac{1}{n_0}\sum_{i=1}^{n} Y_i(1-D_i)\\ &= \frac{\sum_{i=1}^{n} Y_i D_i}{\sum_{i=1}^{n} D_i} - \frac{\sum_{i=1}^{n} Y_i(1-D_i)}{\sum_{i=1}^{n}(1-D_i)}\end{aligned}$$

(f) Similarly,

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} Y_i(1-D_i)}{\sum_{i=1}^{n}(1-D_i)}$$

(g) Remember that $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1\bar{D}_n$. We want to show $\mathbb{E}[\hat{\beta}_0|D_1,...,D_n]=\beta_0$

$$\begin{aligned}\mathbb{E}[\hat{\beta}_0|D_1,...,D_n] &= \mathbb{E}[\bar{Y}_n|D_1,...,D_n] - \mathbb{E}[\hat{\beta}_1|D_1,...,D_n]\bar{D}_n\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i|D_1,...,D_n] - \mathbb{E}[\hat{\beta}_1|D_1,...,D_n]\bar{D}_n\\ &= \frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1 D_i+\mathbb{E}[U_i|D_1,...,D_n]) - \mathbb{E}[\hat{\beta}_1|D_1,...,D_n]\bar{D}_n\\ &= \beta_0+\beta_1\bar{D}_n-\beta_1\bar{D}_n\\ &= \beta_0\end{aligned}$$

Note that I skipped the details in proving $\mathbb{E}[\hat{\beta}_1|D_1,...,D_n]=\beta_1$. By the Law of Iterated Expectations, conditional unbiasedness implies unconditional unbiasedness. Thus $\mathbb{E}[\hat{\beta}_0] = \mathbb{E}[\mathbb{E}[\hat{\beta}_0|D_1,...,D_n]] = \beta_0$

Question 3

(a) Let D be a binary decision random variable, with $D_i = 1$ if voter i voted yes. Then we know $\mathbb{E}[D] = 0.65$ and $X = \sum_{i=1}^{200} D_i$. Thus,

$$\mathbb{E}[X] = \sum_{i=1}^{200} \mathbb{E}[D_i] = 200 \times 65\% = 130$$

(b) We have

$$\begin{aligned}
\mathrm{Var}[X] = \mathrm{Var}\left[\sum_{i=1}^{200} D_i\right] &= \sum_{i=1}^{200} \mathrm{Var}(D_i) \\
&= \sum_{i=1}^{200} \mathbb{E}[D_i](1 - \mathbb{E}[D_i]) \\
&= (200)(0.65)(0.35) \\
&= 45.5
\end{aligned}$$

(c) By CLT, we have

$$\sqrt{n}\Big(\bar{D}_n - \mathbb{E}[D]\Big) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(D))$$

Given $n = 200$, we know

$$\sqrt{200}\Big(\frac{1}{200}\sum_{i=1}^{200} D_i - \mathbb{E}[D]\Big) \approx \mathcal{N}(0, \mathrm{Var}(D))$$

Plugging in $\mathbb{E}[D] = 0.65$ and $\mathrm{Var}[D] = 45.5$, we can approximate the distribution of X

$$X = \sum_{i=1}^{200} D_i \approx \mathcal{N}(130, 45.5)$$

Therefore,

$$\Pr(X \leq 115) = F_X(115), \quad \text{where } F_X \text{ is the CDF of } X$$

Note that you can also normalize to get

$$\Pr(X \leq 115) = \Pr(Z \leq T_{200}) = \Phi(T_{200}) \approx 1.3\%$$
$$\text{where } T_{200} = \left(\frac{115/200 - 0.65}{\sqrt{(0.65)(0.35)/200}}\right) = \left(\frac{115 - 130}{\sqrt{200(0.65)(0.35)}}\right)$$

Question 4

(a) Using the definition of variance, we have

$$\begin{aligned}
\text{Var}\left[X\right] &= \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}\left[X\right] + \mathbb{E}\left[X\right]^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}\left[X\right]^2 + \mathbb{E}\left[X\right]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2
\end{aligned}$$

(b) Notice that $\mathbb{E}\left[X\right] = 0$ implies $\text{Var}\left[X\right] = \mathbb{E}\left[X^2\right]$. Using the hint to define $W \equiv X^2$, then the null hypothesis becomes $H_0 : \mathbb{E}\left[W\right] = \sigma^2$. So the test statistic is simply the one we have seen many times before

$$T_n = \frac{\bar{W}_n - \mathbb{E}\left[W\right]}{\sqrt{\hat{\sigma}^2_W/n}}$$
$$\text{where } \bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i, \quad \hat{\sigma}^2_W = \frac{1}{n-1}\sum_{i=1}^{n}(W_i - \bar{W}_n)^2$$

(c) Using the standard argument (i.e. CLT plus the Slutsky theorem), we can show that

$$\frac{\bar{W}_n - \mathbb{E}\left[W\right]}{\sqrt{\text{Var}\left[W\right]/n}} \xrightarrow{d} \mathcal{N}(0,1) \text{ by CLT}$$
$$T_n = \frac{\bar{W}_n - \mathbb{E}\left[W\right]}{\sqrt{\text{Var}\left[W\right]/n}\sqrt{\hat{\sigma}^2_W/\text{Var}\left[W\right]}} \xrightarrow{d} \mathcal{N}(0,1) \text{ by Slutsky}$$

(d) Assuming n is sufficiently large, given a two-sided test and a normal limiting distribution, we know that the critical value is simply

$$C_\alpha = Z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

You can check that $\Pr(|T_n| \geq C_\alpha) = \alpha$