Econ 210A: Midterm Solutions *

Lloyd Han

May 6, 2014

1. (a) We have here that using the law of total probability

$$E[X] = (-1)\Pr(X=-1) + (0)\Pr(X=0) + (1)\Pr(X=1) \tag{1}$$
$$= (-1)\left(\frac{1}{8}+0+\frac{1}{16}+0+\frac{1}{8}+0\right) + (0)\left(0+\frac{1}{16}+0+\frac{3}{16}+0+0\right) + (1)\left(\frac{1}{8}+0+0+\frac{1}{4}+\frac{1}{16}+0\right) \tag{2}$$
$$= \frac{1}{8} \tag{3}$$

(b) Using the definition

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] \tag{4}$$

Let us first tackle the first term

$$E[XY] = ((-1)\cdot 1)\left(\frac{1}{8}+0\right) + ((-1)\cdot 2)\left(0+\frac{1}{8}\right) + ((-1)\cdot 3)\left(\frac{1}{16}+0\right) \tag{5}$$
$$+ ((0)\cdot 1)(\cdot) + ((0)\cdot 2)(\cdot) + ((0)\cdot 3)(\cdot) \tag{6}$$
$$+ ((1)\cdot 1)\left(\frac{1}{8}+\frac{1}{4}\right) + ((1)\cdot 2)\left(0+\frac{1}{16}\right) + ((1)\cdot 3)(0+0) \tag{7}$$
$$= \frac{-1}{16} \tag{8}$$

where the $(\cdot)$ is because we know that it is multiplied by zero. Now we can compute

$$E[Y] = (1)\left(\frac{1}{8}+0+\frac{1}{8}+0+\frac{3}{16}+\frac{1}{4}\right) + (2)\left(0+\frac{1}{16}+0+\frac{1}{8}+0+\frac{1}{16}\right) + (3)\left(\frac{1}{16}+0+0+0+0+0\right) \tag{9}$$
$$= \frac{11}{8} \tag{10}$$

Thus combining we find

$$\mathrm{Cov}(X,Y) = \frac{-1}{16} - \frac{1}{8}\frac{11}{8} \tag{11}$$
$$= \frac{-15}{64} \tag{12}$$

(c) Let us first compute $E[Y|X, Z=1]$. Now this will be a function of X. Thus we can compute it

*Please let me know of any typos and errors! lloydhan@uchicago.edu for the 3 possible X values

$$E[Y|X=-1,Z=1] = (1)\left(\frac{0}{0+\frac{1}{8}+0}\right) + (2)\left(\frac{\frac{1}{8}}{0+\frac{1}{8}+0}\right) + (3)\left(\frac{0}{0+\frac{1}{8}+0}\right) = 2 \tag{13}$$

$$E[Y|X=0,Z=1] = (1)\left(\frac{\frac{3}{16}}{\frac{3}{16}+0+0}\right) + (2)\left(\frac{0}{\frac{3}{16}+0+0}\right) + (3)\left(\frac{0}{\frac{3}{16}+0+0}\right) = 1 \tag{14}$$

$$E[Y|X=1,Z=1] = (1)\left(\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{16}+0}\right) + (2)\left(\frac{\frac{1}{16}}{\frac{1}{4}+\frac{1}{16}+0}\right) + (3)\left(\frac{0}{\frac{1}{4}+\frac{1}{16}+0}\right) = \frac{6}{5} \tag{15}$$

$$\tag{16}$$

Note above I used Bayes rule, in other words I used that

$$\Pr(Y=1|X=-1,Z=1) = \frac{\Pr(Y=1,X=-1|Z=1)}{\Pr(X=-1,Z=1)} \tag{17}$$

We don't know the left hand side of Eqn. 17 but we do know the right hand side from our table! Combining we have

$$E[Y|X,Z=1] = \begin{cases} 2 & X=-1 \\ 1 & X=0 \\ \frac{6}{5} & X=1 \end{cases} \tag{18}$$

Now we know

$$\begin{aligned}
E\left[E[Y|X,Z=1]|Z=1\right] &= 2\Pr(X=-1|Z=1) + 1\Pr(X=0|Z=1) + \frac{6}{5}\Pr(X=1|Z=1) && (19)\\
&= 2\frac{\left(0+\frac{1}{8}+0\right)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} + 1\frac{\left(\frac{3}{16}+0+0\right)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} + \frac{6}{5}\frac{\left(\frac{1}{4}+\frac{1}{16}+0\right)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} && (20)\\
&= \frac{13}{10} && (21)
\end{aligned}$$

Now for the RHS we have

$$\begin{aligned}
E[Y|Z=1] &= (1)\frac{\left(0+\frac{3}{16}+\frac{1}{4}\right)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} + (2)\frac{\left(\frac{1}{8}+0+\frac{1}{16}\right)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} + (3)\frac{(0+0+0)}{\frac{1}{8}+\frac{3}{16}+\frac{1}{4}+\frac{1}{16}} && (22)\\
&= \frac{13}{10} && (23)
\end{aligned}$$

verified!

2. (a) We know that

$$\begin{aligned}
E[X_i^2] &= E[X_i^2] - E[X_i]^2 + E[X_i]^2 && (24)\\
&= \mathrm{Var}(X_i) + E[X_i]^2 && (25)\\
&= \sigma^2 + \mu^2 && (26)
\end{aligned}$$

as desired.

(b) We can compute

$$\begin{aligned}
E\left[A_n^1\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] && (27)\\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2\right] && (28)\\
&= \sigma^2 + \mu^2 && (29)
\end{aligned}$$

thus clearly it is biased!

(c) We can compute

$$E\left[A_n^2\right] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] \tag{30}$$
$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} X_i^2 + 2\sum_{i\neq j} X_i X_j\right] \tag{31}$$
$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} E[X_i^2] + 2\sum_{i\neq j} E[X_i X_j]\right) \tag{32}$$
$$= \frac{1}{n^2}\left(n(\sigma^2+\mu^2) + 2\frac{n(n-1)}{2}E[X_i X_j]\right) \tag{33}$$
$$= \frac{1}{n^2}\left(n(\sigma^2+\mu^2) + 2\frac{n(n-1)}{2}E[X_i]^2\right) \tag{34}$$
$$= \frac{1}{n^2}\left(n(\sigma^2+\mu^2) + n(n-1)\mu^2\right) \tag{35}$$
$$= \mu^2 + \frac{\sigma^2}{n} \tag{36}$$

thus clearly it is also biased! However, the bias goes to zero as $n \to \infty$.

(d) To prove consistency we use the WLLN. We know that

$$\text{plim}\left(A_n^1\right) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \tag{37}$$
$$= \text{plim}\left(\overline{X_n^2}\right) \tag{38}$$
$$= E[X_i^2] \tag{39}$$
$$= \mu^2 + \sigma^2 \tag{40}$$

Thus it is not consistent.

(e) Here we again use the WLLN in conjunction with the Continous Mapping Theorem.

$$\text{plim}\left(A_n^2\right) = \text{plim}\left(\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^2\right) \tag{41}$$
$$= \text{plim}\left(\left[\bar{X}_n\right]^2\right) \tag{42}$$
$$= \text{plim}\left(\bar{X}_n\right)^2 \tag{43}$$
$$= E[X_i]^2 \tag{44}$$
$$= \mu^2 \tag{45}$$

Thus it is consistent! In the third equality we used the Continuous Mapping Theorem.

3. (a) In the conditional expectation model we have that

$$\alpha + \beta X = E[Y|X] \tag{46}$$

by definition. In other words β is the best linear predictor of Y in the mean squared error sense.

(b) No! The linear regression we do here only represents that the variables X and Y are correlated. To say anything about the causal effect, we need to take a stance on the true data generating process

(c) Given the conditional expectation model we have

$$\begin{aligned}
\mathrm{Cov}\left[X, U\right] &= \mathrm{Cov}\left[X, Y - \alpha - \beta X\right] && (47)\\
&= \mathrm{Cov}[X, Y] - 0 - \beta \mathrm{Var}[X] && (48)\\
&= \mathrm{Cov}[X, Y] - \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]} \mathrm{Var}[X] && (49)\\
&= 0 && (50)
\end{aligned}$$

as desired.

(d) We know that

$$\begin{aligned}
\hat{\beta} &= \frac{\widehat{\mathrm{Cov}[X, Y]}}{\widehat{\mathrm{Var}[X]}} && (51)\\
&= \frac{\sum_{i=1}^{n} \left[(X_i - \bar{X})(Y_i - \bar{Y})\right]}{\sum_{i=1}^{n} (X_i - \bar{X})^2} && (52)\\
&= \frac{\sum_{i=1}^{n} (X_i Y_i) - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} X_i\right)^2} && (53)\\
&= \frac{100 - \frac{200}{20}}{20 - \frac{100}{20}} && (54)\\
&= 6 && (55)
\end{aligned}$$

as desired.

(e) We know that

$$\begin{aligned}
\hat{\alpha} &= \frac{1}{n} \sum_{i=1}^{n} Y_i - \hat{\beta} \frac{1}{n} \sum_{i=1}^{n} X_i && (56)\\
&= \frac{20}{20} - 6 \frac{10}{20} && (57)\\
&= -2 && (58)
\end{aligned}$$

as desired.

(f) The variance of U is given by

$$\begin{aligned}
\widehat{\mathrm{Var}(U_i)} &= \frac{1}{n-2} \sum_{i=1}^{n} U_i^2 && (59)\\
&= \frac{170}{18} && (60)
\end{aligned}$$

(g) Remember from notes or Wooldridge that

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}\right) &= \frac{\frac{1}{n-2} \sum_{i=1}^{n} U_i^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} && (61)\\
&= \frac{\frac{170}{18}}{20 - \frac{100}{20}} && (62)\\
&= \frac{17}{27} && (63)
\end{aligned}$$

as desired.

(h) Remember also that

$$\operatorname{Var}(\hat{\alpha}) = \frac{1}{n}\operatorname{Var}\left(\hat{\beta}\right)\sum x_i^2 \tag{64}$$
$$= \frac{10}{18 \cdot 20}\left(1 + \frac{1/4}{(1/20)15}\right) \tag{65}$$
$$= \frac{17}{27} \tag{66}$$

(i) We know that

$$E[Y|X = 10] = 10(\hat{\beta}) + \hat{\alpha} \tag{67}$$
$$= 58 \tag{68}$$

as desired.

(j) We know that $\hat{\beta}$ is 18 standard deviations from zero. In other words we have

$$\frac{\hat{\beta} - 0}{\widehat{\mathrm{SD}(\hat{\beta})}} = \frac{6}{\sqrt{\frac{17}{27}}} \gg Z_{.05} \approx 1.64 \tag{69}$$

Thus this is clearly going to be significant at the 95% level.

(k) We know that

$$R^2 = 1 - \frac{SSR}{TSS} \tag{70}$$
$$= 1 - \frac{170}{680} \tag{71}$$
$$= \frac{51}{68} \tag{72}$$
$$= \frac{3}{4} \tag{73}$$

as desired.

(l) To form a 95% confidence interval, we can use the CLT. Now we know

$$\Pr\left(-Z_{97.5} \leq \frac{\hat{\alpha} - \alpha}{\widehat{\mathrm{SD}(\hat{\alpha})}} \leq Z_{97.5}\right) = .05 \tag{74}$$

Thus using this we can form

$$-\sqrt{\frac{17}{27}}(1.96) - 2 \leq \alpha \leq \sqrt{\frac{17}{27}}(1.96) - 2 \tag{75}$$

as our α confidence interval.

4. (a) We can compute

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n (x_i^3(\beta x_i + U_i))}{\sum_{i=1}^n x_i^4} - \beta\right) \tag{76}$$
$$= \sqrt{n}\left(\frac{\sum_{i=1}^n x_i^3 U_i}{\sum_{i=1}^n x_i^4}\right) \tag{77}$$
$$= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i^3 U_i}{\frac{1}{n}\sum_{i=1}^n x_i^4} \tag{78}$$

as desired.

(b) We use the WLLN here

$$\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^4\right) = \text{plim}\left(\overline{x_n^4}\right) \tag{79}$$

$$= E[x_i^4] \tag{80}$$

(c) Here we use the LIE

$$E[x_i^3U_i] = E[E[x_i^3U_i|x_i]] \tag{81}$$

$$= E[x_i^3E[U_i|x_i]] \tag{82}$$

$$= E[x_i^3(0)] \tag{83}$$

$$= 0 \tag{84}$$

as desired.

(d) We know by the CLT that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^3U_i - E[x_i^3U_i]\right) \sim N\left(0, \text{Var}\left(x_i^3U_i\right)\right) \tag{85}$$

Now from the previous part we know that $E[x_i^3U_i] = 0$. Thus we also know that

$$\text{Var}\left(x_i^3U_i\right) = E[x_i^6U_i^2] - \left(E[x_i^3U_i]\right)^2 \tag{86}$$

$$= E[x_i^6U_i^2] \tag{87}$$

Thus we find

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^3U_i \sim N\left(0, E[x_i^6U_i^2]\right) \tag{88}$$

(e) Now we can simply just rescale to get

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^3U_i}{\frac{1}{n}\sum_{i=1}^{n}x_i^4} \tag{89}$$

$$\sim \frac{1}{E[x_i^4]}N\left(0, E[x_i^6U_i^2]\right) \tag{90}$$

$$\sim N\left(0, \frac{E[x_i^6U_i^2]}{\left(E[x_i^4]\right)^2}\right) \tag{91}$$

as desired.

(f) We note that under the null hypothesis $\beta = 0$. Now we know that we can form our test statistic using the limiting distribution we derived in part (e). Now we need to replace the variance with our sample estimator. In other words we know that

$$\frac{\sum x_i^6U_i^2}{\left(\sum x_i^4\right)^2} \xrightarrow{p} \frac{E[x_i^6U_i^2]}{\left(E[x_i^4]\right)^2} \tag{92}$$

Thus we have our test statistic T_n given by

$$T_n = \frac{\sqrt{n}\tilde{\beta}_n\sum x_i^4}{\sqrt{\sum x_i^6U_i^2}} \tag{93}$$

Now we know that $T_n \sim N(0,1)$. Thus to determine our critical value for significance level α for a two sided test we have

$$\Pr\left(|T_n| \geq Z_{1-\frac{\alpha}{2}}\right) = \alpha \tag{94}$$

Thus we find that our critical value is given by

$$\left|\tilde{\beta}_n\right| \geq Z_{1-\frac{\alpha}{2}} \sqrt{\sum x_i^6 U_i^2} \frac{1}{\sum x_i^4 \sqrt{n}} \tag{95}$$