Econometrics (ECON 21020) – Winter 2020
Problem Set #3 Solutions

SW 4.3

(a) The coefficient value 9.6 estimates the change in the mean of AWE for each additional year of age. The coefficient value 696.7 is the intercept of the regression line, it estimates the mean of AWE at age 0.

(b) SER is in the same units as the dependent variable (Y, or AWE in this example). Thus SER is measured in dollars per week.

(c) R^2 is unit-free.

(d) The regression predicts average earnings of \$936.70 for a 25-year-old worker, and \$1,128.70 for a 45-year-old worker.

(e) No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

(f) No. The distribution of earnings is positively skewed and has a large kurtosis.

(g) The average value of AWE in the sample is \$1,096.06.

SW 4.4

The expected excess return on an asset $(R - R_f)$ given the market portfolio $(R_m - R_f)$ is proportional to the excess return on the market portfolio:

$$\begin{aligned} R - R_f &= \beta\left(R_m - R_f\right) + \varepsilon \\ 0 &= \mathbf{E}\left[\varepsilon | R_m - R_f\right] \end{aligned}$$

with $\beta > 0$.

(a) Suppose $\beta > 1$. Then,

$$\begin{aligned} \mathrm{Var}\left(R - R_f\right) &= \mathrm{Var}\left(\beta\left(R_m - R_f\right) + \varepsilon\right) \\ &= \mathrm{Var}\left(\beta\left(R_m - R_f\right)\right) + \mathrm{Var}\left(\varepsilon\right) + 2\mathrm{Cov}\left(\beta\left(R_m - R_f\right), \varepsilon\right) \\ &= \beta^2 \mathrm{Var}\left(R_m - R_f\right) + \mathrm{Var}\left(\varepsilon\right) + 0 \\ &\geq \beta^2 \mathrm{Var}\left(R_m - R_f\right) \\ &> \mathrm{Var}\left(R_m - R_f\right) \qquad \because \beta > 1. \end{aligned}$$

(b) Suppose $\beta < 1$.

$$\begin{aligned} \mathrm{Var}\left(R - R_f\right) &= \beta^2 \mathrm{Var}\left(R_m - R_f\right) + \mathrm{Var}\left(\varepsilon\right) \\ \mathrm{Var}\left(R - R_f\right) - \mathrm{Var}\left(R_m - R_f\right) &= \underbrace{\left(\beta^2 - 1\right)}_{<0} \mathrm{Var}\left(R_m - R_f\right) + \mathrm{Var}\left(\varepsilon\right) \end{aligned}$$

If $\mathrm{Var}\left(R_m - R_f\right) < \mathrm{Var}\left(\varepsilon\right)/(1 - \beta^2)$,

$$\mathrm{Var}\left(R - R_f\right) - \mathrm{Var}\left(R_m - R_f\right) > 0.$$

company	estimated return
Wal-Mart	0.0464
Kellogg	0.054
Waste Management	0.0578
Verizon	0.0578
Microsoft	0.073
Best Buy	0.0844
Bank of America	0.1262

(c) We have

$$R_f = 0.035 \qquad \text{and} \qquad R_m = 0.073.$$

From the estimated β, the return on a particular asset is estimated to be

$$\begin{aligned}\hat{R}_i &= R_f + \hat{\beta}_i \left(R_m - R_f\right) \\ &= 0.038\hat{\beta}_i + 0.035.\end{aligned}$$

Thus,

SW 4.7

We have

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$$

Thus,

$$\begin{aligned}\mathbf{E}\left[\hat{\beta}_0\right] &= \mathbf{E}\left[\bar{Y}_n - \hat{\beta}_1 \bar{X}_n\right] \\ &= \mathbf{E}\left[\left(\beta_0 + \beta_1 \bar{X}_n + \frac{1}{n}\sum_{i=1}^{n} U_i\right) - \hat{\beta}_1 \bar{X}_n\right] \\ &= \beta_0 + \mathbf{E}\left[\mathbf{E}\left[(\beta_1 - \hat{\beta}_1)\bar{X}_n | X_1, ..., X_n\right]\right] + \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[U_i] \\ &= \beta_0\ ,\end{aligned}$$

Since $E[U_i] = 0$ and $E[\hat{\beta}_1 | X_1, ..., X_n] = \beta_1$.

SW 4.11

(a) If $\beta_0 = 0$, then

$$Y_i = \beta_1 X_i + U_i.$$

The sample analogue of

$$\mathbf{E}\left[(Y - \beta_1 X)^2\right]$$

is

$$\frac{1}{n}\sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2\ .$$

By solving

$$\min_{b_1} \sum_{i=1}^{n} (Y_i - b_1 X_i)^2 \ ,$$

the F.O.C. is

$$-2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 X_i\right) X_i = 0 \ .$$

Hence we have that

$$\hat{\beta}_1 = \frac{\sum_i X_i Y_i}{\sum_i {X_i}^2} \ .$$

SW 4.12

(a)

$$R^2 = \frac{\sum_i \left(\hat{Y}_i - \bar{Y}_n\right)^2}{\sum_i \left(Y_i - \bar{Y}_n\right)^2} = \frac{\sum_i \left(\hat{\beta}_0 + \hat{\beta}_1 X_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{X}_n\right)^2}{\sum_i \left(Y_i - \bar{Y}_n\right)^2} \qquad \because \sum_i \hat{U}_i = 0$$

$$= \frac{\sum_i \left(\hat{\beta}_1 \left(X_i - \bar{X}_n\right)\right)^2}{\sum_i \left(Y_i - \bar{Y}_n\right)^2} = \hat{\beta}_1^2 \frac{\sum_i \left(X_i - \bar{X}_n\right)^2}{\sum_i \left(Y_i - \bar{Y}_n\right)^2}$$

$$= \underbrace{\frac{\left(\sum_j \left(X_j - \bar{X}_n\right)\left(Y_j - \bar{Y}_n\right)\right)^2}{\left(\sum_j \left(X_j - \bar{X}_n\right)^2\right)^2}}_{=\hat{\beta}_1^2} \frac{\sum_i \left(X_i - \bar{X}_n\right)^2}{\sum_i \left(Y_i - \bar{Y}_n\right)^2}$$

$$= \frac{\left(\sum_j \left(X_j - \bar{X}_n\right)\left(Y_j - \bar{Y}_n\right)\right)^2}{\sum_j \left(X_j - \bar{X}_n\right)^2 \sum_i \left(Y_i - \bar{Y}_n\right)^2} = \left(\frac{\frac{1}{n-1}\sum_j \left(X_j - \bar{X}_n\right)\left(Y_j - \bar{Y}_n\right)}{\sqrt{\frac{1}{n-1}\sum_j \left(X_j - \bar{X}_n\right)^2 \frac{1}{n-1}\sum_i \left(Y_i - \bar{Y}_n\right)^2}}\right)^2$$

$$= \left(r_{XY}\right)^2$$

(b) Note that by repeating **a.**,

$${R_{X \text{on} Y}}^2 = {r_{XY}}^2 = {R_{Y \text{on} X}}^2.$$

(c)

$$\hat{\beta}_1 = \frac{\sum_i \left(X_i - \bar{X}_n\right)\left(Y_i - \bar{Y}_n\right)}{\sum_i \left(X_i - \bar{X}_n\right)^2}$$

$$= \frac{\sum_i \left(X_i - \bar{X}_n\right)\left(Y_i - \bar{Y}_n\right)}{\sqrt{\sum_i \left(X_i - \bar{X}_n\right)^2 \sum_i \left(Y_i - \bar{Y}_n\right)}} \frac{\sqrt{\sum_i \left(Y_i - \bar{Y}_n\right)^2}}{\sqrt{\sum_i \left(X_i - \bar{X}_n\right)^2}}$$

$$= \frac{\frac{1}{n-1}\sum_i \left(X_i - \bar{X}_n\right)\left(Y_i - \bar{Y}_n\right)}{\sqrt{\frac{1}{n-1}\sum_i \left(X_i - \bar{X}_n\right)^2 \frac{1}{n-1}\sum_i \left(Y_i - \bar{Y}_n\right)}} \frac{\sqrt{\frac{1}{n-1}\sum_i \left(Y_i - \bar{Y}_n\right)^2}}{\sqrt{\frac{1}{n-1}\sum_i \left(X_i - \bar{X}_n\right)^2}}$$

$$= r_{XY} \frac{s_Y}{s_X}.$$

SW 5.1

(a) The 95% confidence interval for β_1 is $-5.82 \pm 1.96 * 2.21$.

(b) Our test-statistic is

$$T_n = \left| \frac{(\hat{\beta}_1 - 0)}{\hat{\sigma}_{\hat{\beta}_1}/\sqrt{n}} \right| = \left| \frac{-5.82}{2.21} \right| = 2.6335 \ .$$

Hence the p-value is $2(1 - \Phi(T_n)) \approx 0.0084$. Since the p-value is less than 0.01 we can reject the null hypothesis at the 5% and 1$ significance levels.

(c) Our test-statistic is

$$T_n = \left| \frac{(\hat{\beta}_1 - (-5.6))}{\hat{\sigma}_{\hat{\beta}_1}/\sqrt{n}} \right| = \left| \frac{0.22}{2.21} \right| = 0.10 \ .$$

Hence the p-value is $2(1 - \Phi(T_n)) \approx 0.92$. Since the p-value is larger than 0.1, we cannot reject the null at the 5% level. By the "duality" of hypothesis testing and confidence intervals which we discussed in class, since we cannot reject, this value must be contained in the 95% confidence interval.

(d) The 99% confidence interval for β_0 is $520.4 \pm 2.58 * 20.4$.

SW 5.5

(a) The test scores for students in small classes are higher by 13.9 points on average. Given the standard deviation of test score is 75 points, the gap of 13.9 points is equivalent with 18.54% of one standard deviation.

(b) The estimated effect of class size on test score is statistically significant since the t statistics is

$$\frac{13.9}{2.5} = 5.56.$$

Thus, the null hypothesis of no effect is rejected at the level 0.05 test.

(c) The confidence interval is

$$[13.9 - 2.58 \cdot 2.5, 13.9 + 2.58 \cdot 2.5] = [7.45, 20.35].$$

SW 5.6

(a) The question asks whether the variability in test scores in large classes is the same as the variability in small classes. It is hard to say. On the one hand, teachers in small classes might able so spend more time bringing all of the students along, reducing the poor performance of particularly unprepared students. On the other hand, most of the variability in test scores might be beyond the control of the teacher.

(b) No. Equation 5.3 is valid for heteroskedastic errors. Since homoskedastic errors are a special case of heteroskedastic errors, the validity of the confidence interval constructed in 5.5 (c) is not compromised.

Problem 2

(a) Because PC is a binary variable with possible outcome of zero and one,

$$\hat{\beta}_0 = \frac{1}{\sum_i (1 - PC_i)} \sum_{i:PC_i=0} Col\_GPA_i = 2.5$$

$$\hat{\beta}_1 = \frac{1}{\sum_i PC_i} \sum_{i:PC_i=1} Col\_GPA_i - \frac{1}{\sum_i (1 - PC_i)} \sum_{i:PC_i=0} Col\_GPA_i = 3.5 - 2.5 = 1$$

(b) Our estimator for the asymptotic variance of $\hat{\beta}_1$ is given by

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\frac{1}{n-2} \sum_{i=1}^n \left(PC_i - \overline{PC}_n\right)^2 \hat{U}_i^2}{\left(\frac{1}{n} \sum_{i=1}^n \left(PC_i - \overline{PC}_n\right)^2\right)^2}$$

From the information given,

$$\frac{1}{33} \sum_{i:PC_i=0} (Col\_GPA_i - 2.5)^2 = 0.62$$

$$\frac{1}{52} \sum_{i:PC_i=1} (Col\_GPA_i - 3.5)^2 = 0.47$$

From this we get that

$$\sum_{i:PC_i=0} \hat{U}_i^2 = 33 \cdot 0.62$$

$$\sum_{i:PC_i=1} \hat{U}_i^2 = 52 \cdot 0.47$$

Since the predicted value $\widehat{Col\_GPA}_i$ is just the sample average for each subgroup:

$$\widehat{Col\_GPA}_i = \begin{cases} 3.5, & \text{if } PC_i = 1 \\ 2.5, & \text{if } PC_i = 0\ . \end{cases}$$

Then,

$$\begin{aligned}
\sum_i \left(PC_i - \overline{PC}_n\right)^2 \hat{U}_i^2 &= \sum_{i:PC_i=1} \left(PC_i - \overline{PC}_n\right)^2 \hat{U}_i^2 + \sum_{i:PC_i=0} \left(PC_i - \overline{PC}_n\right)^2 \hat{U}_i^2 \\
&= \sum_{i:PC_i=1} (1 - 53/87)^2 \hat{U}_i^2 + \sum_{i:PC_i=0} (0 - 53/87)^2 \hat{U}_i^2 \\
&= \frac{34^2}{87^2} \sum_{i:PC_i=1} \hat{U}_i^2 + \frac{53^2}{87^2} \sum_{i:PC_i=0} \hat{U}_i^2 \\
&= \frac{34^2}{87^2} \cdot 52 \cdot 0.47 + \frac{53^2}{87^2} \cdot 33 \cdot 0.62 \\
&\approx 11.3258 \\
\sum_i \left(PC_i - \overline{PC}_n\right)^2 &= \sum_{i:PC_i=1} (1 - 53/87)^2 + \sum_{i:PC_i=0} (0 - 53/87)^2 \\
&= 53 \cdot \frac{34^2}{87^2} + 34 \cdot \frac{53^2}{87^2} \\
&\approx 20.7126
\end{aligned}$$

Hence

$$\hat{\sigma}_{\hat{\beta}_1}^2 \approx \frac{11.3258/85}{(20.7126/87)^2} \approx 2.3508 \ .$$

The test statistics is

$$\frac{\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)}{\hat{\sigma}_{\hat{\beta}_1}} \approx \frac{\sqrt{87}\,(1-0)}{1.5332} \approx 6.0835.$$

and the p-value is 0.0000. We can reject the null hypothesis at significance level of 0.01.

(c) The resulting OLS estimates are

$$\begin{aligned}
\hat{\alpha}_0 &= 3.5 \\
\hat{\alpha}_1 &= -1
\end{aligned}$$

Problem 3

(a)

$$\begin{aligned}
\mathbf{E}[Y|X] &= \beta_0 + \beta_1 X + \mathbf{E}[U|X] \\
&= \beta_0 + \beta_1 X \\
\Pr\{Y = 1|X\} &= \mathbf{E}\left[\mathbf{1}_{\{Y=1\}}|X\right] \\
&= \mathbf{E}\left[Y|X\right] = \beta_0 + \beta_1 X
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{Var}(Y|X) &= \mathbf{E}\left[(Y - \mathbf{E}[Y|X])^2 |X\right] \\
&= \mathbf{E}\left[Y^2 - 2Y\mathbf{E}[Y|X] + (\mathbf{E}[Y|X])^2 |X\right] \\
&= \mathbf{E}\left[Y^2|X\right] - 2\mathbf{E}[Y|X]\mathbf{E}[Y|X] + (\mathbf{E}[Y|X])^2 \\
&= \mathbf{E}\left[Y^2|X\right] - (\mathbf{E}[Y|X])^2 = \mathbf{E}[Y|X] - (\mathbf{E}[Y|X])^2 \\
&= \beta_0 + \beta_1 X - (\beta_0 + \beta_1 X)^2 \\
&= \beta_0 (1 - \beta_0) + \beta_1 (1 - 2\beta_0) X - {\beta_1}^2 X^2
\end{aligned}$$

(c)

$$\begin{aligned}
\mathrm{Var}(U|X) &= E\left[(U - \mathbf{E}[U|X])^2 |X\right] \\
&= \mathbf{E}\left[U^2|X\right] \\
&= \mathbf{E}\left[(\beta_0 + \beta_1 X + U - \beta_0 - \beta_1 X)^2 |X\right] \\
&= \mathbf{E}\left[(Y - \mathbf{E}[Y|X])^2 |X\right] = \mathrm{Var}(Y|X).
\end{aligned}$$

The model is heteroskedastic since the conditional variance of U given X is a function of X.

(d) As we see from part (a), the conditional mean of $E[Y|X] = \beta_0 + \beta_1 X$. If X can take any value in the real line, then when $\beta_1 \neq 0$, $E[Y|X]$ can as well, which contradicts the fact that Y is assumed to be binary.

Problem 4

(a) We have 420 observations.

(b) **i.** *income* measures average district income measured in dollars.

ii. The mean of *avginc* is 15.3166 and the standard deviation of *avginc* is 7.2173.

iii. The mean of *income* is 15316.59 and the standard deviation of *income* is 7217.282. Because of the linearity between *avginc* and *income*, both mean and standard deviation of *income* must be 1000 times the mean and standard deviation of *avginc*.

(c) **i.** The mean math score is 653.3426.

ii. 57.86% of the districts have an average class size of 20 or fewer students. The mean math score in districts with average class size of 20 or fewer students is 655.7177.

iii. 42.14% of the districts have an average class size of more than 20 students. The mean math score in districts with average class size of more than 20 students is 650.0819.

iv. Since all districts can either have an average class size of more than 20 students or of 20 or fewer students, the categorization above is exhuastive. Thus, we would see the sample analogue of

$$\mathbf{E}[Math] = \mathbf{E}[Math|Small]\Pr\{Small\} + \mathbf{E}[Math|Large]\Pr\{Large\},$$

which is

$$653.3426 = 0.05785 \cdot 655.7177 + 0.4214 \cdot 650.0819.$$

v. The test statistics is

$$\frac{\overline{Math}_{\text{small}} - \overline{Math}_{\text{large}} - (\mathbf{E}[Math|Small] - \mathbf{E}[Math|Large])}{\sqrt{\widehat{sd}\left(\overline{Math}_{\text{small}}\right)^2 + \widehat{sd}\left(\overline{Math}_{\text{large}}\right)^2}}$$

and the null hypothesis is

$$\mathbf{E}[Math|Small]-\mathbf{E}[Math|Large] = 0 \quad \text{v.} \quad \mathbf{E}[Math|Small]-\mathbf{E}[Math|Large] \neq 0.$$

The test is to reject the null if the absolute value of the test statistics is bigger than 1.64 and to not reject the null otherwise. The realized value of the test statistics is

$$\frac{655.7177 - 650.0819 - 0}{\sqrt{\frac{19.3456^2}{243} + \frac{17.4430^2}{177}}} \approx 3.1218.$$

We reject the null.

vi. The covariance between *avginc* and mean math score is 94.7795. The covariance between *income* and mean math score is 94779.5. The second covariance is 1000 times the first covariance because of the linearity between *avginc* and *income*.

vii. Both correlations are 0.6994 since the effect of the change in unit on the covariances is cancelled out by the difference in Var(*avginc*) and Var(*income*).