Cornell University
Department of Economics

Econ 620 – Spring 2007
Instructor: Professor Kiefer
TA: Jae-ho Yun (jy238@cornell.edu)

Problem Set # 1: Solution Key

1.

$$\begin{aligned}
\widehat{\beta}_1 &= \sum \frac{(x_i-\overline{x})y_i}{(x_i-\overline{x})^2} = \sum \frac{(x_i-\overline{x})(\beta_0+\beta_1 x_i+\beta_2 x_i^2+\varepsilon_i)}{(x_i-\overline{x})^2} \\
&= \beta_0 \sum \frac{(x_i-\overline{x})}{(x_i-\overline{x})^2} + \beta_1 \sum \frac{(x_i-\overline{x})x_i}{(x_i-\overline{x})^2} + \beta_2 \sum \frac{(x_i-\overline{x})x_i^2}{(x_i-\overline{x})^2} \\
&\quad + \sum \frac{(x_i-\overline{x})\varepsilon}{(x_i-\overline{x})^2} \\
&= \beta_1 + \beta_2 \sum \frac{(x_i-\overline{x})x_i^2}{(x_i-\overline{x})^2} + \sum \frac{(x_i-\overline{x})\varepsilon}{(x_i-\overline{x})^2}
\end{aligned}$$

$$\text{Note: } \sum \frac{(x_i-\overline{x})}{(x_i-\overline{x})^2} = 0, \ \sum \frac{(x_i-\overline{x})x_i}{(x_i-\overline{x})^2} = 1$$

Take the conditional expectation;

$$E(\widehat{\beta}_1|x_i) = \beta_1 + \beta_2 \sum \frac{(x_i-\overline{x})x_i^2}{(x_i-\overline{x})^2}$$

$$\text{Note: } E(\sum \frac{(x_i-\overline{x})\varepsilon}{(x_i-\overline{x})^2} \mid x_i) = 0$$

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

Next, the conditional variance is;

$$\begin{aligned}
Var(\widehat{\beta}_1|x_i) &= E[(\widehat{\beta}_1 - E(\widehat{\beta}_1|x_i))^2 \mid x_i] \\
&= E[(\sum \frac{(x_i-\overline{x})\varepsilon}{(x_i-\overline{x})^2})^2 \mid x_i] \\
&\quad \text{(cross product terms are all zero)} \\
&= \frac{\sigma^2}{\sum (x_i-\overline{x})^2}
\end{aligned}$$

2. i) The correlation between y and z is 1 if a>0, -1 if a<0 and 0 if a=0.

   ii) The support of the joint distribution of y and z can be represented by the following set;

   $\{(x,y) : z = ay,\ y \in R\}$ (It is a straight line)

   iii) The correlation between y and x ($x = y^2$) is zero, since $Cov(y, y^2) = E(y * y^2) - E(y)E(y^2) = E(y^3) = 0$

   iv) The support is $\{(x,y) : x = y^2,\ y \in R\}$

   v) In the above problem, clearly, there are dependence between x and y. However, our linear relationship measure, correlation does not capture this relationship.

3. (a) OLS is BLUE if ε_i are homoskedastic with mean zero.

$$\begin{aligned}
\text{Mean:} \quad & E(\varepsilon_i) = \frac{1}{2} \times (-1) + \frac{1}{2} \times (-1) = 0 \text{ for i=1,2} \\
\text{Variance:} \quad & E(\varepsilon_i^2) = \frac{1}{2} \times (-1)^2 + \frac{1}{2} \times (-1)^2 = 1 \text{ for i=1,2} \\
\text{Covarince} \quad : \quad & Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) - E(\varepsilon_i)E(\varepsilon_j) = 0 \text{ for i} \neq \text{j}
\end{aligned}$$

   Hence, our OLS is BLUE.

   (b)

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (x_i \beta + \varepsilon_i)}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2} \\
&= 1 + \frac{1 \times \varepsilon_1 + 2 \times \varepsilon_2}{1^2 + 2^2}
\end{aligned}$$

   We have the following exact distribution of $\widehat{\beta}$

Prob.	.25	.25	.25	.25
$\widehat{\beta}$	2/5	4/5	6/5	8/5

   (c) The alternative estimator

$$\begin{aligned}
\beta^* &= \frac{\sum y_i}{\sum x_i} = \frac{\sum (x_i \beta + \varepsilon_i)}{\sum x_i} = \beta + \frac{\sum \varepsilon_i}{\sum x_i} \\
&= 1 + \frac{\varepsilon_1 + \varepsilon_2}{1 + 2}
\end{aligned}$$

   β^* is clearly unbiased.
   Hence, the exact distribution of β^* is

Prob.	.25	.50	.25
β^*	1/3	1	5/3

(d) The exact variance of the two estimators:

$$\begin{aligned} V(\widehat{\beta}) &= V(\frac{\varepsilon_1 + 2\varepsilon_2}{5}) = \frac{1}{25}(V(\varepsilon_1) + 4V(\varepsilon_2)) = \frac{1}{5} \\ V(\beta^*) &= V(\frac{\varepsilon_1 + \varepsilon_2}{3}) = \frac{1}{9}(V(\varepsilon_1) + V(\varepsilon_2)) = \frac{2}{9} \end{aligned}$$

Hence, $V(\beta^*) > V(\widehat{\beta})$, which is consistent with $\widehat{\beta}$ being BLUE.

4. (a) Recall that

$$\begin{aligned} \widehat{\beta} &= \sum \frac{(x_i - \overline{x})(y_i - \overline{y})}{(x_i - \overline{x})^2} \\ \widehat{\alpha} &= \overline{y} - \widehat{\beta}\overline{x} \\ \text{and } \sum (x_i - \overline{x})(y_i - \overline{y}) &= \sum x_i y_i - n\overline{x}\overline{y} \\ \sum (x_i - \overline{x})^2 &= \sum x_i^2 - n\overline{x}^2 \end{aligned}$$

Hence,

$$\begin{aligned} \widehat{\beta} &= \frac{30}{60} = 0.5 \\ \widehat{\alpha} &= \frac{440}{22} - 0.5 \times \frac{220}{22} = 15 \end{aligned}$$

(b) R^2 is defined as the ratio of the explained sum of squares(ESS) to total sum of squares(TSS).

$$R^2 = \frac{\widehat{\beta}^2 \sum (x_i - \overline{x})^2}{\sum (y_i - \overline{y})^2} = \widehat{\beta}^2 \frac{\sum x_i^2 - n\overline{x}^2}{\sum y_i^2 - n\overline{y}^2} = 0.5^2 \times \frac{60}{8900 - 22 \times 20^2} = 0.15$$

(c) By the normality assumption, we know that

$$\widehat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum (x_i - \overline{x})^2})$$

Moreover,

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2(n-2)$$

where $S^2 = \frac{1}{n-2} \sum e_i^2$. We can also show that $\widehat{\beta}$ and S^2 are independent each other. Then,

$$\frac{\frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum(x_i-\overline{x})^2}}}}{\sqrt{\frac{\frac{(n-2)S^2}{\sigma^2}}{(n-2)}}} = \frac{\widehat{\beta}-\beta}{\frac{S}{\sqrt{\sum(x_i-\overline{x})^2}}} \sim t(n-2)$$

We want to reject the null hypothesis if

$$t = \left| \frac{\widehat{\beta}-\beta}{\frac{S}{\sqrt{\sum(x_i-\overline{x})^2}}} \right| > t_{0.975}(20)$$

under the null hypothesis. On the other hand,

$$\begin{aligned} S^2 &= \frac{1}{n-2}\sum e_i^2 = \frac{1}{n-2}\sum(y_i - \widehat{\alpha} - \widehat{\beta}x_i)^2 \\ &= \frac{1}{n-2}\sum(y_i^2 + \widehat{\alpha}^2 + \widehat{\beta}^2 x_i^2 - 2\widehat{\alpha}y_i + 2\widehat{\alpha}\widehat{\beta}x_i - 2\widehat{\beta}x_i y_i)^2 \\ &= 4.25 \end{aligned}$$

Hence, the test statistics is given by

$$t = \left| \frac{0.5-0}{\sqrt{\frac{4.25}{60}}} \right| = 1.8787$$

Since $t_{0.975}(20) = 2.086$,we cannot reject the null hypothesis.

5. i)

$$\begin{aligned} b_1 &= (X_1' M_2 X_1)^{-1} X_1' M_2 y \\ &= (X_1' M_2 X_1)^{-1} X_1' M_2 (X_1\beta_1 + \ X_2\beta_2 + \varepsilon) \\ &= (X_1' M_2 X_1)^{-1} X_1' M_2 X_1 \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 X_2 \beta_2 \\ &\quad + (X_1' M_2 X_1)^{-1} X_1' M_2 \varepsilon \\ &= \beta_1 + 0 + (X_1' M_2 X_1)^{-1} X_1' M_2 \varepsilon \end{aligned}$$

Take expectation.

$$\begin{aligned} E(b_1) &= \beta_1 + E[(X_1' M_2 X_1)^{-1} X_1' M_2 \varepsilon] \\ &= \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 E[\varepsilon] \\ &= \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 X_1 \gamma \\ &= \beta_1 + \gamma \end{aligned}$$

Therefore, b_1 is biased.

ii)

$$\begin{aligned} b_2 &= (X_2' M_1 X_2)^{-1} X_2' M_1 y \\ &= (X_2' M_1 X_2)^{-1} X_2' M_1 (X_1\beta_1 + \ X_2\beta_2 + \varepsilon) \\ &= (X_2' M_1 X_2)^{-1} X_2' M_1 X_1 \beta_1 + (X_2' M_1 X_2)^{-1} X_2' M_1 X_2 \beta_2 \\ &\quad + (X_2' M_1 X_2)^{-1} X_2' M_1 \varepsilon \\ &= 0 + \beta_2 + (X_2' M_1 X_2)^{-1} X_2' M_1 \varepsilon \end{aligned}$$

Take expectation.

$$\begin{aligned} E(b_2) &= \beta_2 + (X_2' M_1 X_2)^{-1} X_2' M_1 E(\varepsilon) \\ &= \beta_2 + (X_2' M_1 X_2)^{-1} X_2' M_1 X_1 \gamma \\ (Note &: M_1 X_1 = 0) \\ &= \beta_2 + 0 \\ &= \beta_2 \end{aligned}$$

Therefore, b_2 is unbiased.

6. Let A be a orthogonal projection matrix onto the space spanned by a column of ones.

   (a) $R^2 = R_1^2 = R_2^2 = \frac{(x'Ay)^2}{(x'Ax)(y'Ay)}$

   (b) $\widehat{\beta}_1 = \frac{x'Ay}{x'Ax}$, $\widehat{\beta}_2 = \frac{x'Ay}{y'Ay}$, and we have $\widehat{\beta}_1\widehat{\beta}_2 = R^2$

   (c) From $\widehat{\beta}_1 = \frac{x'Ay}{x'Ax}$ and $R^2 = 1 - \frac{e'e}{y'Ay}$, we have
   $t_1 = \frac{\widehat{\beta}_1}{\sqrt{(x'Ax)^{-1}(e_1'e_1')/(n-2)}} = \frac{x'Ay}{\sqrt{(x'Ax)(y'Ay)(1-R^2)/(n-2)}}$
   Note that x and y are symmetric in the above formula. This proves $t_1 = t_2$.