ECON 4261: Introduction to Econometrics
Fall 2009

Problem Set 2 – Answer Key

Exercise 1 Suppose you model a variable Y depending on a non-stochastic variable X, according to the relation $Y_i = \beta X_i + u_i$. The errors are iid with mean zero and variance σ^2. Considering estimating β by the slope of a line between the origin and one out of the n plotted observations (X_i, Y_i).

(a) Is this estimator unbiased?

Yes. Note that in our case $\widehat{\beta}$ = slope between $(0,0)$ and $(X_i, Y_i) = \frac{Y_i}{X_i}$. Then

$$\begin{aligned} E(\widehat{\beta}) &= E\left(\frac{Y_i}{X_i}\right) = E\left(\frac{\beta X_i + u_i}{X_i}\right) \\ &= E\left(\frac{\beta X_i}{X_i}\right) + E\left(\frac{u_i}{X_i}\right) \\ &= \underbrace{E(\beta)}_{\beta} + \frac{1}{X_i}\underbrace{E(u_i)}_{=0} \\ &= \beta. \end{aligned}$$

(b) Find the variance of this estimator.

Again, by construction

$$\begin{aligned} var(\widehat{\beta}) &= var\left(\frac{Y_i}{X_i}\right) = var\left(\frac{\beta X_i + u_i}{X_i}\right) \\ &= var\left(\beta + \frac{u_i}{X_i}\right) = var\left(\frac{u_i}{X_i}\right) \\ &= \frac{1}{X_i^2} var(u_i) \\ &= \frac{\sigma^2}{X_i^2}. \end{aligned}$$

(c) Which of the observations would you choose to calculate the estimate? Justify your answer.

Looking at the answer in (b), we would pick the observation with the greatest X_i, since this would minimize the variance of the estimator.

Exercise 2 You are given 30 pairs of observations (X_i, Y_i) which are to be represented by the following model $Y_i = \beta_0 + \beta_1 X_i + u_i$, and the errors are iid with mean zero and variance σ^2. Suppose you know that $R^2 = 0.25$, $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 49$ and $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = 49$. Calculate the OLS estimate for β_1 (assume is positive).

From class we got the following formula:

$$
\begin{aligned}
R^2 &= \frac{(\widehat{\beta}_1)^2 \sum_{i=1}^{n}(X_i - \overline{X})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\
0.25 &= \frac{(\widehat{\beta}_1)^2 (49)}{(49)} \\
0.25 &= (\widehat{\beta}_1)^2 \\
&\Rightarrow \widehat{\beta}_1 = 0.5
\end{aligned}
$$

Exercise 3 (*Regression without any regressor*) Suppose you are given the model $Y_i = \beta_1 + u_i$. Use OLS to find the estimator of β_1. What is its variance and RSS? What is the coefficient of determination? (R^2)

To find the OLS estimator, we follow the same approach as in class:

$$
\begin{aligned}
& \min \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \\
\Leftrightarrow\ & \min_{\widehat{\beta}} \sum_{i=1}^{n}(Y_i - \widehat{\beta} X_i)^2
\end{aligned}
$$

In order to minimize this function, we take the First Order (Necessary) Condition:

$$
\begin{aligned}
\frac{\partial(\sum_{i=1}^{n}(Y_i-\widehat{\beta}X_i)^2)}{\partial\widehat{\beta}} &= \sum_{i=1}^{n} -2X_i(Y_i-\widehat{\beta}X_i)=0 \\
&\Rightarrow \sum_{i=1}^{n} X_iY_i = \widehat{\beta}\sum_{i=1}^{n} X_i^2 \\
&\Rightarrow \widehat{\beta} = \frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}
\end{aligned}
$$

Note that this is only a necessary condition for a minimum. In order to guarantee that this is actually a minimum (and not a maximum, for example) we need to check the Second Order (Sufficient) Condition:

$$\frac{\partial^2(\sum_{i=1}^{n}(Y_i-\widehat{\beta}X_i)^2)}{\partial\widehat{\beta}^2} = 2\sum_{i=1}^{n} X_i^2 \geq 0$$

Hence we are sure that $\widehat{\beta}$ minimizes the sum of square residuals.

Exercise 4 Suppose you estimate the consumption function

$$Y_i = \alpha_1 + \alpha_2 X_i + u_{1i}$$

and the savings function

$$Z_i = \beta_1 + \beta_2 X_i + u_{2i}$$

where Y is consumption, X is income, Z is saving and $X = Y + Z$, that is income is equal to consumption plus savings.

(a) What is the relationship between $\hat{\alpha}_2$ and $\hat{\beta}_2$? ($\hat{\alpha}_2$ and $\hat{\beta}_2$ are the OLS estimators for α_2 and β_2) Show your calculations.

Note that OLS estimators are given by

$$\hat{\alpha}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_2 = \frac{\sum x_i z_i}{\sum x_i^2}$$

Using the relation given in the question, $x_i = y_i + z_i$ we obtain

$$\hat{\beta}_2 = \frac{\sum x_i z_i}{\sum x_i^2} = \frac{\sum x_i(x_i - y_i)}{\sum x_i^2} = \frac{\sum x_i^2}{\sum x_i^2} - \frac{\sum x_i y_i}{\sum x_i^2} = 1 - \hat{\alpha}_2$$

(b) Will the residual sum of squares, RSS, be the same for the two models? Explain.

In the deviation form, RSS from the first model, RSS_1, is $\sum(y_i - \hat{\alpha}_2 x_i)^2$ and RSS from the second model, RSS_2, is $\sum(z_i - \hat{\beta}_2 x_i)^2$. Again using $x_i = y_i + z_i$ we get:

$$RSS_2 = \sum(z_i - \hat{\beta}_2 x_i)^2 = \sum(x_i - y_i - (1 - \hat{\alpha}_2)x_i)^2 = \sum(-y_i + \hat{\alpha}_2 x_i)^2 = RSS_1$$

Exercise 5 Consider the following models:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_{1i}$$

$$(Y_i - X_{2i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_{2i}$$

Rewrite the second model as:

$$Y_i = \beta_1 + (1 + \beta_2)X_{2i} + \beta_3 X_{3i}$$

Now, two models have the same dependent and independent variables, i.e. we are regressing Y on a constant, X_2 and X_3 thus estimated coefficients can be compared.

(a) Will the OLS estimates of α_1 and β_1 be the same?

Yes, the OLS estimates for α_1 and β_1 will be the same.

(b) Will OLS estimates of α_3 and β_3 be the same?

Yes, the OLS estimates for α_3 and β_3 will be the same because both are the estimated coefficient of X_3 in a regression Y on X_2 and X_3.

(c) What is the relationship between α_2 and β_2?

No, the OLS estimates for α_2 and β_2 will not be the same. The relationship is $\hat{\alpha}_2 = 1 + \hat{\beta}_2$.

(d) Can you compare the R^2 of the two models? Why or why not?

No, because the dependent variables in the two models are different.

Exercise 6 Suppose you have the following model $Y_i = \beta_1 + \beta_2 X_i + u_i$ where u_i satisfies the assumptions of the classical linear regression model. Suppose you have fixed values for X's: $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, $X_4 = 4$, $X_5 = 5$ and $X_6 = 6$. An econometrician with no calculator and an aversion to arithmetic proposes to estimate the slope of the linear relation between Y and X by:

$$\hat{\beta}_2 = \frac{1}{8}(Y_6 + Y_5 - Y_2 - Y_1)$$

Show that this estimator is unbiased, i.e. $E(\hat{\beta}_2) = \beta_2$

Note that OLS estimator can be written as

$$\hat{\beta}_2 = \frac{1}{8}\left[\beta_1 + \beta_2 X_6 + u_6 + \beta_1 + \beta_2 X_5 + u_5 - (\beta_1 + \beta_2 X_2 + u_2) - (\beta_1 + \beta_2 X_1 + u_1)\right]$$

Taking expectations of both sides

$$E[\hat{\beta}_2] = E\left(\frac{1}{8}\left[(\beta_2 X_6 + u_6) + (\beta_2 X_5 + u_5) - (\beta_2 X_2 + u_2) - (\beta_2 X_1 + u_1)\right]\right)$$

Because X values are constant we obtain

$$E[\hat{\beta}_2] = \frac{1}{8}\left[(\beta_2 X_6 + E(u_6)) + (\beta_2 X_5 + E(u_5)) - (\beta_2 X_2 + E(u_2)) - (\beta_2 X_1 + E(u_1))\right]$$

which is equal to

$$E[\hat{\beta}_2] = \frac{1}{8}\left[\beta_2 X_6 + \beta_2 X_5 - \beta_2 X_2 - \beta_2 X_1\right]$$

since $E[u_i] = 0$ for all i. Substituting the X values

$$E[\hat{\beta}_2] = \frac{1}{8}\left[6\beta_2 + 5\beta_2 - 2\beta_2 - 1\beta_2\right] = \beta_2$$

Exercise 7 Let $\hat{\beta}_{YX}$ and $\hat{\beta}_{XY}$ represent OLS estimates of the slopes in the regression of Y on X and X on Y, respectively. Show that

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

where r is the coefficient of correlation between X and Y.

The OLS estimate of the slope from the regression Y on X is given by:

$$\hat{\beta}_{YX} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

and the OLS estimate of the slope from the regression Y on X is given by:

$$\hat{\beta}_{XY} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2}$$

Multiplying the two, we obtain the expression for r^2,squared sample correlation coefficient,

$$r^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}$$

Exercise 8 In class we showed that the OLS estimator of β_1 could be written as

$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$ or $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$

Prove that $\widehat{\beta}_1$ can also be written as

$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})Y_i}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$ or $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$.

Note that if we have

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \\
&= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\left(\frac{\sum Y_i}{n}\right)}{\sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2} \\
&= \frac{\sum_{i=1}^{n} X_i Y_i - \overline{X}\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - 2n\overline{X}\left(\frac{\sum X_i}{n}\right) + n\overline{X}^2} \\
&= \frac{\sum_{i=1}^{n}(X_i Y_i - \overline{X}Y_i)}{\sum_{i=1}^{n}(X_i^2 - 2\overline{X}X_i + \overline{X}^2)} \\
&= \frac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ and we prove the first part.}
\end{aligned}$$

To prove the second part, we follow a similar approach,

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^n X_iY_i - n\overline{XY}}{\sum_{i=1}^n X_i^2 - n\overline{X}^2} \\
&= \frac{\sum_{i=1}^n X_iY_i - n\overline{Y}\left(\frac{\sum X_i}{n}\right)}{\sum_{i=1}^n X_i^2 - 2n\overline{X}^2 + n\overline{X}^2} \\
&= \frac{\sum_{i=1}^n X_iY_i - \overline{Y}\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - 2n\overline{X}\left(\frac{\sum X_i}{n}\right) + n\overline{X}^2} \\
&= \frac{\sum_{i=1}^n (X_iY_i - \overline{Y}X_i)}{\sum_{i=1}^n (X_i^2 - 2\overline{X}X_i + \overline{X}^2)} \\
&= \frac{\sum_{i=1}^n X_i(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} \text{ and this completes the proof.}
\end{aligned}$$