**ECO 441 Fall 2009**

**Solutions-3**

1) The linear regression model with single regressor is

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

Taking the average of the simple regression yields

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u} \tag{2}$$

Subtracting equation (2) from equation (1)

$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u}) \tag{3}$$

We know that the estimate of $\beta_1$ is

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \tag{4}$$

Substitute (3) in (4)

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})[\beta_1 (X_i - \bar{X}) + (u_i - \bar{u})]}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \beta_1 \frac{\sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

Since $\sum (X_i - \bar{X})\bar{u} = 0$, $\hat{\beta}_1$ turns out to be

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (X_i - \bar{X}) u_i}{\sum (X_i - \bar{X})^2}$$

Dividing the numerator and the denominator of $\frac{\sum (X_i - \bar{X}) u_i}{\sum (X_i - \bar{X})^2}$ by $1/N$ yields

$$\hat{\beta}_1 = \beta_1 + \frac{Cov(X, u_i)}{Var(X)}$$

$Cov(X, u_i) = 0 \Leftrightarrow E(u_i|X_i) = 0$. Therefore if $E(u_i|X_i) = 0$ does not hold, $Cov(X, u_i) \neq 0$ and $\frac{Cov(X, u_i)}{Var(X)} \neq 0$.

This will lead $E(\hat{\beta}_1)$ to differ from $\beta_1$.

2) By equation (6.15) in the text, we know

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

Thus, the values of $\bar{R}^2$ are $0.175, 0.189$ *and* $0.193$ for columns (1)-(3).

3) a) Workers with college degrees earn $5.46/hour$ more, on average, than workers with only high school degrees.

b) Men earn $2.64/hour$ more, on average, than women.

4) a) On average, a worker earns 0.29/hour more for each year he ages.

b) Sally's earnings prediction is $4.40 + 5.48 * 1 - 2.62 * 1 + 0.29 * 29 = 15.67/hour$. Betsy's earnings prediction is $4.40 + 5.48 * 1 - 2.62 * 1 + 0.29 * 34 = 17.12/hour$. The difference is 1.45.

5) a) 23.4 or 23,400.

b) In this case $\Delta BDR = 1$ and $\Delta Hsize = 100$. The resulting expected change in price is 23.4+0.156*100=39.0 or 39,000.

c) The loss is 48,800.

d) $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) \Rightarrow R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$, thus $R^2 = 0.727$.

6) a) There are other important determinants of a country's crime rate, including demographic characteristics of the population.

b) Suppose that the crime rate is positively affected by the fraction of young males in the population, and that counties with high crime rates tend to hire more police. In this case, the size of the police force is likely to be positively correlated with the fraction if young males in the population leading to a positive value for the omitted variable bias so that $\hat{\beta}_1 > \beta_1$.

7) For omitted variable bias to occur, two conditions must be true: $X$ (the included regressor) is correlated with the omitted variable, and the omitted variable is a determinant of the dependent variable. Since $X_1$ and $X_2$ are uncorrelated, the estimator of $\beta_1$ does not suffer from omitted variable bias.

8) a) $-0.056$

b) $-0.005$

c) The coefficient has fallen substantially and is no longer significant. Therefore, it seems that result in (a) did suffer from omitted variable bias.

d) The regression in (b) fits the data much better as indicated by the $R^2, \bar{R}^2$ and $SER$. The $R^2$ and $\bar{R}^2$ are similar because the number of observations is large.

e) Students with a "dadcoll=1" complete 0.494 more years of education, on average, than students with "dadcoll=0".

f) These terms capture the opportunity cost of attending college. As $STWMFG$ increases, foregone wages increase, so that, on average, college attendance declines. The negative sign on the coefficient is consistent with this. As $CUE80$ increases, it is more difficult to find a job, which lowers the opportunity cost of attending college, so that college attendance increases. The positive sign on the coefficient is consistent with this. However, these coefficients are not statistically significant (**due to small sample size**) and therefore, makes it impossible to make a valid/strong inference.

g) Bob's predicted years of completed schooling is 14.927.

h) Jim's predicted years of completed schooling is 14.916.