Exercise 1 – The General Linear Model : Answers

1. Given the following information on 67 pairs of values on Y and X

$\Sigma_i (Y_i - \overline{Y})^2 = 1.16$
$\Sigma_i (X_i - \overline{X})^2 = 100$
$\Sigma_i (Y_i - \overline{Y})(X_i - \overline{X}) = -9$

a) find the OLS coefficient estimate from a regression of Y on X.

Using $\hat{b} = \dfrac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} = \dfrac{\sum_i x_i Y_i}{\sum_i x_i^2} = \dfrac{-9}{100}$

So $\hat{\beta} = -0.09$

b) Suppose that Y now also depends on another variable Z and that
$\Sigma_i (Z_i - \overline{Z})^2 = 100 \quad \Sigma_i (X_i - \overline{X})(Z_i - \overline{Z}) = -80$
$\Sigma_i (Y_i - \overline{Y})(Z_i - \overline{Z}) = 10$

Find the coefficients of X and Z in the OLS regression of Y on X and Z. (Hint use the fact that the set of regressors in their mean deviation form = $\Sigma_i x_i^2$ which is equivalent in matrix terms to X'X and use matrix algebra to solve).

We know that the OLS slope parameters can be obtained from estimating the model in mean deviation form (see Exercise 0)

Let the mean deviation version of the model $Y_i = b_1X_i + b_2Z_i + u_i$ be written as

$$y_i = b_1x_i + b_2z_i + u_i$$

where $y_i = Y_i - \overline{Y}_i \quad z_i = \left(Z_i - \overline{Z}_i \right)$ and $U_i = u_i$ since the mean of the residual term is assumed to be zero

Easier to solve this 3 variable model using matrix algebra

Let $w = [x \quad : \quad z]$ where x is the N_x1 vector of observations on the X variable in mean deviation form

Hence $x = \begin{bmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ : \\ X_N - \overline{X} \end{bmatrix}$

and model becomes y = wb + u

OLS gives $\hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = (w'w)^{-1} w'y$

$$\hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 100 & -80 \\ -80 & 100 \end{bmatrix} \begin{bmatrix} -9 \\ 10 \end{bmatrix} = \frac{1}{100^2 - (-80)^2} \begin{bmatrix} 100 & -80 \\ -80 & 100 \end{bmatrix} \begin{bmatrix} -9 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} 100/360 & -8/360 \\ -8/360 & 10/360 \end{bmatrix} \begin{bmatrix} -9 \\ 10 \end{bmatrix}$$

so $\hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} -0.028 \\ 0.078 \end{bmatrix}$

Why does the coefficient of X in part (b) differ from that in part (a)? What conclusions about the estimated relationship can you draw?

Hence the size of the estimates coefficient on the X variable increased compared with the estimate from the original 2 variable model. This confirms the negative covariance between X and Z that is implicit in the variance-covariance matrix of explanatory variables.

Since lecture notes show that $\hat{b}_1 = (x_1'x_1)^{-1} x_1'y - (x_1'x_1)^{-1}(x_1'X_2)\hat{\beta}_2$

If $\hat{\beta}_2 > 0$ since $(X'X)^{-1} > 0$ then $\hat{\beta}_1^{3\,\text{var}} > \hat{\beta}_1^{2\,\text{var}}$ if and only if X'Z <0

2. A regression of total cost on output produces the following coefficients and standard errors

Total $Cost_i$ = 141.77 + 63.48$Output_i$ - 12.96$Output_i^2$ + 0.94$Output_i^3$
(6.38) (4.78) (0.99) (0.06)

Do the results conform to the theoretical expectations about typical marginal and average cost curves?

Microeconomic theory tells us that total cost curves usually display increasing and then decreasing returns to scale. This shape can be captured by a cubic (3rd degree polynomial)

where the constant is an estimate of the fixed cost (should therefore be positive)

Can see from above that this is true

We also know that the marginal cost and average cost curves should be U-shaped each with a positive minimum value
Since the MC curve can be found by differentiating the total cost curve with respect to output

$$MC = \frac{dTC}{dQ} = \beta_1 + 2\beta_2 Q + 3\beta_3 Q^2$$

and for any quadratic to be U-shaped (rather than Λ-shaped) then need $\beta_3>0$

which is satisfied in the output above

(Similarly the AC curve $AC = \frac{TC}{Q} = \frac{\beta_0}{Q} + \beta_1 + \beta_2 Q + \beta_3 Q^2$

needs $\beta_2<0$ and $\beta_3>0$ to be Λ-shaped)

The level of output at which MC is minimised is given by

$$\frac{dMC}{dQ} = 0 = 2\beta_2 Q + 6\beta_3 Q \qquad \text{so } Q^* = -\frac{\beta_2}{3\beta_3} \qquad \text{at the minimum}$$

Since Q* (by assumption) is >0 and we have established that $\beta_3>0$ then $\beta_2<0$ is a necessary condition for the minimum output to exist and this is therefore consistent with the existence of Λ-shaped MC and AC curves

To ensure that the minimum output value of the MC curve >0

$$MC = \frac{dTC}{dQ} = \beta_1 + 2\beta_2 Q + 3\beta_3 Q^2 > 0 \quad \text{for all Q}$$

sub in min Q value from above

$$MC = \frac{dTC}{dQ} = \beta_1 + 2\beta_2\left(-\frac{\beta_2}{3\beta_3}\right) + 3\beta_3\left(-\frac{\beta_2}{3\beta_3}\right)^2 = \frac{3\beta_3\beta_1 - \beta_2^2}{3\beta_3} > 0$$

so >0 iff $3\beta_3\beta_1 >= \beta_2^2$

which will be the case iff $\beta_1>0$

So all the estimated values in the output above are consistent with the requirements of cost curves

3. Given $(X'X)^{-1}\hat{\beta}$ = X'y, write down the normal equations in a 3 variable model estimated by least squares.

In the 3 variable model $\underset{(n_x k)}{X} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ : & : & : \\ 1 & X_{1N} & X_{2N} \end{bmatrix}$

So

$$\underset{(k_x k)}{X'X} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ : & : & : \\ 1 & X_{1N} & X_{2N} \end{bmatrix}\begin{bmatrix} 1 & 1 & .. & 1 \\ X_{11} & X_{12} & .. & X_{1N} \\ X_{21} & X_{22} & .. & X_{2N} \end{bmatrix} = \begin{bmatrix} N & \Sigma_i X_{1i} & \Sigma_i X_{2i} \\ \Sigma_i X_{1i} & \Sigma_i X_{1i}^2 & \Sigma_i X_{1i}X_{2i} \\ \Sigma_i X_{2i} & \Sigma_i X_{1i}X_{2i} & \Sigma_i X_{2i}^2 \end{bmatrix}$$

and $(X'X)^{-1}\hat{\beta}$ = X'y becomes $\begin{bmatrix} N & \Sigma_i X_{1i} & \Sigma_i X_{2i} \\ \Sigma_i X_{1i} & \Sigma_i X_{1i}^2 & \Sigma_i X_{1i}X_{2i} \\ \Sigma_i X_{2i} & \Sigma_i X_{1i}X_{2i} & \Sigma_i X_{2i}^2 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_i y_i \\ \Sigma_i X_{1i} y_i \\ \Sigma_i X_{2i} y_i \end{bmatrix}$

multiplying through gives

$$N\hat{\beta}_0 + \hat{\beta}_1 \Sigma_i X_{1i} + \hat{\beta}_2 \Sigma_i X_{2i} = \Sigma_i y_i \qquad (1)$$

$$\hat{\beta}_0 \Sigma_i X_{1i} + \hat{\beta}_1 \Sigma_i X_{1i}^2 + \hat{\beta}_2 \Sigma_i X_{1i}^2 = \Sigma_i X_{1i} y_i \qquad (2)$$

$$\hat{\beta}_0 \Sigma_i X_{2i} + \hat{\beta}_1 \Sigma_i X_{1i}X_{2i} + \hat{\beta}_2 \Sigma_i X_{2i}^2 = \Sigma_i X_{2i} y_i \qquad (3)$$

$$(1) \div N \Rightarrow \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 = \bar{Y} \qquad (4)$$

so in the 3 variable model the OLS fitted line passes through the hyperplane of means

Sub. (4) into (2)

$$(\bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2)\Sigma_i X_{1i} + \hat{\beta}_1 \Sigma_i X_{1i}{}^2 + \hat{\beta}_2 \Sigma_i X_{1i}{}^2 = \Sigma_i X_{1i} y_i$$

$$\Rightarrow \hat{\beta}_1\left(\Sigma_i X_{1i}{}^2 - \bar{X}_1 \Sigma_i X_{1i} \right) + \hat{\beta}_2\left(\Sigma_i X_{1i}{}^2 - \bar{X}_2 \Sigma_i X_{1i} \right) = \Sigma_i X_{1i} y_i - \bar{Y} \Sigma_i X_{1i}$$

Using the fact that $\Sigma_i X_{1i}{}^2 - \bar{X}_1 \Sigma_i X_{1i} = \Sigma_i X_{1i}{}^2 - N \bar{X}_1{}^2 = \Sigma_i (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)$

(2) becomes

$$\hat{\beta}_1 \Sigma_i (X_{1i} - \bar{X}_1)^2 + \hat{\beta}_2 \Sigma_i (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = \Sigma_i (X_{1i} - \bar{X}_1)(Y_i - \bar{Y})$$

and (3) becomes

$$\hat{\beta}_1 \Sigma_i (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) + \hat{\beta}_2 \Sigma_i (X_{2i} - \bar{X}_2)^2 = \Sigma_i (X_{2i} - \bar{X}_2)(Y_i - \bar{Y})$$

Solving simultaneously and using mean deviation notation

$$\hat{\beta}_1 = \frac{\Sigma x_1 y \Sigma x_2^2 - \Sigma x_2 y \Sigma x_1 x_2}{\Sigma x_1{}^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \qquad (5)$$

$$\hat{\beta}_2 = \frac{\Sigma x_2 y \Sigma x_1^2 - \Sigma x_1 y \Sigma x_1 x_2}{\Sigma x_1{}^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \qquad (6)$$

From (6) if x_1 was absent from the model then the ols estimate of b_2 would be

$$\tilde{\beta}_2 = \frac{\Sigma x_2 y}{\Sigma x_2^2}$$

Since the correlation coefficient squared between x_1 and x_2 is $r^2_{x_1 x_2} = \dfrac{\Sigma (x_1 x_2)^2}{\Sigma x_1^2 \Sigma x_2^2}$

then (6) can be written as $\hat{\beta}_2 = \dfrac{\tilde{\beta}_2}{1 - r^2_{x_1 x_2}} - \dfrac{\beta_{yx_1} \beta_{x_1 x_2}}{1 - r^2_{x_1 x_2}}$

so the slope estimate in the 3 variable regression contains a correction to the slope estimate from the simple 2 variable regression that accounts dfor the effect of the additional variable on both y and x_2

If the correlation between x_1 and x_2 is zero then the slope estimates in the 3 variable and the 2 variable model are the same

4. Show that the R2 can be interpreted as the square of the correlation coefficient between the actual and fitted values of the dependent variable.

From exercise 0 we know that the slope coefficients from a model can always be obtained by estimating the model in mean deviation form

$$y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots \hat{\beta}_k x_{ki} \qquad (1)$$

Can write (1) in a more compact matrix form using the following notation

Let the mean deviation matrix A be given by

$$A = I_N - \left(\frac{1}{N}\right) ii'$$

where I is the identity matrix of order N and i is an N_x1 column vector of ones $\quad i = \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix}$

multiplication by its transpose gives an N_xN matrix of ones

so

$$A = \begin{bmatrix} 1 & 0 & & \\ 0 & 1 & & \\ & & .. & \\ & & & 1 \end{bmatrix} - \frac{1}{N} \begin{bmatrix} 1 & 1 & .. & 1 \\ 1 & 1 & & \\ : & & .. & \\ 1 & & & 1 \end{bmatrix}$$

has the properties that it is both symmetric (A=A') and idempotent ($A=A^2$)

and any matrix that is pre-multiplied by this matrix A will have the property that the resulting elements will be in mean deviation form

Proof by example: Consider the vector $x = \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix}$

Then $Ax = \left(\begin{bmatrix} 1 & 0 & & \\ 0 & 1 & & \\ & & .. & \\ & & & 1 \end{bmatrix} - \frac{1}{N} \begin{bmatrix} 1 & 1 & .. & 1 \\ 1 & 1 & & \\ : & & .. & \\ 1 & & & 1 \end{bmatrix} \right) * \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix}$

$$= \left(\begin{bmatrix} 1 & 0 & & \\ 0 & 1 & & \\ & & .. & \\ & & & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix} \right) - \frac{1}{N} \begin{bmatrix} 1 & 1 & .. & 1 \\ 1 & 1 & & \\ : & & .. & \\ 1 & & & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix}$$

Since multiplying any matrix (or vector) by the identity matrix leaves the vector unchanged (equivalent to multiplication by one)

$$Ax = \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix} - \frac{1}{N} \begin{bmatrix} 1 & 1 & .. & 1 \\ 1 & 1 & & \\ : & & .. & \\ 1 & & & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_N \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{x} \\ : \\ \bar{x} \end{bmatrix} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ : \\ x_N - \bar{x} \end{bmatrix}$$

(note that a single mean value can always be written as $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{N} i' x$)

To show that the R^2 is the square of the correlation coefficient between the actual and fitted values

Given $R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}' X' A X \hat{\beta}}{Y' A Y}$

$$= \frac{\hat{Y}'A\hat{Y}}{Y'AY} \qquad \text{since } \hat{Y} = X\hat{\beta}$$

Since

$Y = \hat{Y} + \hat{u}$ then $AY = A\hat{Y} + A\hat{u} = A\hat{Y} + \hat{u}$

(since the mean of OLS residuals is zero then $A\hat{u} = \hat{u}$)

So $\hat{Y}'AY = \hat{Y}'A\hat{Y} + \hat{Y}'\hat{u} = \hat{Y}'A\hat{Y}$

(since $\hat{Y}'\hat{u} = \hat{\beta}'X'\hat{u} = 0$ using the algebra of least squares (lecture notes))

Hence $R^2 = \dfrac{\hat{Y}'A\hat{Y}}{Y'AY} = \dfrac{\hat{Y}'AY}{Y'AY}$

Now the square of the correlation coefficient

$$r^2_{y\hat{y}} = \frac{\left[\sum_i (\hat{Y}-\bar{\hat{Y}})(Y-\bar{Y})\right]^2}{\sum_i (\hat{Y}-\bar{\hat{Y}})^2 \sum_i (Y-\bar{Y})^2} = \frac{\sum_i (\hat{Y}-\bar{\hat{Y}})^2 \sum_i (\hat{Y}-\bar{\hat{Y}})^2}{\sum_i (\hat{Y}-\bar{\hat{Y}})^2 \sum_i (Y-\bar{Y})^2} = \frac{\sum_i (\hat{Y}-\bar{\hat{Y}})^2}{\sum_i (Y-\bar{Y})^2} = \frac{\hat{Y}'A\hat{Y}}{Y'AY} = R^2$$

Note It follows that the OLS slope coefficients can always be obtained if the regression is run in mean deviation form

Given $y = X\hat{\beta} + \hat{u}$

Multiply by the transformation matrix A

$$Ay = AX\hat{\beta}_2 + A\hat{u}$$

Note that multiply by A eliminates 1st coefficient in B vector (that on the constant) since

$$Ay = \left[Ai \mid AX_2\right]\begin{bmatrix} \hat{\beta}1 \\ \hat{B}2 \end{bmatrix} + A\hat{u} = AX_2\,\hat{B}2 + A\hat{u}$$

since Ai=0

Let Ay=y* and AX=X* and since $A\hat{u} = \hat{u}$

$$y^* = X^*\hat{\beta}_2 + \hat{u}$$

Pre-multiply by X*'

$$X^{*\prime}y^* = X^{*\prime}X^*\hat{\beta}_2 + X^{*\prime}\hat{u}$$

which since $X^{*\prime}\hat{u} = 0$ gives

$$\hat{\beta}_2 = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

ie the slope coefficients if OLS is run in mean deviation form

5 Given the following sets of information, say whether your believe an error has been committed during the course of the estimation process.

a) $R^2_{y.123} = 0.95$ $R^2_{y.1234} = 0.9$

b) $\bar{R}^2_{y.123} = 0.86$ $\bar{R}^2_{y.1234} = 0.82$

a) There must be an error since we know that the unadjusted R^2 will not fall if an extra variable is added to the model

Proof:
Partition the model such that

$$Y = x_1 b_1 + X_2\,\beta_2 + u$$

Where x_1 is an N_x1 vector b_1 is a scalar
X_2 is an $N_x(k\text{-}1)$ matrix β_2 is a (k-1)x1 vector

and $X = [x_1 : X_2]$

The OLS normal equations $X'X\hat{\beta} = X'y$ can now be written as

$$\begin{bmatrix} x_1'x_1 & x_1'X_2 \\ X_2'x_1 & X_2'X_2 \end{bmatrix}\begin{bmatrix} \hat{b}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} x_1'y \\ X_2'y \end{bmatrix} \qquad (1)$$

2nd row can be written

$$X_2'x_1\hat{b}_1 + (X_2'X_2)\hat{\beta}_2 = X_2'y$$

so

$$\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y - (X_2'X_2)^{-1}(X_2'x_1)\hat{b}_1$$

$$\hat{\beta}_2 = \tilde{\beta}_2 - (X_2'X_2)^{-1}(X_2'x_1)\hat{b}_1 \text{ (2)} \qquad (2)$$

(where $\tilde{\beta}_2 = (X_2'X_2)^{-1}X_2'y$ is the OLS estimates if y is regressed on X_2 alone)

The OLS residual from model (1) is

$$\hat{u} = y - x_1\hat{b}_1 - X_2\hat{\beta}_2 \qquad (3)$$

Sub. (2) into (3)

$$\hat{u} = y - x_1\hat{b}_1 - X_2\left[\tilde{\beta}_2 - (X_2'X_2)^{-1}X_2'x_1\hat{b}_1\right]$$

$$= y - X_2\tilde{\beta}_2 + X_2(X_2'X_2)^{-1}X_2'x_1\hat{b}_1 - x_1\hat{b}_1 \qquad (4)$$

Let $\hat{e} = y - X_2\tilde{\beta}_2$

Be the vector of OLS residuals when y is regressed on X_2 alone

and let $X_2(X_2'X_2)^{-1}X_2'x_1\hat{b}_1 - x_1\hat{b}_1 = \left[X_2(X_2'X_2)^{-1}X_2' - I\right]x_1\hat{b}_1$

$$= -M_2x_1\hat{b}_1$$

where M_2 is the residual maker matrix $\left[I - X_2(X_2'X_2)^{-1}X_2'\right]$

so now (4) can be written as

$$\hat{u} = \hat{e} - M_2 x_1 \hat{b}_1$$

$$= \hat{e} - x_1^* \hat{b}_1 \qquad (5)$$

(with $x_1^* = M_2 x_1$ being the vector of residuals from a regression of x_1 on X_1)

The residual sum of squares

$$\hat{u}'\hat{u} = (\hat{e} - x_1^* \hat{b}_1)'(\hat{e} - x_1^* \hat{b}_1)$$

$$= (\hat{e}'\hat{e} - 2\hat{b}_1 x_1^* \hat{e})' + \hat{b}_1^2 (x_1^{*'} x_1^*) \qquad (5')$$

Given we know that the OLS estimates can always be obtained by first netting out the influence of X_2 from both y and x_1 – Frisch-Waugh Theorem) then

$$\hat{b}_1 = (x_1^{*'} x_1^{*'})^{-1} x_1^{*'} y^*$$

and since $= \hat{e} = y^* = M_2 y$ then $x_1^{*'} \hat{e} = x_1^{*'} y^* = \hat{b}_1 (x_1^{*'} x_1^*)$ (from (6)) (7)

Sub. (7) into (5')

$$\hat{u}'\hat{u} = \hat{e}'\hat{e} - \hat{b}_1^2 (x_1^{*'} x_1^*) \qquad (8)$$

Since each term on the second term on the rhs >=0

Then the RSS from the equation containing more variables <= the RSS from the model with less variables

Hence

$$\hat{u}'\hat{u} <= \hat{e}'\hat{e}$$

and this implies that the R^2 = 1 –(RSS/TSS) will never fall when another variable is added to a regression model

Q.E.D.

b) Given $\bar{R}^2_{y.123} = 0.86$ $\qquad$ $\bar{R}^2_{y.1234} = 0.82$

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

Then it is possible that the $\bar{R}^2$ can fall if the t ratio on the extra variable <=1 (but this is conditional on the sample size, N)

Since (1) $\Rightarrow \frac{1-\bar{R}^2}{(N-1)} = \frac{(1-R^2)}{(N-k)}$ for any k (k includes the constant)

then $$\frac{1-\bar{R}^2_{123}(N-k)}{(N-1)} = (1-R^2_{123}) \quad (2)$$

and $$\frac{1-\bar{R}^2_{1234}(N-k)}{(N-1)} = (1-R^2_{1234}) \quad (3)$$

Since part a) shows that the addition of extra variables $\rightarrow \uparrow R^2$ then

$(1-R^2_{1234}) < (1-R^2_{123})$

or

$$1-\bar{R}^2_{1234}(N-5) < 1-\bar{R}^2_{123}(N-4) \quad (4)$$

$\therefore$ (1-0.82)*((N-5)<(1-0.86)*(N-4)

0.18*(N-5) < 0.14*(N-4)

N-5 < 0.777(N-4)

N < 8.5

So can be sure that an error has been made unless the sample size falls below 8

[To show that the $\bar{R}^2$ increases when the t value on the variable > 1

Let $\bar{R}_F^2 = 1 - \frac{RSS/(N-k)}{TSS/(N-1)} = 1 - \frac{\hat{u}'\hat{u}/(N-k)}{y'Ay/(N-1)}$ (1)

be the adjusted R^2 when the extra variable x_k is added to the model

Let $\bar{R}_k^2 = 1 - \frac{RSS/(N-k+1)}{TSS/(N-1)} = 1 - \frac{\hat{u}'_k \hat{u}_k/(N-k+1)}{y'Ay/(N-1)}$ (2)

be the adjusted R^2 when the extra variable x_k is not in the model

$$\therefore \bar{R}^2_F - \bar{R}^2_k = (1) - (2)$$

$$= \frac{\hat{u}'_k \hat{u}_k/(N-k+1)}{y'Ay/(N-1)} - \frac{\hat{u}'\hat{u}/(N-k)}{y'Ay/(N-1)} = \left[\frac{(N-1)}{(N-k+1)}\right]\frac{\hat{u}'_k \hat{u}_k}{y'Ay} - \left[\frac{(N-1)}{(N-k)}\right]\frac{\hat{u}'\hat{u}}{y'Ay}$$

this difference will be >0 iff the ratio of these two terms >1

$$\Rightarrow \frac{\hat{u}'_k \hat{u}_k/(N-k+1)}{\hat{u}'\hat{u}/(N-k)} > 1 \qquad (3)$$

But we know that the RSS from the full model

$$\hat{u}'\hat{u} = \hat{u}_k'\hat{u}_k - \hat{b}_k^2 (x_k^{*'}Ax_k^*)$$

where the 2nd term on the right hand side is the contribution of x_k to the explained sum of squares (in mean deviation form)– see lecture notes –

So (3) becomes $\Rightarrow \frac{\hat{u}'\hat{u} + \hat{b}_k^2 (x_k^{*'}Ax_k^*)/(N-k+1)}{\hat{u}'\hat{u}/(N-k)} > 1$

$$\Rightarrow \frac{\left[\hat{u}'\hat{u} + \hat{b}_k^2 (x_k^{*'}Ax_k^*)\right](N-k)}{\left[(N-k)\hat{u}'\hat{u} + \hat{u}'\hat{u}\right]} > 1$$

since we know the estimated RSS $\hat{u}'\hat{u} = s^2(N-k)$

$$\Rightarrow \frac{\left[\hat{u}'\hat{u} + \hat{b}_k^2 (x_k^{*'} A x_k^*)\right]}{\left[s^2 + \hat{u}'\hat{u}\right]} > 1$$

and since each term in the numerator is a "squared" scalar value must be positive then the fraction can only be >1 iff

$$\Rightarrow \frac{\left[\hat{b}_k^2 (x_k^{*'} A x_k^*)\right]}{\left[s^2\right]} > 1$$

but this is just the square of the estimated t value on the variable x_k. Hence the $\bar{R}^2$ increases when the t value on the variable > 1]

6. Consider the multiple regression model y = XB + u Suppose the independent variables are subject to a linear transformation Z=XΛ where Λ is a diagonal matrix of constants. Show that the residuals from the regression of y on Z are the same as the residuals from a regression of y on X. Compare the coefficient estimates from the 2 regressions.

This is a general proof of the result that rescaling a variable in a model multiplies the original ols estimate by the reciprocal value of the rescaling constant

The transformation matrix appears

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & .. & \\ 0 & & & \lambda_k \end{bmatrix}$$

Given

(1) y = XB + u (2) y = Zγ +v

OLS residuals from both models are

$\hat{u} = y - X\hat{\beta}$ $\hat{v} = y - Z\hat{\gamma}$

$= y - X(X'X)^{-1}X'y$ $= y - Z(Z'Z)^{-1}Z'y$

$= [I – X'(X'X)^{-1}X']y$ $= [I - Z(Z'Z)^{-1}Z']y$

$= [I – X\Lambda(\Lambda'X'X\Lambda)^{-1}X']y$
$= [I – X('X'X)^{-1}X']y$
$= \hat{u}$

so the residuals are the same in both specifications
(and if the residuals are the same then so must be the R^2 values)

From (2) the coefficient vector $\hat{\gamma} = (Z'Z)^{-1}Z'y$
$= \Lambda'^{-1}(X'X)^{-1}\Lambda^{-1}\Lambda'X'y$
$= \Lambda^{-1}(X'X)^{-1}\Lambda'^{-1}\Lambda'X'y$

(using properties of matrix inverses $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$)

$= \Lambda^{-1}(X'X)^{-1}X'y$

$$\hat{\gamma} = \Lambda^{-1}\hat{\beta} \qquad (3)$$

hence the original OLS coefficients are rescaled by the inverse of the rescaling constants contained in the matrix Λ

(Note that the predicted values in the two models are the same since
$\hat{y} = Z\hat{\gamma} = X\Lambda\Lambda^{-1}\hat{\beta} = X\hat{\beta}$)

Show that if instead just one independent variable is multiplied by a constant, λ, then the corresponding regression coefficient is multiplied by 1/ λ and all other coefficients are unchanged.

If the variable to be transformed is X_j n this case the transformation matrix looks like

$$\Lambda = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \lambda_j & \\ 0 & & & 1 \end{bmatrix}$$

ie a diagonal matrix with ones down the main diagonal except for the jth element which contains the constant of multiplication for the j^{th} variable

Since the inverse of a diagonal matrix is also diagonal with the reciprocal of each original element on the new main diagonal then

$$\Lambda^{-1} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1/\lambda_j & \\ 0 & & & 1 \end{bmatrix}$$

So using the result in (3) it follows that then the corresponding regression coefficient is multiplied by 1/a and all other coefficients are unchanged.

Show the effect of adding a constant to one of the right hand side variables. Show that the result implies that for a variable entered in logs the least squares coefficient is independent of the units in which the variable is measured.

If a constant is added to one of the rhs variables then

$$\Lambda = \begin{bmatrix} 1 & 0 & \lambda_j & 0 \\ & 1 & 0 & 0 \\ & & 1 & \\ 0 & & & 1 \end{bmatrix}$$

ie an upper triangular matrix (zeros everywhere below the main diagonal) with the additive constant λ in the jth column of the 1st row and zeros everywhere else above the main diagonal and ones along the main diagonal

It can be shown (eg Johnston & DiNardo) that

i) the determinant of an upper triangular matrix equals the product of the elements on the main diagonal (so in this case the determinant equals 1)

ii) the inverse of an upper triangular matrix is also upper triangular

$$\Lambda^{-1} = \frac{1}{|\Lambda|}\begin{bmatrix} a_{11} & a_{12} & ..a_{1j} & ..a_{1k} \\ & a_{22} & ..a_{2j} & ..a_{2k} \\ & & a_{jj} & : \\ 0 & & & a_{kk} \end{bmatrix} = \Lambda^{-1} = \begin{bmatrix} a_{11} & a_{12} & ..a_{1j} & ..a_{1k} \\ & a_{22} & ..a_{2j} & ..a_{2k} \\ & & a_{jj} & : \\ 0 & & & a_{kk} \end{bmatrix}$$

where a_{ij} is the relevant adjoint element based on the appropriate cofactor

Hence can show that the cofactors on the main diagonal are all equal to one and the other adjoint elements will all be zero except for a_{1j} which = $-\lambda_j$

So

$$\Lambda^{-1} = \begin{bmatrix} 1 & 0 & -\lambda_j & 0 \\ & 1 & 0 & 0 \\ & & 1 & \\ 0 & & & 1 \end{bmatrix}$$

and $\hat{\gamma} = \Lambda^{-1}\hat{\beta}$ becomes

$$\hat{\gamma} = \begin{bmatrix} 1 & 0 & -\lambda_j & 0 \\ & 1 & 0 & 0 \\ & & 1 & \\ 0 & & & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_j \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 - \lambda_j \hat{\beta}_j \\ \hat{\beta}_2 \\ \hat{\beta}_j \\ \hat{\beta}_k \end{bmatrix}$$

So the OLS estimates of the slope coefficients are unchanged when one variable is transformed by an additive constant, but the OLS estimate of the constant term is reduced by $\lambda_j \hat{\beta}_j$

An example of how this result is sometimes seen in practice concerns logarithmic transformations of variables. Consider a log-linear model

$LogY_i = \beta_1 + \beta_2 LogX_{2i} + \ldots \beta_k LogX_{ki} + u_i$

then if X2i is transformed by λX_{2i} then $Log\ \lambda X_{2i} = Log\ \lambda + LogX_{2i}$

Using the above result, the OLS estimate of the coefficient β_2 will be unchanged but the estimate of the constant will become $\hat{\beta}_1 - \hat{\beta}_2\, Log\lambda$

So when estimating a model in logarithmic form the OLS estimates are invariant to the units of measurement.

7. In the earnings function literature, two specifications are commonly estimated:

$$Yi = a_1 + a_2Ed_i + a_3Age + u_1 \qquad (1)$$

and

$$Yi = b_1 + b_2Ed_i + b_3Experience + u_2 \qquad (2)$$

where Y is (log) earnings, Ed. Is years of education and Experience is years spent in the labour market after leaving school. Estimates of b_2 are typically found to be around twice as large as those for a_2. How can you explain this?

(Hint: Age= Years of Work Experience + Years of Education + constant)

Since Age=Experience+Education+constant
then the 2nd equation is a transformation of the first

Experience =Age-Education-constant

So in (2)

$$
\begin{aligned}
Y_i &= b_1 + b_2Ed_i + b_3(\text{Age-Education-constant}) + u_i \\
&= (b_1 - b_3\text{constant}) + (b_2-b_3)Ed_i + b_3\text{Age} + u_i \\
&= a_1 \qquad\qquad + a_2 \qquad + a_3 \qquad + u_i
\end{aligned}
$$

We know from question 6 that a variable transformation will leave the residuals in an OLS regression unchanged so $u_1 = u_2$

And the coefficients $a_2=b_2-b_3$ so $a_2<b_2$ (assuming $b_3>0$) Hence it is possible that the estimate of a_2 will be much less than that of b_2

(and $a_3 = b_2 - a_2$)

8. Given the equation

$$y_t = b_0 + b_1x_{1t} + b_2x_{2t} + \ldots.\ b_kx_{kt} + e_t \qquad t = 1, 2\ldots T$$

can be written in matrix form as $y = i\beta_0 + XB + e$ (1)

where y and e are Tx1 vectors, i is a Tx1 column vector of ones, X is an Txk matrix of observations on k independent variables, β_0 is the coefficient on the constant term and B is a vector of coefficients on the other k independent variables

Show that the first difference of this equation

$$\Delta y_t = b_1\Delta x_{1t} + b_2\Delta x_{2t} + \ldots.\ b_k\Delta x_{kt} + \Delta e_t \qquad t = 2\ldots T$$

can be written in matrix terms as $Ay = AXB + Ae$ (2)

where A is a (T-1)xT matrix that satisfies the condition $Ai = 0$

What is the expected value of the OLS coefficients estimated using equation (2) ?
What can you say about the variance of these estimates?

The condition $Ai = 0$ and the requirement that $Ay=\Delta y$ mean that

$$\underset{(T-1)xT}{A}\underset{Tx1}{\begin{bmatrix} y_1 \\ y_2 \\ : \\ y_T \end{bmatrix}} = \underset{(T-1)x1}{\begin{bmatrix} y_2 - y_1 \\ y_3 - y_2 \\ : \\ y_T - y_{T-1} \end{bmatrix}}$$

ie can transform a vector by multiplying by the matrix A and get a vector of 1st differences is satisfied by the following transformation matrix

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ : & 0 & 1 & \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

ie an upper triangular matrix with -1 on the main diagonal with the exception of the last row (note also that Ai = 0 since

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ : & 0 & 1 & \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ : \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ : \\ 0 \\ 0 \end{bmatrix}$$

Pre-multiplying (2) by A gives

$$\begin{aligned} Ay &= Ai\beta_0 + AX\beta + Ae \\ &= AX\beta + Ae \end{aligned}$$

which is the model in 1st differences

Given the following information (you can find the data set *ps1.*dta on the course website), from a regression of the log of hourly earnings on a set of explanatory variables, calculate a) the percentage difference in pay between

1) men and women
2) union members and non-union workers
3) graduates and non-graduates
4) ethnic minority female graduate union members with 10 years experience and male white, non-union, no qualifications with 10 years experience

```
reg lhw female ethnic union grad intermediate exper exper2

      Source |       SS       df       MS              Number of obs =   17321
-------------+------------------------------           F(  7, 17313) = 1027.74
       Model |   1810.1063     7  258.586614           Prob > F      =  0.0000
```

```
    Residual |  4356.05643 17313  .251606101           R-squared     =  0.2936
-------------+------------------------------           Adj R-squared =  0.2933
       Total |  6166.16272 17320  .356014014           Root MSE      =   .5016

------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2808364    .007658   -36.67   0.000    -.2958468    -.265826
      ethnic |   .0463221   .0191437     2.42   0.016     .0087985    .0838456
       union |   .1666773    .008328    20.01   0.000     .1503535    .1830011
        grad |   .8020626   .0138183    58.04   0.000     .7749773    .8291479
       inter |   .2721199   .0102153    26.64   0.000     .2520969    .2921428
       exper |   .0441852   .0011788    37.48   0.000     .0418746    .0464958
      exper2 |  -.0007928   .0000256   -30.94   0.000     -.000843   -.0007425
       _cons |   1.276704   .0155655    82.02   0.000     1.246194    1.307215
------------------------------------------------------------------------------

Note:
Female =1 if female , = 0 if male
ethnic =1 if individual is from an ethnic minority = 0 otherwise
union =1 if member of trade union, = 0 otherwise
grad=1 if have a degree, = 0 otherwise
intermed=1 if intermediate qualifications = 0 otherwise
```

Hint: read "The Interpretation of Dummy Variables in Semilogarithmic Equations", by R. Halvorsen and R. Palmquist, American Economic Review, 1980, Vol. 70, June, No. 3, pp. 474-475. – available through JSTOR.
http://uk.jstor.org/view/00028282/di950074/95p0167n/0?currentResult=00028282%2bdi950074%2b95p0167n%2b0%2c01%2b19800600%2b9993%2b80199399&searchID=8258cb3a.10651042250&frame=noframe&sortOrder=SCORE&userID=86db14f1@rhbnc.ac.uk/028258cb3a005031220e&dpi=3&viewContent=Article&config=jstor

Given a semi-log wage equation

$$LnW_i = a + bX_i + cD_i + u_i \qquad \text{(A)}$$

where

X is a continuous variable

D is a dummy variable = 1 if attribute satisfied

= 0 otherwise

The coefficient on a continuous variable in a semi-log model

$$\hat{b} = \frac{dLnW}{dX} = \frac{dW/W}{dX} \qquad (1)$$

= % change in wages wrt a small unit change in X (divided by 100)

Since a dummy variable is dichotomous the derivative does not exist

The discrete equivalent to (1) is the proportionate change in W when the dummy variable goes from zero to one

$$g = \frac{W_{D=1} - W_{D=0}}{W_{D=0}}$$

$$\Rightarrow W_{D=1} = (1+g)W_{D=0}$$

Taking natural logs

$Ln(W_{D=1}) = Ln(1+g) + Ln(W_{D=0})$
$Ln(W_{D=1}) - Ln(W_{D=0}) = Ln(1+g)$ (2)

Now the coefficient c in (A) measures the difference in log wages with D= 1 or D=0
Which from (2) equals Ln(1+g)

Hence the proportionate change in wages $g = \exp(c)-1 \neq c$

Expanding $c = Ln(1+g)$

$C = g - 1/2g^2 + 1/3g^3 - \ldots.$

So for small g then $c \approx g$

But for large (+ve) g then c < (true) g ie estimated coefficient under-estimates true % effect

and for large (-ve) g then c > (true) g

Suggests need to transform the semi-log estimate for any coefficient ≈ > 0.25 (in absolute value) otherwise will mis-measure the proportionate effect of the dummy variable

ie use the transformation

$$\exp(c)-1=g$$

Given the estimates in the problem set then the coefficient on the female dummy variable gives the average difference in the log of hourly pay between men and women – other things equal.

With $\hat{\beta}_{female} = 0.28$ this suggests that women earn around 28% less than men

However a more accurate estimate is given using the conversion formula

% difference = $\exp\left(\hat{\beta}_{female}\right) - 1$ = exp(-0.28) -1 = -0.24

so that women earn around 24% less than men

These difference become more noticeable the large the estimate in absolute value

Hence the graduate coefficient suggests a differential of around 80% but the actual estimated percentage difference is given by

$$\exp(0.802) - 1 = 1.23$$

ie 123% more