REVIEW (MULTIVARIATE LINEAR REGRESSION)

- Explain/Obtain the LS estimator (β) of the vector of coefficients (b)

- Explain/Obtain the variance-covariance matrix of β

- Both in the bivariate case (two regressors)

- and in the multivariate case (k regressors)

- Decompose the residual sum of squares

- The coefficient of the multiple correlation

- Information criteria

BIVARIATE REGRESSION

Consider the following bivariate regression

$$Y_i = b_1 + b_2 X_i + e_i,\ i = 1, \ldots, N$$

The above expression can be written in a matrix as

$$\underset{(Nx1)}{Y} = \underset{(Nx2)}{X}\underset{(2x1)}{b} + \underset{(Nx1)}{e},$$

or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

SUM OF THE SQUARED ERRORS

The sum of the squared errors can be written as

$$\sum_{i=1}^{N} e_i^2 = e'e = \begin{bmatrix} e_1 & e_2 & \cdots & e_N \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$= e_1^2 + e_2^2 + \cdots + e_N^2$$

The vector of the errors is given by

$$\underset{(Nx1)}{e} = \underset{(Nx1)}{Y} - \underset{(Nx2)}{X}\underset{(2x1)}{b}$$

$X'X$ FOR THE BIVARIATE CASE

$X'X$ is given by

$$\underset{(2x2)}{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix}$$
$$= \begin{bmatrix} N & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

From the above equation it follows that

$$(X'X)^{-1} = \begin{bmatrix} N & \sum_{i=1}^{N} X_i \\ \sum X & \sum X^2 \end{bmatrix}^{-1}$$
$$= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix}}{N \sum X^2 - (\sum X)^2}$$
$$= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix}}{N \sum x^2}$$

$X'Y$ FOR THE BIVARIATE CASE

Similarly, $X'Y$ is given by

$$\underset{(2x1)}{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

LS ESTIMATOR OF THE SLOPE COEFFICIENT

The LS estimator β is

$$\begin{aligned}
\beta &= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = (X'X)^{-1}X'Y \\
&= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix} \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}}{N \sum x^2} \\
&= \frac{\begin{bmatrix} \sum X^2 \sum Y - \sum X \sum XY \\ N \sum XY - \sum X \sum Y \end{bmatrix}}{N \sum x^2}
\end{aligned}$$

Thus, since $N \sum xy = N \sum XY - \sum X \sum Y$:

$$\beta_2 = \frac{N \sum XY - \sum X \sum Y}{N \sum x^2} = \frac{\sum xy}{\sum x^2}$$

LS ESTIMATOR OF THE CONSTANT

$$\beta = \frac{\left[\begin{array}{c} \sum X^2 \sum Y - \sum X \sum XY \\ N \sum XY - \sum X \sum Y \end{array} \right]}{N \sum x^2}$$

Thus

$$\beta_1 = \frac{\sum X^2 \sum Y - \sum X \sum XY}{N \sum x^2}$$

It can be shown that

$$\beta_1 = \overline{Y} - \beta_2 \overline{X}$$

LS ESTIMATOR β

k VARIABLES

Consider the: i) $Nx1$ vector of the Y's, ii) Nxk matrix of the regressors (X's), iii) $Nx1$ vector of the parameters (b's) and, iv) $Nx1$ vector of the errors (e's)

$$\underset{(Nx1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \underset{(Nxk)}{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2N} & \cdots & X_{kN} \end{bmatrix},$$

$$\underset{(kx1)}{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}, \underset{(Nx1)}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

The multiple regression is

$$Y_i = b_1 + b_2 X_{2i} + \cdots + b_k X_{ki} + e_i$$

In a matrix form it can be written as

$$\underset{(Nx1)}{Y} = \underset{(Nxk)}{X} \underset{(kx1)}{b} + \underset{(Nx1)}{e}$$

The LS estimator of the vector b, as in the bivariate case, is given by $\underset{(kx1)}{\beta} = \underset{(kxk)}{(X'X)^{-1}} \underset{(kx1)}{X'Y}$

VARIANCE-COVARIANCE MATRIX OF THE ERRORS

2 ERRORS

Consider the $2x2$ matrix ee'

$$\underset{(2x2)}{ee'} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \begin{bmatrix} e_1 & e_2 \end{bmatrix} = \begin{bmatrix} e_1^2 & e_1e_2 \\ e_1e_2 & e_2^2 \end{bmatrix}$$

Since $E(e_i) = 0$, $E(e_i^2) = \sigma^2$ (the errors are homoskedastic) and $E(e_1e_2) = 0$ (the errors are serially uncorrelated) the variance-covariance matrix of the vector of errors is given by

$$\underset{(2x2)}{E(ee')} = E \begin{bmatrix} e_1^2 & e_1e_2 \\ e_1e_2 & e_2^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I$$

where I is the identity matrix

VARIANCE-COVARIANCE MATRIX OF THE ERRORS

N ERRORS

If we have N observations:

$$\underset{(NxN)}{ee'} = \begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix}$$

Accordingly, the variance-covariance matrix of the errors, $\underset{(NxN)}{E(ee')}$, is

$$\begin{aligned} \underset{(NxN)}{E(ee')} &= E\begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I \end{aligned}$$

VARIANCE-COVARIANCE MATRIX OF THE LS ESTIMATOR VECTOR β

BIVARIATE CASE

The $2x2$ variance covariance matrix of the vector β is defined as $\underset{(2x2)}{Var(\beta)} = E[\underset{(2x1)}{(\beta - b)}\underset{(1x2)}{(\beta - b)'}]$

Note that if β was a scalar with expected value b then $Var(\beta) = E(\beta - b)^2$

Since β is a $2x1$ vector:

$$\underset{(2x2)}{Var(\beta)} = \begin{bmatrix} Var(\beta_1) & Cov(\beta_1, \beta_2) \\ Cov(\beta_1, \beta_2) & Var(\beta_2) \end{bmatrix}$$

It can be shown that

$$\underset{(2x2)}{Var(\beta)} = \sigma^2 (X'X)^{-1}$$

VARIANCE OF THE ESTIMATOR OF THE SLOPE COEFFICIENT

Recall that

$$(X'X)^{-1} = \left[\begin{array}{cc} \sum X^2 & -\sum X \\ -\sum X & N \end{array} \right] / N \sum x^2$$

Thus, $\underset{(2x2)}{Var(\beta)} = \sigma^2 (X'X)^{-1}$ is given by

$$\begin{aligned} \underset{(2x2)}{Var(\beta)} &= \left[\begin{array}{cc} Var(\beta_1) & Cov(\beta_1, \beta_2) \\ Cov(\beta_1, \beta_2) & Var(\beta_2) \end{array} \right] \\ &= \sigma^2 \left[\begin{array}{cc} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{array} \right] / N \sum x^2 \end{aligned}$$

Moreover, the above expression implies

$$Var(\beta_2) = \frac{\sigma^2 N}{N \sum x^2} = \frac{\sigma^2}{\sum x^2},$$

VARIANCE OF THE ESTIMATOR OF THE CONSTANT

Since

$$\underset{(2x2)}{Var(\beta)} = \sigma^2 \begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix} / N \sum x^2$$

It can be shown that

$$Var(\beta_1) = \sigma^2 [\frac{1}{N} + \frac{\overline{X}^2}{\sum x^2}].$$

BIVARIATE CASE: COVARIANCE BETWEEN THE TWO ESTIMATORS

$$\underset{(2x2)}{Var(\beta)} = \sigma^2 \begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix} / N \sum x^2$$

Finally,

$$Cov(\beta_1, \beta_2) = -\sigma^2 \frac{\sum X}{N \sum x^2} = -\sigma^2 \frac{\overline{X}}{\sum x^2}.$$

VARIANCE-COVARIANCE MATRIX OF THE LS VECTOR β-MULTIVARIATE CASE

For the multiple linear regression we also have

$$\underset{(kxk)}{Var(\beta)} = \sigma^2 \underset{(kxk)}{(X'X)^{-1}}$$

In other words, in this case we have k^2 unknowns: The k variances and the $k(k-1)$ covariances.

DECOMPOSITION OF RESIDUAL SUM OF SQUARES (RSS)

It can be shown that

$$\overbrace{\sum y^2}^{\text{TSS}} = \overbrace{\beta' X' X \beta - N\overline{Y}^2}^{\text{ESS}} + \overbrace{\widehat{e}'\widehat{e}}^{\text{RSS}}$$

TSS: total sum of squares; ESS: explained sum of squares; RSS: residual sum of squares

R^2

We define the R^2 of the regression (the coefficient of the multiple correlation) as

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS-RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

The adjusted R^2 ($\overline{R}^2$) is given by

$$\overline{R}^2 = 1 - \frac{\text{RSS}(N-k)}{\text{TSS}(N-1)}$$

INFORMATION CRITERIA

Next we define two information criteria:

The Schwarz and Akaike information criteria (SIC, AIC respectively):

$$
\begin{aligned}
\text{SIC} &= \ln \frac{\widehat{e}'\widehat{e}}{N} + \frac{k}{N} \ln(N), \\
\text{AIC} &= \ln \frac{\widehat{e}'\widehat{e}}{N} + \frac{2k}{N}
\end{aligned}
$$

the preffered model is the one with the minimum information criterion.

SUMMARY (MULTIPLE LINEAR REGRESSION)

- We obtained the least square estimator of the vector of coefficients:

$$\beta = (X'X)^{-1}X'Y$$

- In the bivariate case this gives us:

$$\begin{aligned}\beta_1 &= \overline{Y} - \beta_2\overline{X}, \\ \beta_2 &= \frac{\sum xy}{\sum x^2}\end{aligned}$$

- We obtained the variance-covariance matrix of β:

$$Var(\beta) = \sigma^2(X'X)^{-1}$$

- In the bivariate case this gives us

$$\begin{aligned} Var(\beta_1) &= \sigma^2[\frac{1}{N} + \frac{\overline{X}^2}{\sum x^2}], \\ Var(\beta_2) &= \frac{\sigma^2}{\sum x^2}, \\ Cov(\beta_1, \beta_2) &= \frac{-\sigma^2\overline{X}}{\sum x^2}, \end{aligned}$$

- We decomposed the TSS:

$$\overbrace{\sum y^2}^{\text{TSS}} = \overbrace{\beta' X' X \beta - N\overline{Y}^2}^{\text{ESS}} + \overbrace{\hat{e}'\hat{e}}^{\text{RSS}}$$

- We defined the R^2 of the regression:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

- We defined two information criteria:

$$\begin{aligned} \text{SIC} &= \ln \frac{\hat{e}'\hat{e}}{N} + \frac{k}{N} \ln(N), \\ \text{AIC} &= \ln \frac{\hat{e}'\hat{e}}{N} + \frac{2k}{N} \end{aligned}$$

PROOFS FROM WEEK 7 ONWARDS

ONLY FOR THE EC5501

SUM OF THE SQUARED ERRORS

The sum of the squared errors can be written as

$$\begin{aligned}\sum_{i=1}^{N} e_i^2 &= e'e = \begin{bmatrix} e_1 & e_2 & \cdots & e_N \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \\ &= e_1^2 + e_2^2 + \cdots + e_N^2\end{aligned}$$

The vector of the errors is given by

$$\underset{(Nx1)}{e} = \underset{(Nx1)}{Y} - \underset{(Nx2)}{X}\underset{(2x1)}{b}$$

The sum of the squared errors, $\sum_{i=1}^{N} e_i^2$, can be written as

$$\begin{aligned}\sum_{i=1}^{N} e_i^2 &= e'e = (Y - Xb)'(Y - Xb) = \\ &= (Y' - b'X')(Y - Xb) = \\ &= Y'Y - b'X'Y - Y'Xb + b'X'Xb\end{aligned}$$

SUM OF THE SQUARED ERRORS

$$\sum_{i=1}^{N} e_i^2 = e'e = Y'Y - b'X'Y - Y'Xb + b'X'Xb$$

Next note that since $\underset{(1xN)}{Y'}\ \underset{(Nx2)}{X}\ \underset{(2x1)}{b}$ is a scalar we have: $(Y'Xb)' = b'X'Y = Y'Xb$. Thus

$$\sum_{i=1}^{N} e_i^2 = e'e = Y'Y - 2b'X'Y + b'X'Xb$$

LS ESTIMATOR OF THE VECTOR b

$$\sum_{i=1}^{N} e_i^2 = e'e = Y'Y - 2b'X'Y + b'X'Xb$$

The least squares (LS) principle is to choose the vector of the parameters b in order to minimize $\sum_{i=1}^{N} e_i^2$.

Thus we take the first derivative of $e'e$ with respect to b and set it equal to zero:

$$\frac{\partial(e'e)}{\partial b} = -2'X'Y + 2X'X\beta = 0$$

The above equation implies that the LS estimator of b, denoted by β, is given by

$$X'X\beta = X'Y \Rightarrow \beta = (X'X)^{-1}X'Y$$

This is a system of two equations and two unknowns (β_1,β_2) because $\underset{(2xN)}{X'}\underset{(Nx2)}{X}$ and $\underset{(2xN)}{X'}\underset{(Nx1)}{Y}$.

$X'X$ FOR THE BIVARIATE CASE

$X'X$ is given by

$$\underset{(2x2)}{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix}$$

$$= \begin{bmatrix} N & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

From the above equation it follows that

$$(X'X)^{-1} = \begin{bmatrix} N & \sum_{i=1}^N X_i \\ \sum X & \sum X^2 \end{bmatrix}^{-1}$$

$$= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^N X_i \\ -\sum X & N \end{bmatrix}}{N \sum X^2 - (\sum X)^2}$$

$$= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^N X_i \\ -\sum X & N \end{bmatrix}}{N \sum x^2}$$

$X'Y$ FOR THE BIVARIATE CASE

Similarly, $X'Y$ is given by

$$\underset{(2x1)}{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

LS ESTIMATOR OF THE SLOPE COEFFICIENT

The LS estimator β is

$$\begin{aligned}
\beta &= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = (X'X)^{-1}X'Y \\
&= \frac{\begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix} \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}}{N \sum x^2} \\
&= \frac{\begin{bmatrix} \sum X^2 \sum Y - \sum X \sum XY \\ N \sum XY - \sum X \sum Y \end{bmatrix}}{N \sum x^2}
\end{aligned}$$

Thus, since $N \sum xy = N \sum XY - \sum X \sum Y$:

$$\beta_2 = \frac{N \sum XY - \sum X \sum Y}{N \sum x^2} = \frac{\sum xy}{\sum x^2}$$

LS ESTIMATOR OF THE CONSTANT

$$\beta = \frac{\begin{bmatrix} \sum X^2 \sum Y - \sum X \sum XY \\ N \sum XY - \sum X \sum Y \end{bmatrix}}{N \sum x^2}$$

Further

$$\begin{aligned} \beta_1 &= \frac{\sum X^2 \sum Y - \sum X \sum XY}{N \sum x^2} \\ &= \frac{\sum X^2 \overline{Y} - \overline{X} \sum XY}{\sum x^2} \end{aligned}$$

Next, in the numerator we $\pm N\overline{Y}\overline{X}^2$ to get

$$\begin{aligned} \beta_1 &= \frac{\overline{Y}(\sum X^2 - N\overline{X}^2) - \overline{X}(\sum XY - N\overline{Y}\,\overline{X})}{\sum x^2} \\ &= \frac{\overline{Y} \sum x^2 - \overline{X} \sum xy}{\sum x^2} = \overline{Y} - \beta_2 \overline{X} \end{aligned}$$

PROPERTIES OF THE LS RESIDUALS

From the least squares methodology we have

$$(X'X)\beta = X'Y$$

Since $Y = X\beta + \widehat{e}$, it follows that

$$\begin{aligned} (X'X)\beta &= X'(X\beta + \widehat{e}) \\ &\Rightarrow X'\widehat{e} = 0 \end{aligned}$$

In other words

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} \widehat{e}_1 \\ \widehat{e}_2 \\ \vdots \\ \widehat{e}_N \end{bmatrix} = \begin{bmatrix} \sum \widehat{e}_i \\ \sum X_i \widehat{e}_i \end{bmatrix} = 0$$

That is, the residuals, $\widehat{e}_i$, have mean zero: $\sum \widehat{e}_i = 0$, and are orthogonal to the regressor X_i: $\sum X_i \widehat{e}_i = 0$

Recall also that $\overline{Y} = \beta_1 + \beta_2 \overline{X}$. In other words $(\overline{X}, \overline{Y})$ satisfy the estimated linear relationship

The predicted values for Y, denoted by $\widehat{Y}$, are given by $\widehat{Y} = X\beta$. Thus

$$\widehat{Y}'\widehat{e} = (X\beta)'\widehat{e} = \beta' X'\widehat{e} = 0$$

In other words, the vector of regression values for Y is uncorrelated with $\widehat{e}$

EXPECTED VALUE OF THE LS ESTIMATOR β

Recall that LS vector estimator β is

$$\beta = (X'X)^{-1}X'Y$$

Since $Y = Xb + e$, it follows that

$$\begin{aligned}\beta &= (X'X)^{-1}X'(Xb + e) \\ &= (X'X)^{-1}X'Xb + (X'X)^{-1}X'e \\ &= b + (X'X)^{-1}X'e \qquad (1)\end{aligned}$$

Taking expectation from both sides of the above equation gives

$$E(\beta) = b + (X'X)^{-1}X'E(e) = b$$

since $E(e) = 0$. That is β is an unbiased estimator of b

VARIANCE-COVARIANCE MATRIX OF THE LS ESTIMATOR VECTOR β

BIVARIATE CASE

The $2x2$ variance covariance matrix of the vector β is defined as $\underset{(2x2)}{Var(\beta)} = E[\underset{(2x1)}{(\beta - b)}\underset{(1x2)}{(\beta - b)'}]$

Note that if β was a scalar with expected value b then $Var(\beta) = E(\beta - b)^2$

Since β is a $2x1$ vector:

$$\underset{(2x2)}{Var(\beta)} = \begin{bmatrix} Var(\beta_1) & Cov(\beta_1, \beta_2) \\ Cov(\beta_1, \beta_2) & Var(\beta_2) \end{bmatrix}$$

From equation (1) we have

$$\beta - b = (X'X)^{-1}X'e$$

Hence,

$$\begin{aligned} \underset{(2x1)}{E[(\beta - b)}\underset{(1x2)}{(\beta - b)'}] &= E\{(X'X)^{-1}X'e[(X'X)^{-1}X'e]'\} \\ &= E[(X'X)^{-1}X'ee'X(X'X)^{-1}] \end{aligned}$$

since $[(X'X)^{-1}]' = (X'X)^{-1}$ and $(X')' = X$

Next, we have

$$\underset{(2x1)}{E[(\beta - b)}\underset{(1x2)}{(\beta - b)'}] = [(X'X)^{-1}X'E(ee')X(X'X)^{-1}]$$

Using the fact that $E(ee') = \sigma^2 I$ we get

$$\begin{aligned} \underset{(2x2)}{Var(\beta)} &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

VARIANCE OF THE ESTIMATOR OF THE CONSTANT

Recall that

$$(X'X)^{-1} = \begin{bmatrix} \sum X^2 & -\sum X \\ -\sum X & N \end{bmatrix} / N \sum x^2$$

Thus, $\underset{(2x2)}{Var(\beta)} = \sigma^2 (X'X)^{-1}$ is given by

$$\begin{aligned} \underset{(2x2)}{Var(\beta)} &= \begin{bmatrix} Var(\beta_1) & Cov(\beta_1, \beta_2) \\ Cov(\beta_1, \beta_2) & Var(\beta_2) \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} \sum X^2 & -\sum_{i=1}^{N} X_i \\ -\sum X & N \end{bmatrix} / N \sum x^2 \end{aligned}$$

Thus,

$$Var(\beta_1) = \sigma^2 \frac{\sum X^2}{N \sum x^2},$$

$$Var(\beta_1) = \sigma^2 \frac{\sum X^2}{N \sum x^2},$$

Using $\sum x^2 = \sum X^2 - N\overline{X}^2 \Rightarrow \sum X^2 = \sum x^2 + N\overline{X}^2$, we get

$$\begin{aligned} Var(\beta_1) &= \sigma^2 \frac{\sum x^2 + N\overline{X}^2}{N \sum x^2} \\ &= \sigma^2 [\frac{1}{N} + \frac{\overline{X}^2}{\sum x^2}]. \end{aligned}$$

DECOMPOSITION OF RESIDUAL SUM OF SQUARES (RSS)

Recall that

$$Y = X\beta + \widehat{e} = \widehat{Y} + \widehat{e}$$

In other words we decompose the Y vector into:

the part explained by the regression, $\widehat{Y} = X\beta$ (the predicted values of Y),

and the unexplained part $\widehat{e}$ (the residuals)

Thus

$$\begin{aligned}\sum Y^2 &= Y'Y = (\widehat{Y} + \widehat{e})'(\widehat{Y} + \widehat{e}) \\ &= \widehat{Y}'\widehat{Y} + \widehat{e}'\widehat{e} = \sum \widehat{Y}^2 + \sum_{i=1}^{N} \widehat{e}_i^2\end{aligned}$$

since $\widehat{Y}'\widehat{e} = \widehat{e}'\widehat{Y} = 0$ (from the first order conditions).
Note that $\widehat{Y}'\widehat{Y} = (X\beta)'X\beta = \beta' X' X\beta$

$$\sum Y^2 = \widehat{Y}'\widehat{Y} + \widehat{e}'\widehat{e} = \beta' X' X \beta + \widehat{e}'\widehat{e}$$

Recall that $\sum y^2 = \sum Y^2 - N\overline{Y}^2 = N\widehat{Var}(Y)$. Thus

$$\overbrace{\sum y^2}^{\text{TSS}} = \overbrace{\beta' X' X \beta - N\overline{Y}^2}^{\text{ESS}} + \overbrace{\widehat{e}'\widehat{e}}^{\text{RSS}}$$

TSS: total sum of squares; ESS: explained sum of squares; RSS: residual sum of squares