ECON 351* -- NOTE 12

OLS Estimation of the Multiple (Three-Variable) Linear Regression Model

This note derives the Ordinary Least Squares (OLS) coefficient estimators for the ***three-variable*** **multiple linear regression model**.

- The **population regression equation, or PRE**, takes the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (1)$$

where u_i is an iid random error term.

- The **OLS sample regression equation (OLS-SRE)** for equation (1) can be written as

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i = \hat{Y}_i + \hat{u}_i \qquad (i = 1, ..., N). \quad (2)$$

where the $\hat{\beta}_j$ are the OLS estimators of the corresponding population regression coefficients β_j (j = 0, 1, 2),

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \qquad (i = 1, ..., N)$$

are the **OLS residuals**, and

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} \qquad (i = 1, ..., N)$$

are the **OLS *estimated* (or *predicted*) values of Y_i**.

The function $f(X_{1i}, X_{2i}) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ is called the **OLS sample regression function** (or **OLS-SRF**).

1. The OLS Estimation Criterion

The **OLS coefficient estimators** are those formulas (or expressions) for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ that ***minimize* the sum of squared residuals RSS** for any given sample of size N.

The **OLS estimation criterion** is therefore:

$$\underset{\{\hat{\beta}_j\}}{\text{Minimize}}\ \text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right)^2 \qquad \textbf{(3)}$$

Interpretation of the $\text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ function:

- The ***knowns*** in the $\text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ function are the **sample observations** $\left(Y_i, 1, X_{1i}, X_{2i}\right)$ for i = 1, …, N. In other words, the N sample values of the observable variables Y, X_1, X_2 are taken as known (or given).

- The ***unknowns*** in the $\text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ function are therefore the **coefficient estimators** $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

- For purposes of deriving the OLS coefficient estimators, the $\text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ function is interpreted as a function of the three unknowns $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

2. The OLS Normal Equations: Derivation of the FOCs

STEP 1: Re-write the $RSS(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ **function** in (3) as follows:

$$RSS(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} f(\hat{u}_i) \qquad \text{where} \qquad f(\hat{u}_i) = \hat{u}_i^2$$

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \qquad \textbf{(3)}$$

Note: The function $f(\hat{u}_i) = \hat{u}_i^2$ is a function of $\hat{u}_i$, and $\hat{u}_i$ is in turn a function of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

STEP 2: Partially differentiate the $RSS(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ **function** in (3) with respect to $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$:

- Using the **chain rule of differentiation**, each partial derivative of the $RSS(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ function takes the general form

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{df}{d\hat{u}_i} \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j}. \qquad (4)$$

- Using the **power rule of differentiation**, the derivative $df/d\hat{u}_i$ is

$$\frac{df}{d\hat{u}_i} = \frac{d(\hat{u}_i^2)}{d\hat{u}_i} = 2\hat{u}_i.$$

The partial derivatives $\partial RSS/\partial \hat{\beta}_j$ for j = 0, 1, 2 are therefore

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{df}{d\hat{u}_i} \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} 2\hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} = 2\sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} \qquad j = 0, 1, 2. \qquad (5)$$

- Since the i-th residual is $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$, the partial derivatives $\partial \hat{u}_i \big/ \partial \hat{\beta}_j$ for j = 0, 1, 2 are:

$$\frac{\partial \hat{u}_i}{\partial \hat{\beta}_0} = -1; \qquad \frac{\partial \hat{u}_i}{\partial \hat{\beta}_1} = -X_{1i}; \qquad \frac{\partial \hat{u}_i}{\partial \hat{\beta}_2} = -X_{2i}.$$

- Substitute the partial derivatives $\partial \hat{u}_i \big/ \partial \hat{\beta}_j$ for j = 0, 1, 2 into equation (5):

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} \qquad j = 0, 1, 2. \tag{5}$$

The partial derivatives $\partial RSS \big/ \partial \hat{\beta}_j$ for j = 0, 1, 2 thus take the form:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{N} \hat{u}_i (-1) = -2 \sum_{i=1}^{N} \hat{u}_i \tag{6.1}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{N} \hat{u}_i (-X_{1i}) = -2 \sum_{i=1}^{N} X_{1i} \hat{u}_i \tag{6.2}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_2} = 2 \sum_{i=1}^{N} \hat{u}_i (-X_{2i}) = -2 \sum_{i=1}^{N} X_{2i} \hat{u}_i \tag{6.3}$$

<u>STEP 3</u>: Obtain the first-order conditions (FOCs) for a minimum of the RSS function by setting the partial derivatives (6.1)-(6.3) equal to zero, then dividing each equation by −2, and finally setting $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$:

- $$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{N} \hat{u}_i \tag{6.1}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 \quad \Rightarrow \quad -2 \sum_{i=1}^{N} \hat{u}_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} \hat{u}_i = 0 \tag{7.1}$$

$$\Rightarrow \quad \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \tag{8.1}$$

- $$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{N} X_{1i} \hat{u}_i \tag{6.2}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \quad \Rightarrow \quad -2 \sum_{i=1}^{N} X_{1i} \hat{u}_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} X_{1i} \hat{u}_i = 0 \tag{7.2}$$

$$\Rightarrow \quad \sum_{i=1}^{N} X_{1i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \tag{8.2}$$

- $$\frac{\partial RSS}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{N} X_{2i} \hat{u}_i \tag{6.3}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 \quad \Rightarrow \quad -2 \sum_{i=1}^{N} X_{2i} \hat{u}_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} X_{2i} \hat{u}_i = 0 \tag{7.3}$$

$$\Rightarrow \quad \sum_{i=1}^{N} X_{2i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \tag{8.3}$$

<u>STEP 4</u>: Rearrange each of the equations (8.1)-(8.3) to put them in the conventional form of the OLS normal equations. Thus, taking summations and rearranging terms, we obtain the **OLS normal equations**:

- $$\sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right) = 0 \qquad (8.1)$$

$$\sum_{i=1}^{N} Y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} - \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} = 0$$

$$- N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} - \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} = -\sum_{i=1}^{N} Y_i$$

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} = \sum_{i=1}^{N} Y_i \qquad \textbf{(N1)}$$

- $$\sum_{i=1}^{N} X_{1i}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right) = 0 \qquad (8.2)$$

$$\sum_{i=1}^{N}\left(X_{1i}Y_i - \hat{\beta}_0 X_{1i} - \hat{\beta}_1 X_{1i}^2 - \hat{\beta}_2 X_{1i}X_{2i}\right) = 0$$

$$\sum_{i=1}^{N} X_{1i}Y_i - \hat{\beta}_0 \sum_{i=1}^{N} X_{1i} - \hat{\beta}_1 \sum_{i=1}^{N} X_{1i}^2 - \hat{\beta}_2 \sum_{i=1}^{N} X_{1i}X_{2i} = 0$$

$$-\hat{\beta}_0 \sum_{i=1}^{N} X_{1i} - \hat{\beta}_1 \sum_{i=1}^{N} X_{1i}^2 - \hat{\beta}_2 \sum_{i=1}^{N} X_{1i}X_{2i} = -\sum_{i=1}^{N} X_{1i}Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{N} X_{1i} + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^{N} X_{1i}X_{2i} = \sum_{i=1}^{N} X_{1i}Y_i \qquad \textbf{(N2)}$$

- $$\sum_{i=1}^{N} X_{2i}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right) = 0 \qquad (8.3)$$

$$\sum_{i=1}^{N}\left(X_{2i}Y_i - \hat{\beta}_0 X_{2i} - \hat{\beta}_1 X_{2i}X_{1i} - \hat{\beta}_2 X_{2i}^2\right) = 0$$

$$\sum_{i=1}^{N} X_{2i}Y_i - \hat{\beta}_0 \sum_{i=1}^{N} X_{2i} - \hat{\beta}_1 \sum_{i=1}^{N} X_{2i}X_{1i} - \hat{\beta}_2 \sum_{i=1}^{N} X_{2i}^2 = 0$$

$$-\hat{\beta}_0 \sum_{i=1}^{N} X_{2i} - \hat{\beta}_1 \sum_{i=1}^{N} X_{2i}X_{1i} - \hat{\beta}_2 \sum_{i=1}^{N} X_{2i}^2 = -\sum_{i=1}^{N} X_{2i}Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{N} X_{2i} + \hat{\beta}_1 \sum_{i=1}^{N} X_{2i}X_{1i} + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i}^2 = \sum_{i=1}^{N} X_{2i}Y_i \qquad \textbf{(N3)}$$

<u>RESULT</u>: Assemble the **three OLS normal equations (N1)-(N3)**:

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} = \sum_{i=1}^{N} Y_i \qquad \textbf{(N1)}$$

$$\hat{\beta}_0 \sum_{i=1}^{N} X_{1i} + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^{N} X_{1i} X_{2i} = \sum_{i=1}^{N} X_{1i} Y_i \qquad \textbf{(N2)}$$

$$\hat{\beta}_0 \sum_{i=1}^{N} X_{2i} + \hat{\beta}_1 \sum_{i=1}^{N} X_{2i} X_{1i} + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i}^2 = \sum_{i=1}^{N} X_{2i} Y_i \qquad \textbf{(N3)}$$

- The **OLS normal equations (N1)-(N3)** constitute ***three linear equations* in the *three unknowns*** $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

- **Solution** of the **OLS normal equations (N1)-(N3)** yields **explicit expressions (or formulas)** for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$; these expressions are the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ of the partial regression coefficients β_0, β_1, and β_2 respectively.

3. Expressions for the OLS Coefficient Estimators

- The expressions (formulas) for the OLS estimators are most conveniently written in ***deviation-from-means form***, which uses lower case letters to denote the deviations of the sample values of each observable variable from their respective sample means. Thus, define the deviations-from-means of Y_i, X_{1i}, and X_{2i} as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_{1i} \equiv X_{1i} - \overline{X}_1; \qquad x_{2i} \equiv X_{2i} - \overline{X}_2;$$

where

$\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ is the sample mean of the Y_i values;

$\overline{X}_1 = \Sigma_i X_{1i} / N = \dfrac{\Sigma_i X_{1i}}{N}$ is the sample mean of the X_{1i} values;

$\overline{X}_2 = \Sigma_i X_{2i} / N = \dfrac{\Sigma_i X_{2i}}{N}$ is the sample mean of the X_{2i} values.

- The **OLS *slope* coefficient estimators** $\hat{\beta}_1$ and $\hat{\beta}_2$ in **deviation-from-means form** are:

$$\hat{\beta}_1 = \frac{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{1i} y_i\right) - \left(\Sigma_i x_{1i} x_{2i}\right)\left(\Sigma_i x_{2i} y_i\right)}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2}; \tag{9.2}$$

$$\hat{\beta}_2 = \frac{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i} y_i\right) - \left(\Sigma_i x_{1i} x_{2i}\right)\left(\Sigma_i x_{1i} y_i\right)}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2}. \tag{9.3}$$

- The **OLS *intercept* coefficient estimator** $\hat{\beta}_0$ is:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2 . \tag{9.1}$$

4. The OLS Variance-Covariance Estimators

- An ***unbiased*** **estimator of the error variance σ^2**

- For the ***general*** **multiple linear regression model** with K regression coefficients, an ***unbiased*** **estimator of the error variance σ^2** is the degrees-of-freedom-adjusted estimator

$$\hat{\sigma}^2 = \frac{\Sigma_i \hat{u}_i^2}{(N-K)} = \frac{RSS}{(N-K)}$$

where K = k + 1 = the total number of regression coefficients in the PRF.

- For the ***three-variable*** **multiple linear regression model** (such as regression equation (1) above) for which **K = 3**, the unbiased estimator of the error variance σ^2 is therefore

$$\hat{\sigma}^2 = \frac{\Sigma_i \hat{u}_i^2}{(N-3)} = \frac{RSS}{(N-3)}. \qquad (10)$$

where (N – 3) is the degrees of freedom for the residual sum of squares RSS in the OLS-SRE (2). $\hat{\sigma}^2$ is an unbiased estimator of σ^2 because it can be shown that $E(\sum_i \hat{u}_i^2) = E(RSS) = (N-3)\sigma^2$.

- The error variance estimator $\hat{\sigma}^2$ is used to obtain ***unbiased*** **estimators** of the ***variances*** **and** ***covariances*** **of the OLS coefficient estimators** $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

- Formulas for the ***variances* and *covariances* of the *slope* coefficient estimators** $\hat{\beta}_1$ **and** $\hat{\beta}_2$ in the three-variable multiple regression model

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2 \Sigma_i x_{2i}^2}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2};$$

$$\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2 \Sigma_i x_{1i}^2}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2};$$

$$\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{\sigma^2 \Sigma_i x_{1i} x_{2i}}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2}.$$

- ***Unbiased* estimators of the *variances* of the *slope* coefficient estimators** $\hat{\beta}_1$ **and** $\hat{\beta}_2$ are obtained by substituting the unbiased estimator $\hat{\sigma}^2$ for the unknown error variance σ^2 in the formulas for $\mathrm{Var}(\hat{\beta}_1)$ and $\mathrm{Var}(\hat{\beta}_2)$:

$$\hat{\mathrm{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2 \Sigma_i x_{2i}^2}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2}; \tag{11.1}$$

$$\hat{\mathrm{Var}}(\hat{\beta}_2) = \frac{\hat{\sigma}^2 \Sigma_i x_{1i}^2}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2}; \tag{11.2}$$

- Similarly, an ***unbiased* estimator of the *covariance* between the slope coefficient estimators** $\hat{\beta}_1$ **and** $\hat{\beta}_2$ is obtained by substituting the unbiased estimator $\hat{\sigma}^2$ for the unknown error variance σ^2 in the formula for $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)$:

$$\hat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = \frac{\hat{\sigma}^2 \Sigma_i x_{1i} x_{2i}}{(\Sigma_i x_{1i}^2)(\Sigma_i x_{2i}^2) - (\Sigma_i x_{1i} x_{2i})^2}. \tag{11.3}$$

❑ **<u>Interpretive formula</u>** for the ***variances* of the OLS *slope* coefficient estimators $\hat{\beta}_j$, j = 1, 2, …, k**

Consider the general multiple linear regression equation given by the PRE

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_j X_{ji} + \cdots + \beta_k X_{ki} + u_i \qquad (12.1)$$

OLS estimation of the PRE in (11) yields the OLS SRE

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_j X_{ji} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \qquad (12.2)$$

- The **formula for $\mathbf{Var(\hat{\beta}_j)}$** for j = 1, 2, …, k can be written as

$$\mathbf{Var(\hat{\beta}_j) = \frac{\sigma^2}{TSS_j\left(1 - R_j^2\right)} \qquad for\ j = 1, 2, \ldots, k} \qquad \mathbf{(13)}$$

where

$TSS_j \equiv \sum_{i=1}^{N} x_{ji}^2 \equiv \sum_{i=1}^{N}\left(X_{ji} - \bar{X}_j\right)^2 \equiv$ the total sample variation in the regressor X_j;

$R_j^2 \equiv$ the R^2 from the OLS regression of regressor X_j on all the other K−1 regressors in (12.1), including the intercept. That is, R_j^2 measures the proportion of the total sample variation in X_j that is explained by the *other* regressors in the PRE. Alternatively, R_j^2 measures the degree of linear dependence between the sample values X_{ji} of the regressor X_j and the sample values of the other regressors in regression equation (12.1).

This formula for $Var(\hat{\beta}_j)$ is given in J. Kmenta, *Elements of Econometrics*, 2nd edition (1986), pp. 437-438.

- **Determinants of $\mathbf{Var(\hat{\beta}_j)}$.**

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{TSS}_j\left(1 - R_j^2\right)} \qquad \text{for } j = 1, 2, \ldots, k \qquad \textbf{(13)}$$

Three factors determine $\text{Var}(\hat{\beta}_j)$:

1. the **error variance** $\boldsymbol{\sigma^2}$;
2. the **total sample variation in $\mathbf{X_j}$, $\mathbf{TSS_j}$**;
3. the **degree of linear dependence between $\mathbf{X_j}$ and the *other* regressors** in the model, as measured by $\mathbf{R_j^2}$.

- **$\mathbf{Var(\hat{\beta}_j)}$ is *smaller***:

(1) the ***smaller* is $\boldsymbol{\sigma^2}$**, the **error variance** in the true model;

(2) the ***larger* is $\mathbf{TSS_j}$**, the **total sample variation in the regressor $\mathbf{X_j}$**;

$$\text{TSS}_j \equiv \sum_{i=1}^{N} x_{ji}^2 \equiv \sum_{i=1}^{N}\left(X_{ji} - \bar{X}_j\right)^2 \text{ is } \textbf{\textit{larger}}$$

- the ***larger* the values of $\mathbf{x_{ji}^2 = (X_{ji} - \bar{X}_j)^2}$** for i = 1, …, N, meaning the greater the sample variation in the X_{ji} values around their sample mean;
- the ***larger* is N**, the size of the estimation sample;

(3) the ***smaller (closer to 0)* is $\mathbf{R_j^2}$**, the ***lower* the degree of linear dependence** between the sample values X_{ji} of the regressor X_j and the sample values of the other regressors in the PRE.

- Conversely, $\mathbf{Var(\hat{\beta}_j)}$ **is *larger***:

(1) the ***larger*** **is σ^2**, the **error variance** in the true model;

(2) the ***smaller*** **is TSS_j**, the **total sample variation in the regressor X_j**;

$$TSS_j \equiv \sum_{i=1}^{N} x_{ji}^2 \equiv \sum_{i=1}^{N} \left(X_{ji} - \overline{X}_j\right)^2 \text{ is } \textit{\textbf{smaller}}$$

- the ***smaller*** **the values of $\mathbf{x_{ji}^2 = (X_{ji} - \overline{X}_j)^2}$** for i = 1, …, N, meaning the greater the sample variation in the X_{ji} values around their sample mean;
- the ***smaller*** **is N**, the size of the estimation sample;

(3) the ***larger (closer to 1)*** **is $\mathbf{R_j^2}$**, the ***greater*** **the degree of linear dependence** between the sample values X_{ji} of the regressor X_j and the sample values of the other regressors in the PRE.

Note: **Assumption A8**, the absence of perfect multicollinearity, rules out the value 1 for R_j^2.

If $\mathbf{R_j^2 = 1}$, then the sample values of the regressor X_j exhibit an exact linear dependence -- are perfectly multicollinear -- with one or more of the other regressors in the model, in which case **it is *impossible* to compute** either

(1) the **OLS estimate $\mathbf{\hat{\beta}_j}$** of the slope coefficient β_j associated with the regressor X_j, or

(2) the estimated value of $Var(\hat{\beta}_j)$, the ***estimated*** **variance of $\mathbf{\hat{\beta}_j}$**.

If $\mathbf{R_j^2 < 1}$, then multicollinearity is simply a question of degree: the closer to 1 is the sample value of R_j^2, the larger is $Var(\hat{\beta}_j)$.

$$\mathbf{Var(\hat{\beta}_j) \rightarrow \infty \quad \text{as} \quad R_j^2 \rightarrow 1}$$

5. Computational Properties of the OLS-SRE (2)

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i = \hat{Y}_i + \hat{u}_i \qquad (i = 1, ..., N). \qquad (2)$$

5.1 The **OLS-SRE passes through the *point of sample means*** $\left(\overline{Y}, \overline{X}_1, \overline{X}_2\right)$; i.e.,

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}_1 + \hat{\beta}_2 \overline{X}_2 . \qquad \textbf{(C1)}$$

Proof of (C1): Follows directly from dividing OLS normal equation (N1) by N.

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} = \sum_{i=1}^{N} Y_i \qquad (N1)$$

$$\frac{N\hat{\beta}_0}{N} + \hat{\beta}_1 \frac{\sum_{i=1}^{N} X_{1i}}{N} + \hat{\beta}_2 \frac{\sum_{i=1}^{N} X_{2i}}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} \quad \text{dividing (N1) by N}$$

$$\hat{\beta}_0 + \hat{\beta}_1 \overline{X}_1 + \hat{\beta}_2 \overline{X}_2 = \overline{Y} .$$

5.2 The ***sum*** **of the *estimated*** $\mathbf{Y_i}$**'s** (the $\hat{\mathbf{Y}}_\mathbf{i}$**'s**) ***equals*** the ***sum*** **of the *observed*** $\mathbf{Y_i}$'s; or the **sample *mean*** **of the *estimated*** $\mathbf{Y_i}$**'s** (the $\hat{\mathbf{Y}}_\mathbf{i}$**'s**) ***equals*** the **sample *mean*** **of the *observed*** $\mathbf{Y_i}$'s.

$$\sum_{i=1}^{N}\hat{Y}_i = \sum_{i=1}^{N} Y_i \quad \textit{or} \quad \frac{\sum_{i=1}^{N}\hat{Y}_i}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} \quad \textit{or} \quad \overline{\hat{Y}} = \overline{Y} \qquad \textbf{(C2)}$$

<u>Proof of (C2)</u>:

1. Substitute $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}_1 - \hat{\beta}_2\overline{X}_2$ in the expression for $\hat{Y}_i$:

$$\begin{aligned}
\hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} \\
&= \left(\overline{Y} - \hat{\beta}_1\overline{X}_1 - \hat{\beta}_2\overline{X}_2\right) + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} \\
&= \overline{Y} + \hat{\beta}_1\left(X_{1i} - \overline{X}_1\right) + \hat{\beta}_2\left(X_{2i} - \overline{X}_2\right) \\
&= \overline{Y} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} \qquad \text{since } x_{1i} \equiv \left(X_{1i} - \overline{X}_1\right),\; x_{2i} \equiv \left(X_{2i} - \overline{X}_2\right)
\end{aligned}$$

2. Now sum the above equation over the sample:

$$\begin{aligned}
\sum_{i=1}^{N}\hat{Y}_i &= N\overline{Y} + \hat{\beta}_1\sum_{i=1}^{N} x_{1i} + \hat{\beta}_2\sum_{i=1}^{N} x_{2i} \\
&= N\overline{Y} \qquad \text{because } \sum_{i=1}^{N} x_{1i} = \sum_{i=1}^{N} x_{2i} = 0.
\end{aligned}$$

3. Thus, dividing both sides of the above equation by N, we obtain the result

$$\sum_{i=1}^{N}\hat{Y}_i = N\overline{Y} \Rightarrow \frac{\sum_{i=1}^{N}\hat{Y}_i}{N} = \overline{Y} \quad \textit{or} \quad \frac{\sum_{i=1}^{N}\hat{Y}_i}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} \quad \textit{or} \quad \overline{\hat{Y}} = \overline{Y} \quad \textit{or} \quad \sum_{i=1}^{N}\hat{Y}_i = \sum_{i=1}^{N} Y_i\,.$$

5.3 The ***sum*** **of the OLS residuals $\hat{u}_i$ (i = 1, …, N) equals *zero***, or the ***sample mean*** **of the OLS residuals $\hat{u}_i$ equals *zero***.

$$\sum_{i=1}^{N} \hat{u}_i = 0 \qquad \textit{or} \qquad \bar{\hat{u}} = \frac{\sum_{i=1}^{N} \hat{u}_i}{N} = 0. \qquad \textbf{(C3)}$$

Proof of (C3): An immediate implication of equation (7.1).

$$\sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \hat{u}_i = 0 \qquad (7.1)$$

5.4 The **OLS residuals $\hat{u}_i$ (i = 1, …, N)** are ***uncorrelated with*** the **sample values of the regressors X_{1i} and X_{2i} (i = 1, …, N)**: i.e.,

$$\sum_{i=1}^{N} X_{1i} \hat{u}_i = 0 \qquad \textit{and} \qquad \sum_{i=1}^{N} X_{2i} \hat{u}_i = 0. \qquad \textbf{(C4)}$$

Proof of (C4): An immediate implication of equations (7.2) and (7.3).

$$\sum_{i=1}^{N} X_{1i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} X_{1i} \hat{u}_i = 0 \qquad (7.2)$$

$$\sum_{i=1}^{N} X_{2i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} X_{2i} \hat{u}_i = 0 \qquad (7.3)$$

5.5 The **OLS residuals** $\hat{\mathbf{u}}_{\mathbf{i}}$ **(i = 1, …, N)** ***are uncorrelated with*** the ***estimated*** **values of $\mathbf{Y_i}$**, the $\hat{\mathbf{Y}}_{\mathbf{i}}$ **values (i = 1, …, N)**: i.e.,

$$\sum_{i=1}^{N} \hat{Y}_i \hat{u}_i = 0 . \qquad \textbf{(C5)}$$

Proof of (C5): Follows from properties (C3) and (C4) above.

1. The $\hat{Y}_i$ are given by the following expression for the OLS sample regression function (the OLS-SRF):

$$\hat{Y}_i = \hat{\beta}_0 + \beta_1 X_{1i} + \hat{\beta}_2 X_{2i} .$$

2. Multiply the above equation by $\hat{u}_i$:

$$\hat{Y}_i \hat{u}_i = \hat{\beta}_0 \hat{u}_i + \hat{\beta}_1 X_{1i} \hat{u}_i + \hat{\beta}_2 X_{2i} \hat{u}_i .$$

3. Summing both sides of the above equation over the sample gives the result

$$\sum_{i=1}^{N} \hat{Y}_i \hat{u}_i = \hat{\beta}_0 \sum_{i=1}^{N} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{N} X_{1i} \hat{u}_i + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} \hat{u}_i$$
$$= 0$$

because

$$\sum_{i=1}^{N} \hat{u}_i = 0 \qquad \text{by property (C3)}$$

and

$$\sum_{i=1}^{N} X_{1i} \hat{u}_i = \sum_{i=1}^{N} X_{2i} \hat{u}_i = 0 \qquad \text{by property (C4).}$$