ECON 351* -- NOTE 13

Goodness-of-Fit in the Multiple Linear Regression Model

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

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad (1)$$

where u_i is an iid random error term.

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

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

where

(1) the **OLS *estimated* (or *predicted*) values of Y_i**, or the OLS sample regression function (OLS-SRF), are

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

(2) the **OLS *residuals*** are

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

1. The OLS Decomposition Equation

1.1 General Form of the OLS Decomposition Equation

- For the OLS sample regression equation (OLS-SRE)

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

the **OLS decomposition equation** is

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 \qquad \Leftrightarrow \qquad \text{TSS = ESS + RSS}$$

where

(1) $\sum_{i=1}^{N} y_i^2 \equiv \sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2 \equiv \text{TSS} \equiv$ the **Total Sum of Squares**

(2) $\sum_{i=1}^{N} \hat{y}_i^2 \equiv \sum_{i=1}^{N} \left(\hat{Y}_i - \overline{Y}\right)^2 \equiv \text{ESS} \equiv$ the **Explained Sum of Squares**

(3) $\sum_{i=1}^{N} \hat{u}_i^2 \equiv \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 \equiv \text{RSS} \equiv$ the **Residual Sum of Squares**

- **Interpretative Formula for ESS**

$$\text{ESS} \equiv \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i + \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ki} y_i$$

where $x_{ji} \equiv X_{ji} - \overline{X}_j$, j = 1, 2, …, k.

Implication: ESS = 0 if $\hat{\beta}_1 = 0$ and $\hat{\beta}_2 = 0$ and … $\hat{\beta}_k = 0$
if $\hat{\beta}_j = 0$ for all j = 1, …, k.

1.2 Derivation of OLS Decomposition Equation

We derive the OLS decomposition equation for the simplest case, that is the three-variable multiple regression model for which the OLS-SRE is

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

<u>STEP 1</u>: Write the OLS-SRE (2) in *deviation-from-means* form.

1. Substitute for $\hat{\beta}_0$ in equation (2) the formula $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2$:

$$Y_i = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i .$$

2. Re-arrange by subtracting $\bar{Y}$ from both sides and collecting terms in $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$Y_i - \bar{Y} = \hat{\beta}_1\left(X_{1i} - \bar{X}_1\right) + \hat{\beta}_2\left(X_{2i} - \bar{X}_2\right) + \hat{u}_i$$

or

$$y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{u}_i = \hat{y}_i + \hat{u}_i ,$$

where

$$\hat{y}_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$

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

3. Therefore, since $\hat{y}_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$, we have the result that

$$y_i = \hat{y}_i + \hat{u}_i . \qquad (3)$$

<u>STEP 2</u>: Square both sides of equation (3).

$$y_i = \hat{y}_i + \hat{u}_i. \tag{3}$$

$$y_i^2 = (\hat{y}_i + \hat{u}_i)^2 = \hat{y}_i^2 + \hat{u}_i^2 + 2\hat{y}_i\hat{u}_i. \tag{4}$$

<u>STEP 3</u>: Sum equation (4) over the sample observations, from i = 1 to i = N.

1. $$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 + 2\sum_{i=1}^{N} \hat{y}_i\hat{u}_i .$$

2. But the computational properties of the OLS-SRE imply that $\sum_{i=1}^{N} \hat{y}_i\hat{u}_i = 0$.

$$\begin{aligned} \sum_{i=1}^{N} \hat{y}_i\hat{u}_i &= \sum_{i=1}^{N} \left(\hat{Y}_i - \overline{Y}\right)\hat{u}_i \\ &= \sum_{i=1}^{N} \left(\hat{Y}_i\hat{u}_i - \overline{Y}\hat{u}_i\right) \\ &= \sum_{i=1}^{N} \hat{Y}_i\hat{u}_i - \overline{Y}\sum_{i=1}^{N} \hat{u}_i \\ &= 0 \end{aligned}$$

because

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

and

$\sum_{i=1}^{N} \hat{Y}_i\hat{u}_i = 0$ by computational property (C5).

- **<u>Result</u>:**

$$\boldsymbol{\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2} \tag{5}$$

i.e, **TSS = ESS + RSS**

2. Computational Formula for RSS

A convenient **computational formula** for the **residual sum of squares RSS** is:

$$\text{RSS} = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} \hat{y}_i^2 = \sum_{i=1}^{N} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i \qquad \textbf{(6)}$$

Derivation of equation (6) for RSS

1. The i-th OLS residual $\hat{u}_i$ can be written in deviations-from-means form as

$$\begin{aligned}
\hat{u}_i &= Y_i - \hat{Y}_i \\
&= Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} && \text{since } \hat{Y}_i = \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \\
&= Y_i - \left(\overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2\right) - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} && \text{since } \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2 \\
&= \left(Y_i - \overline{Y}\right) - \hat{\beta}_1\left(X_{1i} - \overline{X}_1\right) - \hat{\beta}_2\left(X_{2i} - \overline{X}_2\right) \\
&= y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}
\end{aligned}$$

2. Multiplying both sides of the above equation by $\hat{u}_i$, we obtain

$$\begin{aligned}
\hat{u}_i^2 &= \left(y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right)\hat{u}_i \\
&= y_i \hat{u}_i - \hat{\beta}_1 x_{1i} \hat{u}_i - \hat{\beta}_2 x_{2i} \hat{u}_i
\end{aligned}$$

3. Summing both sides of the above equation over the sample yields

$$\begin{aligned}
\Sigma_i \hat{u}_i^2 &= \Sigma_i y_i \hat{u}_i - \hat{\beta}_1 \Sigma_i x_{1i} \hat{u}_i - \hat{\beta}_2 \Sigma_i x_{2i} \hat{u}_i \\
&= \Sigma_i y_i \hat{u}_i
\end{aligned}$$

since the OLS normal equations imply that $\sum_i x_{1i} \hat{u}_i = 0$ and $\sum_i x_{2i} \hat{u}_i = 0$.

4. Setting $\hat{u}_i = y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}$ in the above equation yields the result

$$\begin{aligned}
\Sigma_i \hat{u}_i^2 &= \Sigma_i y_i \left(y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right) \\
&= \sum_i y_i^2 - \hat{\beta}_1 \sum_i x_{1i} y_i - \hat{\beta}_2 \sum_i x_{2i} y_i = \sum_i y_i^2 - \sum_i \hat{y}_i^2
\end{aligned}$$

where $\sum_i \hat{y}_i^2 = \hat{\beta}_1 \sum_i x_{1i} y_i + \hat{\beta}_2 \sum_i x_{2i} y_i$ = ESS.

3. The Coefficient of Determination -- R^2

3.1 Definition of R^2

1. Start with the **OLS decomposition equation**:

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 \qquad \Leftrightarrow \qquad \text{TSS} = \text{ESS} + \text{RSS}$$

2. Divide both sides of the OLS decomposition equation by TSS = $\sum_{i=1}^{N} y_i^2$:

$$1 = \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} + \frac{\sum_i \hat{u}_i^2}{\sum_i y_i^2} \tag{7.1}$$

or

$$1 = \frac{\text{ESS}}{\text{TSS}} + \frac{\text{RSS}}{\text{TSS}} \tag{7.2}$$

3. The **coefficient of determination R^2** is defined as:

$$R^2 \equiv \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = 1 - \frac{\sum_i \hat{u}_i^2}{\sum_i y_i^2} \qquad \text{from equation (7.1)}$$

or

$$R^2 \equiv \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} \qquad \text{from equation (7.2)}$$

3.2 Alternative Formula for R^2

1. Start with the OLS decomposition equation (5)

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 \qquad \textbf{(5)}$$

and expression (6) above for $\text{RSS} = \sum_{i=1}^{N} \hat{u}_i^2$

$$\text{RSS} = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i\ . \qquad \textbf{(6)}$$

2. Substitute the right-hand side of equation (6) for $\sum_{i=1}^{N} \hat{u}_i^2$ in the decomposition equation (5):

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i\ .$$

3. Subtract $\sum_{i=1}^{N} y_i^2$ from both sides of the above equation:

$$0 = \sum_{i=1}^{N} \hat{y}_i^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i\ .$$

4. Solve the above equation for $\sum_{i=1}^{N} \hat{y}_i^2$:

$$\sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i + \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i \equiv \text{ESS}. \qquad (8)$$

❑ **<u>Result</u>:** Substitute the expression for $ESS \equiv \sum_{i=1}^{N} \hat{y}_i^2$ given by equation (8) into the definition of R^2 to obtain the following expression for R^2:

$$R^2 \equiv \frac{ESS}{TSS} \equiv \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = \frac{\hat{\beta}_1 \sum_i x_{1i} y_i + \hat{\beta}_2 \sum_i x_{2i} y_i}{\sum_i y_i^2}. \qquad \textbf{(9.1)}$$

In general, for the general multiple linear regression model with k = K – 1 non-constant regressors, the expression for R^2 is:

$$R^2 \equiv \frac{ESS}{TSS} \equiv \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = \frac{\hat{\beta}_1 \sum_i x_{1i} y_i + \hat{\beta}_2 \sum_i x_{2i} y_i + \cdots + \hat{\beta}_k \sum_i x_{ki} y_i}{\sum_i y_i^2}. \qquad \textbf{(9.2)}$$

3.3 Interpretation of R^2: The Values of R^2

❑ **<u>What Does R^2 Measure?</u>**

R^2 = the proportion of the total sample variation of the dependent variable Y that is explained by the sample regression function, i.e., by the values of the regressors X_{1i}, X_{2i}, …, X_{ki}.

❑ **<u>The Values of R^2</u>**

R^2 values lie in the closed unit interval [0, 1]; i.e., $\mathbf{0 \le R^2 \le 1}$.

❑ Interpreting the Values of R^2

- ***Rule 1*:** The ***closer*** **is the value of R^2 to 1**, the ***better*** **the goodness-of-fit** of the OLS-SRE to the sample data.

  - The upper limiting value $R^2 = 1$ corresponds to a **perfect fit** of the OLS-SRE to the sample data.

    $$R^2 = 1 \quad \Rightarrow \quad \frac{\text{ESS}}{\text{TSS}} = 1 \quad \Rightarrow \quad \text{ESS} = \text{TSS} \quad \Rightarrow \quad \text{RSS} = \sum_i \hat{u}_i^2 = 0.$$

  - But since $\hat{u}_i^2 \geq 0$ for all i, $\text{RSS} = \sum_i \hat{u}_i^2 = 0$ if and only if

    $$\hat{u}_i = 0 \; \forall \; i = 1, ..., N.$$

  - Therefore, a perfect fit of the OLS-SRE means that

    $$\hat{u}_i = 0 \; \forall \; i = 1, ..., N \quad \underline{\textit{or}} \quad Y_i = \hat{Y}_i \quad \forall \; i = 1, ..., N.$$

- ***Rule 2*****:** The ***closer*** **is the value of R^2 to 0**, the ***worse*** **the goodness-of-fit** of the OLS-SRE to the sample data.

  - The lower limiting value $R^2 = 0$ corresponds to the **worst possible fit** of the OLS-SRE to the sample data.

$$R^2 = 0 \quad \Rightarrow \quad \frac{ESS}{TSS} = 0 \quad \Rightarrow \quad ESS = 0 \quad \Rightarrow \quad TSS = RSS.$$

  - When does ESS = 0? ESS = 0 when

$$ESS \equiv \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i + \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ki} y_i = 0.$$

  - A **sufficient condition for ESS = 0** is thus that **all slope coefficient estimates equal zero**: i.e.,

$$\hat{\beta}_j = 0 \quad \forall \ j = 1, 2, ..., k \quad \Leftrightarrow \quad \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_k = 0.$$

  - Finally, since

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$$

and

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2 - \cdots - \hat{\beta}_k \overline{X}_k,$$

it follows that $\hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_k = 0$ means that

$$\hat{Y}_i = \hat{\beta}_0 = \overline{Y} \qquad \forall \ i = 1,..., N.$$

The reason is that $\hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_k = 0$ implies that

(1) $\hat{Y}_i = \hat{\beta}_0$ since $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$

(2) $\hat{Y}_i = \hat{\beta}_0 = \overline{Y}$ since $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2 - \cdots - \hat{\beta}_k \overline{X}_k$.

3.4 Limitations of R^2

The R^2 can be used to compare the goodness-of-fit of alternative sample regression equations only if the regression models satisfy **two conditions**.

(1) The models must have the ***same*** **regressand**, or ***same*** **dependent variable**.

Reason: TSS, ESS, and RSS depend on the units in which the regressand Y_i is measured.

(2) The models must have the ***same*** **number of regressors and regression coefficients --** i.e., the ***same*** **value of K**.

Reason: Adding additional regressors to a regression equation – i.e., increasing the value of K – always increases the value of R^2.

- **ESS** is an ***increasing*** **function** of the number of regressors K.
- **RSS** is a ***decreasing*** **function** of the number of regressors K.
- Therefore, $\mathbf{R^2}$ is an ***increasing*** **function of the number of regressors K**.

4. The Adjusted R^2

4.1 Definition of Adjusted R^2

$$\bar{R}^2 \equiv 1 - \frac{RSS/(N-K)}{TSS/(N-1)} = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

where

$\hat{\sigma}^2 = \dfrac{RSS}{N-K}$ = the unbiased estimator of the error variance σ^2;

$s_Y^2 = \dfrac{TSS}{N-1} = \dfrac{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{N-1}$ = the sample variance of the Y_i values.

4.2 Relationship Between R^2 and Adjusted R^2

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{N-1}{N-K}.$$

(1) For values of K > 1, $\bar{R}^2 < R^2$.

(2) $\bar{R}^2$ can be negative, even though R^2 is non-negative.

4.3 Guidelines for Using Adjusted R^2

1. $\bar{R}^2$ can be used to compare the goodness-of-fit of two regression models **only if the models have the *same* regressand**.

2. $\bar{R}^2$ **should never be the sole criterion** for choosing between two or more sample regression equations.

5. The ANOVA Table for the OLS SRE

5.1 The General ANOVA Table

The **OLS sample regression equation (OLS-SRE)** is written as

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

The **Analysis-of-Variance (ANOVA) table** for the OLS SRE in a multiple linear regression model takes the following general form.

Source of variation	**SS**	**df**	**MSS = SS/df**
The regression function (explained)	$ESS = \sum_{i=1}^{N} \hat{y}_i^2$	K – 1	$\frac{ESS}{K-1} = \frac{\sum_i \hat{y}_i^2}{K-1}$
The residuals (unexplained)	$RSS = \sum_{i=1}^{N} \hat{u}_i^2$	N – K	$\frac{RSS}{N-K} = \frac{\sum_i \hat{u}_i^2}{N-K}$
Total sample variation of Y_i	$TSS = \sum_{i=1}^{N} y_i^2$	N – 1	

Definitions:

K $\equiv$ the ***total*** **number** of estimated **regression coefficients** in the OLS-SRE.

Thus, **k = K – 1** = the **number** of estimated ***slope*** **coefficients** in the OLS-SRE.

Interpretative Expression for ESS:

$$ESS \equiv \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i + \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ki} y_i$$

Interpretative Expression for RSS:

$$RSS \equiv \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} \hat{y}_i^2 = \sum_{i=1}^{N} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} y_i - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} y_i - \cdots - \hat{\beta}_k \sum_{i=1}^{N} x_{ki} y_i$$

5.2 The ANOVA F-statistic

The **ANOVA table** yields an ***F-statistic*** that is used to test the ***joint significance* of *all* the *slope* coefficients** in a multiple linear regression model.

- The ***unrestricted* PRE** is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad (i = 1, ..., N) \qquad (1)$$

- The **null and alternative hypotheses** are:

$$H_0: \beta_j = 0 \text{ for all } j = 1, \ldots, k \quad \Leftrightarrow \quad \beta_1 = 0 \textit{ and } \beta_2 = 0 \ldots \textit{ and } \beta_k = 0$$

$$H_1: \beta_j \neq 0 \text{ for } j = 1, \ldots, k \quad \Leftrightarrow \quad \beta_1 \neq 0 \textit{ and/or } \beta_2 \neq 0 \ldots \textit{ and/or } \beta_k \neq 0$$

The ***null* hypothesis H_0** says that ***all* slope coefficients** are ***jointly* equal to zero**.

The ***alternative* hypothesis H_1** says that ***some or all* of the slope coefficients** are ***not* equal to zero**.

- The ***restricted* PRE** corresponding to the **null hypothesis H_0** is obtained by substituting into the unrestricted PRE (1) the coefficient restrictions specified by H_0. That is, set $\beta_1 = 0$ *and* $\beta_2 = 0$ … *and* $\beta_k = 0$ in regression equation (1); this yields the restricted model:

$$Y_i = \beta_0 + u_i \qquad (i = 1, ..., N) \qquad (10)$$

Note: OLS estimation of regression equation (10) yields the ***restricted* sample regression equation**

$$Y_i = \tilde{\beta}_0 + \tilde{u}_i \qquad (i = 1, ..., N)$$

where

$\tilde{\beta}_0 = \bar{Y} = \dfrac{\sum_{i=1}^{N} Y_i}{N}$ = the **sample mean of the *observed* Y_i values** and

$\tilde{u}_i = Y_i - \tilde{\beta}_0 = Y_i - \bar{Y} = y_i$ = the ***restricted* OLS residuals** (i = 1, …, N).

- The **<u>ANOVA F-statistic</u>** is **the ratio of (1)** the **MSS** (mean sum-of-squares) **for the sample regression function to (2)** the **MSS for the residuals**:

$$\text{ANOVA}-F_0 = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2 \Big/ (K-1)}{\sum_{i=1}^{N}\hat{u}_i^2 \Big/ (N-K)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2 \Big/ (K-1)}{\hat{\sigma}^2}.$$

Note that the denominator of ANOVA-F_0 is the OLS estimator of σ^2:

$$\hat{\sigma}^2 = \frac{\text{RSS}}{(N-K)} = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{(N-K)}$$

- The **<u>null distribution of F_0</u>** – i.e., the distribution of ANOVA-F_0 under the null hypothesis H_0: $\beta_j = 0$ for all j = 1, …, k – is the **F[K–1, N–K] distribution**:

$$F_0 = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} \sim F[K-1, N-K] \quad \text{under } H_0\text{: } \beta_j = 0 \;\; \forall \;\; j = 1, \ldots, k$$

- **<u>Decision Rule -- Formulation 1</u>:**

Let $F_\alpha[K-1, N-K]$ = the ***α-level critical value*** of the $F[K-1, N-K]$ distribution.

Retain $\mathbf{H_0}$ at significance level α if $F_0 \le F_\alpha[K-1, N-K]$.

Reject $\mathbf{H_0}$ at significance level α if $F_0 > F_\alpha[K-1, N-K]$.

- **<u>Decision Rule -- Formulation 2</u>:**

Retain $\mathbf{H_0}$ at significance level α if the p-value for $F_0 \ge \alpha$.

Reject $\mathbf{H_0}$ at significance level α if the p-value for $F_0 < \alpha$.

- **<u>Alternative Formula for the ANOVA F-statistic</u>:**

Recall that the ANOVA F-statistic is written as

$$\text{ANOVA} - F_0 = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} \sim F[K-1, N-K].$$

Recall the definition of the R^2 for the unrestricted OLS SRE (2):

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

To obtain the alternative formula for ANOVA-F_0, divide the numerator and denominator of ANOVA-F_0 by TSS:

$$\text{ANOVA} - F_0 = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} = \frac{\text{ESS}/\text{TSS}/(K-1)}{\text{RSS}/\text{TSS}/(N-K)} = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}.$$

- ❑ **<u>Result</u>:** The **ANOVA F-statistic** can be calculated using either of **two equivalent formulas**:

$$\text{ANOVA} - F_0 = \frac{\text{ESS}/(K-1)}{\text{RSS}/(N-K)} = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}.$$

Note: Either formula allows the ANOVA F-statistic to be computed using only OLS estimates of the ***unrestricted*** **model given by equation (1) --** i.e., using only results for the ***unrestricted*** **OLS-SRE (2)**.