ECON 351* -- NOTE 17

F-Tests of Linear Coefficient Restrictions: A General Approach

1. Introduction

- The **population regression equation (PRE)** for the general multiple linear regression model takes the form:

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

where u_i is an iid (independently and identically distributed) random error term.

PRE (1) constitutes the *unrestricted* model.

- 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

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

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

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

- The **<u>OLS decomposition equation</u>** for the OLS-SRE (2) is

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

$$\begin{array}{ccccc} TSS & = & ESS_1 & + & RSS_1 \\ (N-1) & & (K-1) & & (N-K) \end{array}$$

where the figures in parentheses are the respective ***degrees of freedom*** for the **Total Sum of Squares (TSS)**, the **Explained (Regression) Sum of Squares (ESS_1)** and the **Residual (Error) Sum of Squares (RSS_1)**.

2. The ANOVA F-Test

<u>What is it?</u> The **ANOVA F-test** is

- a ***joint* test** of ***exclusion* restrictions** on ***all* K – 1 slope coefficients**
- a **test** of the ***joint* significance** of ***all* the slope coefficients** in a linear regression model.

<u>The Null and Alternative Hypotheses</u>: A test of the joint significance of all the $K-1 = k$ slope coefficient estimates $\hat{\beta}_j$ ($j = 1, 2, ..., k$) in OLS-SRE (2) consists of a test of the **null hypothesis**

$$H_0: \quad \beta_j = 0 \quad \forall\ j = 1, 2, ..., k$$
$$\beta_1 = 0 \ \textit{and} \ \beta_2 = 0 \ \textit{and} \ \ldots \ \beta_k = 0$$

against the **alternative hypothesis**

$$H_1: \quad \beta_j \neq 0 \qquad j = 1, 2, ..., k.$$
$$\beta_1 \neq 0 \ \textit{and/or} \ \beta_2 \neq 0 \ \textit{and/or} \ \ldots \ \beta_k \neq 0$$

<u>The *Unrestricted* Model</u> is the model corresponding to the alternative hypothesis H_1. It is simply the ***<u>unrestricted</u>* PRE** given by equation (1):

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

Number of free (unrestricted) regression coefficients in (1) is $\mathbf{K = K_1}$.

- From **the *analysis-of-variance table* for the OLS-SRE (2)**, it follows that the ANOVA F-statistic

$$\text{ANOVA} - F_0 = \frac{ESS_1/(K-1)}{RSS_1/(N-K)} = \frac{ESS_1}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.1)}$$

has the **F[K–1, N–K] distribution under $\mathbf{H_0}$**.

- Since $ESS_1 = TSS - RSS_1$, the ANOVA F-statistic can also be written as

$$\text{ANOVA} - F_0 = \frac{(TSS - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(TSS - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.2)}$$

The *Restricted* Model is the model corresponding to the null hypothesis H_0. It is obtained by imposing on the PRE of the unrestricted model (1) the coefficient restrictions specified by H_0. Accordingly, **setting $\beta_1 = \beta_2 = ... = \beta_k = 0$ in equation (1)** yields the ***restricted*** **PRE**:

$$Y_i = \beta_0 + u_i \qquad (4)$$

Number of free (unrestricted) regression coefficients in (4) is $\mathbf{K_0 = 1}$.

OLS estimation of (4) yields the ***restricted*** **OLS-SRE**

$$Y_i = \tilde{\beta}_0 + \tilde{u}_i \qquad (5)$$

where $\tilde{\beta}_0$ denotes the ***restricted*** **OLS estimator of the intercept coefficient β_0** and $\tilde{u}_i$ denotes **the i-th *restricted* OLS residual**.

- **Proposition:** It can be shown that the restricted OLS-SRE (5) has an Explained Sum of Squares $ESS_0 = 0$. Hence, the **OLS decomposition equation for the *restricted* OLS-SRE** takes the form:

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \tilde{u}_i^2$$

$$\downarrow \qquad \downarrow$$

$$\mathbf{TSS = RSS_0}$$

$$\mathbf{(N-1) \quad (N-1)}$$

Proof:

1. From the restricted SRE (5), the ***restricted* OLS residuals** are

$$\tilde{u}_i = Y_i - \tilde{\beta}_0$$

2. Summing these restricted residuals over the sample yields:

$$\sum_i \tilde{u}_i = \sum_i Y_i - N\tilde{\beta}_0$$

3. But the sum of these OLS residuals must equal zero by computational property (C3); setting $\sum_i \tilde{u}_i = 0$ implies that

$$\sum_i \tilde{u}_i = \sum_i Y_i - N\tilde{\beta}_0 = 0 \quad \Rightarrow \quad \begin{aligned} \sum_i Y_i - N\tilde{\beta}_0 &= 0 \\ N\tilde{\beta}_0 &= \sum_i Y_i \\ \tilde{\beta}_0 &= \sum_i Y_i / N \\ &= \bar{Y} \end{aligned}$$

where $\bar{Y}$ is the sample mean value of the Y_i values over i = 1, …, N.

4. The ***restricted* OLS residuals** $\tilde{u}_i = Y_i - \tilde{\beta}_0$ are therefore

$$\tilde{u}_i = Y_i - \tilde{\beta}_0 = Y_i - \bar{Y} = y_i \qquad \text{where} \qquad y_i = Y_i - \bar{Y} \text{ by definition.}$$

5. The ***squared restricted* residuals** are thus

$$\tilde{u}_i^2 = y_i^2,$$

so that the ***restricted* residual sum of squares for the *restricted* SRE** is

$$\sum_i \tilde{u}_i^2 = \sum_i y_i^2,$$

i.e.,

$$\mathbf{RSS_0 = TSS}.$$

- **Implication:** The result RSS_0 = TSS implies that the ANOVA F-statistic (3.2)

$$\text{ANOVA} - F_0 = \frac{(TSS - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(TSS - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.2)}$$

can be written equivalently as:

$$F = \frac{(RSS_0 - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.3)}$$

- **Summary:** We have derived ***three*** **different forms of the ANOVA F-statistic** for testing the joint significance of ***all*** the ***slope*** **coefficient estimates** -- i.e., for testing the **null hypothesis**

H_0: $\beta_j = 0 \quad \forall \; j = 1, 2, ..., k$
$\beta_1 = 0$ *and* $\beta_2 = 0$ *and* … $\beta_k = 0$

against the **alternative hypothesis**

H_1: $\beta_j \neq 0 \qquad j = 1, 2, ..., k.$
$\beta_1 \neq 0$ *and/or* $\beta_2 \neq 0$ *and/or* … $\beta_k \neq 0$

- **The three different forms of the ANOVA F-statistic are:**

$$\text{ANOVA} - F_0 = \frac{ESS_1/(K-1)}{RSS_1/(N-K)} = \frac{ESS_1}{RSS_1}\frac{(N-K)}{(K-1)} \qquad (3.1)$$

$$\text{ANOVA} - F_0 = \frac{(TSS - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(TSS - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad (3.2)$$

$$F = \frac{(RSS_0 - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.3)}$$

- **What is special about form (3.3) of the F-statistic?**

All three forms of the F-statistic can be used to perform the ANOVA F-test of the joint significance of all K−1 = k slope coefficients.

But ***only*** **form (3.3) of the F-statistic** can be used to perform a test of ***any*** **set of *linear* coefficient restrictions** in a multiple linear regression model. So the **F-statistic (3.3) is called a *general* F-statistic**.

- **Converting form (3.3) of the ANOVA F-statistic to the *general* F-statistic**

$$F = \frac{(RSS_0 - RSS_1)/(K-1)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-1)} \qquad \textbf{(3.3)}$$

Numerator **degrees of freedom of ANOVA F-statistic in (3.3):**

$\mathbf{RSS_0}$ has degrees-of-freedom = $df_0 = N - 1$

$\mathbf{RSS_1}$ has degrees-of-freedom = $df_1 = N - K$

$\therefore \quad df_0 - df_1 = N - 1 - (N - K) = N - 1 - N + K = K - 1$

Denominator **degrees of freedom of ANOVA F-statistic in (3.3):**

$\mathbf{RSS_1}$ has $df_1 = N - K$

<u>RESULT</u>: The ANOVA F-statistic in (3.3) is just **a special case of the *general* F-statistic**

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{df_1}{(df_0 - df_1)}$$

where $df_0 - df_1 =$ the **number of independent coefficient restrictions specified by $\mathbf{H_0}$**.

3. General F-Statistic for Testing Linear Coefficient Restrictions

3.1 Form 1 of the General F-Statistic

$$F = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-K_0)} \qquad \textbf{(F1)}$$

where:

RSS_0 = the ***residual sum of squares*** for the ***restricted*** **OLS-SRE**;
RSS_1 = the ***residual sum of squares*** for the ***unrestricted*** **OLS-SRE**;
K_0 = the ***number of free regression coefficients*** in the ***restricted*** **model**;
K = the ***number of free regression coefficients*** in the ***unrestricted*** **model**;
$K - K_0$ = the **number of *independent linear coefficient restrictions*** specified by the null hypothesis H_0;
$N - K$ = the ***degrees of freedom* for RSS_1**, the ***unrestricted*** **RSS**.

Form 1 of General F-Statistic: Alternative Formula

$$F = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-K_0)} \qquad \textbf{(F1)}$$

Alternatively, the general F-statistic (F1) can be written as

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{df_1}{(df_0 - df_1)} \qquad \textbf{(F1)}$$

where:

$df_0 = N - K_0$ = the ***degrees of freedom* for RSS_0**, the ***restricted*** **RSS**;
$df_1 = N - K$ = the ***degrees of freedom* for RSS_1**, the ***unrestricted*** **RSS**;
$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$.

Note: $df_0 - df_1 = K - K_0$ = the **number of *independent linear coefficient restrictions* specified by the null hypothesis H_0.**

Interpretation of Form 1 of General F-Statistic

- The ***restricted*** **OLS-SRE** estimated under the null hypothesis

  H_0: the β_j (j = 1, ..., k) satisfy $df_0 - df_1 = K - K_0$ independent linear coefficient restrictions

  takes the form

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \cdots + \tilde{\beta}_k X_{ki} + \tilde{u}_i = \tilde{Y}_i + \tilde{u}_i \quad (i = 1, \ldots, N) \quad (7.1)$$

where:

- the $\tilde{\beta}_j$ (j = 1, ..., k) are the ***restricted*** **OLS coefficient estimates**

- the $\tilde{Y}_i$ (i = 1, ..., N) are the ***restricted*** **predicted values** of the dependent variable Y, the ***restricted*** **OLS-SRF**

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

- the $\tilde{u}_i$ (i = 1, ..., N) are the ***restricted*** **OLS residuals**

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

- the **OLS decomposition equation** for the ***restricted*** **OLS-SRE** is

$$\underset{(N-1)}{TSS} = \underset{(K_0-1)}{ESS_0} + \underset{(N-K_0)}{RSS_0} \qquad (7.2)$$

- the ***restricted*** **R-squared** for the ***restricted*** **OLS-SRE** is

$$R_R^2 = \frac{ESS_0}{TSS} = 1 - \frac{RSS_0}{TSS}. \qquad (7.3)$$

❑ The ***unrestricted*** **OLS-SRE** estimated under the alternative hypothesis

H_1: the β_j (j = 1, ..., k) *do not* satisfy the $df_0 - df_1 = K - K_0$ independent linear coefficient restrictions specified by H_0

takes the form

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i = \hat{Y}_i + \hat{u}_i \quad (i = 1, \ldots, N) \quad (8.1)$$

where:

- the $\hat{\beta}_j$ (j = 1, ..., k) are the ***unrestricted*** **OLS coefficient estimates**

- the $\hat{Y}_i$ (i = 1, ..., N) are the ***unrestricted*** **predicted values** of the dependent variable Y, the ***unrestricted*** **OLS-SRF**

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

- the $\hat{u}_i$ (i = 1, ..., N) are the ***unrestricted*** **OLS residuals**

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

- the **OLS decomposition equation** for the ***unrestricted*** **OLS-SRE** is

$$\underset{(N-1)}{TSS} = \underset{(K-1)}{ESS_1} + \underset{(N-K)}{RSS_1} \qquad (8.2)$$

- the ***unrestricted*** **R-squared** for the ***unrestricted*** **OLS-SRE** is

$$R_U^2 = \frac{ESS_1}{TSS} = 1 - \frac{RSS_1}{TSS}. \qquad (8.3)$$

- **Compare the OLS decomposition equations for the *restricted* and *unrestricted* OLS-SREs**.

$$\mathbf{TSS = ESS_0 + RSS_0}. \qquad \text{[for } \textbf{\textit{restricted}} \textbf{ SRE]} \qquad (7.2)$$

$$\mathbf{TSS = ESS_1 + RSS_1}. \qquad \text{[for } \textbf{\textit{unrestricted}} \textbf{ SRE]} \qquad (8.2)$$

- Since the Total Sum of Squares (TSS) is the same for both decompositions, it follows that

$$ESS_0 + RSS_0 = ESS_1 + RSS_1. \qquad (9)$$

- Subtracting first RSS_1 and then ESS_0 from both sides of equation (9) allows equation (9) to be rewritten as:

$$RSS_0 - RSS_1 = ESS_1 - ESS_0 \qquad \mathbf{(10)}$$

where

$RSS_0 - RSS_1$ = the **increase in RSS** attributable to ***imposing* the restrictions** specified by the null hypothesis H_0;

$ESS_1 - ESS_0$ = the **increase in ESS** attributable to ***relaxing* the restrictions** specified by the null hypothesis H_0.

NOTE: **Imposing one or more linear coefficient restrictions** on the regression coefficients β_j (j = 0, 1, ..., k) always ***increases*** (or **leaves unchanged**) the ***residual* sum of squares**, and hence always ***reduces*** (or **leaves unchanged**) the ***explained* sum of squares**. Consequently,

$$RSS_0 \geq RSS_1 \iff ESS_1 \geq ESS_0$$

so that

$$RSS_0 - RSS_1 \geq 0 \iff ESS_1 - ESS_0 \geq 0.$$

In other words, **both sides of equation (10) are *always non-negative***.

3.2 Form 2 of the General F-Statistic

$$F = \frac{(R_U^2 - R_R^2)/(K - K_0)}{(1 - R_U^2)/(N - K)} = \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}\frac{(N - K)}{(K - K_0)} \qquad \textbf{(F2)}$$

where:

R_R^2 = the ***R-squared value*** for the ***restricted*** **OLS-SRE**;
R_U^2 = the ***R-squared value*** for the ***unrestricted*** **OLS-SRE**;
K_0 = the ***number of free regression coefficients*** in the ***restricted*** **model**;
K = the ***number of free regression coefficients*** in the ***unrestricted*** **model**;
$K - K_0$ = the number of ***independent linear coefficient restrictions*** specified by the null hypothesis H_0;
$N - K$ = the ***degrees of freedom*** **for** $\mathbf{RSS_1}$, the ***unrestricted*** **RSS**.

Form 2 of the General F-Statistic: Alternative Formula

$$F = \frac{(R_U^2 - R_R^2)/(K - K_0)}{(1 - R_U^2)/(N - K)} = \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}\frac{(N - K)}{(K - K_0)} \qquad \textbf{(F2)}$$

Alternatively, the general F-statistic (F2) can be written as

$$F = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} = \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}\frac{df_1}{(df_0 - df_1)} \qquad \textbf{(F2)}$$

where:

$df_0 = N - K_0$ = the ***degrees of freedom*** **for** $\mathbf{RSS_0}$, the ***restricted*** **RSS**;
$df_1 = N - K$ = the ***degrees of freedom*** **for** $\mathbf{RSS_1}$, the ***unrestricted*** **RSS**;
$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$.

Note: $df_0 - df_1 = K - K_0$ = the **number of *independent linear coefficient restrictions* specified by the null hypothesis** $\mathbf{H_0}$**.**

Derivation of Form 2 of the General F-Statistic

The second form of the general F-statistic can be obtained by dividing the numerator and denominator of the F-statistic (F1) by TSS; this yields

$$F = \frac{[RSS_0 - RSS_1]}{RSS_1}\frac{(N-K)}{(K-K_0)} = \frac{\left[\frac{RSS_0}{TSS} - \frac{RSS_1}{TSS}\right]}{\frac{RSS_1}{TSS}}\frac{(N-K)}{(K-K_0)} \qquad (11)$$

- Equation (7.3) shows that $ESS_0/TSS = R_R^2$ and $RSS_0/TSS = 1 - R_R^2$:

$$R_R^2 = \frac{ESS_0}{TSS} = 1 - \frac{RSS_0}{TSS} \quad \Rightarrow \quad \frac{RSS_0}{TSS} = 1 - R_R^2. \qquad (7.3)$$

- Equation (8.3) shows that $ESS_1/TSS = R_U^2$ and $RSS_1/TSS = 1 - R_U^2$:

$$R_U^2 = \frac{ESS_1}{TSS} = 1 - \frac{RSS_1}{TSS} \quad \Rightarrow \quad \frac{RSS_1}{TSS} = 1 - R_U^2. \qquad (8.3)$$

- Equations (7.3) and (8.3) thus imply that:

$$\frac{RSS_0}{TSS} - \frac{RSS_1}{TSS} = 1 - R_R^2 - (1 - R_U^2) = 1 - R_R^2 - 1 + R_U^2 = R_U^2 - R_R^2. \qquad (12)$$

- Substituting from (8.3) and (12) into equation (11) yields the second form of the general F-statistic:

$$F = \frac{(R_U^2 - R_R^2)/(K-K_0)}{(1-R_U^2)/(N-K)} = \frac{(R_U^2 - R_R^2)}{(1-R_U^2)}\frac{(N-K)}{(K-K_0)} \qquad \textbf{(F2)}$$

Interpretation of the General F-statistics (F1) and (F2)

$$F = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{(K-K_0)} \qquad \textbf{(F1)}$$

$$F = \frac{(R_U^2 - R_R^2)/(K - K_0)}{(1 - R_U^2)/(N-K)} = \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}\frac{(N-K)}{(K-K_0)} \qquad \textbf{(F2)}$$

- The **F-statistic (F1)** in effect determines whether ***imposing*** **the coefficient restrictions** specified by the null hypothesis H_0 significantly ***increases*** **the residual sum of squares RSS.**

- The **F-statistic (F2)** in effect determines whether ***imposing*** **the coefficient restrictions** specified by the null hypothesis H_0 **significantly *reduces* the coefficient of determination, R^2.**

- Note that since $ESS_1 \geq ESS_0$ (and $RSS_0 \geq RSS_1$), the ***unrestricted*** $\mathbf{R^2}$ **is *greater than or equal to*** **the *restricted*** $\mathbf{R^2}$; i.e.,

  $$R_U^2 \geq R_R^2 .$$

- The ***sample values*** **of the F-statistics (F1) and (F2) are always *equal*** **and *non-negative***.

4. Outline of the General F-Test Procedure

The procedure for testing a set of $df_0 - df_1 = K - K_0$ independent linear coefficient restrictions consists of **four basic steps**.

<u>Step 1</u>: Estimate the *unrestricted* model by OLS to obtain the ***unrestricted* OLS SRE** and associated statistics.

<u>Step 2</u>: Estimate the *restricted* model by OLS, after imposing on the unrestricted PRE the linear coefficient restrictions specified by the null hypothesis. This yields the ***restricted* OLS SRE** and associated statistics.

<u>Step 3</u>: Compute the *sample value* of either one of the *general F-statistics*.

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{df_1}{(df_0 - df_1)} \qquad \text{(F1)}$$

$$F = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} = \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}\frac{df_1}{(df_0 - df_1)} \qquad \text{(F2)}$$

where:

$df_0 = N - K_0$ = the ***degrees of freedom* for RSS_0**, the ***<u>restricted</u>* RSS**;
$df_1 = N - K$ = the ***degrees of freedom* for RSS_1**, the ***<u>unrestricted</u>* RSS**;
$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$
= the ***number* of linear coefficient restrictions specified by H_0**.

<u>Step 4</u>: Apply the conventional *decision rule*.

(1) If $\mathbf{F_0 > F_\alpha(K-K_0, N-K)}$, or if the **p-value for $F_0 < \alpha$**, ***reject* the coefficient restrictions** specified by the null hypothesis H_0 **at the 100α% significance level**.

(2) If $\mathbf{F_0 \le F_\alpha(K-K_0, N-K)}$, or if the **p-value for $F_0 \ge \alpha$**, ***retain (do not reject)* the coefficient restrictions** specified by the null hypothesis H_0 **at the 100α% significance level**.

5. Properties of the Restricted and Unrestricted Coefficient Estimators

The statistical properties of the ***restricted*** **OLS coefficient estimators** $\tilde{\beta}_j$ (j = 0, 1, ..., k) and the ***unrestricted*** **OLS coefficient estimators** $\hat{\beta}_j$ (j = 0, 1, ..., k) depend on whether the linear coefficient restrictions are ***true*** **or** ***false***.

5.1. If the linear coefficient restrictions specified by the null hypothesis H_0 are TRUE, then

(1) the ***restricted*** **coefficient estimators** $\tilde{\beta}_j$ (j = 0, 1, ..., k) are **unbiased** and have ***smaller*** **variances** than (i.e., are **efficient** relative to) the unrestricted coefficient estimators $\hat{\beta}_j$ -- i.e.,

$$E\left(\tilde{\beta}_j\right) = \beta_j \quad \text{and} \quad Var\left(\tilde{\beta}_j\right) \le Var\left(\hat{\beta}_j\right) \quad \text{for all j = 0, 1, ..., k;}$$

whereas

(2) the ***unrestricted*** **coefficient estimators** $\hat{\beta}_j$ (j = 0, 1, ..., k) are also **unbiased**, but have ***larger*** **variances** than (i.e., are **inefficient** relative to) the restricted coefficient estimators $\tilde{\beta}_j$ -- i.e.,

$$E\left(\hat{\beta}_j\right) = \beta_j \quad \text{and} \quad Var\left(\tilde{\beta}_j\right) \le Var\left(\hat{\beta}_j\right) \quad \text{for all j = 0, 1, ..., k.}$$

Implication: If the **linear coefficient restrictions** specified by the null hypothesis **are retained** by an F-test, then the ***restricted*** **coefficient estimators** $\tilde{\beta}_j$ **(j = 0, 1, ..., k) are preferred** because they have ***smaller*** **variances** than the unrestricted coefficient estimators $\hat{\beta}_j$ (j = 0, 1, ..., k). ***Both*** **estimators are** ***unbiased*** **if the coefficient restrictions are** ***true***.

5.2. If the linear coefficient restrictions specified by the null hypothesis H_0 are FALSE, then

(1) the ***restricted*** **coefficient estimators** $\tilde{\beta}_j$ (j = 0, 1, ..., k) are ***biased***, although they still have ***smaller*** **variances** than the unrestricted coefficient estimators $\hat{\beta}_j$ -- i.e.,

$$E(\tilde{\beta}_j) \neq \beta_j \quad \text{and} \quad Var(\tilde{\beta}_j) \leq Var(\hat{\beta}_j) \quad \text{for all } j = 0, 1, ..., k;$$

whereas

(2) the ***unrestricted*** **coefficient estimators** $\hat{\beta}_j$ (j = 0, 1, ..., k) are ***unbiased***, though they have ***larger*** **variances** than the restricted coefficient estimators $\tilde{\beta}_j$.

$$E(\hat{\beta}_j) = \beta_j \quad \text{and} \quad Var(\tilde{\beta}_j) \leq Var(\hat{\beta}_j) \quad \text{for all } j = 0, 1, ..., k.$$

Implication: If the **linear coefficient restrictions** specified by the null hypothesis **are rejected** by an F-test, then the ***unrestricted*** **coefficient estimators** $\hat{\beta}_j$ **(j = 0, 1, ..., k) are preferred** because they are ***unbiased***, even though they have larger variances than the restricted coefficient estimators $\tilde{\beta}_j$ (j = 0, 1, ..., k).

ECON 351* -- Addendum to NOTE 17

General F-Tests of Linear Coefficient Restrictions: A Numerical Example

1. An Example

Multiple Linear Regression Model 1

The PRE (population regression equation) for Model 1 is:

$$\text{price}_i = \beta_0 + \beta_1\text{wgt}_i + \beta_2\text{mpg}_i + \beta_3\text{wgt}_i^2 + \beta_4\text{mpg}_i^2 + \beta_5\text{wgt}_i\text{mpg}_i + u_i \qquad \textbf{(1)}$$

$K = 6$; $K - 1 = 5$.

- ***Marginal effect of wgt$_i$***

$$\frac{\partial\,\text{price}_i}{\partial\,\text{wgt}_i} = \frac{\partial\,E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial\,\text{wgt}_i} = \frac{\partial\,E(\text{price}_i \mid \bullet)}{\partial\,\text{wgt}_i} = \beta_1 + 2\beta_3\text{wgt}_i + \beta_5\text{mpg}_i\,.$$

- ***Marginal effect of mpg$_i$***

$$\frac{\partial\,\text{price}_i}{\partial\,\text{mpg}_i} = \frac{\partial\,E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial\,\text{mpg}_i} = \frac{\partial\,E(\text{price}_i \mid \bullet)}{\partial\,\text{mpg}_i} = \beta_2 + 2\beta_4\text{mpg}_i + \beta_5\text{wgt}_i\,.$$

Proposition 1 to Test

The ***marginal* effect of *mpg_i* on *$price_i$* is *zero***: i.e., ***mpg_i* has no effect on *$price_i$***; or **car *$price_i$* is unrelated to fuel efficiency** as measured by ***mpg_i***.

Examine the above expression for the **marginal effect of *mpg_i* on *$price_i$***. We see that a sufficient condition for the proposition to be true for all cars (regardless of their values for wgt_i and mpg_i) is that the **three coefficients β_2, β_4 and β_5 all equal *zero***.

If **$\beta_2 = 0$ *and* $\beta_4 = 0$ *and* $\beta_5 = 0$**, then $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0 \;\; \forall\, i$.

Null and Alternative Hypotheses

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i \qquad \textbf{(1)}$$

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \frac{\partial E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial \text{mpg}_i} = \frac{\partial E(\text{price}_i \mid \bullet)}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i .$$

- **H_0: $\beta_2 = 0$ *and* $\beta_4 = 0$ *and* $\beta_5 = 0$**

$$\Rightarrow \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0 \quad \forall\, i$$

- **H_1: $\beta_2 \neq 0$ *and/or* $\beta_4 \neq 0$ *and/or* $\beta_5 \neq 0$**

$$\Rightarrow \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i$$

- **H_0: $\beta_2 = 0$ *and* $\beta_4 = 0$ *and* $\beta_5 = 0$** $\Rightarrow \dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$

  ***Restricted* model** corresponding to H_0: set $\beta_2 = 0$ and $\beta_4 = 0$ and $\beta_5 = 0$ in the unrestricted model given by PRE (1).

  $$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + u_i \qquad \textbf{(2)}$$

  $K_0 = 3$; $K_0 - 1 = 2$; $N - K_0 = 74 - 3 = 71$;

  $RSS_0 = 384779934$; $df_0 = N - K_0 = 71$; $R_R^2 = 0.3941$.

  <u>**The OLS SRE for the *Restricted* Model – Model 2**</u>

```
. regress price wgt wgtsq

  Source |       SS       df       MS               Number of obs =      74
---------+------------------------------            F(  2,    71) =   23.09
   Model |  250285462     2   125142731             Prob > F      =  0.0000
Residual |  384779934    71  5419435.69             R-squared     =  0.3941
---------+------------------------------            Adj R-squared =  0.3770
   Total |  635065396    73  8699525.97             Root MSE      =  2328.0

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     wgt |  -7.273097   2.691747     -2.702   0.009      -12.64029   -1.905906
   wgtsq |   .0015142   .0004337      3.491   0.001       .0006494     .002379
   _cons |    13418.8   3997.822      3.357   0.001       5447.372    21390.23
------------------------------------------------------------------------------
```

- **H_1: $\beta_2 \neq 0$ *and/or* $\beta_4 \neq 0$ *and/or* $\beta_5 \neq 0$**

$$\Rightarrow \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i$$

***Unrestricted* model** corresponding to H_1: is PRE (1).

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i \qquad \textbf{(1)}$$

$K = 6$; $K - 1 = 5$; $N - K = 74 - 6 = 68$;

$RSS_1 = 326680563$; $df_1 = N - K = 68$; $R^2_U = 0.4856$.

<u>The OLS SRE for the *Unrestricted* Model – Model 1</u>

```
. regress price wgt wgtsq mpg mpgsq wgtmpg

  Source |       SS       df       MS                  Number of obs =      74
---------+------------------------------               F(  5,    68) =   12.84
   Model |   308384833     5  61676966.6               Prob > F      =  0.0000
Residual |   326680563    68  4804125.93               R-squared     =  0.4856
---------+------------------------------               Adj R-squared =  0.4478
   Total |   635065396    73  8699525.97               Root MSE      =  2191.8

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     wgt |  -31.88985   9.148215     -3.486   0.001      -50.14483   -13.63487
   wgtsq |   .0034574   .0008629      4.007   0.000       .0017355    .0051792
     mpg |  -3549.495   1126.464     -3.151   0.002      -5797.318   -1301.672
   mpgsq |   38.74472   12.62339      3.069   0.003       13.55514     63.9343
  wgtmpg |   .5421927   .1971854      2.750   0.008       .1487154    .9356701
   _cons |   92690.55   25520.53      3.632   0.001       41765.12      143616
------------------------------------------------------------------------------
```

2. The General F-Statistic

- ***Usage:*** Can be used **to test *any linear* restrictions on the regression *coefficients*** in a linear regression model.

- ***Formula 1*** for the general F-statistic is:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N - K)} \sim F[df_0 - df_1, df_1]$$

…(F1)

where:

RSS_0 = the ***residual sum of squares*** for the ***restricted*** **OLS-SRE**;
RSS_1 = the ***residual sum of squares*** for the ***unrestricted*** **OLS-SRE**;
K_0 = the ***number of free regression coefficients*** in the ***restricted*** **model**;
K = the ***number of free regression coefficients*** in the ***unrestricted*** **model**;
$K - K_0$ = the **number of *independent linear coefficient restrictions*** specified by the null hypothesis H_0;
$N - K$ = the ***degrees of freedom*** **for RSS_1**, the ***unrestricted*** **RSS**;
$df_0 = N - K_0$ = the ***degrees of freedom*** **for RSS_0**, the ***restricted*** **RSS**;
$df_1 = N - K$ = the ***degrees of freedom*** **for RSS_1**, the ***unrestricted*** **RSS**;
$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$.

Note: $df_0 - df_1 = K - K_0$ = the **number of *independent linear coefficient restrictions* specified by the null hypothesis H_0.**

- ***Formula 2*** for the **general F-statistic** is:

$$F = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} = \frac{(R_U^2 - R_R^2)/(K - K_0)}{(1 - R_U^2)/(N - K)} \sim F[df_0 - df_1, df_1] \quad \textbf{(F2)}$$

where:

R_R^2 = the ***R-squared value*** for the ***restricted*** **OLS-SRE**;
R_U^2 = the ***R-squared value*** for the ***unrestricted*** **OLS-SRE**.

3. Performing a General F-Test of Exclusion Restrictions

Calculate the sample value of the general F-statistic using Formula 1

$RSS_0 = 384779934$; $df_0 = N - K_0 = 71$; $R^2_R = 0.3941$.
$RSS_1 = 326680563$; $df_1 = N - K = 68$; $R^2_U = 0.4856$.

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$
$$= \frac{(384779934 - 326680563)/(71 - 68)}{326680563/68}$$
$$= \frac{58099371/3}{326680563/68}$$
$$= \mathbf{4.031}$$

Calculate the sample value of the general F-statistic using Formula 2

$$F_0 = \frac{(R^2_U - R^2_R)/(df_0 - df_1)}{(1 - R^2_U)/df_1}$$
$$= \frac{(0.4856 - 0.3941)/3}{(1 - 0.4856)/68}$$
$$= \frac{0.0915/3}{0.5144/68}$$
$$= \mathbf{4.032}$$

<u>Result</u>: $\mathbf{F_0 = 4.032}$

Decision Rule -- Formulation 1

Let $F_\alpha[df_0 - df_1, df_1]$ = the ***α-level critical value*** of the $F[df_0 - df_1, df_1]$ distribution.

***Retain* H₀** at significance level α if $F_0 \le F_\alpha[df_0 - df_1, df_1]$.

***Reject* H₀** at significance level α if $F_0 > F_\alpha[df_0 - df_1, df_1]$.

- ***Critical values of F[3, 68]***

  $\alpha = 0.05$: $F_{0.05}[3, 68] = 2.79$
  $\alpha = 0.01$: $F_{0.01}[3, 68] = 4.08$

```
. display invFtail(3, 68, 0.05)
2.794891

. display invFtail(3, 68, 0.01)
4.083395
```

Inference:

- F_0 = **4.032 > 2.79 =** $F_{0.05}[3, 68]$ ⇒ ***reject*** **H_0** at **5%** significance level
- F_0 = **4.032 < 4.08 =** $F_{0.01}[3, 68]$ ⇒ ***retain*** **H_0** at **1%** significance level

Decision Rule -- Formulation 2

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

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

- **p-value for F_0 = 0.0106**

```
. display Ftail(3, 68, 4.032)
.01062773
```

Inference:

- p-value for F_0 = **0.0106 < 0.05** ⇒ ***reject*** **H_0** at **5%** significance level
- p-value for F_0 = **0.0106 > 0.01** ⇒ ***retain*** **H_0** at **1%** significance level

Computing this General F-Test with *Stata*

The *Stata* ***test*** command computes general F-tests of any set of coefficient restrictions on linear regression models.

The OLS SRE for the Unrestricted Model – Model 1

```
. regress price wgt wgtsq mpg mpgsq wgtmpg

  Source |       SS       df       MS                  Number of obs =      74
---------+------------------------------               F(  5,    68) =   12.84
   Model |   308384833     5  61676966.6               Prob > F      =  0.0000
Residual |   326680563    68  4804125.93               R-squared     =  0.4856
---------+------------------------------               Adj R-squared =  0.4478
   Total |   635065396    73  8699525.97               Root MSE      =  2191.8

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     wgt |  -31.88985   9.148215     -3.486   0.001      -50.14483   -13.63487
   wgtsq |   .0034574   .0008629      4.007   0.000       .0017355    .0051792
     mpg |  -3549.495   1126.464     -3.151   0.002      -5797.318   -1301.672
   mpgsq |   38.74472   12.62339      3.069   0.003       13.55514     63.9343
  wgtmpg |   .5421927   .1971854      2.750   0.008       .1487154    .9356701
   _cons |   92690.55   25520.53      3.632   0.001       41765.12      143616
------------------------------------------------------------------------------

. test mpg mpgsq wgtmpg
 ( 1)  mpg = 0.0
 ( 2)  mpgsq = 0.0
 ( 3)  wgtmpg = 0.0

       F(  3,    68) =    4.03
            Prob > F =    0.0106
```

4. Outline of the General F-Test Procedure

Once the null hypothesis H_0 and the alternative hypothesis H_1 have been formulated, the procedure for testing a set of $df_0 - df_1 = K - K_0$ independent linear coefficient restrictions consists of **four basic steps**.

<u>**Step 1**</u>**: Estimate by OLS the *unrestricted* model implied by the alternative hypothesis H_1** to obtain the ***unrestricted* OLS SRE** and associated statistics:

$$RSS_1; \quad df_1 = N - K; \quad R_U^2 = ESS_1/TSS = 1 - RSS_1/TSS.$$

<u>**Step 2**</u>**: Estimate by OLS the *restricted* model implied by the null hypothesis H_0**, after imposing on the unrestricted PRE the linear coefficient restrictions specified by the null hypothesis. This yields the ***restricted* OLS SRE** and associated statistics:

$$RSS_0; \quad df_0 = N - K_0; \quad R_R^2 = ESS_0/TSS = 1 - RSS_0/TSS.$$

<u>**Step 3**</u>**: Compute the *sample value* of *either of the general F-statistics*.**

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} \sim F[df_0 - df_1, df_1]$$

where:

$df_0 = N - K_0 =$ the ***degrees of freedom* for RSS_0**, the <u>***restricted***</u> **RSS**;
$df_1 = N - K =$ the ***degrees of freedom* for RSS_1**, the <u>***unrestricted***</u> **RSS**;
$df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0$
= the ***number* of coefficient restrictions** specified by H_0.

<u>**Step 4**</u>**: Apply the conventional decision rule**:

(1) If $\mathbf{F_0 > F_\alpha(K-K_0, N-K)}$, or if the **p-value for $F_0 < \alpha$**, ***reject* the coefficient restrictions** specified by the null hypothesis H_0 **at the *100α% significance level***;

(2) If $\mathbf{F_0 \le F_\alpha(K-K_0, N-K)}$, or if the **p-value for $F_0 \ge \alpha$**, ***retain (do not reject)* the coefficient restrictions** specified by the null hypothesis H_0 **at the *100α% significance level***.

<u>Proposition 2 to Test</u>

The ***marginal*** **effect of** ***mpg_i*** **on** ***$price_i$*** **is** ***zero*** for cars that weigh **3,000 pounds** and have fuel efficiency of **23 miles per gallon**: i.e., ***mpg_i*** **has no effect on** ***$price_i$*** when ***wgt_i*** **=3000** and ***mpg_i*** **= 23**; or **car** ***$price_i$*** **is unrelated to fuel efficiency** as measured by ***mpg_i*** when ***wgt_i*** **=3000** and ***mpg_i*** **= 23**.

Null and Alternative Hypotheses

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i \qquad \textbf{(1)}$$

$$\frac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \bullet)}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \,.$$

- **H_0:** $\dfrac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$ when ***wgt_i*** **=3000** and ***mpg_i*** **= 23**

$$\Rightarrow \quad \beta_2 + 2\beta_4 23 + \beta_5 3000 = 0 \qquad \textit{or} \qquad \beta_2 + 46\beta_4 + 3000\beta_5 = 0$$

(number of restrictions = 1)

- **H_1:** $\dfrac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \neq 0$ when ***wgt_i*** **=3000** and ***mpg_i*** **= 23**

$$\Rightarrow \quad \beta_2 + 2\beta_4 23 + \beta_5 3000 \neq 0 \qquad \textit{or} \qquad \beta_2 + 46\beta_4 + 3000\beta_5 \neq 0$$

OLS Estimation of Model 1 in *Stata*

```
. regress price wgt wgtsq mpg mpgsq wgtmpg

  Source |       SS       df       MS                  Number of obs =      74
---------+------------------------------               F(  5,    68) =   12.84
   Model |  308384833     5  61676966.6               Prob > F      =  0.0000
Residual |  326680563    68  4804125.93               R-squared     =  0.4856
---------+------------------------------               Adj R-squared =  0.4478
   Total |  635065396    73  8699525.97               Root MSE      =  2191.8

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     wgt |  -31.88985   9.148215     -3.486   0.001      -50.14483   -13.63487
   wgtsq |   .0034574   .0008629      4.007   0.000       .0017355    .0051792
     mpg |  -3549.495   1126.464     -3.151   0.002      -5797.318   -1301.672
   mpgsq |   38.74472   12.62339      3.069   0.003       13.55514     63.9343
  wgtmpg |   .5421927   .1971854      2.750   0.008       .1487154    .9356701
   _cons |   92690.55   25520.53      3.632   0.001       41765.12      143616
------------------------------------------------------------------------------
```

Appropriate Test Statistics for Testing Proposition 2

- **t-statistic:**

$$t\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right) = \frac{\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000}{\widehat{se}\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right)}$$

Null distribution of $t\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right)$ under H_0 is $t[N - K] = t[74 - 6] = t[68]$.

- **F-statistic:**

$$F\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right) = \frac{\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right)^2}{\widehat{Var}\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right)}$$

Null distribution of $F\left(\hat\beta_2 + 2\hat\beta_4 23 + \hat\beta_5 3000\right)$ under H_0 is $F[1, N - K] = F[1, 74 - 6] = F[1, 68]$.

Note: Both the t-statistic and the F-statistic are computed using **only *unrestricted* coefficient estimates of Model 1** computed under the alternative hypothesis H_1.

<u>Using *Stata* to Compute the Two-Tail t-test:</u> Use the ***Stata*** **lincom** command.

- **H_0:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$ when ***wgt***

- **H_1:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \neq 0$ when wgt_i **=3000** and mpg_i **= 23**

```
. lincom _b[mpg] + 2*_b[mpgsq]*23 + _b[wgtmpg]*3000

 ( 1)  mpg + 46 mpgsq + 3000 wgtmpg = 0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -140.6595   90.73004    -1.55   0.126    -321.7085    40.38948
------------------------------------------------------------------------------

. return list

scalars:
                r(df) =  68
                r(se) =  90.73004185185361
          r(estimate) =  -140.6594967166902

. display r(estimate)/r(se)
-1.5503079

. display 2*ttail(r(df), abs(r(estimate)/r(se)))
.12570961

. display 2*ttail(68, abs(r(estimate)/r(se)))
.12570961
```

<u>Using *Stata* to Compute the F-test:</u> Use the ***Stata*** **test** command.

- **H_0:** $\dfrac{\partial\,\text{price}_i}{\partial\,\text{mpg}_i} = \beta_2 + 2\beta_4\text{mpg}_i + \beta_5\text{wgt}_i = 0$ when ***wgt***

- **H_1:** $\dfrac{\partial\,\text{price}_i}{\partial\,\text{mpg}_i} = \beta_2 + 2\beta_4\text{mpg}_i + \beta_5\text{wgt}_i \neq 0$ when $\boldsymbol{wgt_i}$ **=3000** and $\boldsymbol{mpg_i}$ **= 23**

```
. test mpg + 2*mpgsq*23 + wgtmpg*3000 = 0

 ( 1)  mpg + 46 mpgsq + 3000 wgtmpg = 0

       F(  1,    68) =    2.40
            Prob > F =    0.1257

. return list

scalars:
              r(drop) =  0
              r(df_r) =  68
                 r(F) =  2.403454450376304
                r(df) =  1
                 r(p) =  .1257096143054587

. display Ftail(r(df), r(df_r), r(F))
.12570961

. display Ftail(1, 68, r(F))
.12570961

. display sqrt(r(F))
1.5503079
```

Proposition 3 to Test

The ***marginal*** **effect of *mpg_i* on *$price_i$* is *zero*** for cars that weigh **3,000 pounds** and have fuel efficiency of **18 miles per gallon**: i.e., ***mpg_i* has no effect on *$price_i$*** when ***wgt_i* =3000** and ***mpg_i* = 18**; or **car *$price_i$* is unrelated to fuel efficiency** as measured by ***mpg_i*** when ***wgt_i* =3000** and ***mpg_i* = 18**.

Null and Alternative Hypotheses

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i \qquad \textbf{(1)}$$

$$\frac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \bullet)}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \,.$$

- **H_0:** $\dfrac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$ when ***wgt_i* =3000** and ***mpg_i* = 18**

$$\Rightarrow \quad \beta_2 + 2\beta_4 18 + \beta_5 3000 = 0 \qquad \textit{or} \qquad \beta_2 + 36\beta_4 + 3000\beta_5 = 0$$

(number of restrictions = 1)

- **H_1:** $\dfrac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \neq 0$ when ***wgt_i* =3000** and ***mpg_i* = 18**

$$\Rightarrow \quad \beta_2 + 2\beta_4 18 + \beta_5 3000 \neq 0 \qquad \textit{or} \qquad \beta_2 + 36\beta_4 + 3000\beta_5 \neq 0$$

<u>Using *Stata* to Compute the Two-Tail t-test:</u> Use the ***Stata*** **lincom** command.

- **H_0:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$ when $\boldsymbol{wgt_i}$ **=3000** and $\boldsymbol{mpg_i}$ **= 18**
- **H_1:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \neq 0$ when $\boldsymbol{wgt_i}$ **=3000** and $\boldsymbol{mpg_i}$ **= 18**

```
. lincom _b[mpg] + 2*_b[mpgsq]*18 + _b[wgtmpg]*3000

 ( 1)  mpg + 36 mpgsq + 3000 wgtmpg = 0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -528.1067   162.9029    -3.24   0.002    -853.1744    -203.039
------------------------------------------------------------------------------

. return list

scalars:
                r(df) =  68
                r(se) =  162.9029293974137
          r(estimate) =  -528.1067029465044

. display r(estimate)/r(se)
-3.241849

. display (r(estimate)/r(se))^2
10.509585

. display 2*ttail(r(df), abs(r(estimate)/r(se)))
.00184092
```

<u>Using *Stata* to Compute the F-test:</u> Use the ***Stata*** **test** command.

- **H_0:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = 0$ when ***wgt***

- **H_1:** $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i \neq 0$ when wgt_i **=3000** and mpg_i **= 18**

```
. test mpg + 2*mpgsq*18 + wgtmpg*3000 = 0

 ( 1)  mpg + 36 mpgsq + 3000 wgtmpg = 0

       F(  1,    68) =   10.51
            Prob > F =    0.0018

. return list

scalars:
              r(drop) =  0
              r(df_r) =  68
                 r(F) =  10.50958510526579
                r(df) =  1
                 r(p) =  .0018409212648082

. display Ftail(r(df), r(df_r), r(F))
.00184092

. display Ftail(1, 68, r(F))
.00184092

. display sqrt(r(F))
3.241849
```