ECON 351* -- NOTE 18

F-Tests of Exclusion Restrictions on Regression Coefficients: Numerical Examples 1

- Suppose the ***unrestricted* model** is given by the general PRE

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

The specific example we use is the model of car prices given by the PRE

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 weight_i^2 + \beta_3 mpg_i + u_i \qquad \textbf{(1)}$$

where

$price_i$ = the price of the i-th car (in US dollars);
$weight_i$ = the weight of the i-th car (in pounds);
mpg_i = the fuel efficiency of the i-th car (in miles per gallon);
N = 74 = the number of observations in the estimation sample.

Data Source: Stata-format data set **auto.dta** supplied with *Stata Release 10*.

The (partial) **marginal effects** of the **two explanatory variables *$weight_i$* and *mpg_i*** in regression equation (1) are:

$$\frac{\partial\, price_i}{\partial\, weight_i} = \frac{\partial\, E(price_i \mid weight_i, mpg_i)}{\partial\, weight_i} = \beta_1 + 2\beta_2 weight_i \qquad \textbf{(1.1)}$$

$$\frac{\partial\, price_i}{\partial\, mpg_i} = \frac{\partial\, E(price_i \mid weight_i, mpg_i)}{\partial\, mpg_i} = \beta_3 \qquad \textbf{(1.2)}$$

- The ***unrestricted*** **OLS-SRE** is written in general as

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(1*)}$$

or for our specific example as

$$price_i = \hat{\beta}_0 + \hat{\beta}_1 weight_i + \hat{\beta}_2 weight_i^2 + \hat{\beta}_3 mpg_i + \hat{u}_i \qquad (i = 1, \ldots, N) \qquad \textbf{(1*)}$$

where the $\hat{\beta}_j$ (j = 0, …, 3) are the OLS coefficient estimates and the $\hat{u}_i$ (i = 1, …, N) are the OLS residuals.

- Recall from *Note 17* the **two alternative forms of the general F-statistic**:

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

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

$df_0 = N - K_0$ = ***degrees of freedom*** for the ***restricted*** **RSS, RSS_0**;

$df_1 = N - K_1 = N - K$ = ***degrees of freedom*** for the ***unrestricted*** **RSS, RSS_1**;

$df_0 - df_1 = N - K_0 - (N - K_1) = N - K_0 - N + K_1 = K_1 - K_0 = K - K_0$
= the ***number of independent coefficient restrictions*** **specified by the null hypothesis H_0**.

- The ***null distribution*** **of F**: $\mathbf{F \sim F[df_0 - df_1, df_1]}$ under H_0.

<u>TEST 1</u>: a test of the *joint* significance of *all* the *slope* coefficients.

♦ The ***null* hypothesis** is:

H_0: $\beta_j = 0 \quad \forall \; j = 1, 2, 3;$ $\beta_1 = 0$ *and* $\beta_2 = 0$ *and* $\beta_3 = 0$.

♦ The ***alternative* hypothesis** is:

H_1: $\beta_j \neq 0 \quad j = 1, 2, 3;$ $\beta_1 \neq 0$ *and/or* $\beta_2 \neq 0$ *and/or* $\beta_3 \neq 0$.

• The ***unrestricted* model** corresponding to the alternative hypothesis H_1 is simply PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 weight_i^2 + \beta_3 mpg_i + u_i \qquad \textbf{(1)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K = K_1 = 4}$.

(1) The **OLS decomposition equation for the *unrestricted* OLS-SRE** is

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

$TSS = 635065396.0$ *with* $df = N - 1 = 74 - 1 = 73$
$ESS_1 = 262753599.0$ *with* $df = K - 1 = 4 - 1 = 3$
$\mathbf{RSS_1 = 372311797.0}$ ***with*** $\mathbf{df_1 = N - K = 74 - 4 = 70}$

(2) The $\mathbf{R^2}$ **for the *unrestricted* OLS-SRE** is

$$R_U^2 = \frac{ESS_1}{TSS} = \mathbf{0.4137}.$$

- The ***restricted* model** corresponding to the null hypothesis H_0 is obtained by setting $\beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$ in the unrestricted model (1):

$$Y_i = \beta_0 + u_i \tag{2}$$

$$\text{price}_i = \beta_0 + u_i \tag{2}$$

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

(1) The **OLS decomposition equation for the *restricted* OLS-SRE** is

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

$TSS = 635065396.0$ *with* $df = N - 1 = 74 - 1 = 73$
$ESS_0 = 0.0$ *with* $df = K_0 - 1 = 1 - 1 = 0$
$\mathbf{RSS_0 = 635065396.0}$ ***with*** $\mathbf{df_0 = N - K_0 = 74 - 1 = 73}$

(2) The $\mathbf{R^2}$ **for the *restricted* OLS-SRE** is

$$R_R^2 = \frac{ESS_0}{TSS} = \mathbf{0.0}.$$

- The ***sample value* of the F-statistic** is calculated by substituting in the formula

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

the values:

$RSS_0 = 635065396.0$ and $df_0 = N - K_0 = 74 - 1 = 73$
$RSS_1 = 372311797.0$ and $df_1 = N - K = 74 - 4 = 70$
$df_0 - df_1 = K - K_0 = 73 - 70 = 3.$

Thus, the ***sample value* of the F-statistic** is computed as follows:

$$\begin{aligned} F_0 &= \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \\ &= \frac{(635065396 - 372311797)/(73 - 70)}{372311797/70} \\ &= \frac{262753599/3}{372311797/70} \\ &= \frac{87584533.0}{5318739.957} \\ &= 16.467 \end{aligned}$$

Result: $\mathbf{F_0 = \underline{16.467}}$

- ***Null distribution* of the F-statistic**: $\mathbf{F_0 \sim F[3,70]}$ under H_0.

- The ***critical values* of the F[3,70]-distribution** are:

  for $\alpha = 0.05$: $F_\alpha[3,70] = F_{0.05}[3,70] = 2.75$.
  for $\alpha = 0.01$: $F_\alpha[3,70] = F_{0.01}[3,70] = 4.10$.

- **Inference at α = 0.05 (5% significance level)**:

  Since $F_0 = 16.467 > 2.75 = F_{0.05}[3,70]$, ***reject*** $\mathbf{H_0}$ at the ***5%*** significance level.

- **Inference at α = 0.01 (1% significance level)**:

  Since $F_0 = 16.467 > 4.10 = F_{0.01}[3,70]$, ***reject*** $\mathbf{H_0}$ at the ***1%*** significance level.

<u>TEST 2</u>: a test of the *joint* significance of the slope coefficients β_1 and β_2.

♦ The ***null* hypothesis** is:

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

Interpretation of H_0: This null hypothesis says that the **partial marginal effect of the explanatory variable *weight$_i$* equals *zero* for all values of *weight$_i$*** – i.e., that the explanatory variable ***weight$_i$* is unrelated to mean car prices for cars of the *same* fuel efficiency** (the same value of ***mpg$_i$***).

♦ The ***alternative* hypothesis** is:

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

Interpretation of H_1: This alternative hypothesis says that the **partial marginal effect of the explanatory variable *weight$_i$* *does not* equal *zero* for all values of *weight$_i$*** – i.e., that the explanatory variable ***weight$_i$* is related to mean car prices for cars of the *same* fuel efficiency** (the same value of ***mpg$_i$***).

• The ***<u>unrestricted</u>* <u>model</u>** corresponding to the alternative hypothesis H_1 is simply PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 weight_i^2 + \beta_3 mpg_i + u_i \qquad \textbf{(1)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K = K_1 = 4}$.

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

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

TSS = 635065396.0 *with* df = N –1 = 74 – 1 = 73
ESS_1 = 262753599.0 *with* df = K –1 = 4 – 1 = 3
RSS_1 = 372311797.0 *with* df_1 = N – K = 74 – 4 = 70

(2) The $\mathbf{R^2}$ **for the *unrestricted* OLS-SRE** is

$$R_U^2 = \frac{ESS_1}{TSS} = \mathbf{0.4137}.$$

- The ***restricted* model** corresponding to the null hypothesis H_0 is obtained by setting $\beta_1 = 0$ and $\beta_2 = 0$ in the unrestricted model (1):

$$Y_i = \beta_0 + \beta_3 X_{3i} + u_i \qquad \textbf{(3)}$$

$$price_i = \beta_0 + \beta_3 mpg_i + u_i \qquad \textbf{(3)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K_0 = 2}$.

(1) The **OLS decomposition equation for the *restricted* OLS-SRE** is

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

$TSS = 635065396.0$ *with* $df = N - 1 = 74 - 1 = 73$
$ESS_0 = 139449474.0$ *with* $df = K_0 - 1 = 2 - 1 = 1$
$\mathbf{RSS_0 = 495615923.0}$ ***with*** $\mathbf{df_0 = N - K_0 = 74 - 2 = 72}$

(2) The $\mathbf{R^2}$ **for the *restricted* OLS-SRE** is

$$R_R^2 = \frac{ESS_0}{TSS} = \mathbf{0.2196}.$$

- The ***sample value* of the F-statistic** is calculated by substituting in the formula

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

the values:

$RSS_0 = 495615923.0$ and $df_0 = N - K_0 = 74 - 2 = 72$
$RSS_1 = 372311797.0$ and $df_1 = N - K = 74 - 4 = 70$
$df_0 - df_1 = K - K_0 = 72 - 70 = 2.$

Thus, the ***sample value* of the F-statistic** is:

$$
\begin{aligned}
F_0 &= \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \\
&= \frac{(495615923 - 372311797)/(72 - 70)}{372311797/70} \\
&= \frac{123304126/2}{372311797/70} \\
&= \frac{61652063.0}{5318739.957} \\
&= 11.592
\end{aligned}
$$

Result: $\mathbf{F_0} = \underline{\mathbf{11.592}}$

- ***Null distribution* of the F-statistic**: $\mathbf{F_0 \sim F[2,70]}$ under H_0.

- The ***critical values* of the F[2,70]-distribution** are:

  for $\alpha = 0.05$: $F_\alpha[2,70] = F_{0.05}[2,70] = 3.14.$
  for $\alpha = 0.01$: $F_\alpha[2,70] = F_{0.01}[2,70] = 4.95.$

- **Inference at α = 0.05 (5% significance level)**:

  Since $F_0 = 11.592 > 3.14 = F_{0.05}[2,70]$, ***reject*** $\mathbf{H_0}$ at the ***5%*** significance level.

- **Inference at α = 0.01 (1% significance level)**:

  Since $F_0 = 11.592 > 4.95 = F_{0.01}[2,70]$, ***reject*** $\mathbf{H_0}$ at the ***1%*** significance level.

<u>TEST 3</u>: a test of the *individual* significance of the slope coefficient β_3.

♦ The ***null* hypothesis** is:

$$H_0: \quad \beta_3 = 0$$

Interpretation of H_0: This null hypothesis says that the **partial marginal effect of the explanatory variable *mpg_i* equals *zero*** – i.e., that the explanatory variable ***mpg_i* is *unrelated* to mean car prices for cars of the *same* weight** (the same value of ***$weight_i$***).

♦ The ***alternative* hypothesis** is:

$$H_1: \quad \beta_3 \neq 0$$

Interpretation of H_1: This alternative hypothesis says that the **partial marginal effect of the explanatory variable *mpg_i does not* equal *zero*** – i.e., that the explanatory variable ***mpg_i* is related to mean car prices for cars of the *same* weight** (the same value of ***$weight_i$***).

• The ***unrestricted* model** corresponding to the alternative hypothesis H_1 is simply PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 weight_i^2 + \beta_3 mpg_i + u_i \qquad \textbf{(1)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K = K_1 = 4}$.

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

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

TSS = 635065396.0 *with* df = N –1 = 74 – 1 = 73
ESS_1 = 262753599.0 *with* df = K –1 = 4 – 1 = 3
RSS_1 = 372311797.0 *with* df_1 = N – K = 74 – 4 = 70

(2) The $\mathbf{R^2}$ **for the *unrestricted* OLS-SRE** is

$$R_U^2 = \frac{ESS_1}{TSS} = \mathbf{0.4137}.$$

- The ***restricted* model** corresponding to the null hypothesis H_0 is obtained by setting $\beta_3 = 0$ in the unrestricted model (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad \textbf{(4)}$$

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 weight_i^2 + u_i \qquad \textbf{(4)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K_0 = 3}$.

(1) The **OLS decomposition equation for the *restricted* OLS-SRE** is

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

TSS = 635065396.0 *with* df = N −1 = 74 − 1 = 73
ESS_0 = 250285462.0 *with* df = K_0 −1 = 3 − 1 = 2
$\mathbf{RSS_0}$ = 384779934.0 ***with*** **$\mathbf{df_0 = N - K_0}$ = 74 − 3 = 71**

(2) The $\mathbf{R^2}$ **for the *restricted* OLS-SRE** is

$$R_R^2 = \frac{ESS_0}{TSS} = \mathbf{0.3941}.$$

- The ***sample value* of the F-statistic** is calculated by substituting in the formula

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

the values:

$RSS_0 = 384779934.0$ and $df_0 = N - K_0 = 74 - 3 = 71$
$RSS_1 = 372311797.0$ and $df_1 = N - K = 74 - 4 = 70$
$df_0 - df_1 = K - K_0 = 71 - 70 = 1.$

Thus, the ***sample value* of the F-statistic (F1)** is:

$$
\begin{aligned}
F_0 &= \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \\
&= \frac{(384779934 - 372311797)/(71 - 70)}{372311797/70} \\
&= \frac{12468137/1}{372311797/70} \\
&= \frac{12468137.0}{5318739.957} \\
&= 2.3442
\end{aligned}
$$

Result: $\mathbf{F_0 = \underline{2.3442}}$

- ***Null distribution* of the F-statistic**: $\mathbf{F_0 \sim F[1,70]}$ under H_0.

- The ***critical values* of the F[1, 70]-distribution** are:

for $\alpha = 0.10$: $F_\alpha[1,70] = F_{0.10}[1,70] = 2.78.$
for $\alpha = 0.05$: $F_\alpha[1,70] = F_{0.05}[1,70] = 3.99.$
for $\alpha = 0.01$: $F_\alpha[1,70] = F_{0.01}[1,70] = 7.04.$

- **Inference at $\alpha = 0.05$ (5% significance level)**:

Since $F_0 = 2.3442 < 3.99 = F_{0.05}[1,70]$, ***retain*** $\mathbf{H_0}$ at the ***5%*** significance level.

- **Inference at $\alpha = 0.10$ (10% significance level)**:

Since $F_0 = 2.3442 < 2.78 = F_{0.10}[1,70]$, ***retain*** $\mathbf{H_0}$ at the ***10%*** significance level.

- **<u>Alternative formula for the F-statistic</u> for testing H_0: $\beta_3 = 0$ against H_1: $\beta_3 \neq 0$:**

$$F(\hat{\beta}_3) = \frac{\left(\hat{\beta}_3 - \beta_3\right)^2}{\hat{Var}(\hat{\beta}_3)} \sim F[1, N-K] = F[1, 74-4] = F[1, 70].$$

- The ***sample value* of** $\mathbf{F(\hat{\beta}_3)}$ under H_0: $\beta_3 = 0$ is calculated by substituting into this formula the following values:

$\hat{\beta}_3 = -124.7675$;
$\beta_3 = 0$;
$\hat{Var}(\hat{\beta}_3) = \left[\hat{se}(\hat{\beta}_3)\right]^2 = 81.490^2 = 6640.6431$.

Thus, the ***sample value* of** $\mathbf{F(\hat{\beta}_3)}$ under H_0: $\beta_3 = 0$ is:

$$F_0(\hat{\beta}_3) = \frac{\left(\hat{\beta}_3 - \beta_3\right)^2}{\hat{Var}(\hat{\beta}_3)} = \frac{(-124.7675 - 0)^2}{6640.6431} = \frac{15566.929}{6640.6431} = 2.3442 = F_0.$$

<u>Result</u>: $F_0(\hat{\beta}_3)$ **=** F_0 using formula (F1) = **<u>2.3442</u>**.

- **A <u>two-tailed t-test</u> of H_0: $\beta_3 = 0$ vs. H_1: $\beta_3 \neq 0$:**

  The **t-statistic** is:

$$t(\hat{\beta}_3) = \frac{\hat{\beta}_3 - \beta_3}{\hat{se}(\hat{\beta}_3)} \sim t[N-K] = t[74-4] = t[70].$$

- The ***sample value*** of this t-statistic under H_0: $\beta_3 = 0$ is calculated by substituting into this formula the following values:

$$\hat{\beta}_3 = -124.7675;$$
$$\beta_3 = 0;$$
$$\hat{se}(\hat{\beta}_3) = 81.490.$$

  Thus, the ***sample value* of $t(\hat{\beta}_3)$** under H_0: $\beta_3 = 0$ is:

$$t_0(\hat{\beta}_3) = \frac{\hat{\beta}_3 - \beta_3}{\hat{se}(\hat{\beta}_3)} = \frac{-124.7675 - 0}{81.490} = \frac{-124.7675}{81.490} = \underline{\mathbf{-1.5311}}.$$

- ***Null distribution*** **of the t-statistic**: $\mathbf{t_0 \sim t[70]}$ under H_0.

- The ***two-tailed critical values*** **of the t[N−K] = t[70]-distribution** are:

  for $\alpha = 0.10$, $\alpha/2 = 0.05$: $t_{\alpha/2}[70] = t_{0.05}[70] = 1.667.$
  for $\alpha = 0.05$, $\alpha/2 = 0.025$: $t_{\alpha/2}[70] = t_{0.025}[70] = 1.994.$
  for $\alpha = 0.01$, $\alpha/2 = 0.005$: $t_{\alpha/2}[70] = t_{0.005}[70] = 2.648.$

- **Inference at α = 0.10 (10% significance level)**:

  Since $\left|t_0(\hat{\beta}_3)\right| = 1.531 < 1.668 = t_{\alpha/2}[70] = t_{0.05}[70]$, ***retain*** **$H_0$** at the ***10%*** significance level.

(1) Relationship between the two test statistics $\mathbf{t_0(\hat{\beta}_3)}$ and $\mathbf{F_0(\hat{\beta}_3)}$:

$$F_0(\hat{\beta}_3) = \frac{\left(\hat{\beta}_3 - \beta_3\right)^2}{\hat{Var}(\hat{\beta}_3)} = \left[\frac{\hat{\beta}_3 - \beta_3}{\sqrt{\hat{Var}(\hat{\beta}_3)}}\right]^2 = \left[\frac{\hat{\beta}_3 - \beta_3}{\hat{se}(\hat{\beta}_3)}\right]^2 = \left[t_0(\hat{\beta}_3)\right]^2.$$

<u>*Example -- Test 3*</u>*:* For this particular test,

$$\left[t_0(\hat{\beta}_3)\right]^2 = (-1.5311)^2 = 2.3442 = F_0(\hat{\beta}_3).$$

(2) The ***null distributions* of the two test statistics** are related according to a similar equality.

$$\left(t[N-K]\right)^2 \sim F[1, N-K].$$

i.e., the square of a t[N−K] distribution has the F[1, N−K] distribution.

- <u>***Implication 1***</u>***:*** The ***square* of the α/2 critical value of the t[N−K] distribution** ***equals*** the **α-level critical value of the F[1, N−K] distribution**;

  $$\left(t_{\alpha/2}[N-K]\right)^2 = F_\alpha[1,\ N-K].$$

  <u>*Example -- Test 3*</u>*:* For α = 0.10 (the 10% significance level)

  $F_\alpha[1,70] = F_{0.10}[1,70] = 2.779$; $t_{\alpha/2}[70] = t_{0.05}[70] = 1.667$.

  $$\therefore \left(t_{0.05}[70]\right)^2 = 1.667^2 = 2.779 = F_{0.10}[1,70].$$

- <u>***Implication 2***</u>***:***

  the ***two-tailed* p-value for $\mathbf{t_0(\hat{\beta}_3)}$** = the **p-value for $\mathbf{F_0(\hat{\beta}_3)}$**.

  <u>*Example -- Test 3*</u>*:*

  the ***two-tailed p-value* for $\mathbf{t_0(\hat{\beta}_3)}$ = 0.1303**
  the ***p-value* for** $F_0(\hat{\beta}_3)$ **= 0.1303**.

<u>TEST 4</u>: Testing a *restricted* model against an *unrestricted* model.

Consider **two alternative LOG-LOG (double-log) models** for car prices:

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + \alpha_3 (\ln w_i)^2 + \alpha_4 (\ln m_i)^2 + \alpha_5 (\ln w_i)(\ln m_i) + u_i \quad \textbf{(5)}$$

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + u_i \quad \textbf{(6)}$$

where:

$\ln p_i$ = the natural log of $price_i$ for the i-th car (in US dollars);
$\ln w_i$ = the natural log of $weight_i$ for the i-th car (in pounds);
$\ln m_i$ = the natural log of mpg_i for the i-th car (in miles per gallon);
N = 74 = the number of observations in the estimation sample.

- **Regression equation (5)** is an example of a ***variable* elasticity model**.
- **Regression equation (6)** is an example of a ***constant* elasticity model**.

Compare expressions for the **elasticity of *$price_i$* wrt *$weight_i$*** in models (5) and (6).

$$\frac{\partial \ln p_i}{\partial \ln w_i} = \frac{\partial E(\ln p_i \mid \ln w_i, \ln m_i)}{\partial \ln w_i} = \alpha_1 + 2\alpha_3 \ln w_i + \alpha_5 \ln m_i \qquad \text{in model (5).}$$

$$\frac{\partial \ln p_i}{\partial \ln w_i} = \frac{\partial E(\ln p_i \mid \ln w_i, \ln m_i)}{\partial \ln w_i} = \alpha_1 = \text{ a constant} \qquad \text{in model (6).}$$

Compare expressions for the **elasticity of *$price_i$* wrt *mpg_i*** in models (5) and (6).

$$\frac{\partial \ln p_i}{\partial \ln m_i} = \frac{\partial E(\ln p_i \mid \ln w_i, \ln m_i)}{\partial \ln m_i} = \alpha_2 + 2\alpha_4 \ln m_i + \alpha_5 \ln w_i \qquad \text{in model (5).}$$

$$\frac{\partial \ln p_i}{\partial \ln m_i} = \frac{\partial E(\ln p_i \mid \ln w_i, \ln m_i)}{\partial \ln m_i} = \alpha_2 = \text{ a constant} \qquad \text{in model (6).}$$

Question: Which model of car prices would you choose, model (5) or model (6)? Which model of car prices provides a better representation of the sample data?

Strategy: Address this question by ***testing* the *coefficient restrictions* that model (6) imposes on model (5)**.

- Which set of coefficient restrictions on model (5) will yield model (6)?

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + \alpha_3 (\ln w_i)^2 + \alpha_4 (\ln m_i)^2 + \alpha_5 (\ln w_i)(\ln m_i) + u_i \quad (i = 1, ..., N) \quad \textbf{(5)}$$

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + u_i \quad (i = 1, ..., N) \quad \textbf{(6)}$$

- By inspection (comparing models (5) and (6)), it can be seen that the following set of ***three* coefficient exclusion restrictions** on model (5) will yield model (6):

  $\alpha_3 = 0$ *and* $\alpha_4 = 0$ *and* $\alpha_5 = 0$.

- Test these three restrictions using an **F-test**.

- ***Decision criterion*** for choosing between models (5) and (6).

  If the F-test ***retains*** these three restrictions, **choose the *restricted* model (6)**.

  If the F-test ***rejects*** these three restrictions, **choose the *unrestricted* model (5)**.

<u>The F-Test Procedure</u>

<u>Step 1</u>: Formulate the *null* hypothesis H_0 and the *alternative* hypothesis H_1.

- The ***null* hypothesis** is:

  H_0: $\alpha_j = 0 \quad \forall\ j = 3, 4, 5;$ $\qquad \alpha_3 = 0$ *and* $\alpha_4 = 0$ *and* $\alpha_5 = 0$.

  Implies that the ***true* model** (the ***true* PRE**) is **model (6)**:

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + u_i \qquad \textbf{(6)}$$

- The ***alternative* hypothesis** is:

  H_1: $\alpha_j \neq 0 \quad j = 3, 4, 5;$ $\qquad \alpha_3 \neq 0$ *and/or* $\alpha_4 \neq 0$ *and/or* $\alpha_5 \neq 0$.

  Implies that the ***true* model** (the ***true* PRE**) is **model (5)**:

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + \alpha_3 (\ln w_i)^2 + \alpha_4 (\ln m_i)^2 + \alpha_5 (\ln w_i)(\ln m_i) + u_i \qquad \textbf{(5)}$$

Step 2: Formulate and estimate the *unrestricted* model corresponding to the *alternative* hypothesis H_1; save the values of RSS_1 and $df_1 = N - K$.

- The ***unrestricted* model** corresponding to the ***alternative* hypothesis H_1** is simply PRE (5):

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + \alpha_3 (\ln w_i)^2 + \alpha_4 (\ln m_i)^2 + \alpha_5 (\ln w_i)(\ln m_i) + u_i \qquad \textbf{(5)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K = K_1 = 6}$.

- The ***unrestricted* OLS SRE** obtained by OLS estimation of PRE (5) is written as:

$$\ln p_i = \hat{\alpha}_0 + \hat{\alpha}_1 \ln w_i + \hat{\alpha}_2 \ln m_i + \hat{\alpha}_3 (\ln w_i)^2 + \hat{\alpha}_4 (\ln m_i)^2 + \hat{\alpha}_5 (\ln w_i)(\ln m_i) + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(5*)}$$

(1) The **OLS decomposition equation for the *unrestricted* OLS-SRE** is

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

$TSS = 11.223533$ *with* $df = N - 1 = 74 - 1 = 73$

$ESS_1 = 5.6732234$ *with* $df = K - 1 = 6 - 1 = 5$

$\mathbf{RSS_1 = 5.5503097}$ *with* $\mathbf{df_1 = N - K = 74 - 6 = 68}$

(2) The $\mathbf{R^2}$ **for the *unrestricted* OLS-SRE** is

$$R_U^2 = \frac{ESS_1}{TSS} = \frac{5.6732234}{11.223533} = \mathbf{0.5055}.$$

<u>Step 3</u>: Formulate and estimate the *restricted* model corresponding to the *null* hypothesis H_0; save the values of RSS_0 and $df_0 = K - K_0$.

- The ***<u>restricted</u>* <u>model</u>** corresponding to the ***null* hypothesis H_0** is obtained by setting $\alpha_3 = 0$ *and* $\alpha_4 = 0$ *and* $\alpha_5 = 0$ in the unrestricted model (6):

$$\ln p_i = \alpha_0 + \alpha_1 \ln w_i + \alpha_2 \ln m_i + u_i \qquad \textbf{(6)}$$

Number of free (unrestricted) regression coefficients is $\mathbf{K_0 = 3}$.

- The ***restricted* OLS SRE** obtained by OLS estimation of PRE (6) is written as:

$$\ln p_i = \tilde{\alpha}_0 + \tilde{\alpha}_1 \ln w_i + \tilde{\alpha}_2 \ln m_i + \tilde{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(6*)}$$

(1) The **OLS decomposition equation for the *restricted* OLS-SRE** is

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

TSS = 11.223533 *with* df = N −1 = 74 −1 = 73
ESS_0 = 3.4323171 *with* df = K_0 −1 = 3−1 = 2
RSS_0 = 7.7912160 *with* **$df_0 = N - K_0 = 74 - 3 = 71$**

(2) The **R^2 for the *restricted* OLS-SRE** is

$$R_R^2 = \frac{ESS_0}{TSS} = \frac{3.4323171}{11.223533} = \mathbf{0.3058}.$$

<u>Step 4</u>: Compute the *sample value F_0* of the F-statistic.

- The ***sample value* of the F-statistic** is calculated by substituting in the formula

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

the values:

$$\begin{aligned} RSS_0 &= 7.7912160 \quad \text{and} \quad df_0 = N - K_0 = 74 - 3 = 71 \\ RSS_1 &= 5.5503097 \quad \text{and} \quad df_1 = N - K = 74 - 6 = 68 \\ & \qquad\qquad df_0 - df_1 = 71 - 68 = 3. \end{aligned}$$

Thus, the ***sample value* of the F-statistic** is:

$$\begin{aligned} F_0 &= \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \\ &= \frac{(7.7912160 - 5.5503097)/(71 - 68)}{5.5503097/68} \\ &= \frac{(7.7912160 - 5.5503097)/3}{5.5503097/68} \\ &= 9.15154. \end{aligned}$$

<u>Result</u>: **F_0 = <u>9.15154</u>**.

- ***Null distribution* of the F-statistic** is **F[3,68]**, the F-distribution with

  ***numerator* degrees of freedom** = $df_0 - df_1 = 71 - 68 = 3$;

  ***denominator* degrees of freedom** = $df_1 = N - K = 74 - 6 = 68$.

That is, $F \sim F[df_0 - df_1, df_1] = F[K - K_0, N - K] = F[3,68]$ under H_0.

<u>Step 5</u>: Apply the F-test *decision rule*, either Formulation 1 or Formulation 2.

- **<u>Decision Rule -- Formulation 1</u>:** At significance level α,

  - If $\mathbf{F_0 > F_\alpha[df_0 - df_1, df_1] = F_\alpha[K - K_0, N - K] = F_\alpha[3, 68]}$, ***reject*** the coefficient restrictions specified by the ***null* hypothesis $\mathbf{H_0}$** at the ***100α% significance level*** -- i.e., ***reject*** the ***restricted* model (6)**.

  - If $\mathbf{F_0 \le F_\alpha[df_0 - df_1, df_1] = F_\alpha[K - K_0, N - K] = F_\alpha[3, 68]}$, ***retain*** the coefficient restrictions specified by the ***null* hypothesis $\mathbf{H_0}$** at the ***100α% significance level*** -- i.e., ***retain*** the ***restricted* model (6)**.

- Obtain the ***critical values* of the F[3,68]-distribution**.

  The ***5%*** and ***1% critical values* of the F[3,68]-distribution** are:

  for $\alpha = 0.05$: $F_\alpha[3, 68] = F_{0.05}[3,68] = 2.74$.
  for $\alpha = 0.01$: $F_\alpha[3, 68] = F_{0.01}[3,68] = 4.08$.

- **Inference at α = 0.05 (the *5%* significance level)**:

  Since $F_0 = 9.1515 > 2.74 = F_{0.05}[3,68]$, ***reject* $\mathbf{H_0}$** at the ***5%*** significance level.

- **Inference at α = 0.01 (the *1%* significance level)**:

  Since $F_0 = 9.1515 > 4.08 = F_{0.01}[3,68]$, ***reject* $\mathbf{H_0}$** at the ***1%*** significance level.

<u>Conclusion</u>: The ***restricted* model (6) is *rejected*** against the ***unrestricted* model (5)** at both the ***5%*** and ***1%*** significance levels.

Choose the *unrestricted* model (5) over the restricted model (6).

The sample evidence indicates that the coefficient restrictions that model (6) imposes on model (5) are not true. Therefore, the ***restricted* OLS coefficient estimates of model (6) are likely <u>biased</u>**, whereas the ***unrestricted* OLS coefficient estimates of model (5) are <u>unbiased</u>**.

- **<u>Decision Rule -- Formulation 2</u>:** At significance level α,

  - If the **p-value for $F_0 < \alpha$**, ***reject*** the coefficient restrictions specified by the ***null* hypothesis H_0** at the ***100α%* significance level** -- i.e., ***reject*** the ***restricted* model (6)**.

  - If the **p-value for $F_0 \geq \alpha$**, ***retain*** the coefficient restrictions specified by the ***null* hypothesis H_0** at the ***100α%* significance level** -- i.e., ***retain*** the ***restricted* model (6)**.

- **Compute the *p-value of F_0*, the sample value of the F-statistic**.

  Stata command:

  ```
  display Ftail(3, 68, F0)
  ```

  The ***p-value of F_0* = 0.00003624**.

- **Inference at α = 0.05 (the *5%* significance level)**:

  Since p-value of F_0 = 0.00003624 < 0.05, ***reject* H_0** at the ***5%*** significance level.

- **Inference at α = 0.01 (the *1%* significance level)**:

  Since p-value of F_0 = 0.00003624 < 0.01, ***reject* H_0** at the ***1%*** significance level.

<u>Conclusion</u>: The ***restricted* model (6) is *rejected*** against the ***unrestricted* model (5)** at both the 5% and 1% significance levels.

Choose the *unrestricted* model (5) over the restricted model (6).

The sample evidence indicates that the coefficient restrictions that model (6) imposes on model (5) are not true. Therefore, the ***restricted* OLS coefficient estimates of model (6) are likely <u>biased</u>**, whereas the ***unrestricted* OLS coefficient estimates of model (5) are <u>unbiased</u>**.