ECON 351* -- Addendum to NOTE 17

General F-Tests of Linear Coefficient Restrictions: An Example

1. An Example

Multiple Linear Regression Model 1

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

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \beta_5 \text{mpg}_i^2 + \beta_6 \text{wgt}_i \text{mpg}_i + u_i . \qquad \textbf{(1)}$$

$k = 6; \quad k - 1 = 5.$

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

$$\frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \frac{\partial E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial \text{wgt}_i} = \frac{\partial E(\text{price}_i \mid \bullet)}{\partial \text{wgt}_i} = \beta_2 + 2\beta_3 \text{wgt}_i + \beta_6 \text{mpg}_i .$$

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

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \frac{\partial E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial \text{mpg}_i} = \frac{\partial E(\text{price}_i \mid \bullet)}{\partial \text{mpg}_i} = \beta_4 + 2\beta_5 \text{mpg}_i + \beta_6 \text{wgt}_i .$$

Proposition 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_4, β_5 and β_6 all equal *zero***.

If $\beta_4 = 0$ ***and*** $\beta_5 = 0$ ***and*** $\beta_6 = 0$, then $\dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_4 + 2\beta_5 \text{mpg}_i + \beta_6 \text{wgt}_i = 0 \ \forall\ i.$

Null and Alternative Hypotheses

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \beta_5 \text{mpg}_i^2 + \beta_6 \text{wgt}_i \text{mpg}_i + u_i . \quad \textbf{(1)}$$

$$\frac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i)}{\partial\, \text{mpg}_i} = \frac{\partial\, E(\text{price}_i \mid \bullet)}{\partial\, \text{mpg}_i} = \beta_4 + 2\beta_5 \text{mpg}_i + \beta_6 \text{wgt}_i .$$

- **H_0: $\beta_4 = 0$ *and* $\beta_5 = 0$ *and* $\beta_6 = 0$** $\Rightarrow \dfrac{\partial\, \text{price}_i}{\partial\, \text{mpg}_i} = \beta_4 + 2\beta_5 \text{mpg}_i + \beta_6 \text{wgt}_i = 0$

  ***Restricted* model** corresponding to H_0: set $\beta_4 = 0$ and $\beta_5 = 0$ and $\beta_6 = 0$ in the unrestricted model given by PRE (1).

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + u_i \quad \textbf{(2)}$$

$k_0 = 3$; $k_0 - 1 = 2$; $N - k_0 = 74 - 3 = 71$.

$RSS_0 = 384779934$; $df_0 = N - k_0 = 71$; $R_R^2 = 0.3941$.

The OLS SRE for the *Restricted* Model – Model 2

```
. 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: $\beta_4 \neq 0$ *and/or* $\beta_5 \neq 0$ *and/or* $\beta_6 \neq 0$** $\Rightarrow \dfrac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_4 + 2\beta_5 \text{mpg}_i + \beta_6 \text{wgt}_i$

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

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \beta_5 \text{mpg}_i^2 + \beta_6 \text{wgt}_i \text{mpg}_i + u_i \qquad \textbf{(1)}$$

$k = 6$; $k - 1 = 5$; $N - k = 74 - 6 = 68$.

$RSS_1 = 326680563$; $df_1 = N - k = 68$; $R_U^2 = 0.4856$.

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
------------------------------------------------------------------------------
```

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:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(k - k_0)}{RSS_1/(N-k)} \sim F[df_0 - df_1, df_1] \qquad \textbf{(F1)}$$

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**.
$df_0 = N - k_0$ = 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:

$$F = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} = \frac{(R_U^2 - R_R^2)/(k - k_0)}{(1 - R_U^2)/(N-k)} \sim F[df_0 - df_1, df_1]. \qquad \textbf{(F2)}$$

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**.

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 $\mathbf{H_0}$ at significance level α if $F_0 \le F_{\alpha}[df_0 - df_0, df_1]$.

Reject $\mathbf{H_0}$ at significance level α if $F_0 > F_{\alpha}[df_0 - df_0, df_1]$.

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

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

Decision Rule -- Formulation 2

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

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

- **p-value for $\mathbf{F_0}$ = 0.0106**

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

Inference:

- p-value for F_0 = **0.0106 < 0.05** ⇒ ***reject*** $\mathbf{H_0}$ at **5%** significance level
- p-value for F_0 = **0.0106 > 0.01** ⇒ ***retain*** $\mathbf{H_0}$ 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***.

5. Properties of the Restricted and Unrestricted Coefficient Estimators

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

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

whereas

(2) the ***unrestricted*** **coefficient estimators** $\hat{\beta}_j$ (j = 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.,

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

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

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

whereas

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

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

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