ECON 452* -- The Skinny on NOTE 10

Testing Linear Coefficient Restrictions in Linear Regression Models: The Fundamentals

This note outlines the fundamentals of statistical inference in linear regression models.

- **In scalar notation**, the population regression equation, or PRE, for the linear regression model is written in general as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i \qquad \forall\, i \tag{1.1}$$

or

$$Y_i = \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ij} + u_i \qquad \forall\, i \tag{1.2}$$

or

$$Y_i = \sum_{j=0}^{j=k} \beta_j X_{ij} + u_i\,, \quad X_{i0} = 1\ \forall\, i \qquad \forall\, i \tag{1.3}$$

where

$Y_i \equiv$ the i-th population value of the regressand, or dependent variable;

$X_{ij} \equiv$ the i-th population value of the j-th regressor, j = 1, …, k;

$\beta_j \equiv$ the partial slope coefficient of X_{ij}, j = 1, …, k;

$u_i \equiv$ the i-th population value of the unobservable random error term.

- **In vector-matrix notation**, the population regression equation, or PRE, **for a sample of N observations** on a linear regression model can be written as:

$$y = X\beta + u \tag{2}$$

where

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \text{ the } N \times 1 \text{ regressand vector}$$

= the N×1 <u>column</u> vector of observed sample values of the regressand, or dependent variable, Y_i (i = 1, ..., N);

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix} = \text{ the } N \times 1 \text{ error vector}$$

= the N×1 <u>column</u> vector of unobserved random error terms u_i (i = 1, ..., N) corresponding to each of the N sample observations.

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix} = \text{ the } N\times K \text{ regressor matrix}$$

= the N×K matrix of observed sample values of the K = k + 1 regressors X_{i0}, X_{i1}, X_{i2}, ..., X_{ik} (i = 1, ..., N), where the first regressor is a constant equal to 1 for all observations ($X_{i0} = 1 \; \forall \; i = 1, ..., N$).

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \text{ the } K\times 1 \text{ regression coefficient vector}$$

= the K×1 or (k+1)×1 column vector of unknown partial regression coefficients β_j, j = 0, 1, ..., k.

- **Statistical inference** consists of both

1. **testing hypotheses** on the regression coefficient vector β and
2. **constructing confidence intervals** for the individual elements of β.

1. Assumption A6: The Error Normality Assumption

In order to perform statistical inference in the linear regression model, it is necessary to specify the form of the probability distribution of the error vector u in population regression equation (1). The normality assumption does this.

❑ Scalar Formulation of the Error Normality Assumption A6

The random error terms u_i are ***independently*** **and** ***identically*** **distributed** as the ***normal*** **distribution** with

1. **zero conditional means**

$$E\left(u_i \middle| x_i^T\right) = E(u_i) = 0 \qquad \forall\ i$$

2. **constant conditional variances**

$$\mathrm{Var}\left(u_i \middle| x_i^T\right) = E\left(u_i^2 \middle| x_i^T\right) = E\left(u_i^2 \middle| 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma^2 > 0 \qquad \forall\ i$$

3. **zero conditional covariances**

$$\mathrm{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = E\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \qquad \forall\ i \neq s$$

- A compact way of stating error normality assumption A6 is:

  conditional on x_i^T, the u_i are iid as $N(0, \sigma^2)$ **(A6.1)**

  where

  "iid" means "*independently* and *identically* distributed"

  $N(0, \sigma^2)$ denotes a normal distribution with zero mean and variance σ^2.

  Even more briefly, we can say that

  $u_i \mid x_i^T$ are iid as $N(0, \sigma^2)$. **(A6.2)**

❑ Matrix Formulation of the Error Normality Assumption A6

The **N×1 error vector u** has a ***multivariate*** **normal distribution** with

1. a zero conditional mean vector

$E(u\,|\,X) = \underline{0}$ where $\underline{0}$ is an N×1 vector of zeros

2. a constant scalar diagonal covariance matrix V(u)

$V(u\,|\,X) = E(uu^T\,|\,X) = \sigma^2 I_N$ where I_N is the N×N identity matrix

- A compact way of stating the error normality assumption in matrix terms is:

$$u\,|\,X \sim N(\underline{0}, \sigma^2 I_N) \qquad \textbf{(A6)}$$

where $N(\cdot\,,\cdot)$ here denotes the N-variate normal distribution.

❑ Implications of Assumption A6 for the Distribution of the Regressand Vector y

- **Linearity Property of Normal Distribution:** Any linear function of a normally distributed random variable is itself normally distributed.

- **y is a linear function of u:** The PRE $y = X\beta + u$ states that the regressand vector y is a linear function of the error vector u.

- ***Implication:*** Since u is normally distributed by assumption A6 and y is a linear function of u by assumption A1, the linearity property of the normal distribution implies that

$$y \mid X \sim N\left(X\beta, \sigma^2 I_N\right).$$

That is, the **regressand vector y has an N-variate *normal* distribution** with

(1) **conditional mean vector** equal to $E\left(y \mid X\right) = X\beta$

and

(2) **conditional covariance matrix** equal to $V\left(y \mid X\right) = \sigma^2 I_N$.

❑ Implications of Assumption A6 for the Distribution of the OLS Coefficient Estimator $\hat{\beta}$

- $\hat{\beta}$ **is a linear function of y.** Conditional on the regressors X, the OLS coefficient estimator $\hat{\beta}$ is a linear function of the regressand vector y:

$$\hat{\beta}_{OLS} = \hat{\beta} = \left(X^T X\right)^{-1} X^T y$$

- ***Implication:*** Since y is normally distributed by implication of assumption A6 and $\hat{\beta}$ is a linear function of y, the linearity property of the normal distribution implies that

$$\hat{\beta} \,\middle|\, X \sim N\left(\beta, \sigma^2 (X^T X)^{-1}\right). \qquad (3)$$

That is, the **OLS coefficient estimator $\hat{\beta}$ has a K-variate *normal* distribution** with

(1) **conditional mean vector** equal to $E\left(\hat{\beta} \,\middle|\, X\right) = \beta$

and

(2) **conditional covariance matrix** equal to $V\left(\hat{\beta} \,\middle|\, X\right) = \sigma^2 (X^T X)^{-1}$.

2. Formulation of Linear Equality Restrictions on β

The general hypothesis to be tested is that the coefficient vector β satisfies a set of q independent linear restrictions, where $q < K$. We formulate this general hypothesis in vector-matrix form, since this corresponds to the way in which econometric software such as *Stata* is written.

The **null hypothesis H_0** is written in general as:

$$H_0: \quad R\beta = r \quad \Leftrightarrow \quad R\beta - r = \underline{0}$$

The **alternative hypothesis H_1** is written in general as:

$$H_1: \quad R\beta \neq r \quad \Leftrightarrow \quad R\beta - r \neq \underline{0}$$

In H_0 and H_1 above:

R = a $q \times K$ matrix of specified constants;

β = the $K \times 1$ coefficient vector;

r = a $q \times 1$ vector of specified constants;

$\underline{0}$ = a $q \times 1$ null vector, i.e., a $q \times 1$ vector of zeros.

- The q×K restrictions matrix R takes the form

$$R = \begin{bmatrix} r_{10} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{20} & r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{q0} & r_{q1} & r_{q2} & \cdots & r_{qk} \end{bmatrix}$$

where

r_{mj} = the constant on coefficient β_j in the m-th linear restriction, m = 1, …, q.

- The q×1 restrictions vector r takes the form

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

where

r_m = the constant term in the m-th linear restriction, m = 1, …, q.

- The matrix-vector product Rβ is a q×1 vector of linear functions of the regression coefficients β_0, β_1, β_2, … , β_k:

$$R\beta = \underset{(q\times K)}{\begin{bmatrix} r_{10} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{20} & r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{q0} & r_{q1} & r_{q2} & \cdots & r_{qk} \end{bmatrix}} \underset{(K\times 1)}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix}}$$

- The null and alternative hypotheses can therefore be written as follows:

$$H_0\text{: } R\beta = r \quad \Rightarrow \quad \begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

$$H_1\text{: } R\beta \neq r \quad \Rightarrow \quad \begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix} \neq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

Some Specific Examples

Consider the linear regression model given by the PRE

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i \qquad (i = 1, \ldots, N) \qquad (4)$$

Test 1

The null and alternative hypotheses are:

H_0: $\beta_2 = 0$ one linear restriction on coefficient vector β

H_1: $\beta_2 \neq 0$

- The restrictions matrix R in this case is the 1×5 row vector:

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

- The restrictions vector r is in this case the scalar 0 since there is only one restriction specified by the null hypothesis H_0:

$$r = 0.$$

- The matrix-vector product $R\beta$ in this case is:

$$R\beta = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = 0\beta_0 + 0\beta_1 + 1\beta_2 + 0\beta_3 + 0\beta_4 = \beta_2$$

- The null hypothesis H_0: $R\beta = r$ is therefore the single equation:

H_0: $\beta_2 = 0$

Test 2

The PRE is again

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i \qquad (i = 1, \ldots, N) \qquad (4)$$

The null and alternative hypotheses are:

H_0: $\beta_1 = 0$ and $\beta_2 = 0$ two linear restrictions on coefficient vector β

H_1: $\beta_1 \neq 0$ and/or $\beta_2 \neq 0$

- The restrictions matrix R in this case is the 2×5 row vector:

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

- The restrictions vector r is in this case the 2×1 column vector of zeros:

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- The matrix-vector product $R\beta$ in this case is:

$$R\beta = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0\beta_0 + 1\beta_1 + 0\beta_2 + 0\beta_3 + 0\beta_4 \\ 0\beta_0 + 0\beta_1 + 1\beta_2 + 0\beta_3 + 0\beta_4 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

- The null hypothesis H_0: $R\beta = r$ is therefore the matrix equation:

H_0: $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ which says "$\beta_1 = 0$ and $\beta_2 = 0$"

Test 3

The PRE is again

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i \qquad (i = 1, \ldots, N) \qquad (4)$$

The null and alternative hypotheses are:

H_0: $\beta_1 = \beta_3$ and $\beta_2 = -\beta_4$ *or* $\beta_1 - \beta_3 = 0$ and $\beta_2 + \beta_4 = 0$ $(q = 2)$

H_1: $\beta_1 \neq \beta_3$ and/or $\beta_2 \neq \beta_4$ *or* $\beta_1 - \beta_3 \neq 0$ and/or $\beta_2 + \beta_4 \neq 0$

- The restrictions matrix R in this case is the 2×5 row vector:

$$R = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- The restrictions vector r is in this case the 2×1 column vector of zeros:

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- The matrix-vector product $R\beta$ in this case is:

$$R\beta = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0\beta_0 + 1\beta_1 + 0\beta_2 - 1\beta_3 + 0\beta_4 \\ 0\beta_0 + 0\beta_1 + 1\beta_2 + 0\beta_3 + 1\beta_4 \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_3 \\ \beta_2 + \beta_4 \end{bmatrix}$$

- The null hypothesis H_0: $R\beta = r$ is therefore the matrix equation:

$$H_0: \quad \begin{bmatrix} \beta_1 - \beta_3 \\ \beta_2 + \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$ which says "$\beta_1 - \beta_3 = 0$ and $\beta_2 + \beta_4 = 0$"

Test 4

The PRE is again

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i \qquad (i = 1, \ldots, N) \qquad (4)$$

The null and alternative hypotheses are:

H_0: $\beta_1 + 2\beta_2 = \beta_3 + 2\beta_4$ *or* $\beta_1 + 2\beta_2 - \beta_3 - 2\beta_4 = 0$ $(q = 1)$

H_1: $\beta_1 + 2\beta_2 \neq \beta_3 + 2\beta_4$ *or* $\beta_1 + 2\beta_2 - \beta_3 - 2\beta_4 \neq 0$

- The restrictions matrix R in this case is the 1×5 row vector:

$$R = \begin{bmatrix} 0 & 1 & 2 & -1 & -2 \end{bmatrix}$$

- The restrictions vector r is in this case the 1×1 scalar 0:

$$r = 0$$

- The matrix-vector product $R\beta$ in this case is the 1×1 scalar:

$$R\beta = \begin{bmatrix}0 & 1 & 2 & -1 & -2\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{bmatrix} = \begin{bmatrix}0\beta_0 + 1\beta_1 + 2\beta_2 - 1\beta_3 - 2\beta_4\end{bmatrix}$$

$$= \beta_1 + 2\beta_2 - \beta_3 - 2\beta_4$$

- The null hypothesis H_0: $R\beta = r$ is therefore the equation:

H_0: $\beta_1 + 2\beta_2 - \beta_3 - 2\beta_4 = 0$

3. The Three Principles of Hypothesis Testing

- Given the null hypothesis H_0: $R\beta - r = \underline{0}$ and the alternative hypothesis H_1: $R\beta - r \neq \underline{0}$, there are **two alternative sets of parameter estimates** of the PRE $y = X\beta + u$ that one might use to compute a test statistic.

  1. The ***restricted*** **parameter estimates** computed under H_0: $R\beta - r = \underline{0}$, which are denoted as follows:

     $\tilde{\beta}$ = the ***restricted*** OLS estimator of β;

     $\tilde{u} = y - X\tilde{\beta}$ = the ***restricted*** OLS residual vector;

     $RSS_0 = RSS_R = RSS(\tilde{\beta}) = \tilde{u}^T\tilde{u} = \sum_{i=1}^{N}\tilde{u}_i^2$
     = the ***restricted*** residual sum of squares;

     $df_0 = N - (K - q) = N - K + q$ = the degrees of freedom for RSS_0;

     $\tilde{\sigma}^2 = RSS_0/df_0 = RSS_0/N - (K - q)$ = the ***restricted*** OLS estimator of σ^2;

     $R_R^2 = ESS_0/TSS = 1 - (RSS_0/TSS)$ = the ***restricted*** R-squared.

2. The ***unrestricted*** **parameter estimates** computed under H_1: $R\beta - r \neq \underline{0}$, which are denoted as follows:

   $\hat{\beta}$ = the ***unrestricted*** OLS estimator of β;

   $\hat{u} = y - X\hat{\beta}$ = the ***unrestricted*** residual vector;

   $RSS_1 = RSS_U = RSS(\hat{\beta}) = \hat{u}^T\hat{u} = \sum_{i=1}^{N}\hat{u}_i^2$

   = the ***unrestricted*** residual sum of squares;

   $df_1 = N - K$ = the degrees of freedom for RSS_1;

   $\hat{\sigma}^2 = RSS_1/N - K$ = the ***unrestricted*** OLS estimator of σ^2.

   $R_U^2 = ESS_1/TSS = 1 - (RSS_1/TSS)$ = the ***unrestricted*** R-squared.

- The computation of hypothesis tests of linear coefficient restrictions can be performed in general in three different ways:

  1. using ***only*** **the *unrestricted* parameter estimates** of the model;
  2. using ***only*** **the *restricted* parameter estimates** of the model;
  3. using ***both*** **the *restricted* and *unrestricted* parameter estimates** of the model.

- These three options correspond to the **three fundamental principles of hypothesis testing**.

  1. The **Wald principle** of hypothesis testing computes hypothesis tests using ***only*** **the *unrestricted* parameter estimates** of the model computed under the alternative hypothesis H_1.

  2. The **Lagrange Multiplier (LM) principle** of hypothesis testing computes hypothesis tests using ***only*** **the *restricted* parameter estimates** of the model computed under the null hypothesis H_0.

  3. The **Likelihood Ratio (LR) principle** of hypothesis testing computes hypothesis tests using ***both*** the ***restricted*** **parameter estimates** of the model computed under the null hypothesis H_0 ***and*** the ***unrestricted*** **parameter estimates** of the model computed under the alternative hypothesis H_1.

4. Likelihood Ratio F-Tests of Linear Coefficient Restrictions

❑ Null and Alternative Hypotheses

- The **null hypothesis** is that the regression coefficient vector β satisfies a set of q independent linear coefficient restrictions:

$$H_0: \quad R\beta = r \quad \Leftrightarrow \quad R\beta - r = \underline{0}$$

- The **alternative hypothesis** is that the regression coefficient vector β does not satisfy the set of q independent linear coefficient restrictions specified by H_0:

$$H_1: \quad R\beta \neq r \quad \Leftrightarrow \quad R\beta - r \neq \underline{0}$$

❑ **The Likelihood Ratio F-Statistic:** can be written in either of two equivalent forms.

1. <u>Form 1 of the LR F-statistic</u> is expressed in terms of the restricted and unrestricted residual sums of squares, RSS_0 and RSS_1:

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

$$F_{LR} = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{q} \qquad \textbf{(F1)}$$

where:

RSS_0 = the ***residual sum of squares*** for the ***<u>restricted</u>*** **OLS-SRE**;
$df_0 = N - K_0$ = the ***degrees of freedom*** **for RSS_0**, the ***<u>restricted</u>*** **RSS**;
$K_0 = K - q$ = the ***number of free regression coefficients*** in the ***<u>restricted</u>*** **model**;

RSS_1 = the ***residual sum of squares*** for the ***<u>unrestricted</u>*** **OLS-SRE**;
$df_1 = N - K$ = the ***degrees of freedom*** **for RSS_1**, the ***<u>unrestricted</u>*** **RSS**;
$K = k + 1$ = the ***number of free regression coefficients*** in the ***<u>unrestricted</u>*** **model**;

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

Note: The value of q is calculated as follows:

$$q = df_0 - df_1 = N - K_0 - (N - K) = N - K_0 - N + K = K - K_0.$$

2. <u>Form 2 of the LR F-statistic</u> is expressed in terms of the restricted and unrestricted R-squared values, R_R^2 and R_U^2:

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

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

where:

R_R^2 = the ***R-squared value*** for the ***<u>restricted</u>*** **OLS-SRE**;
$K_0 = K - q$ = the ***number of free regression coefficients*** in the ***<u>restricted</u>*** **model**;
$df_0 = N - K_0 = N - (K - q) = N - K + q$ = the ***degrees of freedom*** **for RSS_0**, the ***<u>restricted</u>*** **RSS**;

R_U^2 = the ***R-squared value*** for the ***<u>unrestricted</u>*** **OLS-SRE**;
$K = k + 1$ = the ***number of free regression coefficients*** in the ***<u>unrestricted</u>*** **model**;
$df_1 = N - K$ = the ***degrees of freedom*** **for RSS_1**, the ***<u>unrestricted</u>*** **RSS**;

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

❑ Null distribution of the LR F-statistic

Under error normality assumption A6, the LR F-statistic F_{LR} is distributed under H_0 (i.e., assuming the null hypothesis H_0 is true) as F[q, N−K], the F distribution with q numerator degrees of freedom and N−K denominator degrees of freedom:

$$F_{LR} \sim F[q, N-K] \quad \text{under } H_0\text{: } R\beta = r.$$

5. Wald F-Tests of Linear Coefficient Restrictions

- The **Wald F-Test is Based on the Wald Principle of Hypothesis Testing**

The **Wald principle** of hypothesis testing computes hypothesis tests using ***only* the *unrestricted* parameter estimates** of the model computed under the alternative hypothesis H_1: $R\beta \neq r$. These unrestricted parameter estimates can be denoted as $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)$.

- **<u>General Wald F-statistic</u>.** The general Wald F-statistic is obtained by simply dividing the general Wald statistic W in (10) by q, the number of independent linear coefficient restrictions specified by the null hypothesis H_0: $R\beta = r$:

$$F_{WALD} = \frac{1}{q} W = \frac{\left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{\hat{\beta}} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \tag{9}$$

where:

W = the **general Wald statistic** given below;

$\hat{\beta}$ = a ***consistent unrestricted* estimator of β**, such as the OLS estimator;

$\hat{V}_{\hat{\beta}}$ = a ***consistent* estimator of $V_{\hat{\beta}}$**.

The ***general Wald test statistic* W** for testing the null hypothesis H_0: $R\beta = r$ against the alternative hypothesis H_1: $R\beta \neq r$ takes the form

$$W = \left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{\hat{\beta}} R^T\right)^{-1} \left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2[q] \text{ under } H_0 \qquad (10)$$

where

$\hat{\beta}$ = a ***consistent unrestricted* estimator of β**, such as the OLS estimator;

$\hat{V}_{\hat{\beta}}$ = a ***consistent* estimator of $\mathbf{V}_{\hat{\beta}}$**;

$\chi^2[q]$ = the **chi-square distribution** with **q degrees of freedom**.

Note: Both the coefficient estimator $\hat{\beta}$ and the coefficient covariance matrix estimator $\hat{V}_{\hat{\beta}}$ used in the general Wald statistic W must be ***consistent***, and are computed using only ***unrestricted* estimates** of the linear regression model under the alternative hypothesis H_1: $R\beta \neq r$.

- **Null distribution of Wald-F Statistic:** With the error normality assumption A6, the null distribution of the general Wald-F statistic -- that is, the distribution of the Wald-F statistic if the null hypothesis H_0 is true -- is $F[q, N-K]$, the central F distribution with q numerator degrees of freedom and N−K denominator degrees of freedom.

  The short way of saying this is:

  $$F_{WALD} = \frac{1}{q}W \sim F[q, N-K] \quad \text{under } H_0\text{: } R\beta = r \tag{11}$$

  where

  $F[q, N-K]$ = the F-distribution with q numerator degrees of freedom and N−K denominator degrees of freedom.

  Notes:

  1. The null distribution of the F_{WALD} statistic is exactly F[q, N−K] only if the error normality assumption A6 is true.
  2. However, even if the normality assumption A6 is not true, the null distribution of the F_{WALD} statistic is still approximately F[q, N−K] under fairly general conditions.

- **Common Form of the Wald F-statistic**. In practice, the most common form of the Wald F-statistic is that obtained by using the OLS coefficient covariance matrix estimator in place of $\hat{V}_{\hat{\beta}}$ in (9) and (10):

$$F_W = \frac{1}{q} W_{OLS} = \frac{\left(R\hat{\beta} - r\right)^T \left(R\,\hat{V}_{OLS}\,R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q} \tag{12}$$

where

$\hat{\beta} = \hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y =$ the unrestricted OLS estimator of β;

$\hat{V}_{OLS}\left(\hat{\beta}\right) = \hat{V}_{OLS} = \hat{\sigma}^2\left(X^T X\right)^{-1} =$ the OLS estimator of $V_{\hat{\beta}}$;

$\hat{\sigma}^2 = \dfrac{RSS_1}{N-K} = \dfrac{\hat{u}^T\hat{u}}{N-K} = \dfrac{\sum_{i=1}^{N}\hat{u}_i^2}{N-K} =$ the unrestricted OLS estimator of σ^2;

$W_{OLS} = \left(R\hat{\beta} - r\right)^T \left(R\,\hat{V}_{OLS}\,R^T\right)^{-1}\left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2[q]$ under H_0.

- **Null distribution of the F_W Statistic:** With the error normality assumption A6, the null distribution of the F_W statistic (12) – that is, the distribution of the Wald-F statistic if the null hypothesis H_0 is true – is $F[q, N-K]$, the F distribution with q numerator degrees of freedom and N−K denominator degrees of freedom.

  The short way of saying this is:

$$F_W = \frac{1}{q} W_{OLS} \sim F[q, N-K] \quad \text{under } H_0\text{: } R\beta = r \tag{13}$$

  where $F[q, N-K]$ = the F-distribution with q numerator degrees of freedom and N−K denominator degrees of freedom.

- **Notes on Computation of F_W**

- The Wald F-statistic F_W in (12) is computed using only the ***unrestricted*** **OLS coefficient estimates** $\hat{\beta}$ and the OLS estimate $\hat{V}_{OLS}$ of the variance-covariance matrix of $\hat{\beta}$.

- Both the ***unrestricted*** **OLS coefficient estimator** $\hat{\beta}$ and the **OLS covariance matrix estimator** $\hat{V}_{OLS}$ are ***unbiased*** **and** ***consistent*** under the assumptions of the classical linear regression model.

6. Relationship Between Wald and LR F-Tests

❑ The Wald and LR F-Statistics

$$F_W = \frac{1}{q} W_{OLS} = \frac{\left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{OLS} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \sim F[q, N-K] \text{ under } H_0$$

$$F_{LR} = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(N-K)} = \frac{(RSS_0 - RSS_1)}{RSS_1}\frac{(N-K)}{q} \sim F[q, N-K] \text{ under } H_0$$

❑ Key Result

The key to understanding the relationship between the Wald F-statistic F_W and the LR F-statistic F_{LR} is the following important result (given without the tedious proof):

The quadratic form $\Phi(\hat{\beta})$ defined as

$$\Phi(\hat{\beta}) = \left(R\hat{\beta} - r\right)^T \left(R\,(X^T X)^{-1} R^T\right)^{-1} \left(R\hat{\beta} - r\right)$$

can be shown to equal the difference between the restricted and unrestricted residual sums of squares

$$RSS_0 - RSS_1 = \tilde{u}^T\tilde{u} - \hat{u}^T\hat{u}\ .$$

That is,

$$\Phi(\hat{\beta}) = \left(R\hat{\beta} - r\right)^T \left(R\,(X^TX)^{-1}R^T\right)^{-1}\left(R\hat{\beta} - r\right) = \tilde{u}^T\tilde{u} - \hat{u}^T\hat{u} = RSS_0 - RSS_1 \qquad (14)$$

❑ Rewrite the F_W Statistic

- Use the result (14) and the formula for $\hat{\sigma}^2_{OLS}$ to rewrite the Wald F-statistic F_W.

1. Rewrite the Wald F-statistic F_W as follows

Substitute for $\hat{V}_{OLS}$ in the formula for F_W the expression

$$\hat{V}_{OLS} = \hat{\sigma}^2\left(X^TX\right)^{-1}$$

This gives

$$F_W = \frac{\left(R\hat{\beta} - r\right)^T \left(R\,\hat{V}_{OLS}\,R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q}$$

$$= \frac{\left(R\hat{\beta} - r\right)^T \left(R\,\hat{\sigma}^2_{OLS}(X^TX)^{-1}R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q}$$

$$= \frac{\left(R\hat{\beta} - r\right)^T \left(\hat{\sigma}^2_{OLS}R(X^TX)^{-1}R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q}$$

$$= \frac{\left(R\hat{\beta} - r\right)^T \left(R(X^T X)^{-1} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q \hat{\sigma}^2_{OLS}}$$

$$= \frac{\left(R\hat{\beta} - r\right)^T \left(R(X^T X)^{-1} R^T\right)^{-1} \left(R\hat{\beta} - r\right) \big/ q}{\hat{\sigma}^2_{OLS}} \qquad (15)$$

2. Now substitute for $\hat{\sigma}^2_{OLS}$ in (15) the expression

$$\hat{\sigma}^2_{OLS} = \frac{RSS_1}{N-K} = \frac{\hat{u}^T \hat{u}}{N-K}.$$

This allows us to rewrite the F_W statistic as

$$F_W = \frac{\left(R\hat{\beta} - r\right)^T \left(R(X^T X)^{-1} R^T\right)^{-1} \left(R\hat{\beta} - r\right) \big/ q}{\hat{\sigma}^2_{OLS}}$$

$$= \frac{\left(R\hat{\beta} - r\right)^T \left(R(X^T X)^{-1} R^T\right)^{-1} \left(R\hat{\beta} - r\right) \big/ q}{\hat{u}^T \hat{u} \big/ (N-K)}.$$

3. Finally, use result (14) above to replace the quadratic form in the numerator of F_W, namely $\left(R\hat{\beta}-r\right)^T\left(R\,(X^TX)^{-1}R^T\right)^{-1}\left(R\hat{\beta}-r\right)$, with the equivalent difference between the restricted residual sum of squares $\tilde{u}^T\tilde{u}$ and the unrestricted residual sum of squares $\hat{u}^T\hat{u}$. This permits the F_W statistic to be written as:

$$F_W = \frac{\left(R\hat{\beta}-r\right)^T\left(R(X^TX)^{-1}R^T\right)^{-1}\left(R\hat{\beta}-r\right)\big/q}{\hat{u}^T\hat{u}/(N-K)}$$

$$= \frac{\left(\tilde{u}^T\tilde{u} - \hat{u}^T\hat{u}\right)\big/q}{\hat{u}^T\hat{u}\big/(N-K)} \tag{16.1}$$

$$= \frac{\left(RSS_0 - RSS_1\right)\big/q}{RSS_1\big/(N-K)} \tag{16.2}$$

where $RSS_0 = \tilde{u}^T\tilde{u}$ = the restricted residual sum of squares and $RSS_1 = \hat{u}^T\hat{u}$ = the unrestricted residual sum of squares.

- **<u>Result</u>:** The Wald F-statistic F_W can be written in terms of the restricted and unrestricted residual sums of squares as

$$F_W = \frac{\left(R\hat{\beta}-r\right)^T\left(R\,\hat{V}_{OLS}\,R^T\right)^{-1}\left(R\hat{\beta}-r\right)}{q} = \frac{\left(RSS_0 - RSS_1\right)\big/q}{RSS_1\big/(N-K)}. \tag{17}$$

❑ The F_W and F_{LR} Statistics are Equal

$$F_W = \frac{(R\hat{\beta} - r)^T \left(R\hat{V}_{OLS} R^T\right)^{-1}(R\hat{\beta} - r)}{q} = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(N-K)} = F_{LR}.$$

❑ Tests Based on the F_W and F_{LR} Statistics are Equivalent

The Wald F-statistic F_W and the LR F-statistic F_{LR} yield equivalent or identical tests of H_0: $R\beta = r$ against H_1: $R\beta \neq r$.

This equivalence follows from two facts:

1. **The two test statistics F_W and F_{LR} are *equal***; that is, they yield identical calculated sample values of the F-statistic.

$$F_W = F_{LR}$$

2. **The two test statistics F_W and F_{LR} have *identical null distributions***, namely the F[q, N−K] distribution.

$$F_W \sim F[q, N-K] \quad \text{under} \quad H_0\text{: } R\beta = r$$

and

$$F_{LR} \sim F[q, N-K] \quad \text{under} \quad H_0\text{: } R\beta = r.$$

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

$$F_W = F_{LR} \sim F[q, N - K] \quad \text{under} \quad H_0\text{: } R\beta = r.$$