Testing Multiple Restrictions – The Wald and F Test

We'll be concerned here with testing more general hypotheses than those seen to date. Also concerned with constructing interval predictions from our regression model.

Examples

- $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$; $H_0: \boldsymbol{\beta} = \boldsymbol{0}$ *vs.* $H_A: \boldsymbol{\beta} \neq \boldsymbol{0}$
- $\log(Q) = \beta_1 + \beta_2\log(K) + \beta_3\log(L) + \varepsilon$

  $H_0: \beta_2 + \beta_3 = 1$ *vs.* $H_A: \beta_2 + \beta_3 \neq 1$
- $\log(q) = \beta_1 + \beta_2\log(p) + \beta_3\log(y) + \varepsilon$

  $H_0: \beta_2 + \beta_3 = 0$ *vs.* $H_A: \beta_2 + \beta_3 \neq 0$

If we can obtain one model from another by imposing restrictions on the parameters of the first model, we say that the 2 models are "***Nested***".

We'll be concerned with (several) possible restrictions on $\boldsymbol{\beta}$, in the usual model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ; \quad \boldsymbol{\varepsilon} \sim N[0\,, \sigma^2 I_n]$$

(X is non-random ; $rank(X) = k$)

Let's focus on ***linear restrictions***:

$$r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k = q_1$$
$$r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k = q_2$$
$$. \qquad\qquad (J \text{ restrictions})$$
$$.$$
$$r_{J1}\beta_1 + r_{J2}\beta_2 + \cdots + r_{Jk}\beta_k = q_J$$

Some (many?) of the $r_{ij}'s$ may be zero.

- Combine these J restrictions:

$$\underset{(J\times k)(k\times 1)}{R\boldsymbol{\beta}} = \underset{(J\times 1)}{\boldsymbol{q}} \quad ; \quad R \text{ and } q \text{ are known, \& non-random}$$

Examples

1. $\beta_2 = \beta_3 = \cdots = \beta_k = 0$

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad ; \quad \boldsymbol{q} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

2. $\beta_2 + \beta_3 = 1$

$$R = [0 \quad 1 \quad 1 \quad 0 \quad \cdots \quad 0] \quad ; \quad q = 1$$

3. $\beta_3 = \beta_4$; and $\beta_1 = 2\beta_2$

$$R = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & \cdots & 0 \\ 1 & -2 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \;;\; q = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Suppose that we just estimate the model by LS, and get $\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{y}$.
- It is very unlikely that $R\boldsymbol{b} = \boldsymbol{q}$!
- Denote $\boldsymbol{m} = R\boldsymbol{b} - \boldsymbol{q}$.
- Clearly, $\boldsymbol{m}$ is a $(J \times 1)$ ***random vector***.
- Let's consider the sampling distribution of $\boldsymbol{m}$:

$\boldsymbol{m} = R\boldsymbol{b} - \boldsymbol{q}$; it is a *linear* function of $\boldsymbol{b}$.

If the errors in the model are Normal, then $\boldsymbol{b}$ is Normally distributed, & hence $\boldsymbol{m}$ is Normally distributed.

$E[\boldsymbol{m}] = RE[\boldsymbol{b}] - \boldsymbol{q} = R\boldsymbol{\beta} - \boldsymbol{q}$ (What assumptions used?)

So, $E[\boldsymbol{m}] = \boldsymbol{0}$; iff $R\boldsymbol{\beta} = \boldsymbol{q}$

Also, $V[\boldsymbol{m}] = V[R\boldsymbol{b} - \boldsymbol{q}] = V[R\boldsymbol{b}] = RV[\boldsymbol{b}]R'$

$$= R\sigma^2(X'X)^{-1}R' = \sigma^2 R(X'X)^{-1}R'$$

(What assumptions used?)

So, $\boldsymbol{m} \sim N[\boldsymbol{0}\,, \sigma^2 R(X'X)^{-1}R']$.

Let's see how we can use this information to ***test*** if $R\boldsymbol{\beta} = \boldsymbol{q}$. (Intuition?)

Definition: The ***Wald Test Statistic*** for testing $H_0: R\boldsymbol{\beta} = \boldsymbol{q}$ *vs.* $H_A: R\boldsymbol{\beta} \neq \boldsymbol{q}$ is: $W = \boldsymbol{m}'[V(\boldsymbol{m})]^{-1}\boldsymbol{m}$.

So, *if H_0 is true*:

$$W = (R\boldsymbol{b} - \boldsymbol{q})'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})$$

$$= (R\boldsymbol{b} - \boldsymbol{q})'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})/\sigma^2 .$$

Because $\boldsymbol{m} \sim N[\boldsymbol{0}\,, \sigma^2 R(X'X)^{-1}R']$, then *if H_0 is true*:

$W \sim \chi^2_{(J)}$; provided that σ^2 *is known*.

Notice that:

- This result is valid only *asymptotically* if σ^2 is unobservable, and we replace it with *any consistent estimator*.
- We would reject H_0 if $W > critical\ value$. (*i.e.*, when $\boldsymbol{m} = R\boldsymbol{b} - \boldsymbol{q}$ is sufficiently "large".)

The F-statistic

In the Wald statistic formula, we will replace the unknown σ^2 with s^2, and divide by the number of restrictions, J. Provided that $\varepsilon \sim N$, the F-statistic will follow an F-distribution with J and $n - k$ degrees of freedom.

$$F = \frac{(R\boldsymbol{b}-\boldsymbol{q})'\left[R(X'X)^{-1}R'\right]^{-1}(R\boldsymbol{b}-\boldsymbol{q})/J}{s^2} \sim F_{(J,(n-k))}$$

A More Intuitive Formulation of the F-Statistic

Let R_U^2 be the R^2 from the full, unrestricted model under the alternative hypothesis. Let R_R^2 be the R^2 from the model obtained by imposing the restrictions in the null hypothesis. Then, the F-statistic may be written as:

$$F = \frac{(R_U^2 - R_R^2)/J}{(1 - R_U^2)/(n-k)}$$

Why do we use *this particular test* for linear restrictions?

This *F*-test is **Uniformly Most Powerful**.

Another point to note –

$$\left(t_{(v)}\right)^2 = F_{(1,v)}$$

Restricted Least Squares Estimation:

If we test the validity of certain linear restrictions on the elements of $\boldsymbol{\beta}$, and we can't reject them, how might we incorporate the restrictions (*information*) into the estimator?

Definition: The "Restricted Least Squares" (RLS) estimator of $\boldsymbol{\beta}$, in the model, $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, is the vector, $\boldsymbol{b}_*$, which minimizes the sum of the squared residuals, subject to the constraint(s) $R\boldsymbol{b}_* = \boldsymbol{q}$.

The formula for the restricted least squares estimator is:

$$\boxed{\boldsymbol{b}_* = \boldsymbol{b} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})}$$

- RLS = LS + "Adjustment Factor".
- What if R$\boldsymbol{b}$ = $\boldsymbol{q}$?
- What are the properties of this RLS estimator of $\boldsymbol{\beta}$?

Theorem: The RLS estimator of $\boldsymbol{\beta}$ is *Unbiased* if $R\boldsymbol{\beta} = \boldsymbol{q}$ is TRUE.

Otherwise, the RLS estimator is *Biased*.

Proof:

$$E(\boldsymbol{b}_*) = E(\boldsymbol{b}) - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(\boldsymbol{R}E(\boldsymbol{b}) - \boldsymbol{q})$$
$$= \boldsymbol{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{\beta} - \boldsymbol{q}) \; .$$

So, if $R\boldsymbol{\beta} = \boldsymbol{q}$, then $E(\boldsymbol{b}_*) = \boldsymbol{\beta}$.

Theorem: The covariance matrix of the RLS estimator of $\boldsymbol{\beta}$ is

$$V(\boldsymbol{b}_*) = \sigma^2(X'X)^{-1}\{I - R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}\}$$

Theorem: The matrix, $V(\boldsymbol{b}) - V(\boldsymbol{b}_*)$, is *at least* positive semi-definite.

- This tells us that the variability of the RLS estimator is no more than that of the LS estimator, ***whether or not the restrictions are true***.
- Generally, the RLS estimator will be "more precise" than the LS estimator.

- So, ***if the restrictions are true***, the RLS estimator, $\boldsymbol{b}_*$, is more efficient than the LS estimator, $\boldsymbol{b}$, of the coefficient vector, $\boldsymbol{\beta}$.

Also note the following:

- ***If the restrictions are false***, and we consider MSE($\boldsymbol{b}$) and MSE($\boldsymbol{b}_*$), then the relative efficiency can go either way.
- ***If the restrictions are false***, not only is $\boldsymbol{b}_*$ biased, it's also ***inconsistent***.

In practice:

- Estimate the unrestricted model, using LS.
- Test $H_0: R\boldsymbol{\beta} = \boldsymbol{q}$ *vs.* $H_A: R\boldsymbol{\beta} \neq \boldsymbol{q}$.
- If the null hypothesis can't be rejected, re-estimate the model with RLS.

Otherwise, retain the LS estimates.