Topic 3: Inference and Prediction

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 ; $rank(X) = k$)

To begin with, 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:

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

- We'll assume that $rank(R) = J \ (< k)$.

- No conflicting or redundant restrictions.
- What if $J = k$?

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. $\beta_2 + \beta_3 = 1$

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

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

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

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

<u>Notice that</u>:

- 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 Wald test is a *very general* testing procedure – other testing problems.
- Wald test statistic always constructed using an estimator that ***ignores*** the restrictions being tested.
- As we'll see, *for this particular* testing problem, we can modify the Wald test slightly, and obtain a test that is **exact in finite samples**, and has excellent power properties.

What is the F-statistic?

To derive this test statistic, we need a ***preliminary result***.

Definition:

Let $x_1 \sim \chi^2_{(v_1)}$ and $x_2 \sim \chi^2_{(v_2)}$ ***and independent***

Then

$$F = \frac{\left[\frac{x_1}{v_1}\right]}{\left[\frac{x_2}{v_2}\right]} \sim F_{(v_1, v_2)} \quad ; \quad \text{Snedecor's F-Distribution}$$

Note:

- $\left(t_{(v)}\right)^2 = F_{(1,v)}$; *Why does this make sense?*
- $v_1 F_{(v_1,v_2)} \xrightarrow{d} \chi^2_{(v_1)}$; *Explanation?*

Let's proceed to our main result, which involves the statistic, $F = \left(\frac{W}{J}\right)\left(\frac{\sigma^2}{s^2}\right)$.

Theorem:

$F = \left(\frac{W}{J}\right)\left(\frac{\sigma^2}{s^2}\right) \sim F_{(J,\ (n-k))}$, if the Null Hypothesis $H_0: R\boldsymbol{\beta} = \boldsymbol{q}$ is true.

Proof:

$$F = \frac{(R\boldsymbol{b} - \boldsymbol{q})'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})}{\sigma^2}\left(\frac{1}{J}\right)\left(\frac{\sigma^2}{s^2}\right)$$

$$= \frac{(R\boldsymbol{b} - \boldsymbol{q})'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})\ /J}{\left[\frac{(n-k)s^2}{\sigma^2}\right]/(n-k)} = \left(\frac{N}{D}\right)$$

where $D = \left[\frac{(n-k)s^2}{\sigma^2}\right]/(n-k) = \chi^2_{(n-k)}/(n-k)$.

Consider the numerator:

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

Suppose that H_0 is TRUE, so that $= \boldsymbol{q}$, and then

$$(R\boldsymbol{b} - \boldsymbol{q}) = (R\boldsymbol{b} - R\boldsymbol{\beta}) = \boldsymbol{R}(\boldsymbol{b} - \boldsymbol{\beta}) \ .$$

Now, recall that

$$\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{y} = (X'X)^{-1}X'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon} \ .$$

So, $\quad R(\boldsymbol{b} - \boldsymbol{\beta}) = R(X'X)^{-1}X'\boldsymbol{\varepsilon}$,

and $\quad N = [R(X'X)^{-1}X'\boldsymbol{\varepsilon}]'[\sigma^2 R(X'X)^{-1}R']^{-1}[R(X'X)^{-1}X'\boldsymbol{\varepsilon}]/J$

$$= \left({}^1/_J\right)({}^{\boldsymbol{\varepsilon}}/_{\sigma})'[Q]({}^{\boldsymbol{\varepsilon}}/_{\sigma}) \ ,$$

where $\quad Q = X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$,

and $\quad ({}^{\boldsymbol{\varepsilon}}/_{\sigma}) \sim N[\mathbf{0}\ , I_n]$.

Now, $\quad ({}^{\boldsymbol{\varepsilon}}/_{\sigma})'[Q]({}^{\boldsymbol{\varepsilon}}/_{\sigma}) \sim \chi^2_{(r)}$ if and only if Q ***is idempotent***, where

$$r = rank(Q) \ .$$

Easy to check that Q is idempotent.

So, $\ rank(Q) = tr.(Q)$

$$= tr.\{X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\}$$

$$= tr.\{(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X\}$$

$$= tr.\{R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}\}$$

$$= \{[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'\}$$

$$= tr.(I_J) = J \ .$$

So, $\ N = \left({}^1/_J\right)({}^{\boldsymbol{\varepsilon}}/_{\sigma})'[Q]({}^{\boldsymbol{\varepsilon}}/_{\sigma}) = \chi^2_{(J)}/J$.

- In the construction of F we have a ratio of 2 Chi-Square statistics, each divided by their degrees of freedom.
- Are N and D ***independent***?

- The Chi-Square statistic in *N* is: $({}^{\boldsymbol{\varepsilon}}/_{\sigma})'[Q]({}^{\boldsymbol{\varepsilon}}/_{\sigma})$.
- The Chi-Square statistic in *D* is: $\left[\frac{(n-k)s^2}{\sigma^2}\right]$ (*see bottom of slide 13*)

Re-write this:

$$\left[\frac{(n-k)s^2}{\sigma^2}\right] = \frac{(n-k)}{\sigma^2}\left({}^{\boldsymbol{e}'\boldsymbol{e}}/_{(n-k)}\right) = \left({}^{\boldsymbol{e}'\boldsymbol{e}}/_{\sigma^2}\right)$$

$$= ({}^{M\boldsymbol{\varepsilon}}/_{\sigma})'({}^{M\boldsymbol{\varepsilon}}/_{\sigma}) = ({}^{\boldsymbol{\varepsilon}}/_{\sigma})'M({}^{\boldsymbol{\varepsilon}}/_{\sigma}) .$$

So, we have

$$({}^{\boldsymbol{\varepsilon}}/_{\sigma})'[Q]({}^{\boldsymbol{\varepsilon}}/_{\sigma}) \quad \text{and} \quad ({}^{\boldsymbol{\varepsilon}}/_{\sigma})'M({}^{\boldsymbol{\varepsilon}}/_{\sigma}) .$$

These two statistics are ***independent*** if and only if $MQ = 0$.

$MQ = [I - X(X'X)^{-1}X']\, X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$

$= Q - X(X'X)^{-1}X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$

$= Q - X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$

$= Q - Q = 0$.

So, if H_0 is TRUE, our statistic, *F* is the ratio of 2 *independent* Chi-Square variates, each divided by their degrees of feeedom.

This implies that, if H_0 is TRUE,

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

What assumptions have been used ? *What if H_0 is FALSE* ?

Implementing the test –

- Calculate *F* .

- Reject $H_0: R\boldsymbol{\beta} = \boldsymbol{q}$ in favour of $H_A: R\boldsymbol{\beta} \neq \boldsymbol{q}$ if $> c_\alpha$.

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

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

Another point to note –

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

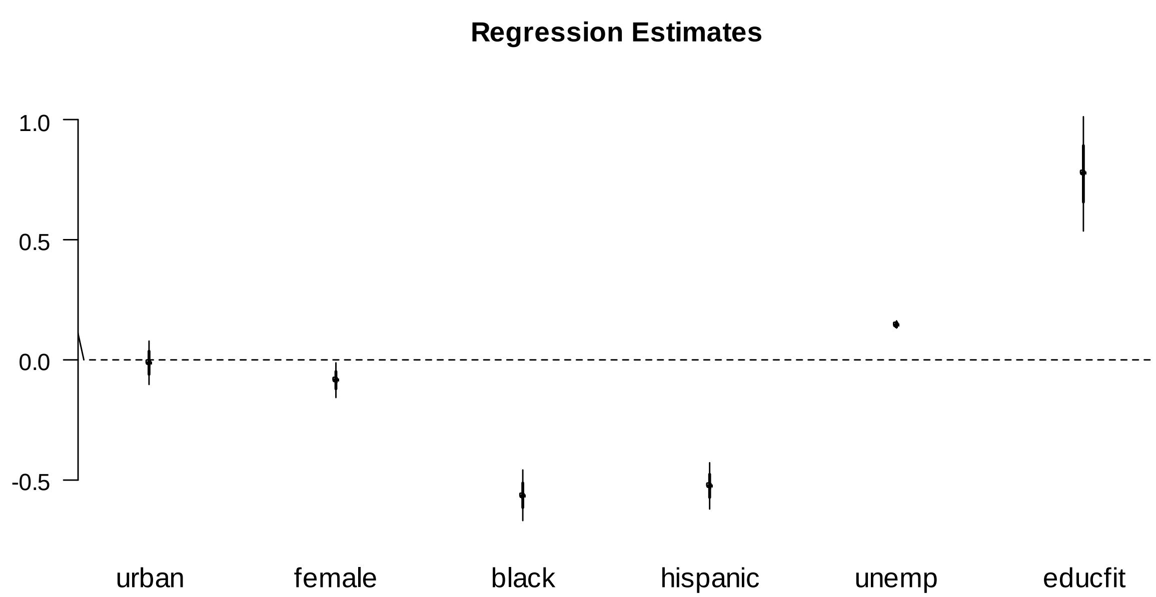
Consider $t_{(n-k)} = (b_i - \beta_i)/\left(s.e.(b_i)\right)$

Then, $\left(t_{(n-k)}\right)^2 \sim F_{(1,(n-k))}$; t-test is **UMP** against **1-sided alternatives**

Example

Let's return to the Card (1993) data, used as an example of I.V.

Recall the results of the IV estimation:




```
resiv = lm(wage ~ urban + gender + ethnicity + unemp + educfit)
summary(resiv)
```

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.053604   1.675314  -1.226   0.2203
urbanyes         -0.013588   0.046403  -0.293   0.7697
genderfemale     -0.086700   0.036909  -2.349   0.0189 *
ethnicityafam    -0.566524   0.051686 -10.961  < 2e-16 ***
ethnicityhispanic -0.529088  0.048429 -10.925  < 2e-16 ***
unemp             0.145806   0.006969  20.922  < 2e-16 ***
educfit           0.774340   0.120372   6.433 1.38e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$(n - k) = (4739 - 7) = 4732$

```
Residual standard error: 1.263 on 4732 degrees of freedom
Multiple R-squared:  0.1175,    Adjusted R-squared:  0.1163
F-statistic:   105 on 6 and 4732 DF,  p-value: < 2.2e-16
```

Let's test the hypothesis that *urban* and *gender* are jointly insignificant.

H_0: $\beta_2 = \beta_3 = 0$ vs. H_A: *At least one of these coeffs.* $\neq 0$. ($J = 2$)

Let's see R-code for calculating the F-stat from the formula:

$$F = \frac{(R\boldsymbol{b} - \boldsymbol{q})'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})/J}{s^2} = (R\boldsymbol{b} - \boldsymbol{q})'[Rs^2(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})/J$$

```
R = matrix(c(0,0,1,0,0,1,0,0,0,0,0,0,0,0),2,7)
> R
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    1    0    0    0    0    0
[2,]    0    0    1    0    0    0    0
```

```
b = matrix(resiv$coef,7,1)
> b
             [,1]
[1,] -2.05360353
[2,] -0.01358775
[3,] -0.08670020
[4,] -0.56652448
[5,] -0.52908814
[6,]  0.14580613
[7,]  0.77433967
```

```
q = matrix(c(0,0),2,1)
> q
     [,1]
[1,]    0
[2,]    0
```

```
m = R%*%b - q
> m
             [,1]
[1,] -0.01358775
[2,] -0.08670020
```

invert $s^2(X'X)^{-1}$

```
F = t(m) %*% solve(R %*% vcov(resiv) %*% t(R)) %*% m
```

transpose *multiply*

```
> F
          [,1]
[1,] 5.583774
```

Is this F-stat “large”?

```
> 1 - pf(F,2,4732)
            [,1]
[1,] 0.003783159
```

Should we be using the F-test?

```
Wald = 2*F
> 1 - pchisq(Wald,2)
            [,1]
[1,] 0.003758353
```

Why are the p-values from the Wald and F-test so similar?

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) $R\boldsymbol{b}_* = \boldsymbol{q}$.

- Let's obtain the expression for this new estimator, and derive its sampling distribution.
- Set up the Lagrangian: $\mathcal{L} = (\boldsymbol{y} - X\boldsymbol{b}_*)'(\boldsymbol{y} - X\boldsymbol{b}_*) + 2\boldsymbol{\lambda}'(R\boldsymbol{b}_* - \boldsymbol{q})$
- Set $(\partial\mathcal{L}/\partial\boldsymbol{b}_*) = \mathbf{0}$; $(\partial\mathcal{L}/\partial\boldsymbol{\lambda}) = \mathbf{0}$, and solve ……….

$$\mathcal{L} = \boldsymbol{y}'\boldsymbol{y} + \boldsymbol{b}_*'X'X\boldsymbol{b}_* - 2\boldsymbol{y}'X\boldsymbol{b}_* + 2\boldsymbol{\lambda}'(R\boldsymbol{b}_* - \boldsymbol{q})$$

$$(\partial\mathcal{L}/\partial\boldsymbol{b}_*) = 2X'X\boldsymbol{b}_* - 2X'\boldsymbol{y} + 2R'\boldsymbol{\lambda} = \mathbf{0} \qquad \textbf{[1]}$$

$$(\partial\mathcal{L}/\partial\boldsymbol{\lambda}) = 2(R\boldsymbol{b}_* - \boldsymbol{q}) = \mathbf{0} \qquad \textbf{[2]}$$

From **[1]**:

$$R'\boldsymbol{\lambda} = X'(\boldsymbol{y} - X\boldsymbol{b}_*)$$

So, $R(X'X)^{-1}R'\boldsymbol{\lambda} = R(X'X)^{-1}X'(\boldsymbol{y} - X\boldsymbol{b}_*)$

or, $\boldsymbol{\lambda} = [R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'(\boldsymbol{y} - X\boldsymbol{b}_*)$ **[3]**

Inserting **[3]** into **[1]**, and dividing by "2":

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

So, $(X'X)\boldsymbol{b}_* = X'\boldsymbol{y} - R'[R(X'X)^{-1}R']^{-1}R(\boldsymbol{b} - \boldsymbol{b}_*)$

or,

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

or, using **[2]**:

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

- RLS = LS + "Adjustment Factor".
- What if R$\boldsymbol{b}$ = $\boldsymbol{q}$?
- Interpretation of this?
- 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

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

Proof:

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

$$= \{I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\}\boldsymbol{b} + \boldsymbol{\alpha}$$

where

$$\boldsymbol{\alpha} = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\boldsymbol{q} \qquad \textit{(non-random)}$$

So, $V(\boldsymbol{b}_*) = AV(\boldsymbol{b})A'$, where $A = \{I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\}$.

That is, $V(\boldsymbol{b}_*) = AV(\boldsymbol{b})A' = \sigma^2 A(X'X)^{-1}A'$ (*assumptions*?)

Now let's look at the matrix, $A(X'X)^{-1}A'$.

$$A(X'X)A' = \{I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\}\,(X'X)^{-1}$$

$$\times \{I - R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}\}$$

$$= (X'X)^{-1} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

$$-2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

$$= (X'X)^{-1}\{I - R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}\} .$$

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

(*What assumptions have we used to get this result*?)

We can use this result immediately to establish the following……

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

Proof:

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

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

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

So, $V(\boldsymbol{b}) - V(\boldsymbol{b}_*) = \sigma^2\Delta$, say

where $\Delta = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$.

This matrix is **square**, **symmetric**, and of **full rank**. So, Δ is at least **p.s.d.**.

- 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.
- When will the RLS and LS estimators have the same variability?
- In addition, we know that the RLS estimator is unbiased ***if the restrictions are true***.
- 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***.

So, it's a good thing that that we know how to construct the UMP test for the validity of the restrictions on the elements of $\boldsymbol{\beta}$!

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.

Example: **Cobb-Douglas Production Function**[1]

```
>
cobbdata=read.csv("http://home.cc.umanitoba.ca/~godwinrt/7010/co
bb.csv")
> attach(cobbdata)
> res = lm(log(y) ~ log(k) + log(l))
> summary(res)

Call:
lm(formula = log(y) ~ log(k) + log(l))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.8444     0.2336   7.896 7.33e-08 ***
log(k)       0.2454     0.1069   2.297   0.0315 *
log(l)       0.8052     0.1263   6.373 2.06e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2357 on 22 degrees of freedom
Multiple R-squared: 0.9731,     Adjusted R-squared: 0.9706
F-statistic: 397.5 on 2 and 22 DF,  p-value: < 2.2e-16
```

What's this?

Let's get the SSE from this regression, for later use:

```
> sum(res$residuals^2)
[1] 1.22226
```

SSE = 1.22226

Test the hypothesis of constant returns to scale:

$$H_0: \beta_2 + \beta_3 = 1 \quad \text{vs.} \quad H_A: \beta_2 + \beta_3 \neq 1$$

```
> R = matrix(c(0,1,1),1,3)
> R
     [,1] [,2] [,3]
[1,]    0    1    1
```

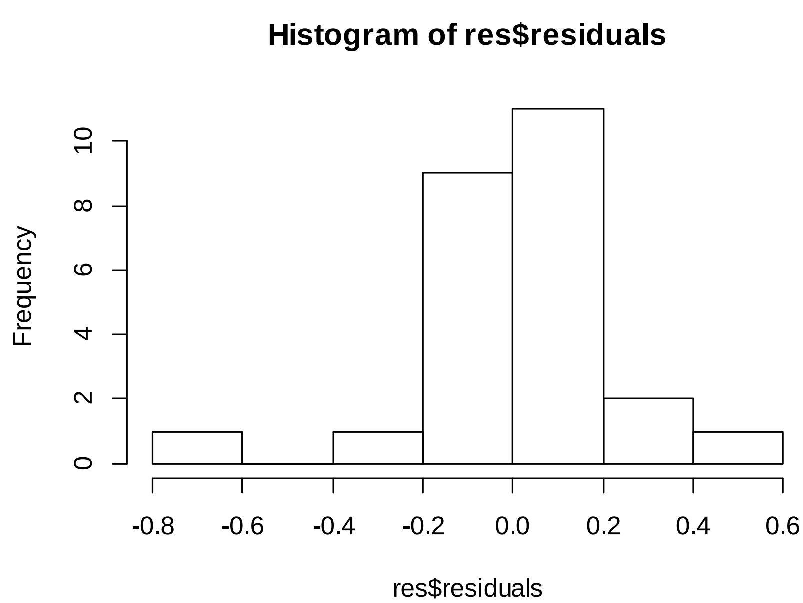
[1] The data are from table F7.2, Greene, 2012

```
> b = matrix(res$coef,3,1)
> b
          [,1]
[1,] 1.8444157
[2,] 0.2454281
[3,] 0.8051830

> q = 1

> m = R%*%b - q
> m
           [,1]
[1,] 0.05061103

> F = t(m) %*% solve(R %*% vcov(res) %*% t(R)) %*% m
> F
         [,1]
[1,] 1.540692

> 1 - pf(F,1,22)
          [,1]
[1,] 0.2275873
```

Cannot reject at 10% sig. level

Are the residuals normally distributed?

```
> hist(res$residuals)
```



```
> library(tseries)
> jarque.bera.test(res$residuals)
        Jarque Bera Test

data:  res$residuals
X-squared = 5.5339, df = 2, p-value = 0.06285
```

Might want to use Wald test instead!

F-test "supported" the validity of the restriction on the coefficients, so now *impose this restriction* of CRTS. Use RLS:

$$\log(Q/L) = \beta_1 + \beta_2 \log(K/L) + \varepsilon$$

```
> rlsres = lm(log(y/l) ~ log(k/l))
> summary(rlsres)
Call:
lm(formula = log(y/l) ~ log(k/l))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.0950     0.1189  17.615 7.55e-15 ***
log(k/l)      0.2893     0.1020   2.835  0.00939 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2385 on 23 degrees of freedom
Multiple R-squared: 0.2589,     Adjusted R-squared: 0.2267
F-statistic: 8.036 on 1 and 23 DF,  p-value: 0.009387

> sum(rlsres$residuals^2)
[1] 1.307857
```

SSE = 1.307857

Form the LS and RLS results for this particular application, note that

$$\boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - X\boldsymbol{b})'(\boldsymbol{y} - X\boldsymbol{b}) = 1.22226$$

$$\boldsymbol{e}_*'\boldsymbol{e}_* = (\boldsymbol{y} - X\boldsymbol{b}_*)'(\boldsymbol{y} - X\boldsymbol{b}_*) = 1.307857$$

So, $\boldsymbol{e}_*'\boldsymbol{e}_* > \boldsymbol{e}'\boldsymbol{e}$.

- In fact this inequality will *always hold* .
- What's the intuition behind this?

Note that:

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

$$= \boldsymbol{e} + X(X'X)R'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b} - \boldsymbol{q})$$

Now, recall that $X'\boldsymbol{e} = \boldsymbol{0}$.

So,

$$\boldsymbol{e}_*'\boldsymbol{e}_* = \boldsymbol{e}'\boldsymbol{e} + (R\boldsymbol{b} - \boldsymbol{q})'A(R\boldsymbol{b} - \boldsymbol{q}) \ ,$$

where:

$$A = [R(X'X)^{-1}R']^{-1}R(X'X)^{-1}(X'X)(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}$$

$= [R(X'X)^{-1}R']^{-1}$; this matrix has **full rank**, and is **p.d.s.**

So, $\boldsymbol{e}_*'\boldsymbol{e}_* > \boldsymbol{e}'\boldsymbol{e}$, because $(R\boldsymbol{b} - \boldsymbol{q})'A(R\boldsymbol{b} - \boldsymbol{q}) > 0$.

This last result also gives us an alternative (convenient) way of writing the formula for the F-statistic:

$$(\boldsymbol{e}_*'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}) = (R\boldsymbol{b} - \boldsymbol{q})'A(R\boldsymbol{b} - \boldsymbol{q})$$

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

Recall that:

$$F = \frac{(R\boldsymbol{b}-\boldsymbol{q})'[R(X'X)^{-1}R']^{-1}(R\boldsymbol{b}-\boldsymbol{q})/J}{s^2}$$

So, clearly,

$$F = \frac{(\boldsymbol{e}_*'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e})/J}{s^2} = \frac{(\boldsymbol{e}_*'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e})/J}{\boldsymbol{e}'\boldsymbol{e}/(n-k)}$$

For the last example:

$J = 1$; $(n - k) = (25 - 3) = 22$

$(\boldsymbol{e}_*'\boldsymbol{e}_*) = 1.307857$; $(\boldsymbol{e}'\boldsymbol{e}) = 1.22226$

So, $F = \frac{(1.307857 - 1.22226)/1}{1.22226/22} = 1.54070$ ✓

In Retrospect

- Now we can see why $R^2 \uparrow$ when we add any regressor to our model (and $R^2 \downarrow$ when we delete any regressor).
- Deleting a regressor is equivalent to imposing a zero restriction on one of the coefficients.
- The residual sum of squares $\uparrow$ and so $R^2 \downarrow$.

Exercise: use the R^2 from the unrestricted and restricted model to calculate F.

Estimating the Error Variance

We have considered the RLS estimator of $\boldsymbol{\beta}$. What about the corresponding estimator of the variance of the error term, σ^2 ?

Theorem:

Let $\boldsymbol{b}_*$ be the RLS estimator of $\boldsymbol{\beta}$ in the model,

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

and let the corresponding residual vector be $\boldsymbol{e}_* = (\boldsymbol{y} - X\boldsymbol{b}_*)$. Then the following estimator of σ^2 is *unbiased*, if the restrictions, $R\boldsymbol{\beta} = \boldsymbol{q}$, are satisfied: $s_*^2 = (\boldsymbol{e}_*'\boldsymbol{e}_*)/(n - k + J)$.

See if you can prove this result!