Econometrics I - Multiple Hypothesis Testing

Ryan T. Godwin

University of Manitoba

So far, we have seen the z and t test, and a few tests to do with IV. In this chapter, we will take a general approach to testing a set of restrictions, and consider the consequences of imposing such restrictions on an estimated model. Some examples of **multiple restrictions** that we might want to test:

1. $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$
   $$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \quad \text{vs.} \quad H_A : \beta_1 \neq 0, \beta_2 \neq 0, \ldots, \beta_k \neq 0$$

2. $$\log(Y) = \beta_1 + \beta_2 \log(K) + \beta_3 \log(L) + \epsilon$$
   $$H_0 : \beta_2 + \beta_3 = 1 \quad \text{vs.} \quad H_A : \beta_2 + \beta_3 \neq 1$$

3. $$\log(q) = \beta_1 + \beta_2 \log(p) + \beta_3 \log(y) + \epsilon$$
   $$H_0 : \beta_2 + \beta_3 = 0 \quad \text{vs.} \quad H_A : \beta_2 + \beta_3 \neq 0$$

If the null hypothesis is true, then this implies a *restricted model*. If we can obtain one model from another by imposing restrictions on the parameters of the first model, we say that the two models are "nested".

We'll be concerned with (several) possible restrictions on β, in the usual model:

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

and let's return to our simplifying assumption that X is non-random. Let's focus on J *linear restrictions*:

$$\begin{gathered} r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k = q_1 \\ r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k = q_2 \\ \vdots \\ r_{J1}\beta_1 + r_{J2}\beta_2 + \cdots + r_{Jk}\beta_k = q_J \end{gathered}$$

where many of the r_{jk}'s will likely be zero. We can combine these J restrictions:

$$R\boldsymbol{\beta} = \boldsymbol{q}$$

where R and $\boldsymbol{q}$ are known and non-random. We'll assume that $\text{rank}(R) = J < k$, so that there are no conflicting or redundant restrictions.
Question: What if $J = k$?

Exercise

Take a set of linear restrictions and write them in terms of R and $\boldsymbol{q}$.

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

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

2. $\beta_2 + \beta_3 = 1$.

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

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

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

Suppose that we 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 the *difference* between what is estimated, and what is hypothesized as:

$$\boldsymbol{m} = R\boldsymbol{b} - \boldsymbol{q}$$

$\boldsymbol{m}$ is a $(J \times 1)$ *random* vector. Let's consider the sampling distribution of $\boldsymbol{m}$. It is a linear function of $\boldsymbol{b}$. If the errors in the model are Normal, then $\boldsymbol{b}$ is Normally distributed, and hence $\boldsymbol{m}$ is Normally distributed as well. Now, to the expected value:

$$E[\boldsymbol{m}] = RE[\boldsymbol{b}] - \boldsymbol{q} = R\boldsymbol{\beta} - \boldsymbol{q}$$

So,

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

In addition, the covariance matrix of $\boldsymbol{m}$ is:

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

Question: What assumptions were used to derive the expected value and variance of $\boldsymbol{m}$?
So, the full sampling distribution of $\boldsymbol{m}$ is:

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

Let's see how we can use this sampling distribution to test if $R\boldsymbol{\beta} = \boldsymbol{q}$.

Wald test

Wald Test Statistic. 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:

$$\begin{aligned} W &= (R\boldsymbol{b} - \boldsymbol{q})' \left[\sigma^2 R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q}) \\ &= (R\boldsymbol{b} - \boldsymbol{q})' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q})/\sigma^2 \end{aligned}$$

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

$$W \sim \chi^2_{(J)}$$

provided that σ^2 is *known*.

Note 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 Wald test is a very general testing procedure and is used in other testing problems.
- The Wald test statistic is always constructed using an estimator that *ignores* the restrictions being tested.

As we'll see in the next section, 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.

F-test statistic and its distribution

F-distribution. Let $x_1 \sim \chi^2_{(v_1)}$, $x_2 \sim \chi^2_{(v_2)}$, and x_1 and x_2 be independent. χ^2 denotes the chi-square distribution, and v_1 and v_2 denotes degrees of freedom. Then:

$$F = \frac{\left[\frac{x_1}{v_1}\right]}{\left[\frac{x_2}{v_2}\right]} \sim F_{(v_1, v_2)}$$

$F_{(v_1, v_2)}$ denotes Snedecor's F-distribution, with degrees of freedom v_1 and v_2.

The issue with the Wald test, is that when we replace σ^2 with s^2 (For example), the Chi-square distribution only *approximately* describes the test statistic. The F-distribution tells us the exact distribution of the test statistic, which depends on the sample size n. As n grows to infinity however, the Wald test and F-test become identical.

The F-statistic can be derived from the Wald statistic by replacing σ^2 with s^2, and dividing by the number of restrictions J in the null hypothesis:

$$F = \left(\frac{W}{J}\right)\left(\frac{\sigma^2}{s^2}\right)$$

So, if H_0 is true, the statistic F is the ratio of two independent Chi-square variables, each divided by their degrees of freedom. This implies that, if H_0 is true:

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

In the proof of the distribution of the F-test statistic, it is important to note that this result relies on A.6, the Normality of the error term. If the errors are not Normal, the F-test statistic does not follow the F distribution.

The F-test is used for testing a set of linear restrictions because it is *uniformly most powerful.*

Finally, note that:

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

Implementing the F-test

One way of implementing the F-test is to estimate the population model by LS and then calculate the F statistic according to the formula:

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

and then reject the null hypothesis if the p-value from $F_{J,(n-k)}$ is less than the significance level. We will see a more convenient and intuitive way to perform an F-test, but first, let us consider a simple example.

Simple F-test in a Cobb-Douglas model

Suppose that we want to estimate the simple Cobb-Douglas model:

$$\log Y = \beta_1 + \beta_2 \log K + \beta_3 \log L + \epsilon$$

Data is from table F7.2, Greene, 2012:

```r
cobbdata <- read.csv("http://home.cc.umanitoba.ca/~godwinrt/7010/cobb.csv")
mod1 <- lm(log(y) ~ log(k) + log(l), data = cobbdata)
summary(mod1)
```

```
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 is the F-statistic of 397.5 for?

Let's get the RSS (residual sum of squares, $\boldsymbol{e}'\boldsymbol{e}$ for later use:

```
sum(mod1$residuals ^ 2)
```

```
[1] 1.22226
```

Now, test the hypothesis of *constant returns to scale*:

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

The R and $\boldsymbol{q}$ associated with this null hypothesis are:

$$R = \left[\begin{array}{ccc} 0 & 1 & 1 \end{array}\right] \quad ; \quad q = 1$$

We can calculate the F-test statistic from equation 1 in R using (this is *not* the way you will do it in practice):

```r
R <- matrix(c(0, 1, 1), 1, 3)
b <- matrix(mod1$coef, 3, 1)
q <- 1
m <- R %*% b - q
Fstat <- t(m) %*% solve(R %*% vcov(mod1) %*% t(R)) %*% m
Fstat
```

```
         [,1]
[1,] 1.540692
```

and then calculate the p-value using:

```r
1 - pf(Fstat, 1, 22)
```

```
          [,1]
[1,] 0.2275873
```

With a p-value of 0.23 we fail to reject the null hypothesis. However, the F-test statistic only follows the F-distribution if ϵ is Normally distributed. Test the Normality assumption using the Jarque-Bera test:

```
install.packages("tseries")
library(tseries)
jarque.bera.test(mod1$residuals)
```

```
  Jarque Bera Test

data:  mod1$residuals
X-squared = 5.5339, df = 2, p-value = 0.06285
```

With a p-value of 0.063, the Normality assumption is dicey, and we may want to use the *Wald* test instead. The Wald test statistic can be calculated from the F-stat by: $W = J \times F$. In this example, there is only one restriction ($J = 1$) so that the Wald and F-statistic coincide. The Wald statistic is asymtptoically (in other words, approximately) Chi-square distributed even if A.6 is violated (as long the LS estimators are asymptotically Normal). To get the p-value from the Chi-square distribution we can use:

```r
1 - pchisq(Fstat, 1)
```

```
          [,1]
[1,] 0.2145148
```

In this instance, the Wald and F-statistics are similar and our decision to "fail to reject" does not change. The Wald and F-tests have supported the validity of the *restrictions* specified in the null hypothesis. We could *impose* the restriction of constant returns to scale, by estimating the *restricted* model:

$$\log(Y/L) = \beta_1 + \beta_2 \log(K/L) + \epsilon$$

In the next section, we consider (i) the benefit of estimating a "restricted" model, and (ii) the cost of using a restricted model, when the restrictions (null hypothesis) are false. To do this, we develop the idea of the *restricted least squares* (RLS) estimator.

Restricted Least Squares

The RLS estimator is developed here, in order to:

- Examine the benefits (in terms of the properties of the estimator) of imposing restrictions on a model
- Examine the costs of imposing false restrictions (again in terms of properties like unbiasedness, efficiency, and consistency)

The basic idea is that the null hypothesis implies *restrictions* on the initial model under the *alternative hypothesis*. If we fail to reject H_0, then the restrictions could be imposed on the initial model simply by substituting the values for (say) β from the null. A model obtained by imposing parameter restrictions on a larger un-restricted model is often said to be "nested".

Nested models

Model B is "nested" in model A if B can be obtained by imposing parameter restrictions on A.

Many tests in econometrics can be interpreted as comparing the "fit" or "performance" of the restricted (nested model, under H_0) to that of the unrestricted model (under H_A). The F-test is no exception. We will see that we can calculate the F-statistic by estimating two models and comparing their RSS or R^2. Before we do so, we develop the RLS estimator.

Restricted least squares estimator (RLS)

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

We will obtain the expression for this new estimator, and derive its sampling distribution. Set up the Lagrangian:

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

For the first order conditions, set $(\partial\mathcal{L}/\partial\boldsymbol{b}_*) = \mathbf{0} \quad ; \quad (\partial\mathcal{L}/\partial\lambda) = \mathbf{0}$, and solve:

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

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

$$(\partial\mathcal{L}/\partial\boldsymbol{\lambda}) = 2 (R\boldsymbol{b}_* - \boldsymbol{q}) = \mathbf{0} \tag{3}$$

From equation 2

$$\begin{aligned} R'\boldsymbol{\lambda} &= X'\left(\boldsymbol{y} - X\boldsymbol{b}_*\right) \\ R\left(X'X\right)^{-1}R'\boldsymbol{\lambda} &= R\left(X'X\right)^{-1}X'\left(\boldsymbol{y} - X\boldsymbol{b}_*\right) \\ \boldsymbol{\lambda} &= \left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}X'\left(\boldsymbol{y} - X\boldsymbol{b}_*\right) \end{aligned} \tag{4}$$

Inserting equation 4 into equation 2, and dividing by 2:

$$\begin{aligned} \left(X'X\right)\boldsymbol{b}_* &= X'\boldsymbol{y} - R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}X'\left(\boldsymbol{y} - X\boldsymbol{b}_*\right) \\ \left(X'X\right)\boldsymbol{b}_* &= X'\boldsymbol{y} - R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(\boldsymbol{b} - \boldsymbol{b}_*\right) \\ \boldsymbol{b}_* &= \left(X'X\right)^{-1}X'\boldsymbol{y} - \left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\boldsymbol{b} - R\boldsymbol{b}_*\right) \end{aligned} \tag{5}$$

and finally substituting equation 3 into equation 5, the RLS estimator is:

$$\boldsymbol{b}_* = \boldsymbol{b} - (X'X)^{-1} R' \left[R (X'X)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q}) \tag{6}$$

From the above formula, we can see that RLS = LS + "adjustment factor".

Questions:

What if R$\boldsymbol{b} = \boldsymbol{q}$? What is the interpretation of this?

We will use the RLS formula in order to derive the *statistical properties* of RLS estimation, thus informing us the benefits of imposing true restrictions, and costs of imposing false ones.

RLS estimator is unbiased iff H_0 is true.

The RLS estimator of $\boldsymbol{\beta}$ is unbiased iff $R\boldsymbol{\beta} = \boldsymbol{q}$ is true. Otherwise, the RLS estimator is *biased.*

Proof.

$$
\begin{aligned}
E\left(\boldsymbol{b}_*\right) &= E(\boldsymbol{b}) - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(\boldsymbol{R}E(\boldsymbol{b}) - \boldsymbol{q}\right) \\
&= \boldsymbol{\beta} - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\boldsymbol{\beta} - \boldsymbol{q}\right)
\end{aligned}
$$

So, if $R\boldsymbol{\beta} = \boldsymbol{q}$, then $E\left(\boldsymbol{b}_*\right) = \boldsymbol{\beta}$. Similarly, the LS estimator is only *consistent* if the restrictions (null hypothesis) are true.

The cost of imposing false restrictions (for example “dropping” a variable that is actually significant) is incurring bias and inconsistency in the LS estimator.

The potential cost of estimating a restrictive model is very high. The potential benefit, however, is an improvement in *efficiency*. That is, the RLS estimator has smaller variance than the LS estimator (regardless of the veracity of the restrictions). Intuitively, if restrictions are imposed on a model, than there are fewer parameters to estimate. The same amount of information (the sample size) can focus on estimating fewer parameters; this translates to smaller standard errors, narrower confidence intervals, etc. Another intuitive way to think of it is as follows. The null hypothesis contains information about the parameters. Using this information while estimating the model improves efficiency.

Below, we prove that the RLS estimators has lower variance than the LS estimators (this does not contradict the GM theorem since the RLS is estimating a *different* model from the LS estimator). The proof does not require that the *restrictions* are actually *true*.

The RLS estimator has smaller variance than the LS estimator.

Proof:
First, we derive the covariance matrix of the RLS estimator of $\boldsymbol{\beta}$.

$$\begin{aligned}
\boldsymbol{b}_* &= \boldsymbol{b} - (X'X)^{-1} R' \left[R (X'X)^{-1} R' \right]^{-1} (R\boldsymbol{b} - \boldsymbol{q}) \\
&= \left\{ I - (X'X)^{-1} R' \left[R (X'X)^{-1} R' \right]^{-1} R \right\} \boldsymbol{b} + \boldsymbol{\alpha}
\end{aligned}$$

where $\boldsymbol{\alpha}$ is non-random and $\boldsymbol{\alpha} = (X'X)^{-1} R' \left[R (X'X)^{-1} R' \right]^{-1} \boldsymbol{q}$.
So, $V(\boldsymbol{b}_*) = AV(\boldsymbol{b})A'$, where
$A = \left\{ I - (X'X)^{-1} R' \left[R (X'X)^{-1} R' \right]^{-1} R \right\}$. That is,
$V(\boldsymbol{b}_*) = AV(\boldsymbol{b})A' = \sigma^2 A (X'X)^{-1} A'$.

Now, examine $A\left(X'X\right)^{-1}A'$:

$$
\begin{aligned}
A\left(X'X\right)A' &= \left\{I - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\right\}\left(X'X\right)^{-1} \\
&\quad \times \left\{I - R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1}\right\} \\
&= \left(X'X\right)^{-1} + \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \\
&\quad \times R\left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1} \\
&\quad - 2\left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}\left\{I - R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1}\right\}
\end{aligned}
$$

So,

$$
\begin{aligned}
V\left(\boldsymbol{b}_*\right) &= \sigma^2\left(X'X\right)^{-1}\left\{I - R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}\right\} \\
&= \sigma^2\left(X'X\right)^{-1} - \sigma^2\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1} \\
&= V(\boldsymbol{b}) - \sigma^2\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}
\end{aligned}
$$

So, $V(\boldsymbol{b}) - V\left(\boldsymbol{b}_*\right) = \sigma^2\Delta$, where $\Delta = \left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}$. Δ is square, symmetric, and of full rank. So, Δ is at least positive semi-definite.

The above proof 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.

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, $\boldsymbol{\beta}$. If the restrictions are false, and we consider the MSE, then the relative efficiency can go either way.

The RLS estimator that we have discussed in this section (equation 6) is used to determine the consequences of imposing restrictions on a model, whether the restrictions are true or false. In practice, equation 6 is not used to effect RLS estimation. Rather, the restrictions are simply imposed on the model (such as dropping a regressor), and the model is re-estimated using LS. In practice, we:

- 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, substitute the restrictions into the model and re-estimate using LS.
- Otherwise, retain the initial estimates under the unrestricted model.

Testing by comparing unrestricted and restricted models

We can rewrite the F-test statistic formula in equation 1 in terms of the residuals or R^2 from the restricted and unrestricted model. This may be a more intuitive way of looking at the F-test.
Let $\boldsymbol{e}_*$ be the residuals from RLS estimation. By substituting in equation 6 for $\boldsymbol{b}_*$, we can write $\boldsymbol{e}_*$ as:

$$\begin{aligned}\boldsymbol{e}_* &= (\boldsymbol{y} - X\boldsymbol{b}_*) = \boldsymbol{y} - X\boldsymbol{b} + X\left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q}) \\ &= \boldsymbol{e} + X\left(X'X\right) R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q})\end{aligned}$$

Recalling that $X'\boldsymbol{e} = \boldsymbol{0}$:

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

where A is full rank and is positive semi-definite, and is:

$$\begin{aligned}A &= \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1}\left(X'X\right)\left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \\ &= \left[R\left(X'X\right)^{-1} R'\right]^{-1}\end{aligned}$$

Question:

$\boldsymbol{e}_*'\boldsymbol{e}_* > \boldsymbol{e}'\boldsymbol{e}$, because $(R\boldsymbol{b}-\boldsymbol{q})'A(R\boldsymbol{b}-\boldsymbol{q}) > 0$. This inequality will always hold. What's the intuition behind this?

Now, take the difference:

$$\begin{aligned}(\boldsymbol{e}_*{}'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}) &= (R\boldsymbol{b}-\boldsymbol{q})'A(R\boldsymbol{b}-\boldsymbol{q}) \\ &= (R\boldsymbol{b}-\boldsymbol{q})'\left[R\left(X'X\right)^{-1}R'\right]^{-1}(R\boldsymbol{b}-\boldsymbol{q})\end{aligned} \tag{7}$$

Recalling that:

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

we can rewrite the F-statistic formula in an alternative and more convenient form:

$$F = \frac{\left(\boldsymbol{e}_*{}'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}\right)/J}{s^2} = \frac{\left(\boldsymbol{e}_*{}'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}\right)/J}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \tag{8}$$

In retrospect, we can see further why R^2 increases when we add any regressor to the model: deleting a regressor is equivalent to imposing a zero restriction on one of the coefficients. The RSS increases and so R^2 decreases. In fact, using the definition for R^2, we can also rewrite the F-test statistic as:

$$F = \frac{\left(R^2 - R_*^2\right)/J}{\left(1 - R^2\right)/\left(n - k\right)} \tag{9}$$

where R_*^2 is from the restricted model.

Equations 8 and 9 provide an intuitive interpretation of the F-test (which apply to many other tests used in econometrics): if the restricted model (under the null hypothesis) “fits” the data much more poorly than the unrestricted model (under the alternative hypothesis), then the F-statistic will be large and the restrictions (null) will be rejected.

F-test in Cobb-Douglas again

We revisit the example from section 13 applying the version of the F-test from equation 8. To accomplish this, estimate two models: one under the null hypothesis (restricted model), and one under the alternative hypothesis (unrestricted model):

```
cobbdata <- read.csv("http://home.cc.umanitoba.ca/~godwinrt/7010/cobb.csv")
unrestricted <- lm(log(y) ~ log(k) + log(l), data = cobbdata)
restricted <- lm(log(y/l) ~ log(k/l), data = cobbdata)
```

Notice that, to implement the restriction of CRTS, the left-hand-side variable has changed. This means that equation 9, and some "canned" commands in Stata and R (for example `anova()`) for performing the F-test, will not work!

We can use the residuals from the two models, however:

```
RSS <- sum(unrestricted$residuals ^ 2)
RSSstar <- sum(restricted$residuals ^ 2)
```

and calculate the F-stat according to equation 8:

```
Fstat2 <- ((RSSstar - RSS) / 1) / (RSS / 22)
Fstat2
```

```
> Fstat2
[1] 1.540692
```

where we notice that the F-statistic is the same as previously calculated.

Testing for differences

Dummy variables are important and are used very commonly in economics research. Although a dummy variable can take many values, in economics it is almost always refers to a binary indicator variable. Typically the values that the dummy variable can take are 0 or 1, where each value corresponds to a certain quality, or to membership in a group. A common convention is to name the dummy variable so that a value of 1 indicates “yes”, and a value of 0 indicates “no”. For example, the variable `foreign` might equal 1 when a firm is foreign-owned, and 0 when domestically owned.

Dummy variables provide a convenient way to test for *differences* between groups, regions, countries, types of firms, moments in time, *treatment* status, etc.

Let D denote our dummy variable. Then, consider a model of the form:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k D_i + \epsilon_i \tag{10}$$

we could then think of testing:

$$H_0 : \beta_k = 0 \quad \text{vs.} \quad H_A : \beta_k \neq 0$$

using a t-test or z-test. Rejection of H_0 implies there is a particular type of *structural change* in the model. In particular, the dummy variable in equation 10 allows the *mean* of y_i to differ depending on the value of D_i, that is:

$$\mathrm{E}\left(y_i | D_i = 1\right) - \mathrm{E}\left(y_i | D_i = 0\right) = \beta_k$$

So, rejection of H_0 implies that D_i has a significant effect on the mean of y_i. More generally, consider a model of the form:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_{k-1}\left(D_i \times x_{i2}\right) + \beta_k D_t + \epsilon_i \tag{11}$$

where we could use an F or Wald test for:

$$H_0 : \beta_{k-1} = \beta_k = 0 \quad \text{vs.} \quad H_A : \text{Not}\, H_0$$

Rejection of H_0 implies a different type of structural change in the model: a shift in the mean and one of the marginal effects. Allowing the dummy variable to interact with the "x" variables, such as in model 11, allows for variables to have different marginal effects depending on regions, times, or treatment status. For example, in model 11, the change in y associated with a change in x_2 is:

$$\frac{\partial y_i}{\partial x_{i2}} = \beta_2 + D_i \beta_{k-1}$$

At the extreme, we could allow the dummy variable to interact fully with every variable in the model:

$$\boldsymbol{y} = X\boldsymbol{\beta}_1 + DX\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{12}$$

Testing the joint significance of every term that includes D_i ($H_0 : \boldsymbol{\beta_2} = 0$) is equivalent to the Chow test (Gregory Chow, 1960). Note the parameter estimates from model 12 can also be obtained by fitting two separate models for the two separate subsamples (as defined by the dummy variable).

Life Expectancy - F-test.

This example follows Greene (2012), example 6.10 (pg. 173). The WHO released a report in 2000 that gained much attention, and was controversial. It suggested that health care expenditure by OECD countries had more of an impact on life expectancy than spending by non-OECD countries.

The dependent variable is `DALE` - disability-adjusted life expectancy. Consider a simple model:

$$DALE = \beta_1 + \beta_2 HEXP + \beta_3 HC3 + \beta_4 HC3^2 + \epsilon \tag{13}$$

where `HEXP` is health expenditure, and `HC3` is a measure of educational attainment. Download data from Greene, choose to use only the year 1997, and delete missing values (to follow Greene):

```
health <- read.csv("http://www.stern.nyu.edu/~wgreene/Text/
    Edition7/TableF6-3.csv")
health <- subset(health, health$YEAR == 1997)
health <- na.omit(health)
```

Next create the squared term (we can't square inside `lm()`):

```r
health$HC3sq <- health$HC3^2
```

Estimate model 13, and view the results:

```r
mod1 <- lm(DALE ~ HEXP + HC3 + HC3sq, data = health)
summary(mod1)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.237039   2.439645  10.345  < 2e-16 ***
HEXP         0.006291   0.001056   5.960 1.23e-08 ***
HC3          7.930951   0.902432   8.788 9.73e-16 ***
HC3sq       -0.438988   0.077157  -5.690 4.85e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
    1

Residual standard error: 6.984 on 187 degrees of freedom
Multiple R-squared:  0.6824,  Adjusted R-squared:  0.6773
F-statistic: 133.9 on 3 and 187 DF,  p-value: < 2.2e-16
```

Now, consider a model that has a dummy variable `OECD` that *interacts* with all of the regressors:

$$\begin{aligned} DALE = &\beta_1 + \beta_2 HEXP + \beta_3 HC3 + \beta_4 HC3^2 + \beta_5 OECD \\ &+ \beta_6 (OECD \times HEXP) + \beta_7 (OECD \times HC3) \\ &+ \beta_8 (OECD \times HC3^2) + \epsilon \end{aligned} \tag{14}$$

Model 14 allows for the effects of both health expenditure and education to differ, depending on whether the country is OECD or not. Estimate this model in R:

```
mod2 <- lm(DALE ~ HEXP + HC3 + HC3sq + OECD + OECD*HEXP +
    OECD*HC3 + OECD*HC3sq,
  data = health)
summary(mod2)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 26.812047   2.651604  10.112  < 2e-16 ***
HEXP         0.009551   0.001934   4.938 1.77e-06 ***
HC3          7.043319   1.074319   6.556 5.46e-10 ***
HC3sq       -0.373804   0.096020  -3.893 0.000139 ***
OECD        15.916263  25.678970   0.620 0.536149    
HEXP:OECD   -0.006872   0.002619  -2.624 0.009435 ** 
HC3:OECD    -0.866530   6.261351  -0.138 0.890082    
HC3sq:OECD  -0.011024   0.378032  -0.029 0.976768    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.86 on 183 degrees of freedom
Multiple R-squared:  0.7002,  Adjusted R-squared:  0.6887 
F-statistic: 61.06 on 7 and 183 DF,  p-value: < 2.2e-16
```

Now, test the hypothesis that there is *no difference* between OECD and non-OECD countries (we are testing the joint significance of all interaction terms):

$$H_0 : \beta_5 = 0, \beta_6 = 0, \beta_7 = 0, \beta_8 = 0 \quad \text{vs.} \quad H_A : \text{ not } H_0$$

To calculate the F-test, we can use the R-squared from the unrestricted and restricted models:

$$F = \frac{\left(R^2 - R_*^2\right)/J}{\left(1 - R^2\right)/\left(n - k\right)} = \frac{(0.7002 - 0.6824)/4}{(1 - 0.7002)/183} = 2.716$$

While this formula for the F-test statistic highlights its dependence on the "fit" of the models under the null and alternative hypotheses, we can easily perform this test in R using the `anova()` function:

```r
anova(mod1, mod2)
```

```
Analysis of Variance Table

Model 1: DALE ~ HEXP + HC3 + HC3sq
Model 2: DALE ~ HEXP + HC3 + HC3sq + OECD + OECD * HEXP +
    OECD * HC3 +
    OECD * HC3sq
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    187 9121.8
2    183 8611.0  4    510.83 2.714 0.0314 *
```

We get a similar F-stat of 2.714, with associated p-value of 0.0314. We reject the null hypothesis at 5% significance. It would appear that OECD countries might differ from non-OECD. However, this difference is being driven by the variable `HEXP`. We would fail to reject the null that the effect of education on life expectancy is the same for both OECD and non-OECD, and could perhaps drop the interaction terms $OECD \times HC3$ and $OECD \times HC3^2$ from the model.