**Economics 522A, Spring 2007**

**Lecture Note 10: F Test and R-squared**

Readings: Ruud 10.3.4, 11.1-11.2. (Also, skim 11.3-11.5)

## Recap: Linear Restrictions on $\beta$

$R\beta$: a linear combination of elements of $\beta$.

Last time we noted:

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

and

$$\frac{(R\beta - R\hat{\beta})'\left[R(X'X)^{-1}R'\right]^{-1}(R\beta - R\hat{\beta})}{\sigma^2} \sim \chi^2_{k-m}.$$

Also recall that

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{n-k},$$

and this is independent of $\hat{\beta}$ and $X$.

Consider:

$$\frac{(R\beta - R\hat{\beta})'\left[R(X'X)^{-1}R'\right]^{-1}(R\beta - R\hat{\beta})/(k-m)}{s^2}.$$

We can write this as:

$$\frac{(R\beta - R\hat{\beta})'\left[R(X'X)^{-1}R'\right]^{-1}(R\beta - R\hat{\beta})}{\sigma^2} \cdot \frac{1}{k-m} \cdot \frac{\sigma^2}{(n-k)s^2} \cdot (n-k).$$

This is the ratio of two independent chi-squared random variables, each divided by their degrees of freedom, so it is distributed as:

$$\frac{\chi^2_{k-m}/(k-m)}{\chi^2_{n-k}/(n-k)} \sim F_{k-m,n-k}.$$

## F Test and Confidence Regions

Suppose we wish to test:

$$H_0: \quad R\beta = r,$$

vs.

$$H_1: \quad R\beta \neq r.$$

Here $r$ is a $(k-m) \times 1$ vector of constants.

Example:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So the null hypothesis is that

$$R\beta = \begin{pmatrix} \beta_1 \\ \beta_2 + \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

F Statistic

$$\hat{F} = \frac{(r - R\hat{\beta})' \left[R(X'X)^{-1}R'\right]^{-1} (r - R\hat{\beta})/(k-m)}{s^2}.$$

(Note: we could have used $R\hat{\beta} - r$ instead of $r - R\hat{\beta}$ and gotten the same thing.)

Under $H_0 : R\beta = r$,

$$\hat{F} \sim F_{k-m,n-k}.$$

For given significance level $\alpha$, let $c$ satisfy:

$$P(F_{k-m,n-k} \leq c) = 1 - \alpha.$$

The F test rejects the null hypothesis if $\hat{F} > c$.

Heuristically: if $H_0$ is true, then $r - R\hat{\beta}$ should be close to 0, and $P(reject)$ will be only $\alpha$. If $H_1$ holds, then $r - R\hat{\beta}$ will typically be different from 0, so $\hat{F}$ will have a distribution shifted to the right. Thus, $P(reject) > \alpha$.

Confidence Interval

Recall from LN4, Result 1: one way to form a confidence interval is to invert a test statistic. The set of parameter values such that we do not reject the null hypothesis at signficance level $\alpha$, is a $(1-\alpha)$ confidence region.

In this case, we look for values of $r$ such that the F test at the $\alpha$ level would not reject:

$$CI = \left\{ \gamma \in Col(R) \mid \frac{(\gamma - R\hat{\beta})' \left[R(X'X)^{-1}R'\right]^{-1} (\gamma - R\hat{\beta})/(k-m)}{s^2} \leq c \right\}.$$

Here $\gamma \in Col(R)$ means any $\gamma$ that can be written as a linear combination of the columns of $R$. In other words, any $\gamma$ of the form

$$\gamma = R\pi$$

for some $k \times 1$ vector $\pi$.

Example: connection to t test

Suppose

$$R = [1, 0, \ldots, 0], \quad r = 0.$$

So

$$R\beta = [1, 0, \ldots, 0] \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \beta_1.$$

And suppose

$$H_0: \quad \beta_1 = 0,$$

vs.

$$H_1: \quad \beta_1 \neq 0.$$

Here $(k - m) = 1$, and $R\hat{\beta} = \hat{\beta}_1$.

First, note that

$$\begin{aligned} R(X'X)^{-1}R' &= [1, 0, \ldots, 0](X'X)^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= [1, 0, \ldots, 0] \begin{bmatrix} (X'X)^{-1}_{11} \\ (X'X)^{-1}_{21} \\ \vdots \\ (X'X)^{-1}_{k1} \end{bmatrix} \\ &= (X'X)^{-1}_{11}. \end{aligned}$$

So

$$\hat{F} = \frac{(\hat{\beta}_1)^2}{s^2 (X'X)^{-1}_{11}} = \left( \frac{\hat{\beta}_1}{\sqrt{s^2 (X'X)^{-1}_{11}}} \right)^2 = \left( \frac{\hat{\beta}_1}{SE_1} \right)^2,$$

which is just the square of the standard t statistic for testing $\beta_1 = 0$.

So we can use either the t statistic, or the F statistic, and reach the same conclusions regarding this hypothesis.

**R-Squared**

When reporting results from an OLS regression, it is common to also report a number called the coefficient of determination, or $R^2$, defined as follows:

Consider regressing $y_i$ on just a constant (so the OLS coefficient amounts to the sample mean of $y_i$), vs. regressing $y_i$ on a whole vector of regressors (which includes a constant term, but also other $x$ variables).

Let $\bar{y}$ be the sample average of the $y_i$, and let $\hat{y}_i = x_i'\hat{\beta}$ be the fitted values using all the regressors.

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}.$$

This measures how much of the "total variation" in the $y_i$ can be accounted for by the fitted values, and will be a number between 0 and 1.

Note: the "R" in $R^2$ is not the same as the restriction matrix $R$. This is an unfortunate clash of notation, but usually will be clear from context.

There are other equivalent ways of calculating $R^2$. Can show that

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2},$$

where $e_i = y_i - x_i'\hat{\beta}$. Also, can show that $R^2$ equals the square of the sample correlation between $y_i$ and $\hat{y}_i$.

Note: a low $R^2$ does not mean the model is "wrong"! If $\sigma^2$ is large, then we could have that $\sum_i e_i^2$ is large relative to the total variation in the $y_i$s.

Also, adding additional regressor variables will never decrease $R^2$, because the OLS fitting algorithm always tries to minimize $\sum_i e_i^2$. So the fact that adding an additional regressor, say $x_{i,k+1}$, leads to lower $R^2$, should not necessarily be taken as evidence that $\beta_{k+1} \neq 0$.

There is a connection between the coefficient of determination and the F test of linear restrictions. Suppose we want to test the hypothesis that

$$H_0: \quad \beta_2 = 0, \beta_3 = 0, \ldots, \beta_k = 0.$$

This says that *all* coefficients except for the intercept term are zero.

The F statistic for this restriction can be shown to equal:

$$\hat{F} = \frac{(n-k)}{(k-1)} \cdot \frac{R^2}{1 - R^2}.$$

So the high values of $R^2$ lead to high values of $\hat{F}$, but if there are a large number of explanatory variables (so $k$ is large), then this reduces $\hat{F}$.