

Lecture 14

①

Testing Multiple Linear Restrictions: The F-test

See Section 4.5 in your textbook.

Consider the following unrestricted model with k independent variables:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad (1)$$

Suppose you want to test the following exclusion restrictions

$$H_0: \beta_{k-q+1} = 0, \dots, \beta_k = 0.$$

Then the restricted model is

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{k-q} x_{k-q} + u \quad (2)$$

Now collect the unrestricted sum of squared residuals, SSR_{ur} , from the first equation as

$$SSR_{ur} = \sum_1^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2$$

(2)

where $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ represent the OLS estimates obtained from applying ordinary least squares to the unrestricted equation (1). When we apply ordinary least squares to the restricted model (2) we obtain the Restricted Least Squares estimates $\tilde{\beta}_0, \tilde{\beta}_1, \dots, \tilde{\beta}_{k-q}$ and the restricted sum of squared residuals, SSR_r , computed as

$$SSR_r = \sum_1^N (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 X_{i1} - \dots - \tilde{\beta}_{k-q} X_{i,k-q})^2.$$

(Note that, in general, $\tilde{\beta}_0 \neq \hat{\beta}_0, \tilde{\beta}_1 \neq \hat{\beta}_1, \dots, \tilde{\beta}_{k-q} \neq \hat{\beta}_{k-q}$.)

Then to test the null hypothesis, we compute the following F-statistic which under H_0 follows an F-distribution with q numerator degrees of freedom and $N-k-1$ denominator degrees of freedom in repeated sampling.

(3)

$$F_{q, N-k-1} = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(N-k-1)}$$

An equivalent R^2 form of the F-statistic is (this follows from the definition of R^2 and the fact that $SST = SSR + SSE$.)

$$F_{q, N-k-1} = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(N-k-1)}$$

R_{ur}^2 is the coefficient of determination in the unrestricted model R_r^2 is the coefficient of determination in the restricted model.

If the observed F-value has a p-value $> \alpha$ ($\alpha = 0.05$ usually), we accept the null hypothesis of joint insignificance of the excluded variables.

If the observed F-value has a p-value $< \alpha$,

(4)

we reject the null hypothesis and accept the alternative hypothesis that one or more of the proposed excluded variables explains a statistically significant ~~proportion~~ proportion of the variation in the dependent variable.

A good example is presented in your textbook, pp. 143-144.

The unrestricted model is

$$\log(\hat{\text{salary}}) = 11.0 + 0.06894 \text{ years} + 0.0126 \text{ gamesyr} \\ (0.29) \quad (0.0121) \quad (0.0026)$$

$$+ 0.00098 \text{ baug} + 0.0144 \text{ hwnsyr} \\ (0.00110) \quad (0.0161)$$

$$+ 0.0108 \text{ rbi'syr} \\ (0.0072)$$

$$N = 353, \text{ SSR}_{ur} = 183.186, R_{ur}^2 = 0.6278$$

(5)

The restricted model is

$$\log(\tilde{\text{salary}}) = 11.22 + 0.0713 \text{ years} + 0.0202 \text{ gamesyr}$$

(0.11) (0.0125) (0.0013)

$$N = 353, \text{ SSR}_r = 198.311, R_r^2 = 0.5971.$$

Now suppose that we want to test the joint statistical significance of the variables bang , hrunsyr , and rbisyr , i.e.

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0.$$

The corresponding F-value is

$$F = \frac{(198.311 - 183.186) / 3}{183.186 / (353 - 5 - 1)} = 9.55$$

using the R^2 form of the F-statistic gives you the same answer, up to minute rounding error,

$$F = \frac{(0.6278 - 0.5971) / 3}{(1 - 0.6278) / (353 - 5 - 1)} = 9.55$$

6

Now is $F = 9.55$ statistically significant at, say, the $\alpha = 0.05$ level of statistical significance?

Well $F_{3,347,0.05} \doteq 2.60$. And since $9.55 > 2.60$

we at least know that the p -value associated with the observed F -value of 9.55 is less than

0.05 . Fortunately, computer programs like

SAS can easily give you the exact p -value

in that it uses numerical integration to get the probability $\Pr(F > 9.55)$ given

numerator degrees of freedom of 3 and denominator

degrees of freedom of 347.