Week 10 - F-Tests and Single Linear Combinations of Parameters

<u>F-tests: testing a group of variables</u>

Example. Suppose we have a model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + u$. Let our hypotheses be $H_0 : \beta_5 = 0, \beta_6 = 0;\ \ H_a : H_0 \text{ is false}$

- When conducting a joint hypothesis test, we say that our H_a is just "H_0 is false", as there might be too many different ways that H_0 could fail so it's not convenient to list them all.

If we test whether $\beta_5 = 0$ at a 5% significance level and then whether $\beta_6 = 0$ at a 5% significance level, the probability of *rejecting* H_0 would = $Pr[\beta_5 \neq 0 \text{ or } \beta_6 \neq 0]$.

However, *t_5 and t_6 might be correlated*.

- If t_5 and t_6 are perfectly correlated, the probability of rejecting H_0 = $Pr[\beta_5 \neq 0] = 5\%$
- If t_5 and t_6 are independent, the probability of rejecting H_0 = $1 - Pr[\beta_5 = 0] * Pr[\beta_6 = 0]$ = $1 - (0.95)(0.95) = 9.75\%$

So by using sequential t-tests, we get a greater-or-equal probability of rejecting H_0. We do not know the probability of type I error.

<u>F-tests</u>

An ***F-test*** tests whether a group of variables have no effect on the dependent variable (after all other variables are controlled for) - i.e., the ***overall significance*** of a group of variables.

An F-test compares two models:

- The ***unrestricted*** model, which is just the complete model
  - $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + u$
- The ***restricted*** model, which applies the restrictions set out in the null hypothesis
  - $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$

The F-test sees if these two models are substantially different. If there is a substantial difference, then the F-statistic will be more extreme than a critical F-value, and we will reject the null.

Assuming that errors are homoskedastic, the ***F-statistic*** is: $F_{STAT} = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k)}$, $F_{STAT} > 0$

Where:

- RSS_R is the residual sum of squares in the restricted model
- RSS_U is the residual sum of squares in the unrestricted model
- q is the number of restrictions in the null hypothesis
- k is the number of variables in the unrestricted model
    - In our example, q = 2 and k = 6

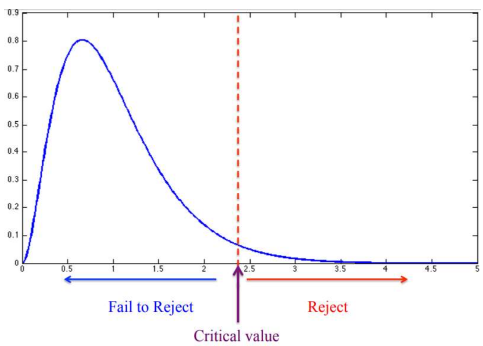
Since $TSS_R = TSS_U$, we can divide the top and bottom by TSS and get another formula for F_{STAT}:

$$F_{STAT} = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-k)}$$

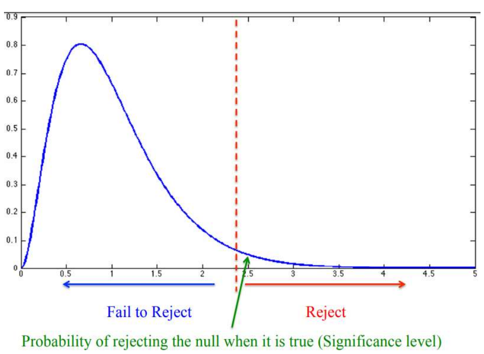
- If the restricted model is $\beta_1 + u$, we are not expecting any variation, therefore we are not explaining any variation, therefore $R_R^2=0$

Under the null, the F-statistic is approximately F(q, n-k) distributed. This approximation is exact if the data are normally distributed or if n → ∞.

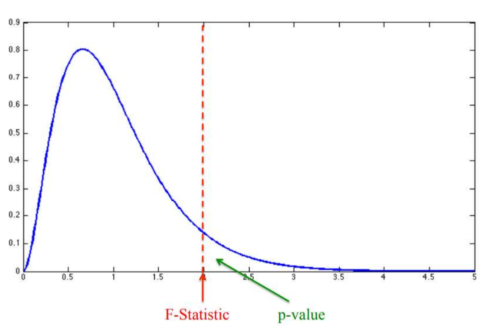
- If you use STATA, it will calculate n for you. Otherwise, F-tables will assume that n = ∞



The shape of the F distribution is always right-skewed (positive). We reject the null if F_{STAT} > critical value.

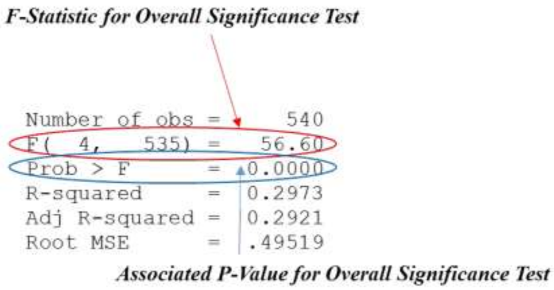


The area under the curve after the critical value is the significance level: the probability of rejecting the null when it is true. This is similar to normal t-tests.



The area under the curve after the F-statistic is the p-value. If the p-value is lower than the significance level, we reject the null.

STATA output, in the right panel.



Note that the F-test is a *two-sided test* i.e. we can only test if multiple population parameters *equal* to something.

Interestingly, if you conduct a t-test and square the t_{STAT}, you get the F_{STAT}.

Special case of F-test: whether q explanatory variables have coefficients of 0

- If H_0 is rejected we say that the restrictions in H_0 are jointly statistically significant
- If H_0 is not rejected we say that the restrictions in H_0 are jointly statistically insignificant, which justifies dropping them from the model
- It is possible that a group of variables are each individually insignificant but are jointly significant
- Another important special case of F-test is to test whether all of the slope parameters are jointly significant
    - H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$

- Although testing exclusion restrictions is the most important application of the F-test, F-test can be used to test general forms of linear restrictions
    - E.g.

$$H_0: \beta_2 = 1, \beta_3 = 0, \beta_4 = 0$$

$$H_0: \beta_2 + \beta_3 = 1, \beta_4 = 0$$

$$H_0: \beta_2 = 0, \beta_3 = 1, \beta_4 - 2\beta_5 = 4$$

$$H_0: \beta_1 = 0, \beta_2 - \beta_3 = 1, \beta_4 + \beta_5 = 3$$

Single hypothesis involving more than one parameter / linear combinations of parameters

Example. Suppose we have a model, $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$.

Let's say we want to test if the effect of β_2 is the same as the effect of β_3.

So, $H_0 : \beta_2 = \beta_3;\ H_a : \beta_2 \neq \beta_3$

We can rewrite the hypotheses as so: $H_0 : \beta_2 - \beta_3 = 0;\ H_a : \beta_2 - \beta_3 \neq 0$

The t_{STAT} is: $\frac{b_2 - b_3}{se(b_2 - b_3)}$

However, this is difficult to calculate, as $se(b_2 - b_3) \neq se(b_2) - se(b_3)$.

1. Define a new variable $\varphi = \beta_2 - \beta_3$.
    a. Thus, $H_0 : \varphi = 0;\ H_a : \varphi \neq 0$
    b. t_{STAT} is $\frac{\hat{\varphi}}{se(\hat{\varphi})}$
2. Estimate the transformed model:

    Since $\beta_2 = \varphi + \beta_3$, $\quad Y = \beta_1 + (\varphi + \beta_3)X_2 + \beta_3 X_3 + u$

    $$Y = \beta_1 + (\varphi)X_2 + \beta_3(X_2 + X_3) + u$$
3. Do an F-test on φ.

Week 11 - Multivariate Transformation

Data transformation

The multivariate model assumes linearity: $E[Y_i|X_{ki}] = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki}\ for\ all\ k$

It can be that some transformation of Y is linear in some transformations of X's:

$log(Production) = \beta_1 + \beta_2 log(labour) + \beta_3 log(capital) + u$

Taking logs

- Changes in logs can be interpreted as percentage changes (e.g. elasticities)
- Logs help up deal with a variable for which different observations are on very different scales
- Recap:

$y = b_1 + b_2 X_{2i} + b_3 X_{3i} + ... + b_k X_{ki}$	Δx_2 by 1 ➔ Δy-hat by b_2, holding all other regressors constant
$ln(y) = b_1 + b_2 X_{2i} + b_3 X_{3i} + ... + b_k X_{ki}$	Δx_2 by 1 ➔ Δy-hat by $(100*b_2)$%, holding all other regressors constant ● Retransformation bias ● "Ln" just means "divide by 100%" on the same side
$y = b_1 + b_2 ln(X_{2i}) + b_3 X_{3i} + ... + b_k X_{ki}$	Δx_2 by 1% ➔ Δy-hat by $(b_2/100)$, holding all other regressors constant ● All values of $x_2 > 0$ ● "Ln" just means "divide by 100%" on the same side
$ln(y) = b_1 + b_2 ln(X_{2i}) + b_3 X_{3i} + ... + b_k X_{ki}$	Δx_2 by 1% ➔ Δy-hat by b_2% ● Retransformation bias ● All values of $x_2 > 0$

- Cobb-Douglas production function:

  $Y = AK^{\alpha}L^{\beta}$

  $lnY = lnA + \alpha lnK + \beta lnL$

  Regression output is given by $\widehat{lnY} = \widehat{lnA} + \hat{\alpha} lnK + \hat{\beta} lnL$
- We can get the unbiased predicted value of lnY from this model for a given K and L
- However, to get an unbiased predicted value of Y, we account for retransformation bias:

  $\hat{Y} = e^{(\widehat{lnY})} \times e^{(se)^2/2}$

Polynomials

- We can also transform explanatory variables into polynomials
- We care about $\Delta\hat{Y}/\Delta X$ for very small amounts of X - this is the ***marginal effect***
- Most common is quadratic, $Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \dots$
    - Can also have cubic, $Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3 + \dots$
    - Marginal effect is the slope of this model at X
- Be wary of how the relationship between the explanatory variables will affect the predicted change in Y.
    - E.g. for $Y = \beta_1 + \beta_2 X + \beta_3 X^2 + u$:

○ If X changes from 1 to 2, predicted change in Y:

$$(\beta_1 + \beta_2 * 2 + \beta_3 * 2^2) - ((\beta_1 + \beta_2 * 1 + \beta_3 * 1^2) = \beta_2 + 3\beta_3$$

○ If X changes from 10 to 11, predicted change in Y:

$$(\beta_1 + \beta_2 * 11 + \beta_3 * 11^2) - ((\beta_1 + \beta_2 * 10 + \beta_3 * 10^2) = \beta_2 + 21\beta_3$$

- Use the polynomial transformation when you have a good reason to believe that the population model follows this form. Otherwise, we might have a model that fits the data in the sample very well, but is nowhere near the population.
    - Holding all the other regressors fixed, does this polynomial form correctly capture how the marginal effect of Y *on* X changes?
- See how Y changes with respect to X's, and compare that rate of change with the polynomial plot