ECON-122
Introduction to Econometrics

Agnieszka Postepska

August 2nd, 2011

Testing multiple parameters

- In applications we often want to test hypothesis about more than one parameter
- an example of such a test would be to test whether in a model with two regressors, x_1 and x_2, the corresponding coefficients, β_1 and β_2 are equal, i.e. do both of the $x's$ have the same impact on the dependent variable y
- let's consider the following example:

$$ln(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

where:

$$\begin{aligned} jc &= \text{number of years of attending a two-year college} \\ univ &= \text{number of years of attending a four-year college} \\ jc &= \text{months in the workforce} \end{aligned}$$

And the hypothesis of interest is: $H_0: \ \beta_1 = \beta_2$

Testing multiple parameters - cont.

- the null says that another year at junior college and another year at a university lead to the same percentage change in *wage*
- the alternative is: $H_1: \ \beta_1 < \beta_2$ (why one-tailed test?)
- we cannot just use the t-statistic for $\hat{\beta}_1$ and $\hat{\beta}_2$ - we need to construct a new test statistic
- rewrite the null and the alternative as:

$$H_0: \ \beta_1 - \beta_2 = 0 \quad \text{and} \quad \beta_1 - \beta_2 < 0$$

- the corresponding t-statistic is:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$

Testing multiple parameters - cont.

- the testing procedure is just as before - we chose a significance level, find the critical value, and decide whether we reject the null
- the only difficulty is to compute the standard error of the difference of the coefficients
- we need to do some math to figure out how to get it
- recall that:

$$Var(\hat{\beta}_1 - \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)$$

- therefore, since standard deviation is just a square root of the above and standard error is an unbiased estimator for standard deviation:

$$se(\hat{\beta}_1 - \hat{\beta}_2) = [se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 - 2Cov(\hat{\beta}_1, \hat{\beta}_2)]^{\frac{1}{2}}$$

- how do we actually compute the covariance? luckily most of the statistical softwares have options to compute it so we don't have compute it by hand (look up matrix-covariance matrix for $\hat{\beta}$)

Testing multiple parameters - cont.

- alternative way, which delivers the answer immediately
- first define new parameter $\theta_1 = \beta_1 - \beta_2$ and then estimate the following model:

$$\begin{aligned} ln(wage) &= \beta_0 + (\theta_1 - \beta_2)jc + \beta_2 univ + \beta_3 exper + u \\ &= \beta_0 + \theta_1 jc + \beta_2(jc + univ) + \beta_3 exper + u \end{aligned}$$

- now we have the standard error of $(\beta_1 - \beta_2)$ explicitly estimated as it is the standard error of the estimate of θ_1
- note that the only reason we estimated the second model is to get this estimate directly
- this approach always work when we want to test linear hypothesis about two (or more) parameters in the model-the trick is to rewrite the model so that the combination of coefficients of interest appears as a single coefficient

Testing multiple (joint) linear restrictions: **F Test**

- consider the following model:

$$ln(w_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

- we know how to test whether schooling has no effect on wages - simple t test: $H_0 : \beta_1 = 0$
- to test whether experience has effect - we can't do the same because now $H_0 : \beta_2 = 0$ **and** $\beta_3 = 0$
- why not multiple t-tests?
  - computational cost
  - different purpose: in our example t-test on β_2 would tell us whether *experience* has effect on *wage* holding *experience squared* constant
- to be able to test these joint hypotheses we need a new test statistic with a known distribution under the joint hypotheses

F-test cont.

- we already know that the null in this multiple restrictions test will be that all of the variables included in the test have no effect on y
- therefore, the alternative is just the opposite: at least one variable (so one x) have and effect on y
- note that it is enough that only one tested coefficient is significantly different from zero for the null to be rejected
- so the exercise we are doing here is essentially testing the restricted model (model with some variables excluded) against the unrestricted model (model with all variables)
- formally, if we are testing whether q parameters jointly have no effect on y, the null and the alternative become:

$$\begin{aligned} H_0: \ & \beta_{k-q} = \beta_{k-q+1} = ... = \beta_k = 0 \\ H_1: \ & \text{at least one of } \beta_{k-q}, \beta_{k-q+1}, ..., \beta_k \neq 0 \end{aligned}$$

where the order of variable is arbitrary - we are imposing q restrictions on chosen q variables out of all k regressors

Restricted versus unrestricted model

- Restricted model, so the model under the null hypothesis:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_{i(k-q-1)} x_{i(k-q-1)} + u_i$$

- whereas unrestricted model is as before:

$$y_i = \beta_0 + \beta_1 x_i 1 + ... + \beta_{ik} x_{ik} + u_i$$

- so we can run these two regressions and compare the results: if additional regressors do not add to the model, then residual sum of squared should not change: $SSR_{restricted} = SSR_{unrestricted}$
- so, if $SSR_{restricted} > SSR_{unrestricted}$ then we reject the null, otherwise we cannot reject the null
- this observation is the basis for constructing the test-statistics and under A1-A6 we have the following result for F test:

$$F_{q,n-k-1} \equiv \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n-k-1)} \sim F_{q,n-k-1}$$

F-statistic in detail

- very easy to compute - *SSR* is given in every OLS regression
- F-stat is always positive (why?)
- test procedure (same as in t-tets):
  - specify null and alternative
  - choose significance level (α) and find critical value given the chosen significance level
  - compute test statistic
  - decide whether to reject the null at the chosen level of significance (reject if $F > c_\alpha$)
- note that $c_{10} < c_5 < c_1$ so if you reject the null at 1 percent you will always reject it at 5 and 10 percent (that applies to t test as well)
- in tables with critical values - denominator degrees of freedom are in the columns, whereas numerator degrees of freedom in the rows
- if you reject the null, you say that the variables are **jointly significant**, if you fail to reject the null, you say that the variables are **jointly insignificant**

F-statistic: alternative formulation

- we can also look at R^2 (instead of looking at SSR) to see whether additional variables have explanatory power
- in other words - do we gain anything in terms of R^2 by adding the variables into the model: $R^2_{unrestircted} \geq R^2_{restricted}$
- if $R^2_{unrestircted} - R^2_{restricted}$ is close to zero then we can suspect it is insignificant, if this difference is large than this is an evidence that it is significant

F-test:

Given an unrestricted and a restricted model, under A1-A6, then:

$$\frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n-k-1)} \sim F_{q,n-k-1}$$

Significance of the regression

- now we will test whether any of the regressors help to explain y
- so the null and alternative become:

$$\begin{aligned} H_0: \;& \beta_1 = \beta_2 = ... = \beta_k = 0 \\ H_1: \;& \text{at least one of } \beta_j \neq 0 \text{ where } j = 1, 2, ..., k \end{aligned}$$

- rejecting the null implies that the specification does not do any better than the mean at predicting y
- now the restricted model becomes:

$$\begin{aligned} y_i &= \beta_0 + u_i \\ \rightarrow R^2_{restricted} &= 0 \end{aligned}$$

- the F statistic of this test is reported automatically by Stata
- notice that you we might have a low R^2 and a large value of F statistic (model is significant but doesn't explain much of the variation in y)

Testing general linear restrictions - general form of F-test

- we can test any combination of linear restrictions
- same procedure as for exclusion test
- R^2 form of the test statistic cannot be applied in this case (usually we have to transform our dependent variable so that the R^2 from the restricted model explains variation in the new dependent variable, not the original one - look at the example)
- examples can be found in application - for example when testing elasticities