[STAT-19] Testing nested models

Miguel-Angel Canela
Associate Professor, IESE Business School

Nested models

The definition of **nested models** given here is general enough to cover a wide range of models, beyond those of this section. We say that two models are nested when one is a particular case of the other, resulting from constraining some of the parameters of the general model. You will find in the literature that, frequently, the objective of the statistical analysis is to choose between two nested models.

The concept of nested models is easily understood in an example of multiple linear regression. Imagine the linear regression models associated to the equations

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon, \qquad Y = \beta_0 + \beta_1 X_1 + \epsilon.$$

It is clear that the **restricted model** (right) is obtained from the **unrestricted model** (left) by constraining $\beta_2 = \beta_3 = 0$. When we talk about "comparing" these two models, we mean testing, in the unrestricted model, the null $H_0 : \beta_2 = \beta_3 = 0$. The equalities in this null hypothesis are called **restrictions** or constraints. I present in this lecture an F test for linear restrictions in the context of OLS estimation. There is a similar approach for maximum likelihood estimation, the **likelihood ratio test**, which we will use profusely in the Econometrics and Multivariate Stats courses.

Testing linear restrictions

Any set of linear restrictions can be tested in a simple way. The idea of the test is as follows. The OLS estimates are the solution of an optimization problem. In the restricted model, the search for the minimum is carried out in a smaller set. For instance, in the example of the preceding section, setting $\beta_2 = \beta_3 = 0$ restricts the search to a two-dimensional space, instead of the four-dimensional space of the unrestricted model. This implies that the residual sum of squares SSR increases when introducing the restrictions and, therefore, R-squared decreases. The test statistic can be written in terms of either SSR or R-squared.

Let the subscripts r and u refer to the restricted to the unrestricted model, respectively. Then, the test statistic is

$$F = \frac{\dfrac{\mathrm{SSR_r} - \mathrm{SSR_u}}{\mathrm{df_r} - \mathrm{df_u}}}{\dfrac{\mathrm{SSR_u}}{\mathrm{df_u}}} = \frac{\dfrac{R^2_\mathrm{u} - R^2_\mathrm{r}}{\mathrm{df_r} - \mathrm{df_u}}}{\dfrac{1 - R^2_\mathrm{u}}{\mathrm{df_u}}}.$$

Under the null (the two restrictions), this statistic has an $F(\mathrm{df_r} - \mathrm{df_u}, \mathrm{df_u})$ distribution, which is used to calculate the corresponding P-value. Note that, in the above example, the unrestricted model has $\mathrm{df_u} = n - 4$, and the restricted model $\mathrm{df_r} = n - 2$, so the distribution of the F statistic is $F(2, n - 4)$.

We typically regard this F test as a test on ΔR^2. In certain fields, it is customary to report two or more nested models, testing the increase in R-squared due to the additional variables. This is sometimes called **hierarchical regression**, not to be confounded with the hierarchical linear models which are another name of the multilevel models that will be seen in the Econometrics course.

Particular cases

This F test gives as particular cases the tests of lecture 18. When testing only one term, it is equivalent to the corresponding t test that comes in the coefficients table. When testing all the terms except the intercept, it is the F test associated to the ANOVA table. More specifically, in the 4-terms unrestricted model used above as an example:

- The default F test applies to the null $\beta_1 = \beta_2 = \beta_3 = 0$.
- The t test for the null $H_0 : \beta_j = 0$ $(j = 1, 2, 3)$ is equivalent to the corresponding F test for this restriction. The square of the t statistic equals the F statistic.

Example 1. Models for the influence of education on wages are frequently used in Econometrics courses. We already saw one of these examples in lecture 17. The `wage1` data set is a subset of a bigger data set used in the Wooldridge's textbook. It contains 526 observations on wages of workers. The variables included are:

- `wage`, 1976 wages in US work force, in US dollars per hour.
- `educ`, years of schooling. `educ = 12` corresponds to complete high school education.
- `exper`, years of potential experience.
- `tenure`, years with current employer.

A standard approach would be to run a regression of wages, in log scale (`lwage`) on the three explanatory variables. Table 1 is the table of coefficients. The R-squared is 0.316 ($F = 80.4$, $P < 0.001$). The coefficient of `educ` is significant, so the conclusion is easy.

TABLE 1. Linear regression results (Example 1)

Coefficient	Estimate	Std. error	t value	p-value
Intercept	0.284	0.104	2.79	0.007
educ	0.092	0.007	12.6	0.000
exper	0.004	0.002	2.39	0.017
tenure	0.0221	0.003	7.13	0.000

I take first the 3-regressor model as the unrestricted model and the model without `exper`, whose coefficient looks less significant than the other two, as the restricted model. Table 2 is the new table of coefficients.

TABLE 2. First restricted model

Coefficient	Estimate	Std. error	t value	p-value
Intercept	0.404	0.092	4.41	0.000
educ	0.087	0.007	12.4	0.000
tenure	0.026	0.003	9.63	0.000

R-squared falls to 0.308 ($\Delta R^2 = 0.008$). The F statistic |df $= (1, 522)$) for the change in R-squared is

$$F = \frac{0.316 - 0.308}{(1 - 0.316)/522} = 5.72,$$

which is the square of the t statistic associated to the coefficient of `exper` in Table 1 ($t = 2.39$).

The p-values associated are the same.

In a second comparison, I take as the restricted model the **null model** obtained by dropping all the regressors, with $R^2 = 0$. The equation includes only the intercept, which equals the mean log wages (Table 3). Now, $F = 80.39$ ($P < 0.001$), which is, precisely, the F statistic of the unrestricted model.

TABLE 3. Second restricted model

Coefficient	Estimate	Std. error	t value	p-value
Intercept	1.623	0.023	70.0	0.000

In my third exercise, I test two coefficients (this test is not included in the default regression output). I take as the restricted model the equation obtained by dropping `exper` and `tenure` (Table 4). Now, $R^2 = 0.186$ ($\Delta R^2 = 0.130$, $F = 49.7$, $P < 0.001$).

TABLE 4. Third restricted model

Coefficient	Estimate	Std. error	t value	p-value
Intercept	0.584	0.097	5.60	0.000
educ	0.083	0.008	10.9	0.000

¶ Source: JM Wooldridge (2013), *Introductory Econometrics — A Modern Approach*, South-Western College Publishing. □

Homework

A. That experience has a positive effect on wages seems evident, but that the effect is linear is unclear. Common sense tells us that the first years are more relevant, and the effect gets weaker when workers have experience enough. A simple way to account with the curvilinear effect of the experience on the wages is to include a squared term in the equation. Try this, and test the two terms related to the experience in one shot.