Lecture 02

Hypothesis testing

Classical hypothesis testing

1. Formulate two competing hypotheses;
2. Derive a test statistic and its sampling distribution;
3. Derive the decision rule to choose one of the hypotheses.

Example 1:

Does an increase in government spending (G) lead to an increase in the interest rate (i)?

Simple model:

$$i_t = \beta_1 + \beta_2 \cdot G_{t-1} + u_t$$

assuming that $u_t \sim N(0, \sigma^2)$.

- In advance, choose the model.

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

- Then, find the test statistic:
  - t-test, given that u_t follows a normal distribution, and consequently so does i_t .
- Find the rejection region:
  - The areas in which the value of the test statistic will lead to the rejection of the null hypothesis.

The regression will give you the estimated parameter $\hat{\beta}_2$ and the standard error (SE) for the estimator $\hat{\sigma}_{\hat{\beta}_2}$.

- The test statistic will then be:

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_2^{H_0}}{\hat{\sigma}_{\hat{\beta}_2}},$$

where $\hat{\beta}_2^{H_0}$ is the hypothesized value for the estimated parameter (in this example, $\hat{\beta}_2^{H_0} = 0$).

- The t statistic measures how many standard deviations away the estimated values are from what was hypothesized.
  - If that distance $\hat{\beta}_2 - \hat{\beta}_2^{H_0}$ were normalized by the standard deviation, the statistic would follow a normal distribution;
  - Since it is divided by an estimate of the standard deviation ($\hat{\sigma}_{\hat{\beta}_2}$), it follows Student's t-distribution.
- If the test statistic that was calculated falls in the rejection region, the null hypothesis should be rejected; otherwise, one fails to reject the null hypothesis.
  - For instance, with $N = 25$ (small sample, hence the t-test) and $\alpha = 0.05$ - *i. e.*, a 95% confidence -, the null hypothesis should be rejected in case $|t| > 2.069$:

- 2.069 is the critical value for a two-tailed t-test, with $N - k = 23\ df$ (25 minus the number of estimated parameters, β_1 and β_2).

Example 2:

Were the hypotheses:

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 > 0$$

a one-tailed t-test should be performed.

- For the same α and N (and consequently, the same number of degrees of freedom), the critical value would be 1.714 .

Joint Hypotheses testing

(Unrestricted) Model:

$$y_i = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u_i$$

Hypotheses:

$$\begin{aligned} H_0 : &\quad \beta_3 = \beta_4 = 0 \\ H_1 : &\quad \beta_3 \neq 0 \ \lor \ \beta_4 \neq 0 \end{aligned}$$

Restricted Model [apply the null hypothesis]:

$$y_i = \beta_1 + \beta_2 x_2 + u_i$$

1. Estimate the unrestricted model and save the sum of squared residuals (RSS_{UR});
2. Estimate the restricted model and save the sum of squared residuals (RSS_{R});
3. Calculate the test statistic:

$$F = \frac{\dfrac{RSS_{\mathrm{R}} - RSS_{\mathrm{UR}}}{\text{Num. of restrictions}}}{\dfrac{RSS_{\mathrm{UR}}}{N - k}}$$

- If the null hypothesis is indeed true, $RSS_{\mathrm{R}} - RSS_{\mathrm{UR}} = 0$;
- The numerator is the average (normalization) of the sum of squared residuals *per* restriction;
- The distribution of the numerator is unknown; however, once it is divided by the denominator, it assumes a known distribution: F-distribution;
- The number of restrictions is the difference between the number of parameters (β_i's) in the unrestricted and restricted models:
  - If there is only one restriction, it is also possible to perform a t-test;
- k in the denominator refers to the unrestricted model: it is the number of parameters in the unrestricted model;
- An alternative formula for the test statistic is:

$$F = \frac{\dfrac{R^2_{\mathrm{UR}} - R^2_{\mathrm{R}}}{\text{Num. of restrictions}}}{\dfrac{1 - R^2_{\mathrm{UR}}}{N - k}}.$$

- The R^2 is [coefficient of determination] of each model;

- More info.

4. Find the critical value F_{crit}, with the number of restrictions as the degrees of freedom for the numerator (ν_1), and $N-k$ as the degrees of freedom for the denominator (ν_2);
5. If $F > F_{\text{crit}}$, reject the null hypothesis.

Example:

Another example of joint hypotheses test (also solvable with an F-test):

$$\begin{aligned} H_0 : &\quad \beta_2 + 2\beta_3 = 4 \\ H_1 : &\quad \beta_2 + 2\beta_3 \neq 4 \end{aligned}$$

Imposing restriction:

$$\beta_2 = 4 - 2\beta_3, \qquad \text{under } H_0$$

submitting into the model:

$$y_i = \beta_1 + (4 - 2\beta_3)x_2 + \beta_3 x_3 + \beta_4 x_4 + u_i$$

and isolating β_i's:

$$y_i - 4x_2 = \beta_1 + \beta_3(x_3 - 2x_2) + \beta_4 x_4 + u_i$$

one obtains the restricted model.

Rewriting:

$$\frac{\tilde{y} = \beta_1 + \beta_3 \tilde{x} + \beta_4 x_4 + u}{\begin{aligned} \tilde{y} = &\quad y - 4x_2 \\ \tilde{x} = &\quad x_3 - 2x_2 \end{aligned}}$$

Example:

Model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$$

Hypotheses:

$$\begin{aligned} H_0 : &\quad \beta_1 - \beta_3 = 1 \\ &\wedge \beta_2 + 2\beta_4 = 6 \\ H_1 : &\quad \beta_1 - \beta_3 \neq 1 \\ &\vee \beta_2 + 2\beta_4 \neq 6 \end{aligned}$$

The restricted model is obtained by submitting both conditions in the null hypothesis.

Then substitute:

$$\beta_1 = 1 + \beta_3 \qquad \beta_2 = 6 - 2\beta_4$$

and obtain the restricted model:

$$y_i = (1 + \beta_3) + (6 - 2\beta_4)x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$$

Rewriting:

$$\tilde{y} = \beta_3\tilde{x}_3 + \beta_4\tilde{x}_4 + u_i$$

$$\begin{aligned} \tilde{y} &= \quad y_i - 1 - 6x_{2i} \\ \tilde{x}_3 &= \quad 1 + x_{3i} \\ \tilde{x}_4 &= \quad x_{4i} - 2x_{2i} \end{aligned}$$

Application:

Cobb-Douglas production function: $y_i = AL_i^{\alpha}K_i^{\beta}e^{u_i}$;

- y_i: output;
- A: technology;
- L: labor;
- K: capital.

When $\alpha + \beta = 1 \ \rightarrow$ constant returns to scale.

Putting to test the existence of constant returns to scale:

$$\begin{aligned} H_0: &\quad \alpha + \beta = 1 \\ H_1: &\quad \alpha + \beta \neq 1 \end{aligned}$$

Linearizing the model:

$$\ln y_i = \ln A + \alpha \ln L_i + \beta \ln K_i + u_i$$

Plugging in the condition in the null hypothesis ($\alpha = 1 - \beta$):

$$\ln y_i - \ln L_i = \ln A + \beta(\ln K_i - \ln L_i) + u_i,$$

whose parameters to estimate are $\ln A$ (which might as well be renamed) and β.

Succinctly:

$$\ln \tilde{y} = A' + \beta\tilde{x} + u_i$$

$$\begin{aligned} \tilde{y} &= \quad \ln y_i - \ln L_i \\ \tilde{x} &= \quad \ln K_i - \ln L_i \\ A' &= \quad \ln A \end{aligned}$$