Slides for the course

Statistics and econometrics

Part 9: Hypothesis testing

European University Institute

Andrea Ichino

September 11, 2014

Outline

The logic of classical hypothesis testing

Small sample distribution of the OLS estimator

A list of possible hypotheses and the correspondent tests

Confidence intervals

Linear combinations of parameters

The F-test

Section 1

The logic of classical hypothesis testing

The question we want to address now

We are now interested in testing hypothesis concerning the parameters of the PRF, using the estimator that we have constructed and analysed in the previous sections.

Here are some examples of hypotheses that we may want to test

- $\beta_j = 0$;
- $\beta_j = q$ where q is any real number;
- $\beta_j \leq q$ where q is any real number, including 0;
- $\beta_j = \beta_h$;
- $\beta_j^2 - 2\beta_j\beta_i = 0$
- $r(\beta) = q$ where $r(.)$ is any function of the parameters.

What does it mean to test an hypothesis in statistics

- Define the “null hypothesis" H_0 to be tested on a parameter.
- Construct a “test statistic" and find its distribution under H_0.
- Compute the test statistic in the specific sample at our disposal.
- Using the theoretical distribution of the test statistic establish the probability of observing its observed value if H_0 is true.
- If this probability is “sufficiently small" reject H_0.
- The “significance" of the test is the threshold level of probability that we consider sufficiently low to conclude that it is unlikely that the observed test statistics could have originated under H_0.
- The “p-value" of the test is the smallest significance level at which H_0 would actually be rejected given the sample. Note that the p-value is a probability

Power of a test, Type I and type II errors

The significance of a test measures the probability of rejecting H_0 when it is true,

- This is the probability of “Type I" decision errors.

But given a specific alternative H_1, we are interested also in the probability of failing to reject H_0 when H_1 is in fact true

- This is the probability of “Type II" decision errors.

The “power of a test" with respect to a specific alternative, is 1 minus the probability of Type II errors:

- the probability of not rejecting the alternative when it is true

Computing the power of a test requires defining a specific alternative and the distribution of the test statistic under H_1

For given significance and alternative, we are interested in finding the test with the highest power.

Section 2

Small sample distribution of the OLS estimator

How can we find a distribution for the OLS parameters

If we have a large sample and we can use asymptotic results, using

- the Central Limit Theorem
- The Delta Method

we can say that:

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \text{Normal}\left(0, \frac{\sigma^2}{Var(x)}\right) \tag{1}$$

without making distributional assumptions on X.

But if we are in a small sample setting, we need distributional assumptions on X to say how $\hat{\beta}_{OLS}$ is distributed

Note that in the case of $\hat{\beta}_{ML}$ the small sample distribution is normal because we have assumed normality to construct the estimator.

The assumption of Normality of U

The Classical Linear Model Assumption is Normality:

- MLR 6: In the population U is independent of X and is distributed normally with zero mean and variance $\sigma^2 I_n$

$$U \sim \text{Normal}(0, \sigma^2 I_n) \tag{2}$$

Note that this implies

$$Y \sim \text{Normal}(X\beta, \sigma^2 I_n) \tag{3}$$

Discussion of the small sample assumption of Normality.

From the distribution of U to the distribution of $\hat{\beta}$

We know that

$$\hat{\beta} \quad = \quad \beta + (X'X)^{-1}X'U \tag{4}$$

using 2 it is easy to see that

$$\hat{\beta} \sim \text{Normal}(\beta, \sigma^2(X'X)^{-1}) \tag{5}$$

And for a single PRF parameter we have that the standardized distribution

$$\frac{\hat{\beta}_j - \beta}{sd(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{\frac{\sigma}{\sqrt{SST_j(1-R_j^2)}}} \sim \text{Normal}(0, 1) \tag{6}$$

From the distribution of U to that of $\hat{\beta}$ (cont.)

In practice, we do not know σ and we have to use its estimate $\hat{\sigma} = \frac{\hat{U}'\hat{U}}{n-k-1}$ so that:

$$\frac{\hat{\beta}_j - \beta}{\hat{sd}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{\frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}} \sim t_{n-k-1} \tag{7}$$

where t_{n-k-1} denotes a "t distribution" with $n-k-1$ degrees of freedom.

Section 3

A list of possible hypotheses and the correspondent tests

$H_0 : \beta_j = 0$ against the one sided alternative $H_1 : \beta_j > 0$

The simplest testable hypothesis is that X_j has positive effect on Y

$$H_0 : \beta_j = 0 \qquad \text{against} \qquad H_1 : \beta_j > 0 \tag{8}$$

The test statistic for this hypothesis and its distribution under H_0 are

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} \sim t_{n-k-1} \tag{9}$$

We reject H_0 if in our sample

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} > c \tag{10}$$

where the critical level $c > 0$ is such that

$$Pr(\tau > c|H_0) = s \qquad \text{with} \qquad \tau \sim t_{n-k-1} \tag{11}$$

s is the significance level (e.g. $s = 0.01$ or $s = 0.05$). The p-value is:

$$p = Pr(\tau > t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)}|H_0) \tag{12}$$

$H_0 : \beta_j = 0$ against the one sided alternative $H_1 : \beta_j < 0$

Similarly we can test that X_j has a negative effect on Y

$$H_0 : \beta_j = 0 \qquad \text{against} \qquad H_1 : \beta_j < 0 \tag{13}$$

The test statistic for this hypothesis and its distribution unde H_0 are

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} \sim t_{n-k-1} \tag{14}$$

We reject H_0 if in our sample

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} < -c \tag{15}$$

where the critical level $-c < 0$ is such that

$$Pr(\tau < -c|H_0) = s \qquad \text{with} \qquad \tau \sim t_{n-k-1} \tag{16}$$

s is the significance level (e.g. $s = 0.01$ or $s = 0.05$). The p-value is:

$$p = Pr(\tau < t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)}|H_0) \tag{17}$$

$H_0 : \beta_j = 0$ against a two sided alternative $H_1 : \beta_j \neq 0$

More generally we can test that X_j has a non zero effect on Y

$$H_0 : \beta_j = 0 \qquad \text{against} \qquad H_1 : \beta_j \neq 0 \tag{18}$$

The test statisticand its distribution under H_0 are again

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} \sim t_{n-k-1} \tag{19}$$

We reject H_0 if in our sample

$$|t_{\hat{\beta}_j}| = \left| \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} \right| > c \tag{20}$$

where the critical level c is such that

$$Pr(|\tau| > c|H_0) = 0.5s \qquad \text{with} \qquad \tau \sim t_{n-k-1} \tag{21}$$

s is the significance level (e.g. $s = 0.01$ or $s = 0.05$). The p-value is:

$$p = 2Pr(\tau > |t_{\hat{\beta}_j}| = \left| \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} \right| |H_0) \tag{22}$$

$H_0 : \beta_j = k$ against the two sided alternative $H_1 : \beta_j \neq k$

In this case we test that the effect of X_j has a specific size:

$$H_0 : \beta_j = k \qquad \text{against} \qquad H_1 : \beta_j \neq k \tag{23}$$

The test statistic and its distribution under H_0 are again

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - k}{\hat{sd}(\hat{\beta}_j)} \sim t_{n-k-1} \tag{24}$$

We reject H_0 if in our sample

$$|t_{\hat{\beta}_j}| = \left| \frac{\hat{\beta}_j - k}{\hat{sd}(\hat{\beta}_j)} \right| > c \tag{25}$$

where the critical level c is such that

$$Pr(|\tau| > c|H_0) = \frac{1}{2}s \qquad \text{with} \qquad \tau \sim t_{n-k-1} \tag{26}$$

s is the significance level (e.g. $s = 0.01$ or $s = 0.05$). The p-value is:

$$p = 2Pr(\tau > |t_{\hat{\beta}_j}| = \left| \frac{\hat{\beta}_j - k}{\hat{sd}(\hat{\beta}_j)} \right| |H_0) \tag{27}$$

Section 4

Confidence intervals

What if we care about intervals of the parameter?

Consider the interval $\{-\lambda_\Phi, \lambda_\Phi\}$ defined by the equation:

$$Pr\left(-\lambda_\Phi < \frac{\hat{\beta}_j - \beta_j}{\hat{sd}(\hat{\beta}_j)} < \lambda_\Phi\right) = \Phi \qquad (28)$$

The limits $\{-\lambda_\Phi, \lambda_\Phi\}$ can be computed using the fact that

$$\frac{\hat{\beta}_j - \beta}{\hat{sd}(\hat{\beta}_j)} \sim t_{n-k-1}$$

Rearranging 28:

$$Pr\left(\hat{\beta}_j - \lambda_\Phi \hat{sd}(\hat{\beta}_j) < \beta < \hat{\beta}_j + \lambda_\Phi \hat{sd}(\hat{\beta}_j)\right) = \Phi \qquad (29)$$

which says that with probability Φ the interval $\{\hat{\beta}_j \pm \lambda_\Phi \hat{sd}(\hat{\beta}_j)\}$ contains the parameter β.

Confidence interval in large sample

In large sample, when the *t* distribution approximates normal distribution, a realiable approximation of the 95% confidence interval is

$$Pr\left(\hat{\beta}_j - 1.96\hat{sd}(\hat{\beta}_j) < \beta < \hat{\beta}_j + 1.96\hat{sd}(\hat{\beta}_j)\right) = 0.95 \qquad (30)$$

which means that with 95% probability an interval of two standard deviations around the estimate contains the parameter.

Section 5

Linear combinations of parameters

A more complex kind of hypothesis

There are situations in which we are interested in testing a slightly more complicated hypothesis:

$$H_0 : \beta_j = \beta_k \qquad \text{against} \qquad H_1 : \beta_j \neq \beta_k \tag{31}$$

The test statistic for this hypothesis and its distribution under H_0 are again

$$t_{\hat{\beta}_j, \hat{\beta}_k} = \frac{\hat{\beta}_j - \hat{\beta}_k}{\hat{sd}(\hat{\beta}_j - \hat{\beta}_k)} \sim t_{n-k-1} \tag{32}$$

and we could follow the usual procedure to test the hypothesis

A more complex kind of hypothesis (cont.)

What is slighlty more problematic in this case is the computation of

$$\hat{sd}(\hat{\beta}_j - \hat{\beta}_k) = \sqrt{[\hat{sd}(\hat{\beta}_j]^2 + [\hat{sd}(\hat{\beta}_k]^2 - 2\hat{Cov}(\hat{\beta}_j, \hat{\beta}_k)} \quad (33)$$

Given that $Var(\hat{\beta}|X) = \hat{\sigma}^2(X'X)^{-1}$ we have all the ingredients to compute the test statistics. But there is a simpler alternative.

Rearranging the PRF to test linear combinations

Consider the population regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{34}$$

and suppose that we want to to test

$$H_0 : \beta_1 = \beta_2 \qquad \text{against} \qquad H_1 : \beta_1 \neq \beta_2 \tag{35}$$

If we add and subtract $\beta_2 x_1$ in 34, we get:

$$\begin{aligned} y &= \beta_0 + (\beta_1 - \beta_2)x_1 + \beta_2(x_2 + x_1) + u \qquad (36) \\ y &= \beta_0 + \theta x_1 + \beta_2(x_2 + x_1) + u \end{aligned}$$

and we can now test with the standard procedure:

$$H_0 : \theta = 0 \qquad \text{against} \qquad H_1 : \theta \neq 0 \tag{37}$$

Note that the estimates of the coefficients on x_2 in 34 and on $(x_2 + x_1)$ in 36 must be numerically identical.

Section 6

The F-test

Multiple linear restrictions

Consider the unrestricted regression in matrix form

$$Y = X_1\beta_1 + X_2\beta_2 + U_{ur} \tag{38}$$

- X_1 is a $n \times k_1 + 1$ matrix;
- β_1 is $k_1 + 1$ vector of parameters;
- X_2 is a $n \times k_2$ matrix;
- β_2 is k_2 vector of parameters;

Suppose that we want to test the following joint hypothesis on the β_2:

$$H_0 : \beta_2 = 0 \qquad \text{against} \qquad H_1 : \beta_2 \neq 0 \tag{39}$$

In which sense and why testing the joint hypothesis is different than the testing the k_2 separate hypotheses on the β_2 parameters?

The F test statistics

Consider the restricted regression

$$Y = X_1\beta_1 + U_r \tag{40}$$

and the unrestricted PRF 38.

A natural starting point to construct a test statistic for the joint hypothesis is to see by how much the Sum of Squared Residuals (SSR) increases going from the restricted to the unrestricted PRF

The F test statistics (cont.)

The F statistic is built around this idea:

$$F = \frac{\frac{(SSR_r - SSR_{ur})}{k_2}}{\frac{SSR_{ur}}{n-k-1}} \sim F_{k_2, n-k-1} \qquad (41)$$

where k_2 is the number of restrictions (the dimension of X_2) and k is the total number of parameters.

The F statistic is distributed accordint to an F distribution because it can be shown to be the ratio of two χ^2 distributions.

Note that the numerator of F is always positive and it is larger, the larger the reduction of SSR delivered by the unrestricted PRF.

The F test statistics (cont.)

We reject H_0, if our sample gives

$$|F| \geq c \tag{42}$$

where the critical level c is such that

$$Pr(f > c|H_0) = s \qquad \text{with} \qquad f \sim F_{k_2, n-k-1} \tag{43}$$

s is the significance level (e.g. $s = 0.01$ or $s = 0.05$). The p-value is:

$$p = Pr(f > F|H_0) \tag{44}$$

Note that the F statistics can be constructed not only for exclusion restrictions but also for more complicated linear restrictions, as long as we can specify the restricted and unrestricted PRF.

The "R-squared" form of the F test

In some cases it may be convenient to exploit the fact that

$$SSR_r = (1 - R_r^2) \quad (45)$$

$$SSR_{ur} = (1 - R_{ur}^2) \quad (46)$$

and therefore the F statistics can be expressed as a function of the R-squared of the restricted and unrestricted distribution:

$$F = \frac{\frac{(R_{ur}^2 - R_r^2)}{k_2}}{\frac{1 - R_{ur}^2}{n-k-1}} \sim F_{k_2, n-k-1} \quad (47)$$

This form of the test is completely equivalent but more convenient for computational purposes.

The F test and the overall significance of a regression

Most packages report the F test for the joint hypothesis that all the regressors have no effect:

$$H_0 : \beta = 0 \qquad \text{against} \qquad H_1 : \beta \neq 0 \tag{48}$$

In this case the restricted PRF is

$$y = \beta_0 + U_r \tag{49}$$

and the F test is

$$F = \frac{\frac{(R^2)}{k}}{\frac{1-R^2}{n-k-1}} \sim F_{k,n-k-1} \tag{50}$$

because the R-squared of the restricted PRF is zero.

This F test provides the same info of the R-squared statistic, but it is framed to allows for a test on the significance of all the regressors.