Slides for the course

Statistics and econometrics

Part 10: Asymptotic hypothesis testing

European University Institute

Andrea Ichino

September 8, 2014

Outline

Why do we need large sample hypothesis testing?

The “trinity” of large sample testing strategies

- The Wald test
- The Lagrance Multiplier (or Score) test
- The Likelihood Ratio test

Large sample testing and regression

- Wald test in the context of regression
- The Lagrange multiplier test in the context of linear regression
- The Likelihood Ratio Test in the context of regression
- Final comment on the “trinity of tests”

Section 1

Why do we need large sample hypothesis testing?

No need of distributional assumptions

More than precision, the advantage of a large sample for testing is that no distributional assumption on the outcome Y is needed.

In particular we do not have to assume MLR 6: normality of $U|X$.

This is particularly important from a methodological point of view because it allows us to use all the machinery of regression analysis also in cases where normality is clearly a wrong assumption:

- Discrete dependent variables
- Limited dependent variables
- “Conditional on positive" models
- Duration analysis
- Count data analysis

Thanks to large samples, econometrics becomes considerably simpler!

Section 2

The “trinity” of large sample testing strategies

Justification of the three tests in a nutshell

$$H_0 : \theta = \theta_0 \tag{1}$$

1. Wald test
   - Under the null, the normalised distance between $\hat{\theta}$ and θ_0 should be small.
2. Lagrange multiplier (or Score) test
   - Under the null, the score $l_\theta(X, \theta_0)$ evaluated at θ_0 should be close to zero .
3. Likelihood ratio test
   - The ratio between the likelihood $L(X, \hat{\theta})$ evaluated at at the estimate $\hat{\theta}$ and the likelihood $L(X, \theta_0)$ evaluated at θ_0 should be close to 1.

Figure 3.1 in Engle's Handbook chapter is a useful way to think at the three tests

Subsection 1

The Wald test

The logic of the Wald test in detail

We have established that

$$\sqrt{n}(\hat{\theta}_n - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, \Omega) \tag{2}$$

where (see previous slides on asymptotics):

- $\Omega = \frac{1}{\mathcal{I}_1(\theta)}$ in the case of ML;

Therefore, under the null,

$$W = \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\Omega}} \quad \xrightarrow{d} \quad \text{Normal}(0, 1) \tag{3}$$

or, taking squares

$$W^2 = \frac{n(\hat{\theta}_n - \theta)^2}{\Omega} \quad \xrightarrow{d} \quad \chi^2_1 \tag{4}$$

The logic of the Wald test in detail (comments)

- Given a random sample, estimates of $\sqrt{\Omega}$ must be used. :

$$\hat{\Omega} = \frac{1}{\hat{\mathcal{I}}_n(\hat{\theta}^{ML})} = \frac{n}{\sum_{i=1}^{n} l_{\theta\theta}(X, \hat{\theta}^{ML})} \tag{5}$$

- The null $H_0 : \theta = \theta_0$ is rejected at a significance s when:

$$\hat{W}^2 > \chi^2_{1,s} \tag{6}$$

where $\chi^2_{1,s}$ is the critical value c of the χ^2_1 such that

$$Pr(W^2 > c|H_0) = s \qquad \text{with} \qquad W^2 \sim \chi^2_1 \tag{7}$$

The p-value is $p = Pr(\chi^2_1 > \hat{W}^2)$

Subsection 2

The Lagrance Multiplier (or Score) test

The logic of the LM test in detail

It can be shown that under the null:

$$\sqrt{n}(l_\theta(X,\theta)) \quad \xrightarrow{d} \quad \text{Normal}(0, \frac{1}{\mathcal{I}_1(\theta)}) \tag{8}$$

Intuitively, $l_\theta(X,\theta)$ is the average of iid random variables (the ln of the derivatives of the pdf of each observation) with mean zero and variance $\frac{1}{\mathcal{I}_1(\theta)}$

The logic of the LM test in detail (cont.)

Therefore, under the null,

$$LM = \frac{\sqrt{n}(l_\theta(X,\theta_0)}{\sqrt{\frac{1}{\mathcal{I}_1(\theta_0)}}} \quad \xrightarrow{d} \quad \text{Normal}(0,1) \tag{9}$$

or, taking squares

$$LM^2 = \frac{n(l_\theta(X,\theta_0)^2}{\frac{1}{\mathcal{I}_1(\theta_0)}} \quad \xrightarrow{d} \quad \chi^2_1 \tag{10}$$

Note that LM test statistics can be computed without having to find the ML estimate.

The name of the test comes from the fact that implicitly we are maximizing the likelihood subject to the constraint $\theta = \theta_0$

$$\max L(X,\theta) + \lambda(\theta - \theta_0) \quad \Rightarrow \quad \lambda = L_\theta \tag{11}$$

The logic of the LM test in detail (comments)

- Estimates of the Information at θ_0 must be used:

$$\frac{1}{\hat{\mathcal{I}}_n(\theta_0)} = -\frac{n}{\sum_{i=1}^{n} l_{\theta\theta}(X, \theta_0)} \tag{12}$$

- The null $H_0 : \theta = \theta_0$ is rejected at a significance s when:

$$\hat{L}M^2 > \chi^2_{1,s} \tag{13}$$

where $\chi^2_{1,s}$ is the critical value c of the χ^2_1 such that

$$Pr(LM^2 > c|H_0) = s \qquad \text{with} \qquad LM^2 \sim \chi^2_1 \tag{14}$$

The p-value is $p = Pr(\chi^2_1 > \hat{L}M^2)$

Subsection 3

The Likelihood Ratio test

The logic of the LR test in detail

Under $H_0 : \theta = \theta_0$, expand the log likelihood $l(\theta_0)$ around $\hat{\theta}^{ML}$

$$l(\theta_0) = l(\hat{\theta}^{ML}) + l_\theta(\hat{\theta}^{ML})(\theta_0 - \hat{\theta}^{ML}) - \frac{1}{2}(\hat{\theta}^{ML} - \theta_0)^2 l_{\theta\theta}(\tilde{\theta}) \qquad (15)$$

where $\tilde{\theta}$ is some intermediate point between $\hat{\theta}^{ML}$ and θ_0, and

$$-\frac{1}{N} l_{\theta\theta}(\tilde{\theta}) \quad \xrightarrow{p} \quad \mathcal{I}_1(\theta) \qquad (16)$$

Note that $l_\theta(\hat{\theta}^{ML}) = 0$ and therefore we can write:

$$LR = 2(l(\theta_0) - l(\hat{\theta}^{ML})) = \frac{(\sqrt{N}(\hat{\theta}^{ML} - \theta_0))^2}{(\mathcal{I}_1(\theta))^{-1}} \sim \chi^2_1 \qquad (17)$$

where the LHS is the Likelihood Ratio Test

The logic of the LR test in detail (comments)

- The null $H_0 : \theta = \theta_0$ is rejected at a significance s when:

$$\widehat{LR}^2 > \chi^2_{1,s} \tag{18}$$

where $\chi^2_{1,s}$ is the critical value c of the χ^2_1 such that

$$Pr(LR^2 > c|H_0) = s \qquad \text{with} \qquad LR^2 \sim \chi^2_1 \tag{19}$$

The p-value is $p = Pr(\chi^2_1 > \widehat{LR}^2)$

Section 3

Large sample testing and regression

A general formulation of an hypothesis concerning β

Given the PRF

$$Y = X\beta + U \tag{20}$$

let's now consider the most general formulation of an hypothesis concerning β:

$$H_0 : r(\beta) = q \qquad \text{against} \qquad H_1 : r(\beta) \neq q \tag{21}$$

where $r(.)$ is any function of the parameters and $r(\beta) - q$ is a $\rho \times 1$ vector, if ρ is the number of restrictions.

So H_0 and H_1 are systems of ρ equations if there are ρ restrictions.

An example

Example :

$$H_0 : r(\beta) = R\beta = q \qquad \text{against} \qquad H_1 : r(\beta) = R\beta \neq q \tag{22}$$

where R is a $\rho \times k + 1$ matrix which charaterize the ρ restrictions on the parameters that we would like to test.

Exercise on the specification of a set of restrictions

Suppose that you are estimating the log of a Cobb Douglas production function in which output depends on labor and capital and you want to test:

- constant returns to scale;
- the return to one unit of labor is twice the return to one unit of capital;
- there exist neutral technological progress/regress.

What is R for these restrictions?

Subsection 1

Wald test in the context of regression

The logic of the Wald test

If the restrictions are valid the quantity $r(\hat{\beta}) - q$ should be close to 0 while otherwise it should be far away from 0.

The Wald form of the test statistic that captures this logic is

$$W = [r(\hat{\beta}) - q]'[Var(r(\hat{\beta}) - q)]^{-1}[r(\hat{\beta}) - q] \tag{23}$$

In other words we want to evaluate how far away from 0 is $r(\hat{\beta}) - q$ after normalizing it by its average variability.

Note that W is a scalar.

Implementation of the test

If $r(\hat{\beta}) - q$ is normally distributed, under H_0

$$W \sim \chi^2_\rho \tag{24}$$

where the number of degrees of freedom ρ is the number of restrictions to be tested.

The difficulty in computing the test statistics is how to determine the variance at the denominator.

The Variance of the Wald Test Statistic

Using the Delta Method in a setting in which $h(\hat{\beta}) = r(\hat{\beta}) - q$

$$Var[r(\hat{\beta}) - q] = \left[\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}}\right] [Var(\hat{\beta})] \left[\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}}\right]' \tag{25}$$

where note that $\left[\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}}\right]$ is a $\rho \times k + 1$ matrix and therefore $Var[r(\hat{\beta}) - q]$ is a $\rho \times \rho$ matrix.

Back to the example

Going back to the example in which $r(\hat{\beta}) - q = R\hat{\beta} - q$

$$Var[R\hat{\beta} - q] = R[Var(\hat{\beta})]R' \tag{26}$$

and the Wald test is

$$W = [R\hat{\beta} - q]'[RVar(\hat{\beta})R']^{-1}[R\hat{\beta} - q] \tag{27}$$

and $Var(\hat{\beta})$ is in practice estimated by substituting the sample counterparts of the asymptotic variance-covariance matrice

Exercise: Wald test and simple restrictions

Consider again the unrestricted regression in matrix form

$$Y = X_1\beta_1 + X_2\beta_2 + U_{ur} \tag{28}$$

where

- X_1 is a $n \times 2$ matrix including the constant;
- β_1 is dimension 2 vector of parameters;
- X_2 is a $n \times 1$ matrix;
- β_2 is dimension 1 vector of parameters;

and suppose that we want to test the following joint hypothesis on the β_2 parameters:

$$H_0 : \beta_2 = 0 \qquad \text{against} \qquad H_1 : \beta_2 \neq 0 \tag{29}$$

What is R in this case?

Exercise: Wald test and simple restriction (cont.)

It is easy to verify that in this case the Wald test is

$$
\begin{aligned}
W &= [R\hat{\beta} - q]'[RVar(\hat{\beta})R']^{-1}[R\hat{\beta} - q] \qquad (30)\\
&= \frac{\hat{\beta}_2^2}{Var(\hat{\beta}_2)}
\end{aligned}
$$

which is the square of a standard t-test, and is distributed as a χ^2 distribution

A drawback of the Wald Test

The Wald test is a general form of a large sample test that requires the estimation of the unrestricted model.

There are cases in which this may be difficult or even impossible.

An alternative large sample testing procedure is the Lagrange Multiplier test, which requires instead only the estimation of the restricted model.

A third alternative is the Likelihood Ratio test, which requires the estimation of both the restricted and the unrestricted models, on an equal basis.

Subsection 2

The Lagrange multiplier test in the context of linear regression

The logic of the test

In the simple context of linear regression we can define a LM test for multiple exclusion restrictions. Consider again the unrestricted regression in matrix form

$$Y = X_1\beta_1 + X_2\beta_2 + U_{ur} \tag{31}$$

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

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

$$H_0 : \beta_2 = 0 \qquad \text{against} \qquad H_1 : \beta_2 \neq 0 \tag{32}$$

The logic of the test (cont.)

Suppose that you estimate the restricted PRF

$$Y = X_1\beta_{r1} + U_r \tag{33}$$

where the subscript r indicates that the population parameters and unobservables of this restricted equation may differ from the corresponding one of the unrestricted PRF.

It is intuitive to hypothesize that in the auxiliary regression

$$\hat{U}_r = X_1\gamma_1 + X_2\gamma_2 + V \tag{34}$$

if the restrictions in the primary PRF are valid then

$$H_0 : \beta_2 = 0 \qquad \Rightarrow \qquad \gamma_2 = 0 \tag{35}$$

The logic of the test (cont.)

Let the R-squared of the auxiliary regression 34 be R_U^2 and consider the statistics

$$LM = nR_U^2 \tag{36}$$

If the restrictions are satisfied, this statistics should be close to zero because;

- X_1 is by construction orthogonal to U_R and therefore $\gamma_1 = 0$;
- and $\gamma_2 = 0$ if the restrictions are satisfied.

Since, given k_2 exclusion restrictions:

$$LM = nR_U^2 \sim \chi^2_{k_2} \tag{37}$$

we can use the Classical testing procedure to test H_0.

Subsection 3

The Likelihood Ratio Test in the context of regression

Derivation of the LR test

Continuing with the regression framework used to discuss the LM test, the LR test is derived as follows:

- Get the ML estimate of the unrestricted regression and retrieve the unrestricted normalized log likelihood l_U
- Get the ML estimate of the restricted regression and retrieve the restricted normalized log likelihood l_R
- Then given k_2 the test statisic is

$$LR = 2(l_R - l_U) \sim \chi^2_{k_2} \tag{38}$$

Subsection 4

Final comment on the “trinity of tests”

Relationship between the three tests

From the Handbook Chapter 13 by Robert Engle

“ These three general principles have a certain symmetry which has revolutionized the teaching of hypothesis tests and the development of new procedures

Essentially, the Lagrance Multiplier approach starts at the null and asks whether movement toward the alternative would be an improvement ...

... while the Wald approace starts at the alternative and considers movement toward the null

The Likelihood ratio method compares the two hypothesis directly on an equal basis.

Figure 3.1 in Engle’s chapter is a useful way to think at the three tests.