APEC 8212 Econometric Analysis II: Recitation 6

March 2, 2012

1 Basic intuitions of the classical hypothesis tests

Consider endogenous random scalar y and exogenous random scalar x such that the density of y is generated by the following model f

$$\mathrm{p}(y \,|\, x) = f(y \,|\, x;\, \theta_0)$$

where $\theta_0 \in \Theta \subset \mathbb{R}^K$. For any $\theta \in \Theta$ and $\theta \neq \theta_0$, it's true that

$$\begin{aligned}
\int_y \log\left(\frac{\mathrm{p}(y \,|\, x)}{f(y \,|\, x;\, \theta)}\right) \mathrm{p}(y \,|\, x)\mathrm{d}y &= \mathrm{E}_y\left[\log\left(\mathrm{p}(y \,|\, x)\right) \,|\, x\right] - \mathrm{E}_y\left[\log\left(f(y \,|\, x;\, \theta)\right) \,|\, x\right] \\
&= \mathrm{E}_y\left[\log\left(f(y \,|\, x;\, \theta_0)\right) \,|\, x\right] - \mathrm{E}_y\left[\log\left(f(y \,|\, x;\, \theta)\right) \,|\, x\right] \\
&\geq 0
\end{aligned}$$

Thus the true value θ_0 maximizes $\mathrm{E}_y\left[\log\left(f(y \,|\, x;\, \theta)\right) \,|\, x\right]$.

Given the data $\{y_i,\, x_i\}$, define the log likelihood function as

$$\mathrm{L}(\theta) = \sum_{i=1}^{N} \mathrm{L}_i(\theta) = \sum_{i=1}^{N} \log f(y_i \,|\, x_i;\, \theta)$$

the maximum likelihood estimator (MLE) is estimated by solving the following problem

$$\max_{\theta \in \Theta} \mathrm{L}(\theta) \tag{1}$$

Define the score of the log likelihood for observation i as

$$\mathrm{s}_i(\theta) = \nabla \mathrm{L}_i(\theta)$$

and the negative Hessian of the log likelihood of observation i as

$$\Pi_i(\theta) = -\nabla \mathrm{s}_i(\theta)$$

The basic idea of the hypothesis test is given the null

$$\mathrm{H}_0 :\, \theta_0 = 0 \tag{2}$$

we want to know if we could reject, or fail to reject H_0, based on the data estimator θ. This is done by comparing the difference between θ and θ_0. There are three classical approaches to do this: Wald, Score (Lagrangian Multiplier or LM), and Likelihood Ratio (LR). The basic idea is the same, but each of them uses a special measure of the difference between θ and θ_0. Specifically, the measure of difference for Wald test is defined directly on θ, for Score/LM test is defined on the score $\{\mathrm{s}_i(\theta)\}$, and for the LR test is defined on the log likelihood function $\mathrm{L}(\theta)$.

2 More details

2.1 Wald test

Wald test is based on the estimation of the unconstrained model, i.e., (1) is solved without any constraint. Let the solution be $\hat{\theta}$. By the first order Taylor expansion

$$\sum_{i=1}^{N} \mathrm{s}_i(\hat{\theta}) \approx \sum_{i=1}^{N} \mathrm{s}_i(\theta_0) - \left(\sum_{i=1}^{N} \Pi_i(\theta_0)\right)(\hat{\theta} - \theta_0) \tag{3}$$

Because of FOC of (1) implies

$$\nabla \mathrm{L}(\hat{\theta}) = \sum_{i=1}^{N} \mathrm{s}_i(\hat{\theta}) = 0 \tag{4}$$

Thus (3) can be rewritten as

$$\sqrt{N}(\hat{\theta} - \theta_0) = \left(N^{-1}\sum_{i=1}^{N} \Pi_i(\theta_0)\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N} \mathrm{s}_i(\theta_0)\right) = (\mathrm{E}\Pi_i(\theta_0))^{-1}\left(N^{-1/2}\sum_{i=1}^{N} \mathrm{s}_i(\theta_0)\right)$$

Wooldridge p478-479 (13.5.1) shows that

$$\mathrm{Var}\left(\mathrm{s}_i(\theta_0)\right) = \mathrm{E}\left(\mathrm{s}_i(\theta_0)\mathrm{s}_i(\theta_0)'\right) = \mathrm{E}\Pi_i(\theta_0) \tag{5}$$

Then by the central limit theorem

$$\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} \mathrm{N}(0, (\mathrm{E}\Pi_i(\theta_0))^{-1}) \tag{6}$$

By Wooldridge Lemma 3.7 (p41), we can construct the Wald statistic as

$$\xi_W = N(\hat{\theta} - \theta_0)' \left(\mathrm{E}\Pi(\theta_0)\right)(\hat{\theta} - \theta_0) \sim \chi^2_K$$

Therefore, in order to do the Wald test, we need to estimate the unconstrained model (1) to estimate $\hat{\theta}$, then

$$\xi_W = N\hat{\theta}' \left(\mathrm{E}\Pi(\hat{\theta})\right)\hat{\theta} \sim \chi^2_K$$

2.2 Score/LM test

Score/LM test is based on the estimation of the constrained model

$$\max_{\theta \in \Theta} \sum_{i=1}^{N} \mathrm{L}_i(\theta)$$

subject to

$$\theta = \theta_0 \,(= 0)$$

Let λ be the Lagrangian multiplier on the constraint, and $\tilde{\theta}$ the solution, we have FOC as

$$\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta}) = \lambda \tag{7}$$

If the constraint is not binding, i.e., the null hypothesis is true, then $\lambda = 0$. Then we can test by null by testing if λ is statistically significantly different from 0. Because by (4) $\sum_{i=1}^{N} \mathrm{s}_i(\hat{\theta}) = 0$, we are equivalently testing the difference between $\tilde{\theta}$ and $\hat{\theta}$ measured by the score.

Similar as (3) we have

$$\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta}) \approx \sum_{i=1}^{N} \mathrm{s}_i(\theta_0) - \left(\sum_{i=1}^{N} \Pi_i(\theta_0)\right)(\tilde{\theta} - \theta_0)$$

Then use (3) and (4) we have

$$N^{-1/2}\lambda \approx \left(N^{-1}\sum_{i=1}^{N} \Pi_i(\theta_0)\right)\sqrt{N}(\hat{\theta} - \tilde{\theta}) = (\mathrm{E}\Pi(\theta_0))\sqrt{N}(\hat{\theta} - \tilde{\theta}) \tag{8}$$

Because it's constrained that $\tilde{\theta} = \theta_0$, by (6) we have

$$\sqrt{N}(\hat{\theta} - \tilde{\theta}) \overset{a}{\sim} \mathrm{N}(0,\, (\mathrm{E}\Pi(\theta_0))^{-1}) \tag{9}$$

Then (8) implies

$$N^{-1/2}\lambda \overset{a}{\sim} \mathrm{N}(0,\, \mathrm{E}\Pi(\theta_0))$$

The Score/LM statistic is constructed as

$$\xi_{LM} = N^{-1}\lambda' \left(\mathrm{E}\Pi(\theta_0)\right)^{-1} \lambda \sim \chi^2_K$$

According to (4) and (7), ξ_{LM} is the quadratic form of the Lagrangian multiplier λ, or the difference between the scores evaluated at $\tilde{\theta}$ and $\hat{\theta}$. This is where its name comes from.

Notice by (5)

$$\begin{aligned} \xi_{LM} &= N^{-1}\lambda' \left(\mathrm{E}\Pi(\theta_0)\right)^{-1} \lambda \\ &= N^{-1}\lambda' \left(\mathrm{E}\left(\mathrm{s}(\theta_0)\mathrm{s}(\theta_0)'\right)\right)^{-1} \lambda \\ &\approx \left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})'\right)\left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})\mathrm{s}_i(\tilde{\theta})'\right)^{-1}\left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})\right) \end{aligned}$$

Consider the linear projection of 1 on $\mathrm{s}_i(\tilde{\theta})$, then the uncentered R_u^2 is given by

$$R_u^2 = \frac{\mathbf{1}'M_s\mathbf{1}}{\mathbf{1}'\mathbf{1}}$$

where $\mathbf{1}$ is the $N \times 1$ column vector of 1 and M_s is the projection matrix defined on $\boldsymbol{s}(\tilde{\theta})$ which is the $N \times K$ matrix where the ith row is $\mathrm{s}_i(\tilde{\theta})'$, then

$$R_u^2 = \frac{\left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})'\right)\left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})\mathrm{s}_i(\tilde{\theta})'\right)^{-1}\left(\sum_{i=1}^{N} \mathrm{s}_i(\tilde{\theta})\right)}{N}$$

thus

$$\xi_{LM} = NR_u^2 \sim \chi^2_K$$

Therefore, to do the Score/LM test, we need to estimate the constrained model to calculate the score $\mathrm{s}_i(\tilde{\theta})$, then the Score/LM statistic is the uncentered R_u^2 from regressing 1 on $\mathrm{s}_i(\tilde{\theta})$.

2.3 LR test

If both the unconstrained and constrained model are estimated, the LR statistic is simply

$$\xi_{LR} = 2(\mathrm{L}(\hat{\theta}) - \mathrm{L}(\tilde{\theta})) \sim \chi^2_K$$

It is very convenient to be computed.

Consider the second order Taylor expansion of the log likelihood function evaluated at $\hat{\theta}$ and $\tilde{\theta}$

$$\mathrm{L}(\hat{\theta}) \approx \mathrm{L}(\theta_0) + (\hat{\theta} - \theta_0)' \sum_{i=1}^{N} \mathrm{s}_i(\theta_0) - \frac{1}{2}(\hat{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\hat{\theta} - \theta_0)$$

$$\mathrm{L}(\tilde{\theta}) \approx \mathrm{L}(\theta_0) + (\tilde{\theta} - \theta_0)' \sum_{i=1}^{N} \mathrm{s}_i(\theta_0) - \frac{1}{2}(\tilde{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\tilde{\theta} - \theta_0)$$

Then

$$\xi_{LR} = 2(\hat{\theta} - \tilde{\theta})' \left(\sum_{i=1}^{N} \mathrm{s}_i(\theta_0) \right) - (\hat{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\hat{\theta} - \theta_0) + (\tilde{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\tilde{\theta} - \theta_0)$$

By (3) and (4)

$$\sum_{i=1}^{N} \mathrm{s}_i(\theta_0) = \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\hat{\theta} - \theta_0)$$

Then

$$\begin{aligned}
\xi_{LR} &= 2(\hat{\theta} - \tilde{\theta})' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\hat{\theta} - \theta_0) - (\hat{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\hat{\theta} - \theta_0) + (\tilde{\theta} - \theta_0)' \left(\sum_{i=1}^{N} \Pi_i(\theta_0) \right) (\tilde{\theta} - \theta_0) \\
&= \sqrt{N}(\hat{\theta} - \tilde{\theta})' \left(N^{-1} \sum_{i=1}^{N} \Pi_i(\theta_0) \right) \sqrt{N}(\hat{\theta} - \tilde{\theta}) \\
&= \sqrt{N}(\hat{\theta} - \tilde{\theta})' \left(\mathrm{E}\Pi(\theta_0) \right) \sqrt{N}(\hat{\theta} - \tilde{\theta})
\end{aligned}$$

Then by (9)

$$\xi_{LR} \sim \chi^2_K$$

3 Null hypothesis in the general form

Now consider the null hypothesis as

$$\mathrm{H}_0 : g(\theta_0) = 0$$

The first order Taylor expansion of g is

$$g(\theta) \approx g(\theta_0) + \nabla g(\theta_0)(\theta - \theta_0)$$

then

$$\sqrt{N} g(\theta) \approx \nabla g(\theta_0) \sqrt{N}(\theta - \theta_0)$$

Therefore, the variance-covariance of g evaluated at any consistent estimator θ can be estimated consistently from the variance-covariance of $\theta - \theta_0$, which we already discussed in the last section.

If the model is linear, we can always estimate the unconstrained model and use Wald test. Usually, it's not easy to impose the constraint if it's nonlinear (no matter the model is linear or nonlinear). In other words, the unconstrained model is easier to estimate. Thus Wald test is prefered when the constraint is nonlinear. If the constrained model is also easy to estimate, LR test would be prefered because it's convenient to compute the LR statistic. Score/LM test would be prefered when the constrained model is easy to estimate, but the unconstrained model is hard to estimate.

For more details and some interesting examples, read Engle 1984, "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", *Handbook of Econometrics*, Volume II, Chapter 13.