TA section 3

April 24th, 2009 by Yang

1. Test

What is a test? Before asking this, why do we estimate an econometric model?

We want to find what the true parameter is. Uncertainty (or deficiency of models and data) makes us decide with probability less than 1. There are two ways to do this.

1. *Test*: Set up a hypothesis on the parameter, and decide whether the hypothesis seems correct.
2. *Confidence Region* (Confidence Interval): Report the set of parameter values that are highly likely to be correct.

Here we focus on a test, while confidence region will be discussed later in the lecture.

2. Size and Power of the test

Let Θ be the parameter space. In many cases, we set up the null hypothesis $H_0 : \theta = \theta_0$ where $\theta_0 \in \Theta$. Consider the alternative hypothesis $H_1 : \theta = \theta_1$ where $\theta_1 \in \Theta$ and $\theta_1 \neq \theta_0$. A test is a criterion on when we accept/reject H_0. The size of the test is defined as follows.

$$\begin{aligned} \text{SIZE} &= \Pr(\text{the test rejects } H_0 | H_0 \text{ is true}) \\ &= \Pr(\text{the test rejects } H_0 | \theta = \theta_0) \end{aligned}$$

The power of the test is defined as follows.

$$\begin{aligned} \text{POWER} &= \Pr(\text{the test rejects } H_0 | H_0 \text{ is false}) \\ &= \Pr(\text{the test rejects } H_0 | H_1 \text{ is true}) \\ &= \Pr(\text{the test rejects } H_0 | \theta = \theta_1) \end{aligned}$$

We have to make the size as small as possible, and at the same time the power as high as possible. We care more about the size, so control the size at a fixed level. Unfortunately, the size may not be 0. Fixing the size, we want to maximize the power.

In some cases, $H_0 : \theta \in \Theta_0$ or $H_1 : \theta \in \Theta_1$ or both, where Θ_0 and Θ_1 are not singleton. In those cases, we define

$$\begin{aligned} \text{SIZE} &= \sup_{\theta_0 \in \Theta_0} \Pr(\text{the test rejects } H_0 | \theta = \theta_0) \\ \text{Power function } \gamma(\theta) &= \Pr(\text{the test rejects } H_0 | \theta) \end{aligned}$$

3. Wald, LM, and LR tests

Consider a null hypothesis $H_0 : \theta = \theta_0$ where $\theta_0 \in \Theta$. There are three test statistics based on χ^2 distribution.

1. The *Wald* statistic looks at $\widehat{\theta}_n - \theta_0$.
2. The *LM* (Langrange Multiplier) statistic looks at $\frac{\partial}{\partial\theta} Q_n(\theta_0)$.
3. The *LR* (Likelihood Ratio) statistic looks at $Q_n(\theta_0) - Q_n(\widehat{\theta}_n)$.

Under appropriate conditions, we have shown that if θ_0 is a true parameter,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, B_0^{-1}\Omega_0 B_0^{-1}\right)$$
$$\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0) \xrightarrow{d} N(0, \Omega_0)$$

Recall that when $X \sim N(0, \Sigma)$, we have $AX \sim N(0, A\Sigma A')$. Therefore,

$$\widehat{\Omega}_n^{-1/2}\widehat{B}_n\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, I_d\right)$$
$$\widetilde{\Omega}_n^{-1/2}\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0) \xrightarrow{d} N(0, I_d)$$

where $\widehat{\Omega}_n \xrightarrow{p} \Omega_0$, $\widehat{B}_n \xrightarrow{p} B_0$, $\widetilde{\Omega}_n \xrightarrow{p} \Omega_0$ and d denotes the dimension of θ. Recall that when $X \sim N(0, I_k)$, we have $X'X \sim \chi^2_k$ which is a χ^2 distribution with k degrees of freedom. So

$$n(\widehat{\theta}_n - \theta_0)'\widehat{B}_n\widehat{\Omega}_n^{-1}\widehat{B}_n(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_d$$
$$n\frac{\partial}{\partial\theta'} Q_n(\theta_0)\widetilde{\Omega}_n^{-1}\frac{\partial}{\partial\theta} Q_n(\theta_0) \xrightarrow{d} \chi^2_d$$

The first is called the *Wald* statistic, and the second is called the *LM* statistic.[1] Here are some remarks. First, the Wald statistic can be constructed using the estimate of the asymptotic variance. After estimating $\widehat{\theta}_n$ using the model, $\widehat{\Omega}_n$ and $\widehat{B}_n$ can be estimated using $\widehat{\theta}_n$. Using these estimates, we can calculate the Wald statistic. Second, the LM statistic can be obtained without estimation of $\widehat{\theta}_n$. We calculate $\widetilde{\Omega}_n$ using θ_0. Also, $\frac{\partial}{\partial\theta} Q_n(\cdot)$ is evaluated at θ_0. Third, we obtained the above results assuming that θ_0 is a true parameter. If this does not hold, those statistics may not follow a χ^2_d distribution. They are expected to diverge to infinity when H_0 is false. Based on the argument made in the previous section, we control the size at α, and maximize the power. The way to do this is to reject if the statistics are greater than $\chi^2_{d,1-\alpha}$, and to accept otherwise, where $\chi^2_{d,1-\alpha}$ denotes the $1-\alpha$ quantile of the χ^2_d distribution.

[1] We can show that under an additional assumption,

$$2n(Q_n(\theta_0) - Q_n(\widehat{\theta}_n)) = \widehat{c}_n LM + o_p(1)$$

for some $\widehat{c}_n$. So the LHS divided by $\widehat{c}_n$ is called the *LR* statistic.

Before moving on to linear IV example, let us see how we can extend the Wald statistic to the nonlinear hypothesis case. Define $H_0 : h(\theta) = 0$, which means that $H_0 : \theta \in \Theta_0 \equiv \{\theta : h(\theta) = 0\}$. The Wald statistic now looks at $h(\widehat{\theta}_n) - h(\theta_0)$ for any $\theta_0 \in \Theta_0$. By Δ-method,

$$\sqrt{n}(h(\widehat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \frac{\partial}{\partial \theta'} h(\theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0)$$

Define $H := \frac{\partial}{\partial \theta'} h(\theta_0)$, and use $h(\theta_0) = 0$, then if θ_0 is a true parameter,

$$\sqrt{n} h(\widehat{\theta}_n) \xrightarrow{d} N\left(0, HB_0^{-1}\Omega_0 B_0^{-1}H'\right)$$

Repeat the similar procedure to get

$$nh(\widehat{\theta}_n)'\left(\widehat{H}\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}'\right)^{-1} h(\widehat{\theta}_n) \xrightarrow{d} \chi_r^2$$

where $\widehat{H} \xrightarrow{p} H$, $\widehat{\Omega}_n \xrightarrow{p} \Omega_0$, $\widehat{B}_n \xrightarrow{p} B_0$, and r denotes the dimension of $h(\cdot)$. The LM statistic generalized to the nonlinear hypothesis case is more complicated, but the intuition is similar. Now we have to estimate $\widetilde{\theta}_n$ to obtain the LM statistic, which is a difference from the original case. But as noted in the lecture note, it sometimes makes the function simpler so that it is easier to estimate compared to $\widehat{\theta}_n$.

4. Linear IV example

Consider again linear IV model. Assume conditional homoskedasticity, so that 2SLS estimator is the most efficient. Suppose $H_0 : \beta = \beta_0$. Under H_0,

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N\left(0, \sigma^2\left[Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'\right]^{-1}\right)$$

Recall that the asymptotic variance can be estimated by

$$\widehat{\sigma}^2\left[\frac{1}{n}\sum_{i=1}^n x_iz_i'\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n z_ix_i'\right]^{-1}$$

where $\widehat{\sigma}^2$ is a consistent estimate of σ^2 with

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - x_i'\widehat{\beta}_n)^2$$

So in a similar way as in the previous section,

$$n(\widehat{\beta}_n - \beta_0)'\left(\widehat{\sigma}^2\left[\frac{1}{n}\sum_{i=1}^n x_iz_i'\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n z_ix_i'\right]^{-1}\right)^{-1}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} \chi_d^2$$

or

$$\frac{n}{\widehat{\sigma}^2}(\widehat{\beta}_n - \beta_0)'\left(\frac{1}{n}\sum_{i=1}^n x_iz_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_ix_i'\right)(\widehat{\beta}_n - \beta_0) \xrightarrow{d} \chi_d^2$$

Using matrix notation and cancelling n out, we have

$$\frac{(\widehat{\beta}_n-\beta_0)'X'Z(Z'Z)^{-1}Z'X(\widehat{\beta}_n-\beta_0)}{\widehat{\sigma}_n^2}\xrightarrow{d}\chi_d^2$$

or equivalently,

$$\frac{(\widehat{\beta}_n-\beta_0)'X'P_zX(\widehat{\beta}_n-\beta_0)}{\widehat{\sigma}_n^2}\xrightarrow{d}\chi_d^2$$

The LHS is defined as the Wald statistic. This is easy to calculate. Recall that this holds when β_0 is the true parameter value, while the Wald statistic goes to infinity if H_0 is false. So after calculating the Wald statistic, we reject H_0 if it is greater than $\chi^2_{d,1-\alpha}$, and accept otherwise. As we have discussed, this maximizes the power of the Wald test, controlling the size of the test at α.

Let us turn to the LM statistic in this example. Recall that the stochastic criterion function was defined as

$$Q_n(\beta):=\frac{1}{2}\left\|A_n\frac{1}{n}\sum_{i=1}^n z_i(y_i-x_i'\beta)\right\|^2$$

To make simple, use matrix notation and $A_n'A_n=(Z'Z/n)^{-1}$, then this can be written as

$$Q_n(\beta)=\frac{1}{2n}(Y-X\beta)'Z(Z'Z)^{-1}Z'(Y-X\beta)$$

Since

$$\frac{\partial}{\partial\beta}Q_n(\beta)=-\frac{1}{n}X'Z(Z'Z)^{-1}Z'(Y-X\beta)$$

we use β_0 to construct

$$\frac{\partial}{\partial\beta}Q_n(\beta_0)=-\frac{1}{n}X'Z(Z'Z)^{-1}Z'(Y-X\beta_0)$$

Also we know that

$$\Omega_0=\sigma^2Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'$$

and thus

$$\widetilde{\Omega}_n=\widetilde{\sigma}^2\frac{1}{n}\sum_{i=1}^n x_iz_i'\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n z_ix_i'=\frac{\widetilde{\sigma}^2}{n}X'Z(Z'Z)^{-1}Z'X$$

where $\widetilde{\sigma}^2=\frac{1}{n}(y_i-x_i'\beta_0)^2$ is constructed using β_0. Therefore, the LM statistic would be

$$\begin{aligned}&n\frac{\partial}{\partial\beta'}Q_n(\beta_0)\widetilde{\Omega}_n^{-1}\frac{\partial}{\partial\beta}Q_n(\beta_0)\\&=\frac{(Y-X\beta_0)'Z(Z'Z)^{-1}Z'X\left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'(Y-X\beta_0)}{\widetilde{\sigma}^2}\xrightarrow{d}\chi_d^2\end{aligned}$$

where the convergence takes place when H_0 is true. Note again that this can be calculated directly from the data and the null hypothesis, without estimating $\widehat{\beta}_n$. A criterion on rejection and acceptance would be determined similarly as in the Wald statistic case.