Economics 522A, Spring 2007

Lecture Note 24: Hypothesis Testing with Generalized Method of Moments

(Reading: Ruud Ch. 22)

1 Testing Restrictions on Parameters

We will continue to use the notation of LN23. Let us consider a hypothesis on the parameters θ, captured by a function $r(\theta)$. We will assume that r is j dimensional, where $j \leq k = dim(\theta)$. Assume that r is twice continuously differentiable, and that its partial derivative matrix

$$R(\theta) = \nabla_\theta r(\theta)$$

has rank j. The null hypothesis is:

$$H_0 : \quad r(\theta_0) = 0.$$

Recall that the GMM estimator $\hat{\theta}$ solves

$$\min_\theta Q_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^n g(w_i, \theta)\right]' B_n \left[\frac{1}{n}\sum_{i=1}^n g(w_i, \theta)\right].$$

1.1 Wald Test

By the delta method,

$$\sqrt{n}(r(\hat{\theta}) - r(\theta)) \xrightarrow{\text{d}} N(0, R(\theta_0)VR(\theta_0)'),$$

where V is the asymptotic variance matrix of $\hat{\theta}$.

Let $\hat{V}_n \xrightarrow{\text{p}} V$, so we have a consistent estimator of V, and let $\hat{R}_n \xrightarrow{\text{p}} R(\theta_0)$, so we have a consistent estimator of $R(\theta_0)$. (Usually, we can use $\hat{R}_n = R(\hat{\theta})$.)

Then we can form a Wald statistic

$$W = nr(\hat{\theta})' \left[\hat{R}_n' \hat{V}_n \hat{R}_n\right]^{-1} r(\hat{\theta}).$$

Under H_0,

$$W \xrightarrow{\text{d}} \chi^2_j.$$

1.2 LM Test

The original GMM estimator $\hat{\theta}$ does not impose the restriction that $r(\hat{\theta}) = 0$. Let us also consider the restricted estimator:

$$\hat{\theta}_R = \arg \min_{\theta : r(\theta)=0} Q_n(\theta).$$

Conisder the gradient of the objective function:

$$\nabla_\theta Q_n(\theta) = 2\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g(w_i,\theta)\right]' B_n \left[\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta)\right] = 2\cdot \hat{G}_n(\theta)' B_n \hat{g}_n(\theta).$$

If the null hypothesis is correct, we would expect that

$$\nabla_\theta Q_n(\hat{\theta}_R) \approx 0.$$

The logic is much the same as the "score" or LM test in the case of maximum likelihood: see how close the gradient of the objective function is to 0, when evaluated at the restricted estimate:

$$LM = n\cdot \hat{g}_n(\hat{\theta}_R)' B_n \hat{G}_n(\hat{\theta}_R)\left[\hat{G}_n(\hat{\theta}_R)' B_n \hat{G}_n(\hat{\theta}_R)\right]^{-1}\hat{G}_n(\hat{\theta}_R)' B_n \hat{g}_n(\hat{\theta}_R).$$

This has the same asymptotic distribution as the Wald test.

1.3 Test Based on the Objective Function

There is also an analog to the likelihood ratio. The idea is simply to compare the objective function under the unrestricted and the restricted estimates. The "distance difference" test statistic has the form:

$$DD = n\cdot\left[Q_n(\hat{\theta}_R) - Q_n(\hat{\theta})\right].$$

Again this has the same asymptotic distribution as the Wald and LM tests.

1.4 Alternative Wald Test

We can formulate a Wald-type test in a different manner, by simply comparing the unrestricted and restricted estimates of θ:

$$W_2 = n\cdot(\hat{\theta} - \hat{\theta}_R)' \hat{V}_n^{-1}(\hat{\theta} - \hat{\theta}_R).$$

This can be shown to have the same asymptotic distribution as the other tests.

1.5 Some Handy Equivalences

Suppose that θ is just-identified, so that the dimension of g is equal to the dimension of θ ($m = k$), and it is possible to set

$$\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta) = 0.$$

Then the GMM objective function attains a minimum at zero:

$$Q_n(\hat{\theta}) = 0.$$

So the DD test statistic becomes simply:

$$DD = n\left[Q_n(\hat{\theta}_R)\right].$$

In addition, the gradient term

$$\hat{G}_n(\hat{\theta}_R) = \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(w_i, \hat{\theta}_R)$$

will be $k \times k$, and typically invertible, so the expression for the LM statistic can be simplified to:

$$LM = n\hat{g}_n(\hat{\theta}_R)' B_n \hat{g}_n(\hat{\theta}_R) = nQ_n(\hat{\theta}_R) = DD.$$

Suppose that the moment function $g(w_i, \theta)$ is *linear* in θ, as is the case for linear IV estimation. Then it can be shown that

$$DD = LM = W_2.$$

2 Testing Moment Conditions

We now consider a different problem: testing whether some of the moment conditions are valid.

To get some intuition, consider the special case of linear IV estimation. Recall that the model is

$$y_i = x_i'\gamma + u_i,$$

where x_i has dimension k, and let z_i be an $m \times 1$ vector satisfying:

$$E[z_i u_i] = 0.$$

Suppose we are confident that the first k elements of z_i are valid instruments:

$$\begin{aligned} E[z_{i1}u_i] &= 0 \\ E[z_{i2}u_i] &= 0 \\ &\vdots \\ E[z_{ik}u_i] &= 0 \end{aligned}$$

However, we are not sure whether the remaining instrumental variables are valid:

$$\begin{aligned} E[z_{i,k+1}u_i] &\stackrel{?}{=} 0 \\ &\vdots \\ E[z_{im}u_i] &\stackrel{?}{=} 0 \end{aligned}$$

One possibility is to use the first k elements of z_i to estimate θ, and then test whether the remaining moment conditions are satisfied. Let

$$z_i^{(1)} = \begin{pmatrix} z_{i1} \\ \vdots \\ z_{ik} \end{pmatrix}, \quad z_i^{(2)} \begin{pmatrix} z_{i,k+1} \\ \vdots \\ z_{im} \end{pmatrix}.$$

Let $\hat{\theta}$ be the linear IV estimate based on the first set of instrumental variables:

$$\hat{\gamma} = \left(\sum_{i=1}^{n} z_i^{(1)} x_i' \right)^{-1} \sum_{i=1}^{n} z_i^{(1)} y_i.$$

If the remaining orthogonality conditions are valid, we should have

$$\frac{1}{n} \sum_{i=1}^{n} z_i^{(2)} (y_i - x_i' \hat{\gamma}) \xrightarrow{\text{p}} 0,$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^{(2)} (y_i - x_i' \hat{\gamma}) \xrightarrow{\text{d}} N(0, L).$$

So if we have a consistent estimator of the asymptotic variance $\hat{L} \xrightarrow{\text{p}} L$, we can form a Wald statistic

$$\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^{(2)} (y_i - x_i' \hat{\gamma}) \right]' \hat{L}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^{(2)} (y_i - x_i' \hat{\gamma}) \right].$$

Under the null hypothesis that the orthogonality conditions hold, this will converge in distribution to a χ^2_{m-k} random variable.

Now, we want to generalize this idea to any overidentified GMM setting. To express the problem compactly, let us partition the moment function $g(w_i, \theta)$ into two vector-valued functions:

$$g(w_i, \theta) = \begin{pmatrix} g_1(w_i, \theta) \\ g_2(w_i, \theta) \end{pmatrix},$$

where g_1 is $m_1 \times 1$, g_2 is $m_2 \times 1$, and $m_1 + m_2 = m$.

Let us also assume that $m_1 = k$, so that if we only used $g_1(w_i, \theta)$ as the basis for estimating θ, we would have a just-identified GMM problem. (More generally, we could have $m_1 > k$, but this would change some of the results below.)

We assume that

$$E[g_1(w_i, \theta)] = 0.$$

We want to test whether in addition,

$$E[g_2(w_i, \theta)] = 0.$$

Suppose that the second set of moment conditions do *not* hold. Then, we would have

$$E[g_2(w_i, \theta)] = \psi,$$

where $\psi \neq 0$.

The trick is to view ψ as an additional set of parameters, and test whether $\psi = 0$. So we think of the augmented parameter vector as $(\theta', \psi')'$, and consider an augmented moment condition:

$$E\left[\left(\begin{array}{c} g_1(w_i, \theta) \\ g_2(w_i, \theta) - \psi \end{array}\right)\right] = 0.$$

First, consider the unrestricted GMM estimator based on the augumented moment condition. The unrestricted GMM estimator for this moment condition would find estimated values $\hat{\theta}$ and $\hat{\psi}$ that set the sample analog of the above equation close to 0. In fact, it is just-identified, because the augmented parameter (θ, ψ) has dimension $k + m_2 = m$. So the minimization problem

$$\min_{\theta,\psi} \left[\frac{1}{n}\sum_{i=1}^{n}\left(\begin{array}{c} g_1(w_i, \theta) \\ g_2(w_i, \theta) - \psi \end{array}\right)\right]' B_n \left[\frac{1}{n}\sum_{i=1}^{n}\left(\begin{array}{c} g_1(w_i, \theta) \\ g_2(w_i, \theta) - \psi \end{array}\right)\right]'$$

can be solved exactly, by first choosing $\hat{\theta}$ to solve

$$\frac{1}{n}\sum_{i=1}^{n} g_1(w_i, \theta) = 0,$$

then choosing ψ by setting

$$\frac{1}{n}\sum_{i=1}^{n} g_2(w_i, \hat{\theta}) - \psi = 0.$$

In other words, we will have

$$\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n} g_2(w_i, \hat{\theta}).$$

This is a valid solution, because it will lead to $Q_n(\hat{\theta}, \hat{\psi}) = 0$.

The restricted estimator, which imposes the restriction that $\psi = 0$, would set the sample analog of

$$E\left[\left(\begin{array}{c} g_1(w_i, \theta) \\ g_2(w_i, \theta) \end{array}\right)\right]$$

close to zero. This is just the original GMM estimator using all the moment conditions.

From here, we can use the test statistics outlined above to test whether $\psi = 0$. A particularly convenient choice is the DD test. Notice that the unrestricted estimator has $Q_n(\hat{\theta}, \hat{\psi}) = 0$. So the DD test is just:

$$DD = n \cdot Q_n(\hat{\theta}_R),$$

where $\hat{\theta}_R$ is the GMM estimator using all the original moment conditions. By our result above, this is also the LM test statistic. This test is sometimes called the Hansen J test.

This test has an interesting interpretation. It would be the same if we had rearranged the elements of $g(w_i, \theta)$ and tested a *different* set of $m-k$ moment conditions. So we don't have to specify which of the $m-k$ moment conditions to test, and we have an "omnibus" test of the validity of all the moment conditions. On the other hand, rejecting the null hypothesis does not tell us which of the moment conditions are invalid.

We also have to be careful about interpreting violations of the moment conditions. In the case of linear IV, a violation could indicate that some of the instruments are not orthogonal to the omitted variables. But it could also indicate a misspecification of the functional form of the main equation (which would lead to lack of orthogonality between z_i and and the "residual" $(y_i - x_i'\gamma)$). So in the linear IV case, the test is really a joint test of the specification *and* the orthogonality of z_i with any omitted variables in the main equation.

3 Hausman Tests

A useful strategy for testing model specifications was proposed by Hausman (1978). He suggested comparing to parameter estimators: one which is consistent and asymptotically normal under the more general model, and one which is consistent only under the restricted specification.

For example, suppose we want to estimate an equation

$$y_i = x_i'\gamma + u_i$$

and we believe that the x_i are uncorrelated with any omitted variables, so that OLS would be consistent.

Suppose we have valid instrumental variables z_i. If $E[x_i u_i] = 0$, we can just do OLS, but if $E[x_i u_i] \neq 0$, then OLS would not be consistent for γ, but 2SLS would be consistent. So we can compare $\hat{\gamma}_{OLS}$ to $\hat{\gamma}_{2SLS}$. We could use a test statistic:

$$H = n \cdot (\hat{\gamma}_{OLS} - \hat{\gamma}_{2SLS})'\hat{V}^{-1}(\hat{\gamma}_{OLS} - \hat{\gamma}_{2SLS}).$$

Here, $\hat{V}^{-1}$ needs to be a consistent estimator of the generalized inverse of the asymptotic variance of $\sqrt{n}(\hat{\gamma}_{OLS} - \hat{\gamma}_{2SLS})$. (We need to talk about generalized inverses, because in some applications, the asymptotic variance may be singular.)