TA section 5

May 8th, 2009 by Yang

## 1. Derivation of the J-test

In the model of GMM estimation, the $J$-statistic is defined as follows.

$$J_n := n\left(\frac{1}{n}\sum_{i=1}^{n} g(W_i,\widehat{\theta}_n)'\right) S_n \left(\frac{1}{n}\sum_{i=1}^{n} g(W_i,\widehat{\theta}_n)\right) \xrightarrow{d} \chi^2_{k-d}$$

where $\widehat{\theta}_n$ is an efficient GMM estimator, and $S_n \xrightarrow{p} V_0^{-1} := Eg(W_i,\theta_0)g(W_i,\theta_0)'$. It should be satisfied that $\widehat{\theta}_n$ is obtained using a consistent estimator of the optimal weight matrix. Nevertheless, it is not necessary to assume that $\widehat{\theta}_n$ is obtained using $S_n$, nor that $S_n$ is estimated using $\widehat{\theta}_n$. To see this in the linear IV example, recall question 1 from the problem set 4. Let $\widehat{\Omega}_1 \xrightarrow{p} \Omega$ and $\widehat{\Omega}_2 \xrightarrow{p} \Omega$. Define $\widehat{\beta}_n$ as the GMM estimator that minimizes

$$Q_n(\beta) = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)z_i'\right)\widehat{\Omega}_1^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\beta)\right)$$

Then we can prove that

$$J_n := n\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)z_i'\right)\widehat{\Omega}_2^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\widehat{\beta}_n)\right) \xrightarrow{d} \chi^2_{k-d}$$

using the same procedure, by defining

$$D_n := I_k - C'\left(\frac{1}{n}Z'X\right)\left(\frac{1}{n}X'Z\widehat{\Omega}_1^{-1}\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}X'Z\widehat{\Omega}_1^{-1}(C')^{-1}\right)$$

From the above argument, we can see many ways to get a $J$-statistic, using different $\widehat{\beta}_n$ and $\widehat{\Omega}_2$. Numerical values vary, but the asymptotic properties will remain the same. A natural way to get the $J$-statistic in the empirical analysis is

$$J_n = n\cdot\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)z_i'\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)^2 z_i z_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\widehat{\beta}_n)$$

where $\widehat{\beta}_n$ is an (feasible) efficient GMM estimator. Another way is

$$J_n = n\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)z_i'\right) S_n \left(\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\widehat{\beta}_n)\right)$$

where $S_n \xrightarrow{p} \Omega^{-1}$ is used to estimate an (feasible) efficient GMM estimator $\widehat{\beta}_n$. Under the assumption of conditional homoskedasticity, this leads us to

$$J_n = \frac{1}{\widehat{\sigma}_n^2}(Y - X\widehat{\beta}_n)'Z(Z'Z)^{-1}Z'(Y - X\widehat{\beta}_n)$$

**2. Validity of the J-test**

The J-test tests the null hypothesis $H_0 : Eg(W_i, \theta_0) = 0$ which consists of $k$ equations. In the linear IV example, it tests $H_0 : Ez_i\varepsilon_i = 0$. The intuition behind this is the following. If all the moment conditions that are used in the 2SLS estimation are true, the sample analogue of those will be close to 0, and thus $J_n$ will be bounded in probability. More specifically, $J_n \xrightarrow{d} \chi^2_{k-d}$ where $k$ is the number of instruments and $d$ is the dimension of $\beta_0$. If at least one of the moment conditions is false, their sample analogue would be different from 0, and thus $J_n$ is expected to diverge to infinity in probability.

But there are two problems as mentioned in the lecture note. First, the J-test has very low power against some alternative hypotheses. Suppose, for example, $Ez_i\varepsilon_i = Ez_ix_i'\gamma \neq 0$ for some nonstochastic $d \times 1$ vector $\gamma$. In this case, we can show that $J_n \xrightarrow{d} \chi^2_{k-d}$.[1] In question 2 in the same problem set, we tested this alternative. In case 1,

$$Ez_i\varepsilon_i = c = \begin{pmatrix} \delta \\ \vdots \\ \delta \end{pmatrix} = \pi\frac{\delta}{\eta} = Ez_ix_i'\frac{\delta}{\eta}$$

so we see that $Ez_i\varepsilon_i = Ez_ix_i'\gamma \neq 0$ holds for $\gamma = \delta/\eta$. It was verified in this case that the rejection probability is the same as the significance level for any choice of $\delta$. This property follows from the fact that the $J$-statistic has the same limit distribution under $H_0$ and $Ez_i\varepsilon_i = Ez_ix_i'\gamma \neq 0$. Therefore we would fail to reject false $H_0$, so end up with using incorrect moment conditions in the second stage. In this case, $\widehat{\beta}_n$ may be inconsistent with high probability, so the second stage test will reject $H_0$ many times even though it is true. This distorts the size of the whole procedure. In the Monte Carlo simulation, we observed that the conditional null rejection probability of the Wald test goes up very quickly as $\delta$ gets further from 0 when the above case is used.

Second, the finite sample approximation of the J-test is very poor. $J_n$ converges in distribution to $\chi^2_{k-d}$ but very slowly, so when the sample size is small, it may reject true $H_0$ too often, or may not reject false $H_0$ many times.

[1] The proof uses a slight modification of that of question 1 in the problem set 4. A sketch of proof is as follows. Although $\overline{g}_n(\beta_0) \xrightarrow{p} Ez_i\varepsilon_i \neq 0$, we have

$$C'\overline{g}_n(\widehat{\beta}_n) = D_nC'\overline{g}_n(\beta_0) \xrightarrow{p} [I_k - R(R'R)^{-1}R']C'Ez_i\varepsilon_i$$

Note that by definition of $R$, the RHS becomes 0, from the assumption that $Ez_i\varepsilon_i = Ez_ix_i'\gamma$. Now use the CLT to obtain

$$C'\sqrt{n}\overline{g}_n(\widehat{\beta}_n) \approx D_nC'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i\varepsilon_i - Ez_i\varepsilon_i) \xrightarrow{d} DC'\overline{N}$$

where $\overline{N} \sim N(0, \Sigma)$ and $\Sigma := Var(z_i\varepsilon_i)$. Since $CDC'\Sigma CDC' = CDC'$, we have

$$J_n = n\overline{g}_n(\widehat{\beta}_n)'\widehat{\Omega}^{-1}\overline{g}_n(\widehat{\beta}_n) \xrightarrow{d} \chi^2_m$$

where $m := tr(CDC'\Sigma) = k - d$. In the last derivation, we use the theorem that if $X \sim N(0, \Sigma)$ and $A\Sigma A = A$, then $X'AX \sim \chi^2_{tr(A\Sigma)}$. To check $CDC'\Sigma CDC' = CDC'$ and $tr(CDC'\Sigma) = k - d$, use the definition of $\Sigma$, $C$, $D$ and $R$, and the assumption $Ez_i\varepsilon_i = Ez_ix_i'\gamma$.

**3. Estimation of the LAD model**

There is no analytic solution of the least absolute deviations (LAD) regression model.

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i' \beta|$$

So we have to use numerical method to find it. There might be some routines that do this work, available online. But there is no built-in MATLAB function doing it. So we use a minimization routine by specifying the criterion function that has to be minimized. It is simple. Write in the main routine

```
beta_init = zeros(k,1);
beta_hat = fminsearch('lad_criterion',beta_init,[],y,x);
```

and make a criterion function `lad_criterion.m` as follows.

```
function RET = lad_criterion(beta,y,x);
RET = mean(abs(y - x * beta));
```

where `y` is a $n \times 1$ vector of $y_i$, and `x` is a $n \times k$ matrix of $x_i'$. The above function returns the value of the criterion function

$$Q_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i' \beta|$$

Then the function `fminsearch` evaluates the above function at lots of $\beta$ starting from `beta_init`, and returns $\widehat{\beta}$ that minimizes the criterion function. Note that the criterion function is not continuously differentiable, but convex in $\beta$, so there is no problem for the routine to find a global minimum.