Moment Testing with non-ML Estimators

M. Arellano
May 1991

The null hypothesis under test is given by

[1] $$H_0 : E[\varphi(w_i,\bar{\theta})] = 0$$

where φ is a px1 vector of functions and θ is a kx1 vector of parameters. The expectation is taken with respect to the distribution of the random vector w_i and $\bar{\theta}$ denotes the true value.

1. The Estimators

Let $\hat{\theta}$ be the minimiser of a differentiable criterion function $c(\theta)$ so that

$$\partial c(\hat{\theta})/\partial\theta = 0$$

If $c(\theta)$ is an M-estimator criterion of the form

$$c(\theta) = \Sigma_{i=1}^{n}\ell(w_i,\ \theta,\ \hat{\tau}_n)$$

based on a sample of size n, then

$$\frac{\partial c(\theta)}{\partial\theta} = \Sigma_{i=1}^{n}q(w_i,\ \theta,\hat{\tau}_n)$$

where

$$q(w_i,\ \theta,\hat{\tau}_n) = \partial\ell(w_i,\ \theta,\hat{\tau}_n)/\partial\theta$$

On the other hand, if $c(\theta)$ is a GMM-criterion of the form

$$c(\theta) = \frac{1}{2}\ (\Sigma_{i=1}^{n}m(w_i,\theta))'A_n(\Sigma_{i=1}^{n}m(w_i,\theta))$$

then
$$\frac{\partial c(\theta)}{\partial\theta} = \left[\ \Sigma_{i=1}^{n}\frac{\partial m(w_i,\theta)}{\partial\theta'}\ \right]' A_n\ \Sigma_{i=1}^{n}m(w_i,\theta)$$

$$= D_n'A_n\Sigma_{i=1}^{n}m(w_i,\theta)$$

$$= \Sigma_{i=1}^{n}q(w_i,\theta,\hat{\tau}_n)$$

where $D_n = \Sigma_{i=1}^{n}\partial m(w_i,\theta)/\partial\theta'$ and $q(w_i,\theta,\hat{\tau}_n) = D_n'A_n m(w_i,\theta)$

In both classes of criteria, $\hat{\tau}_n$ represents some preliminary estimator used in the definition of the first order condition for the estimator of θ. The conclusion of this discussion is that the representation of the FOC's given by

[2] $$\frac{\partial c(\theta)}{\partial \theta} = \Sigma_{i=1}^{n} q(w_i, \theta, \hat{\tau}_n)$$

includes most estimators in use.

Here I assume that

[3] $$n^{-\frac{1}{2}} \Sigma_i \; q(w_i, \bar{\theta}, \hat{\tau}_n) = n^{-\frac{1}{2}} \Sigma_i q(w_i, \bar{\theta}, \text{ plim } \hat{\tau}_n) + o_p(1)$$

so that alternative consistent estimators of $\bar{\tau} = \text{plim } \hat{\tau}_n$ leave the asymptotic distribution of $\partial c(\bar{\theta})/\partial\theta$ unchanged,
and

[4] $$n^{-\frac{1}{2}} \Sigma_i \; q(w_i, \bar{\theta}, \bar{\tau}) \overset{d}{\to} N(0, B)$$

where $B = E(\bar{q}_i \bar{q}_i')$ and

$$\bar{q}_i = q(w_i, \bar{\theta}, \bar{\tau})$$

I also assume that

[5] $$n^{-1} \frac{\partial^2 c(\bar{\theta})}{\partial\theta\partial\theta'} \overset{p}{\to} H > 0$$

With these assumptions, we can obtain an asymptotic normality result for $n^{\frac{1}{2}}(\hat{\theta}-\bar{\theta})$ in the usual way:

$$0 = n^{-\frac{1}{2}} \frac{\partial c(\hat{\theta})}{\partial\theta} = n^{-\frac{1}{2}} \frac{\partial c(\bar{\theta})}{\partial\theta} + n^{-1} \frac{\partial^2 c(\bar{\theta})}{\partial\theta\partial\theta'} n^{\frac{1}{2}} (\hat{\theta}-\bar{\theta}) + o_p(1)$$

so that

$$n^{\frac{1}{2}}(\hat{\theta}-\bar{\theta}) = - H^{-1} n^{-\frac{1}{2}} \frac{\partial c(\bar{\theta})}{\partial\theta} + o_p(1)$$

or also

[6] $$n^{\frac{1}{2}} (\hat{\theta}-\bar{\theta}) = - H^{-1} n^{-\frac{1}{2}} \Sigma_i \; \bar{q}_i + o_p(1)$$

Thus

[7] $$n^{\frac{1}{2}} (\hat{\theta}-\bar{\theta}) \overset{d}{\to} N \; (0, H^{-1}B \; H^{-1})$$

Here we do not necessarily assume that H and B coincide up to a scalar factor as it would happen with efficient criterion functions.

2. The Tests

We want to test H_0 by testing the significance of $n^{-1}\Sigma_i\varphi(w_i,\hat{\theta}) \approx 0$

We use the following approximation:

$$n^{-\frac{1}{2}} \Sigma_i\hat{\varphi}_i = n^{-\frac{1}{2}} \Sigma_i\bar{\varphi}_i + (\frac{1}{n} \Sigma_i \frac{\partial\bar{\varphi}_i}{\partial\theta'}) n^{-\frac{1}{2}} (\hat{\theta}-\bar{\theta}) + o_p(1)$$

where $\hat{\varphi}_i = \varphi(w_i,\hat{\theta})$ and $\bar{\varphi}_i = \varphi(w_i,\bar{\theta})$.

Also using [6]:

$$n^{-\frac{1}{2}} \Sigma_i\hat{\varphi}_i = n^{-\frac{1}{2}} \Sigma_i\bar{\varphi}_i - (n^{-1}\Sigma_i\frac{\partial\bar{\varphi}_i}{\partial\theta'}) H^{-1} n^{-\frac{1}{2}} \Sigma_i\bar{q}_i + o_p(1)$$

Let us denote $E(\frac{\partial\bar{\varphi}_i}{\partial\theta'}) = \Phi$, so that also

[8] $$n^{-\frac{1}{2}} \Sigma_i\hat{\varphi}_i = n^{-\frac{1}{2}} \Sigma_i(\bar{\varphi}_i-\Phi H^{-1}\bar{q}_i) + o_p(1)$$

so $$n^{-\frac{1}{2}} \Sigma_i\hat{\varphi}_i \xrightarrow{d} N(0,V)$$

where

[9] $$V = E [(\bar{\varphi}_i-\Phi H^{-1}\bar{q}_i)(\bar{\varphi}_i-\Phi H^{-1}\bar{q}_i)']$$

A consistent estimator of V replaces E by $n^{-1}\Sigma_i$, $\bar{\varphi}_i$ and $\bar{q}_i$ by $\hat{\varphi}_i$ and $\hat{q}_i = q(w_i,\hat{\theta},\hat{\tau}_n)$, and Φ and H by consistent estimators. In particular,

[10] $$\hat{\Phi} = n^{-1}\Sigma_i \frac{\partial\varphi(w_i,\hat{\theta})}{\partial\theta'}$$

[11] $$\hat{H} = n^{-1} \frac{\partial^2 c(\hat{\theta})}{\partial\theta\partial\theta'}$$

Then we have

[12] $$\hat{V} = n^{-1} \Sigma_i (\hat{\varphi}_i-\hat{\Phi} \hat{H}^{-1} \hat{q}_i)(\hat{\varphi}_i-\hat{\Phi} \hat{H}^{-1}\hat{q}_i)'$$

Now the test statistic can be calculated as:

[13] $$m = n^{-1}(\Sigma_i\hat{\varphi}_i)'\hat{V}^{-1}(\Sigma_i\hat{\varphi}_i) \rightarrow \chi^2_p \text{ under } H_0$$

3. Remarks

Alternative estimators of V can be considered in special cases:

(i) If $c(\theta)$ is an efficient criterion function we can use an estimator of B in order to estimate H.

(ii) If $c(\theta)$ is an ML criterion we can use the fact that

$$\Phi = E\left(\frac{\partial\varphi(w_i,\bar{\theta})}{\partial\theta'}\right) = -E\left[\varphi(w_i,\bar{\theta})\ \frac{\partial \log f(w_i,\bar{\theta})}{\partial\theta'}\right]$$

where $f(w_i,\bar{\theta})$ is the density of the data, in order to obtain alternative estimates of Φ (see Lancaster (1984)).

The form of the statistic given in [12] and [13] corresponds to the original variant of the Information Matrix test as proposed by White (1982) in the case in which $\varphi(w_i,\theta)$ indicates the information matrix equivalence. The m_2 statistic in Arellano & Bond (1991) is also a test of the form [13].

References

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.

Chesher, A.D., (1984): "Testing for Neglected Heterogeneity", *Econometrica*, 52, 865-872.

Lancaster, T. (1984): "The Covariance Matrix of the Information-Matrix Test", *Econometrica*, 52, 1051-1053.

Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047-1070.

Orme, C. (1990): "The Small-Sample Performance of the Information-Matrix Test", *Journal of Econometrics*, 46, 309-331.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-25.