Ec241a **Econometrics**

Spring 2004 **UC Berkeley Department of Economics**

Maximum Likelihood Estimation V:
Information Matrix Tests (Chesher, Lancaster, 1984)

For the duration model testing for unobserved heterogeneity was a complex problem. A simpler approach was suggested by Chesher (1984). It is based on the Information Matrix Equality. The information matrix equality says that

$$\mathcal{I}(\theta_0) = -E\left[\frac{\partial^2 \ln f}{\partial\theta\partial\theta'}(Z,\theta_0)\right] = E\left[\frac{\partial \ln f}{\partial\theta}(Z,\theta_0)\cdot\frac{\partial \ln f}{\partial\theta}(Z,\theta_0)'\right].$$

Note that this equality only holds at the true values of the parameters. We have already used this when we were considering different estimators for the variance of maximum likelihood estimators. This is actually the second in a series of Bartlett equalities, the first is the well known result that the score has expectation zero at the true values of the parameters:

$$E\left[\frac{\partial \ln f}{\partial\theta}(Z,\theta_0)\right] = 0.$$

To see the connection with the previous discussion of the duration model test recall that the score function that was the basis of the test had the form:

$$\mathcal{S}_\eta(y,\lambda_0,\eta\to 0) = -2y\lambda_0 + y^2\lambda_0^2.$$

Now consider the information matrix equality for λ, under the null hypothesis. First,

$$\mathcal{S}(y,\lambda) = \frac{1}{\lambda} - y.$$

Second,

$$\mathcal{H}(y,\lambda) = -\frac{1}{\lambda^2},$$

so that the information matrix (scalar in this case) equality is

$$E\Big[\mathcal{S}(Y,\lambda_0)^2 + \mathcal{H}(Y,\lambda_0)\Big] = E\left[Y^2 - 2\frac{Y}{\lambda_0}\right] = 0,$$

This expression is proportional to the score for the unobserved heterogeneity test. Chesher's contribution is that he discovered that this is true in a much more general sense in two ways. First, we can calculate the score for tests for unobserved heterogeneity using other distributions. Instead of taking a discrete two point mixture, we can use a continuous distribution, for example a log normal distribution. If we look at the score for the test that the variance of that distribution is zero, we end up with the <u>exact</u> same score function. Second, we can do this for different parameters and different models. In each case testing whether a particular parameter has zero variance corresponds to testing the appropriate part of the information matrix equality.

So let Z be a random variable with probability density (mass) function

$$f(z|\theta_0, \theta_1),$$

where θ_0 is a scalar and θ_1 a vector of parameters. Suppose θ_0 can take on both positive and negative values (if not, we can reparametrize the model). Let

$$g(v|\sigma^2),$$

be the probability density function of a scalar random variable V with mean zero and variance σ^2. Now consider the more general model for Z with probability density function

$$f(z|\theta_0, \sigma^2, \theta_1) = \int_v f(z|\theta + v, \theta_1) \cdot g(v|\sigma^2) dv.$$

The parameter θ_0 that was fixed in the null model has in the generalized model a distribution with mean θ_0 and variance σ^2. To test the model we can do a score or Lagrange multiplier test of the null hypothesis $\sigma^2 = 0$ against the alternative hypothesis that $\sigma^2 > 0$.

Chesher shows that for a large, flexible class of distributions $g(\cdot)$, and for general choices of models $f(\cdot)$, the score function that is the basis of the score test for testing the null that $\sigma^2 = 0$, is proportional to

$$\left(\frac{\partial \ln f}{\partial \theta_0}(z; \theta_0, \theta_1)\right)^2 + \frac{\partial^2 \ln f}{\partial \theta_0^2}(z; \theta_0, \theta_1),$$

which has expectation zero by the information matrix equality.

More generally, with a K-dimensional parameter vector θ, there are in principle $(K+1)\cdot K/2$ different elements of the information matrix that can be tested. (Note that although the information matrix is of dimension $K \times K$, the matrix is symmetric, so we cannot test all elements.) Let V be the $N \times M$ matrix with (n,m)th element equal to

$$V_{nm} = \frac{\partial \ln f}{\partial \theta_i}(z_n;\hat{\theta}) \cdot \frac{\partial \ln f}{\partial \theta_j}(z_n;\hat{\theta}) + \frac{\partial^2 \ln f}{\partial \theta_j \partial \theta_j}(z;\hat{\theta}),$$

for some pair (i,j). A little loosely we can still think of this as testing for heterogeneity in the original coefficients. So the alternative hypothesis is that some elements of θ are heterogenous, and we are testing their variances (or covariances) by looking at the question whether the average of V_{nm} is close to zero. We can test any set of these information matrix equalities, up to $K \times (K+1)/2$, as long as we avoid adding columns that are linear combinations of other columns.

Lancaster (1984) suggested a simple way to carry out such tests. In addition to V define the $N \times K$ matrix S with typical element

$$S_{ij} = \frac{\partial \ln f}{\partial \theta_j}(z_i,\hat{\theta}).$$

Each of the K columns in this matrix is a vector of score functions. Then regress a vector of ones, ι, on the $N \times (M+K)$ matrix $W = [S\ V]$. The regression estimates are

$$\hat{\beta}_{ols} = (W'W)^{-1}W'\iota.$$

The (uncentered) R^2 from this regression is

$$R^2 = \frac{\sum_{i=1}^N (1 - W_i\hat{\beta}_{ols})^2}{\sum_{i=1}^N 1^2} = \frac{\iota' W(W'W)^{-1}W'\iota}{\iota'\iota}.$$

The score test statistic turns out to be equal to N times this R^2, or

$$IT = \iota' W(W'W)^{-1}W'\iota.$$

(See the Chesher and Lancaster papers for details.) Under the null hypothesis this test statistic has a chi-squared distribution with degrees of freedom equal to M.

Example

Let us look at another example of such a test. This particular case was actually analyzed by White (1980) (see also White 1982) before the general form of the information matrix test. Consider the linear regression model with

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i,$$

with scalar X_i and

$$\varepsilon_i | X_i \sim \mathcal{N}(0, \sigma^2).$$

First consider the score function:

$$\mathcal{S}(y, x|\beta_0, \beta_1, \sigma^2) = \frac{\partial \ln f}{\partial(\beta_0, \beta_1, \sigma^2)'}(y, x|\beta_0, \beta_1, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2}(y - \beta_0 - x\beta_1) \\ \frac{x}{\sigma^2}(y - \beta_0 - x\beta_1) \\ -\frac{1}{2\sigma^2} + \frac{(y-\beta_0-x\beta_1)^2}{2\sigma^4} \end{pmatrix}.$$

Now consider the $(1, 1)$ element of the information matrix equality:

$$IME(y, x, \beta_0, \beta_1, \sigma^2)_{11} = \mathcal{S}_{\beta_0}(y, x|\beta_0, \beta_1, \sigma^2)^2 + \mathcal{H}_{\beta_0\beta_0}(y, x|\beta_0, \beta_1, \sigma^2)$$

$$= \frac{1}{\sigma^4}(y - \beta_0 - x\beta_1)^2 - \frac{1}{\sigma^2}.$$

Note that this is proportional to the third element of the score function, $\frac{\partial \ln f}{\partial \sigma^2}(\beta_0, \beta_1, \sigma^2)$, so we cannot use this in the test.

Instead, consider the $(2, 2)$ element of the information matrix equality:

$$IME(y, x|\beta_0, \beta_1, \sigma^2) = \mathcal{S}_{\beta_1}(y, x|\beta_0, \beta_1, \sigma^2)^2 + \mathcal{H}_{\beta_1\beta_1}(y, x|\beta_0, \beta_1, \sigma^2)$$

$$= \frac{x^2}{\sigma^4}(y - \beta_0 - x\beta_1)^2 - \frac{x^2}{\sigma^2}.$$

Thus, one possibility for the information matrix is to use the $(2, 2)$ element. The matrix Z would in this case be the $N \times 4$ matrix with ith row equal to

$$W_{i\cdot} = (\mathcal{S}_i' \;\; \mathcal{S}_{\beta_1}^2 + \mathcal{H}_{\beta_1\beta_1}) = \begin{pmatrix} \frac{1}{\sigma^2}(y_i - \beta_0 - x_i\beta_1) \\ \frac{x_n}{\sigma^2}(y_i - \beta_0 - x_i\beta_1) \\ -\frac{1}{2\sigma^2} + \frac{(y_i-\beta_0-x_i\beta_1)^2}{2\sigma^4} \\ \frac{x_i^2}{\sigma^4}(y_i - \beta_0 - x_i\beta_1)^2 - \frac{x_i^2}{\sigma^2} \end{pmatrix}'.$$

Let us inspect the form of the $(2,2)$ element of the information matrix equality in more detail. It can be written as

$$IME(y,x|\beta_0,\beta_1,\sigma^2) = \mathcal{S}_{\beta_1}^2 + \mathcal{H}_{\beta_1\beta_1} = \frac{x^2\varepsilon^2}{\sigma^4} - \frac{x^2}{\sigma^2}.$$

The test thus compares the average value of $x^2\varepsilon^2$ to the average value of $x^2\sigma^2$. Under homoskedasticity this should be zero. The test therefore can be interpreted, and this was the way it was proposed by White in this context, as a test for heteroskedasticity.

Note that in addition to, or instead of, the $(2,2)$ element we can also use other elements of the information matrix equality in the test in this example, including diagonal and off-diagonal elements. $\square$

References

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