Lecture 5

Three ML Test Procedures: LR, W and LM

Likelihood Ratio Test

Let θ be the parameter vector, $\hat{\theta}$ be the ML estimator, and $\hat{\theta}_R$ be the restricted ML estimator.

The likelihood ratio is $\lambda = \dfrac{L(\hat{\theta}_R)}{L(\hat{\theta})}$

where L(.) denotes the likelihood function.

Then $-2\ln\lambda \sim \chi^2_J$

where J denotes the number of restrictions imposed by the null hypothesis.

Wald

$$H_0 = \mathbf{c}(\theta) = \mathbf{q}$$

where $\mathbf{c}(\theta)$ is a Jx1 vector of functions of θ

Then

$$W = \left[\mathbf{c}(\hat{\theta}) - \mathbf{q}\right]' \left(Var\left(\mathbf{c}(\hat{\theta}) - \mathbf{q}\right)\right)^{-1} \left[\mathbf{c}(\hat{\theta}) - \mathbf{q}\right]$$

under H_0 $W \sim \chi^2_J$

Note that: $\left(Var\left(\mathbf{c}(\hat{\theta}) - \mathbf{q}\right)\right) = C\, Var(\hat{\theta})\, C'$

where $C = \left[\dfrac{\partial \mathbf{c}(\hat{\theta})}{\partial \hat{\theta}}\right]_{JXK}$

(assuming Jx1 is the dimension of $\mathbf{c}(\theta)$ and θ is Kx1).

Also note: If $\mathbf{c}(\theta) = R\theta$ is linear then

$$W = \left(R\hat{\theta} - \mathbf{q}\right)' \left(RVar(\hat{\theta})R'\right)^{-1} \left(R\hat{\theta} - \mathbf{q}\right)$$

Lagrangian Multiplier(LM) (sometimes called the Efficient Score Test)

The constrained log likelihood objective function is:

$$\ln L^*(\theta) = \ln L(\theta) + \lambda'[\mathbf{c}(\theta) - \mathbf{q}]$$

where λ is the Jx1 vector of Lagrangian coefficients.

$$\frac{\partial \ln L^*}{\partial \theta} = \frac{\partial \ln L(\theta)}{\partial \theta} + \mathbf{C}'\lambda = \mathbf{0}$$

$$\frac{\partial \ln L^*}{\partial \lambda} = \mathbf{c}(\theta) = \mathbf{0}$$

where $\mathbf{C}' = \left[\dfrac{\partial \mathbf{c}(\theta)}{\partial \theta}\right]_{JXK}$

$\therefore$ at the restricted maximum,

$$\frac{\partial \ln L(\hat{\Phi}_R)}{\partial \hat{\Phi}} = -\hat{C}'\hat{\lambda} = \hat{g}_R$$

where $\hat{g}_R = \left(\frac{\partial \ln L(\hat{\Phi}_R)}{\partial \hat{\Phi}}\right)_{\Phi=\hat{\Phi}_R}$

If the restrictions are valid, at least within the range of sampling variability, $\hat{g}_R = \mathbf{0}.$

Under H_0: LM = $\hat{g}_R'\left[I(\hat{\Phi}_R)\right]^{-1}\hat{g}_R$.

In terms of implementation, let $\hat{g}_{iR}$ denote the gradient of the i-th observation of the likelihood.

Then

$$\hat{g}_R = \sum_{i=1}^{N}\hat{g}_{iR} = \hat{G}_{Ri}'$$

where $\hat{G}_R$ is the NxK matrix with the i-th row equal to $\hat{g}_{iR}$ and **i** is an nx1 column of ones. If we use the BHHH estimator to form a consistent estimate of the variance-covariance matrix of $\hat{\Phi}$,

then $\left[I(\hat{\Phi}_R)\right]^{-1} = \left[\hat{G}_R'\hat{G}_R\right]^{-1}$ and

LM = $\mathbf{i}'\hat{G}_R\left[\hat{G}_R'\hat{G}_R\right]^{-1}\hat{G}_R'\mathbf{i}$.

Note: Since **i'i** = n, LM = n($\mathbf{i}'\hat{G}_R\left[\hat{G}_R'\hat{G}_R\right]^{-1}\hat{G}_R'\mathbf{i}$ ./n**)**

= n R_i^2

where R_i^2 is the uncentered multiple correlation coefficient in a linear regression of a column of ones or (i.e. regressing **i** on $\hat{G}_R$ and obtaining uncentered R^2) the derivatives of the log-likelihood function computed at the restricted estimator. This form shows up in a lot of different contexts in econometric testing.

<u>An Illustration of various ML Test Procedures</u>

An Example:Greene

(See SAS program example that Carter Hill wrote to compute various test statistics)

Consider the data in Table 4.1 and the Gamma distribution:

<u>Unrestricted Model</u>:

$$f(y_i, X_i, \beta, \rho) = \frac{(\beta + X_i)^{-\rho}}{\Gamma(\rho)} y_i^{\rho-1}\exp\{-y_i/(\beta + X_i)\}$$

Now let $H_0 : \rho = 1$

$H_1 : \rho \neq 1$

$$\ln L(\beta, \rho) = -\rho\sum_{1}^{N}\ln(\beta + X_i) - n\ln\Gamma(\rho) + (\rho - 1)\sum_{1}^{N}\ln y_i - \sum_{1}^{N}\frac{y_i}{(\beta + X_i)}$$

$$\frac{\partial \ln L}{\partial \beta} = -\rho \sum_1^N \frac{1}{(\beta + X_i)} + \sum_1^N \frac{y_i}{(\beta + X_i)^2}$$

$$\frac{\partial \ln L}{\partial \rho} = -\sum_1^N \ln(\beta + X_i) - n\frac{\Gamma'(\rho)}{\Gamma(\rho)} + \sum_1^N \ln y_i$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \rho \sum_1^N \frac{1}{(\beta + X_i)^2} - 2\sum_1^N \frac{y_i}{(\beta + X_i)^3}$$

$$\frac{\partial^2 \ln L}{\partial \rho^2} = -\frac{N\left[\Gamma(\rho)\Gamma''(\rho) - (\Gamma'(\rho))^2\right]}{(\Gamma'(\rho))^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \rho} = -\sum_1^N \frac{1}{(\beta + X_i)}$$

Restricted Model

Let $\rho = 1$

$$f(y_i, X_i, \beta, \rho = 1) = \frac{1}{(\beta + X_i)} \exp\{-y_i / (\beta + X_i)\}$$

$$\ln L(\beta) = \ln \prod_{i=1}^N \frac{1}{(\beta + X_i)} \exp\{-y_i / (\beta + X_i)\}$$

$$= -\sum_1^N \ln(\beta + X_i) - \sum_1^N \frac{y_i}{(\beta + X_i)}$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_1^N \frac{1}{(\beta + X_i)} + \sum_1^N \frac{y_i}{(\beta + X_i)^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \sum_1^N \frac{1}{(\beta + X_i)^2} - 2\sum_1^N \frac{y_i}{(\beta + X_i)^3}$$

Now how do we go about maximizing these two likelihood functions. We can use the **Gauss-Newton Method.**

Let $\mathbf{g}(\Phi)$ be the gradient vector for the likelihood function $L(\Phi)$. At the maximum likelihood estimator $\mathbf{g}(\hat{\Phi}) = \mathbf{0}$

Expand this set of equations in to a second-order Taylor series expansion around the true parameter vector Φ:

$$\mathbf{g}(\hat{\Phi}) = \mathbf{g}(\Phi) + H(\Phi)(\hat{\Phi} - \Phi) + \frac{1}{2} T(\Phi^*, \hat{\Phi}, \Phi) = \mathbf{0}$$

Assuming T(.) is negligible then:

$$\mathbf{g}(\theta) = -H(\theta)(\hat{\theta} - \theta)$$

$$(\hat{\theta} - \theta) = -[H(\theta)]^{-1}\mathbf{g}(\theta)$$

$$\hat{\theta} = \theta - [\mathbf{H}(\theta)]^{-1}\mathbf{g}(\theta)$$

The method of Newton then suggests explaining the unknown θ with an initial estimate $\tilde{\theta}$ and obtain a second round estimate

$$\breve{\theta} = \tilde{\theta} - [\mathbf{H}(\tilde{\theta})]^{-1}\mathbf{g}(\tilde{\theta})$$

and **continue to iterate**.

Note: The Newton two-step estimate $\breve{\theta}$ is asymptotically efficient without iteration as long as $\tilde{\theta}$ is a consistent estimator of θ.

In the case of the Gamma distribution we have:

$$\begin{pmatrix}\breve{\beta}\\ \breve{\rho}\end{pmatrix} = \begin{pmatrix}\hat{\beta}\\ \hat{\rho}\end{pmatrix} - \begin{bmatrix} -\hat{\rho}\sum_1^N \frac{1}{(\hat{\beta}+X_i)^2} - 2\sum_1^N \frac{y_i}{(\hat{\beta}+X_i)^3} & . \\ -\sum_1^N \frac{1}{(\hat{\beta}+X_i)} & -Ntrigamma(\rho) \end{bmatrix}^{-1} \begin{pmatrix} -\hat{\rho}\sum_1^N \frac{1}{(\hat{\beta}+X_i)} + \sum_1^N \frac{y_i}{(\hat{\beta}+X_i)^2} \\ -\sum_1^N \ln(\hat{\beta}+X_i)^2 - \frac{N\Gamma'(\hat{\rho})}{\Gamma(\hat{\rho})} + \sum_1^N \ln(y_i) \end{pmatrix}$$

Go to Hill's expmle4.sas program.