Lecture 4

Estimating the Asymptotic Covariance Matrices of Maximum Likelihood Estimators

The asymptotic variance-covariance matrix of MLEs can be estimated by

$$\left[I\left(\underset{\sim}{\theta}\right)\right]^{-1}_{\underset{\sim}{\theta}=\underset{\sim}{\hat{\theta}}_{ML}} = \left[-E\left(\frac{\partial^2 \ln L}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right)\right]^{-1}_{|\underset{\sim}{\theta}=\underset{\sim}{\hat{\theta}}_{ML}} .$$

However, this estimator may not be available because of the difficulty in calculating the necessary expectations to get the information matrix. There do, however, exist two alternative consistent estimators.

Alternative 1 (Direct use of Hessian)

$$\left[\hat{I}\left(\underset{\sim}{\hat{\theta}}\right)\right]^{-1} = \left[-H\left(\underset{\sim}{\hat{\theta}}\right)\right]^{-1} = \left[-\frac{\partial^2 \ln L\left(\underset{\sim}{\theta}\right)}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right]^{-1}_{|\underset{\sim}{\theta}=\underset{\sim}{\hat{\theta}}}$$

where $\text{plim}\, \underset{\sim}{\hat{\theta}} = \underset{\sim}{\theta}$.

Of course, $\underset{\sim}{\hat{\theta}}$ could be chosen as the maximum likelihood estimator $\underset{\sim}{\hat{\theta}}_{ML}$.

Alternative 2 (BHHH or outer product of gradients)

$$\left[\hat{\hat{I}}\left(\underset{\sim}{\hat{\theta}}\right)\right]^{-1} = \left[\sum_{i=1}^{N} \underset{\sim}{\hat{g}_i}\, \underset{\sim}{\hat{g}_i'}\right]^{-1} = \left[\hat{G}\hat{G}'\right]^{-1}$$

where $\hat{G} = \left[\underset{\sim}{\hat{g}_1} \quad \underset{\sim}{\hat{g}_2} \quad \cdots \quad \underset{\sim}{\hat{g}_N}\right]$ and $\underset{\sim}{\hat{g}_i} = \dfrac{\partial \ln L\left(\underset{\sim}{X_i};\underset{\sim}{\theta}\right)}{\partial \underset{\sim}{\theta}}_{|\underset{\sim}{\theta}=\underset{\sim}{\theta}_{ML}}$

The first alternative requires the derivation of the second derivations of the log likelihood function (i.e., the Hessian matrix) while the second alternative only requires the derivation of the first derivatives of the log likelihood function (i.e., the gradient vector).

Alternative 3 (QMLE/Sandwich Variance-Covariance Matrix)

For the **Type 2 ML estimator** (see Davidson and McKinnon (2004, p. 406)) it can be shown that

$$Var(\underset{n\to\infty}{p\lim}\, n^{1/2}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta_0})) = H(\underset{\sim}{\theta_0})^{-1} I(\underset{\sim}{\theta_0}) H(\underset{\sim}{\theta_0})^{-1}.$$

Therefore, a consistent estimate of the variance-covariance matrix of the Type 2 ML estimator is

$$V\hat{a}r_{QMLE}(\hat{\underset{\sim}{\theta}}) = \left[H\left(\hat{\underset{\sim}{\theta}}\right)\right]^{-1} \hat{G}\hat{G}' \left[H(\hat{\underset{\sim}{\theta}})\right]^{-1}.$$

(A Type 2 ML estimator is defined as a solution to the first order conditions for the likelihood function, namely, $\underset{\sim}{g}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}}) = \underset{\sim}{0}$, where $\underset{\sim}{g}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})$ is the gradient (score) vector of the likelihood function evaluated at $\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}$.) In the case that the likelihood function has been properly specified, this estimator has little to recommend it since it is more tedious to calculate than the BHHH estimate. However, unlike the information matrix estimate or the Hessian and BHHH estimates, it is still valid even when the information matrix equality does not hold. Since this equality does not tend to hold when the likelihood function is misspecified, the QMLE estimate of the covariance matrix can still be consistent under general circumstances. See White (1982) and Gourieroux, Monfort, and Trognon (1984) for further discussion. Thus, when the proper specification of the likelihood function is in doubt, one would be well served to compare the inferences drawn for the conventional ML covariance estimates with the inferences drawn when using the QMLE covariance matrix.

Asymptotic Inference for Nonlinear Functions of Parameters - The Delta Method

A) The Univariate Delta method

Let $\hat{\theta}$ be a consistent, asymptotically normal estimate of θ. Then,

Theorem: If $\sqrt{N}\left(\hat{\theta}-\theta\right)\overset{asy}{\rightarrow}N\left(0,\sigma^2\right)$ and if $c(\theta)$ is a continuous function of θ not dependent on N, then the distribution of $c\left(\hat{\theta}\right)$ is given by

$$\sqrt{N}\left(c\left(\hat{\theta}\right)-c(\theta)\right)\overset{asy}{\rightarrow}N\left(0,[c'(\theta)]^2\sigma^2\right) \text{ where } c'(\theta)=\frac{dc(\theta)}{d\theta}.$$

Obviously, a consistent estimate of $Var\left(c\left(\hat{\theta}\right)\right)$ is $\left[c'\left(\hat{\theta}\right)\right]^2\hat{\sigma}^2$.

B) The Multivariate Delta method

Let $\underset{\sim}{\theta}$ be a $K\times 1$ vector of parameters. Furthermore, let $\underset{\sim}{\hat{\theta}}$ be a consistent, asymptotically normal estimate of $\underset{\sim}{\theta}$. Then,

Theorem: If $\sqrt{N}\left(\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta}\right)\overset{asy}{\rightarrow}N\left(\underset{\sim}{0},\Sigma\right)$ and if $\underset{\sim}{c}\left(\underset{\sim}{\theta}\right)$ is a $J\times 1, J\leq K$, continuous function of $\underset{\sim}{\theta}$ not dependent on N, then

$$\sqrt{N}\left(\underset{\sim}{c}\left(\underset{\sim}{\hat{\theta}}\right)-\underset{\sim}{c}\left(\underset{\sim}{\theta}\right)\right)\overset{asy}{\rightarrow}N\left(\underset{\sim}{0},\ C\left(\underset{\sim}{\theta}\right)'\Sigma C\left(\underset{\sim}{\theta}\right)\right)$$

$$\text{where } C\left(\underset{\sim}{\theta}\right)=\begin{bmatrix}\dfrac{dc_1\left(\underset{\sim}{\theta}\right)}{d\underset{\sim}{\theta}} & \dfrac{dc_2\left(\underset{\sim}{\theta}\right)}{d\underset{\sim}{\theta}} & \cdots & \dfrac{dc_J\left(\underset{\sim}{\theta}\right)}{d\underset{\sim}{\theta}}\end{bmatrix}.$$

Obviously, a consistent estimate of $Var\left(\underset{\sim}{c}\left(\underset{\sim}{\hat{\theta}}\right)\right)$ is given by $C\left(\underset{\sim}{\hat{\theta}}\right)'\hat{\Sigma}C\left(\underset{\sim}{\hat{\theta}}\right)$, where $\hat{\Sigma}$ is a consistent estimator of Σ and $C(\underset{\sim}{\hat{\theta}})$ is a consistent estimator of $C(\underset{\sim}{\theta})$.