Lecture 3

Maximum Likelihood Estimation for the Classical Normal Linear Regression Model, Examination of MVU Efficiency of Estimators, and the "Super" Theorem of Maximum Likelihood Estimation

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad ; \underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I)$$

$$L(\underset{\sim}{y} / X, \underset{\sim}{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} * \exp\left(\frac{-\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}}{2\sigma^2}\right)$$

$$l = \ln L(y / X, \underset{\sim}{\beta}, \sigma^2) = -\frac{N}{2} * \ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\beta})' * (\underset{\sim}{y} - X\underset{\sim}{\beta})$$

$$\frac{\partial l}{\partial \underset{\sim}{\beta}} = \frac{\partial}{\partial \underset{\sim}{\beta}}\left[-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})\right]$$

$$= -\frac{1}{2\sigma^2} * \frac{\partial}{\partial \underset{\sim}{\beta}}\left[\underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}'X'\underset{\sim}{y} - \underset{\sim}{y}'X'\underset{\sim}{\beta} + \underset{\sim}{\beta}'X'X\underset{\sim}{\beta}\right]$$

$$= -\frac{1}{2\sigma^2} * \left[\frac{\partial}{\partial \underset{\sim}{\beta}}\left(-2\underset{\sim}{y}'X\underset{\sim}{\beta}\right) + \frac{\partial}{\partial \underset{\sim}{\beta}}\left(\underset{\sim}{\beta}'X'X\underset{\sim}{\beta}\right)\right]$$

$$= -\frac{1}{2\sigma^2} * \left[-2X'\underset{\sim}{y} + 2X'X\underset{\sim}{\beta}\right]$$

$$= \frac{1}{\sigma^2} * \left[X'\underset{\sim}{y} - X'X\underset{\sim}{\beta}\right]$$

F.O.C. with respect to $\underset{\sim}{\beta}$:

$$\frac{1}{\hat{\sigma}^2} * \left[X'\underset{\sim}{y} - X'X\underset{\sim}{\hat{\beta}}\right] = \underset{\sim}{0}$$

$$\left[X'\underset{\sim}{y} - X'X\underset{\sim}{\hat{\beta}}\right] = \underset{\sim}{0}$$

$$X'\underset{\sim}{y} = X'X\underset{\sim}{\hat{\beta}} \qquad \text{;(Normal Equation of OLS)}$$

$$\underset{\sim}{\hat{\beta}} = \left(X'X\right)^{-1}X'\underset{\sim}{y}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2}\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4}\left(\underset{\sim}{y} - X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y} - X\underset{\sim}{\beta}\right)$$

F.O.C. with respect to σ^2:

$$-\frac{N}{2}\left(\frac{1}{\hat{\sigma}^2}\right)+\frac{1}{2\hat{\sigma}^4}\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)=0$$

$$-\frac{N}{2}+\frac{1}{2\hat{\sigma}^2}\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)=0$$

$$\frac{1}{2\hat{\sigma}^2}\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)=\frac{N}{2}$$

$$\hat{\sigma}^2=\frac{\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)}{N}=\frac{\underset{\sim}{e}'\underset{\sim}{e}}{N} \qquad \text{; where } \underset{\sim}{e}=\left(\underset{\sim}{y}-\underset{\sim}{\hat{y}}\right)=\left(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}}\right)$$

$$H(\underset{\sim}{\beta},\sigma^2)=\begin{bmatrix}\dfrac{\partial^2 l}{\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}'} & \dfrac{\partial^2 l}{\partial\underset{\sim}{\beta}\,\partial\sigma^2}\\ \dfrac{\partial^2 l}{\partial\sigma^2\,\partial\underset{\sim}{\beta}'} & \dfrac{\partial^2 l}{(\partial\sigma^2)^2}\end{bmatrix}_{(K+1)*(K+1)}$$

Let’s look at the parts of the Hessian:

$$\frac{\partial^2 l}{\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}'}=\frac{\partial l}{\partial\underset{\sim}{\beta}'}\left[\frac{\partial l}{\partial\underset{\sim}{\beta}}\right]$$

$$=\frac{\partial l}{\partial\underset{\sim}{\beta}'}\left[\left(X'\underset{\sim}{y}-X'X\underset{\sim}{\beta}\right)\left(\frac{1}{\sigma^2}\right)\right]$$

$$=-\frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial\sigma^2}=\frac{N}{2\sigma^4}-\frac{1}{\sigma^6}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)$$

$$\frac{\partial^2 l}{\partial\underset{\sim}{\beta}\,\partial\sigma^2}=-\frac{1}{\sigma^4}X'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)$$

$$\frac{\partial^2 l}{\partial\sigma^2\,\partial\underset{\sim}{\beta}'}=\left[-\frac{1}{\sigma^4}X'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right]'$$

$$=-\frac{1}{\sigma^4}\underset{\sim}{y}'X-X'X\underset{\sim}{\beta}$$

$$=-\frac{1}{\sigma^4}\left(\underset{\sim}{y}'-X'\underset{\sim}{\beta}\right)'X$$

$$I(\beta,\sigma^2) = -E\left[H(\underset{\sim}{\beta},\sigma^2) \right]$$

$$= -E\begin{bmatrix} \dfrac{-X'X}{\sigma^2} & -\dfrac{1}{\sigma^4}X'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) \\ -\dfrac{1}{\sigma^4}\left(\underset{\sim}{y}'-X'\underset{\sim}{\beta}\right)'X & \dfrac{N}{2\sigma^4}-\dfrac{1}{\sigma^6}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) \end{bmatrix}$$

$$= -\begin{bmatrix} \dfrac{-X'X}{\sigma^2} & -\dfrac{1}{\sigma^4}X'E\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) \\ -\dfrac{1}{\sigma^4}E\left(\underset{\sim}{y}'-X'\underset{\sim}{\beta}\right)'X & E\left[\dfrac{N}{2\sigma^4}-\dfrac{1}{\sigma^6}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right] \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{X'X}{\sigma^2} & \dfrac{1}{\sigma^4}X'\underset{\sim}{0} \\ \dfrac{1}{\sigma^4}\underset{\sim}{0}'X & \dfrac{N}{2\sigma^4}-\dfrac{1}{\sigma^6}E\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{X'X}{\sigma^2} & \underset{\sim}{0} \\ \underset{\sim}{0}' & -\dfrac{N}{2\sigma^4}+\dfrac{1}{\sigma^6}N\sigma^2 \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{X'X}{\sigma^2} & \underset{\sim}{0} \\ \underset{\sim}{0}' & \dfrac{N}{2\sigma^4} \end{bmatrix}$$

Then,

$$CRLB = \left[I\left(\underset{\sim}{\beta},\sigma^2\right)\right]^{-1}$$

$$= \begin{bmatrix} \sigma^2\left(X'X\right)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0}' & \dfrac{2\sigma^4}{N} \end{bmatrix}$$

What about the MVU efficiency of $\hat{\beta}$ and $\hat{\sigma}^2$?

It can be shown that

$$Var(\hat{\beta}) = E\left[\left(\underset{\sim}{\hat{\beta}}-E\left(\underset{\sim}{\hat{\beta}}\right)\right)\left(\underset{\sim}{\hat{\beta}}-E\left(\underset{\sim}{\hat{\beta}}\right)\right)'\right]_{K*K}$$

$$= E\left[\left(\underset{\sim}{\hat{\beta}}-\underset{\sim}{\beta}\right)\left(\underset{\sim}{\hat{\beta}}-\underset{\sim}{\beta}\right)'\right]$$

$$= \sigma^2\left(X'X\right)^{-1}$$

And $\hat{\underset{\sim}{\beta}}$ has a variance-covariance matrix that achieves the CRLB for estimation of $\underset{\sim}{\beta}$.

Therefore, by CR THM, $\hat{\underset{\sim}{\beta}}$ is a MVU estimation of $\underset{\sim}{\beta}$. (Uniqueness is not guaranteed)

But using the complete-sufficient statistics approach verifies that $\hat{\underset{\sim}{\beta}}$ is the MVU estimation of $\underset{\sim}{\beta}$.

What about the MVU efficiency of $S^2 = \hat{\sigma}^2 \frac{N}{(N-K)}$?

THM: $\frac{N\hat{\sigma}^2}{\sigma^2} = \frac{(N-K)S^2}{\sigma^2} \sim \chi^2_{N-K}$

Proof: See proof using Quadratic form theory in back of Hogg, McKean, Craig.

Then, $E(S^2) = \sigma^2$ by algebraic manipulation of above THM. Therefore, S^2 is a MVU candidate. Likewise, from the above theorem, it can be shown that

$Var(S^2) = \frac{2\sigma^4}{(N-K)} > \frac{2\sigma^4}{N}$. Then, $Var(S^2)$ does not achieve the CRLB for

estimation of σ^2 and the MVU efficiency of S^2 is still undetermined. However, using the complete-sufficient statistics approach of Mathematical Statistics, it can be shown that S^2 is the MVU estimation of σ^2 in the classical normal linear regression model.

<u>Concerning Asymptotic Properties of MLE Estimators of Parameters of Regular Densities (The “Super” Theorem of ML Estimation)</u>

THM: Let $L(\underset{\sim}{X};\underset{\sim}{\theta})$ be the likelihood function of a regular density. (For definition of Regular; see Greene for example) Under certain regularity conditions, we have the Maximum Likelihood Estimator, say $\hat{\underset{\sim}{\theta}}_{ML}$,

(i) Consistency: $P\lim \hat{\underset{\sim}{\theta}}_{ML} = \underset{\sim}{\theta}$

(ii) Asymptotic Normality:

$$\hat{\underset{\sim}{\theta}}_{ML} \xrightarrow{asy} N\left(\underset{\sim}{\theta}, \{I(\underset{\sim}{\theta})\}^{-1}\right) \quad \text{;where } I(\underset{\sim}{\theta}) = -E[H(\underset{\sim}{\theta})]$$

[Note: The above is a somewhat ”lax” statement. More properly, we should say

$$\sqrt{N}\left(\hat{\underset{\sim}{\theta}}_{ML} - \underset{\sim}{\theta}\right) \overset{asy}{\sim} N(\underset{\sim}{\theta}, V) \text{ where } V = p\lim_{N\to\alpha}\left[\frac{I\left(\hat{\underset{\sim}{\theta}}_{ML}\right)}{N}\right]^{-1} = \lim_{N\to\alpha}\left[\frac{I(\underset{\sim}{\theta})}{N}\right]^{-1}$$

Therefore, a finite approximation of V is $\hat{V} = \dfrac{I\left(\hat{\underset{\sim}{\theta}}_{ML}\right)}{N}$. Then, a finite approximation of the variance covariance matrix of $\hat{\underset{\sim}{\theta}}_{ML}$ is $\dfrac{\hat{V}}{N} = I\left(\hat{\underset{\sim}{\theta}}_{ML}\right)^{-1}$.]

(iii) Asymptotic Efficiency: $Var(\hat{\theta}_{ML})$ achieves the "Asymptotic" Cramer-Rao Lower Bound in infinite samples. This lower bound is denoted by

$$V = \underset{N \to \alpha}{p\lim}\left[\frac{I\left(\hat{\underset{\sim}{\theta}}_{ML}\right)}{N}\right]^{-1} = \lim_{N \to \alpha}\left[\frac{I\left(\underset{\sim}{\theta}\right)}{N}\right]^{-1} .$$

(iv) Invariance: Let c(.) be a continuous function. Then the maximum likelihood estimator of $c(\theta)$ is $c(\hat{\theta}_{ML})$.

For a discussion of the regularity conditions of the above theorem see Econometric Analysis, 5th ed. by William H. Greene (Prentice-Hall, 2003), p. 474. For a proof of the theorem see Greene, pp. 476 – 480.