MAXIMUM-LIKELIHOOD AND THE CLASSICAL LINEAR MODEL

Consider the classical regression model

(1) $$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + \cdots + x_{tk}\beta_k + \varepsilon_t,$$

where $t = 1, \ldots, T$ is the index for T successive observations. Let us assume that the disturbances ε_t are distributed normally, independently and identically with $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = \sigma^2$ for all t. The equation above can be written in summary form as

(2) $$\begin{aligned} y_t &= x_{t.}\beta + \varepsilon_t, \\ \text{where} \quad x_{t.} &= [x_{t1}, x_{t1} \ldots x_{t1}], \\ \text{and} \quad \beta &= [\beta_1, \beta_2, \cdots, \beta_k]'; \end{aligned}$$

and the set of T such equations can be compiled as

(3) $$y = X\beta + \varepsilon.$$

Let us assume that the disturbances ε_t, which are the elements of the vector $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t]'$, are distributed independently and identically according to a normal distribution

(4) $$N(\varepsilon_t; 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-1}{2\sigma^2}(y_t - x_{t.}\beta)^2 \right\}.$$

Then, if the the vectors $x_{t.}$ are taken as data, the observations $y_t; t = 1, \ldots, T$ have density functions $N(y_t; x_t.\beta, \sigma^2)$ which are of the same form as those of the disturbances, and the likelihood function of β and σ^2, based on the sample, is

(5) $$L = \prod_{T=1}^{T} N(y_t; x_{t.}\beta, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left\{ \frac{-1}{2\sigma^2}(y - X\beta)'(y - X\beta) \right\}.$$

The logarithm of this function

(6) $$L^*(\beta, \sigma) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta);$$

and, in order to pursue the theory of maximum-likelihood estimation, we need to find the first and second-order derivatives of L^* with respect to its arguments which are the unknown parameters. The requisite derivatives are as follows:

$$\frac{\partial L^*}{\partial \beta} = \frac{1}{\sigma^2}(y - X\beta)'X, \tag{7}$$

$$\frac{\partial L^*}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta), \tag{8}$$

$$\frac{\partial(\partial L^*/\partial \beta)'}{\partial \beta} = -\frac{1}{\sigma^2}X'X, \tag{9}$$

$$\frac{\partial(\partial L^*/\sigma^2)'}{\partial \sigma^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta), \tag{10}$$

$$\frac{\partial(\partial L^*/\beta)'}{\partial \sigma^2} = -\frac{1}{\sigma^4}X'(y - X\beta). \tag{11}$$

To find the maximum-likelihood estimator of β, we set the derivative of equation (7) to zero. This gives a first-order condition for the maximisation of the log-likelihood function; and the solution of the equation is the estimator

$$\tilde{\beta} = (X'X)^{-1}X'y. \tag{12}$$

This is nothing but the ordinary least-squares estimator of β which is usually denoted by $\hat{\beta}$.

To find the maximum-likelihood estimator of σ^2, we set the derivative of equation (8) to zero. Then we multiply the resulting first-order condition by a factor of $2\sigma^4/T$. Rearranging the result gives a maximum-likelihood estimating equation in the form of

$$\sigma^2(\beta) = \frac{1}{T}(y - X\beta)'(y - X\beta). \tag{13}$$

Setting $\beta = \tilde{\beta}$ gives an estimator for σ^2 which differs from the usual unbiased estimator which is

$$\hat{\sigma}^2 = \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}). \tag{14}$$

However, the differences between $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ vanish as $T \to \infty$.

The theory of maximum-likelihood estimation indicates that, if $\tilde{\theta}$ is the maximum-likelihood estimator, which is obtained by evaluating the first-order condition $\partial L^*/\partial\theta$ for the maximisation of the log-likelihood function $L^*(\theta) = \ln L(\theta)$, then $\sqrt{T}(\tilde{\theta} - \theta)$ has the limiting normal distribution $N(0, M)$, where

$$M = -\text{plim}\left\{\frac{1}{T}\frac{\partial(\partial L^*/\partial\theta)'}{\partial\theta}\right\}^{-1}. \tag{15}$$

Here the derivatives may be evaluated at the point of the maximum-likelihood estimates which is know to tend in probability to the true parameter value θ as the size of the sample increases.

In the present case, we need to find the limits derivatives under (9) (10) and (11) scaled by the factor T^{-1}. With $\beta = \hat{\beta} = (X'X)^{-1}X'y$, the numerator of equation (11) becomes $X'(y - X\hat{\beta}) = X'e = 0$; and so the term can be set to zero. When the LHS of equation (10) is divided by T and β is set to $\tilde{\beta}$, it will be recognised that the equation incorporates the expression $(y-X\tilde{\beta})'(y-X\tilde{\beta})/T = \tilde{\sigma}^2$ which stands from the consistent maximum-likelihood estimator of σ^2. The latter is subject to the condition that $\text{plim}(T \to \infty)\tilde{\sigma}^2 = \sigma^2$. Finally, it will be recognised that, under standard assumptions, $\text{plim}(T^{-1}X'X) = M_{xx}$ is a matrix of finite-valued constants. The conclusion is that

$$\text{plim}\frac{1}{T}\frac{\partial(\partial L^*/\partial\beta)'}{\partial\beta} = -\frac{1}{\sigma^2}M, \tag{16}$$

$$\text{plim}\frac{1}{T}\frac{\partial(\partial L^*/\sigma^2)'}{\partial\sigma^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^4}, \tag{17}$$

$$\text{plim}\frac{1}{T}\frac{\partial(\partial L^*/\beta)'}{\partial\sigma^2} = 0. \tag{18}$$

It follows that, in the case of the classical linear regression model, we have

$$M = \begin{bmatrix} M_{\beta\beta} & M_{\beta\sigma} \\ M_{\sigma\beta} & M_{\sigma\sigma} \end{bmatrix} = \begin{bmatrix} M_{xx}/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{bmatrix}^{-1}. \tag{19}$$

This gives the asymptotic dispersion matrix for $\sqrt{T}\tilde{\beta}$ and $\sqrt{T}\tilde{\sigma}^2$. In finite samples, the dispersion matrix for $\tilde{\beta}$ and $\tilde{\sigma}^2$ may be approximated by

$$D\begin{pmatrix} \tilde{\beta} \\ \tilde{\sigma}^2 \end{pmatrix} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/T \end{bmatrix}. \tag{20}$$

This may be compared with the dispersion matrix of the ordinary least-squares estimates under conditions of normality, which is given by

$$D\begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/(T-k) \end{bmatrix}. \tag{21}$$