Method of Maximum Likelihood Estimation

The method of maximum likelihood estimation (MLE) is an alternative to the method of least squares. This method is based on the assumption that the random disturbance term is independently, identically and normally distributed with zero mean and constant variance.

Model: $Y_i = \alpha + \beta X_i + u_i$

Assumption: $u_i \sim IIN(0, \sigma^2)$

Hence, $Y_i \sim IN(\alpha + \beta X_i, \sigma^2)$

This means that Y_i are independently and normally distributed with mean $\alpha + \beta X_i$ and variance σ^2

Hence, the joint probability density function of $Y_1, Y_2, ..., Y_n$ is:

$$f(Y_1, Y_2, ..., Y_n | \alpha + \beta X_i, \sigma^2) = f_1(Y_1 | \alpha + \beta X_i, \sigma^2) \times f_2(Y_2 | \alpha + \beta X_i, \sigma^2) \times .... \times f_n(Y_n | \alpha + \beta X_i, \sigma^2)$$

with

$$f_i(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i - \alpha - \beta X_i)^2}{2\sigma^2}}$$

Hence, $f(Y_1, Y_2, ..., Y_n | \alpha + \beta X_i, \sigma^2) = \dfrac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}}$

Hence, the Likelihood Function is: $LF(\alpha, \beta, \sigma^2) = \dfrac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}}$

$$\Rightarrow \quad ln(LF) = -n\,ln(\sigma) - \frac{n}{2} ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}$$

$$\Rightarrow \quad ln(LF) = -\frac{n}{2} ln(\sigma^2) - \frac{n}{2} ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}$$

If $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\sigma}^2$ are the MLE of α, β and σ^2 respectively, we have

$$ln(LF) = -\frac{n}{2} ln(\tilde{\sigma}^2) - \frac{n}{2} ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \tilde{\alpha} - \tilde{\beta} X_i)^2}{\tilde{\sigma}^2}$$

From the first-order condition of optimization with respect to $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\sigma}^2$ we get,

$\dfrac{\partial ln(LF)}{\partial \tilde{\alpha}} = -\dfrac{1}{\tilde{\sigma}^2} \sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)(-1) = 0$ Or, $\sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i) = 0$ (1)

$\dfrac{\partial ln(LF)}{\partial \tilde{\beta}} = -\dfrac{1}{\tilde{\sigma}^2} \sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)(-X_i) = 0$ Or, $\sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)(X_i) = 0$ (2)

$$\frac{\partial ln(LF)}{\partial \tilde{\sigma}^2} = -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum(Y_i - \tilde{\alpha} - \tilde{\beta}X_i)^2 = 0 \text{ Or, } \tilde{\sigma}^2 = \frac{1}{n}\sum(Y_i - \tilde{\alpha} - \tilde{\beta}X_i)^2 = \frac{1}{n}\sum \hat{u}_i^2 \quad (3)$$

Thus, from (1) and (2), we get the same estimator of the coefficient as the OLS estimators. This indicates that, in the linear regression model, the maximum likelihood estimators would be the same as the OLS estimators and possess all the desirable properties. In other words, because the OLS estimators are unbiased, consistent and efficient, so are the maximum likelihood estimators.

From (3) we get the maximum likelihood estimator of the variance of the random disturbance term,

$$\tilde{\sigma}^2 = \frac{1}{n}\sum(Y_i - \tilde{\alpha} - \tilde{\beta}X_i)^2 = \frac{1}{n}\sum \hat{u}_i^2$$

Hence, $E(\tilde{\sigma}^2) = \frac{1}{n}E\left(\sum \tilde{u}_i^2\right) = \frac{(n-2)\sigma^2}{n} = \sigma^2 - \frac{2}{n}\sigma^2 \neq \sigma^2$

Thus, $\tilde{\sigma}^2$ is a biased estimator of σ^2.

However, in case of large sample, as $n \to \infty$, $\left(\sigma^2 - \frac{2}{n}\sigma^2\right) \to \sigma^2$. This results in an unbiased estimator of σ^2 in large sample.

Thus the method of MLE is a large sample method and has broader application as it can also be applied to the regression models that are non-linear in parameters. Nonetheless, in general, the method of OLS is applied in estimating regression models as it is easy to apply, but both the methods result in identical estimators of the coefficients. Further, the OLS estimator of the variance of the random disturbance term is unbiased even in small sample.