Maximum Likelihood Estimators

2021-09-08

OLS Estimators are Maximum Likelihood (ML) Estimators

Given the assumptions of the strong classical model (**A.1–A.7**), we can show that the least squares estimators are also the maximum likelihood estimators. Recall that the likelihood is the probability of a set of parameters, computed as the joint density of the data, given a set of observations under a particular probability distribution.

The joint density of the errors is:

$$\prod_{i=1}^{n} f(\epsilon_i;\, 0, \sigma_\epsilon^2) = (2\pi\sigma_\epsilon^2)^{-n/2} \times e^{-\frac{1}{2\sigma_\epsilon^2}\sum \epsilon_i^2}$$

Using assumptions **A.5** (independence of errors) and **A.7** (conditional normality of errors), and the fact that ϵ_i is a linear function of y_i, we can write the likelihood of the parameters given the observations and the normal probability distribution of **y** as,

$$\mathcal{L}\left(\beta_0, \beta_1, \ldots, \beta_k, \sigma_\epsilon^2 \mid \mathbf{y}, n\right) = (2\pi\sigma_\epsilon^2)^{-n/2} \times e^{-\frac{1}{2\sigma_\epsilon^2}\sum \left(\mathbf{y}-\mathbf{X}\beta\right)^2}$$

Or, in log-likelihood form,

$$\log \mathcal{L}\left(\beta_0, \beta_1, \ldots, \beta_k, \sigma_\epsilon^2 \mid \mathbf{y}, n\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum \left(\mathbf{y}-\mathbf{X}\beta\right)^2$$

We can again use calculus (beyond the scope of this course) to optimize this function.

Differentiating this expression with respect to each of the parameters, we get:

$$\begin{aligned}
\frac{\partial \log \mathcal{L}}{\partial \beta_0} &= \frac{1}{\sigma_\epsilon^2} \sum (\mathbf{y} - \mathbf{X}\beta) \\
\frac{\partial \log \mathcal{L}}{\partial \beta_1} &= \frac{1}{\sigma_\epsilon^2} \sum x_{1i}(\mathbf{y} - \mathbf{X}\beta) \\
&\vdots \\
\frac{\partial \log \mathcal{L}}{\partial \beta_k} &= \frac{1}{\sigma_\epsilon^2} \sum x_{ki}(\mathbf{y} - \mathbf{X}\beta) \\
\frac{\partial \log \mathcal{L}}{\partial \sigma_\epsilon^2} &= -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \sum (\mathbf{y} - \mathbf{X}\beta)^2
\end{aligned} \tag{1}$$

We can then set each of these equal to zero and solve.

Solving these equations, we find that the maximum likelihood estimators for the regression coefficients are equivalent to the OLS estimators of these parameters. We also find that,

$$\begin{aligned}
\hat{\sigma}_\epsilon^2 &= \frac{1}{n} \sum (\mathbf{y} - \mathbf{X}\beta)^2 \\
&= \frac{1}{n} \sum (e_i)^2
\end{aligned}$$

Thus the maximum likelihood estimate for the error variance is not the same as the OLS estimate for error variance (the OLS version divides the sum of the errors by $n-2$).

Inference with the ML Estimators

The maximum likelihood estimators for the coefficients (which are the same as the OLS estimators) possess several asymptotic (large sample) properties; most importantly normality. Because of this, with a large sample size, the sampling distribution for ML estimates will also be approximately t-distributed with $n-k-1$ degrees of freedom.

References