THE STATISTICAL PROPERTIES OF THE OLS ESIMATOR: UNBIASEDNESS AND EFFICIENCY

Some Statistical Properties of the Estimator

The expectation or mean vector of $\hat{\beta}$, and its dispersion matrix as well, may be found from the expression

$$\text{(1)} \qquad \begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta+\varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned}$$

The expectation is

$$\text{(2)} \qquad \begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon)\\ &= \beta.\end{aligned}$$

Thus $\hat{\beta}$ is an unbiased estimator. The deviation of $\hat{\beta}$ from its expected value is $\hat{\beta} - E(\hat{\beta}) = (X'X)^{-1}X'\varepsilon$. Therefore the dispersion matrix, which contains the variances and covariances of the elements of $\hat{\beta}$, is

$$\text{(3)} \qquad \begin{aligned}D(\hat{\beta}) &= E\Big[\{\hat{\beta} - E(\hat{\beta})\}\{\hat{\beta} - E(\hat{\beta})\}'\Big]\\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

The Gauss–Markov theorem asserts that $\hat{\beta}$ is the unbiased linear estimator of least dispersion. This dispersion is usually characterised in terms of the variance of an arbitrary linear combination of the elements of $\hat{\beta}$, although it may also be characterised in terms of the determinant of the dispersion matrix $D(\hat{\beta})$. Thus

(4) If $\hat{\beta}$ is the ordinary least-squares estimator of β in the classical linear regression model, and if β^* is any other linear unbiased estimator of β, then $V(q'\beta^*) \geq V(q'\hat{\beta})$ where q is any constant vector of the appropriate order.

Proof. Since $\beta^* = Ay$ is an unbiased estimator, it follows that $E(\beta^*) = AE(y) = AX\beta = \beta$ which implies that $AX = I$. Now let us write $A = (X'X)^{-1}X' + G$. Then $AX = I$ implies that $GX = 0$. It follows that

$$\text{(5)} \qquad \begin{aligned}D(\beta^*) &= AD(y)A'\\ &= \sigma^2\{(X'X)^{-1}X' + G\}\{X(X'X)^{-1} + G'\}\\ &= \sigma^2(X'X)^{-1} + \sigma^2 GG'\\ &= D(\hat{\beta}) + \sigma^2 GG'.\end{aligned}$$

Therefore, for any constant vector q of order k, there is the identity

$$\begin{aligned} V(q'\beta^*) &= q'D(\hat{\beta})q + \sigma^2 q'GG'q \\ &\geq q'D(\hat{\beta})q = V(q'\hat{\beta}); \end{aligned} \tag{6}$$

and thus the inequality $V(q'\beta^*) \geq V(q'\hat{\beta})$ is established.

Estimating the Variance of the Disturbance

The principle of least squares does not, of its own, suggest a means of estimating the disturbance variance $\sigma^2 = V(\varepsilon_t)$. However it is natural to estimate the moments of a probability distribution by their empirical counterparts. Given that $e_t = y_t - x_{t.}\hat{\beta}$ is an estimate of ε_t, it follows that $T^{-1}\sum_t e_t^2$ may be used to estimate σ^2. However, it transpires that this is biased. An unbiased estimate is provided by

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{T-k}\sum_{t=1}^{T} e_t^2 \\ &= \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned} \tag{7}$$

The unbiasedness of this estimate may be demonstrated by finding the expected value of $(y - X\hat{\beta})'(y - X\hat{\beta}) = y'(I - P)y$. Given that $(I - P)y = (I - P)(X\beta + \varepsilon) = (I - P)\varepsilon$ in consequence of the condition $(I - P)X = 0$, it follows that

$$E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = E(\varepsilon'\varepsilon) - E(\varepsilon'P\varepsilon). \tag{8}$$

The value of the first term on the RHS is given by

$$E(\varepsilon'\varepsilon) = \sum_{t=1}^{T} E(e_t^2) = T\sigma^2. \tag{9}$$

The value of the second term on the RHS is given by

$$\begin{aligned} E(\varepsilon'P\varepsilon) &= \text{Trace}\big\{E(\varepsilon'P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon'P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon\varepsilon'P)\big\} \\ &= \text{Trace}\big\{E(\varepsilon\varepsilon')P\big\} = \text{Trace}\big\{\sigma^2 P\big\} = \sigma^2\text{Trace}(P) \\ &= \sigma^2 k. \end{aligned} \tag{10}$$

The final equality follows from the fact that $\text{Trace}(P) = \text{Trace}(I_k) = k$. Putting the results of (9) and (10) into (8), gives

$$E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = \sigma^2(T - k); \tag{11}$$

and, from this, the unbiasedness of the estimator in (7) follows directly.