Yet Another Proof of the Gauss-Markov Theorem — DA Freedman

Statistics 215 — September 2006

Suppose X is fixed (not random), $n \times p$, of full rank p. Suppose

$$Y = X\beta + \epsilon \tag{1}$$

where

$$E(\epsilon) = 0_{n\times 1}\,, \quad \operatorname{cov}(\epsilon) = \sigma^2 I_{n\times n}. \tag{2}$$

THEOREM. GAUSS-MARKOV. The OLS estimator is BLUE.

The acronym BLUE stands for Best Linear Unbiased Estimator, i.e., the one with the smallest variance. Let $\gamma = c'\beta$, where c is $p\times 1$: the parameter γ is a linear combination of the components of β. Examples would include β_1, or $\beta_2 - \beta_3$. The OLS estimator for γ is $\hat\gamma = c'\hat\beta = c'(X'X)^{-1}X'Y$. This is unbiased by (2), and $\operatorname{var}(\hat\gamma) = \sigma^2 c'(X'X)^{-1}c$. Let $\tilde\gamma$ be another linear unbiased estimator for γ. Then $\operatorname{var}(\tilde\gamma) \ge \operatorname{var}(\hat\gamma)$, and $\operatorname{var}(\tilde\gamma) = \operatorname{var}(\hat\gamma)$ entails $\tilde\gamma = \hat\gamma$. That is what the theorem says.

Proof. Recall that X is fixed. Since $\tilde\gamma$ is by assumption a linear function of Y, there is an $n\times 1$ vector d with $\tilde\gamma = d'Y = d'X\beta + d'\epsilon$. Then $E(\tilde\gamma) = d'X\beta$ by (2). Since $\tilde\gamma$ is unbiased, $d'X\beta = c'\beta$ for all β. Therefore,

$$d'X = c'. \tag{3}$$

Caution: c is $p \times 1$ but d is $n \times 1$

It will help to have another formula for $\operatorname{var}(\hat\gamma)$. Since

$$\hat\gamma = c'\beta + c'(X'X)^{-1}X'\epsilon,$$

we have

$$\begin{aligned}\operatorname{var}(\hat\gamma) &= \sigma^2\|c'(X'X)^{-1}X'\|^2\\ &= \sigma^2\|X(X'X)^{-1}c\|^2\\ &= \sigma^2\|X(X'X)^{-1}X'd\|^2\\ &= \sigma^2\|Hd\|^2\end{aligned}$$

where H is the hat matrix. On the other hand, $\tilde\gamma = d'X\beta + d'\epsilon$, so $\operatorname{var}(\tilde\gamma) = \sigma^2\|d\|^2$. But

$$d = Hd + (I - H)d,$$

the summands being orthogonal since H is symmetric and idempotent. In particular, $\|d\|^2 = \|Hd\|^2 + \|(I-H)d\|^2$ and $\|d\|^2 \ge \|Hd\|^2$, the inequality being strict unless

$$d = Hd = X(X'X)^{-1}X'd = X(X'X)^{-1}c,$$

i.e., $\tilde\gamma = \hat\gamma$.