Notes on the Gauss-Markov theorem

DA Freedman 15 November 2004

The OLS regression model is
\[ Y = X\beta + \epsilon, \]
where \( Y \) is an \( n \times 1 \) vector of observable random variables, \( X \) is an \( n \times p \) matrix of observable random variables with rank \( p < n \), and \( \epsilon \) is an \( n \times 1 \) vector of unobservable random variables, IID with mean 0 and variance \( \sigma^2 \), independent of \( X \). We can weaken the assumptions on \( \epsilon \), to
\[ E(\epsilon | X) = 0_{n \times 1}, \quad \text{cov}(\epsilon | X) = \sigma^2 I_{n \times n}. \tag{*} \]

**Vector Version of Gauss-Markov.** Assume \((*)\). Suppose \( X \) is fixed (not random). The OLS estimator is BLUE.

The acronym BLUE stands for Best Linear Unbiased Estimator, i.e., the one with the smallest covariance matrix. If \( \hat{\beta} \) is the OLS estimator and \( \tilde{\beta} \) is another linear estimator that is unbiased, then \( \text{cov}(\tilde{\beta}) \geq \text{cov}(\hat{\beta}) \), i.e., \( \text{cov}(\tilde{\beta}) - \text{cov}(\hat{\beta}) \) is a non-negative definite matrix; furthermore, \( \text{cov}(\tilde{\beta}) = \text{cov}(\hat{\beta}) \) implies \( \tilde{\beta} = \hat{\beta} \). That is what the matrix version of the theorem says.

Proof. Recall that \( X \) is fixed. A linear estimator \( \tilde{\beta} \) must be of the form \( MY \), where \( M \) is a \( p \times n \) matrix. Since \( MY = MX\beta + M\epsilon \) and \( E(M\epsilon) = ME(\epsilon) = 0_{n \times 1} \), unbiasedness means that \( MX\beta = \beta \) for all \( \beta \). Thus, \( MX = I_p \times p \), and \( X'M' = I_p \times p \) as well. Furthermore, \( MY = \beta + M\epsilon \).

For \( \hat{\beta}_{\text{OLS}} \), we have \( M = M_0 \) with \( M_0 = (X'X)^{-1}X' \). Let \( \Delta = M - M_0 \). Then
\[
\Delta X = MX - M_0X \\
= MX - (X'X)^{-1}X'X \\
= I_p \times p - I_p \times p = 0_{p \times p}.
\]

So \( \Delta M_0' = \Delta X (X'X)^{-1} = 0_{p \times p} \), and \( M_0\Delta' = 0_{p \times p} \) too. As noted above, \( \text{E}(M\epsilon) = 0 \). And \( E(\epsilon\epsilon') = \sigma^2 I_{n \times n} \). Therefore,
\[
\text{cov}(MY) = \text{cov}(M\epsilon) \\
= E(M\epsilon\epsilon'M) \\
= \sigma^2 MM' \\
= \sigma^2 (M_0 + \Delta)(M_0 + \Delta)' \\
= \sigma^2 (M_0M_0' + \Delta\Delta' + \Delta M_0' + M_0\Delta') \\
= \sigma^2 (M_0M_0' + \Delta\Delta') = \text{cov}(\tilde{\beta}) + \sigma^2 \Delta\Delta' .
\]

Since \( \Delta\Delta' \) is non-negative definite, \( \text{cov}(\tilde{\beta}) \geq \text{cov}(\hat{\beta}) \). Finally, \( \text{cov}(\tilde{\beta}) = \text{cov}(\hat{\beta}) \) implies \( \tilde{\beta} = \hat{\beta} \) because \( \Delta\Delta' = 0_{p \times p} \) implies \( \Delta = 0_{p \times n} \): look at the diagonal of \( \Delta\Delta' \). This completes the proof.
Discussion. *Statistical Models* has the “single-contrast” version of the theorem, which starts with an estimator for the scalar parameter $c' \beta$. The vector version, on the other hand, starts with an estimator for the vector parameter $\beta$. The vector version implies the single-contrast version: take the given contrast $c$; adjoin $p - 1$ linearly independent contrasts; the vector theorem is invariant under linear re-parameterizations of the column space. (The details of this argument, however, may not be entirely transparent.) By a somewhat more direct argument, the single-contrast version implies the vector version: $c' \text{cov}(\hat{\beta}) c \geq \text{var}(c' \hat{\beta}) = c' \text{cov}(\hat{\beta}) c$ for all $c$, i.e., $\text{cov}(\hat{\beta}) \geq \text{cov}(\hat{\beta})$. 