The Gauss-Markov Theorem

In the linear model \( Y = X\beta + \varepsilon \) with \( E\varepsilon = 0 \) and \( \text{Var}\varepsilon = \sigma^2 I \), if \( c' \in C(X') \), then \( \hat{c}'_{\text{OLS}} \) is the (uniformly over all \( EY \in C(X) \) and \( \sigma^2 \)) Best Linear Unbiased Estimator of \( c'\beta \).

(\( \hat{c}'_{\text{OLS}} \) is the \( v'Y \) that among all such linear combinations of the entries of \( Y \) with mean \( c'\beta \) has the smallest variance.)

Proof. Write \( c'(XX)'X = \rho' \) so that

\[ \hat{c}'_{\text{OLS}} = \rho'Y \]

First note that \( \rho = P_X\rho \). Why? \( \rho = X(X'X)^{-1}X'c \) so that \( \rho \in C(X) \) and \( P_X \) is the projection matrix onto \( C(X) \).

Suppose \( v \) is such that \( Ev'Y = c'\beta \ \forall \beta \). This is

\[ v'X\beta = c'\beta \ \forall \beta \]

which implies that

\[ v'X = c' \]

Consider the variance of \( v'Y \).

\[
\text{Var}(v'Y) = \text{Var}(v'Y - \rho'Y + \rho'Y) = \text{Var}((v' - \rho')Y + \rho'Y) = \text{Var}((v' - \rho')Y) + \text{Var}(\rho'Y) + 2\text{Cov}((v' - \rho')Y, \rho'Y)
\]

Now \( \text{Var}((v' - \rho')Y) \geq 0 \) and thus if we can show that the covariance term above is 0, we will be done. But

\[
\text{Cov}((v' - \rho')Y, \rho'Y) = (v' - \rho')(\text{Var}Y)\rho = \sigma^2 (v' - \rho')\rho = \sigma^2 (v'P_X - \rho'P_X)\rho = \sigma^2 (v'X(X'X)^{-1}X' - \rho')\rho = \sigma^2 (c'(X'X)^{-1}X' - \rho')\rho = \sigma^2 (\rho' - \rho')\rho = 0
\]