The Gauss-Markov Theorem

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

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

Proof. Write $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \boldsymbol{\rho}'$ so that

$$\widehat{\mathbf{c}'\boldsymbol{\beta}}_{\mathrm{OLS}} = \boldsymbol{\rho}'\mathbf{Y}$$

First note that $\boldsymbol{\rho} = \mathbf{P}_{\mathbf{X}}\boldsymbol{\rho}$. Why? $\boldsymbol{\rho} = \mathbf{X}\left((\mathbf{X}'\mathbf{X})^{-}\right)'\mathbf{c}$ so that $\boldsymbol{\rho} \in C(\mathbf{X})$ and $\mathbf{P}_{\mathbf{X}}$ is the projection matrix onto $C(\mathbf{X})$.

Suppose $\mathbf{v}$ is such that $\mathrm{E}\mathbf{v}'\mathbf{Y} = \mathbf{c}'\boldsymbol{\beta}\ \forall\boldsymbol{\beta}$. This is

$$\mathbf{v}'\mathbf{X}\boldsymbol{\beta} = \mathbf{c}'\boldsymbol{\beta}\ \forall\boldsymbol{\beta}$$

which implies that

$$\mathbf{v}'\mathbf{X} = \mathbf{c}'$$

Consider the variance of $\mathbf{v}'\mathbf{Y}$.

$$\begin{aligned}
\mathrm{Var}(\mathbf{v}'\mathbf{Y}) &= \mathrm{Var}(\mathbf{v}'\mathbf{Y} - \boldsymbol{\rho}'\mathbf{Y} + \boldsymbol{\rho}'\mathbf{Y}) \\
&= \mathrm{Var}((\mathbf{v}' - \boldsymbol{\rho}')\mathbf{Y} + \boldsymbol{\rho}'\mathbf{Y}) \\
&= \mathrm{Var}\left((\mathbf{v} - \boldsymbol{\rho})'\mathbf{Y}\right) + \mathrm{Var}(\boldsymbol{\rho}'\mathbf{Y}) + 2\mathrm{Cov}\left((\mathbf{v} - \boldsymbol{\rho})'\mathbf{Y}, \boldsymbol{\rho}'\mathbf{Y}\right)
\end{aligned}$$

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

$$\begin{aligned}
\mathrm{Cov}\left((\mathbf{v} - \boldsymbol{\rho})'\mathbf{Y}, \boldsymbol{\rho}'\mathbf{Y}\right) &= (\mathbf{v} - \boldsymbol{\rho})'(\mathrm{Var}\mathbf{Y})\boldsymbol{\rho} \\
&= \sigma^2(\mathbf{v} - \boldsymbol{\rho})'\boldsymbol{\rho} \\
&= \sigma^2(\mathbf{v} - \boldsymbol{\rho})'\mathbf{P}_{\mathbf{X}}\boldsymbol{\rho} \\
&= \sigma^2(\mathbf{v}'\mathbf{P}_{\mathbf{X}} - \boldsymbol{\rho}'\mathbf{P}_{\mathbf{X}})\boldsymbol{\rho} \\
&= \sigma^2\left(\mathbf{v}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' - \boldsymbol{\rho}'\right)\boldsymbol{\rho} \\
&= \sigma^2\left(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' - \boldsymbol{\rho}'\right)\boldsymbol{\rho} \\
&= \sigma^2(\boldsymbol{\rho}' - \boldsymbol{\rho}')\boldsymbol{\rho} \\
&= 0
\end{aligned}$$

□