ELS ex 3.3a: Gauss Markov theorem

Prove the Gauss-Markov theorem: the least squares estimate of a parameter $a^T\beta$ has variance no bigger than that of any other linear unbiased estimate of $a^T\beta$.

KNOWN:

$$\underset{N\times 1}{Y} = \underset{N\times(p+1)}{X}\underset{(p+1)\times 1}{\beta} + \underset{N\times 1}{\varepsilon}, \quad E(\varepsilon)=0, \quad Cov(\varepsilon) = \sigma_\varepsilon^2 I, \text{ thus } E(Y) = X\beta$$

$$Cov(Y) = (X^TX)^{-1}\sigma_\varepsilon^2 \quad N\times N$$

also written as $Var(Y)$

$$\underset{LS}{\hat{\theta}} = a^T\hat{\beta} = a^T(X^TX)^{-1}X^TY, \quad E(\hat{\theta}) = a^T\beta \text{ unbiased}$$

NEW CANDIDATE

$$\tilde{\theta} = c^TY \quad \text{where } c^T = a^T(X^TX)^{-1}X^T + \underbrace{\delta^T}_{\text{extra term}}$$

trick!

Need that $\tilde{\theta}$ is unbiased, that is

$$E(\tilde{\theta}) = E(c^TY) = E(a^T(X^TX)^{-1}X^TY + \delta^TY)$$

$$= a^T\underbrace{(X^TX)^{-1}X^TX}_{I}\beta + \delta^TX\beta = a^T\beta + \delta^TX\beta$$

$$E(\tilde{\theta}) = a^T\beta \iff \delta^TX = 0 \quad \text{requirement}$$

$\rightarrow$ now we have an unbiased $\tilde{\theta}$, but not the LS.

Now turn to the variance and compare to $Var(\hat{\theta})$.

$$\underset{1\times 1}{Var(\tilde{\theta})} = Ver(c^TY) = \underset{1\times N}{c^T}\overbrace{\underset{N\times N}{Var(Y)}}^{\sigma_\varepsilon^2 I}\underset{N\times 1}{c} = \sigma_\varepsilon^2 \underset{1\times N}{c^T}c$$

$$= \sigma_\varepsilon^2\left[(a^T(X^TX)^{-1}X^T + \delta^T)\underbrace{(a^T(X^TX)^{-1}X^T + \delta^T)^T}_{X(X^TX)^{-1}a + \delta}\right]$$

$$
\begin{aligned}
= \sigma_\varepsilon^2 \Big(& a^T (\mathbf{X}^T\mathbf{X})^{-1} \overbrace{\mathbf{X}^T\mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1}}^{I} a \\
&+ a^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \delta \\
&+ \delta^T \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} a \\
&+ \delta^T \delta \Big)
\end{aligned}
$$

$$\overset{0}{\overset{\shortparallel}{(\delta^T \mathbf{X})^T}} = \mathbf{X}^T \delta = 0$$

$$\delta^T \mathbf{X} = 0$$

$$\sum_{i=1}^{N} \delta_i^2 \geq 0$$

$$Var(\tilde{\theta}) = \underbrace{\sigma_\varepsilon^2 \, a^T (\mathbf{X}^T\mathbf{X})^{-1} a}_{Var(\hat{\theta})} + \sigma_\varepsilon^2 \cdot \underbrace{\delta^T \delta}_{\geq 0}$$

$$\underline{Var(\tilde{\theta}) \geq Var(\hat{\theta})}$$