Recall we set up the MLR model in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

Using this formulation, we can write the OLS estimates of our coeffecients as

$$\widehat{\boldsymbol{\beta}} = \arg\min_b (\mathbf{Y} - \mathbf{X}\mathbf{b})^T(\mathbf{Y} - \mathbf{X}\mathbf{b}) \tag{2}$$

If $(\mathbf{X}^T\mathbf{X})$ is invertible, then we have the *unique* solutions of $\widehat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{3}$$

We now look to investigate the bias and variance of this estimate. Let $C = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Using this, we see

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) = \mathbb{E}(C\mathbf{Y}|X) = C\mathbb{E}(\mathbf{Y}|\mathbf{X}) = C\mathbf{X}\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

Using a similar arguement, we have

$$Var(\mathbf{Y}|\mathbf{X}) = CVar(\mathbf{Y}|\mathbf{X})C^T = CI_{\sigma^2}C^T = \sigma^2 CC^T$$

Where

$$CC^T = \Big[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Big]\Big[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Big]^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}$$

So, under the assumption that $e \sim \mathcal{N}(0, I_{\sigma^2})$ we have

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}) \tag{4}$$

Theorem 1. *(Gauss - Markov) Let $\ell = (\ell_1, \dots, \ell_{p+1} \in \mathbb{R}^{p+1}$ and consider all the linear combinations of the form $\ell^T\boldsymbol{\beta}$. Then for any $\ell \in \mathbb{R}^{p+1}$*

- *The OLS estimate of $\ell^T\boldsymbol{\beta}$ is $\ell\widehat{\boldsymbol{\beta}}$*
- *Amoung all unbiased, linear, estimates of $\ell^T\boldsymbol{\beta}$, $\ell^T\widehat{\boldsymbol{\beta}}$ has the smallest variance. In this case, we call $\widehat{\boldsymbol{\beta}}$ the **B**est **L**inear **U**nbiased **E**stimator (**BLUE**)*

Proof.

- Assume there is an unbiased linear estimator $\widetilde{\boldsymbol{\beta}}$ with lower variance.
- Write $\widehat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{Y}$ and $\widetilde{\boldsymbol{\beta}} = (\mathbf{A} + \mathbf{C})\mathbf{Y}$
- Show $Var(\widehat{\boldsymbol{\beta}}) - Var(\widetilde{\boldsymbol{\beta}})$ is positive definite
- Show $\mathbb{E}(\widetilde{\boldsymbol{\beta}}|X) = \boldsymbol{\beta}$ implies $\mathbf{A}\mathbf{X} = 0$ and with this $Var(\widetilde{\boldsymbol{\beta}}|\mathbf{X})$ implies $\mathbf{A}\mathbf{A}^T = 0$
- Show $Var(\widetilde{\boldsymbol{\beta}}) = Var(\widehat{\boldsymbol{\beta}}) + \sigma^2\mathbf{A}\mathbf{A}^T$ implies $-\sigma^2\mathbf{A}\mathbf{A}^T P.D.$
- Show contradiction of P.D. and conclude.

$\square$