$X \in \mathbb{R}^{n \times (p+1)}$ ①

$Y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2 I_n)$

$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = P_X Y \quad \Rightarrow \hat{Y}_i = (P_X Y)_i = \sum_{j=1}^{n} P_{X_{ij}} Y_j$

$$\mathrm{cov}(Y_i, \hat{Y}_i) = \mathrm{cov}(Y_i, \sum_{j=1}^{n} P_{X_{ij}} Y_j) = \sum_{j=1}^{n} A_{ij} \mathrm{cov}(Y_i, Y_j)$$

$$\mathrm{cov}(Y_i, Y_j) = \begin{cases} \sigma^2 & : i = j \\ 0 & : i \neq j \end{cases} \Rightarrow \mathrm{cov}(Y_i, \hat{Y}_i) = \sum_{j=1}^{n} P_{X_{ij}} \mathrm{cov}(Y_i, Y_j) = P_{X_{ii}} \sigma^2$$

$$\mathrm{cov}(e) = \mathrm{cov}((I - P_X)Y) = (I - P_X)\mathrm{cov}(Y)(I - P_X)^T = (I - P_X)\sigma^2 I_n (I - P_X)^T = \sigma^2 (I - P_X)(I - P_X)^T = \sigma^2 (I - P_X)$$

②

$\hat{\beta}_\Sigma = \underset{\beta}{\mathrm{argmin}} \|Y - X\beta\|^2_\Sigma = \mathrm{argmin} [(Y - X\beta)^T \Sigma^{-1} (Y - X\beta)]$ ①

define: $L(\beta) = [(Y - X\beta)^T \Sigma^{-1} (Y - X\beta)]$, for $\varepsilon \sim N(0, \Sigma) \Rightarrow Y \sim N(X\beta, \Sigma)$ we get:

$$\Rightarrow \nabla_\beta L(\beta) = -2X^T \Sigma^{-1} (Y - X\beta) = 0 \Rightarrow X^T \Sigma^{-1} Y = X^T \Sigma^{-1} X\beta \Rightarrow \hat{\beta}_\Sigma = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

if $\Sigma = I_n$ then $\hat{\beta}_\Sigma = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y = (X^T I^{-1} X)^{-1} X^T I^{-1} Y = (X^T X)^{-1} X^T Y$ meaning optimal OLS

$\hat{\beta}_\Sigma = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} (X\beta + \varepsilon) = (X^T \Sigma^{-1} X)^{-1} [X^T \Sigma^{-1} X\beta + X^T \Sigma^{-1} \varepsilon]$ ②

$= \beta + (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \varepsilon$

$$\Rightarrow E(\hat{\beta}_\Sigma) = E(\beta) + \underbrace{E[(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \varepsilon]}_{= E(\varepsilon) = 0} = \beta \rightarrow E(\hat{\beta}_\Sigma) = \beta$$

$\hat{\beta}_\Sigma = \beta + (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \varepsilon$, for $A = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}$, the random part is $A\varepsilon \Rightarrow \mathrm{cov}(\hat{\beta}_\Sigma) = \mathrm{cov}(A\varepsilon) = A\,\mathrm{cov}(\varepsilon) A^T$

$= A \Sigma A^T = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} (\Sigma) [(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}]^T = (X^T \Sigma^{-1} X)^{-1}$

②

$$\mathrm{cov}(\hat{\beta}_\Sigma) - \mathrm{cov}(\hat{\beta}) = (X^T \Sigma^{-1} X)^{-1} - \sigma^2 (X^T X)^{-1}$$

if $\Sigma \neq \sigma^2 I$ the there is no optimal OLS, and the GLS est. $\hat{\beta}_\Sigma$ is better and therefore $\mathrm{cov}(\hat{\beta}_\Sigma) - \mathrm{cov}(\hat{\beta}) \preceq 0$)

2

$$E(\hat{\theta}) = E(c^T Y) = c^T E[Y] = c^T X \beta = a^T \beta \Rightarrow c^T X = a^T \quad \text{or} \quad X^T c = a$$

$$Var(\hat{\theta}) = c^T cov(Y) c = \sigma^2 c^T c$$

3

c

$$L(\beta) = \|Y - X\beta\|^2 + \lambda \|\beta\|^2 = (Y - X\beta)^T (Y - X\beta) + \lambda \beta^T \beta = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta + \lambda \beta^T \beta$$

$$\Rightarrow \nabla L(\beta) = -2X^T Y + 2X^T X \hat{\beta} + 2\lambda \hat{\beta} = 0 \Rightarrow X^T X \hat{\beta} + \lambda \hat{\beta} = X^T Y \Rightarrow (X^T X + \lambda I)\hat{\beta} = X^T Y \Rightarrow \hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y$$

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y = (X^T X + \lambda I)^{-1} X^T (X\beta + \varepsilon) = (X^T X + \lambda I)^{-1} \left[X^T X \beta + X^T \varepsilon \right]$$ a

$$E(\varepsilon) = 0 \Rightarrow E(X^T \varepsilon) = 0 \Rightarrow E(\hat{\beta}) = (X^T X + \lambda I)^{-1} X^T X \beta$$

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T (X\beta + \varepsilon) \Rightarrow cov(\hat{\beta}) = (X^T X + \lambda I)^{-1} X^T cov(\varepsilon) X (X^T X + \lambda I)^{-1}$$ b

$$= \sigma^2 (X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1}$$

$$X^T X + \lambda I = U(D + \lambda I) U^T \Rightarrow cov(\hat{\beta}) = \sigma^2 U (D + \lambda I)^{-1} D (D + \lambda I)^{-1} U^T = \sigma^2 U \left[Diag\left(\frac{d_i}{(d_i + \lambda)^2} \right) \right] U^T$$