Assignment 2

TOC

题目 1

1. Consider a multiple linear regression model

$$Y = X\beta + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0, \qquad \operatorname{cov}(\varepsilon) = \sigma^2 I,$$

where Y is $n \times 1$ and X is $n \times (k+1)$ of full column rank. Partition $X\beta$ as

$$X\beta = (X_1\ X_2)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}.$$

(1) Find the LSE of β_2.
(2) Find its variance–covariance matrix $\operatorname{cov}(\hat{\beta}_2)$.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \operatorname{cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$$

(1) 求 β_2 的最小二乘估计（LSE）$\hat{\beta}_2$

最小二乘估计（LSE）是最小化残差平方和:

$$\min_{\beta_1,\beta_2} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \min_{\beta_1,\beta_2} \|\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2\|^2$$

对 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ 求导并令导数为 $\mathbf{0}$，得到**正规方程**:

$$\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$$

将 $\boldsymbol{X} = (\boldsymbol{X}_1 \ \boldsymbol{X}_2)$、$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ 代入，分块展开正规方程:

$$\begin{pmatrix} \boldsymbol{X}_1^T \\ \boldsymbol{X}_2^T \end{pmatrix} (\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{\beta}_2) = \mathbf{0}$$

进一步拆分为两个方程:

$$\begin{cases} \boldsymbol{X}_1^T\boldsymbol{Y} = \boldsymbol{X}_1^T\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2 & (1) \\ \boldsymbol{X}_2^T\boldsymbol{Y} = \boldsymbol{X}_2^T\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{\beta}_2 & (2) \end{cases}$$

从方程 (1) 解出 $\boldsymbol{\beta}_1$:

$$\boldsymbol{\beta}_1 = (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{Y} - (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2$$

将 $\boldsymbol{\beta}_1$ 代入方程 (2):

$$\begin{aligned} \boldsymbol{X}_2^T\boldsymbol{Y} &= \boldsymbol{X}_2^T\boldsymbol{X}_1\left[(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{Y} - (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2\right] + \boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{\beta}_2 \\ &= \boldsymbol{X}_2^T\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{Y} - \boldsymbol{X}_2^T\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{\beta}_2 \end{aligned}$$

将含 $\boldsymbol{\beta}_2$ 的项移到左侧，其余项移到右侧:

$$\left[\boldsymbol{X}_2^T\boldsymbol{X}_2 - \boldsymbol{X}_2^T\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\right]\boldsymbol{\beta}_2 = \boldsymbol{X}_2^T\boldsymbol{Y} - \boldsymbol{X}_2^T\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{Y}$$

定义**投影矩阵** $\boldsymbol{P}_1 = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$（$\boldsymbol{P}_1$ 是向 $\boldsymbol{X}_1$ 列空间投影的矩阵），则 $\boldsymbol{I}_n - \boldsymbol{P}_1$ 是向 $\boldsymbol{X}_1$ 列空间正交补投影的矩阵。

因此方程可简化为：

$$\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2 \cdot \boldsymbol{\beta}_2 = \boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{Y}$$

由于 $\boldsymbol{X}$ 满秩，$\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2$ 可逆，最终解得：

$$\hat{\boldsymbol{\beta}}_2 = \left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{Y}$$

(2) 求 $\hat{\beta}_2$ 的方差-协方差矩阵 $\mathrm{cov}(\hat{\beta}_2)$

已知 $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$，因此 $\hat{\boldsymbol{\beta}}_2$ 是 $\boldsymbol{Y}$ 的线性变换：

$$\hat{\boldsymbol{\beta}}_2 = \boldsymbol{A}\boldsymbol{Y}, \quad \text{其中 } \boldsymbol{A} = \left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)$$

对于线性变换 $\hat{\boldsymbol{\beta}}_2 = \boldsymbol{A}\boldsymbol{Y}$，方差-协方差矩阵满足：

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}_2) = \boldsymbol{A} \cdot \mathrm{cov}(\boldsymbol{Y}) \cdot \boldsymbol{A}^T$$

$$\begin{aligned}\mathrm{cov}(\hat{\boldsymbol{\beta}}_2) &= \boldsymbol{A} \cdot \sigma^2\boldsymbol{I}_n \cdot \boldsymbol{A}^T \\ &= \sigma^2\boldsymbol{A}\boldsymbol{A}^T\end{aligned}$$

注意到 $\boldsymbol{I}_n - \boldsymbol{P}_1$ 是**幂等矩阵**（即 $(\boldsymbol{I}_n - \boldsymbol{P}_1)^2 = \boldsymbol{I}_n - \boldsymbol{P}_1$），因此：

$$\boldsymbol{A}\boldsymbol{A}^T = \left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1) \cdot (\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}$$

由幂等性，$(\boldsymbol{I}_n - \boldsymbol{P}_1)^2 = \boldsymbol{I}_n - \boldsymbol{P}_1$，因此：

$$\boldsymbol{A}\boldsymbol{A}^T = \left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\left[\boldsymbol{X}_2^T(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\right]^{-1}$$

最终化简为：

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}_2) = \sigma^2 \left[\boldsymbol{X}_2^T (\boldsymbol{I}_n - \boldsymbol{P}_1) \boldsymbol{X}_2 \right]^{-1}$$

题目2

Suppose that $\mu \equiv \mathbb{E}(Y) = X\beta$ and $\mathrm{Cov}(Y) = \sigma^2 I$ in the linear regression model, and let $\tilde{\phi} = c^T Y$ be any unbiased linear estimator of $\phi = t^T \beta$, where t is an arbitrary vector. Prove that

$$\mathrm{Var}(\tilde{\phi}) \geq \mathrm{Var}(\hat{\phi})$$

where $\hat{\phi} = t^T \hat{\beta}$, and $\hat{\beta} = (X^T X)^{-1} X^T Y$ is the least squares estimator of β.

解答

由 $\hat{\phi} = t^T \hat{\beta}$，根据方差的性质:

$$\mathrm{Var}(\hat{\phi}) = t^T \mathrm{Var}(\hat{\beta}) t$$

已知 $\hat{\beta} = (X^T X)^{-1} X^T Y$，且 $\mathrm{Cov}(Y) = \sigma^2 I$，因此:

$$\mathrm{Var}(\hat{\beta}) = (X^T X)^{-1} X^T \cdot \sigma^2 I \cdot X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

代入得:

$$\mathrm{Var}(\hat{\phi}) = \sigma^2 t^T (X^T X)^{-1} t$$

分析 $\tilde{\phi} = c^T Y$ 的无偏性:

因为 $\tilde{\phi}$ 是 $\phi = t^T \beta$ 的无偏估计，故:

$$\mathbb{E}(\tilde{\phi}) = \mathbb{E}(c^T Y) = c^T \mathbb{E}(Y) = c^T X \beta = t^T \beta$$

由于上式对任意 β 成立，因此:

$$c^T X = t^T \quad (\text{即 } X^T c = t)$$

计算 $\mathrm{Var}(\tilde{\phi})$:

$$extVar(\tilde{\phi}) = \mathrm{Var}(c^T Y) = c^T \mathrm{Cov}(Y) c = \sigma^2 c^T c$$

对 c 进行分解:

令

$$c = X(X^T X)^{-1} t + d,$$

其中 d 是满足 $X^T d = 0$ 的向量（由 $X^T c = t$ 可验证）:

$$X^T d = X^T c - X^T X (X^T X)^{-1} t = t - t = 0.$$

展开 $\sigma^2 c^T c$:

将 $c = X(X^T X)^{-1} t + d$ 代入:

$$\begin{aligned}
\sigma^2 c^T c &= \sigma^2 \left[\left(X(X^T X)^{-1} t + d \right)^T \left(X(X^T X)^{-1} t + d \right) \right] \\
&= \sigma^2 \left[t^T (X^T X)^{-1} X^T X (X^T X)^{-1} t + t^T (X^T X)^{-1} X^T d + d^T X (X^T X)^{-1} t + d^T d \right]
\end{aligned}$$

由 $X^T d = 0$，可知:

$$t^T (X^T X)^{-1} X^T d = 0, \quad d^T X (X^T X)^{-1} t = 0,$$

因此:

$$\sigma^2 c^T c = \sigma^2 t^T (X^T X)^{-1} t + \sigma^2 d^T d$$

比较 $\mathrm{Var}(\tilde{\phi})$ 与 $\mathrm{Var}(\hat{\phi})$:

结合

$$\mathrm{Var}(\hat{\phi}) = \sigma^2 t^T (X^T X)^{-1} t,$$

得到:

$$\mathrm{Var}(\tilde{\phi}) = \mathrm{Var}(\hat{\phi}) + \sigma^2 d^T d$$

由于 $d^T d \geq 0$（当且仅当 $d = 0$ 时取等号），故:

$$\mathrm{Var}(\tilde{\phi}) \geq \mathrm{Var}(\hat{\phi})$$

当且仅当 $d = 0$（即 $c = X(X^TX)^{-1}t$）时，等号成立。

题目3

Consider n independent observations coming from the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

where ε is assumed to be distributed as $N(0, \sigma^2)$.

a) Find the test statistic for the (simultaneous) test $H_0 : \beta_1 = \beta_3,\ \beta_2 = 0$;
b) Derive its limiting distribution under H_0.

问题a：构造检验统计量

首先将原假设 $H_0 : \beta_1 = \beta_3,\ \beta_2 = 0$ 表示为线性约束形式 $C\beta = 0$，其中：

- 参数向量 $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$；
- 约束矩阵

$$C = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

接下来，通过带约束的最小二乘估计推导检验统计量：

无约束下，

$$\hat{\beta}_{\text{full}} = (X^TX)^{-1}X^TY,$$

对应的残差平方和为

$$\mathrm{SSE}_R = \| Y - X\hat{\beta}_{\text{full}} \|_2^2,$$

（R 表示无约束/备择模型）。

约束 $C\beta = 0$ 下构造拉格朗日函数：

$$\mathcal{L}(\beta,\lambda) = \|Y - X\beta\|_2^2 + \lambda^T(C\beta).$$

对 β, λ 求导并设为 0:

$$\begin{cases} \nabla_\beta \mathcal{L} = -2X^T(Y - X\beta) + C^T\lambda = 0 \\ \nabla_\lambda \mathcal{L} = C\beta = 0 \end{cases}$$

由第一式得

$$X^TX\,\beta = X^TY - \tfrac{1}{2}C^T\lambda,$$

左乘 $(X^TX)^{-1}$:

$$\beta = (X^TX)^{-1}X^TY - \tfrac{1}{2}(X^TX)^{-1}C^T\lambda. \tag{1}$$

将 (1) 代入 $C\beta = 0$ 得:

$$C(X^TX)^{-1}X^TY - \tfrac{1}{2}C(X^TX)^{-1}C^T\lambda = 0,$$

解得拉格朗日乘数:

$$\lambda = 2\left[C(X^TX)^{-1}C^T\right]^{-1}C(X^TX)^{-1}X^TY.$$

代回 (1) 得到带约束估计 $\tilde{\beta}$，对应残差平方和:

$$\mathrm{SSE}_F = \|Y - X\tilde{\beta}\|_2^2.$$

无约束估计满足

$$\hat{\beta}_{\text{full}} \sim N(\beta,\ \sigma^2(X^TX)^{-1}),$$

因此

$$C\hat{\beta}_{\text{full}} \sim N\left(C\beta,\ \sigma^2C(X^TX)^{-1}C^T\right).$$

在 $H_0: C\beta = 0$ 下:

$$
C\hat{\beta}_{\text{full}} \sim N\left(0,\ \sigma^2 C(X^TX)^{-1}C^T\right),
$$

从而二次型：

$$
(C\hat{\beta}_{\text{full}})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\beta}_{\text{full}}) \sim \chi^2(2).
$$

另一边：

$$
\text{SSE}_R \sim \sigma^2\chi^2(n-4),
$$

且 SSE_R 与 $C\hat{\beta}_{\text{full}}$ 独立。

因此检验统计量为：

$$
F = \frac{\frac{1}{2}(C\hat{\beta}_{\text{full}})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\beta}_{\text{full}})}{\frac{1}{n-4}\text{SSE}_R} = \frac{(C\hat{\beta}_{\text{full}})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\beta}_{\text{full}})}{2s^2},
$$

其中

$$
s^2 = \frac{\text{SSE}_R}{n-4}
$$

是 σ^2 的无偏估计。

问题 b：原假设下的极限分布

在 H_0 下：

- **分子：**

$$
(C\hat{\beta}_{\text{full}})^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta}_{\text{full}}) \sim \chi^2(2)
$$

- **分母：**

$$
s^2 = \frac{\text{SSE}_R}{n-4}, \qquad \frac{\text{SSE}_R}{\sigma^2} \sim \chi^2(n-4)
$$

因此

$$F \sim F(2,\, n-4).$$

当 $n \to \infty$ 时:

$$F(2, n-4) \implies \frac{1}{2}\chi^2(2),$$

因为此时

$$\frac{\mathrm{SSE}_R}{n-4} \to \sigma^2.$$

题目4

Multiple linear regression model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$$

Hypothesis testing:

$$H_0 : C\boldsymbol{\beta} = \boldsymbol{h} \qquad \text{vs.} \qquad H_1 : C\boldsymbol{\beta} \neq \boldsymbol{h}$$

where C is an $m \times (k+1)$ full-rank matrix and $\boldsymbol{h}$ is an m-vector.

---

(1) 约束 H_0 下最小二乘估计 β

构造拉格朗日函数（$\boldsymbol{\lambda}$ 为 m 维乘子）：

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \boldsymbol{\lambda}^T(C\boldsymbol{\beta} - \boldsymbol{h})$$

对 $\boldsymbol{\beta}$ 求导并设为 0:

$$\nabla_{\beta}\mathcal{L} = -2X^T(\boldsymbol{y} - X\boldsymbol{\beta}) + C^T\boldsymbol{\lambda} = 0$$

整理得

$$X^T X \,\boldsymbol{\beta} = X^T \boldsymbol{y} - \frac{1}{2} C^T \boldsymbol{\lambda} \tag{1}$$

再乘以 $(X^T X)^{-1}$ 得

$$\boldsymbol{\beta} = (X^T X)^{-1} X^T \boldsymbol{y} - \frac{1}{2} (X^T X)^{-1} C^T \boldsymbol{\lambda}. \tag{2}$$

对 $\boldsymbol{\lambda}$ 求导:

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L} = C\boldsymbol{\beta} - \boldsymbol{h} = 0 \tag{3}$$

将 (2) 代入 (3):

$$C(X^T X)^{-1} X^T \boldsymbol{y} - \frac{1}{2} C(X^T X)^{-1} C^T \boldsymbol{\lambda} = \boldsymbol{h}.$$

令

$$\hat{\boldsymbol{\beta}}_{\text{full}} = (X^T X)^{-1} X^T \boldsymbol{y},$$

则

$$\boldsymbol{\lambda}^T = 2 \left[C(X^T X)^{-1} C^T \right]^{-1} \left(C\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{h} \right).$$

代回 (2) 得到 **约束最小二乘估计**:

$$\hat{\boldsymbol{\beta}}_{\text{restricted}} = \hat{\boldsymbol{\beta}}_{\text{full}} - (X^T X)^{-1} C^T \left[C(X^T X)^{-1} C^T \right]^{-1} (C\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{h}).$$

(2) F-检验统计量与原分布

无约束残差平方和:

$$\text{SSE}_{\text{full}} = \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_{\text{full}}\|_2^2.$$

约束模型残差平方和:

$$\mathrm{SSE}_{\mathrm{restricted}} = \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_{\mathrm{restricted}}\|_2^2.$$

利用残差正交性可得:

$$\mathrm{SSE}_{\mathrm{restricted}} = \mathrm{SSE}_{\mathrm{full}} + (C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h}),$$

因此:

$$\mathrm{SSE}_{\mathrm{restricted}} - \mathrm{SSE}_{\mathrm{full}} = (C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h}).$$

构造 F-统计量

在 H_0 下:

$$C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h} \sim N\left(0, \sigma^2 C(X^TX)^{-1}C^T\right),$$

故

$$\frac{1}{\sigma^2}(\mathrm{SSE}_{\mathrm{restricted}} - \mathrm{SSE}_{\mathrm{full}}) \sim \chi^2(m).$$

同时

$$\frac{1}{\sigma^2}\mathrm{SSE}_{\mathrm{full}} \sim \chi^2(n-k-1),$$

且两者独立(因为$s^2 \perp\!\!\!\perp \hat{\beta}$)

于是 F-**统计量**:

$$F = \frac{\left[(C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h})^T \left[C(X^TX)^{-1}C^T\right]^{-1} (C\hat{\boldsymbol{\beta}}_{\mathrm{full}} - \boldsymbol{h})\right]/m}{s^2},$$

其中

$$s^2 = \frac{\mathrm{SSE}_{\mathrm{full}}}{n-k-1}.$$

由 F 分布的定义:

$$F \sim F\big(m,\, n-k-1\big).$$

当

$$F > F_\alpha(m, n-k-1)$$

时拒绝 H_0。

题目5

Consider the linear regression model, $Y = X\beta + \varepsilon$, where $\mathbb{E}\varepsilon = 0$ and $\mathrm{cov}(\varepsilon) = \sigma^2\Sigma$, in which Σ is a known positive definite matrix. Find the BLUE (Best Linear Unbiased Estimator) for β and derive its variance-covariance matrix.

解答

因为 Σ 是**已知正定矩阵**，存在一个对称正定矩阵 $\Sigma^{1/2}$，满足

$$\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T, \qquad \Sigma^{-1/2} = (\Sigma^{1/2})^{-1}.$$

定义变换矩阵

$$P = \Sigma^{-1/2}.$$

对模型进行线性变换：

$$\tilde{Y} = PY, \qquad \tilde{X} = PX, \qquad \tilde{\varepsilon} = P\varepsilon.$$

则：

- $\mathbb{E}(\tilde{\varepsilon}) = 0$,
- $\mathrm{cov}(\tilde{\varepsilon}) = \sigma^2 I_n$。

因此变换后的模型

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

就是经典线性模型。

变换后模型的 OLS 为:

$$\hat{\beta}_{\text{OLS}} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y}.$$

代入 $\tilde{X} = PX = \Sigma^{-1/2}X$、$\tilde{Y} = PY = \Sigma^{-1/2}Y$:

$$\hat{\beta}_{\text{OLS}} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y.$$

这就是原模型下的加权最小二乘估计（WLS），记作:

$$\hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y.$$

$$\mathbb{E}(\hat{\beta}) = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\mathbb{E}(Y) = \beta.$$

因此 $\hat{\beta}$ 是无偏估计。

设任意线性无偏估计 $\tilde{\beta} = CY$。

$$\mathbb{E}(\tilde{\beta}) = CX\beta = \beta \quad \forall\, \beta \quad \Longrightarrow \quad CX = I.$$

令

$$C = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1} + D.$$

代入无偏性 $CX = I$ 得:

$$DX = 0.$$

因为 $\operatorname{cov}(Y) = \sigma^2\Sigma$:

$$\operatorname{cov}(\tilde{\beta}) = \sigma^2 C\Sigma C^T.$$

代入 $C = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1} + D$ 展开，并利用 $DX = 0$ 得:

$$\operatorname{cov}(\tilde{\beta}) = \sigma^2(X^T\Sigma^{-1}X)^{-1} + \sigma^2 D\Sigma D^T.$$

由于 Σ 正定:

$$D\Sigma D^T \text{ 非负定}.$$

因此：

$$\operatorname{cov}(\tilde{\beta}) \geq \operatorname{cov}(\hat{\beta}) = \sigma^2 (X^T \Sigma^{-1} X)^{-1}.$$

只有当 $D = 0$（即 $\tilde{\beta} = \hat{\beta}$）时取等号，因此
$\hat{\beta}$ 是 BLUE。

由 $\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$：

$$\begin{aligned}\operatorname{cov}(\hat{\beta}) &= (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \cdot \sigma^2 \Sigma \cdot \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} \\ &= \sigma^2 (X^T \Sigma^{-1} X)^{-1}.\end{aligned}$$

β 的 BLUE（加权最小二乘估计）为

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y.$$

其方差-协方差矩阵为

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