高级计量经济学

理论经济学博士课程 2023-2024

Lecture 6: Statistical Properties of OLS

Davidson, R. & MacKinnon, J. (2009). Econometrics Theory and Methods. Oxford University Press.

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室　粤海校区汇文楼1510
E-mail　huangjp@szu.edu.cn
Website　https://huangjp.com

多变量分布

Multivariate Distribution

$\boldsymbol{x} = (X_1, X_2, \ldots, X_n)^\top$ 为随机向量，其联合密度函数为 $f(x_1, x_2, \ldots, x_n)$

- 期望值

$$E[\boldsymbol{x}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{bmatrix} = \boldsymbol{\mu}$$

- 方差-协方差矩阵

$$\mathrm{Var}[\boldsymbol{x}] = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma^2_{X_n} \end{bmatrix} = E[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

OLS 假设

线性模型与 OLS 假设

完整的线性回归模型可以写成

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \mid \boldsymbol{X} \sim \mathrm{IID}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

independently and identically distributed
在给出 $\boldsymbol{X}$ 的条件下，u_t 服从独立同分布，
其均值为 0，方差为 σ^2，任意 u_s 与 u_t 的协方差为 0。

这个简洁的表述中包含以下假设：

1. 线性 linearity：$\boldsymbol{y}$ 是 $\boldsymbol{\beta}$ 的线性函数
2. 解释变量 $\boldsymbol{X}$ 可以是随机的，此时 $(y_t, x_{t1}, \ldots, x_{tk})$ 为同一个联合分布的独立样本
3. 外生性 exogeneity：$E[u_t \mid \boldsymbol{X}] = 0$ $\Rightarrow E[\boldsymbol{y} \mid \boldsymbol{X}] = \boldsymbol{X\beta}$
4. 同方差性 homoskedasticity：$\mathrm{Var}[u_t \mid \boldsymbol{X}] = \sigma^2$
   无自相关 non-autocorrelation：$\mathrm{Cov}[u_s, u_t \mid \boldsymbol{X}] = 0$

⇒ 在外生性成立时，
$\mathrm{Var}[u_t \mid \boldsymbol{X}] = \sigma^2 \Leftrightarrow E[u_t^2 \mid \boldsymbol{X}] = \sigma^2$
$\mathrm{Cov}[u_s, u_t \mid \boldsymbol{X}] = 0 \Leftrightarrow E[u_s u_t \mid \boldsymbol{X}] = 0$

为了保证 OLS 估计量 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ 存在，我们还需要假设：

右侧两个条件可以简洁地写成 $E[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$

5. $\boldsymbol{X}$ 列满秩：$\mathrm{rank}(\boldsymbol{X}) = k$

外生性条件

The Exogeneity Condition

$E[u_t \mid \boldsymbol{X}] = 0$ 被称为严格外生性（strict exogeneity），在横截面数据中容易被满足，但在时间序列数据中往往过强。一个较弱的条件是前定性（predeterminedness）$E[u_t \mid \boldsymbol{X}_t] = 0$。

由严格外生性可以导出：

1. $E[u_t] = E_{\boldsymbol{X}}\big[E[u_t \mid \boldsymbol{X}]\big] = 0$
2. $E[\boldsymbol{u}\boldsymbol{u}^\top] = E_{\boldsymbol{X}}\big[E[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}]\big] = E[\sigma^2 \boldsymbol{I}] = \sigma^2 \boldsymbol{I}$

此处假设同方差性和无自相关

因此，$\boldsymbol{X}$ 为随机变量时的结论也适用于 $\boldsymbol{X}$ 为固定变量时。

（条件）均值独立

(Conditional) Mean Independence

均值独立性

当 $E[Y \mid X] = E[Y]$ 时，称 Y 均值独立于 X。

由定义可知，均值独立性不是对称的概念， Y 均值独立于 X 不等于 X 均值独立于 Y。

定理：

$$X \text{ 与 } Y \text{ 独立} \quad \underset{\not\Leftarrow}{\Rightarrow} \quad X \text{ 与 } Y \text{ 均值独立} \quad \underset{\not\Leftarrow}{\Rightarrow} \quad X \text{ 与 } Y \text{ 线性不相关}$$

$$X \perp\!\!\!\perp Y \qquad E[X \mid Y] = E[X], \quad E[Y \mid X] = E[Y] \qquad \mathrm{Cov}[X, Y] = 0 \Leftrightarrow E[XY] = E[X]E[Y]$$

详细信息可参考 https://www.econometrics.blog/post/why-econometrics-is-confusing-part-ii-the-independence-zoo/

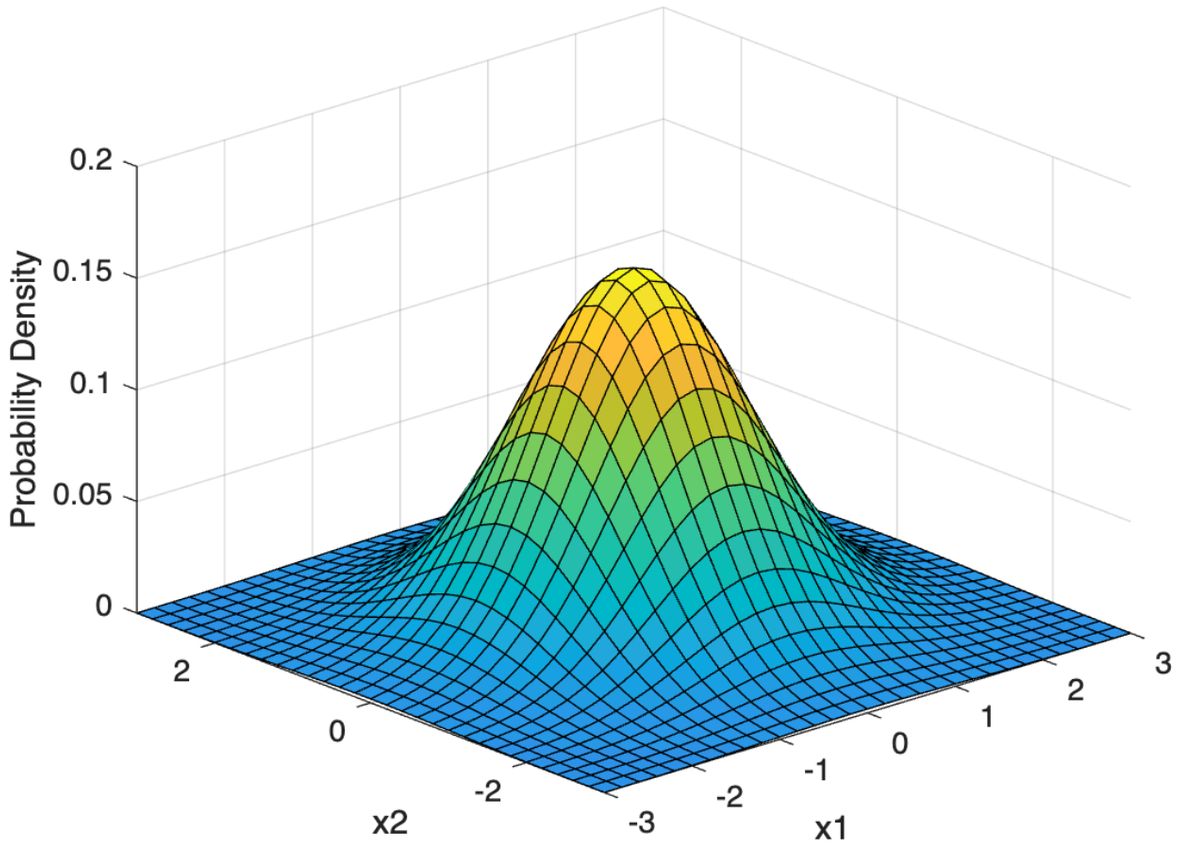
若严格外生性 $E[u_t \mid \boldsymbol{X}] = 0$ 成立，则误差项 $\boldsymbol{u}$ 均值独立于解释变量 $\boldsymbol{X}$。

若 $\boldsymbol{X} \perp\!\!\!\perp \mu_t$，则 $E[u_t \mid \boldsymbol{X}] = E[u_t]$。此时可以直接假设 $E[u_t] = 0$。

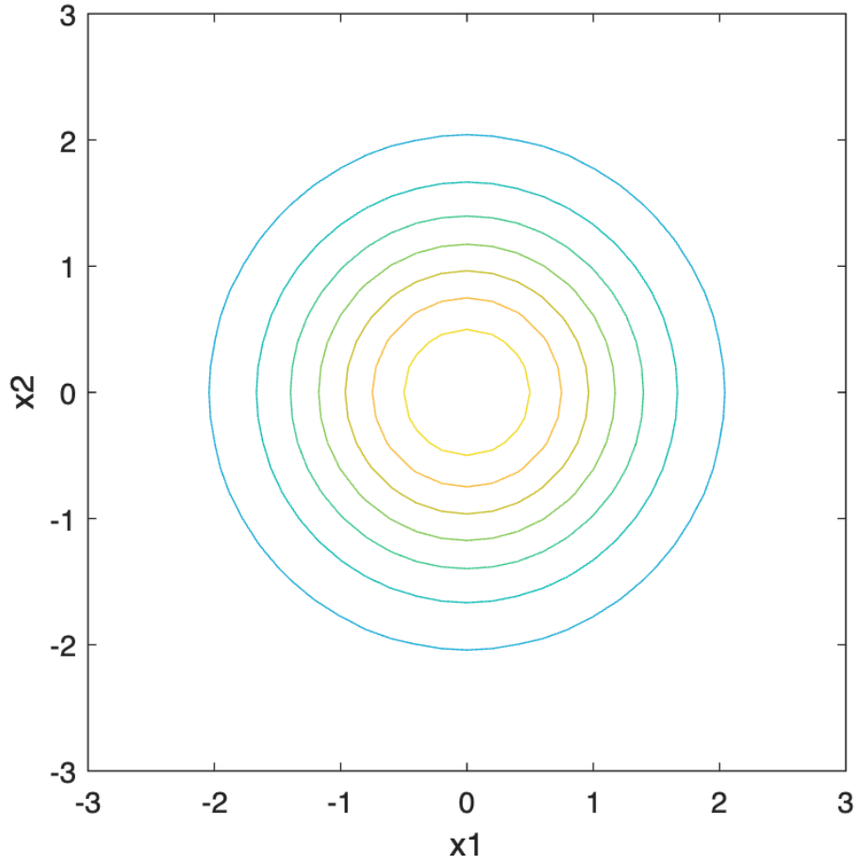
同方差无自相关性条件

The Homoskedasticity & Non-autocorrelation Condition

$\mathrm{Var}[\boldsymbol{u}\,|\boldsymbol{X}] = \sigma^2\boldsymbol{I}$ 常被称为 spherical error，或称误差项服从 spherical distribution（球面分布）。



pdf of bivariate $N(\mathbf{0}, \boldsymbol{I})$



contour plot

2D → 3D: circle → sphere

$\hat{\beta}$的非偏性和一致性

非偏性

Unbiasedness

非偏性

若参数 θ 的真实值是 θ_0，$\hat{\theta}$ 是 θ 的一个估计量。则称 $E[\hat{\theta}] - \theta_0$ 为估计量 $\hat{\theta}$ 的偏差（bias）或偏误。当 $E[\hat{\theta}] - \theta_0 = 0$，即

$$E[\hat{\theta}] = \theta_0$$

时，$\hat{\theta}$ 是非偏的。

估计量的一致性

若参数 θ 的真实值是 θ_0，$\hat{\theta}^n$ 是基于含有 n 个观测值样本的估计量。则当

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}^n = \theta_0 \quad \text{即当 } n \to \infty \text{ 时 } \hat{\theta} \xrightarrow{p} \theta$$

时，$\hat{\theta}^n$ 是一致的。

非偏性与一致性

$(x_1, x_2, \ldots, x_n)$ 为独立随机样本，且每个 x_i 的总体均值为 $E[x_i] = \mu$。

已知在适当的条件下（例如总体方差有限，或服从同分布），样本均值 $\bar{x}$ 是 μ 的非偏统计量和一致估计量（大数法则）。

考虑下面三个 μ 的估计量：

$$\hat{\mu}_1 = \frac{1}{n+1}\sum_{i=1}^{n} x_i \qquad = \frac{n}{n+1}\bar{x} \xrightarrow{p} \mu$$

$$\hat{\mu}_2 = \frac{1.01}{n}\sum_{i=1}^{n} x_i \qquad = 1.01\bar{x} \xrightarrow{p} 1.01\mu$$

$$\hat{\mu}_3 = 0.01x_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} x_i \qquad = 0.01x_1 + 0.99\bar{x} \xrightarrow{p} 0.01x_1 + 0.99\mu$$

其中仅 $\hat{\mu}_1$ 满足一致性，仅 $\hat{\mu}_3$ 满足非偏性。

$\hat{\boldsymbol{\beta}}$ 的非偏性

假设 $\boldsymbol{\beta}$ 的真实值是 $\boldsymbol{\beta}_0$。$\boldsymbol{\beta}$ 的 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 为

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \\ &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u}) \\ &= \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}\end{aligned}$$

$$\Rightarrow \quad E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0 + E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}]$$

根据迭代期望法则，右侧第二项可以写成

$$\begin{aligned}E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}] &= E_{\boldsymbol{X}}\left[E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u} \mid \boldsymbol{X}]\right] \\ &= E_{\boldsymbol{X}}\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top E[\boldsymbol{u} \mid \boldsymbol{X}]\right]\end{aligned}$$

因此，当外生性条件 $E[\boldsymbol{u} \mid \boldsymbol{X}] = \mathbf{0}$ 成立时，

$$E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}] = E_{\boldsymbol{X}}[\mathbf{0}] = \mathbf{0}$$

$$\Rightarrow \quad E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0 \qquad (\hat{\boldsymbol{\beta}} \text{ 是 } \boldsymbol{\beta} \text{ 的非偏估计量})$$

时间序列数据往往不满足外生性，而只满足前定性。此时 $\hat{\boldsymbol{\beta}}$ 有偏。
详见 D&M (2021) pp.90-92

$\hat{\boldsymbol{\beta}}$ 的一致性

已知 $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}$

我们的目标是把右侧第二项的概率极限分解为

$$\left(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \left(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right)$$

$\boldsymbol{X}^\top \boldsymbol{X}$ 是 $k \times k$ 矩阵，其 (i, j) 要素是

$$\boldsymbol{x}_i^\top \boldsymbol{x}_j = \sum_{t=1}^{n} x_{ti} x_{tj}$$

一般情况下不存在 $\operatorname*{plim}_{n\to\infty} \boldsymbol{x}_i^\top \boldsymbol{x}_j$，但有可能存在 $\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{x}_i^\top \boldsymbol{x}_j$。

因为 $\boldsymbol{X}^\top \boldsymbol{X} = \sum_{i=t}^{n} \boldsymbol{X}_t^\top \boldsymbol{X}_t$，则根据大数定律，$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{X}_t^\top \boldsymbol{X}_t = E[\boldsymbol{X}_t^\top \boldsymbol{X}_t]$，因此

需要假设 $\boldsymbol{X}_t$ 是 IID 的随机样本

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} = E[\boldsymbol{X}_t^\top \boldsymbol{X}_t] \equiv \boldsymbol{Q} \quad \Rightarrow \quad \operatorname*{plim}_{n\to\infty} \frac{1}{n} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \boldsymbol{Q}^{-1}$$

书中将 $\boldsymbol{Q}$ 记为 $\boldsymbol{S}_{\boldsymbol{X}^\top \boldsymbol{X}}$

Theorem D.14 (Greene, 2020)
积的概率极限 = 概率极限之积，即

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{X}^n = \boldsymbol{A}, \operatorname*{plim}_{n\to\infty} \boldsymbol{Y}^n = \boldsymbol{B} \;\Rightarrow\; \operatorname*{plim}_{n\to\infty} \boldsymbol{X}^n \boldsymbol{Y}^n = \boldsymbol{A}\boldsymbol{B}$$

Theorem D.14 (Greene, 2020)
逆矩阵的概率极限

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{W}^n = \boldsymbol{\Omega} \;\Rightarrow\; \operatorname*{plim}_{n\to\infty} (\boldsymbol{W}^n)^{-1} = \boldsymbol{\Omega}^{-1}$$

$\hat{\boldsymbol{\beta}}$ 的一致性

接下来我们讨论 $\operatorname{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u}$。

因为 $\boldsymbol{X}^\top \boldsymbol{u} = \sum_{i=t}^{n} \boldsymbol{X}_t^\top u_t$，则根据大数法则，$\operatorname{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top u_t = E[\boldsymbol{X}_t^\top u_t]$，因此

$$\operatorname{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u} = E[\boldsymbol{X}_t^\top u_t]$$

如果假设前定性 $E[u_t \mid \boldsymbol{X}_t] = 0$，则根据 LIE，

$$\begin{aligned} E[\boldsymbol{X}_t^\top u_t] &= E_{\boldsymbol{X}_t}\left[E[\boldsymbol{X}_t^\top u_t \mid \boldsymbol{X}_t]\right] = E_{\boldsymbol{X}_t}\left[\boldsymbol{X}_t^\top \cdot E[u_t \mid \boldsymbol{X}_t]\right] \\ &= E_{\boldsymbol{X}_t}[\boldsymbol{X}_t^\top \cdot 0] = \mathbf{0} \end{aligned}$$

因此，$\operatorname{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u} = E[\boldsymbol{X}_t^\top u_t] = \mathbf{0}$。

当**样本为 IID**，且满足**前定性**条件时（不需要外生性条件），$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$ 右侧第二项的概率极限为 $\mathbf{0}$，即 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的一致估计量。

关于非偏性与一致性的一些讨论

- 非偏性适用于任何大小的样本，而一致性仅适用于大样本。
- 保证 $\hat{\boldsymbol{\beta}}$ 的非偏性需要假设外生性，而保证 $\hat{\boldsymbol{\beta}}$ 的一致性只需要假设前定性。虽然前定性条件弱于外生性条件，但这并不代表非偏性可以推导出一致性，它们是两个不同的概念。非偏性只和期望值有关，而一致性是分布上的性质。
- 解释变量中包含被解释变量的滞后项（lag）的时间序列模型不满足外生性，但满足前定性。因此其 OLS 估计量有偏但一致。
- 非偏性并不一定是最好的性质。有时我们需要在非偏但方差很大的估计量和略偏但方差很小的估计量中进行选择。
- 不满足一致性有两种情况：一种是概率极限不存在（不收敛），另一种是概率极限存在但不收敛于真实的参数值。

$\hat{\beta}$ 的方差与协方差

随机向量的协方差矩阵

Covariance Matrix of Random Vectors

$\boldsymbol{x} = (X_1, X_2, \ldots, X_n)^\top$ 为随机向量时，其方差-协方差矩阵（或习惯性称为协方差矩阵）是

$$
\begin{aligned}
\mathrm{Var}[\boldsymbol{x}] &= \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix} \\
&= E\left[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^\top\right] \\
&= E[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top
\end{aligned}
$$

若 $\boldsymbol{\mu} = E[\boldsymbol{x}] = \mathbf{0}$，则 $\mathrm{Var}[\boldsymbol{x}] = E[\boldsymbol{x}\boldsymbol{x}^\top]$。

协方差矩阵是半正定矩阵

Covariance Matrices are Positive Semidefinite

$k \times k$ 矩阵 $\boldsymbol{A}$ 是正定（半正定）矩阵 $\Leftrightarrow \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \underset{(\geq)}{>} 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$。

- 对任意矩阵 $\boldsymbol{B}$，$\boldsymbol{B}^\top \boldsymbol{B}$ 是半正定矩阵。如果 $\boldsymbol{B}$ 为列满秩，则 $\boldsymbol{B}^\top \boldsymbol{B}$ 是正定矩阵。
- 如果 $\boldsymbol{A}$ 是正定矩阵，则 $\boldsymbol{B}^\top \boldsymbol{A}\boldsymbol{B}$ 是半正定。如果 $\boldsymbol{B}$ 为列满秩，则 $\boldsymbol{B}^\top \boldsymbol{A}\boldsymbol{B}$ 是正定。
- 如果 $\boldsymbol{A}$ 是对称正定矩阵，则存在 $k \times k$ 矩阵 $\boldsymbol{B}$ 满足 $\boldsymbol{A} = \boldsymbol{B}^\top \boldsymbol{B}$，但 $\boldsymbol{B}$ 不是唯一的。（Cholesky 分解）

考虑随机向量 $\boldsymbol{x}$ 的任意线性结合 $\boldsymbol{w}^\top \boldsymbol{x}$:

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{w}^\top \boldsymbol{x}] &= E[\boldsymbol{w}^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{w}] - E[\boldsymbol{w}^\top \boldsymbol{x}]E[\boldsymbol{x}^\top \boldsymbol{w}] \\
&= \boldsymbol{w}^\top E[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{w} - \boldsymbol{w}^\top E[\boldsymbol{x}]E[\boldsymbol{x}^\top]\boldsymbol{w} \qquad \boldsymbol{w} \text{ 非随机且已知} \\
&= \boldsymbol{w}^\top \mathrm{Var}[\boldsymbol{x}]\boldsymbol{w} \geq 0 \qquad \boldsymbol{w}^\top \boldsymbol{x} \text{ 为 scalar}
\end{aligned}$$

因此，$\mathrm{Var}[\boldsymbol{x}]$ 是半正定。

大多数情况下 $\mathrm{Var}[\boldsymbol{u}]$ 是正定矩阵，只有当 $\boldsymbol{w}^\top \boldsymbol{x} = 0$ 时等号才会成立。

$\hat{\boldsymbol{\beta}}$ 的协方差矩阵

假设同方差性和无自相关性 $\mathrm{Var}[\boldsymbol{u} \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$。当外生性 $E[\boldsymbol{u} \mid \boldsymbol{X}] = \boldsymbol{0}$ 成立时，$E[\boldsymbol{u}\boldsymbol{u}^\top | \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$。

$\hat{\boldsymbol{\beta}}$ 的协方差矩阵是

$$
\begin{aligned}
\mathrm{Var}[\hat{\boldsymbol{\beta}}] &= E\left[(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}])^\top\right] \\
&= E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top\right] \\
&= E\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right] \qquad \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u} \\
&= E_{\boldsymbol{X}}\left[E\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \mid \boldsymbol{X}\right]\right] \\
&= E_{\boldsymbol{X}}\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top E[\boldsymbol{u}\boldsymbol{u}^\top | \boldsymbol{X}]\, \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right] \\
&= \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}
\end{aligned}
$$

以上结论也可以针对 $\mathrm{Var}[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}]$ 导出。

如果 σ^2 已知且其真实值是 σ_0^2，则 $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma_0^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$。如果 σ^2 未知，则需要对其进行估计。

预测误差的方差

令 $\gamma = \boldsymbol{\omega}^\top \boldsymbol{\beta}$，$\boldsymbol{\omega}$ 已知。γ 的估计量为 $\hat{\gamma} = \boldsymbol{\omega}^\top \hat{\boldsymbol{\beta}}$。

$\hat{\gamma}$ 的协方差是

$$\begin{aligned}
\mathrm{Var}[\hat{\gamma}] = \mathrm{Var}[\boldsymbol{\omega}^\top \hat{\boldsymbol{\beta}}] &= E[\boldsymbol{\omega}^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{\omega}] \\
&= \boldsymbol{\omega}^\top E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top] \boldsymbol{\omega} \\
&= \boldsymbol{\omega}^\top \mathrm{Var}[\hat{\boldsymbol{\beta}}] \boldsymbol{\omega} \\
&= \boldsymbol{\omega}^\top \sigma_0^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\omega}
\end{aligned}$$

上面的结论可以用来推导预测误差的方差。假设我们用样本外的观测值 $\boldsymbol{X}_s$ 预测 y_s，此时预测值为 $\hat{y}_s = \boldsymbol{X}_s \hat{\boldsymbol{\beta}}$。当外生性成立时，$\hat{\boldsymbol{\beta}}$ 为非偏，因此 $\hat{y}_s$ 也非偏。预测误差的方差是

$$\begin{aligned}
\mathrm{Var}[y_s - \hat{y}_s] = E[(y_s - \boldsymbol{X}_s \hat{\boldsymbol{\beta}})^2] &= E[(\boldsymbol{X}_s \boldsymbol{\beta}_0 + u_s - \boldsymbol{X}_s \hat{\boldsymbol{\beta}})^2] \\
&= E[u_s^2] + E[(\boldsymbol{X}_s \boldsymbol{\beta}_0 - \boldsymbol{X}_s \hat{\boldsymbol{\beta}})^2] \quad \text{因假设了外生性} \\
&= \sigma_0^2 + \mathrm{Var}[\boldsymbol{X}_s \hat{\boldsymbol{\beta}}] \\
&= \sigma_0^2 + \boldsymbol{X}_s \mathrm{Var}[\hat{\boldsymbol{\beta}}] \boldsymbol{X}_s^\top \\
&= \sigma_0^2 + \sigma_0^2 \boldsymbol{X}_s (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_s^\top
\end{aligned}$$

$\mathrm{Var}[X] = E[X^2] - E[X]^2$

$\hat{\beta}$ 的有效性

估计量的有效性

Efficiency of Estimator

对同一参数可以有不同的估计量，且估计量的估计精确度一般也不同。如果估计量 a 的精确度大于估计量 b，我们说 a 比 b 更有效（more efficient）。有效一词可以理解为更有效地利用样本所提供的信息进行估计。

估计量 a 的估计精确度和 $\mathrm{Var}[a]^{-1}$ 成正比，因此 a 比 b 更有效可以表达为 $\mathrm{Var}[a]^{-1} > \mathrm{Var}[b]^{-1}$，或 $\mathrm{Var}[a] < \mathrm{Var}[b]$。

- 标量：a 比 b 更有效 $\Leftrightarrow$ $\mathrm{Var}[a] < \mathrm{Var}[b]$。
- 向量：$\boldsymbol{a}$ 比 $\boldsymbol{b}$ 更有效 $\Leftrightarrow$ $(\mathrm{Var}[\boldsymbol{b}] - \mathrm{Var}[\boldsymbol{a}])$ 是非零半正定矩阵。

$\boldsymbol{a}$ 比 $\boldsymbol{b}$ 更有效意味着 $\boldsymbol{a}$ 中的每个要素，以及任意要素的任意线性结合都至少和 $\boldsymbol{b}$ 中的一样有效。即对任意的权重 $\boldsymbol{\omega}$，$\gamma_a = \boldsymbol{\omega}^\top \boldsymbol{a}$ 应当至少和 $\gamma_b = \boldsymbol{\omega}^\top \boldsymbol{b}$ 一样有效，即

$$\begin{aligned} \mathrm{Var}[\gamma_a] \le \mathrm{Var}[\gamma_b] \quad &\Leftrightarrow \quad \boldsymbol{\omega}^\top \mathrm{Var}[\boldsymbol{a}]\, \boldsymbol{\omega} \le \boldsymbol{\omega}^\top \mathrm{Var}[\boldsymbol{b}]\, \boldsymbol{\omega} \\ &\Leftrightarrow \quad \boldsymbol{\omega}^\top (\mathrm{Var}[\boldsymbol{b}] - \mathrm{Var}[\boldsymbol{a}])\, \boldsymbol{\omega} \ge 0 \end{aligned}$$

$\hat{\boldsymbol{\beta}}$ 的有效性

已知在**外生性**条件下，$\boldsymbol{\beta}$ 的 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 是非偏的。同时 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{y}$ 的要素的线性结合，因此我们说 $\hat{\boldsymbol{\beta}}$ 是线性估计量。

这里考虑 $\boldsymbol{\beta}$ 的另一个线性估计量 $\tilde{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y}$。（$\boldsymbol{A}$ 是 $\boldsymbol{X}$ 的函数）

如果定义 $\boldsymbol{C} = \boldsymbol{A} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$，则有

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} + \boldsymbol{C}\boldsymbol{y} = \hat{\boldsymbol{\beta}} + \boldsymbol{C}\boldsymbol{y}$$

如果我们要求 $\tilde{\boldsymbol{\beta}}$ 为非偏估计量，则需要保证 $E[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0$，即

$$\begin{aligned} E[\tilde{\boldsymbol{\beta}}] = E[\boldsymbol{A}\boldsymbol{y}] &= E[\boldsymbol{A}(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u})] \\ &= E[\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}_0] + E[\boldsymbol{A}\boldsymbol{u}] = \boldsymbol{\beta}_0 \end{aligned}$$

因为 $\boldsymbol{A}$ 是 $\boldsymbol{X}$ 的函数，根据外生性和 LIE 可得 $E[\boldsymbol{A}\boldsymbol{u}] = \boldsymbol{0}$。因此 $\tilde{\boldsymbol{\beta}}$ 非偏需要保证 $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{I}$，或等价的 $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{O}$。

$\hat{\boldsymbol{\beta}}$ 的有效性

当满足**外生性**和 $\boldsymbol{CX} = \boldsymbol{O}$ 时，$\hat{\boldsymbol{\beta}}$ 和 $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \boldsymbol{Cy}$ 都是 $\boldsymbol{\beta}$ 的线性非偏估计量。

$$\begin{aligned}\mathrm{Var}[\tilde{\boldsymbol{\beta}}] &= \mathrm{Var}[\hat{\boldsymbol{\beta}} + \boldsymbol{Cy}] \\ &= \mathrm{Var}[\hat{\boldsymbol{\beta}}] + \mathrm{Var}[\boldsymbol{Cy}] + 2\mathrm{Cov}[\hat{\boldsymbol{\beta}}, \boldsymbol{Cy}]\end{aligned}$$

$\boldsymbol{CX} = \boldsymbol{O} \Rightarrow \boldsymbol{Cy} = \boldsymbol{Cu} \Rightarrow E[\boldsymbol{Cy}] = \boldsymbol{0}$。因此，

$$\begin{aligned}\mathrm{Cov}[\hat{\boldsymbol{\beta}}, \boldsymbol{Cy}] &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\boldsymbol{Cy})^\top] \\ &= E[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{C}^\top] \\ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top E[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{C}^\top \\ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\sigma_0^2\boldsymbol{I}\boldsymbol{C}^\top = \boldsymbol{O}\end{aligned}$$

此处使用了条件
$E[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \sigma^2\boldsymbol{I}$

因此，$\mathrm{Var}[\tilde{\boldsymbol{\beta}}] - \mathrm{Var}[\hat{\boldsymbol{\beta}}] = \mathrm{Var}[\boldsymbol{Cy}]$。已知协方差矩阵是半正定，因此 $\hat{\boldsymbol{\beta}}$ 比 $\tilde{\boldsymbol{\beta}}$ 更有效。

Gauss-Markov Theorem

当线性回归模型 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ 满足外生性条件 $E[\boldsymbol{u} \mid \boldsymbol{X}] = \boldsymbol{0}$ 和同方差无自相关条件 $E[\boldsymbol{uu}^{\top} \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$ 时，**OLS 估计量 $\hat{\boldsymbol{\beta}}$ 是线性非偏估计量当中最有效的**。即对任意线性非偏估计量 $\tilde{\boldsymbol{\beta}}$，协方差矩阵之差

$$\mathrm{Var}[\tilde{\boldsymbol{\beta}}] - \mathrm{Var}[\hat{\boldsymbol{\beta}}]$$

是半正定矩阵。

$\hat{\boldsymbol{\beta}}$ 也被称为 BLUE (Best Linear Unbiased Estimator) 。

残差 $\hat{u}$ 与误差项 u

误差项的估计量

Estimator of the Error Term

在线性模型 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ 中，$\boldsymbol{\beta}$ 和 $\boldsymbol{u}$ 是未知的。我们可以将 OLS 残差 $\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 当作误差项 $\boldsymbol{u}$ 的估计量，并用它来估计 σ^2。

下面是 $\hat{\boldsymbol{u}}$ 的一些统计学性质：

- 从 $\hat{\boldsymbol{\beta}}$ 的一致性可以推出 $\hat{\boldsymbol{u}}$ 的一致性，即 $\operatorname{plim}_{n\to\infty} \hat{\boldsymbol{u}} = \boldsymbol{u}$。
- 从外生性可以推出 $E[\hat{u}_t \mid \boldsymbol{X}] = \mathbf{0}$。
- 从同方差无自相关性可以推出 $\operatorname{Var}[\hat{u}_t \mid \boldsymbol{X}] < \operatorname{Var}[u_t \mid \boldsymbol{X}]$。

$$E[\hat{u}_t \mid \boldsymbol{X}] = \mathbf{0}$$

从残差的定义可得 $\hat{\boldsymbol{u}} = \boldsymbol{M}_X \boldsymbol{y} = \boldsymbol{M}_X \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{M}_X \boldsymbol{u} = \boldsymbol{M}_X \boldsymbol{u}$，因此

$$\begin{aligned}\hat{u}_t &= u_t - \boldsymbol{X}_t (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u} \\ &= u_t - \sum_{s=1}^{n} \boldsymbol{X}_t (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_s^\top u_s\end{aligned}$$

$\boldsymbol{M}_X \boldsymbol{u}$ 的第 t 行

因此，

$$\begin{aligned}E[\hat{u}_t \mid \boldsymbol{X}] &= E[u_t \mid \boldsymbol{X}] - \sum_{s=1}^{n} \boldsymbol{X}_t (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_s^\top E[u_s \mid \boldsymbol{X}] \\ &= 0\end{aligned}$$

$\text{Var}[\hat{u}_t \mid \boldsymbol{X}] < \text{Var}[u_t \mid \boldsymbol{X}]$

从 $\hat{\boldsymbol{u}}$ 的协方差矩阵可得

$$\begin{aligned}\text{Var}[\hat{\boldsymbol{u}} \mid \boldsymbol{X}] &= \text{Var}[\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{u} \mid \boldsymbol{X}] = E[\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{u}\boldsymbol{u}^{\top}\boldsymbol{M}_{\boldsymbol{X}} \mid \boldsymbol{X}] \\ &= \boldsymbol{M}_{\boldsymbol{X}}E[\boldsymbol{u}\boldsymbol{u}^{\top}|\boldsymbol{X}]\,\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{M}_{\boldsymbol{X}}\sigma^2\boldsymbol{I}\boldsymbol{M}_{\boldsymbol{X}} \\ &= \sigma^2\boldsymbol{M}_{\boldsymbol{X}}\end{aligned}$$

令 $\boldsymbol{e}_t$ 为第 t 要素为 1 其他要素都为 0 的向量。则 $\boldsymbol{P}_{\boldsymbol{X}}$ 的第 t 对角要素可以表达为

$$h_t = \boldsymbol{e}_t^{\top}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{e}_t = \boldsymbol{e}_t^{\top}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{e}_t = \|\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{e}_t\|^2 \geq 0$$

如果回归模型包含常数项，则 $h_t > 0$。又因为 $\boldsymbol{I} = \boldsymbol{P}_{\boldsymbol{X}} + \boldsymbol{M}_{\boldsymbol{X}}$,

$$\boldsymbol{e}_t = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{e}_t + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{e}_t \quad \Rightarrow \quad h_t = \|\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{e}_t\|^2 \leq \|\boldsymbol{e}_t\|^2 = 1$$

$\text{Var}[\hat{u}_t \mid \boldsymbol{X}]$ 是 $\sigma^2\boldsymbol{M}_{\boldsymbol{X}}$ 的第 t 对角要素，因此可以表达为

$$\text{Var}[\hat{u}_t \mid \boldsymbol{X}] = \sigma^2(1 - h_t) < \sigma^2 = \text{Var}[u_t \mid \boldsymbol{X}]$$

σ^2 的估计

Estimating σ^2

如果误差项的方差 σ^2 未知，我们就需要去估计它。根据矩估计法，可以用误差项的样本方差 $\frac{1}{n}\sum_{t=1}^{n} u_t^2$ 作为估计量。但是误差项无法观察，我们只能考虑用残差替代。

最简单的MM估计量是 $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2$。

$$E[\hat{\sigma}^2 \mid \boldsymbol{X}] = \frac{1}{n}\sum_{t=1}^{n} E[\hat{u}_t^2 \mid \boldsymbol{X}]$$

$$= \frac{1}{n}\sum_{t=1}^{n} \mathrm{Var}[\hat{u}_t \mid \boldsymbol{X}] = \frac{1}{n}\sum_{t=1}^{n} (1 - h_t)\sigma^2$$

因为 $\sum_{t=1}^{n} h_t = k$ （S.2.6, p.80），可得

$$E[\hat{\sigma}^2 \mid \boldsymbol{X}] = \frac{n-k}{n}\sigma^2 < \sigma^2$$

因此，$\hat{\sigma}^2$ 有偏，不适合作为 σ^2 的估计量。而 $s^2 = \frac{1}{n-k}\sum_{t=1}^{n} \hat{u}_t^2$ 则是一个非偏估计量。s 被称作回归标准误（the standard error of regression）。我们可以用 s^2 估计 $\hat{\boldsymbol{\beta}}$ 的协方差矩阵，即

$$\widehat{\mathrm{Var}}[\hat{\boldsymbol{\beta}}] = s^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

均方误差

均方误差

Mean Squared Error

我们在讨论 $\hat{\boldsymbol{\beta}}$ 的估计精确度时假设了 $\hat{\boldsymbol{\beta}}$ 的非偏性。如果一个估计量 $\tilde{\boldsymbol{\beta}}$ 有偏，那就无法用它的协方差矩阵 $\mathrm{Var}[\tilde{\boldsymbol{\beta}}]$ 作为估计精确度的指标。

更具普遍性的估计精确度指标是均方误差（mean squared error, MSE）：

$$
\begin{aligned}
\mathrm{MSE}(\tilde{\boldsymbol{\beta}}) &= E[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top] \\
&= E[(\tilde{\boldsymbol{\beta}} - E[\tilde{\boldsymbol{\beta}}] + E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)(\tilde{\boldsymbol{\beta}} - E[\tilde{\boldsymbol{\beta}}] + E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)^\top] \\
&= E[(\tilde{\boldsymbol{\beta}} - E[\tilde{\boldsymbol{\beta}}])(\tilde{\boldsymbol{\beta}} - E[\tilde{\boldsymbol{\beta}}])^\top] + E[(E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)(E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)^\top] \\
&= \mathrm{Var}[\tilde{\boldsymbol{\beta}}] + (E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)(E[\tilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0)^\top
\end{aligned}
$$

因此，对于单一参数估计量，$\mathrm{MSE} = \mathrm{Var} + \mathrm{bias}^2$。当估计量是非偏时，$\mathrm{MSE} = \mathrm{Var}$，因此我们可以用协方差矩阵衡量非偏估计量的估计精确度。对于一致但有偏的估计量，则应该用 MSE。