高级计量经济学

Lecture 6: Statistical Properties of OLS (2)

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

$\hat{\beta}$ 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 $\boldsymbol{\mu} = E[\boldsymbol{x}] = \boldsymbol{0}$，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 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

$\hat{\boldsymbol{\beta}}$ 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

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

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 $\mathrm{Var}[a] < \mathrm{Var}[b]$。

- Scalars：a 比 b 更有效 $\Leftrightarrow$ $\mathrm{Var}[a] < \mathrm{Var}[b]$。
- Vectors：$\boldsymbol{a}$ 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 $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{O}$。

$\hat{\beta}$ 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 $E[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$ 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 (<u>B</u>est <u>L</u>inear <u>U</u>nbiased <u>E</u>stimator) 。

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 $E[\hat{u}_t \mid \boldsymbol{X}] = \boldsymbol{0}$。
- 从同方差无自相关性可以推出 $\mathrm{Var}[\hat{u}_t \mid \boldsymbol{X}] < \mathrm{Var}[u_t \mid \boldsymbol{X}]$。

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

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

$$\begin{aligned}\hat{u}_t &= u_t - \boldsymbol{X}_t(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u} \qquad \boldsymbol{M}_X\boldsymbol{u} \text{ 的第 } t \text{ 行} \\ &= u_t - \sum_{s=1}^{n} \boldsymbol{X}_t(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}_s^\top u_s\end{aligned}$$

因此，

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

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

从 $\hat{\boldsymbol{u}}$ 的协方差矩阵可得

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

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

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

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

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

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

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

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 $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2$。

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

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。