

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

$\mathbf{x} = (X_1, X_2, \dots, X_n)^\top$ 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

$$\begin{aligned}\text{Var}[\mathbf{x}] &= \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix} \\ &= E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^\top] \\ &= E[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top\end{aligned}$$

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

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 $\Leftrightarrow \underset{(\geq)}{x^\top A x} > 0$ for all $x \neq \mathbf{0}$.

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

考虑随机向量 x 的任意线性结合 $w^\top x$:

$$\begin{aligned}\text{Var}[w^\top x] &= E[w^\top x x^\top w] - E[w^\top x] E[x^\top w] \\ &= w^\top E[x x^\top] w - w^\top E[x] E[x^\top] w \\ &= w^\top \text{Var}[x] w \geq 0\end{aligned}$$

w 非随机且已知

$w^\top x$ 为 scalar

因此, $\text{Var}[x]$ 是半正定。

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

$\hat{\beta}$ 的协方差矩阵

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

$\hat{\beta}$ 的协方差矩阵是

$$\begin{aligned}\text{Var}[\hat{\beta}] &= E[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])^\top] \\&= E[(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)^\top] \\&= E[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u} \mathbf{u}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}] \\&= E_{\mathbf{X}} \left[E[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u} \mathbf{u}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mid \mathbf{X}] \right] \\&= E_{\mathbf{X}} [(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E[\mathbf{u} \mathbf{u}^\top \mid \mathbf{X}] \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}] \\&= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}\end{aligned}$$
$$\hat{\beta} = \beta_0 + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u}$$

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

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

预测误差的方差

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

$\hat{\gamma}$ 的协方差是

$$\begin{aligned}\text{Var}[\hat{\gamma}] &= \text{Var}[\omega^\top \hat{\beta}] = E[\omega^\top (\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)^\top \omega] \\ &= \omega^\top E[(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)^\top] \omega \\ &= \omega^\top \text{Var}[\hat{\beta}] \omega \\ &= \omega^\top \sigma_0^2 (X^\top X)^{-1} \omega\end{aligned}$$

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 X_s 预测 y_s , 此时预测值为 $\hat{y}_s = X_s \hat{\beta}$ 。当外生性成立时, $\hat{\beta}$ 为非偏, 因此 \hat{y}_s 也非偏。预测误差的方差是

$$\begin{aligned}\text{Var}[y_s - \hat{y}_s] &= E[(y_s - X_s \hat{\beta})^2] = E[(X_s \beta_0 + u_s - X_s \hat{\beta})^2] \\ &= E[u_s^2] + E[(X_s \beta_0 - X_s \hat{\beta})^2] \quad \text{因假设了外生性} \\ &= \sigma_0^2 + \text{Var}[X_s \hat{\beta}] \\ &= \sigma_0^2 + X_s \text{Var}[\hat{\beta}] X_s^\top \\ &= \sigma_0^2 + \sigma_0^2 X_s (X^\top X)^{-1} X_s^\top\end{aligned}$$

$\hat{\beta}$ 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 $\text{Var}[a]^{-1}$ 成正比，因此 a 比 b 更有效可以表达为 $\text{Var}[a]^{-1} > \text{Var}[b]^{-1}$ ，或 $\text{Var}[a] < \text{Var}[b]$ 。

- Scalars: a 比 b 更有效 $\Leftrightarrow \text{Var}[a] < \text{Var}[b]$ 。
- Vectors: a 比 b 更有效 $\Leftrightarrow (\text{Var}[b] - \text{Var}[a])$ 是非零半正定矩阵。

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

$$\begin{aligned}\text{Var}[\gamma_a] \leq \text{Var}[\gamma_b] &\Leftrightarrow \omega^\top \text{Var}[a] \omega \leq \omega^\top \text{Var}[b] \omega \\ &\Leftrightarrow \omega^\top (\text{Var}[b] - \text{Var}[a]) \omega \geq 0\end{aligned}$$

$\hat{\beta}$ 的有效性

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

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

如果定义 $C = A - (X^T X)^{-1} X^T$, 则有

$$\tilde{\beta} = Ay = (X^T X)^{-1} X^T y + Cy = \hat{\beta} + Cy$$

如果我们要求 $\tilde{\beta}$ 为非偏估计量, 则需要保证 $E[\tilde{\beta}] = \beta_0$, 即

$$\begin{aligned} E[\tilde{\beta}] &= E[Ay] = E[A(X\beta_0 + u)] \\ &= E[AX\beta_0] + E[Au] = \beta_0 \end{aligned}$$

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

$\hat{\beta}$ 的有效性

当满足外生性和 $CX = \mathbf{0}$ 时, $\hat{\beta}$ 和 $\tilde{\beta} = \hat{\beta} + Cy$ 都是 β 的线性非偏估计量。

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

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

$$\begin{aligned}\text{Cov}[\hat{\beta}, Cy] &= E[(\hat{\beta} - \beta_0)(Cy)^\top] \\ &= E[(X^\top X)^{-1}X^\top uu^\top C^\top] \\ &= (X^\top X)^{-1}X^\top E[uu^\top]C^\top \\ &= (X^\top X)^{-1}X^\top \sigma_0^2 I C^\top = \mathbf{0}\end{aligned}$$

此处使用了条件
 $E[uu^\top | X] = \sigma^2 I$

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

Gauss-Markov Theorem

当线性回归模型 $y = X\beta + u$ 满足外生性条件 $E[u | X] = \mathbf{0}$ 和同方差无自相关条件 $E[uu^\top | X] = \sigma^2 I$ 时，OLS 估计量 $\hat{\beta}$ 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 $\tilde{\beta}$ ，协方差矩阵之差

$$\text{Var}[\tilde{\beta}] - \text{Var}[\hat{\beta}]$$

是半正定矩阵。

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

下面是 \hat{u} 的一些统计学性质:

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

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

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

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

因此，

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

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

从 \hat{u} 的协方差矩阵可得

$$\begin{aligned}\text{Var}[\hat{u} | X] &= \text{Var}[M_X u | X] = E[M_X u u^\top M_X | X] \\ &= M_X E[u u^\top | X] M_X = M_X \sigma^2 I M_X \\ &= \sigma^2 M_X\end{aligned}$$

令 e_t 为第 t 要素为 1 其他要素都为 0 的向量。则 P_X 的第 t 对角要素可以表达为

$$h_t = e_t^\top P_X e_t = e_t^\top P_X P_X e_t = \|P_X e_t\|^2 \geq 0$$

如果回归模型包含常数项，则 $h_t > 0$ 。又因为 $I = P_X + M_X$,

$$e_t = P_X e_t + M_X e_t \Rightarrow h_t = \|P_X e_t\|^2 \leq \|e_t\|^2 = 1$$

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

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

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

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

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

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

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

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

均方误差

均方误差

Mean Squared Error

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

$$\begin{aligned}\text{MSE}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta_0)(\tilde{\beta} - \beta_0)^\top] \\ &= E[(\tilde{\beta} - E[\tilde{\beta}] + E[\tilde{\beta}] - \beta_0)(\tilde{\beta} - E[\tilde{\beta}] + E[\tilde{\beta}] - \beta_0)^\top] \\ &= E[(\tilde{\beta} - E[\tilde{\beta}])(\tilde{\beta} - E[\tilde{\beta}])^\top] + E[(E[\tilde{\beta}] - \beta_0)(E[\tilde{\beta}] - \beta_0)^\top] \\ &= \text{Var}[\tilde{\beta}] + (E[\tilde{\beta}] - \beta_0)(E[\tilde{\beta}] - \beta_0)^\top\end{aligned}$$

因此，对于单一参数估计量， $\text{MSE} = \text{Var} + \text{bias}^2$ 。当估计量是非偏时， $\text{MSE} = \text{Var}$ ，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。