计量经济学

第 7 讲：OLS 估计的小样本推断

授课人：刘　岩

中山大学商学院

2025 年 3 月 24 日

本讲内容

1. OLS 估计方差无偏性证明
2. 多元正态分布的基本性质
3. 小样本 OLS 估计的分布

本节内容

方差估计无偏性的证明

结论 在线性回归基本假设 1–3 下，给定模型 OLS 估计所得残差向量 $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$，则

$$s^2 = \frac{SSR_{OLS}}{N-K} = \frac{\boldsymbol{e}^\top \boldsymbol{e}}{N-K}$$

为模型误差项方差 σ^2 的无偏估计量，即 $\mathbb{E}[s^2] = \sigma^2$

证明 将证明 $\mathbb{E}[s^2|\boldsymbol{X}] = \sigma^2$，进而得到 $\mathbb{E}[s^2] = \sigma^2$。首先，首先注意到 $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} = (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y}$；定义矩阵 $\boldsymbol{M} \equiv \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$，则有 $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$，故 $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{Y} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) = \boldsymbol{M}\boldsymbol{\varepsilon}$；同时，易验证 $\boldsymbol{M}^2 = \boldsymbol{M}$，由此知 $\boldsymbol{e}^\top\boldsymbol{e} = \boldsymbol{\varepsilon}^\top\boldsymbol{M}\boldsymbol{\varepsilon}$，故 $\mathbb{E}[\boldsymbol{e}^\top\boldsymbol{e}|\boldsymbol{X}] = \mathbb{E}[\boldsymbol{\varepsilon}^\top\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}]$。

方差估计无偏性的证明

继续 其次，令 m_{ij} 为 $\boldsymbol{M}$ 的第 i 行、第 j 列元素，则有 $\boldsymbol{\varepsilon}^{\intercal}\boldsymbol{M}\boldsymbol{\varepsilon}=\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_i\varepsilon_j m_{ij}$，取期望可得

$$\mathbb{E}[\boldsymbol{\varepsilon}^{\intercal}\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}]=\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\varepsilon_i\varepsilon_j m_{ij}|\boldsymbol{X}]=\sum_{i=1}^{N}\sum_{j=1}^{N}m_{ij}\mathbb{E}[\varepsilon_i\varepsilon_j|\boldsymbol{X}]=\sum_{i=1}^{N}\sigma^2 m_{ii}=\sigma^2\mathrm{tr}(\boldsymbol{M})$$

最后，由矩阵求迹 $\mathrm{tr}(\cdot)$ 的性质可知

$$\begin{aligned}\mathrm{tr}(\boldsymbol{M})=\mathrm{tr}(\boldsymbol{I}_N-\boldsymbol{X}(\boldsymbol{X}^{\intercal}\boldsymbol{X})^{-1}\boldsymbol{X}^{\intercal})&=\mathrm{tr}(\boldsymbol{I}_N)-\mathrm{tr}(\boldsymbol{X}(\boldsymbol{X}^{\intercal}\boldsymbol{X})^{-1}\boldsymbol{X}^{\intercal})\\&=N-\mathrm{tr}((\boldsymbol{X}^{\intercal}\boldsymbol{X})^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{X})=N-\mathrm{tr}(\boldsymbol{I}_K)=N-K\end{aligned}$$

故 $\mathbb{E}[s^2|\boldsymbol{X}]=\frac{(N-K)\sigma^2}{N-K}=\sigma^2$。

本节内容

协方差矩阵：矩阵表示

- 给定随机向量 $\boldsymbol{X} = [X_1, \ldots, X_n]^\intercal$，其协方差矩阵 $\mathrm{var}(\boldsymbol{X})$ 定义为

$$\mathrm{var}(\boldsymbol{X}) = \boldsymbol{\Sigma}_X \equiv \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \cdots & \mathrm{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \cdots & \mathrm{cov}(X_n, X_n) \end{bmatrix}$$

- $\boldsymbol{X}$ 的期望记为 $\boldsymbol{\mu}_X$，则

$$\boldsymbol{\Sigma}_X = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu}_X)(\boldsymbol{X} - \boldsymbol{\mu}_X)^\intercal\right] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\intercal] - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^\intercal$$

随机向量的仿射变换

给定 n-维随机向量 $\boldsymbol{X} = [X_1, \ldots, X_n]^\top$，假设有 $m \times n$ 常数变换矩阵 $\boldsymbol{A}$，以及 $m \times 1$ 常数平移向量 (translation vector) $\boldsymbol{c}$

- $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{c}$ 是一个 m-维随机向量，称为 $\boldsymbol{X}$ 的仿射变换 (affine transformation)
  - 当 $\boldsymbol{c} = \boldsymbol{0}$ 时，仿射变换退化为线性变换
- 期望的仿射变换：$\boldsymbol{\mu}_Y = \mathbb{E}\boldsymbol{Y} = \mathbb{E}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{c} = \boldsymbol{A}\mathbb{E}\boldsymbol{X} + \boldsymbol{c} = \boldsymbol{A}\boldsymbol{\mu}_X + \boldsymbol{c}$
- $\boldsymbol{Y}$ 的协方差矩阵 $\boldsymbol{\Sigma}_Y$ 可写为

$$
\begin{aligned}
\boldsymbol{\Sigma}_Y &= \mathbb{E}[\boldsymbol{Y}\boldsymbol{Y}^\top] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^\top \\
&= \mathbb{E}[(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{c})(\boldsymbol{X}^\top\boldsymbol{A}^\top + \boldsymbol{c}^\top)] - (\boldsymbol{A}\boldsymbol{\mu}_X + \boldsymbol{c})(\boldsymbol{\mu}_X^\top\boldsymbol{A}^\top + \boldsymbol{c}^\top) \\
&= \boldsymbol{A}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^\top\right)\boldsymbol{A}^\top \\
&= \boldsymbol{A}\boldsymbol{\Sigma}_X\boldsymbol{A}^\top
\end{aligned}
$$

多元正态分布的仿射变换

- n 维正态分布：$\boldsymbol{X} = [X_1, \ldots, X_n] \sim \mathcal{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$，令 $\boldsymbol{x} = [x_1, \ldots, x_n]^\intercal$，其密度函数为

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}_X}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_X)^\intercal \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_X) \right\}$$

- 考虑仿射变换 $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{c}$，$\boldsymbol{A}$ 为 $m \times n$ 常数变换矩阵，$\boldsymbol{c}$ 为 $m \times 1$ 常数向量，则 $\boldsymbol{Y}$ 服从 m 元正态分布，即 $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y) = \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}_X + \boldsymbol{c}, \boldsymbol{\Sigma}_X)$

结论 若 $\boldsymbol{X}$ 服从多维正态分布，且对某两个 X_i, X_j 有 $\operatorname{cov}(X_i, X_j) = 0$，则 X_i, X_j 相互独立

例 考虑 2 维标准正态 $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_2)$，则 $Y_1 = X_1 + X_2, Y_2 = X_1 - X_2$ 相互独立

本节内容

OLS 估计系数的分布

- 假设 DGP 中系数向量真实值为 $\boldsymbol{\beta}_0$，OLS 估计为 $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}$，则在假设 1–3 之下 $\mathbb{E}[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|X] = \boldsymbol{0}$ 且

$$
\begin{aligned}
\operatorname{var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = \operatorname{var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|\boldsymbol{X}) &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \underbrace{\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}]}_{=\sigma^2\boldsymbol{I}} \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}
\end{aligned}
$$

- 进一步引入假设 4，则条件在 $\boldsymbol{X}$ 上时，$\hat{\boldsymbol{\beta}}$ 的随机性完全来自 $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$，且为 $\boldsymbol{\varepsilon}$ 的仿射变换，故同样服从多元正态分布，即

$$
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1})
$$

OLS 估计系数的假设检验：单系数情形

- 给定 $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 | \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1})$，似乎可以对任意 $\beta_k, k = 0, \ldots, K-1$ 进行假设检验
  - 如果 σ^2 确实是已知参数，则 $\hat{\beta}_k - \beta_{k,0} \sim \mathcal{N}(0, \sigma^2[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{k+1,k+1})$，其中 $[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{k+1,k+1}$ 表示矩阵 $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ 对角线上第 $k+1$ 个元素，此时

$$z_k \equiv \frac{\hat{\beta}_k - \beta_{k,0}}{\sqrt{\sigma^2[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{k+1,k+1}}} \sim \mathcal{N}(0,1)$$

    是一个直接可用的检验统计量：给定原假设下 $\beta_{k,0}$ 的取值，由样本 $\{Y_i, \boldsymbol{X}_i\}_{i=1}^N$ 可直接计算分子和分母取值，并且知晓该统计量服从标准正态分布

- 问题在于，σ^2 也是模型未知参数
- 简单的想法：用样本统计量 $s^2 = \boldsymbol{e}^\top \boldsymbol{e}/(N-K)$ 代替 σ^2

单系数检验的 t-统计量

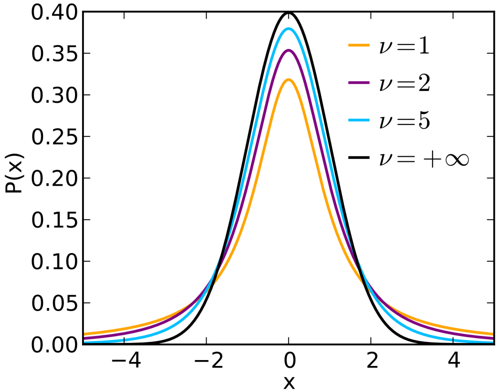
- 对回归系数 β_k，构造如下 t-统计量

$$t_k \equiv \frac{\hat{\beta}_k - \beta_{k,0}}{\sqrt{s^2[(\mathbf{X}^\top\mathbf{X})^{-1}]_{k+1,k+1}}}, \quad k = 0, \ldots, K-1$$

结论 假设 1–4 成立时，给定样本 $\mathbf{X}$ 的条件下，统计量 $t_k|\mathbf{X}$ 服从自由度为 $N-K$ 的 t-分布

- t-分布的密度函数可以直接写出，但没有必要掌握（记住）
- 给定两个相互独立的随机变量 W, Z，其中 $W \sim \mathcal{N}(0,1)$ 而 $Z \sim \chi^2(N-K)$，即 Z 服从自由度为 $N-K$ 的 χ^2 分布，则 $U = \frac{W}{\sqrt{Z/(N-K)}}$ 服从自由度为 $N-K$ 的 student t-分布，记作 $t(N-K)$ 分布
- 自由度为 L 的 χ^2 分布，念卡方分布，记作 $\chi^2(L)$ 分布，定义为 L 个相互独立的标准正态分布随机变量的平方和

Student t-分布密度函数：自由度 $\nu = 1, 2, 5$



当 $\nu \to \infty$ 时，t-分布密度函数趋近于标准正态分布，即 t 分布当自由度趋近于无穷时，分布函数收敛到标准正态分布

单系数检验的 t-统计量：证明

证明 t_k 可写为

$$t_k \equiv \frac{\hat{\beta}_k - \beta_{k,0}}{\sqrt{\sigma^2[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{k+1,k+1}} \cdot \sqrt{\frac{s^2}{\sigma^2}}} = \frac{z_k}{\sqrt{\frac{\boldsymbol{e}^\top \boldsymbol{e}/(N-K)}{\sigma^2}}}$$

注意到条件在 $\boldsymbol{X}$ 下时，分子 $z_k \sim \mathcal{N}(0,1)$，故只需证明分母与分子独立且为 $\chi^2(N-K)$ 分布。

为此，注意到 $z_k|\boldsymbol{X}$ 通过 $\hat{\beta}_k$ 是 $\boldsymbol{\varepsilon}$ 的线性变换，而 $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{\varepsilon}$ 也是 $\boldsymbol{\varepsilon}$ 的线性变换，故 $(\hat{\boldsymbol{\beta}}, \boldsymbol{e})|\boldsymbol{X}$ 为联合正态分布，于是由 $\mathrm{cov}(\hat{\boldsymbol{\beta}}, \boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{0}_{K,N}$（最后证明）说明二者无相关性，进而相互独立。这进一步意味着 z_k 与 $\boldsymbol{e}^\top \boldsymbol{e}$ 独立，故分子与分母独立。而 $\frac{1}{\sigma^2}\boldsymbol{e}^\top \boldsymbol{e} = \frac{1}{\sigma^2}\boldsymbol{\varepsilon}^\top \boldsymbol{M}\boldsymbol{\varepsilon}$ 可证明服从 $\chi^2(N-K)$ 分布（不做要求）。

单系数检验的 t-统计量：证明

继续 最后验证 $\mathrm{cov}(\hat{\boldsymbol{\beta}}, \boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{0}_{K,N}$。事实上，该条件协方差定义为 $\mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\boldsymbol{e}^\top|\boldsymbol{X}]$，而 $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}$，$\boldsymbol{e}^\top = \boldsymbol{\varepsilon}^\top\boldsymbol{M}$，故

$$\begin{aligned}\mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\boldsymbol{e}^\top|\boldsymbol{X}] &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}]\boldsymbol{M} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\sigma^2\boldsymbol{I})\boldsymbol{M} \\ &= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) = \boldsymbol{0}_{K,N}\end{aligned}$$