计量经济学

第 6 讲：OLS 估计的小样本性质

授课人：刘　岩

中山大学商学院

2025 年 3 月 17 日

本讲内容

本节内容

线性回归模型：小样本 vs 大样本

考虑线性回归模型

$$Y_i = \boldsymbol{X}_i^{\intercal}\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, N$$

其中 $\boldsymbol{X}_i$ 为 K 个解释变量的列向量

- 小样本 (small sample)：给定有限样本量，$N < \infty$，且通常而言 $N \leq 100$
  - 重点在于样本量 N 有限且给定，满足 $N > K$
  - 此时可以简单将 $\boldsymbol{X} = \left[\boldsymbol{X}_1^{\intercal}, \ldots, \boldsymbol{X}_N^{\intercal}\right]^{\intercal}$ 视作给定的 $N \times K$ 矩阵
  - $\boldsymbol{X}_i$ 的随机性不重要，DGP 随机性焦点在 $\{\varepsilon_i\}$，不过为与大样本理论衔接，依然假设 $\{X_i\}$ 随机
- 大样本 (large sample)：样本量趋于无穷 $N \to \infty$
  - 这是一个理论假定 (theoretic postulate)，现实应用中样本量都是有限的，但不妨碍考虑 $N \to \infty$ 的情况
  - 此时 $\boldsymbol{X}_i$ 的随机性质就变得非常关键，需要一系列额外假设

小样本回归的标准假设

考虑小样本线性回归模型

$$Y_i = \boldsymbol{X}_i^{\intercal}\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, N < \infty$$

相应的标准假设如下：

1. 外生性：$\mathbb{E}[\varepsilon_i|\boldsymbol{X}] = \mathbb{E}[\varepsilon_i|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N] = 0, i = 1, \ldots, N$
   - $\mathbb{E}[W|Z]$ 表示 W 条件在 Z 上的期望，称为条件期望 (conditional expectation)
   - 外生性意味着 $\mathbb{E}\varepsilon_i = 0$ 以及 $\mathbb{E}[X_{ki}\varepsilon_i] = 0, k = 0, \ldots, K-1, i = 1, \ldots, N$
2. 无多重共线性：矩阵 $\boldsymbol{X}$ 列满秩，即 $\text{rank}(\boldsymbol{X}) = K < N$
   - 多重共线性 (multi-colinearity)：K 个变量之间线性相关，即某个线性组合为常数，方差为 0
   - 此时 $\boldsymbol{X}^{\intercal}\boldsymbol{X}$ 为可逆矩阵：注意到 $\text{rank}(\boldsymbol{X}) = K$ 时 $\boldsymbol{X}^{\intercal}\boldsymbol{X}$ 为正定矩阵，故可逆

小样本回归的标准假设

3. **球面误差**：误差项方差为常数 $\mathbb{E}[\varepsilon_i^2|\boldsymbol{X}] = \sigma^2 > 0, i = 1, \ldots, N$，以及不同观测间误差项无相关性 $\mathbb{E}[\varepsilon_i \varepsilon_j|\boldsymbol{X}] = 0, i, j = 1, \ldots, N, i \neq j$
   - 又称为同方差 (homoskedasticity) 假设，其矩阵形式为 $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}] = \sigma^2 \boldsymbol{I}_N$
   - 上述假设弱于 $\{\varepsilon_i\}$ 为 iid 序列；若 $\{\varepsilon_i\}$ iid，且与 $\boldsymbol{X}$ 独立，则自动满足假设 1 和 3
4. **误差正态分布**：$\{\varepsilon_i\}_{i=1}^N \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$，且 ε_i 与 $\boldsymbol{X}$ 相互独立，$i = 1, \ldots, N$
   - 该假设同时蕴含假设 1 和 3，是线性回归模型经典 (classical) 理论的主要假设

注 对于线性回归模型小样本经典理论而言，$\boldsymbol{X}$ 的随机性并不重要，上述假设中 ε_i 关于 $\boldsymbol{X}$ 的条件期望，均可换为 ε_i 的无条件期望，称为固定回归变量 (fixed regressor)

本节内容

线性回归模型的拟合值

考虑给定样本 $\{Y_i, \boldsymbol{X}_i\}_{i=1}^N$ 下的线性回归模型 $Y_i = \boldsymbol{X}_i^\intercal \boldsymbol{\beta} + \varepsilon_i, i = 1, \ldots, N$

- 给定 OLS 估计 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$，则

$$\hat{Y}_i = \boldsymbol{X}_i^\intercal \hat{\boldsymbol{\beta}}, \quad i = 1, \ldots, N$$

称为 Y_i 的 OLS 拟合值 (fitted value)，或预测值 (prediction/predicted value)

- $e_i \equiv Y_i - \hat{Y}_i = Y_i - \boldsymbol{X}_i^\intercal \hat{\boldsymbol{\beta}}$ 称为 OLS 残差 (residual)
- 残差向量 $\boldsymbol{e} = [e_1, \ldots, e_N]^\intercal$ 满足

$$\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}, \qquad \boldsymbol{X}^\intercal \boldsymbol{e} = \boldsymbol{0}$$

后一等式表示残差向量 $\boldsymbol{e}$ 与 $\boldsymbol{X}$ 的所有列向量正交

残差平方和与方差分解

- 给定任意的拟合值 $\{\hat{Y}_i\}$ 及相应的残差 $\{e_i\}$，可定义残差平方和 (sum of squared residuals) 为

$$SSR = \sum_{i=1}^{N} e_i^2 = \boldsymbol{e}^\top \boldsymbol{e}$$

- 对于 OLS 拟合值及残差，有 $Y_i = \hat{Y}_i + e_i$ 或 $\boldsymbol{Y} = \hat{\boldsymbol{Y}} + \boldsymbol{e}$，且 $\hat{\boldsymbol{Y}}^\top \boldsymbol{e} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{e} = \boldsymbol{0}$，故因变量样本变动 (sample variation) 满足：

$$\sum_{i=1}^{N} Y_i^2 = \boldsymbol{Y}^\top \boldsymbol{Y} = (\hat{\boldsymbol{Y}} + \boldsymbol{e})^\top (\hat{\boldsymbol{Y}} + \boldsymbol{e}) = \hat{\boldsymbol{Y}}^\top \hat{\boldsymbol{Y}} + \hat{\boldsymbol{e}}^\top \hat{\boldsymbol{e}} = \sum_{i=1}^{N} \hat{Y}_i^2 + SSR_{OLS}$$

上述结果又称为因变量的方差分解 (variance decomposition)

线性回归模型的 R^2

- OLS 估计下，$\boldsymbol{Y}^\top\boldsymbol{Y}=\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}}+\hat{\boldsymbol{e}}^\top\hat{\boldsymbol{e}}$，故可定义线性回归的 R^2 为

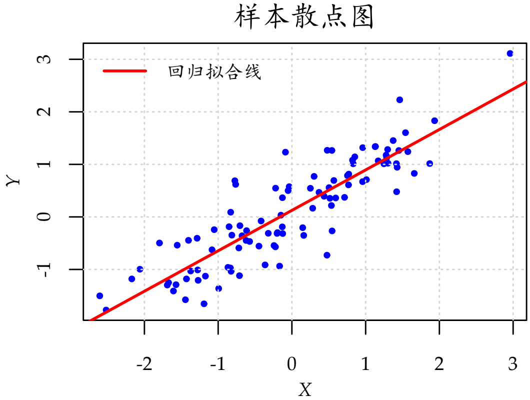
$$R^2=\frac{\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}}}{\boldsymbol{Y}^\top\boldsymbol{Y}}=1-\frac{\hat{\boldsymbol{e}}^\top\hat{\boldsymbol{e}}}{\boldsymbol{Y}^\top\boldsymbol{Y}}=1-\frac{SSR_{OLS}}{\boldsymbol{Y}^\top\boldsymbol{Y}}$$

  R^2 反映了线性回归模型对数据样本的整体拟合情况，或模型的解释力 (explanatory power)

- 又由 $\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}},\hat{\boldsymbol{e}}^\top\hat{\boldsymbol{e}}\geq 0$，可知 $R^2\in[0,1]$
- 对非 OLS 估计，也可以类似定义 $R^2=1-\frac{SSR}{\boldsymbol{Y}^\top\boldsymbol{Y}}$，但此时 R^2 可能为负值，即 $SSR>\boldsymbol{Y}^\top\boldsymbol{Y}$，这在工具变量估计 (2SLS) 中很常见

线性回归与拟合值示例

DGP: $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.5^2), X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), Y_i = 0.2 + 0.8X_i + \varepsilon_i, N = 100$



回归模型 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ 估计值为 $\hat{\beta}_0 = 0.1248, \hat{\beta}_1 = 0.7696, R^2 = 0.7762$

本节内容

OLS 系数估计的无偏性

- 假设线性回归模型的数据生成过程 (DGP) 中，回归系数的真实值为 $\boldsymbol{\beta}_0$，即 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$，则回归系数的 OLS 估计满足

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}$$

- 上式两端取期望，可得 $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0 + \mathbb{E}[(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}]$
- 由<u>全期望律</u> (law of total expectation)，即 $\mathbb{E}[W] = \mathbb{E}_Z\{\mathbb{E}[W|Z]\}$，可知

$$\mathbb{E}[(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}] = \mathbb{E}_X\{\mathbb{E}[(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}|\boldsymbol{X}]\}$$

- 在假设 1 下，由条件期望的<u>提出性</u> (pulling out)，即对任意函数 $f(\cdot)$ 有 $\mathbb{E}[f(Z)W|Z] = f(Z)\mathbb{E}[W|Z]$，得

$$\mathbb{E}_X\{\mathbb{E}[(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}|\boldsymbol{X}]\} = \mathbb{E}_X\{(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}]\} = \mathbb{E}_X\{\boldsymbol{0}\} = \boldsymbol{0}$$

OLS 系数估计的无偏性

结论 在假设 1–2 下，线性回归模型系数的 OLS 估计 $\hat{\boldsymbol{\beta}}$ 是模型系数真实值 $\boldsymbol{\beta}_0$ 的无偏估计，即

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0$$

证明 前页计算可知

$$\mathbb{E}[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon}] = \boldsymbol{0}$$

故 $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0 + \mathbb{E}[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon}] = \boldsymbol{\beta}_0$

方差的无偏估计

在假设 1–3 下，可证明 OLS 估计得到的残差平方和 SSR_{OLS} 及相应的

$$s^2 \equiv \frac{SSR_{OLS}}{N-K}$$

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