Problem Set

1 Finite-Sample Properties of OLS

1.1 新增样本点

▷ **问题**. 给定简单线性模型

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

其中

$$\mathbf{y} = (y_1, \cdots, y_n)',\ \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},\ \boldsymbol{\beta} = (\beta_0, \beta_1)',\ \boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'.$$

设该模型在 Gauss-Markov 假设下，OLS 估计量为 $\mathbf{b} = (b_0, b_1)'$. 若新增一个样本点

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1},$$

求 $\mathrm{Var}(\hat{y}_{n+1} - y_{n+1}|\mathbf{X}, x_{n+1})$，其中 $\hat{y}_{n+1} = b_0 + x_{n+1}b_1$. 在给定 $\mathbf{X}$ 的情况下，求出当 x_{n+1} 为何值时， $\mathrm{Var}(\hat{y}_{n+1} - y_{n+1}|\mathbf{X}, x_{n+1})$ 有最小值.

▷ **解答**.

$$\begin{aligned}\hat{y}_{n+1} - y_{n+1} &= (b_0 + x_{n+1}b_1) - (\beta_0 + x_{n+1}\beta_1 + \varepsilon_{n+1}) \\ &= (b_0 - \beta_0) + x_{n+1}(b_1 - \beta_1) - \varepsilon_{n+1}\end{aligned}$$

记 $\bar{x} = n^{-1}\sum_{i=1}^n x_i$. 因为

$$\begin{aligned}\mathrm{Var}(\mathbf{b}|\mathbf{X}, x_{n+1}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} &= \sigma^2 \begin{pmatrix} n & \sum\limits_{i=1}^n x_i \\ \sum\limits_{i=1}^n x_i & \sum\limits_{i=1}^n x_i^2 \end{pmatrix}^{-1} \\ &= \frac{\sigma^2}{\sum\limits_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} n^{-1}\sum\limits_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.\end{aligned}$$

所以

$$\mathrm{Var}(b_0 - \beta_0|\mathbf{X}, x_{n+1}) = \frac{\sigma^2 \sum\limits_{i=1}^n x_i^2}{n\sum\limits_{i=1}^n (x_i - \bar{x})^2},$$

$$\mathrm{Var}[x_{n+1}(b_1 - \beta_1)|\mathbf{X}, x_{n+1}] = \frac{\sigma^2 x_{n+1}^2}{\sum\limits_{i=1}^n (x_i - \bar{x})^2},$$

$$\mathrm{Var}(\varepsilon_{n+1}|\mathbf{X}, x_{n+1}) = \sigma^2,$$

$$\mathrm{Cov}[b_0-\beta_0, x_{n+1}(b_1-\beta_1)|\mathbf{X}, x_{n+1}] = -\frac{\sigma^2\bar{x}x_{n+1}}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2},$$
$$\mathrm{Cov}(b_0-\beta_0, \varepsilon_{n+1}|\mathbf{X}, x_{n+1}) = \mathrm{Cov}[x_{n+1}(b_1-\beta_1), \varepsilon_{n+1}|\mathbf{X}, x_{n+1}] = 0.$$

由此可得

$$\begin{aligned}\mathrm{Var}(\hat{y}_{n+1}-y_{n+1}|\mathbf{X}, x_{n+1}) &= \frac{\sigma^2\sum\limits_{i=1}^{n}x_i^2 + n\sigma^2x_{n+1}^2 - 2n\sigma^2\bar{x}x_{n+1}}{n\sum\limits_{i=1}^{n}(x_i-\bar{x})^2} + \sigma^2 \\ &= \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\left(x_{n+1}^2 - 2\bar{x}x_{n+1} + \frac{1}{n}\sum_{i=1}^{n}x_i^2\right) + \sigma^2.\end{aligned}$$

当 $x_{n+1}=\bar{x}$ 时，$\mathrm{Var}(\hat{y}_{n+1}-y_{n+1}|\mathbf{X}, x_{n+1})$ 有最小值.

1.2 增加解释变量个数会提高 R^2

▷ **问题**. 证明对简单线性模型

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times K}\boldsymbol{\beta}_{K\times 1} + \boldsymbol{\varepsilon}_{n\times 1},$$

进行 OLS 回归时，使用 $K-1$ 个解释变量的 R^2 小于等于使用 K 个解释变量时的的 R^2.

▷ **解答**. 使用 K 个自变量时，OLS 估计量 $\mathbf{b}$ 最小化了残差平方和

$$\mathbf{b} = \arg\min_{\hat{\boldsymbol{\beta}}\in\mathbb{R}^K}\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\|^2.$$

因此，对任意 $\hat{\boldsymbol{\beta}}\in\mathbb{R}^K$, 有

$$\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\|^2 \geq \|\mathbf{y}-\mathbf{X}\mathbf{b}\|^2 = SSR_K.$$

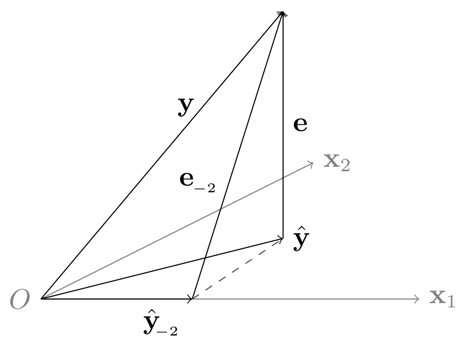
设使用 $K-1$ 个自变量时对应的 OLS 估计量为 $\mathbf{b}_{-K}\in\mathbb{R}^{K-1}$. 令 $\mathbf{b}_{-K}^* = (\mathbf{b}_{-K}, 0)$, 利用上式得到

$$SSR_{K-1} = \|\mathbf{y}-\mathbf{X}\mathbf{b}_{-K}\|^2 = \|\mathbf{y}-\mathbf{X}\mathbf{b}_{-K}^*\|^2 \geq SSR_K.$$

由 R^2 的表达式

$$R^2 = 1 - \frac{SSR}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}$$

可知使用 $K-1$ 个解释变量的 R^2 小于等于使用 K 个解释变量时的的 R^2. 该结果的几何解释如下图所示.



这幅图展示了 $K=2$ 的低维情形，其中 $\mathbf{x}_1, \mathbf{x}_2$ 是 $\mathbf{X}$ 的列向量，$\mathbf{e}=\mathbf{y}-\mathbf{X}\mathbf{b}$. 根据直角三角形的斜边长度大于直角边这一性质可以得到

$$SSR_1 = \|\mathbf{e}_{-2}\|^2 \geq \|\mathbf{e}\|^2 = SSR_2.$$

除非 $\mathbf{y}$ 在 $\mathbf{x}_2$ 上的投影为 0，以上不等式严格成立。沿着这个思路，我们可以给出另一种证明。因为

$$\|\mathbf{e}_{-K}\|^2 = \|(\mathbf{e}_{-K}-\mathbf{e})+\mathbf{e}\|^2 = \|\mathbf{e}_{-K}-\mathbf{e}\|^2 + \|\mathbf{e}\|^2 + 2(\mathbf{e}_{-K}-\mathbf{e})'\mathbf{e},$$

故只需证明 $(\mathbf{e}_{-K}-\mathbf{e})'\mathbf{e}=0$. 设使用 K 个自变量时回归时，零化子 $\mathbf{M}=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$，使用 $K-1$ 个自变量时对应的零化子为 $\mathbf{M}_{-K}=\mathbf{I}_n-\mathbf{X}_{-K}(\mathbf{X}_{-K}'\mathbf{X}_{-K})^{-1}\mathbf{X}_{-K}'$. 注意到

$$\mathbf{X}'\mathbf{e}=\mathbf{0}_{K\times 1} \implies \mathbf{X}_{K-1}'\mathbf{e}=\mathbf{0}_{(K-1)},$$

我们有

$$(\mathbf{e}_{-K}-\mathbf{e})'\mathbf{e} = (\mathbf{y}'\mathbf{M}_{-K}'-\mathbf{y}'\mathbf{M}')\mathbf{e} = \mathbf{y}'(\mathbf{M}_{-K}'\mathbf{e}-\mathbf{M}'\mathbf{e}).$$

一方面，

$$\mathbf{M}_{-K}'\mathbf{e} = (\mathbf{I}_n-\mathbf{X}_{-K}(\mathbf{X}_{-K}'\mathbf{X}_{-K})^{-1}\mathbf{X}_{-K}')\mathbf{e}=\mathbf{e},$$

另一方面，

$$\mathbf{M}'\mathbf{e}=\mathbf{M}\mathbf{M}\mathbf{y}=\mathbf{M}\mathbf{y}=\mathbf{e},$$

故可得

$$(\mathbf{e}_{-K}-\mathbf{e})'\mathbf{e}=\mathbf{y}'(\mathbf{e}-\mathbf{e})=0.$$

于是我们证明了

$$\|\mathbf{e}_{-K}\|^2 = \|\mathbf{e}_{-K}-\mathbf{e}\|^2+\|\mathbf{e}\|^2 \geq \|\mathbf{e}\|^2.$$

1.3 模型误设：加入无关变量

▷ **问题**. 已知 DGP（数据生成过程）

$$\mathbf{y} = \underset{(n\times K_1)}{\mathbf{X}_1}\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}.$$

若使用如下模型进行估计

$$\mathbf{y} = \underset{(n\times K_1)}{\mathbf{X}_1}\boldsymbol{\beta}_1+\underset{(n\times K_2)}{\mathbf{X}_2}\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}^*,$$

将得到对应的 OLS 估计量 $\mathbf{b}=(\mathbf{b}_1,\mathbf{b}_2)$. 在 Gauss-Markov 假定满足的条件下，证明 $\mathrm{E}[\mathbf{b}_1|\mathbf{X}_1,\mathbf{X}_2]=\boldsymbol{\beta}_1$，$\mathrm{Var}[\mathbf{b}_1|\mathbf{X}_1,\mathbf{X}_2]\geq\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$.

▷ **解答**. 设 $\mathbf{M}=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$，$\mathbf{M}_2=\mathbf{I}_n-\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$. 用 $\mathbf{X}_1'\mathbf{M}_2$ 左乘

$$\mathbf{y}=\mathbf{X}_1\mathbf{b}_1+\mathbf{X}_2\mathbf{b}_2+\mathbf{e}.$$

得到

$$\mathbf{X}_1'\mathbf{M}_2\mathbf{y}=\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\mathbf{b}_1+\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_2\mathbf{b}_2+\mathbf{X}_1'\mathbf{M}_2\mathbf{e}. \tag{1}$$

因为 $\mathbf{X}'\mathbf{e}=\mathbf{0} \implies \mathbf{X}_1'\mathbf{e}=\mathbf{X}_2'\mathbf{e}=\mathbf{0}$，类似 1.2 可以得到

$$\mathbf{M}_2\mathbf{e}=(\mathbf{I}_n-\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2')\mathbf{e}=\mathbf{e}-\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{e}=\mathbf{e},$$

从而有 $\mathbf{X}_1'\mathbf{M}_2\mathbf{e}=\mathbf{X}_1'\mathbf{e}=\mathbf{0}$. 又由 $\mathbf{M}_2\mathbf{X}_2=\mathbf{0}$ 知 $\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_2\mathbf{b}_2=\mathbf{0}$. 于是 (1) 式简化为

$$\mathbf{X}_1'\mathbf{M}_2\mathbf{y}=\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\mathbf{b}_1.$$

为证明 $\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ 可逆，我们可以使用反证法. 设矩阵 $\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ 的秩小于 K_1，则存在非零向量 $\mathbf{v}\in\mathbb{R}^{K_1}$，使得

$$\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\mathbf{v}=\mathbf{0},$$

从而有

$$\mathbf{v}'\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\mathbf{v}=(\mathbf{M}_2\mathbf{X}_1\mathbf{v})'(\mathbf{M}_2\mathbf{X}_1\mathbf{v})=0 \implies \mathbf{M}_2\mathbf{X}_1\mathbf{v}=\mathbf{0}.$$

由 $\mathbf{M}_2\mathbf{X}_2=\mathbf{0}$ 知 $\mathrm{Im}\mathbf{X}_2\subseteq\mathrm{Ker}\mathbf{M}_2$. 注意到 $\dim\mathrm{Im}\mathbf{X}_2=\dim\mathrm{Ker}\mathbf{M}_2=K_2$，因此有 $\mathrm{Im}\mathbf{X}_2=\mathrm{Ker}\mathbf{M}_2$. 故存在 $\mathbf{w}\in\mathbb{R}^{K_1}$，使得

$$\mathbf{X}_1\mathbf{v}=\mathbf{X}_2\mathbf{w}\in\mathrm{Im}\mathbf{X}_1\cap\mathrm{Im}\mathbf{X}_2.$$

由 $\mathbf{X}_1$ 列满秩知 $\mathbf{X}_1\mathbf{v}\neq\mathbf{0}$，故 $\dim(\mathrm{Im}\mathbf{X}_1\cap\mathrm{Im}\mathbf{X}_2)\geq 1$，从而

$$\dim(\mathrm{Im}\mathbf{X}_1\cup\mathrm{Im}\mathbf{X}_2)=\dim\mathrm{Im}\mathbf{X}_1+\dim\mathrm{Im}\mathbf{X}_2-\dim(\mathrm{Im}\mathbf{X}_1\cap\mathrm{Im}\mathbf{X}_2)\leq K_1+K_2-1,$$

这与 $(\mathbf{X}_1,\mathbf{X}_2)$ 列满秩矛盾！这样就证明了 $\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ 可逆. 由此解出

$$\mathbf{b}_1=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}.$$

于是我们可以将 $\mathbf{b}_1-\boldsymbol{\beta}_1$ 表示成 $\mathbf{X}_1$，$\mathbf{X}_2$ 和 $\boldsymbol{\varepsilon}$ 的函数

$$\begin{aligned}
\mathbf{b}_1-\boldsymbol{\beta}_1&=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}-\boldsymbol{\beta}_1\\
&=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2(\mathbf{X}\boldsymbol{\beta}_1+\boldsymbol{\varepsilon})-\boldsymbol{\beta}_1\\
&=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{X}\boldsymbol{\beta}_1+(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\boldsymbol{\varepsilon}-\boldsymbol{\beta}_1\\
&=\boldsymbol{\beta}_1+(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\boldsymbol{\varepsilon}-\boldsymbol{\beta}_1\\
&=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\boldsymbol{\varepsilon}\\
&=\mathbf{C}\boldsymbol{\varepsilon}\qquad(\text{令 }\mathbf{C}=(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2).
\end{aligned}$$

故

$$\mathrm{E}[\mathbf{b}_1-\boldsymbol{\beta}_1|\mathbf{X}_1,\mathbf{X}_2]=\mathrm{E}[\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}_1,\mathbf{X}_2]=\mathbf{C}\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}_1,\mathbf{X}_2]=0,$$

即 $\mathrm{E}[\mathbf{b}_1|\mathbf{X}_1,\mathbf{X}_2]=\boldsymbol{\beta}_1$. 下面证明一个引理：设 $\mathbf{A},\mathbf{B}$ 正定，

$$\mathbf{A}\geq\mathbf{B} \implies \mathbf{B}^{-1}\geq\mathbf{A}^{-1}.$$

事实上，这可由恒等式

$$\mathbf{B}^{-1}-\mathbf{A}^{-1}=\mathbf{A}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}=\mathbf{A}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}+\mathbf{A}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{B}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{A}^{-1}$$

直接得到. 因为

$$\begin{aligned}
\mathrm{Var}[\mathbf{b}_1|\mathbf{X}_1,\mathbf{X}_2] &= \mathrm{Var}[\mathbf{b}_1-\boldsymbol{\beta}_1|\mathbf{X}_1,\mathbf{X}_2] \\
&= \mathrm{Var}[\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}_1,\mathbf{X}_2] \\
&= \mathbf{C}\mathrm{Var}[\boldsymbol{\varepsilon}|\mathbf{X}_1,\mathbf{X}_2]\mathbf{C}' \\
&= \sigma^2(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{M}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2'\mathbf{X}_1)^{-1} \\
&= \sigma^2(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1},
\end{aligned}$$

要证明

$$\sigma^2(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1} \geq \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1},$$

只需证明

$$\mathbf{X}_1'\mathbf{X}_1 \geq \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1.$$

设投影矩阵 $\mathbf{P}_2 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2' = \mathbf{I}_n - \mathbf{M}$，则只需证

$$\mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1 = (\mathbf{P}_2\mathbf{X}_1)'(\mathbf{P}_2\mathbf{X}_1) \geq \mathbf{O}.$$

上式确实成立，因此我们证明了 $\mathrm{Var}[\mathbf{b}_1|\mathbf{X}_1,\mathbf{X}_2] \geq \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$.