

附录 1-4

附录 1：拓展贝叶斯信息准则（EBIC）

在处理逻辑回归模型与线性模型的估计任务时，本文分别通过最小化如下的拓展贝叶斯信息准则（EBIC）（Chen 和 Chen，2008）来确定参数 λ_n 和 m_n ，即

$$\text{EBIC}(\lambda_{ni}, m_n) = -2l_n(\gamma) + (\hat{p} + \hat{q})m_n \log(n) + 2(\hat{p} + \hat{q})m_n \phi \log(\hat{p} + \hat{q}), \quad i = 1, 2;$$

$$\text{EBIC}(\lambda_{n3}) = \log(RSS_{\lambda_{n3}}) + \hat{p}(\log(n) + \theta \log(\hat{p}))/n。$$

其中， $l_n(\gamma)$ 为给定数据 $(D_i, \mathbf{X}_i, \mathbf{Z}_i)$ ， $i = 1, 2, \dots, n$ 时，通过 γ 的组 Lasso 或者适应性组 Lasso 估计量计算的极大似然函数值。 $\hat{p} + \hat{q}$ 为 $\hat{f}_{nt} \neq 0$ 的个数，其中 $t = 1, 2, \dots, p + q$ ， $\phi > 1 - 1/(2\kappa)$ ， $0 < \kappa < 1/3$ 。 $RSS_{\lambda_{n3}}$ 是给定 λ_{n3} 的条件下，计算出的残差平方和， \hat{p} 是第二步得到参数估计量 $\hat{\beta}$ 的非零元素个数， $0 < \theta < 1$ 。特别地，当变量个数小于样本量，采取最小化贝叶斯信息准则（BIC）， $\text{EBIC}(\lambda_n) = \log(RSS_{\lambda_n}) + \hat{p} \log(n)/n$ 。当事件发生比对数满足线性简约模型的假设条件时，DS-ALIVE 自动选择 $m_n = 1$ ，此时退化为 Zhong 等（2022）利用逻辑回归建模的特例。在此情形下，蒙特卡洛模拟结果证实了提出的 DS-ALIVE 方法与双选下基于逻辑回归简约模型所得的处理效应估计量，在有限样本性质方面有相同的表现。

附录 2：引理 3.1 的证明

引理 S1. 设 x_1, x_2, \dots, x_n 是独立同分布且均值为 0 的随机变量三角阵，假设 $M_n = (\text{E}x_1^2)^{1/2} / (\text{E}|x_1|^3)^{1/3} > 0$ ，对于以较慢的速度 $b_n \rightarrow \infty$ ， $n^{1/6}M_n/b_n \rightarrow 1$ 。则对 $0 < \delta < 1$ ， $n^{1/6}M_n/b_n \rightarrow 1$ ，有

$$\left| \frac{\mathbb{P}(|S_n/V_n| \leq x)}{2(1 - \Phi(x))} - 1 \right| \ll \frac{A}{b_n^3} \rightarrow 0$$

其中， $S_n = \sum_{i=1}^n x_i$ ， $V_n = \sum_{i=1}^n x_i^2$ ， Φ 为标准正态分布的累计分布函数， A 是一个正值常数。

该引理来源于 Belloni 等（2014）以及 Fan 和 Zhong（2018）。

引理 3.1 的证明：注意到 $\hat{D}_i = \frac{\exp(\sum_{j \in A^*} \hat{f}_{nj}(x_{ij}) + \sum_{k \in A^*} \hat{f}_{nk}(z_{ik}))}{1 + \exp(\sum_{j \in A^*} \hat{f}_{nj}(x_{ij}) + \sum_{k \in A^*} \hat{f}_{nk}(z_{ik}))}$ ，进而有

$$\begin{aligned} & \hat{D}_i - D_i^* \\ &= \frac{\exp(\sum_{k \in A^*} \hat{f}_{nk}(z_{ik}) + \sum_{j \in A^*} \hat{f}_{nj}(x_{ij}))}{1 + \exp(\sum_{k \in A^*} \hat{f}_{nk}(z_{ik}) + \sum_{j \in A^*} \hat{f}_{nj}(x_{ij}))} - \frac{\exp(\sum_{k \in A} f_k(z_{ik}) + \sum_{j \in A} f_j(x_{ij}))}{1 + \exp(\sum_{k \in A} f_k(z_{ik}) + \sum_{j \in A} f_j(x_{ij}))} \\ & \ll \sum_{k \in A^*} [\hat{f}_{nk}(z_{ik}) - f_k(z_{ik})] + \sum_{j \in A^*} [\hat{f}_{nj}(x_{ij}) - f_j(x_{ij})] + \\ & \quad \left(\sum_{k \in A^*} [\hat{f}_{nk}(z_{ik}) - f_k(z_{ik})] + \sum_{j \in A^*} [\hat{f}_{nj}(x_{ij}) - f_j(x_{ij})] \right)^2 (1 + o(1)) \\ & \approx \sum_{j, k \in A^*} [\hat{f}_{nk}(z_{ik}) + \hat{f}_{nj}(x_{ij})] - \sum_{j, k \in A} [f_k(z_{ik}) + f_j(x_{ij})] \end{aligned}$$

进一步地，由引理 3.1 关于参数估计相合性结论，有

$$\begin{aligned}
 & \left| \sum_{j,k \in A} \left[\hat{f}_{nk}(z_{ik}) + \hat{f}_{nj}(x_{ij}) \right] - \sum_{j,k \in A} \left[f_k(z_{ik}) + f_j(x_{ij}) \right] \right| \\
 &= \left| \sum_{i \in A} \sum_{\sigma=1}^{m_n} (\hat{\gamma}_{i\sigma} \psi_{\sigma}(T_{ii})) - \sum_{i \in A} \sum_{\sigma=1}^{m_n} (\gamma_{i\sigma} \psi_{\sigma}(T_{ii})) \right| \\
 &= \left| \sum_{i \in A \cup A} \sum_{\sigma=1}^{m_n} (\hat{\gamma}_{i\sigma} - \gamma_{i\sigma}) \psi_{\sigma}(T_{ii}) \right| \\
 &\leq \sqrt{2s(1+o(1))m_n \sum_{i \in A \cup A} \sum_{\sigma=1}^{m_n} (\hat{\gamma}_{i\sigma} - \gamma_{i\sigma})^2} \\
 &= \sqrt{2s(1+o(1))m_n n^{-\frac{2d-1}{2d+1}} \log n} \\
 &= O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right)
 \end{aligned}$$

其中，利用到中心化 B 样条 $|\psi_{\sigma}(T_{ii})|, 2$ 的性质。

附录 3：假设 (A) – (D)

令 $\mathbf{U} = (\mathbf{u}_1^T, \mathbf{u}_2^T, \mathbf{L}, \mathbf{u}_n^T)^T$ 。对于 $t=1, 2, \mathbf{L}, (p+q)m_n$ 。 \mathbf{u}_t 是 \mathbf{U} 的第 t 列。下面给出引理 3.1 所需的假设 (i) – (iv)：

(A) 中心化 B 样条基函数满足，基矩阵 \mathbf{u}_t 是有界的，即存在 C_{u_t} 使得 $\|\mathbf{u}_t\|_2 \leq \sqrt{n}C_{u_t}$ ，对于 $t=1, 2, \mathbf{L}, (p+q)m_n$ 均成立。

(B) 存在一个给定序列 Q_n ，定义 $C = \left\{ \boldsymbol{\delta} \in R^{pm_n} : \|\boldsymbol{\delta}\|_2 \neq 0, \|\boldsymbol{\delta}\|_2 \leq N_n, \left| \left\{ j : \|\boldsymbol{\delta}_j\|_0 \neq 0 \right\} \right| = O(s_n) \right\}$ ，对 $\boldsymbol{\delta} \in C$ ， $\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T, \mathbf{L}, \boldsymbol{\delta}_p^T)$ ，存在 $c_4 \in (0, 0.5)$ 和 c_5, c_6 使得 $c_5 c_4^{2s_n} m_n^{-1} \boldsymbol{\delta}^T \mathbf{U}^T \mathbf{U} \boldsymbol{\delta} / (n \|\boldsymbol{\delta}\|_2) \geq c_6 m_n^{-1}$ 。

(C) 令 $d_i^0 = \sum_{j=1}^p \sum_{\sigma=1}^{m_n} \gamma_{j\sigma}^0 \psi_{\sigma}(x_{ij}) + \sum_{k=1}^q \sum_{\sigma=1}^{m_n} \gamma_{k\sigma}^0 \psi_{\sigma}(z_{ik})$ ， $\mu_{D_i} = \exp(d_i^0) / (1 + \exp(d_i^0))$ 和 $\varepsilon_i = D_i - \mu_{D_i}$ 。假设误差 ε_i 服从次高斯分布，即，对于 $c_7 = \log 2$ ，使得对于 $i=1, 2, \mathbf{L}, n$ ，如下不等式成立 $P(|\varepsilon_i| \geq t) \leq 2 \exp(-c_7 t^2)$ ， $\forall t > 0$ 。

(D) 当 $n \rightarrow \infty$ ，序列 $c_{f,n} \rightarrow 0$ ，对所有 $t \in A$ ，真实非零函数 f_t 满足 $\min_{t \in A} \|f_t\|_2 \geq c_{f,n}$ 。

附录 4：定理 3.1 的证明

定理 3.1 的证明：由式 (10) 有，

$$\begin{aligned}
 \hat{\alpha}_{DA} &= (\hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}^T \mathbf{M}_I \mathbf{Y} \\
 &= (\hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}^T \mathbf{M}_I (\alpha_0 \mathbf{D}^* + \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{v}) \\
 &= (\hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}^T \mathbf{M}_I (\alpha_0 (\hat{\mathbf{D}} + \mathbf{D}^* - \hat{\mathbf{D}}) + \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{v}) \\
 &= \alpha_0 + (\hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}^T \mathbf{M}_I (\alpha_0 (\mathbf{D}^* - \hat{\mathbf{D}}) + \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{v})
 \end{aligned}$$

整理得

$$\sqrt{n}(\hat{\alpha}_{DA} - \alpha_0) = \left(\frac{\hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}}}{n} \right)^{-1} \frac{\hat{\mathbf{D}}^T \mathbf{M}_I (\alpha_0 (\mathbf{D}^* - \hat{\mathbf{D}}) + \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{v})}{\sqrt{n}} @T_1^{-1} \cdot T_2$$

为了符号简捷，首先定义 $\boldsymbol{\varsigma} = \hat{\mathbf{D}} - \mathbf{D}^*$ 。对向量 $\mathbf{d} \in \mathbb{R}^n$ ，令 $\hat{\boldsymbol{\beta}}_{\mathbf{d}}(A) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{d} - \mathbf{X} \boldsymbol{\beta}\|^2$ ，其中，对于 $j \in A$ ，有 $\beta_j = 0$ ，记 $\hat{k}_{\mathbf{d}} = \|\hat{\boldsymbol{\beta}}_{\mathbf{d}}(A)\|_0$ 。

首先处理 T_1 项，

$$\begin{aligned} T_1 &= \hat{\mathbf{D}}^T \mathbf{M}_I \hat{\mathbf{D}} / n = (\hat{\mathbf{D}} - \mathbf{D}^* + \mathbf{D}^*)^T (\mathbf{M}_I - \mathbf{M}_I + \mathbf{M}_I) (\hat{\mathbf{D}} - \mathbf{D}^* + \mathbf{D}^*) / n \\ &= \frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*}{n} + \frac{\boldsymbol{\varsigma}^T \mathbf{M}_I \boldsymbol{\varsigma}}{n} + \frac{2\boldsymbol{\varsigma}^T \mathbf{M}_I \mathbf{D}^*}{n} + \frac{\mathbf{D}^{*T} (\mathbf{M}_I - \mathbf{M}_I) \mathbf{D}^*}{n} \\ &\quad @ \frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*}{n} + T_{11} + T_{12} + T_{13} \end{aligned}$$

首先对 T_{11} 项分解如下

$$T_{11} = \frac{\boldsymbol{\varsigma}^T \mathbf{M}_I \boldsymbol{\varsigma}}{n} = \frac{\boldsymbol{\varsigma}^T \boldsymbol{\varsigma}}{n} - \frac{\boldsymbol{\varsigma}^T \mathbf{P}_I \boldsymbol{\varsigma}}{n}$$

其中，第一项根据引理 3.1，有 $\boldsymbol{\varsigma}^T \boldsymbol{\varsigma} / n = O_p(n^{\frac{4(d-1)}{2d+1}} \log^2 n)$ 。由引理 S1，首先控制 $\|\mathbf{X}^T \mathbf{v} / \sqrt{n}\|_\infty$ 界。由条件三和条件四，可得

$$\|\mathbf{X}^T \mathbf{v} / \sqrt{n}\|_\infty = O_p(\sqrt{\log(p \vee n)}) \quad (\text{A.1})$$

因此有

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}_\varsigma(I^\wedge)\|_1 &\asymp \sqrt{\hat{k}_\varsigma} \|\hat{\boldsymbol{\beta}}_\varsigma(I^\wedge)\|_2 \\ &= \sqrt{\hat{k}_\varsigma} \|(\mathbf{X}_I^T \mathbf{X}_I / n)^{-1} \mathbf{X}_I^T \boldsymbol{\varsigma} / n\|_2 \\ &\asymp \sqrt{\hat{k}_\varsigma} \phi_{\min}^{-1}(\sqrt{\hat{k}_\varsigma}) \sqrt{\hat{k}_\varsigma} \|\mathbf{X}_I^T \boldsymbol{\varsigma}\|_\infty / \sqrt{n} \\ &= O_p\left(\sqrt{\frac{\log(p \vee n)}{n}}\right) \end{aligned} \quad (\text{A.2})$$

推导出

$$\begin{aligned} \left| \frac{\boldsymbol{\varsigma}^T \mathbf{P}_I \boldsymbol{\varsigma}}{n} \right| &= \left| \frac{\boldsymbol{\varsigma}^T \mathbf{X}_I (\mathbf{X}_I^T \mathbf{X}_I)^{-1} \mathbf{X}_I^T \boldsymbol{\varsigma}}{n} \right| = \left| \frac{\boldsymbol{\varsigma}^T \mathbf{X}_I \hat{\boldsymbol{\beta}}_\varsigma(I^\wedge)}{n} \right| \asymp \|\hat{\boldsymbol{\beta}}_\varsigma(I^\wedge)\|_1 \|\mathbf{X}_I^T \boldsymbol{\varsigma} / n\|_\infty \\ &= O_p\left(s \sqrt{\frac{\log(p \vee n)}{n}}\right) \cdot O_p\left(\sqrt{\frac{\log(p \vee n)}{n}}\right) = O_p\left(\frac{\log(p \vee n)}{n}\right) \end{aligned}$$

因此，

$$T_{11} = O_p\left(n^{\frac{4(d-1)}{2d+1}} \log^2 n\right) + O_p\left(\frac{\log(p \vee n)}{n}\right)$$

接下来处理 T_{12} 项，将 T_{12} 项写成

$$T_{12} = \frac{2\boldsymbol{\varsigma}^T \mathbf{M}_I \mathbf{D}^*}{n} = \frac{2\boldsymbol{\varsigma}^T \mathbf{D}^*}{n} - \frac{2\boldsymbol{\varsigma}^T \mathbf{P}_I \mathbf{D}^*}{n}$$

由 $D_i^* \in [0, 1]$ 有界性和引理 3.1，有

$$\left| \frac{2\boldsymbol{\varsigma}^T \mathbf{D}^*}{n} \right| \asymp \left| \frac{2\sum_{i=1}^n (\hat{D}_i - D_i^*)}{n} \right| = O_p\left(n^{\frac{2d-2}{2d+1}} \log n\right)$$

类似于 T_{11} 项的处理，由式 (A.2) 和条件三，有

$$\left| \frac{\boldsymbol{\varsigma}^T \mathbf{P}_I \mathbf{D}^*}{n} \right| \asymp \|\hat{\boldsymbol{\beta}}_\varsigma(I^\wedge)\|_1 \cdot \|\mathbf{X}_I^T \mathbf{D}^* / n\|_\infty = O_p\left(s \sqrt{\frac{\log(p \vee n)}{n}}\right) \cdot O_p(1) = O_p\left(\sqrt{\frac{\log(p \vee n)}{n}}\right)$$

因此，有

$$|T_{12}| = O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) + O_p \left(\sqrt{\frac{\log(p \vee n)}{n}} \right)$$

为分析 T_{13} 项，首先聚焦于 $\mathbf{M}_{I^*} - \mathbf{M}_I$ 。假定 $\|\mathbf{M}_{I^*} - \mathbf{M}_I\|_\infty$ 存在， $\mathbf{M}_{I^*} - \mathbf{M}_I$ 最大值出现在其第 (k, l) 个元素上，标记为 m_Δ 。由 $\mathbf{M}_{I^*} - \mathbf{M}_I = \mathbf{P}_{I^*} - \mathbf{P}_I$ ，有

$$m_\Delta = \sum_{j \in I} \sum_{i \in I} x_{ki} \mathbf{A}_{ij} x_{lj} - \sum_{j \in I^*} \sum_{i \in I^*} x_{ki} \mathbf{A}_{ij}^0 x_{lj}, \text{ 其中, } \mathbf{A} = (\mathbf{X}_I^T \mathbf{X}_I)^{-1} \text{ 和 } \mathbf{A}^0 = (\mathbf{X}_{I^*}^T \mathbf{X}_{I^*})^{-1}。对于任$$

意小的 $\delta > 0$ ，有

$$\begin{aligned} P(|m_\Delta| > \delta) &= P\left(\left|\sum_{j \in I} \sum_{i \in I} x_{ki} \mathbf{A}_{ij} x_{lj} - \sum_{j \in I^*} \sum_{i \in I^*} x_{ki} \mathbf{A}_{ij}^0 x_{lj}\right| > \delta\right) \\ &,, P\left(\left|\sum_{i, j \in I \cap I^c} (x_{ki} \mathbf{A}_{ij} x_{lj} - x_{ki} \mathbf{A}_{ij}^0 x_{lj})\right| > \delta/3\right) \\ &+ P\left(\left|\sum_{i, j \in I \cap I^c} (x_{ki} \mathbf{A}_{ij} x_{lj} - x_{ki} \mathbf{A}_{ij}^0 x_{lj})\right| > \delta/3\right) \\ &+ P\left(\left|\sum_{i, j \in I^c \cap I} (x_{ki} \mathbf{A}_{ij} x_{lj} - x_{ki} \mathbf{A}_{ij}^0 x_{lj})\right| > \delta/3\right) \\ &\rightarrow 0 \end{aligned} \quad (\text{A.3})$$

其中，鉴于 $\mathbf{A}_{ij} = \mathbf{A}_{ij}^0$ ， $i, j \in I \cap I^*$ ，第一项趋于 0。由 $P(\hat{\mathbf{I}} = \mathbf{I}) \rightarrow 1$ ，可得第二项及第三项亦趋近于 0。所以， $m_\Delta = o_p(1)$ 。由条件三与式 (A.3)，得

$$T_{13} = \frac{\mathbf{D}^{*T}(\mathbf{M}_{I^*} - \mathbf{M}_I)\mathbf{D}^*}{n},, \frac{\mathbf{D}^{*T}\mathbf{D}^*}{n} o_p(1) = O_p(1) \cdot o_p(1) = o_p(1)$$

$$\text{进而有, } T_1 = \frac{\mathbf{D}^{*T}\mathbf{M}_I\mathbf{D}^*}{n} + o_p(1)。$$

接下来处理 T_2 项，

$$\begin{aligned} T_2 &= \frac{\hat{\mathbf{D}}^T \mathbf{M}_{I^*} [(\mathbf{D}^* - \hat{\mathbf{D}})\alpha_0 + \mathbf{X}\beta_0 + \mathbf{v}]}{\sqrt{n}} \\ &= \frac{(\hat{\mathbf{D}} - \mathbf{D}^* + \mathbf{D}^*)^T (\mathbf{M}_{I^*} - \mathbf{M}_I + \mathbf{M}_I) [(\mathbf{D}^* - \hat{\mathbf{D}})\alpha_0 + \mathbf{X}\beta_0 + \mathbf{v}]}{\sqrt{n}} \\ &= \frac{(\boldsymbol{\varsigma} + \mathbf{D}^*)^T (\mathbf{M}_{I^*} - \mathbf{M}_I + \mathbf{M}_I) [-\boldsymbol{\varsigma}\alpha_0 + \mathbf{X}\beta_0 + \mathbf{v}]}{\sqrt{n}} \\ &= \frac{\mathbf{D}^{*T}\mathbf{M}_I\mathbf{v}}{\sqrt{n}} - \frac{\alpha_0 \boldsymbol{\varsigma}^T \mathbf{M}_{I^*} \boldsymbol{\varsigma}}{\sqrt{n}} + \frac{\boldsymbol{\varsigma}^T \mathbf{M}_{I^*} \mathbf{X}\beta_0}{\sqrt{n}} + \frac{\boldsymbol{\varsigma}^T \mathbf{M}_{I^*} \mathbf{v}}{\sqrt{n}} - \frac{\alpha_0 \mathbf{D}^{*T} \mathbf{M}_{I^*} \boldsymbol{\varsigma}}{\sqrt{n}} + \frac{\mathbf{D}^{*T} \mathbf{M}_{I^*} \mathbf{X}\beta_0}{\sqrt{n}} + \frac{\mathbf{D}^{*T} (\mathbf{M}_{I^*} - \mathbf{M}_I) \mathbf{v}}{\sqrt{n}} \\ &\textcircled{=} \frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{v}}{\sqrt{n}} + T_{21} + T_{22} + T_{23} + T_{24} + T_{25} + T_{26} \end{aligned}$$

$$\text{由 } T_{11} \text{ 可知, } |T_{21}| = O_p \left(n^{-\frac{2(d-1)}{2d+1}} \log^2 n \right) + O_p \left(\frac{\log(p \vee n)}{\sqrt{n}} \right)。$$

由 $\hat{I}_2 \subseteq \hat{I}^*$ 和引理 3.1，有

$$\begin{aligned}
 & \left\| \mathbf{M}_{I^*} \mathbf{D}^* / \sqrt{n} \right\|, \left\| \mathbf{M}_{I_2} \mathbf{D}^* / \sqrt{n} \right\| \\
 &= \left\| (\mathbf{D}^* - \mathbf{P}_{I_2} \mathbf{D}^*) / \sqrt{n} \right\| \\
 &, \left\| (\mathbf{D}^* - \hat{\mathbf{D}}) / \sqrt{n} \right\| \\
 &= O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right)
 \end{aligned} \tag{A.4}$$

由 post-Lasso 结论，可得

$$\begin{aligned}
 & \left\| \mathbf{M}_{I^*} \mathbf{X} \boldsymbol{\beta}_0 / \sqrt{n} \right\|_2 - \left\| \mathbf{M}_{I^*} \alpha_0 \mathbf{D}^* / \sqrt{n} \right\|_2, \left\| \mathbf{M}_{I^*} (\alpha_0 \mathbf{D}^* + \mathbf{X} \boldsymbol{\beta}_0) / \sqrt{n} \right\|_2 \\
 &, \left\| \mathbf{M}_{I_1} (\alpha_0 \mathbf{D}^* + \mathbf{X} \boldsymbol{\beta}_0) / \sqrt{n} \right\|_2 \\
 &= \left\| (\alpha_0 \mathbf{D}^* + \mathbf{X} \boldsymbol{\beta}_0 - \mathbf{P}_{I_1} (\alpha_0 \mathbf{D}^* + \mathbf{X} \boldsymbol{\beta}_0)) / \sqrt{n} \right\|_2 \\
 &, \left\| (\mathbf{X} \hat{\boldsymbol{\beta}}_y(I_1) - \alpha_0 \mathbf{D}^* - \mathbf{X} \boldsymbol{\beta}_0) / \sqrt{n} \right\|_2 \\
 &= O_p \left(\sqrt{s \log(p \vee n) / n} \right)
 \end{aligned} \tag{A.5}$$

其中，依据 post-Lasso 的收敛速度（Belloni 和 Chernozhukov, 2013），可确认上述推导中的最后一个等号得以成立。

结合式（A.4）和式（A.5），有

$$\left\| \mathbf{M}_{I^*} \mathbf{X} \boldsymbol{\beta}_0 / \sqrt{n} \right\| = O_p \left(\sqrt{\log(p \vee n) / n} \right) + O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right)$$

因此，由柯西施瓦茨不等式，有

$$\begin{aligned}
 |T_{22}| &= \left| \frac{\boldsymbol{\varsigma}^T \mathbf{M}_{I^*} \mathbf{X} \boldsymbol{\beta}_0}{\sqrt{n}} \right|, \sqrt{n} \left\| \frac{\hat{\mathbf{D}} - \mathbf{D}^*}{\sqrt{n}} \right\|_2 \cdot \left\| \frac{\mathbf{M}_{I^*} \mathbf{X} \boldsymbol{\beta}_0}{\sqrt{n}} \right\|_2 \\
 &= \sqrt{n} O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \cdot \left[O_p \left(\sqrt{s \log(p \vee n) / n} \right) + O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \right] \\
 &= O_p \left(n^{-\frac{2d-2}{2d+1}} \log^2(p \vee n) \right) + O_p \left(n^{-\frac{3(2d+3)}{2(2d+1)}} \log^2 n \right)
 \end{aligned}$$

接下来处理 T_{23} 项，将其拆分为

$$T_{23} = \frac{\boldsymbol{\varsigma}^T \mathbf{M}_{I^*} \mathbf{v}}{\sqrt{n}} = \frac{\boldsymbol{\varsigma}^T \mathbf{v}}{\sqrt{n}} - \frac{\boldsymbol{\varsigma}^T \mathbf{P}_{I^*} \mathbf{v}}{\sqrt{n}}$$

其中，第一项由引理 3.1，有

$$\frac{\boldsymbol{\varsigma}^T \mathbf{v}}{\sqrt{n}} = \left\| \hat{\mathbf{D}}_i - \mathbf{D}_i^* \right\|_\infty \frac{\sum_{i=1}^n v_i}{\sqrt{n}} = O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \cdot O_p(1) = O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right)$$

另外，由中心极限定理，有 $\frac{\sum_{i=1}^n v_i}{\sqrt{n}} = O_p(1)$ 成立。类似于 T_{11} 处理方法，由式（A.1）和式（A.2）

有

$$\begin{aligned}
 & \left| \frac{\boldsymbol{\zeta}^T \mathbf{P}_I \mathbf{v}}{\sqrt{n}} \right|, \left\| \hat{\boldsymbol{\beta}}_{\boldsymbol{\zeta}}(I^{\wedge}) \right\|_1 \cdot \left\| \frac{\mathbf{X}_I^T \mathbf{v}}{\sqrt{n}} \right\|_{\infty} \\
 &= O_p \left(s \sqrt{\frac{\log(p \vee n)}{n}} \right) \cdot O_p \left(\sqrt{\log(p \vee n)} \right) \\
 &= O_p \left(\frac{s \log(p \vee n)}{\sqrt{n}} \right)
 \end{aligned}$$

因此，

$$T_{23} = O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) + O_p \left(\frac{\log(p \vee n)}{\sqrt{n}} \right)$$

接下来处理 T_{24} 项，由柯西施瓦茨不等式、式 (A.4) 和引理 3.1，有

$$\begin{aligned}
 |T_{24}| &= \left| \frac{\mathbf{D}^{*T} \mathbf{M}_I \boldsymbol{\zeta}}{\sqrt{n}} \right| \\
 &\leq \sqrt{n} \left\| \frac{\hat{\mathbf{D}} - \mathbf{D}^*}{\sqrt{n}} \right\|_2 \cdot \left\| \frac{\mathbf{M}_I \mathbf{D}^*}{\sqrt{n}} \right\|_2 \\
 &= \sqrt{n} O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \\
 &= O_p \left(n^{\frac{3(2d+3)}{2(2d+1)}} \log^2 n \right)
 \end{aligned}$$

结合柯西不等式，式 (A.4) 和式 (A.5)，有

$$\begin{aligned}
 |T_{25}| &= \left| \frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{X} \boldsymbol{\beta}_0}{\sqrt{n}} \right| \\
 &\leq \sqrt{n} \left\| \frac{\mathbf{M}_I \mathbf{X} \boldsymbol{\beta}_0}{\sqrt{n}} \right\| \left\| \frac{\mathbf{M}_I \mathbf{D}^*}{\sqrt{n}} \right\| \\
 &= \sqrt{n} \left[O_p \left(\sqrt{\frac{s \log(p \vee n)}{n}} \right) + O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \right] O_p \left(n^{-\frac{2d-2}{2d+1}} \log n \right) \\
 &= O_p \left(n^{-\frac{2d-2}{2d+1}} \log^2(p \vee n) \right) + O_p \left(n^{\frac{3(2d+3)}{2(2d+1)}} \log^2 n \right)
 \end{aligned}$$

类似于 T_{13} 的处理方法，有

$$|T_{26}| = \left| \frac{\mathbf{D}^{*T} (\mathbf{M}_I - \mathbf{M}_I) \mathbf{v}}{\sqrt{n}} \right| \leq \left| \frac{\mathbf{D}^{*T} \mathbf{v}}{\sqrt{n}} \right| o_p(1) = \left| \frac{\sum_{i=1}^n v_i}{\sqrt{n}} \right| o_p(1) = O_p(1) o_p(1) = o_p(1)$$

因此，有

$$T_2 = \frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{v}}{\sqrt{n}} + o_p(1)$$

进而，

$$\sqrt{n}(\hat{\alpha}_{\text{DA}} - \alpha_0) = \left(\frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*}{n} + o_p(1) \right)^{-1} \left(\frac{\mathbf{D}^{*T} \mathbf{M}_I \mathbf{v}}{\sqrt{n}} + o_p(1) \right)$$

由弱大数定理得 $\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^* / n = E(\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*) + o_p(1)$ ，由于 $\sum_{i=1}^n D_i^* \mathbf{M}_I (i,j) v_j / n$ 是独立同分布，其均值为 0，方差为 $\sigma_n^2 = (E(\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*))^{-1} E(\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^* v_j^2) (E(\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*))^{-1}$ ，由中心极限定理和 Slutsky 定理得 $\sigma_n^{-1} \sqrt{n}(\hat{\alpha}_{DA} - \alpha_0) \xrightarrow{d} N(0,1)$ 。如果条件同方差满足，即 $Var(v_i | \mathbf{X}_i) = \sigma_v^2$ 成立时，有 $\sigma_n^2 = (E(\mathbf{D}^{*T} \mathbf{M}_I \mathbf{D}^*))^{-1} \sigma_v^2$ 。

参考文献

- [1] Belloni A, Chernozhukov V, Hansen C. Inference on Treatment Effects after Selection among High-dimensional Controls [J]. The Review of Economic Studies, 2014, 81(2): 608-650.
- [2] Belloni A, Chernozhukov V. Least Squares after Model Selection in High-dimensional Sparse Models[J]. Bernoulli, 2013, 19: 521-547.
- [3] Chen J, Chen Z. Extended Bayesian Information Criterion for Model Selection with Large Model Space[J]. Biometrika, 2008, 94: 759-771.
- [4] Fan Q, Zhong W. Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective [J]. Journal of Business and Economic Statistics, 2018, 36:388-399.
- [5] Zhong W, Zhou W, Fan Q, et al. Dummy Endogenous Treatment Effect Estimation Using High Dimensional Instrumental Variable[J]. The Canadian Journal of Statistics, 2022, 50(3): 795-819.

附表 1 模拟情形一中各种估计量的偏误和均方误

n	方法	模型 1		模型 2	
		偏误	均方误	偏误	均方误
n=100	OLS	0.102	0.022	0.105	0.021
	2SLS	-0.037	0.066	0.025	0.076
	LIVE	0.036	0.017	0.114	0.038
	DS-LIVE	-0.011	0.015	0.106	0.036
	NAIVE	-0.031	0.017	0.039	0.015
	ALIVE	0.029	0.015	0.024	0.015
	DS-ALIVE	-0.010	0.014	-0.012	0.013
n=300	OLS	0.110	0.020	0.107	0.018
	2SLS	-0.033	0.042	0.022	0.055
	LIVE	0.039	0.013	0.113	0.029
	DS-LIVE	-0.007	0.010	0.111	0.029
	NAIVE	-0.032	0.012	-0.024	0.011
	ALIVE	0.026	0.009	0.019	0.009
	DS-ALIVE	-0.007	0.008	-0.001	0.009

附表 2 模拟情形二中各种估计量的偏误和均方误

n	方法	模型 1		模型 2	
		偏误	均方误	偏误	均方误
n=100	OLS	0.122	0.026	0.122	0.024
	2SLS-L	-0.042	0.071	0.034	0.054
	LIVE	0.031	0.016	0.122	0.039
	DS-LIVE	0.005	0.014	0.100	0.036
	NAIVE	0.039	0.015	0.040	0.014
	ALIVE	0.033	0.013	0.021	0.013
	DS-ALIVE	0.005	0.012	0.006	0.013
n=300	OLS	0.121	0.022	0.123	0.022
	2SLS-L	-0.024	0.042	-0.022	0.038
	LIVE	0.024	0.010	0.118	0.029
	DS-LIVE	0.005	0.010	0.103	0.027
	NAIVE	0.026	0.011	0.034	0.010
	ALIVE	0.025	0.010	0.023	0.011
	DS-ALIVE	0.004	0.009	-0.003	0.009

附表 3 简约模型 1 下各种估计量的偏误和均方误

方法	p+q=200, n=100		p+q=300, n=100		p+q=400, n=100		p+q=500, n=100		p+q=600, n=100	
	偏误	均方误	偏误	均方误	偏误	均方误	偏误	均方误	偏误	均方误
OLS	0.114	0.024	0.115	0.025	0.117	0.025	0.119	0.025	0.111	0.024
2SLS-L	-0.033	0.069	-0.032	0.070	-0.031	0.071	-0.030	0.063	-0.031	0.067
LIVE	0.021	0.015	0.023	0.014	0.025	0.015	0.023	0.015	0.023	0.015
DS-LIVE	0.007	0.012	0.007	0.013	0.007	0.014	0.007	0.013	0.007	0.014
NAIVE	0.028	0.016	0.028	0.016	0.031	0.016	0.031	0.016	0.041	0.015
ALIVE	0.019	0.016	0.020	0.014	0.019	0.014	0.022	0.014	0.021	0.014
DS-ALIVE	0.007	0.010	0.007	0.010	0.007	0.010	0.007	0.010	0.008	0.010

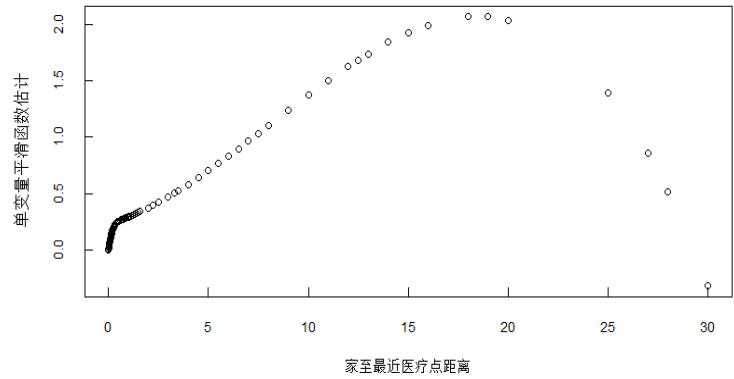
附表 4 简约模型 2 下各种估计量的偏误和均方误

方法	p+q=200, n=100		p+q=300, n=100		p+q=400, n=100		p+q=500, n=100		p+q=600, n=100	
	偏误	均方误	偏误	均方误	偏误	均方误	偏误	均方误	偏误	均方误
OLS	0.115	0.022	0.111	0.021	0.116	0.023	0.115	0.024	0.112	0.023
2SLS-L	0.036	0.049	0.041	0.071	0.041	0.060	0.042	0.068	0.046	0.061
LIVE	0.111	0.035	0.112	0.036	0.117	0.037	0.110	0.034	0.112	0.036
DS-LIVE	0.104	0.036	0.112	0.037	0.104	0.036	0.106	0.037	0.112	0.035
NAIVE	0.039	0.013	0.038	0.013	0.057	0.015	0.055	0.014	0.060	0.015
ALIVE	0.020	0.014	0.024	0.013	0.025	0.015	0.029	0.015	0.032	0.013
DS-ALIVE	0.007	0.012	0.008	0.012	0.007	0.011	0.007	0.012	0.007	0.012

附表 5 描述统计量

变量	均值	中位数	最小值	最大值
收入	7397	10	1	1029999
字词测试得分	2.370	4	0	10
体重	122.700	120	108	260
性别	0.498	0	0	1
岁数	45.780	46	18	96
健康状况	0.686	1	0	1
是否居于城市	0.438	1	0	1
文化程度	2.069	2	1	7
是否中部	0.291	0	0	1
是否西部	0.283	0	0	1
是否慢性病	0.163	0	0	1
每周锻炼身体次数	2.195	0	0	50
是否有高污染企业	0.238	0	0	1
是否有水冲设施	0.368	0	0	1
至最近医疗设施距离	1.653	0.500	0.005	30
是否回收垃圾	0.126	0	0	1
是否用自来水	0.739	1	0	1
是否吸烟	0.275	0	0	1
是否用清洁能源	0.674	1	0	1
是否喝酒	0.143	0	0	1

注：数据来源于中国家庭追踪调查（CFPS）



附图 1 家到最近医疗点距离与单变量函数的估计之间的关系图

附表 6 身体健康状态对收入或字词测试得分的影响

方法	收入			字词测试得分		
	系数	标准误	<i>t</i> 值	系数	标准误	<i>t</i> 值
OLS	0.031	0.038	0.800	-0.085	0.067	-1.262
2SLS-L	-0.381**	0.161	-2.360	-3.738*	2.262	-1.652
LIVE	-0.384**	0.163	-2.363	-5.269**	2.256	-2.335
DS-LIVE	-0.395**	0.163	-2.423	-5.370**	2.258	-2.377

NAIVE	-0.413**	0.162	-2.545	-3.865*	2.308	-1.674
ALIVE	-0.414**	0.163	-2.548	-7.433***	2.006	-3.704
DS-ALIVE	-0.423***	0.162	-2.606	-8.190***	2.011	-4.071

注：***, **, *分别表示 0.01, 0.05 和 0.1 的显著性水平。