应用概率统计　第 37 卷
第 6 期　2021 年 12 月

Chinese Journal of Applied Probability and Statistics
Dec., 2021, Vol. 37, No. 6, pp. 551-568
doi: 10.3969/j.issn.1001-4268.2021.06.001

超高维部分线性模型的 PRAR 变量选择 *

杨　鑫

(北京工业大学理学部统计与数据科学学院, 北京, 100124)

李冰月

(北京计算技术及应用研究所, 北京, 100854)

田　萍⋆

(许昌学院数理学院, 许昌, 461000)

摘　要： 本文考虑超高维部分线性模型, 其中参数向量维数是样本量的指数阶. 基于 profile 最小二乘方法和保留正则化方法, 本文提出了新的变量选择方法用来解决超高维部分线性模型的变量选择问题. 在一定的正则条件下, 证明了所得估计量的符号相合性. 通过数值模拟和实例分析, 将该方法与 Lasso、SIS-Lasso、自适应 Lasso 方法进行对比, 发现所提方法在恢复线性部分参数向量符号方面明显优于其它方法.

关键词： 部分线性模型; 变量选择; 高维数据; Lasso; 符号相合性; 保留正则化

中图分类号： O212.7

英文引用格式： YANG X, LI B Y, TIAN P. Profile regularization after retention variable selection for ultrahigh dimensional partially linear models [J]. Chinese J Appl Probab Statist, 2021, 37(6): 551–568. (in Chinese)

§1. 引　　言

部分线性模型是 Engle 等 [1] 研究电力需求与温度变化间的关系时提出的, 假定响应变量与部分协变量呈线性关系, 同时与另一部分协变量存在未知的非线性关系, 从而部分线性模型既保留了线性模型易于解释的优点, 又含有非参数回归的稳健性特点, 在应用上具有更大的灵活性.

随着信息技术和计算机的高速发展, 积累的数据日益增多, 涉及范围也越来越广, 高维数据逐渐被人们所认识, 并成为热门的一个研究方向. 高维线性模型的变量选择取得了丰富的研究成果, 统计学家提出了一些基于惩罚函数的估计方法, 如 Lasso [2]、SCAD [3] 和自适应 Lasso [4] 等. 在超高维情况下, Fan 和 Lv [5] 首次提出了确定性独立筛选方法, 根据边际相关性进行变量选择. 在此基础上, Weng [6] 提出保留正则化方法, 根据边际相关性选出部分重要变量, 然后对剩余变量施加 Lasso 惩罚, 达到变量选择的目的. Wang [7] 研究了超

*国家自然科学基金面上项目 (批准号: 11871001、11971001) 和北京市自然科学基金项目 (批准号: 1182003) 资助.

⋆通讯作者, E-mail: tianping@xcu.edu.cn.

本文 2019 年 12 月 17 日收到, 2020 年 5 月 10 日收到修改稿.

高维数据下向前回归方法, 证明该方法具有模型选择相合性. Cheng 等 [8] 进一步提出贪婪的向前回归方法, 并证明识别全部重要变量所需的迭代次数更小.

然而, 对部分线性模型的变量选择进行研究的文献相对较少, 李锋等 [9] 研究了自适应 Lasso 在部分线性模型中的表现. Ni [10] 用样条逼近非参数函数, 构造了 SCAD 惩罚最小二乘估计. Wang 等 [11] 研究了 profile 似然推断, 构造了似然比统计量, 并研究了其渐近分布. Zhu 等 [12] 基于 profile 最小二乘方法, 构造了参数向量和非参数函数的相合估计. Müller 和 van de Geer [13] 利用 Lasso 方法和光滑函数构造了参数向量和非参数函数双惩罚最小二乘估计. Ma 和 Huang [14] 用多项式样条逼近非参数函数, 构造了参数向量的 Lasso 惩罚最小二乘估计. 超高维情况下的文献更少, Liang 等 [15] 将向前回归方法推广到部分线性模型中, 提出 profile 向前回归方法, 研究了模型选择的相合性, 并证明了所提方法能够以更快的收敛速度包含真实模型. 赖秋楠等 [16] 针对超高维部分线性模型, 提出了 profile 贪婪向前回归变量筛选方法, 在一定正则条件下, 证明了所提方法具有筛选相合性. Li 等 [17] 将向前回归方法用于超高维变系数部分线性模型中, 并证明该方法具有模型选择相合性.

本文考虑超高维部分线性模型的变量选择问题, 其中参数向量的维数是样本量的指数阶. 基于 profile 最小二乘方法, 本文提出 profile regularization after retention (PRAR) 方法, 其主要思想是先利用 profile 方法将部分线性模型转化为线性模型, 根据边际相关系数估计保留一部分重要变量, 然后对其余变量进行 Lasso 惩罚, 从而得到线性部分参数向量的部分惩罚最小二乘估计. 并理论证明了估计量的符号相合性. 与 Lasso、SIS-Lasso 和自适应 Lasso 方法相比, 数值模拟显示出 PRAR 方法在恢复参数向量符号方面效果最好, 符号恢复比例依概率趋向于 1. 在实例分析中, PRAR 方法不仅能够选出最简单的模型, 而且预测误差是最小的.

本文结构如下: 第 2 节介绍部分线性模型, 并构造了参数向量的 PRAR 估计. 第 3 节研究估计量的渐近性质, 得到参数向量的估计具有符号相合性. 第 4 节通过数值模拟和实例分析研究所提方法的优良性. 第 5 节给出了定理的证明.

§2. 模型与方法

设 $\{(\boldsymbol{X}_i, U_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自 $(\boldsymbol{X}, U, Y)$ 的独立同分布的样本, 本文考虑的部分线性模型为

$$Y_i = \boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\beta} + g(U_i) + \varepsilon_i, \qquad 1 \leqslant i \leqslant n, \tag{1}$$

其中 $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^{\mathsf{T}}$ 为 p 维协变量, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^{\mathsf{T}}$ 是未知的 p 维参数向量, $g(\cdot)$ 是未知的非参数光滑函数, 通常假定 $\boldsymbol{\beta}$ 稀疏, 不失一般性, 假定 U_i 在 $[0,1]$ 上取值, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, 且与 $\boldsymbol{X}_i$ 和 U_i 独立.

为了便于说明, 引入一些符号表示. 对任一矩阵 $\boldsymbol{A}$, 定义投影阵 $P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}$. 对任一集合 A, 用 $|A|$ 表示集合中元素的个数. 对任一 k 维向量 $\boldsymbol{w}$, 若 $K \subseteq \{1, 2, \cdots, k\}$,

则 $\boldsymbol{w}_K=\{w_j, j\in K\}$ 表示 $|K|$ 维向量. 定义向量范数 $\|\boldsymbol{w}\|_1=\sum\limits_{i=1}^{k}|w_i|$, $\|\boldsymbol{w}\|_\infty=\max\limits_{1\leqslant i\leqslant k}|w_i|$, $\|\boldsymbol{w}\|_2=\sqrt{\boldsymbol{w}^\intercal\boldsymbol{w}}$. 对任一矩阵 $\boldsymbol{M}_{k_1\times k_2}$, 若 $K_1\subseteq\{1,2,\cdots,k_1\}$, $K_2\subseteq\{1,2,\cdots,k_2\}$, 记 $\boldsymbol{M}_{K_1K_2}=\{M_{ij}, i\in K_1, j\in K_2\}$ 表示 $|K_1|\times|K_2|$ 矩阵. $\boldsymbol{M}_{K_2}=\{M_{ij}, i=1,2,\cdots,k_1, j\in K_2\}$ 表示 $k_1\times|K_2|$ 矩阵. 定义矩阵范数

$$\|\boldsymbol{M}\|_\infty=\max_{1\leqslant i\leqslant k_1}\sum_{j=1}^{k_2}|M_{ij}|,\qquad \|\boldsymbol{M}\|_2=[\lambda_{\max}(\boldsymbol{M}^\intercal\boldsymbol{M})]^{1/2}.$$

不指明范数时, 本文均表示 2 范数. 当 $k_1=k_2=k$ 时, $\lambda_{\min}(\boldsymbol{M})$ 和 $\lambda_{\max}(\boldsymbol{M})$ 分别为 $\boldsymbol{M}$ 的最小特征值和最大特征值.

对模型 (1) 两边关于 U_i 求条件期望得 $g(U_i)=\mathsf{E}(Y_i\,|\,U_i)-\sum\limits_{j=1}^{p}\mathsf{E}(X_{ij}\,|\,U_i)\beta_j$, 代入式 (1) 整理有

$$\widetilde{Y}_i=\widetilde{\boldsymbol{X}}_i^\intercal\boldsymbol{\beta}+\varepsilon_i,\qquad 1\leqslant i\leqslant n, \tag{2}$$

其中 $\widetilde{Y}_i=Y_i-\mathsf{E}(Y_i\,|\,U_i)$, $\widetilde{\boldsymbol{X}}_i=\boldsymbol{X}_i-\mathsf{E}(\boldsymbol{X}_i\,|\,U_i)=(\widetilde{X}_{i1},\widetilde{X}_{i2},\cdots,\widetilde{X}_{ip})^\intercal$, 因此可将部分线性模型转化为线性模型. 记 $G_0(u)=\mathsf{E}(Y_i\,|\,U_i=u)$, $G_j(u)=\mathsf{E}(X_{ij}\,|\,U_i=u)$, $1\leqslant j\leqslant p$, $\widehat{G}_j(u)$ 为 $G_j(u)$ 的局部线性估计. 令 $K_h(\cdot)=h^{-1}K(\cdot/h)$, 其中 $K(\cdot)$ 为核函数, h 为带宽. 则有

$$\widehat{G}_0(u)=\sum_{k=1}^{n}W_{nk}(u)Y_k,\qquad \widehat{G}_j(u)=\sum_{k=1}^{n}W_{nk}(u)X_{kj},\quad 1\leqslant j\leqslant p,$$

其中

$$\begin{aligned}W_{ni}(u)&=\frac{K_h(U_i-u)[S_{n,2}(u)-(U_i-u)S_{n,1}(u)]}{S_{n,0}(u)S_{n,2}(u)-[S_{n,1}(u)]^2},\\ S_{n,j}(u)&=\sum_{i=1}^{n}K_h(U_i-u)(U_i-u)^j,\qquad j=0,1,2.\end{aligned}$$

记 $\widehat{Y}_i=Y_i-\widehat{G}_0(U_i)$, $\widehat{\boldsymbol{X}}_i=\boldsymbol{X}_i-\widehat{\boldsymbol{G}}(U_i)$, 其中 $\widehat{\boldsymbol{G}}(U_i)=(\widehat{G}_1(U_i),\widehat{G}_2(U_i),\cdots,\widehat{G}_p(U_i))^\intercal$, 设计阵 $\widehat{\boldsymbol{X}}=(\widehat{\boldsymbol{X}}_1,\widehat{\boldsymbol{X}}_2,\cdots,\widehat{\boldsymbol{X}}_n)^\intercal$, $\widehat{\boldsymbol{Y}}=(\widehat{Y}_1,\widehat{Y}_2,\cdots,\widehat{Y}_n)^\intercal$, 那么边际相关系数估计为

$$\widehat{\beta}_j^M=\sum_{i=1}^{n}\frac{(\widehat{X}_{ij}-\overline{X}_j)}{\sum\limits_{k=1}^{n}(\widehat{X}_{kj}-\overline{X}_j)^2}\widehat{Y}_i,\qquad 1\leqslant j\leqslant p, \tag{3}$$

其中 $\overline{X}_j=n^{-1}\sum\limits_{i=1}^{n}\widehat{X}_{ij}$. 给定阈值 γ_n, 可得集合 $\widehat{R}=\{1\leqslant j\leqslant p:|\widehat{\beta}_j^M|\geqslant\gamma_n\}$, 由于 $\widehat{R}$ 中包含了与 Y 边际相关系数较强的变量, 故称其为保留集. 当存在重要变量与响应变量边际不相关时, 这些变量将不会进入 $\widehat{R}$ 中, 为解决此问题, 考虑对 $\widehat{R}^c$ 中的变量施加惩罚, 定义部分线性模型参数向量的 PRAR 估计为

$$\check{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}}-\widehat{\boldsymbol{X}}\boldsymbol{\beta}\big\|^2+\lambda_n\big\|\boldsymbol{\beta}_{\widehat{R}^c}\big\|_1\Big\},$$

其中 λ_n 为调节参数, $\widehat{R}^c$ 为 $\widehat{R}$ 的余集. 上式既可使 $\widehat{R}$ 内变量系数的估计精度不受惩罚函数的影响, 又可以对 $\widehat{R}^c$ 内的变量系数施加惩罚, 使部分系数压缩为 0, 达到变量选择的目的.

当阈值 γ_n 取值较小或者变量中存在与 Y 的相关性较强的不重要变量时, $\widehat{R}$ 中将会误选不重要变量. 令 $Q=\{j\in\widehat{R}^c,\check{\beta}_j\neq 0\}$, 定义 $\boldsymbol{\beta}$ 的 PRAR_+ 估计

$$\widetilde{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}_{(\widehat{R}\cup Q)^c}=\mathbf{0}}{\operatorname{argmin}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}}-\widehat{\boldsymbol{X}}_{\widehat{R}}\boldsymbol{\beta}_{\widehat{R}}-\widehat{\boldsymbol{X}}_Q\boldsymbol{\beta}_Q\big\|^2+\lambda_n^*\big\|\boldsymbol{\beta}_{\widehat{R}}\big\|_1\Big\},$$

其中 λ_n^* 为调节参数, $\mathbf{0}$ 表示零向量, 在不同的地方其维数可能不同. 上式只是对 $\widehat{R}$ 内的变量进行筛选, 所以保证 Q 中对应的系数估计的精度不受惩罚函数的影响.

§3. 主要结果

记 $\{\beta_j^M,1\leqslant j\leqslant p\}$ 为真实的边际相关系数. 为了方便条件 (C7) 和证明, 令模型 (2) 中响应变量 $\widetilde{Y}$ 为 V_0, 协变量 $\widetilde{X}_j$ 为 V_j, $1\leqslant j\leqslant p$. 记 $M_j(u)$ 是 V_j 的矩母函数, $0\leqslant j\leqslant p$. 重要变量集 $S=\{1\leqslant j\leqslant p:\beta_j\neq 0\}$, 它的基数 $s=|S|$. 为了证明估计量的渐近性质, 需要以下正则条件.

(C1) 随机误差项 $\varepsilon_i\sim\mathrm{N}(0,\sigma^2)$.

(C2) 记 $\widetilde{\boldsymbol{X}_i}$ 的协方差阵为 Σ, 假设存在常数 $0<C_{\min}<C_{\max}<\infty$, 使得 $2C_{\min}<\lambda_{\min}(\Sigma)\leqslant\lambda_{\max}(\Sigma)<2^{-1}C_{\max}$.

(C3) 假设 $\|\boldsymbol{\beta}\|\leqslant C_{\boldsymbol{\beta}}$, 其中 $C_{\boldsymbol{\beta}}>0$ 为常数. 并且对某些常数 $C_1>0$, $\xi_{\min}>0$, 有 $\min\limits_{j\in S}|\beta_j|\geqslant C_1n^{-\xi_{\min}}$.

(C4) 存在正常数 ξ 和 ξ_0, 使得 $\ln p=O(n^{\xi})$, $s=O(n^{\xi_0})$, 并且 $0<\xi<(1-2k)/4$, 其中 $0<k<1/4$, $\xi+2\xi_0+3\xi_{\min}<1$.

(C5) 记 $b_n=n^{-4/5}$, $c_n=n^{-2/5}\ln n$, 则权重函数 $W_{nk}(\cdot)$ 满足

(i) $\max\limits_{1\leqslant k\leqslant n}\sum\limits_{i=1}^n W_{nk}(U_i)=O_p(1)$,

(ii) $\max\limits_{1\leqslant i,k\leqslant n}W_{nk}(U_i)=O_p(b_n)$,

(iii) $\max\limits_{1\leqslant i\leqslant n}\sum\limits_{k=1}^n W_{nk}(U_i)I(|U_i-U_k|>c_n)=O_p(c_n)$.

(C6) 存在 C_2, 使得对任意 $u_1,u_2\in[0,1]$, 有 $\max\limits_{0\leqslant j\leqslant p}|G_j(u_1)-G_j(u_2)|\leqslant C_2|u_1-u_2|$.

(C7) 假设对所有 $0\leqslant u\leqslant t_0/\sigma_v$, $\max\limits_{0\leqslant j\leqslant p}\mathsf{E}[\exp(u|V_j|)]<\infty$, 其中 t_0,σ_v 为正常数, 矩母函数 $M_j(u)$ $(0\leqslant j\leqslant p)$ 满足

$$\max_{0\leqslant j\leqslant p}\sup_{0\leqslant u\leqslant t_0}\Big|\frac{\mathrm{d}^3\ln M_j(u)}{\mathrm{d}u^3}\Big|<\infty.$$

另外假设 $\max\limits_{0\leqslant j\leqslant p}\mathsf{E}|V_j|^{2m}\leqslant\sigma_v^2$, 其中 $m>2$.

(C8) $\|\Sigma\boldsymbol{\beta}\|_\infty=O(n^{(1-2k)/8})$.

(C9) $\boldsymbol{\beta}_S^{\mathsf{T}}\Sigma_{SS}\boldsymbol{\beta}_S=O(1)$.

注记 1 条件 (C1) 主要是为了证明方便, 该条件可以放松到均值为 0 的轻尾分布情形, (C2) – (C4) 为模型选择文献中的常用假设, (C5) – (C6) 为非参数回归估计的标准假设, (C7) 给出了矩条件, (C8) 给出了边际相关系数的上界, (C9) 给出了 $\mathsf{Var}\,(\widetilde{Y})$ 的上界. 那么在给定的正则条件下, 可以得到下面的重要定理.

定理 2 假设条件 (C1) 和 (C4) – (C9) 成立, 则存在常数 $c_1>0$, 使得当 $n\to\infty$ 时, 有

$$\mathsf{P}\Big(\max_{1\leqslant j\leqslant p}|\widehat{\beta}_j^M-\beta_j^M|>c_1n^{-k}\Big)\longrightarrow 0.$$

定理 2 表明边际相关系数估计 $\widehat{\beta}_j^M$ 具有相合性. 根据文献 [6] 的推论 1, 若 $\zeta_n=\|\Sigma_{S^cS}\boldsymbol{\beta}_S\|_\infty$, 阈值满足 $\gamma_n=\zeta_n+c_1n^{-k}$, 则所选的保留集具有确定保留性质, 即 $\mathsf{P}(\widehat{R}\subseteq S)\to 1$. 定义集合 $R=\{j\in S:|\beta_j^M|>\zeta_n+2c_1n^{-k}\}$, 为了证明 PRAR 估计的渐近性质, 需要以下假设条件.

(C10) $\|\{\Sigma_{S^cS}\Sigma_{SS}^{-1}\}_{S\cap R^c}\|_\infty\leqslant 1-\gamma$, 其中 $\gamma\in(0,1]$.

定理 3 若条件 (C1) – (C10) 成立, 且 $\lambda_n=O(n^{-\xi_{\min}-\xi_0/2})$, 那么当 $n\to\infty$ 时, 有

$$\mathsf{P}\big(\mathrm{sign}(\check{\boldsymbol{\beta}})=\mathrm{sign}(\boldsymbol{\beta})\big)\longrightarrow 1.$$

定理 3 说明了 PRAR 方法得到的估计量具有符号相合性, 即该方法不仅能把重要变量和不重要变量分开, 并且对重要变量的系数估计能保证符号一致. 定义 $Z=\{j\in S^c:|\beta_j^M|\geqslant\gamma_n-c_1n^{-k}\}$, 其元素个数 $z=|Z|$, 为了证明 PRAR_+ 估计量的符号相合性, 需要下列的条件.

(C11) $\max\limits_{S\subset Q\subset S\cup Z}\|\{\Sigma_{Q^cQ}(\Sigma_{QQ})^{-1}\}_{S\cap R^c}\|_\infty\leqslant 1-\gamma$.

(C12) $\|\Sigma_{ZS}\Sigma_{SS}^{-1}\|_\infty\leqslant 1-\eta$, 其中 $\eta>0$.

定理 4 若条件 (C1) – (C9) 和 (C11) – (C12) 成立, 且满足

$$\frac{z}{s}\to 0,\qquad \lambda_n=O(n^{-\xi_{\min}-\xi_0/2}),\qquad \lambda_n^*=O(n^{-\xi_{\min}-\xi_0/2}),$$

那么当 $n\to\infty$ 时, 有

$$\mathsf{P}\big(\mathrm{sign}(\widetilde{\boldsymbol{\beta}})=\mathrm{sign}(\boldsymbol{\beta})\big)\longrightarrow 1.$$

定理 4 说明 PRAR_+ 估计具有符号相合性, 且适用于 R 中误选不重要变量情况, 具有稳健性.

§4. 数值模拟和实例分析

考虑部分线性模型 $Y=\boldsymbol{X}^{\intercal}\boldsymbol{\beta}+g(U)+\varepsilon$, 其中 $\boldsymbol{\beta}=(2.5,-2,0,\cdots,0)^{\intercal}$ 为 p 维向量, $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\intercal}\sim\mathrm{N}(0,\Sigma)$, 其中协方差阵 Σ 满足

$$\Sigma=\begin{bmatrix}\Sigma_{11} & 0\\ 0 & I\end{bmatrix},\qquad \text{其中 }\Sigma_{11}=\begin{bmatrix}1 & r_0 & r_1 & r_3\\ r_0 & 1 & r_2 & r_4\\ r_1 & r_2 & 1 & 0\\ r_3 & r_4 & 0 & 1\end{bmatrix}.$$

取非参数函数为 $g(U)=10\sin(2\pi U)$, 其中 U 服从 $[0,1]$ 上的均匀分布, $\varepsilon\sim\mathrm{N}(0,4)$. 对于协方差阵有下面两种参数设置:

(A) $r_0=0.8$, $r_1=-r_2=r_3=-r_4=-0.1$;

(B) $r_0=0.75$, $r_1=r_2=r_3=-r_4=0.2$.

模拟过程中分别生成容量 n 为 100、200、300、400、500 的样本, 且维数 $p=\lfloor 100\cdot\exp(n^{0.2})\rfloor$. 所有的 Lasso 过程由 glmnet 包来实现, 在进行局部线性估计时, 用广义交叉验证法选取带宽, 选择 Epanechnikov 核 $K(x)=0.75(1-x^2)I(|x|\leqslant 1)$. 在 SIS 方法中, 只选前 $\lfloor n/\ln n\rfloor$ 个变量进入模型. 自适应 Lasso 方法的权重选为 $w_j=1/\widehat{\beta}_j^M$, 其中 $\widehat{\beta}_j^M$ 由式 (3) 计算得到. PRAR 方法采用 Weng 等[6] 提出的置换方法选取阈值 γ_n, 若得到的保留集 $\widehat{R}$ 的个数大于 $\lceil\sqrt{n}\,\rceil$, 那么只保留系数较大的 $\lceil\sqrt{n}\,\rceil$ 个变量.

采用符号恢复比例来衡量方法的优劣, 即 200 次模拟中能够只选择出变量 X_1 和 X_2, 并且相应系数估计值恰为一正一负的比例. 模拟结果见表 1 和表 2, 表中 PLasso、PSIS-Lasso 和 PAda-Lasso 分别表示 Lasso、SIS-Lasso 和自适应 Lasso 方法, PRAR 和 PRAR_+ 分别表示本文所提的 PRAR 估计和 PRAR_+ 估计, 它们右下角数字为计算阈值 γ_n 时进行置换的次数, 如 PRAR_{10} 表示对响应变量置换 10 次后得到的估计量.

由表 1 可知, 随着样本量的增加, 每种方法的符号恢复比例都有升高的趋势. Lasso、SIS-Lasso 和自适应 Lasso 方法表现较差, 只有当样本量为 500 时才偶尔恢复 $\boldsymbol{\beta}$ 的符号. 相比而言, PRAR 和 PRAR_+ 方法明显有很大优势, 特别是 PRAR_+ 方法, 当样本量为 500 时, 符号恢复比例趋于 1, 从侧面印证了 PRAR 估计的符号相合性. 另外随着置换次数的增加, 符号恢复比例基本上变化不大, 并不呈现明显的升高或降低趋势.

由表 2 可知, Lasso、SIS-Lasso 以及自适应 Lasso 方法的表现依然很差. PRAR 方法的估计也不理想, 但这并不影响 PRAR_+ 方法的拟合效果. 可以看到随着样本量的增加, PRAR_+ 估计的符号恢复比例趋向于 1, 这说明 PRAR_+ 估计具有稳健性.

下面将 PRAR 方法用于分析一组乳腺癌数据, 其来自文献 [18]. 该数据集从 97 位年龄在 55 岁以下淋巴结阴性乳腺癌患者得到, 包含 97 组样本, 每组样本探测了 24 881 个基因表达水平和 7 个临床危险因素 (年龄、雌激素受体和孕激素受体状态、淋巴细胞渗透、Angioinvasion、病理等级、肿瘤大小). 由于部分数据缺失, 其中 24 188 个基因数据有效.

表 1　(A) 情况下各种方法的符号恢复比例

(n,p)	$(100,1\,323)$	$(200,1\,791)$	$(300,2\,285)$	$(400,2\,750)$	$(500,3\,199)$
PLasso	0.000	0.000	0.000	0.000	0.025
PSIS-Lasso	0.000	0.000	0.000	0.000	0.000
PAda-Lasso	0.000	0.000	0.000	0.000	0.015
PRAR_1	0.030	0.395	0.655	0.530	0.420
PRAR_5	0.030	0.370	0.775	0.750	0.625
PRAR_{10}	0.015	0.355	0.755	0.815	0.735
PRAR_{15}	0.010	0.320	0.735	0.810	0.760
PRAR_{30}	0.010	0.300	0.705	0.855	0.850
PRAR_{+1}	0.030	0.475	0.860	0.920	0.965
PRAR_{+5}	0.030	0.395	0.840	0.950	0.975
PRAR_{+10}	0.015	0.365	0.800	0.975	0.990
PRAR_{+15}	0.010	0.330	0.760	0.935	0.990
PRAR_{+30}	0.010	0.300	0.715	0.970	0.975

表 2　(B) 情况下各种方法的符号恢复比例

(n,p)	$(100,1\,323)$	$(200,1\,791)$	$(300,2\,285)$	$(400,2\,750)$	$(500,3\,199)$
PLasso	0.000	0.000	0.000	0.000	0.005
PSIS-Lasso	0.000	0.000	0.000	0.000	0.000
PAda-Lasso	0.000	0.000	0.000	0.000	0.000
PRAR_1	0.070	0.135	0.050	0.005	0.000
PRAR_5	0.085	0.225	0.100	0.020	0.000
PRAR_{10}	0.065	0.250	0.135	0.035	0.000
PRAR_{15}	0.070	0.275	0.120	0.050	0.010
PRAR_{30}	0.060	0.310	0.185	0.085	0.005
PRAR_{+1}	0.120	0.490	0.770	0.920	0.970
PRAR_{+5}	0.110	0.480	0.835	0.950	0.975
PRAR_{+10}	0.080	0.460	0.785	0.915	0.990
PRAR_{+15}	0.080	0.470	0.805	0.950	0.980
PRAR_{+30}	0.065	0.450	0.810	0.945	0.970

赖秋楠等[16] 在部分线性模型下对该数据集进行研究, 根据其研究结果, 本文考虑使用基因 271 作为响应变量 Y_i, 找出与该基因相关的其他基因, 并且将基因 272 作为非参数函数的协变量 U_i, 而其余基因的表达水平作为线性协变量 $\boldsymbol{X}_i$, 据此构造如下模型

$$Y_i = \boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\beta} + g(U_i) + \varepsilon_i$$

来拟合给定的数据, 其中 ε_i 是随机误差项. 基于本文提出的方法, 对上述模型做变量筛选, 并与 Lasso、SIS-Lasso 和自适应 Lasso 方法进行比较. 首先对数据进行标准化处理, 以消

除各基因表达水平数量级差别的影响, 然后在 97 组样本中随机选取 92 个做为训练集 $\mathscr{R}$, 其余 5 个做为测试集 $\mathscr{Q}$. 与数值模拟不同的是, SIS 步选择 $d=n-1$ 个变量进入模型. 整个过程重复进行 60 次, 通过测试集 $\mathscr{Q}$ 来计算各种方法的平均预测均方误差:

$$\mathrm{APMSE}=\frac{1}{60}\sum_{i=1}^{60}\Big[\frac{1}{5}\sum_{i\in\mathscr{Q}}\Big(\widehat{Y}_i-\sum_{j=1}^{p}\widehat{X}_{ij}\widehat{\beta}_j\Big)^2\Big].$$

除此之外, 计算各种方法所选择的平均模型大小:

$$\mathrm{AMS}=\frac{1}{60}\sum_{i=1}^{60}|M_i|,$$

其中 M_i 表示第 i 次拟合所得的子模型, $|M_i|$ 表示模型 M_i 中元素个数. 所得结果见下表 3, 表中括号内数据表示相应预测均方误差和模型大小的标准差.

表 3　各种方法的预测均方误差及平均模型大小

	PLasso	PSIS-Lasso	PAda-Lasso	PRAR	PRAR_+
APMSE	0.247 (0.169)	0.257 (0.177)	0.246 (0.148)	0.233 (0.193)	0.228 (0.180)
AMS	29.117 (22.444)	28.667 (5.694)	7.067 (14.654)	3.217 (5.551)	2.960 (5.562)

由表 3 可看出 Lasso、SIS-Lasso 和自适应 Lasso 方法的预测均方误差较大, 并且 Lasso 和 SIS-Lasso 方法选择的模型较大. 相比而言, PRAR 和 PRAR_+ 方法相差不大, 不仅选择的模型最简单, 而且平均预测均方误差最小.

§5.　定理证明

本节将给出定理的详细证明. 在开始证明定理之前, 先介绍以下两个引理.

引理 5　假设条件 (C1) 和 (C4)–(C6) 成立, 则有

$$\max_{0\leqslant j\leqslant p}\max_{1\leqslant i\leqslant n}|\widehat{G}_j(U_i)-G_j(U_i)|=o_p(n^{-1/4}\ln^{-1}n).$$

引理 6　令 $\widehat{\Sigma}=n^{-1}\widehat{\boldsymbol{X}}^{\mathsf{T}}\widehat{\boldsymbol{X}}$, $\overline{\Sigma}=n^{-1}\widetilde{\boldsymbol{X}}^{\mathsf{T}}\widetilde{\boldsymbol{X}}$, 集合 $\mathscr{M}\subseteq\{1,2,\cdots,p\}$, 在引理 5 的条件下, 若条件 (C2) 成立, 且 $\check{m}=O(n^{2\xi_0+4\xi_{\min}})$, 则下列不等式成立的概率趋于 1,

$$C_{\min}\leqslant\min_{|\mathscr{M}|\leqslant\check{m}}\lambda_{\min}(\widehat{\Sigma}_{\mathscr{M}\mathscr{M}})\leqslant\max_{|\mathscr{M}|\leqslant\check{m}}\lambda_{\max}(\widehat{\Sigma}_{\mathscr{M}\mathscr{M}})\leqslant C_{\max}.$$

引理 5 和引理 6 的证明详见文献 [15].

定理 2 的证明:　根据引理 5 可知

$$\max_{1\leqslant i\leqslant n}|\widehat{Y}_i-\widetilde{Y}_i|=\max_{1\leqslant i\leqslant n}|\widehat{G}_0(U_i)-G_0(U_i)|=o_p(n^{-1/4}\ln^{-1}n). \tag{4}$$

同理有

$$\max_{1\leqslant j\leqslant p}\max_{1\leqslant i\leqslant n}|\widehat{X}_{ij}-\widetilde{X}_{ij}|=o_p(n^{-1/4}\ln^{-1}n). \tag{5}$$

令

$$\widetilde{\beta}_j^M=\sum_{i=1}^n\frac{(\widetilde{X}_{ij}-\overline{X}'_j)}{\sum\limits_{k=1}^n(\widetilde{X}_{kj}-\overline{X}'_j)^2}\widetilde{Y}_i,\qquad 1\leqslant j\leqslant p,$$

其中 $\overline{X}'_j=n^{-1}\sum\limits_{i=1}^n\widetilde{X}_{ij}$. 根据文献 [6] 的命题 2 可知, 在条件 (C8)–(C9) 下, 对任意 $c_2>0$, 存在常数 $c_3>0$, 使得

$$\mathsf{P}\Big(\max_{1\leqslant j\leqslant p}|\widetilde{\beta}_j^M-\beta_j^M|>c_2n^{-k}\Big)=O\big(p\exp(-c_3n^{(1-2k)/4})\big). \tag{6}$$

记 $S_{j1}=n^{-1}\sum\limits_{i=1}^n(\widehat{X}_{ij}-\overline{X}_j)^2$, $S_{j2}=n^{-1}\sum\limits_{i=1}^n(\widetilde{X}_{ij}-\overline{X}'_j)^2$, $1\leqslant j\leqslant p$, 则有

$$\begin{aligned}&\max_{1\leqslant j\leqslant p}|\widehat{\beta}_j^M-\widetilde{\beta}_j^M|\\ \leqslant&\max_{1\leqslant j\leqslant p}\Big|\sum_{i=1}^n\frac{(\widehat{X}_{ij}-\overline{X}_j)(\widehat{Y}_i-\widetilde{Y}_i)}{nS_{j1}}\Big|+\max_{1\leqslant j\leqslant p}\Big|\sum_{i=1}^n\Big[\frac{(\widehat{X}_{ij}-\overline{X}_j)}{nS_{j1}}-\frac{(\widetilde{X}_{ij}-\overline{X}'_j)}{nS_{j2}}\Big]\widetilde{Y}_i\Big|\\ \doteq&\, I_1+I_2.\end{aligned}$$

由 Cauchy-Schwarz 不等式和式 (4) 可知

$$I_1\leqslant\max_{1\leqslant j\leqslant p}\sqrt{\frac{1}{nS_{j1}}\sum_{i=1}^n(\widehat{Y}_i-\widetilde{Y}_i)^2}=o_p(n^{-1/4}\ln^{-1}n).$$

另一方面

$$\begin{aligned}I_2\leqslant&\max_{1\leqslant j\leqslant p}\Big|\sum_{i=1}^n\frac{(\widehat{X}_{ij}-\widetilde{X}_{ij})\widetilde{Y}_i}{nS_{j1}}\Big|+\max_{1\leqslant j\leqslant p}\Big|\sum_{i=1}^n\frac{(\overline{X}'_j-\overline{X}_j)\widetilde{Y}_i}{nS_{j1}}\Big|\\ &+\max_{1\leqslant j\leqslant p}\Big|\Big(\frac{1}{nS_{j1}}-\frac{1}{nS_{j2}}\Big)\sum_{i=1}^n(\widetilde{X}_{ij}-\overline{X}'_j)\widetilde{Y}_i\Big|\\ \doteq&\, I_{21}+I_{22}+I_{23}.\end{aligned}$$

根据式 (4) 和式 (5), 在条件 (C9) 下, 再次运用 Cauchy-Schwarz 不等式可得

$$\begin{aligned}I_{21}&\leqslant\max_{1\leqslant j\leqslant p}\frac{1}{S_{j1}}\sqrt{\frac{1}{n}\sum_{i=1}^n(\widehat{X}_{ij}-\widetilde{X}_{ij})^2\cdot\frac{1}{n}\sum_{i=1}^n\widehat{Y}_i^2}=o_p(n^{-1/4}\ln^{-1}n),\\ I_{22}&\leqslant\max_{1\leqslant j\leqslant p}\frac{1}{S_{j1}}\sqrt{\frac{1}{n}\sum_{i=1}^n(\overline{X}'_j-\overline{X}_j)^2\cdot\frac{1}{n}\sum_{i=1}^n\widehat{Y}_i^2}=o_p(n^{-1/4}\ln^{-1}n),\end{aligned}$$

$$I_{23} \leqslant \max_{1\leqslant j\leqslant p} \frac{|S_{j2}-S_{j1}|}{S_{j1}S_{j2}} \sqrt{S_{j2}\cdot \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2} = o_p(n^{-1/4}\ln^{-1} n).$$

从而 $I_2 = o_p(n^{-1/4}\ln^{-1} n)$. 由以上讨论可知, 对任意常数 $c>0$, 当 $n\to\infty$ 时, 有

$$\mathsf{P}\Big(\max_{1\leqslant j\leqslant p} |\widehat{\beta}_j^M - \widetilde{\beta}_j^M| > cn^{-1/4}\ln^{-1} n\Big) \longrightarrow 0. \tag{7}$$

由式 (6) 和式 (7) 可知, 存在常数 $c_1>0$, 使得当 $n\to\infty$ 时,

$$\mathsf{P}\Big(\max_{1\leqslant j\leqslant p} |\widehat{\beta}_j^M - \beta_j^M| > c_1 n^{-k}\Big) \longrightarrow 0.$$

定理证毕. □

定理 3 的证明: 定义集合 $\overline{S}^c = \widehat{R}^c\backslash S^c$, PRAR 估计为

$$\check{\boldsymbol\beta} = \operatorname*{argmin}_{\boldsymbol\beta}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol Y} - \widehat{\boldsymbol X}\boldsymbol\beta\big\|^2 + \lambda_n\big\|\boldsymbol\beta_{\widehat{R}^c}\big\|_1\Big\}. \tag{8}$$

可得估计方程 $n^{-1}\widehat{\boldsymbol X}^{\intercal}(\widehat{\boldsymbol Y} - \widehat{\boldsymbol X}\check{\boldsymbol\beta}) = \lambda_n\partial\big\|\check{\boldsymbol\beta}_{\widehat{R}^c}\big\|$, 其中 $\partial\big\|\check{\boldsymbol\beta}_{\widehat{R}^c}\big\|$ 表示 $\big\|\boldsymbol\beta_{\widehat{R}^c}\big\|_1$ 在 $\boldsymbol\beta = \check{\boldsymbol\beta}$ 的偏导数. 定义式 (8) 的 oracle 估计为

$$\overline{\boldsymbol\beta} = \operatorname*{argmin}_{\boldsymbol\beta_{S^c}=\mathbf{0}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol Y} - \widehat{\boldsymbol X}_S\boldsymbol\beta_S\big\|^2 + \lambda_n\big\|\boldsymbol\beta_{\overline{S}^c}\big\|_1\Big\}. \tag{9}$$

显然式 (9) 是一个严格凸问题, 具有唯一解. 因此如果向量 $\overline{\boldsymbol\beta}$ 是式 (9) 的解, 当且仅当

$$\overline{\boldsymbol\beta}_{S^c} = 0, \qquad n^{-1}\widehat{\boldsymbol X}_S^{\intercal}(\widehat{\boldsymbol Y} - \widehat{\boldsymbol X}_S\overline{\boldsymbol\beta}_S) = \lambda_n\partial\big\|\overline{\boldsymbol\beta}_{\overline{S}^c}\big\|, \tag{10}$$

其中 $\partial\big\|\overline{\boldsymbol\beta}_{\overline{S}^c}\big\|$ 表示 $\big\|\boldsymbol\beta_{\overline{S}^c}\big\|_1$ 在 $\boldsymbol\beta_S = \overline{\boldsymbol\beta}_S$ 的偏导数. 由式 (10) 得

$$\overline{\boldsymbol\beta}_S = \big(\widehat{\boldsymbol X}_S^{\intercal}\widehat{\boldsymbol X}_S\big)^{-1}\big(\widehat{\boldsymbol X}_S^{\intercal}\widehat{\boldsymbol Y} - n\lambda_n\partial\big\|\overline{\boldsymbol\beta}_{\overline{S}^c}\big\|\big).$$

显然, 如果 $\overline{\boldsymbol\beta}_S$ 满足

$$\big\|n^{-1}\widehat{\boldsymbol X}_{S^c}^{\intercal}(\widehat{\boldsymbol Y} - \widehat{\boldsymbol X}_S\overline{\boldsymbol\beta}_S)\big\|_\infty < \lambda_n. \tag{11}$$

那么式 (9) 的唯一解 $\overline{\boldsymbol\beta}$ 也满足式 (8). 将 $\overline{\boldsymbol\beta}_S$ 代入式 (11), 有

$$\Big\|\widehat{\boldsymbol X}_{S^c}^{\intercal}\widehat{\boldsymbol X}_S\big(\widehat{\boldsymbol X}_S^{\intercal}\widehat{\boldsymbol X}_S\big)^{-1}\partial\big\|\overline{\boldsymbol\beta}_{\overline{S}^c}\big\| + (n\lambda_n)^{-1}\widehat{\boldsymbol X}_{S^c}^{\intercal}\big[\boldsymbol I_n - \widehat{\boldsymbol X}_S\big(\widehat{\boldsymbol X}_S^{\intercal}\widehat{\boldsymbol X}_S\big)^{-1}\widehat{\boldsymbol X}_S^{\intercal}\big]\widehat{\boldsymbol Y}\Big\|_\infty < 1. \tag{12}$$

令

$$\begin{aligned} &W = \{\check{\boldsymbol\beta}\ \text{唯一, 且}\ \operatorname{sign}(\check{\boldsymbol\beta}) = \operatorname{sign}(\boldsymbol\beta)\},\\ &W_1 = \{\text{式 (8) 有唯一解, 且式 (12) 成立}\},\\ &W_2 = \Big\{\min_{j\in S}|\beta_j| > \big\|\overline{\boldsymbol\beta}_S - \boldsymbol\beta_S\big\|_\infty\Big\}. \end{aligned}$$

由文献 [6] 的式 (A14) 可知, $\mathsf{P}(W) \geqslant \mathsf{P}(W_1) - \mathsf{P}(W_2^c)$. 因此要想证明 $\check{\boldsymbol{\beta}}$ 的符号相合性, 只需证明 $W_1 \xrightarrow{\mathsf{P}} 1$ 且 $W_2^c \xrightarrow{\mathsf{P}} 0$. 下面记

$$\begin{aligned}
&\mathscr{T} = \{S_* : R \subseteq S_* \subseteq S\},\\
&A = \{R \subseteq \widehat{R} \subseteq S\},\\
&B = \Big\{\max_{1\leqslant j\leqslant p} |\widehat{\beta}_j^M - \beta_j^M| \leqslant c_1 n^{-k}\Big\}.
\end{aligned}$$

由文献 [6] 的推论 1 和式 (3.3) 可知, 当 $n \to \infty$ 时, 有 $\mathsf{P}(A) \to 1$, 定理 2 表明 $\mathsf{P}(B) \to 1$. 将式 (12) 重写为

$$\Big\|\widehat{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big(\widehat{\boldsymbol{X}}_S^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| + (n\lambda_n)^{-1}\widehat{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\big(\boldsymbol{I}_n - P_{\widehat{\boldsymbol{X}}_S}\big)\widehat{\boldsymbol{Y}}\Big\|_\infty < 1. \tag{13}$$

记 $\boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)^{\mathsf{T}}$, 即只有第 i 个元素为 1, 其余都为 0, 在不同的地方可以有不同的维数. 则根据式 (5) 与式 (4) 可知

$$\begin{aligned}
&\boldsymbol{e}_i^{\mathsf{T}}\widehat{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big(\widehat{\boldsymbol{X}}_S^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| = \boldsymbol{e}_i^{\mathsf{T}}\widetilde{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| + o_p(1),\\
&\boldsymbol{e}_i^{\mathsf{T}}\widehat{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\big(\boldsymbol{I}_n - P_{\widehat{\boldsymbol{X}}_S}\big)\widehat{\boldsymbol{Y}} = \boldsymbol{e}_i^{\mathsf{T}}\widetilde{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\big(\boldsymbol{I}_n - P_{\widetilde{\boldsymbol{X}}_S}\big)\widetilde{\boldsymbol{Y}} + o_p(1).
\end{aligned}$$

又因为 $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, 因此式 (13) 等价于

$$\Big\|\widetilde{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| + (n\lambda_n)^{-1}\widetilde{\boldsymbol{X}}_{S^c}^{\mathsf{T}}\big(\boldsymbol{I}_n - P_{\widetilde{\boldsymbol{X}}_S}\big)\boldsymbol{\varepsilon}\Big\|_\infty < 1. \tag{14}$$

令

$$\begin{aligned}
&\boldsymbol{F} = \widetilde{\boldsymbol{X}}_{S^c}^{\mathsf{T}} - \Sigma_{S^cS}\Sigma_{SS}^{-1}\widetilde{\boldsymbol{X}}_S^{\mathsf{T}},\\
&\boldsymbol{K}_1 = \Sigma_{S^cS}\Sigma_{SS}^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\|,\\
&\boldsymbol{K}_2 = \boldsymbol{F}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| + (n\lambda_n)^{-1}\boldsymbol{F}\big(\boldsymbol{I}_n - P_{\widetilde{\boldsymbol{X}}_S}\big)\boldsymbol{\varepsilon}.
\end{aligned}$$

则式 (14) 等价于 $\|\boldsymbol{K}_1 + \boldsymbol{K}_2\|_\infty < 1$. 一方面, 当 A 成立时, 根据条件 (C10) 可得

$$\|\boldsymbol{K}_1\|_\infty = \big\|\{\Sigma_{S^cS}\Sigma_{SS}^{-1}\}_{S\cap\widehat{R}^c}\big\|_\infty \leqslant \big\|\{\Sigma_{S^cS}\Sigma_{SS}^{-1}\}_{S\cap R^c}\big\|_\infty \leqslant 1 - \gamma.$$

因此

$$\mathsf{P}(\|\boldsymbol{K}_1\|_\infty \leqslant 1 - \gamma) \longrightarrow 1, \qquad n \to \infty. \tag{15}$$

另一方面

$$\begin{aligned}
\mathsf{P}\Big(\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big) &= \mathsf{P}\Big(\Big\{\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big\} \cap A\Big) + \mathsf{P}\Big(\Big\{\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big\} \cap A^c\Big)\\
&\leqslant \mathsf{P}\Big(\bigcup_{S_1\in\mathscr{T}} \|\boldsymbol{K}_2(S_1)\|_\infty > \frac{\gamma}{2}\Big) + \mathsf{P}(A^c),
\end{aligned}$$

其中 $\boldsymbol{K}_2(S_1)$ 是用 S_1 代替 $\widehat{R}$ 后得到相应的 $\boldsymbol{K}_2$, 对应的式 (9) 的解用 $\overline{\boldsymbol{\beta}}(S_1)$ 来表示, 则

$$\boldsymbol{K}_2(S_1)=\boldsymbol{F}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|+(n\lambda_n)^{-1}\boldsymbol{F}\big(\boldsymbol{I}_n-P_{\widetilde{\mathbf{X}}_S}\big)\boldsymbol{\varepsilon}.$$

根据 $\widetilde{\boldsymbol{X}}_i$ 的定义可知, 期望 $\mathsf{E}(\widetilde{\boldsymbol{X}}_i)=\mathsf{E}[\boldsymbol{X}_i-\mathsf{E}(\boldsymbol{X}_i\,|\,U_i)]=\mathbf{0}$. 则 $\boldsymbol{F}$ 的期望为 $\mathbf{0}$, 协方差阵 $\mathsf{Cov}\,(\boldsymbol{F})=\Sigma_{S^cS^c}-\Sigma_{S^cS}(\Sigma_{SS})^{-1}\Sigma_{SS^c}$, 记 $\Sigma_{S^c|S}$. 对任意 $1\leqslant j\leqslant|S^c|$, 记 $\boldsymbol{F}$ 的第 j 行为 $\boldsymbol{F}_j^{\mathsf{T}}=(F_{j1},F_{j2},\cdots,F_{jn})$, 因为 F_{jk} $(1\leqslant k\leqslant n)$ 为独立的随机变量, 且期望为 0, 方差为 $(\Sigma_{S^c|S})_{jj}$, 由中心极限定理可知

$$\boldsymbol{F}_j^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|+(n\lambda_n)^{-1}\boldsymbol{F}_j^{\mathsf{T}}\big(\boldsymbol{I}_n-P_{\widetilde{\mathbf{X}}_S}\big)\boldsymbol{\varepsilon}\,\big|\,\big(\widetilde{\boldsymbol{X}}_S,\boldsymbol{\varepsilon}\big)\sim\mathrm{N}(0,V_j),\tag{16}$$

其中方差

$$V_j=(\Sigma_{S^c|S})_{jj}\big[\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|+(n\lambda_n)^{-2}\boldsymbol{\varepsilon}^{\mathsf{T}}\big(\boldsymbol{I}_n-P_{\widetilde{\mathbf{X}}_S}\big)\boldsymbol{\varepsilon}\big].$$

由于 Σ 有限, 并且 $\boldsymbol{I}_n-P_{\widetilde{\mathbf{X}}_S}$ 为对称幂等阵, 从而存在 C_3, 使得

$$V_j\leqslant C_3\big[\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{T}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|+(n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2\big].$$

令

$$H=\bigcup_{S_1\in\mathscr{T}}\Big\{\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|+(n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2>\frac{s}{nC_{\min}}+\frac{\sigma^2(1+\sqrt{s/n}\,)}{n\lambda_n^2}\Big\}.$$

则

$$\mathsf{P}\Big(\bigcup_{S_1\in\mathscr{T}}\|\boldsymbol{K}_2(S_1)\|_\infty>\frac{\gamma}{2}\Big)\leqslant\mathsf{P}\Big(\bigcup_{S_1\in\mathscr{T}}\|\boldsymbol{K}_2(S_1)\|_\infty>\frac{\gamma}{2}\,\Big|\,H^c\Big)+\mathsf{P}(H).\tag{17}$$

由于

$$\begin{aligned}\mathsf{P}(H)\leqslant\mathsf{P}\Big(&\bigcup_{S_1\in\mathscr{T}}\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|>\frac{s}{nC_{\min}}\Big)\\&+\mathsf{P}\Big((n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2>\frac{\sigma^2(1+\sqrt{s/n}\,)}{n\lambda_n^2}\Big).\end{aligned}$$

对任意 $S_1\in\mathscr{T}$, 根据文献 [15] 的式 (A.7) 可知, 下式成立的概率趋于 1,

$$\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|\leqslant s\big\|\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\big\|=\frac{s}{n}\|\overline{\Sigma}_{SS}^{-1}\|\leqslant\frac{s}{nC_{\min}}.$$

因此

$$\mathsf{P}\Big(\bigcup_{S_1\in\mathscr{T}}\big(\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\big\|>\frac{s}{nC_{\min}}\Big)\longrightarrow0.$$

另一方面, 由于 $\varepsilon_i\sim\mathrm{N}(0,\sigma^2)$, $1\leqslant i\leqslant n$, 从而 $\|\boldsymbol{\varepsilon}\|^2/\sigma^2\sim\chi^2(n)$, 根据文献 [19] 的式 (54a) 可知

$$\mathsf{P}\Big((n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2>\frac{\sigma^2(1+\sqrt{s/n}\,)}{n\lambda_n^2}\Big)=\mathsf{P}\Big(\frac{\|\boldsymbol{\varepsilon}\|^2}{\sigma^2}>n\big(1+\sqrt{s/n}\,\big)\Big)\leqslant\exp\Big(-\frac{3s}{16}\Big),$$

其中 $s/n < 1/2$, 因此 $\mathsf{P}(H) \to 0$. 由正态分布的尾概率不等式可知

$$\mathsf{P}\Big(\bigcup_{S_1 \in \mathscr{T}} \|\boldsymbol{K}_2(S_1)\|_\infty > \frac{\gamma}{2} \,\Big|\, H^c \Big) \leqslant 2^{s+1}(p-s)\exp\Big(-\frac{\gamma^2}{8V}\Big),$$

其中 $V = s/(nC_{\min}) + \sigma^2(1+\sqrt{s/n}\,)/(n\lambda_n^2)$. 容易求得 $\ln(p-s)+(s+1)\ln 2 = o\big(\gamma^2/(8V)\big)$, 因此存在 $c_* > 0$, 使得对充分大的 n, 有

$$\mathsf{P}\Big(\bigcup_{S_1 \in \mathscr{T}} \|\boldsymbol{K}_2(S_1)\|_\infty > \frac{\gamma}{2} \,\Big|\, H^c \Big) \leqslant \exp(-c_* s).$$

从而可得

$$\mathsf{P}\Big(\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big) \longrightarrow 0, \qquad n \to \infty. \tag{18}$$

根据式 (15) 和 (18) 可得

$$\mathsf{P}\Big(\|\boldsymbol{K}_1 + \boldsymbol{K}_2\|_\infty \leqslant 1 - \frac{\gamma}{2}\Big) \longrightarrow 1, \qquad n \to \infty. \tag{19}$$

从而式 (12) 以概率 1 成立. 下面证明式 (8) 有唯一解, 从而 $W_1 \xrightarrow{\mathsf{P}} 1$, 首先记

$$K = \Big\{ \Big\| \widehat{\boldsymbol{X}}_{S^c}^\intercal \widehat{\boldsymbol{X}}_S \big(\widehat{\boldsymbol{X}}_S^\intercal \widehat{\boldsymbol{X}}_S\big)^{-1} \partial \|\overline{\boldsymbol{\beta}}_{\overline{S}^c}\| + (n\lambda_n)^{-1} \widehat{\boldsymbol{X}}_{S^c}^\intercal \big(\boldsymbol{I}_n - P_{\widehat{\boldsymbol{X}}_S}\big)\widehat{\boldsymbol{Y}} \Big\|_\infty < 1 - \frac{\gamma}{2} \Big\}.$$

则由式 (19) 可知, $K \xrightarrow{\mathsf{P}} 1$. 下面证明当 A 和 K 成立时, 式 (8) 必然有唯一解. 不妨假设存在另一个解 $\overline{\boldsymbol{\beta}}^*$, 令 $\overline{\boldsymbol{\beta}}(\alpha) = \alpha\overline{\boldsymbol{\beta}} + (1-\alpha)\overline{\boldsymbol{\beta}}^*$ $(0 < \alpha < 1)$. 则 $\overline{\boldsymbol{\beta}}(\alpha)$ 也是式 (8) 的解, 从而

$$\begin{aligned}
&\big\|n^{-1}\widehat{\boldsymbol{X}}_{S^c}^\intercal(\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{X}}_S\overline{\boldsymbol{\beta}}(\alpha))\big\|_\infty \\
&\leqslant \alpha\big\|n^{-1}\widehat{\boldsymbol{X}}_{S^c}^\intercal(\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{X}}_S\overline{\boldsymbol{\beta}})\big\|_\infty + (1-\alpha)\big\|n^{-1}\widehat{\boldsymbol{X}}_{S^c}^\intercal(\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{X}}_S\overline{\boldsymbol{\beta}}^*)\big\|_\infty \\
&< \alpha\lambda_n + (1-\alpha)\lambda_n = \lambda_n.
\end{aligned}$$

因此 $\overline{\boldsymbol{\beta}}(\alpha)_{S^c} = \boldsymbol{0}$, 即 $\overline{\boldsymbol{\beta}}(\alpha)$ 也是式 (9) 的解, 由解的唯一性知 $\overline{\boldsymbol{\beta}}(\alpha) = \overline{\boldsymbol{\beta}}$, 从而 $\overline{\boldsymbol{\beta}} = \overline{\boldsymbol{\beta}}^*$, 即式 (8) 有唯一解, 因此 $W_1 \xrightarrow{\mathsf{P}} 1$. 最后证明 $W_2 \xrightarrow{\mathsf{P}} 0$. 首先考虑 $\|\overline{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S\|_\infty$, 简单计算有

$$\begin{aligned}
\overline{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S &= \big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S\big)^{-1}\widehat{\boldsymbol{X}}_S^\intercal\boldsymbol{\varepsilon} + \big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S\big)^{-1}\widehat{\boldsymbol{X}}_S^\intercal\big(\widehat{\boldsymbol{X}}_S - \widetilde{\boldsymbol{X}}_S\big)\boldsymbol{\beta}_S \\
&\quad + \big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S\big)^{-1}\widehat{\boldsymbol{X}}_S^\intercal\big(\widehat{\boldsymbol{Y}} - \widetilde{\boldsymbol{Y}}\big) - \lambda_n\big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S/n\big)^{-1}\partial\|\overline{\boldsymbol{\beta}}_{\overline{S_1}^c}\| \\
&\doteq I_3 + I_4 + I_5 - I_6.
\end{aligned}$$

根据条件 (C3) 和引理 6 可得

$$\begin{aligned}
\|I_4\|_\infty &\leqslant \big\|\big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S/n\big)^{-1}\big\|_2\big\|n^{-1}\widehat{\boldsymbol{X}}_S^\intercal\big(\widehat{\boldsymbol{X}}_S - \widetilde{\boldsymbol{X}}_S\big)\boldsymbol{\beta}_S\big\|_2 \leqslant \frac{C_{\boldsymbol{\beta}}\sqrt{C_{\max}}}{C_{\min}}o_p\big(n^{-1/4}\ln^{-1}n\big), \\
\|I_5\|_\infty &\leqslant \big\|\big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S/n\big)^{-1}\big\|_2\big\|n^{-1}\widehat{\boldsymbol{X}}_S^\intercal\big(\widehat{\boldsymbol{Y}} - \widetilde{\boldsymbol{Y}}\big)\big\|_2 \leqslant \frac{\sqrt{C_{\max}}}{C_{\min}}o_p\big(n^{-1/4}\ln^{-1}n\big),
\end{aligned}$$

$$\|I_6\|_\infty \leqslant \lambda_n s^{1/2}\big\|\big(\widehat{\boldsymbol{X}}_S^\intercal \widehat{\boldsymbol{X}}_S/n\big)^{-1}\big\|_2 = \lambda_n s^{1/2}\lambda_{\max}(\widehat{\Sigma}_{SS}^{-1}) \leqslant \frac{\lambda_n s^{1/2}}{C_{\min}}.$$

另一方面, 由于

$$\boldsymbol{e}_i^\intercal\big(\widehat{\boldsymbol{X}}_S^\intercal\widehat{\boldsymbol{X}}_S\big)^{-1}\widehat{\boldsymbol{X}}_S^\intercal\boldsymbol{\varepsilon} = \boldsymbol{e}_i^\intercal\big(\widetilde{\boldsymbol{X}}_S^\intercal\widetilde{\boldsymbol{X}}_S\big)^{-1}\widetilde{\boldsymbol{X}}_S^\intercal\boldsymbol{\varepsilon} + o_p(1).$$

那么由文献 [6] 的式 (A28) 可知 $\mathsf{P}\big(\|I_3\|_\infty > \sqrt{s/(nC_{\min})}\big) \leqslant 2\exp(-s/(2\sigma^2))$. 因此

$$\mathsf{P}\Big(\big\|\boldsymbol{\beta}_S - \overline{\boldsymbol{\beta}}_S\big\|_\infty \leqslant \frac{\lambda_n s^{1/2}}{C_{\min}} + \sqrt{s/(nC_{\min})}\Big) \geqslant 1 - 2\exp(-c_4 s),$$

其中 $c_4 > 0$, 在条件 (C3) 下很容易验证, 对于充分大的 n,

$$\min_{j\in S}|\beta_j| > \frac{\lambda_n s^{1/2}}{C_{\min}} + \sqrt{s/(nC_{\min})}.$$

因此 $\mathsf{P}(W_2^c) \leqslant 1 - 2\exp(-c_4 s) \to 0$. 综上可知

$$\mathsf{P}\big(\mathrm{sign}(\check{\boldsymbol{\beta}}) = \mathrm{sign}(\boldsymbol{\beta})\big) \longrightarrow 1, \qquad n \to \infty.$$

定理证毕. □

定理 4 的证明: 记不重要变量集 $N = \{j : 1 \leqslant j \leqslant p,\, \beta_j = 0\}$. 定义 $S = \widehat{S}_1 \cup \widehat{S}_2$, $N = \widehat{N}_1 \cup \widehat{N}_2$, 其中 $\widehat{S}_2$ 和 $\widehat{N}_2$ 组成保留集 $\widehat{R} = \widehat{S}_2 \cup \widehat{N}_2$. 首先, 考虑 PRAR 估计

$$\check{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{X}}\boldsymbol{\beta}\big\|^2 + \lambda_n\big(\big\|\boldsymbol{\beta}_{\widehat{S}_1}\big\|_1 + \big\|\boldsymbol{\beta}_{\widehat{N}_1}\big\|_1\big)\Big\}. \tag{20}$$

下面证明 $\mathsf{P}(\check{\boldsymbol{\beta}}_{\widehat{S}_1} \neq \boldsymbol{0}) \to 1$ 和 $\mathsf{P}(\check{\boldsymbol{\beta}}_{\widehat{N}_1} = \boldsymbol{0}) \to 1$. 定义式 (20) 的 oracle 估计

$$\overline{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}_{\widehat{N}_1}=\boldsymbol{0}}{\operatorname{argmin}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{X}}_{\widehat{Q}}\boldsymbol{\beta}_{\widehat{Q}}\big\|^2 + \lambda_n\big\|\boldsymbol{\beta}_{\widehat{S}_1}\big\|_1\Big\}, \tag{21}$$

其中 $\widehat{Q} = S \cup \widehat{N}_2$, 类似定理 3, 要想证明 $\mathsf{P}(\check{\boldsymbol{\beta}}_{\widehat{S}_1} \neq \boldsymbol{0}) \to 1$, 只要证明式 (20) 有唯一解, 且

$$\Big\|\widetilde{\boldsymbol{X}}_{\widehat{Q}^c}^\intercal\widetilde{\boldsymbol{X}}_{\widehat{Q}}\big(\widetilde{\boldsymbol{X}}_{\widehat{Q}}^\intercal\widetilde{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\widehat{Q}}\big\| + (n\lambda_n)^{-1}\widetilde{\boldsymbol{X}}_{\widehat{Q}^c}^\intercal\big(\boldsymbol{I}_n - P_{\widetilde{\boldsymbol{X}}_{\widehat{Q}}}\big)\boldsymbol{\varepsilon}\Big\|_\infty < 1. \tag{22}$$

令

$$\begin{aligned}
\boldsymbol{F} &= \widetilde{\boldsymbol{X}}_{\widehat{Q}^c}^\intercal - \Sigma_{\widehat{Q}^c\widehat{Q}}\Sigma_{\widehat{Q}\widehat{Q}}^{-1}\widetilde{\boldsymbol{X}}_{\widehat{Q}}^\intercal,\\
\boldsymbol{K}_1 &= \Sigma_{\widehat{Q}^c S}\Sigma_{\widehat{Q}\widehat{Q}}^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\widehat{Q}}\big\|,\\
\boldsymbol{K}_2 &= \boldsymbol{F}\widetilde{\boldsymbol{X}}_{\widehat{Q}}\big(\widetilde{\boldsymbol{X}}_{\widehat{Q}}^\intercal\widetilde{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1}\partial\big\|\overline{\boldsymbol{\beta}}_{\widehat{Q}}\big\| + (n\lambda_n)^{-1}\boldsymbol{F}\big(\boldsymbol{I}_n - P_{\widetilde{\boldsymbol{X}}_{\widehat{Q}}}\big)\boldsymbol{\varepsilon}.
\end{aligned}$$

那么, 式 (22) 等价于 $\|\boldsymbol{K}_1 + \boldsymbol{K}_2\|_\infty < 1$. 注意到 $\widehat{Q}$ 是随机的. 下面引入一些符号表示.

$$A = \{R \subseteq \widehat{S}_2 \subseteq S,\, S \subseteq \widehat{Q} \subseteq S \cup Z\}, \qquad B = \{S \subseteq \widehat{Q} \subseteq S \cup Z\},$$

$$A^* = \{(Q, S_2) : S \subseteq Q \subseteq S \cup Z,\, R \subseteq S_2 \subseteq S\}, \qquad C = \{\widehat{N}_2 \subseteq Z\}.$$

从定理 2 不难看出 $\mathsf{P}(A) \to 1$, 因此由条件 (C11) 可得 $\mathsf{P}(\|\boldsymbol{K}_1\|_\infty < 1-\gamma) \geqslant \mathsf{P}(A) \to 1$. 令

$$\begin{aligned} H = \bigcup_{(Q,S_2)\in A^*} \Big\{ & \big(\partial\|\overline{\boldsymbol{\beta}}_Q\|\big)^{\mathsf{T}} \big(\widetilde{\boldsymbol{X}}_Q^{\mathsf{T}} \widetilde{\boldsymbol{X}}_Q\big)^{-1} \partial\|\overline{\boldsymbol{\beta}}_Q\| + (n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2 \\ & > \frac{s+z}{nC_{\min}} + \frac{\sigma^2\big(1+\sqrt{(s+z)/n}\,\big)}{n\lambda_n^2} \Big\}. \end{aligned}$$

这里 $\partial\|\overline{\boldsymbol{\beta}}_Q\|$ 是用 Q 和 S_2 代替式 (21) 的 $\widehat{Q}$ 和 $\widehat{S}_2$ 得到的, 那么

$$\begin{aligned} \mathsf{P}\Big(\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big) &\leqslant \mathsf{P}\Big(\Big\{\|\boldsymbol{K}_2\|_\infty > \frac{\gamma}{2}\Big\} \cap A\Big) + \mathsf{P}(A^c) \\ &\leqslant \mathsf{P}\Big(\Big\{\bigcup_{(Q,S_2)\in A^*} \|\boldsymbol{K}_2(Q,S_2)\|_\infty > \frac{\gamma}{2}\Big\} \cap A\Big) + \mathsf{P}(A^c) \\ &\leqslant \mathsf{P}\Big(\bigcup_{(Q,S_2)\in A^*} \|\boldsymbol{K}_2(Q,S_2)\|_\infty > \frac{\gamma}{2} \,\Big|\, H^c\Big) + \mathsf{P}(H) + \mathsf{P}(A^c). \end{aligned}$$

先看第二项,

$$\begin{aligned} \mathsf{P}(H) \leqslant\; & \mathsf{P}\Big(\bigcup_{(Q,S_2)\in A^*} \big(\partial\|\overline{\boldsymbol{\beta}}_Q\|\big)^{\mathsf{T}} \big(\widetilde{\boldsymbol{X}}_Q^{\mathsf{T}} \widetilde{\boldsymbol{X}}_Q\big)^{-1} \partial\|\overline{\boldsymbol{\beta}}_Q\| > \frac{s+z}{nC_{\min}}\Big) \\ & + \mathsf{P}\Big((n\lambda_n)^{-2}\|\boldsymbol{\varepsilon}\|^2 > \frac{\sigma^2\big(1+\sqrt{(s+z)/n}\,\big)}{n\lambda_n^2}\Big). \end{aligned}$$

类似定理 3 可知 $\mathsf{P}(H) \to 0$, 并且

$$\mathsf{P}\Big(\bigcup_{(Q,S_2)\in A^*} \|\boldsymbol{K}_2(Q,S_2)\|_\infty > \frac{\gamma}{2} \,\Big|\, H^c\Big) \leqslant 2^{s+1+z}(p-s)\exp\Big(-\frac{\gamma^2}{8V}\Big),$$

其中 $V = (s+z)/(nC_{\min}) + \sigma^2\big(1+\sqrt{(s+z)/n}\big)/(n\lambda_n^2)$. 在定理 4 的设置下, 显然式 (22) 成立. 式 (20) 解的唯一性类似定理 3 可得. 接下来考虑 $\|\overline{\boldsymbol{\beta}}_{\widehat{Q}} - \boldsymbol{\beta}_{\widehat{Q}}\|_\infty$, 简单计算可得

$$\begin{aligned} \overline{\boldsymbol{\beta}}_{\widehat{Q}} - \boldsymbol{\beta}_{\widehat{Q}} = \;& \big(\widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \widehat{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1} \widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \boldsymbol{\varepsilon} + \big(\widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \widehat{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1} \widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \big(\widehat{\boldsymbol{X}}_S - \widetilde{\boldsymbol{X}}_S\big)\boldsymbol{\beta}_S \\ & + \big(\widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \widehat{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1} \widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \big(\widehat{\boldsymbol{Y}} - \widetilde{\boldsymbol{Y}}\big) - \lambda_n \big(\widehat{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \widehat{\boldsymbol{X}}_{\widehat{Q}}/n\big)^{-1} \partial\|\overline{\boldsymbol{\beta}}_{\widehat{Q}}\|. \end{aligned}$$

根据 Cauchy-Schwarz 不等式可得

$$\|\overline{\boldsymbol{\beta}}_{\widehat{Q}} - \boldsymbol{\beta}_{\widehat{Q}}\|_\infty \leqslant \big\|\big(\widetilde{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \widetilde{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1} \widetilde{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}} \boldsymbol{\varepsilon}\big\|_\infty + \frac{\lambda_n (s+z)^{1/2}}{C_{\min}}.$$

令 $L_1 = \lambda_n (s+z)^{1/2}/C_{\min} + \sqrt{(s+z)/(nC_{\min})}$, 则

$$\mathsf{P}\big(\|\overline{\boldsymbol{\beta}}_{\widehat{Q}} - \boldsymbol{\beta}_{\widehat{Q}}\|_\infty \geqslant L_1\big)$$

$$
\begin{aligned}
&\leqslant \mathsf{P}\Big(\bigcup_{S\subseteq Q\subseteq S\cup Z}\big\|\big(\widetilde{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}}\widetilde{\boldsymbol{X}}_{\widehat{Q}}\big)^{-1}\widetilde{\boldsymbol{X}}_{\widehat{Q}}^{\mathsf{T}}\boldsymbol{\varepsilon}\big\|_{\infty}\geqslant\sqrt{(s+z)/(nC_{\min})}\Big)+\mathsf{P}(B^c)\\
&\leqslant 2^{z+1}\exp\Big(-\frac{s+z}{2\sigma^2}\Big)+\mathsf{P}(B^c).
\end{aligned}
$$

由于 $\mathsf{P}(B)\geqslant\mathsf{P}(A)\to 1$, 故可得 $\mathsf{P}(\|\overline{\boldsymbol{\beta}}_{\widehat{Q}}-\boldsymbol{\beta}_{\widehat{Q}}\|_\infty\geqslant L_1)\to 0$. 由条件 (C3), 不难证明 $\min\limits_{j\in S}|\beta_j|\geqslant L_1$, 因此

$$
\mathsf{P}\Big(\min_{j\in S}|\beta_j|>\|\overline{\boldsymbol{\beta}}_{\widehat{S}_1}-\boldsymbol{\beta}_{\widehat{S}_1}\|_\infty\Big)\longrightarrow 1. \tag{23}
$$

在定理 4 的条件下, 有 $\mathsf{P}(\check{\boldsymbol{\beta}}_{\widehat{S}_1}\neq\mathbf{0})\to 1$. 对于 PRAR$_+$ 估计, 有

$$
\widetilde{\boldsymbol{\beta}}=\mathop{\mathrm{argmin}}_{\boldsymbol{\beta}_{\widehat{N}_1}=\mathbf{0}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}}-\widehat{\boldsymbol{X}}_{\widehat{Q}}\boldsymbol{\beta}_{\widehat{Q}}\big\|^2+\lambda_n^*\big(\|\boldsymbol{\beta}_{\widehat{S}_2}\|_1+\|\boldsymbol{\beta}_{\widehat{N}_2}\|_1\big)\Big\}. \tag{24}
$$

为了证明 $\mathsf{P}\big(\mathrm{sign}(\widetilde{\boldsymbol{\beta}})=\mathrm{sign}(\boldsymbol{\beta})\big)\to 1$, 只需证明 $\mathsf{P}\big(\mathrm{sign}(\widetilde{\boldsymbol{\beta}}_S)=\mathrm{sign}(\boldsymbol{\beta}_S)\big)\to 1$, 且 $\mathsf{P}\big(\widetilde{\boldsymbol{\beta}}_{\widehat{N}_2}=\mathbf{0}\big)\to 1$. 定义式 (24) 的 oracle 估计

$$
\boldsymbol{\beta}^*=\mathop{\mathrm{argmin}}_{\boldsymbol{\beta}_{\widehat{N}_1}=\mathbf{0},\,\boldsymbol{\beta}_{\widehat{N}_2}=\mathbf{0}}\Big\{\frac{1}{2n}\big\|\widehat{\boldsymbol{Y}}-\widehat{\boldsymbol{X}}_S\boldsymbol{\beta}_S\big\|^2+\lambda_n^*\big\|\boldsymbol{\beta}_{\widehat{S}_2}\big\|_1\Big\}.
$$

令

$$
\begin{aligned}
\widetilde{\boldsymbol{F}}&=\widetilde{\boldsymbol{X}}_{\widehat{N}_2}^{\mathsf{T}}-\Sigma_{\widehat{N}_2S}\Sigma_{SS}^{-1}\widetilde{\boldsymbol{X}}_S^{\mathsf{T}},\\
\widetilde{\boldsymbol{K}}_1&=\Sigma_{\widehat{N}_2S}\Sigma_{SS}^{-1}\partial\|\boldsymbol{\beta}_S^*\|,\\
\widetilde{\boldsymbol{K}}_2&=\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{X}}_S\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\|\boldsymbol{\beta}_S^*\|+(n\lambda_n^*)^{-1}\widetilde{\boldsymbol{F}}\big(\boldsymbol{I}_n-P_{\widetilde{\boldsymbol{X}}_S}\big)\boldsymbol{\varepsilon}.
\end{aligned}
$$

那么由条件 (C12) 可知 $\mathsf{P}(\|\widetilde{\boldsymbol{K}}_1\|_\infty<1-\eta)\geqslant\mathsf{P}(C)\to 1$. 令

$$
\widetilde{H}=\bigcup_{R\subseteq S_2\subseteq S}\Big\{\big(\partial\|\boldsymbol{\beta}_S^*\|\big)^{\mathsf{T}}\big(\widetilde{\boldsymbol{X}}_S^{\mathsf{T}}\widetilde{\boldsymbol{X}}_S\big)^{-1}\partial\|\boldsymbol{\beta}_S^*\|+(n\lambda_n^*)^{-2}\|\boldsymbol{\varepsilon}\|^2>\frac{s}{nC_{\min}}+\frac{\sigma^2(1+\sqrt{s/n}\,)}{n(\lambda_n^*)^2}\Big\}.
$$

记集合 $B^*=\{(N_2,S_2):N_2\subseteq Z,\,R\subseteq S_2\subseteq S\}$, 则有

$$
\begin{aligned}
\mathsf{P}\Big(\big\|\widetilde{\boldsymbol{K}}_2\big\|_\infty>\frac{\eta}{2}\Big)&\leqslant\mathsf{P}\Big(\Big\{\big\|\widetilde{\boldsymbol{K}}_2\big\|_\infty>\frac{\eta}{2}\Big\}\cap A\Big)+\mathsf{P}(A^c)\\
&\leqslant\mathsf{P}\Big(\Big\{\bigcup_{(N_2,S_2)\in B^*}\big\|\widetilde{\boldsymbol{K}}_2(N_2,S_2)\big\|_\infty>\frac{\eta}{2}\Big\}\Big)+\mathsf{P}(A^c)\\
&\leqslant\mathsf{P}\Big(\bigcup_{(N_2,S_2)\in B^*}\|\widetilde{\boldsymbol{K}}_2(N_2,S_2)\|_\infty>\frac{\eta}{2}\,\Big|\,\widetilde{H}^c\Big)+\mathsf{P}(\widetilde{H})+\mathsf{P}(A^c)\\
&\leqslant 2^{z+s+1}z\exp\Big(-\frac{\eta^2}{8\widetilde{V}}\Big)+\exp\Big(-\frac{3s}{16}\Big)+\mathsf{P}(A^c),
\end{aligned}
$$

其中 $\widetilde{V}=s/(nC_{\min})+\sigma^2(1+\sqrt{s/n}\,)/[n(\lambda_n^*)^2]$. 从而下式成立的概率趋于 1,

$$
\Big\|\widehat{\boldsymbol{X}}_{\widehat{N}_2}^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big(\widehat{\boldsymbol{X}}_S^{\mathsf{T}}\widehat{\boldsymbol{X}}_S\big)^{-1}\partial\|\boldsymbol{\beta}_S^*\|+\frac{1}{n\lambda_n}\widehat{\boldsymbol{X}}_{\widehat{N}_2}^{\mathsf{T}}\big(\boldsymbol{I}_n-P_{\widehat{\boldsymbol{X}}_S}\big)\widehat{\boldsymbol{Y}}\Big\|_\infty<1.
$$

式 (24) 解的唯一性类似定理 3 可得, 则 $\mathsf{P}\big(\widetilde{\boldsymbol{\beta}}_{\widehat{N}_2}=\mathbf{0}\big)\to 1$. 最后考虑 $\|\boldsymbol{\beta}_S^*-\boldsymbol{\beta}_S\|_\infty$, 简单计算有

$$\begin{aligned}\big\|\boldsymbol{\beta}_S^*-\boldsymbol{\beta}_S\big\|_\infty&=\Big\|\big(\widehat{\boldsymbol{X}}_S^{\intercal}\widehat{\boldsymbol{X}}_S\big)^{-1}\big(\widehat{\boldsymbol{X}}_S^{\intercal}\widehat{\boldsymbol{Y}}-n\lambda_n^*\partial\big\|\boldsymbol{\beta}_S^*\big\|\big)-\boldsymbol{\beta}_S\Big\|_\infty\\&\leqslant\big\|\big(\widetilde{\boldsymbol{X}}_S^{\intercal}\widetilde{\boldsymbol{X}}_S\big)^{-1}\widetilde{\boldsymbol{X}}_S^{\intercal}\boldsymbol{\varepsilon}\big\|_\infty+\frac{\lambda_n^*s^{1/2}}{C_{\min}}.\end{aligned}$$

令 $L_2=\lambda_n^*s^{1/2}/C_{\min}+\sqrt{s/(nC_{\min})}$, 则存在 c_5, $\mathsf{P}(\|\boldsymbol{\beta}_S^*-\boldsymbol{\beta}_S\|_\infty\leqslant L_2)\geqslant 1-2\exp(-c_5s)$. 由于 $L_1=O(L_2)$,

$$\mathsf{P}\Big(\min_{j\in S}|\beta_j|>\big\|\boldsymbol{\beta}_S^*-\boldsymbol{\beta}_S\big\|_\infty\Big)\geqslant 1-2\exp(-c_5s).$$

因此

$$\mathsf{P}\big(\mathrm{sign}\big(\widetilde{\boldsymbol{\beta}}\big)=\mathrm{sign}(\boldsymbol{\beta})\big)\longrightarrow 1,\qquad n\to\infty.$$

定理证毕. $\square$

参 考 文 献

[1] ENGLE R F, GRANGER C W J, RICE J, et al. Semiparametric estimates of the relation between weather and electricity sales [J]. *J Amer Statist Assoc*, 1986, **81(394)**: 310–320.

[2] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. *J Roy Statist Soc Ser B*, 1996, **58(1)**: 267–288.

[3] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *J Amer Statist Assoc*, 2001, **96(456)**: 1348–1360.

[4] ZOU H. The adaptive lasso and its oracle properties [J]. *J Amer Statist Assoc*, 2006, **101(476)**: 1418–1429.

[5] FAN J Q, LV J C. Sure independence screening for ultrahigh dimensional feature space [J]. *J R Stat Soc Ser B Stat Methodol*, 2008, **70(5)**: 849–911.

[6] WENG H L, FENG Y, QIAO X Y. Regularization after retention in ultrahigh dimensional linear regression models [J]. *Statist Sinica*, 2019, **29(1)**: 387–407.

[7] WANG H S. Forward regression for ultra-high dimensional variable screening [J]. *J Amer Statist Assoc*, 2009, **104(488)**: 1512–1524.

[8] CHENG M Y, FENG S Y, LI G R, et al. Greedy forward regression for variable screening [J]. *Aust N Z J Stat*, 2018, **60(1)**: 20–42.

[9] 李锋, 卢一强, 李高荣. 部分线性模型的 Adaptive LASSO 变量选择 [J]. *应用概率统计*, 2012, **28(6)**: 614–624.

[10] NI H F. Penalized least-squares estimation for regression coefficients in high-dimensional partially linear models [J]. *J Statist Plann Inference*, 2012, **142(2)**: 379–389.

[11] WANG X G, SONG L X, KANG X N. Profile likelihood inferences on the partially linear model with a diverging number of parameters [J]. *Comm Statist Theory Methods*, 2014, **43(1)**: 13–27.

[12] ZHU L P, LI R Z, CUI H J. Robust estimation for partially linear models with large-dimensional covariates [J]. *Sci China Math*, 2013, **56(10)**: 2069–2088.

[13] MÜLLER P, VAN DE GEER S. The partial linear model in high dimensions [J]. *Scand J Stat*, 2015, **42(2)**: 580–608.

[14] MA C, HUANG J. Asymptotic properties of Lasso in high-dimensional partially linear models [J]. *Sci China Math*, 2016, **59(4)**: 769–788.

[15] LIANG H, WANG H S, TSAI C L. Profiled forward regression for ultrahigh dimensional variable screening in semiparametric partially linear models [J]. *Statist Sinica*, 2012, **22(2)**: 531–554.

[16] 赖秋楠, 李玉杰, 李高荣. 超高维部分线性模型的 PGFR 变量筛选 [J]. 应用概率统计, 2017, **33(6)**: 608–624.

[17] LI Y J, LI G R, LIAN H, et al. Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models [J]. *J Multivariate Anal*, 2017, **155**: 133–150.

[18] VAN'T VEER L J, DAI H Y, VAN DE VIJVER M J, et al. Gene expression profiling predicts clinical outcome of breast cancer [J]. *Nature*, 2002, **415(6871)**: 530–536.

[19] WAINWRIGHT M J. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso) [J]. *IEEE Trans Inform Theory*, 2009, **55(5)**: 2183–2202.

Profile Regularization after Retention Variable Selection for Ultrahigh Dimensional Partially Linear Models

YANG Xin

(*College of Statistics and Data Science, Faculty of Science, Beijing University of Technology, Beijing, 100124, China*)

LI Bingyue

(*Beijing Institute of Computer Technology and Applications, Beijing, 100854, China*)

TIAN Ping

(*School of Science, Xuchang University, Xuchang, 461000, China*)

Abstract: In this paper, we consider the ultrahigh dimensional partially linear model, in which the dimension of the parametric vector is exponential order of the sample size. Based on profile least squares and regularization after retention method, we propose a new method to perform variable selection for the ultrahigh dimensional partially linear model. Under certain regularity conditions, it is proved that the estimator achieves sign consistency. Compared with Lasso, SIS-Lasso and adaptive Lasso, it is found that the proposed method is better in terms of recovering the coefficient sign of linear part through the numerical simulation and real data analysis.

Keywords: partially linear model; variable selection; high-dimensional data; Lasso; sign consistency; regularization after retention

2020 Mathematics Subject Classification: 62G05; 62G20