高维单指标门限回归模型的估计方法

李地青¹, 何勇^{2*}, 张新生¹

¹ 复旦大学管理学院, 上海 200433;

² 山东财经大学统计学院, 济南 250014

E-mail: dqli16@fudan.edu.cn, heyong@sdufe.edu.cn, xszhang@fudan.edu.cn

收稿日期: 2018-09-06; 接受日期: 2018-11-28; 网络出版日期: 2019-06-03; * 通信作者

国家自然科学基金 (批准号: 11801316 和 11571080)、全国统计科学研究 (批准号: 2018LY63) 和山东省社科规划基金 (批准号: 16CTJJ01 和 18DSHJ04) 资助项目

摘要 本文考虑存在协变量阈值参数的高维单指标门限回归模型. 本文提出基于 l_1 范数惩罚方法来估计回归系数和阈值参数, 并且提出一种近端梯度算法来检测可能存在的变点. 此外, 在一定的稀疏条件下, 本文得到回归系数估计量 l_1 范数下估计误差和预测误差的非渐近 Oracle 不等式. 最后, 通过数值模拟研究展示所提出方法的有限样本表现.

关键词 变点 协变量阈值 高维数据 单指标模型

MSC (2010) 主题分类 62H12

1 引言

自从文献 [1] 提出了 LASSO (least absolute shrinkage and selection operator) 估计, 过去几十年出现了大量研究高维线性回归模型的文献. 对于高维线性回归方法的一般概述和近期发展, 可参见文献 [2]. 线性回归模型假设预测变量与响应变量之间的回归函数是线性的, 但在实际中, 对高维数据进行建模时, 发现线性假设过于严格, 往往并不适用. 为了解决该问题, 很多学者考虑将正则化惩罚方法与非线性模型相结合, 继而研究高维非线性回归模型. 例如, 统计学家们研究了高维可加回归模型 [3, 4]、单指标模型 [5–8] 和关联结构回归模型 [9, 10] 等.

单指标模型通过引入连接函数将线性回归模型进行推广, 在很多科学领域越来越受欢迎, 包括生物统计、经济学和金融计量经济学等 (参见文献 [11–14]). 单指标模型具有如下形式: $y = f(\mathbf{x}^T \boldsymbol{\beta}) + \epsilon$, 其中 $y \in \mathbb{R}$ 是一维响应变量, $\mathbf{x} \in \mathbb{R}^p$ 是 p 维协变量, $\boldsymbol{\beta} \in \mathbb{R}^p$ 是稀疏的回归系数向量, ϵ 是误差项. 当上述 f 是恒等函数时, 单指标函数就退化成普通线性模型. 单指标模型形式上更加灵活, 具有很强的可解释性, 且能够对变量间的交互效应进行建模. 这些优良性使其成为重要的半参数回归模型.

本文关注高维单指标模型, 并且假设模型中的函数形式 f 是已知的. 具有已知函数形式的单指标模型有重要的统计意义, 例如, 它与相位恢复 (phase retrieval) 问题密切相关, 具体细节可以参见文

英文引用格式: Li D Q, He Y, Zhang X S. Threshold single index regression model from high-dimensional data (in Chinese). Sci Sin Math, 2020, 50: 423–446, doi: 10.1360/SCM-2018-0618

献 [15]. 另一方面, 模型参数的不稳定性在信号处理、质量控制、故障检测和金融等领域非常普遍. 参数变化可能是由政策或资源、人口、社会方面的改变引起的. 忽略参数变化可能导致预测误差偏大, 进而使模型不可靠, 这促使我们考虑由协变量阈值而产生变点的高维单指标模型. 尽管有大量文献研究高维回归模型或者阈值变点模型, 但其中只有少数同时涉及上述两个方面. 文献 [16] 在线性模型下提出了一种信息准则方法, 能够同时进行变点检测和变量选择. 文献 [17] 考虑了结构突变自回归模型, 并将其估计问题转化为回归中的变量选择问题, 提出了基于组 LASSO (group LASSO) 惩罚的估计方法. 文献 [18] 考虑了含有不可观测固定交互效应的面板模型下结构变点的估计问题, 并提出了利用带有自适应性组融合 LASSO (group fused LASSO) 惩罚项的主成分估计方法来检测多个变点. 文献 [19] 在线性模型下提出了一种基于 LASSO 惩罚来检测由于协变量阈值而产生的变点和估计回归系数的方法. 据我们所知, 本文是首次考虑高维情形下存在协变量阈值的单指标模型的文章. 具体地, Y 是响应变量, $\mathbf{X}$ 是一个 p 维协变量, U 是阈值变量, 可能是 $\mathbf{X}$ 的子集. 假设 $\{(y_i, \mathbf{X}_i, U_i), i = 1, \dots, b\}$ 是 $(Y, \mathbf{X}, U)$ 的随机样本. 阈值单指标模型具有如下形式:

$$y_i = f(\mathbf{X}_i^T \boldsymbol{\beta}_0 + \mathbf{X}_i^T \boldsymbol{\delta}_0 I\{U_i \leq \tau_0\}) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

其中参数 $\boldsymbol{\beta}_0$ 和 $\boldsymbol{\delta}_0$ 是未知的 p 维回归系数, τ_0 是潜在的真实阈值参数, 通常假设它的取值范围是 $(0, 1)$. 文献 [20] 曾在经典低维线性回归问题中考虑过同样的模型 (1.1), 其中 $U_i = i/n$. 假设模型中的 f 是单调并连续可微的, 基于这个假设, 估计 (1.1) 中阈值参数的一个直接方法是对变换后的数据 $\{f^{-1}(y_i), \mathbf{X}_i\}$ 使用文献 [19] 提出的方法, 进而解决可能存在变点的稀疏线性回归问题. 但该方法仅适用于误差项 $\epsilon = 0$, 即无噪声的情形. 否则, 对数据的逆变换必将降低阈值参数估计的精度, 导致预测误差过高.

本文提出基于 l_1 范数惩罚的方法来估计回归系数和阈值参数. 由于阈值参数的存在以及链接函数 f 非线性, 所得到的优化问题是非凸的. 我们提出一种近端梯度算法, 通过搜索 τ_0 的可能取值, 比较目标函数的大小, 从而确定变点. 但对固定的 τ 值, 由于函数 f 的非线性, l_1 范数惩罚的优化问题仍然是非凸的, 这使我们无法得到全局最优解. 然而, 根据文献 [15] 的结果可以发现该优化问题的任何驻点在一定条件下都具有最优的统计收敛速度. 因而, 我们提出基于梯度下降的算法, 每步迭代在梯度下降的步骤后, 再进行软阈值计算, 保证其能收敛到驻点.

总体来说, 本文的贡献主要在于我们为高维单指标门限回归模型提出了基于 l_1 范数惩罚的回归系数和阈值参数的估计, 并提出有效的近端梯度算法来求解相应的非凸优化问题. 在一定的稀疏条件下, 我们得到非渐近 Oracle 不等式用来度量预测误差和回归系数的 l_1 范数误差. 模拟结果表明, 我们提出的方法与作为基准的 Oracle 方法表现相近. 所谓 Oracle 方法是指假定真实变点位置或者回归系数的支撑集合已知.

本文剩余内容安排如下. 第 2 节介绍一些定义和符号说明, 并提出基于 l_1 范数惩罚的估计模型, 同时提供算法来解决相应的优化问题. 第 3 节介绍所提方法的理论性质. 第 4 节进行数值模拟来比较所提方法以及一些传统方法的有限样本表现. 第 5 节简要讨论未来相关的研究方向. 定理的具体证明在附录 A 中给出.

2 方法

在介绍 l_1 范数惩罚估计方法前, 我们先给出一些符号说明. 以下符号在本文各处的引用含义是一致的. 令 $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{d \times d}$, 记 $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{ij}|$, $\|\mathbf{A}\|_1 = \sum_{i=1}^d \sum_{j=1}^d |a_{ij}|$. $\{(y_i, \mathbf{X}_i, U_i), i = 1, \dots, n\}$

是独立同分布的样本, 其中 $\mathbf{X}_i \in \mathbb{R}^d$. $\mathbf{X}_i(\tau) = (\mathbf{X}_i^T, \mathbf{X}_i^T I\{U_i < \tau\})$ 为 $2d \times 1$ 维的向量. $\mathbf{X}(\tau)$ 为设计矩阵, 其中第 i 行是 $\mathbf{X}_i^T(\tau)$. 设 $\mathbf{a} = (a_1, \dots, a_m)^T$ 和 $\mathbf{b} = (b_1, \dots, b_m)^T$ 为 $m \times 1$ 维的向量, $|\mathbf{a}|_p$ 表示向量的 p 范数, $\|\mathbf{a}\|_m = ((1/2m) \sum_{i=1}^m a_i^2)^{\frac{1}{2}}$ 表示经验范数, $\langle \mathbf{a}, \mathbf{b} \rangle$ 表示 $\mathbf{a}$ 和 $\mathbf{b}$ 的内积. $\text{soft}(u, v)$ 表示软阈值, 其中 $\text{soft}(u, v) = \text{sign}(u) \max\{|u| - v, 0\}$. 除此之外, 记 $\mathbf{J}(\mathbf{a}) = \{j \in \{1, \dots, m\} : a_j \neq 0\}$ 为向量 $\mathbf{a}$ 的非零坐标集合, $|\mathbf{J}(\mathbf{a})|$ 记为 $\mathbf{J}(\mathbf{a})$ 中元素的个数. 同时记 $\mathcal{M}(\mathbf{a}) = \sum_{j=1}^m I\{a_j \neq 0\}$ 为 $\mathbf{a}$ 中非零元素的个数, 与 $|\mathbf{J}(\mathbf{a})|$ 等价. 定义一个向量 $\mathbf{a}_J \in \mathbb{R}^d$, 在 $\mathbf{J}$ 上, 它的元素与 $\mathbf{a}$ 元素相同; 在 $\mathbf{J}$ 的补集 $\mathbf{J}^C$ 上, 它的元素均为 0. 对实数 c , $\lfloor c \rfloor$ 表示小于 c 的最大整数. 最后记

$$\begin{aligned} h_{(\alpha, \tau)}(\mathbf{x}, u) &= f(\mathbf{x}^T \boldsymbol{\beta} + \mathbf{x}^T \boldsymbol{\delta} I\{u < \tau\}), \\ h_0(\mathbf{x}, u) &= f(\mathbf{x}^T \boldsymbol{\beta}_0 + \mathbf{x}^T \boldsymbol{\delta}_0 I\{u < \tau_0\}), \\ \hat{h}(\mathbf{x}, u) &= f(\mathbf{x}^T \hat{\boldsymbol{\beta}} + \mathbf{x}^T \hat{\boldsymbol{\delta}} I\{u < \hat{\tau}\}), \end{aligned}$$

预测误差记为

$$\|\hat{h} - h_0\|_n = \left\{ \frac{1}{2n} \sum_{i=1}^n [\hat{h}(\mathbf{X}_i, U_i) - h_0(\mathbf{X}_i, U_i)]^2 \right\}^{\frac{1}{2}}.$$

ℓ_1 范数惩罚估计方法 接下来, 本文介绍基于 LASSO 惩罚的方法估计单指标回归模型中的变点. 令 $\boldsymbol{\alpha}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\delta}_0^T)^T$, 模型 (1.1) 可以改写成下式:

$$y_i = f(\mathbf{X}_i(\tau_0)^T \boldsymbol{\alpha}_0) + \epsilon_i, \quad i = 1, \dots, n. \quad (2.1)$$

假设 $\{\epsilon_i : i = 1, \dots, n\}$ 是独立同分布的对称次 Gauss 随机变量, 其均值为 0, 方差为 σ^2 . 令 $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{f} = (f_1, \dots, f_n)^T$, 其中 $f_i = f(\mathbf{X}_i(\tau)^T \boldsymbol{\alpha})$. 残差平方和如下所示:

$$S_n(\boldsymbol{\alpha}, \tau) = \frac{1}{2n} \sum_{i=1}^n [y_i - f(\mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{X}_i^T \boldsymbol{\delta} I\{U_i < \tau\})]^2 = \|\mathbf{y} - \mathbf{f}\|_n^2, \quad (2.2)$$

其中 $\boldsymbol{\alpha} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$. 给定 τ 值, 可以通过求解以下优化问题得到 LASSO 解 $\hat{\boldsymbol{\alpha}}(\tau)$:

$$\hat{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{R}^{2d}}{\text{argmin}} [S_n(\boldsymbol{\alpha}, \tau) + \lambda |\boldsymbol{\alpha}|_1], \quad (2.3)$$

其中 λ 是调节参数, $\mathcal{A}$ 是 $\boldsymbol{\alpha}_0$ 的参数空间. $\hat{\boldsymbol{\alpha}}(\tau)$ 依赖于 τ 的取值, 因而, 变点 τ_0 可由下式估计:

$$\hat{\tau} = \underset{\tau \in \mathbb{T} \subset \mathbb{R}}{\text{argmin}} [S_n(\hat{\boldsymbol{\alpha}}(\tau), \tau) + \lambda |\hat{\boldsymbol{\alpha}}(\tau)|_1], \quad (2.4)$$

其中 $\mathbb{T} = [t_0, t_1]$ 是真实阈值参数 τ_0 的参数空间. 给定 n , 由于 U_i 是离散的, $\mathbf{X}_i I\{U_i < \tau\}$ 在 τ 的一个区间内保持不变, 这意味着 (2.4) 中优化问题的解必定是一个区间. 为简化结果, 本文使用区间的最小值作为估计量. 事实上, 我们可以使用一步最优化同时估计 $\boldsymbol{\alpha}$ 和 τ :

$$(\hat{\boldsymbol{\alpha}}, \hat{\tau}) := \underset{\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{R}^{2d}, \tau \in \mathbb{T} \subset \mathbb{R}}{\text{argmin}} \{S_n(\boldsymbol{\alpha}, \tau) + \lambda |\boldsymbol{\alpha}|_1\}. \quad (2.5)$$

注 2.1 对固定的 τ 值, 由于函数 f 非线性, ℓ_1 范数惩罚的优化问题仍然是非凸的, 因此无法得到全局最优解. 然而, 根据文献 [15] 的结果, 优化问题的任何驻点在适当条件下都具有最优的统计收敛速度.

在优化问题 (2.5) 中, 本文给出 β_0 和 δ_0 的 l_1 范数惩罚, 其中 δ_0 是回归系数的变化量. 事实上, (1.1) 可等价表示为下式:

$$Y_i = \begin{cases} f(\mathbf{X}_i^T \beta_0) + \epsilon_i, & \text{当 } U_i \geq \tau_0 \text{ 时,} \\ f(\mathbf{X}_i^T \beta_1) + \epsilon_i, & \text{当 } U_i < \tau_0 \text{ 时,} \end{cases}$$

其中 $\beta_1 = \beta_0 + \delta_0$. 等价地, 可以对 β_0 和 β_1 施加惩罚. 本文选择对 β_0 和 δ_0 施加惩罚. 当 $\delta_0 = \mathbf{0}$ 时, 模型退化成无变点的单指标模型. 如果 δ 被估计为 $\mathbf{0}$, 则等价于选择了一般的单指标模型.

本文提出近端梯度算法来解决非凸损失函数的 l_1 范数正则化问题, 具体细节在算法 1 中展示. 在算法 1 的第 6 步中, 可以使用 Barzilai-Borwein 谱方法来选择第 t 次迭代的步长 a_t . 特别地, 如果 $t = 0$, 则令 $a_t = 1$, 否则令 $a_t = \langle \delta^{(t)}, g^{(t)} \rangle / \langle \delta^{(t)}, \delta^{(t)} \rangle$ 或者 $a_t = \langle g^{(t)}, g^{(t)} \rangle / \langle \delta^{(t)}, g^{(t)} \rangle$, 其中 $\delta^{(t)} = \alpha^{(t)} - \alpha^{(t-1)}$, $g^{(t)} = \nabla L(\alpha^{(t)}) - \nabla L(\alpha^{(t-1)})$.

算法 1 检测单指标回归模型变点的近端梯度算法

1. **输入:** 样本 $\{(y_i, \mathbf{X}_i, U_i) : i = 1, \dots, n\}$, 调整参数 λ , 更新变量 $\eta > 1$, 常数 $\zeta, a_{\min}, a_{\max}, t_0$ 和 t_1 , 其中 $0 < a_{\min} < 1 < a_{\max}$, $0 < t_0 < t_1 < 1$, 整数 M , 估计精度 ϵ .
 2. **初始化:** 迭代计数器 $t = 0$, 并令 $\alpha^{(0)} \in \mathbb{R}^{2d}$.
 3. **for** τ in $[t_0, t_1]$ **do**
 4. 令 $\mathbf{X}_i(\tau) := (\mathbf{X}_i^T, \mathbf{X}_i^T I\{U_i < \tau\})^T$
 5. **repeat**
 6. 选择适当的迭代步长 a_t .
 7. **repeat**
 8. $\mathbf{u}^{(t)} = \alpha^{(t)} + \frac{1}{na_t} \sum_{i=1}^n [y_i - f(\mathbf{X}_i(\tau) \alpha^{(t)})] f'(\mathbf{X}_i(\tau) \alpha^{(t)}) \mathbf{X}_i(\tau)$.
 9. $\alpha_j^{(t+1)} = \text{soft}(u_j^{(t)}, \frac{\lambda}{a_t})$, 对 $1 \leq j \leq n$.
 10. $a_t = \eta \cdot a_t$.
 11. **until** $\alpha^{(t+1)}$ 满足:
 12. $\phi(\alpha^{(t+1)}) \leq \max \{ \phi(\alpha^{(t+1)}) - \zeta \cdot a_t / 2 \cdot \|\alpha^{(t+1)} - \alpha^{(t)}\|_2^2 : \max(t - M, 0) \leq j \leq t \}$.
 13. 更新迭代计数器 $t = t + 1$.
 14. **until** $\|\alpha^{(t)} - \alpha^{(t-1)}\|_2 / \|\alpha^{(t)}\|_2 \leq \epsilon$
 15. 储存 $\phi(\tau) = \phi(\alpha^{(t)})$, $\hat{\alpha}(\tau) = \alpha^{(t)}$
 16. **end for**
 17. **输出:** $\hat{\tau} = \arg\min_{\tau \in [t_0, t_1]} \phi(\tau)$, $\hat{\alpha} = \hat{\alpha}(\hat{\tau})$.
-

3 理论性质

介绍参数估计量的主要理论性质前, 本文先给出以下的假设.

假设 3.1 (1) 对任意 $\alpha \in \mathcal{A} \subset \mathbb{R}^{2d}$, 假设 $\max_{j=1, \dots, 2d} |\alpha_j| \leq C_1$.

(2) 对任意 $j = 1, \dots, 2d$ 和任意 $\tau \in \mathbb{T}$, 假设 $\|\mathbf{X}^{(j)}(\tau)\|_n \leq C_2$, $\mathbf{X}^{(j)}(\tau)$ 是 $\mathbf{X}(\tau)$ 的第 j 列.

(3) 不存在 $i \neq j$ 使得 $U_i = U_j$, 并假设 $\mathbb{T} = [t_0, t_1]$ 满足 $\min_{i=1, \dots, n} U_i < t_0 < t_1 < \max_{i=1, \dots, n} U_i$.

(4) 函数 f 满足 $f(0) = 0$ 并且连续可微, 且存在 $0 < a < b$, 满足 $f'(x) \in [a, b]$, $\forall x \in \mathbb{R}$.

假设 3.1(1) 要求参数变量的每个分量都是有界的; 假设 3.1(2) 要求每个协变量在 τ 上一致有界; 假设 3.1(3) 假设可能的变点在区间 $[t_0, t_1]$ 内, 这是变点分析的一般假设, 即变点位于远离端点的位置. 除此之外, 假设 3.1(3) 要求 U_i 之间没有结点, 不失一般性, 可以将 U_i 转换为 $U_i = i/n$. 假设 3.1(4) 假设函数 f 是单调并连续可微的.

记 $(\hat{\alpha}, \hat{\tau})$ 为等式 (2.5) 的 LASSO 解, 并且,

$$\lambda = C\sigma\sqrt{\frac{\log 6d}{n}}, \quad (3.1)$$

其中 $C > 2bC_2/\mu$ 和 $\mu \in (0, 1)$ 是固定常数, 我们有如下定理.

定理 3.1 设假设 3.1 成立, 令 μ 为常数, 且 $0 < \mu < 1$, 记 $(\hat{\alpha}, \hat{\tau})$ 为等式 (2.5) 的 LASSO 解, 其中 λ 由等式 (3.1) 给出, 则至少以概率 $1 - (6d)^{1 - \frac{\mu^2 C_2^2}{4b^2 C_2^2}}$ 有

$$\|\hat{h} - h_0\|_n \leq K_1 \sqrt{\lambda \mathcal{M}(\alpha_0)},$$

其中 $K_1 \equiv \sqrt{C_1(4 + 2\mu)}$.

定理 3.1 给出了预测误差的非渐近上界. 假设 $n \rightarrow \infty, d \rightarrow \infty, \lambda \mathcal{M}(\alpha_0) \rightarrow 0$, 则很容易推导出渐近收敛性. 当 $\mathcal{M}(\alpha_0) = o(\sqrt{n/\log d})$ 时, $\lambda \mathcal{M}(\alpha_0) \rightarrow 0$, 这意味着稀疏度 $\mathcal{M}(\alpha_0)$ 可以随着 n 增加. 定理 3.1 中的上界并不是严格的. 与文献 [19] 类似, 本文可以通过一致约束特征值条件, 改进预测误差的非渐近上界以及回归系数的 l_1 范数估计误差的上界.

假设 3.2 给定整数 s ($1 \leq s \leq 2d$)、正数 c_0 和集合 $\mathbb{S} \subset \mathbb{R}$, 我们有下式:

$$\kappa(s, c_0, \mathbb{S}) := \min_{\tau \in \mathbb{S}} \min_{J_0 \subset \{1, \dots, 2d\} | J_0| \leq s} \min_{\gamma \neq 0 | \gamma_{J_0^c}|_1 \leq c_0 | \gamma_{J_0}|_1} \frac{|\mathbf{X}(\tau)\gamma|_2}{\sqrt{n}|\gamma_{J_0}|_2} > 0.$$

上述条件称为一致约束特征值 (uniform restricted eigenvalue, URE) $(s, c_0, \mathbb{S})$ 条件, 文献 [21] 首次介绍了一致约束特征值的条件, 并提出了约束特征值的充分条件. 文献 [19] 给出了更强的条件, 假设约束特征值条件在 τ 上具有一致性. 如果 τ_0 已知, 则假设 3.2 退化为文献 [21] 中的约束特征值假设, 其中 $\mathbb{S} = \{\tau_0\}$.

通过假设 3.2, 本文可以改进定理 3.1 的非渐近上界. 下文将理论结果分为两种情形, 分别为没有变点的情形 ($\delta_0 = 0$) 和存在变点的情形 ($\delta_0 \neq 0$). 首先考虑 $\delta_0 = 0$ 的情形, 这种情形下, 真实模型是没有变点的单指标模型 $y_i = f(\mathbf{X}_i^T \beta_0) + \epsilon_i$.

定理 3.2 给定 $\mu \in (0, 1)$ 和 $\mathcal{M}(\alpha_0) \leq s \leq d$, 假设 3.1 和 3.2 成立, 且 $\kappa = \kappa(s, (1 + \mu)/(1 - \mu), \mathbb{T})$. 此外, 假设 $\delta_0 = 0$. 记 $(\hat{\alpha}, \hat{\tau})$ 为 (2.5) 的 LASSO 解, λ 由 (3.1) 给出, 则至少以概率 $1 - (6d)^{1 - \frac{\mu^2 C_2^2}{4b^2 C_2^2}}$ 有

$$\|\hat{h} - h_0\|_n \leq K_2 \frac{\sigma}{\kappa} \sqrt{\frac{s \log 6d}{n}}, \quad |\hat{\alpha} - \alpha_0|_1 \leq K_2 \frac{\sigma s}{a^2 \kappa^2} \sqrt{\frac{\log 6d}{n}}, \quad \mathcal{M}(\hat{\alpha}) \leq K_2 \frac{b^2 \phi_{\max} s}{a^2 \kappa^2}, \quad (3.2)$$

其中 $K_2 > 0$.

上述定理分别给出了系数 α 的均方预测误差和 l_1 范数下估计的非渐近上界. 值得注意的是, 与定理 3.1 中的预测误差收敛速度 $s\sqrt{\log d/n}$ 相比, 上述定理的收敛速度 ($\sqrt{s \log d/n}$) 更快. 根据文献 [19, 定理 2] 可以看出, 我们的方法与线性模型的 LASSO 估计量具有相同的 Oracle 不等式, 这意味着收敛速度是最优的, 即使 f 是恒等函数, 收敛速度也不会提高. 不同之处在于, 等式 (3.2) 的收敛速度包含了 f' 的下界 a , 这是很直观的结果. 我们可以考虑 $a = 0$ 的极端情形, 即 f 是常数函数. 这时观察值不提供关于 α 的信息, 不可能一致估计 α . 另一种情形是 α 非常大, 这会提高信噪比使 α 容易估计.

接下来讨论存在一个变点的情形. 除上述假设外, 我们还需要给出更多正则化条件.

假设 3.3 假设 $s \geq \mathcal{M}(\alpha_0)$, 对任意满足 $|\tau - \tau_0| > \theta \geq \min_i |U_i - \tau_0|$ 和 $\alpha \in \{\alpha : \mathcal{M}(\alpha) \leq s\}$ 的 θ 和 τ , 存在常数 $c > 0$ 满足

$$\|h_{(\alpha, \tau)} - h_0\|_n^2 > c\theta.$$

假设 3.4 对任意 $\theta \in (0, 1)$, 存在常数 $C < \infty$ 满足

$$\sup_j \sup_{|\tau - \tau_0| < \theta} \frac{1}{n} \sum_{i=1}^n |X_{ij}|^2 |I(U_i < \tau_0) - I(U_i < \tau)| \leq C\theta.$$

假设 3.5 定义 $M_n^2(\theta)$ 如下:

$$M_n^2(\theta) = \frac{1}{n\theta} \sum_{i=a_n}^{b_n} (\mathbf{X}_i^T \boldsymbol{\delta}_0)^2,$$

其中 $a_n = \min\{1, \lfloor n(\tau_0 - \theta) \rfloor\}$, $b_n = \max\{\lfloor n(\tau_0 + \theta) \rfloor, n\}$. 假设 $\theta_0 = \frac{1}{n} \vee K_1 \sqrt{\lambda \mathcal{M}(\alpha_0)}$, 则对任意 $\theta \in [1/n, \theta_0]$, $M_n^2(\theta)$ 是有界的.

假设 3.6 以下不等式成立:

$$\frac{2 + \mu}{1 - \mu} < \frac{6cs}{\kappa^2 a^2}, \quad (3.3)$$

$$\frac{36(1 + \mu)\lambda|\boldsymbol{\delta}_0|_1 s}{(1 - \mu)^2 \kappa^2} < 1, \quad (3.4)$$

$$\left(\frac{1}{1 - \frac{36(1 + \mu)\lambda|\boldsymbol{\delta}_0|_1 s}{(1 - \mu)^2 \kappa^2}} \right)^2 < \frac{24cs(2 + \mu)}{\kappa^2 a^2 (1 - \mu)}, \quad (3.5)$$

$$\frac{\lambda a^2 |\boldsymbol{\delta}_0|_1}{(1 - \mu)c} \left(\frac{1}{1 - \frac{36(1 + \mu)\lambda|\boldsymbol{\delta}_0|_1 s}{(1 - \mu)^2 \kappa^2}} \right)^2 < 1. \quad (3.6)$$

假设 3.3 给出了 τ_0 的可识别条件, 这意味着给定 $s \geq \mathcal{M}(\alpha_0)$, 对任意 $\alpha \in \{\alpha : \mathcal{M}(\alpha) \leq s\}$ 和 τ 并且 $(\alpha, \tau) \neq (\alpha_0, \tau_0)$, 都有 $\|h_{(\alpha, \tau)} - h_0\|_n \neq 0$. 假设 3.3 的条件比一般可识别性条件更强. 当 τ 远离 τ_0 时, 假设 3.3 给出了 h 的偏离率, 它会影响估计误差 $\hat{\tau}$ 的上界. 假设 3.4 刻画了协变量的平滑性, 这个假设要求不能有太多 U_i 堆积在 τ_0 附近, 并且要求在变点附近给定 U_i 后, $\mathbf{X}_i$ 的期望是有界的. 若 $(\mathbf{X}_i, Q_i)$ 是随机向量, 则当如下标准假设条件满足时假设 3.4 成立: Q_i 是连续分布, 并对任意 j , 在 τ_0 附近, $E(|\mathbf{X}_i^{(j)}|^2 | U_i = \tau)$ 是连续有界的. 假设 3.5 在某种程度上来说是固定设计矩阵二阶矩的弱化条件. 由于 $\boldsymbol{\delta}_0 \neq 0$, 结合此假设和假设 3.2 可知, 对任意 $\theta \in [1/n, \theta_0]$, $M_n^2(\theta)$ 是有界的且远离 0. 假设 3.6 中的 4 个不等式虽然形式复杂, 但很容易验证. 特别地, 当 $s\lambda|\boldsymbol{\delta}_0|_1 \rightarrow 0$ 且 $\lambda|\boldsymbol{\delta}_0|_1 \rightarrow 0$ 时, 假设 3.6 中的条件能同时满足, 也就是说 s 和 $|\boldsymbol{\delta}_0|_1$ 可以趋近于无穷. 以下定理给出各估计量误差的非渐近上界.

定理 3.3 若假设 3.1–3.6 成立, 并且在定理 3.2 中, 对于 $0 < \mu < 1$, 有 $\mathbb{S} = \{|\tau - \tau_0| \leq \theta_0\}$, $\mathcal{M}(\alpha_0) \leq s \leq d$, $\kappa = \kappa(s, \frac{2+\mu}{1-\mu}, \mathbb{S})$. 记 $(\hat{\alpha}, \hat{\tau})$ 为 (2.5) 的 LASSO 解, 其中 λ 由 (3.1) 给出. 给定正数 C_3 和 C_4 , 至少以概率 $1 - (6d)^{1 - \frac{\mu^2 C_2^2}{4b^2 C_2^2}} - C_3(6d)^{-C_4}$ 有

$$\begin{aligned} \|\hat{h} - h_0\|_n &\leq K_3 \frac{\sigma}{\kappa a} \sqrt{\frac{s \log 6d}{n}}, \quad |\hat{\alpha} - \alpha_0|_1 \leq K_3 \frac{\sigma s}{\kappa^2 a^2} \sqrt{\frac{\log 6d}{n}}, \\ |\hat{\tau} - \tau_0| &\leq K_3 \frac{\sigma^2 s}{\kappa^2 a^2} \frac{\log 6d}{n}, \quad \mathcal{M}(\hat{\alpha}) \leq K_3 \frac{b^2 \phi_{\max} s}{\kappa^2 a^2}, \end{aligned} \quad (3.7)$$

其中 $K_3 > 0$.

定理 3.3 给出了预测误差、回归系数的 l_1 范数误差和系数稀疏度的非渐近上界. 给定常数 σ, κ, a 和 b , 定理 3.3 中估计量的收敛速度与定理 3.2 中给出的收敛速度相同. 此外, 这个定理给出了变点估计 $\hat{\tau}$ 误差的非渐近上界. 只要 $\text{slog } d/n \rightarrow 0$, $\hat{\tau}$ 可以渐近收敛到 τ_0 . 定理 3.3 是对文献 [19] 的拓展. 与经典线性回归不同, 由于 f 非线性, 因此, 本文的优化问题是非凸的. 尽管存在非凸性, 本文提出的估计量可以在一定条件下与线性模型具有相同的统计收敛速度, 这也是本文的主要贡献.

4 模拟研究

本节介绍我们的方法分别在线性与非线性回归两种情形下的表现. 在非线性回归中, 我们将本文提出的基于 LASSO 惩罚的方法 (记为 “LASSO”) 与 “Inverting” 方法作比较. “Inverting” 方法首先将数据通过非线性函数求逆, 然后转化为线性模型的方法进行估计. 我们采用的非线性函数为 $f(x) = 2x + \cos(x)$, 它的导数以 $a = 1$ 和 $b = 3$ 为界. 此外, 本文也给出了 “Post-LASSO” 方法和两个 Oracle 方法 (“Oracle 1” 和 “Oracle 2”) 的结果. “Post-LASSO” 方法第一步利用 “LASSO” 方法选择变量, 然后对第一步选择的模型使用普通最小二乘法 (ordinary least squares, OLS), 这是提高 “LASSO” 方法估计精度的常用步骤. “Oracle 1” 假设非零系数的回归变量是已知的. “Oracle 2” 假设真实变点 τ_0 和非零系数的回归变量均是已知的.

从模型 (1.1) 中独立抽取样本 $\mathbf{X}_i \sim N(\mathbf{0}, \Sigma)$, $\epsilon_i \sim N(0, 0.5^2)$, 其中 $\Sigma \in \mathbb{R}^{d \times d}$ 是 Toeplitz 矩阵, $\Sigma_{jk} = \rho^{|j-k|}$, $U_i = i/n$. 模拟分为 $\mathbf{X}_i$ 的分量相互独立和相关两种情形, 分别对应 $\rho = 0$ 和 $\rho = 0.5$. 回归系数设置为 $\beta_0 = (1, 0, 1, 0, \dots, 0)$, $\delta_0 = c(0, -1, 1, 0, \dots, 0)$, 其中 $c = 0, 0.5, 1$. 值得注意的是, $c = 0$ 对应没有变点存在的情形, 这时, Oracle 1 与 Oracle 2 相同, 合并记为 Oracle. 此外, 样本量 $n = 200$, 阈值参数为 $\tau_0 = 0.4$, τ 的取值区间为 $\mathbb{T} = [0.15, 0.85]$. 等式 (3.1) 中的常数设为 $C = 2b \max_j \|\mathbf{X}^{(j)}\|_n$, 由此确定调节参数 λ . $\mathbf{X}_i$ 的维数 d 分别设为 100、200 和 300.

我们用均方预测误差、 $\mathcal{M}(\hat{\alpha})$ 均值以及 $\hat{\alpha}$ 和 τ 的 l_1 范数估计误差来比较不同方法的表现. 结果基于 200 次数值模拟得出. 在均方预测误差的计算上, 每次模拟, 我们再产生 400 个新样本 $\{(y_j, \mathbf{X}_j, U_j), j = 1, \dots, 400\}$, 在得到估计量 $\hat{\alpha}$ 和 $\hat{\tau}$ 后, 按下式计算预测误差:

$$\|\hat{h} - h_0\|_n^2 = \frac{1}{400} \sum_{i=1}^{400} [h_0(\mathbf{X}_j, U_j) - \hat{h}(\mathbf{X}_j, U_j)]^2.$$

表 1–3 分别是变量维数 100、200 和 300 时的模拟结果. 从表 1–3 的结果可以看出, 本文所提的 “LASSO” 方法在所有情形中均比 “Inverting” 方法表现得好. 当变量维数 d 增加时, 虽然所有方法的表现都在变差, 但本文所提方法表现得更加稳定, 并且变点位置的估计误差几乎控制在 0.05 内. 当 $c = 0.5$ 时, “LASSO” 方法在 $\rho = 0.5$ 的情形中表现不好, 但它比 “Inverting” 方法性能好. 原因在于信噪比太小几乎无法检测到变点. 当协变量的相关性变大时, 本文所提的 “LASSO” 方法无法恢复非零系数的支持集, 因而, “Post-LASSO” 方法表现不好. 但当协变量相互独立时, “Post-LASSO” 方法性能较好. 此外, “LASSO” 方法几乎可以正确地选出具有非零系数的变量, 因而, “Post-LASSO” 方法的表现几乎与 Oracle 方法的相同.

图 1 展示了不同 τ_0 和维数 d 下变点估计误差和回归系数估计误差. 从图 1 可以看出, 对于不同的 τ_0 和 d , 我们提出的 “LASSO” 方法和 “Post-LASSO” 方法均有较好的表现. 从图 1(a)–1(c) 可以看出, “Post-LASSO” 方法与 Oracle 方法表现很接近. 从图 1(d)–1(f) 看, $\hat{\alpha}$ 的 l_1 范数估计误差随着维数增加而缓慢增加. 尽管如此, “LASSO” 方法比 “Inverting” 方法表现更好, “Post-LASSO” 方法与

表 1 $\tau_0 = 0.4, d = 100, n = 200, m = 400$, 基于 200 次模拟的结果

		$\rho = 0.5$					$\rho = 0$				
跳跃比	估计方法	$\rho = 0.5$					$\rho = 0$				
		$\ \hat{h} - h_0\ _n^2$	$E[\mathcal{M}(\hat{\alpha})]$	$E \hat{\alpha} - \alpha_0 _1$	$E \hat{\tau} - \tau_0 _1$	$\ \hat{h} - h_0\ _n^2$	$E[\mathcal{M}(\hat{\alpha})]$	$E \hat{\alpha} - \alpha_0 _1$	$E \hat{\tau} - \tau_0 _1$		
单指标模型的模拟结果	$c = 1.0$	Inverting	0.86	4.11	1.69	0.01	0.41	4.48	0.80	0.00	0.00
		LASSO	0.42	4.62	1.12	0.00	0.20	4.55	0.55	0.00	0.00
		Post-LASSO	0.02	4.62	0.11	0.00	0.01	4.55	0.10	0.00	0.00
		Oracle 1	0.02	4.00	0.09	0.00	0.01	4.00	0.08	0.00	0.00
		Oracle 2	0.01	4.00	0.09	0.00	0.01	4.00	0.08	0.00	0.00
	$c = 0.5$	Inverting	1.33	2.64	1.17	0.26	0.45	4.37	0.79	0.03	0.01
		LASSO	0.91	3.72	1.06	0.12	0.20	4.52	0.55	0.01	0.00
		Post-LASSO	0.22	3.72	0.63	0.05	0.01	4.53	0.10	0.00	0.00
		Oracle 1	0.01	4.00	0.09	0.01	0.01	4.00	0.08	0.00	0.00
		Oracle 2	0.01	4.00	0.09	0.00	0.01	4.00	0.08	0.00	0.00
	$c = 0.0$	Inverting	0.18	2.18	0.20	—	0.16	2.04	0.24	—	—
		LASSO	0.30	2.29	0.14	—	0.19	2.09	0.16	—	—
线性回归模型的模拟结果		Post-LASSO	0.18	2.29	0.07	—	0.06	2.09	0.05	—	—
		Oracle	0.00	2.00	0.03	—	0.00	2.00	0.03	—	—
	$c = 1.0$	LASSO	0.22	4.79	1.71	0.04	0.10	4.77	0.83	0.00	0.00
		Post-LASSO	0.04	4.79	0.45	0.01	0.01	4.77	0.23	0.00	0.00
		Oracle 1	0.01	4.00	0.19	0.01	0.01	4.00	0.17	0.00	0.00
		Oracle 2	0.01	4.00	0.19	0.00	0.01	4.00	0.17	0.00	0.00
	$c = 0.5$	LASSO	0.35	3.16	1.17	0.26	0.12	4.59	0.81	0.05	0.01
		Post-LASSO	0.14	3.16	1.04	0.19	0.01	4.59	0.27	0.01	0.01
		Oracle 1	0.01	4.00	0.20	0.02	0.01	4.00	0.17	0.01	0.01
		Oracle 2	0.01	4.00	0.19	0.00	0.01	4.00	0.17	0.00	0.00
	$c = 0.0$	LASSO	0.18	2.70	0.24	—	0.14	2.40	0.26	—	—
		Post-LASSO	0.13	2.70	0.22	—	0.11	2.40	0.19	—	—
		Oracle	0.00	2.00	0.06	—	0.00	2.00	0.06	—	—

表 2 $\tau_0 = 0.4, d = 200, n = 200, m = 400$ 基于 200 次循环的模拟结果

		$\rho = 0.5$					$\rho = 0$				
跳跃比	估计方法	$\ \hat{h} - h_0\ _n^2$	$E[\mathcal{M}(\hat{\alpha})]$	$E \hat{\alpha} - \alpha_0 _1$	$E \hat{\tau} - \tau_0 _1$	$\ \hat{h} - h_0\ _n^2$	$E[\mathcal{M}(\hat{\alpha})]$	$E \hat{\alpha} - \alpha_0 _1$	$E \hat{\tau} - \tau_0 _1$		
单指标模型的模拟结果	$c = 1.0$										
	Inverting	1.09	4.06	1.85	0.03	0.49	4.51	0.86	0.00		
	LASSO	0.51	5.29	1.30	0.01	0.24	4.57	0.60	0.00		
	Post	0.03	5.29	0.15	0.00	0.02	4.57	0.10	0.00		
	Oracle 1	0.02	4.00	0.09	0.00	0.02	4.00	0.08	0.00		
	Oracle 2	0.00	4.00	0.09	0.00	0.01	4.00	0.08	0.00		
	$c = 0.5$										
	Inverting	1.51	2.57	1.19	0.29	0.57	4.36	0.86	0.04		
	LASSO	1.19	3.60	1.10	0.17	0.24	4.58	0.60	0.00		
	Post	0.27	3.60	0.77	0.07	0.01	4.58	0.10	0.00		
线性回归模型的模拟结果	$c = 1.0$										
	Oracle 1	0.01	4.00	0.09	0.01	0.01	4.00	0.08	0.00		
	Oracle 2	0.00	4.00	0.09	0.00	0.01	4.00	0.08	0.00		
	$c = 0.0$										
	Inverting	0.15	2.03	0.27	—	0.15	2.12	0.22	—		
	LASSO	0.26	2.21	0.15	—	0.16	2.06	0.18	—		
	Post	0.20	2.21	0.05	—	0.12	2.06	0.04	—		
	Oracle 1&2	0.00	2.00	0.03	—	0.00	2.00	0.03	—		
	LASSO	0.28	5.08	1.88	0.06	0.12	4.88	0.90	0.00		
	Post	0.08	5.08	0.67	0.02	0.01	4.88	0.24	0.00		
	$c = 0.5$										
	Oracle 1	0.01	4.00	0.19	0.01	0.01	4.00	0.17	0.00		
	Oracle 2	0.01	4.00	0.19	0.00	0.01	4.00	0.17	0.00		
	LASSO	0.42	3.03	1.19	0.30	0.18	4.61	0.88	0.08		
	Post	0.13	3.03	1.01	0.19	0.02	4.61	0.32	0.01		
	Oracle 1	0.01	4.00	0.20	0.02	0.01	4.00	0.17	0.01		
	Oracle 2	0.01	4.00	0.19	0.00	0.01	4.00	0.17	0.00		
	LASSO	0.18	2.58	0.25	—	0.14	2.35	0.29	—		
	Post	0.12	2.58	0.20	—	0.10	2.35	0.19	—		
	Oracle 1&2	0.00	2.00	0.06	—	0.00	2.00	0.06	—		

表 3 $\tau_0 = 0.4, d = 300, n = 200, m = 400$ 基于 200 次循环的模拟结果

单指标模型的模拟结果	跳跃比	估计方法	$\rho = 0.5$					$\rho = 0$				
			$\ \hat{h} - h_0\ _n^2$		$E[\mathcal{M}(\hat{\alpha})]$		$E \hat{\alpha} - \alpha_0 _1$	$\ \hat{h} - h_0\ _n^2$		$E[\mathcal{M}(\hat{\alpha})]$		$E \hat{\alpha} - \alpha_0 _1$
			$c = 1.0$	Inverting	1.16	4.16		$c = 0.5$	Inverting	1.90	0.04	
	$c = 1.0$	LASSO	0.54	5.58	1.35	0.01	0.04	0.55	4.73	0.63	0.00	0.00
		Post	0.03	5.58	0.15	0.00	0.00	0.02	4.73	0.11	0.00	0.00
		Oracle 1	0.02	4.00	0.09	0.00	0.00	0.02	4.00	0.08	0.00	0.00
		Oracle 2	0.01	4.00	0.09	0.00	0.00	0.01	4.00	0.08	0.00	0.00
		Inverting	1.38	2.49	1.20	0.29	0.29	0.63	4.36	0.90	0.04	0.04
	$c = 0.5$	LASSO	1.38	3.76	1.12	0.20	0.20	0.29	4.68	0.64	0.01	0.01
		Post	0.34	3.76	0.87	0.08	0.08	0.02	4.68	0.11	0.00	0.00
		Oracle 1	0.01	4.00	0.09	0.01	0.01	0.01	4.00	0.08	0.00	0.00
		Oracle 2	0.01	4.00	0.09	0.00	0.00	0.01	4.00	0.08	0.00	0.00
		Inverting	0.14	2.1	0.22	-	-	0.17	2.05	0.29	-	-
线性回归模型的模拟结果	$c = 1.0$	LASSO	0.17	2.19	0.15	-	-	0.23	2.09	0.19	-	-
		Post	0.11	2.19	0.04	-	-	0.18	2.09	0.05	-	-
		Oracle 1&2	0.00	2.00	0.03	-	-	0.00	2.00	0.03	-	-
		LASSO	0.30	5.56	1.93	0.07	0.07	0.13	5.00	0.95	0.01	0.01
		Post	0.11	5.56	0.86	0.02	0.02	0.01	5.00	0.26	0.00	0.00
	$c = 0.5$	Oracle 1	0.01	4.00	0.18	0.01	0.01	0.01	4.00	0.18	0.00	0.00
		Oracle 2	0.01	4.00	0.18	0.01	0.01	0.01	4.00	0.18	0.00	0.00
		LASSO	0.42	3.17	1.20	0.31	0.31	0.18	4.69	0.93	0.08	0.08
		Post	0.15	3.17	1.09	0.21	0.21	0.03	4.69	0.37	0.02	0.02
		Oracle 1	0.01	4.00	0.19	0.02	0.02	0.01	4.00	0.18	0.01	0.01
	$c = 0.0$	Oracle 2	0.01	4.00	0.19	0.00	0.00	0.01	4.00	0.18	0.00	0.00
		LASSO	0.15	2.54	0.25	-	-	0.11	2.34	0.31	-	-
		Post	0.13	2.54	0.22	-	-	0.07	2.34	0.19	-	-
		Oracle 1&2	0.00	2.00	0.06	-	-	0.00	2.00	0.06	-	-

图 1 (a)–(c) 变点估计误差; (d)–(f) 回归系数的估计误差. $\square$ 表示 “Inverting” 方法; $\triangle$ 表示 “LASSO” 方法; $\times$ 代表 “Post-LASSO”; $\diamond$ 表示 “Oracle 1” 方法; $\blacksquare$ 表示 “Oracle 2” 方法

Oracle 方法表现相差无几. 简单总结一下, 模拟结果验证了我们所提方法的理论结果, 并且方法在不同维数、不同相关性、线性或非线性回归中均有较强的稳健性.

5 总结

本文考虑了可能存在变点的高维单指标模型, 其中变点由协变量阈值产生. 本文提出 l_1 范数惩罚方法用来估计回归系数和阈值参数. 在一定的稀疏条件下, 我们得到了预测误差和回归系数的 l_1 范数估计误差的非渐近性质. 模拟研究展示了所提方法的优势. 在本文中, 单指标模型的连接函数假定已知, 在实际应用中可能受到限制. 我们将会进一步研究连接函数未知的情形, 例如, 首先使用核密度方法估计连接函数, 然后运用本文所提的算法进行估计. 其中涉及的理论分析更为复杂, 因而将此问题留作未来研究方向.

参考文献

- 1 Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc Ser B Stat Methodol*, 1996, 58: 267–288
- 2 Bühlmann P, van de Geer S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York: Springer, 2011
- 3 Meier L, van de Geer S, Bühlmann P. High-dimensional additive modeling. *Ann Statist*, 2009, 37: 3779–3821
- 4 Ravikumar P, Lafferty J, Liu H, et al. Sparse additive models. *J R Stat Soc Ser B Stat Methodol*, 2009, 71: 1009–1030
- 5 Zhu L P, Zhu L X. Nonconcave penalized inverse regression in single-index models with high dimensional predictors. *J Multivariate Anal*, 2009, 100: 862–875

- 6 Foster J C, Taylor J M G, Nan B. Variable selection in monotone single-index models via the adaptive LASSO. *Stat Med*, 2013, 32: 3944–3954
- 7 Luo S, Ghosal S. Forward selection and estimation in high dimensional single index models. *Stat Methodol*, 2016, 33: 172–179
- 8 Radchenko P. High dimensional single index models. *J Multivariate Anal*, 2015, 139: 266–282
- 9 He Y, Zhang X, Zhang L. Variable selection for high dimensional Gaussian copula regression model: An adaptive hypothesis testing procedure. *Comput Statist Data Anal*, 2018, 124: 132–150
- 10 Cai T T, Zhang L. High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Statist Sinica*, 2018, 28: 963–993
- 11 Xue L, Wang Q. Empirical likelihood for single-index varying-coefficient models. *Bernoulli*, 2012, 18: 836–856
- 12 Chang Z, Xue L, Zhu L. On an asymptotically more efficient estimation of the single-index model. *J Multivariate Anal*, 2010, 101: 1898–1901
- 13 Pang Z, Xue L. Estimation for the single-index models with random effects. *Comput Stat Data Anal*, 2012, 56: 1837–1853
- 14 Li G, Peng H, Dong K, et al. Simultaneous confidence bands and hypothesis testing for single-index models. *Statist Sinica*, 2014, 24: 937–955
- 15 Yang Z, Wang Z, Liu L, et al. Sparse nonlinear regression: Parameter estimation under nonconvexity. In: *Proceedings of the 33th International Conference on Machine Learning*. New York: World Scientific Publishing, 2016, 2472–2481
- 16 Wu Y. Simultaneous change point analysis and variable selection in a regression problem. *J Multivariate Anal*, 2008, 99: 2154–2171
- 17 Chan N H, Yau C Y, Zhang R M. Group LASSO for structural break time series. *J Amer Statist Assoc*, 2014, 109: 590–599
- 18 Li D, Qian J, Su L. Panel data models with interactive fixed effects and multiple structural breaks. *J Amer Statist Assoc*, 2016, 111: 1804–1819
- 19 Lee S, Seo M H, Shin Y. The LASSO for high dimensional regression with a possible change point. *J R Stat Soc Ser B Stat Methodol*, 2016, 78: 193–210
- 20 Jin B S, Dong C L, Tan C C, et al. Estimator of a change point in single index models. *Sci China Math*, 2014, 57: 1701–1712
- 21 Bickel P J, Ritov Y, Tsybakov A B. Simultaneous analysis of LASSO and Dantzig selector. *Ann Statist*, 2009, 37: 1705–1732
- 22 Vaart A W V D, Wellner J A. *Weak Convergence and Empirical Processes*. New York: Springer, 1996

附录 A 定理的证明

下面给出主要定理的证明. 为了证明定理 3.1, 附录 A.1 首先给出引理 A.1 和 A.2. 附录 A.2 给出定理 3.2 的证明, 其中引理 A.3 和 A.4 发挥了重要作用. 附录 A.3 展示定理 3.3 的证明. 由于定理 3.3 的证明更加复杂, 因而, 我们先逐步证明引理 A.5–A.9.

附录 A.1 定理 3.1 的证明

根据中值定理可知, 存在 $\{\xi_i : i = 1, \dots, n\}$ 和 $\{\eta_i : i = 1, \dots, n\}$ 满足

$$\begin{aligned} f(\mathbf{X}_i^T(\hat{\tau})\hat{\alpha}) - f(\mathbf{X}_i^T(\tau_0)\alpha_0) &= f'(\xi_i)(\mathbf{X}_i^T(\hat{\tau})\hat{\alpha} - \mathbf{X}_i^T(\tau_0)\alpha_0), \\ f(\mathbf{X}_i^T(\tau)\alpha) - f(\mathbf{X}_i^T(\tau_0)\alpha_0) &= f'(\eta_i)(\mathbf{X}_i^T(\tau)\alpha - \mathbf{X}_i^T(\tau_0)\alpha_0), \end{aligned}$$

其中 ξ_i 位于 $\mathbf{X}_i^T(\hat{\tau})\hat{\alpha}$ 与 $\mathbf{X}_i^T(\tau_0)\alpha_0$ 之间, η_i 位于 $\mathbf{X}_i^T(\tau)\alpha$ 与 $\mathbf{X}_i^T(\tau_0)\alpha_0$ 之间. 令

$$\mathbf{V}_1 = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \epsilon_i \quad \text{和} \quad \mathbf{V}_2(\tau) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i I\{U_i < \tau\} \epsilon_i.$$

对于常数 $\mu \in (0, 1)$, 定义事件

$$\mathbb{A} = \left\{ |\mathbf{V}_1|_\infty \leq \frac{\mu\lambda}{b} \right\} \quad \text{和} \quad \mathbb{B} = \left\{ \sup_{\tau \in \mathbb{T}} |\mathbf{V}_2(\tau)|_\infty \leq \frac{\mu\lambda}{b} \right\}.$$

进一步定义 $\mathbf{J}_0 = \mathbf{J}(\boldsymbol{\alpha}_0)$ 和 $\mathbf{R}_n = \mathbf{R}_n(\boldsymbol{\alpha}_0, \tau_0)$, 其中

$$\mathbf{R}_n(\boldsymbol{\alpha}, \tau) := \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \boldsymbol{\delta} \{I(U_i < \hat{\tau}) - I(U_i < \tau)\}.$$

第一个引理给出预测误差和系数误差的一个基本不等式, 这个引理是确定三个定理中估计误差上界的基础.

引理 A.1 若 $\mathbb{A}$ 和 $\mathbb{B}$ 成立, 则有

$$\|\hat{h} - h_0\|_n^2 + (1 - \mu)\lambda|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|_1 \leq 2\lambda|(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{\mathbf{J}_0}|_1 + \mathbf{R}_n. \quad (\text{A.1})$$

证明 由 (2.5) 可得, 对所有 $\boldsymbol{\alpha} \in \mathbb{R}^{2d}$ 和 $\tau \in \mathbb{T}$, 有

$$S_n(\hat{\boldsymbol{\alpha}}, \hat{\tau}) + \lambda|\hat{\boldsymbol{\alpha}}|_1 \leq S_n(\boldsymbol{\alpha}, \tau) + \lambda|\boldsymbol{\alpha}|_1. \quad (\text{A.2})$$

通过 (2.2) 中 $S_n(\boldsymbol{\alpha}, \tau)$ 的定义可得

$$\begin{aligned} S_n(\hat{\boldsymbol{\alpha}}, \hat{\tau}) - S_n(\boldsymbol{\alpha}, \tau) &= \frac{1}{2n} \|\mathbf{y} - f(\mathbf{X}(\hat{\tau})\hat{\boldsymbol{\alpha}})\|_2^2 - \frac{1}{2n} \|\mathbf{y} - f(\mathbf{X}(\tau)\boldsymbol{\alpha})\|_2^2 \\ &= \|\hat{h} - h_0\|_n^2 - \|h_{(\boldsymbol{\alpha}, \tau)} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i [f(\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}}) - f(\mathbf{X}_i^T(\tau)\boldsymbol{\alpha})]. \end{aligned}$$

定义 $T_n(\boldsymbol{\alpha}, \tau) = \frac{1}{n} \sum_{i=1}^n \epsilon_i [f'(\eta_i) - f'(\xi_i)][\mathbf{X}_i^T(\tau)\boldsymbol{\alpha} - \mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0]$, 我们有

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \epsilon_i [f(\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}}) - f(\mathbf{X}_i^T(\tau)\boldsymbol{\alpha})] \\ &= \frac{1}{n} \sum_{i=1}^n \epsilon_i \{f'(\xi_i)[\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}} - \mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0] - f'(\eta_i)[\mathbf{X}_i^T(\tau)\boldsymbol{\alpha} - \mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0]\} \\ &= \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i)[\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}} - \mathbf{X}_i^T(\tau)\boldsymbol{\alpha}] - \frac{1}{n} \sum_{i=1}^n \epsilon_i [f'(\eta_i) - f'(\xi_i)][\mathbf{X}_i^T(\tau)\boldsymbol{\alpha} - \mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0] \\ &= \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i)[\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}} - \mathbf{X}_i^T(\tau)\boldsymbol{\alpha}] - T_n(\boldsymbol{\alpha}, \tau). \end{aligned}$$

因此,

$$\begin{aligned} &S_n(\hat{\boldsymbol{\alpha}}, \hat{\tau}) - S_n(\boldsymbol{\alpha}, \tau) \\ &= \|\hat{h} - h_0\|_n^2 - \|h_{(\boldsymbol{\alpha}, \tau)} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i)[\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}} - \mathbf{X}_i^T(\tau)\boldsymbol{\alpha}] + T_n(\boldsymbol{\alpha}, \tau) \\ &= \|\hat{h} - h_0\|_n^2 - \|h_{(\boldsymbol{\alpha}, \tau)} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) [\mathbf{X}_i^T \hat{\boldsymbol{\delta}} I(U_i < \hat{\tau}) \\ &\quad - \mathbf{X}_i^T \boldsymbol{\delta} I(U_i < \tau)] + T_n(\boldsymbol{\alpha}, \tau) \end{aligned}$$

$$\begin{aligned}
 &= \|\hat{h} - h_0\|_n^2 - \|h_{(\alpha, \tau)} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\beta} - \beta) - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\delta} - \delta) I(U_i < \hat{\tau}) \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta [I(U_i < \hat{\tau}) - I(U_i < \tau)] + T_n(\alpha, \tau).
 \end{aligned}$$

将上述等式与 (A.2) 结合, 可得

$$\begin{aligned}
 \|\hat{h} - h_0\|_n^2 &\leq \|h_{(\alpha, \tau)} - h_0\|_n^2 - \lambda |\hat{\alpha}|_1 + \lambda |\alpha|_1 + \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\beta} - \beta) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta [I(U_i < \hat{\tau}) - I(U_i < \tau)] - T_n(\alpha, \tau) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\delta} - \delta) I(U_i < \hat{\tau}).
 \end{aligned} \tag{A.3}$$

事件 $\mathbb{A}$ 和 $\mathbb{B}$ 成立时, 我们有

$$\frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\beta} - \beta) \leq \mu \lambda \sum_{j=1}^d (\hat{\beta}_j - \beta_j)$$

和

$$\frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\delta} - \delta) I(U_i < \hat{\tau}) \leq \mu \lambda \sum_{j=1}^d (\hat{\delta}_j - \delta_j).$$

于是, (A.3) 可以写成

$$\|\hat{h} - h_0\|_n^2 \leq \|h_{(\alpha, \tau)} - h_0\|_n^2 - \lambda |\hat{\alpha}|_1 + \lambda |\alpha|_1 + \mu \lambda |\hat{\alpha} - \alpha|_1 + \mathbf{R}_n(\alpha, \tau) - T_n(\alpha, \tau), \tag{A.4}$$

其中

$$|\hat{\alpha}_j - \alpha_{0j}| + |\alpha_{0j}| - |\hat{\alpha}_j| = 0, \quad j \notin \mathbf{J}_0. \tag{A.5}$$

令 $(\alpha, \tau) = (\alpha_0, \tau_0)$, 则有 $T_n(\alpha_0, \tau_0) = 0$, 因而, 在事件 $\mathbb{A}$ 和 $\mathbb{B}$ 下, 可得

$$\begin{aligned}
 \|\hat{h} - h_0\|_n^2 + (1 - \mu) \lambda |\hat{\alpha} - \alpha_0|_1 &\leq \lambda (|\hat{\alpha} - \alpha_0|_1 + |\alpha_0|_1 - |\hat{\alpha}|_1) + \mathbf{R}_n(\alpha_0, \tau_0) \\
 &= \lambda (|(\hat{\alpha} - \alpha_0)_{\mathbf{J}_0}|_1 + |\alpha_{0\mathbf{J}_0}|_1 - |\hat{\alpha}_{\mathbf{J}_0}|_1) + \mathbf{R}_n \\
 &\leq 2\lambda |(\hat{\alpha} - \alpha_0)_{\mathbf{J}_0}|_1 + \mathbf{R}_n.
 \end{aligned}$$

(A.1) 得证. □

引理 A.2 证明了事件 $\mathbb{A} \cap \mathbb{B}$ 的概率趋近于 1.

引理 A.2 假设 $\{\epsilon_i : i = 1, \dots, n\}$ 是独立同分布的对称亚 Gauss 随机变量, 其均值为 0, 代方差为 σ^2 , 则

$$\mathbb{P}(\mathbb{A} \cap \mathbb{B}) \geq 1 - 6d \exp \left\{ -\frac{n\mu^2\lambda^2}{4\sigma^2 b^2 C_2^2} \right\}. \tag{A.6}$$

证明 若 $\{\epsilon_i : i = 1, \dots, n\}$ 是独立同分布的对称亚 Gauss 随机变量, 其均值为 0, 代方差为 σ^2 , 则 $V_{1j} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{ij} \epsilon_i$ 是中心亚 Gauss 随机变量, 其代方差的上界为

$$\frac{\sigma^2}{n^2} \sum_{i=1}^n \mathbf{X}_{ij}^2 \leq \frac{2\sigma^2}{n} \left(\frac{1}{2n} \sum_{i=1}^n \mathbf{X}_{ij}^2 \right) \leq \frac{2\sigma^2 C_2^2}{n}.$$

通过次 Gauss 随机变量的方差定义可知,

$$\mathbb{P}\left(|V_{1j}| > \sqrt{\frac{2}{n}} \sigma C_2 t\right) \leq 2 \exp\left(-\frac{t^2}{2}\right), \quad \forall t > 0. \quad (\text{A.7})$$

令 $t = \frac{\mu\lambda}{b} \sqrt{\frac{n}{2\sigma^2 C_2^2}}$, 则有

$$\mathbb{P}\left(|V_{1j}| > \frac{\mu\lambda}{b}\right) \leq 2 \exp\left(-\frac{n\mu^2 \lambda^2}{4\sigma^2 b^2 C_2^2}\right).$$

因而,

$$\mathbb{P}(\mathbb{A}^C) = \mathbb{P}\left(|V_1|_\infty > \frac{\mu\lambda}{b}\right) \leq \sum_{j=1}^d \mathbb{P}\left(|V_{1j}| > \frac{\mu\lambda}{b}\right) \leq 2d \exp\left(-\frac{n\mu^2 \lambda^2}{4\sigma^2 b^2 C_2^2}\right). \quad (\text{A.8})$$

接下来考虑事件 $\mathbb{B}$. 根据假设 3.1(3) 可知, U_i 之间没有结点. 因此, 根据 U_i 的大小重新对 i 排序, $\sum_{i=1}^n \mathbf{X}_{ij} I\{U_i < \tau\} \epsilon_i$ 为部分和. 若 ϵ_i 是独立且对称的亚 Gauss 随机变量, 则 $\mathbf{X}_{ij} \epsilon_i$ 也是独立且对称的亚 Gauss 变量. 因而, 根据 Lévy 不等式 (参见文献 [22, 第 431 页]), 我们有

$$\begin{aligned} \mathbb{P}\left(\sup_{\tau \in \mathbb{T}} |V_{2j}(\tau)| > \frac{\mu\lambda}{b}\right) &\leq \mathbb{P}\left(\sup_{k \leq n} \left| \frac{1}{n} \sum_{i=1}^k f'(\xi_i) \mathbf{X}_{ij} \epsilon_i \right| > \frac{\mu\lambda}{b}\right) \\ &\leq 2\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^k f'(\xi_i) \mathbf{X}_{ij} \epsilon_i \right| > \frac{\mu\lambda}{b}\right) \\ &\leq 4 \exp\left(-\frac{n\mu^2 \lambda^2}{4\sigma^2 b^2 C_2^2}\right). \end{aligned}$$

因此,

$$\mathbb{P}(\mathbb{B}^C) = \mathbb{P}\left(\sup_{\tau \in \mathbb{T}} |V_2(\tau)|_\infty > \mu\lambda\right) \leq \sum_{j=1}^d \mathbb{P}\left(\sup_{\tau \in \mathbb{T}} |V_{2j}(\tau)| > \mu\lambda\right) \leq 4d \exp\left(-\frac{n\mu^2 \lambda^2}{4\sigma^2 b^2 C_2^2}\right). \quad (\text{A.9})$$

结合 (A.8) 和 (A.9) 可得

$$\mathbb{P}(\mathbb{A} \cap \mathbb{B}) \geq 1 - \mathbb{P}(\mathbb{A}^C) - \mathbb{P}(\mathbb{B}^C) \geq 1 - 6d \exp\left(-\frac{n\mu^2 \lambda^2}{4\sigma^2 b^2 C_2^2}\right).$$

证毕. □

定理 3.1 的证明 根据引理 A.1, 在 $\mathbb{A} \cap \mathbb{B}$ 成立时, 我们有

$$\|\hat{h} - h_0\|_n^2 \leq 2\lambda |(\hat{\alpha} - \alpha_0)_{J_0}|_1 + \mathbf{R}_n,$$

其中

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta_0 \{I(U_i < \hat{\tau}) - I(U_i < \tau_0)\} = \sum_{j=1}^d \delta_{0j} [V_{2j}(\hat{\tau}) - V_{2j}(\tau_0)].$$

因此, 根据事件 $\mathbb{B}$ 可得

$$|\mathbf{R}_n| \leq 2\mu\lambda|\boldsymbol{\delta}_0|_1. \quad (\text{A.10})$$

此外, 根据假设 3.1 可得

$$|\boldsymbol{\delta}_0|_1 \leq \alpha_{\max}\mathcal{M}(\boldsymbol{\alpha}_0) \leq C_1\mathcal{M}(\boldsymbol{\alpha}_0)$$

和

$$|(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0}|_1 \leq 2\alpha_{\max}\mathcal{M}(\boldsymbol{\alpha}_0) \leq 2C_1\mathcal{M}(\boldsymbol{\alpha}_0).$$

结合以上式子和引理 A.2, 定理 3.1 得证. $\square$

附录 A.2 定理 3.2 的证明

本节给出定理 3.2 的证明. 首先给出一个引理导出关于 LASSO 估计量 $\hat{\boldsymbol{\alpha}}$ 稀疏性的 Oracle 不等式.

引理 A.3 若假设 3.1(4) 成立且事件 $\mathbb{A} \cap \mathbb{B}$ 成立, 则有

$$\mathcal{M}(\hat{\boldsymbol{\alpha}}) \leq \frac{2b^2\phi_{\max}}{(1-\mu)^2\lambda^2} \|\hat{\mathbf{h}} - \mathbf{h}_0\|_n^2. \quad (\text{A.11})$$

证明 根据 KKT (Karush-Kuhn-Tucker) 条件, (2.3) 中的 LASSO 解 $\hat{\boldsymbol{\alpha}}(\tau)$ 满足

$$\nabla L(\hat{\boldsymbol{\alpha}}(\tau)) + \lambda \partial|\hat{\boldsymbol{\alpha}}(\tau)|_1 = 0,$$

其中 $\nabla L(\boldsymbol{\alpha}(\tau))$ 是 $L(\boldsymbol{\alpha}(\tau)) := S_n(\boldsymbol{\alpha}(\tau), \tau)$ 的梯度, $\partial|\boldsymbol{\alpha}(\tau)|_1$ 是 $|\boldsymbol{\alpha}(\tau)|_1$ 的次梯度. 上述的表达式也可以写成如下形式:

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} = \lambda \text{sign}(\hat{\beta}_j(\tau)), & \text{若 } \hat{\beta}_j(\tau) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} \leq \lambda, & \text{若 } \hat{\beta}_j(\tau) = 0, \\ \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} = \lambda \text{sign}(\hat{\delta}_j(\tau)), & \text{若 } \hat{\delta}_j(\tau) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} \leq \lambda, & \text{若 } \hat{\delta}_j(\tau) = 0. \end{cases}$$

事件 $\mathbb{A} \cap \mathbb{B}$ 成立时, 我们有

$$\left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} \right| \leq \mu\lambda \quad \text{和} \quad \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} \right| \leq \mu\lambda. \quad (\text{A.12})$$

对任意 $\tau \in \mathbb{T}$ 和任意 $j \in \{1, \dots, d\}$. 结合上述式子可得

$$\begin{cases} \left| \frac{1}{n} \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} \right| \geq (1-\mu)\lambda, & \text{若 } \hat{\beta}_j(\tau) \neq 0, \\ \left| \frac{1}{n} \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} \right| \geq (1-\mu)\lambda, & \text{若 } \hat{\delta}_j(\tau) \neq 0. \end{cases}$$

令 F 为 $n \times 1$ 维的向量, 它的第 i 列元素为 $[f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)]f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))$, 则

$$\begin{aligned} \frac{1}{n^2} F^T \mathbf{X}(\hat{\tau}) \mathbf{X}^T(\hat{\tau}) F &= \frac{1}{n^2} \sum_{j=1}^d \left\{ \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} \right\}^2 \\ &\quad + \frac{1}{n^2} \sum_{j=1}^d \left\{ \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} \right\}^2 \\ &\geq \frac{1}{n^2} \sum_{j: \hat{\beta}_j(\tau) \neq 0} \left\{ \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} \right\}^2 \\ &\quad + \frac{1}{n^2} \sum_{j: \hat{\delta}_j(\tau) \neq 0} \left\{ \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)] f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)) X_{ij} I\{U_i < \tau\} \right\}^2 \\ &\geq (1-\mu)^2 \lambda^2 \left(\sum_{j: \hat{\beta}_j(\tau) \neq 0} 1 + \sum_{j: \hat{\delta}_j(\tau) \neq 0} 1 \right) \\ &\geq (1-\mu)^2 \lambda^2 \mathcal{M}(\hat{\boldsymbol{\alpha}}). \end{aligned}$$

将 $\mathbf{X}(\hat{\tau}) \mathbf{X}^T(\hat{\tau})/n$ 的最大特征值记为 $\phi_{\max}$. 注意到

$$\begin{aligned} \frac{1}{n^2} F^T \mathbf{X}(\hat{\tau}) \mathbf{X}^T(\hat{\tau}) F &\leq \phi_{\max} \frac{1}{n} F^T F \leq 2\phi_{\max} \frac{1}{2n} \sum_{i=1}^n [f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0) - \mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau)]^2 [f'(\mathbf{X}_i^T(\tau)\hat{\boldsymbol{\alpha}}(\tau))]^2 \\ &\leq 2\phi_{\max} b^2 \|\hat{h} - h_0\|_n^2. \end{aligned}$$

结合上述式子, 引理得证. $\square$

引理 A.4 若 $\delta_0 = 0$, 假设 3.2 成立, 且 $\mu \in (0, 1)$, $\mathcal{M}(\boldsymbol{\alpha}_0) \leq s \leq d$, $\kappa = \kappa(s, \frac{1+\mu}{1-\mu}, \mathbb{T})$. 记 $(\hat{\boldsymbol{\alpha}}, \hat{\tau})$ 为 (2.5) 的 LASSO 解, 其中 λ 由 (3.1) 给出, 则至少以概率 $1 - (6d)^{1-\frac{\mu^2 C^2}{4b^2 C_2^2}}$ 有

$$\|\hat{h} - h_0\|_n \leq \frac{2\sqrt{2}C\sigma}{a\kappa} \sqrt{\frac{s \log 6d}{n}}, \quad |\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|_1 \leq \frac{8C\sigma s}{(1-\mu)\kappa^2 a^2} \sqrt{\frac{\log 6d}{n}}, \quad \mathcal{M}(\hat{\boldsymbol{\alpha}}) \leq \frac{16b^2 \phi_{\max} s}{(1-\mu)^2 a^2 \kappa^2}.$$

证明 根据引理 A.1 可得

$$\|\hat{h} - h_0\|_n^2 + (1-\mu)\lambda |\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|_1 \leq 2\lambda |(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0}|_1 + \mathbf{R}_n,$$

其中 $\delta_0 = 0$ 意味着 $R_n = 0$, 因而,

$$(1-\mu)\lambda |\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|_1 \leq 2\lambda |(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0}|_1,$$

等价于

$$|(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0^c}|_1 \leq \frac{1+\mu}{1-\mu} |(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0}|_1.$$

根据假设 3.2 可得 $\kappa = \kappa(s, \frac{1+\mu}{1-\mu}, \mathbb{T}) > 0$. 若 $\delta_0 = 0$, 则 $\mathbf{X}(\hat{\tau})\boldsymbol{\alpha}_0 = \mathbf{X}(\tau_0)\boldsymbol{\alpha}_0$, 可得

$$\begin{aligned} \kappa^2 |(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)_{J_0}|_2^2 &\leq \frac{1}{n} |\mathbf{X}(\hat{\tau})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)|_2^2 = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i^T(\hat{\tau})\hat{\boldsymbol{\alpha}} - \mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{[f'(\xi_i)]^2} [f(\mathbf{X}_i^T(\hat{\tau})\boldsymbol{\alpha}) - f(\mathbf{X}_i^T(\tau_0)\boldsymbol{\alpha}_0)]^2 \right\} \end{aligned}$$

$$\leq \frac{2}{a^2} \|\hat{h} - h_0\|_n^2. \quad (\text{A.13})$$

结合引理 A.1 和 (A.13) 可得

$$\|\hat{h} - h_0\|_n^2 \leq 2\lambda |(\hat{\alpha} - \alpha_0)_{J_0}|_1 \leq 2\lambda\sqrt{s} |(\hat{\alpha} - \alpha_0)_{J_0}|_2 \leq \frac{2\sqrt{2}\lambda\sqrt{s}}{a\kappa} \|\hat{h} - h_0\|_n.$$

根据 (3.1) 定义的 λ 可得

$$\|\hat{h} - h_0\|_n \leq \frac{2\sqrt{2}\lambda\sqrt{s}}{a\kappa} = \frac{2\sqrt{2}C\sigma}{a\kappa} \sqrt{\frac{s \log 6d}{n}}.$$

引理的第一个结论得证.

结合上述式子可得

$$\begin{aligned} |\hat{\alpha} - \alpha_0|_1 &\leq |(\hat{\alpha} - \alpha_0)_{J_0}|_1 + |(\hat{\alpha} - \alpha_0)_{J_0^c}|_1 \leq |(\hat{\alpha} - \alpha_0)_{J_0}|_1 + \frac{1+\mu}{1-\mu} |(\hat{\alpha} - \alpha_0)_{J_0}|_1 \\ &\leq \frac{2\sqrt{s}}{1-\mu} |(\hat{\alpha} - \alpha_0)_{J_0}|_2 \leq \frac{2\sqrt{2}\sqrt{s}}{(1-\mu)a\kappa} \|\hat{h} - h_0\|_n \\ &\leq \frac{8C\sigma s}{(1-\mu)\kappa^2 a^2} \sqrt{\frac{\log 6d}{n}}. \end{aligned}$$

引理的第二个结论得证.

最后, 根据引理 A.4, 可得

$$\mathcal{M}(\hat{\alpha}) \leq \frac{2b^2\phi_{\max}}{(1-\mu)^2\lambda^2}, \quad \|\hat{h} - h_0\|_n^2 \leq \frac{16b^2\phi_{\max}s}{(1-\mu)^2a^2\kappa^2}.$$

因此, 在事件 $\mathbb{A} \cap \mathbb{B}$ 成立时, 引理的第三个结论得证. $\square$

定理 3.2 的证明 根据引理 A.4, 令

$$K_2 = \max\left(2\sqrt{2}C, \frac{8}{1-\mu}, \frac{16}{(1-\mu)^2}\right),$$

定理得证. $\square$

附录 A.3 定理 3.3 的证明

不失一般性, 假设 $U_i = i/n$. 对于常数 $\theta > 0$, 定义事件

$$\mathbb{C}(\theta) = \left\{ \sup_{|\tau - \tau_0| < \theta} \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i \mathbf{X}_i^T \boldsymbol{\delta}_0 [I(U_i < \hat{\tau}) - I(U_i < \tau)] \right| \leq \frac{\lambda\sqrt{\theta}}{b} \right\}.$$

下一个引理给出事件 $\mathbb{A} \cap \mathbb{B} \cap [\bigcap_{j=1}^m \mathbb{C}(\theta_j)]$ 的概率下界.

引理 A.5 假设 $\{\epsilon_i : i = 1, \dots, n\}$ 的定义与引理 A.2 相同. 给定 m 和正数 $\theta_1, \dots, \theta_m$, 且 $M_n(\theta_i) > 0$, $i = 1, \dots, m$, 其中 $M_n(\theta)$ 定义如假设 3.5 所示, 则有

$$\mathbb{P}\left(\mathbb{A} \cap \mathbb{B} \cap \left[\bigcap_{j=1}^m \mathbb{C}(\theta_j)\right]\right) \geq 1 - 6d \exp\left\{-\frac{n\mu^2\lambda^2}{4\sigma^2b^2C_2^2}\right\} - 4 \sum_{j=1}^m \exp\left\{-\frac{n\lambda^2}{M_n^2(\theta_j)b^2\sigma^2}\right\}.$$

证明 如引理 A.2 所示, 利用 Lévy 不等式, 我们有

$$\begin{aligned} P(\mathbb{C}(\theta_j)^C) &\leq P\left(\sup_{|\tau-\tau_0|<\theta} \left|\frac{1}{n} \sum_{i=1}^n \epsilon_i \mathbf{X}_i^T \boldsymbol{\delta}_0 [I(U_i < \hat{\tau}) - I(U_i < \tau)]\right| > \frac{\lambda\sqrt{\theta}}{b}\right) \\ &\leq 2P\left(\left|\frac{1}{n} \sum_{i=[n(\tau_0-\theta_j)]}^{[n(\tau_0+\theta_j)]} \epsilon_i \mathbf{X}_i^T \boldsymbol{\delta}_0\right| > \frac{\lambda\sqrt{\theta}}{b}\right). \end{aligned}$$

若 $\{\epsilon_i : i = 1, \dots, n\}$ 是独立同分布的亚 Gauss 随机变量, 其均值为 0, 代方差为 σ^2 , 则 $\frac{1}{n} \sum_{i=a_n}^{b_n} \mathbf{X}_i^T \boldsymbol{\delta}_0 \epsilon_i$ 是中心化的亚 Gauss 随机变量, 其代方差的上界为

$$\frac{\sigma^2}{n^2} \sum_{i=a_n}^{b_n} [\mathbf{X}_i^T \boldsymbol{\delta}_0]^2 \leq \frac{M_n^2(\theta_j) \sigma^2 \theta_j}{n},$$

其中 $a_n = [n(\tau_0 - \theta_j)]$, $b_n = [n(\tau_0 + \theta_j)]$. 因此, 根据 (A.7) 可得

$$P\left(\left|\frac{1}{n} \sum_{i=a_n}^{b_n} \epsilon_i \mathbf{X}_i^T \boldsymbol{\delta}_0\right| > \frac{\lambda\sqrt{\theta_j}}{b}\right) = 2 \exp\left\{-\frac{n\lambda^2}{M_n^2(\theta_j)b^2\sigma^2}\right\}.$$

因而,

$$P(\mathbb{C}(\theta_j)^C) \leq 4 \exp\left\{-\frac{n\lambda^2}{M_n^2(\theta_j)b^2\sigma^2}\right\}.$$

结合引理 A.2 和此结果, 引理得证. $\square$

引理 A.6 若假设 3.3 的条件满足. 令

$$\theta^* = \max\left\{\min_i |U_i - \tau_0|, c^{-1}\lambda(4\alpha_{\max}\mathcal{M}(\boldsymbol{\alpha}_0) + 2\mu|\boldsymbol{\delta}_0|_1)\right\},$$

其中 c 是假设 3.3 中所定义的常数. 在事件 $\mathbb{A}$ 和 $\mathbb{B}$ 成立时, 有

$$|\hat{\tau} - \tau_0| \leq \theta^*.$$

证明 根据 (A.12), 在事件 $\mathbb{A} \cap \mathbb{B}$ 成立时, 有

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\leq \mu\lambda|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|_1, \\ \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) I(U_i < \hat{\tau}) &\leq \mu\lambda|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0|_1. \end{aligned}$$

结合上述不等式和引理 A.1 的证明, 可得

$$\begin{aligned} &S_n(\hat{\boldsymbol{\alpha}}, \hat{\tau}) + \lambda|\hat{\boldsymbol{\alpha}}|_1 - [S_n(\boldsymbol{\alpha}_0, \tau_0) + \lambda|\boldsymbol{\alpha}_0|_1] \\ &= \|\hat{h} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) I(U_i < \hat{\tau}) - \lambda(|\boldsymbol{\alpha}_0|_1 - |\hat{\boldsymbol{\alpha}}|_1) - R_n \\ &\geq \|\hat{h} - h_0\|_n^2 - \mu\lambda|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|_1 - \lambda(|\boldsymbol{\alpha}_0|_1 - |\hat{\boldsymbol{\alpha}}|_1) - R_n \end{aligned}$$

$$\begin{aligned}
&\geq \|\hat{h} - h_0\|_n^2 - 2\lambda|(\hat{\alpha} - \alpha_0)_{J_0}|_1 - R_n \\
&\geq \|\hat{h} - h_0\|_n^2 - [4\lambda\alpha_{\max}\mathcal{M}(\alpha_0) + 2\mu\lambda|\delta_0|_1],
\end{aligned}$$

其中最后一个不等式来自 (A.10). 如果 $|\hat{\tau} - \tau_0| > \theta^*$, 根据假设 3.3 可得

$$S_n(\hat{\alpha}, \hat{\tau}) + \lambda|\hat{\alpha}|_1 - [S_n(\alpha_0, \tau_0) + \lambda|\alpha_0|_1] > 0.$$

这与 (2.4) 相矛盾. 因此定理得证. $\square$

引理 A.7 假设 $|\hat{\tau} - \tau_0| \leq \theta_\tau$, $|\hat{\alpha} - \alpha_0| \leq \theta_\alpha$, 进一步假设 3.4 和 3.2 的条件成立, 其中 $\mathbb{S} = \{|\tau - \tau_0| \leq \theta_\tau\}$, $\mathcal{M}(\alpha_0) \leq s \leq d$, $\kappa = \kappa(s, \frac{2+\mu}{1-\mu}, \mathbb{S})$, $0 < \mu < 1$. 因而, 在 $\mathbb{A}$ 、 $\mathbb{B}$ 和 $\mathbb{C}(\theta_\tau)$ 成立时, 可得

$$\begin{aligned}
\|\hat{h} - h_0\|_n^2 &\leq 3\lambda \left\{ \frac{12\lambda s}{\kappa^2 a^2} \vee \frac{2}{\kappa} \sqrt{C\theta_\alpha \theta_\tau |\delta_0|_1 s} \vee \sqrt{\theta_\tau} \right\}, \\
|\hat{\alpha} - \alpha_0|_1 &\leq \frac{3}{1-\mu} \left\{ \frac{12\lambda s}{\kappa^2 a^2} \vee \frac{2}{\kappa} \sqrt{C\theta_\alpha \theta_\tau |\delta_0|_1 s} \vee \sqrt{\theta_\tau} \right\}.
\end{aligned}$$

证明 在事件 $\mathbb{C}(\theta_\tau)$ 成立时, 有

$$|R_n| = \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta_0 \{I(U_i < \hat{\tau}) - I(U_i < \tau_0)\} \right| \leq \lambda \sqrt{\theta_\tau}.$$

首先考虑 $|(\hat{\alpha} - \alpha_0)_{J_0}|_1 \leq \sqrt{\theta_\tau}$ 的情形. 根据引理 A.1, 在 $\mathbb{A} \cap \mathbb{B}$ 成立时, 可得

$$\|\hat{h} - h_0\|_n^2 \leq 2\lambda|(\hat{\alpha} - \alpha_0)_{J_0}|_1 + R_n \leq 2\lambda\sqrt{\theta_\tau} + \lambda\sqrt{\theta_\tau} = 3\lambda\sqrt{\theta_\tau} \quad (\text{A.14})$$

和

$$|\hat{\alpha} - \alpha_0|_1 \leq \frac{2}{1-\mu}|(\hat{\alpha} - \alpha_0)_{J_0}|_1 + \frac{R_n}{(1-\mu)\lambda} \leq \frac{2\sqrt{\theta_\tau}}{1-\mu} + \frac{\sqrt{\theta_\tau}}{1-\mu} = \frac{3\sqrt{\theta_\tau}}{1-\mu}. \quad (\text{A.15})$$

接下来考虑 $|(\hat{\alpha} - \alpha_0)_{J_0}|_1 > \sqrt{\theta_\tau}$ 的情形. 结合引理 A.1, 在事件 $\mathbb{A}$ 和 $\mathbb{B}$ 成立时, 有

$$\|\hat{h} - h_0\|_n^2 + (1-\mu)\lambda|\hat{\alpha} - \alpha_0|_1 \leq 3\lambda|(\hat{\alpha} - \alpha_0)_{J_0}|_1, \quad (\text{A.16})$$

即

$$|(\hat{\alpha} - \alpha_0)_{J_0^c}|_1 \leq \frac{2+\mu}{1-\mu}|(\hat{\alpha} - \alpha_0)_{J_0}|_1.$$

与引理 A.4 的证明相同, 若假设 3.2 满足 $\text{URE}(s, \frac{2+\mu}{1-\mu}, \mathbb{S})$, 则当 $\delta_0 \neq 0$ 时,

$$\begin{aligned}
\kappa^2|(\hat{\alpha} - \alpha_0)_{J_0}|_2^2 &\leq |\mathbf{X}(\hat{\tau})(\hat{\alpha} - \alpha_0)|_2^2 \\
&\leq \frac{1}{n} \sum_{i=1}^n \{ \mathbf{X}_i^T \hat{\beta} + \mathbf{X}_i^T I(U_i < \hat{\tau}) \hat{\delta} - \mathbf{X}_i^T \beta_0 - \mathbf{X}_i^T I(U_i < \tau_0) \delta_0 \\
&\quad + \mathbf{X}_i^T \delta_0 [I(U_i < \tau_0) - I(U_i < \hat{\tau})] \} \\
&\leq \frac{2}{a^2} \|\hat{h} - h_0\|_n^2 + \frac{1}{n} \sum_{i=1}^n \{ 2\mathbf{X}_i^T(\hat{\tau})(\hat{\alpha} - \alpha_0) \mathbf{X}_i^T \delta_0 [I(U_i < \tau_0) - I(U_i < \hat{\tau})] \} \\
&\leq \frac{2}{a^2} \|\hat{h} - h_0\|_n^2 + 2\theta_\alpha |\delta_0|_1 \sup_j \frac{1}{n} \sum_{i=1}^n |X_{ij}|^2 |I(U_i < \tau_0) - I(U_i < \hat{\tau})|
\end{aligned}$$

$$\leq \frac{2}{a^2} \|\hat{h} - h_0\|_n^2 + 2C\theta_\alpha\theta_\tau|\delta_0|_1,$$

其中第三个不等式由不等式 $(a+b)^2 \leq a^2 + 2(a+b)b$ 可得, 最后一个不等式由假设 3.4 可得. 结合 (A.16) 与上述结果可得

$$\|\hat{h} - h_0\|_n^2 \leq 3\lambda|(\hat{\alpha} - \alpha_0)_{J_0}|_1 \leq 3\lambda\sqrt{s}|(\hat{\alpha} - \alpha_0)_{J_0}|_2 \leq \frac{3\lambda\sqrt{s}}{\kappa} \left(\frac{2}{a^2} \|\hat{h} - h_0\|_n^2 + 2C\theta_\alpha\theta_\tau|\delta_0|_1 \right)^{1/2},$$

其中 $a + b \leq 2a \vee 2b$, 我们有

$$\|\hat{h} - h_0\|_n^2 \leq \frac{36\lambda^2 s}{\kappa^2 a^2} \vee \frac{6\lambda}{\kappa} \sqrt{C\theta_\alpha\theta_\tau|\delta_0|_1 s}. \quad (\text{A.17})$$

结合 (A.14) 与上述上界, 我们得出引理 A.7 的第一个结论. 类似引理 A.4 的证明, 我们有

$$\begin{aligned} |\hat{\alpha} - \alpha_0|_1 &\leq \frac{3}{1-\mu}|(\hat{\alpha} - \alpha_0)_{J_0}|_1 \leq \frac{3}{1-\mu}\sqrt{s}|(\hat{\alpha} - \alpha_0)_{J_0}|_2 \\ &\leq \frac{3\sqrt{s}}{(1-\mu)\kappa} \left(\frac{2}{a^2} \|\hat{h} - h_0\|_n^2 + 2C\theta_\alpha\theta_\tau|\delta_0|_1 \right)^{1/2}. \end{aligned}$$

结合 $a + b \leq 2a \vee 2b$, 可得

$$|\hat{\alpha} - \alpha_0|_1 \leq \frac{36\lambda s}{(1-\mu)\kappa^2 a^2} \vee \frac{6}{(1-\mu)\kappa} \sqrt{C\theta_\alpha\theta_\tau|\delta_0|_1 s}. \quad (\text{A.18})$$

结合 (A.15) 与 (A.18), 我们得出引理 A.7 的第二个结论. $\square$

如果我们结合引理 A.6 和 A.7 的结果, 则下面的引理可以进一步收紧 $|\hat{\tau} - \tau_0|$ 的上界.

引理 A.8 假设 $|\hat{\tau} - \tau_0| \leq \theta_\tau$, $|\hat{\alpha} - \alpha_0| \leq \theta_\alpha$, 进一步假设 3.3 成立, 令 $\tilde{\theta} = \frac{\lambda}{c}[(1+\mu)\theta_\alpha + \sqrt{\theta_\tau}]$, 则 $\mathbb{A}$ 、 $\mathbb{B}$ 和 $\mathbb{C}(\theta_\tau)$ 成立时,

$$|\hat{\tau} - \tau_0| \leq \tilde{\theta}.$$

证明 根据引理 A.1 的证明可得

$$\begin{aligned} S_n(\hat{\alpha}, \hat{\tau}) - S_n(\alpha_0, \tau_0) &= \|\hat{h} - h_0\|_n^2 - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\beta} - \beta_0) - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\delta} - \delta_0) I(U_i < \hat{\tau}) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta_0 [I(U_i < \hat{\tau}) - I(U_i < \tau_0)]. \end{aligned}$$

在 $\mathbb{A}$ 、 $\mathbb{B}$ 和 $\mathbb{C}(\theta_\tau)$ 成立时, 可得

$$\begin{aligned} &\left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\beta} - \beta_0) + \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T (\hat{\delta} - \delta_0) I(U_i < \hat{\tau}) \right| \\ &\leq \mu\lambda|\hat{\beta} - \beta_0|_1 + \mu\lambda|\hat{\delta} - \delta_0|_1 \leq \mu\lambda|\hat{\alpha} - \alpha_0|_1 \leq \mu\lambda\theta_\alpha \end{aligned}$$

和

$$\left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f'(\xi_i) \mathbf{X}_i^T \delta_0 [I(U_i < \hat{\tau}) - I(U_i < \tau_0)] \right| \leq \lambda\sqrt{\theta_\tau}.$$

若 $\tilde{\theta} < |\hat{\tau} - \tau_0| < \theta_\tau$, 根据假设 3.3, 可得

$$S_n(\hat{\alpha}, \hat{\tau}) + \lambda|\hat{\alpha}|_1 - [S_n(\alpha_0, \tau_0) + \lambda|\alpha_0|_1] \geq \|\hat{h} - h_0\|_n^2 - \mu\lambda\theta_\alpha - \lambda\sqrt{\theta_\tau} - \lambda|\alpha_0 - \hat{\alpha}|_1$$

$$\geq c\tilde{\theta} - (1 + \mu)\lambda\theta_{\alpha} - \lambda\sqrt{\theta_{\tau}}.$$

若 $\tilde{\theta} = \frac{\lambda}{c}[(1 + \mu)\theta_{\alpha} + \sqrt{\theta_{\tau}}]$, 则

$$S_n(\hat{\alpha}, \hat{\tau}) + \lambda|\hat{\alpha}|_1 - [S_n(\alpha_0, \tau_0) + \lambda|\alpha_0|_1] \geq 0.$$

这与 (2.5) 矛盾. 因此结果得证. $\square$

引理 A.7 中的上界是 θ_{τ} 和 θ_{α} 的函数. 因此可以迭代应用引理 A.7 和 A.8 来收紧上界. 引理 A.9 证明当样本量增加时, 迭代过程最终可以导出一个与 θ_{τ} 和 θ_{α} 无关的上界.

引理 A.9 若假设 3.1–3.6 成立, 并且假设 3.2 中的 $\mathbb{S} = \{|\tau - \tau_0| \leq \theta^*\}$, $\mathcal{M}(\alpha_0) \leq s \leq d$, $\kappa = \kappa(s, \frac{2+\mu}{1-\mu}, \mathbb{S})$, $0 < \mu < 1$. 记 $(\hat{\alpha}, \hat{\tau})$ 为 (2.5) 的 LASSO 解, 其中 λ 由 (3.1) 给出, 则对有限整数 t^* , 存在一组常数 $\theta_1, \dots, \theta_{t^*}$, 使得对 $t = 1, \dots, t^*$, $M_n(\theta_t) > 0$, 则至少以概率

$$1 - 6d \exp \left\{ -\frac{n\mu^2\lambda^2}{4\sigma^2b^2C_2^2} \right\} - 4 \sum_{t=1}^{t^*} \exp \left\{ -\frac{n\lambda^2}{M_n^2(\theta_t)b^2\sigma^2} \right\},$$

我们有

$$\begin{aligned} \|\hat{h} - h_0\|_n &\leq \frac{6C\sigma}{\kappa a} \sqrt{\frac{s \log 6d}{n}}, \quad |\hat{\alpha} - \alpha_0|_1 \leq \frac{36C\sigma s}{(1-\mu)a^2\kappa^2} \sqrt{\frac{\log 6d}{n}}, \\ |\hat{\tau} - \tau_0| &\leq \frac{24(2+\mu)C^2\sigma^2s \log 6d}{(1-\mu)(ca^2\kappa^2)n}, \quad \mathcal{M}(\hat{\alpha}) \leq \frac{72b^2\phi_{\max}s}{(1-\mu)^2a^2\kappa^2}. \end{aligned}$$

证明 定义 θ_{α}^* 和 θ_{τ}^* 如下:

$$\theta_{\alpha}^* = \frac{36\lambda s}{(1-\mu)a^2\kappa^2} \quad \text{和} \quad \theta_{\tau}^* = \frac{24(2+\mu)}{1-\mu} \frac{\lambda^2 s}{ca^2\kappa^2}.$$

假设

$$\frac{2}{\kappa} \sqrt{C\theta_{\alpha}\theta_{\tau}|\delta_0|_1 s} \vee \sqrt{\theta_{\tau}} \leq \frac{12\lambda s}{\kappa^2 a^2}. \quad (\text{A.19})$$

也就是说引理 A.7 中 $\|\hat{h} - h_0\|_n$ 和 $|\hat{\alpha} - \alpha_0|_1$ 的上界不是 θ_{α} 和 θ_{τ} 的函数. 根据引理 A.7 可知, $|\hat{\alpha} - \alpha_0|_1$ 上界为 θ_{α} . 此外,

$$\|\hat{h} - h_0\|_n^2 \leq \frac{36\lambda^2 s}{\kappa^2 a^2}. \quad (\text{A.20})$$

结合 (A.19) 和引理 A.8, 其中 $\theta_{\alpha} = \theta_{\alpha}^*$, 可得

$$|\hat{\tau} - \tau_0| \leq \frac{\lambda}{c}[(1 + \mu)\theta_{\alpha}^* + \sqrt{\theta_{\tau}}] \leq \frac{\lambda}{c} \left[(1 + \mu) \frac{36\lambda s}{(1-\mu)\kappa^2 a^2} + \frac{12\lambda s}{\kappa^2 a^2} \right] = \frac{24(2+\mu)}{1-\mu} \frac{\lambda^2 s}{ca^2\kappa^2},$$

即 θ_{τ}^* . 如果 (A.19) 不成立, 我们仍旧可以证明, 在有限步迭代应用引理 A.7 和 A.8 后, 可以得到收敛的上界, 并且比 θ_{α}^* 和 θ_{τ}^* 小. 结合引理 A.1 和定理 3.1, 有

$$(1 - \mu)\lambda|\hat{\alpha} - \alpha_0|_1 \leq 2\lambda|(\hat{\alpha} - \alpha_0)_{J_0}|_1 + R_n \leq 4\lambda\alpha_{\max}\mathcal{M}(\alpha_0) + 2\lambda\alpha_{\max}\mathcal{M}(\alpha_0) \leq 6\lambda\alpha_{\max}s.$$

因此,

$$|\hat{\alpha} - \alpha_0|_1 \leq \frac{6\alpha_{\max}s}{1-\mu}.$$

类似地, 根据引理 A.6 可得

$$|\hat{\tau} - \tau_0| \leq \frac{c}{\lambda}(4\alpha_{\max}\mathcal{M}(\alpha_0) + 2\mu|\delta_0|_1) \leq \frac{c}{\lambda}(4\alpha_{\max}s + 2\mu\alpha_{\max}s) \leq \frac{6\lambda\alpha_{\max}s}{c}.$$

设迭代的初始值为

$$\theta_{\alpha}^{(1)} = \frac{6\alpha_{\max}s}{1-\mu}, \quad \theta_{\tau}^{(1)} = \frac{6\lambda\alpha_{\max}s}{c}.$$

如果 $\theta_{\alpha}^{(1)}$ 和 $\theta_{\tau}^{(1)}$ 比 θ_{α}^* 和 θ_{τ}^* 小, 那么不必进行迭代. 否则, 令 $\theta_{\alpha}^{(t)}$ 和 $\theta_{\tau}^{(t)}$ 分别表示 $|\hat{\alpha} - \alpha_0|_1$ 和 $|\hat{\tau} - \tau_0|$ 在第 t 轮迭代时的上界. 当 (A.19) 不成立时, 我们分两种情形计算不动点.

情形 (i)

$$\theta_{\alpha}^{(t)} = \frac{3}{1-\mu}\sqrt{\theta_{\tau}^{(t-1)}}.$$

根据引理 A.8 可得

$$\theta_{\tau}^{(t)} = \frac{\lambda}{c}[(1+\mu)\theta_{\alpha}^{(t)} + \sqrt{\theta_{\tau}^{(t-1)}}] = \frac{2\lambda(2+\mu)}{c(1-\mu)}\sqrt{\theta_{\tau}^{(t-1)}}.$$

简单计算后可得到非零的不动点

$$\theta_{\tau}^{\infty} = \frac{4\lambda^2(2+\mu)^2}{c^2(1-\mu)^2} \quad \text{和} \quad \theta_{\alpha}^{\infty} = \frac{6\lambda(2+\mu)}{c(1-\mu)^2}.$$

此外, (3.3) 满足 $\theta_{\alpha}^{\infty} < \theta_{\alpha}^*$, $\theta_{\tau}^{\infty} < \theta_{\tau}^*$.

情形 (ii)

$$\theta_{\alpha}^{(t)} = \frac{6}{(1-\mu)\kappa}\sqrt{c\theta_{\alpha}^{(t-1)}\theta_{\tau}^{(t-1)}|\delta_0|_1s} = B_1\sqrt{\theta_{\alpha}^{(t-1)}}\sqrt{\theta_{\tau}^{(t-1)}}.$$

此外, 根据引理 A.8 可得

$$\theta_{\tau}^{(t)} = \frac{\lambda}{c}[\sqrt{\theta_{\tau}^{(t-1)}} + (1+\mu)\theta_{\alpha}^{(t)}] = \frac{\lambda}{c}\left[\frac{6(1+\mu)}{(1-\mu)\kappa}\sqrt{c|\delta_0|_1s}\sqrt{\theta_{\alpha}^{(t-1)}} + 1\right]\sqrt{\theta_{\tau}^{(t-1)}}(B_2\sqrt{\theta_{\alpha}^{(t-1)}} + B_3)\sqrt{\theta_{\tau}^{(t-1)}}.$$

如果 $B_1B_2 < 1$, 该式有非零的唯一不动点

$$\theta_{\tau}^{\infty} = \left(\frac{B_3}{1-B_1B_2}\right)^2 = \left(\frac{c^{-1}\lambda}{1-\frac{36(1+\mu)\lambda|\delta_0|_1s}{(1-\mu)^2\kappa^2}}\right)^2 \quad \text{和} \quad \theta_{\alpha}^{\infty} = B_1^2\theta_{\tau}^{\infty} = \frac{36c|\delta_0|_1s}{(1-\mu)^2\kappa^2}\theta_{\tau}^{\infty}.$$

这就是 (3.4) 的情形. 此外, 由 (3.6) 和 (3.5) 可得 $\theta_{\alpha}^{\infty} < \theta_{\alpha}^*$, $\theta_{\tau}^{\infty} < \theta_{\tau}^*$. 在两种情形中, $\theta_{\alpha}^{(t)}$ 和 $\theta_{\tau}^{(t)}$ 都随着 t 的增加而严格减小, 并分别收敛到低于 θ_{α}^* 和 θ_{τ}^* 的 θ_{α}^{∞} 和 θ_{τ}^{∞} . 因此, 在有限的 t^* 步之后, 可得 $|\hat{\alpha} - \alpha_0|_1 \leq \theta_{\alpha}^*$ 和 $|\hat{\tau} - \tau_0| \leq \theta_{\tau}^*$. 类似地, 可以得到 (A.20) 的预测误差界值. 结合 (A.20) 和引理 A.3, 可以得到回归系数估计量稀疏性的上界为

$$\mathcal{M}(\hat{\alpha}) \leq \frac{2b^2\phi_{\max}}{(1-\mu)^2\lambda^2}\|\hat{h} - h_0\|_n^2 \leq \frac{72b^2\phi_{\max}}{(1-\mu)^2\kappa^2a^2}.$$

迭代过程中, 每次应用引理 A.7 和 A.8 都要求 $\mathcal{C}(\theta_t)$ ($t = 1, \dots, t^*$) 成立, 因此, 引理 A.5 得证. $\square$

定理 3.3 的证明 结合假设 3.1 和引理 A.9, 令

$$C_3 = 4t^*, \quad C_4 = \frac{C^2}{b^2 \max_{t=1, \dots, t^*} M_n^2(\theta_t)},$$

$$K_3 = 6C \vee \frac{36C}{1-\mu} \vee \frac{24C^2(2+\mu)}{c(1-\mu)} \vee \frac{72}{(1-\mu)^2},$$

则定理得证. $\square$

Threshold single index regression model from high-dimensional data

Diqing Li, Yong He & Xinsheng Zhang

Abstract Instability of model parameters is a common phenomenon in areas such as signal processing, quality control, fault detection, finance, security, and clinical medicine. In this paper we focus on the high-dimensional single index model with a possible change point due to a covariate threshold. We propose the ℓ_1 -penalized likelihood estimators for regression coefficients as well as the threshold parameter. A proximal gradient algorithm is proposed for detecting the possible change point. Besides, non-asymptotic Oracle inequalities for both the prediction risk and the ℓ_1 -estimation loss for regression coefficients are obtained under certain sparsity condition. At last, thorough simulation studies are conducted to illustrate the empirical performance of the proposed method.

Keywords change point, covariate threshold, high-dimensional data, single index model

MSC(2010) 62H12

doi: 10.1360/SCM-2018-0618