



纵向数据部分线性模型的多方向变量选择

苏瀚, 朱慧姚, 王艺杰, 李高荣*

北京师范大学统计学院, 北京 100875

E-mail: suhan@mail.bnu.edu.cn, zhydzmx@163.com, 202321011018@mail.bnu.edu.cn, ligaorong@bnu.edu.cn

收稿日期: 2025-01-07; 接受日期: 2025-08-07; 网络出版日期: 2025-10-14; * 通信作者

国家自然科学基金 (批准号: 12271046 和 12131006) 资助项目

摘要 本文研究纵向数据下个性化部分线性模型的变量选择问题, 结合多方向收缩惩罚函数, 实现个性化变量选择和异质性协变量的个体分组. 多方向惩罚函数允许针对不同协变量指标得到不同的分组, 提高了个性化估计效率和变量选择准确性. 此外, 本文采用剖面 (profile) 局部线性光滑方法对模型进行估计, 在一定正则条件下建立参数和非参数估计量的渐近性质, 得到变量选择和个体分组的相合性. 本文还给出一种高效的交替方向乘子算法实现所提方法, 并从理论和模拟的角度研究所提算法的收敛性、收敛速度和计算复杂度. 模拟研究表明, 与现有的惩罚方法相比, 所提方法在参数估计、个性化变量选择、异质性个体分组和预测性能等方面均具有更好的表现. 进一步地, 将所提方法用于分析原发性胆汁性肝硬化数据集以证实其优良性.

关键词 纵向数据 个性化建模 部分线性模型 多方向惩罚函数 变量选择

MSC (2020) 主题分类 62J07, 62F07, 62F12

1 引言

随着精准医疗、个性化教育以及市场定位的兴起, 为满足个体特定需求的个性化建模和分析, 近年来逐渐受到学者关注. 例如, 在基因关联疾病的研究中, 某基因 α 对患者 A 是重要指标, 但对患者 B 的影响却不是显著的. 不同患者对基因的敏感性不同, 导致传统“以一代整”的模型不能有效描述个体的异质性, 例如, 对于纵向数据的个性化建模和分析工作, Zhu 和 Qu^[29] 考虑了个性化的非参数回归模型, 通过成对分组惩罚实现对纵向数据轨迹的聚类分析; Tang 等^[14] 考虑了针对回归系数的个性化线性模型, 提出了一种多方向收缩惩罚函数 (multi-directional separation penalty, MDSP) 来识别具有相似协变量效应的个体子群.

本文考虑如下的纵向数据个性化部分线性模型:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_i + \mathbf{z}_{ij}^T \boldsymbol{\alpha} + g(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m, \quad (1.1)$$

英文引用格式: Su H, Zhu H Y, Wang Y J, et al. Multi-directional variable selection of partially linear models with longitudinal data (in Chinese). Sci Sin Math, 2025, 55: 1–34, doi: 10.1360/SSM-2025-0007

其中 $y_{ij} \in \mathbb{R}$ 是响应变量, $\beta_i = (\beta_{i1}, \dots, \beta_{ip})^T$ 是关联协变量 $\mathbf{x}_{ij} = (x_{ij,1}, \dots, x_{ij,p})^T$ 的 p 维个性化 (异质性) 参数向量, $\alpha = (\alpha_1, \dots, \alpha_q)^T$ 是关联协变量 $\mathbf{z}_{ij} = (z_{ij,1}, \dots, z_{ij,q})^T$ 的 q 维群体共享 (同质性) 参数向量, $g(\cdot)$ 是一元未知的连续光滑函数, t_{ij} 是代表时间的一元协变量, 不失一般性, 假设其支撑集为 $[0, 1]$. 随机误差向量 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})^T$ 的均值为 $\mathbf{0}$, 协方差矩阵为 Σ_i .

模型 (1.1) 包含了几类重要的统计模型. 如果 $\beta_i = \beta_0 \in \mathbb{R}^p$, 参数部分回归系数不随个体 i 改变, 则模型 (1.1) 退化为纵向数据部分线性模型. 近年来关于纵向数据部分线性模型的研究已经取得了丰富的成果, 详细讨论参见文献 [7, 11, 19–23, 28]. 如果 $g(t_{ij}) = 0$, 不考虑时间对响应变量的影响, 则模型 (1.1) 退化为纵向数据个性化线性模型, 对于纵向数据个性化回归模型的研究已成为现代统计学领域的热点问题之一. 例如, Tang 等 [14] 研究了该模型的系数估计、多方向变量选择和子群划分问题, Hou 和 Wang [8] 将其进一步推广到纵向数据个性化分位数回归模型的研究中, Yu 等 [25] 研究了多源数据背景下个性化线性模型的变量选择的异质性系数聚类问题, Zhou 和 Jiang [27] 讨论了个性化系数 β_i 具有函数型指标结构情形下的子群分析问题. 对于以上模型的研究工作, 已经涌现了大量的研究成果. 但是模型 (1.1) 的参数估计、变量选择和异质性个体分组问题还没有被研究.

模型 (1.1) 结合了个性化建模和现代半参数回归模型, 其中一个优点是, 通过引入非参数部分 $g(t_{ij})$, 考虑了纵向数据个性化分析中响应变量对于时间的非线性相依关系, 这在临床检验和疾病进程监测等领域的应用中是广泛存在的. 同时, 由于 $g(\cdot)$ 是单变量函数, 模型 (1.1) 平衡了模型灵活性和可解释性, 不会遭受多元非参数模型常见的“维数灾祸”问题. 另外, 模型 (1.1) 还引入了个性化参数 $\{\beta_i\}_{i=1}^p$, 考虑了纵向数据分析中的不同样本个体对应的协变量 $\{\mathbf{x}_{ij}\}_{i \in \{1, \dots, n\}}$ 对于响应变量 $\{Y_{ij}\}_{i \in \{1, \dots, n\}}$ 效应的异质性, 使研究者能够探索对响应变量具有显著影响的个体分组, 因此具有广阔的应用前景.

本文基于个性化数据的多方向收缩惩罚思想和半参数剖面 (profile) 技巧, 给出模型 (1.1) 参数和非参数部分的惩罚估计量, 研究模型 (1.1) 的个性化变量选择方法和异质性个体的子群划分. 在一些正则条件下, 证明参数估计量的渐近性质和非参数估计量的极限分布, 同时建立变量选择和确定分组的相合性. 在算法上, 研究组内相关结构的估计方法, 给出有效的交替方向乘子 (alternating direction method of multipliers, ADMM) 迭代算法以计算多方向收缩惩罚估计的数值解. 最后, 模拟研究表明, 所提方法与已有惩罚估计方法相比, 在参数估计和变量选择等方面均具有更好的表现. 进一步地, 将所提出方法用于原发性胆汁性肝硬化数据集, 目的是实现个性化变量选择、划分异质性个体的分组结构.

本文余下内容的结构如下. 第 2 节讲述模型设定和估计方法, 详细介绍多方向惩罚函数, 并将其推广到多分组情形. 第 3 节建立估计量的理论性质. 第 4 节讨论调节参数、窗宽和异质性分组个数的选取方法, 介绍组内相关结构的估计方法, 给出基于 ADMM 的迭代算法, 并证明算法的收敛速度和计算复杂度. 第 5 节通过模拟研究和实际数据分析验证所提方法在有限样本下的有效性. 第 6 节给出本文的结论并展望未来的工作. 定理及相关引理的证明呈现在附录中.

2 估计方法

本节针对纵向数据个性化部分线性模型 (1.1), 主要关心的是模型中兴趣参数和非参数函数的估计问题. 为了简单, 令 $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$ 是 $m \times 1$ 的响应变量, $\mathbf{x}_{ij} = (x_{ij,1}, \dots, x_{ij,p})^T$ 是具有异质性效应的 p 维协变量, $\mathbf{z}_{ij} = (z_{ij,1}, \dots, z_{ij,q})^T$ 是具有同质性效应的 q 维协变量, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im})^T$

和 $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \dots, \mathbf{z}_{im})^T$ 分别是 $m \times p$ 和 $m \times q$ 的协变量设计矩阵, $\mathbf{T}_i = (t_{i1}, \dots, t_{im})^T$ 为 m 维协变量, 且 $g(\mathbf{T}_i) = (g(t_{i1}), \dots, g(t_{im}))^T$, 则模型 (1.1) 可写为如下矩阵形式:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{Z}_i \boldsymbol{\alpha} + g(\mathbf{T}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n. \quad (2.1)$$

假定参数 $\boldsymbol{\beta}_{(n)} = \text{vec}\{\boldsymbol{\beta}_i\}_{i=1}^n$ 和 $\boldsymbol{\alpha}$ 已知, 并记 $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{(n)})$ 是一个维数为 $p_{\boldsymbol{\theta}} = np + q$ 的参数向量, 其中参数维数 $p_{\boldsymbol{\theta}}$ 随样本量 n 发散. 记 $y_{ij}^* := y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_i - \mathbf{z}_{ij}^T \boldsymbol{\alpha}$, 则模型 (1.1) 可写成如下的纵向数据非参数回归模型:

$$y_{ij}^* = g(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m. \quad (2.2)$$

当假定参数向量 $\boldsymbol{\theta}$ 已知时, 可利用回归样条、光滑样条和局部光滑等非参数方法对未知非参数函数 $g(\cdot)$ 进行估计. Fan [5] 提出的局部线性光滑方法具有很多优良性质, 因此本文采用局部线性光滑方法对非参数函数 $g(\cdot)$ 进行估计. 对 t 的一个小邻域, 由 Taylor 展式, 有

$$g(t_{ij}) \approx g(t) + g'(t)(t_{ij} - t) =: a + b(t_{ij} - t). \quad (2.3)$$

为了获得 a 和 b 的估计, 可极小化下面的加权最小二乘目标函数:

$$\sum_{i=1}^n \sum_{j=1}^m [y_{ij}^* - a - b(t_{ij} - t)]^2 K_h(t_{ij} - t), \quad (2.4)$$

其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ 是一个核函数, $h > 0$ 是窗宽. 经简单计算, 可得 $g(t)$ 的“伪估计”为

$$\tilde{g}(t; \boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_i - \mathbf{z}_{ij}^T \boldsymbol{\alpha}), \quad (2.5)$$

其中

$$w_{ij}(t; h) = \frac{N^{-1}K_h(t_{ij} - t)[S_{n,2}(t) - (t_{ij} - t)S_{n,1}(t)]}{S_{n,0}(t)S_{n,2}(t) - S_{n,1}^2(t)}, \quad (2.6)$$

$$S_{n,l}(t) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m K_h(t_{ij} - t)(t_{ij} - t)^l, \quad l = 0, 1, 2, \quad (2.7)$$

这里 $N = nm$ 为样本量总数. 进一步地, 非参数伪估计 (2.5) 可以表示为如下的向量形式:

$$\tilde{g}(\mathbf{T}_i; \boldsymbol{\theta}) = \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}_i; h)(\mathbf{y}_k - \mathbf{X}_k \boldsymbol{\beta}_k - \mathbf{Z}_k \boldsymbol{\alpha}), \quad (2.8)$$

其中 $\mathbf{W}_k(\mathbf{T}_i; h)$ 是 $m \times m$ 的权重矩阵, 即

$$\mathbf{W}_k(\mathbf{T}_i; h) = \begin{bmatrix} w_{k1}(t_{i1}; h) & w_{k2}(t_{i1}; h) & \dots & w_{km}(t_{i1}; h) \\ w_{k1}(t_{i2}; h) & w_{k2}(t_{i2}; h) & \dots & w_{km}(t_{i2}; h) \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1}(t_{im}; h) & w_{k2}(t_{im}; h) & \dots & w_{km}(t_{im}; h) \end{bmatrix}.$$

进一步地, 考虑纵向数据的组内相关结构, 可构造参数向量 θ 的加权最小二乘目标函数为

$$\begin{aligned} L_{n,m}(\theta) &= \frac{1}{2} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{Z}_i \alpha - \mathbf{X}_i \beta_i - \tilde{g}(\mathbf{T}_i; \theta)]^T \mathbf{V}_i^{-1} [\mathbf{y}_i - \mathbf{Z}_i \alpha - \mathbf{X}_i \beta_i - \tilde{g}(\mathbf{T}_i; \theta)] \\ &= \frac{1}{2} \sum_{i=1}^n [\tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i \alpha - \tilde{\mathbf{X}}_i \beta_{(n)}]^T \mathbf{V}_i^{-1} [\tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i \alpha - \tilde{\mathbf{X}}_i \beta_{(n)}], \end{aligned} \quad (2.9)$$

其中 $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2}$ 表示第 i 个个体的工作协方差矩阵, $\mathbf{A}_i$ 表示 $\mathbf{y}_i$ 的边际方差构成的对角阵, $\mathbf{R}_i$ 是一个 $m \times m$ 的工作相关矩阵, 维数随着重复观测次数 m 发散. 进一步地, $\tilde{\mathbf{y}}_i = \mathbf{y}_i - \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}_i) \mathbf{y}_k$, $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i - \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}_i) \mathbf{Z}_k$, $\tilde{\mathbf{X}}_i = \mathbf{X}_i \mathbf{E}_i - \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}_i) \mathbf{X}_k \mathbf{E}_k$. 为了便于简化目标函数, 这里引入 $p \times np$ 的分块对角阵 $\mathbf{E}_i = \text{diag}(\mathbf{0}, \dots, \mathbf{I}_i, \dots, \mathbf{0})$, 即第 i 个对角位置上是 p 维单位阵 $\mathbf{I}_i$, 其余均为 $\mathbf{0}$. 这时, $\mathbf{E}_i \beta_{(n)} = \beta_i$, 方便从全体系数 $\beta_{(n)}$ 中提取针对第 i 个个体的个性化系数向量 β_i .

本文的主要目的除了对模型中的兴趣参数和非参数函数进行估计外, 更重要的是将具有相似协变量效应的个体进行分组, 即实现有效的个性化建模. 为了解决该问题, 拟采用 Tang 等^[14] 提出的 MDSP 惩罚方法, 通过对个性化参数 β_i ($i = 1, \dots, n$) 进行多方向压缩确定分组, 并利用组内信息同时实现参数估计和变量选择. 针对异质性协变量的第 k ($k = 1, \dots, p$) 个系数向量 $\beta_{\cdot k} = (\beta_{1k}, \dots, \beta_{nk})^T$, 假设存在 B_k 个分组, 即

$$\beta_{ik} = \begin{cases} \gamma_k^{(l)}, & \text{如果 } i \in \mathcal{G}_k^{(l)}, \quad l = 1, \dots, B_k - 1, \\ 0, & \text{如果 } i \in \mathcal{G}_k^{(0)}, \quad i = 1, \dots, n, \end{cases} \quad (2.10)$$

其中 $\gamma_k^{(l)}$ ($l = 1, \dots, B_k - 1$) 是未知的非零子分组同质性效应, 表示第 l 组内个体的共同系数; 而指标集 $\{\mathcal{G}_k^{(l)}\}_{l=0,1,\dots,B_k-1}$ 表示第 k 个异质性的个体分组. 为叙述方便, 下面仅讨论针对每个异质性协变量的所有个体都分为两组: 非零效应组 ($\beta_{ik} = \gamma_k, i \in \mathcal{G}_k$) 和零效应组 ($\beta_{ik} = 0, i \in \mathcal{G}_k^c$).

在 (2.10) 的设定下, 为了同时实现参数估计、变量选择和个体分组, 构造如下带有多方向惩罚函数 $s_\lambda(\cdot, \cdot)$ 的惩罚加权最小二乘目标函数:

$$Q_{n,m}(\alpha, \beta_{(n)}, \gamma) = L_{n,m}(\theta) + S_{\lambda_{n,m}}(\beta_{(n)}, \gamma), \quad (2.11)$$

其中 $L_{n,m}(\theta)$ 的定义见 (2.9), $S_{\lambda_{n,m}}(\beta_{(n)}, \gamma) = \sum_{i=1}^n \sum_{k=1}^p s_\lambda(\beta_{ik}, \gamma_k)$, 且 $\gamma = (\gamma_1, \dots, \gamma_p)^T$ 是未知的非零子分组同质性效应. 进一步地, 多方向惩罚函数定义为

$$s_\lambda(\beta_{ik}, \gamma_k) = \lambda_{n,m} \min(|\beta_{ik}|, |\beta_{ik} - \gamma_k|), \quad (2.12)$$

其中 $\lambda_{n,m} > 0$ 是调节参数. 为了直观展示 MDSP 方法在压缩估计和异质性个体分组方面的表现, 考虑 MDSP 惩罚最小二乘目标函数 $(z - \theta)^2/2 + s_\lambda(\theta, \gamma)$, 其中 θ 是个性化系数真值, z 是 θ 的最小二乘估计, $s_\lambda(\theta, \gamma) = \lambda \min(|\theta|, |\theta - \gamma|)$, 调节参数 $\lambda > 0$. 图 1(a) 展示了 MDSP 惩罚函数, 图 1(b) 展示了 MDSP 门限解.

可以看出, MDSP 惩罚函数是施加在 θ 上的分段凸函数, 不仅可以较弱的信号压缩到 0, 产生稀疏解, 同时还可以为较强的信号提供一个非零的压缩方向 γ , 保护强信号不被向 0 的方向压缩, 得到最终的参数估计, 并且不依赖额外的调节参数选取.

(2.11) 中的多方向惩罚项是一个双重求和形式, 包含针对个体维数和变量维数的求和, 从而提供了如下两个角度来看待模型 (1.1) 的个性化变量选择和异质性个体分组问题. 一方面, 考虑个体指标 i

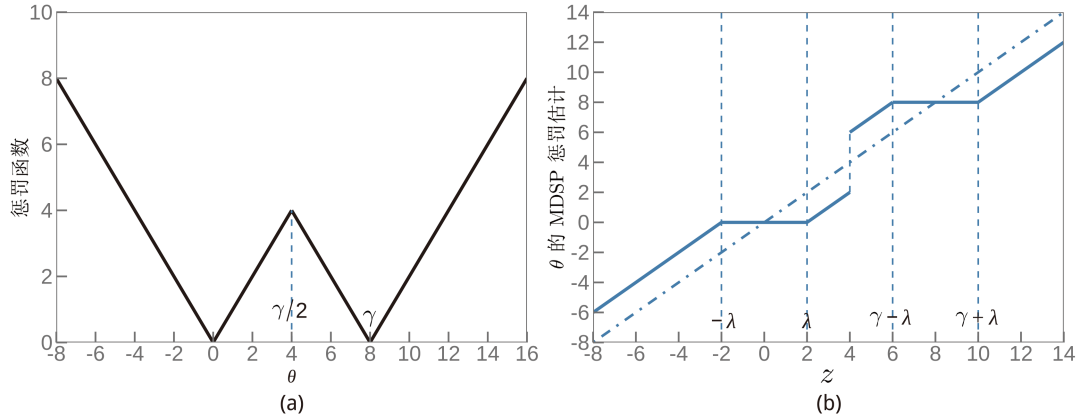


图 1 (网络版彩图) (a) MDSP 惩罚函数 $s_\lambda(\theta, \gamma) = \lambda \min(|\theta|, |\theta - \gamma|)$; (b) MDSP 门限解 (实线), 点断线表示最小二乘估计, 其中 θ 表示个性化系数真值, 调节参数 $\lambda = 2$, 非零压缩方向 $\gamma = 8$

及其对应的 $\mathbf{X}_i$, 给定 $\{\gamma_k\}_{k=1}^p$ 时, MDSP 惩罚项 $\sum_{k=1}^p s_\lambda(\beta_{ik}, \gamma_k)$ 可以实现第 i 个 p 维个性化系数 $\beta_i = (\beta_{i1}, \dots, \beta_{ip})^T$ 的个性化变量选择, 并为每个 β_{ik} 提供了 0 和 γ_k 两个压缩方向. 另一方面, 对于给定的协变量指标 k , MDSP 惩罚项 $\sum_{i=1}^n s_\lambda(\beta_{ik}, \gamma_k)$ 可以实现异质性个体分组. 具体地, MDSP 方法将信号较弱的个体分到零效应组 $\mathcal{G}_k$; 同时将信号较强的个体聚为一个子类, 聚类中心为 γ_k , 即分到非零效应组 $\mathcal{G}_k^c$.

可以发现, MDSP 方法所得估计和分组结果之间会相互影响, 系数估计 $\{\hat{\beta}_{ik}\}_{i=1}^n$ 、压缩方向估计 $\hat{\gamma}_k$ 和个体分组结果都会不断迭代更新以适应目标函数 (2.11). 相较于传统的针对初始系数估计量进行聚类的两阶段个体分组方法, MDSP 方法具有更好的可解释性. 同时, 通过最小化目标函数 (2.11), 可同时估计未知的非零子分组效应 γ_k , 即非零系数 $\{\hat{\beta}_{ik} \neq 0\}$ 的聚类中心, $k = 1, \dots, p$. 注意到, $\hat{\gamma}_k$ 的估计过程利用了具有相似效应的其他个体的信息, 且利用了个性化模型估计量所得分组信息, 可以有效减少估计的偏差和方差, 因此拥有更好的预测性. 相较于传统的基于两两成对惩罚项的融合聚类方法, MDSP 方法更倾向于分离系数差异大的异质性个体, 而非融合, 同时具有更高的计算效率. 传统方法, 如融合 LASSO (least absolute shrinkage and selection operator) 方法 [16], 其两两成对惩罚项 $\sum_{i,j} |\beta_{ik} - \beta_{jk}|$ 会将来自不同子分组的个体作融合, 包含 $O(np^2)$ 个 L_1 惩罚项. 而 MDSP 仅包含 $O(np)$ 个 L_1 惩罚项, 具有更好的计算可行性.

极小化式 (2.11) 可以得到参数 α 和 $\beta_{(n)}$ 的估计, 记为 $\hat{\theta} = \text{vec}(\hat{\alpha}, \hat{\beta}_{(n)})$, 再将 $\hat{\theta}$ 代入“伪估计” (2.5) 中, 可得到非参数函数的估计为

$$\hat{g}(t; \hat{\theta}) = \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h) (y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_i - \mathbf{z}_{ij}^T \hat{\alpha}). \quad (2.13)$$

注 2.1 本文采用了 MDSP 惩罚剖面最小二乘估计的思想, 在参数部分的估计中引入了工作相关矩阵, 建立了参数和非参数估计的理论性质. 理论上, 也可以利用 Wang [17] 与 Yao 和 Li [24] 的方法, 在非参数估计中引入工作协方差矩阵, 提高非参数估计的效果, 但是又会产生估计的累积误差, 并增加计算时间. 因此, 对非参数部分的估计仅考虑独立相关结构.

值得一提的是, 所提方法容易推广到多分组情形, 即每个个性化系数可能存在超过两种不同的异质性子分组效应 (包含零效应). 不失一般性, 假设第 k 个异质性协变量的个性化系数 β_{ik} 存在 $B_k = 3$ 个分组, 并且非零的异质性子分组效应为未知的正向效应 $\gamma_k^+ > 0$ 和负向效应 $\gamma_k^- < 0$. 三分组的广义

多方向惩罚函数定义为

$$s_{\lambda}^*(\beta_{ik}, \gamma_k^+, \gamma_k^-) = \lambda_{n,m} \min(|\beta_{ik}|, |\beta_{ik} - \gamma_k^+|, |\beta_{ik} - \gamma_k^-|), \quad \text{s.t. } \gamma_k^+ > 0, \quad \gamma_k^- < 0. \quad (2.14)$$

与前文中关于 MDSP 惩罚函数的讨论类似, (2.14) 定义的广义多方向惩罚函数给出了 3 个个性化系数的压缩方向 0, γ_k^+ 和 γ_k^- , 可以有效降低多分组情形下个性化建模中的系数估计误差, 提升个性化模型的拟合效果.

3 渐近性质

本节重点研究模型 (1.1) 中参数和非参数函数估计的渐近性质. 为了简单, 首先引入一些记号, 记 $\tilde{\mathbf{U}}_i = [\tilde{\mathbf{Z}}_i \ \tilde{\mathbf{X}}_i]$ 是 $m \times (q + np)$ 的协变量矩阵, 则惩罚目标函数 (2.11) 可写为

$$\begin{aligned} Q_{n,m}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{(n)}, \boldsymbol{\gamma}) &= \frac{1}{2} \sum_{i=1}^n (\tilde{\mathbf{y}}_i - \tilde{\mathbf{U}}_i \boldsymbol{\theta})^T \mathbf{V}_i^{-1} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{U}}_i \boldsymbol{\theta}) + \sum_{i=1}^n \sum_{k=1}^p s_{\lambda_{n,m}}(\beta_{ik}, \gamma_k) \\ &= L_{n,m}(\boldsymbol{\theta}) + S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}, \boldsymbol{\gamma}). \end{aligned} \quad (3.1)$$

对于 $L_{n,m}(\boldsymbol{\theta})$, 作如下运算:

$$\mathbf{G}_{n,m}(\boldsymbol{\theta}) = -\frac{\partial L_{n,m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \sum_{i=1}^n \tilde{\mathbf{U}}_i^T \mathbf{V}_i^{-1} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{U}}_i \boldsymbol{\theta}), \quad (3.2)$$

$$\mathbf{D}_{n,m} = -\frac{\partial \mathbf{G}_{n,m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \sum_{i=1}^n \tilde{\mathbf{U}}_i^T \mathbf{V}_i^{-1} \tilde{\mathbf{U}}_i, \quad (3.3)$$

$$\mathbf{H}_{n,m} = \text{Cov}(\mathbf{G}_{n,m}(\boldsymbol{\theta})) = \sum_{i=1}^n \tilde{\mathbf{U}}_i^T \mathbf{V}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1} \tilde{\mathbf{U}}_i, \quad (3.4)$$

其中 $\boldsymbol{\Sigma}_i = \text{Cov}(\mathbf{y}_i) = \mathbf{A}_i^{1/2} \mathbf{R}_i^0 \mathbf{A}_i^{1/2}$, 这里 $\mathbf{R}_i^0$ 表示真实的相关矩阵. 注意到 $\mathbf{D}_{n,m}$ 和 $\mathbf{H}_{n,m}$ 都是 $np + q$ 维的对称矩阵. 关于组内相关结构, 通常假定对于 $i = 1, \dots, n$, 有 $\mathbf{R}_i^0 = \mathbf{R}^0$ 和 $\mathbf{R}_i = \mathbf{R}$. 若组内相关结构 $\mathbf{R}$ 能被正确指定, 则有 $\mathbf{D}_{n,m} = \mathbf{H}_{n,m}$. 记 $\hat{\mu}_1(\mathbf{T}) = \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}) \mathbf{y}_k$, $\hat{\mu}_2(\mathbf{T}) = \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}) \mathbf{U}_k$, 其中 $\mathbf{U}_k = [\mathbf{Z}_k, \mathbf{X}_k, \mathbf{E}_k]$. 此时, $\tilde{\mathbf{y}}_i$ 和 $\tilde{\mathbf{U}}_i$ 可以重新写为

$$\begin{aligned} \tilde{\mathbf{y}}_i &= \mathbf{y}_i - \hat{\mu}_1(\mathbf{T}_i) = \mathbf{y}_i - \mu_1(\mathbf{T}_i) + \mu_1(\mathbf{T}_i) - \hat{\mu}_1(\mathbf{T}_i) =: \check{\mathbf{y}}_i + \tilde{\mu}_1(\mathbf{T}_i), \\ \tilde{\mathbf{U}}_i &= \mathbf{U}_i - \hat{\mu}_2(\mathbf{T}_i) = \mathbf{U}_i - \mu_2(\mathbf{T}_i) + \mu_2(\mathbf{T}_i) - \hat{\mu}_2(\mathbf{T}_i) =: \check{\mathbf{U}}_i + \tilde{\mu}_2(\mathbf{T}_i), \end{aligned}$$

其中 $\mu_1(\mathbf{T}_i) = \mathbb{E}(\mathbf{y} \mid \mathbf{T}_i)$, $\mu_2(\mathbf{T}_i) = \mathbb{E}(\mathbf{U} \mid \mathbf{T}_i)$, $\tilde{\mu}_1(\mathbf{T}_i) = \mu_1(\mathbf{T}_i) - \hat{\mu}_1(\mathbf{T}_i)$, $\tilde{\mu}_2(\mathbf{T}_i) = \mu_2(\mathbf{T}_i) - \hat{\mu}_2(\mathbf{T}_i)$, $\check{\mathbf{y}}_i = \mathbf{y}_i - \mu_1(\mathbf{T}_i)$, $\check{\mathbf{U}}_i = \mathbf{U}_i - \mu_2(\mathbf{T}_i)$. 另外, 记 $\check{\mathbf{Z}}_i = \mathbf{Z}_i - \mathbb{E}(\mathbf{Z} \mid \mathbf{T}_i)$, $\check{\mathbf{X}}_i = \mathbf{X}_i - \mathbb{E}(\mathbf{X} \mid \mathbf{T}_i)$. 定义 $\check{\mathbf{D}}_{n,m} = \sum_{i=1}^n \check{\mathbf{U}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{U}}_i$, $\check{\mathbf{H}}_{n,m} = \sum_{i=1}^n \check{\mathbf{U}}_i^T \mathbf{V}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1} \check{\mathbf{U}}_i$.

在给出本文主要渐近性质之前, 首先给出如下的正则条件.

(C1) 假设窗宽满足 $h = cN^{-1/5}$, 其中 $N = nm$, 且 $c > 0$ 为某个常数.

(C2) 核函数 $K(\cdot)$ 是对称的概率密度函数, 且在它的支撑集 $[-1, 1]$ 上有界变差.

(C3) 假设 $\sup_{t \in [0,1]} \mathbb{E}[\varepsilon_{ij}^4 \mid t_{ij} = t] < \infty$, $\sup_{t \in [0,1]} \mathbb{E}[x_{ij,k}^4 \mid t_{ij} = t] < \infty$ 和 $\sup_{t \in [0,1]} \mathbb{E}[z_{ij,r}^4 \mid t_{ij} = t] < \infty$, 其中 $x_{ij,k}$ 和 $z_{ij,r}$ 分别表示 $\mathbf{x}_{ij}$ 的第 k 个分量和 $\mathbf{z}_{ij}$ 的第 r 个分量, $k = 1, \dots, p$, $r = 1, \dots, q$.

(C4) 假设协变量 T 的密度函数 $f(t)$ 在 $(0, 1)$ 上连续可微, 且满足

$$0 < \inf_{t \in [0,1]} f(t) \leq \sup_{t \in [0,1]} f(t) < \infty.$$

(C5) 假设 $\mu_{2,r}(t)$ 和 $g(t)$ 在 $(0, 1)$ 上是二次连续可微的, $r = 1, \dots, np + q$, 其中 $\mu_{2,r}(t)$ 是 $\mu_2(t)$ 的第 r 个分量.

(C6) 假设方差函数 $\mathbb{V}\{\varepsilon(t)\}$ 在 $(0, 1)$ 上连续.

(C7) 未知参数 $\tilde{\theta} = \text{vec}(\alpha, \gamma)$ 属于紧空间 $\mathcal{B} \subseteq \mathbb{R}^{p+q}$, 它的真值存在于 $\mathcal{B}$ 内部; 同时对于 $1 \leq i \leq n$, $1 \leq j \leq m$, $\mathbf{p}_{ij} = \text{vec}(\mathbf{z}_{ij}, \mathbf{x}_{ij})$ 属于紧空间 $\mathcal{X} \subset \mathbb{R}^{p+q}$.

(C8) 记 $\mathbf{P}_i = (\check{\mathbf{Z}}_i, \check{\mathbf{X}}_i)$ 且 $\mathbf{P}_{i,k}$ 表示该矩阵的第 k 列, 对 $1 \leq k \leq p + q$, 有

$$\|\mathbf{P}_{i,k}\|_2^2 = O_p(m) \quad \text{和} \quad \sum_{i=1}^n m^{-1} \|\mathbf{P}_{i,k}\|_2^2 = O_p(n).$$

(C9) $m^{-1} \lambda_{\min}(\check{\mathbf{X}}_i^T \check{\mathbf{X}}_i) > C_1$, $(nm)^{-1} \lambda_{\min}(\sum_{i=1}^n \check{\mathbf{Z}}_i^T (\mathbf{I}_m - \check{\mathbf{H}}_{\mathbf{X}_i}) \check{\mathbf{Z}}_i) > C_2$, 其中 C_1 和 C_2 均为正常数, $\check{\mathbf{H}}_{\mathbf{X}_i} = \check{\mathbf{X}}_i (\check{\mathbf{X}}_i^T \check{\mathbf{X}}_i)^{-1} \check{\mathbf{X}}_i^T$, $\lambda_{\min}(\mathbf{M})$ 和 $\lambda_{\max}(\mathbf{M})$ 分别表示正定阵 $\mathbf{M}$ 的最小和最大特征值.

(C10) 当 n 或 m 足够大时, 假设矩阵 $\check{\mathbf{D}}_{n,m}$ 和 $\check{\mathbf{H}}_{n,m}$ 都是正定的.

(C11) 对所有 p 和 q , 记 $\mathbf{\Gamma}_{n,m} = (nm)^{-1} \check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}$, 并满足 $0 < C_3 < \lambda_{\min}(\mathbf{\Gamma}_{n,m}) \leq \lambda_{\max}(\mathbf{\Gamma}_{n,m}) < C_4 < \infty$, 其中 C_3 和 C_4 是正常数.

(C12) 记 $\eta_m = \lambda_{\max}(\mathbf{R}_i^{-1} \mathbf{R}_i^0)$, 满足 $|\eta_m| \leq C_5$, 其中 C_5 是某个正常数.

条件 (C1) 是对窗宽阶数施加的条件, 确保了欠光滑是不需要的, 因此可用数据驱动的方法选择窗宽. 条件 (C2) 对 $K(\cdot)$ 紧支撑集的要求仅是技术上的假定, 在实际应用中通常可以放宽. 条件 (C3) 是基本的矩条件. 条件 (C4) 确保了 $\hat{\mu}_{2,k}(t)$ 和 $\hat{g}(t)$ 的分母依概率 1 有界且下确界大于 0, 结合条件 (C2) 和 (C5), 确保了 $\hat{\mu}_{2,k}(t)$ 和 $\hat{g}(t)$ 具有高阶的收敛速度. 条件 (C6) 是基本假设条件. 条件 (C7) 和 (C10) 是纵向数据分析中常见的假设条件, 且容易满足, 参见文献 [14] 及相关文献. 条件 (C8) 和 (C9) 是推导有界回归理论性质所需要的温和正则条件, 弱于要求 $\check{\mathbf{X}}_i^T \check{\mathbf{X}}_i / m$ 收敛为正常数矩阵, 同时允许相同个体内的协变量是常量. 条件 (C11) 是对信息矩阵的限制, 确保参数估计的相合性. 条件 (C12) 中 η_m 表示真实相关矩阵与工作相关矩阵之间的差距, η_m 有界的条件在实际中容易达到, 只需要真实相关矩阵 $\mathbf{R}_i^0$ 满足某些限制, 详细讨论参见文献 [14].

注 3.1 随着样本量 n 的增加, 个性化系数 $\{\beta_i\}_{i=1}^n$ 的数量随着 n 趋于无穷大, 因此总体参数 θ 的维数 p_θ 也随着 n 趋于无穷大, 如果重复观测次数 m 是固定的, 将导致 MDSP 估计量的渐近 oracle 性质和确定分组相合性难以保证. 为了保证每个个性化系数估计量的相合性, 要求 n 和 m 均趋于无穷大, 类似的假设参见文献 [8, 14, 27].

求解估计方程 $\mathbf{G}_{n,m}(\theta) = 0$ 可得到未加惩罚项的参数估计 $\hat{\theta}^u$. 下面的定理说明了该估计量的相合性依赖于信息矩阵 $\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}$.

定理 3.1 在正则条件 (C1)–(C7), 条件 (C10) 和 (C11) 下, 对任意 $\delta > 0$, 则存在 (3.2) 的解 $\hat{\theta}^u$, 满足

$$\mathbb{P}(p_\theta^{-1/2} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} (\hat{\theta}^u - \theta^0)\|_2 > \delta) < \frac{1}{\delta^2},$$

其中 $p_\theta = np + q$, θ^0 表示参数真值. 此外, 若满足条件 $(\mathcal{C}_a) : \lambda_{\min}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \rightarrow \infty$, 则有

$$\mathbb{P}(p_\theta^{-1/2} \|\hat{\theta}^u - \theta^0\|_2 > \delta) \rightarrow 0.$$

定理 3.1 提供了谱范数意义下, 未加惩罚的估计量 $\hat{\boldsymbol{\theta}}^u$ 的渐近结果, 同时从定理 3.1 的证明可以发现, 当 $n, m \rightarrow \infty$ 时, 可以保证 $\hat{\boldsymbol{\theta}}^u$ 的相合性.

下面给出 oracle 估计的渐近性质, 本文所谓 oracle 估计是指已知真实分组情况下的参数估计. 为叙述简便, 引入一些符号表示. 记 $\{\mathcal{G}_k\}_{k=1,\dots,p}$ 表示针对不同协变量指标的个体分组情况, 即 $\mathcal{G}_k \subset \{1, 2, \dots, n\}$, 并且当 $i \in \mathcal{G}_k$ 时, $\beta_{ik} = \gamma_k$; 当 $i \notin \mathcal{G}_k$ 时, $\beta_{ik} = 0$. 记 $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{(n)})$ 是总体的系数向量, 而 $\boldsymbol{\theta}^0 = \text{vec}(\boldsymbol{\alpha}^0, \boldsymbol{\beta}_{(n)}^0)$ 是其真值, $\boldsymbol{\gamma}^0$ 是 $\boldsymbol{\gamma}$ 的真值. 此外, 将个性化系数 $\boldsymbol{\beta}_i$ 的真值记为 $\boldsymbol{\beta}_i^0 = \text{vec}(\boldsymbol{\beta}_{i, \mathcal{A}_i}^0, \mathbf{0})$, 其中 $\{\mathcal{A}_i\}_{i=1,\dots,n}$ 表示针对每个个体其协变量指标的分组情况. 显然 $\mathcal{A}_i \subset \{1, 2, \dots, p\}$, 且当 $k \in \mathcal{A}_i$ 时, $\beta_{ik}^0 = \gamma_k^0$; 而当 $k \notin \mathcal{A}_i$ 时, 则为 $\beta_{ik}^0 = 0$.

为推导 oracle 估计的理论性质, 假设真实的分组情况已知, 即 $\mathcal{G}_k$ ($1 \leq k \leq p$) 和 $\mathcal{A}_i$ ($1 \leq i \leq n$) 已知. 引入分组变量 $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{ip})^T$, 其中 $\omega_{ik} = 1_{\{i \in \mathcal{G}_k\}} = 1_{\{k \in \mathcal{A}_i\}}$. 则个性化的 oracle 参数 $\boldsymbol{\beta}_i^{\text{or}}$ 可以通过分组变量 $\boldsymbol{\omega}_i$ 与子同质性效应 $\boldsymbol{\gamma}$ 建立起联系, 即 $\boldsymbol{\beta}_i^{\text{or}} = \boldsymbol{\omega}_i \circ \boldsymbol{\gamma}$, “ $\circ$ ” 表示两个相同维数向量对应元素的乘积运算. 进一步地, 可在两个参数空间中建立一种映射: $\mathbb{R}^p(\boldsymbol{\gamma}) \mapsto \mathbb{R}^{np}(\boldsymbol{\beta}_{(n)}^{\text{or}}) : \boldsymbol{\Omega}\boldsymbol{\gamma} = \boldsymbol{\beta}_{(n)}^{\text{or}}$, 其中 $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_n]^T$ 是 $np \times p$ 的矩阵, 且 $\boldsymbol{\Omega}_i = \text{diag}(\boldsymbol{\omega}_i)$ 是对角阵.

这时, $\boldsymbol{\beta}_{(n)}^{\text{or}}$ 满足惩罚项部分 $S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^{\text{or}}, \boldsymbol{\gamma}) = 0$, oracle 估计可通过极小化如下目标函数得到:

$$\text{vec}(\hat{\boldsymbol{\gamma}}^{\text{or}}, \hat{\boldsymbol{\alpha}}^{\text{or}}) = \arg \min_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} \frac{1}{2} \sum_{i=1}^n (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i^{\text{or}} \boldsymbol{\gamma} - \tilde{\mathbf{Z}}_i \boldsymbol{\alpha})^T \mathbf{V}_i^{-1} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i^{\text{or}} \boldsymbol{\gamma} - \tilde{\mathbf{Z}}_i \boldsymbol{\alpha}), \quad (3.5)$$

其中 $\tilde{\mathbf{X}}_i^{\text{or}} = \mathbf{X}_i \boldsymbol{\Omega}_i - \sum_{k=1}^n \mathbf{W}_k(\mathbf{T}_i) \mathbf{X}_k \boldsymbol{\Omega}_k$.

定理 3.2 记 $\eta_m = \max_{1 \leq i \leq n} \{\lambda_{\max}(\mathbf{R}_i^{-1} \mathbf{R}_i^0)\}$. 在正则条件 (C1)–(C12) 下, 由 (3.5) 得到的 oracle 估计 $\hat{\boldsymbol{\theta}}^{\text{or}} = \text{vec}(\hat{\boldsymbol{\alpha}}^{\text{or}}, \hat{\boldsymbol{\gamma}}^{\text{or}})$, 有

$$\eta_m^{-1/2} \|(\check{\mathbf{D}}_{n,m}^{\text{or}})^{1/2} (\hat{\boldsymbol{\theta}}^{\text{or}} - \tilde{\boldsymbol{\theta}}^0)\|_2 \leq O_p(1),$$

其中 $\tilde{\boldsymbol{\theta}}^0 = \text{vec}(\boldsymbol{\alpha}^0, \boldsymbol{\gamma}^0)$, $\check{\mathbf{D}}_{n,m}^{\text{or}} = \sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1} \check{\mathbf{U}}_i^{\text{or}}$, 且 $\check{\mathbf{U}}_i^{\text{or}} = [\mathbf{Z}_i - \mathbb{E}(\mathbf{Z} | \mathbf{T}_i), \mathbf{X}_i \boldsymbol{\Omega}_i - \mathbb{E}(\mathbf{X} \boldsymbol{\Omega} | \mathbf{T}_i)]$. 此外, 若满足条件 $(\mathcal{C}_a^*) : \eta_m^{-1} \lambda_{\min}(\check{\mathbf{D}}_{n,m}^{\text{or}}) \rightarrow \infty$, 则在 L_2 范数意义下, 有 $\hat{\boldsymbol{\theta}}^{\text{or}} \xrightarrow{\mathbb{P}} \tilde{\boldsymbol{\theta}}^0$.

定理 3.2 表明了 oracle 估计的相合性取决于 η_m 和 $\lambda_{\min}(\check{\mathbf{D}}_{n,m}^{\text{or}})$. 定理中的 η_m 代表了工作相关矩阵和真实相关矩阵之间的偏差. 当工作相关矩阵 $\mathbf{R}_i$ 被正确指定时, oracle 估计可通过减少 η_m 来得到额外的估计有效性. 基于上述结论, 可以在 $n, m \rightarrow \infty$ 的假设下建立所提估计量的渐近性质.

定理 3.3 记 $\tau_{nm} = \lambda_{\min}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m})$. 在正则条件 (C1)–(C12) 下, 若 $\frac{\lambda_{n,m}}{\tau_{nm}} \rightarrow 0$ 且 $\frac{\lambda_{n,m}}{\sqrt{\tau_{nm}}} \rightarrow \infty$, 则 MDSP 目标函数 (2.11) 存在一个局部最小值 $\text{vec}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{(n)}, \hat{\boldsymbol{\gamma}})$. 当 $\tau_{nm} \rightarrow \infty$ 时, 有

$$\mathbb{P}\{\text{vec}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{(n)}, \hat{\boldsymbol{\gamma}}) = \text{vec}(\hat{\boldsymbol{\alpha}}^{\text{or}}, \hat{\boldsymbol{\beta}}_{(n)}^{\text{or}}, \hat{\boldsymbol{\gamma}}^{\text{or}})\} \rightarrow 1.$$

推论 3.1 (变量选择的相合性) 在定理 3.3 的条件下, 当 $\tau_{nm} \rightarrow \infty$ 时, 有

$$\mathbb{P}\left(\bigcap_{i=1}^n \{\hat{\mathcal{A}}_i = \mathcal{A}_i\}\right) \rightarrow 1.$$

推论 3.2 (确定分组的相合性) 在定理 3.3 的条件下, 当 $\tau_{nm} \rightarrow \infty$ 时, 有

$$\mathbb{P}\left(\bigcap_{k=1}^p \{\hat{\mathcal{G}}_k = \mathcal{G}_k\}\right) \rightarrow 1.$$

定理 3.3 表明了所提出的 MDSP 估计是渐近于 oracle 估计的, 并且它最大程度地运用了潜在分组结构的信息, 确保了 MDSP 估计能继承 oracle 估计的收敛速度, 同时实现了每个个性化估计量的正确分组. 下面的定理 3.4 给出非参数估计 $\hat{g}(t; \hat{\theta})$ 的渐近正态性.

定理 3.4 在正则条件 (C1)–(C12) 下, 若 $h \rightarrow 0$ 且 $nmh \rightarrow \infty$, 对于任给定的 $t \in (a, b) \subset [0, 1]$, 有

$$\sqrt{nmh} \left(\hat{g}(t; \hat{\theta}) - g(t) - \frac{1}{2} g''(t) h^2 \int u^2 K(u) du \right) \xrightarrow{L} N(0, \sigma^2(t)), \quad (3.6)$$

其中 $\xrightarrow{L}$ 表示依分布收敛, 且 $\sigma^2(t) = \frac{\mathbb{V}(\varepsilon(t))}{f(t)} \int K^2(u) du$.

定理 3.4 建立了非参数估计量 $\hat{g}(t; \hat{\theta})$ 的渐近正态性, 可以用来构造非参数函数的区间估计和逐点置信带.

4 参数选取和算法

本节讨论调节参数 $\lambda_{n,m}$ 和窗宽 h 的选取、异质性分组个数的选取、工作相关矩阵 $\mathbf{R}_i$ 的估计、求解目标函数 (2.11) 的 ADMM 算法以及 ADMM 算法的收敛速度和计算复杂度等问题.

4.1 调节参数和窗宽的选取

在利用 MDSP 惩罚函数进行变量选择时, 调节参数 $\lambda_{n,m}$ 对变量选择的结果具有决定性的作用. 因此, 需要发展数据驱动的方法选取最优的调节参数. 基于 Li 等^[10] 提出的 BIC (Bayesian information criterion) 数据驱动方法, 定义如下修正的 BIC 准则:

$$\text{BIC}(\lambda_{n,m}) = \log(\text{RSS}(\lambda_{n,m})/N) + C_n d(\lambda_{n,m}) \log(N)/N,$$

其中 $\text{RSS}(\lambda_{n,m}) = \sum_{i=1}^n [\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_i - \mathbf{Z}_i \hat{\alpha} - \hat{g}(\mathbf{T}_i)]^T \mathbf{V}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_i - \mathbf{Z}_i \hat{\alpha} - \hat{g}(\mathbf{T}_i)]$; $d(\lambda_{n,m})$ 是自由度, 表示给定调节参数 $\lambda_{n,m}$ 时需要估计的有效参数个数; $C_n > 0$ 是某个正常数, 在数值模拟中, 取 $C_n = \log(\log(np))$. 这时, 可通过极小化 $\text{BIC}(\lambda_{n,m})$ 选取最优的调节参数 $\lambda_{n,m}$.

在利用局部线性光滑方法拟合非参数 $g(\cdot)$ 时, 窗宽 h 控制着非参数估计的光滑度, 窗宽 h 过大, 会使拟合曲线过于光滑而导致欠拟合; 窗宽 h 过小, 则会使拟合曲线震荡加剧而产生过拟合现象. 本文采用广义交叉验证 (generalized cross-validation, GCV) 方法选取最优窗宽, 其中 GCV 目标函数定义为

$$\text{GCV}(h) = \frac{N^{-1}}{(N^{-1} \text{tr}(\mathbf{I}_N - \mathbf{W}(h)))^2} \sum_{i=1}^n [\tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i \hat{\alpha} - \tilde{\mathbf{X}}_i \hat{\beta}_i]^T [\tilde{\mathbf{y}}_i - \tilde{\mathbf{Z}}_i \hat{\alpha} - \tilde{\mathbf{X}}_i \hat{\beta}_i], \quad (4.1)$$

其中 $\mathbf{W}(h)$ 为 (4.9) 定义的 $N \times N$ 维的总权重矩阵. 因此, 可以通过极小化 GCV 函数 (4.1), 选取最优的窗宽 h .

4.2 异质性分组个数的选取

在使用多方向惩罚方法前, 确定异质性分组的个数 B_k 至关重要. 为解决这一问题, 将给出一种数据驱动的方法选择异质性分组的个数. 具体地, 对于异质性协变量的第 k 个系数向量 $\beta_{\cdot k} =$

$(\beta_{1k}, \dots, \beta_{nk})^T$, 假设其存在 B_k 个分组, 为确定分组个数 B_k , 可以定义如下的广义 BIC (generalized BIC, GBIC) 准则:

$$\text{GBIC}(B_k) = \log(\text{RSS}(B_k)/N) + a_N(B_k + q - 1) \log(N)/N, \quad (4.2)$$

其中

$$\text{RSS}(B_k) = \sum_{i=1}^n [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_i \hat{\boldsymbol{\alpha}} - \hat{\mathbf{g}}(\mathbf{T}_i)]^T \mathbf{V}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_i \hat{\boldsymbol{\alpha}} - \hat{\mathbf{g}}(\mathbf{T}_i)];$$

$a_N > 0$ 是依赖于总样本量 $N = nm$ 的正常数, 当 $a_N = 1$ 时 GBIC 准则退化为传统的 BIC 准则. 在 $p_{\boldsymbol{\theta}} = np + q$ 发散的设定下, 选取 $a_N = 2 \log(\log(p_{\boldsymbol{\theta}}))$.

更一般地, 对于存在多维个性化协变量和个性化系数 $\{\boldsymbol{\beta}_k\}_{k=1}^p$ 情形, 即存在多个异质性个体分组待确定时, 可以根据 (4.2) 逐一确定每个异质性个体分组个数, 最终得到所有个性化协变量的异质性分组个数 $\{B_k\}_{k=1}^p$.

4.3 组内相关结构的估计

在惩罚加权最小二乘目标函数 (2.11) 中, 为了刻画纵向数据组内相关结构, 考虑工作协方差矩阵 $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\rho) \mathbf{A}_i^{1/2}$, 其中 $\mathbf{R}_i(\rho)$ 是 $m \times m$ 的未知工作相关矩阵, ρ 为讨厌参数, $\mathbf{A}_i$ 是 $m \times m$ 的对角阵, 其第 j 个对角元素为 $\sigma^2(t_{ij})$, 表示 y_{ij} 的边际方差. Liang 和 Zeger [12] 针对纵向数据广义线性模型, 讨论了几种特定相关结构 $\mathbf{R}_i(\rho)$ 的估计, 并给出了不同结构下讨厌参数 ρ 的矩估计.

为了方便所提方法在实际问题中的应用, 本节也讨论在几种相关结构下工作相关矩阵的估计问题. 记 $\hat{r}_{ij} = (y_{ij} - \hat{g}(t_{ij}) - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_i - \mathbf{z}_{ij}^T \hat{\boldsymbol{\alpha}}) / \hat{\sigma}(t_{ij})$ 为 Pearson 残差, 其中 $\hat{\sigma}(t_{ij})$ 是响应变量估计值的标准差, 即 $\hat{\sigma}(t_{ij}) = \sqrt{\nabla(\hat{y}_{ij})}$, 其中 $\hat{y}_{ij} = \hat{g}(t_{ij}) + \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_i + \mathbf{z}_{ij}^T \hat{\boldsymbol{\alpha}}$.

(1) 可交换阵结构, 满足 $\text{corr}(y_{ij}, y_{ij'}) = \rho, j \neq j'; \text{corr}(y_{ij}, y_{ij'}) = 1, j = j'$. 则讨厌参数 ρ 的矩估计为

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m(m-1)} \sum_{j \neq j'}^m \hat{r}_{ij} \hat{r}_{ij'}.$$

(2) 一阶自回归 (AR(1)) 结构, 满足 $\text{corr}(y_{ij}, y_{ij'}) = \rho^{|j-j'|}$. 这时, 讨厌参数 ρ 的矩估计为

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m-1} \sum_{j \leq m-1} \hat{r}_{ij} \hat{r}_{i,j+1}.$$

(3) s -步相依 (s -dependent) 结构, 满足 $\text{corr}(y_{ij}, y_{i,j+k}) = \rho_k, k = 1, 2, \dots, s; \text{corr}(y_{ij}, y_{i,j+k}) = 0, k > s; \text{corr}(y_{ij}, y_{i,j+k}) = 1, k = 0$. 这时, 该结构的讨厌参数是一个 $s \times 1$ 的向量 $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_s)^T$, 则讨厌参数的矩估计为

$$\hat{\rho}_k = \frac{1}{n} \sum_{i=1}^n \frac{1}{m-k} \sum_{j \leq m-k} \hat{r}_{ij} \hat{r}_{i,j+k}.$$

特别地, 当 $s = 1$ 时所对应的相关矩阵为 MA(1) 结构.

(4) 不确定的相关结构, 即有 $\text{corr}(y_{ij}, y_{ik}) = \rho_{jk}, j \neq k$, 且 $\text{corr}(y_{ij}, y_{ik}) = 1, j = k$. 此时需要估计 $m(m-1)/2$ 个参数. 考虑利用残差 $\hat{\varepsilon}_{ij} = y_{ij} - \hat{g}(t_{ij}) - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_i - \mathbf{z}_{ij}^T \hat{\boldsymbol{\alpha}}$ 对工作协方差阵 $\mathbf{V}_i$ 进行估计. 注

意到不同个体具有的组内相关结构相同, 即 $\mathbf{V}_i = \mathbf{V}$, $i = 1, \dots, n$, 故有如下的工作协方差矩阵估计:

$$\widehat{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^n \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^{\mathrm{T}}, \quad (4.3)$$

其中 $\widehat{\boldsymbol{\varepsilon}}_i = (\widehat{\varepsilon}_{i1}, \dots, \widehat{\varepsilon}_{im})^{\mathrm{T}}$ 是 m 维的残差向量.

4.4 ADMM 算法

为了极小化惩罚加权最小二乘目标函数 (2.11), 下面利用 ADMM 算法给出参数估计的具体迭代公式, 并通过 (4.3) 对组内协方差阵进行估计, 从而增强参数估计的有效性. 首先, 引入一系列限制条件 $\beta_{ij} = \nu_{ij}$, $1 \leq i \leq n$, $1 \leq j \leq p$, 考虑如下新的带有限制条件的优化问题:

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\gamma}} L_{n,m}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + S_{\lambda_{n,m}}(\boldsymbol{\nu}, \boldsymbol{\gamma}), \quad \text{s.t.} \quad \boldsymbol{\beta} = \boldsymbol{\nu}, \quad (4.4)$$

这里, 为了简单, 重新记 $\boldsymbol{\beta}_{np \times 1} \equiv (\beta_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ 和 $\boldsymbol{\nu}_{np \times 1} \equiv (\nu_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$. 下面采用 ADMM 算法求解优化问题 (4.4), 定义增广 Lagrange 函数为

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\gamma}) = L_{n,m}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + S_{\lambda_{n,m}}(\boldsymbol{\nu}, \boldsymbol{\gamma}) + \boldsymbol{\Lambda}^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\nu}) + \frac{\kappa}{2} \|\boldsymbol{\beta} - \boldsymbol{\nu}\|_2^2, \quad (4.5)$$

这里, 为了简单, 记 $\boldsymbol{\Lambda}_{np \times 1} \equiv (\Lambda_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ 是 Lagrange 乘子, κ 是固定的增广参数. 逐步更新 $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$, $\{\boldsymbol{\nu}, \boldsymbol{\gamma}\}$ 和 $\boldsymbol{\Lambda}$, 它们在第 $l+1$ 步的迭代公式如下:

$$\{\boldsymbol{\alpha}^{(l+1)}, \boldsymbol{\beta}^{(l+1)}\} = \arg \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} L_{n,m}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \frac{\kappa}{2} \|\boldsymbol{\beta} - \boldsymbol{\nu}^{(l)} + \kappa^{-1} \boldsymbol{\Lambda}^{(l)}\|_2^2, \quad (4.6)$$

$$\{\boldsymbol{\nu}^{(l+1)}, \boldsymbol{\gamma}^{(l+1)}\} = \arg \min_{\boldsymbol{\nu}, \boldsymbol{\gamma}} S_{\lambda_{n,m}}(\boldsymbol{\nu}, \boldsymbol{\gamma}) + \frac{\kappa}{2} \|\boldsymbol{\beta}^{(l+1)} - \boldsymbol{\nu} + \kappa^{-1} \boldsymbol{\Lambda}^{(l)}\|_2^2, \quad (4.7)$$

$$\boldsymbol{\Lambda}^{(l+1)} = \boldsymbol{\Lambda}^{(l)} + \kappa(\boldsymbol{\beta}^{(l+1)} - \boldsymbol{\nu}^{(l+1)}). \quad (4.8)$$

(4.6) 的优化问题具有如下的显式解:

$$\{\boldsymbol{\alpha}^{(l+1)}, \boldsymbol{\beta}^{(l+1)}\} = (\mathbf{X}^{\mathrm{T}} \widetilde{\mathbf{V}}^{(l)} \mathbf{X}^* + \text{bdiag}\{\mathbf{0}_{q \times q}, \kappa \mathbf{I}_{np}\})^{-1} (\mathbf{X}^{\mathrm{T}} (\mathbf{I} - \mathbf{W}) \mathbf{Y} + \text{vec}\{\mathbf{0}_{q \times 1}, \kappa \boldsymbol{\nu}^{(l)} - \boldsymbol{\Lambda}^{(l)}\}),$$

其中 bdiag 代表分块对角矩阵, $\mathbf{X}^* = (\mathbf{I} - \mathbf{W}) \widetilde{\mathbf{X}}$, 且

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_2 & \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_n & \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_n \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1(\mathbf{T}_1) & \mathbf{W}_2(\mathbf{T}_1) & \dots & \mathbf{W}_n(\mathbf{T}_1) \\ \mathbf{W}_1(\mathbf{T}_2) & \mathbf{W}_2(\mathbf{T}_2) & \dots & \mathbf{W}_n(\mathbf{T}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_1(\mathbf{T}_n) & \mathbf{W}_2(\mathbf{T}_n) & \dots & \mathbf{W}_n(\mathbf{T}_n) \end{bmatrix}, \quad (4.9)$$

这里 $\widetilde{\mathbf{X}}$ 是 $nm \times (q + np)$ 的协变量矩阵, $\mathbf{Y} = \text{vec}(\mathbf{y}_i)_{i=1, \dots, n}$ 是 $nm \times 1$ 的响应变量, $\mathbf{I}$ 是 $nm \times nm$ 的单位阵. $\mathbf{W}$ 是 $nm \times nm$ 的权重矩阵, 用来估计非参数函数, 即

$$g(\mathbf{T})_{np \times 1}^{(l+1)} = \mathbf{W}(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^{(l+1)} - \mathbf{Z} \boldsymbol{\alpha}^{(l+1)}). \quad (4.10)$$

同时, $\widetilde{\mathbf{V}}^{(l)} = \text{bdiag}\{\mathbf{V}^{(l)-1}, \dots, \mathbf{V}^{(l)-1}\}$ 是由 n 个 $\mathbf{V}^{(l)-1}$ 组成的分块对角矩阵, 用于表示数据中的组内相关结构, 其中 $\mathbf{V}^{(l)} = n^{-1} \sum_{i=1}^n \boldsymbol{\varepsilon}_i^{(l)} \boldsymbol{\varepsilon}_i^{(l)\mathrm{T}}$, 且 $\boldsymbol{\varepsilon}_i^{(l)} = \mathbf{y}_i - g^{(l)}(\mathbf{T}_i) - \mathbf{X}_i \boldsymbol{\beta}_i^{(l)} - \mathbf{Z}_i \boldsymbol{\alpha}^{(l)}$ 是迭代第 l 步的估计残差.

考虑 (4.7) 的优化问题. 由于参数 ν 和 γ 不可分离, 所以需要不断迭代更新 ν 和 γ 来实现求解

$$\arg \min_{\nu, \gamma} \sum_{i=1}^n \left\{ \frac{\kappa}{2} (\nu_{ij} - \beta_{ij}^{(l+1)} - \kappa^{-1} \Lambda_{ij}^{(l)})^2 + \lambda_{n,m} \min(|\nu_{ij}|, |\nu_{ij} - \gamma_j|) \right\}. \quad (4.11)$$

给定 γ , 记 $\tilde{\beta}_{ik}^{(l+1)} = \beta_{ik}^{(l+1)} + \kappa^{-1} \Lambda_{ik}^{(l)}$, 则 ν 的更新公式为

$$\nu_{ik}^{(\text{new})} = \begin{cases} \text{sign}(\tilde{\beta}_{ik}^{(l+1)}) \cdot \max \left(0, |\tilde{\beta}_{ik}^{(l+1)}| - \frac{\lambda_{n,m}}{\kappa} \right), & |\tilde{\beta}_{ik}^{(l+1)}| \leq |e_{ik}^{(l+1)}|, \\ \gamma_k^{(\text{old})} + \text{sign}(\tilde{\beta}_{ik}^{(l+1)} - \gamma_k^{(\text{old})}) \cdot \max \left(0, |\tilde{\beta}_{ik}^{(l+1)} - \gamma_k^{(\text{old})}| - \frac{\lambda_{n,m}}{\kappa} \right), & |\tilde{\beta}_{ik}^{(l+1)}| > |e_{ik}^{(l+1)}|, \end{cases}$$

其中 $k = 1, \dots, p, i = 1, \dots, n, e_{ik}^{(l+1)} = \tilde{\beta}_{ik}^{(l+1)} - \gamma_k^{(\text{old})}$, sign 表示符号函数. 再给定 ν , 此时 γ 的优化可以通过沿着每个协变量 $k = 1, \dots, p$ 的一维搜索得到.

综合上面的内容, 下面给出一个迭代算法. 为改进估计的效率, 考虑组内相关结构进而引入 $\tilde{\mathbf{V}}$. 首先, 为获得初始估计, 忽略组内相关结构, 通过 (4.6) 得到参数的初始估计 $\{\alpha^{(0)}, \beta^{(0)}\}$. 然后再利用矩估计方法 (4.3) 得到组内相关性的估计 $\tilde{\mathbf{V}}$, 进一步通过 (4.6) 强化参数估计, 具体的迭代计算步骤如下.

步骤 1 令 $l = 0$, 初始化 $\{\alpha^{(0)}, \beta^{(0)}, \nu^{(0)}, \gamma^{(0)}\}$. 给定 $\lambda_{n,m}$ 和 κ , 设定 $\Lambda = \mathbf{0}$, 并设定停止迭代的偏差 ε_1 和 ε_2 .

步骤 2 利用 (4.3) 计算工作协方差阵 $\tilde{\mathbf{V}}^{(l)}$.

步骤 3 通过 (4.6), 更新 $\{\alpha^{(l+1)}, \beta^{(l+1)}\}$.

步骤 4 通过 (4.7) 以及平行计算 $j = 1, \dots, p$, 更新 $\{\nu_j^{(l+1)}, \gamma_j^{(l+1)}\}$.

步骤 5 更新 $\Lambda^{(l+1)} = \Lambda^{(l)} + \kappa(\beta^{(l+1)} - \nu^{(l+1)})$.

步骤 6 令 $l = l + 1$, 迭代步骤 2-5, 直至 $\{\|\beta^{(l+1)} - \beta^{(l)}\|_2 / (np) + \|\alpha^{(l+1)} - \alpha^{(l)}\|_2 / q + \|\gamma^{(l+1)} - \gamma^{(l)}\|_2 / p\} < \varepsilon_1$ 和 $\|\mathbf{r}^{(l+1)} - \mathbf{r}^{(l)}\|_2 < \varepsilon_2$, 其中 $\mathbf{r}^{(l)} = \beta^{(l)} - \nu^{(l)}$.

使用 ADMM 算法求解目标函数 (2.11) 时, 初始值的选取对于算法的求解非常重要. 当选取的初始值接近真实解时, ADMM 算法可以很快收敛到全局最优解; 当初始值偏离真实解时, 算法也能在有限步迭代内实现收敛. 在实际应用中, 可采用个性化最小二乘估计作为初始值, 或采用较小 $\lambda_{n,m}$ 和随机初始值所得的 MDSP 估计作为初始值.

下面讨论 ADMM 算法的收敛性、收敛速度和计算复杂度. 在给出 ADMM 算法的主要理论性质之前, 首先给出如下的正则条件.

(C1') (强制性条件) 假设 $g_{n,m}(\theta, \nu, \gamma) = L_{n,m}(\theta) + S_{\lambda_{n,m}}(\nu, \gamma)$ 在可行域上具有强制性, 即在可行域 $\{\theta, \nu, \gamma : \beta = \nu\}$ 上, 如果 $\|(\theta^T, \nu^T, \gamma^T)^T\|_2 \rightarrow \infty$, 则有 $g_{n,m}(\theta, \nu, \gamma) \rightarrow \infty$, 并且 $g_{n,m}(\theta, \nu, \gamma)$ 在可行域上有下界.

(C2') (光滑性条件) 假设损失函数 $L_{n,m}(\theta)$ 是 Lipschitz 可微的, 即梯度向量 $\nabla L_{n,m}(\theta)$ 是 Lipschitz 连续的, 其 Lipschitz 常数为 L_l . 假设惩罚函数 $S_{\lambda_{n,m}}(\nu, \gamma)$ 是广义 Lipschitz 可微的, 即其 Clarke 广义梯度 $\nabla^C S_{\lambda_{n,m}}(\nu, \gamma)$ 是 Lipschitz 连续的, 其 Lipschitz 常数为 L_s .

(C3') (Lipschitz 次最小化路径) 对任意的 $\mathbf{b} \in \mathbb{R}^p$, 存在 $\hat{\theta}(\mathbf{b}) = \arg \min_{\theta} \{L_{n,m}(\theta) : \beta = \mathbf{b}\}$ 是 $\mathbf{b}$ 的 Lipschitz 连续函数, 即存在常数 $C_l > 0$, 对任意的 $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^p$, 都有 $\|\hat{\theta}(\mathbf{b}_1) - \hat{\theta}(\mathbf{b}_2)\|_2 \leq C_l \|\mathbf{b}_1 - \mathbf{b}_2\|_2$.

条件 (C1')-(C3') 是使用 ADMM 算法求解非凸、非光滑优化问题的常见条件, 详细讨论参见文献 [3, 4, 18]. 对于优化问题 (4.4) 对应的平方损失函数和 MDSP 惩罚函数, 条件 (C1')-(C3') 是可以被验证的.

定理 4.1 (算法收敛性) 在正则条件 (C1')–(C3') 下, 对于优化问题 (4.4), 存在充分大的 κ , 使得所提 ADMM 算法满足如下性质:

- (1) (原始残差收敛性) $\lim_{l \rightarrow \infty} \|e^{(l)}\|_2^2 = 0$, 其中 $e^{(l)} = \beta^{(l)} - \nu^{(l)}$;
- (2) (对偶残差收敛性) $\lim_{l \rightarrow \infty} \|\nu^{(l+1)} - \nu^{(l)}\|_2^2 = 0$;
- (3) (估计收敛性) 算法所产生的估计序列 $\{\alpha^{(l)}, \beta^{(l)}, \gamma^{(l)}, \nu^{(l)}, \Lambda^{(l)}\}$ 有界, 并且至少有一个极限点 $(\alpha^*, \beta^*, \gamma^*, \nu^*, \Lambda^*)$, 其中每个极限点是增广 Lagrange 函数 (4.5) 的不动点.

定理 4.2 (算法收敛速度) 在正则条件 (C1')–(C3') 下, 记 ADMM 算法所产生的估计序列为 $\{\alpha^{(l)}, \beta^{(l)}, \gamma^{(l)}, \nu^{(l)}, \Lambda^{(l)}\}_{l \geq 0}$, 则有

- (1) (序列最优收敛速度) $\min_{0 \leq i \leq l} (\|\beta^{(i+1)} - \beta^{(i)}\|_2^2 + \|\nu^{(i+1)} - \nu^{(i)}\|_2^2 + \|\gamma^{(i+1)} - \gamma^{(i)}\|_2^2) = o(1/l)$;
- (2) (次梯度最优收敛速度) $\min_{0 \leq i \leq l} \|d^{(i)}\|_2 = o(1/\sqrt{l})$, 其中 $d^{(i)} \in \partial \mathcal{L}(\alpha^{(i)}, \beta^{(i)}, \nu^{(i)}, \gamma^{(i)}, \Lambda^{(i)})$ 为次梯度向量;

(3) (全局收敛) 增广 Lagrange 函数 (4.5) 是一个 Kurdyka-Lojasiewicz (KL) 函数 [1], 且产生的序列 $\{\alpha^{(l)}, \beta^{(l)}, \gamma^{(l)}, \nu^{(l)}\}_{l \geq 0}$ 可以全局收敛到唯一的极限点.

定理 4.1 说明, 求解优化问题 (4.4) 的 ADMM 算法在理论上是收敛的, 并可以收敛到增广 Lagrange 函数的不动点. 定理 4.2 说明, 所提 ADMM 算法的逐步差异序列, 如 $\{\beta^{(l+1)} - \beta^{(l)}\}$, 以次线性速度 $o(1/l)$ 趋于 0, 次梯度范数序列 $\{\|d^{(l+1)}\|_2\}$ 也以 $o(1/\sqrt{l})$ 的速度收敛到 0, 且 ADMM 算法是全局收敛的, 无论初始值如何选取, 产生的序列都可以收敛到唯一的极限点. 定理 4.1 和 4.2 从理论角度讨论了 ADMM 算法的稳定性和收敛性.

最后, 考虑所提 ADMM 算法的计算复杂度问题. 算法主要涉及初始化, 工作协方差矩阵 $\tilde{\mathbf{V}}^{(l)}$ 的计算, 参数 $\{\alpha^{(l)}, \beta^{(l)}, \nu^{(l)}, \gamma^{(l)}, \Lambda^{(l)}\}$ 的更新, 以及收敛性检查. 逐步推导每次迭代的计算成本并进行汇总, 去除低阶项, 可以得到 ADMM 算法的计算复杂度为

$$T(N, p_\theta) = O(N^2 + Np_\theta^2 + p_\theta^3) = O(\max(N^2, Np_\theta^2, p_\theta^3)). \quad (4.12)$$

进一步地, 算法复杂度可进一步简化为 $T(n, m) = O(n^3m + (nm)^2) = O(n^2m \cdot [n \vee m])$.

从 (4.12) 可以看出, 所提 ADMM 算法的计算复杂度的主项为: 求解 (4.6) 时构造系统矩阵的复杂度 Np_θ^2 , 解线性系统的复杂度 p_θ^3 , 以及计算非参数估计 (4.10) 时矩阵乘法的复杂度 N^2 . 在模拟研究中, 进一步给出了不同参数设置下, 所提 ADMM 算法在单次迭代时间、总迭代时间、内存占用情况等方面的表现. 总体来看, 所提算法在稳定性、收敛性、计算时间、敏感性和计算成本等方面均表现良好, 与理论结果相符.

5 数值模拟与实际数据分析

本节通过数值模拟和实际数据分析验证本文所提方法的有效性, 并与文献中方法进行比较. 在数值模拟和实际数据分析中, 利用第 4.1 小节提出的修正 BIC 准则选取最优的调节参数 $\lambda_{n,m}$ 和 GCV 方法选取最优的窗宽 h , 利用第 4.2 小节的方法估计组内相关矩阵, 利用第 4.3 小节提出的 ADMM 算法进行参数估计和变量选择, 在非参数拟合中取 Gauss 核函数. 所有数值模拟结果都是基于 100 次重复实验.

5.1 数值模拟

考虑如下具有 3 个群体变量和 1 个个性化变量的个性化部分线性模型:

$$y_{ij} = \alpha_1 z_{ij,1} + \alpha_2 z_{ij,2} + \alpha_3 z_{ij,3} + \beta_i x_{ij} + g(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m. \quad (5.1)$$

在该模型中, 设置样本量 $n = 50, 100$, 个体重复观测次数 $m = 10, 20$, 个性化系数设置为

$$\beta = (\beta_1, \dots, \beta_n)^T = (\underbrace{\gamma, \dots, \gamma}_{n/2}, \underbrace{0, \dots, 0}_{n/2})^T,$$

这里 γ 是真实的非零子分组同质性效应, 设置为 1 或 2, 而同质性的群体共享参数设为 $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 4$. 非参数函数取 $g(t) = \sin(2\pi t)$, 且 $t \in [0, 1]$, 即 t 从区间 $[0, 1]$ 上的均匀分布产生随机数. 协变量 $z_{ij,1}, z_{ij,2}, z_{ij,3}$ 和 x_{ij} 来自正态分布 $N(0, 1)$, 随机误差 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})^T$ 服从均值为 $\mathbf{0}$ 、协方差阵为 Σ_i 的正态分布, 且不同个体之间独立, 有 $\Sigma_i = \sigma^2 \mathbf{R}(\rho)$, 其中 $\sigma = 1$, $\rho = 0.5$, 组内相关结构 $\mathbf{R}$ 分别取一阶自回归 AR(1) 结构和可交换结构. 主要比较本文所提 MDSP 方法和其他 3 种个性化惩罚变量选择方法的表现, 分别是 LASSO [15]、SCAD (smoothly clipped absolute deviation) [6] 和 MCP (minimax concave penalty) [26] 方法. 具体地, 分别用 LASSO、SCAD 和 MCP 惩罚函数替换目标函数 (2.11) 中的惩罚项 $S_{\lambda_{n,m}}(\beta_{(n)}, \gamma)$, 实现模型 (1.1) 的个性化变量选择. 此外, 还比较了两种非变量选择的模型: 个性化最小二乘模型 (Sub) 和同质性最小二乘模型 (Homo), 后者假定 $\beta_i = \beta$, $i = 1, \dots, n$.

表 1 展示通过 GBIC 准则确定模型 (5.1) 下 MDSP 方法中异质性分组个数的结果, $m = 10, n = 50$, $\gamma = 1$, 组内相关结构为 AR(1). 可以看出 GBIC 先下降后上升, 在 $B_k = 2$ 时取得最小值, 因此选取 MDSP 方法的异质性分组个数 $B_k = 2$, 结果符合本文潜在的模拟设定. 在其他设定下, 通过 GBIC 准则确定异质性分组个数的结果类似. 不失一般性, 本节考虑两分组情形 MDSP 方法的表现情况, 并检验 MDSP 方法对于分组个数错误设定时的稳健性.

表 2 给出个性化系数估计量的均方根误差 (root mean square errors, RMSE), 即 $(np)^{-1/2} \|\hat{\beta} - \beta\|_2$ 和使用 MDSP 方法对非参数部分估计 $\hat{g}(t)$ 的均方根误差, 即 $\{(n_{\text{grid}})^{-1} \sum_{k=1}^{n_{\text{grid}}} [\hat{g}(t_k) - g(t_k)]^2\}^{1/2}$, 其中 n_{grid} 为网格点的数量, $\{t_k, k = 1, \dots, n_{\text{grid}}\}$ 为等距的网格点, 模拟中取区间 $[0, 1]$ 上 100 个等距的点作为网格点. 此外, 图 2 和 3 提供了两种情形下对应的箱线图, 其他模拟情形所得的箱线图类似, 限于篇幅, 不再展示.

从表 2、图 2 和 3 的结果可以得到如下结论.

(1) 本文所提 MDSP 方法相比于其他方法, 在各种情形下都有更小的 RMSE, 而 LASSO, SCAD, MCP 和 Sub 这 4 种变量选择方法表现类似, 但都要好于 Homo 的结果.

(2) 当样本量 n 或重复观测次数 m 增加时, 所有方法的 RMSE 都会有所减小, 并且当 m 增大时, 这种估计提升更为明显; 但是所提 MDSP 方法在 n 和 m 较小时仍能取得很小的 RMSE.

(3) 在非零子分组效应更为显著时 ($\gamma = 2$), 所提 MDSP 方法效果更有效, 对比其他方法至少有 70% ($m = 10$) 和 135% ($m = 20$) 的提升. 这是因为所提 MDSP 方法在估计个性化系数时能够借用不同个体的信息, 保留强信号的同时又能将弱信号压缩为 0.

表 1 MDSP 方法在不同异质性分组个数下的 GBIC 大小

评价指标	异质性分组个数 B_k			
	1	2	3	4
GBIC(B_k)	-5.064	-9.855	-7.634	-3.302

表 2 不同子方差效应 ($\gamma = 1, \gamma = 2$) 下个性化系数估计量的均方根误差 (RMSE) 和 MDSP 方法 $g(t)$ 的均方根误差 (RMSE)

	协方差阵	n	m	方法						
				MDSP	LASSO	SCAD	MCP	Sub	Homo	RMSE($\hat{g}(t)$)
$\gamma = 1$	AR(1)	50	10	0.268	0.340	0.365	0.356	0.403	0.503	0.051
			20	0.213	0.226	0.253	0.230	0.255	0.501	0.063
		100	10	0.234	0.331	0.458	0.425	0.377	0.502	0.047
			20	0.109	0.222	0.415	0.355	0.249	0.501	0.024
$\gamma = 2$	AR(1)	50	10	0.154	0.346	0.341	0.334	0.399	1.003	0.049
			20	0.108	0.216	0.185	0.187	0.246	1.001	0.044
		100	10	0.132	0.336	0.490	0.417	0.377	1.001	0.048
			20	0.061	0.220	0.302	0.226	0.247	1.001	0.031
$\gamma = 1$	可交换阵	50	10	0.225	0.331	0.358	0.348	0.396	0.504	0.059
			20	0.187	0.220	0.247	0.223	0.247	0.502	0.074
		100	10	0.210	0.332	0.461	0.429	0.380	0.502	0.074
			20	0.092	0.223	0.415	0.355	0.248	0.501	0.043
$\gamma = 2$	可交换阵	50	10	0.135	0.319	0.314	0.308	0.384	1.003	0.068
			20	0.117	0.226	0.196	0.199	0.250	1.001	0.065
		100	10	0.127	0.334	0.479	0.404	0.379	1.002	0.070
			20	0.056	0.225	0.319	0.239	0.249	1.001	0.042

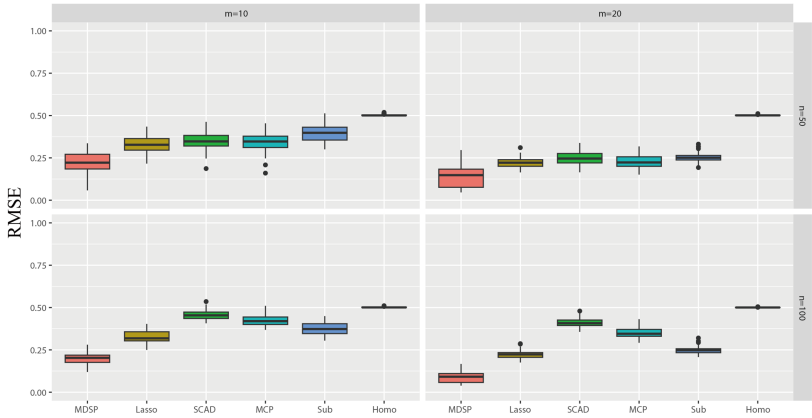


图 2 (网络版彩图) 在 $\gamma = 1$ 和 AR(1) 组内相关结构下, 个性化系数估计 RMSE 的箱线图比较

进一步地, 设定样本量 $n = 100$, 重复观测 m 分别取 10 和 20 时, 对所提 MDSP 方法的变量选择准确性进行比较, 评价指标包括正确筛选变量的比率 (CVSR, 将 β_i 正确分类为 0 或非 0 的比率)、敏感度 (将实际 β_i 中非 0 部分正确分类为非 0 的比率) 和特异性 (将 β_i 中实际为 0 部分正确分类为 0 的比率). 表 3 和 4 分别展示了 AR(1) 和可交换组内结构下, MDSP 方法的变量选择效果与其他方法的比较. 图 4 和 5 直观展示了 AR(1) 相关结构下两种情形 ($\gamma = 1, 2$) 变量选择准确性的箱线图, 可交换相关结构下的结果类似, 此处不再展示. 从表 3 和 4 以及图 4 和 5 的结果容易看出如下结论:

- (1) 对比于其他 3 种选择方法, 所提 MDSP 方法有更高的 CVSR, 同时在敏感性和特异性上也表

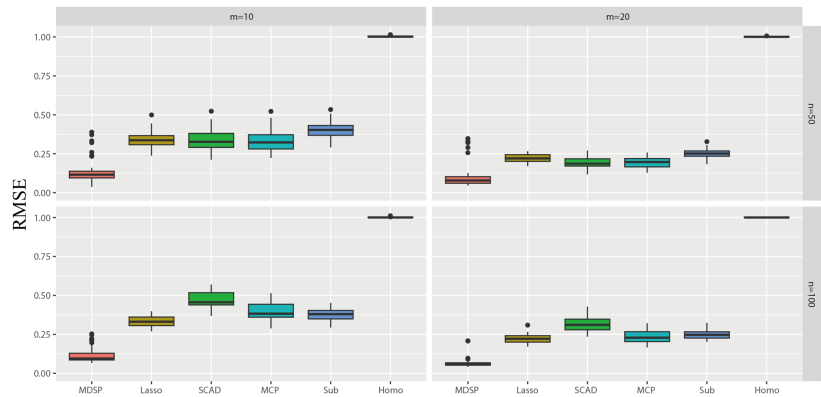


图 3 (网络版彩图) 在 $\gamma = 2$ 和可交换组内相关结构下, 个性化系数估计 RMSE 的箱线图比较

表 3 组内相关结构为 AR(1) 时, 不同子同质性效应下个性化系数估计量 β_i 的变量选择准确性

评价指标		n	m	方法			
				MDSP	LASSO	SCAD	MCP
$\gamma = 1$	CVSR	100	10	0.924	0.882	0.840	0.849
			20	0.980	0.965	0.920	0.943
	敏感度	100	10	0.926	0.884	0.746	0.818
			20	0.983	0.961	0.854	0.920
	特异性	100	10	0.923	0.881	0.933	0.881
			20	0.976	0.969	0.985	0.967
$\gamma = 2$	CVSR	100	10	0.992	0.939	0.951	0.932
			20	1.000	0.982	0.990	0.981
	敏感度	100	10	0.990	0.996	0.966	0.980
			20	1.000	1.000	0.997	0.999
	特异性	100	10	0.995	0.881	0.936	0.884
			20	1.000	0.964	0.983	0.963

现很好, 能够有效识别和剔除弱信号并保留强信号;

(2) 所有方法都对重复观测次数 m 比较敏感, 当 m 增加时, 它们的选择准确性都会提高;

(3) 当非零子分组效应增强 ($\gamma = 2$) 时, 所提 MDSP 方法表现更有效, 准确性可以达到 100%.

本文还对 MDSP 方法对于非参数函数的拟合情况进行了模拟实验. 在模型 (5.1) 设定的基础上, 给定组内相关结构为可交换结构 ($\sigma = 1, \rho = 0.5$) 且非零子分组同质性效应 $\gamma = 2$, 设定样本量 $n = 100$, 重复观测 $m = 10$, 观察非参数曲线的拟合情况. 表 5 展示了不同方法下非参数估计量的 RMSE 结果, 图 6 展示了非参数 $g(t)$ 的拟合效果以及根据理论结果绘制的 95%- 置信带结果.

从表 5 和图 6 可以得到如下结论:

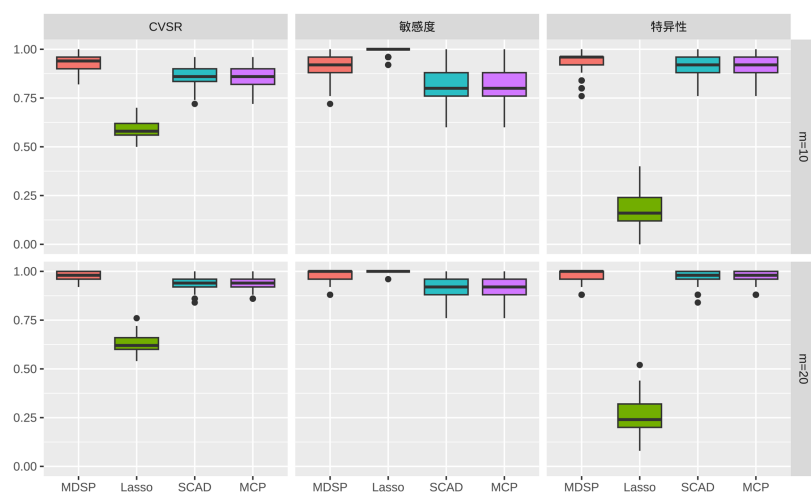
(1) 各方法都能正确拟合非参数函数的变化趋势, 但 MDSP 方法可以达到最小的估计误差 MSE;

(2) MDSP 方法拟合的非参数曲线与真实曲线几乎重合, 尤其在边界点处相比其他方法有更好的拟合效果;

(3) 根据 MDSP 方法, 非参数估计量的渐近分布绘制的置信水平为 95% 的置信带可以覆盖住整条真实曲线.

表 4 组内相关结构为可交换结构时, 不同子同质性效应下个性化系数估计量 β_i 的变量选择准确性

				方法			
评价指标		n	m	MDSP	LASSO	SCAD	MCP
$\gamma = 1$	CVSR	100	10	0.921	0.892	0.843	0.861
			20	0.974	0.963	0.917	0.939
	敏感度	100	10	0.913	0.895	0.747	0.825
			20	0.972	0.960	0.848	0.913
	特异性	100	10	0.928	0.889	0.938	0.897
			20	0.975	0.966	0.986	0.966
$\gamma = 2$	CVSR	100	10	0.993	0.941	0.953	0.936
			20	0.999	0.985	0.992	0.983
	敏感度	100	10	0.993	0.998	0.971	0.983
			20	0.999	0.999	0.997	0.998
	特异性	100	10	0.992	0.883	0.935	0.889
			20	0.999	0.969	0.986	0.967

图 4 (网络版彩图) 在 AR(1) 相关结构, 当 $n = 100$, $m = 10, 20$ 时, 个性化系数估计的分类正确率 (CVSR)、敏感度和特异性的箱线图比较, 其中 $\gamma = 1$

进一步地, 为了测试所提 MDSP 方法在不同异质性个体分组设定下的表现情况, 说明其在分组个数错误指定时的稳健性, 针对模型 (5.1) 产生的数据, 组内相关结构设定为 AR(1), 样本量为 $n = 60$, 重复观测次数为 $m = 10$. 对于系数 β_i , 有如下 3 种设定: (1) 系数只有 1 组非零的同质性效应 ($B_k = 1$, $\beta_i = \gamma = 2$, $i = 1, \dots, n$); (2) 个性化系数有 2 个异质性分组 ($B_k = 2$, $\gamma_0 = 0$, $\gamma_1 = -3$ 或 1); (3) 个性化系数有 3 个异质性分组 ($B_k = 3$, $\gamma_0 = 0$, $\gamma_1 = -3$, $\gamma_2 = 1$). 对 3 种设定都使用第 2 节所提两分组 MDSP 方法进行模型拟合和参数估计, 即假定 $B_k = 2$, 每个异质性协变量的所有个体都分为 2 组: 非零效应组 ($\beta_{ik} = \gamma_k$, $i \in \mathcal{G}_k$) 和零效应组 ($\beta_{ik} = 0$, $i \in \mathcal{G}_k^c$).

图 7 展示了分组个数正确指定和错误指定时, MDSP 方法和个性化最小二乘法的表现情况, 蓝色虚线代表压缩方向的 MDSP 估计 $\hat{\gamma}_{\text{MDSP}}$, 坐标轴上的圆点代表 γ 真值, 红色圆点代表个性化系数的

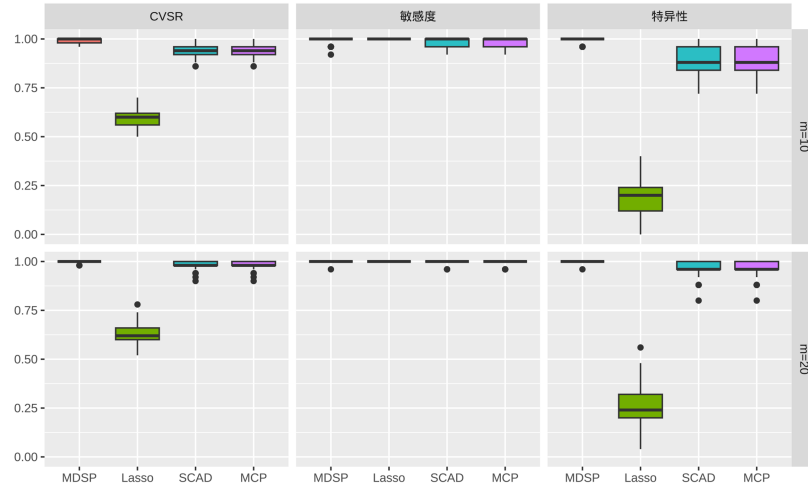


图 5 (网络版彩图) 在 $AR(1)$ 相关结构, 当 $n = 100$, $m = 10, 20$ 时, 个性化系数估计的分类正确率 (CVSR)、敏感度和特异性的箱线图, 其中 $\gamma = 2$

表 5 不同方法的非参数估计 $\hat{g}(t)$ 的 RMSE 结果

评价指标	方法			
	MDSP	LASSO	SCAD	MCP
$RMSE(\hat{g}(t))$	0.041	0.069	0.169	0.170

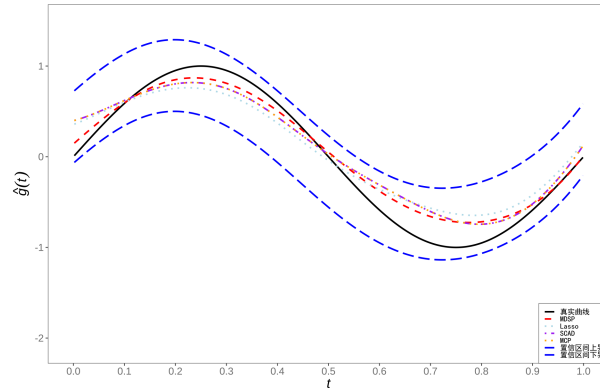


图 6 (网络版彩图) 不同方法下 $\hat{g}(t)$ 拟合情况和 95%- 置信带

MDSP 估计 $\hat{\beta}_{MDSP}$, 白色空心圆点代表个性最小二乘估计 $\hat{\beta}_{OLS}$. 图 7(b) 和 7(c) 对应设定 (2) 的分组个数正确指定的情形, 可以看出所提 MDSP 方法在分组数量正确指定时可以很好地估计出非零同质性效应 γ_1 (蓝色虚线), 将具有两分组的个性化系数分别向 $\gamma_0 = 0$ 和非零的 γ_1 两个方向进行压缩, 得到最终的 MDSP 估计 (红色圆点). 相比于个性化最小二乘方法, MDSP 方法有效减少了个性化系数的估计误差. 图 7(a) 和 7(d) 分别对应设定 (1) 和 (3) 的分组个数错误指定的情形, 说明了所提方法的稳健性. 具体地, 图 7(a) 表明哪怕不存在较弱信号的零效应子分组, MDSP 方法依然可以很好估计出同质性的非零效应, 说明所提 MDSP 方法是兼容系数不存在子分组的同质性设定的. 这是因为 MDSP 的压缩是多方向的, 存在非零的压缩方向, 从而可以减少较强信号的估计偏差. 图 7(d) 表明所

提 MDSP 方法正确识别了三分组中最强的非零方向 $\gamma_1 = -3$, 有效估计出了该分组的个性化系数, 这是因为 $\gamma_1 = -3$ 的信号要比 $\gamma_2 = 1$ 更强. 同时将较弱的信号向 $\gamma_0 = 0$ 进行压缩, 最终识别的两个方向分别是 0 和 -3 .

表 6 进一步报告了分组个数错误指定时, 情形 (1) 和 (3) 下参数估计和变量选择的结果. 可以看出, 对比其他传统惩罚变量选择方法, MDSP 方法有更小的 RMSE 和更高的选择准确性 CVSR, 对于分组个数错误指定情形具有稳健性.

最后, 从数值模拟的角度评估 ADMM 求解算法的稳定性、收敛性、计算成本和敏感度. 所有数值模拟结果采用单机 3.5 GHz CPU、16 GB 内存的实验平台. 表 7 给出了 100 次 Monte Carlo 实验下, ADMM 算法在不同模拟设置下的迭代步数和计算成本情况, 评价指标包括个性化系数估计量 RMSE、迭代收敛步数、单次迭代时间、总迭代时间和内存占用情况, 其余模拟设定与表 2 一致.

根据表 7 的实验结果, 可得如下结论.

(1) 在所有的模拟设定下, 所提 ADMM 算法在大多数情况下可以在 5 步迭代步骤之内完成收敛,

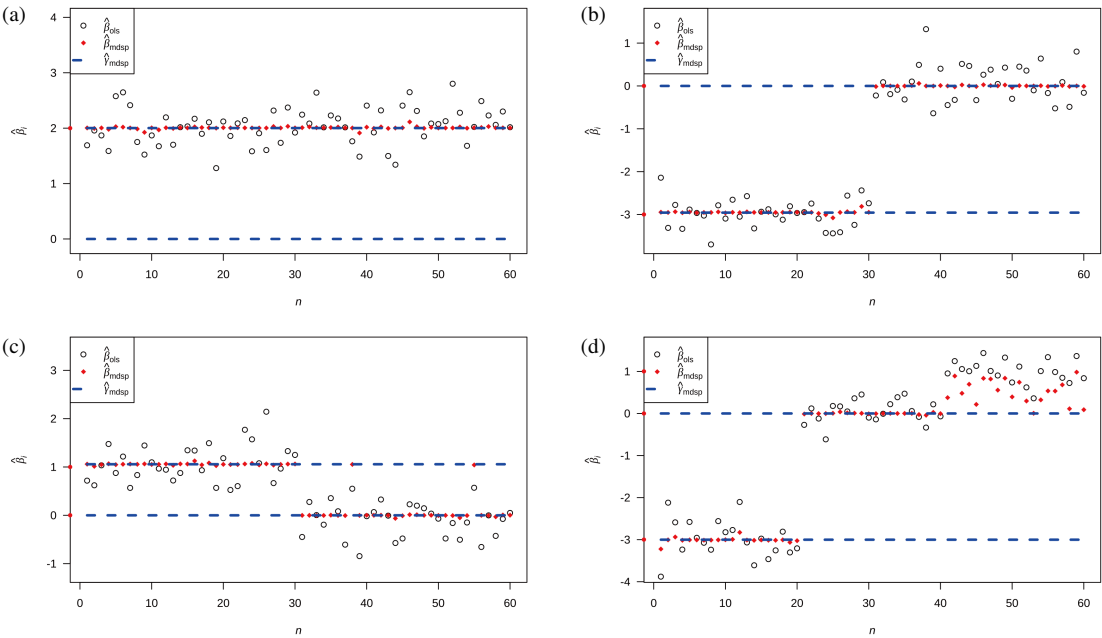


图 7 (网络版彩图) 真实分组情形设置为 (1)–(3) 时, 指定分组个数指定为 $B_k = 2$ (包含零效应分组)、个性化最小二乘与 MDSP 方法的参数估计和变量选择表现情况. (a) 同质性情形 (1), $B_k = 1, \gamma = 2$; (b) 异质性两分组情形 (2), $B_k = 2, \gamma_0 = 0, \gamma_1 = -3$; (c) 异质性两分组情形 (2), $B_k = 2, \gamma_0 = 0, \gamma_1 = 1$; (d) 异质性三分组情形 (3), $B_k = 3, \gamma_0 = 0, \gamma_1 = -3, \gamma_2 = 1$. 样本量 $n = 60$, 重复观测次数 $m = 10$

表 6 真实分组情形为 (1) 和 (3), 分组个数错误指定为 $B_k = 2$ 时, 个性化系数 β_i 在不同方法下的估计结果						
真实分组情形	评价指标	MDSP	OLS	LASSO	SCAD	MCP
$B_k = 1 (\beta_i = 2)$	RMSE	0.195	0.357	0.399	0.583	0.499
	CVSR	0.991	—	0.991	0.986	0.987
$B_k = 3 (\beta_i = 0, -3, 1)$	RMSE	0.222	0.356	0.453	0.453	0.405
	CVSR	0.914	—	0.869	0.898	0.899

表 7 ADMM 算法的算法稳定性、收敛性、计算成本分析结果

	协方差阵	n	m	RMSE	迭代步数统计		计算资源消耗		
					经验均值	经验方差	单次迭代 (s)	总耗时 (s)	内存占用 (%)
$\gamma = 1$	AR(1)	50	10	0.268	4.294	8.451	0.015	0.894	30.3
			20	0.213	8.188	14.264	0.048	3.401	35.7
		100	10	0.234	3.788	3.370	0.097	4.334	39.1
			20	0.109	5.628	8.221	0.367	25.606	43.8
$\gamma = 2$	AR(1)	50	10	0.154	4.024	2.194	0.015	0.878	35.2
			20	0.108	6.450	13.717	0.051	5.528	39.6
		100	10	0.132	3.790	4.997	0.099	4.316	42.1
			20	0.061	5.346	4.932	0.393	22.499	46.8
$\gamma = 1$	可交换阵	50	10	0.225	4.226	4.334	0.016	0.972	31.9
			20	0.187	7.584	24.732	0.049	3.907	34.8
		100	10	0.210	3.806	2.553	0.099	4.196	38.3
			20	0.092	5.564	11.581	0.393	22.896	42.1
$\gamma = 2$	可交换阵	50	10	0.135	4.050	1.393	0.014	0.937	36.0
			20	0.117	6.882	14.146	0.053	3.610	40.0
		100	10	0.127	3.774	1.870	0.100	4.947	43.9
			20	0.056	5.628	6.459	0.380	23.040	47.7

对于样本量 n 、重复观测次数 m 和协方差结构的变化具有良好的稳定性; 即使在小 n 大 m 的情形下, ADMM 算法依然可以在 8 步内完成收敛. 这一点与第 4.4 小节给出的收敛速度的理论结果 (定理 4.1 和 4.2) 相符.

(2) 所提算法收敛步数的波动较低, 经验的标准差基本不大于 4, 基本不存在陷入局部极小值或不收敛的情形, 具有良好的算法稳定性.

(3) 所提算法的计算成本低, 对于计算资源的要求友好. 算法的单次迭代时间和总耗时均较低, 随着 n 和 m 的增加迭代时间会缓慢增加, 并不会快速发散, 计算复杂度较低. 这一点与第 4.4 小节给出的算法复杂度结果相符.

(4) 在具有良好的算法稳定性、收敛性和计算时间的同时, 所提方法享有较低的系数估计量 RMSE, 具有良好的估计性能.

5.2 实际数据分析

本小节将所提 MDSP 方法应用于 R 程序包 `joineRML` 中 1974–1984 年的原发性胆汁性肝硬化 (PBC) 数据集, 该数据集包含了多个实验室结果, 涉及 312 名患者, 其中 140 人已经死亡, 且性别比例至少为 9:1 (女性对男性). 考虑一些基线协变量, 如年龄 (age)、性别 (gender)、药物使用指标 (drug), 这些协变量在研究开始时已经记录. 同时该实验纵向收集了各种生物指标, 包括肝肿大 (hepatauxe)、蜘蛛痣 (spider angioma)、水肿和腹水 (edema and ascites)、血清胆红素 (serum bilirubin)、白蛋白 (albumin)、碱性磷酸酶 (alkaline)、血清谷草转氨酶 (SGOT)、每立方米血小板 (platelets) 和凝血酶原 (prothrombin) 等. 该数据集的胆固醇水平协变量包含了太多缺失数据, 因此选择忽略该变量. 考虑到本文运用个性化建模并进行个体分组, 所选样本不包括重复次数小于 7 次的受试者, 因此, 本次分析

利用 89 名患者共 623 条记录进行分析.

受 Tang 等 [13]、Bo 和 Zhang [2] 研究的启发, 将对数变换后的血清胆红素水平 $\log(\text{serbili})$ 作为响应变量, 将基线变量年龄 (age)、性别 (gender)、药物使用指标 (drug) 设置为群体共享协变量, 具有时间依赖性的 $\log(\text{SGOT})$ 设为非参数协变量. 将其他生物指标, 如白蛋白 (albumin)、碱性磷酸酶 (alkaline)、每立方米血小板 (platelets) 和凝血酶原 (prothrombin), 设置为纵向数据部分线性模型的候选个性化协变量. 表 8 记录了选取不同协变量 (组) 作为个性化变量时, 模型在测试集上预测均方误差的大小.

根据表 8 的结果, 选取每立方米血小板 (platelets) 作为唯一的个性化协变量, 可以建立如下个性化部分线性模型:

$$\begin{aligned} \log(\text{serbili}_{ij}) = & \alpha_1 \text{age}_{ij} + \alpha_2 \text{gender}_{ij} + \alpha_3 \text{drug}_{ij} + \alpha_4 \log(\text{prothrombin}_{ij}) \\ & + \alpha_5 \text{alkaline}_{ij} + \beta_i \log(\text{platelets}_{ij}) + g(\log(\text{SGOT}_{ij})) + \varepsilon_{ij}, \end{aligned}$$

其中 $i = 1, \dots, 89, j = 1, \dots, 7$.

在具体分析中, 将该 PBC 数据集分为训练集 (前 5 次观测) 和测试集 (后 2 次观测), 计算所提 MDSP 方法在测试集上的预测均方误差, 并且与同质性 (Homo) 设定方法和 LASSO 方法进行比较. 同时, 为说明 MDSP 方法能发现更多的数据结构, 还计算了每个个体分组下协变量 $\log(\text{platelets})$ 系数的 p -值. 具体地, 将减去非参数估计量后的 $\log(\text{serbili}_{ij})$ 作为响应变量, 同质性变量和个性化变量不变. 对于模型 (5.1) 的每一个个性化系数 β_i , 考虑假设检验问题 $H_{0i}: \beta_i = 0 \leftrightarrow H_{1i}: \beta_i \neq 0$, 个性化协变量 $\log(\text{platelets})$ 系数的 p -值可以通过构造 Wald 检验统计量计算得到, 后者在原假设下的渐近分布为 χ^2_1 . 实现上, p -值可以通过 R 程序包 `geeglm` 中的函数 `geeglm()` 获得. 表 9 总结了预测的均方误差和个性化变量 $\log(\text{platelets})$ 系数估计的 p -值. 与 Homo 和 LASSO 方法相比, 所提 MDSP 方法分别将预测误差降低了 35% 和 31%. 在变量选择方面, Homo 模型提供的结果并不显著. 然而, 所提 MDSP 方法得到了个性化变量 $\log(\text{platelets})$ 系数的显著 p -值, 这些个体被识别为非零效应分组 $\mathcal{G}^{(\gamma)}$, 而其零效应分组 $\mathcal{G}^{(0)}$ 的 p -值不显著. 相比之下, LASSO 方法并没有实现有效的分组区分, 得到的两个分组 Comp1 和 Comp2 都具有显著性, 没有真正区别个体差异. 这表明所提 MDSP 方法提供了更具信息量的分组结构, 它同时实现了个性化变量选择和划分异质性个体的分组结构.

图 8 展示了不同方法在测试集上的预测情况, 曲线代表响应变量 $\log(\text{serbili})$ 随时间依赖变量 $\log(\text{SGOT})$ 的变化趋势. 显然 MDSP 具有更好的拟合效果, 尤其在拐点处拟合效果好于其他方法.

表 8 选取不同协变量 (组) 作为个性化变量时模型的测试集 RMSE 值

个性化变量	测试集 RMSE
每立方米血小板数 (platelets)	0.532
凝血酶原 (prothrombin)	0.999
碱性磷酸酶 (alkaline)	0.594
碱性磷酸酶 (alkaline)、凝血酶原 (prothrombin)	1.619
每立方血小板数 (platelets)、碱性磷酸酶 (alkaline)	0.703
每立方血小板数 (platelets)、凝血酶原 (prothrombin)	0.533
每立方血小板数 (platelets)、凝血酶原 (prothrombin)、碱性磷酸酶 (alkaline)	1.439

表 9 不同方法下个性化变量 $\log(\text{platelets})$ 系数的 p - 值, 测试集上响应变量 $\log(\text{serbili})$ 的预测误差

评价指标	个性化协变量设置为 $\log(\text{platelets})$				
	Homo	MDSP		LASSO	
		$\mathcal{G}^{(0)}$	$\mathcal{G}^{(\gamma)}$	Comp1	Comp2
$\log(\text{platelets})$ 的 p - 值	0.026	0.764	0.000	0.007	0.000
$\log(\text{serbili})$ 的预测均方误差	0.717	0.532		0.696	

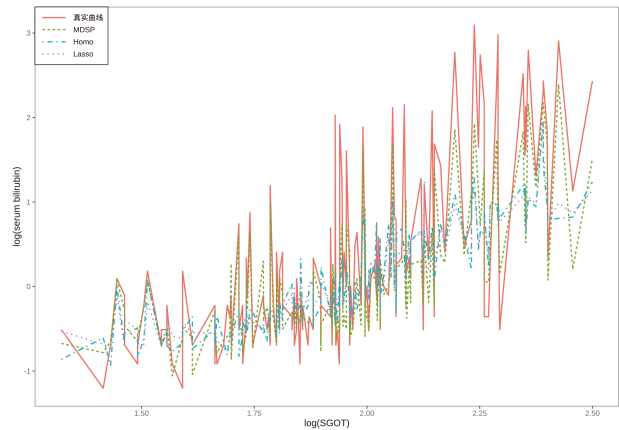


图 8 (网络版彩图) 不同方法下的响应变量预测情况

6 结论与展望

本文在纵向数据下研究了个性化部分线性模型的变量选择问题, 采用多方向惩罚函数 MDSP 和剖面局部线性光滑估计方法, 同时实现了个性化模型参数估计、个性化变量选择和异质性个体分组. 在一定的正则条件下, 建立了参数和非参数估计量的理论性质, 进一步得到了变量选择和个体分组的相合性. 本文还给出了所提方法的 ADMM 算法实现. 大量模拟实验表明本文所提估计方法在参数估计、个性化变量选择、异质性个体划分和预测性能等方面均优于已有方法, 同时具有良好的稳健性. 实际数据分析结果也验证了所提方法的有效性.

以下是未来可能的研究方向: 本文的方法建立在总体参数维数 p_{θ} 发散而个性化异质性协变量维数 p 和同质性协变量维数 q 固定的框架下. 这是因为在实际问题中逐个个体都能观测到的协变量往往是有限的, 因此用于个性化建模的异质性协变量是有限的, 如精准医疗问题中患者的个性化处理变量. 一个有趣的问题是, 发展适用于高维个性化变量的估计和推断方法, 可以结合逐个个体的惩罚最小二乘估计理论和本文的定理结果, 将所提方法推广到 p 发散的情形. 此外, 本文所提方法还可以拓展到其他半参数回归模型中, 如变系数模型和可加模型.

致谢 衷心感谢编委和审稿人所提出的宝贵建议, 他们的意见和建议使得本文有了进一步的提高.

参考文献

1 Attouch H, Bolte J, Svaiter B F. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. [Math Program](#), 2013, 137: 91–129

2 Bo X L, Zhang W P. Subgroup analysis for longitudinal data via semiparametric additive mixed effects model. [J Syst Sci Complex](#), 2023, 36: 2155–2185

- 3 Clarke F H. Optimization and Nonsmooth Analysis. New York: John Wiley & Sons, 1983
- 4 Deng W, Lai M J, Peng Z, et al. Parallel multi-block ADMM with $o(1/k)$ convergence. *J Sci Comput*, 2017, 71: 712–736
- 5 Fan J Q. Design-adaptive nonparametric regression. *J Amer Statist Assoc*, 1992, 87: 998–1004
- 6 Fan J Q, Li R Z. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Amer Statist Assoc*, 2001, 96: 1348–1360
- 7 Fan J Q, Li R Z. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J Amer Statist Assoc*, 2004, 99: 710–723
- 8 Hou Z H, Wang L. Heterogeneous quantile regression for longitudinal data with subgroup structures. *Comput Statist Data Anal*, 2024, 194: 107928
- 9 Knopp K. Infinite Sequences and Series. Chelmsford: Courier Corporation, 1956
- 10 Li G R, Peng H, Zhu L X. Nonconcave penalized M-estimation with a diverging number of parameters. *Statist Sinica*, 2011, 21: 391–419
- 11 Li G R, Yang Y P. Semiparametric Models with Longitudinal Data (in Chinese). Beijing: Science Press, 2015 [李高荣, 杨宜平. 纵向数据半参数模型. 北京: 科学出版社, 2015]
- 12 Liang K Y, Zeger S L. Longitudinal data analysis using generalized linear models. *Biometrika*, 1986, 73: 13–22
- 13 Tang C Y, Zhang W, Leng C. Discrete longitudinal data modeling with a mean-correlation regression approach. *Statist Sinica*, 2019, 29: 853–876
- 14 Tang X W, Xue F, Qu A. Individualized multidirectional variable selection. *J Amer Statist Assoc*, 2021, 116: 1280–1296
- 15 Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Stat Methodol*, 1996, 58: 267–288
- 16 Tibshirani R, Saunders M, Rosset S, et al. Sparsity and smoothness via the fused Lasso. *J R Stat Soc Ser B Stat Methodol*, 2005, 67: 91–108
- 17 Wang N. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 2003, 90: 43–52
- 18 Wang Y, Yin W T, Zeng J S. Global convergence of ADMM in nonconvex nonsmooth optimization. *J Sci Comput*, 2019, 78: 29–63
- 19 Wang Z F, Liu H J, Ding H, et al. Semiparametric regression analysis of doubly censored longitudinal data with informative observation times. *Sci Sin Math*, 2023, 53: 1253–1268
- 20 Wong K Y, Zeng D, Lin D Y. Semiparametric latent-class models for multivariate longitudinal and survival data. *Ann Statist*, 2022, 50: 487
- 21 Xu C, Hadjipantelis P Z, Wang J L. Semi-parametric joint modeling of survival and longitudinal data: The R package JSM. *J Statist Softw*, 2020, 93: 1–29
- 22 Xue L G, Zhu L X. Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika*, 2007, 94: 921–937
- 23 Xue L G, Zhu L X. Empirical likelihood-based inference in a partially linear model for longitudinal data. *Sci China Ser A*, 2008, 51: 115–130
- 24 Yao W X, Li R Z. New local estimation procedure for a non-parametric regression function for longitudinal data. *J R Stat Soc Ser B Stat Methodol*, 2013, 75: 123–138
- 25 Yu T T, Ye S Y, Wang R. High-dimensional variable selection accounting for heterogeneity in regression coefficients across multiple data sources. *Canad J Statist*, 2024, 52: 900–923
- 26 Zhang C H. Nearly unbiased variable selection under minimax concave penalty. *Ann Statist*, 2010, 38: 894–942
- 27 Zhou Y Q, Jiang F. Heterogeneous functional regression for subgroup analysis. *J Comput Graph Statist*, 2025, In press
- 28 Zhu L X, Xue L G. Empirical likelihood confidence regions in a partially linear single-index model. *J R Stat Soc Ser B Stat Methodol*, 2006, 68: 549–570
- 29 Zhu X L, Qu A. Cluster analysis of longitudinal profiles with subgroups. *Electron J Stat*, 2018, 12: 171–193

附录 A 一些引理和主要定理的证明

本附录包含一些必要的引理及其证明, 还包含第 3 节中定理 3.1–3.4 的证明.

引理 A.1 如果正则条件 (C1)–(C11) 成立, 则对任意常数 a 和 b 且 $(a, b) \subset [0, 1]$, 有

$$\sup_{a \leq t \leq b} \mathbb{E}\{|\mu_1(t) - \hat{\mu}_1(t)|^2\} = O(h^4 + n^{-1}m^{-1}h^{-1}),$$

$$\sup_{a \leq t \leq b} \mathbb{E}\{|\mu_{2,k}(t) - \hat{\mu}_{2,k}(t)|^2\} = O(h^4 + n^{-1}m^{-1}h^{-1}), \quad k = 1, \dots, np + q.$$

引理 A.1 的详细证明参见文献 [22], 此处省略.

引理 A.2 如果正则条件 (C1)–(C12) 成立, 当 $n \rightarrow \infty$, $m \rightarrow \infty$ 时, 有

$$\left\| \frac{1}{nm} (\mathbf{D}_{n,m} - \check{\mathbf{D}}_{n,m}) \right\| = O_p(p\theta(n^{-1/2}m^{-1/2}h^{-1/2} + h^2)),$$

其中 $\mathbf{D}_{n,m}$ 和 $\check{\mathbf{D}}_{n,m}$ 的定义分别见 (3.3) 和定理 3.2, $\|\mathbf{M}\|$ 表示矩阵 $\mathbf{M}$ 的 Frobenius 范数.

引理 A.2 的证明类似文献 [11, 引理 3.5.1] 的证明, 此处省略.

引理 A.3 在正则条件 (C1)–(C12) 下, 对任意的 $\delta > 0$ 和 $\mathbf{a} \in \mathbb{R}^{np+q}$, 有

$$\mathbb{P}(|\mathbf{a}^T(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)|^2 > \delta) \leq \delta^{-2} \mathbf{a}^T (\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \mathbf{a}.$$

进一步地, 对于个性化参数估计 $\hat{\boldsymbol{\beta}}_i^u$ 和群体估计 $\hat{\boldsymbol{\alpha}}^u$, 有

$$\begin{aligned} \mathbb{P}(\|\hat{\boldsymbol{\beta}}_i^u - \boldsymbol{\beta}_i^0\|_2 > \delta) &\leq 4^{-1} p \delta^{-4} \eta_m \lambda_{\min}(\check{\mathbf{D}}_{\mathbf{X}_i}^{-1}), \\ \mathbb{P}(\|\hat{\boldsymbol{\alpha}}^u - \boldsymbol{\alpha}^0\|_2 > \delta) &\leq 4^{-1} q \delta^{-4} \eta_m \lambda_{\min}(\check{\mathbf{D}}_{\mathbf{Z}}^{-1}), \end{aligned}$$

其中 $\check{\mathbf{D}}_{\mathbf{X}_i} = \check{\mathbf{X}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{X}}_i$, $\check{\mathbf{D}}_{\mathbf{Z}} = \sum_{i=1}^n \check{\mathbf{Z}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{Z}}_i$, 且 $i = 1, \dots, n$.

证明 对任意的 $\delta > 0$ 和 $\mathbf{a} \in \mathbb{R}^{np+q}$, 有

$$\mathbb{P}(|\mathbf{a}^T(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)|^2 > \delta) \leq \delta^{-2} \mathbf{a}^T (\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \mathbf{a}.$$

为推导个性化参数估计 $\hat{\boldsymbol{\beta}}_i^u$ 的渐近性质, 需先写出矩阵 $\check{\mathbf{D}}_{n,m}$ 的具体形式, 它是一个分块稀疏矩阵, 即

$$\begin{aligned} \check{\mathbf{D}}_{n,m} &= \begin{pmatrix} \check{\mathbf{D}}_{zz}(q \times q) & \check{\mathbf{D}}_{zx}(q \times np) \\ \check{\mathbf{D}}_{xz}(np \times q) & \check{\mathbf{D}}_{xx}(np \times np) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \check{\mathbf{Z}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{Z}}_i & \check{\mathbf{Z}}_1^T \mathbf{V}_1^{-1} \check{\mathbf{X}}_1 & \check{\mathbf{Z}}_2^T \mathbf{V}_2^{-1} \check{\mathbf{X}}_2 & \cdots & \check{\mathbf{Z}}_n^T \mathbf{V}_n^{-1} \check{\mathbf{X}}_n \\ \check{\mathbf{X}}_1^T \mathbf{V}_1^{-1} \check{\mathbf{Z}}_1 & \check{\mathbf{X}}_1^T \mathbf{V}_1^{-1} \check{\mathbf{X}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \check{\mathbf{X}}_2^T \mathbf{V}_2^{-1} \check{\mathbf{Z}}_2 & \mathbf{0} & \check{\mathbf{X}}_2^T \mathbf{V}_2^{-1} \check{\mathbf{X}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{X}}_n^T \mathbf{V}_n^{-1} \check{\mathbf{Z}}_n & \mathbf{0} & \mathbf{0} & \cdots & \check{\mathbf{X}}_n^T \mathbf{V}_n^{-1} \check{\mathbf{X}}_n \end{pmatrix}, \end{aligned}$$

其中 $\check{\mathbf{Z}}_i = \mathbf{Z}_i - \mathbb{E}(\mathbf{Z} | \mathbf{T})$, $\check{\mathbf{X}}_i = \mathbf{X}_i - \mathbb{E}(\mathbf{X} | \mathbf{T})$. 类似地, 可得 $\check{\mathbf{H}}_{n,m}$ 的具体形式为

$$\check{\mathbf{H}}_{n,m} = \begin{pmatrix} \sum_{i=1}^n \check{\mathbf{Z}}_i^T \mathbf{V}_i^{-1} \boldsymbol{\Sigma}_n \mathbf{V}_n^{-1} \check{\mathbf{Z}}_i & \check{\mathbf{Z}}_1^T \mathbf{V}_1^{-1} \boldsymbol{\Sigma}_1 \mathbf{V}_1^{-1} \check{\mathbf{X}}_1 & \cdots & \check{\mathbf{Z}}_n^T \mathbf{V}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{V}_n^{-1} \check{\mathbf{X}}_n \\ \check{\mathbf{X}}_1^T \mathbf{V}_1^{-1} \boldsymbol{\Sigma}_1 \mathbf{V}_1^{-1} \check{\mathbf{Z}}_1 & \check{\mathbf{X}}_1^T \mathbf{V}_1^{-1} \boldsymbol{\Sigma}_1 \mathbf{V}_1^{-1} \check{\mathbf{X}}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{X}}_n^T \mathbf{V}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{V}_n^{-1} \check{\mathbf{Z}}_n & \mathbf{0} & \cdots & \check{\mathbf{X}}_n^T \mathbf{V}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{V}_n^{-1} \check{\mathbf{X}}_n \end{pmatrix}.$$

由矩阵推导运算, 可得 $\mathbf{a}^T(\check{\mathbf{D}}_{n,m}\check{\mathbf{H}}_{n,m}^{-1}\check{\mathbf{D}}_{n,m})^{-1}\mathbf{a} \leq \eta_m \mathbf{a}^T(\check{\mathbf{D}}_{n,m})^{-1}\mathbf{a}$. 同时 $\check{\mathbf{D}}_{n,m}$ 可分解为

$$\check{\mathbf{D}}_{n,m} = \begin{pmatrix} \mathbf{I}_q & \check{\mathbf{D}}_{zx}\check{\mathbf{D}}_{xx}^{-1} \\ \mathbf{0} & \mathbf{I}_{np} \end{pmatrix} \begin{pmatrix} \check{\mathbf{D}}_{zz} - \check{\mathbf{D}}_{zx}\check{\mathbf{D}}_{xx}^{-1}\check{\mathbf{D}}_{xz} & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{D}}_{xx} \end{pmatrix} \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ \check{\mathbf{D}}_{xx}^{-1}\check{\mathbf{D}}_{xz} & \mathbf{I}_{np} \end{pmatrix}.$$

因此, 有

$$\check{\mathbf{D}}_{n,m}^{-1} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ -\check{\mathbf{D}}_{xx}^{-1}\check{\mathbf{D}}_{xz} & \mathbf{I}_{np} \end{pmatrix} \begin{pmatrix} (\check{\mathbf{D}}_{zz} - \check{\mathbf{D}}_{zx}\check{\mathbf{D}}_{xx}^{-1}\check{\mathbf{D}}_{xz})^{-1} & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{D}}_{xx}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{np} - \check{\mathbf{D}}_{zx}\check{\mathbf{D}}_{xx}^{-1} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix}.$$

取 $\mathbf{a} = \text{vec}(\mathbf{0}_q, \mathbf{a}_1, \dots, \mathbf{a}_n)$, 其中 $\mathbf{a}_j \in \mathbb{R}^p$, $1 \leq j \leq n$, 并且当 $j = i$ 时, $\mathbf{a}_j = \mathbf{1}_p$; 而 $j \neq i$ 时, $\mathbf{a}_j = \mathbf{0}_p$. 于是 $\mathbf{a}$ 可以从参数向量 $\boldsymbol{\theta}$ 中提取出个性化参数 β_i . 进一步可得

$$\mathbf{a}^T \check{\mathbf{D}}_{n,m}^{-1} \mathbf{a} = \mathbf{1}_p^T (\check{\mathbf{X}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{X}}_i)^{-1} \mathbf{1}_p \leq p \lambda_{\min}(\check{\mathbf{X}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{X}}_i)^{-1}.$$

因此, 有

$$\begin{aligned} \mathbb{P}(\|\hat{\beta}_i^u - \beta_i^0\|_2 > \delta) &= \mathbb{P}(\|\hat{\beta}_i^u - \beta_i^0\|_2^2 > \delta^2) \leq \mathbb{P}(|\mathbf{a}^T(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)|^2 > 2\delta^2) \\ &\leq 4^{-1}\delta^{-4}\mathbf{a}^T(\check{\mathbf{D}}_{n,m}\check{\mathbf{H}}_{n,m}^{-1}\check{\mathbf{D}}_{n,m})^{-1}\mathbf{a} \leq 4^{-1}\delta^{-4}\eta_m \mathbf{a}^T(\check{\mathbf{D}}_{n,m})^{-1}\mathbf{a} \\ &\leq 4^{-1}p\delta^{-4}\eta_m \lambda_{\min}((\check{\mathbf{X}}_i^T \mathbf{V}_i^{-1} \check{\mathbf{X}}_i)^{-1}). \end{aligned}$$

进一步地, 使用类似的证明过程, 可得群体变量系数估计 $\hat{\boldsymbol{\alpha}}$ 的渐近性质, 此处省略证明过程. $\square$

引理 A.4 在正则条件 (C1)–(C12) 下, 记 $\tau_{nm} = \lambda_{\min}(\check{\mathbf{D}}_{n,m}\check{\mathbf{H}}_{n,m}^{-1}\check{\mathbf{D}}_{n,m})$, 对于任意的 $\delta > 0$, 当 $\tau_{nm} \rightarrow \infty$ 时, 有 $\mathbb{P}(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|_\infty > \delta) \rightarrow 0$.

引理 A.4 提供了关于估计 $\hat{\boldsymbol{\theta}}^u$ 一个更强的相合性, 即在 L_∞ 范数下的相合性. 引理 A.4 的证明与下面定理 3.1 的证明过程类似, 此处省略.

引理 A.5 假设存在一数列 $\{a_i\}_{i=1,\dots,n}$, 它有分类指标集 $\mathcal{G}_l$ ($l = 1, \dots, L$), 且满足对任意的 $i \in \mathcal{G}_l$, 有 $|a_i - b| \leq \epsilon$, 其中 ϵ 是一个小的正常数. 因此, 可得下列目标函数存在一个局部最小值 $\hat{\mathbf{b}}$:

$$S(\mathbf{b} | \mathbf{a}) = \sum_{i=1}^n \left(\min_{1 \leq l \leq L} |a_i - b_l| \right),$$

满足 $\|\hat{\mathbf{b}} - \mathbf{b}\|_\infty \leq 2\epsilon$.

引理 A.5 表明, 通过极小化函数 $S(\mathbf{b} | \mathbf{a})$ 得到的分组中心 $\hat{\mathbf{b}}$ 在无穷范数下是渐近于真实分组中心 $\mathbf{b}$ 的. 引理 A.5 的证明可参见文献 [14, 引理 A.3] 的证明.

引理 A.6 在正则条件 (C3') 下, 存在常数 $C_p > 0$, 使得

$$\|\boldsymbol{\Lambda}^{(i+1)} - \boldsymbol{\Lambda}^{(i)}\|_2 \leq C_p (\|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2).$$

证明 由于 $(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)})$ 是 $\mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\Lambda}^{(i)})$ 的极小值点, 因此有

$$\nabla^C S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}) - \boldsymbol{\Lambda}^{(i)} - \kappa(\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i+1)}) = \mathbf{0},$$

其中 ∇^C 表示 Clarke 广义梯度. 由于 $\boldsymbol{\Lambda}^{(i+1)} = \boldsymbol{\Lambda}^{(i)} + \kappa(\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i+1)})$, 因此 $\boldsymbol{\Lambda}^{(i+1)} = \nabla^C S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)})$, 并且

$$\|\boldsymbol{\Lambda}^{(i+1)} - \boldsymbol{\Lambda}^{(i)}\|_2 = \|\nabla^C [S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}) - S_{\lambda_n, m}(\boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)})]\|_2$$

$$\leq L_s(\|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2),$$

其中正则条件 (C2') 保证了该不等式的成立. $\square$

引理 A.7 ADMM 算法得到的序列满足

$$\begin{aligned} \lim_{i \rightarrow \infty} \|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2^2 &= 0, & \lim_{i \rightarrow \infty} \|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 &= 0, \\ \lim_{i \rightarrow \infty} \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2 &= 0, & \lim_{i \rightarrow \infty} \|\boldsymbol{\Lambda}^{(i+1)} - \boldsymbol{\Lambda}^{(i)}\|_2^2 &= 0. \end{aligned}$$

证明 由条件 (C2') 可知 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ 有下界, 再结合性质 (P3) 可知 $\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) \leq \mathcal{L}(\boldsymbol{\theta}^{(0)}, \boldsymbol{\nu}^{(0)}, \boldsymbol{\gamma}^{(0)}, \boldsymbol{\Lambda}^{(0)})$ 对任意的 $i = 1, 2, \dots$ 成立, 这意味着 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ 也有上界, 因此 $g_{n,m}(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma})$ 也有上界并且 $\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ 在 $i \rightarrow \infty$ 时收敛.

由条件 (C1') 可以得到 $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)})$ 有界. 此外, 通过引理 A.6 的证明过程可知

$$\boldsymbol{\Lambda}^{(i)} = \nabla^C S_{\lambda_{n,m}}(\boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)})$$

也有界. 因此, $\lim_{i \rightarrow \infty} \|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2^2 = 0$, $\lim_{i \rightarrow \infty} \|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 = 0$ 和 $\lim_{i \rightarrow \infty} \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2 = 0$ 显然成立. 因此, 由引理 A.6 可知 $\lim_{i \rightarrow \infty} \|\boldsymbol{\Lambda}^{(i+1)} - \boldsymbol{\Lambda}^{(i)}\|_2^2 = 0$ 也显然成立. 引理 A.7 证毕. $\square$

定理 3.1 的证明 已知 $\hat{\boldsymbol{\theta}}^u$ 是估计方程 $\mathbf{G}_{n,m}(\boldsymbol{\theta}) = 0$ 的解, 在正则条件 (C10) 和引理 A.3 下, 由 Taylor 展式, 有

$$\mathbf{G}_{n,m}(\hat{\boldsymbol{\theta}}^u) = \mathbf{G}_{n,m}(\boldsymbol{\theta}^0) - \mathbf{D}_{n,m}(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0) = 0.$$

记 $\mathbf{G}_{n,m}(\boldsymbol{\theta}^0) = \mathbf{G}_{n,m}$, 对 $\mathbf{G}_{n,m}$ 作如下分解:

$$\begin{aligned} \mathbf{G}_{n,m} &= \sum_{i=1}^n [\check{\mathbf{U}}_i + \check{\boldsymbol{\mu}}_2(\mathbf{T}_i)]^T \mathbf{V}_i^{-1} [\boldsymbol{\varepsilon}_i + \check{\boldsymbol{\mu}}_1(\mathbf{T}_i) - \check{\boldsymbol{\mu}}_2(\mathbf{T}_i) \boldsymbol{\theta}^0] \\ &= \check{\mathbf{G}}_{n,m} + \sum_{i=1}^n \check{\mathbf{U}}_i^T \mathbf{V}_i^{-1} [\check{\boldsymbol{\mu}}_1(\mathbf{T}_i) - \check{\boldsymbol{\mu}}_2(\mathbf{T}_i) \boldsymbol{\theta}^0] + \sum_{i=1}^n \check{\boldsymbol{\mu}}_2(\mathbf{T}_i)^T \mathbf{V}_i^{-1} \boldsymbol{\varepsilon}_i \\ &\quad + \sum_{i=1}^n \check{\boldsymbol{\mu}}_2(\mathbf{T}_i)^T \mathbf{V}_i^{-1} [\check{\boldsymbol{\mu}}_1(\mathbf{T}_i) - \check{\boldsymbol{\mu}}_2(\mathbf{T}_i) \boldsymbol{\theta}^0], \end{aligned}$$

其中 $\check{\mathbf{G}}_{n,m} = \sum_{i=1}^n \check{\mathbf{U}}_i^T \mathbf{V}_i^{-1} \boldsymbol{\varepsilon}_i$. 由引理 A.2, 可得 $\|\mathbf{G}_{n,m} - \check{\mathbf{G}}_{n,m}\|_2 = \sqrt{np_{\boldsymbol{\theta}}} \cdot O_p(h^2 + (nmh)^{-1/2})$. 为方便, 记 $\delta_{nm} = (nmh)^{-1/2}$, 简单计算则有

$$\begin{aligned} \hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0 &= \mathbf{D}_{n,m}^{-1} \mathbf{G}_{n,m} = [\check{\mathbf{D}}_{n,m} + np_{\boldsymbol{\theta}} \cdot O_p(\delta_{nm} + h^2)]^{-1} [\check{\mathbf{G}}_{n,m} + \sqrt{np_{\boldsymbol{\theta}}} \cdot O_p(\delta_{nm} + h^2)] \\ &= [I + np_{\boldsymbol{\theta}} \check{\mathbf{D}}_{n,m}^{-1} \cdot O_p(\delta_{nm} + h^2)]^{-1} \check{\mathbf{D}}_{n,m}^{-1} [\check{\mathbf{G}}_{n,m} + \sqrt{np_{\boldsymbol{\theta}}} \cdot O_p(\delta_{nm} + h^2)] \\ &= \left[I + \left(\frac{\check{\mathbf{D}}_{n,m}}{np_{\boldsymbol{\theta}}} \right)^{-1} \cdot O_p(\delta_{nm} + h^2) \right]^{-1} \check{\mathbf{D}}_{n,m}^{-1} [\check{\mathbf{G}}_{n,m} + \sqrt{np_{\boldsymbol{\theta}}} \cdot O_p(\delta_{nm} + h^2)] \\ &= [I - O_p(\delta_{nm} + h^2) + o(\delta_{nm} + h^2)] \check{\mathbf{D}}_{n,m}^{-1} [\check{\mathbf{G}}_{n,m} + \sqrt{np_{\boldsymbol{\theta}}} \cdot O_p(\delta_{nm} + h^2)] \\ &= \check{\mathbf{D}}_{n,m}^{-1} \check{\mathbf{G}}_{n,m} + \frac{1}{\sqrt{np_{\boldsymbol{\theta}}}} O_p(\delta_{nm} + h^2). \end{aligned}$$

因此, 有 $\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m}(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0) = \check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{G}}_{n,m} + \check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} \cdot (np_{\boldsymbol{\theta}})^{-1/2} O_p(\delta_{nm} + h^2)$. 利用 Chebyshev 不等式和条件 (C11), 可得

$$\mathbb{P}(p_{\boldsymbol{\theta}}^{-1/2} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m}(\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)\|_2 > \delta)$$

$$\begin{aligned}
&= \mathbb{P}(p_{\theta}^{-1/2} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{G}}_{n,m} + \check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} \cdot (np_{\theta})^{-1/2} O_p(\delta_{nm} + h^2)\|_2 > \delta) \\
&\leq p_{\theta}^{-1} \delta^{-2} \mathbb{E} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{G}}_{n,m} + \check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} \cdot (np_{\theta})^{-1/2} O_p(\delta_{nm} + h^2)\|_2^2 \\
&\leq p_{\theta}^{-1} \delta^{-2} \cdot 2 \mathbb{E} (\|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{G}}_{n,m}\|_2^2 + \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} \cdot (np_{\theta})^{-1/2} O_p(\delta_{nm} + h^2)\|_2^2) \\
&= 2\delta^{-2} p_{\theta}^{-1} \mathbb{E} [\text{tr}(\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{G}}_{n,m} \check{\mathbf{G}}_{n,m}^T \check{\mathbf{H}}_{n,m}^{-1/2})] + 2\delta^{-2} p_{\theta}^{-1} \mathbb{E} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} \cdot (np_{\theta})^{-1/2} O_p(\delta_{nm} + h^2)\|_2^2 \\
&\leq 2\delta^{-2} p_{\theta}^{-1} \text{tr}(\check{\mathbf{H}}_{n,m}^{-1/2} \mathbb{E}(\check{\mathbf{G}}_{n,m} \check{\mathbf{G}}_{n,m}^T) \check{\mathbf{H}}_{n,m}^{-1/2}) + 2\delta^{-2} p_{\theta}^{-1} \cdot \lambda_{\max}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{H}}_{n,m}) \cdot (np_{\theta})^{-1} O_p(\delta_{nm}^2 + h^4) \\
&= 2\delta^{-2} + 2\delta^{-2} p_{\theta}^{-1} O_p(\delta_{nm}^2 + h^4) \\
&= 2\delta^{-2} [1 + o_p(1)] \\
&< c_0 \delta^{-2},
\end{aligned}$$

其中 $c_0 > 3$ 是一个常数. 注意到 $\|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} (\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)\|_2 \leq \lambda_{\min}^{1/2}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \|\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0\|_2$, 则可得

$$\begin{aligned}
\mathbb{P}(p_{\theta}^{-1/2} \|\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}\|_2 > \delta) &= \mathbb{P}\{p_{\theta}^{-1/2} \lambda_{\min}^{1/2}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \|\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}\|_2 > \lambda_{\min}^{1/2}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \delta\} \\
&\leq \mathbb{P}\{p_{\theta}^{-1/2} \|\check{\mathbf{H}}_{n,m}^{-1/2} \check{\mathbf{D}}_{n,m} (\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}^0)\|_2 > \lambda_{\min}^{1/2}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \delta\} \\
&< c_0 \lambda_{\min}^{-1}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m}) \delta^{-2}.
\end{aligned}$$

记 $\tau_{nm} = \lambda_{\min}(\check{\mathbf{D}}_{n,m} \check{\mathbf{H}}_{n,m}^{-1} \check{\mathbf{D}}_{n,m})$. 当 $\tau_{nm} \rightarrow \infty$ 时, 有 $\mathbb{P}(p_{\theta}^{-1/2} \|\hat{\boldsymbol{\theta}}^u - \boldsymbol{\theta}\|_2 > \delta) \rightarrow 0$. 综上所述, 完成了定理 3.1 的证明. $\square$

定理 3.2 的证明 记 $\check{\mathbf{H}}_{n,m}^{\text{or}} = \sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1} \check{\mathbf{U}}_i^{\text{or}}$ 和 $\check{\mathbf{D}}_{n,m}^{\text{or}} = \sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1} \check{\mathbf{U}}_i^{\text{or}}$. 对 oracle 估计直接应用定理 3.1, 只需要将定理 3.1 中的 $\check{\mathbf{H}}_{n,m}$ 和 $\check{\mathbf{D}}_{n,m}$ 分别替换为 $\check{\mathbf{H}}_{n,m}^{\text{or}}$ 和 $\check{\mathbf{D}}_{n,m}^{\text{or}}$, 则有

$$\mathbb{P}\{(p+q)^{-1/2} \|(\check{\mathbf{H}}_{n,m}^{\text{or}})^{-1/2} \check{\mathbf{D}}_{n,m}^{\text{or}} (\hat{\boldsymbol{\theta}}^{\text{or}} - \tilde{\boldsymbol{\theta}}^0)\|_2 > \delta\} < \frac{1}{\delta^2}.$$

注意到, 对任意的 $\mathbf{u} \in \mathbb{R}^{p+q}$, 都有

$$\begin{aligned}
\mathbf{u}^T \check{\mathbf{H}}_{n,m}^{\text{or}} \mathbf{u} &= \mathbf{u}^T \left(\sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1} \check{\mathbf{U}}_i^{\text{or}} \right) \mathbf{u} \\
&= \mathbf{u}^T \left(\sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1/2} \mathbf{V}_i^{-1/2} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1/2} \mathbf{V}_i^{-1/2} \check{\mathbf{U}}_i^{\text{or}} \right) \mathbf{u} \\
&= \sum_{i=1}^n (\mathbf{V}_i^{-1/2} \check{\mathbf{U}}_i^{\text{or}} \mathbf{u})^T (\mathbf{V}_i^{-1/2} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1/2}) (\mathbf{V}_i^{-1/2} \check{\mathbf{U}}_i^{\text{or}} \mathbf{u}) \\
&\leq \lambda_{\max}(\mathbf{V}_i^{-1/2} \boldsymbol{\Sigma}_i \mathbf{V}_i^{-1/2}) \sum_{i=1}^n (\mathbf{V}_i^{-1/2} \check{\mathbf{U}}_i^{\text{or}} \mathbf{u})^T (\mathbf{V}_i^{-1/2} \check{\mathbf{U}}_i^{\text{or}} \mathbf{u}) \\
&= \lambda_{\max}(\mathbf{R}_i^{-1} \mathbf{R}_i^0) \mathbf{u}^T \left(\sum_{i=1}^n (\check{\mathbf{U}}_i^{\text{or}})^T \mathbf{V}_i^{-1} \check{\mathbf{U}}_i^{\text{or}} \right) \mathbf{u} \\
&= \eta_m \mathbf{u}^T \check{\mathbf{D}}_{n,m}^{\text{or}} \mathbf{u}.
\end{aligned}$$

经一系列矩阵运算, 可得

$$\begin{aligned}
\mathbf{u}^T \check{\mathbf{H}}_{n,m}^{\text{or}} \mathbf{u} \leq \eta_m \mathbf{u}^T \check{\mathbf{D}}_{n,m}^{\text{or}} \mathbf{u} &\Rightarrow \mathbf{u}^T (\check{\mathbf{D}}_{n,m}^{\text{or}})^{-1/2} \check{\mathbf{H}}_{n,m}^{\text{or}} (\check{\mathbf{D}}_{n,m}^{\text{or}})^{-1/2} \mathbf{u} \leq \eta_m \mathbf{u}^T \mathbf{u}, \\
&\Rightarrow \mathbf{u}^T \mathbf{u} \leq \eta_m \mathbf{u}^T (\check{\mathbf{D}}_{n,m}^{\text{or}})^{1/2} (\check{\mathbf{H}}_{n,m}^{\text{or}})^{-1} (\check{\mathbf{D}}_{n,m}^{\text{or}})^{1/2} \mathbf{u},
\end{aligned}$$

$$\begin{aligned} &\Rightarrow \eta_m^{-1} \mathbf{u}^T (\check{\mathbf{D}}_{n,m}^{\text{or}})^{-1} \mathbf{u} \leq \mathbf{u}^T (\check{\mathbf{H}}_{n,m}^{\text{or}})^{-1} \mathbf{u}, \\ &\Rightarrow \eta_m^{-1} \mathbf{u}^T \check{\mathbf{D}}_{n,m}^{\text{or}} \mathbf{u} \leq \mathbf{u}^T \check{\mathbf{D}}_{n,m}^{\text{or}} (\check{\mathbf{H}}_{n,m}^{\text{or}})^{-1} \check{\mathbf{D}}_{n,m}^{\text{or}} \mathbf{u}, \quad \forall \mathbf{u} \in \mathbb{R}^{p+q}. \end{aligned}$$

因此, 可得 $\|(\check{\mathbf{H}}_{n,m}^{\text{or}})^{-1/2} \check{\mathbf{D}}_{n,m}^{\text{or}} (\hat{\boldsymbol{\theta}}^{\text{or}} - \tilde{\boldsymbol{\theta}}^0)\|_2 \geq \eta_m^{-1/2} \|\check{\mathbf{D}}_{n,m}^{\text{or}} (\hat{\boldsymbol{\theta}}^{\text{or}} - \tilde{\boldsymbol{\theta}}^0)\|_2$. 于是, 得到

$$\mathbb{P}\{\eta_m^{-1/2} \|(\check{\mathbf{D}}_{n,m}^{\text{or}})^{1/2} (\hat{\boldsymbol{\theta}}^{\text{or}} - \tilde{\boldsymbol{\theta}}^0)\|_2 > \delta\} < \frac{c_1}{\delta^2},$$

其中 $c_1 > 0$. 至此完成了定理 3.2 的证明. $\square$

定理 3.3 的证明 重新回顾惩罚加权最小二乘目标函数 (2.11), 且令 $\boldsymbol{\theta}^* = \boldsymbol{\theta}^0 + (\tau_{nm})^{-1/2} \mathbf{u}$, $\boldsymbol{\gamma}^* \in \mathbb{R}^p$, 其中 $\|\mathbf{u}\|_2 = d$. 注意到 $S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^0, \boldsymbol{\gamma}^0) = 0$. 由 Taylor 展式, 有

$$\begin{aligned} D_{n,m}(\mathbf{u}) &= Q_{n,m}(\boldsymbol{\theta}^*, \boldsymbol{\gamma}^*) - Q_{n,m}(\boldsymbol{\theta}^0, \boldsymbol{\gamma}^0) = L_{n,m}(\boldsymbol{\theta}^*) - L_{n,m}(\boldsymbol{\theta}^0) + S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^*, \boldsymbol{\gamma}^*) \\ &= (\tau_{nm})^{-1/2} \mathbf{u}^T \dot{L}_{n,m}(\boldsymbol{\theta}^0) + \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T \ddot{L}_{n,m}(\boldsymbol{\theta}^0) \mathbf{u} + S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^*, \boldsymbol{\gamma}^*) \\ &= -(\tau_{nm})^{-1/2} \mathbf{u}^T \mathbf{G}_{n,m} + \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T \mathbf{D}_{n,m} \mathbf{u} + S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^*, \boldsymbol{\gamma}^*) \\ &=: -I_1 + I_2 + S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^*, \boldsymbol{\gamma}^*), \end{aligned}$$

其中 $\dot{L}_{n,m}(\boldsymbol{\theta})$ 是 $L_{n,m}(\boldsymbol{\theta})$ 的梯度向量, $\ddot{L}_{n,m}$ 是 Jacobi 矩阵. 对于 $D_{n,m}(\mathbf{u})$ 中的第一项, 利用引理 A.2, 可得

$$I_1 = (\tau_{nm})^{-1/2} \mathbf{u}^T \sum_{i=1}^n \tilde{\mathbf{U}}_i^T \mathbf{V}_i (\tilde{\mathbf{y}}_i - \tilde{\mathbf{U}}_i \boldsymbol{\theta}^0) = (\tau_{nm})^{-1/2} \mathbf{u}^T \check{\mathbf{G}}_{n,m} + o_p(1).$$

对于 $D_{n,m}(\mathbf{u})$ 中的第二项, 利用引理 A.3, 可得

$$I_2 = \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T (\mathbf{D}_{n,m} - \check{\mathbf{D}}_{n,m} + \check{\mathbf{D}}_{n,m}) \mathbf{u} = \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T \check{\mathbf{D}}_{n,m} \mathbf{u} + o_p(1).$$

只需要证明 I_2 的主项部分能控制 I_1 的主项部分即可. 注意到

$$\mathbb{P}(|\mathbf{u}^T \check{\mathbf{G}}_{n,m}|^2 > \delta) \leq \delta^{-2} \mathbb{E}(\mathbf{u}^T \check{\mathbf{G}}_{n,m} \check{\mathbf{G}}_{n,m}^T \mathbf{u}) = \delta^{-2} \mathbf{u}^T \mathbb{E}(\check{\mathbf{G}}_{n,m} \check{\mathbf{G}}_{n,m}^T) \mathbf{u} = \delta^{-2} \mathbf{u}^T \check{\mathbf{H}}_{n,m} \mathbf{u}.$$

由此可推断 $\frac{|\mathbf{u}^T \check{\mathbf{G}}_{n,m}|}{\sqrt{\mathbf{u}^T \check{\mathbf{H}}_{n,m} \mathbf{u}}} = O_p(1)$. 想要证明 $(\tau_{nm})^{-1/2} |\mathbf{u}^T \check{\mathbf{G}}_{n,m}| \leq \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T \check{\mathbf{D}}_{n,m} \mathbf{u}$, 只需要证明

$$(\tau_{nm})^{-1/2} \sqrt{\mathbf{u}^T \check{\mathbf{H}}_{n,m} \mathbf{u}} \leq \frac{1}{2} (\tau_{nm})^{-1} \mathbf{u}^T \check{\mathbf{D}}_{n,m} \mathbf{u}.$$

由定理 3.2 的证明可知 $\sqrt{\mathbf{u}^T \check{\mathbf{H}}_{n,m} \mathbf{u}} \leq \sqrt{\eta_m} \sqrt{\mathbf{u}^T \check{\mathbf{D}}_{n,m} \mathbf{u}}$. 这时只需要满足 $\sqrt{\eta_m} \leq \frac{1}{2} \frac{\sqrt{\lambda_{\min}(\check{\mathbf{D}}_{n,m})}}{\sqrt{\tau_{nm}}} \|\mathbf{u}\|_2$, 其中 η_m 是有界量. 上述条件即为 $\eta_m \leq C \|\mathbf{u}\|_2^2$, 其中 C 是某个常数, 当 $\|\mathbf{u}\|_2$ 足够大时显然成立.

因此, 当 $\|\mathbf{u}\|_2$ 足够大时, I_1 能够被 I_2 所控制, 且 I_2 是正值, 同时对惩罚项总有 $S_{\lambda_{n,m}}(\boldsymbol{\beta}_{(n)}^*, \boldsymbol{\gamma}^*) > 0$, 则可得

$$\mathbb{P}\left\{\inf_{\|\mathbf{u}\|_2=d} D_{n,m}(\mathbf{u}) > 0\right\} \rightarrow 1.$$

这表明目标函数以概率趋于 1 存在局部最小值 $\hat{\boldsymbol{\theta}}$, 且该值存在于以 $(\tau_{nm})^{-1/2} d$ 为半径、以 $\boldsymbol{\theta}^0$ 为球心的球体内, 即 $\hat{\boldsymbol{\theta}} \in B(\boldsymbol{\theta}^0, (\tau_{nm})^{-1/2} d)$. 同时, 这也表明 $\hat{\beta}_i$ 的收敛速度都是 $(\tau_{nm})^{-1/2}$. 由引理 A.4

的证明可知, 对任意的正常数 r , 都有 $\mathbb{P}(\|\hat{\beta}_{(n)} - \beta_{(n)}^0\|_\infty > p^{-1}r) \rightarrow 0$. 利用引理 A.7, 在给定 $\|\hat{\beta}_{(n)} - \beta_{(n)}^0\|_\infty \leq p^{-1}r$ 条件下, 函数 $S_{\lambda_{n,m}}(\gamma | \hat{\beta}_{(n)})$ 存在极小值满足 $\gamma \in B(\gamma^0, r)$. 因此, 对任意的常数 $r > 0$, 当 $\tau_{nm} \rightarrow \infty$ 时, 目标函数 $Q_{n,m}(\alpha, \beta_{(n)}, \gamma)$ 总存在局部最小值 $(\hat{\alpha}, \hat{\beta}_{(n)}, \hat{\gamma})$, 且满足

$$\mathbb{P}\left(\bigcap_{1 \leq i \leq n} \{\hat{\beta}_i \in B_{\beta^0}(r)\} \cap \{\hat{\alpha} \in B_{\alpha^0}(r)\} \cap \{\hat{\gamma} \in B_{\gamma^0}(r)\}\right) \rightarrow 1,$$

其中 $B_{\beta^0}(r)$ 表示以 β^0 为球心、以 r 为半径的球, 其他符号同理.

接下来证明, 当 $\tau_{nm} \rightarrow \infty$ 时, 目标函数 $Q_{n,m}(\theta^*, \gamma^*)$ 在区域 $\{\theta^* \in B(\theta^0, (\tau_{nm})^{-1/2}d)\} \cap \{\gamma^* \in B(\gamma^0, (\tau_{nm})^{-1/2}d)\}$ 上是非凸的. 注意到, 若 $\beta_{ik}^0 = \gamma_k^0$, 则有

$$\sup_{\beta_{ik}^* \in B(\beta_{ik}^0), \gamma_k^* \in B(\gamma_k^0)} |\beta_{ik}^* - \gamma_k^*| \leq \sup_{\beta_{ik}^*} |\beta_{ik}^* - \beta_{ik}^0| + \sup_{\gamma_k^*} |\gamma_k^* - \gamma_k^0| + |\beta_{ik}^0 - \gamma_k^0| \leq 2(\tau_{nm})^{-1/2}d \rightarrow 0.$$

同理可得 $\inf_{\beta_{ik}^* \in B(\beta_{ik}^0)} |\beta_{ik}^*| \geq \|\gamma_k^0\| - (\tau_{nm})^{-1/2}d \rightarrow \|\gamma_k^0\|$, 因此有

$$\mathbb{P}\left(\sup_{\beta_{ik}^* \in B(\beta_{ik}^0), \gamma_k^* \in B(\gamma_k^0)} |\beta_{ik}^* - \gamma_k^*| \leq \inf_{\beta_{ik}^* \in B(\beta_{ik}^0)} |\beta_{ik}^*| \right) \rightarrow 1.$$

进一步地, 定义

$$\tilde{S}_{\lambda_{n,m}}(\beta_{(n)}^*, \gamma^*) = \lambda_{n,m} \sum_{k=1}^p \left\{ \sum_{i \in \mathcal{G}_k^c} |\beta_{ik}^*| + \sum_{i \in \mathcal{G}_k} |\beta_{ik}^* - \gamma_k^*| \right\},$$

并记 $\tilde{Q}_{n,m}(\theta, \gamma) = L_{n,m}(\theta) + \tilde{S}_{\lambda_{n,m}}(\beta_{(n)}, \gamma)$, 于是当 τ_{nm} 充分大时, 在 $\{\theta^* \in B(\theta^0, (\tau_{nm})^{-1/2}d)\} \cap \{\gamma^* \in B(\gamma^0, (\tau_{nm})^{-1/2}d)\}$ 处有 $Q_{n,m}(\theta^*, \gamma^*) = \tilde{Q}_{n,m}(\theta^*, \gamma^*)$, 因此 $\arg \min Q_{n,m} = \arg \min \tilde{Q}_{n,m}$. 记

$$\theta^{**} = \theta^0 + \lambda_{n,m}^{-1} \mathbf{u}, \gamma^{**} = \gamma^0 + \lambda_{n,m}^{-1} \mathbf{v},$$

类似有

$$\begin{aligned} D_{n,m}(\mathbf{u}, \mathbf{v}) &= \tilde{Q}_{n,m}(\theta^{**}, \gamma^{**}) - \tilde{Q}_{n,m}(\theta^0, \gamma^0) = L_{n,m}(\theta^{**}) - L_{n,m}(\theta^0) + \tilde{S}_{\lambda_{n,m}}(\beta_{(n)}^{**}, \gamma^{**}) \\ &= \frac{(\tau_{nm})^{1/2}}{\lambda_{n,m}} (\tau_{nm})^{-1/2} \mathbf{u}^T \dot{L}_{n,m}(\theta^0) + \frac{1}{2} \frac{\tau_{nm}}{\lambda_{n,m}^2} (\tau_{nm})^{-1} \mathbf{u}^T \ddot{L}_{n,m}(\theta^0) \mathbf{u} \\ &\quad + \lambda_{n,m} \sum_{k=1}^p \left\{ \sum_{i \in \mathcal{G}_k^c} \lambda_{n,m}^{-1} |u_{ik}| + \sum_{i \in \mathcal{G}_k} \lambda_{n,m}^{-1} |u_{ik} - v_{ik}| \right\}. \end{aligned}$$

因为 $\frac{\lambda_{n,m}}{(\tau_{nm})^{1/2}} \rightarrow \infty$, 所以有 $D_{n,m}(\mathbf{u}, \mathbf{v}) \xrightarrow{\mathbb{P}} D(\mathbf{u}, \mathbf{v})$, 其中 $D(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^p \{\sum_{i \in \mathcal{G}_k^c} |u_{ik}| + \sum_{i \in \mathcal{G}_k} |u_{ik} - v_{ik}|\}$, 在 $\{u_{ik} = 0 \mid i \in \mathcal{G}_k^c; u_{ik} = v_{ik} \mid i \in \mathcal{G}_k\}$ 取最小值. 因为 $D(\mathbf{u}, \mathbf{v})$ 是非凸函数, 所以有 $\arg \min D_{n,m}(\mathbf{u}, \mathbf{v}) \xrightarrow{\mathbb{P}} \arg \min D(\mathbf{u}, \mathbf{v})$, 进而 $\arg \min Q_{n,m}(\hat{\theta}, \hat{\gamma}) \xrightarrow{\mathbb{P}} \arg \min D(\mathbf{u}, \mathbf{v})$. 这也说明了 $\mathbb{P}(\hat{\beta}_{ik} = 0 \mid i \in \mathcal{G}_k^c) \rightarrow 1$ 和 $\mathbb{P}(\hat{\beta}_{ik} = \hat{\gamma}_k \mid i \in \mathcal{G}_k^c) \rightarrow 1$. 综上所述, 完成了定理 3.3 的证明. 根据定理 3.3, 推论 3.1 和 3.2 的证明是自然的. $\square$

定理 3.4 的证明 将参数估计 $\hat{\theta}$ 代入伪估计 $\tilde{g}(t; \theta)$, 得到非参数估计量 $\hat{g}(t; \hat{\theta})$, 即

$$\hat{g}(t; \hat{\theta}) = \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h) (y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_i - \mathbf{z}_{ij}^T \hat{\alpha}), \quad (\text{A.1})$$

其中 $w_{ij}(t, h)$ 和 $S_{n,l}(t)$ 的定义分别见 (2.6) 和 (2.7). 用 $\mathbb{E}_{t \in \mathcal{T}}[\xi_n]$ 表示给定 $t \in \mathcal{T}$ 下 ξ_n 的条件期望. 若 $\mathbb{E}(|\xi_n|^r) = O(a_n^r)$, 则记 $\xi_n = O_r(a_n)$. 利用 Cauchy-Schwarz 不等式, 容易得到 $O_r(a_n)O_r(b_n) = O_{r/2}(a_nb_n)$. 接下来计算 $S_{n,l}(t)$ 的阶数, 利用条件 (C4) 以及对核函数的假设条件 (C1)–(C3), 简单计算有

$$\mathbb{E}_{t \in \mathcal{T}}[S_{n,l}(t)] = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m \int K_h(x-t)(x-t)^l f(x) dx = a_l h^l f(t)(1 + O(h)),$$

其中 $a_l = \int u^l K(u) du$, $l = 0, 1, 2, 3$, 并且 $a_0 = 1$, $a_1 = 0$. 再有

$$S_{n,l}(t) = \mathbb{E}_{t \in \mathcal{T}}[S_{n,l}(t)] + O_4\{\{\mathbb{E}|S_{n,l}(t) - \mathbb{E}_{t \in \mathcal{T}}[S_{n,l}(t)]|^4\}^{1/4}\} = a_l h^l f(t)\{1 + O_4[h + (nmh)^{-1/2}]\}.$$

进一步可以得到

$$\begin{aligned} S_{n,0}(t) &= f(t)\{1 + O_4[h + (nmh)^{-1/2}]\}, \quad S_{n,1}(t) = hO_4\{h + (nmh)^{-1/2}\}, \\ S_{n,2}(t) &= a_2 f(t)h^2\{1 + O_4[h + (nmh)^{-1/2}]\}, \quad S_{n,3}(t) = h^3O_4\{h + (nmh)^{-1/2}\}. \end{aligned}$$

因此, 有 $S_{n,0}(t)S_{n,2}(t) - S_{n,1}^2(t) = a_2 f^2(t)h^2\{1 + O_2[h + (nmh)^{-1/2}]\}$. 对 (A.1) 进行简单计算, 有

$$\begin{aligned} \widehat{g}(t; \widehat{\boldsymbol{\theta}}) - g(t) &= \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)(y_{ij} - \mathbf{x}_{ij}^T \widehat{\boldsymbol{\beta}}_i - \mathbf{z}_{ij}^T \widehat{\boldsymbol{\alpha}}) - g(t) \\ &= \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)[\varepsilon_{ij} + g(t_{ij}) + \mathbf{x}_{ij}^T(\boldsymbol{\beta}_i^0 - \widehat{\boldsymbol{\beta}}_i) + \mathbf{z}_{ij}^T(\boldsymbol{\alpha}^0 - \widehat{\boldsymbol{\alpha}})] - g(t) \\ &= \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)\varepsilon_{ij} + \left[\sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)g(t_{ij}; \boldsymbol{\theta}) - g(t) \right] \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m w_{ij}(t; h)[\mathbf{x}_{ij}^T(\boldsymbol{\beta}_i^0 - \widehat{\boldsymbol{\beta}}_i) + \mathbf{z}_{ij}^T(\boldsymbol{\alpha}^0 - \widehat{\boldsymbol{\alpha}})] \\ &=: I_1 + I_2 + I_3. \end{aligned}$$

对于 I_2 , 根据 Taylor 展式和一些简单计算, 有

$$\begin{aligned} I_2 &= [S_{n,0}(t)S_{n,1}(t) - S_{n,2}^2(t)]^{-1} \left\{ \frac{1}{N} K_h(t_{ij} - t) \sum_{i=1}^n \sum_{j=1}^m [S_{n,2}(t) - (t_{ij} - t)S_{n,1}(t)] \right. \\ &\quad \times \left(g(t) + g'(t)(t_{ij} - t) + \frac{1}{2}g''(t)(t_{ij} - t)^2 + O(|t_{ij} - t|^3) \right) \Big\} - g(t) \\ &= \frac{1}{2}g''(t)(S_{n,0}(t)S_{n,1}(t) - S_{n,2}^2(t))^{-1}(S_{n,2}^2(t) - S_{n,1}(t)S_{n,3}(t)) + o_p(1) \\ &= \frac{1}{2}g''(t)h^2a_2 + O_p(h + (nmh)^{-1/2}). \end{aligned}$$

对于 I_3 , 由定理 3.1, 以及 $\widehat{\boldsymbol{\theta}}$ 是 $\sqrt{nm}$ 相合估计, $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2 = O_p((nm)^{-1/2})$, 利用文献 [28, 引理 2], 容易证明 $I_3 = o_p((nm)^{-1/2})$.

最后, 考虑 I_1 . 随机误差 ε_{ij} 的均值为 0, 方差为 $\sigma^2(t_{ij})$, 现计算 I_1 的方差. 记 $\delta_{nm} = (nmh)^{-1/2}$, 对于权重部分 $w_{ij}(t; h)$, 计算有

$$w_{ij}(t; h) = \frac{1}{N} K_h(t_{ij} - t) \left(\frac{a_2 f(t)h^2(1 + O_4(h + \delta_{nm})) - (t_{ij} - t)h \cdot O_4(h + \delta_{nm})}{a_2 f^2(t)h^2(1 + O_2(h + \delta_{nm}))} \right)$$

$$= \frac{1}{Nf(t)} K_h(t_{ij} - t)(1 + O_2(h + \delta_{nm})).$$

利用独立随机变量的和, 对于 I_1 的方差, 有

$$\begin{aligned} \mathbb{V}(I_1) &= \frac{1}{N^2 f^2(t)} \sum_{i=1}^n \mathbb{V} \left(\sum_{j=1}^m K_h(t_{ij} - t) \varepsilon_{ij} \right) + O_p(h + \delta_{nm}) \\ &= \frac{n}{N^2 f^2(t)} \mathbb{E} \left[\sum_{j=1}^m K_h(t_{ij} - t) \varepsilon_{ij} \right]^2 + O_p(h + \delta_{nm}) \\ &= \frac{n}{N^2 f^2(t)} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2=1}^m K_h(t_{ij_1} - t) \varepsilon_{ij_1} K_h(t_{ij_2} - t) \varepsilon_{ij_2} \right] + O_p(h + \delta_{nm}) \\ &= \frac{nm}{N^2 f^2(t)} \{ \mathbb{E}[K_h(t_{ij} - t) \varepsilon_{ij}]^2 + o_p(1) \} + O_p(h + \delta_{nm}) \\ &= \frac{nm}{N^2 f^2(t)} \mathbb{V}(\varepsilon(t)) \int K_h^2(x - t) f(x) dx + O_p(h + \delta_{nm}) \\ &= \frac{1}{nmh} \frac{\mathbb{V}(\varepsilon(t))}{f(t)} \int K^2(u) du + O_p(h + \delta_{nm}). \end{aligned}$$

综上所述, 利用大数定律可以得到

$$\sqrt{nmh} \left(\hat{g}(t; \hat{\boldsymbol{\theta}}) - g(t) - \frac{1}{2} g''(t) h^2 \int u^2 K(u) du \right) = \sqrt{nmh} I_1 + O_p(h) \xrightarrow{L} N(0, \sigma^2(t)).$$

综上所述, 完成了定理 3.4 的证明. $\square$

定理 4.1 的证明 根据惩罚函数 (2.10)、等价的带约束优化问题 (4.4) 和增广 Lagrange 函数 (4.5) 的定义, 容易证明条件 (C1')–(C3') 成立. 因此, 要证明 ADMM 算法的收敛性, 首先需要证明如下性质 (P1)–(P4) 成立.

(P1) (连续性) 增广 Lagrange 函数 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ 对于 $(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ 是连续函数.

(P2) (有界性) $\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ 有下界, 并且 $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ 是有界的.

(P3) (下降充分性) 存在常数 $C_1 > 0$, 使得对于充分大的 κ , 有下式成立:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}, \boldsymbol{\Lambda}^{(i+1)}) \\ &\geq C_1 (\|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2^2 + \|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2). \end{aligned}$$

(P4) (有界次梯度) 存在常数 $C_2 > 0$ 和次梯度向量 $\mathbf{d}^{(i)} \in \partial \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$, 其中 ∂ 代表次梯度算子, 有下式成立:

$$\|\mathbf{d}^{(i)}\|_2 \leq C_2 (\|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2 + \|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2).$$

在条件 (C2') 和 (C3') 成立时, 性质 (P1) 显然成立. 首先证明性质 (P3) 成立. 考虑到

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}, \boldsymbol{\Lambda}^{(i+1)}) \\ &= \underbrace{\mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})}_{\text{I}} \\ &\quad + \underbrace{\mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}, \boldsymbol{\Lambda}^{(i+1)})}_{\text{II}}. \end{aligned}$$

下面分别给出 I 和 II 两项的下界. 由于 $\boldsymbol{\theta}^{(i+1)}$ 是目标函数 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ 的极小值点, 因此其满足下列最优化条件:

$$\nabla_{\boldsymbol{\beta}} L_{n,m}(\boldsymbol{\theta}^{(i+1)}) + \boldsymbol{\Lambda}^{(i)} + \kappa(\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i)}) = \mathbf{0}, \quad (\text{A.2})$$

$$\nabla_{\boldsymbol{\alpha}} L_{n,m}(\boldsymbol{\theta}^{(i+1)}) = \mathbf{0}. \quad (\text{A.3})$$

因此, 简单计算有

$$\begin{aligned} & \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) \\ &= L_{n,m}(\boldsymbol{\theta}^{(i)}) - L_{n,m}(\boldsymbol{\theta}^{(i+1)}) + (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)})^T \boldsymbol{\Lambda}^{(i)} + \frac{\kappa}{2} (\|\boldsymbol{\beta}^{(i)} - \boldsymbol{\nu}^{(i)}\|_2^2 - \|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2) \\ &= L_{n,m}(\boldsymbol{\theta}^{(i)}) - L_{n,m}(\boldsymbol{\theta}^{(i+1)}) + (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)})^T \boldsymbol{\Lambda}^{(i)} + \frac{\kappa}{2} (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)})^T (\boldsymbol{\beta}^{(i)} - \boldsymbol{\nu}^{(i)} + \boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i)}) \\ &= L_{n,m}(\boldsymbol{\theta}^{(i)}) - L_{n,m}(\boldsymbol{\theta}^{(i+1)}) - \nabla_{\boldsymbol{\beta}} L_{n,m}(\boldsymbol{\theta}^{(i+1)})^T (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}) + \frac{\kappa}{2} \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2^2 \\ &\geq -\frac{L_l}{2} \|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)}\|_2^2 + \frac{\kappa}{2} \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2^2 \\ &\geq \frac{\kappa - ML_l}{2} \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2^2, \end{aligned}$$

其中 $M > 0$ 是一个常数, 第 4 个等式源于 (A.2). 考虑到 (A.3) 的限制和条件 (C2'), 第一个不等式可由下式得到:

$$\begin{aligned} & |L_{n,m}(\boldsymbol{\theta}^{(i)}) - L_{n,m}(\boldsymbol{\theta}^{(i+1)}) - \nabla_{\boldsymbol{\beta}} L_{n,m}(\boldsymbol{\theta}^{(i+1)})^T (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)})| \\ &= |L_{n,m}(\boldsymbol{\theta}^{(i)}) - L_{n,m}(\boldsymbol{\theta}^{(i+1)}) - \nabla_{\boldsymbol{\theta}} L_{n,m}(\boldsymbol{\theta}^{(i+1)})^T (\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)})| \\ &\leq \frac{L_l}{2} \|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)}\|_2^2. \end{aligned}$$

下面说明第二个不等式成立. 由 $\text{diag}(\mathbf{0}_{q \times q}, \mathbf{I}_{p \times p})\boldsymbol{\theta} = (\mathbf{0}_{1 \times q}, \boldsymbol{\beta}^T)^T = (\mathbf{0}_{1 \times q}, \boldsymbol{\nu}^T)^T$ 和条件 (C3') 可得, 存在常数 $M > 0$, 使得

$$\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)}\|_2^2 \leq M \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2^2.$$

综合上述结果, 给出了项 I 的下界.

对于项 II, 当 $\kappa \geq 2L_s$ 时, 有下列不等式成立:

$$\begin{aligned} & \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}, \boldsymbol{\Lambda}^{(i+1)}) \\ &= S_{\lambda_n, m}(\boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}) - S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}) + (\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i)})^T \boldsymbol{\Lambda}^{(i)} - (\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i+1)})^T \boldsymbol{\Lambda}^{(i+1)} \\ &\quad + \frac{\kappa}{2} (\|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 - \|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\nu}^{(i+1)}\|_2^2) \\ &= S_{\lambda_n, m}(\boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}) - S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}) - \nabla^C S_{\lambda_n, m}(\boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)})^T (\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i+1)}) \\ &\quad + \frac{\kappa}{2} \|\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i+1)}\|_2^2 - \frac{1}{\kappa} \|\boldsymbol{\Lambda}^{(i)} - \boldsymbol{\Lambda}^{(i+1)}\|_2^2 \\ &\geq -\frac{L_s}{2} (\|\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i+1)}\|_2^2 + \|\boldsymbol{\gamma}^{(i)} - \boldsymbol{\gamma}^{(i+1)}\|_2^2) + \frac{\kappa}{2} \|\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i+1)}\|_2^2 \\ &\quad - \frac{L_s^2}{\kappa} (\|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2)^2 \\ &\geq \frac{\kappa - 2L_s}{2} (\|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2) \end{aligned}$$

$$\geq 0,$$

其中第一个不等式可以由条件 (C2') 和引理 A.6 得到. 综上所述, 可以得到项 I 与项 II 之和的下界为

$$\begin{aligned} & \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}) - \mathcal{L}(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\nu}^{(i+1)}, \boldsymbol{\gamma}^{(i+1)}, \boldsymbol{\Lambda}^{(i+1)}) \\ & \geq \frac{\kappa - ML_l}{2} \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2^2 + \frac{\kappa - 2L_s}{2} (\|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2^2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2), \end{aligned}$$

其中 $\kappa > \max(ML_l, 2L_s)$, 令 $C_1(\kappa) = \max((\kappa - ML_l)/2, (\kappa - 2L_s)/2)$, 则性质 (P3) 证毕.

接下来, 证明性质 (P2) 成立. 根据引理 A.7 的证明过程, 性质 (P2) 可以验证. 继续完成性质 (P4) 的证明. 由条件 (C2') 可知, 目标函数 $\mathcal{L}$ 是广义可微的, 其梯度满足如下不等式:

$$\begin{aligned} \|\nabla_{(\boldsymbol{\nu}^T, \boldsymbol{\gamma}^T)^T} \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})\|_2 &= \|\nabla^C S_{\lambda_n, m}(\boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}) - \boldsymbol{\Lambda}^{(i)} - \kappa(\boldsymbol{\beta}^{(i)} - \boldsymbol{\nu}^{(i)})\|_2 \\ &= \|\boldsymbol{\Lambda}^{(i)} - \boldsymbol{\Lambda}^{(i-1)}\|_2 \leq L_s (\|\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i-1)}\|_2 + \|\boldsymbol{\gamma}^{(i)} - \boldsymbol{\gamma}^{(i-1)}\|_2), \\ \|\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})\|_2 &= \|\nabla L_{n, m}(\boldsymbol{\theta}^{(i)}) + \boldsymbol{\Lambda}^{(i)} + \kappa(\boldsymbol{\beta}^{(i)} - \boldsymbol{\nu}^{(i)})\|_2 \\ &= \|\nabla L_{n, m}(\boldsymbol{\theta}^{(i)}) - \nabla L_{n, m}(\boldsymbol{\theta}^{(i+1)}) + \kappa(\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)})\|_2 \\ &\leq L_l (\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)}\|_2 + \kappa \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i+1)}\|_2), \\ &\leq (L_l M + \kappa) \|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2, \\ \|\nabla_{\boldsymbol{\Lambda}} \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})\|_2 &= \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\nu}^{(i)}\|_2 = \frac{1}{\kappa} \|\boldsymbol{\Lambda}^{(i)} - \boldsymbol{\Lambda}^{(i-1)}\|_2 \\ &\leq \frac{L_s}{\kappa} (\|\boldsymbol{\nu}^{(i)} - \boldsymbol{\nu}^{(i-1)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2). \end{aligned}$$

因此, 令 $C_2(\kappa) = \max(L_s L_l M + \kappa, L_s/\kappa)$, 则可得

$$\|\nabla \mathcal{L}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})\|_2 \leq C_2(\kappa) (\|\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}\|_2 + \|\boldsymbol{\nu}^{(i+1)} - \boldsymbol{\nu}^{(i)}\|_2 + \|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2),$$

性质 (P4) 证毕.

根据性质 (P1)–(P4), 对于 ADMM 算法所产生的序列 $\{\boldsymbol{\alpha}^{(l)}, \boldsymbol{\beta}^{(l)}, \boldsymbol{\gamma}^{(l)}, \boldsymbol{\nu}^{(l)}\}_{l \geq 0}$, 容易证明上述序列至少有一个极限点, 并且每个极限点都是一个不动点. 性质 (P2) 说明序列 $\{\boldsymbol{\alpha}^{(l)}, \boldsymbol{\beta}^{(l)}, \boldsymbol{\gamma}^{(l)}, \boldsymbol{\nu}^{(l)}\}_{l \geq 0}$ 是子列收敛的, 即

$$\lim_{t \rightarrow \infty} (\boldsymbol{\alpha}^{(l_t)}, \boldsymbol{\beta}^{(l_t)}, \boldsymbol{\gamma}^{(l_t)}, \boldsymbol{\nu}^{(l_t)}, \boldsymbol{\Lambda}^{(l_t)}) = (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \boldsymbol{\nu}^*, \boldsymbol{\Lambda}^*),$$

并且每个极限点是有界的. 根据性质 (P2) 和 (P3), 增广 Lagrange 函数 $\mathcal{L}(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ 单调递减并且有下界, 进而有

$$\lim_{l \rightarrow \infty} \|\boldsymbol{e}^{(l)}\|_2^2 = 0, \lim_{l \rightarrow \infty} \|\boldsymbol{\nu}^{(l+1)} - \boldsymbol{\nu}^{(l)}\|_2^2 = 0.$$

根据性质 (P4), 有 $\lim_{l \rightarrow \infty} \|d^{(l+1)}\|_2 = 0$, 进而对于子列有 $\lim_{l \rightarrow \infty} \|d^{(l_t)}\|_2 = 0$. 根据性质 (P1) 的连续性, 有

$$\mathcal{L}(\boldsymbol{\alpha}^{(l_t)}, \boldsymbol{\beta}^{(l_t)}, \boldsymbol{\nu}^{(l_t)}, \boldsymbol{\gamma}^{(l_t)}, \boldsymbol{\Lambda}^{(l_t)}) = \mathcal{L}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\nu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\Lambda}^*),$$

因此

$$\mathbf{0} \in \partial \mathcal{L}(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \boldsymbol{\nu}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\Lambda}^{(i)}).$$

综上所述, 完成了定理 4.1 的证明. □

定理 4.2 的证明 根据定理 4.1 的结果, 并结合文献 [4, 引理 1.2] 和 [9, 定理 3.3.1], 可以证明 ADMM 算法所生成的解序列具有最优收敛速度. 进一步地, 根据性质 (P4), 可以验证次梯度是有界的, 显然可以证明次梯度序列 $\{\|d^{(i)}\|\}_{i=1}^l$ 的最优收敛速度是 $o(1/\sqrt{l})$. 最后, 由于增广 Lagrange 函数 (4.5) 是下半连续的, 并且属于半代数函数族, 因此其是一个 KL 函数. 结合文献 [1, 定理 2.9], 可以证明所提 ADMM 算法所产生的序列 $\{\alpha^{(l)}, \beta^{(l)}, \gamma^{(l)}, \nu^{(l)}\}_{l \geq 0}$ 可以全局收敛到一个极限点.

综上所述, 完成了定理 4.2 的证明. □

Multi-directional variable selection of partially linear models with longitudinal data

Han Su, Huiyao Zhu, Yijie Wang & Gaorong Li

Abstract In this paper, we investigate the problem of variable selection in personalized partially linear models for longitudinal data, integrating a multi-directional penalty function to achieve individualized variable selection and subgroup identification of heterogeneous covariates. The multi-directional penalty function allows for different groupings based on various covariate indicators, enhancing the efficiency of individualized estimation and the accuracy of variable selection. Moreover, this paper employs a profile local polynomial smoothing method for model estimation. Under some regularity conditions, we establish the asymptotic properties of parametric and nonparametric estimators, achieving variable selection and subgroup identification consistency. In this paper, we also present an efficient ADMM (alternating direction method of multipliers) algorithm to implement the proposed method, and investigate the convergence, convergence rate, and computational complexity of the proposed algorithm from both theoretical and simulation perspectives. Simulation studies show that the proposed method has better performance in terms of parameter estimation and variable selection compared with existing penalized estimators. We further apply the proposed method for analyzing the liver cirrhosis dataset to demonstrate its superior performance.

Keywords longitudinal data, personalized modeling, partially linear model, multi-directional penalty function, variable selection

MSC(2020) 62J07, 62F07, 62F12

doi: 10.1360/SSM-2025-0007