

无交叉多指标分位数回归模型中的估计与推断

董宸^{1,2}, 马舒洁³, 朱利平⁴, 冯兴东^{5*}

1. 上海财经大学经济学院, 上海 200433;
2. 上海财经大学人文学院, 上海 200433;
3. Department of Statistics, University of California at Riverside, Riverside, CA 92521, USA;
4. 中国人民大学统计与大数据研究院, 北京 100872;
5. 上海财经大学统计与管理学院, 上海 200433

E-mail: dong.chen@mail.shufe.edu.cn, shujie.ma@ucr.edu, zhu.liping@ruc.edu.cn, feng.xingdong@mail.shufe.edu.cn

收稿日期: 2018-08-02; 接受日期: 2019-07-05; 网络出版日期: 2020-02-28; * 通信作者

国家自然科学基金(批准号: 11371236, 11571218, 11422107 和 11690012)、上海财经大学创新团队和东北师范大学应用统计教育
部重点实验室(批准号: 130028906)资助项目

摘要 在过去的 30 年中分位数回归模型的研究已十分深入。然而在实际的应用场景中, 由传统估计方法所得到的分位数回归估计量, 经常会在不同分位数水平上出现互相交叉的现象, 这给分位数回归模型的实际应用造成了解释和预测上的困难。为解决这个问题, 本文提出一种带单调约束的半参数多指标分位数回归模型的研究框架。首先将半参数多指标分位数回归模型与充分降维模型相结合, 并利用两者间的联系获得指标估计量的相合估计。之后使用张量积样条方法拟合半参数模型在单调约束条件下的非参数结构。通过数值模拟的方式比较所提方法与现有可行方案所得结果在平均预测误差上的差异, 实验结果和实际案例的结果都验证了本文所提出模型的可行性。

关键词 降维 线性规划 多指标模型 无交叉 分位数回归

MSC (2020) 主题分类 62G08, 62G32, 62G20, 62H12

1 引言

传统的均值回归模型聚焦于预测变量与响应变量条件均值间关系的研究。1978 年, Koenker 和 Bassett^[1] 提出了分位数回归 (quantile regression) 模型来研究预测变量与响应变量条件分位数之间的关系, 进而可以更全面地描绘出预测变量对响应变量条件分布的影响。得益于此, 近些年来, 分位数回归模型在生物学^[2]、金融学^[3] 和经济学^[4] 等研究领域中都诞生了比较成功的应用案例。

若所掌握的信息能够帮助确定合适的参数模型, 人们通常会优先选择参数模型进行研究。然而在较为复杂的研究场合中, 我们可能需要使用非参数模型来尽可能地减少模型设定带来的误差。为此学者们研究了各种非参数分位数回归模型的性质。例如, 在预测变量数较少时, Stone^[5] 提出了 k -近邻

英文引用格式: Dong C, Ma S J, Zhu L P, et al. Estimation and inference for non-crossing multiple-index quantile regression (in Chinese). Sci Sin Math, 2021, 51: 631–658, doi: 10.1360/N012018-00193

的非参数估计模型, Fan 等^[6] 及 Yu 和 Jones^[7] 研究了局部多项式分位数回归模型, Chaudhuri^[8] 和 Koenker^[9] 则建议使用样条方法在这种情形下进行非参数逼近。随着研究场景的复杂化, 当预测变量数目较多时, 介于参数与非参数之间的半参数模型, 如单指标 (single-index) 模型就受到了更为广泛的关注。在模型估计方面, Wu 等^[10] 提出了一种回切 (back-fitting) 估计算法, Kong 和 Xia^[11] 设计了一种自适应 (adaptive) 方法, 而 Ma 和 He^[12] 则选择了一种拟剖面似然 (pseudo-profile likelihood) 的方法来估计单指标分位数回归模型, 这些方法都取得了不错的结果。此外, Zhu 等^[13] 将单指标分位数回归模型的估计问题简化成了一系列线性分位数回归模型与非参数分位数回归问题的结合, Alkenani 和 Yu^[14] 通过引入惩罚项以过滤掉不重要的预测变量, 从而将 Wu 等^[10] 的研究成果推广到了更多的应用场景中。

对于单个分位数水平上的估计量而言, 前文所提及的工作已经得到了很多研究成果。然而当需要同时估计不同分位数水平上的回归估计量时, 利用传统方法所得结果常常会在不同分位数水平上出现彼此间的交叉。这种基于有限样本的估计而产生的交叉现象则给模型的解释带来困难。为了避免这种现象的产生, He^[15] 在单一预测变量模型的估计中添加了线性约束来避免分位数回归曲线的交叉; Dette 和 Volgushev^[16] 在改进了传统的局部近邻算法的基础上, 利用额外引入的分布函数的单调性来确保所估计出的分位数回归曲线无交叉; Chernozhukov 等^[17] 则利用了分布函数的单调性设计了一种两步方法来确保无交叉性。当人们需要研究多元分位数回归模型时, Bondell 等^[18] 针对线性分位数模型和非参数分位数模型提出了一种带约束的无交叉估计方法, 并提议使用光滑样条进行拟合; 此外 Liu 和 Wu^[19] 在线性模型的研究中选取了核估计的方法; Yang 和 Tokdar^[20] 则是利用了先验 Gauss 过程构造了无交叉的线性分位数曲面。

本文研究充分降维模型和半参数多指标分位数回归模型之间的联系, 并利用这种关系构造无交叉的分位数回归估计量。本文模型起源于对非参数分位数回归模型的研究, 引入多指标模型可以丰富分位数回归在预测变量是多维甚至高维时的应用场景。同时, 半参数模型结构也放宽了文献 [18, 19] 中关于线性模型的一些假设条件。总而言之, 本文主要有 3 个贡献。

(1) 本文建立了半参数多指标分位数回归模型与充分降维方法之间的联系。这使得研究人员可以利用现有的充分降维方法^[21-23] 获得指标参数的估计量。同时, 基于该估计量的模型估计依然享有理论最优的收敛速度。

(2) 本文优化了带有单调约束的分位数回归估计算法。利用线性规划的对偶形式将原问题的计算进行了简化, 提高了整个运算过程的速度。同时这一对偶形式使得整个计算过程可并行化, 从而使得运算可以进一步加速。

(3) 本文研究单调约束条件下的半参数估计量的全局收敛性质, 证明所提出的分位数曲面的估计是一致收敛且理论最优的, 此外在给定一些较弱的假设条件下, 还研究了所提出的分位数回归估计量的弱收敛性质, 为后续的统计推断工作提供了可能。

本文的结构安排如下: 第 2 节详细介绍带单调约束的半参数分位数回归模型的两步估计方法。在第 1 步中首先估计指标参数, 在第 2 步中根据该指标参数估计量求得关于无交叉的非参数分位数回归估计量。由于需要同时在多个分位数水平上对分位数回归估计量进行求解, 第 3 节具体介绍所使用的算法。第 4 节重点研究所提出模型估计量的理论性质。之后, 在第 5 节利用数值模拟实验研究所提出模型的有限样本性质, 并利用该模型分析美国汽车销售的数据。第 6 节总结本文的研究工作。相关定理的证明可参见附录 A。本文中算法的 Matlab 代码见 <https://bb9.sufe.edu.cn/bbcswebdav/users/2011000070/Codes/NonCrossing.zip>。

2 模型假设与估计

考虑随机向量 $\mathbf{X} \in \mathbb{R}^p$ 和随机变量 $Y \in \mathbb{R}$, $F_{Y|\mathbf{X}}(y | \mathbf{x})$ 是给定 $\mathbf{X} = \mathbf{x}$ 情形下 Y 的条件分布函数, $Q_{Y|\mathbf{X}}(\tau) = \inf\{y : F_{Y|\mathbf{X}}(y | \mathbf{x}) \geq \tau\}$ 是 Y 在给定 $\mathbf{X} = \mathbf{x}$ 情形下 τ 水平的条件分位数函数. 假设对于任意 $\tau \in (0, 1)$, 存在一未知函数 m 和一载荷矩阵 $\mathbf{B}^0 \in \mathbb{R}^{p \times q}$ 使得 $Q_{Y|\mathbf{X}}(\tau) = m(\mathbf{X}^\top \mathbf{B}^0, \tau)$, 即以下等式成立:

$$\Pr\{Y \leq m(\mathbf{X}^\top \mathbf{B}^0, \tau) | \mathbf{X}\} = \tau, \quad \forall \tau \in (0, 1). \quad (2.1)$$

模型 (2.1) 是很多模型的集合, 当 $p = q = 1$ 时, 模型 (2.1) 是一个非参数分位数回归模型; 当 $p > q = 1$ 时, 模型 (2.1) 退化为一个单指标模型. 在实际应用中, 通常我们会选择 $q = 1, 2$. 在本文中, 为了避免出现分位数估计量彼此交叉的现象, 考虑分位数回归曲面 $Q_{Y|\mathbf{X}}(\tau)$ 或 $m(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 在施加单调约束情形下的估计, 即对于任意两个分位数水平 $\tau_1 \leq \tau_2$, 有 $Q_{Y|\mathbf{X}}(\tau_1) \leq Q_{Y|\mathbf{X}}(\tau_2)$, 或者 $m(\mathbf{X}^\top \mathbf{B}^0, \tau_1) \leq m(\mathbf{X}^\top \mathbf{B}^0, \tau_2)$ 成立.

定理 2.1 我们可以证明以下两个论述等价:

- (1) $Q_{Y|\mathbf{X}}(\tau) = m(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 对任意 $\tau \in (0, 1)$ 均成立;
- (2) $F_{Y|\mathbf{X}}(y | \mathbf{x}) = F_{Y|\mathbf{X}^\top \mathbf{B}^0}(y | \mathbf{x}^\top \mathbf{B}^0)$ 对任意的 $y \in \mathbb{R}^1$ 和 $\mathbf{x} \in \mathbb{R}^p$ 都成立.

定理 2.1 表明条件分位数函数与条件分布函数享有相同的降维结构. 换言之, 当解释变量 \mathbf{X} 维数 p 较高时, 充分降维结构的存在可以使我们用 $q \leq p$ 维线性组合 $\mathbf{X}^\top \mathbf{B}^0$ 去替代原来的 \mathbf{X} 进行分析, 同时该替换并不会损耗任何有关响应变量条件分位数函数或条件分布函数的信息. 易知给定 q 的情形下, 载荷矩阵也即指标参数 \mathbf{B}^0 并不唯一, 因此, 人们转而研究 \mathbf{B}^0 列向量张成的空间, 称为中心降维子空间 $\mathcal{S}(\mathbf{B}^0) = \mathcal{S}_{Y|\mathbf{X}}$. 关于如何确定中心降维子空间的维数 q 以及确保其唯一性, 可以参见充分降维文献 [24–29] 中的相关讨论. 参照 Ma 和 Zhu^[23] 的归一化条件, 假设 \mathbf{B}^0 属于以下集合:

$$\Theta = \left\{ \mathbf{B} = \begin{pmatrix} \mathbf{I}_q \\ \mathbf{B}^* \end{pmatrix} \in \mathbb{R}^{p \times q} \mid \mathbf{B}^* \in \mathbb{R}^{p-q \times q} \right\},$$

其中 \mathbf{I}_q 为 $q \times q$ 单位矩阵.

利用定理 2.1 构建了分位数回归模型 (2.1) 与充分降维模型间的联系. 该定理表明我们可以使用充分降维研究中任一具有 \sqrt{n} 阶收敛速度的系数估计量, 如 Li^[21] 的切片逆回归 (sliced inverse regression) 方法、Zhu 等^[30] 的累积切片估计 (cumulative slicing estimation) 以及 Ma 和 Zhu^[23] 所提出的半参数方法等. 在本文中, 如不特加说明, 统一记 $\hat{\mathbf{B}}$ 为 \mathbf{B}^0 的估计量.

接下来将重点讨论在给定 \mathbf{B}^0 的估计量 $\hat{\mathbf{B}}$ 时对分位数回归曲面 m 的估计. 拟用张量积样条方法去逼近 m , 记 $\hat{\mathbf{U}} = (\mathbf{X}^\top \hat{\mathbf{B}})^\top = (\hat{U}_1, \dots, \hat{U}_q)^\top$, 其中假设 d 阶正则化的 B 样条基底函数为

$$\boldsymbol{\pi}_i(U_i) = \{\pi_{j_i}(U_i), 1 \leq j_i \leq J_{n,i}\}^\top,$$

其节点序列 $\{r_{i,j_i}\}$ 满足 $r_{i,1} = \dots = r_{i,d_i} < r_{i,d_i+1} < \dots < r_{i,J_{n,i}} < r_{i,J_{n,i}+1} = \dots = r_{i,J_{n,i}+d_i}$, $i = 1, \dots, q+1$, 同时假设

$$\frac{\max_{d_i \leq j_i \leq J_{n,i}} |r_{i,j_i+1} - r_{i,j_i}|}{\min_{d_i \leq j_i \leq J_{n,i}} |r_{i,j_i+1} - r_{i,j_i}|} \leq M_i$$

关于 n 一致成立, 且 $0 < \min\{M_1, \dots, M_{q+1}\} \leq \max\{M_1, \dots, M_{q+1}\} < \infty$. 因此, $m(\hat{\mathbf{U}}, \tau)$ 可由以下张

量积样条函数逼近:

$$g_n(\hat{\mathbf{U}}, \tau) = \sum_{j_1=1}^{J_{n,1}} \cdots \sum_{j_q=1}^{J_{n,q}} \sum_{j_{q+1}=1}^{J_{n,q+1}} \lambda_{j_1, \dots, j_q, j_{q+1}} \pi_{j_1}(\hat{U}_1) \cdots \pi_{j_q}(\hat{U}_q) \pi_{j_{q+1}}(\tau) = \boldsymbol{\lambda}^\top \boldsymbol{\Pi}(\hat{\mathbf{U}}, \tau), \quad (2.2)$$

其中 $\boldsymbol{\lambda} = (\lambda_{j_1, \dots, j_q, j_{q+1}} : 1 \leq j_i \leq J_{n,i}, 1 \leq i \leq q+1)^\top$ 是节点系数, $\boldsymbol{\Pi}(\mathbf{U}, \tau) = \boldsymbol{\pi}_1(U_1) \otimes \boldsymbol{\pi}_2(U_2) \otimes \cdots \otimes \boldsymbol{\pi}_q(U_q) \otimes \boldsymbol{\pi}_{q+1}(\tau)$ 是张量积 B 样条基底函数.

为了同时估计不同分位数水平上的无交叉的分位数回归曲面, 令 $g_n(\mathbf{U}, \cdot)$ 关于 τ 单调非降, 使得对于 $0 = \tau_0 < \tau_1 < \cdots < \tau_{\kappa_n} < \tau_{\kappa_n+1} = 1$ 有 $g_n(\mathbf{U}, \tau_1) \leq g_n(\mathbf{U}, \tau_2) \leq \cdots \leq g_n(\mathbf{U}, \tau_{\kappa_n})$ 成立. 以下命题表明了 $g_n(\mathbf{U}, \cdot)$ 关于 τ 的单调性与 B 样条节点系数 $\lambda_{j_1, \dots, j_q, j_{q+1}}$ 间的关系.

命题 2.1 (2.2) 中函数 g_n 关于 τ 是单调非降的一个充分条件是 $\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} \geq \lambda_{j_1, \dots, j_q, j_{q+1}}$, $j_{q+1} = 1, \dots, J_{n,q+1} - 1$. 此外, 当所使用的 B 样条基底函数阶数不超过 2 时, 该条件也是函数 g_n 关于 τ 是单调非降的必要条件.

上述命题中的充分条件在研究中已经得到广泛的使用, Kelly 和 Rice^[31] 首先研究了 B 样条在单调约束条件下的逼近方法, He^[15] 使用该充分条件研究了一元单调分位数回归曲线的估计问题, Leitenstorfer 和 Tutz^[32] 研究了该充分条件在广义加性模型中的应用, Wu 和 Zhang^[33] 则根据该条件设计了一种张量积样条去处理二元 I 型区间删失数据 (bivariate current status data).

综上所述, 假设 (y_i, \mathbf{x}_i^\top) 是 (Y, \mathbf{X}^\top) 的一组随机样本, 通过最小化下述目标函数并考虑命题 2.1 中的单调约束来得到节点系数 $\boldsymbol{\lambda}$ 的估计:

$$L_n(\boldsymbol{\lambda}) = \sum_{i=1}^n \sum_{k=1}^{\kappa_n} \rho_{\tau_k} t(y_i - \boldsymbol{\lambda}^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \hat{\mathbf{B}}, \tau_k)), \quad (2.3)$$

其中 $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ 是“打勾”型损失函数, $0 = \tau_0 < \tau_1 < \cdots < \tau_{\kappa_n} < \tau_{\kappa_n+1} = 1$ 是一系列分位数水平序列. 我们将在下一节中具体讨论有关的算法细节.

3 模型求解与算法实现

首先为了介绍方便, 我们将 (2.3) 向量化. 记 $\boldsymbol{\lambda} = (\lambda_{1,1,\dots,1}, \dots, \lambda_{J_{n,1},J_{n,2},\dots,J_{n,q+1}})^\top$, 并令 $\mathbf{y} = (y_1, \dots, y_n)^\top \otimes \mathbf{1}_{\kappa_n \times 1}$, 其中 $\mathbf{1}_{\kappa_n \times 1}$ 是元素全为 1 的 κ_n 维向量. 设 $\mathbf{z} = (z_{1,1}, \dots, z_{1,\kappa_n}, \dots, z_{n,\kappa_n})^\top$, $\mathbf{z}^+ = \mathbf{z}I(\mathbf{z} \geq 0)$, $\mathbf{z}^- = -\mathbf{z}I(\mathbf{z} < 0)$, $I(A)$ 是一个关于 A 的示性函数. 显然由正、负部的定义可知, $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$ 且 $|\mathbf{z}| = \mathbf{z}^+ + \mathbf{z}^-$. 令 $\boldsymbol{\tau} = \mathbf{1}_{n \times 1} \otimes (\tau_1, \dots, \tau_{\kappa_n})^\top$, 同时记

$$\boldsymbol{\Pi} = (\boldsymbol{\Pi}(\mathbf{x}_1^\top \hat{\mathbf{B}}, \tau_1), \dots, \boldsymbol{\Pi}(\mathbf{x}_1^\top \hat{\mathbf{B}}, \tau_{\kappa_n}), \boldsymbol{\Pi}(\mathbf{x}_2^\top \hat{\mathbf{B}}, \tau_1), \dots, \boldsymbol{\Pi}(\mathbf{x}_2^\top \hat{\mathbf{B}}, \tau_{\kappa_n}), \dots, \boldsymbol{\Pi}(\mathbf{x}_n^\top \hat{\mathbf{B}}, \tau_{\kappa_n}))^\top.$$

在不等式 $\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} \geq \lambda_{j_1, \dots, j_q, j_{q+1}} (j_{q+1} = 1, \dots, J_{n,q+1} - 1)$ 成立时, 最小化 (2.3) 等价于求解以下线性规划问题:

$$\begin{aligned} & \min_{\boldsymbol{\lambda}, \mathbf{z}^+, \mathbf{z}^-} \{ \boldsymbol{\tau}^\top \mathbf{z}^+ + (1 - \boldsymbol{\tau})^\top \mathbf{z}^- \} \\ & \text{s.t. } \mathbf{y} - \boldsymbol{\Pi} \boldsymbol{\lambda} = \mathbf{z}^+ - \mathbf{z}^-, \\ & \quad \mathbf{R}^\top \boldsymbol{\lambda} \geq \mathbf{0}, \\ & \quad \mathbf{z}^+ \geq \mathbf{0}, \quad \mathbf{z}^- \geq \mathbf{0}, \end{aligned} \quad (3.1)$$

其中 $\mathbf{R} = \mathbf{I}_{q_n} \otimes \mathbf{A}$, $q_n = (J_{n,1} J_{n,2} \cdots J_{n,q}) \times (J_{n,1} J_{n,2} \cdots J_{n,q})$, 而 \mathbf{A} 是一个 $J_{n,q+1} \times (J_{n,q+1} - 1)$ 带状矩阵, 其中元素 a_{ij} 当 $i = j$ 时取值 -1 , 当 $i = j + 1$ 时取值 1 , 此外均取值为 0 .

不难看出原问题 (3.1) 需要考虑一个由 $\boldsymbol{\lambda}$ 、 \mathbf{z}^+ 与 \mathbf{z}^- 组成的长向量, 同时还需要考虑维数较高的矩阵 \mathbf{R} 所代表的单调约束, 因此, 我们拟使用其对偶问题来提高计算效率. 定义一个 $\{J_{n,1} J_{n,2} \cdots (J_{n,q+1} + 1)\}$ 维的向量 $\boldsymbol{\vartheta} = (\vartheta_{j_1, j_2, \dots, j_{q+1}})^\top$, 其中元素满足下列等式:

$$\begin{cases} \vartheta_{j_1, \dots, j_q, 1} - \vartheta_{j_1, \dots, j_q, 2} = \lambda_{j_1, \dots, j_q, 1}, \\ \vartheta_{j_1, \dots, j_q, 3} = \lambda_{j_1, \dots, j_q, 2} - \lambda_{j_1, \dots, j_q, 1}, \\ \vdots \\ \vartheta_{j_1, \dots, j_q, (J_{n,q+1}+1)} = \lambda_{j_1, \dots, j_q, (J_{n,q+1})} - \lambda_{j_1, \dots, j_q, (J_{n,q+1}-1)}. \end{cases}$$

在这种情形下, 约束 $\mathbf{R}^\top \boldsymbol{\lambda} \geq \mathbf{0}$ 可被简化为 $\boldsymbol{\vartheta} \geq \mathbf{0}$. 假设存在一个矩阵 \mathbf{K} 使得 $\mathbf{K}\boldsymbol{\vartheta} = \boldsymbol{\lambda}$ 成立, 通过引入 Lagrange 乘子 $\boldsymbol{\alpha}$ 、 $\boldsymbol{\gamma}$ 、 $\boldsymbol{\delta}$ 和 $\boldsymbol{\nu}$, 我们可通过下式求得 (3.1) 的对偶形式:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\nu}) &= \inf_{\boldsymbol{\vartheta}, \mathbf{z}^+, \mathbf{z}^-} \{ \boldsymbol{\tau}^\top \mathbf{z}^+ + (1 - \boldsymbol{\tau})^\top \mathbf{z}^- + \boldsymbol{\alpha}^\top (\mathbf{y} - \boldsymbol{\Pi} \mathbf{K} \boldsymbol{\vartheta} - \mathbf{z}^+ + \mathbf{z}^-) \\ &\quad - \boldsymbol{\gamma}^\top \boldsymbol{\vartheta} - \boldsymbol{\delta}^\top \mathbf{z}^+ - \boldsymbol{\nu}^\top \mathbf{z}^- \} \\ \text{s.t. } \boldsymbol{\vartheta} &\geq \mathbf{0}, \quad \mathbf{z}^+ \geq \mathbf{0}, \quad \mathbf{z}^- \geq \mathbf{0}, \quad \boldsymbol{\gamma} \geq \mathbf{0}, \quad \boldsymbol{\delta} \geq \mathbf{0}, \quad \boldsymbol{\nu} \geq \mathbf{0}. \end{aligned}$$

根据文献 [34] 中的 Karush-Kuhn-Tucher (KKT) 条件, 可以进一步得到

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\nu}) = \begin{cases} \mathbf{y}^\top \boldsymbol{\alpha}, & \text{若 } \mathbf{K}^\top \boldsymbol{\Pi}^\top \boldsymbol{\alpha} + \boldsymbol{\gamma} = \mathbf{0}, \quad \boldsymbol{\tau} - \boldsymbol{\alpha} = \boldsymbol{\delta}, \quad \boldsymbol{\alpha} + \boldsymbol{\tau} - \mathbf{1} = \boldsymbol{\nu}, \\ -\infty, & \text{其他情形.} \end{cases}$$

因此, 原问题 (3.1) 的对偶问题为

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \mathbf{y}^\top \boldsymbol{\alpha} \\ \text{s.t. } &\mathbf{K}^\top \boldsymbol{\Pi}^\top \boldsymbol{\alpha} \leq \mathbf{0}, \\ &\boldsymbol{\tau} - \mathbf{1} \leq \boldsymbol{\alpha} \leq \boldsymbol{\tau}, \end{aligned} \tag{3.2}$$

其中 $\boldsymbol{\alpha}$ 是 $n\kappa_n$ 维向量. 由对偶理论可知, 原问题最优解中每个变量的取值对应其影子价格 (shadow price), 参见文献 [34]. 此外, 得益于约束条件的“箱型”形式, 对偶问题 (3.2) 的求解还可以利用并行计算实现进一步的加速.

4 模型理论研究

在讨论模型估计量理论性质之前, 首先引入以下符号: 对于任意正实数序列 $\{a_n\}_{n=1}^\infty$ 和 $\{b_n\}_{n=1}^\infty$, $a_n \asymp b_n$ 表示 $\lim_{n \rightarrow \infty} (a_n/b_n) = c$, $c > 0$. 对于某矩阵 $\mathbf{G} = (G_{ij})$, 记 $\|\mathbf{G}\|_\infty = \max_{i,j}(|G_{ij}|)$, 其中 G_{ij} 是矩阵的第 i 行 j 列元素. 假设 \mathbf{X} 有紧支撑集 \mathcal{X} , $\mathbf{U} = (\mathbf{X}^\top \mathbf{B}^0)^\top$ 的支撑集为 \mathcal{U} , 为了简化符号, 记 $\mathcal{B} = \{(\mathbf{U}, \tau) : \mathbf{U} \in \mathcal{U}, (\kappa_n + 1)^{-1} \leq \tau \leq 1 - (\kappa_n + 1)^{-1}\}$. 同时对于任意函数 $f(\mathbf{U}, \tau)$, 定义 $\|f\|_\infty = \sup_{\mathcal{B}} |f(\mathbf{U}, \tau)|$. 不失一般性, 假设以下条件成立:

(A1) 函数 $m(\mathbf{U}, \tau)$ 具有连续 a 阶混合偏导数, 即存在

$$\nabla^a m(\mathbf{U}, \tau) = \frac{\partial^a m(\mathbf{U}, \tau)}{\partial U_1^{w_1} \cdots \partial U_q^{w_q} \partial \tau^{w_{q+1}}},$$

这里 $0 \leq w_i \leq a$, $\sum_i w_i = a$ 对 $i = 1, \dots, q+1$ 都成立且有 $a \geq q+1$ 与 $\min(d_1, \dots, d_{q+1}) \geq a+2$ 成立.

(A2) 令 $\{J_n\}_{n=1}^\infty$ 为随着 n 趋于无穷也趋于无穷的发散序列, 此外有 $n^{-1/2} J_n^{3(q+1)} \log n = o(1)$, $\kappa_n^{-2} J_n^{q+3} = o(1)$ 和 $J_n \asymp J_{n,i}$ ($i = 1, \dots, q+1$) 成立. 需要指出的是, 上述条件在分位数回归模型的研究中广为使用, 参见文献 [35].

其中假设 (A1) 是关于分位数回归曲线/曲面光滑性的假定, 假设 (A2) 是关于所使用 B 样条函数光滑性的假定.

假设 $\widehat{\mathbf{B}}$ 是 \mathbf{B}^0 的 \sqrt{n} 阶相合估计, $\widehat{\boldsymbol{\lambda}}(\widehat{\mathbf{B}})$ 是在给定指标参数估计 $\widehat{\mathbf{B}}$ 时, 由第 3 节中的单调约束优化算法得到的节点系数估计. 记 $\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) = \widehat{\boldsymbol{\lambda}}(\widehat{\mathbf{B}})^\top \boldsymbol{\Pi}(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$. 关于 $\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$ 的全局收敛性质, 可证明以下结论成立.

定理 4.1 假设 (A1) 和 (A2) 成立, 在定理 2.1 成立的前提下, 对于 q 指标无交叉分位数回归函数 \widehat{g}_n , 有

$$\sup_{\mathcal{A}} |\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)| = O_p(n^{-1/2} J_n^{1+q/2} + \kappa_n^{-1} J_n^{(q+3)/2} + J_n^{-a+(q+1)/2})$$

成立, 其中 $\mathcal{A} = \{(\mathbf{X}, \tau) : \mathbf{X} \in \mathcal{X}, (\kappa_n + 1)^{-1} \leq \tau \leq 1 - (\kappa_n + 1)^{-1}\}$.

由定理 4.1 的证明过程可知 $\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$ 的收敛速度已经达到理论最优, 即在给定 \mathbf{B}^0 的一个 \sqrt{n} 阶相合估计 $\widehat{\mathbf{B}}$ 前提下, 所得到的非参数分位数回归曲线/曲面估计与给定 \mathbf{B}^0 所得到的非参数分位数回归曲线/曲面估计拥有相同的收敛速度.

在很多应用场合中, 人们往往还关心对未知函数 m 的统计推断结果, 此时需要考察所提估计量 \widehat{g}_n 的弱收敛性质. 假设所研究的分位数水平同在一个紧集之内, 且有以下假设成立.

(A3) 存在常数 $\underline{f} < \bar{f}$ 使得条件密度函数 $f(m(\mathbf{X}^\top \mathbf{B}^0, \tau) | \mathbf{X})$ 满足 $\underline{f} \leq f(m(\mathbf{X}^\top \mathbf{B}^0, \tau) | \mathbf{X}) \leq \bar{f}$ 对于所有 $(\mathbf{X}, \tau) \in \mathcal{C}$ 一致成立, 其中

$$\mathcal{C} = \{(\mathbf{X}, \tau) : \mathbf{X} \in \mathcal{X}, \epsilon_1 \leq \tau \leq 1 - \epsilon_2, \epsilon_1, \epsilon_2 > 0\}.$$

(A4) 矩阵 \mathbf{D}_n 的最小特征根对所有的样本量 n 都大于 0, 其中

$$\mathbf{D}_n = \frac{1}{n} \sum_{i=1}^n \int_{\epsilon_1}^{1-\epsilon_2} f(m(\mathbf{x}_i^\top \mathbf{B}^0, \tau) | \mathbf{x}_i) \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau) \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau)^\top d\tau.$$

这里的假设 (A3) 通常被用于控制响应变量概率密度函数的上下界, 而假设 (A4) 则常用于保证估计量 Bahadur 表达式的存在.

附录的引理 A.1 证明了存在一函数 $g_n^0(\mathbf{U}, \tau) = \boldsymbol{\lambda}^0^\top \boldsymbol{\Pi}(\mathbf{U}, \tau)$ 使得

$$\sup_{\mathbf{U}, \tau} |g_n^0(\mathbf{U}, \tau) - m(\mathbf{U}, \tau)| = O_p(J_n^{-a}).$$

记 $Z_n(\mathbf{X}, \tau) = \widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - g_n^0(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$, 则可证如下定理成立:

定理 4.2 如果假设 (A1)–(A4) 成立, 则 Z_n 收敛于一均值为 0 且协方差核为

$$C_n((\mathbf{X}_1^\top, \tau_1), (\mathbf{X}_2^\top, \tau_2)) = \frac{\boldsymbol{\Pi}(\mathbf{X}_1^\top \mathbf{B}^0, \tau_1)^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}_2^\top \mathbf{B}^0, \tau_2)}{n}$$

的 Gauss 过程, 其中 $(\mathbf{X}_1^\top, \tau_1)^\top, (\mathbf{X}_2^\top, \tau_2)^\top \in \mathcal{C}$, 矩阵 \mathbf{D}_n 已经在假设 (A4) 中给出定义,

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n \int_{\epsilon_1 \leq \tau_1, \tau_2 \leq 1-\epsilon_2} (\min\{\tau_1, \tau_2\} - \tau_1 \tau_2) \mathbf{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_1) \mathbf{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_2)^\top d\tau_1 d\tau_2.$$

如引理 A.1 中所展示的那样, 当 m 足够光滑时, g_n^0 一致趋近于 m . 因为有 $\|\widehat{\mathbf{B}} - \mathbf{B}^0\| = O_p(n^{-1/2})$ 成立, 我们可知 $g_n^0(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$ 充分趋近于 $m(\mathbf{X}^\top \mathbf{B}^0, \tau)$, 其中 $\tau \in (0, 1)$. 结合定理 4.2, 我们可以对函数 m 进行相关的统计推断工作.

5 实证分析

本节将通过数值模拟实验和实际案例分析, 重点研究所提出模型估计量的有限样本性质; 在数值模拟部分设计了 3 种数据生成方案: 方差齐次的单指标模型、方差非齐次的单指标模型和多指标模型, 重点考察所提出方法的平均预测误差; 在汽车销量数据分析的实际案例中, 将展现分位数曲线交叉对模型解释性造成的问题, 并比较不同模型方案的统计推断结果.

5.1 数值模拟

在 Monte Carlo 模拟实验中, 固定预测变量的维数为 $p = 20$, 训练集的样本大小为 $n = 400$. 从标准多元正态分布中生成 \mathbf{x}_i , 同时分别从 (i) 标准正态分布、(ii) 自由度为 3 的 t 分布、(iii) 均值为 1 的指数分布中生成噪声项 ε_i , 选用以下 3 种数据生成模型:

$$y_i = \cos\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) + \exp\left(\frac{-|\mathbf{x}_i^\top \boldsymbol{\beta}|}{2}\right) + \varepsilon_i, \quad (5.1)$$

$$y_i = \exp(-|\mathbf{x}_i^\top \boldsymbol{\beta}|) + \{\exp(-|\mathbf{x}_i^\top \boldsymbol{\beta}|)\}^{1/2} \varepsilon_i, \quad (5.2)$$

$$y_i = \frac{3 \exp(-|\mathbf{x}_i^\top \boldsymbol{\beta}_1|)}{\exp(-|\mathbf{x}_i^\top \boldsymbol{\beta}_2|) + 1} + 0.3 \varepsilon_i. \quad (5.3)$$

在模型 (5.1) 和 (5.2) 中, 令 $\mathbf{B}^0 = \boldsymbol{\beta} = (1, -\mathbf{1}_{1 \times 7}, \mathbf{1}_{1 \times 7}, \mathbf{0}_{1 \times 5})^\top$; 在模型 (5.3) 中, 令 $\mathbf{B}^0 = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, 其中

$$\boldsymbol{\beta}_1 = (1, 0, \mathbf{1}_{1 \times 4} \otimes (8^{-1/2}, -8^{-1/2}), \mathbf{0}_{1 \times 10})^\top, \quad \boldsymbol{\beta}_2 = (0, 1, 8^{-1/2} \otimes \mathbf{1}_{1 \times 8}, \mathbf{0}_{1 \times 10})^\top.$$

用文献 [36] 中基于 $(Y | \mathbf{X})$ 的条件密度函数最小平均方差估计法 (minimum average (conditional) variance estimation based on the conditional density function, dMAVE) 求得关于 \mathbf{B}^0 的 \sqrt{n} 阶估计 $\widehat{\mathbf{B}}$. 在前两个模型中, 比较由下述 5 种方法所计算得到的分位数曲线:

- (i) 所提出模型的无约束估计量 (给定 $\widehat{\mathbf{B}}$), 即从原问题 (3.1) 中去除单调约束 $\mathbf{R}^\top \boldsymbol{\lambda} \geq \mathbf{0}$ 后所求得的样条估计量, 记为 “NEW.WO($\widehat{\mathbf{B}}$)”;
- (ii) 所提出模型的无交叉样条估计量 (给定 $\widehat{\mathbf{B}}$), 记为 “NEW($\widehat{\mathbf{B}}$)”;
- (iii) 所提出模型的无交叉样条估计量 (给定 \mathbf{B}^0), 记为 “NEW(\mathbf{B}^0)”;
- (iv) 文献 [16] 所提出的核估计方法 (给定 $\widehat{\mathbf{B}}$), 记为 “DV($\widehat{\mathbf{B}}$)”;
- (v) 文献 [16] 所提出的核估计方法 (给定 \mathbf{B}^0), 记为 “DV(\mathbf{B}^0)”;
- (vi) 文献 [17] 所提出的重排序方法 (给定 $\widehat{\mathbf{B}}$), 记为 “CFG($\widehat{\mathbf{B}}$)”;
- (vii) 文献 [17] 所提出的重排序方法 (给定 \mathbf{B}^0), 记为 “CFG(\mathbf{B}^0)”.

其中方法 (ii) 是建议在实际应用中使用的方法; 方法 (i) 是本文所设计的多指标分位数回归模型的无

单调约束版本, 希望通过比较前两种方法的差异来研究加入单调性约束对实际拟合结果的影响; 由于在实际应用中我们无法获取 \mathbf{B}^0 , 所以方法 (iii) 被用作数据拟合结果的评比基准, 通过比较方法 (ii) 与 (iii) 中的结果来研究降维矩阵估计 $\hat{\mathbf{B}}$ 对数据拟合结果的影响. Dette 和 Volgushev^[16] 的方法是一类基于核估计的非参数方法, 通过引入一个全局函数分布 G 使所得非参数估计关于 τ 满足单调性的约束. 然而所选分布函数 G 与核估计的窗宽需要依靠人为选取, 所得非参数分位数回归的估计因此存在较大波动; Chernozhukov 等^[17] 也是利用分布函数天然的单调递增性设计出了一种通过对不满足约束条件的估计量进行重新排序, 从而使最终估计满足约束条件的方案.

值得注意的是, 文献 [16,17] 中所提出的模型都只针对一维预测变量的情形, 这里为了将其结论拓展到多维情形, 拟先使用降维模型将多维分位数回归模型转化成无交叉单指标的分位数回归模型, 再使用这两种方法获得单指标分位数回归曲线的估计. 因此, 只在比较单指标模型时比较所提及的 7 种方法, 在多指标模型中我们仅比较方法 (i)–(iii) 的结果.

从标准多元正态分布中生成额外一组样本量为 100 的数据用作预测数据集, 并比较不同方法所得的平均预测误差 (mean prediction error, MPE)

$$\text{MPE}_k = \frac{1}{100} \sum_{i=1}^{100} \{\hat{g}_n(\mathbf{x}_i^\top \hat{\mathbf{B}}, \tau_k) - m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)\}^2.$$

所选分位数水平网格为

$$\tau_k = \{0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.90, 0.95, 0.975\},$$

Monte Carlo 实验的样本量为 500, 我们将 MPE 的均值与标准差列在表 1–3 中.

表 1 的第 1 和 2 列结果体现了添加无交叉约束对于模型估计的重要性, 添加单调约束与无约束的分位数回归估计量之间的差异在厚尾情形下体现得较为明显, 如噪声分布为 t 分布和指数分布时的情形. 利用本文所提出的模型, 人们可以在避免分位数回归曲线交叉的同时提升预测结果; 第 2 和 3 列结果表明了, 尽管在理论上给定降维矩阵估计量 $\hat{\mathbf{B}}$ 与给定真实降维矩阵 \mathbf{B}^0 所得到的估计量收敛速度相同, 但是在有限样本的表现中, 我们还是可以发现 \mathbf{B}^0 估计的准确性依然会影响到预测结果; 比较表 1 的第 2、4、6 列与第 3、5、7 列结果, 我们可以注意到本文所提出的方法与文献 [17] 有更为相近的结果表现, 同时两者的结果在各个分位数水平上也要一致优于文献 [16] 中方法的结果. 由上述分析可知, 在实际研究场景中, 我们可以通过改善关于 \mathbf{B}^0 的估计来提升模型的预测效果, 同时还可以通过添加单调约束来避免分位数回归估计的相互交叉, 进而获得更好的预测结果.

表 2 中异方差的对比结果与表 1 的结果类似, 本文所提出的模型在误差项是正态分布时拥有最优的平均误差结果; 同时对比基于降维矩阵估计量 $\hat{\mathbf{B}}$ 的估计与基于 \mathbf{B}^0 的估计, 我们可以注意到两者间的差距已经非常小. 注意到文献 [17] 中的方法在极端分位数水平上有更稳健的结果, 这主要归因于文献 [17] 的方法利用对回归估计量的重新排序来满足约束条件, 而并不改变已满足约束的估计, 然而当样本量很少时两者间的差异并不明显, 同时两种方法的结果均一致好于文献 [16] 中方法的结果. 总体而言, 由表 2 的结果可知, 使用本文所提出的模型可以获得不错的结果.

从表 3 可以看出本文所提出的模型在多指标情形下也可以给出较好的预测结果. 在二维场合下, 样本点分布更为分散, 因此通过充分利用全局信息来修正分位数回归曲面的交叉现象, 可以得到预测误差更小的结果, 同时更准确的降维矩阵估计也可以进一步提升模型的预测效果.

当人们需要使用分位数回归模型来构建相关的预测区间时, 所选用的方法在分布尾部的表现就显得尤为重要, 而由于样本点在尾部的数目通常十分有限, 因此, 分位数回归函数估计量间出现的交叉

表 1 模型 (5.1) MPE_k 的均值与标准差

噪声项	τ	NEW.WO(\hat{B})	NEW(\hat{B})	NEW(B^0)	DV(\hat{B})	DV(B^0)	CFG(\hat{B})	CFG(B^0)
$\mathcal{N}(0, 1)$	0.025	0.358 (0.011)	0.325 (0.010)	0.117 (0.003)	0.990 (0.017)	0.703 (0.005)	0.300 (0.006)	0.159 (0.003)
	0.050	0.265 (0.008)	0.250 (0.008)	0.071 (0.002)	0.612 (0.012)	0.364 (0.003)	0.255 (0.007)	0.098 (0.002)
	0.100	0.223 (0.007)	0.214 (0.007)	0.052 (0.001)	0.354 (0.008)	0.153 (0.002)	0.220 (0.006)	0.060 (0.001)
	0.250	0.190 (0.006)	0.187 (0.006)	0.036 (0.001)	0.203 (0.006)	0.051 (0.001)	0.190 (0.006)	0.039 (0.001)
	0.500	0.182 (0.005)	0.181 (0.005)	0.033 (0.001)	0.189 (0.006)	0.042 (0.001)	0.182 (0.005)	0.034 (0.001)
	0.750	0.191 (0.005)	0.188 (0.005)	0.036 (0.001)	0.273 (0.009)	0.088 (0.001)	0.194 (0.005)	0.040 (0.001)
	0.900	0.221 (0.005)	0.213 (0.005)	0.054 (0.001)	0.613 (0.016)	0.280 (0.003)	0.230 (0.005)	0.066 (0.001)
	0.950	0.252 (0.006)	0.239 (0.006)	0.073 (0.002)	1.044 (0.021)	0.585 (0.004)	0.273 (0.005)	0.110 (0.002)
	0.975	0.339 (0.008)	0.302 (0.007)	0.119 (0.003)	1.588 (0.026)	1.028 (0.006)	0.336 (0.005)	0.181 (0.003)
	$t(3)$	0.025	0.962 (0.060)	0.709 (0.041)	0.656 (0.041)	1.003 (0.009)	0.862 (0.006)	0.367 (0.007)
	0.050	0.289 (0.011)	0.241 (0.009)	0.192 (0.009)	0.423 (0.007)	0.315 (0.004)	0.185 (0.003)	0.130 (0.003)
	0.100	0.117 (0.002)	0.105 (0.002)	0.052 (0.001)	0.207 (0.005)	0.132 (0.005)	0.110 (0.002)	0.057 (0.001)
	0.250	0.075 (0.001)	0.074 (0.001)	0.021 (0.000)	0.132 (0.003)	0.076 (0.002)	0.074 (0.001)	0.022 (0.000)
	0.500	0.068 (0.001)	0.068 (0.001)	0.016 (0.000)	0.084 (0.001)	0.032 (0.001)	0.068 (0.001)	0.016 (0.000)
	0.750	0.075 (0.001)	0.074 (0.001)	0.022 (0.000)	0.101 (0.001)	0.047 (0.001)	0.076 (0.001)	0.023 (0.000)
	0.900	0.119 (0.002)	0.107 (0.002)	0.054 (0.001)	0.253 (0.003)	0.149 (0.002)	0.112 (0.002)	0.056 (0.001)
	0.950	0.283 (0.012)	0.238 (0.008)	0.204 (0.010)	0.619 (0.006)	0.450 (0.003)	0.191 (0.003)	0.131 (0.002)
	0.975	0.944 (0.067)	0.688 (0.038)	0.708 (0.051)	1.412 (0.010)	1.175 (0.006)	0.381 (0.006)	0.318 (0.005)
	$Exp(1)$	0.025	0.425 (0.005)	0.418 (0.005)	0.141 (0.001)	0.239 (0.004)	0.047 (0.001)	0.318 (0.004)
	0.050	0.322 (0.004)	0.318 (0.004)	0.116 (0.001)	0.228 (0.003)	0.042 (0.001)	0.269 (0.004)	0.100 (0.001)
	0.100	0.237 (0.003)	0.234 (0.003)	0.087 (0.001)	0.212 (0.003)	0.038 (0.001)	0.219 (0.003)	0.080 (0.000)
	0.250	0.191 (0.002)	0.191 (0.002)	0.071 (0.000)	0.204 (0.003)	0.049 (0.001)	0.192 (0.002)	0.073 (0.000)
	0.500	0.207 (0.002)	0.207 (0.002)	0.083 (0.001)	0.253 (0.003)	0.108 (0.001)	0.207 (0.002)	0.083 (0.001)
	0.750	0.247 (0.003)	0.242 (0.003)	0.119 (0.001)	0.278 (0.003)	0.131 (0.002)	0.240 (0.003)	0.116 (0.001)
	0.900	0.368 (0.005)	0.343 (0.005)	0.211 (0.004)	0.720 (0.006)	0.476 (0.004)	0.356 (0.004)	0.215 (0.003)
	0.950	0.563 (0.012)	0.509 (0.010)	0.372 (0.009)	1.802 (0.013)	1.458 (0.010)	0.551 (0.007)	0.399 (0.006)
	0.975	1.074 (0.029)	0.911 (0.025)	0.752 (0.023)	3.712 (0.021)	3.271 (0.016)	0.913 (0.012)	0.750 (0.012)

现象也更加频繁地出现。通过考虑单调性约束，我们能更加有效地利用全局信息来提高估计和预测效果。可以由表 1–3 中的结果观测到这一现象。图 1–3 展示了由传统方法和本文所提方法求得的分位数回归曲面，由此可以观察到通过添加单调约束可以有效去除分位数回归曲面的交叉现象。

5.2 实例分析

本节通过一个实例分析来展现本文所提出的模型在实际案例中的具体应用。研究的数据集由 2004 年中美国某地区汽车和卡车的价格、销量等数据所组成。具体的数据集的描述信息可以在网站 http://www.amstat.org/publications/jse/jse_data_archive.htm 上找到。这个数据集包含 19 个可供研究的指标，包含汽车的品牌、型号、售价、经销商的花费、引擎大小、马力大小、汽车重量、轴距、长度和宽度等。希望研究有关汽车的其余信息对汽车售价的影响。同时不只关心汽车其余信息对汽车售价均值的影响，更感兴趣哪些因素对于高售价的汽车价格影响更大，什么因素更吸引愿意购买低价汽车的用

表 2 模型 (5.2) MPE_k 的均值与标准差

噪声项	τ	NEW.WO($\hat{\mathbf{B}}$)	NEW($\hat{\mathbf{B}}$)	NEW(\mathbf{B}^0)	DV($\hat{\mathbf{B}}$)	DV(\mathbf{B}^0)	CFG($\hat{\mathbf{B}}$)	CFG(\mathbf{B}^0)
$\mathcal{N}(0, 1)$	0.025	0.030 (0.001)	0.026 (0.001)	0.019 (0.001)	0.150 (0.002)	0.129 (0.001)	0.032 (0.001)	0.027 (0.001)
	0.050	0.014 (0.000)	0.013 (0.000)	0.010 (0.000)	0.075 (0.001)	0.061 (0.001)	0.016 (0.000)	0.015 (0.001)
	0.100	0.008 (0.000)	0.008 (0.000)	0.007 (0.000)	0.028 (0.001)	0.021 (0.000)	0.009 (0.000)	0.009 (0.000)
	0.250	0.011 (0.000)	0.011 (0.000)	0.009 (0.000)	0.011 (0.000)	0.008 (0.000)	0.011 (0.000)	0.009 (0.000)
	0.500	0.023 (0.000)	0.023 (0.000)	0.015 (0.000)	0.025 (0.000)	0.013 (0.000)	0.023 (0.000)	0.016 (0.000)
	0.750	0.049 (0.001)	0.049 (0.001)	0.025 (0.000)	0.059 (0.001)	0.029 (0.001)	0.050 (0.001)	0.027 (0.000)
	0.900	0.084 (0.001)	0.083 (0.001)	0.040 (0.000)	0.147 (0.002)	0.082 (0.001)	0.086 (0.001)	0.043 (0.001)
	0.950	0.115 (0.001)	0.113 (0.001)	0.051 (0.001)	0.250 (0.003)	0.160 (0.002)	0.114 (0.001)	0.060 (0.001)
	0.975	0.166 (0.003)	0.161 (0.003)	0.070 (0.001)	0.379 (0.004)	0.271 (0.003)	0.144 (0.002)	0.079 (0.001)
	$t(3)$	0.025	0.118 (0.009)	0.098 (0.007)	0.100 (0.008)	0.187 (0.006)	0.167 (0.002)	0.051 (0.001)
	0.050	0.032 (0.002)	0.030 (0.002)	0.029 (0.002)	0.069 (0.005)	0.055 (0.002)	0.018 (0.001)	0.019 (0.001)
	0.100	0.011 (0.000)	0.010 (0.000)	0.009 (0.000)	0.028 (0.004)	0.021 (0.001)	0.010 (0.000)	0.009 (0.000)
	0.250	0.012 (0.000)	0.012 (0.000)	0.009 (0.000)	0.020 (0.001)	0.017 (0.001)	0.011 (0.000)	0.008 (0.000)
	0.500	0.020 (0.000)	0.020 (0.000)	0.013 (0.000)	0.022 (0.000)	0.011 (0.000)	0.020 (0.000)	0.013 (0.000)
	0.750	0.034 (0.000)	0.034 (0.000)	0.019 (0.000)	0.036 (0.000)	0.016 (0.000)	0.035 (0.000)	0.020 (0.000)
	0.900	0.061 (0.001)	0.060 (0.001)	0.033 (0.000)	0.087 (0.001)	0.045 (0.001)	0.060 (0.001)	0.034 (0.000)
	0.950	0.114 (0.003)	0.110 (0.003)	0.068 (0.002)	0.185 (0.002)	0.121 (0.002)	0.092 (0.001)	0.056 (0.001)
	0.975	0.263 (0.016)	0.232 (0.013)	0.170 (0.012)	0.381 (0.003)	0.295 (0.003)	0.148 (0.002)	0.102 (0.002)
	$Exp(1)$	0.025	0.028 (0.000)	0.029 (0.000)	0.018 (0.000)	0.016 (0.000)	0.004 (0.000)	0.022 (0.000)
	0.050	0.024 (0.000)	0.024 (0.000)	0.016 (0.000)	0.017 (0.000)	0.005 (0.000)	0.021 (0.000)	0.014 (0.000)
	0.100	0.021 (0.000)	0.021 (0.000)	0.015 (0.000)	0.017 (0.000)	0.005 (0.000)	0.020 (0.000)	0.013 (0.000)
	0.250	0.024 (0.000)	0.024 (0.000)	0.016 (0.000)	0.022 (0.000)	0.007 (0.000)	0.024 (0.000)	0.015 (0.000)
	0.500	0.036 (0.000)	0.036 (0.000)	0.023 (0.000)	0.039 (0.001)	0.018 (0.000)	0.036 (0.000)	0.023 (0.000)
	0.750	0.066 (0.001)	0.066 (0.001)	0.040 (0.000)	0.065 (0.001)	0.028 (0.001)	0.067 (0.001)	0.041 (0.000)
	0.900	0.130 (0.001)	0.127 (0.001)	0.078 (0.001)	0.212 (0.002)	0.120 (0.002)	0.132 (0.001)	0.084 (0.001)
	0.950	0.206 (0.003)	0.201 (0.003)	0.123 (0.002)	0.493 (0.004)	0.353 (0.004)	0.211 (0.002)	0.143 (0.002)
	0.975	0.365 (0.009)	0.342 (0.008)	0.226 (0.007)	0.961 (0.007)	0.776 (0.006)	0.331 (0.004)	0.241 (0.004)

户等.

首先需要对数据进行预处理. 删除了数据集中带有缺失项的条目, 同时还删除了经销商花费这一因素, 因为该变量与汽车的销售价格高度相关, 其 Pearson 相关系数达到了 0.999. 同时, 还把汽车型号中皮卡类汽车剔除, 经过处理, 一共有 387 条样本可供分析. 设售价为响应变量, 记作 Y , 预测变量中包含跑车哑变量 (X_1)、SUV 哑变量 (X_2)、旅行车哑变量 (X_3)、小箱车哑变量 (X_4)、四轮驱动型哑变量 (X_5)、后轮驱动型哑变量 (X_6)、引擎大小 (X_7)、汽缸数 (X_8)、马力大小 (X_9)、城区每加仑英里数 (miles per gallon, MPG) (X_{10})、高速 MPG (X_{11})、汽车重量 (X_{12})、轴距 (X_{13})、长度 (X_{14}) 和宽度 (X_{15}).

在确定多指标模型中的指标个数时, 使用了常用的交叉检验法 (参见文献 [24]). 通过这一方法, 我们确定了单指标模型来分析这组数据. 通过计算各预测变量与汽车销售价格之间的 Pearson 相关系数, 我们确定了汽车的马力 X_9 与汽车的售价 Y 之间的相关系数最高 (0.8351), 从这点来看, 汽车马力是

表 3 模型 (5.3) MPE_k 的均值与标准差

噪声项	τ	NEW.WO(\hat{B})	NEW(\hat{B})	NEW(B^0)
$\mathcal{N}(0, 1)$	0.025	0.191 (0.030)	0.131 (0.004)	0.045 (0.001)
	0.050	0.175 (0.026)	0.123 (0.004)	0.039 (0.001)
	0.100	0.153 (0.014)	0.118 (0.004)	0.034 (0.001)
	0.250	0.128 (0.006)	0.117 (0.004)	0.030 (0.001)
	0.500	0.125 (0.004)	0.123 (0.004)	0.028 (0.001)
	0.750	0.154 (0.008)	0.141 (0.004)	0.029 (0.000)
	0.900	0.218 (0.007)	0.183 (0.005)	0.035 (0.001)
	0.950	0.248 (0.008)	0.232 (0.006)	0.043 (0.001)
	0.975	0.260 (0.009)	0.270 (0.007)	0.052 (0.001)
	$t(3)$	0.276 (0.036)	0.222 (0.020)	0.153 (0.018)
$t(3)$	0.025	0.188 (0.019)	0.149 (0.010)	0.079 (0.006)
	0.100	0.135 (0.010)	0.107 (0.008)	0.035 (0.001)
	0.250	0.105 (0.005)	0.101 (0.007)	0.028 (0.001)
	0.500	0.111 (0.008)	0.108 (0.008)	0.026 (0.001)
	0.750	0.157 (0.012)	0.140 (0.008)	0.028 (0.000)
	0.900	0.268 (0.024)	0.209 (0.009)	0.041 (0.001)
	0.950	0.365 (0.041)	0.334 (0.018)	0.099 (0.008)
	0.975	0.444 (0.050)	0.473 (0.041)	0.187 (0.023)
	$Exp(1)$	0.144 (0.012)	0.137 (0.007)	0.040 (0.001)
	0.050	0.135 (0.010)	0.121 (0.007)	0.035 (0.001)
$Exp(1)$	0.100	0.122 (0.008)	0.109 (0.007)	0.030 (0.001)
	0.250	0.112 (0.006)	0.110 (0.006)	0.027 (0.001)
	0.500	0.136 (0.008)	0.129 (0.006)	0.027 (0.000)
	0.750	0.197 (0.017)	0.156 (0.006)	0.033 (0.001)
	0.900	0.322 (0.050)	0.192 (0.007)	0.047 (0.001)
	0.950	0.315 (0.030)	0.238 (0.008)	0.075 (0.002)
	0.975	0.308 (0.017)	0.271 (0.009)	0.117 (0.004)

影响汽车售价诸多因素中十分重要的一项. 因此, 在归一化条件中, 固定汽车马力前的系数为 1, 并基于此条件估计指标系数. 我们使用文献 [36] 中的 dMAVE 方法计算降维矩阵估计量 \hat{B} , 并将结果汇总在表 4 中.

由表 4 中的结果可知, 在 1% 的显著水平上, 美国市场中一款汽车是否是 SUV (X_2), 是否是后轮驱动 (X_6), 同时引擎大小 (X_7)、汽缸数 (X_8) 和重量 (X_{12}) 都对这辆车的售价有着显著的影响. 此外, 在汽车的其他指标大致相同且我们所得到的单指标取值不是很小时 (如引擎较大、汽缸较多或者汽车较重) 时, SUV 车型或者后轮驱动并不能带来价格优势, 而更大的引擎、更多的汽缸和更大的汽车重量 (也往往匹配着更大马力的发动机) 则很可能带来更高的售价.

在分位数回归模型中, 人们经常通过计算某点在 $\tau = 0.025$ 和 $\tau = 0.975$ 上的条件分位数值来快速构造在该点处 95% 的预测区间. 我们通过计算所有点上的去一 (leave-one-out) 样本来得到预测区间, 而具体的比较指标包含平均覆盖概率、预测区间的平均长度和标准差. 所提方法的带约束和无约

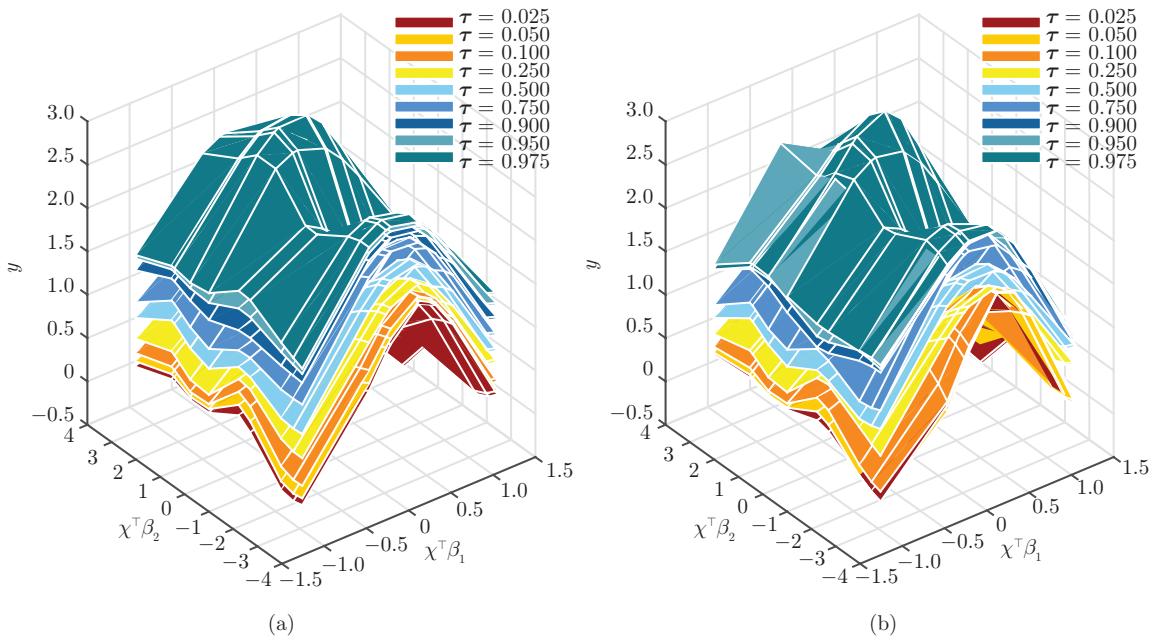


图 1 噪声项为 $\mathcal{N}(0, 1)$ 时的条件分位数回归曲面. (a) 添加无交叉约束的模型结果; (b) 未添加无交叉约束的模型结果

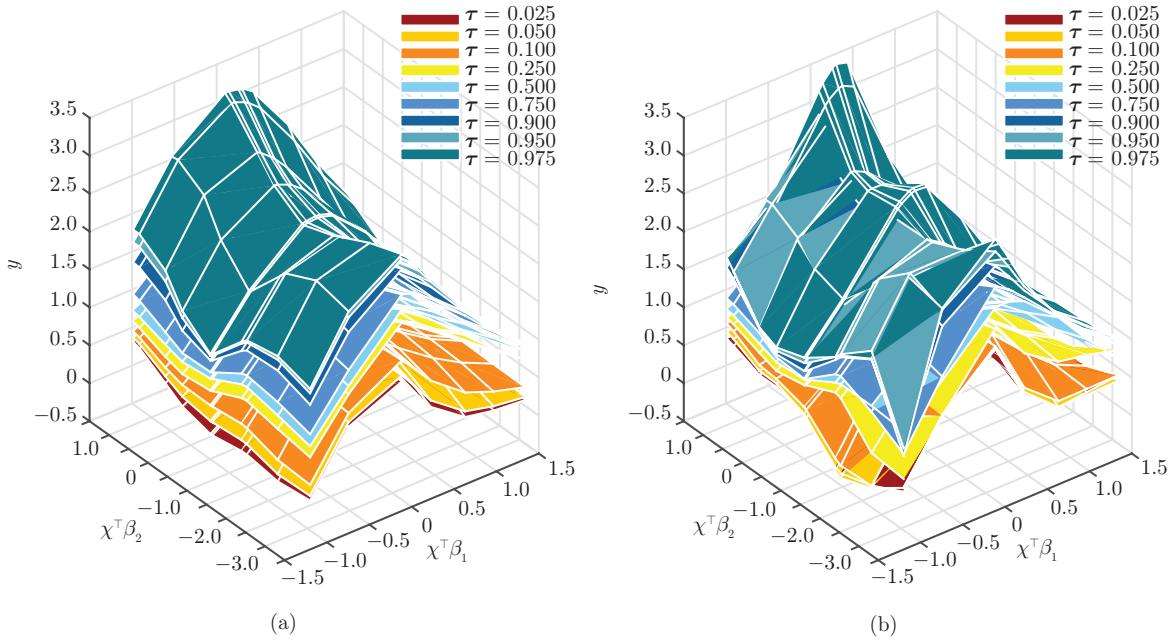


图 2 噪声项为 $t(3)$ 时的条件分位数回归曲面. (a) 添加无交叉约束的模型结果; (b) 未添加无交叉约束的模型结果

由表 5 可见, 本文方法得到的预测区间相较于其对应的无约束方法以及文献 [16, 17] 的两种方法有更高的平均预测覆盖概率。由图 4 可以看出, 无约束分位数回归曲线在尾部产生了分位数交叉现象, 即在无约束分位数曲线的图中 $\tau = 0.025$ 水平处的曲线与 $\tau = 0.950$ 和 $\tau = 0.975$ 水平处的分位数曲线产生了交叉。这意味着分位数曲线间的交叉影响到了模型整体的预测结果, 同时也会给汽车价格不

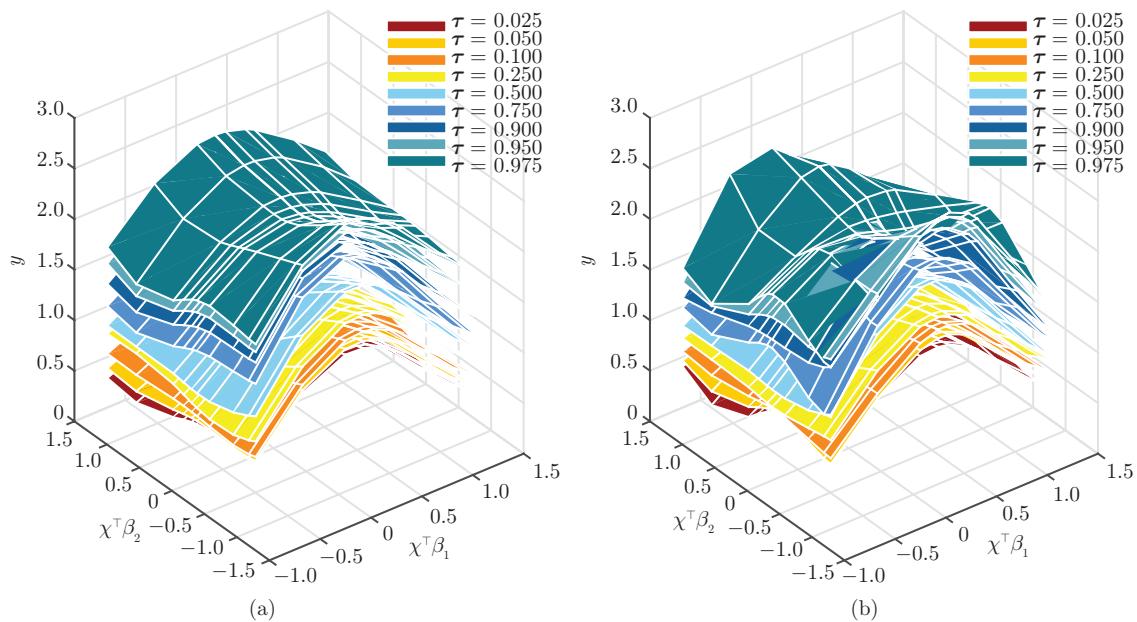


图 3 噪声项为 $\text{Exp}(1)$ 时的条件分位数回归曲面. (a) 添加无交叉约束的模型结果; (b) 未添加无交叉约束的模型结果

同分位数水平上的比较带来困惑. 这些交叉现象在图 4(a) 中清晰可见. 这说明在数据较少的尾部上的无约束估计效果不佳. 而更好地利用了全局信息的带约束方法则给出了较好的结果.

由图 4(b) 可知, 如果单指标取值不是特别小, 随着 $\mathbf{x}^\top \hat{\boldsymbol{\beta}}$ 的增大, 汽车的销售价格会进一步地提高. 而高档车型 (售价在整体汽车售价水平中前 95% 的车型) 的售价则随着这一指标值的增加而持续增加. 因此, 这一指标 $\mathbf{x}^\top \hat{\boldsymbol{\beta}}$ 的计算可以帮助销售部门大致了解汽车的硬件配置所带来的售价区间, 而最终售价的确定 (即在区间中的不同取值) 还将取决于品牌和售后等带来的正负面影响.

表 4 dMAVE 方法求得的汽车参数的指标估计量、标准误与 p - 值

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
估计值	0.4269	-0.7700	-0.1268	-0.5194	0.1564	0.5760	-0.6091
标准误	0.2965	0.2488	0.1363	0.4852	0.1275	0.1663	0.1426
p - 值	0.1507	0.0021	0.3526	0.2851	0.2208	0.0006	0.0000
	$\hat{\beta}_8$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$
估计值	0.3211	-0.1487	0.2864	1.3304	-0.0725	-0.1379	-0.2422
标准误	0.1190	0.2026	0.1774	0.2632	0.1464	0.1108	0.1065
p - 值	0.0073	0.4635	0.1072	0.0000	0.6208	0.2140	0.0235

表 5 单指标模型下 90% 和 95% 预测区间的覆盖概率与平均区间长度 (标准差)

		NEW.WO($\hat{\boldsymbol{B}}$)	NEW($\hat{\boldsymbol{B}}$)	DV($\hat{\boldsymbol{B}}$)	CFG($\hat{\boldsymbol{B}}$)
90%	平均覆盖概率	0.873	0.881	0.835	0.824
	平均预测区间长	1.189 (0.051)	1.182 (0.049)	0.953 (0.034)	0.978 (0.040)
95%	平均覆盖概率	0.917	0.935	0.842	0.879
	平均预测区间长	1.530 (0.066)	1.538 (0.066)	1.000 (0.038)	1.199 (0.053)

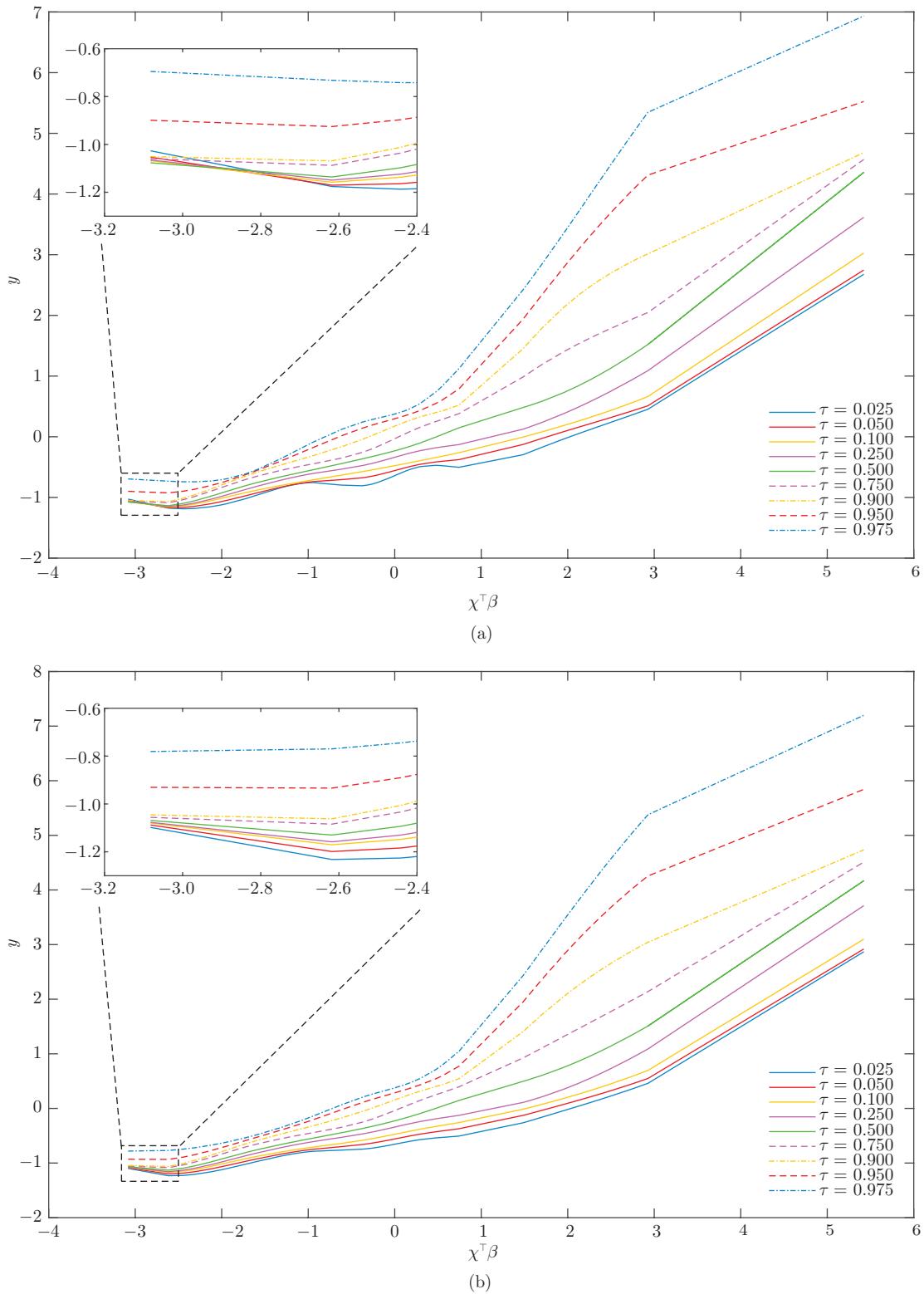


图 4 τ 不同分位数水平上的条件分位数回归曲线. (a) 无约束单指标分位数回归模型的拟合结果; (b) 无交叉单指标分位数回归模型的拟合结果

6 总结

本文研究了在半参数多指标分位数回归模型的框架下, 对无交叉分位数回归函数的估计和统计推断问题, 提出一种两步模型来估计模型中的未知参数, 首先借助已有的充分降维方法给出关于指标参数的相合估计, 在此基础上进一步利用张量积样条模型在满足单调约束的前提下逼近真实分位数回归曲线/曲面。我们还利用定理证明了, 依赖于真实指标参数所估计得到的分位数曲线/曲面估计量与依赖于指标参数估计量所求得的分位数回归曲线/曲面估计量具有相同的收敛速度。同时为了降低计算的复杂度, 我们进一步优化了线性规划问题对偶形式中的约束形式, 使得整个估计算法变得高效且可并行化。

在美国汽车销量数据的分析案例中, 我们展示了运用传统无约束的分位数回归模型去同时估计多个分位数水平上的曲线时, 会产生常见的分位数曲线相交叉的情形。这个现象在削弱了模型解释性的同时也降低了模型预测结果的准确性。通过在传统分位数回归模型中引入合理的关于分位数水平的单调性约束, 可以有效地解决这个问题。同时根据有关的数值模拟结果, 我们提议在分析实际数据以及从事相关的研究工作中, 当可供分析的样本量不是特别大或者我们更关心模型在极端分位数水平上的预测效果时, 应选择使用无交叉估计的分位数回归估计方法。

致谢 感谢两位匿名审稿人提出宝贵的修改建议。

参考文献

- 1 Koenker R, Bassett J G. Regression quantiles. *Econometrica*, 1978, 46: 33–50
- 2 Huang L P, Zhu W Y, Saunders C P, et al. A novel application of quantile regression for identification of biomarkers exemplified by equine cartilage microarray data. *BMC Bioinform*, 2008, 9: 300
- 3 Baur D G, Dimpfl T, Jung R C. Stock return autocorrelations revisited: A quantile regression approach. *J Empir Finance*, 2012, 19: 254–265
- 4 Fitzenberger B, Koenker R, Machado J A. *Economic Applications of Quantile Regression*. Berlin: Springer, 2013
- 5 Stone C J. Consistent nonparametric regression. *Ann Statist*, 1977, 5: 595–620
- 6 Fan J Q, Hu T C, Truong Y K. Robust non-parametric function estimation. *Scand J Statist*, 1994, 21: 433–446
- 7 Yu K M, Jones M C. Local linear quantile regression. *J Amer Statist Assoc*, 1998, 93: 228–237
- 8 Chaudhuri P. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann Statist*, 1991, 19: 760–777
- 9 Koenker R, Ng P, Portnoy S. Quantile smoothing splines. *Biometrika*, 1994, 81: 673–680
- 10 Wu T Z, Yu K M, Yu Y. Single-index quantile regression. *J Multivariate Anal*, 2010, 101: 1607–1621
- 11 Kong E F, Xia Y C. A single-index quantile regression model and its estimation. *Econom Theory*, 2012, 28: 730–768
- 12 Ma S J, He X M. Inference for single-index quantile regression models with profile optimization. *Ann Statist*, 2016, 44: 1234–1268
- 13 Zhu L P, Huang M, Li R Z. Semiparametric quantile regression with high-dimensional covariates. *Statist Sinica*, 2012, 22: 1379–1401
- 14 Alkenani A, Yu K M. Penalized single-index quantile regression. *Int J Stat Probab*, 2013, 2: 12–30
- 15 He X M. Quantile curves without crossing. *Amer Statist*, 1997, 51: 186–192
- 16 Dette H, Volgushev S. Non-crossing non-parametric estimates of quantile curves. *J R Stat Soc Ser B Stat Methodol*, 2008, 70: 609–627
- 17 Chernozhukov V, Fernandez-Val I, Galichon A. Quantile and probability curves without crossing. *Econometrica*, 2010, 78: 1093–1125
- 18 Bondell H D, Reich B J, Wang H. Noncrossing quantile regression curve estimation. *Biometrika*, 2010, 97: 825–838
- 19 Liu Y F, Wu Y C. Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *J Nonparametr Stat*, 2011, 23: 415–437
- 20 Yang Y, Tokdar S T. Joint estimation of quantile planes over arbitrary predictor spaces. *J Amer Statist Assoc*, 2017, 112: 1107–1120
- 21 Li K C. Sliced inverse regression for dimension reduction. *J Amer Statist Assoc*, 1991, 86: 316–327

- 22 Cook R D. Regression Graphics: Ideas for Studying Regressions Through Graphics. New Jersey: John Wiley & Sons, 2008
- 23 Ma Y Y, Zhu L P. Efficient estimation in sufficient dimension reduction. *Ann Statist*, 2013, 41: 250–268
- 24 Xia Y C, Tong H, Li W K, et al. An adaptive estimation of dimension reduction space. *J R Stat Soc Ser B Stat Methodol*, 2002, 64: 363–410
- 25 Ye Z S, Weiss R E. Using the bootstrap to select one of a new class of dimension reduction methods. *J Amer Statist Assoc*, 2003, 98: 968–979
- 26 Cook R D, Li B. Dimension reduction for conditional mean in regression. *Ann Statist*, 2002, 30: 455–474
- 27 Cook R D, Ni L. Sufficient dimension reduction via inverse regression. *J Amer Statist Assoc*, 2005, 100: 410–428
- 28 Zeng P. Determining the dimension of the central subspace and central mean subspace. *Biometrika*, 2008, 95: 469–479
- 29 Bura E, Yang J. Dimension estimation in sufficient dimension reduction: A unifying approach. *J Multivariate Anal*, 2011, 102: 130–142
- 30 Zhu L P, Zhu L X, Feng Z H. Dimension reduction in regressions through cumulative slicing estimation. *J Amer Statist Assoc*, 2010, 105: 1455–1466
- 31 Kelly C, Rice J. Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 1990, 46: 1071–1085
- 32 Leitenstorfer F, Tutz G. Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 2007, 8: 654–673
- 33 Wu Y, Zhang Y. Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *Ann Statist*, 2012, 40: 1609–1636
- 34 Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004
- 35 Koenker R. Quantile Regression. Cambridge: Cambridge University Press, 2005
- 36 Xia Y C. A constructive approach to the estimation of dimension reduction directions. *Ann Statist*, 2007, 35: 2654–2690
- 37 Iske A, Quak E, Floater M B. Tutorials on Multiresolution in Geometric Modelling. Summer School Lecture Notes. Berlin: Springer-Verlag, 2002
- 38 Van der Vaart A W. Asymptotic Statistics. Cambridge: Cambridge University Press, 2000
- 39 DeVore R A, Lorentz G G. Constructive Approximation. Berlin: Springer, 1993
- 40 Knight K. Limiting distributions for L_1 regression estimators under general conditions. *Ann Statist*, 1998, 26: 755–770
- 41 He X M, Shi P D. Monotone B-spline smoothing. *J Amer Statist Assoc*, 1998, 93: 643–650
- 42 He X M, Shao Q M. A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann Statist*, 1996, 24: 2608–2630
- 43 He X M, Shao Q M. On parameters of increasing dimensions. *J Multivariate Anal*, 2000, 73: 120–135
- 44 Arcones M A. The Bahadur-Kiefer representation of L_p regression estimators. *Econom Theory*, 1996, 12: 257–283
- 45 Pollard D. Empirical Processes: Theory and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward/Alexandria: Institute of Mathematical Statistics/American Statistical Association, 1990

附录 A 相关补充说明与理论证明

附录 A.1 张量积 B 样条

为了更好地阐述我们的无交叉非参数分位数回归模型, 这里简要介绍用张量积 B 样条逼近多元函数的方法 (参见文献 [37]). 假设对于正整数 $d_1 \in \mathbb{Z}^+$ 和节点系数 $\{r_1, \dots, r_{n_1+d_1}\}$, 关于 $x \in \mathbb{R}$ 的函数 $f(x)$ 可由以下 B 样条方法作非参数逼近:

$$f(x) = \sum_{k=1}^{n_1} a_k \phi_k(x),$$

其中 $\phi_k(x)$ 为 d 阶 B 样条基底函数. 记 $\hat{f} \in \mathcal{S}_1$, \mathcal{S}_1 为由 $\{\phi_1(x), \dots, \phi_{n_1}(x)\}$ 张成的空间. 考虑一个在 S_1 中节点系数为 $\tau \in \mathbb{R}$ 的函数的样条函数 \hat{g} ,

$$g(x, \tau) = \sum_{k=1}^{n_1} a_k(\tau) \phi_k(x),$$

对于其中的节点函数 $a_k(\tau)$, 类似地, 我们可以用以 $\{t_1, \dots, t_{n_2+d_2}\}$ 为节点的 d_2 阶 B 样条逼近函数

$$a_k(\tau) = \sum_{j=1}^{n_2} b_{k,j} \psi_j(\tau), \quad k = 1, \dots, n_1,$$

其中 $b_{k,j}$ 为样条节点. 记 $a_k \in \mathcal{S}_2$, \mathcal{S}_2 为由 $\{\psi_1(x), \dots, \psi_{n_2}(x)\}$ 张成的空间. 由上述表达式可知,

$$g(x, \tau) = \sum_{k=1}^{n_1} \sum_{j=1}^{n_2} b_{k,j} \psi_j(\tau) \phi_k(x) = \sum_{k=1}^{n_1} \sum_{j=1}^{n_2} b_{k,j} \pi_{j,k}(x, \tau).$$

$g \in \mathcal{S}_1 \otimes \mathcal{S}_2$ 被称为张量积 B 样条函数, 同时记 $\pi(x, \tau) = \psi(\tau) \otimes \phi(x)$. 更一般地, 对于一个 p 维函数 $f(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_k)^\top \in \mathbb{R}^k$, 人们可以用 k 组 B 样条基底组合成张量样条基底 $\Pi(\mathbf{x}) = \pi_1(x_1) \otimes \dots \otimes \pi_k(x_k)$ 对原函数 $f(\mathbf{x})$ 作非参数逼近.

附录 A.2 定理 2.1 的证明

假设 $Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau)$ 可由某些 $m(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 表示.

从 (1) 到 (2), 如果有

$$Q_{Y|\mathbf{X}}(\tau) = Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau), \quad \forall \tau \in (0, 1)$$

成立, 那么我们可以证明

$$\begin{aligned} F_{Y|\mathbf{X}}(y | \mathbf{x}) &= \int_0^1 I(Q_{Y|\mathbf{X}}(\tau) \leq y) d\tau \\ &= \int_0^1 I(Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau) \leq y) d\tau \\ &= F_{Y|\mathbf{X}^\top \mathbf{B}^0}(y | \mathbf{x}^\top \mathbf{B}^0). \end{aligned}$$

从 (2) 到 (1), 假设 $F_{Y|\mathbf{X}}(y | \mathbf{x}) = F_{Y|\mathbf{X}^\top \mathbf{B}^0}(y | \mathbf{x}^\top \mathbf{B}^0)$, 由条件分位数的定义可以得到

$$\Pr(Y \leq Q_{Y|\mathbf{X}}(\tau) | \mathbf{x}) = \Pr(Y \leq Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau) | \mathbf{x}^\top \mathbf{B}^0) = \tau.$$

因为 $\Pr(Y \leq y | \mathbf{x}) = \Pr(Y \leq y | \mathbf{x}^\top \mathbf{B}^0)$ 对于 $y \in \mathbb{R}$ 均成立, 所以有

$$\Pr(Y \leq Q_{Y|\mathbf{X}}(\tau) | \mathbf{x}) = \Pr(Y \leq Q_{Y|\mathbf{X}}(\tau) | \mathbf{x}^\top \mathbf{B}^0).$$

因此,

$$\Pr(Y \leq Q_{Y|\mathbf{X}}(\tau) | \mathbf{x}^\top \mathbf{B}^0) = \Pr(Y \leq Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau) | \mathbf{x}^\top \mathbf{B}^0).$$

进而由分位数 $Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau)$ 的定义可知,

$$Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau) \leq Q_{Y|\mathbf{X}}(\tau).$$

类似地,

$$Q_{Y|\mathbf{X}}(\tau) \leq Q_{Y|\mathbf{X}^\top \mathbf{B}^0}(\tau).$$

至此定理 2.1 证毕.

附录 A.3 命题 2.1 的证明

记 $\{r_{1,q+1}, r_{2,q+1}, \dots, r_{J_{n,q+1},q+1}\}$ 是函数 $g_n(\mathbf{U}, \tau)$ 在 τ 方向上用 B 样条拟合的节点序列, d_{q+1} 是所用样条基底函数的阶数. 由 B 样条基底函数的性质可知, g_n 关于 τ 的偏导数为

$$\frac{\partial g_n(U_1, \dots, U_q, \tau)}{\partial \tau} = \sum_{j_1=1}^{J_{n,1}} \cdots \sum_{j_q=1}^{J_{n,q}} \sum_{j_{q+1}=1}^{J_{n,q+1}-1} \frac{(d_{q+1}-1)(\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} - \lambda_{j_1, \dots, j_q, j_{q+1}})}{r_{(j_{q+1}+d_{q+1}), q+1} - r_{(j_{q+1}+1), q+1}} \\ \times \pi_{j_1}(U_1) \cdots \pi_{j_q}(U_q) \pi_{j_{q+1}}^{(d_{q+1}-1)}(\tau),$$

其中 $\{\pi_{j_{q+1}}^{(d_{q+1}-1)}(\tau) : 1 \leq j_{q+1} \leq J_{n,q+1}-1\}$ 是 $d_{q+1}-1$ 阶的 B 样条基底函数. 因为 $\pi_{j_1}(U_1), \dots, \pi_{j_q}(U_q)$ 和 $\pi_{j_{q+1}}^{(d_{q+1}-1)}(\tau)$ 均是非负函数, 条件 $\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} - \lambda_{j_1, \dots, j_q, j_{q+1}} \geq 0$ 因此是 $\partial g_n(U_1, \dots, U_q, \tau)/\partial \tau \geq 0$ 的充分条件.

另一方面, 对于一些 $\tau_0 \in [r_{s_{q+1},q+1}, r_{s_{q+1}+1,q+1}]$, 由 B 样条的性质知, 当 $s_{q+1} - (d_{q+1}-1) + 1 \leq j_{q+1} \leq s_{q+1}$ 时, $\pi_{j_{q+1}}(\tau_0) = 0$ 成立. 当 $\pi_{j_{q+1}}(\tau_0)$ 是二阶及以下的基底函数时,

$$\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} - \lambda_{j_1, \dots, j_q, j_{q+1}} \geq 0$$

是 $\partial g_n(U_1, \dots, U_q, \tau)/\partial \tau \geq 0$ 的必要条件. 同时存在 $c_1, c_2 > 0$ 使得不等式

$$\lambda_{j_1, \dots, j_q, (j_{q+1}+1)} - \lambda_{j_1, \dots, j_q, j_{q+1}} \geq c_1 > 0$$

当且仅当 $\partial g_n(U_1, \dots, U_q, \tau)/\partial \tau \geq c_2 > 0$ 时成立.

附录 A.4 定理 4.1 的证明

为了充分研究本文所提出非参数估计量的相合性和有效性, 首先需要研究样条估计量 $\hat{g}_n(\mathbf{x}^\top \mathbf{B}, \tau) = \hat{\boldsymbol{\lambda}}(\mathbf{B})^\top \boldsymbol{\Pi}(\mathbf{x}^\top \mathbf{B}, \tau)$ 的全局收敛性, 其中 $\hat{\boldsymbol{\lambda}}(\mathbf{B})$ 是在给定矩阵 \mathbf{B} 的前提下, 由优化问题 (3.1) 得到的样条节点估计量, 这里的 \mathbf{B} 为任意满足 $\|\mathbf{B} - \mathbf{B}^0\|_\infty \leq Cn^{-1/2}$ 的降维矩阵. 由于降维矩阵估计量 $\hat{\mathbf{B}}$ 很明显落在这个区域中, 于是, 我们重点研究基于 $\hat{\mathbf{B}}$ 的样条函数估计量 $\hat{g}_n(\mathbf{x}^\top \hat{\mathbf{B}}, \tau) = \hat{\boldsymbol{\lambda}}(\hat{\mathbf{B}})^\top \boldsymbol{\Pi}(\mathbf{x}^\top \hat{\mathbf{B}}, \tau)$. 为了简化符号, 令 $\mathbf{U} = \mathbf{X}^\top \mathbf{B}^0$, $\tilde{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}}(\mathbf{B}^0)$ 是基于降维矩阵 \mathbf{B}^0 的节点系数估计量, 因此记 $\tilde{g}_n(\mathbf{U}, \tau) = \tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Pi}(\mathbf{U}, \tau) = \tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Pi}(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 是基于 \mathbf{B}^0 的样条函数估计量. 首先, 证明下述两个引理成立.

引理 A.1 假设条件 (A1) 和 (A2) 成立, 存在一个张量积底样条函数

$$g_n^0(\mathbf{U}, \tau) = \boldsymbol{\lambda}^{0\top} \boldsymbol{\Pi}(\mathbf{U}, \tau) = \sum_{j_1=1}^{J_{n,1}} \cdots \sum_{j_q=1}^{J_{n,q}} \sum_{j_{q+1}=1}^{J_{n,q+1}} \lambda_{j_1, \dots, j_q, j_{q+1}}^0 \pi_{j_1}(U_1) \cdots \pi_{j_q}(U_q) \pi_{j_{q+1}}(\tau)$$

与非降 $\lambda_{j_1, \dots, j_q, 1}^0 \leq \cdots \leq \lambda_{j_1, \dots, j_q, J_{n,q+1}}^0$ 或非升 $\lambda_{j_1, \dots, j_q, 1}^0 \leq \cdots \leq \lambda_{j_1, \dots, j_q, J_{n,q+1}}^0$ 节点系数序列, 使得 $\sup_{\mathbf{B}} |m(\mathbf{U}, \tau) - g_n^0(\mathbf{U}, \tau)| = O_p(J_n^{-a})$ 存在, 其中 $\pi_{j_1}(U_1), \dots, \pi_{j_q}(U_q), \pi_{j_{q+1}}(\tau)$ 分别是 d_1, \dots, d_q, d_{q+1} 阶的样条基底函数.

证明 不失一般性, 只证明两指标模型带单调非降约束的情形, 多指标与单调非增的情形可以类似证明. 令

$$h = \max \left\{ \max_{d_1 \leq j_1 \leq J_{n,1}} (r_{j_1+1,1} - r_{j_1,1}), \max_{d_2 \leq j_2 \leq J_{n,2}} (r_{j_2+1,2} - r_{j_2,2}), \max_{d_3 \leq j_3 \leq J_{n,3}} (r_{j_3+1,3} - r_{j_3,3}) \right\}.$$

定义 $\mathcal{J}_i = \{j_i \mid s_i + 1 - d_i \leq j_i \leq s_i, i = 1, 2, 3\}$. 由样条函数的性质可知, 存在 $\xi_1 < \dots < \xi_{J_{n,1}}$, $\zeta_1 < \dots < \zeta_{J_{n,2}}$, $\eta_1 < \dots < \eta_{J_{n,3}}$, 使得

$$\begin{aligned} g_n^0(\mathbf{U}, \tau) &= \sum_{j_1=1}^{J_{n,2}} \sum_{j_2=1}^{J_{n,2}} \sum_{j_3=1}^{J_{n,3}} m(\xi_{j_1}, \zeta_{j_2}, \eta_{j_3}) \pi_{j_1}(U_1) \pi_{j_2}(U_2) \pi_{j_3}(\tau) \\ &= \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \in \mathcal{J}_2} \sum_{j_3 \in \mathcal{J}_3} m(\xi_{j_1}, \zeta_{j_2}, \eta_{j_3}) \pi_{j_1}(U_1) \pi_{j_2}(U_2) \pi_{j_3}(\tau) \end{aligned}$$

关于 τ 是单调非降的, 其中 $s_1 \in \{1, \dots, J_{n,1} + d_1\}$, $s_2 \in \{1, \dots, J_{n,2} + d_2\}$, $s_3 \in \{1, \dots, J_{n,3} + d_3\}$. 对于任意存在于 $[r_{1,s_1}, r_{1,s_1+1}) \times [r_{2,s_2}, r_{2,s_2+1}) \times [r_{3,s_3}, r_{3,s_3+1})$ 的 $(\tilde{\mathbf{U}}, \tilde{\tau})$, 由样条基底函数的性质可知,

$$m(\tilde{\mathbf{U}}, \tilde{\tau}) = \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \in \mathcal{J}_2} \sum_{j_3 \in \mathcal{J}_3} m(\tilde{U}_1, \tilde{U}_2, \tilde{\tau}) \pi_{j_1}(\tilde{U}_1) \pi_{j_2}(\tilde{U}_2) \pi_{j_3}(\tilde{\tau}),$$

其中 $\tilde{U}_1 = \mathbf{x}^\top \tilde{\beta}_1$, $\tilde{U}_2 = \mathbf{x}^\top \tilde{\beta}_2$. 基于上述结论可知,

$$\begin{aligned} |m(\tilde{\mathbf{U}}, \tilde{\tau}) - g_n^0(\tilde{\mathbf{U}}, \tilde{\tau})| &\leq \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \in \mathcal{J}_2} \sum_{j_3 \in \mathcal{J}_3} |m(\tilde{U}_1, \tilde{U}_2, \tilde{\tau}) - m(\xi_{j_1}, \zeta_{j_2}, \eta_{j_3})| \pi_{j_1}(\tilde{U}_1) \pi_{j_2}(\tilde{U}_2) \pi_{j_3}(\tilde{\tau}) \\ &\leq \max_{\mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3} |m(\tilde{U}_1, \tilde{U}_2, \tilde{\tau}) - m(\xi_{j_1}, \zeta_{j_2}, \eta_{j_3})|. \end{aligned}$$

定义函数 $\Omega(f; h) = \max\{|f(\mathbf{U}, \tau) - f(\mathbf{U}', \tau')| : |U_1 - U'_1| \leq h, |U_2 - U'_2| \leq h, |\tau - \tau'| \leq h\}$, $\mathbf{U} = (U_1, U_2)^\top$, $\mathbf{U}' = (U'_1, U'_2)^\top$, 我们可以证明 $\Omega(f; h)$ 是关于 h 的单调且具有次可加性的函数, 即有 $\Omega(g; h_1) \leq \Omega(f; h_1 + h_2) \leq \Omega(f; h_1) + \Omega(f; h_2)$ 对于任意 $h_1 > 0$ 和 $h_2 > 0$ 都成立. 关于这方面的详细证明可参见文献 [33, 引理 2.19].

选取

$$\begin{aligned} \xi_{j_1} &= \begin{cases} r_{1,1} + \frac{(j_1 - 1)(r_{d_1+1,1} - r_{d_1,1})}{d_1}, & j_1 = 1, \dots, d_1, \\ r_{j_1,1}, & j_1 = d_1 + 1, \dots, J_{n,1}, \end{cases} \\ \zeta_{j_2} &= \begin{cases} r_{1,2} + \frac{(j_2 - 1)(r_{d_2+1,2} - r_{d_2,2})}{d_2}, & j_2 = 1, \dots, d_2, \\ r_{j_2,2}, & j_2 = d_2 + 1, \dots, J_{n,2}, \end{cases} \\ \eta_{j_3} &= \begin{cases} r_{1,3} + \frac{(j_3 - 1)(r_{d_3+1,3} - r_{d_3,3})}{d_3}, & j_3 = 1, \dots, d_3, \\ r_{j_3,3}, & j_3 = d_3 + 1, \dots, J_{n,3}, \end{cases} \end{aligned}$$

则有下列不等式始终成立:

$$\begin{cases} \xi_{j_1+1} - \xi_{j_1} > 0, & |\xi_{j_1} - r_{j_1,1}| \leq h, \quad j_1 = 1, \dots, J_{n,1}, \\ \zeta_{j_2+1} - \zeta_{j_2} > 0, & |\zeta_{j_2} - r_{j_2,2}| \leq h, \quad j_2 = 1, \dots, J_{n,2}, \\ \eta_{j_3+1} - \eta_{j_3} > 0, & |\eta_{j_3} - r_{j_3,3}| \leq h, \quad j_3 = 1, \dots, J_{n,3}. \end{cases}$$

因此对于任意 $\tilde{U}_i \in [r_{s_i,i}, r_{s_i+1,i})$, $j_i \in \mathcal{J}_i$, $|\tilde{U}_i - r_{j_i,i}| \leq d_i h$ ($i = 1, 2, 3$) 成立.

记 $d = \min\{d_1, d_2, d_3\}$, 则根据上述结论有 $|\tilde{U}_1 - \xi_{j_1}| \leq (d+1)h$, $|\tilde{U}_2 - \zeta_{j_2}| \leq (d+1)h$ 和 $|\tilde{\tau} - \eta_{j_3}| \leq (d+1)h$ 成立. 因此,

$$\sup_{\mathcal{B}} |m(\mathbf{U}, \tau) - g_n^0(\mathbf{U}, \tau)| \leq \Omega(m; (d+1)h) \leq (d+1)\Omega(m; h), \quad (\text{A.1})$$

其中最后一个不等号来源于 $\Omega(g; h)$ 关于 h 的次可加性.

记 $G_{d_1, d_2, d_3} = \{\boldsymbol{\lambda}^0^\top \mathbf{\Pi}(\mathbf{U}, \tau), \lambda_{j_1, j_2, 1}^0 \leq \dots \leq \lambda_{j_1, j_2, J_{n, 3}}^0, 1 \leq j_1 \leq J_{n, 1}, 1 \leq j_2 \leq J_{n, 2}\}$ 为阶数 d_1, d_2 和 d_3 的张量积底样条函数集. 根据 (A.1), 可得 $\|m - g_n^0\|_\infty = \|m - G_{d_1, d_2, d_3}\|_\infty \leq (d+1)\Omega(m; h)$. 记 m 与 G_{d_1, d_2, d_3} 间的距离为 $l(m, G_{d_1, d_2, d_3})$. 对于任意 $g \in G_{d_1, d_2, d_3}$, 因为 g 具有有界的偏导数, 可以证明

$$l(m, G_{d_1, d_2, d_3}) = l(m - g, G_{d_1, d_2, d_3}) \leq (d+1)\Omega(m - g, h),$$

可以注意到的是, 如果记 $\mathcal{H} = \{(U_1, U'_1, U_2, U'_2, \tau, \tau') \mid |U_1 - U'_1| \leq h, |U_2 - U'_2| \leq h, |\tau - \tau'| \leq h\}$, 那么,

$$\begin{aligned} \Omega(m - g, h) &= \max_{\mathcal{H}} \|(m - g)(U_1, U_2, \tau) - (m - g)(U'_1, U'_2, \tau')\| \\ &\leq \left\| \frac{\partial(m - g)}{\partial U_1} \right\|_\infty h + \left\| \frac{\partial(m - g)}{\partial U_2} \right\|_\infty h + \left\| \frac{\partial(m - g)}{\partial \tau} \right\|_\infty h. \end{aligned}$$

由此可以得到

$$l(m, G_{d_1, d_2, d_3}) \leq (d+1)h \left\{ l\left(\frac{\partial m}{\partial U_1}, G_{d_1-1, d_2, d_3}\right) + l\left(\frac{\partial m}{\partial U_2}, G_{d_1, d_2-1, d_3}\right) + l\left(\frac{\partial m}{\partial \tau}, G_{d_1, d_2, d_3-1}\right) \right\},$$

其中

$$\begin{aligned} G_{d_1-1, d_2, d_3} &= \left\{ \frac{\partial g}{\partial U_1}, g \in G_{d_1, d_2, d_3} \right\}, \\ G_{d_1, d_2-1, d_3} &= \left\{ \frac{\partial g}{\partial U_2}, g \in G_{d_1, d_2, d_3} \right\}, \\ G_{d_1, d_2, d_3-1} &= \left\{ \frac{\partial g}{\partial \tau}, g \in G_{d_1, d_2, d_3} \right\}. \end{aligned}$$

类似地, 可以得到

$$\begin{aligned} l(m, G_{d_1, d_2, d_3}) &\leq c|h|^a \left\{ \sum_{\Theta_w} \left\| \frac{\partial^a m}{\partial U_1^{w_1} \partial U_2^{w_2} \partial \tau^{w_0}} \right\|_\infty \right\} \\ &\leq c|h|^a \max_{\Theta_w} \left\| \frac{\partial^a m}{\partial U_1^{w_1} \partial U_2^{w_2} \partial \tau^{w_{q+1}}} \right\|_\infty, \end{aligned}$$

其中 $\Theta_w = \{0 \leq w_i \leq a, \sum_i w_i = a, i = 1, \dots, q+1\}$. 因此, 由上述结论和条件 (A2) 结论可证引理 A.1 成立. \square

记 $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ 是随机向量 \mathbf{U} 的一组随机样本, 根据文献 [38, 定理 5.7], 可得 $\|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\| = o_p(1)$. 记 $b(\mathbf{u}_i, \tau) = \boldsymbol{\lambda}^0^\top \mathbf{\Pi}(\mathbf{u}_i, \tau) - m(\mathbf{u}_i, \tau)$,

$$\begin{aligned} D_n(\tilde{\boldsymbol{\lambda}}) &= \sum_{k=1}^{\kappa_n} \sum_{i=1}^n \rho_{\tau_k} \{y_i - \tilde{\boldsymbol{\lambda}}^\top \mathbf{\Pi}(\mathbf{u}_i, \tau_k)\} - \sum_{k=1}^{\kappa_n} \sum_{i=1}^n \rho_{\tau_k} \{y_i - m(\mathbf{u}_i, \tau_k)\} \\ &= \sum_{k=1}^{\kappa_n} \sum_{i=1}^n \{\rho_{\tau_k} (y_i - m(\mathbf{u}_i, \tau_k)) - b(\mathbf{u}_i, \tau_k) - (\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau_k)\} - \rho_{\tau_k} (y_i - m(\mathbf{u}_i, \tau_k)), \end{aligned}$$

由 $\tilde{\boldsymbol{\lambda}}$ 的定义有

$$D_n(\tilde{\boldsymbol{\lambda}}) - D_n(\boldsymbol{\lambda}^0) = \sum_{k=1}^{\kappa_n} \sum_{i=1}^n \rho_{\tau_k} (y_i - m(\mathbf{u}_i, \tau_k)) - b(\mathbf{u}_i, \tau_k) - (\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau_k))$$

$$-\sum_{k=1}^{\kappa_n} \sum_{i=1}^n \rho_{\tau_k}(y_i - m(\mathbf{u}_i, \tau_k) - b(\mathbf{u}_i, \tau_k)) \leq 0. \quad (\text{A.2})$$

定义

$$\begin{aligned} \bar{D}_n(\tilde{\boldsymbol{\lambda}}) = \kappa_n \sum_{i=1}^n \mathbb{E} \int & \{ \rho_\tau(y_i - m(\mathbf{u}_i, \tau) - b(\mathbf{u}_i, \tau) - (\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau)) \\ & - \rho_\tau(y_i - m(\mathbf{u}_i, \tau)) \} d\tau. \end{aligned}$$

我们首先可以证明以下引理成立.

引理 A.2 令 $c_n = (n^{1/2} \kappa_n J_n^{-(q+1)/2} + n J_n^{(1-q)/2}) \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|$, 如果条件 (A1) 和 (A2) 存在, 对于满足 $\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\| = o(1)$ 的 $\boldsymbol{\lambda}_n$, 有

$$\begin{aligned} c_n^{-1} & \left| D_n(\boldsymbol{\lambda}_n) - D_n(\boldsymbol{\lambda}^0) - (\bar{D}_n(\boldsymbol{\lambda}_n) - \bar{D}_n(\boldsymbol{\lambda}^0)) \right. \\ & \left. + \sum_{k=1}^{\kappa_n} \sum_{i=1}^n (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau_k) \{ \tau_k - I(y_i - m(\mathbf{u}_i, \tau_k) \leq 0) \} \right| \\ & = O_p(1). \end{aligned}$$

证明 记 $\boldsymbol{\theta}_n = (\theta_{j_1, j_2, j_3})_{1 \leq j_1 \leq J_{n,1}, 1 \leq j_2 \leq J_{n,2}, 1 \leq j_3 \leq J_{n,3}} = \boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0$, 在这个证明中考虑 $\tau = 0.5$ 的情形, 其他分位数水平上的证明与之类似. 设

$$\begin{aligned} R_i(\mathbf{u}_i, \tau_k) & = |y_i - m(\mathbf{u}_i, \tau_k) - b(\mathbf{u}_i, \tau_k) - \boldsymbol{\theta}_n^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau_k)| - |y_i - m(\mathbf{u}_i, \tau_k) - b(\mathbf{u}_i, \tau_k)| \\ & \quad + \boldsymbol{\theta}_n^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau_k) \{ \tau_k - I(y_i - m(\mathbf{u}_i, \tau_k) \leq 0) \}, \end{aligned} \quad (\text{A.3})$$

定义 $L_n = \|\boldsymbol{\theta}_n\|$, 对于 $\tau \in [\tau_k - 1, \tau_k]$ 和 $\tau_k \in [r_{q+1, s_3-1}, r_{q+1, s_3}]$, 由条件 (A1) 与 B 样条基底函数的性质, 有以下结论以概率趋近于 1 成立:

$$\begin{aligned} \left| \sum_{i=1}^n \{R_i(\mathbf{u}_i, \tau_k) - R_i(\mathbf{u}_i, \tau)\} \right| & \leq \sum_{i=1}^n 5 |\boldsymbol{\theta}_n^\top \{\boldsymbol{\Pi}(\mathbf{u}_i, \tau_k) - \boldsymbol{\Pi}(\mathbf{u}_i, \tau)\}| \\ & \leq 5 \sum_{i=1}^n \sum_{j_1=1}^{J_{n,1}} \sum_{j_2=1}^{J_{n,2}} \sum_{j_3=1}^{J_{n,3}} |\theta_{j_1, j_2, j_3}| \pi_{j_1}(u_{1,i}) \pi_{j_2}(u_{2,i}) \sup_{j_3} |\pi_{j_3}(\tau_k) - \pi_{j_3}(\tau)| \\ & \leq 5n \sqrt{L_n^2 J_n^3 \left(\sup_{j_1} \mathbb{E}\{\pi_{j_1}(u_1)\} \sup_{j_2} \mathbb{E}\{\pi_{j_2}(u_2)\} \right)^2} \sup_{j_3} |\pi_{j_3}(\tau_k) - \pi_{j_3}(\tau)| \\ & \leq Cn J_n^{-1/2} L_n \kappa_n^{-1}, \end{aligned} \quad (\text{A.4})$$

其中 $0 < C < \infty$ (对于 q 指标模型, (A.4) 最后一个表达式中的速率为 $Cn J_n^{(1-q)/2} L_n \kappa_n^{-1}$). 同理, 有下述结论以概率趋近于 1 成立:

$$\begin{aligned} & \left| \sum_{i=1}^n \sum_{k=1}^{\kappa_n} R_i(\mathbf{u}_i, \tau_k) - \kappa_n \sum_{i=1}^n \int R_i(\mathbf{u}_i, \tau) d\tau \right| \\ & \leq \left| \sum_{i=1}^n \sum_{k=1}^{\kappa_n} R_i(\mathbf{u}_i, \tau_k) - \frac{\kappa_n}{\kappa_n + 1} \sum_{i=1}^n \sum_{k=1}^{\kappa_n} R_i(\mathbf{u}_i, \tau_k) \right| + \frac{\kappa_n}{\kappa_n + 1} \sum_{i=1}^n \sup_{1 \leq k \leq \kappa_n + 1} |R_i(\mathbf{u}_i, \tau_k)| + O(n J_n^{-1/2} L_n) \end{aligned}$$

$$\leq 2 \sum_{i=1}^n \sup_{1 \leq k \leq \kappa_n+1} |R_i(\mathbf{u}_i, \tau_k)| + O(n J_n^{-1/2} L_n),$$

且

$$\sum_{i=1}^n \sup_{1 \leq k \leq \kappa_n+1} |R_i(\mathbf{u}_i, \tau_k)| \leq \sum_{i=1}^n \sup_{1 \leq k \leq \kappa_n+1} 3\mathbb{E}|\boldsymbol{\theta}_n^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau_k)| = O(n L_n J_n^{-1/2}),$$

因此,

$$\left| \sum_{i=1}^n \sum_{k=1}^{\kappa_n} R_i(\mathbf{u}_i, \tau_k) - \kappa_n \sum_{i=1}^n \int R_i(\mathbf{u}_i, \tau) d\tau \right| = O_p(n J_n^{(1-q)/2} L_n) \quad (\text{A.5})$$

成立. 进一步地, 对于某个足够大的 n , 存在常数 $0 < C' < \infty$ 和 $0 < C'' < \infty$, 使得

$$\begin{aligned} \text{Var} \left\{ \int R_i(\mathbf{u}_i, \tau) d\tau \right\} &\leq \mathbb{E} \left\{ \int 3|\boldsymbol{\theta}_n^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau)| d\tau \right\}^2 \\ &\leq 9C' J_n^{-2} \mathbb{E} \left\{ \sum_{j_1} \sum_{j_2} \sum_{j_3} |\theta_{j_1, j_2, j_3}| \pi_{j_1}(u_{1i}) \pi_{j_2}(u_{2i}) \right\}^2 \\ &\leq C'' J_n^{-3} L_n^2, \end{aligned}$$

其中第二个不等式来源于 [39, 第五章, 定理 4.2]. 在 q 指标模型中, 最后速率的表达式为 $C'' J_n^{-(q+1)} L_n^2$. 由 Chebyshev 不等式可得

$$\left| \kappa_n \sum_{i=1}^n \int R_i(\mathbf{u}_i, \tau) d\tau - \kappa_n \sum_{i=1}^n \mathbb{E} \int R_i(\mathbf{u}_i, \tau) d\tau \right| = O_p(n^{1/2} \kappa_n J_n^{-(q+1)/2} L_n). \quad (\text{A.6})$$

因此, 引理 A.2 的结论可由 (A.5) 和 (A.6) 得到. \square

定理 4.1 的证明 令 Y 给定预测变量 \mathbf{X} 的条件分布函数为 $F(t | \mathbf{X}) = \Pr(Y \leq t | \mathbf{X})$, 其概率密度函数为 $f(t | \mathbf{X})$. 对满足 $\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\| = o(1)$ 的序列 $\boldsymbol{\lambda}_n$, 根据 Knight 恒等式^[40]

$$\rho_\tau(x-y) - \rho_\tau(x) = y\{I(x \leq 0) - \tau\} + \int_0^y \{I(x \leq t) - I(x \leq 0)\} dt,$$

可以得到

$$\begin{aligned} \bar{D}_n(\boldsymbol{\lambda}_n) - \bar{D}_n(\boldsymbol{\lambda}^0) &= \kappa_n \sum_{i=1}^n \mathbb{E} \left[\int_0^1 \int_{b(\mathbf{u}_i, \tau)}^{b(\mathbf{u}_i, \tau) + (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau)} \{F(t + m(\mathbf{u}_i, \tau) | \mathbf{x}_i) \right. \\ &\quad \left. - F(m(\mathbf{u}_i, \tau) | \mathbf{x}_i)\} dt d\tau \right] \\ &\sim \kappa_n \sum_{i=1}^n \mathbb{E} \left\{ \int_0^1 f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) b(\mathbf{u}_i, \tau) \boldsymbol{\Pi}(\mathbf{u}_i, \tau)^\top (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0) d\tau \right\} \\ &\quad + \frac{\kappa_n}{2} \sum_{i=1}^n (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbb{E} \left\{ \int_0^1 f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) \boldsymbol{\Pi}(\mathbf{u}_i, \tau) \boldsymbol{\Pi}(\mathbf{u}_i, \tau)^\top d\tau \right\} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0). \quad (\text{A.7}) \end{aligned}$$

利用与求 (A.5) 类似的方法可以求得

$$\left| \sum_{i=1}^n \sum_{k=1}^{\kappa_n} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{u}_i, \tau_k) \{ \tau_k - I(y_i - m(\mathbf{u}_i, \tau_k) \leq 0) \} \right|$$

$$\begin{aligned} & -\kappa_n \sum_{i=1}^n \int (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau) \{ \tau - I(y_i - m(\mathbf{u}_i, \tau) \leq 0) \} d\tau \Big| \\ & = O_p(n J_n^{-1/2} L_n). \end{aligned} \quad (\text{A.8})$$

由引理 A.2、(A.2) 和 (A.8) 的结论可知, 以下结论以趋近于 1 的概率成立:

$$\begin{aligned} D_n(\boldsymbol{\lambda}_n) - D_n(\boldsymbol{\lambda}^0) & = -\kappa_n \sum_{i=1}^n \int (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau) \{ \tau - I(y_i - m(\mathbf{u}_i, \tau) \leq 0) \} d\tau \\ & \quad + \overline{D}_n(\boldsymbol{\lambda}_n) - \overline{D}_n(\boldsymbol{\lambda}^0) + O(c_n) \\ & \leq 0. \end{aligned}$$

结合 (A.7) 可知, 存在常数 $0 < C < \infty$ 使得下式成立:

$$\begin{aligned} & \left| \kappa_n \sum_{i=1}^n \int (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau) \{ \tau - I(y_i - m(\mathbf{u}_i, \tau) \leq 0) \} d\tau \right| \\ & + \left| \kappa_n \sum_{i=1}^n \mathbb{E} \left\{ \int f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) b(\mathbf{u}_i, \tau) \mathbf{\Pi}(\mathbf{u}_i, \tau)^\top (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0) d\tau \right\} \right| + C c_n \\ & \geq \kappa_n \sum_{i=1}^n 2^{-1} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbb{E} \left\{ \int_0^1 f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) \mathbf{\Pi}(\mathbf{u}_i, \tau) \mathbf{\Pi}(\mathbf{u}_i, \tau)^\top d\tau \right\} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0). \end{aligned} \quad (\text{A.9})$$

令 $\mu_{j_1, j_2, j_3}(\mathbf{u}_i, \tau) = \int \pi_{j_1}(u_1) \pi_{j_2}(u_2) \pi_{j_3}(\tau) \{ \tau - I(y_i - m(\mathbf{u}_i, \tau) \leq 0) \} d\tau$. 因为 $\mathbb{E}\{\mu_{j_1, j_2, j_3}(\mathbf{u}_i, \tau)\} = 0$ 和 $\text{Var}\{\mu_{j_1, j_2, j_3}(\mathbf{u}_i, \tau)\} \asymp J_n^{-5}$, 可以由弱大数定律得知, 对于任意的 j_1 和 j_2 , 有

$$\left| \sum_{i=1}^n \mu_{j_1, j_2, j_3}(\mathbf{u}_i, \tau_k) \right| = O_p(n^{1/2} J_n^{-5/2}).$$

因此, 对于满足 $\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\| = o(1)$ 的 $\boldsymbol{\lambda}_n$, 以下结论以趋近于 1 的概率成立:

$$\begin{aligned} & \left| \kappa_n \sum_{i=1}^n \int (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbf{\Pi}(\mathbf{u}_i, \tau) \{ \tau - I(y_i - m(\mathbf{u}_i, \tau) \leq 0) \} d\tau \right| \\ & = \left| \kappa_n \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \in \mathcal{J}_2} \sum_{j_3 \in \mathcal{J}_3} \theta_{j_1, j_2, j_3} \left\{ \sum_{i=1}^n \mu_{j_1, j_2, j_3}(\mathbf{u}_i, \tau) \right\} \right| \\ & \leq C_1 \kappa_n n^{1/2} J_n^{-1} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|, \end{aligned}$$

其中 $0 < C_1 < \infty$ (对于 q 指标模型, 该收敛速率为 $C_1 \kappa_n n^{1/2} J_n^{-q/2} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|$). 此外, 由引理 A.1 的结论可得

$$\left| \kappa_n \sum_{i=1}^n \mathbb{E} \left\{ \int f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) b(\mathbf{u}_i, \tau) \mathbf{\Pi}(\mathbf{u}_i, \tau)^\top (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0) d\tau \right\} \right| \leq C_2 \kappa_n n J_n^{-a} J_n^{-3/2} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|$$

(对于 q 指标模型, 该收敛速率为 $C_2 \kappa_n n J_n^{-a} J_n^{-(q+1)/2} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|$). 此外, 根据文献 [39, 定理 5.4.2] 可知, 存在一个常数 $0 < C_3 < \infty$ 使得

$$\kappa_n \sum_{i=1}^n 2^{-1} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)^\top \mathbb{E} \left\{ \int_0^1 f(m(\mathbf{u}_i, \tau) | \mathbf{x}_i) \mathbf{\Pi}(\mathbf{u}_i, \tau) \mathbf{\Pi}(\mathbf{u}_i, \tau)^\top d\tau \right\} (\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0)$$

$$\geq C_3 n \kappa_n J_n^{-3} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|^2$$

成立 (对于 q 指标模型, 该收敛速率为 $C_3 n \kappa_n J_n^{-(q+1)} \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}^0\|^2$). 因此以趋近于 1 的概率有

$$\begin{aligned} C_3 n \kappa_n J_n^{-3} \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\|^2 &\leq C_1 \kappa_n n^{1/2} J_n^{-1} \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\| \\ &+ C_2 J_n^{-a-3/2} \kappa_n n \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\| + C(n^{1/2} \kappa_n J_n^{-3/2} + n J_n^{-1/2}) \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\|. \end{aligned}$$

整理后可得

$$\begin{aligned} \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\| &= O_p\{n^{-1} \kappa_n^{-1} J_n^3 (n^{1/2} \kappa_n J_n^{-1} + J_n^{-a-3/2} \kappa_n n + n^{1/2} \kappa_n J_n^{-3/2} + n J_n^{-1/2})\} \\ &= O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}) \end{aligned}$$

(对于 q 指标模型, 该收敛速率为 $O_p(n^{-1/2} J_n^{1+q/2} + \kappa_n^{-1} J_n^{(q+3)/2} + J_n^{-a+(q+1)/2})$). 将其代入非参函数的表达式中有

$$\begin{aligned} \|\tilde{g}_n(\mathbf{U}, \tau) - g_n^0(\mathbf{U}, \tau)\|_\infty &\leq \sup_{\mathcal{B}} \sum_{j_1 \in \mathcal{J}_1} \sum_{j_2 \in \mathcal{J}_2} \sum_{j_3 \in \mathcal{J}_3} (\tilde{\lambda}_{j_1, j_2, j_3} - \lambda_{j_1, j_2, j_3}^0) \pi_{j_1}(U_1) \pi_{j_2}(U_2) \pi_{j_3}(\tau) \\ &\asymp \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\| = O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}). \end{aligned}$$

根据引理 A.1 的结论可得

$$\begin{aligned} \|\tilde{g}_n(\mathbf{X}^\top \mathbf{B}^0, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)\|_\infty &\leq \|\tilde{g}_n(\mathbf{X}^\top \mathbf{B}^0, \tau) - g_n^0(\mathbf{X}^\top \mathbf{B}^0, \tau)\|_\infty \\ &+ \|g_n^0(\mathbf{X}^\top \mathbf{B}^0, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)\|_\infty \\ &= O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}). \end{aligned} \quad (\text{A.10})$$

进一步地, 由 (A.10) 结论可得

$$\begin{aligned} \sup_{\mathcal{A}} |\hat{g}_n(\mathbf{X}^\top \hat{\mathbf{B}}, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)| &\leq \sup_{\mathcal{A}} \{L_1 \|(\mathbf{X}^\top \hat{\mathbf{B}}, \tau) - (\mathbf{X}^\top \mathbf{B}^0, \tau)\| + |\hat{g}_n(\mathbf{X}^\top \mathbf{B}^0, \tau) - \tilde{g}(\mathbf{X}^\top \mathbf{B}^0, \tau)| \\ &+ |\tilde{g}(\mathbf{X}^\top \mathbf{B}^0, \tau) - g_n^0(\mathbf{X}^\top \mathbf{B}^0, \tau)| + |g_n^0(\mathbf{X}^\top \mathbf{B}^0, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)|\} \\ &= \sup_{\mathcal{A}} |\hat{g}_n(\mathbf{X}^\top \mathbf{B}^0, \tau) - \tilde{g}(\mathbf{X}^\top \mathbf{B}^0, \tau)| \\ &+ O_p(n^{-1/2}) + O(J_n^{-a}) + O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}) \\ &= \sup_{\mathcal{A}} |\{\hat{\beta}_{s_1, s_2, s_3}(\hat{\boldsymbol{\lambda}}(\hat{\mathbf{B}})) - \tilde{\beta}_{s_1, s_2, s_3}\}^\top \boldsymbol{\varphi}_{s_1, s_2, s_3}(\mathbf{X}^\top \mathbf{B}^0, \tau)| \\ &+ O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}), \end{aligned} \quad (\text{A.11})$$

其中 L_1 是 Lipschitz 常数, 使得 $\|\Pi(\mathbf{X}^\top \hat{\mathbf{B}}, \tau) - \Pi(\mathbf{X}^\top \mathbf{B}^0, \tau)\| \leq L_1 \|(\mathbf{X}^\top \hat{\mathbf{B}}, \tau) - (\mathbf{X}^\top \mathbf{B}^0, \tau)\|$ 成立.

令 $I_{s_1, s_2, s_3} = [t_{1, s_1}, t_{1, s_1+1}] \times [t_{2, s_2}, t_{2, s_2+1}] \times [t_{3, s_3}, t_{3, s_3+1}]$, $s_i = 1, \dots, J_{n,i}-1$, 其中 t_{i, s_i}^* ($i = 1, 2, 3$) 是各子区间上的中点, 根据 B 样条函数的性质, 将张量积底样条函数 $\boldsymbol{\lambda}^\top \Pi(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 重新写作为 $\boldsymbol{\beta}_{s_1, s_2, s_3}(\boldsymbol{\lambda})^\top \boldsymbol{\varphi}_{s_1, s_2, s_3}(\mathbf{X}^\top \mathbf{B}^0, \tau)$, 其中 $(\mathbf{X}^\top \mathbf{B}^0, \tau) \in I_{s_1, s_2, s_3}$, $\boldsymbol{\beta}_{s_1, s_2, s_3}(\boldsymbol{\lambda}) \in \mathbb{R}^a$,

$$\begin{aligned} \boldsymbol{\varphi}_{s_1, s_2, s_3}(\mathbf{X}^\top \mathbf{B}^0, \tau) &= \left(1, \frac{2\{(\mathbf{X}^\top \mathbf{B}^0)_1 - t_{1, s_1}^*\}}{t_{1, s_1+1} - t_{1, s_1}}, \dots, \left[\frac{2\{(\mathbf{X}^\top \mathbf{B}^0)_1 - t_{1, s_1}^*\}}{t_{1, s_1+1} - t_{1, s_1}} \right]^{w_1-1} \right) \\ &\otimes \left(1, \frac{2\{(\mathbf{X}^\top \mathbf{B}^0)_2 - t_{2, s_2}^*\}}{t_{2, s_2+1} - t_{2, s_2}}, \dots, \left[\frac{2\{(\mathbf{X}^\top \mathbf{B}^0)_2 - t_{2, s_2}^*\}}{t_{2, s_2+1} - t_{2, s_2}} \right]^{w_2-1} \right) \end{aligned}$$

$$\otimes \left(1, \frac{2(\tau - t_{3,s_3}^*)}{t_{3,s_3+1} - t_{3,s_3}}, \dots, \left[\frac{2(\tau - t_{3,s_3}^*)}{t_{3,s_3+1} - t_{3,s_3}} \right]^{w_3-1} \right),$$

这里 $(\mathbf{X}^\top \mathbf{B}^0)_1$ 和 $(\mathbf{X}^\top \mathbf{B}^0)_2$ 分别是向量 $\mathbf{X}^\top \mathbf{B}^0$ 中的前两个元素. 考虑对 $\widehat{\lambda}(\widehat{\mathbf{B}})$ 中第 (s_1, s_2, s_3) 个元素加入一些扰动, 则根据 B 样条基底函数的性质, 一共有 $\sum_i w_i = a$ 个 $\beta(\lambda)$ 受到了影响.

令

$$\begin{aligned}\boldsymbol{\theta}(\boldsymbol{\lambda}) &= (\{\beta_{s_1-w_1, s_2-w_2, s_3-w_3}(\boldsymbol{\lambda}) - \widetilde{\beta}_{s_1-w_1, s_2-w_2, s_3-w_3}\}^\top, \dots, \{\beta_{s_1, s_2, s_3}(\boldsymbol{\lambda}) - \widetilde{\beta}_{s_1, s_2, s_3}\}^\top)^\top, \\ \phi(\mathbf{X}^\top \mathbf{B}^0, \tau) &= (\varphi_{s_1-w_1, s_2-w_2, s_3-w_3}(\mathbf{X}^\top \mathbf{B}^0, \tau), \dots, \varphi_{s_1, s_2, s_3}(\mathbf{X}^\top \mathbf{B}^0, \tau))^\top.\end{aligned}$$

根据文献 [41, 引理 A.1] 的结论可知, 可以给出关于优化问题 $\min_{\boldsymbol{\lambda}} \sum_{i=1}^n \sum_{k=1}^{\kappa_n} \rho_{\tau_k}(y_i - \boldsymbol{\lambda}^\top \Pi(\mathbf{x}_i^\top \widehat{\mathbf{B}}, \tau_k))$ 的梯度最优条件, 即

$$\sum_{i \in \mathcal{S}_{s_1, s_2, s_3}} \sum_{k=1}^{\kappa_n} \psi_{\tau_k}(e_i - \boldsymbol{\theta}(\widehat{\lambda}(\widehat{\mathbf{B}}))^\top \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) + \tilde{r}_n(\mathbf{x}_i^\top \widehat{\mathbf{B}}, \tau_k)) \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) = O(1),$$

其中 $\mathcal{S}_{s_1, s_2, s_3} = \{i : (\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) \in I_{s_1, s_2, s_3}\}$, $e_i = y_i - m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)$, $\tilde{r}_n(\mathbf{x}_i^\top \widehat{\mathbf{B}}, \tau_k) = m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) - \tilde{g}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) + \widehat{\lambda}(\widehat{\mathbf{B}})^\top \{\Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) - \Pi(\mathbf{x}_i^\top \widehat{\mathbf{B}}, \tau_k)\}$. 根据 (A.10), 可得

$$\sup_{\mathbf{x} \in \mathcal{A}} \|\tilde{r}_n(\mathbf{x}^\top \widehat{\mathbf{B}}, \tau_k)\| = O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}).$$

令 M_n 为集合 $\mathcal{S}_{s_1, s_2, s_3}$ 中元素的个数. 考虑 $\boldsymbol{\theta} \in \Theta_L = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq L(\log n/M_n)^{1/2}\}$, $L > 0$, 则有

$$\begin{aligned}\mathbf{A}_{s_1, s_2, s_3}(\boldsymbol{\theta}) &= \sum_{i \in \mathcal{S}_{s_1, s_2, s_3}} \sum_{k=1}^{\kappa_n} \psi_{\tau_k}(e_i - \boldsymbol{\theta}^\top \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) + \tilde{r}_n(\mathbf{x}_i^\top \widehat{\mathbf{B}}, \tau_k)) \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k), \\ \mathbf{Q}_{s_1, s_2, s_3} &= M_n \sum_{i \in \mathcal{S}_{s_1, s_2, s_3}} \sum_{k=1}^{\kappa_n} f(m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) | \mathbf{x}_i) \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) \phi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)^\top.\end{aligned}$$

根据文献 [42, 引理 4.1] 的结论可知,

$$\sup_{\mathcal{A}} \sup_{\boldsymbol{\theta} \in \Theta_L} |\mathbf{A}_{s_1, s_2, s_3}(\boldsymbol{\theta}) + M_n \mathbf{Q}_{s_1, s_2, s_3} \boldsymbol{\theta}(\lambda)| = O_p\left(\left(\frac{\log n}{M_n}\right)^{1/2}\right).$$

假设 $M_n = n/J_n^3$, 根据文献 [41, 定理 1] 的证明过程有

$$\sup_{\mathcal{A}} g |\widehat{\beta}_{s_1, s_2, s_3}(\widehat{\lambda}(\widehat{\mathbf{B}})) - \widetilde{\beta}_{s_1, s_2, s_3} g| = O_p\left(\left(\frac{\log n}{M_n}\right)^{1/2}\right).$$

又因为 $\sup_{\mathcal{A}} |\varphi_{s_1, s_2, s_3}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|^2 \leq a$, 故

$$\sup_{\mathcal{A}} |\widehat{g}_n(\mathbf{X}^\top \mathbf{B}^0, \tau) - \tilde{g}(\mathbf{X}^\top \mathbf{B}^0, \tau)| = O_p\left(\left(J_n^3 \frac{\log n}{n}\right)^{1/2}\right). \quad (\text{A.12})$$

结合 (A.11) 可得

$$\sup_{\mathcal{A}} |\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)| = O_p(n^{-1/2} J_n^2 + \kappa_n^{-1} J_n^{5/2} + J_n^{-a+3/2}).$$

对于一般的 q 指标分位数模型有如下结论:

$$\sup_{\mathcal{A}} |\widehat{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - m(\mathbf{X}^\top \mathbf{B}^0, \tau)| = O_p(n^{-1/2} J_n^{1+q/2} + \kappa_n^{-1} J_n^{(q+3)/2} + J_n^{-a+(q+1)/2}). \quad (\text{A.13})$$

证毕. \square

附录 A.5 定理 4.2 的证明

首先考虑给定 \mathbf{B}^0 情形下, 多指标分位数回归模型估计量 $\tilde{g}(\mathbf{X}^\top \mathbf{B}^0, \tau) = \tilde{\lambda}(\mathbf{B}^0)^\top \Pi(\mathbf{X}^\top \mathbf{B}^0, \tau)$ 的渐近有效性, 其中 $\tilde{\lambda}(\mathbf{B}^0)$ 是由优化问题 $\min_{\lambda} \sum_{i=1}^n \sum_{k=1}^{\kappa_n} \rho_{\tau_k}(y_i - \lambda^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k))$ 估计出的样条系数估计量, 这里简记为 $\hat{\lambda}$. 同时记 $\psi_\tau(u) = \tau - I(u < 0)$,

$$\phi(\mathbf{v}_i, \boldsymbol{\lambda}) = -\frac{1}{\kappa_n} \sum_{k=1}^{\kappa_n} \psi_{\tau_k}(y_i - \boldsymbol{\lambda}^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k),$$

其中 $\mathbf{v}_i = (y_i, \mathbf{x}_i^\top \mathbf{B}^0)^\top$, $i = 1, 2, \dots, n$. 根据引理 A.1 的结论和 Cauchy 不等式可以证明

$$\sum_{i=1}^n \mathbb{E} \|\phi(\mathbf{v}_i, \boldsymbol{\lambda}^0)\|^2 = O_p(n J_n^{q+1}).$$

换言之, 可以证明 $\|\sum_{i=1}^n \phi(\mathbf{v}_i, \boldsymbol{\lambda}^0)\| = O_p(n^{1/2} J_n^{(q+1)/2})$ 成立. 令

$$\eta_i(\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2) = \phi(\mathbf{v}_i, \boldsymbol{\varpi}_1) - \phi(\mathbf{v}_i, \boldsymbol{\varpi}_2) - \mathbb{E}\phi(\mathbf{v}_i, \boldsymbol{\varpi}_1) + \mathbb{E}\phi(\mathbf{v}_i, \boldsymbol{\varpi}_2).$$

根据假设条件 (A2) 和文献 [43, 引理 2.1], 文献 [43] 中的条件 (C4) 和 (C5) 可以满足, 即有

$$\begin{aligned} \sup_{\|\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_2\| \leq C_3 (J_n^{q+1}/n)^{1/2}} \sum_{i=1}^n \mathbb{E} |\boldsymbol{\alpha}^\top \eta_i(\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2)|^2 &= O(n^{1/2} J_n^{2(q+1)}), \\ \sup_{\boldsymbol{\alpha} \in \mathcal{S}_{J_n^{q+1}}} \sup_{\|\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_2\| \leq C_3 (J_n^{q+1}/n)^{1/2}} \sum_{i=1}^n (\boldsymbol{\alpha}^\top \eta_i(\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2))^2 &= O_p(n^{1/2} J_n^{2(q+1)}), \end{aligned}$$

其中 $\boldsymbol{\alpha} \in \mathcal{S}_{J_n^{q+1}}$ 且 $C_3 > 0$ 是常数, 而 \mathcal{S}_p 表示一个 p 维的单位球面上所有的点组成的集合.

根据文献 [44, 引理 4.1] 得知, 对任意的 $s, t \in \mathbb{R}$, 有不等式 $|\psi_\tau(t+s) - \psi_\tau(t)| \leq 5 \min(|s| |t|^{-1}, 1)$ 成立. 因为 B 样条的基底函数在 $(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)$ 处的函数值的上下界依赖于其阶数, 即 $\|\Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)\| \leq \sum_{l=1}^{q+1} d_l$. 可以证明存在一个常数 $C_1 > 0$ 和 K_i ($i = 1, \dots, 6$) 使得

$$\begin{aligned} \mathbb{E} \sup_{\|\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_2\| \leq \delta} \|\eta_i(\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2)\|^2 &= \mathbb{E} \sup_{\|\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_2\| \leq \delta} \left\| \frac{1}{\kappa_n} \sum_{k=1}^{\kappa_n} \{[\psi_{\tau_k}(y_i - \boldsymbol{\varpi}_1^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) - \psi_{\tau_k}(y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k))]\right. \\ &\quad \left. - \{\mathbb{E}\psi_{\tau_k}(y_i - \boldsymbol{\varpi}_1^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) - \mathbb{E}\psi_{\tau_k}(y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k))\}] \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) \right\|^2 \\ &\leq \mathbb{E} \sup_{\|\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_2\| \leq \delta} \left\| \frac{1}{\kappa_n} \sum_{k=1}^{\kappa_n} \left[5 \min \left\{ \frac{|(\boldsymbol{\varpi}_2 - \boldsymbol{\varpi}_1)^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|}{|y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|}, 1 \right\} - \{F_{e^0|\mathbf{x}_i}(m(\mathbf{x}_i^\top \mathbf{B}^0, \tau) \right. \right. \\ &\quad \left. \left. - \boldsymbol{\varpi}_1^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) - F_{e^0|\mathbf{x}_i}(m(\mathbf{x}_i^\top \mathbf{B}^0, \tau) - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k))\} \right] \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) \right\|^2 \\ &\leq \mathbb{E} [K_1 \delta J_n^{q+1} |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|^{-1} I(\delta J_n^{(q+1)/2} \leq |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|)] \\ &\quad + K_1 J_n^{(q+1)/2} I(\delta J_n^{(q+1)/2} > |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|) - K_2 \delta J_n^{q+1}]^2 \\ &\leq K_3 [\mathbb{E} \{K_1 \delta J_n^{q+1} |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|^{-1} I(1 \leq \delta J_n^{(q+1)/2} \leq |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|)\}]^2 \\ &\quad + \mathbb{E} \{K_1 \delta J_n^{q+1} |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|^{-1} I(\delta J_n^{(q+1)/2} \leq 1 \leq |y_i - \boldsymbol{\varpi}_2^\top \Pi(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|)\}^2 \end{aligned}$$

$$\begin{aligned}
& + \mathbb{E}\{K_1\delta J_n^{q+1}|y_i - \boldsymbol{\varpi}_2^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|^{-1} I(\delta J_n^{(q+1)/2} \leq |y_i - \boldsymbol{\varpi}_2^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)| \leq 1)\}^2 \\
& + K_4 J_n^{q+1} \Pr(\delta J_n^{(q+1)/2} + |m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) - \boldsymbol{\varpi}_2^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)|) + (K_2 \delta J_n^{q+1})^2] \\
& \leq K_3 [K_1^2 \delta^2 J_n^{2(q+1)} \{\delta^{-2} J_n^{-(q+1)} I(\delta J_n^{(q+1)/2} \geq 1) + 1 + K_6 \delta^{-1} J_n^{-(q+1)/2}\} \\
& + K_5 \delta J_n^{3(q+1)/2} + K_2^2 \delta^2 J_n^{2(q+1)}] \\
& \leq C_1 n^{1/2} \delta.
\end{aligned}$$

因此证明了文献 [43] 中的条件 (C2) 成立. 至此, 验证了文献 [43] 中所有的假设条件. 现设

$$b^0(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) = \boldsymbol{\lambda}^{0\top} \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) - m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k),$$

根据引理 A.1 可知 $|b^0(\mathbf{X}^\top \mathbf{B}^0, \tau)| = O_p(J_n^{-a})$, 同时

$$\begin{aligned}
& \mathbb{E}\psi_{\tau_k}(y_i - \boldsymbol{\lambda}^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) - \mathbb{E}\psi_{\tau_k}(y_i - \boldsymbol{\lambda}^{0\top} \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) \\
& = f(m(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) | \mathbf{x}_i)(\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) + o(\{(\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\top \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)\}^2).
\end{aligned} \tag{A.14}$$

根据 (A.14) 可以证明存在某个正常数 C_3 使得

$$\begin{aligned}
& \sup_{\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^0\| \leq C_3 n^{-1/2} J_n^{(q+1)/2}} \left| \boldsymbol{\alpha}^\top \sum_{i=1}^n \mathbb{E}(\phi(\mathbf{v}_i, \boldsymbol{\lambda}) - \phi(\mathbf{v}_i, \boldsymbol{\lambda}^0)) - n \boldsymbol{\alpha}^\top \mathbf{D}_n (\boldsymbol{\lambda} - \boldsymbol{\lambda}^0) \right| \\
& = o(n J_n^{(q+1)/2} (n^{-1/2} J_n^{q+1})^2) \\
& = o(J_n^{5(q+1)/2})
\end{aligned}$$

对任意的 $\boldsymbol{\alpha} \in \mathcal{S}_{J_n^{q+1}}$ 成立. 所以我们即可验证 [43] 中的条件 (C3) 成立. 因此, 根据文献 [43, 定理 2.2] 可知,

$$\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0 = -\frac{1}{n \kappa_n} \mathbf{D}_n^{-1} \sum_{i=1}^n \sum_{k=1}^{\kappa_n} \psi_{\tau_k}(y_i - \boldsymbol{\lambda}^{0\top} \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) + \mathbf{r}_n, \tag{A.15}$$

其中 $\|\mathbf{r}_n\| = o_p(n^{-1/2})$. 所以, 对于任意 $\boldsymbol{\alpha} \in \mathcal{S}_{J_n^{q+1}}$, 可以得到

$$\begin{aligned}
& \text{Var} \left\{ \boldsymbol{\alpha}^\top \mathbf{D}_n^{-1} \frac{1}{\kappa_n} \sum_{k=1}^{\kappa_n} \psi_{\tau_k}(y_i - \boldsymbol{\lambda}^{0\top} \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k)) \boldsymbol{\Pi}(\mathbf{x}_i^\top \mathbf{B}^0, \tau_k) \mid \mathbf{x}_i \right\} \\
& = \boldsymbol{\alpha}^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\alpha} + O_p(\kappa_n^{-1} \boldsymbol{\alpha}^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\alpha}) + O_p(J_n^{-a}) + O(\kappa_n^{-2}).
\end{aligned} \tag{A.16}$$

所以, $\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Pi}(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - g_n^0(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)$ 的渐近协方差矩阵是

$$\frac{\boldsymbol{\Pi}(\mathbf{X}^\top \mathbf{B}^0, \tau)^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}^\top \mathbf{B}^0, \tau)}{n}.$$

因此, 可得下列结论:

$$\frac{\tilde{g}_n(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - g_n^0(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau)}{\sqrt{\boldsymbol{\Pi}(\mathbf{X}^\top \mathbf{B}^0, \tau)^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}^\top \mathbf{B}^0, \tau)/n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

此外, 对于 $(\mathbf{X}_{(i)}, \tau_{(i)}), (\mathbf{X}_{(j)}, \tau_{(j)}) \in \mathcal{C}$ 有

$$\frac{\text{ACov}(\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Pi}(\mathbf{X}_{(i)}^\top \mathbf{B}^0, \tau_{(i)}) - g_n^0(\mathbf{X}_{(i)}^\top \mathbf{B}^0, \tau_{(i)}), \tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Pi}(\mathbf{X}_{(j)}^\top \mathbf{B}^0, \tau_{(j)}) g_n^0(\mathbf{X}_{(j)}^\top \mathbf{B}^0, \tau_{(j)}))}{\boldsymbol{\Pi}(\mathbf{X}_{(i)}^\top \mathbf{B}^0, \tau_{(i)})^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}_{(j)}^\top \mathbf{B}^0, \tau_{(j)})/n} \xrightarrow{p} 1,$$

这里用 ACov 指代渐近协方差矩阵. 同样采用类似于上面的证明思路, 并结合文献 [43, 引理 3.3], 得

$$\frac{\text{ACov}(\widehat{g}_n(\mathbf{X}_{(i)}^\top \widehat{\mathbf{B}}, \tau_{(i)}) - g_n^0(\mathbf{X}_{(i)}^\top \widehat{\mathbf{B}}, \tau_{(i)}), \widehat{g}_n(\mathbf{X}_{(j)}^\top \widehat{\mathbf{B}}, \tau_{(j)}) - g_n^0(\mathbf{X}_{(j)}^\top \widehat{\mathbf{B}}, \tau_{(j)}))}{\boldsymbol{\Pi}(\mathbf{X}_{(i)}^\top \mathbf{B}^0, \tau_{(i)})^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}_{(j)}^\top \mathbf{B}^0, \tau_{(j)})/n} \xrightarrow{p} 1.$$

一般而言, 对于任意正整数 k 和 $(\mathbf{X}_{(l)}, \tau_{(l)}) \in \mathcal{C}$ ($l = 1, \dots, k$), 随机向量

$$(Z_n(\mathbf{X}_{(1)}, \tau_{(1)}), Z_n(\mathbf{X}_{(2)}, \tau_{(2)}), \dots, Z_n(\mathbf{X}_{(k)}, \tau_{(k)}))^\top$$

依分布收敛于一个均值为 0、协方差核矩阵为 $\boldsymbol{\Pi}(\mathbf{X}_{(i)}^\top \mathbf{B}^0, \tau_{(i)})^\top \mathbf{D}_n^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n^{-1} \boldsymbol{\Pi}(\mathbf{X}_{(j)}^\top \mathbf{B}^0, \tau_{(j)})/n$ ($1 \leq i \leq j \leq k$) 的多元正态分布.

为了构建 Z_n 的渐近分布, 需要验证同等连续性 (equicontinuity) 条件. 对于任意序列 $\delta_n \rightarrow 0$ 和满足 $\|(\mathbf{X}, \tau) - (\mathbf{X}', \tau')\| \leq \delta_n$ 条件的 $(\mathbf{X}, \tau), (\mathbf{X}', \tau') \in \mathcal{C}$, 有

$$\begin{aligned} & \sup_{\|(\mathbf{X}, \tau)^\top - (\mathbf{X}', \tau')^\top\| \leq \delta_n} |Z_n(\mathbf{X}, \tau) - Z_n(\mathbf{X}', \tau')| \\ &= \sup_{\|(\mathbf{X}, \tau)^\top - (\mathbf{X}', \tau')^\top\| \leq \delta_n} |\{\widehat{\boldsymbol{\lambda}}(\widehat{\mathbf{B}}) - \boldsymbol{\lambda}^0\}^\top \{\boldsymbol{\Pi}(\mathbf{X}^\top \widehat{\mathbf{B}}, \tau) - \boldsymbol{\Pi}(\mathbf{X}'^\top \widehat{\mathbf{B}}, \tau')\}| \\ &= O_p(\delta_n n^{-1/2} J_n^{(q+1)/2}) \end{aligned}$$

成立. 根据条件 (A2), 当 $\delta_n \rightarrow 0$, 有

$$\sup_{\|(\mathbf{X}, \tau)^\top - (\mathbf{X}', \tau')^\top\| \leq \delta_n} |Z_n(\mathbf{X}, \tau) - Z_n(\mathbf{X}', \tau')| \rightarrow 0.$$

根据文献 [45, 定理 10.2], 可以立即证明定理 4.2 的结论.

Estimation and inference for non-crossing multiple-index quantile regression

Chen Dong, Shujie Ma, Liping Zhu & Xingdong Feng

Abstract Though the theoretical properties of quantile regression have been extensively studied in the past three decades, in practice it is not unusual to obtain crossing surfaces by estimating regression functions at different quantile levels with regular approaches. The crossing quantile surfaces are intrinsically uninterpretable. To address this issue, we consider the semiparametric multi-index quantile regression subject to the monotonicity restriction at different quantile levels. We first connect the semiparametric multi-index quantile regression with a dimension-reducible model. Such a relationship allows us to estimate the index coefficients consistently. The B-splines are then used to approximate the nonparametric function under the monotonicity restriction, which numerically corresponds to a constrained linear programming problem. To further improve the efficiency, we estimate the B-spline coefficients based on the dual of the constrained optimization problem. We assess the finite-sample performance of our proposed method through a comprehensive simulation study, and compare the prediction performance of different methods through a real case study.

Keywords dimension reduction, linear programming, multiple-index model, non-crossing, quantile regression

MSC(2020) 62G08, 62G32, 62G20, 62H12

doi: 10.1360/N012018-00193