基于众数回归的部分函数型线性可加模型的稳健估计

余平^{1,2}, 杜江¹, 张忠占^{1*}

1. 北京工业大学应用数理学院, 北京 100124;

2. 山西师范大学数学与计算机科学学院, 临汾 041000

E-mail: ypxye@sina.com, dujiang@bjut.edu.cn, zzhang@bjut.edu.cn

收稿日期: 2017-05-16; 接受日期: 2018-03-20; 网络出版日期: 2019-04-16; * 通信作者

国家自然科学基金 (批准号: 11771032 和 11501018) 资助项目

摘要 本文讨论部分函数型线性可加模型参数的稳健估计, 该模型由经典的可加回归模型和函数型线性模型组合而成. 采用 B-样条基函数对模型中斜率函数和非参数可加函数进行近似, 然后通过最大化众数回归目标函数得到基于众数回归的估计. 在一些正则条件下, 本文给出估计的收敛速度和渐近分布. 最后通过模拟计算和应用实例以表明所提方法的有效性. 模拟结果表明, 该方法不仅具有稳健性, 即不易受污染数据或厚尾分布的影响, 而且在信噪比较大时可以与最小二乘方法有相同的表现.

关键词 众数回归 B-样条 可加模型 函数型线性模型

MSC (2010) 主题分类 62G05, 62N02

1 引言

近年来, 计算工具和存储技术的快速发展, 为我们搜集、存储、传输随时间或者空间连续变化的曲线数据、曲面数据或者三维图像数据提供了方便. 这类数据统称为函数型数据. 函数型数据的概念最早由加拿大统计学家 Ramsay^[1] 于 1982 年提出来. 目前, 函数型数据广泛存在于工程技术、社会科学及自然科学领域, 并且在金融、经济、环境科学、医学等具体学科有着广泛应用, 例如, 人体生长曲线数据, 气象站观测到的某一地区多年的气温数据, 证券交易市场产生的多只股票的分时或日均成交价、收盘价、涨跌幅、交易量、交易额等数据, 心理学研究中的脑电波信号数据, 医学诊断中产生的核磁共振数据等 (参见文献 [2, 3]). 对于这种函数型数据, 传统的统计模型面临巨大的挑战. 一方面, 观测时点密集, 数据具有高维特征; 另一方面, 由于事物或者现象本身具有某种内在的连续性, 所收集到的数据通常还具有多重共线性. 对于这种具有高维特征和共线性的函数型数据, 若采用传统的主成分分析或者偏最小二乘方法来进行降维处理, 可能会造成某些信息的损失, 而且还缺乏可解释性. 为克服这种困难, 统计学者们提出了函数型数据分析 (functional data analysis, FDA) 方法, 并且取得

英文引用格式: Yu P, Du J, Zhang Z Z. Robust estimation for partial functional linear additive model based on modal regression (in Chinese). Sci Sin Math, 2019, 49: 799–814, doi: 10.1360/N012017-00101

了丰富的研究成果. 例如, Ramsay 和 Silverman 在专著 [3, 4] 中对函数型数据分析方法及其应用进行了系统的汇总. Ferraty 和 Vieu 在专著 [2] 中深入研究了函数型数据的非参数回归方法. Horváth 和 Kokoszka 的专著 [5] 对函数型数据的变点识别、独立性检验、函数型时间序列等问题的成果进行了系统的总结. Cardot 等 [6] 利用样条方法研究了函数型线性模型中斜率函数的估计问题并得到了估计的渐近性质. Hall 和 Horowitz [7] 基于函数型主成分分析, 给出了函数型线性模型的最小二乘估计方法, 证明了估计量具有非参数最优收敛速度. Cai 和 Hall [8] 对函数型线性模型的预测问题进行了研究, 证明了预测结果的好坏取决于自变量的协方差函数、斜率函数和样本曲线的光滑程度, 并对预测估计量收敛速度的上下限进行了详细的讨论. Cai 和 Yuan [9] 从再生核 Hilbert 空间的角度, 考虑了函数型协方差算子的核函数和斜率函数所处空间的核函数间的关系, 在较弱的条件下得到了估计与预测的最优收敛速度. Crambes 等 [10] 利用带惩罚的光滑样条方法确保了估计的存在性, 并研究了估计的大样本性质. Lian [11] 介绍了函数型部分线性模型, 采用核估计方法得到了斜率函数和非参数函数估计的收敛速度. Shin [12] 对部分函数型线性回归模型提出了最小二乘估计方法, 并证明了所提参数分量的估计量具有渐近正态性, 斜率函数的估计量具有最优的收敛速度. Müller 和 Stadtmüller [13] 将函数型线性模型推广到广义函数型线性模型并基于拟似然方法进行估计. Peng 等 [14] 提出了变系数部分函数型线性回归模型, 基于样条的方法分别给出斜率系数和变系数函数的收敛速度. 更多关于函数型数据分析的介绍可以参见综述性文献 [15].

上述关于函数型模型的估计方法都集中于均值回归, 基于最小二乘法或似然方法. 众所周知, 均值、中位数和众数是用来度量数据中心的三个重要数字特征. 均值容易受到样本中异常值的影响, 其取值有时可能偏离数据主体, 但是中位数和众数具有抵抗数据中异常点影响的特性, 因此, 二者是稳健的统计量. 由于众数着重反映样本取值的稠密情况, 因此, 在误差具有单峰分布的情形, 利用众数回归模型往往比中位数回归能够得到更窄的置信区间并且具有很好的预测效果. 关于众数回归的研究在参数模型和半参数模型中已有许多研究成果. 例如, Yao 等 [16] 在非参数模型的框架下研究了众数回归的统计推断问题, 严格地证明了此方法具有很高的效率. Zhang 等 [17] 和 Zhao 等 [18] 分别利用局部多项式方法和 B-样条方法把众数回归推广到变系数部分线性模型中. Lv 等 [19] 利用众数回归研究了部分线性可加模型的稳健估计. 更多关于众数回归的结果参见文献 [20–24]. 本文借鉴 Yao 和 Li [20] 众数回归估计的思想, 将众数回归方法应用到一种新的更具有弹性的函数型模型—部分函数型线性可加模型, 该模型也是 Lian [11] 所介绍的函数型部分线性模型的拓展. 据我们目前的了解, 关于函数型数据的众数回归估计还没有相关的研究工作.

本文后续内容结构如下: 第 2 节介绍部分函数型线性可加模型及其估计方法; 第 3 节给出正则条件和大样本性质; 第 4 节提出估计的 EM (expectation-maximization) 迭代算法并讨论调节参数的选择; 第 5 节考虑模型误差服从正态和非正态分布的随机模拟; 第 6 节将所提方法应用到 Tecator 数据中; 第 7 节给出所有定理的证明. 第 8 节给出本文的结论.

2 模型与估计

2.1 部分函数型线性可加模型

考虑如下的部分函数型线性可加模型:

$$Y = g(\mathbf{Z}) + \int_{\mathcal{T}} X(t)\alpha(t)dt + e, \quad (2.1)$$

其中 Y 是定义在概率空间 $(\Omega, \mathcal{B}, P)$ 上的实值随机变量, $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ 是一个 p 维的随机向量, 具有紧的支撑 $\mathcal{Z}$. $X(t)$ 是一个定义在概率空间 $(\Omega, \mathcal{B}, P)$ 上的均值为零的随机过程, 在 $\mathcal{T}$ 上平方可积且二阶矩存在 (即 $E|X(t)|^2 < \infty, \forall t \in \mathcal{T}$), $\alpha(t)$ 为在 $\mathcal{T}$ 上平方可积的斜率函数, e 为随机误差, 且满足 $E(e | X, \mathbf{Z}) = 0$. $g(\mathbf{Z}) = \sum_{j=1}^p g_j(Z_j)$ 为非参数可加函数. 不失一般性, 令 $\mathcal{Z} = [0, 1]^p$, $\mathcal{T} = [0, 1]$, 则 $X(t)$ 为 $H = L^2[0, 1]$ 上随机函数. 为了保证模型的可识别性, 假定 $E(g_j(Z_j)) = 0, j = 1, \dots, p$. 当 $g(\cdot) = 0$ 时, 模型 (2.1) 即为函数型线性回归模型 [5, 7-9]; 当 $\alpha(\cdot) = 0$ 时, 模型 (2.1) 为经典的可加模型 [25]; 当 $p = 1$ 时, 模型 (2.1) 即为函数型部分线性模型 [11].

2.2 众数回归

下面简要介绍众数回归估计方法. 以线性模型 $Y_i = \mathbf{X}_i^T \boldsymbol{\alpha} + \varepsilon_i$ ($i = 1, \dots, n$) 为例, Yao 和 Li [20] 建议通过极大化下面目标函数获得系数向量 $\boldsymbol{\alpha}$ 的估计:

$$Q_h(\boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^n \phi_h(Y_i - \mathbf{X}_i^T \boldsymbol{\alpha}), \quad (2.2)$$

其中 $\phi_h(t) = h^{-1} \phi(\frac{t}{h})$, $\phi(t)$ 是对称核密度函数, h 是控制稳健性的窗宽. 由 (2.2) 获得的估计 $\hat{\boldsymbol{\alpha}}$ 称为众数回归估计. 众数回归估计可以看成是一种 M-估计, 只要合理地选择窗宽 h , 我们就可以获得很好的稳健估计. 关于众数回归和窗宽选择的更多讨论可参见文献 [16, 20].

2.3 估计方法

设 $(X_i, Y_i, \mathbf{Z}_i)$ ($i = 1, \dots, n$) 为来自模型 (2.1) 的一组独立样本. 根据众数回归估计的思想, 模型 (2.1) 的众数回归估计极大化下面的表达式:

$$\frac{1}{n} \sum_{i=1}^n \phi_h \left(Y_i - \sum_{j=1}^p g_j(Z_{ij}) - \int_0^1 X_i(t) \alpha(t) dt \right). \quad (2.3)$$

注意在 (2.3) 中, 可加函数 $g_j(\cdot)$ ($j = 1, \dots, p$) 和斜率函数 $\alpha(\cdot)$ 为未知非参数函数. 对于非参数函数估计常见方法有局部多项式估计和样条估计等. 局部多项式估计属于局部平滑, 样条方法是对函数的全局进行逼近. 本文采用 B-样条对可加函数和斜率函数进行逼近.

首先介绍可加非参数函数 $g_j(\cdot)$ ($j = 1, \dots, p$) 的样条逼近, 斜率函数 $\alpha(\cdot)$ 的样条逼近方法相类似. 令 $\boldsymbol{\pi}_1(\cdot) = (B_1(\cdot), \dots, B_{N_1}(\cdot))^T$ 为一组 B-样条基函数, 其中 $N_1 = v_1 + 1 + k_1$, $v_1 + 1$ 是样条的阶, k_1 是样条基函数的内节点数目. 更多关于 B-样条函数的细节参见文献 [26, 27]. 注意对于不同的函数 $g_j(\cdot)$, 可以采用不同阶数的样条和内节点数进行逼近, 其性质的讨论完全类似. 为简单起见, 我们对每个 $g_j(\cdot)$ 采用相同的样条基和节点, 从而 N_1 也相同. 考虑到可识别性条件 $E(g_j(Z_j)) = 0$ ($j = 1, \dots, p$), 同 Lv 等 [19] 和 Guo 等 [28] 一样, 对 B-样条基函数进行中心化. 定义中心化的样条函数空间 S_n 为

$$S_n \triangleq \left\{ (g_{n1}, \dots, g_{np}) : g_{nj}(z) = \sum_{l=1}^{N_1} \psi_{jl}(z) \gamma_{jl}, (\gamma_{j1}, \dots, \gamma_{jN_1}) \in \mathbb{R}^{N_1}, j = 1, \dots, p \right\}, \quad (2.4)$$

其中 $\psi_{jl}(u) = B_l(u) - \bar{B}_{jl}$ ($j = 1, \dots, p, l = 1, \dots, N_1$), $\bar{B}_{jl} = \frac{1}{n} \sum_{i=1}^n B_l(Z_{ij})$. 在一定的光滑性条件下, $g_j(\cdot)$ 可以由 $g_{nj}(\cdot)$ 逼近:

$$g_j(z) \approx g_{nj}(z) = \sum_{l=1}^{N_1} \psi_{jl}(z) \gamma_{jl} = \boldsymbol{\varphi}_j^T \boldsymbol{\gamma}_j, \quad (2.5)$$

其中 $\varphi_j \triangleq \varphi_j(z) = (\psi_{j1}(z), \dots, \psi_{jN_1}(z))^T$, $\gamma_j = (\gamma_{j1}, \dots, \gamma_{jN_1})^T$ 是样条系数向量.

类似地, 我们引入 $\pi_2(\cdot) = (C_1(\cdot), \dots, C_{N_2}(\cdot))^T$ 为另一组 B-样条基函数用来逼近斜率函数 $\alpha(\cdot)$, 其中 $N_2 = v_2 + 1 + k_2$, $v_2 + 1$ 是样条的阶, k_2 是样条基函数的内节点数目. 在一定的平滑性条件下, 斜率函数 $\alpha(t)$ 可以由下式较好地逼近:

$$\alpha(t) \approx \sum_{l=1}^{N_2} C_l(t) \beta_l = \pi_2(t)^T \beta, \quad (2.6)$$

其中 $\beta = (\beta_1, \dots, \beta_{N_2})$ 为样条系数向量.

进一步地, 令 $\mathbf{W} = \langle X(t), \pi_2(t) \rangle = (\int_0^1 X(t) C_1(t) dt, \dots, \int_0^1 X(t) C_{N_2}(t) dt)^T$, $\mathbf{W}_i = \langle X_i(t), \pi_2(t) \rangle$, $\varphi_{ij} = (\psi_{j1}(Z_{ij}), \dots, \psi_{j(N_1)}(Z_{ij}))^T$, $\Pi_i = (\varphi_{i1}^T, \dots, \varphi_{ip}^T)^T$, $\gamma = (\gamma_1^T, \dots, \gamma_p^T)^T$, $\theta = (\beta^T, \gamma^T)^T$, 其中 φ_{ij} 和 Π_i 分别为 φ_j 和 Π 的样本形式. 类似于 Yu 和 Ruppert^[29], 定义均值函数

$$m(\omega_i; \theta) = \mathbf{W}_i^T \beta + \Pi_i^T \gamma,$$

则均值函数的梯度矩阵为

$$m^{(1)}(\omega, \theta) = \begin{pmatrix} \mathbf{W} \\ \Pi \end{pmatrix}.$$

利用前面的记号, 则 $\alpha(\cdot)$ 和 $g_j(\cdot)$ ($j = 1, \dots, p$) 的众数回归估计 $\hat{\alpha}(\cdot) = \sum_{j=1}^{N_2} \hat{\beta}_j \hat{v}_j(\cdot)$ 和 $\hat{g}_j(\cdot) = \sum_{l=1}^{N_1} \hat{\gamma}_{jl} \psi_{jl}(\cdot)$ ($j = 1, \dots, p$) 可由关于 β 和 γ 极大化如下目标函数得到:

$$Q(\theta) = \sum_{i=1}^n \phi_h(Y_i - m(\omega_i; \theta)) = \sum_{i=1}^n \phi_h(Y_i - \mathbf{W}_i^T \beta - \Pi_i^T \gamma). \quad (2.7)$$

正如 Yao 等^[16]所指出的, 相对于窗宽 h 来说, 核函数的选取并不是特别关键. 因此, 为了计算和讨论性质的方便, 同文献 [18, 22] 一样, 本文中取 $\phi(t)$ 为标准正态密度函数.

3 渐近性质

下面给出上述众数回归估计 $\hat{\alpha}(\cdot)$ 和 $\hat{g}_j(\cdot)$ ($j = 1, \dots, p$) 的正则条件和渐近性质. 首先, 令 $K(t, s) = \text{Cov}(X(t), X(s))$ 表示 $X(t)$ 的协方差函数. $\alpha_0(\cdot)$ 和 $g_{0j}(\cdot)$ ($j = 1, \dots, p$) 分别表示 $\alpha(\cdot)$ 和 $g_j(\cdot)$ ($j = 1, \dots, p$) 的真实值, 相应的样条逼近系数向量分别为 β_0 和 γ_{0j} , 同时记 $\theta_0 = (\beta_0^T, \gamma_0^T)^T$, 具体细节也可以见后面的引理 7.1. $a_n \sim b_n$ 表示当 $n \rightarrow \infty$ 时, $|a_n/b_n|$ 有大于 0 的下界和有限上界. c 表示正的常数, 它在不同地方可能取不同的数值. $\|\cdot\|$ 表示函数的 L^2 范数或者向量的 Euclid 范数. $U = \langle X(t), \alpha_0(t) \rangle$. 同时引入下面的记号:

$$F(z, u, h) = E\{\phi_h''(e) \mid \mathbf{Z} = \mathbf{z}, U = u\},$$

$$G(z, u, h) = E\{\phi_h'(e)^2 \mid \mathbf{Z} = \mathbf{z}, U = u\}.$$

令

$$\Sigma_n = \frac{1}{n} \left\{ \sum_{i=1}^n \phi_h''(e_i) \mathbf{W}_i \mathbf{W}_i^T - \left[\sum_{i=1}^n \phi_h''(e_i) \mathbf{W}_i \Pi_i^T \right] \left[\sum_{j=1}^n \phi_h''(e_j) \Pi_j \Pi_j^T \right]^{-1} \right\}$$

$$\times \left[\sum_{k=1}^n \phi_h''(e_k) \Pi_k \mathbf{W}_k^T \right] \Big\}, \quad (3.1)$$

$$\Delta_n = \frac{1}{n} \left\{ \sum_{i=1}^n (\phi_h'(e_i))^2 \mathbf{W}_i \mathbf{W}_i^T - \left[\sum_{i=1}^n (\phi_h'(e_i))^2 \mathbf{W}_i \Pi_i^T \right] \left[\sum_{j=1}^n \phi_h''(e_j) \Pi_j \Pi_j^T \right]^{-1} \right. \\ \left. \times \left[\sum_{k=1}^n \phi_h''(e_k) \Pi_k \mathbf{W}_k^T \right] \right\}, \quad (3.2)$$

$$\mathbf{R}_{nj} = \frac{1}{n} \left\{ \sum_{i=1}^n \phi_h''(e_i) \varphi_{ij} \varphi_{ij}^T - \left[\sum_{i=1}^n \phi_h''(e_i) \varphi_{ij} \mathbf{W}_i^T \right] \left[\sum_{j=1}^n \phi_h''(e_j) \mathbf{W}_j \mathbf{W}_j^T \right]^{-1} \right. \\ \left. \times \left[\sum_{k=1}^n \phi_h''(e_k) \mathbf{W}_k \varphi_{kj}^T \right] \right\}, \quad j = 1, \dots, p, \quad (3.3)$$

$$\mathbf{S}_{nj} = \frac{1}{n} \left\{ \sum_{i=1}^n (\phi_h'(e_i))^2 \varphi_{ij} \varphi_{ij}^T - \left[\sum_{i=1}^n (\phi_h'(e_i))^2 \varphi_{ij} \mathbf{W}_i^T \right] \left[\sum_{j=1}^n \phi_h''(e_j) \mathbf{W}_j \mathbf{W}_j^T \right]^{-1} \right. \\ \left. \times \left[\sum_{k=1}^n \phi_h''(e_k) \mathbf{W}_k \varphi_{kj}^T \right] \right\}, \quad j = 1, \dots, p. \quad (3.4)$$

在下文中, 对于充分大的 n , 假定 Σ_n 和 $\mathbf{R}_{nj}$ 为非奇异的矩阵. 为了得到众数回归估计的渐近性质, 文中还需要下面几个正则条件.

(C1) $E\|X\|^4 \leq c$.

(C2) 以 $K(t, s)$ 为核的算子为正定算子.

(C3) 令 $\omega_{l1}, \dots, \omega_{lk_l}$ ($l = 1, 2$) 分别为 $[0, 1]$ 区间上非参函数 $g_j(\cdot)$ ($l = 1$) 和斜率函数 $\alpha(\cdot)$ ($l = 2$) 的内节点序列. 记 $\omega_{l0} = 0$, $\omega_{l, k_l+1} = 1$, $h_{li} = \omega_{li} - \omega_{l, i-1}$, $h_l = \max_{1 \leq i \leq k_l+1} \{h_{li}\}$, 则存在非零常数 c_l 使得 $\frac{h_l}{\min_{1 \leq i \leq k_l+1} \{h_{li}\}} < c_l$, $\max_{1 \leq i \leq k_l+1} \{h_{l, i+1} - h_{li}\} = o(k_l^{-1})$ ($l = 1, 2$).

(C4) $k_1 \sim k_2 \sim k$, 即分别用来逼近 $g_j(\cdot)$ 和 $\alpha(\cdot)$ 的样条基函数内节点数都与 k 同阶.

(C5) $E(g_{0j}(Z_j)) = 0$, $g_{0j}(Z_j)$ ($j = 1, \dots, p$) 和 $\alpha_0(t)$ 分别在 $[0, 1]$ 上 r ($r \geq 2$) 阶连续可微.

(C6) 随机向量 $\mathbf{Z}$ 的联合密度函数 $f_{\mathbf{Z}}(\mathbf{z})$ 在支撑 $[0, 1]^p$ 上连续, 且存在正的常数 c_3 和 c_4 使得 $c_3 \leq \inf_{\mathbf{z} \in [0, 1]^p} f_{\mathbf{Z}}(\mathbf{z}) \leq \sup_{\mathbf{z} \in [0, 1]^p} f_{\mathbf{Z}}(\mathbf{z}) \leq c_4$.

(C7) $F(u, \mathbf{z}, h)$ 关于变量 $\mathbf{z}$ 和 u 是连续的. 对于任意的 $h > 0$, 有 $F(u, \mathbf{z}, h) < 0$, 其中窗宽 h 为一个常数, 不依赖于 n .

(C8) 随机误差 e 满足 $E(\phi_h'(e) | \mathbf{z}, u) = 0$, 且 $E(\phi_h''^2(e) | \mathbf{z}, u)$ 、 $E(\phi_h'^3(e) | \mathbf{z}, u)$ 和 $E(\phi_h'''(e) | \mathbf{z}, u)$ 关于 $\mathbf{z}$ 和 u 是连续的.

(C9) 假定依趋近于 1 的概率成立: 随机矩阵 $\Omega(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n m^{(1)}(\boldsymbol{\omega}_i; \boldsymbol{\theta}) m^{(1)}(\boldsymbol{\omega}_i; \boldsymbol{\theta})^T$ 在 $\boldsymbol{\theta}_0$ 的邻域其最小特征值的下极限有大于 0 的下界.

(C10) 随机向量 $\mathbf{Z}$ 的分量 $Z_1, \dots, Z_p$ 相互独立.

条件 (C1) 和 (C2) 是函数型线性模型中常见的基本条件, 具体可以参见文献 [6, 7]. (C2) 要求 $K(t, s)$ 正定是为了保证斜率函数的可识别性. 条件 (C3) 是对斜率函数和可加非参数函数的样条内节点的要求, 即 $\omega_{l1}, \dots, \omega_{lk_l}$ 为 c_l -拟均匀的内节点序列. 条件 (C4) 类似于文献 [22, 条件 (C4)], 主要为了计算方便, 使得斜率函数的样条估计中的内节点数和非参数函数样条估计中的内节点数同阶. 条件 (C5) 后半是对非参数可加函数和斜率函数光滑性的要求, 用来保证非参数可加函数的估计取得最佳收敛速度, 可参见文献 [30, 31]. 条件 (C6) 是可加模型中一般性条件, 可参见文献 [28, 30]. (C7) 和

(C8) 是非参数模型中众数回归估计常见的基本条件, 具体可参见文献 [18, 22, 23]. 条件 (C9) 控制回归函数中各部分之间的共线性. 条件 (C10) 是简化 $Z_1, \dots, Z_p$ 间的相关结构, 便于给出可加函数估计的渐近分布.

定理 3.1 假定正则条件 (C1)–(C10) 成立, 当 $n \rightarrow \infty$ 时, $k \rightarrow \infty$ 且 $k/n \rightarrow 0$, 则有

$$\begin{aligned}\|\hat{\alpha}(\cdot) - \alpha_0(\cdot)\|^2 &= O_p(n^{-1}k + k^{-2r}), \\ \|\hat{g}_j(\cdot) - g_{0j}(\cdot)\|^2 &= O_p(n^{-1}k + k^{-2r}), \quad j = 1, \dots, p.\end{aligned}$$

定理 3.2 假定正则条件 (C1)–(C10) 成立, $n/(k^{2r+1}) = o(1)$, 则对任意固定 $t \in (0, 1)$, 任意固定的 $\mathbf{z} = (z_1, \dots, z_p) \in (0, 1)^p$, 当 $n \rightarrow \infty$ 时, 有

$$\sqrt{\frac{n}{k_2}}(\hat{\alpha}(t) - \alpha^*(t)) \xrightarrow{d} N(0, \Xi(t)), \quad (3.5)$$

$$\sqrt{\frac{n}{k_1}} \begin{pmatrix} \hat{g}_1(z_1) - g_1^*(z_1) \\ \vdots \\ \hat{g}_p(z_p) - g_p^*(z_p) \end{pmatrix} \xrightarrow{d} N(0, \Psi), \quad (3.6)$$

其中 $\alpha^*(\cdot) = B_2(\cdot)^T \beta_0$, $g_j^*(\cdot) = \varphi_j^T \gamma_{0j}$, $\Xi(t) = \lim_{n \rightarrow \infty} \frac{1}{k_2} \boldsymbol{\pi}_2^T(t) \Sigma_n^{-1} \Delta_n \Sigma_n^{-1} \boldsymbol{\pi}_2(t)$, $\Psi = \text{diag}[\psi_1(z_1), \dots, \psi_p(z_p)]$, $\psi_j(z_j) = \lim_{n \rightarrow \infty} \frac{1}{k_1} \boldsymbol{\varphi}_j^T(z_j) \mathbf{R}_{nj}^{-1} \mathbf{S}_{nj} \mathbf{R}_{nj}^{-1} \boldsymbol{\varphi}_j(z_j)$, $\boldsymbol{\varphi}_j(z_j) = (\psi_{j1}(z_j), \dots, \psi_{jN_1}(z_j))^T$ ($j = 1, \dots, p$), Σ_n 、 Δ_n 、 $\mathbf{R}_{nj}$ 和 $\mathbf{S}_{nj}$ 分别如 (3.1)–(3.4) 所示.

定理 3.1 给出了斜率函数和非参数可加函数估计的收敛速度, 当 $k \sim n^{\frac{1}{2r+1}}$ 时, 两个估计都达到 Stone^[32] 所建立的一元非参数回归估计的最优速度. 定理 3.2 中结论 (3.6) 给出同文献 [33] 类似形式的渐近分布, 但文献 [33] 考虑的是广义可加模型的惩罚样条估计, 其模型和估计方法都与本文不同. 注意, 在定理 3.2 中, $\hat{\alpha}(t) - \alpha_0(t) = (\alpha^*(t) - \alpha_0(t)) + (\hat{\alpha}(t) - \alpha^*(t))$. 由第 7 节中引理 7.1 知, $\alpha^*(t) - \alpha_0(t) = O(k_2^{-r})$. 由定理 3.1 可知, $\alpha^*(t) - \alpha_0(t)$ 被 $\hat{\alpha}(t) - \alpha^*(t)$ 所控制. 从而可以利用 $\hat{\alpha}(t) - \alpha^*(t)$ 的渐近性质描述 $\hat{\alpha}(t) - \alpha_0(t)$ 的渐近性质. 例如, 利用 $\Xi(t)$ 的相合估计构造 $\alpha_0(t)$ 的逐点置信带. 类似的结论对于非参数函数 $g_{0j}(\cdot)$ ($j = 1, \dots, p$) 也成立.

4 估计算法和调节参数的选取

本节将首先讨论如何选择窗宽, 然后给出样条的阶和节点的选取办法, 最后拓展 Yao 等^[16] 提出的 MEM (modal expectation-maximization) 算法, 提出部分函数型线性可加模型的两步估计方法.

4.1 窗宽的选择

由于窗宽 h 控制着众数回归估计的稳健性, 因此, 如何选择最佳的 h 是非常关键的工作. 为方便窗宽的选择, 同 Yang 和 Yang^[23] 一样, 假定 e 和 $(X(\cdot), \mathbf{Z})$ 相互独立. 根据 Zhu 等^[22]、Yang 和 Yang^[23] 的建议, 我们选取最优窗宽 h_{opt} :

$$h_{\text{opt}} = \arg \min_h r(h) \triangleq \arg \min_h \frac{G(h)F^{-2}(h)}{\sigma^2}, \quad (4.1)$$

其中 $\sigma^2 = E(e^2)$, $G(h) = E[\phi'_h(e)]^2$, $F(h) = E[\phi''_h(e)]$. 从上面的表示 (4.1) 中可知, h_{opt} 是不依赖于 n 、只依赖于 e 的条件误差分布. 在实际中, e 的条件分布一般是不知道的, 借助 Yao 等^[16] 的思想, 我们可以先利用下面两式获得 $G(h)$ 和 $F(h)$ 的估计:

$$\hat{G}(h) = \frac{1}{n} \sum_{i=1}^n [\phi'_h(\hat{e}_i)]^2, \quad \hat{F}(h) = \frac{1}{n} \sum_{i=1}^n \phi''_h(\hat{e}_i).$$

此时 $r(h)$ 的估计为 $\hat{r}(h) = \frac{\hat{G}(h)\hat{F}^{-2}(h)}{\hat{\sigma}^2}$, 其中 $\hat{e}_i = Y_i - \mathbf{W}_i^T \hat{\beta} - \mathbf{\Pi}_i^T \hat{\gamma}$, $\hat{\beta}$ 、 $\hat{\gamma}$ 和 $\hat{\sigma}$ 是初始的领航估计 (pilot estimates), 如中位数回归估计 (least absolute deviation, LAD) 或者复合分位数估计 (composite quantile regression, CQR). 虽然有了 $r(h)$ 的估计, 但实际当中我们很难得到 h 的显式解. 借助于 Zhao 等^[18] 和 Zhu 等^[22] 格子点搜索方法可以获得 h_{opt} , h 的一些可能格子点可以取为 $h = 0.5\hat{\sigma} \times 1.02^j$, $j = 0, \dots, k$, 其中 k 可以取为 50 或者 100. 在后面第 5 节的实际模拟中, 选取 $k = 100$.

4.2 样条的阶和节点的选择

在样条逼近中, 样条的阶和节点数选取非常重要. 由于高阶样条带来的复杂交互效应以及共线性问题, 我们建议选择低阶样条, 如线性、二次或立方样条. 在后面实际模拟中, 我们选取立方样条 (即 $h_1 = h_2 = 3$) 和等距节点. 类似于文献 [34], 本文采用下面的 BIC (Bayesian information criterion) 准则来选取内节点的个数 k_1 和 k_2 :

$$\text{BIC}(k_1, k_2) = \log \left(\sum_{i=1}^n \phi_h(Y_i - \mathbf{\Pi}_i^T \hat{\gamma}_{(k_1)} - \mathbf{W}_i^T \hat{\beta}_{(k_2)}) \right) + \frac{\log n}{n} (p(h_1 + k_1 + 1) + (h_2 + k_2 + 1)), \quad (4.2)$$

其中 $h_1 = h_2 = 3$, $\hat{\gamma}_{(k_1)}$ 和 $\hat{\beta}_{(k_2)}$ 通过极大化 (2.7) 得到.

4.3 算法

直接极大化 (2.7), 计算很可能比较缓慢. 类似于 Yao 等^[16] 和 Zhu 等^[22], 我们采用下面修正的 MEM 估计算法. 具体地, 记 $\zeta_i = (\mathbf{W}_i^T, \mathbf{\Pi}_i^T)^T$.

步骤 1 (E-step) 由下式更新 $\pi(i | \theta^{(m)})$:

$$\pi(i | \theta^{(m)}) = \frac{\phi_h(Y_i - \zeta_i^T \theta^{(m)})}{\sum_{i=1}^n \phi_h(Y_i - \zeta_i^T \theta^{(m)})}.$$

步骤 2 (M-step) 更新 θ 的估计 $\theta^{(m+1)}$:

$$\theta^{(m+1)} = \arg \max_{\theta} \sum_{i=1}^n (\pi(i | \theta^{(m)}) \log \phi_h(Y_i - \zeta_i^T \theta)) = (\zeta^T \mathbf{V} \zeta)^{-1} \zeta^T \mathbf{V} \mathbf{Y},$$

其中 $\zeta = (\zeta_1, \dots, \zeta_n)^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{V}$ 是一个 $n \times n$ 的对角矩阵, 其第 i 个对角元素为 $\pi(i | \theta^{(m)})$.

步骤 3 重复迭代步骤 1 和 2 直至算法收敛.

5 模拟研究

本节通过数值试验研究文中所提出的稳健众数回归估计方法 (简记为 MR) 在有限样本下的实际表现, 并与最小二乘 B-样条估计 (简记为 LS) 和中位数 B-样条估计 (简记为 LAD) 做比较. 在模拟

中, 样本容量 n 分别取 100、200 和 400. 从以下模型产生数据:

$$Y = g_1(Z) + g_2(U) + \int_0^1 \alpha(t)X(t)dt + e,$$

其中非参数可加函数部分: $g_1(Z) = 2\sin(2\pi Z)$, $g_2(U) = 8U(1 - U^2)$, $Z \sim U(0, 1)$, $U \sim U(-1, 1)$. 函数型线性部分产生与文献 [35] 中的相类似, 即

$$\alpha(t) = \sqrt{2}\sin\left(\frac{\pi t}{2}\right) + 3\sqrt{2}\sin\left(\frac{3\pi t}{2}\right), \quad X(t) = \sum_{j=1}^{50} \xi_j v_j(t),$$

ξ_j 服从均值为 0、方差为 $\lambda_j = 10((j - 0.5)\pi)^{-2}$ 的正态分布, $v_j(t) = \sqrt{2}\sin((j - 0.5)\pi t)$. 取 e 与 (Z, U) 独立. 另外在模拟中, 类似于 Zhao 等 [24], 考虑以下 4 种情形的随机误差: (1) $e \sim N(0, \sigma^2)$, σ 分别取 0.5、0.6、0.7、0.8、0.9 和 1.0; (2) $e \sim t(3)$; (3) $e \sim 0.9N(0, 1) + 0.1N(0, 10^2)$ (简记为 MN). 在模拟中, 我们选取 3 次 B-样条和等距节点, 内节点数按照第 4 节 (4.2) 标准进行选取.

另外, 类似于 Peng 等 [14], 我们用下面根均方误差 (RASE) 来评价 $g_1(\cdot)$ 、 $g_2(\cdot)$ 和 $\alpha(\cdot)$ 估计的精度:

$$\begin{aligned} \text{RASE}_1 &= \left\{ \sum_{k=1}^{n_1} \frac{(\hat{g}_1(z_k) - g_1(z_k))^2}{n_1} \right\}^{1/2}, \\ \text{RASE}_2 &= \left\{ \sum_{k=1}^{n_2} \frac{(\hat{g}_2(u_k) - g_2(u_k))^2}{n_2} \right\}^{1/2}, \\ \text{RASE}_3 &= \left\{ \sum_{k=1}^{n_3} \frac{(\hat{\alpha}(t_k) - \alpha(t_k))^2}{n_3} \right\}^{1/2}, \end{aligned}$$

其中 $\{z_k, k = 1, \dots, n_1\}$ 、 $\{u_k, k = 1, \dots, n_2\}$ 和 $\{t_k, k = 1, \dots, n_3\}$ 分别是计算 $\hat{g}_1(\cdot)$ 、 $\hat{g}_2(\cdot)$ 和 $\hat{\alpha}(\cdot)$ 函数值的等距格子点. 在模拟中, 取格子点数 $n_1 = n_2 = n_3 = 200$.

重复模拟 500 次, 表 1-4 给出 3 种误差分布下的 RASE_j ($j = 1, 2, 3$) 的均值、中位数和方差, 图 1-3 则显示了正态分布误差情形下 RASE_j ($j = 1, 2, 3$) 的均值随 σ 和样本量的变化情况. 由表 1-4 和图 1-3 可以看出, (1) 给定的误差分布下, RASE 的均值、中位数和方差随着样本量 n 的增加而减小; (2) 当误差来自正态分布 $N(0, \sigma^2)$ 时, 总体上 LS 估计表现最好, 在多数情况下 MR 估计介于 LS 估

表 1 $N(0, 0.5^2)$ 误差下 RASE_j ($j = 1, 2, 3$) 的均值、中位数和方差

n	RASE	LS			LAD			MR		
		均值	中位数	方差	均值	中位数	方差	均值	中位数	方差
100	RASE ₁	0.105	0.074	0.009	0.114	0.088	0.010	0.106	0.076	0.009
	RASE ₂	0.057	0.037	0.003	0.064	0.044	0.003	0.058	0.038	0.003
	RASE ₃	0.044	0.030	0.002	0.070	0.049	0.005	0.046	0.030	0.002
200	RASE ₁	0.047	0.032	0.002	0.051	0.038	0.002	0.048	0.033	0.002
	RASE ₂	0.026	0.019	0.001	0.030	0.023	0.001	0.026	0.019	0.001
	RASE ₃	0.021	0.015	0.000	0.032	0.022	0.001	0.022	0.016	0.000
400	RASE ₁	0.021	0.014	0.000	0.022	0.015	0.000	0.021	0.014	0.000
	RASE ₂	0.013	0.009	0.000	0.015	0.012	0.000	0.013	0.010	0.000
	RASE ₃	0.011	0.009	0.000	0.015	0.011	0.000	0.012	0.009	0.000

表 2 $N(0,1)$ 误差下 $RASE_j$ ($j = 1, 2, 3$) 的均值、中位数和方差

n	RASE	LS			LAD			MR		
		均值	中位数	方差	均值	中位数	方差	均值	中位数	方差
100	RASE ₁	0.144	0.115	0.012	0.177	0.148	0.017	0.186	0.153	0.019
	RASE ₂	0.103	0.082	0.007	0.131	0.103	0.010	0.143	0.117	0.012
	RASE ₃	0.170	0.112	0.030	0.273	0.188	0.073	0.298	0.193	0.112
200	RASE ₁	0.069	0.053	0.003	0.084	0.066	0.004	0.087	0.068	0.004
	RASE ₂	0.047	0.037	0.001	0.063	0.054	0.002	0.065	0.055	0.002
	RASE ₃	0.079	0.056	0.006	0.122	0.081	0.014	0.127	0.087	0.015
400	RASE ₁	0.030	0.023	0.001	0.036	0.029	0.001	0.037	0.031	0.001
	RASE ₂	0.023	0.019	0.000	0.030	0.026	0.000	0.030	0.026	0.000
	RASE ₃	0.040	0.028	0.001	0.056	0.038	0.003	0.056	0.038	0.003

表 3 $t(3)$ 误差下 $RASE_j$ ($j = 1, 2, 3$) 的均值、中位数和方差

n	RASE	LS			LAD			MR		
		均值	中位数	方差	均值	中位数	方差	均值	中位数	方差
100	RASE ₁	0.271	0.205	0.083	0.223	0.178	0.029	0.246	0.196	0.042
	RASE ₂	0.219	0.168	0.059	0.172	0.141	0.016	0.196	0.163	0.020
	RASE ₃	0.520	0.327	0.509	0.379	0.261	0.143	0.433	0.279	0.219
200	RASE ₁	0.119	0.099	0.008	0.096	0.079	0.004	0.103	0.085	0.005
	RASE ₂	0.102	0.082	0.010	0.077	0.063	0.003	0.083	0.070	0.003
	RASE ₃	0.241	0.150	0.068	0.161	0.101	0.034	0.160	0.109	0.028
400	RASE ₁	0.059	0.048	0.002	0.044	0.038	0.001	0.044	0.037	0.001
	RASE ₂	0.050	0.041	0.001	0.035	0.030	0.001	0.036	0.031	0.001
	RASE ₃	0.113	0.075	0.016	0.075	0.052	0.005	0.077	0.054	0.006

表 4 混合正态 (MN) 误差下 $RASE_j$ ($j = 1, 2, 3$) 的均值、中位数和方差

n	RASE	LS			LAD			MR		
		均值	中位数	方差	均值	中位数	方差	均值	中位数	方差
100	RASE ₁	0.768	0.612	0.405	0.210	0.177	0.024	0.220	0.175	0.089
	RASE ₂	0.667	0.530	0.305	0.167	0.141	0.017	0.161	0.138	0.014
	RASE ₃	1.794	1.101	4.404	0.351	0.235	0.140	0.351	0.225	0.190
200	RASE ₁	0.339	0.278	0.060	0.093	0.076	0.005	0.091	0.073	0.005
	RASE ₂	0.312	0.247	0.061	0.075	0.064	0.003	0.074	0.061	0.003
	RASE ₃	0.839	0.530	0.938	0.153	0.109	0.020	0.144	0.099	0.021
400	RASE ₁	0.169	0.143	0.014	0.046	0.037	0.001	0.045	0.037	0.001
	RASE ₂	0.151	0.123	0.012	0.035	0.028	0.001	0.035	0.028	0.001
	RASE ₃	0.418	0.284	0.172	0.077	0.051	0.005	0.071	0.047	0.005

图 1 $N(0, \sigma^2)$ 误差下 $g_1(\cdot)$ 三种估计的平均 RASE: LS- 实线、LAD- 虚线和 MR- 点线

图 2 $N(0, \sigma^2)$ 误差下 $g_2(\cdot)$ 三种估计的平均 RASE: LS- 实线、LAD- 虚线和 MR- 点线

图 3 $N(0, \sigma^2)$ 误差下 $\alpha(\cdot)$ 三种估计的平均 RASE: LS- 实线、LAD- 虚线和 MR- 点线

计和 LAD 估计之间, 而对于较大信噪比 (σ 较小) 的情形, LS 估计和 MR 估计表现大致相同; (3) 当误差来自厚尾分布或有异常值时, LAD 估计和 MR 估计表现比较稳健, 而 LS 估计表现比较差, 特别是当误差出现异常值的情形, LAD 估计和 MR 估计的优势更加明显, 我们也计算了标准 Cauchy 分布误差的情形, 表现同 $t(3)$ 相类似, 为了节省空间, 故而略去。

6 实例分析

下面将本文所提众数回归估计方法应用于 Tecator 数据的分析, 并与最小二乘估计作简单比较. 该数据集包含 215 个肉质品观测样本, 每一个肉质品样本包含 100 个由 Tecator 近红外分析仪在波段 850 至 1050 纳米范围内记录的吸收光谱, 以及由化学方法测定的肉质品的脂肪、蛋白质和水分的含量指标. 数据可以从 <http://lib.stat.cmu.edu/datasets/tecator> 下载, 关于数据详细描述也可参见文献 [36, 37]. 我们用函数型预测变量 $X(t)$ 表示吸收光谱, 响应变量 Y 表示中心化的脂肪含量, 实值变量 Z 和 U 分别表示蛋白质含量和水分含量.

此外, 同 Aneiros-Pérez 和 Vieu [36] 一样, 我们将样本分成两部分: $I_1 = \{(X_i, Y_i, Z_i, U_i), i = 1, \dots, 165\}$ 号样本为训练集, $I_2 = \{(X_i, Y_i, Z_i, U_i), i = 166, \dots, 215\}$ 号样本为检验集. 训练集用于模型中参数的估计, 检验集用于验证预测效果. 预测效果采用如下相对预测均方误差准则 (MSEP) [36] 来评价:

$$\text{MSEP} = \frac{1}{50} \sum_{i \in I_2} \frac{(Y_i - \hat{Y}_i)^2}{\text{Var}_{I_2}(Y_i)}.$$

考虑用本文所提出的模型 (2.1) 进行预测, 其中包含关于 U 和 Z 的两个可加未知函数. 图 4 为用最小二乘回归代替本文考虑的众数回归所得模型 (2.1) 的残差的 Q-Q 图. 显然, 残差分布明显偏离

图 4 模型 (2.1) 的最小二乘拟合残差的 Q-Q 图

正态, 并有离群值. 对残差进行 Shapiro-Wilk 正态性检验, p 值小于 0.001. 这启发我们使用具有稳健性的众数回归. 利用最小二乘法拟合模型得到的 MSEF 为 0.021, 而用众数回归拟合的模型进行预测, 所得 MSEF 为 0.016. 这表明, 利用该模型达到了相当好的预测精度, 说明脂肪含量可以用蛋白质含量和水分含量很好地预测, 从而可以在今后的实验中节省实验成本, 除非要求很高的精度. 这也表明本文提出的模型和估计方法在实际中是有效的.

7 主要结论及其证明

引理 7.1 函数 $g_{0j}(z)$ ($j = 1, \dots, p$) 和 $\alpha_0(t)$ 满足条件 (C3), k_1 和 k_2 满足定理 3.1 中 k 的条件, 则分别存在样条逼近 $g_j^*(z) = \varphi_j^T \gamma_{0j}$ 和 $\alpha^*(t) = \pi_2^T(t) \beta_0$ 使得

$$\sup_{z \in [0,1]} |g_{0j}(z) - g_j^*(z)| = O_p(k_1^{-r} + k_1^{1/2} n^{-1/2}), \quad \sup_{t \in [0,1]} |\alpha_0(t) - \alpha^*(t)| = O_p(k_2^{-r} + k_2^{1/2} n^{-1/2}),$$

进一步地, 若内节点数 $k_1 \sim k_2 \sim n^{\frac{1}{2r+1}}$, 则

$$\sup_{z \in [0,1]} \|g_{0j}(z) - g_j^*(z)\| = O_p(n^{-\frac{r}{2r+1}}), \quad \sup_{t \in [0,1]} |\alpha_0(t) - \alpha^*(t)| = O_p(n^{-\frac{r}{2r+1}}),$$

其中 $\gamma_{0j} = (\gamma_{0j,1}, \dots, \gamma_{0j,N_1})^T$ ($j = 1, \dots, p$), $\beta_0 = (\beta_{01}, \dots, \beta_{0N_2})^T$.

引理 7.1 的证明可参见文献 [26, 引理 1] 的证明.

定理 3.1 的证明 令 $\gamma_0 = (\gamma_{01}^T, \dots, \gamma_{0p}^T)^T$, $\theta_0 = (\beta_0^T, \gamma_0^T)^T$, $\delta = k^{-r} + k^{\frac{1}{2}} n^{-\frac{1}{2}}$, $T_1 = \delta^{-1}(\beta - \beta_0)$, $T_2 = \delta^{-1}(\gamma - \gamma_0)$, $T = (T_1^T, T_2^T)^T$.

首先将证明, 对于任意给定的 $\epsilon > 0$, 存在一个充分大的正数 L 使得当 n 充分大时,

$$P\left\{\sup_{\|T\|=L} Q(\theta_0 + \delta T) < Q(\theta_0)\right\} \geq 1 - \epsilon. \quad (7.1)$$

由 ϵ 的任意性, (7.1) 意味着依趋于 1 的概率, $Q(\boldsymbol{\theta})$ 在球 $\{\boldsymbol{\theta}_0 + \delta \mathbf{T} : \|\mathbf{T}\| \leq L\}$ 的内部存在一个局部最大值点 $\hat{\boldsymbol{\theta}}$, 使得 $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(\delta)$. 利用 Taylor 公式简单计算有

$$\begin{aligned} Q(\boldsymbol{\theta}_0 + \delta \mathbf{T}) - Q(\boldsymbol{\theta}_0) &= -\delta \sum_{i=1}^n \phi'_h(e_i + R_{1i} + R_{2i})\Omega_i + \frac{1}{2}\delta^2 \sum_{i=1}^n \phi''_h(e_i + R_{1i} + R_{2i})\Omega_i^2 \\ &\quad - \frac{1}{6}\delta^3 \sum_{i=1}^n \phi'''_h(\zeta_i)\Omega_i^3 \\ &\triangleq A_1 + A_2 + A_3, \end{aligned}$$

其中 ζ_i 介于 $(e_i + R_{1i} + R_{2i})$ 和 $(e_i + R_{1i} + R_{2i} - \delta\Omega_i)$ 之间,

$$R_{1i} = \sum_{j=1}^p g_{0j}(Z_{ij}) - \boldsymbol{\Pi}_i^T \boldsymbol{\gamma}_0, \quad R_{2i} = \langle X_i(t), \alpha_0(t) - \boldsymbol{\pi}_2(t)^T \boldsymbol{\beta}_0 \rangle, \quad \Omega_i = \mathbf{W}_i^T T_1 + \boldsymbol{\Pi}_i^T T_2.$$

利用条件 (C1)、引理 7.1 和 Hölder 不等式知, $|R_{1i}| = O_p(\delta)$, $|R_{2i}| = O_p(\delta)$.

再次利用 Taylor 公式可得

$$\sum_{i=1}^n \phi'_h(e_i + R_{1i} + R_{2i})\Omega_i = \sum_{i=1}^n \left\{ \phi'_h(e_i) + \phi''_h(e_i)(R_{1i} + R_{2i}) + \frac{1}{2}\phi'''_h(\tilde{e}_i)(R_{1i} + R_{2i})^2 \right\} \Omega_i,$$

其中 $\tilde{e}_i$ 介于 e_i 和 $e_i + R_{1i} + R_{2i}$ 之间. 利用条件 (C6) 和 (C8), 可得 $\|A_1\| = O_p(n\delta^2\|\mathbf{T}\|)$. 对于 A_2 , 简单计算可得

$$A_2 = O_p(E(F(\mathbf{z}, u, h)n\delta^2\|\mathbf{T}\|^2)).$$

因此, 对于充分大的正整数 L , 当 $\|\mathbf{T}\| = L$ 时, A_1 被 A_2 一致控制. 类似地, 利用条件 (C8) 和 $\delta = o(1)$, 当 $\|\mathbf{T}\| = L$ 时, 有 $\delta\|\mathbf{T}\| \rightarrow 0$, 简单计算有 $A_3 = O_p(n\delta^3\|\mathbf{T}\|^3)$. 故当 $\|\mathbf{T}\| = L$ 时, A_2 一致控制 A_3 . 由上可知, 选择充分大的 L , (7.1) 成立, 即存在着局部最大值点 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\boldsymbol{\gamma}}$ 使得 $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(\delta)$, $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta)$.

下面考虑 $\|\hat{\alpha}(t) - \alpha_0(t)\|$. 记 $R_1(t) = \alpha_0(t) - \mathbf{B}_2^T(t)\boldsymbol{\beta}_0$, 则有

$$\begin{aligned} \|\hat{\alpha}(t) - \alpha_0(t)\|^2 &= \int_0^1 (\mathbf{B}_2^T(t)\hat{\boldsymbol{\beta}} - \alpha_0(t))^2 dt \\ &= \int_0^1 (\mathbf{B}_2^T(t)\hat{\boldsymbol{\beta}} - \mathbf{B}_2^T(t)\boldsymbol{\beta}_0 + R_1(t))^2 dt \\ &\leq 2 \int_0^1 \{\mathbf{B}_2^T(t)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}^2 dt + 2 \int_0^1 R_1(t)^2 dt \\ &= 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + 2 \int_0^1 R_1(t)^2 dt, \end{aligned}$$

其中 $\mathbf{H} = \int_0^1 \mathbf{B}_1(t)\mathbf{B}_1^T(t)dt$. 调用 $\|\mathbf{H}\| = O(1)$ (参见文献 [22]) 和 $\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| = O_p(\delta)$, 简单计算可得

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(\delta^2).$$

此外, 根据 $\int_0^1 R_1(t)^2 dt = O_p(\delta^2)$ 可知, 定理 3.1 中第 1 部分结论成立. 同理可证定理 3.1 中第 2 部分结论成立. 证毕. $\square$

定理 3.2 的证明 由前面定理 3.1 的证明知, 当 $n \rightarrow \infty$ 时, 依趋于 1 的概率 $Q(\theta)$ 在 $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$ 处达到最大值. 因而, 用 Taylor 公式简单计算, 有

$$\begin{aligned} -\frac{1}{n} \frac{\partial Q(\theta)}{\partial \beta} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}} &= \frac{1}{n} \sum_{i=1}^n \phi'_h \{Y_i - \mathbf{W}_i^T \hat{\beta} - \Pi_i \hat{\gamma}\} \mathbf{W}_i \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \phi'_h(e_i) + \phi''_h(e_i) U_i + \frac{1}{2} \phi'''_h(\check{e}_i) U_i^2 \right\} \mathbf{W}_i = 0, \end{aligned} \quad (7.2)$$

$$\begin{aligned} -\frac{1}{n} \frac{\partial Q(\hat{\theta})}{\partial \gamma} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}} &= \frac{1}{n} \sum_{i=1}^n \phi'_h \{Y_i - \mathbf{W}_i^T \hat{\beta} - \Pi_i \hat{\gamma}\} \Pi_i^T \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \phi'_h(e_i) + \phi''_h(e_i) U_i + \frac{1}{2} \phi'''_h(\check{e}_i) U_i^2 \right\} \Pi_i = 0, \end{aligned} \quad (7.3)$$

其中

$$U_i = R_{1i} + R_{2i} - \mathbf{W}_i^T (\hat{\beta} - \beta_0) - \Pi_i^T (\hat{\gamma} - \gamma_0),$$

$\check{e}_i$ 和 $\check{e}_i$ 都介于 e_i 和 $Y_i - \mathbf{W}_i^T \hat{\beta} - \Pi_i \hat{\gamma}$ 之间, R_{1i} 和 R_{2i} 定义见定理 3.1 中的证明过程. 令

$$\begin{aligned} \Phi_n &= \frac{1}{n} \sum_{i=1}^n \phi''_h(e_i) \Pi_i \Pi_i^T, \\ \Gamma_n &= \frac{1}{n} \sum_{i=1}^n \phi''_h(e_i) \Pi_i \mathbf{W}_i^T, \\ \Lambda_n &= \frac{1}{n} \sum_{i=1}^n \Pi_i [\phi'_h(e_i) + \phi''_h(e_i) (R_{1i} + R_{2i})]. \end{aligned}$$

由 (7.3) 简单计算可得

$$\hat{\gamma} - \gamma_0 = (\Phi_n + o_p(1))^{-1} (\Lambda_n + \Gamma_n (\beta_0 - \hat{\beta})). \quad (7.4)$$

根据 Σ_n 和 Δ_n 的定义, 把 (7.4) 代入 (7.2) 中, 利用条件 (C1)、(C8) 和引理 7.1 简单计算可得

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta_0) &= \frac{1}{\sqrt{n}} (\Sigma_n)^{-1} \left\{ \sum_{i=1}^n \phi'_h(e_i) \left[\mathbf{W}_i - \left(\sum_{k=1}^n \phi''_h(e_k) \mathbf{W}_k \Pi_k^T \right) \right. \right. \\ &\quad \left. \left. \times \left(\sum_{j=1}^n \phi''_h(e_j) \Pi_j \Pi_j^T \right)^{-1} \Pi_i \right] \right\} + o_p(1). \end{aligned} \quad (7.5)$$

由于 $\hat{\alpha}(t) - \alpha^*(t) = \mathbf{B}_1^T(t) (\hat{\beta} - \beta_0)$, 则对任意固定的 $t \in (0, 1)$, 当 $n \rightarrow \infty$ 时, 利用 Slutsky 定理和多元正态分布的性质可得

$$\sqrt{\frac{n}{k_2}} (\hat{\alpha}(t) - \alpha^*(t)) \xrightarrow{d} N(0, \Xi(t)), \quad (7.6)$$

其中

$$\Xi(t) = \lim_{n \rightarrow \infty} \frac{1}{k_2} \pi_2^T(t) \Sigma_n^{-1} \Delta_n \Sigma_n^{-1} \pi_2(t).$$

类似可证 $g_{0j}(\cdot)$ ($j = 1, \dots, p$) 的渐近正态性. 由于 $\hat{g}_j(z_j) - g_j^*(z_j) = \varphi_j^T(z_j) (\hat{\gamma}_j - \gamma_{0j})$, 对于任意固定的 $z_j \in [0, 1]$, 当 $n \rightarrow \infty$ 时, 可得

$$\sqrt{\frac{n}{k_1}} (\hat{g}_j(z_j) - g_j^*(z_j)) \xrightarrow{d} N(0, \psi_j(z_j)), \quad j = 1, \dots, p, \quad (7.7)$$

其中

$$\psi_j(z_j) = \lim_{n \rightarrow \infty} \frac{1}{k_1} \boldsymbol{\varphi}_j^T(z_j) \mathbf{R}_{nj}^{-1} \mathbf{S}_{nj} \mathbf{R}_{nj}^{-1} \boldsymbol{\varphi}_j(z_j), \quad \boldsymbol{\varphi}_j(z_j) = (\psi_{j1}(z_j), \dots, \psi_{jN_1}(z_j))^T.$$

证毕. $\square$

8 结论

本文考虑了一种新的更具有弹性的部分函数型线性可加模型的众数回归估计问题, 利用 B-样条函数逼近可加非参数函数和斜率函数, 得到估计量的相合性和渐近正态性. 模拟表明, 众数估计方法可以较处理好函数型数据中响应变量具有厚尾特征或者异常值的情况. 相应估计方法可以用到其他函数型模型, 如部分函数型线性模型^[12,35]、变系数部分函数型线性模型^[14]等. 此外, 本文只考虑函数一元函数型预测变量, 进一步可以推广到基于众数回归估计的多元函数型预测变量的变量选择问题.

致谢 感谢匿名审稿人给出的宝贵意见和建议.

参考文献

- 1 Ramsay J O. When the data are functions. *Psychometrika*, 1982, 47: 379–396
- 2 Ferraty F, Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer-Verlag, 2006
- 3 Ramsay J O, Silverman B W. *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer-Verlag, 2002
- 4 Ramsay J O, Silverman B W. *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag, 2005
- 5 Horváth L, Kokoszka P. *Inference for Functional Data with Applications*. New York: Springer-Verlag, 2012
- 6 Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. *Statist Sinica*, 2003, 13: 571–592
- 7 Hall P, Horowitz J L. Methodology and convergence rates for functional linear regression. *Ann Statist*, 2007, 35: 70–91
- 8 Cai T T, Hall P. Prediction in functional linear regression. *Ann Statist*, 2006, 34: 2159–2179
- 9 Cai T T, Yuan M. Minimax and adaptive prediction for functional linear regression. *J Amer Statist Assoc*, 2012, 107: 1201–1216
- 10 Crambes C, Kneip A, Sarda P. Smoothing splines estimators for functional linear regression. *Ann Statist*, 2009, 37: 35–72
- 11 Lian H. Functional partial linear model. *J Nonparametr Stat*, 2011, 23: 115–128
- 12 Shin H. Partial functional linear regression. *J Statist Plann Inference*, 2009, 139: 3405–3418
- 13 Müller H G, Stadtmüller U. Generalized functional linear models. *Ann Statist*, 2005, 3: 774–805
- 14 Peng Q Y, Zhou J J, Tang N S. Varying coefficient partially functional linear regression models. *Statist Papers*, 2016, 57: 827–841
- 15 Wang J L, Chiou J M, Müller H G. Functional data analysis. *Annu Rev Stat Appl*, 2016, 3: 257–295
- 16 Yao W, Lindsay B G, Li R. Local modal regression. *J Nonparametr Stat*, 2012, 24: 647–663
- 17 Zhang R, Zhao W, Liu J. Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. *J Nonparametr Stat*, 2013, 25: 523–544
- 18 Zhao W, Zhang R, Liu J, et al. Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Ann Inst Statist Math*, 2014, 66: 165–191
- 19 Lv J, Yang H, Guo C. Variable selection in partially linear additive models for modal regression. *Comm Statist Simulation Comput*, 2017, 46: 5646–5665
- 20 Yao W, Li L. A new regression model: Modal linear regression. *Scand J Statist*, 2014, 41: 656–671
- 21 Yang H, Yang J. A robust and efficient estimation and variable selection method for partially linear single-index models. *J Multivariate Anal*, 2014, 129: 227–242
- 22 Zhu H, Lv Z, Yu K, et al. Robust variable selection in partially varying coefficient single-index model. *J Korean Statist Soc*, 2015, 44: 45–57
- 23 Yang J, Yang H. A robust penalized estimation for identification in semiparametric additive models. *Statist Probab Lett*, 2016, 110: 268–277
- 24 Zhao W, Zhang R, Liu J, et al. Robust adaptive estimation for semivarying coefficient models. *Statist Probab Lett*, 2015, 97: 132–141

- 25 Buja A, Hastie T, Tibshirani R. Linear smoothers and additive models. *Ann Statist*, 1989, 17: 453–510
- 26 Huang J, Horowitz J L, Wei F. Variable selection in nonparametric additive models. *Ann Statist*, 2010, 38: 2282–2313
- 27 Schumaker L L. *Spline Functions: Basic Theory*. New York: John Wiley & Sons, 1981
- 28 Guo J, Tang M, Tian M, et al. Variable selection in high-dimensional partially linear additive models for composite quantile regression. *Comput Statist Data Anal*, 2013, 65: 56–67
- 29 Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. *J Amer Statist Assoc*, 2002, 97: 1042–1054
- 30 Horowitz J L, Lee S. Nonparametric estimation of an additive quantile regression model. *J Amer Statist Assoc*, 2005, 100: 1238–1249
- 31 Zhou J, Chen M. Spline estimators for semi-functional linear model. *Statist Probab Lett*, 2012, 82: 505–513
- 32 Stone C J. Optimal global rates of convergence for nonparametric regression. *Ann Statist*, 1982, 10: 1040–1053
- 33 Yoshida T, Naito K. Asymptotics for penalised splines in generalised additive models. *J Nonparametr Stat*, 2014, 26: 269–289
- 34 Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann Statist*, 2011, 39: 305–332
- 35 Yu P, Zhang Z, Du J. A test of linearity in partial functional linear regression. *Metrika*, 2016, 79: 953–969
- 36 Aneiros-Pérez G, Vieu P. Semi-functional partial linear regression. *Statist Probab Lett*, 2006, 76: 1102–1110
- 37 Kim H J. A best linear threshold classification with scale mixture of skew normal populations. *Comput Statist*, 2015, 30: 1–28

Robust estimation for partial functional linear additive model based on modal regression

Ping Yu, Jiang Du & Zhongzhan Zhang

Abstract In this paper, we present a new robust estimation method based on modal regression for partial functional linear additive regression model, which combines the classic additive regression model with the functional linear regression model. Both the slope function and nonparametric additive functions are approximated by B-splines, and thus estimated through maximizing the modal objective function. Under some regularity conditions, we give the convergence rates and asymptotic normality of the estimators. Finally, simulation studies and real data analysis are conducted to investigate the performance of the proposed methodologies, where it turns out that the estimators obtained are not only robust against outliers or heavy-tail error distributions, but also highly efficient as least squares estimators when the signal-noise ratio is high.

Keywords modal regression, B-spline, additive model, functional linear model

MSC(2010) 62G05, 62N02

doi: 10.1360/N012017-00101