

# 单指标分位回归模型估计的 MM 算法\*

郭媛媛<sup>1</sup>, 杨雪梅<sup>2</sup>, 孙志华<sup>1,3†</sup>

(1 中国科学院大学数学科学学院, 北京 100049; 2 华北电力大学数理学院, 北京 102206;  
3 中国科学院大数据挖掘与知识管理重点实验室, 北京 100190)  
(2019 年 5 月 14 日收稿; 2019 年 10 月 9 日收修改稿)

Guo Y Y, Yang X M, Sun Z H. MM algorithm of the estimation of single-index quantile regression[J]. Journal of University of Chinese Academy of Sciences, 2021, 38(3): 289-296.

**摘 要** 单指标分位回归模型是一类重要的半参数模型, 具有降维的优点的同时保留了非参数分位回归模型的稳健性。但现有的单指标分位回归模型的估计程序大部分都是通过内点法来实现。对单指标分位回归模型估计程序的 MM (majorize-minimize) 算法进行研究。首先找到目标函数的优化函数, 然后通过最小化优化函数来得到估计, 再逐步迭代至收敛。数值模拟和实证研究表明 MM 算法在单指标分位回归模型的估计中具有较好的稳定性, 能够得到比较准确的估计结果, 且相比于内点法, 计算效率更高, 耗时更短。

**关键词** 单指标分位回归模型; MM 算法; 替代函数; 计算效率

中图分类号: O212.4 文献标志码: A doi:10.7523/j.issn.2095-6134.2021.03.001

## MM algorithm of the estimation of single-index quantile regression

GUO Yuanyuan<sup>1</sup>, YANG Xuemei<sup>2</sup>, SUN Zhihua<sup>1,3</sup>

(1 School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;  
2 School of Mathematics and Physics, North China Electric Power University, Beijing 102206, China;  
3 Key Laboratory of Big Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract** The single-index quantile regression model is an important semiparametric model with the merit of dimensionality reduction. Furthermore, it retains the robustness of a nonparametric model. For most existing estimating procedures of single-index quantile regression models, the estimators are obtained via minimizing the objective functions by the interior point method. In this paper, we investigate the MM (majorize-minimize) algorithm of the single index quantile regression model estimating procedure. We first construct the majorize function of the objective function and then minimize the substituted majorize function to find the estimators. Our numerical simulations and empirical study show that for the considered model, the MM algorithm has good stability and can yield more accurate estimation. Compared with the interior point method, the MM algorithm is more efficient and takes less time.

**Keywords** single-index quantile regression model; MM-algorithm; surrogate function; computational efficiency

\* 国家自然科学基金(11971045)、中央高校基本科研业务费专项资金和中国科学院大数据挖掘与知识管理重点实验室开放课题资助  
† 通信作者, E-mail: sunzh@amss.ac.cn

分位回归模型具有稳健的特点,并且能够对响应变量的分布做出精细的描述,因此获得很多学者的关注。目前,分位回归方法已经成为分析数据的一个非常重要的工具,广泛地应用在金融、医疗、生物研究等领域。分位回归模型的研究和应用可参考文献[1-4]及相关文献。参数分位回归模型是在实际中应用广泛的一类回归模型<sup>[5]</sup>。但很多时候无法对参数分位回归模型进行正确的设定。基于误定的参数分位回归的统计推断结果经常是不可信的,文献[6-7]阐述了参数分位回归需要进行检验的必要性。非参数分位回归不存在误定的风险,但当样本量比较小且协变量比较多时,非参数分位回归方法可能会受到维数祸根的问题的困扰。

对分位回归,构建目标函数时用到的损失函数  $\rho_\tau(r) = \tau r - I_{|r| < 0}$  具有不光滑的特点,从而使得求解目标函数的最小值比较困难,且可能出现多个最小值点的情况,参见文献[5,8-11]。一种解决上面问题的方法是将分位回归模型的求解问题转化为线性规划问题,再利用单纯形法或内点法进行计算。不管是单纯形法,还是内点法,运算效率都不能令人满意。2000年,Hu和Lange<sup>[12]</sup>提出一种新的用于求解分位回归问题的算法,即MM算法。MM算法概念简单,易于执行,且数值稳定,比内点法拥有更强的数值计算能力。文献[11]对4种求解分位回归问题的算法,即内点法、MM算法、坐标下降法和ADMM算法进行比较研究,验证了MM算法具有数值稳定和计算效率高的特点。

单指标分位回归模型具有降维的效果,同时保持了非参数分位回归的稳健性,其估计问题的研究吸引了很多研究者的兴趣。文献[13-14]提出基于两步迭代的估计方法,文献[15-16]进一步提出不需要迭代的估计方法,文献[17-18]又提出基于贝叶斯方法的估计方法,文献[19-21]探讨单指标分位回归模型的变量选择以及加权复合单指标分位回归模型的估计。然而这些文献所提的估计方法都基于内点法来实现,内点法在计算分位回归模型时,计算效率低、耗时久,尤其在样本量较大的情况下,这种缺点更为明显。MM算法在求解分位回归模型的估计时比较高效和便捷,这在文献[22-24]中均有体现,但是没有文献研究单指标分位回归模型的MM算法,故本文研究单指标分位回归模型估计的MM算法。

我们借鉴文献[12]的方法,对单指标分位回归模型的每一步迭代程序中目标函数构建其替代函数,从而将复杂的优化问题简单化。然后,基于优化函数再进行求解计算得到估计值。我们构建的优化函数是光滑的,并能够保证每次迭代目标函数是下降的。数值模拟和实例分析结果表明基于MM算法的估计程序具有较好的稳定性,能够得到比较准确的估计结果,并且相较于传统的内点算法具有更强的数值计算能力,用时更短。

## 1 单指标分位回归模型的估计介绍

对于给定的分位数  $\tau \in (0,1)$ , 在给定  $\mathbf{x}$  的条件下,响应变量  $y$  的  $\tau$  分位数  $\theta_\tau(\mathbf{x})$  与协变量  $\mathbf{x}$  之间的关系如下:

$$\theta_\tau(\mathbf{x}) = g(\mathbf{x}^T \boldsymbol{\gamma}),$$

其中  $\mathbf{x} \in \mathbb{R}^d$  是  $d$  维协变量,  $g(\cdot)$  表示未知的一元联系函数。另外  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)^T$  为未知的单指标向量,满足  $\|\boldsymbol{\gamma}\| = 1$  且  $\gamma_1 > 0$ ,  $\|\cdot\|$  表示 Euclidean 范数。这个约束条件是为了模型的可识别性<sup>[25]</sup>,已广泛应用在有关单指标模型的文献中。

本文采用局部线性方法对  $\boldsymbol{\gamma}$  和  $g(\cdot)$  进行估计,详细内容可参考文献[13],具体算法如下:

**step 1** 采用文献[26]提出的平均导数法得到  $\boldsymbol{\gamma}$  的初值,记为  $\hat{\boldsymbol{\gamma}}^{(0)}$ , 并对其进行标准化,使其满足  $\|\hat{\boldsymbol{\gamma}}\| = 1$  且  $\hat{\gamma}_1 > 0$ 。

**step 2** 对给定的  $\hat{\boldsymbol{\gamma}}$ , 考虑下面的最小化问题:

$$\min_{(a_j, b_j)} \sum_{i=1}^n \rho_\tau(y_i - a_j - b_j(\mathbf{x}_i - \mathbf{x}_j)^T \hat{\boldsymbol{\gamma}}) w_{ij}, \quad (1)$$

其中  $\rho_\tau(r) = r(\tau - I(r \leq 0))$ ,  $w_{ij} = \frac{K_h(\mathbf{x}_i^T \boldsymbol{\gamma} - \mathbf{x}_j^T \boldsymbol{\gamma})}{\sum_{i=1}^n K_h(\mathbf{x}_i^T \boldsymbol{\gamma} - \mathbf{x}_j^T \boldsymbol{\gamma})}$ ,  $K(\cdot)$  为核函数且  $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ 。

记所得估计为  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ 。带宽的选择为  $h_\tau = h_m \{\tau(1-\tau)/\phi(\Phi^{-1}(\tau))^2\}^{1/5}$ , 其中  $\phi(\cdot)$ ,  $\Phi(\cdot)$  分别为标准正态分布的概率密度函数和分布函数,  $h_m = \left\{ \frac{\left[ \int K^2(v) dv \right] [\text{var}(y | \mathbf{x}^T \boldsymbol{\gamma} = u)]}{n \left[ \int v^2 K(v) dv \right]^2 \left[ \frac{d^2}{du^2} E(y | \mathbf{x}^T \boldsymbol{\gamma} = u) \right]^2 [f_{u_0}(u)]} \right\}^{1/5}$ ,

为均值回归时的最优带宽,详细推导过程见文献

[27]。

**step 3** 基于 step 1 得到的估计  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ , 进一步考虑下面的问题:

$$\min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_{\tau}(y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)^T \gamma) w_{ij}, \quad (2)$$

其中  $w_{ij}$  和  $h_{\tau}$  直接使用上一步计算的结果。记所得估计为  $\tilde{\gamma}$ 。

**step 4** 重复 step 1 和 step 2 直至收敛。

在上述算法中,  $\hat{\gamma}$  标准化的方法为:  $\gamma = \text{sign}_1 \tilde{\gamma} / \|\tilde{\gamma}\|$ , 其中  $\text{sign}_1$  指  $\gamma$  的第一个分量的符号函数, 只有第一步  $\hat{\gamma}$  需要标准化。最终  $g(\cdot)$  在  $u$  处的估计结果为  $\hat{g}(\cdot; h; \hat{\gamma}) = \hat{a}$ , 其中

$$(\hat{a}, \hat{b}) = \underset{a, b}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}(y_i - a - b(\mathbf{x}_i^T \hat{\gamma} - u)) K_h(\mathbf{x}_i^T \hat{\gamma} - u). \quad (3)$$

## 2 估计程序的MM算法

### 2.1 MM算法介绍

下面介绍MM算法的基本思想。假设需要最小化的目标函数为  $L(\theta): \mathbb{R}^p \rightarrow \mathbb{R}$ ,  $\theta^k$  为第  $k$  步的迭代值。MM算法每次迭代分两步来进行。首先, 构造目标函数的优化函数  $Q(\theta | \theta^k): \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$  满足

$$Q(\theta^k | \theta^k) = L(\theta^k), \\ Q(\theta | \theta^k) \geq L(\theta) \quad \forall \theta. \quad (4)$$

然后, 对优化函数  $Q(\theta | \theta^k)$  进行最小化, 得到下一步的迭代值  $\theta^{k+1}$ , 则有

$$Q(\theta^{k+1} | \theta^k) \leq Q(\theta^k | \theta^k). \quad (5)$$

综合式(4)和式(5), 可知  $L(\theta^{k+1}) \leq L(\theta^k)$ 。这种下降趋势保证了MM算法具有显著的数值稳定性。

观察目标函数式(1)~式(3), 可以发现它们均为非光滑函数, 因而不易得到最优解, 故借用文献[12]提出的MM算法的思想来处理这个问题。其主要的处理方式如下: 首先给  $\rho_{\tau}(r)$  加一扰动  $\varepsilon$ , 得到其近似函数

$$\rho_{\tau}^{\varepsilon}(r) = \rho_{\tau}(r) - \frac{\varepsilon}{2} \ln(\varepsilon + |r|).$$

然后, 在  $r^k$  处用下式来优化  $\rho_{\tau}^{\varepsilon}(r)$

$$\zeta_{\tau}^{\varepsilon}(r | r^k) = \frac{1}{4} \left[ \frac{r^2}{\varepsilon + |r^k|} + (4\tau - 2)r + c \right],$$

其中,  $c$  是满足  $\zeta_{\tau}^{\varepsilon}(r^k | r^k) = \rho_{\tau}^{\varepsilon}(r^k)$  的常数。下面

给出单指标分位回归模型估计程序中最小化问题(1)、(2)和(3)的基于MM算法的详细求解过程。

### 2.2 非参数部分的估计

首先, 将式(1)中的目标函数修正为

$$L_{\varepsilon}^j(\theta) = \sum_{i=1}^n \rho_{\tau}^{\varepsilon}(r_{ij}(\theta)) w_{ij}, j = 1, 2, \dots, n,$$

其中:  $r_{ij}(\theta) = y_i - a_j - b_j(\mathbf{x}_i - \mathbf{x}_j)^T \hat{\gamma}$ ,  $\theta$  为参数  $(a_j, b_j)$ 。进一步构造近似函数  $L_{\varepsilon}^j(\theta)$  在第  $k$  次迭代值  $\theta^k$  处的优化函数:

$$Q_{\tau, j}^{\varepsilon}(\theta | \theta^k) = \sum_{i=1}^n \zeta_{\tau}^{\varepsilon}(r_{ij} | r_{ij}^k) w_{ij}, j = 1, 2, \dots, n. \quad (6)$$

下面考虑第  $k+1$  次迭代值的确定问题。MM算法是通过优化函数  $Q_{\tau}^{\varepsilon}(\theta | \theta^k)$  的最小化来得到  $\theta$  的估计: 使  $Q_{\tau}^{\varepsilon}(\theta | \theta^k)$  达到最小的  $\theta$  值即为下一次迭代值  $\theta^{k+1}$ 。由于残差项  $r$  为  $\theta$  的线性函数, 对式(6)求导, 令其为零并化简整理可得

$$\sum_{i=1}^n w_{ij} \left( \frac{Y_i}{(\varepsilon + |r_{ij}^k|)} + (2\tau - 1) \right) X_i = \sum_{i=1}^n \frac{w_{ij}}{(\varepsilon + |r_{ij}^k|)} X_i X_i^T \theta, j = 1, 2, \dots, n,$$

其中  $Y_i = y_i$ ,  $X_i = (1, (\mathbf{x}_i - \mathbf{x}_j)^T \hat{\gamma})^T$ ,  $r_{ij}^k = y_i - a_j^k - b_j^k(\mathbf{x}_i - \mathbf{x}_j)^T = Y_i - X_i^T \theta^k$ , 故  $\theta^{k+1}$  的表达式:

$$\theta^{k+1} = \left( \sum_{i=1}^n \frac{w_{ij}}{(\varepsilon + |r_{ij}^k|)} X_i X_i^T \right)^{-1} \sum_{i=1}^n w_{ij} \left( \frac{Y_i}{(\varepsilon + |r_{ij}^k|)} + (2\tau - 1) X_i \right). \quad (7)$$

再通过式(7)的迭代, 逐渐逼近目标函数  $L_{\varepsilon}^j(\theta)$  的最小值。

由此, 可以将MM算法总结为如下步骤:

- 1) 选择迭代初始值  $\theta^0$  和一个较小的正常数  $\varepsilon$ , 置  $k = 0$ ;
- 2) 计算  $r_{ij}^k$ , 根据式(7)确定第  $k+1$  次迭代值  $\theta^{k+1}$ ;
- 3) 令  $k = k + 1$ , 判断是否满足收敛准则, 若满足收敛准则, 即: 当

$$\left| \frac{Q_{\tau}^{\varepsilon}(\theta^k | \theta^k) - Q_{\tau}^{\varepsilon}(\theta^{k+1} | \theta^k)}{Q_{\tau}^{\varepsilon}(\theta^k | \theta^k)} \right| < \delta$$

或者  $Q_{\tau}^{\varepsilon}(\theta^k | \theta^k) - Q_{\tau}^{\varepsilon}(\theta^{k+1} | \theta^k) < \delta$  时, 可终止迭代, 其中  $\delta$  是预先取定的足够小的数。否则返回2)继续迭代, 直到满足收敛准则。

2.3 单指标向量的估计

首先定义式(2)中目标函数的近似函数:

$$L_{\tau}^{\varepsilon}(\boldsymbol{\gamma}) = \sum_{j=1}^n \sum_{i=1}^n \boldsymbol{\rho}_{\tau}^{\varepsilon}(y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\gamma}) w_{ij}.$$

在  $\boldsymbol{\gamma}^k$  处的优化函数可以构建为

$$R_{\tau}^{\varepsilon}(\boldsymbol{\gamma} | \boldsymbol{\gamma}^k) = \sum_{j=1}^n \sum_{i=1}^n \frac{1}{4} \left[ \frac{r_{ij}(\boldsymbol{\gamma})^2}{\varepsilon + |r_{ij}^k|} + (4\tau - 2)r_{ij}(\boldsymbol{\gamma}) + c \right] w_{ij},$$

其中:  $r_{ij}(\boldsymbol{\gamma}) = y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\gamma}$ ,  $c$  是满足使  $R_{\tau}^{\varepsilon}(\boldsymbol{\gamma}^k | \boldsymbol{\gamma}^k) = L_{\tau}^{\varepsilon}(\boldsymbol{\gamma}^k)$  的常数. 然后, 对  $R_{\tau}^{\varepsilon}(\boldsymbol{\gamma} | \boldsymbol{\gamma}^k)$  关于  $\boldsymbol{\gamma}$  求导并令其为 0, 得到  $\boldsymbol{\gamma}^{k+1}$  为

$$\boldsymbol{\gamma}^{k+1} = \left( \sum_{j=1}^n \sum_{i=1}^n \frac{w_{ij}}{(\varepsilon + |r_{ij}^k|)} \mathbf{X}_{ij} \mathbf{X}_{ij}^T \right)^{-1} \sum_{j=1}^n \sum_{i=1}^n w_{ij} \left( \frac{Y_{ij}}{(\varepsilon + |r_{ij}^k|)} + (2\tau - 1) \right) \mathbf{X}_{ij}. \quad (8)$$

其中  $\mathbf{X}_{ij} = \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)$ ,  $Y_{ij} = y_i - \hat{a}_j$ ,  $r_{ij}^k = y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\gamma}^k$ . 按照 2.2 节所给出的 MM 算法将  $\boldsymbol{\theta}$  替换为  $\boldsymbol{\gamma}$ ,  $Q_{\tau}^{\varepsilon}(\cdot)$  替换为  $R_{\tau}^{\varepsilon}(\cdot)$ , 对式(8)的迭代, 即可得到  $\boldsymbol{\gamma}$  的估计.

最后, 可将基于 MM 算法的单指标模型的估计总结为如下步骤:

1) 参考第 1 节 step 1 所提供的方法, 得到  $\boldsymbol{\gamma}$  的初始估计;

2) 对给定的  $\hat{\boldsymbol{\gamma}}$ , 按照 2.2 节计算出  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ ;

3) 对于给定的  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ , 按照 2.3 节计算出  $\hat{\boldsymbol{\gamma}}$ ;

4) 重复 2)、3) 步骤, 直至收敛.

2.4 联系函数  $g(\cdot)$  的估计

对于联系函数的估计, 式(6)中目标函数式的近似函数可定义为

$$L_{1,\tau}^{\varepsilon}(a, b) = \sum_{i=1}^n \rho_{\tau}^{\varepsilon}(y_i - a - b(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - u)) K_h(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - u).$$

按照 2.2 节方法构建  $L_{1,\tau}^{\varepsilon}(a, b)$  的优化函数, 求解可得

$$(\hat{a}, \hat{b}) = \left( \sum_{i=1}^n \frac{w_i}{(\varepsilon + |r_i^k|)} \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} \sum_{i=1}^n w_i \left( \frac{Y_i}{(\varepsilon + |r_i^k|)} + (2\tau - 1) \right) \mathbf{X}_i.$$

则  $\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) = \hat{a}$ , 其中,  $w_i = K_h(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - u)$ ,  $\mathbf{X}_i = (1, \mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - u)^T$ ,  $Y_i = y_i$ ,  $r_i^k = y_i - \hat{a} - \hat{b}(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - u)$ .

3 数值模拟

对于近似函数  $\rho_{\tau}^{\varepsilon}(r)$  中的扰动  $\varepsilon$ , 按照文献 [12] 给出的建议, 选取  $\varepsilon$  满足  $\varepsilon n | \ln \varepsilon | = \delta$ .

3.1 模拟 1

借鉴文献[13]模拟 1 的模型设置, 考虑模型

$$y = \sin\left(\frac{\pi(u - A)}{C - A}\right) + 0.1Z,$$

其中  $u = \mathbf{x}^T \boldsymbol{\gamma}$ ,  $\mathbf{x} = (x_1, x_2, x_3)^T$ ,  $\boldsymbol{\gamma} = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ ,

$$A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}, C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}; x_i \text{ i. i. d. } \sim$$

$U(0, 1)$ ,  $i = 1, 2, 3$ ;  $Z \sim N(0, 1)$ , 且  $x_i$  与  $Z$  相互独立. 取样本容量  $n = 200$ , 在分位数  $\tau$  分别取 0.25、0.5、0.75 的情况下估计  $\boldsymbol{\gamma}$ , 分别模拟 100.

选择高斯核函数  $K(\nu) = \frac{1}{\sqrt{2\pi}} e^{-\nu^2/2}$ , 迭代精度  $\delta$

取  $1 \times 10^{-6}$ , 带宽  $h_{\tau}$  的选取如前所述.

类似参考文献[13, 19, 28]所用的评价标准, 用 Bias(样本偏差)、SE(样本标准差)、MSE(样本均方误差)、CI(95%的置信区间)、CP(真值落在置信区间的比例)评价单指标向量部分估计  $\hat{\boldsymbol{\gamma}}$  的估计精度, 用 ASE 和 AAE 评价联系函数部分  $\hat{g}(\cdot)$  的估计精度, 并给出 ASE 与 AAE 的 Mean(样本均值)与 SE(样本均方误差), 其中 ASE 和 AAE 的定义如下:

$$ASE = n^{-1} \sum_{i=1}^n [\hat{g}(u_i) - g(u_i)]^2,$$

$$AAE = n^{-1} \sum_{i=1}^n |\hat{g}(u_i) - g(u_i)|.$$

在不同的分位数下, 计算上述评价指标值, 所得结果列于表 1 和表 2, 可以看出, 用 MM 算法计算单指标分位回归模型, 无论是单指标向量的估计还是联系函数的估计, 都有比较好的结果, 且在较小和较大的分位数下依然有良好的表现, 这表明本文所提出的计算方法是有效的. 将该估计结果与文献[13]中关于该模型的模拟结果进行对比, 可以看出, 用 MM 算法计算单指标分位回归模型, 与内点法相比, 估计的精度是相似的, 估计系数的偏差大小都在  $10^{-3} \sim 10^{-2}$ , 估计系数的样本标准差数量级均为  $10^{-2}$ , 但在计算效率上, 本文所提出的方法大大优于内点法, 这将在模拟 3 中展示和说明.



表 1 模拟 1 中不同的  $\tau$  下,估计  $\hat{\gamma}$  的 Bias、SE、MSE、CI、CP

Table 1 The Bias、SE、MSE、CI、CP of $\hat{\gamma}$ under different choices of $\tau$ in simulation 1						
	评价 标准	Bias	SE	MSE	CI	CP
$\tau=0.25$	$\hat{\gamma}_1$	0.001 9	0.075 5	0.005 7	(0.431 1,0.727 4)	0.94
	$\hat{\gamma}_2$	-0.007 5	0.079 4	0.006 3	(0.414 1,0.725 5)	0.93
	$\hat{\gamma}_3$	-0.003 6	0.095 2	0.009 1	(0.387 0,0.760 3)	0.96
$\tau=0.5$	$\hat{\gamma}_1$	0.001 7	0.038 8	0.001 5	(0.503 0, 0.655 1)	0.94
	$\hat{\gamma}_2$	0.001 0	0.037 8	0.001 4	(0.502 0, 0.650 4)	0.94
	$\hat{\gamma}_3$	0.005 9	0.034 9	0.001 2	(0.514 7, 0.651 8)	0.94
$\tau=0.75$	$\hat{\gamma}_1$	-0.006 6	0.025 8	0.000 7	(0.520 1,0.621 2)	0.95
	$\hat{\gamma}_2$	0.002 7	0.024 6	0.000 6	(0.531 8,0.628 3)	0.93
	$\hat{\gamma}_3$	0.006 1	0.006 0	0.000 6	(0.536 6,0.630 2)	0.93

表 2 模拟 1 中不同  $\tau$  的选择下, $\hat{g}(u)$  的 ASE、AAE 的 Mean、SE

Table 2 Outcomes of $\hat{g}(u)$ for the models under different choices of $\tau$ in simulation 1						
$\hat{g}(u)$	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
	Mean	SE	Mean	SE	Mean	SE
ASE	0.142 1	0.020 1	0.043 6	0.004 8	0.060 7	0.006 1
AAE	0.340 9	0.021 1	0.175 1	0.009 1	0.168 6	0.009 3

3.2 模拟 2

借鉴文献[13]模拟 2 的模型设置,考虑模型

$$y = 10\sin(0.75u) + \sqrt{1 + \sin(u)}Z,$$

其中  $u = \mathbf{x}^T \boldsymbol{\gamma}, \mathbf{x} = (x_1, x_2)^T, \boldsymbol{\gamma} = (1, 2)^T / \sqrt{5}, x_i \text{ i. i. d } \sim N(0, 0.25^2), i = 1, 2; Z \sim N(0, 1)$ , 且  $x_i$  与  $Z$  相互独立。取样本容量  $n = 400$ , 在分位数  $\tau$  分别取 0.25、0.5、0.75 的情况下估计  $\boldsymbol{\gamma}$ , 分别模拟 100 次。其他设置同模拟 1, 所得结果列于表 3 和表 4, 可以看出新提出的算法在方差的分布不相同时, 依然具有良好的估计结果, 具有稳健性。将该估计结果与文献[13]中对应模型的模拟结果对比, 依然可以发现, 用 MM 算法计算单指标分位回归模型, 与内点法相比, 估计的精度是相似的, 再次表明我们所提出的方法是有效的。

3.3 与内点法比较

由前两个模拟可以发现, MM 算法在单指标分位回归估计的计算问题中表现良好, 接下来比较 MM 算法与内点法的计算效率, 考虑如下 3 个模型:

- 模型 1:  $Y = \exp(X_1' \boldsymbol{\gamma}_1) + \varepsilon_1;$
- 模型 2:  $Y = 5\cos(X_2' \boldsymbol{\gamma}_2) + \varepsilon_2;$

表 3 模拟 2 中不同的  $\tau$  下,估计  $\hat{\gamma}$  的 Bias、SE、MSE、CI、CP

Table 3 The Bias、SE、MSE、CI、CP of $\hat{\gamma}$ under different choices of $\tau$ in simulation 2						
	评价 标准	Bias	SE	MSE	CI	CP
$\tau = 0.25$	$\hat{\gamma}_1$	-0.014 1	0.118 1	0.014 1	(0.201 5,0.664 5)	0.97
	$\hat{\gamma}_2$	-0.000 2	0.060 8	0.003 7	(0.775 3,1.013 9)	0.93
$\tau = 0.5$	$\hat{\gamma}_1$	-0.000 5	0.114 9	0.013 2	(0.221 3, 0.671 9)	0.93
	$\hat{\gamma}_2$	-0.006 2	0.057 3	0.003 3	(0.775 8, 1.000 3)	0.95
$\tau = 0.75$	$\hat{\gamma}_1$	-0.004 4	0.108 4	0.011 7	(0.230 2,0.655 3)	0.95
	$\hat{\gamma}_2$	-0.003 1	0.056 3	0.003 1	(0.780 9,1.001 7)	0.92

表 4 模拟 2 中不同  $\tau$  的选择下, $\hat{g}(u)$  的 ASE、AAE 的 Mean、SE

Table 4 Outcomes of $\hat{g}(u)$ for the models under different choices of $\tau$ in simulation 2						
$\hat{g}(u)$	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
	Mean	SE	Mean	SE	Mean	SE
ASE	0.054 4	0.047 1	0.029 7	0.027 1	0.051 3	0.050 4
AAE	0.385 3	0.176 6	0.269 6	0.126 9	0.367 1	0.204 0

模型 3:  $Y = 5\sin(X_3' \boldsymbol{\gamma}_3) + \exp(-16(X_3' \boldsymbol{\gamma}_3)^2) + \varepsilon_3.$

其中  $\boldsymbol{\gamma}_1 = (1, 1, 1)^T / \sqrt{3}, X_{1,i} \sim N(0, 1), i = 1, 2, 3, \varepsilon_1 \sim N(0, 1); \boldsymbol{\gamma}_2 = (2, 1, 2)^T / 3, X_{2,i} \sim U(0, 1), i = 1, 2, 3, \varepsilon_2 \sim N(0, 1); \boldsymbol{\gamma}_3 = (1, 2)^T / \sqrt{5}, X_{3,i} \sim N(0, 1), i = 1, 2, \varepsilon_3 \sim \text{Exp}(1)$ 。对每个模型, 取样本量分别为 100、200, 模拟 100 次, 用 MM 算法和内点法进行估计计算, 并用 100 次模拟的总体运行时间与平均迭代次数比较算法的运行效率, 所得结果列于表 5、表 6。

表 5  $n=100$ , 模型 1、2、3 的估计结果比较

Table 5 Estimation comparison among models 1, 2, and 3 with  $n=100$

		总体运行时间/h	平均迭代次数
model 1	MM	0.43	66.45
	IP	2.20	62.26
model 2	MM	0.30	54.83
	IP	0.65	17.27
model 3	MM	0.26	41.32
	IP	0.54	17.36

注: IP 指内点算法。

可以发现 MM 算法所用的时间远远少于内点法, 且随着样本量的增大, 这种计算效率上的优势更加明显。这是由于用内点法解决分位回归

表 6  $n=200$ ,模型 1、2、3 的估计结果比较  
Table 6 Estimation comparison among models  
1, 2, and 3 with  $n=200$

		总体运行时间/h	平均迭代次数
model 1	MM	1. 51	56. 97
	IP	25. 16	52. 99
model 2	MM	1. 09	50. 73
	IP	10. 98	22. 37
model 3	MM	0. 95	37. 98
	IP	8. 01	20. 03

问题,是将目标函数及约束条件转化为线性规划问题,再用内点法来求解该问题,但转化之后的线性规划问题,协变量维数与样本量的大小有着正相关的关系,这种方法极大地增加了算法的计算量与所用时间。本文第 3.2 节 中的问题转化为线性规划后,协变量的维数为  $2n+p$ ,第 3.3 节 中的问题转化为线性规划后,维数为  $2n^2+p$ ,具体转化方法及维数的增加量可见文献[11]。而 MM 算法只需对  $p$  维矩阵做运算,故两种方法的计算效率随样本量的增加会产生越来越大的差距。

4 实例分析

在这一节,应用本文所提的估计方法对 Boston 房屋数据集进行分析,Boston 房屋数据集包含 506 个样本。该数据集及其详细介绍均可在 R 包 mlbench (<http://cran.r-project.org/>) 找到。该数据集的响应变量 MEDV(自住房屋房价的中位数)为右删失数据,当其大于 50 000 时,按 50 000 来记录,故比较适合用分位回归来分析。该数据集已被很多文献分析过,借鉴文献[13-14,19] 的设置,选取 MEDV(自住房屋房价的中位数)为响应变量,同时选取 RM(每间住宅的房屋数目),TAX(每一万美元的不动产税率),PTRATIO(城镇中学生与教师的比例),LSTAT(该地区房东收入较低的百分比)4 个协变量,对 TAX 与 LSTAT 进行对数变换,并对响应变量做中心化处理,构造如下模型

$$\hat{\epsilon}_{\tau}(\text{MEDV} \mid \text{RM}, \text{TAX}, \text{PTRATIO}, \text{LSTAT}) = g(\gamma_1 \text{RM} + \gamma_2 \log(\text{TAX}) + \gamma_3 \text{PTRATIO} + \gamma_4 \log(\text{LSTAT})).$$

用本文提出的方法对该问题进行估计,计算在不同分位数下系数的估计值,并采用 bootstrap 方法估计标准差,方法如下,具体细节可参考文献[13,29]。

1) 计算每个样本的误差,  $\hat{\epsilon}_{\tau,i} = y_i - \hat{g}_{\tau}(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}})$ . 再对误差进行中心化;

- 2)对  $\{\hat{\epsilon}_{\tau,i}\}$  进行重抽样得到  $\{\epsilon_{\tau,i}^*\}$  ;
- 3)产生新的响应变量  $y_i^* = \hat{g}_{\tau}(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}}) + \epsilon_{\tau,i}^*$  ;
- 4)根据新的样本  $(\mathbf{x}_i, y_i^*)$ , 得到新的估计。

重复模拟 100 次计算标准差,所得结果列于表 7。从表 7 可以发现, RM 的系数在不同的分位数下皆为正,这表明每栋房屋的房间数量越多,房价就越高且收入越多的家庭更加在意每栋房屋的房间数量;log(TAX)的系数为负且随分位数逐渐变大,这表明不动产的税率越高,房价越低且收入较低的家庭更加在意不动产税率的大小;PTRATIO 的系数为负且随分位数变化较小,这表明学生与教师的比例越大,即教师资源越匮乏,房价越低且低收入家庭与高收入家庭对教育的重视程度是同样大的;log(LSTAT) 系数为负且随分位数逐渐变小,这表明一个地区低收入人群所占的百分比越高,房价越低且收入较高的家庭更加在意一个地区的低收入人群比例。

表 7 波士顿数据集在单指标分位回归模型下的  
系数估计及标准差估计

Table 7 Coefficient estimation and standard deviation  
estimation of Boston data set under the single-index  
quantile regression mode

$\tau$	RM	log(TAX)	PTRATIO	log(LSTAT)
0. 1	0. 304 0 (0. 038 2)	-0. 610 7 (0. 051 2)	-0. 064 4 (0. 010 0)	-0. 567 2 (0. 052 3)
0. 25	0. 390 4 (0. 042 0)	-0. 508 3 (0. 045 1)	-0. 060 8 (0. 009 7)	-0. 533 3 (0. 048 7)
0. 5	0. 484 3 (0. 038 4)	-0. 446 7 (0. 044 6)	-0. 071 3 (0. 008 6)	-0. 614 6 (0. 043 8)
0. 75	0. 639 7 (0. 052 4)	-0. 338 8 (0. 067 1)	-0. 069 5 (0. 010 2)	-0. 592 0 (0. 061 9)
0. 9	0. 672 8 (0. 071 8)	-0. 078 0 (0. 089 6)	-0. 055 2 (0. 012 9)	-0. 810 6 (0. 076 8)

注:括号中数据为标准差估计。

图 1 展示了不同分位数下  $g(\cdot)$  的估计,其中,横坐标为估计所得的  $\hat{\boldsymbol{\gamma}}$  与协变量  $\mathbf{X}$  的线性组合,纵坐标为对应的联系函数的估计值  $\hat{g}(\cdot)$ ,散点为响应变量  $Y$  的实际观测值。观察图 1,可以发现散点(真实观测值)都紧紧围绕在估计函数  $\hat{g}(\cdot)$  的周围且估计曲线随分位数的增大而上移,这表明本文提出的方法应用在 Boston 数据集上是有效的。此外,还可以发现,在不同分位数下,所估计单指标  $\mathbf{x}^T \boldsymbol{\gamma}_{\tau}$  的取值范围不同,这说明采用分位回归分析该问题是非常有必要的。

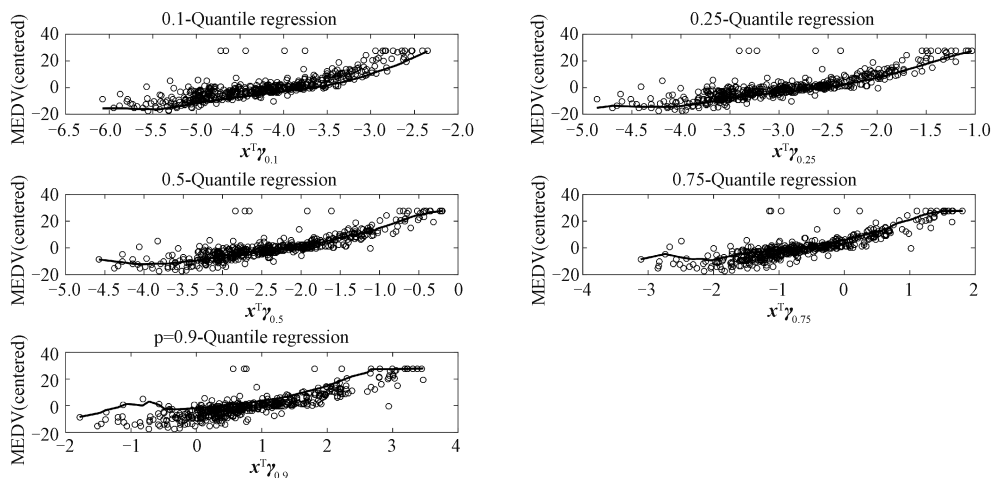


图 1 联系函数  $g(u)$  及在不同的  $\tau$  下,  $\hat{g}_\tau(u)$  的估计

Fig. 1 The link function  $g(u)$  and the estimation of  $\hat{g}_\tau(u)$  under different choices of  $\tau$

5 结论

本文研究单指标分位回归模型估计方法的 MM 算法。相比于内点法,MM 算法极大地缩短了计算时间,提高了运算效率。此外,本文给出单指标分位回归模型在 MM 算法下的参数估计公式,在每次迭代过程中,将协变量与响应变量的观测值直接代入公式,即可得到参数的估计值,避免了每次迭代都要优化目标函数的麻烦。

参考文献

[ 1 ] Zhao X, Wang W, Liu L, et al. A flexible quantile regression model for medical costs with application to Medical Expenditure Panel Survey Study [J]. Statistics in Medicine, 2018, 37(17):2645-2666.

[ 2 ] Petrella L, Raponi V. Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress[J]. Journal of Multivariate Analysis, 2019, 173:70-84.

[ 3 ] Tan X Z, Gan T Y, Chen S, et al. Modeling distributional changes in winter precipitation of Canada using Bayesian spatiotemporal quantile regression subjected to different teleconnections [J]. Climate Dynamics, 2019, 52: 2105-2124.

[ 4 ] Xu B, Lin B. Investigating the differences in CO<sub>2</sub> emissions in the transport sector across Chinese provinces: evidence from a quantile regression model (Article) [J]. Journal of Cleaner Production, 2018, 27(175):109-122.

[ 5 ] 姜成飞. 分位数回归方法综述[J]. 科技信息, 2013, 30(25): 185-240.

[ 6 ] He X, Zhu L X. A lack-of-fit test for quantile regression[J]. Journal of the American Statistical Association, 2003, 98(464):1013-1022.

[ 7 ] Mercedes C A, César S S, Wenceslao G M. A lack-of-fit test for quantile regression models with high-dimensional covariates [J]. Computational Statistics and Data Analysis, 2015, 33(88):128-138.

[ 8 ] 关静. 分位数回归理论及其应用[D]. 天津:天津大学, 2009.

[ 9 ] 朱平芳, 张征宇. 无条件分位数回归: 文献综述与应用实例[J]. 统计研究, 2012, 29(3): 88-96.

[ 10 ] 陈建宝, 丁军军. 分位数回归技术综述[J]. 统计与信息论坛, 2008, 23(3): 89-96.

[ 11 ] Gao J Y. Computation in quantile and composite quantile regression models with or without regularization [D]. Edmonton: University of Alberta, 2015.

[ 12 ] Hunter D R, Lange K. Quantile regression via an MM algorithm [J]. Journal of Computational and Graphical Statistics, 2000, 9(1): 60-77.

[ 13 ] Wu T Z, Yu K M, Yu Y. Single-index quantile regression [J]. Journal of Multivariate Analysis, 2010, 40(101): 1607-1621.

[ 14 ] Kong E F, Xia Y C. A single-index quantile regression model and its estimation[J]. Econometric Theory, 2012, 28(4): 730-768.

[ 15 ] Kuruwita C N. Non-iterative estimation and variable selection in the single-index quantile regression model [J]. Communications in Statistics-Simulation and Computation, 2016, 45(10):3615-3628.

[ 16 ] Christou E. A Non-iterative method for fitting the single index quantile regression model with uncensored and censored data [D]. Pennsylvania: The Pennsylvania State University, 2016.

[ 17 ] Hu Y A, Gramacy R B, Lian H. Bayesian quantile regression for single-index models[J]. Statistics and Computing, 2013, 23(4):437-454.

[ 18 ] Alshaybawee T, Midi H, Alhamzawi R. Bayesian elastic net single index quantile regression [J]. Journal of Applied

- Statistics, 2017, 44(5):853-871.
- [19] Jiang R, Qian W M, Zhou Z G. Weighted composite quantile regression for single-index models[J]. Journal of Multivariate Analysis, 2016, 148: 34-48.
- [20] Liu H L, Yang H. Estimation and variable selection in single-index composite quantile regression[J]. Communications in Statistics-Simulation and Computation, 2017, 46(9):7022-7039.
- [21] Zhong W, Zhu L P, Li R Z, et al. Regularized quantile regression and robust feature screening for single index models[J]. Statistica Sinica, 2016, 26(1):69-95.
- [22] Jiang Y L. Nonparametric quantile regression models via majorization minimization-algorithm[J]. Statistics and Its Interface, 2014, 7(2):235-240.
- [23] Xie S Y, Wan A T K, Zhou Y. Quantile regression methods with varying-coefficient models for censored data[J]. Computational Statistics and Data Analysis, 2015, 88:154-172.
- [24] Chen Y L, Tang M L, Tian M Z. Semiparametric hierarchical composite quantile regression[J]. Communications in statistics: theory and methods, 2015, 50(4-6):996-1012.
- [25] Lin W, Kulasekera K B. Identifiability of single-index models and additive-index models[J]. Biometrika Journal, 2007, 31(94):496-501.
- [26] Chaudhuri P, Doksum K, Samarov A. On average derivative quantile regression[J]. Annals of Statistics, 1997, 25(2):715-744.
- [27] Yu K M, Jones M C. Local linear quantile regression[J]. Journal of the American Statistic Association, 1998, 93(441):228-237.
- [28] Lv Y Z, Zhang R Q, Zhao W H, et al. Quantile regression and variable selection of partial linear single-index model[J]. Ann Inst Stat Math, 2015, 67(2):375-409.
- [29] De Gooijer J G, Zerom D. On additive conditional quantiles with high-dimensional covariates[J]. Journal of the American Statistic Association, 2003, 98(116):135-146.