

精确线性约束条件下众数回归模型的参数估计

王照良, 张天乙, 张航

(河南理工大学 数学与信息科学学院, 河南 焦作 454000)

摘要:研究了众数线性回归模型中回归系数具有精确线性约束条件下的估计问题.提出了拉格朗日乘子的众数估计量,证明了所得估计量的相合性及渐近正态性,并给出了一个求解最大化目标函数的迭代算法.最后通过数值模拟研究了所提方法的良好有限样本性质.

关键词:回归模型;最小二乘估计;众数;精确线性约束;拉格朗日乘子

中图分类号:O212.7

文献标志码:A

文章编号:1000-2367(2025)05-0090-07

设 $\mathbf{X} \in \mathbf{R}^p$ 是解释变量, $Y \in \mathbf{R}$ 是响应变量, 那么线性模型

$$Y = \mathbf{X}^T \boldsymbol{\beta} + \epsilon \quad (1)$$

是诸多回归关系中最基本的一个. 在模型(1)中, $\boldsymbol{\beta}$ 是回归系数, ϵ 是随机误差. 假定 $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ 是来自 $(\mathbf{X}, Y)$ 的一个独立同分布的样本. 通常假定 $E(\epsilon | \mathbf{X}) = 0$, 此时最小二乘估计法广泛应用于估计模型参数 $\boldsymbol{\beta}$, 即寻找 $\boldsymbol{\beta}$ 使得下式 $\sum_{i=1}^n (Y_i - \mathbf{X}_i^T \boldsymbol{\beta})^2$ 达到最小. 为了实现对数据中经常出现的异常值和重尾误差的稳健性, 越来越多的工作考虑分位数回归, 此时假定 $P(\epsilon < 0 | \mathbf{X}) = \tau$, 其中 $\tau \in (0, 1)$ 为指定的分位数水平. 通过最小化以下的目标函数来估计参数 $\boldsymbol{\beta}$, $\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}_i^T \boldsymbol{\beta})$, 其中 $\rho_\tau(t) = t(\tau - I(t < 0))$, $I(\cdot)$ 为示性函数.

当随机误差 ϵ 的分布为对称分布时, 利用均值回归或分位数回归得到的估计量具有优越的统计性质. 但在现实生活中很多数据是有偏的. 如果误差分布为非对称或重尾时, 仍使用均值或分位数回归模型, 所得到的参数估计量会存在较大的偏差且往往不具有良好的统计性质. 相比于均值和分位数, 众数有许多优良的性质, 它捕捉的是最有可能的值. 在许多应用中, 如工资、价格、计量经济学中的支出、生物医学应用中的血液和脑脊液中的生物标志物, 众数是中心趋势的最有效度量. 因此, 众数回归的研究有着十分重要的理论和应用价值.

SAGER 等^[1]首次提出了众数回归的概念, 考虑用众数回归模型来描述条件分布的中心, 引起了诸多学者的研究兴趣. COLLOMB 等^[2]研究了在条件众数下众数估计量的一致性. LEE^[3]提出了线性众数回归模型, 研究了此模型的强相合性. KEMP 等^[4]研究了半参数众数回归模型中估计量的渐近性质, 证明了估计量的相合性. ZHAO 等^[5]考虑了基于众数回归的经验似然估计. YAO 等^[6]提出了基于非参数模型的局部众数回归, 证明了参数的渐近性质. ZHANG 等^[7]又将众数回归的方法运用到部分线性变系数模型上. 以上关于众

收稿日期: 2024-05-28; 修回日期: 2024-07-02.

基金项目: 教育部人文社会科学研究项目(20YJC910010); 河南省测绘科学与技术“双一流”学科创建项目(GC-CYJ202428; GCCYJKT202539); 河南理工大学博士基金(B2020-37).

作者简介(通信作者): 王照良(1983—), 男, 河南开封人, 河南理工大学副教授, 博士, 研究方向为半参数统计建模与高维数据分析, E-mail: wangzhaoliang@hpu.edu.cn.

引用本文: 王照良, 张天乙, 张航. 精确线性约束条件下众数回归模型的参数估计[J]. 河南师范大学学报(自然科学版), 2025, 53(5): 90-96. (Wang Zhaoliang, Zhang Tianyi, Zhang Hang. Estimation for the modal regression models with exact linear constraints[J]. Journal of Henan Normal University(Natural Science Edition), 2025, 53(5): 90-96. DOI:10.16366/j.cnki.1000-2367.2024.05.28.0001.)

数回归模型的研究说明了对非对称、重尾具有很好的稳健性,而且所得估计同最小二乘相比,在正态误差分布下几乎不会损失效率.重要的一点,YAO等^[8]指出当数据有偏时,在相同的置信水平下众数回归可以提供比均值回归和分位数回归更短的预测区间.

实际数据分析中,除了样本信息以外,很多时候还能从其他途径比如前期实验或经验研究中获取一些信息.显然如果将这些信息纳入到模型的估计中,将会提升估计量的精度.这些信息常见的表现形式有等式约束、不等式约束、随机约束等.因此,对于带约束的回归模型的研究,从理论和实践上都有十分重要的意义.国内外已有众多学者对带约束的参数回归模型和非参数回归模型进行了研究.例如,RAO等^[9]研究了带有等式约束的线性回归模型的参数估计,提出了约束最小二乘估计.申维^[10]考虑了带约束的非线性回归模型.许文源等^[11]对带有线性不等式约束的情形进行了研究.OZKALE^[12]研究了带有随机约束的线性回归模型的岭估计问题.

但目前还没有文献研究带有约束条件的众数回归模型参数估计问题.本文将研究具有精确约束的众数回归模型参数估计问题,即对于模型(1),假定 $\text{Mode}(\epsilon | \mathbf{X}) = 0$,且参数 $\boldsymbol{\beta}$ 满足如下精确约束条件

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{d}, \quad (2)$$

其中 $\mathbf{d}$ 是 $k \times 1$ 的已知向量, $\mathbf{H}$ 是 $k \times p$ 的已知行满秩矩阵.本文提出了估计方法,证明了所提估计量的理论性质,并通过随机模拟研究了所提方法的有限样本性能.

1 方法和渐近性质

1.1 方法

在线性模型(1)中,众数回归模型假定 $\text{Mode}(\epsilon | \mathbf{X}) = 0$,即

$$\text{Mode}(Y | \mathbf{X}) = \mathbf{X}^T \boldsymbol{\beta}. \quad (3)$$

如果式(3)成立,YAO等^[8]提出通过最大化基于核的目标函数

$$L_h(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \phi_h(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) \quad (4)$$

来估计参数 $\boldsymbol{\beta}$,其中 $\phi_h(t) = h^{-1} \phi(t/h)$ 且 $\phi(t)$ 是关于原点对称的核密度函数.为了计算方便,本文假设 $\phi(t)$ 是标准正态密度函数,即 $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.基于此核函数的选择,MEM迭代算法有显示形式的结果,参看式(12).需要强调的是本文中的所有渐近结果对其他核依然成立.

记给定 $\mathbf{X}$ 时随机误差 ϵ 的条件密度为 $f_\epsilon(\epsilon | \mathbf{X})$.如果 $f_\epsilon(\epsilon | \mathbf{X})$ 是关于原点对称的,则式(4)中 $\boldsymbol{\beta}$ 的估计与传统的最小二乘回归的系数估计没有区别,因为此时 $\text{Mode}(Y | \mathbf{X}) = E(Y | \mathbf{X})$.但是当 $f_\epsilon(\epsilon | \mathbf{X})$ 不是对称的,则众数回归中的系数估计与均值回归的系数估计存在区别.例如存在一种可能给定 $\mathbf{X}$ 时 Y 的众数是 $\mathbf{X}$ 的线性函数,但传统的均值回归可能是非线性的,参看 YAO等^[8]的例1.众数回归模型具有以下特点:(i)众数回归捕捉的是最有可能的值,即给定 $\mathbf{X}$ 时 Y 的条件分布的众数,当条件分布是非对称的时候,条件众数更有效.(ii)基于(i),众数回归模型可以得到精度更高的预测.(iii)当对于厚尾的条件误差分布时,众数回归比传统回归更稳健,通过调节参数 h 的选取自适应数据.

对于众数线性回归模型,当感兴趣的参数 $\boldsymbol{\beta}$ 满足精确的线性约束条件(2),本文提出通过最大化如下的拉格朗日辅助目标函数

$$Q_h(\boldsymbol{\beta}) = L_h(\boldsymbol{\beta}) + \boldsymbol{\lambda}^T (\mathbf{H}\boldsymbol{\beta} - \mathbf{d}) = \frac{1}{n} \sum_{i=1}^n \phi_h(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) + \boldsymbol{\lambda}^T (\mathbf{H}\boldsymbol{\beta} - \mathbf{d}) \quad (5)$$

来估计参数 $\boldsymbol{\beta}$,其中 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)^T$ 是拉格朗日乘子.记通过最大化式(5)得到参数的估计量为 $\hat{\boldsymbol{\beta}}_R$,即 $\hat{\boldsymbol{\beta}}_R = \arg \max_{\boldsymbol{\beta}} Q_h(\boldsymbol{\beta})$.该最值问题没有显示解,1.3小节给出了一个改进的MEM迭代算法来求解该最值问题.

1.2 理论性质

为了得到参数估计量 $\hat{\boldsymbol{\beta}}_R$ 的大样本性质,首先给出如下正则性条件:

(A1) 对任意的 x , $f_{\epsilon}^{(v)}(t | \mathbf{X} = x)$, $v = 0, 1, 2, 3$ 在原点附近连续, 有 $f'_{\epsilon}(0 | \mathbf{X} = x) = 0$.

(A2) 定义 $\mathbf{J} = E[f''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T]$, $\mathbf{K} = E[f'''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i]$ 和 $\mathbf{L} = E[f_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T]$. 假设 $n^{-1} \sum_{i=1}^n f''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T = \mathbf{J} + o_p(1)$, $n^{-1} \sum_{i=1}^n f'''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i = \mathbf{K} + o_p(1)$ 和 $n^{-1} \sum_{i=1}^n f_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T = \mathbf{L} + o_p(1)$, 其中 $-\mathbf{J}$ 是正定矩阵.

(A3) $n^{-1} \sum_{i=1}^n \|\mathbf{X}_i\|^4 = O_p(1)$.

(A4) 当 $n \rightarrow \infty$ 时, $h \rightarrow 0$ 且 $nh^5 \rightarrow \infty$.

下面定理 1 给出了参数估计量 $\hat{\boldsymbol{\beta}}_R$ 的收敛速度, 说明了 $\hat{\boldsymbol{\beta}}_R$ 的相合性.

定理 1 假设正则性条件(A1)~(A4)成立, 当 $n \rightarrow \infty$ 时, 则 $\|\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\| = O_p\{h^2 + (nh^3)^{-1/2}\}$, 其中 $\boldsymbol{\beta}_0$ 是众数回归函数(1)式中系数的真实值.

证明 令 $a_n = h^2 + (nh^3)^{-1/2}$. 为了证明定理 1, 只需证明对任意给定 $\xi > 0$, 存在一个相当大的常数 c 使得

$$P\left\{\sup_{\|\boldsymbol{\mu}\|=c} Q_h(\boldsymbol{\beta}_0 + a_n \boldsymbol{\mu}) < Q_h(\boldsymbol{\beta}_0)\right\} \geq 1 - \xi, \quad (6)$$

其中 $Q_h(\boldsymbol{\beta})$ 为式(5)中定义的目标函数, $\boldsymbol{\beta}_0$ 是参数的真实值. 式(6)成立意味着在集合 $\{\boldsymbol{\beta}_0 + a_n \boldsymbol{\mu} : \|\boldsymbol{\mu}\| \leq c\}$ 中至少以概率 $1 - \xi$ 存在一个局部最大值点. 这等价于存在一个局部最大值点 $\hat{\boldsymbol{\beta}}_R$ 使得 $\|\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\| = O_p(a_n)$.

为了证明式(6), 首先注意到

$$\phi'_h(t) = -\frac{t}{h^3} \phi\left(\frac{t}{h}\right), \phi''_h(t) = \frac{1}{h^3} \left(\frac{t^2}{h^2} - 1\right) \phi\left(\frac{t}{h}\right) \text{ 和 } \phi'''_h(t) = \frac{1}{h^4} \left\{ \frac{3t}{h} - \left(\frac{t}{h}\right)^3 \right\} \phi\left(\frac{t}{h}\right). \quad (7)$$

结合式(7)并利用 Taylor 展开可得

$$\begin{aligned} Q_h(\boldsymbol{\beta}_0 + a_n \boldsymbol{\mu}) - Q_h(\boldsymbol{\beta}_0) &= a_n \mathbf{K}_n^{*T} \boldsymbol{\mu} + \frac{a_n^2}{2} \boldsymbol{\mu}^T \mathbf{J}_n \boldsymbol{\mu} - \\ &\frac{a_n^3}{6nh^4} \sum_{i=1}^n \phi'''_h\left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}^*}{h}\right) (\mathbf{X}_i^T \boldsymbol{\mu})^3 = M_1 + M_2 + M_3, \end{aligned} \quad (8)$$

其中 $\|\boldsymbol{\mu}\| = c$ 且 $\boldsymbol{\beta}^*$ 满足 $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \leq ca_n$, $\mathbf{K}_n^*$ 和 $\mathbf{J}_n$ 分别为

$$\begin{aligned} \mathbf{K}_n^* &= \frac{\partial Q_h(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} = -\frac{1}{n} \sum_{i=1}^n \phi'_h(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0) \mathbf{X}_i + \mathbf{H}^T \boldsymbol{\lambda} = \mathbf{K}_n + \mathbf{H}^T \boldsymbol{\lambda} = \\ &[\mathbf{I}_p - \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{H}] \mathbf{K}_n = -P_H \mathbf{K}_n \\ \mathbf{J}_n &= \frac{\partial^2 Q_h(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{1}{n} \sum_{i=1}^n \phi''_h(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0) \mathbf{X}_i \mathbf{X}_i^T. \end{aligned}$$

其中 $P_H = \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{H} - \mathbf{I}_p$, $\mathbf{I}_p$ 为 $p \times p$ 单位矩阵. 令 $\mathbf{D} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ 和 $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. 结合条件(A2), 由 Taylor 展开以及 $\phi(t)$ 的对称性, 可得 $\mathbf{J}_n$ 和 $\mathbf{K}_n^*$ 的期望和方差

$$\begin{aligned} E(\mathbf{J}_n | \mathbf{D}) &= \frac{1}{n} \sum_{i=1}^n f''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T \{1 + o_p(1)\} = \mathbf{J} \{1 + o_p(1)\}, \\ \text{Var}(\mathbf{J}_n | \mathbf{D}) &= O_p\{(nh^5)^{-1}\}, \\ E(\mathbf{K}_n^* | \mathbf{D}) &= -P_H \frac{h^2}{2n} \sum_{i=1}^n f'''_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \{1 + o_p(1)\} = -\frac{h^2}{2} P_H \mathbf{K} \{1 + o_p(1)\}, \\ \text{Cov}(\mathbf{K}_n^* | \mathbf{D}) &= \frac{1}{n^2 h^3} \nu_2 \sum_{i=1}^n f_{\epsilon}(0 | \mathbf{X}_i) \mathbf{X}_i \mathbf{X}_i^T \{1 + o(1)\} = \frac{1}{nh^3} \nu_2 P_H \mathbf{L} P_H \{1 + o_p(1)\}, \end{aligned} \quad (9)$$

其中 $\nu_2 = \int t^2 \phi^2(t) dt$, 矩阵 $\mathbf{J}$, $\mathbf{K}$ 和 $\mathbf{L}$ 由条件(A2)给出. 式(9)中计算矩阵方差时, 计算的是矩阵每个元素的方差. 利用对任意随机变量 X , 有 $X = E(X) + O_p(\{\text{Var}(X)\}^{1/2})$ 成立这一结果, 结合条件(A4)与式(9), 显然可得 $\mathbf{J}_n = \mathbf{J} + o_p(1)$. 同时 $\mathbf{K}_n = O_p(a_n)$, 因此 $M_1 = O_p(a_n^2)$. 注意到 $M_2 = 0.5 a_n^2 \boldsymbol{\mu}^T \mathbf{J} \boldsymbol{\mu} \{1 + o_p(1)\}$. 因为

$\phi^{(4)}(t)$ 有界且 $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \leq ca_n$, 所以 $\phi'''(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}^*}{h}) = \phi'''(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0}{h})(1 + o_p(1))$. 注意到 $\phi'''(t) = (3t - t^3)\phi(t)$, 由 Taylor 展开和 $\phi(t)$ 的对称性, 可得:

$$E\{\phi'''(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0}{h}) \mid \mathbf{D}\} = O_p(h^4) \text{ 和 } \text{Var}\{\phi'''(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0}{h}) \mid \mathbf{D}\} = O_p(h).$$

因为 $nh^5 \rightarrow \infty$, 可得 $M_3 = o_p(a_n^2)$. 对于任意的 $\xi > 0$, 只要选择足够大的 c , 就可使式(8)中的 M_2 以不小于 $1 - \xi$ 的概率控制 M_1 和 M_3 . 因为 $-\mathbf{J}$ 为正定矩阵, 所以通过选取足够大的 c 可以保证 $\mathbf{Q}_h(\boldsymbol{\beta}_0 + a_n \boldsymbol{\mu}) - \mathbf{Q}_h(\boldsymbol{\beta}_0) < 0$ 以不小于 $1 - \xi$ 的概率成立. 因此式(6)成立, 从而定理 1 得证.

下面定理 2 给出了估计量 $\hat{\boldsymbol{\beta}}_R$ 的渐近正态性.

定理 2 在与定理 1 相同的假设下, 估计量 $\hat{\boldsymbol{\beta}}_R$ 有如下的渐近正态性

$$\sqrt{nh^3}[\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 - \frac{h^2}{2}\mathbf{J}^{-1}P_H\mathbf{K}\{1 + o_p(1)\}] \xrightarrow{D} N(0, \nu_2\mathbf{J}^{-1}P_H\mathbf{L}P_H\mathbf{J}^{-1}),$$

其中“ $\xrightarrow{D}$ ”表示依分布收敛, $\nu_2 = \int t^2 \phi^2(t) dt$, $P_H = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H} - \mathbf{I}_p$, 矩阵 $\mathbf{J}$, $\mathbf{K}$ 和 $\mathbf{L}$ 由条件(A2)给出.

证明 根据 $\hat{\boldsymbol{\beta}}_R$ 的定义, 显然 $\hat{\boldsymbol{\beta}}_R$ 满足一阶条件 $\partial \mathbf{Q}_h(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = 0$. 由 Taylor 展开可得:

$$0 = \frac{\partial \mathbf{Q}_h(\hat{\boldsymbol{\beta}}_R)}{\partial \boldsymbol{\beta}} = \mathbf{K}_n^* + (\mathbf{J}_n + \mathbf{R}_n)(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) = -P_H\mathbf{K}_n + (\mathbf{J}_n + \mathbf{R}_n)(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0), \quad (10)$$

其中 $\mathbf{R}_n = -\frac{1}{2nh^4} \sum_{i=1}^n [\phi'''(\frac{Y_i - \boldsymbol{\beta}^{*T} \mathbf{X}_i}{h}) \{(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)^T \mathbf{X}_i\} \mathbf{X}_i \mathbf{X}_i^T]$, 且 $\boldsymbol{\beta}^*$ 满足 $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \leq \|\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\|$. 利用式(10)可得 $\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 = -(\mathbf{J}_n + \mathbf{R}_n)^{-1} \mathbf{K}_n^*$. 利用定理 1, 有 $\|\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\| = O_p(a_n)$, 其中 $a_n = (nh^3)^{-1/2} + h^2$, 类似于式(8)中 M_3 的证明, 可得 $\mathbf{R}_n = o_p(1)$. 因此基于(9)式, 有

$$\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 = \mathbf{J}^{-1}P_H\mathbf{K}_n\{1 + o_p(1)\}. \quad (11)$$

下面证明 $\sqrt{nh^3} \mathbf{K}_n$ 的渐近正态性. 只需证明对任意的单位向量 $\boldsymbol{\eta} \in \mathbf{R}^p$,

$$\{\boldsymbol{\eta}^T \text{Cov}(\sqrt{nh^3} \mathbf{K}_n) \boldsymbol{\eta}\}^{-1/2} \{\sqrt{nh^3} \boldsymbol{\eta}^T \mathbf{K}_n - \boldsymbol{\eta}^T E(\sqrt{nh^3} \mathbf{K}_n)\} \xrightarrow{D} N(0, 1).$$

令 $U_i = -\frac{1}{\sqrt{nh}} \phi'(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0}{h}) \boldsymbol{\eta}^T \mathbf{X}_i$, 则 $\boldsymbol{\eta}^T \sqrt{nh^3} \mathbf{K}_n = \sum_{i=1}^n U_i$. 下面验证 Lyapunov 条件. 基于式(9), 有 $\text{Cov}(\mathbf{K}_n) =$

$$\frac{1}{nh^3} \nu_2 \mathbf{L} \{1 + o_p(1)\}, \text{ 则}$$

$$\text{Var}(\boldsymbol{\eta}^T \sqrt{nh^3} \mathbf{K}_n) = nh^3 \boldsymbol{\eta}^T \text{Cov}(\mathbf{K}_n) \boldsymbol{\eta} = \nu_2 \boldsymbol{\eta}^T \mathbf{L} \boldsymbol{\eta} + o(1).$$

因此, 只要证明 $nE|U_i|^3 \rightarrow 0$. 注意到 $(\boldsymbol{\eta}^T \mathbf{X}_i)^2 \leq \|\mathbf{X}_i\|^2 \|\boldsymbol{\eta}\|^2 = \|\mathbf{X}_i\|^2$, 利用 $\phi'(\cdot)$ 是有界的, 则 $nE|U_i|^3 \leq O\{(nh^3)^{-1/2}\} \rightarrow 0$. 因此 Lyapunov 条件成立, $\sqrt{nh^3} \mathbf{K}_n$ 满足渐近正态性, 即

$$\sqrt{nh^3} \{\mathbf{K}_n - \frac{h^2}{2} \mathbf{K} \{1 + o_p(1)\}\} \xrightarrow{D} N(0, \nu_2 \mathbf{L}).$$

利用 Slutsky 定理, 结合式(11)可得 $\sqrt{nh^3}[\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 - \frac{h^2}{2} \mathbf{J}^{-1}P_H\mathbf{K}\{1 + o_p(1)\}] \xrightarrow{D} N(0, \nu_2 \mathbf{J}^{-1}P_H\mathbf{L}P_H\mathbf{J}^{-1})$.

定理 2 得证.

定理 2 显示估计量 $\hat{\boldsymbol{\beta}}_R$ 有渐近方差 $\nu_2 \mathbf{J}^{-1}P_H\mathbf{L}P_H\mathbf{J}^{-1}/(nh^3)$, 与 YAO 等^[8]中定理 2.3 相比, 因为 $\mathbf{J}^{-1}P_H\mathbf{L}P_H\mathbf{J}^{-1} \leq \mathbf{J}^{-1}\mathbf{L}\mathbf{J}^{-1}$, 所以考虑了线性约束的估计量 $\hat{\boldsymbol{\beta}}_R$ 更加有效.

1.3 算法

由于最大化式(5)无显示解, 基于 YAO 等^[8]提出了 MEM(modal expectation maximization)算法, 结合拉格朗日乘子法, 本文提出了如下的修正版本的 MEM 算法来获得式(5)的近似解. 具体的算法为:

步骤 0 利用最小二乘估计得到初始值 $\hat{\boldsymbol{\beta}}_R^{(0)}$, 并令 $k = 0$.

步骤 1 E 步:更新权重 $\pi(j \mid \hat{\boldsymbol{\beta}}_R^{(k)})$

$$\pi(j \mid \hat{\boldsymbol{\beta}}_R^{(k)}) = \frac{\phi_h(Y_j - \mathbf{X}_j^T \hat{\boldsymbol{\beta}}_R^{(k)})}{\sum_{i=1}^n \phi_h(Y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_R^{(k)})} \propto \phi_h(Y_j - \mathbf{X}_j^T \hat{\boldsymbol{\beta}}_R^{(k)}), j = 1, \dots, n.$$

步骤 2 M 步:迭代更新 $\hat{\boldsymbol{\beta}}_R^{(k+1)}$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_R^{(k+1)} = \arg \max_{\boldsymbol{\beta}} \sum_{j=1}^n \pi(j \mid \hat{\boldsymbol{\beta}}_R^{(k)}) \ln \phi_h(Y_j - \mathbf{X}_j^T \boldsymbol{\beta}) + \boldsymbol{\lambda}^T (\mathbf{H} \boldsymbol{\beta} - \mathbf{d}) = \\ (\mathbf{D}^T \mathbf{W}_k \mathbf{D})^{-1} (\mathbf{D}^T \mathbf{W}_k \mathbf{Y} - \frac{1}{2} \mathbf{H}^T \boldsymbol{\lambda}), \end{aligned} \tag{12}$$

其中 $\mathbf{D} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T, \mathbf{Y} = (Y_1, \dots, Y_n)^T, \mathbf{W}_k$ 是 $n \times n$ 的对角矩阵,其第 j 个对角线元素为 $\pi(j \mid \hat{\boldsymbol{\beta}}_R^{(k)})$.在式 (12) 中,拉格朗日乘子 $\boldsymbol{\lambda}$ 为 $\boldsymbol{\lambda} = 2(\mathbf{H}(\mathbf{D}^T \mathbf{W}_k \mathbf{D})^{-1} \mathbf{H}^T)^{-1}(\mathbf{H} \hat{\boldsymbol{\beta}}^{(k+1)} - \mathbf{d})$,其中 $\hat{\boldsymbol{\beta}}^{(k+1)} = (\mathbf{D}^T \mathbf{W}_k \mathbf{D})^{-1} \mathbf{D}^T \mathbf{W}_k \mathbf{Y}$

$$\arg \max_{\boldsymbol{\beta}} \sum_{j=1}^n \pi(j \mid \hat{\boldsymbol{\beta}}_R^{(k)}) \ln \phi_h(Y_j - \mathbf{X}_j^T \boldsymbol{\beta}).$$

步骤 3 重复迭代 E 步和 M 步,直至算法收敛为止,记所得的估计为 $\hat{\boldsymbol{\beta}}_R$.

2 数值模拟

本节通过 Monte Carlo 随机模拟来研究第 1 节所提出的受约束的众数回归估计量(记作 MODL-R)的有限样本性质.将所提方法与不考虑约束条件的众数回归(记作 MODL),均值回归的最小二乘估计(记作 LSE),均值回归带约束的最小二乘估计(记作 LSE-R)3 种估计方法进行比较.

为实施模拟,考虑如下的线性回归模型

$$Y = 1 + 3X_1 + 2X_2 + 3X_3 + \sigma(\mathbf{X})\epsilon,$$

其中 $X_j \sim U(0,1), j = 1,2,3, \sigma(\mathbf{X}) = 1 + 2X_1$.关于模型误差,考虑两种情形:(i) $\epsilon \sim N(0,1)$ 和(ii) $\epsilon \sim 0.5N(-1,2.5^2) + 0.5N(1,0.5^2)$.不难看出,对于情形(i)有 $E(\epsilon) = \text{Mode}(\epsilon) = 0$,所以此时 $E(Y \mid \mathbf{X}) = \text{Mode}(Y \mid \mathbf{X}) = 1 + 3X_1 + 2X_2 + 3X_3$,参数 $\boldsymbol{\beta}$ 满足线性约束 $\mathbf{H}_1 \boldsymbol{\beta} = 0$,其中矩阵 $\mathbf{H}_1$ 为

$$\mathbf{H}_1 = \begin{pmatrix} 3 & -1 & 0 & 0 \\ 0 & 0 & 1.5 & -1 \end{pmatrix}.$$

对于情形(ii), $E(\epsilon) = 0$ 和 $\text{Mode}(\epsilon) = 1$,所以此时 $E(Y \mid \mathbf{X}) = 1 + 3X_1 + 2X_2 + 3X_3$ 不等于 $\text{Mode}(Y \mid \mathbf{X}) = 2 + 5X_1 + 2X_2 + 3X_3$.条件均值时参数 $\boldsymbol{\beta}$ 满足线性约束 $\mathbf{H}_1 \boldsymbol{\beta} = 0$,而条件众数时参数 $\boldsymbol{\beta}$ 满足线性约束 $\mathbf{H}_2 \boldsymbol{\beta} = 0$,其中矩阵 $\mathbf{H}_2$ 为

$$\mathbf{H}_2 = \begin{pmatrix} 2.5 & -1 & 0 & 0 \\ 0 & 0 & 1.5 & -1 \end{pmatrix}.$$

在计算众数回归时,需要选择带宽 h ,采用 YAO 等^[8]的方法来选择带宽 h .为了评估估计量的有限样本性能,分别在样本容量 $n = 100, 200$ 和 500 下通过 1 000 次重复模拟计算偏差(Bias),标准差(SD)和均方误差(MSE).模拟旨在研究不同样本容量 n 设置下的各种估计方法的表现.附录表 S1 和表 S2 分别展示了当 $\epsilon \sim N(0,1)$ 和 $\epsilon \sim 0.5N(-1,2.5^2) + 0.5N(1,0.5^2)$ 时的模拟结果.

从表 S1 可以看出,LSE,LSE-R,MODL 和 MODL-R 很好地估计了它们的目标参数,考虑线性约束时的估计均优于忽略约束时的估计.随着样本量 n 的增加,各方法的参数估计均方误差逐渐减小,说明估计量的相合性.在误差分布为对称的正态分布时,众数回归表现不如最小二乘,但总体来看,效率损失不大.

从表 S2 可以看出,当 ϵ 服从混合正态的情况下,最小二乘估计不能正确发现 Y 与 $\mathbf{X}$ 的众数关系,因为情形(ii)的随机误差是非对称的,且均值与众数不同.与表 S1 一样,考虑线性约束时的估计均优于忽略约束时的估计.但最重要的一点,与本例中的其他方法相比,MODL-R 具有更小的标准差,特别是当 $n = 200$ 或 $n = 500$ 时.因此,对于有限样本,MODL-R 不仅具有良好的众数解释,而且在误差非对称时,比最小二乘方法具有更好的估计精度.

为了对估计量的渐近正态性有一个直观的认识,图 1 给出了当 $n=500$ 时,MODL-R 方法 1 000 次计算的参数估计量 $\hat{\beta}_R$ 的直方图,同时给出的还有依据这些估计值计算所得的密度函数非参数核估计曲线(红色实线)和正态分布参数估计曲线(蓝色虚线).从图 1 可知,参数估计量 $\hat{\beta}_R$ 与正态分布接近程度较好.以上模拟结果验证了所提出方法的理论结果,并说明了该方法具有良好的有限样本实施效果.

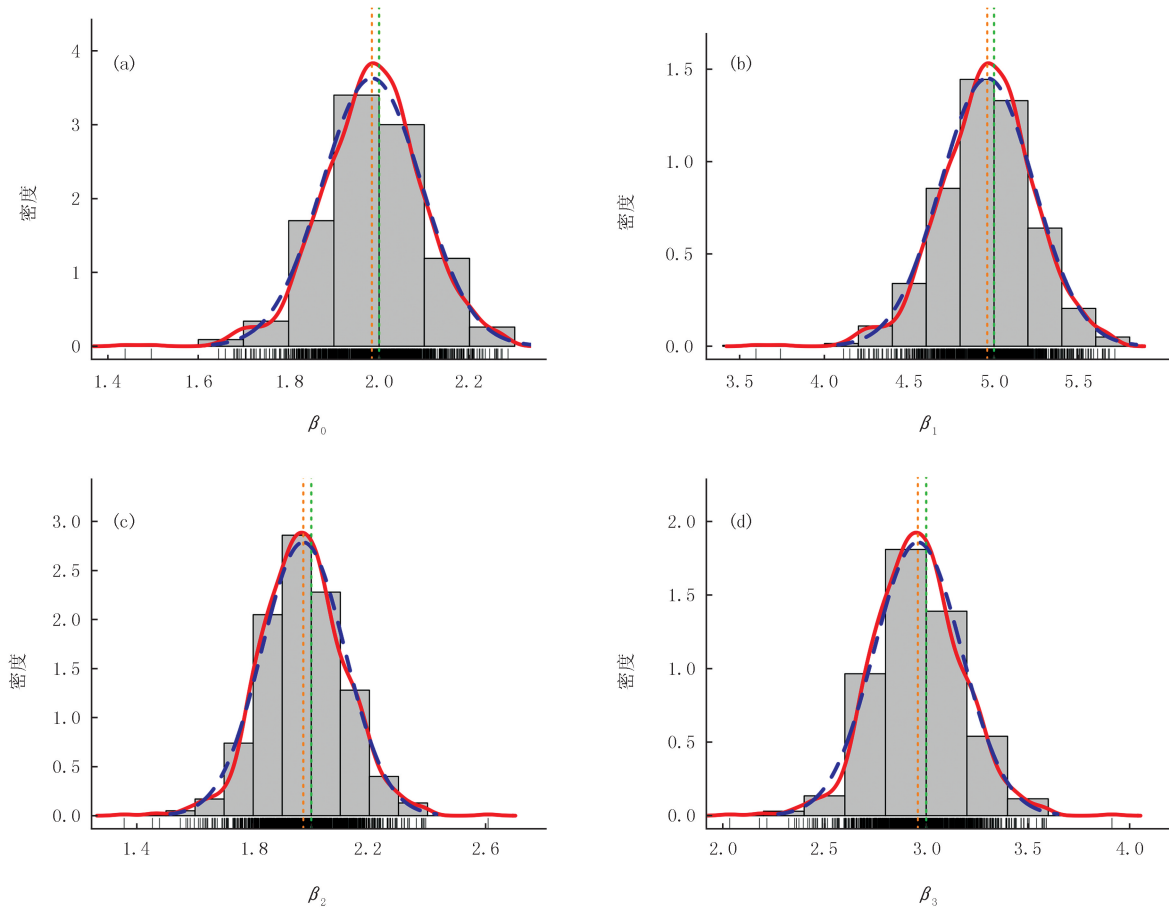


图1 当 $n=500$ 时,基于1 000次模拟计算估计量 $\hat{\beta}_R$ 的直方图及密度估计曲线
Fig.1 Histogram and density estimation curve of estimator $\hat{\beta}_R$ based on 1 000 simulations when $n=500$

3 结 论

本文考虑了精确线性约束条件(2)下众数线性回归模型(1)的参数估计问题.结合拉格朗日乘子法,提出了受约束的众数估计方法,并给出了优化问题的一个迭代算法.在一定的正则性条件下,证明了所提出的参数估计量的理论性质.此外,通过 Monte Carlo 模拟研究了所提出方法的有限样本性能.理论结果和模拟结果均表明所提出的估计量优于其他类型的估计量.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.05.28.0001).

参 考 文 献

[1] SAGER T W, THISTED R A. Maximum likelihood estimation of isotonic modal regression[J]. The Annals of Statistics, 1982, 10(3): 690-707.
[2] COLLOMB G, HARDLE W, HASSANI S. A note on prediction via estimation of the conditional mode function[J]. Journal of Statistical Planning and Inference, 1986, 15(2): 227-236.
[3] LEE M J. Mode regression[J]. Econometrics, 1989, 42(3): 337-349.

[4] KEMP G C R,SILVA J M C S.Regression towards the mode[J].Journal of Econometrics,2012,170(1):92-101.

[5] ZHAO W,ZHANG R,LIU Y,LIU J.Empirical likelihood based modal regression[J].Statistical Papers,2015,5(2):411-430.

[6] YAO W,LINDSAY B G,LI R.Local modal regression[J].Journal of Nonparametric Statistics,2012,24(3):647-663.

[7] ZHANG R,ZHAO W,LIU J.Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression[J].Journal of Multivariate Analysis,2013,122:226-238.

[8] YAO W,LI L.A new regression model:modal linear regression[J].Scandinavian Journal of Statistics,2014,41(3):656-671.

[9] RAO C R,TOUTENBURG H,SHALABH,HEUMANN C.Linear models and generalization:least squares and alternatives[M].New York:Springer-Verlag,2008.

[10] 申维.带约束的非线性回归诊断混合模型分析[J].武汉工业大学学报,1995,17(3):147-152.
SHEN W.Restriction of nonlinear regression diagnostic models with case-weights and mean-shift simultaneously analysis[J].Journal of Wuhan University of Technology,1995,17(3):147-152.

[11] 许文渊,王东谦.带有线性不等式约束的最小二乘[J].系统科学与数学,1984,(4):55-62.
XU W Y,WANG D Q.Least square estimation with linear inequality constraints[J].Journal of Systems Science and Mathematical Sciences,1984,4(1):55-62.

[12] OZKALE M R.A stochastic restricted ridge regression estimator[J].Journal of Multivariate Analysis,2009,100(8):1706-1716.

Estimation for the modal regression models with exact linear constraints

Wang Zhaoliang, Zhang Tianyi, Zhang Hang

(School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: This article considers the estimation problem of regression coefficients in modal linear regression models with exact linear constraints. A mode estimator with Lagrange multipliers is proposed, and the consistency and asymptotic normality of the obtained estimator are established. Furthermore, an iterative algorithm is proposed to solve the maximization objective function. Simulation studies are carried out to illustrate the finite sample performance of the proposed method.

Keywords: regression model; least squares estimation; mode; exact linear constraints; Lagrange multiplier

[责任编辑 陈留院 杨浦]

附 录

表 S1 当 $\varepsilon \sim N(0,1)$ 时的模拟结果
Tab. S1 Simulation results when $\varepsilon \sim N(0,1)$

方法	参数真值	$n=100$			$n=200$			$n=500$		
		Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
LSE	$\beta_0=1$	0.024 6	0.614 5	0.378 2	-0.012 1	0.409 5	0.167 8	-0.004 2	0.266 4	0.071 0
	$\beta_1=3$	-0.030 2	0.751 3	0.565 4	0.005 6	0.532 7	0.283 8	-0.005 5	0.328 5	0.107 9
	$\beta_2=2$	-0.001 1	0.717 3	0.514 5	-0.002 3	0.514 2	0.264 4	0.000 0	0.314 0	0.098 6
	$\beta_3=3$	-0.027 5	0.735 2	0.541 3	0.018 2	0.512 6	0.263 1	0.004 1	0.330 0	0.108 9
LSE-R	$\beta_0=1$	-0.001 5	0.181 5	0.033 0	-0.002 2	0.126 9	0.016 1	-0.002 4	0.078 4	0.006 2
	$\beta_1=3$	-0.004 5	0.544 6	0.296 6	-0.006 5	0.380 8	0.145 0	-0.007 1	0.235 2	0.055 4
	$\beta_2=2$	-0.002 8	0.302 7	0.091 7	0.004 1	0.214 1	0.045 9	0.000 9	0.134 3	0.018 0
	$\beta_3=3$	-0.004 2	0.454 1	0.206 2	0.006 1	0.321 2	0.103 2	0.001 3	0.201 4	0.040 6
MODL	$\beta_0=1$	0.051 7	0.854 5	0.732 8	-0.013 3	0.650 6	0.423 4	0.000 9	0.497 1	0.247 1
	$\beta_1=3$	-0.013 6	1.247 8	1.557 3	0.017 6	1.053 0	1.109 1	-0.013 6	0.845 1	0.714 3
	$\beta_2=2$	-0.027 0	1.046 1	1.095 0	0.017 7	0.820 4	0.673 4	-0.003 3	0.613 2	0.376 0
	$\beta_3=3$	-0.018 9	1.066 1	1.136 9	-0.002 1	0.806 8	0.650 9	0.005 9	0.619 4	0.383 7
MODL-R	$\beta_0=1$	0.012 4	0.306 3	0.094 0	-0.006 6	0.263 8	0.069 6	-0.006 4	0.2016	0.040 7
	$\beta_1=3$	0.037 2	0.919 0	0.846 0	-0.019 9	0.791 5	0.626 8	-0.019 3	0.604 9	0.366 3
	$\beta_2=2$	-0.004 5	0.443 4	0.196 6	0.001 3	0.344 6	0.118 8	0.008 2	0.253 8	0.064 5
	$\beta_3=3$	-0.006 7	0.665 1	0.442 4	0.001 9	0.516 9	0.267 2	0.012 2	0.380 7	0.145 1

表 S2 当 $\varepsilon \sim 0.5N(-1,2.5^2) + 0.5N(1,0.5^2)$ 时的模拟结果
Tab. S2 Simulation results when $\varepsilon \sim 0.5N(-1,2.5^2) + 0.5N(1,0.5^2)$

方法	参数真值	$n=100$			$n=200$			$n=500$		
		Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
LSE	$\beta_0=1$	0.098 0	1.293 4	1.682 4	0.015 2	0.934 9	0.167 8	-0.004 2	0.266 4	0.071 0
	$\beta_1=3$	-0.082 2	1.553 5	2.420 1	-0.007 8	1.122 1	0.283 8	-0.005 5	0.328 5	0.107 9
	$\beta_2=2$	-0.056 4	1.502 8	2.261 6	-0.008 5	-0.005 9	0.264 4	0.000 0	0.314 0	0.098 6
	$\beta_3=3$	-0.057 0	1.592 8	2.540 3	0.013 5	-0.008 8	0.263 1	0.004 1	0.330 0	0.108 9
LSE-R	$\beta_0=1$	0.000 2	0.372 0	0.138 4	0.002 5	0.126 9	0.016 1	-0.002 4	0.078 4	0.006 2
	$\beta_1=3$	0.000 6	1.116 0	1.245 4	0.007 4	0.380 8	0.145 0	-0.007 1	0.235 2	0.055 4
	$\beta_2=2$	-0.005 9	0.625 0	0.390 6	-0.005 9	0.214 1	0.045 9	0.000 9	0.134 3	0.018 0
	$\beta_3=3$	-0.008 8	0.937 4	0.878 9	-0.008 8	0.321 2	0.103 2	0.001 3	0.201 4	0.040 6
MODL	$\beta_0=2$	-0.259 1	0.667 1	0.512 2	-0.159 9	0.650 6	0.423 4	0.000 9	0.497 1	0.247 1
	$\beta_1=5$	0.161 7	0.798 3	0.663 4	0.057 4	1.053 0	1.109 1	-0.013 6	0.845 1	0.714 3
	$\beta_2=2$	-0.028 8	0.780 3	0.609 7	0.010 3	0.820 4	0.673 4	-0.003 3	0.613 2	0.376 0
	$\beta_3=3$	-0.016 9	0.758 3	0.575 2	0.020 3	0.806 8	0.650 9	0.005 9	0.619 4	0.383 7
MODL-R	$\beta_0=2$	-0.020 5	0.222 1	0.049 8	-0.021 3	0.263 8	0.069 6	-0.006 4	0.201 6	0.040 7
	$\beta_1=5$	-0.051 2	0.555 3	0.311 0	-0.053 3	0.791 5	0.626 8	-0.019 3	0.604 9	0.366 3
	$\beta_2=2$	-0.106 7	0.320 4	0.114 0	-0.046 7	0.344 6	0.118 8	0.008 2	0.253 8	0.064 5
	$\beta_3=3$	-0.160 0	0.480 6	0.256 6	-0.070 1	0.516 9	0.267 2	0.012 2	0.380 7	0.145 1