

变系数模型的局部加权组合分位数估计 *

解其昌

吕秀梅

(山东工商学院经济学院, 烟台, 264005)

(重庆工商大学财政金融学院, 重庆, 400067)

摘要

变系数模型是经典线性模型的推广, 它以更加灵活的形式来模拟变量间的非线性关系. 采用局部加权组合分位数方法来估计模型的系数函数. 推导出估计量的局部Bahadur表示以及渐近正态性. 构造一个二次规划, 给出了最优权重的选择方法. 对于非正态误差分布, 理论分析和数值模拟表明局部加权组合分位数比局部最小二乘估计有更高的效率; 而对于正态误差分布, 局部加权组合分位数与局部最小二乘估计有着几乎同样的效率. 通过Monte Carlo模拟和实证分析, 检验估计量的有限样本性质, 结果与理论相一致.

关键词: 变系数模型, 渐近正态性, 渐近相对效率, 局部加权组合分位数估计.

学科分类号: O212.2.

§1. 引言

随着现代数据分析技术的不断发展以及计算能力的快速提升, 传统的参数模型已经不能完全满足人们的需求. 如何开拓变量之间的隐式结构, 成为科学的新主题. 在这种背景下, 推广线性模型和增加它们的灵活应用成为必然趋势. 这期间许多非线性模型和它们变换的模型已经快速发展起来了. 其中, 最受关注的模型之一就是变系数模型(Hastie和Tibshirani, 1993)

$$Y = \alpha_0(U) + \mathbf{X}^\top \boldsymbol{\alpha}(U) + \varepsilon, \quad (1.1)$$

这里 \top 是转置运算, Y 是反应变量, $\alpha_0(U)$ 与 $\boldsymbol{\alpha}(U) = (\alpha_1(U), \dots, \alpha_p(U))^\top$ 是关于协变量 U 的未知光滑系数函数, $\mathbf{X} = (X_1, \dots, X_p)^\top$ 是协变量, p 是大于等于1的正整数, 及 ε 是随机误差. 从方程(1.1)中可以看出变系数模型允许系数依赖于某些协变量.

过去几十年里, 变系数模型已经在理论及应用中得到了迅速发展. 在理论方面, 局部多项式和光滑样条常用来估计模型(1.1). 当系数函数有不同光滑度时, Fan和Zhang(1999)提出了一个两步局部多项式法来估计变系数模型. Cai等(2000)发展了局部多项式技术并且建立了变系数模型估计的渐近理论. 基于样条近似, Huang等(2004)给出了系数函数估计的收敛率. Wang等(2008)进一步讨论了样条近似变系数模型方法并把它应用于变量选择

*教育部基金(109140)和山东工商学院博士启动基金(521014306203)资助.

本文2013年12月30日收到, 2014年6月24日收到修改稿.

doi: 10.3969/j.issn.1001-4268.2014.06.007

上. 关于这两种方法的更多详细介绍可参考Fan和Zhang (2008)以及其中的文献. 在应用方面, 变系数模型已经被广泛用于金融计量、生物医药、政治科学和环境工程等方面. 例如, Hong和Lee (2004)考虑变系数模型在金融计量方面的应用; Cederman和Rao (2001)用变系数模型来研究国际冲突; 以及Cai和Tiwari (2000)把变系数模型用于环境科学问题的检验. 事实上, 不像其它非参数模型会遭到“维度祸根”的影响, 变系数模型引起了人们的极大兴趣.

上述文献中所提到的估计方法几乎都是基于最小二乘法(最小化残差平方和或加权残差平方和). 虽然该方法容易操作且由它所得到的估计服从渐近正态分布, 但是由于最小二乘回归的不稳健性使得这些估计对一些异常值点(outlying points)非常敏感. 同时, 当误差项远离正态分布时, 最小二乘估计的效率会明显降低. 一个有效克服上述问题的备选方法就是使用分位数回归(Koenker和Bassett, 1978). 近来, 已经有几位作者采用分位数估计方法来研究变系数模型(1.1). 使用局部多项式近似技术, Honda (2004)以及Cai和Xu (2008)分别研究了独立数据和时间序列数据的变系数模型分位数估计. Kim (2007)应用样条来近似系数函数并且得到了分位数估计变系数模型的收敛率. 尽管这些文献提供了使用分位数方法研究变系数模型的基本理论, 然而它们只考虑单个特定分位点上的模型估计并且假设随机误差在这个特殊分位点的逆累积分布值为零. 受这个思想的启发, 我们检验跨越不同分位点的变系数模型(1.1)的估计并且释放随机误差在分位点逆累积分布值为零这一强假设.

本文考虑变系数模型的局部加权组合分位数估计. 我们使用局部多项式来近似每一个系数函数. 通过最小化一个加权组合分位数处罚函数, 得到了变系数模型的稳健估计, 并且给出了估计量的局部Bahadur表示以及渐近正态性. 此外, 对局部加权组合分位数与局部最小二乘估计的渐近相对效率(ARE)进行了评估. 结果显示ARE值严格依赖于误差分布. 特别, 当误差为标准正态分布时, 局部加权组合分位数几乎与局部最小二乘估计有着同样的效率. 然而, 对于非正态误差分布, 局部加权组合分位数比局部最小二乘估计具有更高的效率. 采用Monte Carlo模拟来演示估计量的性质, 模拟结果表明局部加权组合分位数估计比局部最小二乘更稳健. 此外, 实证研究的结论也与理论相一致.

§2. 局部加权组合分位数估计

假设 $\{U_i, \mathbf{X}_i, Y_i\}_{i=1}^n$ 是来自于模型(1.1)的一系列随机样本, 这里 $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$. 通过Taylor展式, 系数函数可以被局部近似为

$$\alpha_j(U) \approx \alpha_j(u) + \alpha'_j(u)(U - u), \quad j = 0, 1, \dots, p, \quad (2.1)$$

其中 U 是 u 临域内的任意一点. 对 $j = 0, 1, \dots, p$, 定义 $a_j = \alpha_j(u)$ 及 $b_j = \alpha'_j(u)$.

在最小二乘(Ls)框架下, 把 a_j 和 b_j 的估计分别记为 \hat{a}_j 和 \hat{b}_j , 则 $\{(\hat{a}_j, \hat{b}_j)\}_{j=0}^p$ 最小化

$$\sum_{i=1}^n \left[Y_i - a_0 - b_0(U_i - u) - \sum_{j=1}^p \{a_j - b_j(U_i - u)\} X_{ij} \right]^2 K_h(U_i - u), \quad (2.2)$$

这里 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $h > 0$ 是窗宽以及 $K(\cdot)$ 是一个核函数.

方程(2.2)所给出的局部最小二乘回归拥有很多良好性质. 例如, 漐近极大极小意义上的高统计效率(Fan和Gijbels, 1992)及自动修正边界影响等(Fan, 1993). 然而, 当误差分布没有有限的二阶矩时, 局部最小二乘估计将不再一致(Fan和Gijbels, 1996). 虽然Honda (2004)以及Cai和Xu (2008)使用局部分位数代替局部最小二乘回归来研究变系数模型并且证明了局部分位数比局部最小二乘估计更稳健, 但是局部分位数回归只能反映估计量在某一独立特定分位点的性质. 为了进一步挖掘局部分位数估计的性质, 我们考虑跨越不同分位点的局部加权分位数回归.

对于给定的正整数 M , 令 $\tau_k = k/(M+1)$ 为第 k 个分位点, $k = 1, \dots, M$. 那么, 变系数模型(1.1)的第 τ_k 个分位数方程为

$$Q_{\tau_k}(U, \mathbf{X}) = c_{\tau_k} + \alpha_0(U) + \mathbf{X}^\top \boldsymbol{\alpha}(U), \quad (2.3)$$

其中 $c_{\tau_k} = F_\varepsilon^{-1}(\tau_k)$ 以及 $F_\varepsilon(\cdot)$ 是 ε 的分布函数.

采用式(2.1)的近似方法, 局部加权组合分位数(Wcq)估计为

$$\begin{aligned} (\hat{\mathbf{a}}_0^{*\top}, \hat{b}_0, \hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top)^\top &= \arg \min_{\mathbf{a}_0^*, b_0, \mathbf{a}, \mathbf{b}} \sum_{k=1}^M w_k \left\{ \sum_{i=1}^n \rho_{\tau_k} (Y_i - a_{0k} - b_0(U_i - u) \right. \\ &\quad \left. - \mathbf{X}_i^\top \mathbf{a} - \mathbf{X}_i^\top (U_i - u) \mathbf{b}) K_h(U_i - u) \right\}, \end{aligned} \quad (2.4)$$

其中 $\rho_{\tau_k}(v) = \tau_k v - v I(v \leq 0)$, $I(\cdot)$ 是示性函数, $\mathbf{a}_0^* = (a_{01}, \dots, a_{0M})^\top$, $a_{0k} = a_0 + c_{\tau_k}$, $\mathbf{a} = (a_1, \dots, a_p)^\top$, $\mathbf{b} = (b_1, \dots, b_p)^\top$ 以及 $w_k \geq 0$ 是一系列权满足 $\sum_{k=1}^M w_k = 1$.

从表达式(2.4)可以看出局部加权组合分位数回归是来自于不同分位点的最小化组合目标 ρ -函数, 它对不同分位点的目标函数进行加权平均. 事实上, Zou和Yuan (2008)考虑了等度权($\{w_k = 1/M\}_{k=1}^M$)时的线性模型组合分位数估计并且给出了估计量的漐近分布. Kai等(2010)把等度权组合分位数方法推广到非参数模型上. 近来, Jiang等(2014)给出了非线性模型的加权组合分位数估计. 本文讨论非等度权的变系数模型组合分位数估计.

§3. 漐近理论

为了方便表示, 我们先给出一些符号表示. 定义 $\mu_j = \int u^j K(u) du$, $v_j = \int u^j K^2(u) du$, $\gamma_{nh} = (nh)^{-1/2}$ 和 $\psi_{\tau_k}(s) = \tau_k - I(s \leq 0)$. 使 $r_i(u) = \alpha_0(U_i) - \alpha_0(u) - \alpha'_0(u)(U_i - u) + \mathbf{X}_i^\top (\boldsymbol{\alpha}(U_i) - \boldsymbol{\alpha}(u) - \boldsymbol{\alpha}'(u)(U_i - u))$, $\varepsilon_{i,k}^*(u) = \varepsilon_i - c_{\tau_k} + r_i(u)$, $\tau_{kk'} = \min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}$,

$t = \sum_{k,k'}^M w_k w_{k'} \tau_{kk'}$ 和 $\boldsymbol{\theta} = \gamma_{nh}^{-1} \{ (\mathbf{a}_0^* - \boldsymbol{\alpha}_0^*(u))^\top, (\mathbf{a} - \boldsymbol{\alpha}(u))^\top, h(b_0 - \alpha'_0(u)), h(\mathbf{b} - \boldsymbol{\alpha}'(u))^\top \}^\top$.

这里 $\boldsymbol{\alpha}_0^*(u) = (\alpha_{01}(u), \dots, \alpha_{0M}(u))^\top$ 与 $\alpha_{0k}(u) = \alpha_0(u) + c_{\tau_k}$. 进一步, 让 $U_{ih} = (U_i - u)/h$, $\Pi(u) = \mathbb{E}[\mathbf{X}\mathbf{X}^\top | U = u]$, $\mathbf{C} = \text{diag}(w_1 f_\varepsilon(c_{\tau_1}), \dots, w_M f_\varepsilon(c_{\tau_M}))$, $\tilde{c} = \sum_{k=1}^M w_k f_\varepsilon(c_{\tau_k})$ 和 \mathbf{T} 是一个 $M \times M$ 矩阵, 其第 (k, k') 个元素为 $w_k w_{k'} \tau_{kk'}$. 此外, 令

$$\begin{aligned}\mathbf{D}(u) &= \mathbb{E} \left[\begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \tilde{c} \mathbf{X} \mathbf{X}^\top \end{pmatrix} \middle| U = u \right], \quad \mathbf{B}(u) = \mathbb{E} \left[\begin{pmatrix} 1 & \mathbf{0} \\ 0 & \mathbf{X} \mathbf{X}^\top \end{pmatrix} \middle| U = u \right], \\ \boldsymbol{\Sigma}(u) &= \mathbb{E} \left[\begin{pmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & t \mathbf{X} \mathbf{X}^\top \end{pmatrix} \middle| U = u \right],\end{aligned}$$

以及 $\mathbf{X}_{i,k} = \{e_k^\top, \mathbf{X}_i^\top, U_{ih}, \mathbf{X}_i^\top U_{ih}\}^\top$, 这里 e_k 是第 k 个元素为 1 其余元素为 0 的 M 维向量.

为了证明局部加权组合分位数估计的渐近性质, 我们给出下列假设条件. 这些假设已经被广泛使用于局部最小二乘回归中(见Fan和Zhang, 1999; Cai等, 2000以及Fan和Zhang, 2008等).

假设 1 随机变量 U 存在有界支撑 Ω 并且其密度函数 $f_U(\cdot)$ 为正且二阶导连续.

假设 2 在 $u \in \Omega$ 上, 系数函数 $\alpha_0(\cdot)$ 和 $\boldsymbol{\alpha}(\cdot)$ 有连续的二阶导数.

假设 3 对所有的 $u \in \Omega$, $\mathbf{D}(u)$ 和 $\mathbf{B}(u)$ 非奇异.

假设 4 $f_\varepsilon(\cdot)$ 有界且不等于 0, 以及其二阶导数连续且一致有界.

假设 5 核函数 $K(\cdot)$ 是有界和非负的, 并且存在关于 0 对称的紧支撑.

假设 6 存在 $\delta^* > \delta > 2$, 使得 $\mathbb{E}\|\mathbf{X}_i\|^{2\delta^*} < \infty$ 且有 $nh^{\delta/(\delta-2)} \rightarrow \infty$.

在给出局部加权组合分位数估计的渐近性质之前, 我们先阐述一些必要的引理, 其中引理3.1参见Fan和Zhang (1999), 引理3.2与3.3来源于Cai和Xu (2008).

引理 3.1 设 $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ 是独立的随机向量, 其中 $Y_i, i = 1, \dots, n$ 是一列数值随机变量. 进一步, 使 $\mathbb{E}|Y|^s < \infty$ 且 $\sup_x \int |y|^s f(x, y) dy < \infty$, 这里 f 表示向量 (\mathbf{X}, Y) 的联合密度. 设 $K(\cdot)$ 是具有有界支撑并且满足Lipschitz条件的一个有界正函数. 假设对 $\epsilon < 1 - s^{-1}$ 有 $n^{2\epsilon-1}h \rightarrow \infty$, 则

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n [K_h(\mathbf{X}_i - \mathbf{x}) Y_i - \mathbb{E}\{K_h(\mathbf{X}_i - \mathbf{x}) Y_i\}] \right| = O_p(\{\log(1/h)/nh\}^{1/2}).$$

引理 3.2 设 $\mathcal{V}_n(\Delta)$ 为一个向量函数并且满足:

(i) 对任意 $\lambda \geq 1$, $-\Delta^\top \mathcal{V}_n(\lambda \Delta) \geq -\Delta^\top \mathcal{V}_n(\Delta)$;

(ii) 对 $0 < \mathcal{E} < \infty$, $\sup_{\|\Delta\| \leq \mathcal{E}} \|\mathcal{V}_n(\Delta) + \mathbf{G}\Delta - \mathbf{B}_n\| = o_p(1)$, 这里 \mathbf{B}_n 是一个随机向量满足 $\|\mathbf{B}_n\| = O_p(1)$ 以及 \mathbf{G} 是一个非随机矩阵;

(iii) Δ_n 是一个随机向量满足 $\|\mathcal{V}_n(\Delta_n)\| = o_p(1)$;

那么, $\|\Delta_n\| = O_p(1)$ 并且 $\Delta_n = \mathbf{G}^{-1}\mathbf{B}_n + o_p(1)$.

从 $\boldsymbol{\theta}$ 的定义知 $Y_i - a_{0k} - b_0(U_i - u) - \mathbf{X}_i^\top \{\mathbf{a} + \mathbf{b}(U_i - u)\} = \varepsilon_{i,k}^*(u) - \gamma_{nh}\boldsymbol{\theta}^\top \mathbf{X}_{i,k}$, 于是最小化方程(2.4)等价于

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{k=1}^M w_k \left\{ \sum_{i=1}^n \rho_{\tau_k}(\varepsilon_{i,k}^*(u) - \gamma_{nh}\boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}) \right\}.$$

对某个 $0 < \mathcal{E} < \infty$, 定义 $\Gamma_{\mathcal{E}} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq \mathcal{E}\}$,

$$\mathcal{V}_n(\boldsymbol{\theta}) = \gamma_{nh} \sum_{k=1}^M w_k \left[\sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u) - \gamma_{nh}\boldsymbol{\theta}^\top \mathbf{X}_{i,k}) \mathbf{X}_{i,k} K(U_{ih}) \right]$$

和

$$\mathcal{Q}_{nk}(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \psi_{\tau_k}(\varepsilon_{i,k}^*(u) - \gamma_{nh}\boldsymbol{\theta}^\top \mathbf{X}_{i,k}) - \psi_{\tau_k}(\varepsilon_{i,k}^*(u)) \right\} \mathbf{X}_{i,k} K(U_{ih}).$$

引理 3.3 若使 $\hat{\boldsymbol{\delta}} = \arg \min_{\boldsymbol{\delta}} \sum_{i=1}^n w_i \rho_{\tau_k}(Y_i - \boldsymbol{\delta}^\top \mathbf{X}_i)$, 这里 $w_i > 0$, 则有

$$\left\| \sum_{i=1}^n w_i \mathbf{X}_i \psi_{\tau_k}(Y_i - \hat{\boldsymbol{\delta}}^\top \mathbf{X}_i) \right\| \leq \dim(\mathbf{X}) \max_{i \leq n} \|w_i \mathbf{X}_i\|.$$

引理 3.4 在假设1-6条件下, 有

$$\sup_{\boldsymbol{\theta} \in \Gamma_{\mathcal{E}}} \left\| \mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0}) - \mathbb{E}[\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0})] \right\| = o_p(1).$$

证明: 从 $\mathcal{Q}_{nk}(\boldsymbol{\theta})$ 的定义知 $\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0}) = \gamma_{nh} \sum_{k=1}^M w_k \mathcal{Q}_{nk}(\boldsymbol{\theta})$ 并且注意到

$$\begin{aligned} \gamma_{nh} \mathcal{Q}_{nk}(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n I(\varepsilon_{i,k}^*(u) \leq 0) \mathbf{X}_{i,k} K_h(U_i - u) \\ &\quad - \frac{1}{n} \sum_{i=1}^n I(\varepsilon_{i,k}^*(u) \leq \gamma_{nh}\boldsymbol{\theta}^\top \mathbf{X}_{i,k}) \mathbf{X}_{i,k} K_h(U_i - u) \\ &= \mathcal{L}_1 - \mathcal{L}_2. \end{aligned}$$

由于 \mathcal{L}_1 和 \mathcal{L}_2 都是独立同分布随机变量与核函数乘积的和形式, 所以从引理3.1得 $\mathcal{L}_1 = \mathbb{E}(\mathcal{L}_1) + O_p(\log^{1/2}(1/h)/\sqrt{nh})$ 以及 $\mathcal{L}_2 = \mathbb{E}(\mathcal{L}_2) + O_p(\log^{1/2}(1/h)/\sqrt{nh})$. 于是,

$$\gamma_{nh} \mathcal{Q}_{nk}(\boldsymbol{\theta}) = \gamma_{nh} \mathbb{E}[\mathcal{Q}_{nk}(\boldsymbol{\theta})] + O_p(\log^{1/2}(1/h)/\sqrt{nh}).$$

这样, $\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0}) = \gamma_{nh} \sum_{k=1}^M w_k \mathbb{E}[\mathcal{Q}_{nk}(\boldsymbol{\theta})] + O_p(\log^{1/2}(1/h)/\sqrt{nh}) = \mathbb{E}[\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0})] + o_p(1)$, 即证. \square

引理 3.5 如果假设1-6被满足, 则

$$\left\| \mathbb{E}[\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0})] + f_U(u) \mathbf{S}(u) \boldsymbol{\theta} \right\| = o_p(1),$$

这里 $\mathbf{S}(u) = \text{diag}\{\mathbf{D}(u), \tilde{c}\mu_2 \mathbf{B}(u)\}$.

证明：根据迭代期望法则，有

$$\begin{aligned}
 & \mathbb{E}[\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0})] \\
 &= \gamma_{nh} \sum_{k=1}^M w_k \mathbb{E}[\mathcal{Q}_{nk}(\boldsymbol{\theta})] = \gamma_{nh} \sum_{k=1}^M w_k \mathbb{E}[\mathbb{E}(\mathcal{Q}_{nk}(\boldsymbol{\theta}) | U, \mathbf{X})] \\
 &= \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n \{F_\varepsilon(c_{\tau_k} - r_i(u)) - F_\varepsilon(c_{\tau_k} - r_i(u) + \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k})\} \mathbf{X}_{i,k} K(U_{ih}) \\
 &= -\mathbf{S}_n^*(u)\boldsymbol{\theta} + o(1),
 \end{aligned}$$

其中 $\mathbf{S}_n^*(u) = (nh)^{-1} \sum_{k=1}^M w_k \sum_{i=1}^n f_\varepsilon(c_{\tau_k}) \mathbf{X}_{i,k} \mathbf{X}_{i,k}^\top K(U_{ih})$.

进一步，从引理3.1知 $\mathbf{S}_n^*(u) \xrightarrow{p} \mathbb{E}[\mathbf{S}_n^*(u)] = f_U(u)\mathbf{S}(u) + O_p(h^2)$. 于是，

$$\mathbb{E}[\mathcal{V}_n(\boldsymbol{\theta}) - \mathcal{V}_n(\mathbf{0})] \xrightarrow{p} -f_U(u)\mathbf{S}(u)\boldsymbol{\theta},$$

即该引理成立. \square

引理 3.6 在假设1-6条件下，得 $\|\mathcal{V}_n(\mathbf{0})\| = O_p(1)$.

证明：首先，定义

$$\mathcal{V}_n^*(\mathbf{0}) = \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k} \{\varepsilon_i - c_{\tau_k}\} \mathbf{X}_{i,k} K(U_{ih}),$$

且注意

$$\text{Cov}(\psi_{\tau_k}(\varepsilon_i - c_{\tau_k}), \psi_{\tau_{k'}}(\varepsilon_j - c_{\tau_{k'}})) = \begin{cases} \tau_{kk'}, & \text{当 } i = j; \\ 0, & \text{当 } i \neq j. \end{cases}$$

其次，经过简单的代数运算知

$$\text{Var}(\mathcal{V}_n^*(\mathbf{0})) \rightarrow f_U(u)\text{diag}(v_0 \Sigma(u), tv_2 \mathbf{B}(u)) = O(1).$$

最后，从

$$\mathcal{V}_n(\mathbf{0}) = \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k} \{\varepsilon_i - c_{\tau_k} + r_i(u)\} \mathbf{X}_{i,k} K(U_{ih})$$

得

$$\begin{aligned}
 & \text{Var}(\mathcal{V}_n(\mathbf{0}) - \mathcal{V}_n^*(\mathbf{0}) | U, \mathbf{X}) \\
 & \leq \frac{M^2}{nh} \sum_{i=1}^n \mathbf{X}_{i,k} \mathbf{X}_{i,k}^\top K^2(U_{ih}) \max_k \{F_\varepsilon(c_{\tau_k} + |r_i(u)|) - F_\varepsilon(c_{\tau_k})\} = o_p(1).
 \end{aligned}$$

因此， $\text{Var}(\mathcal{V}_n(\mathbf{0}) - \mathcal{V}_n^*(\mathbf{0})) = o_p(1)$ ，这表明 $\text{Var}(\mathcal{V}_n(\mathbf{0})) = O_p(1)$ ，即 $\|\mathcal{V}_n(\mathbf{0})\| = O_p(1)$.

\square

下面定理给出了局部加权组合分位数估计的Bahadur表示.

定理 3.1 在假设1-6条件下, 有

$$\hat{\boldsymbol{\theta}} = \frac{\mathbf{S}^{-1}(u)}{\sqrt{nh}f_U(u)} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u)) \mathbf{X}_{i,k} K(U_{ih}) + o_p(1).$$

证明: 对于固定的 $\tau_k \in (0, 1)$ 和 $\lambda \geq 1$, 知

$$\begin{aligned} & -\boldsymbol{\theta}^\top \gamma_{nh} \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \lambda \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} \mathbf{X}_{i,k} K(U_{ih}) \\ &= \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \lambda \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} (-\gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}) \end{aligned}$$

是一个关于 λ 的增函数. 因此,

$$\begin{aligned} & \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \lambda \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} (-\gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}) \\ & \geq \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} (-\gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}). \end{aligned}$$

上式暗含着

$$\begin{aligned} & \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \lambda \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} (-\gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}) \\ & \geq \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k} \{ \varepsilon_{i,k}^*(u) - \gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k} \} (-\gamma_{nh} \boldsymbol{\theta}^\top \mathbf{X}_{i,k}) K(U_{ih}), \end{aligned}$$

即 $-\boldsymbol{\theta}^\top \mathcal{V}_n(\lambda \boldsymbol{\theta}) \geq -\boldsymbol{\theta}^\top \mathcal{V}_n(\boldsymbol{\theta})$. 于是, 引理3.2的条件(i)满足.

接下来, 如果使 $\mathbf{B}_n = \mathcal{V}_n(\mathbf{0})$ 和 $\mathbf{G} = f_U(u)\mathbf{S}(u)$, 则从引理3.4和3.5知

$$\sup_{\|\boldsymbol{\theta}\| \leq \mathcal{E}} \|\mathcal{V}_n(\boldsymbol{\theta}) + \mathbf{G}\boldsymbol{\theta} - \mathbf{B}_n\| = o_p(1)$$

成立, 并且由引理3.6得 $\|\mathbf{B}_n\| = O_p(1)$. 这样, 引理3.2的条件(ii)被满足.

从引理3.3知

$$\begin{aligned} \|\mathcal{V}_n(\hat{\boldsymbol{\theta}})\| & \leq M \dim(\gamma_{nh} \mathbf{X}_{i,k}) \max_{i \leq n} \|\gamma_{nh} K(U_{ih}) \mathbf{X}_{i,k}\| \\ & = O(1) \max_{i \leq n} \|\gamma_{nh} K(U_{ih}) \mathbf{X}_{i,k}\|. \end{aligned} \tag{3.1}$$

对任意 $\tilde{\epsilon} > 0$ 和常数 $C > 0$ 以及 $\delta^* > \delta > 2$, 由Markov不等式有

$$\begin{aligned} \mathsf{P} \left(\max_{i \leq n} \left\| \frac{K(U_{ih}) \mathbf{X}_{i,k}}{\sqrt{nh}} \right\| > \tilde{\epsilon} n^{1/\delta-1/2} h^{-1/2} \right) & \leq n \mathsf{P} \left(\left\| \frac{K(U_{ih}) \mathbf{X}_{i,k}}{\sqrt{nh}} \right\| > \tilde{\epsilon} n^{1/\delta-1/2} h^{-1/2} \right) \\ & \leq C \frac{\mathsf{E} \|K(U_{ih}) \mathbf{X}_{i,k}\|^{\delta^*}}{\tilde{\epsilon}^{\delta^*} n^{\delta^*/\delta-1}}. \end{aligned}$$

应用Cauchy-Schwarz不等式知

$$\mathbb{E}\|K(U_{ih})\mathbf{X}_i\|^{\delta^*} \leq \sqrt{\mathbb{E}|K(U_{ih})|^{2\delta^*}}\sqrt{\mathbb{E}\|\mathbf{X}_i\|^{2\delta^*}}.$$

因为

$$\begin{aligned}\mathbb{E}|K(U_{ih})|^{2\delta^*} &= \int \left|K\left(\frac{U_i - u}{h}\right)\right|^{2\delta^*} f(U_i) dU_i \\ &= h \int |K(t)|^{2\delta^*} f(u + ht) dt = O(h),\end{aligned}$$

所以

$$\mathbb{P}\left(\max_{i \leq n} \left\|\frac{K(U_{ih})\mathbf{X}_{i,k}}{\sqrt{nh}}\right\| > \tilde{\epsilon}n^{1/\delta-1/2}h^{-1/2}\right) \leq O(\sqrt{h})O(n^{1-\delta^*/\delta}) = o(1). \quad (3.2)$$

从假设条件6及式(3.2)知

$$\max_{i \leq n} \|\gamma_{nh}K(U_{ih})\mathbf{X}_{i,k}\| = o_p(n^{1/\delta-1/2}h^{-1/2}) = o_p(1). \quad (3.3)$$

最后, 结合式(3.1)和式(3.3)得 $\|\mathcal{V}_n(\hat{\theta})\| = O(1)o_p(1) = o_p(1)$. 因此, 根据引理3.2, 有

$$\hat{\theta} = \mathbf{G}^{-1}\mathbf{B}_n + o_p(1) = \frac{\mathbf{S}^{-1}(u)}{\sqrt{nh}f_U(u)} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u)) \mathbf{X}_{i,k} K(U_{ih}) + o_p(1),$$

即证该定理. \square

下面的定理给出了局部加权组合分位数估计的渐近正态性.

定理 3.2 在假设1-6条件下, 有

$$\sqrt{nh} \begin{pmatrix} \hat{\mathbf{a}}_0^* - \boldsymbol{\alpha}_0^*(u) \\ \hat{\mathbf{a}} - \boldsymbol{\alpha}(u) \end{pmatrix} - \frac{\mu_2 h^2}{2} \begin{pmatrix} \boldsymbol{\alpha}_0''(u) \\ \boldsymbol{\alpha}''(u) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\{\mathbf{0}, \boldsymbol{\Upsilon}(u)\},$$

这里 $\xrightarrow{\mathcal{D}}$ 表示依分布收敛以及 $\boldsymbol{\Upsilon}(u) = v_0 f_U^{-1}(u) \mathbf{D}^{-1}(u) \boldsymbol{\Sigma}(u) \mathbf{D}^{-1}(u)$.

证明: 由于 $\mathbf{S}(u) = \text{diag}(\mathbf{D}(u), \tilde{c}\mu_2 \mathbf{B}(u))$ 是一个拟对角矩阵, 所以从定理3.1知

$$\sqrt{nh} \begin{pmatrix} \hat{\mathbf{a}}_0^* - \boldsymbol{\alpha}_0^*(u) \\ \hat{\mathbf{a}} - \boldsymbol{\alpha}(u) \end{pmatrix} = f_U^{-1}(u) \mathbf{D}^{-1}(u) \mathbf{W}_n(u) + o_p(1), \quad (3.4)$$

其中 $\mathbf{W}_n(u) = \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n K(U_{ih}) \psi_{\tau_k}(\varepsilon_i - c_{\tau_k} + r_i(u)) (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top$.

记 $\mathbf{W}_n^\#(u) = \mathbf{W}_n(u) - \mathbf{W}_n^*(u)$, 这里

$$\mathbf{W}_n^*(u) = \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n K(U_{ih}) \psi_{\tau_k}(\varepsilon_i - c_{\tau_k}) (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top.$$

这样, 由 $\mathbf{W}_n(u)$ 和 $\mathbf{W}_n^*(u)$ 的定义得

$$\mathbf{W}_n^\#(u) = \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n K(U_{ih}) \{I(\varepsilon_i \leq c_{\tau_k}) - I(\varepsilon_i \leq c_{\tau_k} - r_i(u))\} (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top.$$

于是,

$$\begin{aligned} \mathbb{E}[\mathbf{W}_n^\#(u) | U, \mathbf{X}] &= \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n K(U_{ih}) \{F_\varepsilon(c_{\tau_k}) - F_\varepsilon(c_{\tau_k} - r_i(u))\} (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top \\ &= \gamma_{nh} \sum_{k=1}^M w_k \sum_{i=1}^n K(U_{ih}) r_i(u) f_\varepsilon(c_{\tau_k}) \{1 + o_p(1)\} (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top. \end{aligned}$$

一方面, 通过简单的矩阵运算知

$$\mathbb{E}[\mathbf{W}_n^\#(u)] \rightarrow \sqrt{nh} \frac{\mu_2 h^2}{2} f_U(u) \mathbf{D}(u) \begin{pmatrix} \boldsymbol{\alpha}_0''(u) \\ \boldsymbol{\alpha}''(u) \end{pmatrix} + O_p(\sqrt{nh} h^2). \quad (3.5)$$

此外, 令 $\hbar_{ik}^a = c_{\tau_k} - r_i(u)$ 和 $\hbar_{ik}^b = c_{\tau_k}$, 则有

$$\{\psi_{\tau_k}(\varepsilon_i - c_{\tau_k} + r_i(u)) - \psi_{\tau_k}(\varepsilon_i - c_{\tau_k})\}^2 = I(\lambda_{ik}^* < \varepsilon_i < \lambda_{ik}^\#),$$

这里 $\lambda_{ik}^* = \min(\hbar_{ik}^a, \hbar_{ik}^b)$ 和 $\lambda_{ik}^\# = \max(\hbar_{ik}^a, \hbar_{ik}^b)$.

于是,

$$\begin{aligned} &\mathbb{E}[\{\psi_{\tau_k}(\varepsilon_i - c_{\tau_k} + r_i(u)) - \psi_{\tau_k}(\varepsilon_i - c_{\tau_k})\}^2 K^2(U_{ih}) (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top (\mathbf{e}_k^\top, \mathbf{X}_i^\top)] \\ &= \mathbb{E}[\{F_\varepsilon(\lambda_{ik}^\#) - F_\varepsilon(\lambda_{ik}^*)\} K^2(U_{ih}) (\mathbf{e}_k^\top, \mathbf{X}_i^\top)^\top (\mathbf{e}_k^\top, \mathbf{X}_i^\top)] = O(h^3), \end{aligned}$$

这表明 $\text{Var}(\mathbf{W}_n^\#(u)) = o_p(1)$.

另一方面, 直接计算得 $\mathbb{E}[\mathbf{W}_n^*(u)] = 0$ 以及 $\text{Var}[\mathbf{W}_n^*(u)] \rightarrow v_0 f_U(u) \boldsymbol{\Sigma}(u)$.

应用 Cramér-Wold 定理, 知 $\mathbf{W}_n^*(u)$ 满足中心极限定理, 即

$$\mathbf{W}_n^*(u) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, v_0 f_U(u) \boldsymbol{\Sigma}(u)). \quad (3.6)$$

因此, 结合式(3.4), (3.5)和(3.6)推出该定理. \square

接下来, 定义 $\alpha_0(u)$ 和 $\boldsymbol{\alpha}(u)$ 的局部加权组合分位数估计分别为 $\hat{\alpha}_0(u) = M^{-1} \sum_{k=1}^M \hat{a}_{0k}$ 和 $\hat{\boldsymbol{\alpha}}(u) = \hat{\mathbf{a}}$.

定理 3.3 在假设1-6条件下, 有

$$\sqrt{nh} \left(\hat{\alpha}_0(u) - \alpha_0(u) - \frac{1}{M} \sum_{k=1}^M c_{\tau_k} - \frac{\mu_2 h^2}{2} \alpha_0''(u) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left\{ 0, R_1(M) \frac{v_0}{f_U(u)} \right\}$$

以及

$$\sqrt{nh} \left(\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\mu_2 h^2}{2} \boldsymbol{\alpha}''(u) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left\{ \mathbf{0}, R_2(M) \frac{v_0}{f_U(u)} \Pi^{-1}(u) \right\},$$

其中 $R_1(M) = M^{-2} \sum_{k=1}^M \sum_{k'=1}^M \tau_{kk'} f_\varepsilon^{-1}(c_{\tau_k}) f_\varepsilon^{-1}(c_{\tau_{k'}})$ 与 $R_2(M) = t/\tilde{c}^2$.

证明: 由定理3.1知

$$\gamma_{nh}^{-1}(\hat{\alpha}_{0k} - \alpha_{0k}(u)) = \frac{f_U^{-1}(u)}{\sqrt{nh}f_\varepsilon(c_{\tau_k})} \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u))K(U_{ih}) + o_p(1).$$

于是,

$$\gamma_{nh}^{-1}\left(\hat{\alpha}_0(u) - \alpha_0(u) - \frac{1}{M} \sum_{k=1}^M c_{\tau_k}\right) = \frac{f_U^{-1}(u)}{M\sqrt{nh}} \sum_{k=1}^M f_\varepsilon^{-1}(c_{\tau_k}) \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u))K(U_{ih}) + o_p(1).$$

类似于定理3.2的证明知

$$\begin{aligned} & \frac{1}{\sqrt{nh}} \sum_{k=1}^M f_\varepsilon^{-1}(c_{\tau_k}) \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u))K(U_{ih}) \\ & \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\mu_2 h^2 M f_U(u)}{2} \sqrt{nh} \alpha_0''(u), v_0 f_U(u) M^2 R_1(M)\right). \end{aligned}$$

因此, $\hat{\alpha}_0(u)$ 的渐近正态性成立.

同理,

$$\gamma_{nh}^{-1}(\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u)) = \frac{(\bar{c}\Pi(u))^{-1}}{\sqrt{nh}f_U(u)} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u))\mathbf{X}_i K(U_{ih}) + o_p(1)$$

以及

$$\frac{1}{\sqrt{nh}} \sum_{k=1}^M w_k \sum_{i=1}^n \psi_{\tau_k}(\varepsilon_{i,k}^*(u))\mathbf{X}_i K(U_{ih}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\mu_2 h^2}{2} \bar{c}\Pi(u) f_U(u) \sqrt{nh} \boldsymbol{\alpha}''(u), t v_0 f_U(u) \Pi(u)\right).$$

这样, 由上述两式直接获得 $\hat{\boldsymbol{\alpha}}(u)$ 的渐近正态性. \square

从定理3.3很容易看出, 局部加权组合分位数估计是 \sqrt{nh} 收敛的. 并且令人惊讶的是定理3.3所得到的误偏与局部最小二乘估计的相同, 但是协方差矩阵却不同, 见Zhang和Lee (2000)对模型(1.1)的局部最小二乘估计的误偏与方差.

§4. 最优权选择

方程(2.4)表明局部加权组合分位数估计依赖于分位点个数 M 和权重 $\{w_k\}_{k=1}^M$. Zou和Yuan (2008), Kai等(2010)以及Jiang等(2014)指出如何选择一个最优的 M 有待研究. 然而, 通常使用一个小的数值 M , 例如 $M = 9$ 能够获得组合分位数估计的潜在有效性. 同时, 使用太大的 M 值, 不仅有有可能不会提高估计量的有效性而且还会增加模型的计算时间.

由定理3.3知, 系数函数估计的方差为 $v_0 f_U^{-1}(u) \Pi^{-1}(u) R_2(M)$, 权重的选择应当使其达到最小. 由于 v_0 , $f_U^{-1}(u)$ 和 $\Pi^{-1}(u)$ 均不涉及 w_k , 因此只需最小化 $R_2(M)$ 来选择权重即可. 下面我们采用类似于文Bradic等(2011)的方法来选择最优权重. 记 $\mathbf{w} = (w_1, \dots, w_M)^\top$ 和 $\boldsymbol{\Gamma}$ 是

一个 $M \times M$ 矩阵并且第 (k, k') 位置上的元素是 $\tau_{kk'}$. 局部加权组合分位数估计的最优权重 $\mathbf{w}_+^{\text{opt}}$ 可以通过解如下二次规划问题

$$\mathbf{w}_+^{\text{opt}} = \begin{cases} \min & \mathbf{w}^\top \mathbf{\Gamma} \mathbf{w}, \\ \text{s.t.} & \mathbf{f}^\top \mathbf{w} = 1, \end{cases} \quad 0 \leq \mathbf{w} < \infty$$

来计算, 这里 $\mathbf{f} = (f_\varepsilon(c_{\tau_1}), \dots, f_\varepsilon(c_{\tau_M}))^\top$.

表1 当 $M = 11$ 时, 不同误差分布下最优权选择

分位点	1/12	2/12	3/12	4/12	5/12	6/12	7/12	8/12	9/12	10/12	11/12
t(3)	0.00	0.05	0.11	0.13	0.14	0.14	0.14	0.13	0.11	0.05	0.00
e(2)	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\mathcal{B}(1, 3)$	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.07
$\mathcal{G}(3, 2)$	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.09	0.05	0.05	0.06
$\mathcal{N}(0, 1)$	0.19	0.09	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.09	0.19
Cauchy	0.00	0.00	0.00	0.05	0.30	0.30	0.30	0.05	0.00	0.00	0.00
Laplace	0.00	0.00	0.00	0.00	0.27	0.45	0.27	0.00	0.00	0.00	0.00
Mixture	0.00	0.15	0.12	0.10	0.09	0.09	0.09	0.10	0.12	0.15	0.00
Log $\mathcal{N}(0, 1)$	0.63	0.31	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00

通过解上述二次规划问题, 表1给出了各种误差分布在 $M = 11$ 时的最优权. 这些误差分布包括: 正态分布 $\mathcal{N}(0, 1)$, 自由度为3的t分布 $t(3)$, Beta分布 $\mathcal{B}(1, 3)$, Gamma分布 $\mathcal{G}(3, 2)$, 参数为2的指数分布 $e(2)$, 对数正态分布 $\text{Log}\mathcal{N}(0, 1)$, 混合正态分布 Mixture $0.1\mathcal{N}(0, 25) + 0.9\mathcal{N}(0, 1)$, Laplace分布 Laplace 以及 Cauchy 分布 Cauchy.

从表1中, 很容易看出不同误差分布所需要的有效最优权数以及各分位点所对应的权值大小. 当误差分布是 $e(2)$ 时, 只需要1个分位点; Laplace和 $\text{Log}\mathcal{N}(0, 1)$ 分布分别需要3个和4个分位点; $\mathcal{B}(1, 3)$ 和 Cauchy 分布需要5个分位点; $t(3)$ 和 Mixture 分布需要9个分位点; 而对于 $\mathcal{N}(0, 1)$ 和 $\mathcal{G}(3, 2)$ 分布来说, 所有的11个分位点都需要. 由于一些权值等于0, 相对于使用等度权来说, 采用最优权值 $\mathbf{w}_+^{\text{opt}}$ 能够有效减少计算量. 此外, 对不同分位点赋予不同的权值能够使重要的分位点对估计量产生更大作用, 同时也能够减少次要分位点造成的影响.

§5. 演近相对效率

为了演示估计量的有效性, 本节检验局部加权组合分位数与局部最小二乘估计 $\alpha(u)$ 的演近相对效率 (ARE). 我们把 $\hat{\alpha}^{\text{Wcq}}(u)$ 和 $\hat{\alpha}^{\text{Ls}}(u)$ 相应的定义为 $\alpha(u)$ 的局部加权组合分位数 (Wcq) 估计和局部最小二乘 (Ls) 估计. 通过计算演近均方误差 MSE 来测量它们的有效性.

从Zhang和Lee(2000)可知 $\hat{\alpha}^{\text{Ls}}(u)$ 的渐近MSE为

$$\begin{aligned}\text{MSE}^{\text{Ls}}(h, u) &= \mathbb{E}\|\hat{\alpha}^{\text{Ls}}(u) - \alpha(u)\|^2 \\ &= \frac{\mu_2^2\|\alpha''(u)\|^2}{4}h^4 + \frac{\sigma^2v_0}{f_U(u)}\text{tr}\{\Pi^{-1}(u)\}\frac{1}{nh},\end{aligned}$$

这里 $\text{tr}(\cdot)$ 表示矩阵的迹以及 σ^2 是 ε 的方差. 通过最小化 $\text{MSE}^{\text{Ls}}(h, u)$, 得到局部最小二乘回归的最优窗宽为

$$h_{\text{opt}}^{\text{Ls}}(u) = n^{-1/5} \left[\frac{\sigma^2 v_0 \text{tr}\{\Pi^{-1}(u)\}}{\mu_2^2 \|\alpha''(u)\|^2 f_U(u)} \right]^{1/5}. \quad (5.1)$$

由定理3.3知 $\hat{\alpha}^{\text{Wcq}}(u)$ 的渐近MSE为

$$\begin{aligned}\text{MSE}^{\text{Wcq}}(h, u) &= \mathbb{E}\|\hat{\alpha}^{\text{Wcq}}(u) - \alpha(u)\|^2 \\ &= \frac{\mu_2^2\|\alpha''(u)\|^2}{4}h^4 + \frac{v_0 R_2(M)}{f_U(u)}\text{tr}\{\Pi^{-1}(u)\}\frac{1}{nh}.\end{aligned}$$

于是, 局部加权组合分位数回归的最优窗宽为

$$h_{\text{opt}}^{\text{Wcq}}(u) = n^{-1/5} \left[\frac{v_0 R_2(M) \text{tr}\{\Pi^{-1}(u)\}}{\mu_2^2 \|\alpha''(u)\|^2 f_U(u)} \right]^{1/5}. \quad (5.2)$$

定理 5.1 当 $n \rightarrow \infty$ 时, 局部加权组合分位数和局部最小二乘估计 $\alpha(u)$ 的渐近相对效率为

$$\text{ARE}(u) = \frac{\text{MSE}^{\text{Ls}}\{h_{\text{opt}}^{\text{Ls}}(u), u\}}{\text{MSE}^{\text{Wcq}}\{h_{\text{opt}}^{\text{Wcq}}(u), u\}} \rightarrow (R_2(M)\sigma^{-2})^{-4/5}.$$

证明: 把 $h_{\text{opt}}^{\text{Ls}}(u)$ 和 $h_{\text{opt}}^{\text{Wcq}}(u)$ 分别代入 $\text{MSE}^{\text{Ls}}(h, u)$ 与 $\text{MSE}^{\text{Wcq}}(h, u)$ 中, 然后由 $\text{ARE}(u)$ 的定义即得该定理. \square

表2 不同误差分布的渐近相对效率 $\text{ARE}(u)$ 值

误差分布	渐近相对效率 $\text{ARE}(u)$ 值						
	$M = 5$	$M = 9$	$M = 11$	$M = 13$	$M = 19$	$M = 29$	$M = 99$
$\mathcal{N}(0, 1)$	0.93	0.97	0.97	0.98	0.99	0.99	0.98
t(3)	4.06	4.10	4.12	4.12	4.12	4.12	4.12
Laplace	2.74	2.88	2.90	2.92	2.95	2.99	3.01
Mixture	5.39	5.45	5.45	5.47	5.47	5.47	5.47
Log $\mathcal{N}(0, 1)$	26.52	37.57	41.27	44.16	50.62	57.25	72.99

从定理5.1可以看出, $\text{ARE}(u)$ 严格依赖于误差项的分布. 表2给出了一些常见误差分布的 $\text{ARE}(u)$ 值. 由表中的统计结果可以发现, 局部最小二乘估计在误差项服从 $\mathcal{N}(0, 1)$ 时

有着最佳估计. 并且, 在 $\mathcal{N}(0, 1)$ 分布时, 无论 M 值的大小, $ARE(u)$ 均接近于1. 这表明局部加权组合分位数估计与局部最小二乘估计非常接近. 然而, 对于其它的四种分布来说, 局部加权组合分位数估计比局部最小二乘估计有着更高的效率, 即大的 $ARE(u)$ 值, 尤其 $Log\mathcal{N}(0, 1)$ 分布的效率达到十位数值以上. 除了 $Log\mathcal{N}(0, 1)$ 分布外, 对于相同分布而不同的 M 值, $ARE(u)$ 值几乎相差无几. 这说明使用一个小的 M 值与使用一个大的 M 值所得到的局部加权组合分位数估计不会相差太大.

§6. 窗宽选择

在局部多项式近似理论中, 窗宽承担着减小估计值误偏的重要角色. 尽管我们已经推导出了局部加权组合分位数估计的最优理论窗宽(5.2)式, 然而由于该式中存在着许多未知量, 直接使用这个最优理论窗宽 $h_{opt}^{Wcq}(u)$ 非常困难. 下面我们介绍一个备选方法来选择窗宽. 从(5.1)式和(5.2)式知

$$h_{opt}^{Wcq}(u) = (R_2(M)\sigma^{-2})^{1/5} h_{opt}^{Ls}(u) \quad (6.1)$$

成立. 因此, 我们采用上述关系来计算 $\hat{h}_{opt}^{Wcq}(u)$. 具体来说, 首先使用一些存在的窗宽选择准则来选择 $\hat{h}_{opt}^{Ls}(u)$, 例如Akaike信息准则(AIC), Bayes信息准则(BIC), 交叉核实准则(CV)和广义交叉核实准则(GCV)等. 本文使用Cai等(2000)提出的多重交叉核实准则(MCV)来选择 $\hat{h}_{opt}^{Ls}(u)$. 其基本思想是用长度为 $n - qm$ ($q = 1, \dots, Q$)的 Q 个子序列估计未知系数函数, 然后基于估计的模型来计算下一个长度为 m 部分的一步预测误差. 最优窗宽是通过最小化 Q 个子样本平均预测误差

$$APE(h) = Q^{-1} \sum_{q=1}^Q APE_q(h)$$

来得到的, 这里 $APE_q(h) = m^{-1} \sum_{i=n-qm+1}^{n-qm+m} \left\{ Y_i - \hat{\alpha}_0(U_i) - \sum_{j=1}^p \hat{\alpha}_j(U_i) X_{ij} \right\}^2$.

在计算中, 我们使用 $Q = 4$ 和 $m = \lfloor 0.1n \rfloor$, 其中 $\lfloor s \rfloor$ 表示不超过 s 的最大整数.

其次, 构造 σ^2 和 $R_2(M)$ 的估计. 事实上, 从 $R_2(M)$ 的表达式知, 计算 $R_2(M)$ 的核心是估计 $f_\varepsilon(F_\varepsilon^{-1}(\tau_k))$. 接下来, 讨论 σ^2 和 $f_\varepsilon(F_\varepsilon^{-1}(\tau_k))$ 的估计.

(i) 使用式(2.2)得到 $\hat{\alpha}_0(u), \hat{\alpha}_1(u), \dots, \hat{\alpha}_p(u)$ 并计算残差

$$\hat{\varepsilon}_i = Y_i - \hat{\alpha}_0(U_i) - \hat{\alpha}_1(U_i)X_{i1} - \cdots - \hat{\alpha}_p(U_i)X_{ip}.$$

(ii) 计算方差 σ^2 的估计 $\hat{\sigma}^2 = (n - p - 1)^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^2$.

(iii) 取 $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ 的样本 τ_k 分位点得到 $F_\varepsilon^{-1}(\tau_k)$ 的估计 $\hat{F}_\varepsilon^{-1}(\tau_k)$.

(iv) 使用非参数密度函数估计 $f_\varepsilon(s)$, 即 $\hat{f}_\varepsilon(s) = (nb)^{-1} \sum_{i=1}^n K((s - \hat{\varepsilon}_i)/b)$, 其中 $b > 0$ 是窗宽以及 $K(\cdot)$ 是非负核函数.

(v) 将 $\hat{F}_\varepsilon^{-1}(c_{\tau_k})$ 代入到 $\hat{f}_\varepsilon(s)$ 中得 $\hat{f}_\varepsilon(\hat{F}_\varepsilon^{-1}(\tau_k))$.

本文使用Gaussian核函数且选择窗宽

$$b = 0.9 \min \left\{ \text{sd}(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n), \frac{\text{IQR}(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)}{1.34} \right\} \times n^{-1/5},$$

这里 $\text{sd}(\cdot)$ 和 $\text{IQR}(\cdot)$ 分别代表样本标准差和四分位数间距.

最后, 根据(6.1)式得到 $\hat{h}_{\text{opt}}^{\text{Wcq}}(u)$.

§7. Monte Carlo模拟

这一节应用Monte Carlo模拟方法来评价局部加权组合分位数估计的有限样本性质. 我们采用Koenker和Park (1996)的内点迭代算法来计算局部Wcq估计. 在整个模拟过程中, Epanechnikov核函数 $K(z) = 0.75(1 - z^2)I(|z| < 1)$ 被使用.

考虑如下模型

$$Y = \alpha_0(U) + \alpha_1(U)X_1 + \alpha_2(U)X_2 + \varepsilon,$$

这里 $\alpha_0(u) = \exp(2u - 1)$, $\alpha_1(u) = 8u(1 - u)$ 和 $\alpha_2(u) = 2\sin^2(2\pi u)$; 协变量 U 服从 $[0, 1]$ 上的均匀分布 $U \sim \mathcal{U}[0, 1]$, 并且与 (X_1, X_2) 相互独立, 其中 X_1 和 X_2 服从相关系数为 $2^{-1/2}$ 的标准正态分布. 事实上, 上述模型及其协变量的生成过程与文Cai等(2000)相同. 我们考虑四种不同的误差分布, 它们是正态分布 $\mathcal{N}(0, 1)$, 自由度为3的t分布 $t(3)$, 混合正态分布Mixture $0.1\mathcal{N}(0, 25) + 0.9\mathcal{N}(0, 1)$ 以及Laplace分布Laplace.

为了评价估计量的有限样本性质, 我们计算 $\hat{\alpha}_j(\cdot)$ 的均方误差平方根(RASE), 其定义为

$$\text{RASE}_j = \left\{ \frac{1}{\tilde{n}} \sum_{l=1}^{\tilde{n}} \{ \hat{\alpha}_j(u_l) - \alpha_j(u_l) \}^2 \right\}^{1/2}, \quad j = 0, 1, 2,$$

这里 $\hat{\alpha}_j(\cdot)$ 是局部加权组合分位数(Wcq)或局部最小二乘(Ls)估计及 $\{u_l = 0.02(l-1) + 0.02 : 1 \leq l \leq \tilde{n} = 49\}$ 是一系列格点.

图1给出了样本容量 $n = 400$ 并且误差分布为 $\mathcal{N}(0, 1)$ 及Mixture的典型数据集, 使用局部加权组合分位数和局部最小二乘拟合系数函数的情形. 为了评价 M 值的影响, 我们也给出了 $M = 5, 11$ 和 99 时的局部加权组合分位数估计系数函数的拟合曲线. 从图1的(a), (c)和(e)中, 可以看出当误差项服从 $\mathcal{N}(0, 1)$ 时, 局部加权组合分位数与局部最小二乘对真实系数函数的拟合几乎一样好; 同时, $M = 5, 11$ 和 99 的局部加权组合分位数估计对系数函数的拟合也没有太大的差别. 然而, 从图1的(b), (d)和(f)中可以发现当误差项服从Mixture时, 局部最小二乘估计对真实系数函数的拟合明显要比局部加权组合分位数估计表现的差; 同时也可看出, $M = 99$ 时的局部加权组合分位数拟合没有明显地呈现出比 $M = 5$ 和 $M = 11$ 时的好. 这也证实了前文的结果, 使用一个大的 M 值不一定能够获得一个更优估计.

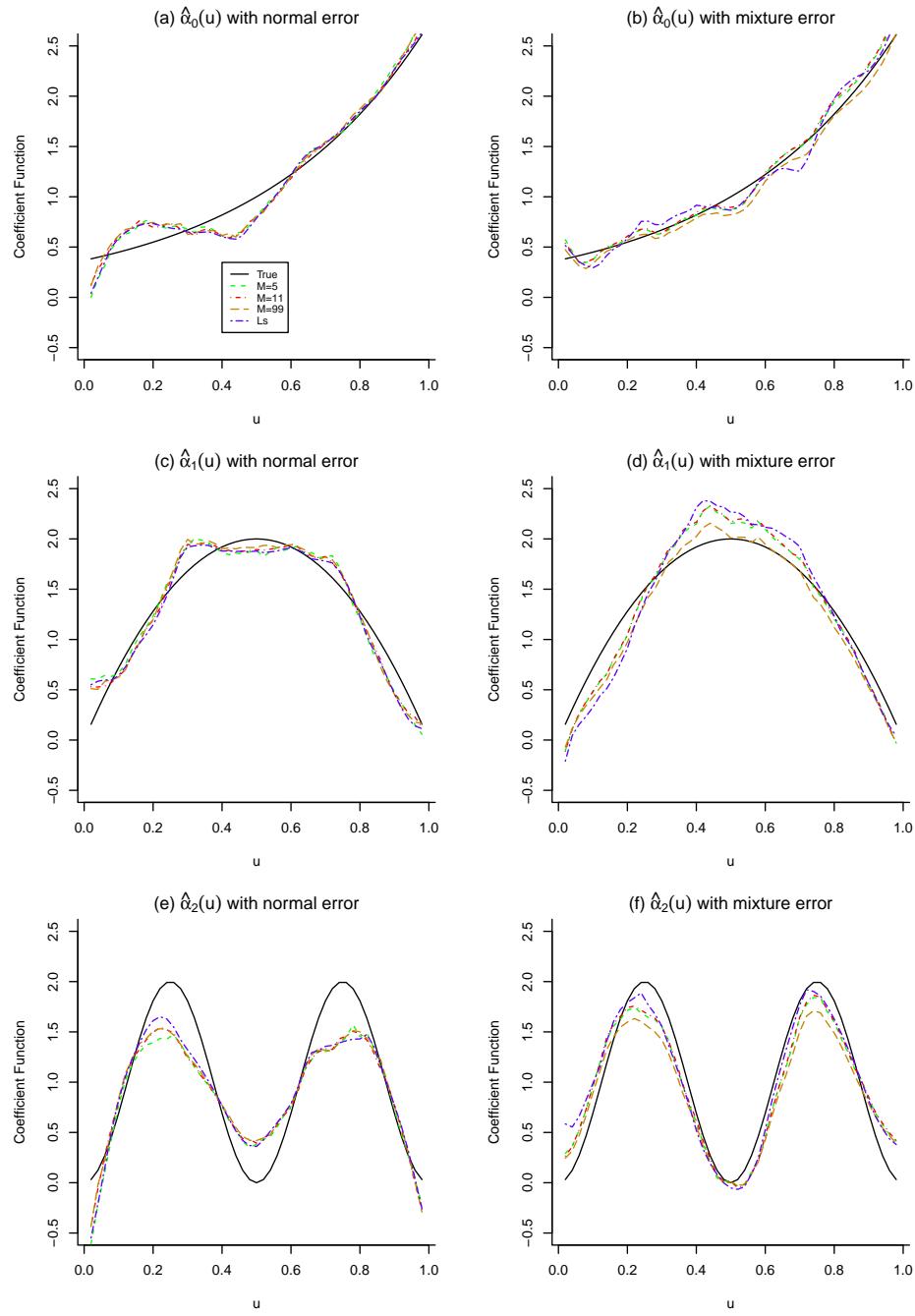


图1 一个典型数据集的系数函数估计

当 $M = 9$ 并且误差分布为Mixture时, 考查局部最小二乘估计和局部加权组合分位数估计的抽样性质. 我们生成1000个随机样本并进行500次的随机抽样估计. 同时, 每一次的随机抽样选取的样本容量为 $n = 500$. 经过500次的重复模拟, 局部加权组合分位数和局部最小二乘估计 $\{\hat{\alpha}_j(u)\}_{j=0}^2$ 的拟合曲线描绘在图2中. 从该图中很明显地看出, 局部加权组合

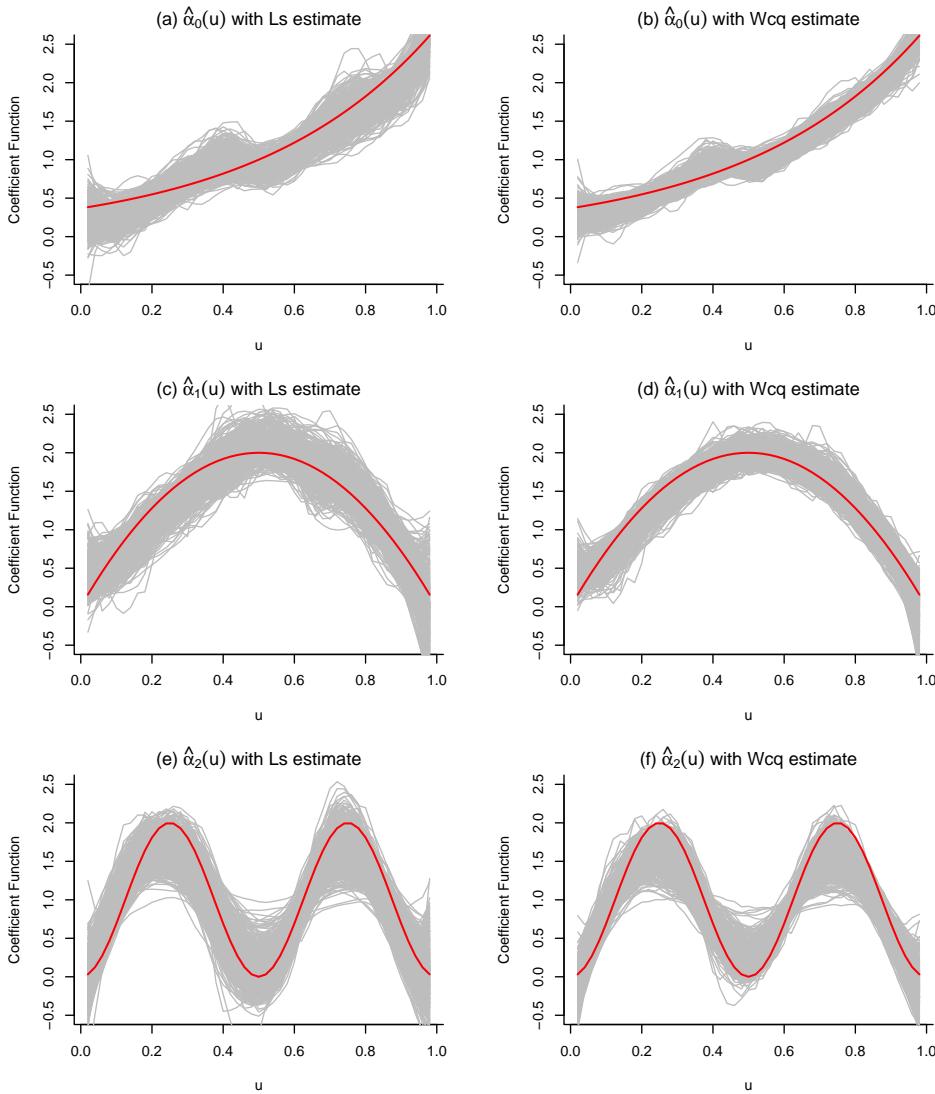


图2 模拟500次的系数函数估计

分位数估计的拟合曲线带要比局部最小二乘的窄. 这个结果表明局部加权组合分位数估计更稳定并且估计值更接近真实值.

图3给出的是 $M = 9$, 样本容量为 $n = 400$ 和 $n = 800$ 时, 四种不同误差分布($\mathcal{N}(0, 1)$, Mixture, $t(3)$ 和Laplace)的500次模拟RASE统计结果. 图3中的四个子图对应着四个不同的误差分布. 其中每一幅子图给出的是样本容量为 $n = 400$ 和 $n = 800$ 时, 每一个系数函数的500次模拟RASE均值. 同时, 浅灰色的条体和深灰色的条体分别表示的是局部加权组合分位数估计和局部最小二乘估计的RASE均值. 从图3可以看出只有 $\mathcal{N}(0, 1)$ 误差分布时, 每一个系数函数的局部加权组合分位数与局部最小二乘估计的RASE均值接近; 然而对于其它重尾误差分布来说, 局部加权组合分位数估计的每个系数函数有着更小的RASE均值.

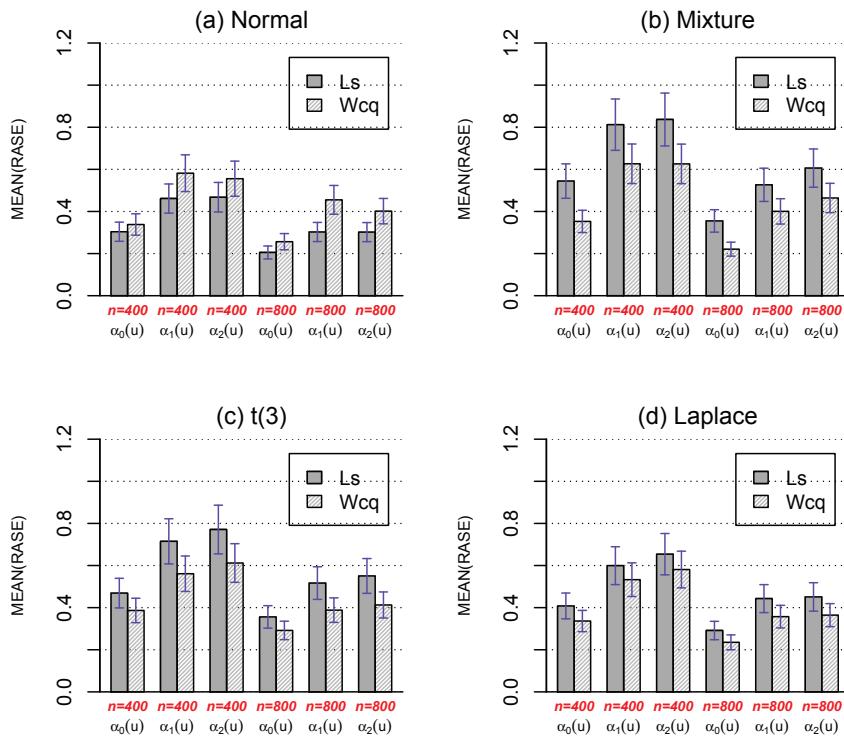


图3 Ls与Wcq估计系数函数的RASE

由图3的每一幅子图, 可以发现随着样本的不断增大, 局部加权组合分位数估计系数函数的RASE均值会逐渐减小.

§8. 实证研究

本节使用局部加权组合分位数方法来分析香港1994年1月1日到1995年11月31日间的环境数据集. 我们研究该样本期间内每周五的污染物水平与就医总人数之间的相互关系. 取就医总人数的对数作为反应变量(Y), 以及协变量包括二氧化硫水平(X_1), 二氧化氮水平(X_2)和粉尘水平(X_3). 我们采用如下的变系数模型来分析这个数据集:

$$Y = \alpha_0(u) + \alpha_1(u)X_1 + \alpha_2(u)X_2 + \alpha_3(u)X_3 + \varepsilon,$$

其中 u 表示时间并被标准化在 $[0, 1]$ 区间上. 需要说明的是这个数据集和模型已经被Fan和Zhang (1999)所检验.

我们分别使用局部加权组合分位数(Wcq)与局部最小二乘(Ls)来拟合上述模型. 在进行局部Wcq回归时, 我们考查了 $M = 5, 11, 21$ 的系数函数估计. 由于这三个 M 值的Wcq估计非常相似, 因此下文仅提供 $M = 11$ 时的计算结果. 按照第6节介绍的方法, 我们得到 $\hat{\boldsymbol{w}}_+^{\text{opt}} = (0.00, 0.25, 0.01, 0.00, 0.17, 0.09, 0.08, 0.17, 0.00, 0.00, 0.22)^\top$ 以及 $\hat{h}_{\text{opt}}^{\text{Wcq}} = 0.37$.

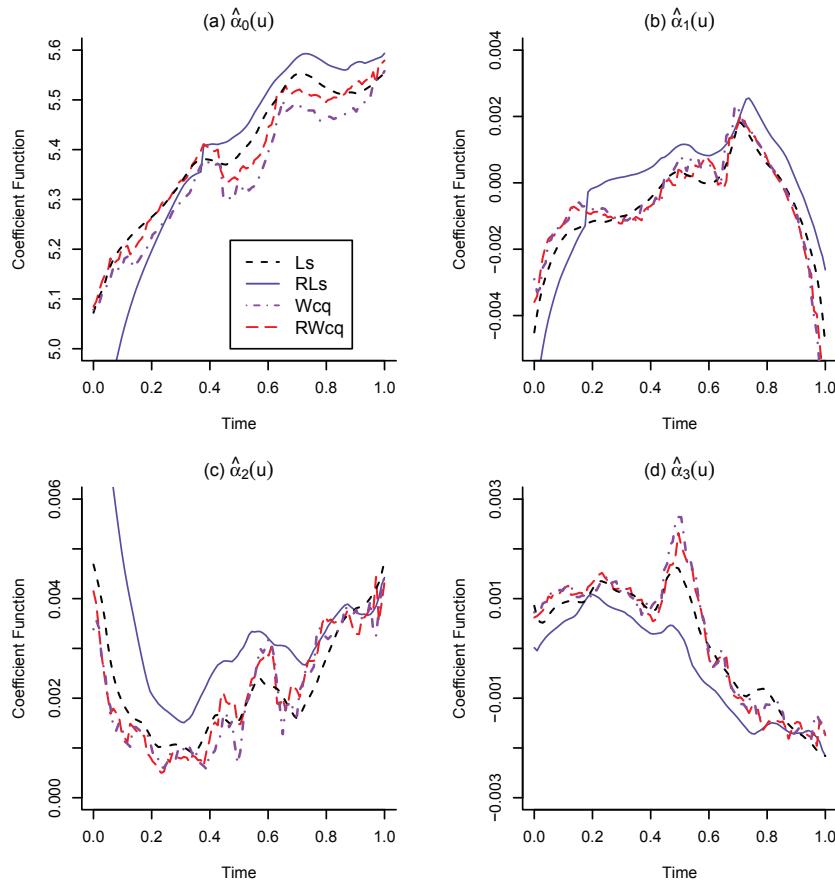


图4 系数函数拟合趋势

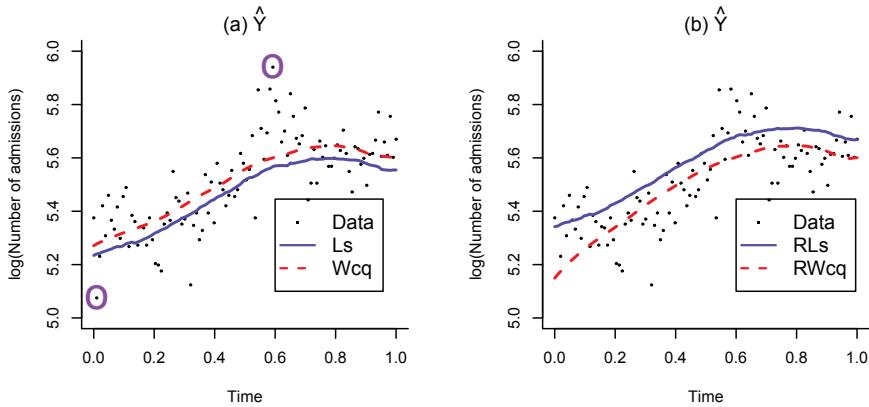


图5 样本散点图及就医总人数对数的预期估计

图4描绘出了估计的系数函数形状. 从图中可以清晰地发现所有估计出的系数函数都随时间变化并且局部Ls与局部Wcq估计的总体趋势基本一致. 此外, 图5(a)给出了拟合的曲线形状 $\hat{Y} = \hat{\alpha}_0(u) + \hat{\alpha}_1(u)\bar{X}_1 + \hat{\alpha}_2(u)\bar{X}_2 + \hat{\alpha}_3(u)\bar{X}_3$, 这里 \bar{X}_i 是 X_i 的均值.

接下来, 我们人为地移动第68和第34个反应变量的观测值(图5(a)中的“o”)来讨论局部Wcq回归的稳健性。具体来说, 我们把第68和第34个反应变量的观测值从(5.94, 5.07)移成更为极端的情形(7.94, 4.07), 然后使用局部Ls和Wcq来重新拟合这个数据集, 为了区分把它们分别记为RLs与RWcq。图5(b)给出了拟合曲线的形状。图4演示了系数函数的估计结果。从图中可以看出受这两个人造极端值的影响, 局部Ls发生了明显的变化, 而局部Wcq估计几乎没发生什么改变。因此, 实证结论表明局部Wcq比Ls回归更稳健。

此外, 我们也比较了本文的方法与局部分位数方法的区别, 如有需要请联系作者。

参 考 文 献

- [1] Bradic, J., Fan, J. and Wang, W., Penalized composite quasi-likelihood for ultrahigh dimensional variable selection, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(3) (2011), 325–349.
- [2] Cai, Z., Fan, J. and Li, R., Efficient estimation and inferences for varying-coefficient models, *Journal of the American Statistical Association*, **95**(451)(2000), 888–902.
- [3] Cai, Z. and Tiwari, R.C., Application of a local linear autoregressive model to BOD time series, *Environmetrics*, **11**(3)(2000), 341–350.
- [4] Cai, Z. and Xu, X., Nonparametric quantile estimations for dynamic smooth coefficient models, *Journal of the American Statistical Association*, **103**(484)(2008), 1595–1608.
- [5] Cederman, L.E. and Rao, M.P., Exploring the dynamics of the democratic peace, *Journal of Conflict Resolution*, **45**(6)(2001), 818–833.
- [6] Fan, J., Local linear regression smoothers and their minimax efficiencies, *The Annals of Statistics*, **21**(1)(1993), 196–216.
- [7] Fan, J. and Gijbels, I., Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, **20**(4)(1992), 2008–2036.
- [8] Fan, J. and Gijbels, I., *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall, 1996.
- [9] Fan, J. and Zhang, W., Statistical estimation in varying coefficient models, *The Annals of Statistics*, **27**(5)(1999), 1491–1518.
- [10] Fan, J. and Zhang, W., Statistical methods with varying coefficient models, *Statistics and Its Interface*, **1**(1)(2008), 179–195.
- [11] Jiang, J., Jiang, X. and Song, X., Weighted composite quantile regression estimation of DTARCH models, *The Econometrics Journal*, **17**(1)(2014), 1–23.
- [12] Hastie, T. and Tibshirani, R., Varying-coefficient models, *Journal of the Royal Statistical Society: Series B (Methodological)*, **55**(4)(1993), 757–796.
- [13] Honda, T., Quantile regression in varying coefficient models, *Journal of Statistical Planning and Inference*, **121**(1)(2004), 113–125.
- [14] Hong, Y. and Lee, T., Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models, *The Review of Economics and Statistics*, **86**(3)(2004), 840–840.
- [15] Huang, J., Wu, C. and Zhou, L., Polynomial spline estimation and inference for varying coefficient models with longitudinal data, *Statistica Sinica*, **14**(3)(2004), 763–788.

- [16] Kai, B., Li, R. and Zou, H., Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72(1)**(2010), 49–69.
- [17] Kim, M.O., Quantile regression with varying coefficients, *The Annals of Statistics*, **35(1)**(2007), 92–108.
- [18] Koenker, R. and Bassett, G., Regression quantiles, *Econometrica*, **46(1)**(1978), 33–50.
- [19] Koenker, R. and Park, B.J., An interior point algorithm for nonlinear quantile regression, *Journal of Econometrics*, **71(1-2)**(1996), 265–283.
- [20] Wang, L., Li, H. and Huang, J., Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements, *Journal of the American Statistical Association*, **103(484)**(2008), 1556–1569.
- [21] Zhang, W. and Lee, S., Variable bandwidth selection in varying-coefficient models, *Journal of Multivariate Analysis*, **74(1)**(2000), 116–134.
- [22] Zou, H. and Yuan, M., Composite quantile regression and the oracle model selection theory, *The Annals of Statistics*, **36(3)**(2008), 1108–1126.

Local Weighted Composite Quantile Estimating for Varying Coefficient Models

XIE QICHANG

(School of Economics, Shandong Institute of Business and Technology, Yantai, 264005)

LÜ XIUMEI

(School of Finance, Chongqing Technology and Business University, Chongqing, 400067)

A generalization of classical linear models is varying coefficient models, which offer a flexible approach to modeling nonlinearity between covariates. A method of local weighted composite quantile regression is suggested to estimate the coefficient functions. The local Bahadur representation of the local estimator is derived and the asymptotic normality of the resulting estimator is established. Comparing to the local least squares estimator, the asymptotic relative efficiency is examined for the local weighted composite quantile estimator. Both theoretical analysis and numerical simulations reveal that the local weighted composite quantile estimator can obtain more efficient than the local least squares estimator for various non-normal errors. In the normal error case, the local weighted composite quantile estimator is almost as efficient as the local least squares estimator. Monte Carlo results are consistent with our theoretical findings. An empirical application demonstrates the potential of the proposed method.

Keywords: Varying coefficient models, asymptotic normality, asymptotic relative efficiency, local weighted composite quantile estimating.

AMS Subject Classification: 62J08.