文章编号：0253-374X(2014)11-1750-05

DOI:10.11908/j.issn.0253-374x.2014.11.020

变系数模型的核权二次推断函数方法

李静茹，钱伟民

（同济大学 数学系，上海 200092）

摘要：利用二次推断函数的思想，以一系列基矩阵的线性组合来逼近工作相关矩阵，建立了纵向数据变系数模型的核权二次推断函数(QIF)，基于此得到函数系数的局部线性估计，并证明了估计的渐近性质. 实际中样本量和适合的窗宽并不能保证独立结构拟合的核估计总是最好的，通过适当扩大窗宽，在局部范围内引入样本相关性，既可以提高估计的效果，又不会造成“过拟合”的现象. 模拟中也给出了一种选取 QIF 窗宽的一种实用方法.

关键词：变系数模型；纵向数据；核；二次推断函数

中图分类号：O212.7　　**文献标志码**：A

Kernel Quadratic Inference Function Method for Varying-coefficient Model

LI Jingru, *QIAN Weimin*

(Department of Mathematics, Tongji University, Shanghai 200092, China)

Abstract: Following the idea of the quadratic inference function (QIF), a kernel quadratic function method for varying-coefficient model with longitudinal data was proposed by using local polynomial smoothing method and approximating the working correlation with a serious of basic matrices in the generalized estimation equation. The asymptotic normality of the estimators of the coefficient functions was proved. This method improved the performance of the estimators by widening the bandwidth in order to plug in the correlation within subjects to the local area, which won't lead to an "over fitting" phenomenon. An applied method was also proposed to choose QIF bandwidth in the simulation.

Key words: varying-coefficient model; longitudinal data; kernel; quadratic inference function

1　引言

非参数回归模型由于其形式自由、对数据的假定要求小、稳健性高等优点而越来越受到重视. 但常用的非参数估计方法在估计多元的非参数回归函数时需要大量数据，估计极不稳定，人们称这种现象为“维数祸根”. 对于高维数据近年来半参数回归分析受到广泛关注，其中变系数模型是一个研究的热点，它具有结构简单、容易解释、应用广泛等特点.

本文讨论变系数模型：

$$y_{ij}=\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}(t_{ij})+\varepsilon_{ij},\ i=1,\cdots,N,\ j=1,\cdots,n_i \tag{1}$$

式中：y_{ij}为一维响应变量；$\boldsymbol{x}_{ij}$为 p 维协变量；t_{ij}为一维协变量；$\boldsymbol{\beta}(\cdot)$为 p 维未知函数系数，$E(\varepsilon_{ij})=0$，$E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{\mathrm{T}})=\Sigma_i$，其中 $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1},\cdots\varepsilon_{in_i})^{\mathrm{T}}$，$\Sigma_i$ 未知，即同一个体的不同观测存在相关性，且相关结构未知.

变系数模型是一般线性模型的推广，由 Hastie 和 Tibshirani[1] 提出. 对于函数项系数的估计，主要方法有核估计最小二乘方法；光滑样条补偿最小二乘法(Wahba[2])；局部多项式法[3-5]等. Hoover 等[6] 将变系数模型推广到纵向数据的分析，给出了函数项系数的局部多项式估计.

在非参数纵向数据的局部模型中，考虑相关结构十分重要. Wang[7]、Lin 等[8]证明了使用真正的相关结构的核光滑样条方法要比使用独立结构得到的估计更有效. 但在实际中方差结构通常是未知的，经验上估计非构造的相关结构很困难，存在可能非正定、不可逆和冗余参数多等对估计至关重要的问题. 采用 Liang 等[9] 提出的使用工作相关矩阵的广义估计方程方法，又会产生大量的不必要的待估参数. Qu 等[10] 提出了二次推断函数方法，用一系列基矩阵的线性组合来逼近工作相关矩阵. 这一方法的好

收稿日期：2013-09-30

第一作者：李静茹(1979—)，女，讲师，博士生，主要研究方向为应用统计，E-mail:jingru_li@tongji.edu.cn

通讯作者：钱伟民(1962—)，男，理学博士，教授，博士生导师，主要研究方向为应用统计. E-mail:wmqian2003@aliyun.com

处是可以把线性组合中的系数视为冗余参数，不予理睬，而通过最小化二次推断函数，直接获得参数的估计. Qu 和 Li[11] 将这一思想应用于纵向数据变系数模型，通过惩罚样条方法得到函数系数的估计.

本文利用了二次推断函数的思想，使用局部多项式(一阶)光滑法建立了纵向数据变系数模型的核权二次推断函数，基于此得到函数系数的估计，并证明了估计的渐近性质. 在随机模拟中对核权二次推断函数估计与非构造协方差结构的最小二乘估计做了比较. Lin 和 Carroll[12] 指出，最渐近有效的核估计是在完全忽略样本相关性得到的. 但在实际中充分大的样本容量和趋向于 0 的窗宽都不易达到，模拟中会发现独立结构下的估计并不总是最好的. 由此可以给出了一种选取核权二次推断函数方法窗宽的一种方法，以及如何确定拟合的工作相关结构.

2　估计方法

利用 Taylor 展开 $\beta_k(t_{ij})=\beta_k(t_0)+\beta_k'(t_0)(t_{ij}-t_0)+o(h)$，模型(1)可化为

$$y_{ij}=\sum_{k=1}^{p}x_{ij}^k(\beta_k(t_0)+\beta_k'(t_0)(t_{ij}-t_0))+\varepsilon_{ij} \quad (2)$$

记 $\boldsymbol{Y}_i=(y_{i1},\cdots,y_{in_i})^{\mathrm{T}}$，$\boldsymbol{X}_i$ 为 $n_i\times 2p$ 矩阵，其第 j 行为 $(\boldsymbol{x}_{ij}^{\mathrm{T}},\boldsymbol{x}_{ij}^{\mathrm{T}}(t_{ij}-t_0))$，$\boldsymbol{\beta}=(\beta_1(t_0),\cdots,\beta_p(t_0),\beta_1'(t_0),\cdots,\beta_p'(t_0))^{\mathrm{T}}$，则式(2)可改写为

$$\boldsymbol{Y}_i=\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{\varepsilon}_i \quad (3)$$

利用 Wedderburn[13] 的拟似然方程(quasi-likelihood equation)及 Taylor 展开的局部性，考虑如下估计方程

$$\sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\boldsymbol{V}_i^{-1}\boldsymbol{K}_i^{1/2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})=0 \quad (4)$$

其中 $\boldsymbol{K}_i=\mathrm{diag}(K_h(t_{i1}-t_0),\cdots,K_h(t_{in_i}-t_0))$，$K_h(\cdot)=K(\cdot/h)/h$，$K(\cdot)$为核函数. $\boldsymbol{V}_i$ 为 $\boldsymbol{Y}_i$ 的协方差矩阵，实际中通常未知，基于样本方差的经验估计又往往是不可信的. Liang 等[9] 介绍的广义估计方程中建议使用 $\boldsymbol{A}_i^{1/2}\boldsymbol{R}\boldsymbol{A}_i^{1/2}$ 近似代替 $\boldsymbol{V}_i$，$\boldsymbol{A}_i$ 为 $\boldsymbol{Y}_i$ 的对角边缘方差矩阵，$\boldsymbol{R}$ 称为工作相关矩阵，含有少量冗杂函数. 在线性模型下，如果 $\boldsymbol{R}$ 被误指定，回归参数的估计仍然是相合的，只是有效性不够. 在此基础上 Qu 等[10] 提出一种二次推断函数的方法，可以有效提高参数的估计效率. 首先用一组基矩阵的线性组合来逼近 $\boldsymbol{R}^{-1}$，即令 $\boldsymbol{R}^{-1}\approx\sum_{r=1}^{m}\alpha_r\boldsymbol{M}_r$，$\boldsymbol{M}_r$ 为已知对称基矩阵，$\alpha_1,\cdots\alpha_m$ 为未知常数. 基矩阵族是一个足够充分的族，至少适合逼近常用的相关结构. 关于 $\boldsymbol{M}_r$ 的选取可进一步参见 Qu 等[10,14]. 把 $\boldsymbol{V}_i$ 和 $\boldsymbol{R}^{-1}$ 的近似形式代入估计方程(4)可得

$$\sum_{r=1}^{m}\sum_{i=1}^{N}\alpha_r\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\boldsymbol{A}_i^{-1/2}\boldsymbol{M}_r\boldsymbol{A}_i^{-1/2}\boldsymbol{K}_i^{1/2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\beta)=0 \quad (5)$$

估计方程(5)是下面估计函数向量的 $\boldsymbol{g}_N(\boldsymbol{\beta})$ 线性组合，

$$\boldsymbol{g}_N(\boldsymbol{\beta})=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}_i(\boldsymbol{\beta})=\frac{1}{N}\begin{pmatrix}\sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\boldsymbol{A}_i^{-1/2}\boldsymbol{M}_1\boldsymbol{A}_i^{-1/2}\boldsymbol{K}_i^{1/2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\ \vdots\\ \sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\boldsymbol{A}_i^{-1/2}\boldsymbol{M}_m\boldsymbol{A}_i^{-1/2}\boldsymbol{K}_i^{1/2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\end{pmatrix} \quad (6)$$

由于 $\boldsymbol{g}_N(\boldsymbol{\beta})$ 的维数为 $2pm$，大于待估参数的维数 $2p$，因此不能令每个估计函数都等于零来解 $\boldsymbol{\beta}$. 取而代之的是让 $\boldsymbol{g}_N(\boldsymbol{\beta})$ 中的元素尽可能地接近于 0，即通过最小化二次函数 $\hat{\boldsymbol{\beta}}=\arg\min_{\beta}\boldsymbol{g}_N\boldsymbol{\Omega}^{-1}\boldsymbol{g}_N$ 得到 $\boldsymbol{\beta}$ 的估计. 其中 $\boldsymbol{\Omega}=\mathrm{var}(\boldsymbol{g}_N)$，实际中 $\boldsymbol{\Omega}$ 通常是未知的，可以 $\boldsymbol{C}_N=\sum_{i=1}^{N}\boldsymbol{g}_i\boldsymbol{g}_i^{\mathrm{T}}/N^2$ 来估计，这样就可以写出基于样本的核权二次推断函数(QIF)，即 $Q_N(\boldsymbol{\beta})=\boldsymbol{g}_N^{\mathrm{T}}\boldsymbol{C}_N^{-1}\boldsymbol{g}_N$，此处 $\boldsymbol{C}_N$ 可逆的条件是 N 大于等于 $\boldsymbol{g}_i$ 的维数($2pm$). 最小化 $Q_N(\boldsymbol{\beta})$即可得到回归参数的估计：

$$\hat{\boldsymbol{\beta}}=\arg\min_{\beta}\boldsymbol{g}_N\boldsymbol{C}_N^{-1}\boldsymbol{g}_N. \quad (7)$$

利用二次推断函数的好处是可以把表达式(5)中的系数视为冗杂参数，不予理会，而直接估计回归参数. 由核权二次推断函数 $Q_N(\boldsymbol{\beta})$得到关于估计方程

$$\dot{\boldsymbol{Q}}_N(\boldsymbol{\beta})=2\dot{\boldsymbol{g}}_N^{\mathrm{T}}\boldsymbol{C}_N^{-1}\boldsymbol{g}_N-\boldsymbol{g}_N^{\mathrm{T}}\boldsymbol{C}_N^{-1}\dot{\boldsymbol{C}}_N\boldsymbol{C}_N^{-1}\boldsymbol{g}_N=0 \quad (8)$$

其中 $\dot{\boldsymbol{g}}_N=\partial\boldsymbol{g}_N/\partial\boldsymbol{\beta}$ 为 $2pm\times 2p$ 矩阵. $\dot{\boldsymbol{C}}_N$ 为三维数组 $(\partial\boldsymbol{C}_N/\partial\boldsymbol{\beta}_1,\cdots,\partial\boldsymbol{C}_N/\partial\boldsymbol{\beta}_{2p})$，利用 Newton-Raphson 方法解方程(8)，需 $Q_N(\boldsymbol{\beta})$关于 $\boldsymbol{\beta}$ 的二阶导数 $\ddot{\boldsymbol{Q}}_N(\boldsymbol{\beta})$，经直接计算易知，$\ddot{\boldsymbol{Q}}_N(\boldsymbol{\beta})$可以由 $2\dot{\boldsymbol{g}}_N^{\mathrm{T}}\boldsymbol{C}_N^{-1}\dot{\boldsymbol{g}}_N$ 来近似. 由此建立迭代关系：

$$\hat{\boldsymbol{\beta}}^{(j+1)}=\hat{\boldsymbol{\beta}}^{(j)}-\ddot{\boldsymbol{Q}}_N^{-1}(\hat{\boldsymbol{\beta}}^{(j)})\dot{\boldsymbol{Q}}_N(\hat{\boldsymbol{\beta}}^{(j)}) \quad (9)$$

3　渐近性质

假设以下条件成立：

条件 1：系数函数 $\beta_k(\cdot)$有连续二阶导数，$k=1,\cdots,p$.

条件 2：t_{ij}，$i=1,\cdots N$，$j=1,\cdots n_i$ 独立同分布，密

度函数 $f(t)$有连续一阶导数.

条件 3:$\boldsymbol{x}_{ij}, i=1,\cdots N, j=1,\cdots n_i$ 为有界设计点列,且有 $\sum_{i=1}^{N}\sum_{j=1}^{n_i}\|\boldsymbol{x}_{ij}\|^2/N=O(1)$.

条件 4:核函数 $K(\cdot)$为有界、对称、有有界紧支撑的概率密度函数.

条件 5:当 $N\to\infty$时,$h\to 0$,$Nh\to\infty$且 $Nh^5=O(1)$.

条件 5*:当 $N\to\infty$时,$h\to 0$,$Nh\to\infty$且 $Nh^5=o(1)$.

为表述方便,引入下列记号:记 $\boldsymbol{M}^{(r)}=\boldsymbol{A}_i^{-1/2}\boldsymbol{M}_r\boldsymbol{A}_i^{-1/2}, r=1,\cdots,m$; 表示矩阵 $\boldsymbol{M}^{(r)}$的第(k,l)元;$\Sigma_{i,(k,l)}$表示矩阵 $\boldsymbol{\Sigma}_i$ 的第(k,l)元;$\boldsymbol{H}=\mathrm{diag}(1,\cdots 1,h,\cdots h)$ 为由 p 个 1 和 p 个 h 构成的对角阵,$\boldsymbol{H}_{\mathrm{b}}=\mathrm{diag}(\boldsymbol{H},\cdots,\boldsymbol{H})$ 为由 m 个 $\boldsymbol{H}$ 构成的分块对角阵;$\mu_\lambda=\int t^\lambda K(t)\mathrm{d}t$,$\nu_\lambda=\int t^\lambda K^2(t)\mathrm{d}t$;$\boldsymbol{0}$ 表示所有元素均为 0 的向量;$\boldsymbol{\Omega}_1^{(r)}=\sum_{i=1}^{N}\sum_{l=1}^{n_i}M_{(l,l)}^{(r)}\boldsymbol{x}_{il}\boldsymbol{x}_{il}^{\mathrm{T}}/N$,$\boldsymbol{\Omega}_2^{(r,s)}=\sum_{i=1}^{N}\sum_{l=1}^{n_i}\Sigma_{i,(l,l)}M_{(l,l)}^{(r)}M_{(l,l)}^{s}\boldsymbol{x}_{il}\boldsymbol{x}_{il}^{\mathrm{T}}/N$;$\boldsymbol{\Omega}_2$ 为 $m\times m$ 块分块矩阵,每一块都有 $2p\times 2p$ 矩阵,其第(r,s)块矩阵为 $\begin{pmatrix}\nu_0 & \nu_1\\ \nu_2 & \nu_3\end{pmatrix}\otimes\boldsymbol{\Omega}_2^{(r,s)}$;$\boldsymbol{A}=f(t_0)\begin{pmatrix}\mu\otimes\Omega_1^{(0)}\\ \vdots\\ \mu\otimes\Omega_1^{(0)}\end{pmatrix}$,其中 $\mu=\begin{pmatrix}\mu_2\\ \mu_3\end{pmatrix}$.

定理 1 假设条件 1~5 成立,则有

$$\sqrt{Nh}(\boldsymbol{H}_{\mathrm{b}}^{-1}\boldsymbol{g}_N-h^2\boldsymbol{A}\boldsymbol{\beta}''(t_0))\xrightarrow{d}N(\boldsymbol{0},f(t_0)\boldsymbol{\Omega}_2),\tag{10}$$

其中 $\boldsymbol{\beta}''(t_0)=(\beta_1''(t_0),\cdots,\beta_p''(t_0)^{\mathrm{T}}$. 进一步地如果条件 5 换为条件 5*,式(10)可简化为

$$\sqrt{Nh}\boldsymbol{g}_N\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{C}_0),\tag{11}$$

其中 $\boldsymbol{C}_0=f(t_0)\boldsymbol{H}_{\mathrm{b}}\boldsymbol{\Omega}_2\boldsymbol{H}_{\mathrm{b}}$.

定理 2 假设条件 1~4,5* 成立,最小化核权二次推断函数 $Q_N(\beta)$得到的函数系数估计渐近服从正态分布,即

$$\sqrt{Nh}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{d}N(\boldsymbol{0},(\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{C}_0\boldsymbol{g}_0)^{-1}).\tag{12}$$

定理 3 假设条件 1~5 成立,最小化核权二次推断函数 $Q_N(\boldsymbol{\beta})$得到的多项式回归参数的估计存在,且有 $\hat{\boldsymbol{\beta}}\to\boldsymbol{\beta}$,a. s. ,$N\to\infty$.

定理证明略去.

综上,使用局部多项式拟合构造了核权二次推断函数,可以证明由此得到的函数系数具有渐近正态性和强相合性. 与 Qu 等[11]的使用惩罚样条方法所得到的的结论是类似的. 与样条方法相比,局部多项式拟合的计算速度略慢,但却具有同时估计系数函数及其各阶导数的优势. 样条方法通过控制节点数和惩罚系数来控制拟合的误差与光滑度,而局部多项式估计是选取合适的窗宽来完成这一任务的,两者均是非参数回归的主要方法. 同时局部多项式估计时线性估计类中的最佳估计,它具有几个吸引人的特点,例如它有好的最小最大性质,可适用于各种设计,如随机设计和固定设计等;它容易解释、实施并适应于导数的估计等.

4 随机模拟

数据生成模型类似于 Qu 等[11]中的例 2,具体如下:

$$y_{ij}=\beta_0(t_{ij})+\sum_{s=1}^{3}X_s^{(k)}(t_{ij})\beta_k(t_{ij})+\varepsilon_{ij},$$
$$i=1,\cdots,50,j=1,\cdots,n_i$$
$$\beta_0(t)=12+20\sin(t\pi/60),$$
$$\beta_1(t)=2-3\cos[(t-25)\pi/15],$$
$$\beta_2=6-0.2t,\beta_3(t)=-4(20-t)^3/1\,000.$$

(1) t 取{0,1,……,31}除 0 点外每个观测都有 60% 的可能缺失. 在不缺失的点上,t 的观测值为 $t+\Delta t$,$\Delta t\sim U(-0.5,0.5)$.

(2) $X_i^{(1)}\sim U(t/10,2+t/10)$,$X_i^{(2)}\sim N(0.1+X_i^{(1)}/2+X_i^{(1)})$,$X_i^{(3)}\sim B(1,0.6)$

(3) $\varepsilon_i\sim MV(0,\boldsymbol{\Sigma}_i)$,$\boldsymbol{\Sigma}_i$ 为对角线为 2,其余位置为 1.6 的 n_i 阶方阵.

最小化核权二次推断函数 $Q_N(\beta)$,利用迭代公式(9)计算 $\hat{\beta}$,取独立结构意义下广义核权最小二乘估计为初值,一般迭代 4 次左右即可达到收敛要求.

基矩阵选择如下:$\boldsymbol{M}_1$ 为单位阵,$\boldsymbol{M}_2$ 是除对角线为 0,其他元素为 1 的矩阵,$\boldsymbol{M}_3$ 是对角线两侧为 1,其他位置为 0 的矩阵,$\boldsymbol{M}_4$ 是第(1,1)元、(n_i,n_i)元为 1,其他为 0 的矩阵. 在 $Q_N(\boldsymbol{\beta})$中仅使用 $\boldsymbol{M}_1$,就相当于使用了独立结构的工作相关矩阵来进行拟合,其估计就等价于独立结构下广义核权最小二乘估计$(\hat{\beta}^{(0)})$;同时使用 $\boldsymbol{M}_1$、$\boldsymbol{M}_2$ 则相当于用等相关结构来拟合,此时的估计记为 $\hat{\beta}_{ex}$;同时使用 $\boldsymbol{M}_1$、$\boldsymbol{M}_3$、$\boldsymbol{M}_4$ 则相当于用一阶自回归结构拟合,其估计记为 $\hat{\beta}_{\mathrm{AR(1)}}$;而同时使用所有基矩阵,就相当于把以上各种结构

都考虑在内了，此处称之为综合结构，估计为 $\hat{\beta}_{\mathrm{QIF4}}$.

采用用平均绝对偏差 MADE，即 $M_{\mathrm{ADE}}=\sum_{j=0}^{30}\sum_{p=1}^{4}(31)^{-1}|\hat{\beta}_k(t_{ij})-\beta_k(t_{ij})|/\mathrm{range}(\beta_k)$ 来衡量估计的精确度，其中 range(·)表示函数的取值范围大小. 用比值 $R=M_{\mathrm{ADE},\hat{\beta}_I}/M_{\mathrm{ADE},\hat{\beta}_{II}}$ 来比较两个估计 $\hat{\beta}_I$ 和 $\hat{\beta}_{II}$ 的优劣. 令 $R_1=M_{\mathrm{ADE},\hat{\beta}_{(0)}}/M_{\mathrm{ADE},\hat{\beta}_V}$，$R_2=M_{\mathrm{ADE},\hat{\beta}_{ex}}/M_{\mathrm{ADE},\hat{\beta}_V}$，$R_3=M_{\mathrm{ADE},\hat{\beta}_{\mathrm{AR}(1)}}/M_{\mathrm{ADE},\hat{\beta}_V}$，$R_4=M_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF}_4}}/M_{\mathrm{ADE},\hat{\beta}_V}$. 图 1 给出了在固定窗宽下，分别使用以上 4 种相关结构，用 QIF 方法得到的估计与非构造协方差结构下的广义最小二乘估计 $\hat{\beta}_V$ 的 M_{ADE} 比值，其中

$$\hat{\beta}_V=\Big(\sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{K}_i^{1/2}\boldsymbol{X}_i\Big)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i^{1/2}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{K}_i^{1/2}\boldsymbol{Y}_i$$

$\hat{\boldsymbol{V}}_i=\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathrm{T}}$. 从图中可以看出，4 种结构下绝大多数的估计都明显好于 $\hat{\boldsymbol{\beta}}_V$. 对应图 1a—1d，每次模拟样本量 $N=100$.

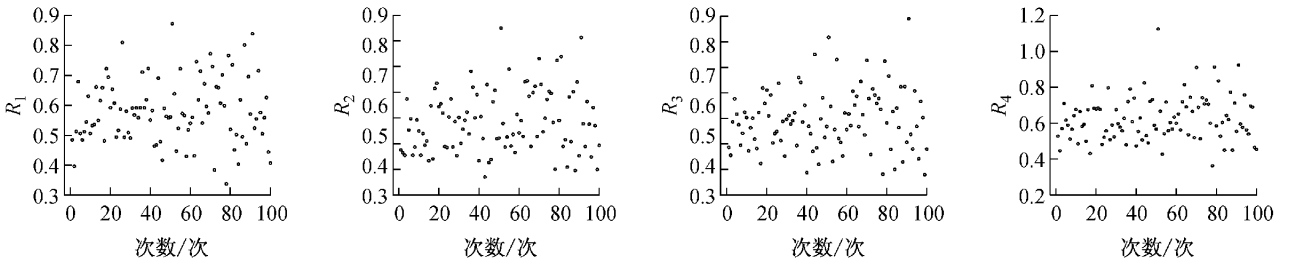

图 1　固定窗宽 $h=1.2$ 时，100 次模拟，$\hat{\boldsymbol{\beta}}^{(0)}$，$\hat{\boldsymbol{\beta}}_{ex}$，$\hat{\boldsymbol{\beta}}_{\mathrm{AR}(1)}$，$\hat{\boldsymbol{\beta}}_{\mathrm{QIF4}}$ 分别与 $\hat{\boldsymbol{\beta}}_V$ 的 $\mathrm{M}_{\mathrm{ADE}}$ 比值散点图

Fig. 1　Scatter diagrams for the MADE's ratios of $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\boldsymbol{\beta}}_{ex}$, $\hat{\boldsymbol{\beta}}_{AR(1)}$, $\hat{\boldsymbol{\beta}}_{\mathrm{QIF4}}$ to $\hat{\boldsymbol{\beta}}_V$ respectively in 100 simulations with fixed $h=1.2$

Lin 等[12]指出，最渐进有效的核估计是在完全忽略样本相关性得到的，被称为“工作独立”的方法，即在独立工作相关下得到的估计. 因此理论上 $\hat{\boldsymbol{\beta}}^{(0)}$ 是最渐进有效的. 但在实际中，样本量很难达到充分大，窗宽 h 的选取也不能太小，否则会出现“过拟合”现象. 表 1 分别给出了在不同样本量和不同 h 下，$\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}/\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}$ 的值. 可以发现，对每组数据总可以找到合适的 h 使得指定相关结构的估计好于独立结构的估计.

表 1　$\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{ex}}$、$\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}$ 分别与 $\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}$ 的比值

Tab. 1　The ratios of $\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{ex}}$, $\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}$ to $\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}$ respectively

R		h					
		0.4	0.8	1.2	1.6	2.0	2.4
$R=\frac{\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{ex}}}{\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}}$ $N=100$	1	1.078	0.98	1.00	1.04	1.00	0.98
	2	1.09	0.94	0.85	0.80	0.87	1.03
	3	1.25	1.05	1.06	0.96	0.92	0.96
	4	1.16	1.21	1.05	0.99	0.97	1.02
	5	0.99	0.99	0.89	0.98	1.11	1.25
$R=\frac{\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}}{\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}}$ $N=100$	1	1.10	0.99	0.94	0.99	1.10	1.21
	2	1.10	0.95	0.91	0.98	1.13	1.25
	3	0.95	0.90	0.93	0.96	1.06	1.15
	4	1.06	1.05	0.98	0.99	1.06	1.16
	5	1.06	0.93	0.99	1.13	1.30	1.44

实际应用中，h 的选择经常使用交叉验证的方法，即选择 h 使得 $C_{\mathrm{V}}=\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{-i}(t_{ij}))/Nn_i$ 达到最小，其中 $\beta^{-i}(\cdot)$ 为去掉第 i 个个体数据所得到的估计. 表 2 是在 50 个样本下，随机的 20 次试验中 $\hat{\boldsymbol{\beta}}_{\mathrm{QIF4}}$ 与 $\hat{\boldsymbol{\beta}}^{(0)}$ 的 $\mathrm{M}_{\mathrm{ADE}}$ 比值，在各自的估计中采用了由交叉验证(cross validation，CV)方法选择的最好 h. 从中可以发现，多数情况下使 C_{V} 最小的 QIF 窗宽 h_{QIF4} 大于等于独立结构下的最好窗宽 h_0，$\hat{\beta}_{\mathrm{QIF4}}$ 的表现好于 $\hat{\beta}^{(0)}$ 或者是效果相当. 直观

表 2　$R=\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}/\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}$ 及用 CV 方法选出的两种估计的最好 h

Tab. 2　$R=\mathrm{M}_{\mathrm{ADE},\hat{\beta}_{\mathrm{QIF4}}}/\mathrm{M}_{\mathrm{ADE},\hat{\beta}^{(0)}}$ and the best h chosen through CV method

序号	R	h_0	h_{QIF4}	序号	R	h_0	h_{QIF4}
1	1.03	1.53	1.80	11	0.86	1.53	1.80
2	1.01	1.53	1.62	12	1.11	1.35	1.80
3	0.93	1.53	1.80	13	1.00	1.62	1.71
4	0.89	1.35	1.35	14	0.89	1.44	1.71
5	1.28	1.62	1.53	15	0.92	1.62	1.80
6	0.93	1.53	1.80	16	1.04	1.80	1.53
7	0.95	1.44	1.80	17	1.01	1.53	1.62
8	1.20	0.99	0.90	18	0.90	1.44	1.80
9	1.17	1.80	1.71	19	1.03	1.62	1.44
10	1.15	1.26	1.35	20	0.97	1.62	1.80

上的解释是当 h 增大时,每个个体对估计有贡献的观测个数就会增加,而同一个体的不同观测存在相关性,此时采用 QIF 估计就会比较好. 这也给出了一个寻找 QIF 方法窗宽的启示. 由于 CV 方法计算耗费大,可以仅对相对简单的独立结构通过交叉验证的方法选取 h. 在此基础上,在一定范围内向上搜索,找到使 C_V 最小的 h,可使计算量减小一半.

众所周知,核估计中窗宽 h 的选择至关重要,h 选择过大,会使估计的偏差加大,h 选择过小虽然可以减少估计的偏差,但却增加了估计的方差,造成"过拟合"的现象. 通过以上模拟可以看到,使用核权二次推断函数方法,选择合适的 h,在局部范围内适当引入了数据间的相关结构,既可以保证估计的效果足够令人满意,也不会造成"过拟合"的结果.

参考文献:

[1] Hastie T J, Tibshirani R J. Varying-coefficient models (with discussion)[J]. Journal of the Royal Statistical Society Series B, 1993, 55(4): 757.

[2] Wahba G. Spline models for observational data [M]. Philadelphia: SIAM, 1990.

[3] Stone C J. Consistent nonparametric kernel regression[J]. Annals of Statistics, 1997, 5(4):595.

[4] Cleveland W S. Robust locally weighted regression and smoothing scatterplots[J]. Journal of the American Statistical Association, 1979, 74:828.

[5] Fan J. Design-adaptive nonparametric regression[J]. Journal of the American Statistical Association, 1992, 87:998.

[6] Hoover D R, Rice J A, Wu C O, *et al*. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data[J]. *Biomatrika*, 1998, 85:809

[7] Wang N. Marginal nonparametric kernel regression accounting for within-subjuct correlation[J]. *Biometrika*, 2003, 90:43.

[8] Lin X, Wang N, Welsh A H, et al. Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data[J]. *Biometrika*, 2004, 91:177.

[9] Liang K Y , Zegger S L. Longitudinal data analysis using generalized linear models[J]. *Biometika*, 1986, 73: 12.

[10] Qu A, Lindsay B G, Li B. Improving geralised estimating equations using quadratic inference functions[J]. *Biometrika*, 2000, 87: 823.

[11] Qu A, Li R. Quadratic inference function for varying-coefficient models with longitudinal data [J]. Biometrics, 2006, 62:379.

[12] Lin X, Carroll R J. Nonparametric function estimation for clustered data when the predictor is measured without/with error [J]. Journal of the American Statistical Association, 2000, 95:520.

[13] Wedderburn R W M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method[J]. *Biometrika*, 1974, 61:439.

[14] Qu A, Lindsay B G. Building adaptive estimating equations when inverse-of-covariance estimation is difficult[J]. Journal of the Royal Statistical Society Series B, 2003, 65:127.