线性约束下纵向数据部分线性模型的估计

冯彬娟,童 画,袁德美
重庆工商大学 数学与统计学院,重庆 400067

摘 要:目的 研究纵向数据部分线性模型的参数和未知回归函数的估计问题。**方法** 考虑在一些统计应用中,模型参数通常带有一定的约束,提出一种基于约束最小二乘与二次光滑局部线性估计的方法。该方法首先利用 profile 最小二乘法和 Lagrange 乘数法得到参数和回归函数的约束,即 profile 最小二乘估计量;再结合改进的二次光滑局部线性估计方法得到约束条件下模型的最终估计,并在一定正则条件下,证明了所构造的参数和回归函数估计量的渐近正态性;同时,通过数值模拟得到了有约束和无约束两种情况下参数分量的偏差、标准差和均方误差,并绘制了两种情况下回归函数的拟合曲线,验证了上述方法的有效性。**结果** 模拟结果表明:相对于不考虑约束条件的估计量,考虑约束条件的估计量具有更高的估计精度;回归函数的拟合曲线展现出了良好的拟合效果,进一步验证了所提出估计方法的有效性。**结论** 在实际研究中,通常可以获取参数分量的一些额外信息,充分利用这些信息能够提高估计的准确性;与无约束的估计方法相比,带有约束的估计方法能使估计的效率得到提高。

关键词:纵向数据;部分线性模型;约束估计;profile 最小二乘;二次光滑估计
中图分类号:O212.7 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2024.0006.011

Estimation of Partially Linear Models for Longitudinal Data under Linear Constraints

FENG Binjuan, TONG Hua, YUAN Demei
School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China

Abstract: Objective The estimation of parameters and unknown regression functions in partially linear models for longitudinal data was investigated. **Methods** Considering that model parameters often have certain constraints in some statistical applications, a method based on constrained least squares and quadratic smoothing local linear estimation was proposed. This method first utilized profile least squares and the Lagrange multiplier method to obtain constrained profile least squares estimators for parameters and regression functions. Then, combined with an improved quadratic smoothing local linear estimation method, the final estimation of the model under constraints was obtained. Under certain regularity conditions, the asymptotic normality of the constructed parameters and the estimators of the regression function was proved. Meanwhile, through numerical simulations, the biases, standard deviations, and mean square errors of parameter components under both constrained and unconstrained situations were obtained. The fitting curves of regression functions under both situations were plotted to verify the effectiveness of the proposed method. **Results** Simulation results showed that compared with estimators without considering constraints, estimators considering constraints had higher estimation accuracy. The fitting curves of regression functions demonstrated good fitting effects, further confirming the effectiveness of the proposed estimation method. **Conclusion** In practical research, additional information about parameter components

收稿日期:2023-08-06 **修回日期:**2023-09-21 **文章编号:**1672-058X(2024)06-0087-07
基金项目:重庆市自然科学基金(CSTB2022NSCQ-MSX1370).
作者简介:冯彬娟(1997—),女,四川广安人,硕士生,从事纵向数据分析与应用研究.
通讯作者:袁德美(1966—),男,四川南部人,教授,从事概率论极限理论、统计抽样分布渐近理论研究. Email: yuandemei@163.com.
引用格式:冯彬娟,童画,袁德美.线性约束下纵向数据部分线性模型的估计[J].重庆工商大学学报(自然科学版),2024,41(6):87-93.
FENG Binjuan, TONG Hua, YUAN Demei. Estimation of partially linear models for longitudinal data under linear constraints[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(6): 87-93.

can often be obtained, and fully utilizing this information can improve estimation accuracy. Compared with unconstrained estimation methods, constrained estimation methods can improve estimation efficiency.

Keywords: longitudinal data; partially linear models; constrained estimation; profile least squares; quadratic smoothing estimation

1 引 言

从20世纪80年代开始,半参数回归模型在研究和应用领域中变得越来越重要。最早由Engle等[1]提出部分线性模型,用来研究气候对电力需求的影响,这一模型结合了线性模型和非参数模型的优势,相较于非参数模型,在保持灵活性的同时还能提供更强的解释性。

近年来,纵向数据的相关研究已经成为统计学领域的热点内容之一。纵向数据,指的是对同一组个体进行反复多次观测,这些观测可以在不同的时间点或空间上进行。个体间的观测数据往往彼此独立,但同一个体的不同观测数据可能存在相关性。在对纵向数据进行统计建模的研究中,学者们通常关注两个方面:一方面是如何从众多变量中筛选出影响重要的变量,以增强模型的解释与预测能力,另一方面是如何使用更稳健的估计方法来估计模型中的参数。

关于部分线性模型在纵向数据中的应用,Zeger等[2]最早引了部分线性模型,应用于研究艾滋病病人体内所含有的CD4细胞浓度随时间变化的情况。此后,许多学者对该模型的估计方法进行了研究,其中包括Lin等[3]利用profile核估计方程研究该模型;He等[4]结合回归样条估计和M估计算法对模型进行估计,并证明了估计量的渐近性;Xue等[5]利用经验似然推断的方法研究了该模型;Fan等[6]则基于profile最小二乘法和局部多项式估计的新方法讨论了该模型的估计问题,并给出了模型变量选择的新方法;牟婷[7]在研究中使用Cholesky分解和Profile最小二乘估计方法提出了一种新的估计方法,并得到了估计的大样本性质。

值得注意的是,尽管经典局部线性回归估计方法在最小最大性质和适用性等方面具有优点,然而在部分线性模型中,该方法存在一些局限性。其中一个限制是它只能在目标点的局部区域内进行直线段的拟合,忽略了相当部分的信息,如未有效利用包括直线段斜率在内的其他信息。在此启发下,He等[8]在独立数据下提出了二次光滑局部线性估计方法;李生彪[9]在这一方法的基础上,利用二次光滑估计方法研究了纵向数据模型的估计问题,并对该方法的估计效果进行了验证;Prasangika等[10]将二次光滑估计推广到了时间序列数据的非参数回归模型中。

另外,在实际问题研究中,各种模型的参数往往具有先验信息,这些信息可以通过约束条件来表达,利用这些附加信息内容能够使参数估计达到更好的效果。现阶段已有许多文献资料研究了带有约束条件模型的估计问题,如王秀丽[11]在高维数据下研究了半参数回归模型在约束条件下的统计推断问题,证明了约束估计的相合性和渐近正态性;郭佳佳[12]在线性约束和随机约束下研究了带有测量误差的半参数回归模型的估计问题。

综上,学者们对纵向数据的研究一直在不断深入,但对于经典局部线性回归估计方法,在部分线性模型中的运用存在局限。为了解决这些问题,一些改进的方法被提出,如二次光滑估计,但目前的一些结果大都是基于独立数据[13-14],且未考虑先验信息的存在。本文结合已有研究以及现有理论基础,利用约束最小二乘与改进的二次光滑局部线性估计相结合的估计方法,研究该模型的约束估计问题。

假设有 n 个观测个体,第 i 个个体观测的次数记 m_i 次,记 t_{ij} 为第 i 个个体第 j 次观测时刻,则 y_{ij} 和 $\boldsymbol{x}_{ij}=(x_{ij1},\cdots,x_{ijp})$ 分别是第 i 个个体在时间 t_{ij} 的响应变量和协变量,则纵向数据可记为

$$\{(t_{ij},y_{ij},\boldsymbol{x}_{ij}^{\mathrm{T}}),1\leqslant i\leqslant n,1\leqslant j\leqslant m_i\}$$

本文考虑线性约束下纵向数据部分线性模型:

$$\begin{cases}y_{ij}=\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+g(t_{ij})+\varepsilon_{ij}\\ \boldsymbol{A}\boldsymbol{\beta}=\boldsymbol{d}\end{cases}\tag{1}$$

其中,$(\boldsymbol{x}_{ij}^{\mathrm{T}},t_{ij})\in R^p\times\mathbf{R}$ 是一些设计点列,既可以是固定的,也可以是随机的;随机误差 ε_{ij} 与 $(\boldsymbol{x}_{ij},t_{ij})$ 相互独立;$\boldsymbol{\beta}$ 是未知的 p 维参数向量;$g(\cdot)$ 是未知的回归函数;记 $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1},\varepsilon_{i2},\cdots,\varepsilon_{im_i})^{\mathrm{T}}$,$\{\boldsymbol{\varepsilon}_i\}$ 相互独立,且 $E(\varepsilon_{ij})=0$,$\mathrm{Cov}(\boldsymbol{\varepsilon}_i)=\sum_{m_i\times m_i}$,$\mathrm{Var}(\varepsilon_{ij})=\sigma^2$;$\boldsymbol{A}$ 为已知的 $q\times p$ 阶行满秩的常数矩阵;$\boldsymbol{d}$ 为已知的 q 维常数向量;$\{m_i\}$ 是有界正整数。

2 估计方法

2.1 profile最小二乘估计

由模型式(1)可知,在不考虑约束条件下有

$$y_{ij}=\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+g(t_{ij})+\varepsilon_{ij}\tag{2}$$

首先假定 $\boldsymbol{\beta}$ 已知,记 $y_{ij}^{*}=y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}$,则式(2)可改写成

$$y_{ij}^{*}=g(t_{ij})+\varepsilon_{ij}$$

转化为非参数回归模型,采用局部线性拟合方法估计

$g(t)$,在 t_0 的一个小领域内点 t 处进行泰勒公式展开得

$$g(t)\approx g(t_0)+g'(t_0)(t-t_0)\equiv a+b(t-t_0)$$

并极小化下式可得到未知函数 $g(\cdot)$ 在 t_0 的初始估计 $\breve{g}(t_0)$。

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}[y_{ij}^{*}-a-b(t_{ij}-t_0)]^2\omega(t_{ij})K_h(t_{ij}-t_0)$$

其中,$K_h(\cdot)$ 是核函数,$K_h(\cdot)=h^{-1}K(\cdot/h)$,$\omega(t_{ij})$ 是权函数,h 是窗宽。

为了方便表示,本文用矩阵符号。考虑平衡纵向数据情形,设 $m_1=m_2=\cdots=m_i=m$,其中 $i=1,2,\cdots,n$。记

$$\boldsymbol{H}_{t_0}=\begin{pmatrix}1 & \cdots & 1 & \cdots\\ h^{-1}(t_{11}-t_0) & \cdots & h^{-1}(t_{1m}-t_0) & \cdots h^{-1}(t_{nm}-t_0)\end{pmatrix}^{\mathrm{T}}$$

$$\boldsymbol{X}=(\boldsymbol{x}_1^{\mathrm{T}},\cdots,\boldsymbol{x}_n^{\mathrm{T}})^{\mathrm{T}},\boldsymbol{x}_i=(x_{i1},\cdots,x_{im})^{\mathrm{T}}$$

$$\boldsymbol{Y}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_n^{\mathrm{T}})^{\mathrm{T}},\boldsymbol{y}_i=(y_{i1},\cdots,y_{im})^{\mathrm{T}}$$

$$\boldsymbol{\varepsilon}=(\boldsymbol{\varepsilon}_1^{\mathrm{T}},\cdots,\boldsymbol{\varepsilon}_n^{\mathrm{T}})^{\mathrm{T}},\boldsymbol{\varepsilon}_i=(\varepsilon_{i1},\cdots,\varepsilon_{im})^{\mathrm{T}}$$

$$\boldsymbol{W}=\mathrm{diag}(\omega(t_{11}),\cdots,\omega(t_{1m}),\omega(t_{21})\cdots\omega(t_{nm}))$$

$$\boldsymbol{W}_{t_0}=\mathrm{diag}(\omega(t_{11})K_h(t_{11}-t_0),\cdots,\omega(t_{nm})K_h(t_{nm}-t_0))$$

其中,$\boldsymbol{H}_{t_0}$ 是 $nm\times 2$ 矩阵,$\boldsymbol{W}_{t_0}$ 是 $nm\times nm$ 矩阵。因此回归函数 $g(\cdot)$ 的估计为

$$\breve{g}(t_0)=(1,0)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \quad (3)$$

记 $S(t_0)=(1,0)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}$,再令 $\boldsymbol{G}=(\boldsymbol{g}_1^{\mathrm{T}},\cdots,\boldsymbol{g}_n^{\mathrm{T}})^{\mathrm{T}},\boldsymbol{g}_i=(g(t_{i1}),\cdots,g(t_{im}))^{\mathrm{T}}$,则式(2)表示为 $\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{G}+\boldsymbol{\varepsilon}$,则 $\boldsymbol{G}$ 的初始估计为

$$\breve{\boldsymbol{G}}=\boldsymbol{S}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})$$

其中,$\boldsymbol{S}$ 是 $nm\times nm$ 光滑矩阵,它仅依赖于观测值 t_{ij},可以根据已有的文献指定。在本文中结合局部线性回归估计,可取光滑矩阵 $\boldsymbol{S}=(S(t_{11}),\cdots,S(t_{1m}),S(t_{21}),\cdots,S(t_{nm}))^{\mathrm{T}}$,即

$$\boldsymbol{S}=\begin{pmatrix}(1,0)(\boldsymbol{H}_{t_{11}}^{\mathrm{T}}\boldsymbol{W}_{t_{11}}\boldsymbol{H}_{t_{11}})^{-1}\boldsymbol{H}_{t_{11}}^{\mathrm{T}}\boldsymbol{W}_{t_{11}}\\ \vdots\\ (1,0)(\boldsymbol{H}_{t_{1m}}^{\mathrm{T}}\boldsymbol{W}_{t_{1m}}\boldsymbol{H}_{t_{1m}})^{-1}\boldsymbol{H}_{t_{1m}}^{\mathrm{T}}\boldsymbol{W}_{t_{1m}}\\ (1,0)(\boldsymbol{H}_{t_{21}}^{\mathrm{T}}\boldsymbol{W}_{t_{21}}\boldsymbol{H}_{t_{21}})^{-1}\boldsymbol{H}_{t_{21}}^{\mathrm{T}}\boldsymbol{W}_{t_{21}}\\ \vdots\\ (1,0)(\boldsymbol{H}_{t_{nm}}^{\mathrm{T}}\boldsymbol{W}_{t_{nm}}\boldsymbol{H}_{t_{nm}})^{-1}\boldsymbol{H}_{t_{nm}}^{\mathrm{T}}\boldsymbol{W}_{t_{nm}}\end{pmatrix}$$

将 $\breve{\boldsymbol{G}}$ 代入模型 $\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{G}+\boldsymbol{\varepsilon}$ 中得到

$$(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}=(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad (4)$$

对于模型式(4),利用 profile 最小二乘法,可得到参数 $\boldsymbol{\beta}$ 的估计为

$$\hat{\boldsymbol{\beta}}=[\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{S})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{X}]^{-1}\times\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{S})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y} \quad (5)$$

记 $\tilde{\boldsymbol{X}}=(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{X}$,$\tilde{\boldsymbol{Y}}=(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}$,式(5)可表示为 $\hat{\boldsymbol{\beta}}=(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{Y}}$,将式(3)的 $\boldsymbol{\beta}$ 替换成 $\hat{\boldsymbol{\beta}}$,则可以得到回归系数 $g(\cdot)$ 的估计为

$$\hat{g}(t_0)=(1,0)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \quad (6)$$

结合 profile 最小二乘的理论以及相关计算可得到未知函数 $g(\cdot)$ 的一阶导数 $g^{(1)}(\cdot)$ 在 t_0 处的估计为

$$\hat{g}^{(1)}(t_0)=\left(0,\frac{1}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \quad (7)$$

在此基础上,已经获得了参数和未知函数的无约束估计量,下面考虑在约束条件下对模型参数进行估计的问题。

2.2　约束条件下模型的局部线性估计

对于在线性约束 $\boldsymbol{A}\boldsymbol{\beta}=\boldsymbol{d}$ 下模型的估计,利用 Lagrange 乘子法构造如下辅助函数:

$$F(\boldsymbol{\beta},\boldsymbol{\lambda})=\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}-\breve{g}(t_{ij}))^2+2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{d}) \quad (8)$$

对式(8)求偏导,并令导数为 0 可得到:

$$\begin{cases}\dfrac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}}=-2\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{Y}}+2\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}}\boldsymbol{\beta}+2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}=0\\ \dfrac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}=2(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{d})=0\end{cases} \quad (9)$$

根据式(9)计算可得

$$\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}-(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}$$

因为 $\boldsymbol{A}\boldsymbol{\beta}=\boldsymbol{d}$,所以

$$\boldsymbol{d}=\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{A}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}$$

假设 $\boldsymbol{A}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}$ 的逆矩阵存在,则 $\boldsymbol{\lambda}$ 的估计为

$$\hat{\boldsymbol{\lambda}}=[\boldsymbol{A}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{d})$$

则约束条件 $\boldsymbol{A}\boldsymbol{\beta}=\boldsymbol{d}$ 下 $\boldsymbol{\beta}$ 的最小二乘估计为

$$\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{A}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{A}^{\mathrm{T}}]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{d}) \quad (10)$$

设 $\boldsymbol{F}=\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\tilde{\boldsymbol{X}}$,式(10)可以表示为

$$\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{d}) \quad (11)$$

通过将式(6)和式(7)的 $\hat{\boldsymbol{\beta}}$ 替换成 $\tilde{\boldsymbol{\beta}}$,可得到约束条件下回归函数 $g(\cdot)$ 和其一阶导数 $g^{(1)}(\cdot)$ 的估计:

$$\hat{g}_r(t_0)=(1,0)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}(\boldsymbol{Y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})$$

$$\hat{g}_r^{(1)}(t_0)=\left(0,\frac{1}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}(\boldsymbol{Y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})$$

2.3 约束条件下模型的二次光滑估计

结合目标点 t_0 处局部拟合直线截距值 $\hat{g}_r(t_0)$ 和其导数估计值 $\hat{g}_r^{(1)}(t_0)$,可得到模型式(1)在目标点处的二次光滑估计值,参照文献[14]可得的 $g(\cdot)$ 的最终估计:

$$\tilde{g}(t_0)=\int_Q[\hat{g}_r(u)+\hat{g}_r^{(1)}(u)(t_0-u)]K_h(t_0-u)\,\mathrm{d}u=\int_Q\left[\left(1,\frac{t_0-u}{h}\right)(\boldsymbol{H}_u^{\mathrm{T}}\boldsymbol{W}_u\boldsymbol{H}_u)^{-1}\boldsymbol{H}_u^{\mathrm{T}}\boldsymbol{W}_u\times(\boldsymbol{Y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})\right]K_h(t_0-u)\,\mathrm{d}u \tag{12}$$

He 等[8]详细说明了改进的二次光滑局部线性回归估计方法具有边界效应。因此,本文仅讨论纵向数据下内点的情况。设观测的时间区间为 $[0,T]$,目标点 $t_0\in[2h,T-2h]$,则边界区间为 $[0,2h)$ 和 $(T-2h,T]$。

3 估计渐近性质

引入以下记号:$\boldsymbol{A}\otimes\boldsymbol{B}$ 表示矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的 Kronecker 乘积,$\boldsymbol{A}^{\otimes 2}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$。记 $\mu_l=\int u^lK(u)\,\mathrm{d}u$,$\nu_l=\int u^lK^2(u)\,\mathrm{d}u$,$l=0,1,2$,$\boldsymbol{\Gamma}=\begin{pmatrix}1&\mu_1\\\mu_1&\mu_2\end{pmatrix}$,$\boldsymbol{\Gamma}^*=\begin{pmatrix}1&\nu_1\\\nu_1&\nu_2\end{pmatrix}$,$S_n(t_0)=\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0}$。

为研究估计量的渐近性质,沿用文献[9]中的假设条件:

A1 核函数 $K(\cdot)$ 是一连续的概率密度函数,具有有界支撑,为了简化计算,可假设核函数具有对称性,即 $K(x)=K(-x)$,权函数 $\omega(\cdot)$ 为随机的连续函数。

A2 $n\to\infty$,$nh^8\to 0$ 且 $nh^2/(\log n)^2\to\infty$。

A3 回归函数 $g(\cdot)$ 在内点处存在有界的四阶连续导数。

A4 $t_{ij}(1\leqslant i\leqslant n,1\leqslant j\leqslant m)$ 有连续密度 $f(t)$,并令 $\nu(t_0)=\mathrm{E}[\boldsymbol{x}^{\mathrm{T}}(t)\mid t=t_0]$。

A5 $E\int_0^\infty[x-\nu(t)]^{\otimes 2}\omega(t)f(t)\,\mathrm{d}t$ 非奇异。

这些假设条件是非参数统计研究中常用的条件。考虑到核函数是对称的,则显然 $\mu_0=1$,对于奇数 l,有 $\mu_1=\nu_1=0$。令

$$K_1(z)=\int K(z-u)K(u)\,\mathrm{d}u$$

$$K_2(z)=\int u(z-u)K(z-u)K(u)\,\mathrm{d}u$$

$$V(K)=\int[K_1(z)-K_2(z)/\mu_2]^2\mathrm{d}z \tag{13}$$

定理1 在假设 A1—A5 成立的条件下,$\tilde{\boldsymbol{\beta}}$ 是渐近正态性的,即

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{L}N(0,\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}^{\mathrm{T}})$$

其中,

$$\boldsymbol{C}=(\boldsymbol{I}-\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A})\boldsymbol{D}^{-1}$$

$$\boldsymbol{D}=E\int_0^\infty[x(t)-v(t)]^{\otimes 2}\omega(t)f(t)\,\mathrm{d}t$$

$$\boldsymbol{B}=E\left\{\int_0^\infty[x(t)-v(t)]\boldsymbol{\varepsilon}\omega(t)f(t)\,\mathrm{d}t\right\}^{\otimes 2}$$

定理2 在假设 A1—A5 成立的条件下,$\tilde{g}(\cdot)$ 是渐近正态性的,即

$$\sqrt{nh}\left[\tilde{g}(t_0)-g(t_0)-\frac{1}{4}g^{(4)}(t_0)(\mu_2^2-\mu_4)h^4\right]\xrightarrow{L}N(0,\sum(t_0))$$

其中,$\sum(t_0)=[f(t_0)]^{-1}\sigma^2V(K)$。

4 定理证明

给出下面几个引理。

引理1 在假设 A1—A5 成立的条件下,有

$$\boldsymbol{S}_n(t_0)=nf(t_0)\omega(t_0)\otimes\boldsymbol{\Gamma}[1+o_p(1)]$$

证明 由矩阵计算可得

$$\boldsymbol{S}_n(t_0)=\begin{pmatrix}S_{n,0}&S_{n,1}\\S_{n,1}&S_{n,2}\end{pmatrix},l=0,1,2$$

其中,

$$S_{n,l}=\sum_{i=1}^n\sum_{j=1}^m\omega(t_{ij})K_h(t_{ij}-t_0)\left(\frac{t_{ij}-t_0}{h}\right)^l$$

当 $h\to 0$,$nh\to\infty$ 时,有

$$ES_{n,l}=nE\left[\int_0^\infty\omega(t)K_h(t-t_0)\left(\frac{t-t_0}{h}\right)^lf(t_0)\,\mathrm{d}t\right]=n\omega(t_0)f(t_0)\mu_l[1+o_p(1)] \tag{14}$$

将式(14)代入 $S_{n,l}=ES_{n,l}+O_p(\sqrt{\mathrm{Var}(S_{n,l})})$,即可得引理1成立。

引理2 在假设 A1—A5 成立的条件下,有

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{X}=nf(t_0)\omega(t_0)v(t_0)\otimes(1,0)^{\mathrm{T}}[1+o_p(1)] \tag{15}$$

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}G=nf(t_0)\omega(t_0)g(t_0)\otimes(1,0)^{\mathrm{T}}[1+o_p(1)]$$

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{\varepsilon}=(1,0)^{\mathrm{T}}o_p(1) \tag{16}$$

证明 仅证明式(15),其余类似证明,由于

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{X}=\begin{pmatrix}\sum_{i=1}^n\sum_{j=1}^m\omega(t_{ij})K_h(t_{ij}-t_0)\boldsymbol{x}_{ij}^{\mathrm{T}}\\\sum_{i=1}^n\sum_{j=1}^m\omega(t_{ij})K_h(t_{ij}-t_0)\boldsymbol{x}_{ij}^{\mathrm{T}}\left(\frac{t_{ij}-t_0}{h}\right)\end{pmatrix}$$

类似于 $S_{n,l}$,可求得

$$\sum_{i=1}^n\sum_{j=1}^m\omega(t_{ij})K_h(t_{ij}-t_0)\boldsymbol{x}_{ij}^{\mathrm{T}}\left(\frac{t_{ij}-t_0}{h}\right)^l=n\omega(t_0)f(t_0)v(t_0)\mu_l[1+o_p(1)],l=0,1,2$$

则有

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{X}=\begin{pmatrix} nf(t_0)\omega(t_0)v(t_0)\mu_0[1+o_p(1)] \\ nf(t_0)\omega(t_0)v(t_0)\mu_1[1+o_p(1)] \end{pmatrix}$$

又因为对称密度核函数 $K(\cdot)$ 下,有 $\mu_0=1,\mu_1=0$,将 $\mu_0=1,\mu_1=0$ 代入上式,可得到式(15)成立。

引理 3　在假设 A1—A5 成立的条件下,有

$$\frac{1}{n}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}}\xrightarrow{P}\boldsymbol{D} \tag{17}$$

$$\frac{1}{n}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{G}=o_p(1) \tag{18}$$

证明　仅证明式(17),其他类似可证。由引理 1 和引理 2 的式(15)得到

$$(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{X}=v(t)\otimes(1,0)^{\mathrm{T}}[1+o_p(1)]$$

从而有

$$\boldsymbol{S}\boldsymbol{X}=(v(t_{11}),\cdots,v(t_{1m}),v(t_{21}),\cdots,v(t_{nm}))^{\mathrm{T}}[1+o_p(1)]$$

进而有

$$\frac{1}{n}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}}=$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\omega(t_{ij})[\boldsymbol{x}_{ij}-v(t_{ij})]^{\otimes 2}[1+o_p(1)]=$$

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}[\boldsymbol{x}_i-v(t)]^{\otimes 2}\omega_i(t)f(t)\,\mathrm{d}t[1+o_p(1)]$$

其中,$\boldsymbol{D}=\sum_{i=1}^{n}\int_0^{\infty}[\boldsymbol{x}_i-v(t)]^{\otimes 2}\omega_i(t)f(t)\,\mathrm{d}t$。由大数定理可得式(17)成立。

引理 4　在假设 A1—A5 成立的条件下,有

$$\mathrm{Bias}(\tilde{g}(t_0))=\frac{1}{4}g^{(4)}(t_0)(\mu_2^2-\mu_4)h^4+o_p(h^4)$$

$$\mathrm{Var}(\tilde{g}(t_0))=[nhf(t_0)]^{-1}\sigma^2V(K)[1+o_p(1)]$$

证明　类似文献[9]定理 1、定理 2 的证明,这里省略。

4.1　定理 1 证明

由式(11)知

$$\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}-\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{d})-\boldsymbol{\beta}=$$

$$(\boldsymbol{I}-\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$

所以

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})=(\boldsymbol{I}-\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A})(\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))$$

又

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=\sqrt{n}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})(\boldsymbol{G}+\boldsymbol{\varepsilon})=$$

$$\sqrt{n}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{G}+$$

$$\sqrt{n}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\varepsilon}=B_1+B_2 \tag{20}$$

由引理 3 的式(18)可知

$$B_1=o_p(1)$$

由引理 2 的式(16)可知

$$(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}[1+o_p(1)] \tag{21}$$

由式(17)和式(21)可得

$$\boldsymbol{B}_2=\frac{1}{\sqrt{n}}\boldsymbol{D}^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\varepsilon}[1+o_p(1)]$$

由于

$$\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\varepsilon}=\sum_{i=1}^{n}\sum_{j=1}^{m}[\boldsymbol{x}_{ij}-v(t_{ij})]\omega_{ij}\varepsilon_{ij}[1+o_p(1)]=$$

$$\sum_{i=1}^{n}\int_0^{\infty}[\boldsymbol{x}_i-v(t)]\omega_i(t)\boldsymbol{\varepsilon}_i f(t)\,\mathrm{d}t[1+o_p(1)]$$

由 Slutsky 定理和中心极限定理得

$$\frac{1}{\sqrt{n}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\varepsilon}\xrightarrow{L}N(0,\boldsymbol{B})$$

其中,$\boldsymbol{B}=E\left\{\int_0^{\infty}[x(t)-v(t)]\boldsymbol{\varepsilon}\omega(t)f(t)\,\mathrm{d}t\right\}^{\otimes 2}$。故有

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{L}N(0,\boldsymbol{D}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}) \tag{22}$$

注意到 $\boldsymbol{D}^{-1}$ 的对称性,由式(19)和式(22)完成定理 1 的证明。

4.2　定理 2 证明

将式(12)的 u 换成 t,有

$$\tilde{g}(t_0)=\int_Q\left(1,\frac{t_0-t}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\times$$

$$(\boldsymbol{Y}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})K_h(t_0-t)\,\mathrm{d}t=$$

$$\int_Q\left(1,\frac{t_0-t}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\times$$

$$[\boldsymbol{G}+\boldsymbol{X}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})+\boldsymbol{\varepsilon}]K_h(t_0-t)\,\mathrm{d}t$$

其中,$\int_Q$ 是定积分,Q 是变量 t 的允许范围。从而

$$E(\tilde{g}(t_0))=\int_Q\left(1,\frac{t_0-t}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\times$$

$$\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}[\boldsymbol{G}+\boldsymbol{X}\boldsymbol{E}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})]K_h(t_0-t)\,\mathrm{d}t$$

由引理 1、引理 2 的式(15)和定理 1 可得到 $\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\|=O_p(n^{-\frac{1}{2}})$,进而有

$$\tilde{g}(t_0)-E(\tilde{g}(t_0))=\int_Q\left(1,\frac{t_0-t}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\times$$

$$\boldsymbol{\varepsilon}K_h(t_0-t)\,\mathrm{d}t+O_p(n^{-\frac{1}{2}})$$

因为

$$M=\int_Q\left(1,\frac{t_0-t}{h}\right)(\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\boldsymbol{H}_{t_0})^{-1}\boldsymbol{H}_{t_0}^{\mathrm{T}}\boldsymbol{W}_{t_0}\times\boldsymbol{\varepsilon}K_h(t_0-t)\,\mathrm{d}t$$

则有

$$\tilde{g}(t_0)-g(t_0)=\tilde{g}(t_0)-E(\tilde{g}(t_0))+E(\tilde{g}(t_0))-g(t_0)=$$

$$M+O_p(n^{-\frac{1}{2}})+\mathrm{Bias}(\tilde{g}(t_0)) \tag{23}$$

令 $u=\frac{t-t_0}{h}, z=\frac{s-t_0}{h}, z_{ij}=\frac{t_{ij}-t_0}{h}$,经过计算可得

$$M=\frac{1}{nf(t_0)}\int_Q\left(1,\frac{t_0-t}{\mu_2 h}\right)\boldsymbol{U}\boldsymbol{\varepsilon}K_h(t_0-t)\,\mathrm{d}t=$$
$$\frac{1}{nf(t_0)}\sum_{i=1}^{n}\sum_{j=1}^{m}\int_Q\left[K_h(t_{ij}-t)+\frac{1}{\mu_2}\frac{(t_0-t)}{h}\frac{(t_{ij}-t)}{h}K_h(t_{ij}-t)\right]\varepsilon_{ij}K_h(t_0-t)\,\mathrm{d}t=$$
$$\frac{1}{nf(t_0)}\sum_{i=1}^{n}\sum_{j=1}^{m}\int_Q\left[\frac{1}{h}K(z_{ij}-u)-\frac{1}{\mu_2 h}u(z_{ij}-u)K(z_{ij}-u)\right]\varepsilon_{ij}K(u)\,\mathrm{d}u$$

其中,

$$\boldsymbol{U}=\begin{pmatrix}K_h(t_{11}-t) & \cdots & K_h(t_{nm}-t)\\ \frac{t_{11}-t}{h}K_h(t_{11}-t) & \cdots & \frac{t_{nm}-t}{h}K_h(t_{nm}-t)\end{pmatrix}$$

由式(13)、Slutsky 定理和中心极限定理可得 M 的渐近分布为 $N(0,(nhf(t_0))^{-1}\sigma^2V(K))$,进一步由引理 4 及式(23)完成定理 2 的证明。

5 数值模拟

为了进一步说明所提出估计方法在有限样本下的表现情况,本节使用 R 软件进行数值模拟。考虑以下模型:

$$y_{ij}=x_{ij1}\beta_1+x_{ij2}\beta_2+g(t_{ij})+\varepsilon_{ij}$$

其中,$\beta_1=1,\beta_2=4,x_{ij1}\sim N(0,1),x_{ij2}\sim N(0,1),g(t_{ij})=t+2\exp(\cos(2t))$,$t_{ij}$ 服从[0,1]的均匀分布,$m_i=m=3$,即所有个体的观测次数相同。同一个个体观测是相关的,随机误差 ε_{ij} 服从 $N(0,\sigma^2)$,$\sigma^2=0.5$,且组内相关系数 $\rho_{\varepsilon_{ij},\varepsilon_{ik}}=\rho=0.6(j\neq k)$,其协方差矩阵由相关系数 ρ 和 σ^2 决定。核函数选择标准的 Gaussian 核函数 $K(u)=\frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$,窗宽的选取采用交叉验证法。另外假定模型的约束条件为 $\beta_1+\beta_2=5$。

当模拟的个体数目为 30、50、100,模拟次数为 1 000 次时,分别将忽略约束条件的估计 $\hat{\beta}$ 和考虑约束条件的估计 $\tilde{\beta}$ 的估计偏差 R_{Bias}、标准差 R_{SD} 以及均方误差 R_{MSE} 记录在表 1。当样本容量为 $n=100$ 时,不带约束条件和带约束条件的 $g(t_{ij})$ 拟合曲线分别如图 1 和图 2 所示。

表 1 忽略约束条件和考虑约束条件下参数分量的估计结果

Table 1 Estimation results for parametric components with constraints ignored and constraints considered

β	n	$\hat{\beta}$			$\tilde{\beta}$		
		R_{Bias}	R_{SD}	R_{MSE}	R_{Bias}	R_{SD}	R_{MSE}
1	30	0.006 39	0.083 73	0.007 82	0.005 37	0.079 12	0.006 85
	50	0.005 41	0.072 10	0.006 04	0.004 74	0.067 35	0.005 27
	100	0.002 79	0.055 82	0.003 94	0.002 57	0.050 39	0.003 19
4	30	0.007 75	0.088 21	0.008 52	0.005 98	0.077 13	0.006 74
	50	0.005 81	0.079 23	0.007 02	0.003 92	0.066 92	0.005 23
	100	0.003 38	0.059 04	0.004 09	0.001 73	0.044 37	0.002 96

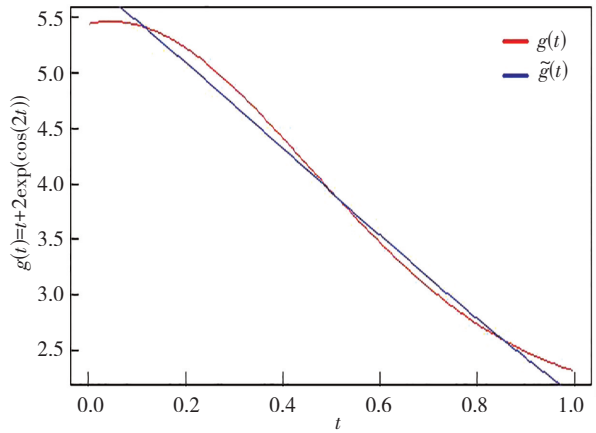

图 1 不带约束条件下 $g(t_{ij})$ 的拟合曲线图

Fig. 1 Fitted curve of $g(t_{ij})$ with constraints

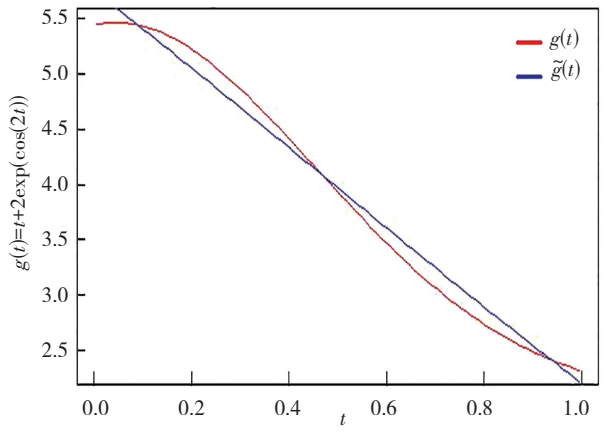

图 2 带约束条件下 $g(t_{ij})$ 的拟合曲线图

Fig. 2 Fitted curve of $g(t_{ij})$ with constraints

表 1 的结果显示:随着模拟中样本容量的增加,两种情况下的效果都逐渐改善,表明它们都具有渐近无偏性。此外,当样本容量相同时,从偏差、标准差和均方误差这 3 个指标可以看出:带有约束条件的参数估计结果的偏差、标准差和均方误差都小于不带约束条件参数的估计结果,说明带有约束条件的参数估计在精度上表现更好,提高了估计效果。

从图 1 和图 2 可以看出:回归函数估计结果的拟合效果较好,表明本文提出的估计方法是有效的。同时,通过比较图 1 和图 2 可以发现,在相同的样本容量下,带有约束条件的回归函数拟合效果优于没有约束条件的回归函数拟合效果。因此,约束估计的表现优于无约束估计。如果回归参数向量具有某些先验信息,可以利用这些先验信息来提高参数估计的有效性。

6　结论与展望

针对线性约束下纵向数据部分线性模型的估计问题,对其先采用 profile 最小二乘和 Lagrange 乘数法得到了模型式(1)中参数和回归函数的约束 profile 最小二乘估计量;再结合改进的二次光滑估计思想给出了最终的估计,并在适当的正则条件下得到了新估计的渐近正态性,详细地给出了重要定理 1 和定理 2 的证明;最后通过 R 软件进行了数值模拟研究,分析了带约束条件下估计量相对于无约束条件下估计量的优势。

近年来,纵向数据研究工作不断深入,理论和应用研究都有所进展。然而,仍有一些问题需要进一步研究。其中有两个值得关注的研究方向:第一,纵向数据建模分析过程中的变量选择问题。变量选择对于模型的解释能力和估计精度具有重要影响,因此需要仔细选择适合的变量,以建立一个稳健的模型。第二,本文只研究了线性约束下模型的估计问题,对于其他的约束情况,如随机约束,研究模型的估计方法也具有重要实际意义。

参考文献(References):

[1] ENGLE R F, GRANGER C W J, RICE J, et al. Semiparametric estimates of the relation between weather and electricity scales[J]. Journal of the American Statistical Association, 1986, 81(394): 247-269.

[2] ZEGER S L, DIGGLE P J. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters[J]. Biometrics, 1994, 50(3): 689-699.

[3] LIN X H, CARROL R J. Semiparametric regression for clustered data using generalized estimating equations[J]. Journal of the American Statistical Association, 2001, 96(455): 1045-1056.

[4] HE X M, ZHU Z Y, FUNG W K. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure[J]. Biometrika, 2002, 89(3): 579-590.

[5] XUE L G, ZHU L X. Empirical likelihood-based inference in a partially linear model for longitudinal data[J]. Science in China (Series A: Mathematics), 2008, 51(1): 115-130.

[6] FAN J, LI R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis[J]. Journal of the American Statistical Association, 2004, 99(467): 710-723.

[7] 牟婷. 纵向数据下部分线性模型的局部估计法[D]. 上海: 华东师范大学, 2014.
MOU Ting. The local estimation method for partial linear model for longitudinal data[D]. Shanghai: East China Normal University, 2014.

[8] HE H, HUANG LI S. Double-smoothing for bias reduction in local linear regression[J]. Journal of Statistical Planning and Inference, 2009, 139(3): 1056-1072.

[9] 李生彪. 纵向数据下部分线性模型的二次光滑估计[J]. 延边大学学报(自然科学版), 2019, 45(3): 201-207.
LI Sheng-biao. Double smoothing estimation for partial linear model with longitudinal data[J]. Journal of Yanbian University (Natural Science Edition), 2019, 45(3): 201-207.

[10] PRASANGIKA K D, TANG W, YAO Z, et al. Double smoothing local linear estimation in nonlinear time series[J]. Communication in Statistics-Theory and Methods, 2023, 52(5): 1385-1399.

[11] 王秀丽. 高维半参数回归模型的统计推断及相关问题研究[D]. 山东 曲阜: 曲阜师范大学, 2019.
WANG Xiu-li. Research on statistical inference and related issues of high-dimensional semiparametric regression model[D]. Qufu Shandong: Qufu Normal University, 2019.

[12] 郭佳佳. 约束条件下带有测量误差的部分线性变系数模型的估计[D]. 重庆: 重庆工商大学, 2022.
GUO Jia-jia. Estimation of partially linear varying-coefficient model with measurement error under constraint conditions[D]. Chongqing: Chongqing Technology and Business University, 2022.

[13] TANG W, ZUO G X, HE H. Double-smoothing for varying coefficient models[J]. Journal of Nonparametric Statistics, 2011, 23(4): 917-926.

[14] HE H, TANG W, ZUO G X. Statistical inference in the partial linear models with the double smoothing local linear regression method[J]. Journal of Statistical Planning and Inference, 2014, 146(1): 102-112.

责任编辑:李翠薇