变系数模型变窗宽局部 M-估计的渐近正态性*

吴小腊, 刘万荣, 李泽华
(湖南师范大学 数学与计算机科学学院, 长沙 410081)

摘 要:局部多项式回归方法已证明是一个有效的非参数回归方法,有优于流行核方法的特点,如设计的自适应性和高的渐进效率。然而,它的一个缺点是缺乏稳健性。M-型估计是达到所需稳健性的一种自然预期。而且常窗宽对齐次待估曲线是合理的,但对更复杂的曲线如非齐次的,异方差的曲线则失去灵活性,为了完全做到用模型数据去估计参数,本文结合上述两种方法,对变系数模型的系数参数进行估计,并在其中嵌入一个变窗宽加以提高,得到了估计的渐进正态性。

关键词:变系数模型;局部 M-估计;变窗宽
中图分类号:O212.1 **文献标识码**:A **文章编号**:1672-6693(2008)01-0050-04

考虑变系数模型

$$Y = \sum_{j=1}^{p} \beta_j(T) X_j + \varepsilon \tag{1}$$

其中 Y 是实值因变量,$X = (X_1,\cdots,X_p)^{\mathrm{T}}$ 是 p 维随机向量,T 是一维随机变量,$\beta_j(\cdot)(j = 1,\cdots,p)$ 是具有相同光滑程度的未知函数,ε 是随机误差,且满足 $E(\varepsilon \mid T,X) = 0$,$\mathrm{Var}(\varepsilon \mid T,X) = \sigma^2(T)$。

变系数模型是由 Hastie 和 Tibshirani(1993)首先提出的,它是经典线性模型的一种有用的推广,它已被广泛地应用于经济学、生物医学、流行病学、环境科学等学科中,有关模型(1)式中函数系数 $\beta_j(\cdot)$ 的估计,已有几种方法,当 $\beta_j(\cdot)(j = 1,\cdots,p)$ 具有相同的光滑度时,文献[1]给出的局部最小平方方法是一种简单而有用的方法,所得到的估计量是最优的,Hastie(1993)在文献[2]中给出了光滑样条和核方法;Fan(2000),Tsang(2001)提出了局部多项式和光滑样条的两步估计方法[3-4],使得光滑参数的选择可以因系数的不同而不同。唐庆国等在文献[5]中提出了一步估计方法用以估计变系数模型中具有不同光滑度的未知函数。卢一强(2003)通过 B 样条来近似模型(1)式中的$\beta_j(\cdot)$,在有重复观测的情形下讨论了 B 样条 M-估计的收敛速度[6]。

稳健的变窗宽局部线性回归继承了局部多项式回归的优点并且克服了最小二乘方法缺乏稳健性的缺点,变窗宽的嵌入使得局部 M-估计能成功地处理空间非齐次性曲线、异方差性及非均匀设计密度。本文用变窗宽局部 M-估计对变系数模型的系数参数进行估计,在合适的正规条件下,所提的估计是存在的且是渐近正态的。

1 变系数模型变窗宽 M-估计

假设来自于模型(1)式的观测值(T_i,X_i,Y_i),$1 \leqslant i \leqslant n$ 是独立同分布的,$\beta_1(t)$,$\beta_2(t)$,$\cdots$,$\beta_p(t)$ 有连续二阶导数,则对给定的点 t_0,在 t_0 的某一邻域里有 $\beta_j(t) \approx \beta_j(t_0) + \beta'_j(t_0)(t - t_0) \equiv b_j + c_j \dfrac{T_i - t_0}{h_n}$,变系数模型变窗宽 M-估计定义为下列问题的解,极小化

$$\sum_{i=1}^{n} \rho\Big(Y_i - \sum_{j=1}^{p} \Big[\Big(b_j + c_j \frac{T_i - t_0}{h_0}\Big) X_{ij}\Big]\Big) \alpha(T_i) K\Big[\frac{T_i - t_0}{h_n} \alpha(T_i)\Big] \tag{2}$$

其中 $K(\cdot)$ 是一核函数,$\rho(\cdot)$ 是一给定的抗异常值的函数,h_n 是一趋于零的正数列,$\alpha(\cdot)$ 是反映在每一数据

* 收稿日期:2007-05-25 修回日期:2007-09-17
资助项目:国家自然科学基金(No. 30230210)
作者简介:吴小腊(1982-),女,湖南益阳人,硕士研究生,研究方向为数理统计。

点磨光程度可变化的非负连续函数,$\frac{h_n}{\alpha(T_i)}$ 为变窗宽。

极小化(2)式,得下列局部方程

$$\sum_{i=1}^{n}\Psi(Y_i-\sum_{j=1}^{p}[(b_j+c_j\frac{T_i-t_0}{h_n})X_{ij}])\alpha(T_i)K[\frac{T_i-t_0}{h_n}\alpha(T_i)]X_i=0 \tag{3}$$

$$\sum_{i=1}^{n}\Psi(Y_i-\sum_{j=1}^{p}[(b_j+c_j\frac{T_i-t_0}{h_n})X_{ij}])\alpha(T_i)K[\frac{T_i-t_0}{h_n}\alpha(T_i)]\frac{T_i-t_0}{h_n}X_i=0 \tag{4}$$

$\Psi(\cdot)$ 是 $\rho(\cdot)$ 的导函数,$X_i=(X_{i1},X_{i2},\cdots,X_{ip})^{\mathrm{T}}$,$X_{ij}(1\leqslant i\leqslant n,1\leqslant j\leqslant p)$ 表示 X_i 的第 j 个分量的观测值,$\beta_j(t_0)$ 和 $\beta'_j(t_0)$ 的变窗宽局部 M-估计分别定义为(3)、(4)式的解 $\hat{b}_j$ 和 $\frac{\hat{c}_j}{h_n}$。

为了方便,首先给出如下记号。

$$\varepsilon_i=Y_i-\sum_{j=1}^{p}\beta_j(T_i)X_{ij},\ \alpha_j=\alpha(T_i),\ \alpha^*=\min_t\alpha(t)>0,\ K_i=K[\frac{T_i-t_0}{h_n}\alpha(T_i)],$$

$$\varphi(X_i,T_i)=E(\Psi^2(\varepsilon_i)\mid X_i,T_i),\ V(X_i,T_i)=E(\Psi'(\varepsilon_i)\mid X_i,T_i),$$

$$v_\lambda=\int K(u)u^\lambda\mathrm{d}u,\ \omega_\lambda=\int K^2(u)u^\lambda\mathrm{d}u,\lambda\geqslant 0,$$

$$\Omega_{\lambda_1,\lambda_2}=\begin{bmatrix}X_{11}^{\lambda_1}X_{11}^{\lambda_2},X_{11}^{\lambda_1}X_{12}^{\lambda_2} & \cdots & X_{11}^{\lambda_1}X_{1p}^{\lambda_2}\\ \square & \ddots & \square\\ X_{1p}^{\lambda_1}X_{11}^{\lambda_2},X_{1p}^{\lambda_1}X_{12}^{\lambda_2} & \cdots & X_{1p}^{\lambda_1}X_{1p}^{\lambda_2}\end{bmatrix},\bar{\Omega}_{\lambda_1,\lambda_2}=\begin{bmatrix}x_{11}^{\lambda_1}x_{11}^{\lambda_2},x_{11}^{\lambda_1}x_{12}^{\lambda_2} & \cdots & x_{11}^{\lambda_1}x_{1p}^{\lambda_2}\\ \square & \ddots & \square\\ x_{1p}^{\lambda_1}x_{11}^{\lambda_2},x_{1p}^{\lambda_1}x_{12}^{\lambda_2} & \cdots & x_{1p}^{\lambda_1}x_{1p}^{\lambda_2}\end{bmatrix}$$

$$\mu_{\lambda_1,\lambda_2}=\int V(x,t_0)\bar{\Omega}_{\lambda_1,\lambda_2}f(x,t_0)\mathrm{d}x,\ \gamma_{\lambda_1,\lambda_2}=\int\varphi(x,t_0)\bar{\Omega}_{\lambda_1,\lambda_2}f(x,t_0)\mathrm{d}x,$$

$$\beta''(t_0)=(\beta''_1(t_0),\beta''_2(t_0),\cdots,\beta''_p(t_0))^{\mathrm{T}},\ \beta_0=(\beta_1(t_0),\cdots,\beta_p(t_0),h_n\beta'_1(t_0),\cdots,h_n\beta'_p(t_0))^{\mathrm{T}},$$

$$\hat{\beta}=(\hat{b}_1,\cdots,\hat{b}_p,\hat{c}_1,\cdots,\hat{c}_p)^{\mathrm{T}}=(\hat{\beta}_1(t_0),\cdots,\hat{\beta}_p(t_0),h_n\hat{\beta}'_1(t_0),\cdots,h_n\hat{\beta}'_p(t_0))^{\mathrm{T}},$$

$$R(T_i)=\sum_{j=1}^{p}(\beta_j(T_i)-\beta_j(t_0)-\beta'_j(t_0)(T_i-t_0))X_{ij},$$

$$A=\begin{pmatrix}v_0\mu_{1,1} & \frac{v_1\mu_{11}}{\alpha(t_0)}\\ \frac{v_1\mu_{11}}{\alpha(t_0)} & \frac{v_2\mu_{11}}{\alpha^2(t_0)}\end{pmatrix},\ D=\begin{pmatrix}nh_n\alpha(t_0)\gamma_{1,1}\omega_0 & nh_n\gamma_{1,1}\omega_1\\ nh_n\gamma_{1,1}\omega_1 & nh_n\frac{1}{\alpha(t_0)}\gamma_{1,1}\omega_2\end{pmatrix}$$

假设满足下列条件

1) $h_n\to 0$,且 $nh_n\to+\infty(n\to\infty)$;

2) $E(\Psi(\varepsilon_i)\mid X_i,T_i)=0$,且 $\varphi(X_i,T_i)$,$V(X_i,T_i)$ 存在;

3) 设 X 和 T 的联合分布为 $f(x,t)$,函数 $f(x,t)$,$\varphi(x,t)$ 和 $V(x,t)$ 在紧支撑 $X\times T$ 上对 t 有连续的偏导数,且 $f(x,t)>0$,系数函数 $\beta_j(\cdot)(j=1,\cdots,p)$ 及 $\rho(\cdot)$ 有二阶连续导数;

4) 核函数 $K(\cdot)$ 是一对称的,且有紧支撑的概率密度函数,比如 $[-1,1]$;

5) 对任意的 $(x,t)\in X\times T$,$V(x,t)$ 大于零;

6) $\mu_{\lambda_1,\lambda_2}$,$\gamma_{\lambda_1,\lambda_2}$ 里每个元素是有界的,且 $\mu_{1,1}$ 正定;

7) 函数 $\Psi(\cdot)$ 在 t_0 的邻域里关于 t 一致满足:当 $\delta\to 0$ 时

$$E[\sup_{|z|\leqslant\delta}|\Psi(\varepsilon+z)-\Psi(\varepsilon)-\Psi'(\varepsilon)z|\mid x,t]=o(\delta),E[\sup_{|z|\leqslant\delta}|\Psi'(\varepsilon+z)-\Psi'(\varepsilon)z|\mid x,t]=o(1)$$

2 系数参数估计的渐近性

定理 在条件1)~7)满足下,(3)、(4)式存在解,也就是说 $\hat{\beta}$ 存在,如果 $\hat{\beta}$ 是 β_0 一致估计,则有

$$\sqrt{nh_n}\left[(\hat{\beta}-\beta_0)-0.5h_n^2(\alpha(t_0))^{-2}A^{-1}\begin{pmatrix}v_2\mu_{1,1}a''(t_0)\\ \alpha(t_0)v_3\mu_{1,1}a''(t_0)\end{pmatrix}\right]\xrightarrow{d}N(0,A^{-1}DA)$$

成立。

在证明定理前,先给出如下引理。

引理 1　以下约定 $\lambda_1,\lambda_2=0$ 或 1。令 $N_{\lambda_1,\lambda_2}=\sum_{i=1}^{n}\Psi(\varepsilon_i)\alpha_i X_i^{\lambda_1}K_i(T_i-t_0)^{\lambda_2}$,则有 $N_{\lambda_1,\lambda_2}=O_p(n^{\frac{1}{2}}h_n^{\lambda_2+\frac{1}{2}})$。

证明 因为 (T_i,X_i,Y_i),$1\leqslant i\leqslant n$ 是独立同分布的。

$$EN_{\lambda_1,\lambda_2}=nE[E\Psi(\varepsilon_1)\alpha_1X_1^{\lambda_1}K(\frac{T_1-t_0}{h_n}\alpha(T_1))(T_1-t_0)^{\lambda_2}\mid X_1,T_1]=0$$

$$\mathrm{Var}N_{\lambda_1,\lambda_2}=nh_n^{2\lambda_2+1}\frac{1}{(\alpha(t_0))^{2\lambda_2-1}}\gamma_{\lambda_1,\lambda_1}\omega_{2\lambda_2}(1+o(1))$$

所以 $N_{\lambda_1,\lambda_2}=O_p(n^{\frac{1}{2}}h_n^{\lambda_2+\frac{1}{2}})$。　　证毕

引理 2　令 $S_{\lambda_1,\lambda_2}=\sum_{i=1}^{n}\Psi'(\varepsilon_i)R(T_i)\alpha_iX_i^{\lambda_1}K_i(T_i-t_0)^{\lambda_2}$,则

$$S_{\lambda_1,\lambda_2}=0.5nh_n^{2\lambda_2+3}\frac{1}{(\alpha(t_0))^{2\lambda_2+2}}v_{\lambda_2}+2\mu_{\lambda_1,1}\beta''(t_0)(1+o(1))。$$

证明　$$R(T_i)=\sum_{j=1}^{p}(\beta_j(T_i)-\beta_j(t_0)-\beta'_j(t_0)(T_i-t_0))X_{ij},\ |T_i-t_0|<\frac{h_n}{\alpha^*}$$

对 $\beta_j(T_i)$ 应用泰勒展开式,有

$$R(T_i)=\sum_{j=1}^{p}(\frac{\beta''_j(t_0)}{2}(T_i-t_0)^2+o(h_n^2))X_{i,j}=X_i^{\mathrm{T}}(0.5\beta''(t_0)(T_i-t_0)^2)+o(h_n^2)\tag{5}$$

将(5) 式代入 S_{λ_1,λ_2} 的表达式中,类似引理 1 的证法,可得引理 2。　　证毕

引理 3　令 $U_{n,\lambda}=\sum_{i=1}^{n}\Psi'(\varepsilon_i)\alpha_iX_iX_i^{\mathrm{T}}K_i(T_i-t_0)^{\lambda}$,则 $U_{n,\lambda}=nh_n^{\lambda+1}\frac{1}{(\alpha(t_0))^{\lambda}}v_\lambda\mu_{1,1}(1+o_p(1))$。

证明方法同引理 1。

引理 4　令 $F_{\lambda_1,\lambda_2}=\sum_{i=1}^{n}[\Psi(\varepsilon_i+\hat{\eta}_i)-\Psi(\varepsilon_i)-\Psi'(\varepsilon_i)\hat{\eta}_i]\alpha_iX_i^{\lambda_1}K_i(T_i-t_0)^{\lambda_2}$,假设 $\hat{\beta}$ 是 β_0 的一致估计的条件下,则有 $F_{\lambda_1,\lambda_2}=o_p(nh_n^{\lambda_2+1})[h_n^2+(\hat{\beta}-\beta_0)]$ 成立。

证明 令

$$\hat{\eta}_i=R(T_i)-\sum_{j=1}^{p}[\hat{\beta}_j(t_0)-\beta_j(t_0)+(\hat{\beta}'_j(t_0)-\beta'_j(t_0))\frac{T_i-t_0}{h_n}]X_{ij}=R(T_i)-\begin{pmatrix}X_i^{\mathrm{T}}\\ \frac{T_i-t_0}{h_n}X_i^{\mathrm{T}}\end{pmatrix}(\hat{\beta}-\beta_0)$$

因而有　$$|\hat{\eta}_i|\leqslant 0.5C(h_n^2+(\hat{\beta}-\beta_0)),$$

$$E|F_{\lambda_1,\lambda_2}|\leqslant nE|\Psi(\varepsilon_1+\hat{\eta}_1)-\Psi(\varepsilon_1)-\Psi'(\varepsilon_1)\hat{\eta}_1\|\alpha_1\|\|X_1{}^{\lambda_1}\|\|K_1\|(T_1-t_0)^{\lambda_2}|\leqslant$$
$$o(h_n^2+(\hat{\beta}-\beta_0))nE|\|\alpha_1\|\|X_1{}^{\lambda_1}\|\|K_1\|(T_1-t_0)^{\lambda_2}|$$

由条件 7) 及类似引理 1 的证明方法得引理 5 的结论。　　证毕

证明 (定理) 由(3)、(4) 式知

$$\sum_{i=1}^{n}\Psi(Y_i-\sum_{j=1}^{p}[(\hat{\beta}_j(t_0)+\hat{\beta}'_j(t_0)(T_i-t_0))X_{ij}])\alpha_iK_iA_i=0\tag{6}$$

其中 $A_i=\begin{pmatrix}X_i\\ \frac{T_i-t_0}{h_n}X_i\end{pmatrix}$,即　$$\sum_{i=1}^{n}\Psi'(\varepsilon_i)\hat{\eta}_i\alpha_iK_iA_i=\begin{pmatrix}S_{1,0}\\ h_n^{-1}S_{1,1}\end{pmatrix}-\begin{pmatrix}U_{n,0} & h_n^{-1}U_{n,1}\\ h_n^{-1}U_{n,1} & h_n^{-2}U_{n,2}\end{pmatrix}(\hat{\beta}-\beta_0)\tag{7}$$

由(6) 式及前面所给的记号知

$$\hat{\beta}-\beta_0=\begin{pmatrix}U_{n,0} & h_n^{-1}U_{n,1}\\ h_n^{-1}U_{n,1} & h_n^{-2}U_{n,2}\end{pmatrix}^{-1}\left(\begin{pmatrix}F_{1,0}\\ h_n^{-1}F_{1,1}\end{pmatrix}+\begin{pmatrix}N_{1,0}\\ h_n^{-1}N_{1,1}\end{pmatrix}+\begin{pmatrix}S_{1,0}\\ h_n^{-1}S_{1,1}\end{pmatrix}\right)\tag{8}$$

由(8) 式及引理 1 ~ 4 有

$$\sqrt{nh_n}\left[(\hat{\beta}-\beta_0)-0.5h_n^2(\alpha(t_0))^{-2}A^{-1}\begin{pmatrix} v_2\mu_{1,1}\beta''(t_0) \\ \alpha(t_0)v_3\mu_{1,1}\beta''(t_0)\end{pmatrix}\right]=A^{-1}(nh_n)^{-\frac{1}{2}}(N_{1,0},h_n^{-1}N_{1,1})^{\mathrm{T}}+o_p(1)$$

下面求$(nh_n)^{-\frac{1}{2}}(N_{1,0},h_n^{-1}N_{1,1})^{\mathrm{T}}$的渐近分布，设$d_1$和$d_2$是任意两个不同时为0的$p\times 1$向量，记$D_n=(nh_n)^{-\frac{1}{2}}(d_1^{\mathrm{T}}N_{1,0}+h_n^{-1}d_2^{\mathrm{T}}N_{1,1})$，下面证明$D_n$是渐近正态的，由引理1易知

$$ED_n=0,\mathrm{Var}D_n=(nh_n)^{-1}(d_1^{\mathrm{T}},d_2^{\mathrm{T}})\begin{pmatrix} nh_n\alpha(t_0)\gamma_{1,1}\omega_0 & nh_n\gamma_{1,1}\omega_1 \\ nh_n\gamma_{1,1}\omega_1 & nh_n\dfrac{1}{\alpha(t_0)}\gamma_{1,1}\omega_2\end{pmatrix}(d_1^{\mathrm{T}},d_2^{\mathrm{T}})^{\mathrm{T}}$$

独立同分布的中心极限定理

$$D_n\xrightarrow{d}N(0,(d_1^{\mathrm{T}},d_2^{\mathrm{T}})D(d_1^{\mathrm{T}},d_2^{\mathrm{T}})^{\mathrm{T}}),(nh_n)^{-\frac{1}{2}}(N_{1,0},N_{1,1},h_n^{-1})\xrightarrow{d}N(0,D) \quad (9)$$

联合(6)～(9)式有定理的结论成立。

参考文献：

[1] CLEVELAND W S, GROSSSE E, SHYU W M. Local Regression Models[A]. CHAMBERS J M, HASTIE T J. Statistical Models in S. Pacific Grove[C]. California: Wadsworth/Brooks-Cole, 1992.

[2] HASTIE T J, TIBISHIRANI R J. Varying-coefficient Models[J]. J Royal Statist Soc B, 1993, 55: 757-796.

[3] FAN J, ZHANG J T. Two-step Estimation of Functional Linear Models with Applications to Longitudinal Data[J]. J Royal Statist Soc Ser B, 2000, 62(1), 303-322.

[4] CHIANG C T, RICE J A, WC C O. Smoothing Spline Estimation for Varying Coefficient Models with Repeatedl Measure Dependent Varibles[J]. J Amer Statist Ass, 2001, 96(454): 605-619.

[5] 唐庆国，王金德．变系数模型中的一步估计法[J]．中国科学A辑，2005，35(1)：23-38.

[6] 卢一强，曾林蕊．变系数模型B样条M估计的收敛性[J]．应用概率统计，2003(19)：415-423.

[7] FAN J, GIBELS I. Variable Bandwidth and Local Linear Regression Smoothers[J]. Ann Statist, 1992, 20: 2008-2036.

[8] FAN J, JIANG J. Variable Bandwidth and One Step Local M-estimator[J]. Science in China, 2000, 43: 65-80.

[9] FAN J, GIJBELS I. Data-driven Bandwidth Selection in Local Polonomial Fitting: Variable Bandwidth and Spatial Adaption[J]. J Royal Statist Soc B, 1995, 57: 371-394.

[10] 张日权，王静龙．部分线性模型的M-估计[J]．应用数学学报，2005, 28: 151-157.

Variable Bandwidth and Local M-estimation of Varying Coefficient Models

WU Xiao-la, LIU Wan-rong, LI Ze-hua

(College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China)

Abstract: Local polynomial regression methods have been demonstrated as effective nonparametric smoothers. They have advantages over popular kernel methods, in terms of the ability of design adaptation and high asymptotic efficiency. Moreover, the local polynomial regression smoothers can adapt to almost all regression settings and cope very well with the edge effects. A drawback of these local regression estimators is, however , lack of robustness, and M-type of regression estimators are natural candidates for achieving desirable robustness properties. In this paper the variable bandwidth and one step local M approach is employed to estimate the coefficient functions in varying coefficient models. The proposed method inherits the advantages of local polynomial regression and overcomes the shortcoming in lack of robustness of least-squares techniques. The use of variable bandwidth enhances the flexibility of resulting local M-estimators and makes them possible to cope well with spatially inhomogeneous curves, heteroscedastic errors and nonuniform design densities. Under appropriate regularity conditions, it is shown that the proposed estimators exit and are asymptotically normal. This paper focus on establishing joint asymptotic normality of the nonparametric M-type estimators of coefficient functions and its associated derivation based on local linear regression smoothers implemented with variable bandwidth.

Key words: varying coefficient models; local M-estimation; variable bandwidth

(责任编辑　黄　颖)