第 46 卷 第 2 期 兰州理工大学学报 Vol.46 No.2
2020 年 4 月 Journal of Lanzhou University of Technology Apr.2020

文章编号：1673-5196(2020)02-0161-05

基于众数回归的变系数模型统一变量选择

夏亚峰，贾馨懿

（兰州理工大学 理学院，甘肃 兰州 730050）

摘要：研究了众数回归下变系数模型的统一变量选择问题.利用 B 样条基函数近似非参数部分，在众数回归下建立 SCAD 惩罚函数同时选择变系数模型中的重要变量并且识别具有常数效应的协变量，在一定条件下，证明惩罚估计量相合性和稀疏性，通过数值模拟评估所提出的变量选择方法的有效性.

关键词：众数回归；变系数模型；变量选择；SCAD 惩罚

中图分类号：O212.7 文献标志码：A

A unified variable selection approach for varying coefficient models based modal regression

XIA Ya-feng, JIA Xin-yi

(School of Science, Lanzhou Univ. of Tech., Lanzhou 730050, China)

Abstract: A unified variable selection in varying coefficient model with modal regression is studied. B spline basis function is adopted to approximate the nonparametric part, and SCAD penalty function is established under modal regression. The proposed procedure using SCAD penalty function to select important variables simultaneously for the varying coefficient model can also identify the covariates with constant effects. Under certain conditions, the concoherence and sparseness of penalty estimation are proved. Finally, the effectiveness of the proposed variable selection method is evaluated by numerical simulation.

Key words: Modal regression; varying coefficient model; variable selection; SCAD penalty

假定$\{Y_i, X_i, U_i\}$是来自如下变系数模型的一组独立同分布样本

$$\boldsymbol{Y} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}(U) + \varepsilon \tag{1}$$

式中：$U \in [0,1]$为指标变量；$\boldsymbol{Y}$ 为响应变量；$\boldsymbol{X} = (X_1, \cdots, X_d)^{\mathrm{T}} \in \mathbf{R}^d$ 为 d 维协变量；$\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \cdots, \beta_d(\cdot))^{\mathrm{T}}$为未知的 d 维函数向量；ε 为模型随机误差；其中 $E(\varepsilon|U=u)=0$，$\mathrm{Var}(\varepsilon|U=u)=\sigma^2(u)$.关于变系数模型(1)的变量选择问题已经有了较多的研究.Wang 等[1]结合组 LASSO 和样条基函数近似的思想研究了重要变量的选择问题，Wang 等[2]提出了基于核光滑和自适应组 LASSO 的 KLASSO 方法，并证明了 KLASSO 方法的理论性质，Xue 等[3]讨论了在超高维下重要变量的选择问题.

关于稳健变量选择的研究.Noh 等[4]在分位数框架下基于 B 样条近似和组 SCAD 研究了变系数模型的变量选择，Wang 等[5]结合自适应组 LASSO 和 Walsh-average 回归研究了变量选择问题，Zhao 等[6]结合自适应 LASSO 以及L_1 和L_2 损失函数研究了稳健变量选择问题，Yang 等[7]结合自适应组 LASSO 和复合分位数回归讨论了稳健变量选择问题，Hohsuk 等[8]在分位数框架下利用正则化估计方法与 B 样条近似方法研究了变系数模型的变量选择问题.

关于变系数模型统一变量选择的研究. Hu 等[9]基于局部多项式方法和自适应组 LASSO 提出了能识别常系数和变系数的惩罚估计方法，Wang 等[10]结合 B 样条近似和组 SCAD 提出了能识别常系数和变系数的惩罚估计方法，Tang[11]运用 B 样条方法与分位数回归的方法研究了变系数模型变量选择的统一方法.

在本文中基于众数回归的稳健性和有效性，结合 B 样条近似和 SCAD 惩罚函数研究了变系数模

收稿日期：2018-07-09
基金项目：国家自然科学基金(61663024)
作者简介：夏亚峰(1963-)，男，甘肃天水人，教授.

型的统一变量选择问题.在一定条件下,惩罚估计能同时选择变系数模型中的重要变量而且能识别具有常数效应的协变量.此外,本文建立了所得估计量的相合性和稀疏性.

1 变量选择

利用文献[12]众数回归估计思想并结合B样条基函数逼近非参数的方法研究模型(1)的统一变量选择问题.

B样条基函数近似非参数函数系数 $\boldsymbol{\beta}_l(\cdot)$:

$$\boldsymbol{\beta}_l(u) \approx \sum_{k=1}^{K_n} \gamma_{lk} \boldsymbol{B}_k(u) = B(u)^{\mathrm{T}} \boldsymbol{\gamma}_l$$

$$l = 1, \cdots, d$$

其中:$\boldsymbol{\gamma}_l = (\gamma_{l1}, \cdots, \gamma_{lK_n})^{\mathrm{T}}$ 是B样条系数向量,$\boldsymbol{B}(u) = (B_1, \cdots, B_{K_n})^{\mathrm{T}}$ 是阶数为 $M+1$ 的B样条基函数,$K_n = k_n + m + 1$ 表示基函数的个数,k_n 是内节点的个数.根据B样条基函数的性质:

$$\|\boldsymbol{\gamma}_l\|_R^2 = \boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{R} \boldsymbol{\gamma}_l$$

$$\boldsymbol{R} = \int \boldsymbol{B}(u) \boldsymbol{B}(u)^{\mathrm{T}} d(u), \quad \|\boldsymbol{\gamma}_l\|_D^2 = \boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{D} \boldsymbol{\gamma}_l$$

$$l = 1, \cdots, d$$

$$\boldsymbol{D} = \sum_{k=2}^{n} v(U_k) v(U_k)^{\mathrm{T}}$$

$$\boldsymbol{v}(U_k) = (B_1(U_k) - B_1(U_{k-1}), \cdots, B_1(U_k) - B_1(U_{k-1}))^{\mathrm{T}}$$

令 $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\mathrm{T}}, \cdots, \boldsymbol{\gamma}_d^{\mathrm{T}})^{\mathrm{T}}$,则在众数回归框架下,$\gamma$ 的估计 $\hat{\gamma}$ 可以通过极大化式(2)获得

$$Q(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \varphi_h(\boldsymbol{Y}_i - \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{\gamma}) \tag{2}$$

式中:$\boldsymbol{W}_i = I_d \otimes \boldsymbol{B}(U_i) \boldsymbol{X}_i$;$\varphi(t)$是核函数;$h$ 是窗宽,可根据Kai等[13]的方法选择出最优的窗宽.对于变系数模型统一变量选择问题,建立SCAD惩罚函数同时选择变系数模型中的重要变量并且识别具有常数效应的协变量,定义如下的惩罚估计方程:

$$L(\boldsymbol{\gamma}) = Q(\boldsymbol{\gamma}) - n \sum_{l=1}^{d} p_{\lambda_1}(\|\boldsymbol{\gamma}_l\|_R) - n \sum_{l=1}^{d} p_{\lambda_2}(\|\boldsymbol{\gamma}_l\|_D) \tag{3}$$

式中:$p_\lambda(\cdot)$是SCAD惩罚函数;λ_1 和λ_2 是调节参数,可采用BIC准则选取最优参数.式(3)中第一个惩罚函数$p_{\lambda_1}(\cdot)$是用来控制变系数模型的稀疏性,第二个惩罚函数$p_{\lambda_2}(\cdot)$是用来识别变系数模型的常系数协变量.通过极大化目标函数(3),可以得到稀疏估计并实现统一变量选择.

记$\beta_0(\cdot)$为模型中参数真值,不失一般性,定义 $\boldsymbol{\beta}(u) = (\boldsymbol{\beta}_v^{\mathrm{T}}(u), \boldsymbol{\beta}_c^{\mathrm{T}}, \boldsymbol{\beta}_z^{\mathrm{T}})^{\mathrm{T}}$,其中 $\boldsymbol{\beta}_{vd}(u) = (\beta_1(u), \cdots, \beta_{d_1}(u))^{\mathrm{T}}$ 是d_1 维变系数向量,$\boldsymbol{\beta}_c = (\beta_{d_1+1}, \cdots, \beta_s)^{\mathrm{T}}$ 是$s-d_1$ 维常系数向量,$\boldsymbol{\beta}_z = 0$ 是 $d-s$ 维零向量.令

$$F(x, u, h) = E(\varphi_h''(\varepsilon) \mid X = x, U = u)$$

$$G(x, u, h) = E(\varphi_h'(\varepsilon)^2 \mid X = x, U = u)$$

记

$$a_n = \max\{|p_{\lambda_1}'(\|\boldsymbol{\gamma}_{l0}\|_R)|, |p_{\lambda_2}'(\|\boldsymbol{\gamma}_{l0}\|_D)| : \boldsymbol{\gamma}_{l0} \neq 0\}$$

$$b_n = \max\{|p_{\lambda_1}''(\|\boldsymbol{\gamma}_{l0}\|_R)|, |p_{\lambda_2}''(\|\boldsymbol{\gamma}_{l0}\|_D)| : \boldsymbol{\gamma}_{l0} \neq 0\}$$

其中:$\boldsymbol{\gamma}_{l0}$为B样条函数空间中逼近 $\boldsymbol{\beta}_{l0}(\cdot)$的最佳样条系数向量.

2 渐近性质

为了研究所提出方法的渐近性质,给出如下正则条件:

1) 指标变量U具有有界的支撑Ω,其密度函数$f_U(\cdot)$取值大于0并具有连续的二阶导数,不失一般性,假定Ω为单位区间[0,1];

2) $\beta_l(u), l = 1, \cdots, d$ 为区间[0,1]上的 r 阶连续导数,其中 $r \geqslant 2$;

3) 对所有的(X_i, U_i),误差ε_i 的密度函数满足$0 < f_i^{\varepsilon}(\cdot) < \infty$,且具有一致连续的导数;

4) $F(x, u, h)$和 $G(x, u, h)$关于(x, u)连续,且$\forall h > 0, F(x, u, h) < 0$;

5) $E(\varphi_h'(\varepsilon) \mid x, u) = 0$, $E(\varphi_h''(\varepsilon)^2 \mid x, u)$, $E(\varphi_h'(\varepsilon)^3 \mid x, u)$和 $E(\varphi_h'''(\varepsilon) \mid x, u)$关于$(x, u)$连续;

6) $\liminf\limits_{n \to \infty} \liminf\limits_{\|\gamma_l\|_R \to 0+} \lambda_1^{-1} p_{\lambda_1}'(\|\boldsymbol{\gamma}_l\|_R) > 0$

$$l = s+1, \cdots, d$$

$$\liminf_{n \to \infty} \liminf_{\|\gamma_l\|_D \to 0+} \lambda_2^{-1} p_{\lambda_2}'(\|\boldsymbol{\gamma}_l\|_D) > 0$$

$$l = d_1+1, \cdots, s$$

下面的定理给出了估计$\hat{\boldsymbol{\beta}}(\cdot)$的相合性以及统一变量选择的稀疏性.

定理1 假设正则条件1)~6)成立,且节点数k_n 满足$k_n = O(n^{1/(2r+1)})$,如果$b_n \to 0$ 则有

$$\|\hat{\boldsymbol{\beta}}_l(\cdot) - \boldsymbol{\beta}_{l0}\| = O_p(n^{1/(2r+1)} + a_n)$$

$$l = 1, \cdots, d$$

定理2 假设正则条件1)~6)成立,且节点数k_n 满足$k_n = O(n^{1/(2r+1)})$.当 $n \to \infty$,

$$\lambda_{\max} = \max\{\lambda_1, \lambda_2\} \to 0$$

且

$$\lambda_{\min} = \min\{\lambda_1, \lambda_2\}, \quad n^{\frac{r}{2r+1}} \lambda_{\min} \to \infty$$

那么

1) $p(\hat{\boldsymbol{\beta}}_l=\mathbf{0})\to 1, l=s+1,\cdots,d$.

2) 对于每个 $l=d_1+1,\cdots,s$，存在一个常数 $\hat{\boldsymbol{\beta}}_l$，使得 $p(\hat{\boldsymbol{\beta}}_l(u)=\hat{\boldsymbol{\beta}}_l)\to 1$.

3 估计算法与调节参数的选择

3.1 估计算法

损失函数 $Q(\gamma)$ 在原点不可导，SCAD 惩罚函数在原点也不可导，直接极大化目标函数(3)是比较困难的，因而利用局部二次近似逼近来获得惩罚估计.对于给定的初始估计 $\hat{\boldsymbol{\gamma}}_l^{(0)}$ 满足 $\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R>0$ 以及 $\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D>0, l=1,\cdots,d$，那么惩罚函数可以由如下形式近似：

$$p_{\lambda 1}(\|\boldsymbol{\gamma}_l\|_R)\approx p_{\lambda 1}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)+\frac{1}{2}\{p'_{\lambda 1}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R\}\times\{\|\boldsymbol{\gamma}_l\|_R^2-\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R^2$$

$$p_{\lambda 2}(\|\boldsymbol{\gamma}_l\|_D)\approx p_{\lambda 2}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)+\frac{1}{2}\{p'_{\lambda 2}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D\}\times\{\|\boldsymbol{\gamma}_l\|_D^2-\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D^2\}$$

并令

$$E_1=\mathrm{diag}\{(p'_{\lambda 1}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R)\boldsymbol{R},\cdots,(p'_{\lambda 1}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_R)\boldsymbol{R},$$

$$E_2=\mathrm{diag}\{(p'_{\lambda 2}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)\boldsymbol{D},\cdots,(p'_{\lambda 2}(\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)/\|\hat{\boldsymbol{\gamma}}_l^{(0)}\|_D)\boldsymbol{D}\}$$

由 EM 算法可获得目标函数的惩罚估计：

Step 1(E-step)：由下式更新 $\pi(i|\boldsymbol{\gamma}^{(m)})$

$$\pi(i\mid\boldsymbol{\gamma}^{(m)})=\frac{\varphi_h(\boldsymbol{Y}_i-\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{\gamma}^{(m)})}{\sum_{i=1}^{n}\varphi_h(\boldsymbol{Y}_i-\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{\gamma}^{(m)})}\quad i=1,\cdots,n$$

Step 2(M-step)：更新 $\hat{\boldsymbol{\gamma}}^{(m+1)}$

$$\hat{\boldsymbol{\gamma}}^{(m+1)}=\arg\max_{\gamma}\sum_{i=1}^{n}\{\pi(i\mid\boldsymbol{\gamma}^{(m)})\log\varphi_h(\boldsymbol{Y}_i-\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{\gamma})\}-\frac{n}{2}\boldsymbol{\gamma}^{\mathrm{T}}(\boldsymbol{E}_1+\boldsymbol{E}_2)\boldsymbol{\gamma}=(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{S}\boldsymbol{W}_i+n(\boldsymbol{E}_1+\boldsymbol{E}_2))^{-1}\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{S}\boldsymbol{Y}$$

其中：$\boldsymbol{S}$ 是以 $\pi(i|\boldsymbol{\gamma}^{(m)})$ 为其第 i 个对角元素的 $n\times n$ 对角矩阵.

Step 3：重复迭代 Step 1 和 Step 2 直至收敛，记算法收敛所得的估计为 $\hat{\gamma}$.

3.2 调节参数的选择

为实现上面的估计算法，节点的个数 k_n，SCAD 惩罚函数中的调节参数 a,λ_1,λ_2 都需要合适的选取.根据 Fan 等[14]，本文采用 $a=3.7$，同时基于 Xue 等[3]选择节点的思想，采用等距节点且节点个数 $k_n=[n^{1/(2m+3)}]$，采用 BIC 准则选取最优的 $\lambda=(\lambda_1,\lambda_2)$.

$$\mathrm{BIC}(\lambda)=\log(n^{-2}\sum_{i=1}^{n}\varphi_h(\boldsymbol{Y}_i-\boldsymbol{W}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}))+df_v\frac{\log(n/k_n)}{n/k_n}+df_c\frac{\log(n)}{n}$$

其中：df_v 代表模型估计中识别出的变系数部分的个数；df_c 代表模型估计中识别出的常系数部分的个数.

4 数值模拟

通过数值模拟检验变系数模型的惩罚估计方法在有限样本下的模型估计和变量选择的有效性.

对于变系数模型

$$Y_i=\sum_{l=1}^{10}\boldsymbol{\beta}_l(U_i)\boldsymbol{X}_{il}+\varepsilon_i,\quad i=1,\cdots,n$$

考虑 $U_i\sim U(0,1)$，$\boldsymbol{\beta}_1=2\sin(2\pi u)$，$\boldsymbol{\beta}_2=8u-9$，$\boldsymbol{\beta}_3=2-3\cos((6u-5)\pi/3)$，$\boldsymbol{\beta}_4=2$，$\boldsymbol{\beta}_5=1$，其余 5 个变量为不相关变量，协变量 $\boldsymbol{X}_{il}(l=1,\cdots 10)$ 均服从多元正态分布，其均值为零且协方差矩阵中 (k,l) 个元素为 $0.8^{|k-l|}$.在模拟中，考虑下面 3 种误差分布：$N(0,0.5)$、$t(3)$ 和污染正态分布(CN)$0.9N(0,0.5)+0.1N(0,15)$，为了衡量估计的效果，采用均方误差平方根(RASE)来评价惩罚估计的拟合效果，其定义为

$$\mathrm{RASE}=\sqrt{n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d}(\hat{\beta}_l(U_i)-\boldsymbol{\beta}_l(U_i))^2}$$

模拟结果见表 1 所列，其中包括多次模拟的 RASE 的平均值，正确识别零系数的平均个数“CZ”，错误识别非零系数的平均个数“IZ”，正确识别变系数的平均个数“CV”，正确识别常系数的平均个数“CC”.

从表 1 可以得到如下结论：

1) 对于给定的模型和误差分布，在正确识别零系数和常系数方面，SCAD 估计得到的值比 LASSO 估计得到的值更接近于真实模型，这说明 SCAD 估计在变量选择以及识别常数协变量方面比 LASSO 估计表现的好.

2) 当模型误差服从正态分布时，SCAD 惩罚方法得到的模型估计相比于其它分布具有最小的

RASE;当模型误差服从厚尾分布 $t(3)$ 或含有异常点(CN)时,SCAD 惩罚方法得到模型估计具有较小的 RASE,优于 LASSO 惩罚方法.

表1 基于不同变量选择方法的模拟结果

Tab.1 Simulated results based on different variable selection methods

n	误差	方法	CZ	IZ	CV	CC	RASE
200	$N(0,0.5)$	LASSO	4.58	0.13	3.00	1.64	0.366 5
		SCAD	4.71	0.05	3.00	2.00	0.333 4
	$t(3)$	LASSO	4.39	0.15	3.00	1.80	0.370 4
		SCAD	4.42	0.11	3.00	1.94	0.349 2
	CN	LASSO	4.20	0.04	3.00	1.81	0.377 7
		SCAD	4.91	0.14	3.00	1.86	0.356 1
400	$N(0,0.5)$	LASSO	4.86	0.12	3.00	1.90	0.249 7
		SCAD	5.00	0.08	2.86	2.00	0.239 4
	$t(3)$	LASSO	4.67	0.10	3.00	1.76	0.258 7
		SCAD	4.53	0.13	3.00	2.00	0.249 6
	CN	LASSO	4.76	0.08	3.00	1.81	0.256 4
		SCAD	4.81	0.07	3.00	2.00	0.247 7

5 定理证明

定理1证明 记 $\delta=n^{-r/(2r+1)}+a_n$ 和 $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0+\delta\boldsymbol{T}$,其中 $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_1^{\mathrm{T}},\cdots,\boldsymbol{\gamma}_d^{\mathrm{T}})^{\mathrm{T}}$,$\boldsymbol{T}=(\boldsymbol{T}_1^{\mathrm{T}},\cdots,\boldsymbol{T}_d^{\mathrm{T}})^{\mathrm{T}}$,先证明对于任意给定的 $\eta>0$,存在一个充分大的 C 使得式(4)成立

$$P\{\sup_{\|T\|=C}L(\boldsymbol{\gamma})<L(\boldsymbol{\gamma}_0)\}\leqslant 1-\eta \tag{4}$$

令 $\Pi(\boldsymbol{\gamma})=\{L(\boldsymbol{\gamma})-L(\boldsymbol{\gamma}_0)\}$,则由 Taylor 展开,有

$$\begin{aligned}\Pi(\boldsymbol{\gamma})=&-\delta\sum_{i=1}^{n}\varphi_h'(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}R(U_i))\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T}+\\&\delta^2\sum_{i=1}^{n}\varphi_h''(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}R(U_i))(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T})^2-\\&\delta^3\sum_{i=1}^{n}\varphi_h'''(\boldsymbol{\zeta}_i)(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T})^3-\\&n\sum_{l=1}^{d}\{p_{\lambda1}(\|\boldsymbol{\gamma}_l\|_R)-p_{\lambda1}(\|\boldsymbol{\gamma}_{l0}\|_R)\}-\\&n\sum_{l=1}^{d}\{p_{\lambda2}(\|\boldsymbol{\gamma}_l\|_D)-p_{\lambda2}(\|\boldsymbol{\gamma}_{l0}\|_D)\}\triangleq\\&J_1+J_2+J_3+J_4+J_5\end{aligned}$$

其中:ζ_i 在 $\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i)$ 和 $\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i)-\delta\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T}$ 之间,$\boldsymbol{R}(u)=(\boldsymbol{R}_1(u),\cdots,\boldsymbol{R}_d(u))^{\mathrm{T}}$,$\boldsymbol{R}_l(u)=\boldsymbol{\beta}_l(u)-\boldsymbol{B}(u)^{\mathrm{T}}\boldsymbol{\gamma}_{l0}$,$l=1,\cdots,d$.

根据正则条件1)～2)以及文献[15]中推论6.21的结论,有 $\|\boldsymbol{R}_l(u)\|=O(K^{-r})$,进一步,由 Taylor 展式可得

$$\begin{aligned}J_1=&-\delta\sum_{i=1}^{n}\varphi_h'(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T}=\\&-\delta\sum_{i=1}^{n}\{\varphi_h'(\boldsymbol{\varepsilon}_i)+\varphi_h''(\boldsymbol{\varepsilon}_i)(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))+\\&\varphi_h'''(\xi_i)(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))^2\}\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T}\end{aligned}$$

其中:ξ_i 位于 $\boldsymbol{\varepsilon}_i$ 和 $\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i)$ 之间.经过简单计算,有

$$\begin{aligned}J_1=&-\delta\sum_{i=1}^{n}\varphi_h'(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{T}=\\&O_p(nK^{-r}\delta\|\boldsymbol{T}\|)=O_p(n\delta^2\|\boldsymbol{T}\|)\end{aligned}$$

对于 J_2,易证

$$J_2=E(F(\boldsymbol{X},U,h))O_p(n\delta^2\|\boldsymbol{T}\|^2)$$

类似的经过计算可得 $J_3=O_p(n\delta^3\|\boldsymbol{T}\|^3)$,根据定理条件 $a_n\to0$,因此 $\delta\to0$,即 $\delta\|\boldsymbol{T}\|\to0$,由此可得 $J_3=o_p(J_2)$,在 $\|\boldsymbol{T}\|=C$ 上,J_3 也被 J_2 一致地控制住.

考虑 J_4,利用 $\|\boldsymbol{\gamma}_l\|_R-\|\boldsymbol{\gamma}_{l0}\|_R\leqslant\|\boldsymbol{\gamma}_l-\boldsymbol{\gamma}_{l0}\|_R<\|\boldsymbol{\gamma}_l-\boldsymbol{\gamma}_{l0}\|$ 和 $b_n\to0$,可得

$$\begin{aligned}J_4=&n\sum_{l=1}^{d}\{p_{\lambda1}(\|\boldsymbol{\gamma}_l\|_R)-p_{\lambda1}(\|\boldsymbol{\gamma}_{l0}\|_R)\}=\\&n\sum_{l=1}^{d}\{p_{\lambda1}(\|\boldsymbol{\gamma}_{l0}+\delta\boldsymbol{T}_l\|_R)-p_{\lambda1}(\|\boldsymbol{\gamma}_{l0}\|_R)\}=\\&n\sum_{l=1}^{d}\{p_{\lambda1}'(\|\boldsymbol{\gamma}_{l0}\|_R)(\|\boldsymbol{\gamma}_{l0}+\delta\boldsymbol{T}_l\|_R-\|\boldsymbol{\gamma}_{l0}\|_R)+\\&\frac{1}{2}p_{\lambda1}''(\|\boldsymbol{\gamma}_{l0}\|_R)(\|\boldsymbol{\gamma}_{l0}+\delta\boldsymbol{T}_l\|_R-\\&\|\boldsymbol{\gamma}_{l0}\|_R)^2(1+o_p(1))\}\leqslant\\&\sum_{l=1}^{s}\{n\delta p_{\lambda1}'(\|\boldsymbol{\gamma}_{l0}\|_R)\|\boldsymbol{T}_l\|+\\&\frac{1}{2}n\delta^2p_{\lambda1}''(\|\boldsymbol{\gamma}_{l0}\|_R)\|\boldsymbol{T}_l\|^2(1+o_p(1))\}\leqslant\\&\sqrt{s}\{n\delta a_n\|\boldsymbol{T}_l\|+n\delta^2b_n\|\boldsymbol{T}_l\|^2\}\end{aligned}$$

根据 $b_n\to0$,在 $\|\boldsymbol{T}\|=C$ 上,J_4 被 J_2 一致地控制住.同样的方法亦可证 J_5 也被 J_2 一致地控制住.

因此,通过选取充分大的 C,可以得到式(4)成立,从而存在局部最大值点满足:

$$\|\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0}\|=O_p(\delta) \tag{5}$$

注意到

$$\begin{aligned}&\|\hat{\boldsymbol{\beta}}_l(\cdot)-\boldsymbol{\beta}_{l0}(\cdot)\|^2=\int_0^1(\hat{\boldsymbol{\beta}}_l(u)-\boldsymbol{\beta}_{l0}(u))^2\mathrm{d}u=\\&\int_0^1(\boldsymbol{B}(u)^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_l-\boldsymbol{B}(u)^{\mathrm{T}}\boldsymbol{\gamma}_{l0}+\boldsymbol{R}_l(u))^2\mathrm{d}u\leqslant\\&2\int_0^1(\boldsymbol{B}(u)^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_l-\boldsymbol{B}(u)^{\mathrm{T}}\boldsymbol{\gamma}_{l0})^2\mathrm{d}u+\\&2\int_0^1(\boldsymbol{R}_l(u))^2\mathrm{d}u=\\&2(\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0})^{\mathrm{T}}\boldsymbol{R}(\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0})+2\int_0^1(\boldsymbol{R}_l(u))^2\mathrm{d}u\end{aligned}$$

其中:$\boldsymbol{R}=\int\boldsymbol{B}(u)\boldsymbol{B}(u)^{\mathrm{T}}\mathrm{d}u$.由于 $\|\boldsymbol{R}\|=O(1)$ 以及

式(5)，有

$$(\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0})^{\mathrm{T}}\boldsymbol{R}(\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0})=O_p(n^{-2r/(2r+1)}+a_n^2)$$

此外，易证

$$\int_0^1(\boldsymbol{R}_l(u))^2\mathrm{d}u=O_p(n^{-2r/(2r+1)})$$

从而得到

$$\|\hat{\boldsymbol{\beta}}_l(\cdot)-\boldsymbol{\beta}_{l0}(\cdot)\|=O_p(n^{-r/(2r+1)}+a_n)$$

定理成立.

定理 2 证明　先证明定理的第一部分.根据 SCAD 惩罚函数的性质，当$\lambda_{\max}\to 0$ 时，$a_n=0$.由定理 1 的结论，只需证明对任意给定充分小的 $\upsilon=Cn^{-r/(2r+1)}$，当 $n\to\infty$时，对任意 γ 满足 $\|\hat{\boldsymbol{\gamma}}_l-\boldsymbol{\gamma}_{l0}\|=O_p(n^{-r/(2r+1)})$，$l=1,\cdots,s$ 以概率 1 使式(6,7)成立.

$$\frac{\partial L(\boldsymbol{\gamma})}{\partial\hat{\boldsymbol{\gamma}}_l}<0,\quad 0<\hat{\boldsymbol{\gamma}}_l<\upsilon,\quad l=s+1,\cdots,d \tag{6}$$

$$\frac{\partial L(\boldsymbol{\gamma})}{\partial\hat{\boldsymbol{\gamma}}_l}>0,\quad -\upsilon<\hat{\boldsymbol{\gamma}}_l<0,\quad l=s+1,\cdots,d \tag{7}$$

类似于定理 1 的证明，可得

$$\begin{aligned}\frac{\partial L(\boldsymbol{\gamma})}{\partial\hat{\boldsymbol{\gamma}}_l}=&\frac{\partial Q(\boldsymbol{\gamma})}{\partial\hat{\boldsymbol{\gamma}}_l}-n(p_{\lambda_1}(\|\hat{\boldsymbol{\gamma}}_l\|_R))'=\\&-\sum_{i=1}^n\{\boldsymbol{W}_{il}\varphi_h'(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))-\\&\sum_{i=1}^n\varphi_h''(\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i))\boldsymbol{W}_{il}\boldsymbol{W}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+\\&\sum_{i=1}^n\varphi'''_h(\varphi_i)\boldsymbol{W}_{il}(\boldsymbol{W}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0))^2-\\&np'_{\lambda_1}(\|\hat{\boldsymbol{\gamma}}_l\|_R)\frac{\boldsymbol{R}\cdot\hat{\boldsymbol{\gamma}}_l}{\|\hat{\boldsymbol{\gamma}}_l\|_R}=\\&-n\lambda_1\left\{\lambda_1^{-1}\cdot p'_{\lambda_1}(\|\hat{\boldsymbol{\gamma}}_l\|_R)\frac{R\cdot\hat{\boldsymbol{\gamma}}_l}{\|\hat{\boldsymbol{\gamma}}_l\|_R}+\right.\\&\left.O_p(\lambda_1^{-1}\cdot n^{-r/(2r+1)})\right\}\end{aligned}$$

其中：φ_i 在$Y_i-\boldsymbol{W}_i^{\mathrm{T}}\gamma$ 和$\boldsymbol{\varepsilon}_i+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{R}(U_i)$之间.

由条件 6)，

$$\liminf_{n\to\infty}\liminf_{\|\gamma_l\|_R\to 0+}\lambda_1^{-1}\cdot p'_{\lambda_1}(\|\boldsymbol{\gamma}_l\|_R)>0$$

$$n^{\frac{r}{2r+1}}\lambda_1>n^{\frac{r}{2r+1}}\lambda_{\min}\to\infty$$

因此$\frac{\partial L(\boldsymbol{\gamma})}{\partial\hat{\boldsymbol{\gamma}}_l}$的符号是由$\hat{\boldsymbol{\gamma}}_l$ 决定的，即式(6,7)成立.

进一步，由于 $\sup_u\|\boldsymbol{B}(u)\|=O(1)$和$\hat{\boldsymbol{\beta}}_l(u)=\boldsymbol{B}(u)^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_l$，可得结论成立.对于定理的第二部分，可用类似的证明方法得到.

参考文献：

[1] WANG L F, LI H Z, HUANG J Z. Variable selection for nonparametric varying coefficient models for analysis of repeated measurements [J]. Journal of American Statistical Association, 2008, 103: 1556-1569.

[2] WANG H S, XIA Y C. Shrinkage estimation of the varying coefficient model [J]. Journal of American Statistical Association, 2009, 104: 747-757.

[3] XUE L, QU A. Variable selection in high-dimensional varying-coefficient models with global optimality [J]. Journal of Machine Learning Research, 2012, 13: 1973-1998.

[4] NOH H, CHUNG K, KEILEGOM I. Variable selection of varying coefficient models in quantile regression [J]. Electronic Journal of Statistics, 2012, 6: 1220-1238.

[5] WANG K N, LIN L. Walsh-average based variable selection for varyingcoefficient models [J]. Journal of the Korean Statistical Society, 2015, 44: 95-110.

[6] ZHAO P X, LI G H. Modified SEE variable selection for varying-coefficient instrumental variable models [J]. Statistical Methodology, 2013, 12: 60-70.

[7] YANG H, LV J, GUO C H. Weighted composite quantile regression estimation and variable selection for varying-coefficient models with heteroscedasticity [J]. Journal of the Korean Statistical Society, 2015, 44: 77-94.

[8] HOHSUK N, KWANGHUN C, INGRID V. Variable selection of varying coefficient models in quantile regression [J]. Electronic Journal of Statistics, 2012, 6: 1220-1238.

[9] HU T, XIA Y C. Adaptive semi-varying coefficient model selection [J]. Statistica Sinica, 2012, 22: 575-599.

[10] WANG M Q, SONG L X. Identification for semiparametric varying coefficient partially linear models [J]. Statistics and Probability Letters, 2013, 83: 1311-1320.

[11] TANG Y L, WANG H J, ZHU Z Y, *et al*. A unified variable selection approach for varying coefficient models [J]. Statistica Sinica, 2012, 22: 601-628.

[12] YAO W X, LI L H. A new regression model: modal linear regression [J]. Scandinavian Journal of Statistics, 2014, 41: 656-671.

[13] KAI B, LI R, ZOU H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models [J]. The Annals of Statistics, 2011, 39: 305-332.

[14] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. Journal of the American Statistics Association, 2001, 96: 1348-1360.

[15] SCHUMAKER L. Splines function: basic theory [M]. New York: Wiley, 1981.