文章编号：1673-5196(2008)02-0155-05

缺失数据下半参数回归模型的渐近性质

罗双华[1]，田 萍[2]，蒋红英[3]

(1.兰州理工大学 理学院，甘肃 兰州 730050；2.河南许昌学院 数学系，河南 许昌 461000；3.兰州理工大学 土木工程学院，甘肃 兰州 730050)

摘要：在缺失响应变量的不完全数据下，对半参数回归模型进行研究.利用局部线性回归拟合方法建立缺失数据下半参数回归模型参数分量$\boldsymbol{\beta}$和非参数分量g的局部线性估计$\hat{\boldsymbol{\beta}}_n$，$\hat{g}_n^*(t)$，基于$\hat{\boldsymbol{\beta}}_n$建立$\sigma^2$的估计量$\hat{\sigma}_n^2$.在适当的条件下，证明$\hat{\boldsymbol{\beta}}_n$，$\hat{\sigma}_n^2$的渐近正态性，得到$\hat{g}_n^*(t)$的最优弱收敛速度.

关键词：半参数回归模型；局部线性回归光滑；渐近正态性；缺失数据；最优弱收敛速度

中图分类号：O211.7 **文献标识码**：A

Asymptotic behavior of semiparametric regression model under missing data

LUO Shuang-hua[1]，TIAN Ping[2]，JIANG Hong-ying[3]

(1.School of Science，Lanzhou Univ. of Tech.，Lanzhou 730050，China；2.Dept. of Mathematics，Xuchang University，Xuchang 461000，China；3.College of Civil Engineering，Lanzhou Univ. of Tech.，Lanzhou 730050，China)

Abstract：A semiparametric regressive model was investigated under missing response data. Local linear estimation $\hat{\boldsymbol{\beta}}_n$ and $\hat{g}_n^*(t)$ of parametric component $\boldsymbol{\beta}$ and nonparametric component g in semi-parametric regression model was performed by using local linear regressive fitting. The estimator $\hat{\sigma}_n^2$ of σ^2 was also obtained on the basis of $\hat{\boldsymbol{\beta}}_n$. The asymptotic normality of the estimators $\hat{\boldsymbol{\beta}}_n$，and $\hat{\sigma}_n^2$ was proved on an appropriate condition and the optimal weak convergence rate of the estimator $\hat{g}_n^*(t)$ was obtained，also.

Key words：semiparametric regression model；local linear smoothness；asymptotic normality；missing data；optimal weak convergence rate

对于半参数回归模型：

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + g(T_i) + \varepsilon_i \tag{1}$$

其中$(\boldsymbol{X},\boldsymbol{T})$为取值于$R^p\times[0,1]$上的随机向量，$\boldsymbol{\beta}$为$p\times1$的未知参数向量，$g$为定义于$[0,1]$上的未知函数，$\varepsilon$为随机误差，$E\varepsilon=0$，$E\varepsilon^2=\sigma^2>0$且$(\boldsymbol{X},\boldsymbol{T})$与$\varepsilon$独立.在完全数据下，许多文献已对$\boldsymbol{\beta}$和$g$进行估计并讨论了渐近正态性和最优弱收敛速度.但在实际中，数据常不能完全观测，即有i.i.d.样本$\{\boldsymbol{X}_i=(X_{i1},\cdots,X_{ip})',T_i,Y_i,\delta_i\}_{i=1}^n$，其中$\{\boldsymbol{X}_i,T_i\}_{i=1}^n$全部被观测，当$\delta_i=1$时，$Y_i$被观测，$\delta_i=0$时，$Y_i$缺失，且满足缺失条件(MAR)：

$$P(\delta=1\mid \boldsymbol{X},\boldsymbol{T},\boldsymbol{Y}) = P(\delta=1\mid \boldsymbol{X},\boldsymbol{T}) = p(\boldsymbol{X},\boldsymbol{T})$$

在缺失数据下对非参数回归或半参数回归模型的研究较少[1~4].本文主要基于i.i.d.缺失数据样本$\{\boldsymbol{X}_i=(X_{i1},\cdots,X_{ip})',T_i,Y_i,\delta_i\}$，将局部线性拟合方法[5]用于半参数回归模型，得到$\boldsymbol{\beta}$，g和σ^2的估计量$\hat{\boldsymbol{\beta}}_n$，$\hat{g}_n^*(t)$和$\hat{\sigma}_n^2$的渐近正态性和最优弱收敛速度.

1 模型和估计量

设$\{\boldsymbol{X}_i=(X_{i1},\cdots,X_{ip})',T_i,Y_i,\delta_i\}_{i=1}^n$是一组来自模型(1)的i.i.d的样本.在模型(1)中，首先视$\boldsymbol{\beta}$已知，将它看作非参数回归模型.为得到$g(\boldsymbol{T})$的估计，采用局部线性拟合，即求a和b以最小化

$$\sum_{i=1}^n[Y_i-\boldsymbol{X}_i'\boldsymbol{\beta}-a-b(T_i-t)]^2K_h(T_i-t)\delta_i$$

计算得

$$\hat{g}_n(t)=\hat{a}=\frac{\sum_{i=1}^n U_{ni}(t)(Y_i-\boldsymbol{X}_i'\boldsymbol{\beta})}{\sum_{j=1}^n U_{nj}(t)} \tag{2}$$

收稿日期：2007-01-18

基金项目：甘肃省自然科学基金(071RJZA063)

作者简介：罗双华(1978-)，女，四川遂宁人，硕士，讲师.

式中：

$$U_{nj}(t)=K_h(T_j-t)\delta_j[R_{n,2}(t)-(T_j-t)R_{n,1}(t)]$$

$$R_{nl}(t)=\sum_{i=1}^{n}\delta_i K_h(T_i-t)(T_i-t)^l,\quad l=0,1,2$$

将式(2)进行改写得

$$\hat{g}_n(t)=\sum_{i=1}^{n}W_{ni}(t)(Y_i-\boldsymbol{X}_i'\boldsymbol{\beta})=\hat{g}_{1n}(t)-\hat{\boldsymbol{g}}_{2n}'(t)\boldsymbol{\beta} \tag{3}$$

式中：$\hat{g}_{1n}(t)=\sum_{i=1}^{n}W_{ni}(t)Y_i$ 和 $\hat{\boldsymbol{g}}_{2n}'(t)=\sum_{i=1}^{n}W_{ni}(t)\boldsymbol{X}_i$ 分别为 $g_1(t)=E(Y_1\mid T_1=t)$ 和 $g_2(t)=E(\boldsymbol{X}_1\mid T_1=t)$ 的局部线性估计，$\hat{\boldsymbol{g}}_{2n}'$ 表示向量 $\hat{\boldsymbol{g}}_{2n}$ 的转置.

$$W_{ni}(t)=\frac{U_{ni}(t)}{\sum_{j=1}^{n}U_{nj}(t)}$$

将 $\hat{g}_n(t)$ 代入式(1)可得

$$Y_i=\tilde{X}_i'\boldsymbol{\beta}+\varepsilon_i \tag{4}$$

式中：$\tilde{X}_i=\boldsymbol{X}_i-\sum_{j=1}^{n}W_{nj}(T_i)\boldsymbol{X}_j$

$$\tilde{Y}_i=Y_i-\sum_{j=1}^{n}W_{nj}(T_i)Y_j,\ 0\leqslant \mathrm{i}\leqslant \mathrm{n}$$

记 $\tilde{X}=(\tilde{X}_1,\cdots,\tilde{X}_n)'$，$\tilde{Y}=(\tilde{Y}_1,\cdots,\tilde{Y}_n)'$. 利用式(4)由通常的最小二乘法定义 $\boldsymbol{\beta}$ 的估计为式(5)最小化的解

$$\sum_{i=1}^{n}(\tilde{Y}_i-\tilde{X}_i'\boldsymbol{\beta})^2\delta_i \tag{5}$$

记 $\boldsymbol{V}=\mathrm{Diag}[\delta_1,\cdots,\delta_n]$，可得 $\boldsymbol{\beta}$ 的估计和 σ^2 的估计分别为

$$\hat{\boldsymbol{\beta}}_n=(\tilde{X}'\boldsymbol{V}\tilde{X})^{-1}\tilde{X}'\boldsymbol{V}\tilde{Y} \tag{6}$$

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i-\tilde{X}_i'\hat{\boldsymbol{\beta}}_n)^2\delta_i/p(\boldsymbol{X}_i,T_i) \tag{7}$$

把 $\hat{\boldsymbol{\beta}}_n$ 代入式(3)得

$$\hat{g}_n^*(t)=\sum_{i=1}^{n}W_{ni}(t)(Y_i-\boldsymbol{X}_i'\hat{\boldsymbol{\beta}}_n)=\hat{g}_{1n}(t)-\hat{\boldsymbol{g}}_{2n}'(t)\hat{\boldsymbol{\beta}}_n$$

2 主要结果

本文的主要结果需要的条件如下：

1) T_1 的密度函数 $r(t)$ 满足：

$$0<\inf_{0\leqslant t\leqslant 1}\sup_{0\leqslant t\leqslant 1}r(t)<\infty$$

2) 窗宽满足：

$$\frac{nh^2}{\lg^2 n}\to\infty,\ nh^6\to 0,\ 当\ n\to\infty$$

3) 核函数 $K(\cdot)$ 为连续对称概率密度函数，具有有界变差和有界支撑$[-1,1]$.

4) $g(t)$在$[0,1]$上具有有界的二阶导数，$g_1(t)$ 和 $g_{2j}(t)$有二阶有界导数，$g_{2j}(t)=E(X_{1j}\mid T_1=t)$是 $g_2(t)$的第 j 个分量.

5) $E(Y_1^2\mid T_1=t)$和 $E(X_{1j}^2\mid T_1=t)$，$1\leqslant j\leqslant p$ 都是 t 的有界函数.

6) $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{X}_1-E(\boldsymbol{X}_1\mid T_1))$是正定矩阵.

下面给出本文的主要结果.

定理 1 若条件 1)～6)成立，则

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\xrightarrow{L}N(\boldsymbol{0},\sigma^2p^{-1}(x,t)\boldsymbol{\Sigma}^{-1}) \tag{8}$$

定理 2 若条件 1)～6)成立且 $0<\mathrm{Var}(\varepsilon'\boldsymbol{S}\varepsilon)<\infty$，则

$$\sqrt{n}(\hat{\sigma}_n^2-\sigma^2)\xrightarrow{L}N(0,\mathrm{Var}(\varepsilon'\boldsymbol{S}\varepsilon)) \tag{9}$$

式中：$\boldsymbol{S}=\mathrm{Diag}[\delta_1/p(\boldsymbol{X}_1,T_1),\cdots,\delta_n/p(\boldsymbol{X}_n,T_n)]$.

定理 3 若条件 1)～6)成立，且取

$$h_n=O(n^{-1/5})$$

则 $$\hat{g}_n^*(t)-g(t)=O_p(n^{-2/5})$$

定理 4 若条件 1)～6)成立，则

$$p(x,t)(\hat{\boldsymbol{\beta}}_n-\beta)'\tilde{X}'\mathrm{V}\tilde{X}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})/\hat{\sigma}_n^2\xrightarrow{L}\chi_p^2 \tag{10}$$

此结果可用于关于 $\boldsymbol{\beta}$ 的大样本置信椭球估计以及假设检验.

注 1 上述结果可推广到 $\boldsymbol{T}$ 为多维的情形.

3 定理的证明

首先给出一些记号

$$\boldsymbol{W}_{ni}=(W_{ni}(T_1),W_{ni}(T_2),\cdots,W_{ni}(T_n))$$

$$\boldsymbol{\Sigma}=(\Sigma_{ij})_{p\times p}$$

$$\boldsymbol{W}=(W_{n1},\cdots,W_{nn})'$$

$$\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n\})'$$

$$\boldsymbol{\varepsilon}=(\varepsilon_1,\cdots,\varepsilon_n)'$$

$$\boldsymbol{Y}=(Y_1,\cdots,Y_n)'$$

$$m_{ij}=g_{2j}(T_i)$$

$$h_{ij}=X_{ij}-m_{ij}$$

$$1\leqslant \mathrm{i}\leqslant \mathrm{n},\quad 1\leqslant \mathrm{j}\leqslant \mathrm{p}$$

$$\tilde{G}(\boldsymbol{T})=(g(T_1)-\hat{g}_n(T_1),\cdots,g(T_n)-\hat{g}_n(T_n))'$$

则易见有 $\tilde{X}=(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}$，$\tilde{Y}=(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}$ 且

$$\tilde{Y}=\tilde{X}^{\beta}+\tilde{G}(\boldsymbol{T})+\boldsymbol{\varepsilon} \tag{11}$$

为方便记，用 c 表示绝对正常数，每次出现值可不同，先给出如下若干引理.

引理 1 若条件 1)～6)成立，则有

$$\max_{0\leqslant i\leqslant n}\left|m_{ij}-\sum_{k=1}^{n}W_{nk}(T_i)X_{kj}\right|\to 0$$
$$\mathrm{a.s.}\quad 1\leqslant j\leqslant p$$

注 2　此引理的证明类似于文[6]的引理 4.2.3 可证得.

引理 2　若条件 1)～6)成立,则

$$\frac{1}{n}\tilde{X}'V\tilde{X}\to p(x,t)\boldsymbol{\Sigma},\mathrm{a.s.}\tag{12}$$

证明　因为 $\tilde{X}=(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}$,其中第$(i,j)$元为

$$\tilde{X}_{ij}=X_{ij}-\sum_{s=1}^{n}W_{ns}(T_i)X_{sj}=$$
$$h_{ij}+m_{ij}-\sum_{s=1}^{n}W_{ns}(T_i)X_{sj}\tag{13}$$

所以 $\tilde{X}'V\tilde{X}$ 的第(i,j)元为

$$\sum_{t=1}^{n}\tilde{X}_{ti}\tilde{X}_{tj}\delta_t=\sum_{t=1}^{n}h_{ti}h_{tj}\delta_t+$$
$$\sum_{t=1}^{n}h_{ti}\Big(m_{tj}-\sum_{s=1}^{n}(W_{ns}(T_t)X_{sj})\Big)\delta_t+$$
$$\sum_{t=1}^{n}h_{tj}\Big(m_{ti}-\sum_{s=1}^{n}W_{ns}(T_t)X_{si}\Big)\delta_t+$$
$$\sum_{t=1}^{n}\Big(m_{ti}-\sum_{s=1}^{n}W_{ns}(T_t)X_{si}\Big)\Big(m_{tj}-$$
$$\sum_{s=1}^{n}(W_{ns}(T_t)X_{sj})\Big)\delta_t=\sum_{m=1}^{4}I_{ij}^{(m)}\tag{14}$$

由$(\boldsymbol{X}_i,T_i)$,$1\leqslant i\leqslant n$ 的 i.i.d. 及强大数定律知,对 $1\leqslant i,j\leqslant p$ 有

$$\frac{1}{n}I_{ij}^{(1)}=\frac{1}{n}\sum_{t=1}^{n}h_{ti}h_{tj}\delta_t\to p(x,t)E(h_{1i}h_{1j})=$$
$$p(x,t)\Sigma_{ij},\mathrm{a.s.}\tag{15}$$

式中:$\Sigma_{ij}=\mathrm{Cov}(X_{1i}-E(X_{1i}\mid T_1),X_{1j}-E(X_{1j}\mid T_1))$ 正是 Σ 的第(i,j)元.

$$\frac{1}{n}\sum_{s=1}^{n}|h_{si}\delta_s|\to p(x,t)E|h_{1i}|\leqslant\infty,\mathrm{a.s.}\tag{16}$$

$$E(X_{si}\mid T_s=t)=E(X_{1i}\mid T_1=t)$$
$$1\leqslant s\leqslant n,0\leqslant t\leqslant 1\tag{17}$$

因而由引理 1 及式(16,17)得

$$\frac{1}{n}I_{ij}^{(2)}\leqslant\max_{1\leqslant k\leqslant n}\left|m_{kj}-\sum_{s=1}^{n}W_{ns}(T_k)X_{sj}\right|\times$$
$$\frac{1}{n}\sum_{t=1}^{n}|h_{ti}\delta_t|\to 0,\mathrm{a.s.}$$

$$\frac{1}{n}I_{ij}^{(3)}\leqslant\max_{1\leqslant k\leqslant n}\left|m_{ti}-\sum_{s=1}^{n}W_{ns}(T_t)X_{si}\right|\times$$
$$\frac{1}{n}\sum_{t=1}^{n}|h_{tj}\delta_t|\to 0,\mathrm{a.s.}$$

$$\frac{1}{n}I_{ij}^{(4)}\leqslant\max_{1\leqslant k\leqslant n}\left|m_{ti}-\sum_{s=1}^{n}W_{ns}(T_t)X_{si}\right|\times$$
$$\max_{1\leqslant k\leqslant n}\left|m_{tj}-\sum_{s=1}^{n}W_{ns}(T_t)X_{sj}\right|\to 0,\mathrm{a.s.}$$

由此及式(14,15)得式(12)成立.

引理 3　若条件 1)～6)成立,则

$$R_j=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{ij}\delta_i(g(T_i)-\hat{g}_n(T_i))\xrightarrow{P}0$$

$$S_j=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(m_{ij}-\sum_{k=1}^{n}W_{nk}(T_i)X_{kj}\Big)\delta_i\varepsilon_i\xrightarrow{P}0$$

$$H_j=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(m_{ij}-\sum_{k=1}^{n}W_{nk}(T_i)X_{kj}\Big)\delta_i(g(T_i)-$$
$$\hat{g}_n(T_i))\xrightarrow{P}0$$

证明　此引理的证明类似于文[6]引理 4.2.6 的证明可得到.

引理 4[7]　若条件 1)～4)成立,则

$$\sup_t\left|g(t)-\sum_{i=1}^{n}W_{ni}(t)g(T_i)\right|\leqslant ch^2,\mathrm{a.s.}$$

引理 5[7]　若条件 1)～4)成立,则

$$\max_i\sup_t|W_{ni}(t)|=O((nh_n)^{-1}),\mathrm{a.s.}$$

$$\max_{0\leqslant i\leqslant n}\left|\sum_{k=1}^{n}W_{nk}(T_i)\varepsilon_k\right|\to 0,\mathrm{a.s.}$$
$$1\leqslant j\leqslant p$$

且存在正常数 c 使

$$P\Big(\sup_t\sum_{i=1}^{n}W_{ni}^2(t)\leqslant cn^{-1}h^{-1}\Big)\to 1$$

定理 1 的证明　由式(6,11)显然有

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})=\sqrt{n}[(\tilde{X}'V\tilde{X})^{-1}\tilde{X}'V\tilde{Y}-\boldsymbol{\beta}]$$
$$=\sqrt{n}(\tilde{X}'V\tilde{X})^{-1}(\tilde{X}'V\tilde{G}+\tilde{X}'V\varepsilon)\tag{18}$$

因为 $\tilde{X}'V\tilde{G}$ 的第 j 个分量$(1\leqslant j\leqslant p)$为

$$\sum_{t=1}^{n}\Big(X_{tj}-\sum_{s=1}^{n}W_{ns}(T_t)X_{sj}\Big)(g(T_t)-\hat{g}_n(T_t))\delta_t=$$
$$\sum_{t=1}^{n}h_{tj}(g(T_t)-\hat{g}_n(T_t))\delta_t+$$
$$\sum_{t=1}^{n}\Big(m_{tj}-\sum_{s=1}^{n}W_{ns}(T_t)X_{sj}\Big)(g(T_t)-\hat{g}_n(T_t))\delta_t=$$
$$R_j+H_j$$

所以由引理 3 知:

$$\frac{1}{\sqrt{n}}\tilde{X}'V\tilde{G}\xrightarrow{P}0\tag{19}$$

而引理 2 表明

$$\sqrt{n}(\tilde{X}'V\tilde{X})^{-1/2}\to p^{-1/2}(x,t)\boldsymbol{\Sigma}^{-1/2}$$
$$n(\tilde{X}'V\tilde{X})^{-1}\longrightarrow p^{-1}(x,t)\boldsymbol{\Sigma}^{-1},\mathrm{a.s.}\tag{20}$$

因而$\sqrt{n}(\tilde{X}'V\tilde{X})^{-1}\tilde{X}'V\tilde{G}\xrightarrow{P}0$.故为证式(8),只需证

$\sqrt{n}(\widetilde{X}'V\widetilde{X})^{-1}\widetilde{X}'V\varepsilon \xrightarrow{L} N(\mathbf{0}, \sigma^2 p^{-1}(x,t)\boldsymbol{\Sigma}^{-1})$. 由式(20)和引理 2 知只需证

$$(\widetilde{X}'V\widetilde{X})^{-1/2}\widetilde{X}'V\varepsilon \xrightarrow{L} N(\mathbf{0}, \sigma^2 \boldsymbol{I}_p) \tag{21}$$

即只需证对任意 $p\times 1$ 常数向量 $\boldsymbol{a}=(a_1,\cdots,a_p)'\neq 0$ 有

$$\boldsymbol{a}'(\widetilde{X}'V\widetilde{X})^{-1/2}\widetilde{X}'V\varepsilon \xrightarrow{L} N(\mathbf{0}, \sigma^2 \boldsymbol{a}'\boldsymbol{a}) \tag{22}$$

为此，记 $\boldsymbol{A}_n=(\widetilde{X}'V\widetilde{X})^{-1/2}\widetilde{X}'V=(a_{nij})_{p\times n}$，则由 $\boldsymbol{A}_n\boldsymbol{A}_n'=\boldsymbol{I}_p$ 知：

$$\sum_{k=1}^n a_{nik}a_{njk}=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases} \quad \theta>0 \tag{23}$$

$$\boldsymbol{S}_n \triangleq \boldsymbol{a}'(\widetilde{X}'V\widetilde{X})^{-1/2}\widetilde{X}'V\varepsilon = \sum_{j=1}^n\Big(\sum_{i=1}^p a_{nij}\Big)\varepsilon_j \triangleq \sum_{j=1}^n b_{nj}\varepsilon_j$$

记 $\boldsymbol{F}_{n,j}=\sigma\{(\boldsymbol{X}_1,T_1),\cdots,(\boldsymbol{X}_n,T_n),\varepsilon_1,\cdots,\varepsilon_n\}$，$1\leqslant j\leqslant n$. $\boldsymbol{F}_{n,0}=\sigma\{(\boldsymbol{X}_1,T_1),\cdots,(\boldsymbol{X}_n,T_n)\}$，则由独立性及式(23)易得

$$E(b_{nj}\varepsilon_j\mid \boldsymbol{F}_{n,j-1})=b_{nj}E(\varepsilon_j)=0,\ 1\leqslant j\leqslant n$$

$$\sum_{j=1}^n E(b_{nj}^2\varepsilon_j^2\mid \boldsymbol{F}_{n,j})=\sum_{j=1}^n b_{nj}^2E(\varepsilon_j^2)= \sigma^2\sum_{j=1}^2\Big(\sum_{i=1}^p a_{nij}a_i\Big)^2=\sigma^2\sum_{i=1}^p a_i^2=\sigma^2\boldsymbol{a}'\boldsymbol{a}$$

且对任意 $e>0$ 和任意数 $M>0$

$$\sum_{j=1}^n E(b_{nj}^2\varepsilon_j^2 I(|b_{nj}\varepsilon_j|\geqslant e)\mid \boldsymbol{F}_{n,j-1})\leqslant$$
$$\sum_{j=1}^n E\left\{b_{nj}^2\varepsilon_j^2\left[I\left(|b_{nj}|\geqslant\frac{e}{M}\right)+I(|\varepsilon_j|\geqslant M)\right]\Big|\boldsymbol{F}_{n,j-1}\right\}=$$
$$\sigma^2\sum_{j=1}^n b_{nj}^2 I\left(|b_{nj}|\geqslant\frac{e}{M}\right)+a'aE\varepsilon_1^2I(|\varepsilon_1|\geqslant M)\leqslant$$
$$\sigma^2 a'aI\left(\max_{1\leqslant j\leqslant n}|b_{nj}|\geqslant\frac{e}{M}\right)+a'aE\varepsilon_1^2I(|\varepsilon_1|\geqslant M) \tag{24}$$

因当 M 充分大时式(24)第二项为一充分小量. 由 Dvoretzky 定理即知为证式(22)只需证

$$\max_{1\leqslant j\leqslant n}|b_{nj}|\xrightarrow{P}0 \tag{25}$$

为此，记

$$p^{-1/2}(x,t)\boldsymbol{\Sigma}^{-1/2}=(\gamma_{ij})_{p\times p}$$

$$\boldsymbol{\Delta}_n=\sqrt{n}(\widetilde{X}'V\widetilde{X})^{-1/2}-p^{-1/2}(x,t)\boldsymbol{\Sigma}^{-1/2}=(\delta_{nij})_{p\times p}$$

则由式(20)有 $\max\limits_{1\leqslant i,j\leqslant p}|\delta_{nij}|=o(1)$，a.s.，且显然

$$\max_{1\leqslant i,j\leqslant p}|\gamma_{ij}|\leqslant c<\infty,\ \sqrt{n}a_{nij}=\sum_{k=1}^p(\delta_{nik}+\gamma_{ik})\widetilde{X}_{jk}$$

故由引理 1 和式(13)得

$$\max_{1\leqslant j\leqslant n}|b_{nj}|\leqslant\max_{1\leqslant j\leqslant n}\sum_{k=1}^p\Big|\sum_{i=1}^p(\delta_{nij}+\gamma_{ik})a_i)\widetilde{X}_{jk}\Big|\leqslant$$
$$\frac{c}{\sqrt{n}}\max_{1\leqslant j\leqslant n}\sum_{k=1}^p\Big|h_{jk}+m_{jk}-\sum_{s=1}^n W_{ns}(T_j)X_{sk}\Big|\leqslant$$
$$\max_{1\leqslant j\leqslant n}\sum_{k=1}^p|h_{jk}|+o(1),\quad \text{a.s.}$$

对于充分小得 $e>0$，记 $X_{jk}'=X_{jk}I(|X_{jk}|\leqslant j^{1/2}e^2)$，$h_{ik}'=X_{jk}'-E(X_{jk}'\mid T_j)$，$h_{jk}''=h_{jk}-h_{jk}'$. 则当 n 充分大时得

$$P\{\max_{1\leqslant j\leqslant n}|b_{nj}|\geqslant e\}\leqslant P\Big\{\max_{1\leqslant j\leqslant n}\sum_{k=1}^p|h_{ik}''|\geqslant c\sqrt{n}e\Big\}\leqslant$$
$$\sum_{k=1}^p\sum_{j=1}^n E(X_{jk}^2I(|X_{jk}|\geqslant j^{1/2}e^2))\leqslant$$
$$\frac{c}{\sqrt{n}}\sum_{k=1}^p EX_{1k}^2+c\sum_{k=1}^p E(X_{1k}^2I(|X_{1k}|\geqslant n^{1/4}e^2))\to 0\quad n\to\infty$$

这里用到 $X_{1k},\cdots,X_{nk}$ 同分布，且 $EX_{1k}^2\leqslant c<\infty$，$1\leqslant k\leqslant p$. 由此即得式(25)成立.

定理 2 的证明 因为

$$\hat\sigma_n^2=\frac{1}{n}\sum_{i=1}^n(\widetilde{Y}_i-\widetilde{X}_i\hat{\boldsymbol{\beta}}_n)^2\delta_i/p(\boldsymbol{X}_i,T_i)=$$
$$\frac{1}{n}(\widetilde{Y}-\widetilde{X}\beta)'\boldsymbol{S}(\widetilde{Y}-\widetilde{X}\beta)=$$
$$\frac{1}{n}\varepsilon'V\varepsilon-\frac{1}{n}\varepsilon'V\widetilde{X}(\widetilde{X}'V\widetilde{X})^{-1}\widetilde{X}'S\varepsilon+$$
$$\frac{1}{n}\widetilde{G}'[\boldsymbol{S}-\boldsymbol{S}\widetilde{X}(\widetilde{X}'S\widetilde{X})^{-1}\widetilde{X}'\boldsymbol{S}]\widetilde{G}-$$
$$\frac{2}{n}\widetilde{G}'\boldsymbol{S}\widetilde{X}(\widetilde{X}'\boldsymbol{S}\widetilde{X})^{-1}\widetilde{X}'\boldsymbol{S}\varepsilon+\frac{2}{n}\widetilde{G}'S\varepsilon\triangleq$$
$$J_1-J_2+J_3-2J_4+2J_5 \tag{26}$$

显然有 $\sqrt{n}(J_1-\sigma^2)\xrightarrow{L}N(\mathbf{0},\mathrm{Var}(\varepsilon'\boldsymbol{S}\varepsilon))$ (27)

又类似于式(21)知 $\frac{1}{\sqrt{n}}\widetilde{X}'\boldsymbol{S}\varepsilon=O_p(1)$.

由此和式(19, 20)易见 $\sqrt{n}J_2\xrightarrow{P}0$，$\sqrt{n}J_4\xrightarrow{P}0$.

因为 $\boldsymbol{S}-\boldsymbol{S}\widetilde{X}(\widetilde{X}'\boldsymbol{S}\widetilde{X})^{-1}\widetilde{X}'\boldsymbol{S}\leqslant\boldsymbol{S}$

所以

$$J_3\leqslant\frac{1}{n}\widetilde{G}'\boldsymbol{S}\widetilde{G}=\frac{1}{n}\sum_{i=1}^n(g(T_i)-\hat g_n(T_i))^2\frac{\delta_i}{p(\boldsymbol{X}_i,T_i)}\leqslant\frac{1}{n}\sum_{i=1}^n\Big(g(T_i)-\sum_{k=1}^n W_{nk}(T_i)g(T_k)\Big)^2\frac{1}{p(\boldsymbol{X}_i,T_i)}+$$
$$\frac{1}{n}\sum_{i=1}^n\Big(\sum_{k=1}^n W_{nk}(T_i)\varepsilon_k\Big)^2\frac{1}{p(\boldsymbol{X}_i,T_i)}=J_6+J_7$$

由引理 4 可知 $\sqrt{n}J_6\to 0$.

由引理 5 易得 $\sqrt{n}J_7\xrightarrow{P}0$，故有 $\sqrt{n}J_3\xrightarrow{P}0$.

对于 J_5 有

$$J_5=\frac{1}{n}\sum_{i=1}^{n}(g(T_i)-\hat g_n(T_i))\varepsilon_i\frac{\delta_i}{p(\boldsymbol{X}_i,T_i)}\leqslant$$

$$\frac{1}{n}\sum_{i=1}^{n}(g(T_i)-\sum_{k=1}^{n}W_{nk}(T_i)g(T_k)-$$

$$\sum_{k=1}^{n}W_{nk}(T_i\varepsilon_k)\varepsilon_i\frac{1}{p(\boldsymbol{X}_i,T_i)}\leqslant$$

$$\frac{1}{n}\sum_{i=1}^{n}(g(T_i)-\sum_{k=1}^{n}W_{nk}(T_i)g(T_k))\varepsilon_i\frac{1}{p(\boldsymbol{X}_i,T_i)}-$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(T_i)\varepsilon_k\varepsilon_i\frac{1}{p(\boldsymbol{X}_i,T_i)}=$$

$$\frac{1}{n}\sum_{i=1}^{n}(g(T_i)-\sum_{k=1}^{n}W_{nk}(T_i)g(T_k))\varepsilon_i\frac{1}{p(\boldsymbol{X}_i,T_i)}-$$

$$\frac{1}{n}\sum_{i=1}^{n}W_{ni}(T_i)\varepsilon_i^2\frac{1}{p(\boldsymbol{X}_i,T_i)}-$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k\neq i}W_{nk}(T_i)\varepsilon_k\varepsilon_i\frac{1}{p(\boldsymbol{X}_i,T_i)}\tag{28}$$

对于式(28)可由引理 4 和引理 5 知有 $J_5\xrightarrow{P}0$.所以定理成立.

定理 3 的证明　记 $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_p)'$, $\hat{\boldsymbol{\beta}}_n=(\hat\beta_{n1},\cdots,\hat\beta_{np})'$.由模型(1)求条件期望知 $g(t)=g_1(t)-g_2{}'(t)\boldsymbol{\beta}$, $0\leqslant t\leqslant 1$,于是有

$$\hat g_n^*(t)-g(t)=\hat g_{1n}(t)-g_1(t)-$$
$$\sum_{j=1}^{p}(\hat g_{2nj}(t)-g_{2j}(t))(\hat\beta_{nj}-\beta_j)-$$
$$\sum_{j=1}^{p}(\hat g_{2nj}(t)-g_{2j}(t))\beta_j-\sum_{j=1}^{p}g_{2j}(t)(\hat\beta_{nj}-\beta_j)$$

式中:$\hat g_{2nj}(t)=\sum_{i=1}^{n}W_{ni}(t)X_{ij}$ 是 $\hat g_{2n}(t)$ 的第 j 个分量.因为

$$\hat g_{1n}(t)-g_1(t)=\sum_{i=1}^{n}W_{ni}(t)(Y_i-g_1(T_i))+$$
$$(\sum_{i=1}^{n}W_{ni}(t)g_1(T_i)-g_1(t))\triangleq Q_1+Q_2\tag{29}$$

若记 $Z_i=Y_i-g_1(T_i)$.注意到在给定$\{T_1,\cdots,T_n\}$的条件下 Z_i 与 Z_j 独立($i\neq j$),且 $E(Z_i|T_i)=0$ 有

$$EQ_1^2=E(\sum_{i=1}^{n}W_{ni}^2(t)Z_i^2)+$$
$$E(\sum_{i\neq j}W_{ni}(t)W_{nj}(t)Z_iZ_j)=$$
$$E(\sum_{i=1}^{n}W_{ni}^2(t)E(Z_i^2|T_i))+$$
$$E(\sum_{i\neq j}W_{ni}(t)E(Z_i|T_i)E(Z_j|T_j))=$$
$$E(\sum_{i=1}^{n}W_{ni}^2(t)E(Z_i^2|T_i))$$

由条件 5)有 $E(Z_i^2|T_i)\leqslant E(Y_i^2|T_i)=E(Y_1^2|T_1=t)|_{t=T_i}<c<\infty$,又因 $h=cn^{-1/5}$,且由引理 5 有

$$EQ_1^2\leqslant cn^{-1}h^{-1}=cn^{-4/5}\tag{30}$$

由引理 4 有 $Q_2=O(h^2)$,a.s.由此及式(29,30)可知:

$$\hat g_{1n}(t)-g_1(t)=O(n^{-2/5})\tag{31}$$

同理有

$$\hat g_{2nj}(t)-g_{2j}(t)=O_p(n^{-2/5})\quad 1\leqslant j\leqslant p\tag{32}$$

因此由定理 1 及式(31,32)可得定理成立.

定理 4 的证明　由引理 2、定理 1 和定理 2 得

$$p^{1/2}(x,t)\sqrt{n}(\frac{1}{n}\tilde{X}'V\tilde{X})^{1/2}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})/\hat\sigma_n=$$

$$p^{1/2}(x,t)(\tilde{X}'V\tilde{X})^{1/2}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})/\hat\sigma_n\xrightarrow{L}N(\boldsymbol{0},\boldsymbol{I}_p)$$

故由文[8]中定理 10.2 可得成立.

致谢:本文得到河南省高校青年骨干教师计划(2006141)的资助,在此表示感谢.

参考文献:

[1] 罗双华,庞书侠.缺失数据下局部 M-估计的相合性和渐近正态性 [J].兰州理工大学学报,2006,32(4):145-148.

[2] LIANG H,SUO W.Estimation in partially linear models with missing covariates [J].Journal of the American Statistical Association.2004,99:357-367.

[3] WANG Q H,RAO J.Empirical likelihood for linear regression models under imputation for missing responses [J].The Canadian Journal of Statistics,2001,29:597-608.

[4] WANG Qihua,OLIVER.Semiparametric regression analysis with missing response at random [J].Journal of the American Statistical Association,2004,99:334-345.

[5] FAN J,GIJBELS I.Local polynomial modelling and its applications [M].London:Chapman and Hall,1996.

[6] 柴根象,洪圣岩.半参数回归模型 [M].合肥:安徽教育出版社,1995.

[7] 施云驰,柴根象.半参数回归模型的局部多项式光滑 [J].同济大学学报,2000,28(1):80-83.

[8] ARNOLD S F.Theory of linear models and multivariate analysis [M].New York:John Wiley and Son,1981.