缺失数据情形下半参数回归模型的二阶段估计
Two Step Estimators for Parameters in a Semi-parametric Regression Model with Missing Data

刘　妍[1],李英华[2],秦永松[2],李剑君[2]
LIU Yan[1],LI Ying-hua[2],QIN Yong-song[2],LI Jian-jun[2]

(1. 广西师范大学附属外国语学校,广西桂林　541004;2. 广西师范大学数学科学学院,广西桂林　541004)
(1. Foreign Language School Attached to Guangxi Normal University, Guilin, Guangxi, 541004, China; 2. School of Mathematical Sciences, Guangxi Normal University, Guilin, Guangxi, 541004, China)

摘要:在响应变量有缺失的不完全数据情形下,利用二阶段估计方法得到半参数回归模型 $Y = X'\beta + g(T) + e$ 中参数 β 和非参数 $g(\cdot)$ 的估计,并给出估计渐近正态性的充分条件.

关键词:半参数回归模型　缺失数据　二阶段估计　渐近正态性
中图法分类号:O212.7　　**文献标识码**:A　　**文章编号**:1005-9164(2009)03-0268-06

Abstract: In the case of incomplete samples where the response variable is missing at random, we employ a two-step method to estimate β and $g(\cdot)$ in a semiparametric model $Y = X'\beta + g(T) + e$, and obtain the asymptotic distributions of the estimators.

Key words: semiparametric model, missing data, two step estimator, asymptotic normality

考虑半参数回归模型

$$Y_i = X_i'\beta + g(T_i) + e_i, i = 1,2,\cdots,n, \quad (0.1)$$

其中$\{(X_i,T_i),1\leqslant i\leqslant n\}$独立同分布,$X_i$为$p$维随机向量,$T_i\in[0,1]$,$1\leqslant i\leqslant n$,且$\{X_i,1\leqslant i\leqslant n\}$与$\{T_i,1\leqslant i\leqslant n\}$相互独立,$\beta$为$p$维未知参数,$g(\cdot)$为$[0,1]$上的未知Borel函数,$\{e_i,1\leqslant i\leqslant n\}$为独立同分布随机误差,$Ee_1 = 0$,$0 < \sigma_0^2 = Ee_1{}^2 < \infty$,且$\{e_i,1\leqslant i\leqslant n\}$与$\{(X_i,T_i),1\leqslant i\leqslant n\}$相互独立. 该模型是一类半参数模型,在经济,生物,农业等领域有着较好的应用. 在完全样本情形下,文献[1]采用二阶段估计方法研究模型(0.1)中β,$g(\cdot)$的估计及大样本性质. 然而在实际应用中,由于一些无法避免的原因会导致某些数据无法获得,因此本文在独立不完全样本$\{(X_i,T_i,Y_i,\delta_i),1\leqslant i\leqslant n\}$情形下研究模型(0.1)中$\beta$,$g(\cdot)$的估计及大样本性质,其中$\{(X_i,T_i),1\leqslant i\leqslant n\}$可全部观察到,$\delta_i$为指示$Y_i$是否缺失的指示变量. 当$\delta_i = 1$时,$Y_i$不缺失,当$\delta_i = 0$时,$Y_i$缺失,假定$\{\delta_i,1\leqslant i\leqslant n\}$与$\{T_i,1\leqslant i\leqslant n\}$独立,而且满足缺失机制(MAR):$P(\delta = 1|Y,X,T) = P(\delta = 1|X,T)$.

1　相关条件及引理

给出如下条件:

A1　T_1有密度$r(t)$,且$0 < \inf\limits_{0\leqslant t\leqslant 1} r(t) \leqslant \sup\limits_{0\leqslant t\leqslant 1} r(t) < \infty$.

A2　$g(t)$在$[0,1]$上满足1阶Lipschitz条件.

A3　$E\|X\|^2 < \infty$,其中$E\|X\|^2 = \sum_{j=1}^{p} X_{1j}^2$.

A4　$\Sigma = E[\delta\{X - E(\delta X)/E\delta\}\{X - E(\delta X)/E\delta\}']$为正定阵.

A5　正整数列$\{k = k_n, n\geqslant 1\}$及非负实数$\{v_{ni},1\leqslant i\leqslant n\}$满足(i)$k/(\sqrt{n}\log n)\longrightarrow\infty$,$k/n\longrightarrow 0$,(ii)$\sup_n\{k\max\limits_{1\leqslant i\leqslant k} v_{ni}\} < \infty$,$\max\limits_{k<i\leqslant n} v_{ni} = o(n^{-1})$.

B1　(i)$Eg^2(T) < \infty$,(ii)存在有限函数$M(t)$,使对a. s. $t(\mu)$,存在$\eta = \eta(t) > 0$,当$0 < \rho < \eta$时,有$\int_U |g(u) - g(t)|\mu(du) \leqslant M(t)\rho^\lambda\mu(U)$,其中$\mu$表示$T_1$的分布,$U = \{u: |u - t|\leqslant\rho\}$.

B2　正整数列$\{k = k_n, n\geqslant 1\}$及非负实数$\{v_{ni},1\leqslant i\leqslant n\}$满足(i)$k\longrightarrow\infty$,$k/n\longrightarrow 0$,(ii)对某个实数$0 < \lambda\leqslant 1$,$\lim\limits_{n\longrightarrow\infty}\frac{k^{1+1/2\lambda}}{n} = 0$,(iii)$\sup\limits_n\{k\max\limits_{1\leqslant i\leqslant k} v_{ni}\} < \infty$,

收稿日期:2009-03-05
作者简介:刘　妍(1978-),女,硕士研究生,主要从事数理统计研究。

$\inf_n\{k\sum_{i=1}^n v_{ni}^2\}>0$,(iv) $\sum_{i>k}v_{ni}=O((\sum_{i=1}^n v_{ni}{}^2/n)^{\frac{1}{2}})$.

引理 1 若 $E\|X\|^2<\infty$,则 $n^{-1}R_r\longrightarrow\Sigma$. a. s.

证明 由大数定律和条件 A1 知

$$n^{-1}R_r=\frac{\sum_{i=1}^n\delta_iX_iX_i'}{n}-\frac{\sum_{i=1}^n\delta_iX_i}{n}\cdot\frac{\sum_{i=1}^n\delta_iX_i'}{\sum_{i=1}^n\delta_i}\xrightarrow{a.s.}E\delta XX'-E\delta X\cdot\frac{E\delta X'}{E\delta}=\Sigma.$$

引理 2 设条件 A1 和 A5(i) 成立,则

$$\sup_{0\leqslant t\leqslant 1}|T_{R_{(k,t)}}-t|=O_p(k/n).$$

证明 参见文献[2]中引理 3 的证明.

引理 3 设条件 A1 和 A5 成立,则对 $1\leqslant j\leqslant p$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\cdot\frac{\sum_{s=1}^nW_{ns}(T_i)\delta_s\{g(T_s)-g(T_i)+e_s\}}{\sum_{s=1}^nW_{ns}(T_i)\delta_s}\xrightarrow{P}0.$$

证明 记 $p=E\delta$,则对 i 一致有 $\sum_{s=1}^nW_{ns}(T_i)\delta_s=E\{\delta|T=T_i\}+o_p(n^{-1/4})=p+o_p(n^{-1/4})$,那么

$$\frac{1}{\sum_{s=1}^nW_{ns}(T_i)\delta_s}=\frac{1}{p}+o_p(n^{-1/4}),$$

$$\frac{1}{(\sum_{s=1}^nW_{ns}(T_i)\delta_s)^2}=\frac{1}{p^2}+o_p(n^{-1/4}).$$

考察

$$\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\cdot\frac{\sum_{s=1}^nW_{ns}(T_i)\delta_s\{g(T_s)-g(T_i)+e_s\}}{\sum_{s=1}^nW_{ns}(T_i)\delta_s}=$$

$$\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\cdot\frac{\sum_{s=1}^nW_{ns}(T_i)\delta_s\{g(T_s)-g(T_i)\}}{\sum_{s=1}^nW_{ns}(T_i)\delta_s}+\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\frac{\sum_{s=1}^nW_{ns}(T_i)\delta_se_s}{\sum_{s=1}^nW_{ns}(T_i)\delta_s}\triangleq J_1+J_2.$$

由 $\{(X_i,T_i,\delta_i),1\leqslant i\leqslant n\}$ 与 $\{e_i,1\leqslant i\leqslant n\}$ 独立及 $Ee_i=0$ 得 $EJ_1J_2=0$.

分别考虑 J_2 和 J_1. 由前面的推导知

$$J_2=\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\{p^{-1}+o_p(n^{-1/4})\}\sum_{s=1}^nW_{ns}(T_i)\delta_se_s=p^{-1}\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\sum_{s=1}^nW_{ns}(T_i)\delta_se_s+o_p(n^{-1/4})\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\sum_{s=1}^nW_{ns}(T_i)\delta_se_s\triangleq J_{21}+J_{22}.$$

类似于文献[2]中引理 4 的证明可得 $(EJ_{21}^2)/n\longrightarrow 0$. 又由于

$$|J_{22}|\leqslant o_p(n^{-1/4})\sum_{i=1}^n|\delta_iX_{ij}-E\delta_iX_{ij}|\cdot|\sum_{s=1}^nW_{ns}(T_i)\delta_se_s|=o_p(n^{-1/2})\sum_{i=1}^n|\delta_iX_{ij}-E\delta_iX_{ij}|,$$

故 $\frac{1}{\sqrt{n}}|J_{22}|\leqslant n^{-1}o_p(1)\sum_{i=1}^n|\delta_iX_{ij}-E\delta_iX_{ij}|=o_p(1)$,从而 $\frac{1}{\sqrt{n}}J_2\xrightarrow{P}0$.

$$J_1=p^{-1}\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\sum_{s=1}^nW_{ns}(T_i)\cdot\delta_s(g(T_s)-g(T_i))+\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})\cdot o_p(n^{-1/4})\sum_{s=1}^nW_{ns}(T_i)\delta_s(g(T_s)-g(T_i))\triangleq J_{11}+J_{12},$$

并且

$$EJ_{11}^2=p^{-2}E\{\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})^2[\sum_{s=1}^nW_{ns}(T_i)\delta_s(g(T_s)-g(T_i))]^2\}+p^{-2}E\{\sum_{i_1\neq i_2}(\delta_{i_1}X_{i_1j}-E\delta_{i_1}X_{i_1j})(\delta_{i_2}X_{i_2j}-E\delta_{i_2}X_{i_2j})\cdot\sum_{s=1}^nW_{ns}(T_{i_1})\delta_s(g(T_s)-g(T_{i_1}))\sum_{s=1}^nW_{ns}(T_{i_2})\cdot\delta_s(g(T_s)-g(T_{i_2}))\}\triangleq J_{111}+J_{112}.$$

再类似于文献[2]中引理 4 的证明可得 $\frac{1}{n}J_{111}\longrightarrow 0$. 又 $J_{112}\leqslant o(k^{-2})p^{-2}n^2E|(\delta_1X_{1j}-\delta_1E\delta_1X_{1j})(\delta_2X_{2j}-\delta_2E\delta_2X_{2j})|$,故 $\frac{1}{n}J_{112}\longrightarrow 0$,$\frac{1}{\sqrt{n}}J_{11}\xrightarrow{P}0$.

另外 $J_{12}=\sum_{i=1}^n(\delta_iX_{ij}-E\delta_iX_{ij})o_p(n^{-1/2})$. $|J_{12}|\leqslant o_p(n^{-1/2})\sum_{i=1}^n|\delta_iX_{ij}-E\delta_iX_{ij}|=n^{-1/2}o_p(1)\sum_{i=1}^n|\delta_iX_{ij}-E\delta_iX_{ij}|$,故 $\frac{1}{\sqrt{n}}J_{12}\xrightarrow{P}0$. 引理 3 证明完毕.

引理 4 设 $Y_1,\cdots,Y_n$ iid.,$EY_1^2<\infty$,则

$$\lim_{n\to\infty}\max_{1\leqslant i\leqslant n}\frac{|Y_i|}{\sqrt{n}}=0,\text{a. s.}$$

证明 参见文献[3]中引理 2 的证明.

引理 5 设条件 B1,B2(i ~ iii) 成立,则

$$\sum_{i=1}^kv_{ni}|g(T_{R_{i,t}})-g(t)|/(\sum_{i=1}^kv_{ni}^2)^{1/2}\xrightarrow{P}0.$$

证明 参见文献[4]中引理 2 的证明.

引理 6 设条件 A4,A5 成立,则对任何 $1\leqslant j\leqslant p$,有 $J_n=\max_{1\leqslant i\leqslant n}|\sum_{s=1}^nW_{ns}(T_i)\delta_sX_{sj}-E(\delta_1X_{1j})|\longrightarrow 0$. a. s.

证明 记 $(\delta_sX_{sj})^{(1)}=\delta_sX_{sj}I(|\delta_sX_{sj}|\leqslant s^{1/2})$,$(\delta_sX_{sj})^{(2)}=\delta_sX_{sj}I(|\delta_sX_{sj}|>s^{1/2})$,$(\delta_sX_{sj})^*=$

$(\delta_s X_{sj})^{(1)} - E(\delta_s X_{sj})^{(1)}$, $(\delta_s X_{sj})^{**} = (\delta_s X_{sj})^{(2)} - E(\delta_s X_{sj})^{(2)}$. 则

$$J_n \leqslant \max_{1\leqslant i\leqslant n} \left| \sum_{s=1}^{n} W_{ns}(T_i)(\delta_s X_{sj})^{*} \right| + \max_{1\leqslant i\leqslant n} \left| \sum_{s=1}^{n} W_{ns}(T_i)(\delta_s X_{sj})^{**} \right| \triangleq J_{n1} + J_{n2}.$$

用 $\widetilde{P},\widetilde{E}$ 分别表示在给定 $\{T_1,\cdots,T_n\}$ 的条件下所取的概率和期望. 注意到在给定 $\{T_1,\cdots,T_n\}$ 的条件下, $W_{n1}(T_i)(\delta_1 X_{1j})^{*},\cdots,W_{nn}(T_i)(\delta_n X_{nj})^{*}$ 相互独立. 当 $1\leqslant s\leqslant n$ 时,有

$$\widetilde{E}[W_{ns}(T_i)(\delta_s X_{sj})^{*}] = E[W_{ns}(T_i)((\delta_s X_{sj})^{(1)} - E(\delta_s X_{sj})^{(1)}) | \{T_1,\cdots,T_n\}] = W_{ns}(T_i)E(\delta_s X_{sj})^{*} = 0,$$

而且

$$\max_{1\leqslant i,s\leqslant n} |W_{ns}(T_i)(\delta_s X_{sj})^{*}| \leqslant 2\sqrt{n} \cdot \max_{1\leqslant s\leqslant n} W_{ns}(T_i) \leqslant 2\sqrt{n} \max_{1\leqslant i,s\leqslant n} v_{ns} \leqslant 2\frac{\sqrt{n}}{k} \triangleq b_n,$$

$$B_n^2 = \sum_{i=1}^{n} \widetilde{E}[W_{ns}(T_i)(\delta_s X_{sj})^{*}]^2 \leqslant EX_{sj}^2 \sum_{s=1}^{n} v_{ns}^2 \leqslant C\cdot\frac{1}{k}.$$

由 Bernstein 不等式知,对任意 $\epsilon>0$,有

$$P(J_{n1}\geqslant\epsilon) = E\{\widetilde{P}(J_{n1}\geqslant\epsilon)\} \leqslant E\{\sum_{i=1}^{n}\widetilde{P}(|\sum_{s=1}^{n}W_{ns}(T_i)(\delta_s X_{sj})^{*}|\geqslant\epsilon)\} \leqslant 2\sum_{i=1}^{n}\exp\{-\epsilon^2/(2B_n+2b_n\epsilon)\} \leqslant 2n\exp\{-Ck/\sqrt{n}\}.$$

由 Borel-Cantelli 引理可得 $J_{n1}\longrightarrow 0$. a. s. 由条件 A5 知, $\max_{1\leqslant i,s\leqslant n} W_{ns}(T_i) \leqslant \max_{1\leqslant s\leqslant k} v_{ns} + \max_{k\leqslant s\leqslant n} v_{ns} \leqslant C\cdot\frac{1}{k}$. 注意到 $EX_{sj}^2<\infty \Rightarrow E(\delta_s X_{sj})^2<\infty \Rightarrow \sum_{s=1}^{n} P(|\delta_s X_{sj}|>s^{1/2})<\infty \Rightarrow P(|\delta_s X_{sj}|>s^{1/2},\text{i. o.})=0 \Rightarrow \sum_{i=1}^{n}|\delta_s X_{sj}|I(|\delta_s X_{sj}|>s^{1/2})<\infty$, a. s. 因此

$$J_{n2} \leqslant \max_{1\leqslant i\leqslant n} W_{ns}(T_i) | \sum_{s=1}^{n} |(\delta_s X_{sj})^{(2)}| + \sum_{s=1}^{n} E|(\delta_s X_{sj})^{(2)}|| \leqslant \frac{C}{k} | \sum_{s=1}^{n} |\delta_s X_{sj}| I(|\delta_s X_{sj}|>s^{1/2}) + EX_{sj}^2 \sum_{s=1}^{n} s^{-1/2}| \leqslant \frac{C}{k}\sum_{s=1}^{n}|\delta_s X_{sj}|I(|\delta_s X_{sj}|>s^{1/2}) + \frac{C\sqrt{n}}{k} \longrightarrow 0, \text{a. s.}$$

故引理 6 成立.

2 主要结果

2.1 模型(0.1)中的 β 和 $g(\cdot)$ 的估计

令 $\alpha = Eg(T_i)$, $\epsilon_i = g(T_i) - \alpha + e_i$, $i\geqslant 1$. 则模型(0.1)可以化为 $Y_i = \alpha + X_i'\beta + \epsilon_i$, $1\leqslant i\leqslant n$,其中 $\epsilon_1,\cdots,\epsilon_n$ iid. ,且 $E\epsilon_1 = 0$, $0<\sigma^2 = E\epsilon_1^2 = \mathrm{Var}[g(T)] + \sigma_0^2 < \infty$.

缺失数据情形下,模型(0.1)转化为 $\delta_i Y_i = \delta_i\alpha + \delta_i X_i'\beta + \delta_i\epsilon_i$, $1\leqslant i\leqslant n$. 因而 α 和 β 最小二乘估计(其中 β 的估计为初始估计)为

$$\hat{\alpha}_r = \sum_{i=1}^{n}\{\frac{1}{\sum_{i=1}^{n}\delta_i} - \frac{\sum_{i=1}^{n}\delta_i X'_i}{\sum_{i=1}^{n}\delta_i}\cdot R_r^{-1}\cdot (X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})\}\delta_i Y_i,$$

$$\hat{\beta}_r = R_r^{-1}\cdot\sum_{i=1}^{n}\delta_i(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})\cdot (Y_i - \frac{\sum_{i=1}^{n}\delta_i Y_i}{\sum_{i=1}^{n}\delta_i}),$$

其中

$$R_r = \sum_{i=1}^{n}\delta_i(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})'.$$

用 $\hat{\beta}_r$ 代替模型(0.1)中的 β,得缺失数据下的非参数模型 $\delta_i Y_i \approx \delta_i X_i'\hat{\beta}_r + \delta_i g(T_i) + \delta_i e_i$, $1\leqslant i\leqslant n$,即 $g(t) \approx \frac{E\{\delta(Y - X'\hat{\beta}_r)|T=t\}}{E\{\delta|T=t\}}$, $t\in[0,1]$. 故定义 $g(t)$ 的最近邻估计(最终估计)为 $\widetilde{g}_r(t) = \frac{\sum_{i=1}^{n}\delta_i(Y_i - X_i'\hat{\beta}_r)W_{ni}(t)}{\sum_{i=1}^{n}\delta_i W_{ni}(t)}$,其中 $\{W_{ni}(t),1\leqslant i\leqslant n\}$ 为一列近邻权函数,满足 $W_{nR_{i,t}}(t) = v_{ni}$, $i=1,\cdots,n$, $\{v_{ni},1\leqslant i\leqslant n\}$ 是给定的一组非负实数,满足 $\sum_{i=1}^{n} v_{ni} = 1$, $\{R_{i,t},1\leqslant i\leqslant n\}$ 为 $1,\cdots,n$ 的一个排列,满足 $|T_{R_{1,t}} - t| \leqslant\cdots\leqslant |T_{R_{n,t}} - t|$,并按小足标在前的方式消结.

将 $\widetilde{g}_r(t)$ 代入模型(0.1)中,得到缺失数据下的线性模型 $\delta_i Y_i \approx \delta_i X_i'\beta + \delta_i\widetilde{g}_r(T_i) + \delta_i e_i$, $1\leqslant i\leqslant n$. 从而得到 β 的最小二乘估计(最终估计)为 $\widetilde{\beta}_r = R_r^{-1}\cdot \sum_{i=1}^{n}\delta_i\cdot(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})(Y_i - \widetilde{g}_r(T_i))$.

2.2 估计的渐近正态性

定理 1 设条件 A1,A2,B1(i) 满足,则

$$\sqrt{n}(\hat{\alpha}_r - \alpha) \xrightarrow{L} N(0,\sigma^2(\frac{E\delta X}{E\delta})'\Sigma^{-1}(\frac{E\delta X}{E\delta}) + \sigma^2(E\delta)^{-1}).$$

证明 易知

$$\hat{\alpha}_r - \alpha = \sum_{i=1}^{n}[\frac{1}{\sum_{i=1}^{n}\delta_i} - \frac{\sum_{i=1}^{n}\delta_i X_i'}{\sum_{i=1}^{n}\delta_i}\cdot R_r^{-1}\cdot$$

$(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})]\delta_i\epsilon_i \triangleq \sum_{i=1}^{n} a_{ni}\epsilon_i.$

记 $V_n = \sum_{i=1}^{n} b_{ni}\epsilon_i, b_{ni} = a_{ni}/(\sigma^2 \sum_{i=1}^{n} a_{ni}^2)^{\frac{1}{2}}$. 由引理1可得 $n^{-1}R_r \xrightarrow{a.s.} \Sigma^{-1}$，因此

$$n\sum_{i=1}^{n} a_{ni}^2 = \frac{n}{\sum_{i=1}^{n}\delta_i} + \frac{\sum_{i=1}^{n}\delta_i X_i'}{\sum_{i=1}^{n}\delta_i} nR_r^{-1} \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i}$$

$$\xrightarrow{a.s.} (E\delta)^{-1} + (\frac{E\delta X}{E\delta})' \sum^{-1} (\frac{E\delta X}{E\delta}).$$

只需证明

$$V_n \xrightarrow{L} N(0,1), \qquad (2.1)$$

就可以由上述推导知定理1成立. 为此，记 $F_n = (X_1, \cdots, X_n)$. 由 $\{X_i, 1 \leqslant i \leqslant n\}$ 与 $\{\epsilon_i, 1 \leqslant i \leqslant n\}$ 独立，有 $E(b_{ni}\epsilon_i | F_n) = b_{ni}E\epsilon_i = 0, i = 1, \cdots, n, \sum_{i=1}^{n} E(b_{ni}^2 \epsilon_i^2 | F_n) = \sum_{i=1}^{n} b_{ni}^2 E\epsilon_i^2 = 1$. 对任意 $\epsilon > 0, M > 0$，有

$$\sum_{i=1}^{n} E\{b_{ni}^2\epsilon_i^2 I(|b_{ni}\epsilon_i| \geqslant \epsilon) | F_n\} \leqslant \sigma^2 \sum_{i=1}^{n} E\{b_{ni}^2 I(|b_{ni}| \geqslant \epsilon/M)\} + \sigma^{-2}E\{\epsilon_1^2 I(|\epsilon_1| \geqslant M)\} \leqslant I(\max_{1\leqslant i\leqslant n} |b_{ni}| \geqslant \epsilon/M) + \sigma^2 E\epsilon_1^2 I(|\epsilon_1| \geqslant M).$$

因为当 M 充分大时上式右端第2项为一无穷小量. 由 Dvoretzky 定理知，为证明(2.1)式，只需证明 $\max_{i=1} |b_{ni}| \xrightarrow{a.s.} 0$，即只需要证明 $\sqrt{n} \max_{1\leqslant i\leqslant n} |a_{ni}| \xrightarrow{a.s.} 0$. 利用引理4可得

$$\sqrt{n} \max_{1\leqslant i\leqslant n} |a_{ni}| \leqslant \frac{\sum_{i=1}^{n}\delta_i \| nX_i \|}{\sum_{i=1}^{n}\delta_i} \cdot nR_r^{-1} \cdot (\frac{\max_{1\leqslant i\leqslant n} \| X_i \|}{\sqrt{n}} + \frac{1}{\sqrt{n}} \cdot \frac{\sum_{i=1}^{n}\delta_i \| X_i \|}{\sum_{i=1}^{n}\delta_i}) + \sqrt{n} \cdot \frac{1}{\sum_{i=1}^{n}\delta_i} \xrightarrow{a.s.} 0.$$

故定理1证明完毕.

定理 2 设条件 A3, A4, B1(i) 满足，则 $\sqrt{n}(\hat{\beta}_r - \beta) \xrightarrow{L} N(0, \sigma^2\Sigma^{-1})$.

证明 易知

$$\hat{\beta}_r - \beta = R_r^{-1}\{\sum_{i=1}^{n}(X_i - \frac{E\delta X}{E\delta})\delta_i\epsilon_i - (\frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i} - \frac{E\delta X}{E\delta})\sum_{i=1}^{n}\delta_i\epsilon_i\}.$$

由于 $\{(X_i - \frac{E\delta X}{E\delta})\delta_i\epsilon_i, 1 \leqslant i \leqslant n\}$ 是均值为0，协方差阵为 $\sigma^2\Sigma$ 的 iid. 随机变量序列，故由多维随机变量的中心极限定理和强大数定律可得 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \frac{E\delta X}{E\delta})\delta_i\epsilon_i \xrightarrow{L} N(0, \sigma^2\Sigma)$，$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\epsilon_i \xrightarrow{L} N(0, \sigma^2 E\delta)$，$\frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i} \xrightarrow{a.s.} \frac{E\delta X}{E\delta}$. 再利用引理1和 Slutzky 定理可知定理2成立.

定理 3 设条件 A3, A4, B1, B2 满足，则 $(\tilde{g}_r(t) - g(t))/(\sigma_0{}^2 \sum_{i=1}^{k} v_{ni}^2/E\delta)^{1/2} \xrightarrow{L} N(0, 1)$.

证明 由于

$$\tilde{g}_r(t) - g(t) = \frac{\sum_{i=1}^{n}\delta_i(Y_i - X_i'\hat{\beta}_r)W_{ni}(t)}{\sum_{i=1}^{n}\delta_i W_{ni}(t)} - g(t) = \frac{\sum_{i=1}^{n} W_{ni}(t)\delta_i e_i}{\sum_{i=1}^{n} W_{ni}(t)\delta_i} + \frac{\sum_{i=1}^{n} W_{ni}(t)\delta_i\{g(T_i) - g(t)\}}{\sum_{i=1}^{n} W_{ni}(t)\delta_i} - \frac{\sum_{i=1}^{n} W_{ni}(t)\delta_i X_i'(\hat{\beta}_r - \beta)}{\sum_{i=1}^{n} W_{ni}(t)\delta_i} \triangleq L_{n1} + L_{n2} + L_{n3}.$$

令 $p = E\delta$，先证明 $L_{n1}/(\sigma_0^2 p^{-1} \sum_{i=1}^{k} v_{ni}^2)^{1/2} \xrightarrow{L} N(0, 1)$. 由于 $L_{n1} = p^{-1}\sum_{i=1}^{n} W_{ni}(t)\delta_i e_i + o_p(n^{-1/4}) \cdot \sum_{i=1}^{n} W_{ni}(t)\delta_i e_i \triangleq L_{n11} + L_{n12}$，故只需证明

$$\sum_{i=1}^{n} W_{ni}(t)\delta_i e_i/(\sigma_0^2 p \sum_{i=1}^{k} v_{ni}^2)^{1/2} \xrightarrow{L} N(0,1). \quad (2.2)$$

因 $\sum_{i=1}^{n} W_{ni}(t)\delta_i e_i = \sum_{i=1}^{k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}} + \sum_{i>k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}}$，由 B2 及引理4可得

$$\frac{|\sum_{i>k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}}|}{(\sigma_0^2 p \sum_{i=1}^{k} v_{ni}^2)^{1/2}} \leqslant C \frac{\max_{1\leqslant i\leqslant n} |e_i|}{\sqrt{n}} \xrightarrow{a.s.} 0.$$

故要证明(2.2)式，只需证明

$$Q_n \triangleq \sum_{i=1}^{k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}}/(\sigma_0^2 p \sum_{i=1}^{k} v_{ni}^2)^{1/2} \xrightarrow{L} N(0,1).$$

为此，记

$$W_{nR_{i,t}}^*(t) = v_{ni}^* = \begin{cases} v_{ni}, 1 \leqslant i \leqslant k, \\ 0, k < i \leqslant n. \end{cases}$$

则 $Q_n = \sum_{i=1}^{n} W_{nR_{i,t}}^*(t)\delta_{R_{i,t}} e_i/(\sigma_0{}^2 p \sum_{i=1}^{k} v_{ni}^2)^{\frac{1}{2}}$. 注意到 $\{(X_i, T_i), 1 \leqslant i \leqslant n\}$ 与 $\{e_i, 1 \leqslant i \leqslant n\}$ 独立，$Ee_1 = 0$. 由 Dvoretzky 定理可以证明 $Q_n \xrightarrow{L} N(0,1)$，从而(2.2)式成立，继而有 $L_{n11}(\sigma_0^2 p^{-1} \sum_{i=1}^{k} v_{ni}^2)^{1/2} \xrightarrow{L} N(0, 1)$.

再证明 $L_{n12}/(\sigma_0^2 p^{-1} \sum_{i=1}^{k} v_{ni}^2)^{1/2} = o_p(1)$. 因为

$$\frac{L_{n12}}{(\sigma_0{}^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} = \frac{o_p(n^{-1/4})\sum_{i=1}^{k} v_{ni}\delta_{R_{i,t}} e_i}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} + \frac{o_p(n^{-1/4})\sum_{i>k} v_{ni}\delta_{R_{i,t}} e_i}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} \triangleq L'_{n12} + L''_{n12}.$$

由于

$$E\left[\frac{n^{-1/4}\sum_{i=1}^{k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}}}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}}\right]^2 \leqslant n^{-1/2}\ \frac{\sigma_0^2\sum_{i=1}^{k} v_{ni}^2}{\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2} = n^{-1/2},$$

所以

$$\frac{n^{-1/4}\sum_{i=1}^{k} v_{ni}\delta_{R_{i,t}} e_{R_{i,t}}}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} = o_p(1), L'_{n12} = o_p(1).$$

同时 $|L''_{n12}| \leqslant o_p(n^{-1/4})\ \frac{\max\limits_{1\leqslant i\leqslant n}|e_i|}{\sqrt{n}} = o_p(1)$, $L_{n12}/(\sigma_0^2 p^{-1}\sum\limits_{i=1}^{k} v_{ni}^2)^{1/2} = o_p(1)$，故 $L_{n1}/(\sigma_0^2 p^{-1}\cdot \sum_{i=1}^{k} v_{ni}^2)^{1/2} \xrightarrow{L} N(0,1)$. 又由 B1、B2 及引理 4 和引理 5 得

$$\frac{|L_{n2}|}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} \leqslant \frac{(p^{-1} + o_p(n^{-1/4}))\sum_{i=1}^{k} v_{ni}|g(T_{R_{i,t}}) - g(t)|}{(\sigma_0^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} + C\frac{\max\limits_{1\leqslant i\leqslant n}|g(T_{R_{i,t}}) - g(t)|}{\sqrt{n}} \xrightarrow{P} 0.$$

最后考虑 L_{n3}. 记 $\beta = (\beta_1,\cdots,\beta_p)'$, $\hat\beta_r = (\hat\beta_{r1},\cdots,\hat\beta_{rp})$，则有

$$L_{n3} = p^{-1}\sum_{j=1}^{p}\sum_{i=1}^{n} W_{ni}(t)\delta_i X_{ij}(\hat\beta_{rj} - \beta_j) + o_p(n^{-1/4})\sum_{j=1}^{p}\sum_{i=1}^{n} W_{ni}(t)\delta_i X_{ij}(\hat\beta_{rj} - \beta_j).$$

$$E\left[\sum_{i=1}^{n} W_{ni}(t)(\delta_i X_{ij} - E(\delta_i X_{ij}))\right]^2 = E(\delta_i X_{ij} - E(\delta_i X_{ij}))^2(\sum_{i=1}^{n} v_{ni})^2 \leqslant \frac{C}{k} \longrightarrow 0.$$

于是 $\sum_{i=1}^{n} W_{ni}(t)\delta_i X_{ij} \xrightarrow{P} E(\delta_i X_{ij})$, $1\leqslant j\leqslant p$. 再由定理 2 得 $L_{n3} = O_p(n^{-1/2})$，从而可得

$$\frac{L_{n3}}{(\sigma_0{}^2 p^{-1}\sum_{i=1}^{k} v_{ni}^2)^{1/2}} \xrightarrow{P} 0.$$

综合 L_{n1}, L_{n2}, L_{n3}，即可得定理 3.

定理 4 设条件 A1 ～ A5 满足，则

$$\sqrt{n}(\hat\beta_r - \beta) \xrightarrow{L} N(0, \sigma_0{}^2\Sigma^{-1}).$$

证明 由于

$$\sqrt{n}(\hat\beta_r - \beta) = -\sqrt{n}R_r^{-1}\sum_{i=1}^{n}\delta_i(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})\frac{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i) + e_s\}}{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s} + \sqrt{n}R_r^{-1}\sum_{i=1}^{n}\delta_i(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})\cdot \frac{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s X'_s(\hat\beta_r - \beta)}{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s} + \sqrt{n}R_r^{-1}\sum_{i=1}^{n}\delta_i(X_i - \frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i})e_i \triangleq -nR_r{}^{-1}J_{n1} + nR_r{}^{-1}J_{n2} + J_{n3}.$$

首先考虑 J_{n1}. 由于 J_{n1} 的第 j 个分量为

$$J_{n1}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\delta_i X_{ij} - E\delta_i X_{ij})\cdot \frac{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i) + e_s\}}{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s} - \frac{\sum_{i=1}^{n}\delta_i X_{ij}}{\sum_{i=1}^{n}\delta_i}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\delta_i - E\delta_i)\cdot \frac{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i) + e_s\}}{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(E\delta_i\frac{\sum_{i=1}^{n}\delta_i X_{ij}}{\sum_{i=1}^{n}\delta_i} - E\delta_i X_{ij})\cdot \frac{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i) + e_s\}}{\sum_{s=1}^{n} W_{ns}(T_i)\delta_s} \triangleq J_{n4} - J_{n5} - J_{n6}.$$

由引理 3 可知 $J_{n4} \xrightarrow{P} 0$. 由于 $\frac{\sum_{i=1}^{n}\delta_i X_{ij}}{\sum_{i=1}^{n}\delta_i} \xrightarrow{a.s.} \frac{E\delta_1 X_{1j}}{E\delta_1}$，又由引理 3 可得 J_{n5} 中除去 $\frac{\sum_{i=1}^{n}\delta_i X_{ij}}{\sum_{i=1}^{n}\delta_i}$ 的部分依概率收敛于 0，因而 $J_{n5} \xrightarrow{P} 0$. 再由中心极限定理有 $\sum_{i=1}^{n}\delta_i/n = E\delta + o_p(n^{-1/2})$，因而得 $\frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i} = \frac{E\delta X}{E\delta} + o_p(n^{-1/2})$，从而 $E\delta_i\frac{\sum_{i=1}^{n}\delta_i X_i}{\sum_{i=1}^{n}\delta_i} - E\delta_i X_i = o_p(n^{-\frac{1}{2}})$，所以

$$J_{n6} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} o_p(n^{-1/2})(p^{-1} + o_p(n^{-1/4}))\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i) + e_s\} = \frac{1}{n}\sum_{i=1}^{n} o_p(1)\sum_{s=1}^{n} W_{ns}(T_i)\delta_s\{g(T_s) - g(T_i)\} + \frac{1}{n}\sum_{i=1}^{n} o_p(1)\sum_{s=1}^{n} W_{ns}(T_i)\delta_s e_s \triangleq J_{n61} + J_{n62}.$$

由于对一切 i 有 $|\sum_{s=1}^{n} W_{ns}(T_i)\delta_s(g(T_s) - g(T_i))|$

$\leqslant \sup\limits_{0\leqslant t\leqslant 1}|T_{R_{k,t}}-t|+C\sum_{s>k}v_{ns}=O_p(\frac{k}{n})$，$|J_{n61}|\leqslant o_p(1)\ \frac{1}{n}\sum_{i=1}^{n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_s\{g(T_s)\ -\ g(T_i)\}|\ \leqslant o_p(1)O_p(\frac{k}{n})=o_p(1)$. 又由于$\{e_i,1\leqslant i\leqslant n\}$ iid. 且与$\{(X_i,T_i),1\leqslant i\leqslant n\}$相互独立，$Ee_i=0$，故

$$E(\sum_{s=1}^{n}W_{ns}(T_i)\delta_s e_s)^2=\sum_{s=1}^{n}E(W_{ns}(T_i)^2\delta_s^2e_s^2)\leqslant\sum_{s=1}^{n}v_{ns}^2\sigma_0^2=O(1/k),$$

因此

$$|J_{n62}|\leqslant|\frac{1}{n}\sum_{i=1}^{n}o_p(\frac{1}{\sqrt{k}})|\leqslant o_p(\frac{1}{\sqrt{k}}),$$

所以 $J_{n6}\xrightarrow{P}0$.

再考虑 J_{n2}. 记 $\beta=(\beta_1,\cdots,\beta_p)'$，$\hat{\beta}_r=(\hat{\beta}_1,\cdots,\hat{\beta}_p)'$，可以得到

$$J_{n2}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot\frac{\sum_{s=1}^{n}\sum_{l=1}^{p}W_{ns}(T_i)\delta_sX_{sl}(\hat{\beta}_{rl}-\beta_l)}{\sum_{s=1}^{n}W_{ns}(T_i)\delta_s}.$$

注意到$\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})=0$，利用大数定律和引理6可得

$$|\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\frac{\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}}{\sum_{s=1}^{n}W_{ns}(T_i)\delta_s}|$$

$$\leqslant|\frac{1}{n}\sum_{i=1}^{n}o_p(n^{-1/4})\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\sum_{s=1}^{n}W_{ns}(T_i)\cdot\delta_sX_{sl}|+|p^{-1}\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}\triangleq J_{21}+J_{22},$$

$$J_{21}=o_p(n^{-1/4})\frac{1}{n}\sum_{i=1}^{n}|\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}|\leqslant 2o_p(n^{-1/4})\cdot\frac{1}{n}\sum_{i=1}^{n}|X_i|\ \max_{1\leqslant i\leqslant n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}\ -\ E(\delta X_{1l})|\ =o_p(1),$$

$$J_{22}=p^{-1}|\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot(\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}-E(\delta_1X_{1l}))|\leqslant 2p^{-2}\cdot\frac{1}{n}\sum_{i=1}^{n}|X_i|\max_{1\leqslant i\leqslant n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}-E(\delta_1X_{1l})|\xrightarrow{P}0,$$

最后考虑 J_{n3}. 用类似于证明定理2的方法可以证明 $J_{n3}\xrightarrow{L}N(0,\sigma_0{}^2\Sigma^{-1})$.

综上所述，定理4成立.

注 文献[1]中定理1～4是本文结果($\delta_i=1$)的特例.

参考文献：

[1] 薛留根，韩建国. 半参数回归模型中二阶段估计的渐近性质[J]. 高校应用数学学报：A辑，2001，16(1)：87-94.

[2] 洪圣岩. 一类半参数回归模型的估计理论[J]. 中国科学：A辑，1991，21(12)：1258-1272.

[3] 赵林城，白志东. 非参数回归函数最近邻估计的强相合性[J]. 中国科学：A辑，1984，14(5)：387-393.

[4] 洪圣岩. 最近邻回归估计的渐近正态性[J]. 应用概率统计，1991，7(2)：187-191.

(责任编辑：尹　闯)

(上接第267页 Continue from page 267)

注 当 $\delta_i\equiv 1$ 时，由本文的结果可以推出文献[2]和文献[4]的主要结果. 易知 $\sigma_2^2\leqslant\sigma_1^2$，即附加信息下的估计比未含附加信息的估计更渐近有效.

致谢：

感谢秦永松教授给予的指导和帮助.

参考文献：

[1] 郑忠国. 条件中位数的最近邻估计和它的Bootstrap统计量的渐进性质[J]. 中国科学：A辑，1984，12：1074-1089.

[2] Liu Z J，Tu D S. Kernel method on conditona median estimation [J]. Chinese Science Bulletin，1987，5：642-643.

[3] Xiang X J. A kernel estimation of a conditional quantile [J]. Multivarate Anal，1996，59：206-216.

[4] 秦永松，苏淳. 附加信息时条件分位数的估计及其渐进性质应[J]. 应用数学学报，2000，23(1)：55-62.

[5] 范承华. 缺失数据半参数回归分析[D]. 北京：北京工业大学，2007.

[6] Qin J，Lawless J. Empirical likelihood and general estimating equation[J]. Ann Statist，1994，22：300-325.

[7] Owen A B. Empirical likelihood confidence regions[J]. Ann Statist，1990，18：90-120.

[8] Chow Y S，Teicher H. Probabbility theory[M]. New York：Spring-Verlag，1988.

(责任编辑：尹　闯)