文章编号:1671-9352(2010)08-0123-04

核实数据下响应变量缺失的线性模型均值估计

宇世航
(齐齐哈尔大学理学院, 黑龙江 齐齐哈尔 161006)

摘要:借助核实数据,构造了回归系数的最小二乘估计,然后用借补方法和加权借补方法估计响应变量的均值,最后证明了估计量的渐近正态性。

关键词:随机缺失;线性 EV 模型;加权借补估计;渐近正态性
中图分类号:O212 **文献标志码**:A

Estimators of the mean of linear errors-in-variables models under validation data for missing response data

YU Shi-hang
(Department of Maths, Qiqihar University, Qiqihar 161006, Heilongjiang, China)

Abstract: The least squares estimate for the unknown regression coefficients are constructed with the help of validation data. Then input estimators are used to estimate the mean of the response. It is shown that the proposed estimators are asymptotically normal.

Key words: missing atrandom; linear errors-in-variables model; weighted imputation estimators; asymptotically normal

0 引言

考虑线性模型

$$Y = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon。 \tag{0.1}$$

其中, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}}$ 是 $p \times 1$ 的回归系数向量, $\boldsymbol{X}$ 是 p 维随机向量, ε 是随机误差。

通常把带有变量误差的模型称为 EV(errors-variables)模型,称之为测量误差模型。已有不少文献[1-2]对线性测量误差的线性 EV 模型进行了研究,这种模型假定 $\tilde{\boldsymbol{X}} = \boldsymbol{X} + e$,这里 $\boldsymbol{X}$ 是不能被直接观测的解释变量, $\tilde{\boldsymbol{X}}$ 是一个可观测变量。但是,在一般情况下 $\tilde{\boldsymbol{X}}$ 与 $\boldsymbol{X}$ 之间具有很复杂的关系,例如 $\tilde{\boldsymbol{X}} = \varphi(\boldsymbol{X}, e)$,这里 e 是测量误差且与 $(\boldsymbol{X}, \boldsymbol{Y})$ 独立, φ 是一个任意已知的函数。在这种情况下要对未知参数 $\boldsymbol{\beta}$ 进行有效统计推断是相当困难的。基于替代数据与核实数据的统计推断已引起国内外学者的重视。例如:Sepanski 与 Lee 和 Xue 研究了基于核实数据的非线性 EV 模型[3-4],Wang 在这方面做了许多工作[5-6]。

事实上经常会遇到响应变量缺失的情况,这时通常的统计推断不能直接应用,通常处理缺失数据的方法是对每个缺失的响应变量值进行借补,然后利用标准方法和借补后的完全数据进行研究。对于线性模型在响应变量随机缺失和删失的情形,Wang 和 Qin 分别对这两种情况进行了研究[7-8]。而对协变量带有测量误差且响应变量缺失的情况,目前的文献还较少。本文借助于核实数据,利用借补方法和加权借补方法估计响应变量的均值,并证明了其估计量的渐近正态性。

收稿日期:2009-10-26
基金项目:黑龙江省教育厅科学技术研究项目(11551543)
作者简介:宇世航(1972-),女,硕士,副教授,主要研究方向为数理统计、概率极限理论. Email: qqhrysh@163.com

1 回归系数的估计

假设主要数据$\{(\tilde{\boldsymbol{X}}_i, Y_i, \delta_i)_{i=1}^n\}$是独立同分布(i. i. d)的样本,且与 i. i. d 的核实样本$\{(\boldsymbol{X}_i, \tilde{\boldsymbol{X}}_i)_{i=n+1}^{n+m}\}$相互独立,其中当 Y_i 缺失时 $\delta_i=0$,否则 $\delta_i=1$。进一步假定 $E(\varepsilon|\tilde{\boldsymbol{X}})=0$,令 $u(\tilde{\boldsymbol{X}})=E(\boldsymbol{X}|\tilde{\boldsymbol{X}})$,则基于完全观测数据,模型(0.1)可改写为

$$\delta Y=\delta u^{\mathrm{T}}(\tilde{\boldsymbol{X}})\boldsymbol{\beta}+\delta\eta, \tag{1.1}$$

其中 $\eta=\varepsilon+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}-u^{\mathrm{T}}(\tilde{\boldsymbol{X}})\boldsymbol{\beta}$。

利用核实数据,在 $\tilde{\boldsymbol{X}}=\tilde{x}$ 的情况下用 X 的核回归来估计模型(1.1)的回归系数 $u(\tilde{\boldsymbol{X}})$,$u(\tilde{\boldsymbol{X}})$的估计定义为

$$\hat{u}_m(\tilde{x})=\frac{\sum_{i=n+1}^{n+m}\boldsymbol{X}_i K\left(\frac{\tilde{\boldsymbol{X}}_i-\tilde{x}}{h_m}\right)}{\sum_{i=n+1}^{n+m}K\left(\frac{\tilde{\boldsymbol{X}}_i-\tilde{x}}{h_m}\right)}, \tag{1.2}$$

其中 $K(\cdot)$为核函数,h_m 是收敛于0的窗宽。

定义$\boldsymbol{\beta}$的估计量是

$$S_{m,n}(\boldsymbol{\beta})=\frac{1}{n}\sum_{j=1}^{n}\delta_j(Y_j-\hat{u}_m(\tilde{\boldsymbol{X}}_j)\boldsymbol{\beta})^2+\frac{1}{m}\sum_{i=n+1}^{n+m}\delta_i(Y_i-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})^2 \tag{1.3}$$

的最小值。由式(1.3)可得$\boldsymbol{\beta}$的估计量为

$$\hat{\boldsymbol{\beta}}_{m,n}=\hat{\boldsymbol{\Sigma}}_{m,n}^{-1}\hat{\boldsymbol{A}}_{m,n}, \tag{1.4}$$

这里,$\hat{\boldsymbol{\Sigma}}_{m,n}=\frac{1}{n}\sum_{j=1}^{n}\delta_j\hat{u}_m(\tilde{\boldsymbol{X}}_j)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_j)+\frac{1}{m}\sum_{i=n+1}^{n+m}\delta_i\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}$, $\hat{\boldsymbol{A}}_{m,n}=\frac{1}{n}\sum_{j=1}^{n}\delta_j\hat{u}_m(\tilde{\boldsymbol{X}}_j)Y_j+\frac{1}{m}\sum_{i=n+1}^{n+m}\delta_i\boldsymbol{X}_iY_i$。

令 D^k 是定义在 $\mathbf{R}^P$ 上的连续函数族,并且满足$\frac{\partial^{i_1}}{\partial x_1^{i_1}}\frac{\partial^{i_2}}{\partial x_2^{i_2}}\cdots\frac{\partial^{i_p}}{\partial x_p^{i_p}}f(x_1,\cdots,x_p)$一致有界。对任何向量$\boldsymbol{\alpha}$,用$\|\boldsymbol{\alpha}\|$表示 Euclidean 模。给出正则条件(下列条件统称为条件 A):

$(A\cdot X)E\|\boldsymbol{X}\|^2<\infty$。

$(A\cdot Y)EY^2<\infty$。

$(A\cdot\varepsilon)\sup_{\tilde{x}}E[\varepsilon^2|\tilde{\boldsymbol{X}}=\tilde{x}]<\infty$。

$(A\cdot u)u(\cdot)\in D^k, k>p+1$。

$(A\cdot K)K(\cdot)$是非负有界的 k 阶核函数,且具有有界支撑。

$(A\cdot\tilde{X})$(1)存在一正常数序列 η_n 满足 $nP(f_{\tilde{x}}(\tilde{\boldsymbol{X}}<\eta_n)\to 0$;(2) $f_{\tilde{x}}\in D^k$。

$(A\cdot h_m)$ (1) $mh_m^{2p}\eta_n^2\to\infty$;(2) $mh_m^{2k}\to 0$。

$(A\cdot nm)\frac{n}{m}\to\lambda, \lambda\geqslant 0$ 为常数。

$(A\cdot\boldsymbol{\Sigma})\boldsymbol{\Sigma}=E[\delta u(\tilde{\boldsymbol{X}})u^{\mathrm{T}}(\tilde{\boldsymbol{X}})]+E[\delta\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$为正定。

定理 1.1 假设条件 A 成立,则有

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{d}N(0,\boldsymbol{\Sigma}^{-1}V\boldsymbol{\Sigma}^{\mathrm{T}}),$$

其中 $V=E[\delta u(\tilde{\boldsymbol{X}})u^{\mathrm{T}}(\tilde{\boldsymbol{X}})(Y-u^{\mathrm{T}}(\tilde{\boldsymbol{X}})\boldsymbol{\beta})^2]+\lambda\{E[\delta u(\tilde{\boldsymbol{X}})u^{\mathrm{T}}(\tilde{\boldsymbol{X}})[\delta((\boldsymbol{X}-u^{\mathrm{T}}(\tilde{\boldsymbol{X}}))^{\mathrm{T}}\boldsymbol{\beta})]^2]+E[\delta\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}(Y-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^2]+2E[(Y-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})((\boldsymbol{X}-u^{\mathrm{T}}(\tilde{\boldsymbol{X}}))^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{X}u^{\mathrm{T}}(\tilde{\boldsymbol{X}})]\}$。

这里$\hat{\boldsymbol{\beta}}_{m,n}$的渐近方差由 $\hat{\boldsymbol{V}}=\hat{\boldsymbol{\Sigma}}_{n,m}^{-1}[\hat{\boldsymbol{V}}_1+\hat{\boldsymbol{V}}_2]\hat{\boldsymbol{\Sigma}}_{n,m}^{\mathrm{T}}$一致估计,$\hat{\boldsymbol{\Sigma}}_{n,m}^{-1}$同式(1.4)定义,并且

$$\hat{\boldsymbol{V}}_1=\frac{1}{n}\sum_{j=1}^{n}[\delta_j\hat{u}_m(\tilde{\boldsymbol{X}}_j)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_j)(Y_j-\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_j)\hat{\boldsymbol{\beta}}_{m,n})^2],$$

$$\hat{\boldsymbol{V}}_2=\frac{n}{m^2}\sum_{i=1}^{m}\{[\delta_i\hat{u}_m(\tilde{\boldsymbol{X}}_i)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)[\delta_i((\boldsymbol{X}_i-\hat{u}_i^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i))^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{m,n})]^2]+[\delta_i\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}(Y_i-\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{m,n})^2]+2[(Y_i-\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{m,n})((\boldsymbol{X}_i-\hat{u}_i^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i))^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{m,n})\boldsymbol{X}_i\hat{u}_i^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)]\}。$$

2　响应均值的估计

记响应均值 $EY=\theta$,下面基于上述回归系数的估计量 $\hat{\boldsymbol{\beta}}_{m,n}$,给出 θ 的估计量及其渐近性质。

定义 2.1　$\hat{\theta}_I=\frac{1}{n}\sum_{i=1}^{n}\{\delta_i Y_i+(1-\delta_i)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n}\}$。

定义 2.2　$\hat{\theta}_p=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\hat{P}(\tilde{\boldsymbol{X}}_i)}Y_i+\left(1-\frac{\delta_i}{\hat{P}(\tilde{\boldsymbol{X}}_i)}\right)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n}\right\}$,

这里 $\hat{P}(\tilde{\boldsymbol{X}}_i)=\dfrac{\sum_{i=1}^{n}\delta_i K\left(\dfrac{\tilde{\boldsymbol{X}}_i-\tilde{x}}{b_n}\right)}{\sum_{i=1}^{n}K\left(\dfrac{\tilde{\boldsymbol{X}}_i-\tilde{x}}{b_n}\right)}$为 $P(\tilde{\boldsymbol{X}})=P(\delta_i=1\mid\tilde{\boldsymbol{X}}_i=\tilde{x})$ 的核估计。

定理 2.1　在条件 A 成立下,有

$$\sqrt{n}(\hat{\theta}_I-\theta)\xrightarrow{d}N(0,\boldsymbol{V}_I),$$

其中 $\boldsymbol{V}_I=\boldsymbol{\Sigma}_1+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_3\boldsymbol{\beta}-2\boldsymbol{\Sigma}_4^{\mathrm{T}}\boldsymbol{\beta}\theta+\theta^2+\boldsymbol{\Sigma}_2^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}})\boldsymbol{\Sigma}_2+\boldsymbol{\Sigma}_2^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}})\boldsymbol{\Sigma}_5$。记 $\boldsymbol{\Sigma}_1=E[\delta(Y-u^{\mathrm{T}}(\tilde{\boldsymbol{X}})\boldsymbol{\beta})^2]$, $\boldsymbol{\Sigma}_2=E[(1-\delta)u(\tilde{\boldsymbol{X}})]$, $\boldsymbol{\Sigma}_3=E[u(\tilde{\boldsymbol{X}})u^{\mathrm{T}}(\tilde{\boldsymbol{X}})]$, $\boldsymbol{\Sigma}_4=E[u(\tilde{\boldsymbol{X}})]$, $\boldsymbol{\Sigma}_5=E[\delta u(\tilde{\boldsymbol{X}})]$。这里 θ_I 的渐近方差 $\boldsymbol{V}_I$ 的估计 $\hat{\boldsymbol{V}}_I$ 为

$$\hat{\boldsymbol{V}}_I=\hat{\boldsymbol{\Sigma}}_1+\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_3\hat{\boldsymbol{\beta}}-2\hat{\boldsymbol{\Sigma}}_4^{\mathrm{T}}\hat{\boldsymbol{\beta}}\hat{\theta}+\hat{\theta}^2+\hat{\boldsymbol{\Sigma}}_2^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Sigma}}^{\mathrm{T}})\hat{\boldsymbol{\Sigma}}_2+\hat{\boldsymbol{\Sigma}}_2^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Sigma}}^{\mathrm{T}})\hat{\boldsymbol{\Sigma}}_5,$$

且

$$\hat{\boldsymbol{\Sigma}}_1=\frac{1}{n}\sum_{i=1}^{n}[\delta_i(Y_i-\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}})^2],\ \hat{\boldsymbol{\Sigma}}_2=\frac{1}{n}\sum_{i=1}^{n}[(1-\delta_i)\hat{u}_m(\tilde{\boldsymbol{X}}_i)],$$

$$\hat{\boldsymbol{\Sigma}}_3=\frac{1}{n}\sum_{i=1}^{n}[\hat{u}_m(\tilde{\boldsymbol{X}}_i)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)],\hat{\boldsymbol{\Sigma}}_4=\frac{1}{n}\sum_{i=1}^{n}[\hat{u}_m(\tilde{\boldsymbol{X}}_i)],\hat{\boldsymbol{\Sigma}}_5=\frac{1}{n}\sum_{i=1}^{n}[\delta_i\hat{u}_m(\tilde{\boldsymbol{X}}_i)]。$$

定理 2.2　在条件 A 成立下,当 $nb_n^4\to0$, $nb_n\to\infty$ 时,有

$$\sqrt{n}(\hat{\theta}_p-\theta)\xrightarrow{d}N(0,\boldsymbol{V}_p),$$

其中,$\boldsymbol{V}_P=\boldsymbol{\Sigma}_{1P}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_3\boldsymbol{\beta}-2\boldsymbol{\Sigma}_4^{\mathrm{T}}\boldsymbol{\beta}\theta+\theta^2+\boldsymbol{\Sigma}_{2P}^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}})\boldsymbol{\Sigma}_{2P}+\boldsymbol{\Sigma}_{2P}^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}})\boldsymbol{\Sigma}_{5P}$,且

$$\boldsymbol{\Sigma}_{1P}=E\left(\frac{\delta(Y-u^{\mathrm{T}}(\tilde{\boldsymbol{X}})\boldsymbol{\beta})^2}{P(\tilde{\boldsymbol{X}})}\right)=\frac{\boldsymbol{\Sigma}_1}{E(\delta\mid\tilde{\boldsymbol{X}})},\ \boldsymbol{\Sigma}_{2P}=E\left(u(\tilde{\boldsymbol{X}})-\frac{E(\delta u\mid\tilde{\boldsymbol{X}})}{E(\delta\mid\tilde{\boldsymbol{X}})}\right),\ \boldsymbol{\Sigma}_{5P}=\frac{\boldsymbol{\Sigma}_5}{E(\delta\mid\tilde{\boldsymbol{X}})}。$$

3　定理的证明

定理 1.1 的证明类似于文献[6]中定理 2.1 的证明。

定理 2.1 的证明

$$\begin{aligned}\sqrt{n}(\hat{\theta}_I-\theta)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(\delta_i Y_i+(1-\delta_i)\hat{u}_m^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n})-\theta]=\\&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(Y_i-u^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\boldsymbol{\beta})+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)u^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)(\hat{\boldsymbol{\beta}}_{m,n}-\boldsymbol{\beta})+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(u^{\mathrm{T}}(\tilde{\boldsymbol{X}}_i)\boldsymbol{\beta}-\theta)+\\&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)(\hat{u}_m(\tilde{\boldsymbol{X}}_i)-u(\tilde{\boldsymbol{X}}_i))^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{m,n}=\\&B_{n1}+B_{n2}+B_{n3}+o_P(1),\end{aligned}$$

由中心极限定理,显然有

$$B_{n1}\xrightarrow{d}N(0,\boldsymbol{\Sigma}_1),$$

$$B_{n3}\xrightarrow{d}N(0,\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_3\boldsymbol{\beta}-2\boldsymbol{\Sigma}_4^{\mathrm{T}}\boldsymbol{\beta}\theta+\theta^2)。$$

再由定理 1.1,有

$$B_{n2} \xrightarrow{d} N(0, \boldsymbol{\Sigma}_2^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-T})\boldsymbol{\Sigma}_2)。$$

经简单推导得

$$\mathrm{cov}(B_{n1}, B_{n3}) = 0,\ \mathrm{cov}(B_{n2}, B_{n3}) = 0,\ \mathrm{cov}(B_{n1}, B_{n2}) = \boldsymbol{\Sigma}_2^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-\mathrm{T}})\boldsymbol{\Sigma}_5。$$

定理得证。

定理 2.2 的证明

$$\sqrt{n}(\hat{\theta}_P - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i}{\hat{P}(\widetilde{\boldsymbol{X}}_i)}Y_i + \left(1 - \frac{\delta_i}{\hat{P}(\widetilde{\boldsymbol{X}}_i)}\right)\hat{u}_m^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n} - \theta\right] =$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i}{P(\widetilde{\boldsymbol{X}}_i)}Y_i + \left(1 - \frac{\delta_i}{P(\widetilde{\boldsymbol{X}}_i)}\right)\hat{u}_m^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n} + \right.$$

$$\left.\frac{\delta_i[\hat{P}(\widetilde{\boldsymbol{X}}_i) - P(\widetilde{\boldsymbol{X}}_i)]}{P^2(\widetilde{\boldsymbol{X}}_i)}(Y_i - \hat{u}_m^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\hat{\boldsymbol{\beta}}_{m,n}) - \theta\right] + o_P(1) =$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{P(\widetilde{\boldsymbol{X}}_i)}(Y_i - u^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\boldsymbol{\beta}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{P(\widetilde{\boldsymbol{X}}_i)}\right)u^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)(\hat{\boldsymbol{\beta}}_{m,n} - \boldsymbol{\beta}) +$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(u^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\boldsymbol{\beta} - \theta) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i[\hat{P}(\widetilde{\boldsymbol{X}}_i) - P(\widetilde{\boldsymbol{X}}_i)]}{P^2(\widetilde{\boldsymbol{X}}_i)}(Y_i - u^{\mathrm{T}}(\widetilde{\boldsymbol{X}}_i)\boldsymbol{\beta}) + o_P(1) =$$

$$T_{n1} + T_{n2} + T_{n3} + T_{n4} + o_P(1)。$$

类似文献[7]的证明有 $T_{n4} = o_p(1)$,

由中心极限定理有

$$T_{n1} \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{1P}),$$

$$T_{n3} \xrightarrow{d} N(0, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_3\boldsymbol{\beta} - 2\boldsymbol{\Sigma}_4^{\mathrm{T}}\boldsymbol{\beta}\theta + \theta^2)。$$

再由定理 1.1 有

$$T_{n2} \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{2p}^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-\mathrm{T}})\boldsymbol{\Sigma}_{2p})。$$

由 $$E\left(\frac{\delta(Y - u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta})}{P(\widetilde{\boldsymbol{X}})}(u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta} - \theta)\right) = E\left\{E\left[\left(\frac{\delta(Y - u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta})}{P(\widetilde{\boldsymbol{X}})}(u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta} - \theta)\right) \middle| Y, \widetilde{\boldsymbol{X}}\right]\right\} =$$

$$E\lfloor(Y - u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta})(u^{\mathrm{T}}(\widetilde{\boldsymbol{X}})\boldsymbol{\beta} - \theta) \mid \widetilde{\boldsymbol{X}}\rfloor = 0$$

可知 $\mathrm{cov}(T_{n1}, T_{n3}) = 0$,同理可得 $\mathrm{cov}(T_{n2}, T_{n3}) = 0$, 从而易知 $\mathrm{cov}(T_{n1}, T_{n2}) = \boldsymbol{\Sigma}_{2P}^{\mathrm{T}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-\mathrm{T}})\boldsymbol{\Sigma}_{5P}$。定理得证。

参考文献:

[1] FULLER W A. Measurement error models[M]. New York: Wiley, 1987.

[2] CUI H J, CHEN S X. Empirical likelihood confibence region for parameter in the reeors-in-variables models[J]. Multivariate Anal, 2003, 84(1):101-115.

[3] SEPANSKI J H, LEE L F. Semiparametric estimation of nonlinear error-in-varables models with validation study [J]. Nonparametric Statist, 1995, 4:365-394.

[4] XUE L G. Empirical likelihood inference in nonlinear semiparemetric EV models with validation data[J]. Acta Mathematica Sinica, Chinese Series, 2006, 49(1):145-154.

[5] WANG Q H. Dimension reduction in partly linear error-in-response models with validation data[J]. Multivariate Anal, 2003, 85:234-252.

[6] WANG Q H. Estimation of partial linear error-in-variables models with validation data[J]. Multivariate Anal, 1999, 69:30-64.

[7] WANG Q H, LINTON Oliver. Semiparametric regression analysis with missing response at random[J]. Amer Statist Assoc, 2004, 99:334-345.

[8] QIN G S, JING B Y. Empirical likelihood for censored linear regression[J]. Scan J Statist, 2001, 28:661-673.

(编辑:孙培芹)