doi:10.16055/j.issn.1672-058X.2020.0003.006

响应缺失的部分函数型 EV 回归模型的统计推断*

方 连 娣[1,2]

(1. 铜陵学院 数学与计算机学院, 安徽 铜陵 244000;2. 江苏大学 电气信息工程学院,江苏 镇江 212013)

摘 要:针对响应变量随机缺失且解释变量带有测量误差的部分函数型线性回归模型,讨论了模型中未知参数和未知系数函数的估计问题及其渐近性质;先通过一定方法对缺失数据和带有测量误差的数据进行处理,然后将模型转化为一般的函数型线性回归模型,再利用 Karhumen-Loevez 展开和主成分分析法给出模型的经验形式,最后运用经典的多元统计分析极小化目标函数得到相应未知量的最小二乘估计,并在一定的条件下给出了参数估计量的渐近正态性和斜率函数估计量的收敛速度;从而说明给出的估计量是有效估计,完全观测下的函数型数据统计推断方法可以被推广到不完全观测的情形。

关键词:函数型数据;测量误差;K-L 展开;响应缺失

中图分类号:O212.7 **文献标志码**:A **文章编号**:1672-058X(2020)03-0042-05

0 引 言

近年来,函数型线性回归模型引起了人们的广泛兴趣,很多学者对其进行了研究[1-4]。实际中,我们观测到的数据可能既有函数型的又有向量型的,带有这样混合数据的模型是常有的。鉴于此,文献[5]提出如下部分函数型线性回归模型:

$$Y = Z^{\mathrm{T}}\theta + \int_0^1 X(t)\beta(t)\,\mathrm{d}t + e \tag{1}$$

其中,Y 是实值响应变量,Z 是 q 维随机向量且 $EZ=0$,$E(Z^{\mathrm{T}}Z)<\infty$;$X(t)\in L^2[0,1]$ 是随机过程且 $EX=0$,$E(\|X\|^2)=E\int_0^1 X^2(t)\,\mathrm{d}t<\infty$;$\theta$ 和 $\beta(t)$ 分别是 q 维未知回归参数和未知系数函数,且 $\beta(t)\in L^2[0,1]$,$\|\beta\|^2<\infty$;e 是独立于(X,Z)的随机误差且 $E(e)=0$,$\mathrm{var}(e)=\sigma^2<\infty$。

模型式(1)结合了经典的多元线性模型和函数型线性模型,具有独特的优越性,引起了很多学者的关注,比如文献[6]通过把空间中的函数型数据进行 Karhumen-Loeve 展开(K-L 展开),给出模型式(1)中系数函数的估计量及其渐近性质。文献[7]对模型式(1)提出了岭估计方法,并证明参数分量的岭估计量具有渐近正态性以及斜率函数的岭估计量与最小二乘估计量的收敛速度是一致的。文献[8]在模型误差序列为平稳 α 混合序列情形下,给出了未知参数和斜率函数的估计方法,进一步建立了参数估计量的渐近正态性和斜率函数估计量的收敛速度。

以上成果都是基于完全观测数据下的,而在现实生活中,观测的实验数据往往是不完全的,出现缺失或者带有测量误差。如:在工业试验中,由于出现与试验无关的机械故障造成试验结果的缺失;医学研究中,由于医学仪器的精密度而使得对某项指标的观测往往含有测量误差等。对于缺失数据情形下的函数型线性模型也已取得一定成果。例

收稿日期:2019-09-04;**修回日期**:2019-10-25.

* **基金项目**:安徽省高校科学研究重点项目(KJ2018A0477);安徽省高校优秀拔尖人才培育资助项目(GXYQ2018089).

作者简介:方连娣(1982—),女,安徽枞阳人,副教授,博士研究生,从事统计分析研究.

如文献[9]利用可观测样本曲线去补偿函数型数据缺失的部分;文献[10]在平稳遍历和响应变量随机丢失的情况下研究了非参数回归模型系数函数的估计问题及其大样本性质。而关于解释变量有测量误差或两种不完全观测同时出现情形下的函数型线性模型研究尚少,对模型式(1)在不完全观测下的研究也尚未有见。基于以上的讨论,参考面板数据的相关成果[11-12],本文对线性解释变量 Z 带有测量误差且响应变量 Y 随机缺失的模型式(1)进行研究。具体来说,一方面,真实变量 Z 不能被观测,只能观测到替代变量 W 的值,两变量间存在关系 $\boldsymbol{W}=Z+\boldsymbol{v}$,这里 $\boldsymbol{v}$ 是测量误差向量,与 Z,X 相互独立且满足 $E\boldsymbol{v}=0,\mathrm{cov}(\boldsymbol{v})=\Xi$ 已知;另一方面,响应变量 Y 是随机缺失的,假定随机缺失变量为 δ,若 Y 可观测,则 $\delta=1$,否则 $\delta=0$,且

$$P(\delta=1|X,Y,Z,e)=P(\delta=1|X,Z)$$

这样,模型式(1)的观测数据就是$(X,Y,W;\delta)$,它是不完全。本文中,首先,基于这样的不完全观测,对数据进行预处理,将模型转换为类似于完全观测情形下的模型。然后,再运用 K-L 展开、主成分分析和多元最小二乘估计给出模型中未知参数和未知系数函数的估计量。最后,在一定的条件下给出所有估计量的渐近性质和收敛速度并证明。

1　估　计

基于不完全观测$(X,Y,W;\delta)$,由模型式(1)可得:

$$\delta Y=\delta W^{\mathrm{T}}\theta+\delta\langle X(t),\beta(t)\rangle+\delta e-\delta\boldsymbol{v}^{\mathrm{T}}\theta \tag{2}$$

记 $\tilde{Y}=\delta Y,\tilde{W}=\delta W,\tilde{X}(\cdot)=\delta X(\cdot),\tilde{e}=\delta e,\tilde{\boldsymbol{v}}=\delta\boldsymbol{v},\varepsilon=\delta e-\delta\boldsymbol{v}^{\mathrm{T}}\theta$,则模型式(2)可转化为

$$\tilde{Y}=\tilde{W}^{\mathrm{T}}\theta+\langle\tilde{X}(t),\beta(t)\rangle+\tilde{e}-\tilde{\boldsymbol{v}}^{\mathrm{T}}\theta \tag{3}$$

令 $\tilde{X}$ 的协方差函数为 $\Psi_{\tilde{X}}(s,t)=\mathrm{cov}(\tilde{X}(s),\tilde{X}(t))$,$\Psi_{\tilde{X}}$ 的协方差算子 $\varphi_{\tilde{X}}=\langle\tilde{X}(s),\Psi_{\tilde{X}}(s,t)\rangle$。定义 $\mu_1\geqslant\mu_2\geqslant\cdots\geqslant0$ 和 $\xi_1,\xi_2,\cdots$分别是 $\varphi_{\tilde{X}}$ 非负特征值和相应的特征函数,满足

$$\varphi_{\tilde{X}}(\xi_k)=\mu_k\xi_k(\forall k=1,2,\cdots)$$

再令

$$\Psi_{\tilde{W}}=\mathrm{var}(\tilde{W}),\Psi_{\tilde{W}\tilde{Y}}=\mathrm{cov}(\tilde{W},\tilde{Y})$$

$$\Psi_{\tilde{Y}\tilde{X}}=\mathrm{cov}(\tilde{Y},\tilde{X}(\cdot)),\Psi_{\tilde{W}\tilde{X}}=\mathrm{cov}(\tilde{W},\tilde{X}(\cdot))$$

现在假定$(X_i,Y_i,W_i;\delta_i)_{1\leqslant i\leqslant n}$是来自模型式(1)的 i. i. d. 观测样本,则有 $\tilde{Y}_i=\delta_iY_i,\tilde{W}_i=\delta_iW_i,\tilde{X}_i(\cdot)=\delta_iX_i(\cdot),\tilde{e}_i=\delta_ie_i,\tilde{\boldsymbol{v}}_\iota=\delta_i\boldsymbol{v}_\iota,\varepsilon_i=\delta_i e_i-\delta_i v_\iota^{\mathrm{T}}\theta$。易知 $\Psi_{\tilde{X}}(s,t)$ 和 $\varphi_{\tilde{X}}$ 的估计可分别定义为 $\hat{\Psi}_{\tilde{X}}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i(s)\tilde{X}_i(t)$ 和 $\hat{\phi}_{\tilde{X}}(t)=\langle\tilde{X}(s),\hat{\Psi}_{\tilde{X}}(s,t)\rangle$。相应地,$\hat{\phi}_{\tilde{X}}(t)$的特征值和特征函数$(\hat{\mu}_k,\hat{\xi}_k)$即为$(\mu_k,\xi_k)$的估计$(k=1,2,\cdots)$。同样有

$$\hat{\Psi}_{\tilde{W}}=\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_i\tilde{W}_i^{\mathrm{T}},\hat{\Psi}_{\tilde{W}\tilde{Y}}=\frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_i\tilde{W}_i$$

$$\hat{\Psi}_{\tilde{Y}\tilde{X}}(t)=\frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_i\tilde{X}_i(t),\hat{\Psi}_{\tilde{W}\tilde{X}}(t)=\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_i\tilde{X}_i(t)$$

下面开始考虑未知系数函数和未知参数的估计。由 Karhunen-Loeve 展开可知 $\beta(t)=\sum_{k=1}^{\infty}\rho_k\xi_k(t),\rho_k=\langle\beta(t),\xi_k(t)\rangle;\tilde{X}(t)=\sum_{k=1}^{\infty}\lambda_k\xi_k(t)$,其中 $\lambda_k=\langle\tilde{X}(t),\xi_k(t)\rangle$ 互不相关,且 $E\lambda_k=0,\mathrm{var}(\lambda_k)=\mu_k$。利用主成分分析,给定某个足够大的正整数 $\mathbf{N}$,模型式(3)可写成

$$\tilde{Y}=\tilde{W}\theta+V_N\rho+\tilde{e}-\tilde{\boldsymbol{v}}\theta \tag{4}$$

其中,

$\tilde{Y}=(\tilde{Y}_1,\tilde{Y}_2,\cdots\tilde{Y}_n)^{\mathrm{T}},\tilde{W}=(\tilde{W}_1,\tilde{W}_2\cdots\tilde{W}_n)^{\mathrm{T}},\rho=(\rho_1,\rho_2,\cdots,\rho_N)^{\mathrm{T}},\tilde{e}=(\tilde{e}_1,\tilde{e}_2,\cdots,\tilde{e}_n)^{\mathrm{T}},\boldsymbol{v}=(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n)^{\mathrm{T}}$,
$V_N=\{\langle X_i,\xi_k\rangle\}_{1\leqslant i\leqslant n,1\leqslant k\leqslant N}$

运用经典多元统计分析给出感兴趣未知量的最小二乘估计,极小化如下目标函数

$$F(\theta,\rho)=\|\tilde{Y}-\tilde{W}\theta-V_N\rho\|^2-\theta^{\mathrm{T}}\tilde{\Xi}\theta$$

则有 $\hat{\theta}=[\tilde{W}^{\mathrm{T}}(I-U_N)\tilde{W}-n\tilde{\Xi}]^{-1}\tilde{W}(I-U_N)\tilde{Y}\hat{\rho}=(V_N^{\mathrm{T}}V_N)^{-1}V_N^{\mathrm{T}}(\tilde{Y}-\tilde{W}\hat{\theta})$,其中 $U_N=V_N(V_N^{\mathrm{T}}V_N)^{-1}V_N^{\mathrm{T}},\tilde{\Xi}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\Xi$。即

$$\hat{\theta}=\left(\hat{\Psi}_{\tilde{W}}-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\hat{\Psi}_{\tilde{X}\tilde{W}},\hat{\xi}_k\rangle}{\hat{\mu}_k}-n\tilde{\Xi}\right)^{-1}\times\left(\hat{\Psi}_{\tilde{W}\tilde{Y}}-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\hat{\Psi}_{\tilde{Y}\tilde{X}},\hat{\xi}_k\rangle}{\hat{\mu}_k}\right) \tag{5}$$

其中,

$$\langle\hat{\boldsymbol{\Psi}}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle=\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_i\langle\tilde{X}_i,\hat{\xi}_k\rangle$$

$$\langle\hat{\boldsymbol{\Psi}}_{\tilde{Y}\tilde{X}},\hat{\xi}_k\rangle=\frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_i\langle\tilde{X}_i,\hat{\xi}_k\rangle$$

又知道

$$\hat{\rho}=(\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_N)^{\mathrm{T}}$$

$$\hat{\rho}_k=\frac{\langle\hat{\boldsymbol{\Psi}}_{\tilde{Y}\tilde{X}}-\hat{\theta}^{\mathrm{T}}\hat{\boldsymbol{\Psi}}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle}{\hat{\mu}_k},k=1,2,\cdots,N$$

由此可得 $\hat{\beta}(t)=\sum_{k=1}^{N}\hat{\rho}_k(t)\hat{\xi}_k(t)$。

为进一步得到本文的结果,给出正则条件如下条件 A:

(A1) $E\|X\|^4<\infty$, $Ee^4<\infty$, $E\|\boldsymbol{v}\|^4<\infty$;

(A2) 设 $b>1$,对于 $\forall k\geqslant1$, $c^{-1}k^{-b}\leqslant\mu_k\leqslant ck^{-b}$ 且 $\mu_k-\mu_{k+1}\geqslant ck^{-b-1}$,这里 c 为常数;

(A3) $E\|W\|_{R^q}^4=E[(W^{\mathrm{T}}W)^2]<\infty$,并且对 $\forall k\geqslant1$,对每个 i 都有 $|\langle\boldsymbol{\Psi}_{\tilde{W}\tilde{X}},\xi_k\rangle|\leqslant ck^{-(b+d)}$,这里 $d>\frac{b}{2}+1$,c 为常数;

(A4) 对于 $\forall k$, $E[\lambda_k^4]\leqslant c\mu_k^2$,这里 c 为常数;

(A5) $d>\frac{b}{2}+1$, $\|\rho_k\|\leqslant ck^{-d}$, c 为常数;

(A6) $N\sim n^{\frac{1}{b+d}}$;

(A7) 设 $\eta_{ij}=\tilde{W}_{ij}-\langle f_j,\tilde{W}_i\rangle$, $i=1,2,\cdots,n$, $j=1,2,\cdots,q$,假定对每个 j, $\eta_{1j},\eta_{2j},\cdots,\eta_{nj}$ 是 i. i. d. ,使得 $E[\eta_{1j}\mid\tilde{W}_1,\tilde{W}_2,\cdots,\tilde{W}_n]=0$ 且 $E[\eta_{1j}^2\mid\tilde{W}_1,\tilde{W}_2,\cdots,\tilde{W}_n]=D_{jj}$,这里 D_{jj} 为 $D=E[\eta_1\eta_1^{\mathrm{T}}]=\boldsymbol{\Psi}_{\tilde{W}}-\sum_{k=1}^{\infty}\frac{\langle\boldsymbol{\Psi}_{\tilde{W}\tilde{X}},\xi_k\rangle\langle\boldsymbol{\Psi}_{\tilde{X}\tilde{W}},\xi_k\rangle}{\mu_k}$ 的第 j 个对角线元素,其中 $\eta_i=(\eta_{i1},\eta_{i2},\cdots,\eta_{iq})^{\mathrm{T}}$, $f_j=\sum_{k=1}^{\infty}\frac{\langle\boldsymbol{\Psi}_{\tilde{W}_j\tilde{X}},\xi_k\rangle}{\mu_k}\xi_k$。

进一步地,假定 $\boldsymbol{D}$ 是一个正定阵。

定理 1 当条件 A 成立时,有

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{L}N(0,(\frac{1}{n}\boldsymbol{D}-\tilde{\Xi})^{-1}\Sigma(\frac{1}{n}\boldsymbol{D}-\tilde{\Xi})^{-1})$$

其中 $\Sigma=(\sigma^2+\theta^{\mathrm{T}}\tilde{\Xi}\theta)D+\tilde{\Xi}\theta\theta^{\mathrm{T}}\tilde{\Xi}^{\mathrm{T}}$。

定理 2 条件 A 成立时,有

$$\|\hat{\beta}(t)-\beta(t)\|^2=O_p(n^{-\frac{2d-1}{b+2d}})$$

注:定理 1 和定理 2 说明,在不完全观测下模型式(1)的回归参数的估计也是渐近正态的,系数函数估计的收敛速度仍是如文献[6]的速度,未受到线性回归项的影响。

2 定理证明

在开始定理证明之前,类似于文献[6]给出几个相关引理。

引理 1 假设条件 A 成立,则

$$E\|\hat{\boldsymbol{\Phi}}_{\tilde{X}}-\boldsymbol{\Phi}_{\tilde{X}}\|_\infty^2\leqslant\frac{1}{n}E\|X\|^4$$

$$E\|\hat{\boldsymbol{\Psi}}_{\tilde{W}}-\boldsymbol{\Psi}_{\tilde{W}}\|_\infty^2\leqslant\frac{1}{n}E\|W\|_{R^q}^4$$

$$E\|\hat{\boldsymbol{\Psi}}_{\tilde{W}_j\tilde{X}}-\boldsymbol{\Psi}_{\tilde{W}_j\tilde{X}}\|_\infty^2\leqslant\frac{1}{n}E^{\frac{1}{2}}\|W\|_{R^q}^4EX^4\quad j=1,2,\cdots,q$$

证明 类似于文献[6]中 Lemma A. 2 的证明。

引理 2 设 $\hat{\Gamma}_j(g)=\sum_{k=1}^{\infty}\frac{\langle\hat{\boldsymbol{\Psi}}_{\tilde{W}_j\tilde{X}},\hat{\xi}_k\rangle}{\hat{\mu}_k}\langle\hat{\xi}_k,g\rangle$, $\Gamma_j(g)=\sum_{k=1}^{\infty}\frac{\langle\boldsymbol{\Psi}_{\tilde{W}_j\tilde{X}},\xi_k\rangle}{\mu_k}\langle\xi_k,g\rangle$, $g\in L^2[0,1]$。

若条件 A 成立,则对于 $b>1$ 和 $d>\frac{b}{2}+1$,有

$$\|\hat{\Gamma}_j-\Gamma_j\|=O_p(n^{-\frac{2d-1}{b+2d}}),j=1,2,\cdots,q$$

证明 类似于文献[6]中 Lemma A. 2 的证明。

引理 3 记

$$\hat{\boldsymbol{D}}=\hat{\boldsymbol{\Psi}}_{\tilde{W}}-\sum_{k=1}^{\infty}\frac{\langle\hat{\boldsymbol{\Psi}}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\hat{\boldsymbol{\Psi}}_{\tilde{X}\tilde{W}},\hat{\xi}_k\rangle}{\hat{\mu}_k}$$

$$\Gamma=[\Gamma_h(\boldsymbol{\Psi}_{\tilde{W}_k\tilde{X}})]_{h,k=1,2,\cdots,q},\hat{\Gamma}=[\hat{\Gamma}_h(\hat{\boldsymbol{\Psi}}_{\tilde{W}_k\tilde{X}})]_{h,k=1,2,\cdots,q}$$

则有 $\hat{D}=D+o_p(1)$。

证明 易知 $D=\boldsymbol{\Psi}_{\tilde{W}}-\Gamma$,则由引理 1—引理 2 可得:

$$\|\boldsymbol{D}-\hat{\boldsymbol{D}}\|_\infty\leqslant\|\hat{\boldsymbol{\Psi}}_{\tilde{W}}-\boldsymbol{\Psi}_{\tilde{W}}\|_\infty+\|\hat{\Gamma}-\Gamma\|_\infty=$$

$$\|\hat{\boldsymbol{\Psi}}_{\tilde{W}}-\boldsymbol{\Psi}_{\tilde{W}}\|_\infty+\max_{1\leqslant h\leqslant q}\sum_{k=1}^{q}\|\hat{\Gamma}_h(\hat{\boldsymbol{\Psi}}_{\tilde{W}_k\tilde{X}})-\Gamma_h(\boldsymbol{\Psi}_{\tilde{W}_k\tilde{X}})\|\leqslant$$

$$O_p(n^{-\frac{1}{2}})+O_p(n^{-\frac{2d-1}{2(b+2d)}})=o_p(1)$$

即得 $\|\boldsymbol{D}-\hat{\boldsymbol{D}}\|=o_p(1)$,得证。

定理 1 证明 由式(5)可得:

$$\sqrt{n}(\hat{\theta}-\theta)=\sqrt{n}(\hat{\boldsymbol{D}}-n\tilde{\Xi})^{-1}\times$$

$$\left(\hat{\Psi}_{\tilde{W}\tilde{Y}}-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\hat{\Psi}_{\tilde{Y}\tilde{X}},\hat{\xi}_k\rangle}{\hat{\mu}_k}-(\hat{D}-n\tilde{\Xi})\theta\right)^3=$$

$$\sqrt{n}(\hat{D}-n\tilde{\Xi})^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{W}_i-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\hat{\Psi}_{\tilde{Y}\tilde{X}},\hat{\xi}_k\rangle}{\hat{\mu}_k}\right]\times\right.$$

$$\left.(\langle\tilde{X}_i,\beta(t)\rangle+\tilde{e}-\upsilon^{\mathrm{T}}\theta)+n\tilde{\Xi}\theta\right\}=\left(\frac{1}{n}\hat{D}-\tilde{\Xi}\right)^{-1}$$

$$\frac{1}{\sqrt{n^3}}\sum_{i=1}^{n}\left[\tilde{W}_i-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\tilde{X}_i,\hat{\xi}_k\rangle}{\hat{\mu}_k}\right]\langle\tilde{X}_i,\beta(t)\rangle+$$

$$\left(\frac{1}{n}\hat{D}-\tilde{\Xi}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{n}\left[\tilde{W}_i-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\tilde{X}_i,\hat{\xi}_k\rangle}{\hat{\mu}_k}\right]\times\right.$$

$$\left.(\tilde{e}_i-\upsilon_i^{\mathrm{T}}\theta)+\tilde{\Xi}\theta\right\}+\left(\frac{1}{n}\hat{D}-\tilde{\Xi}\right)^{-1}\frac{1}{\sqrt{n^3}}\times$$

$$\sum_{i=1}^{n}\left[\sum_{k=1}^{\infty}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\tilde{X}_i,\xi_k\rangle}{\hat{\mu}_k}-\sum_{k=1}^{N}\frac{\langle\hat{\Psi}_{\tilde{W}\tilde{X}},\hat{\xi}_k\rangle\langle\tilde{X}_i,\hat{\xi}_k\rangle}{\hat{\mu}_k}\right]\times$$

$$(\tilde{e}_i-\upsilon_i^{\mathrm{T}}\theta)\triangleq\left(\frac{1}{n}\hat{D}-\tilde{\Xi}\right)^{-1}(T_{11}+T_{12}+T_{13})$$

经过一些推导并利用引理 1 和引理 2 可得:

$$T_{11}=o_p(1)$$
$$T_{13}=o_p(1)$$
$$T_{12}\xrightarrow{L}N(0,\Sigma)$$

其中,$\Sigma=(\sigma^2+\theta^{\mathrm{T}}\tilde{\Xi}\theta)D+\tilde{\Xi}\theta\theta^{\mathrm{T}}\tilde{\Xi}^{\mathrm{T}}$。

再由引理 3 可得:

$$\sqrt{n}(\hat{\theta}-\theta)=\left(\frac{1}{n}D-\tilde{\Xi}+o_p(1)\right)^{-1}(T_{12}+o_p(1))\xrightarrow{L}$$

$$N\left(0,\left(\frac{1}{n}D-\tilde{\Xi}\right)^{-1}\Sigma\left(\frac{1}{n}D-\tilde{\Xi}\right)^{-1}\right)$$

故定理 1 得证。

定理 2 证明　由条件 A,可得:

$$\|\hat{\beta}(t)-\beta(t)\|^2=\left\|\sum_{k=1}^{N}\hat{\rho}_k\hat{\xi}_k-\sum_{k=1}^{\infty}\rho_k\xi_k\right\|^2\leqslant$$

$$\left\|\sum_{k=1}^{N}\hat{\rho}_k\hat{\xi}_k-\sum_{k=1}^{N}\rho_k\xi_k\right\|^2+2\left\|\sum_{k=N+1}^{\infty}\rho_k\xi_k\right\|^2\leqslant$$

$$4\sum_{k=1}^{N}(\hat{\rho}_k-\rho_k)^2+8N\sum_{k=1}^{N}\rho_k^2\|\hat{\xi}_k-\xi_k\|^2+2\sum_{k=N+1}^{\infty}\rho_k^2$$

再由定理 1 及文献[6]类似讨论可得定理 2 的结论。

3　结束语

部分函数型线性回归模型具有广泛应用,其带有不完全观测数据的情形更符合实际的应用,对于既带有响应缺失又带有测量误差的复杂情形,通过对不完全数据的预处理,对模型进行转换,然后将完全观测下统计方法推广到不完全观测下的研究,所获得的估计量的渐近性质和收敛速度和完全观测情形下的结果仍然保持了一致。此结果符合文献[4,13]中对相关问题的讨论。

参考文献(References):

[1] CARDOT H, FERRATY F. Functional Linear Model [J]. Statistics and Probability Letters, 1999, 45(1): 11—22

[2] JAMES G M, SILVERMAN B W. Functional Adaptive Model Estimation [J]. Journal of American Statistics Association, 2004,470(100):565—576

[3] WARING M E, LAPANC K L. Overweight in Children and Adolescents in Relation to Attention Deficit Hyperactivity Disorder: Results from a National Sample [J]. Pediatrics,2008,122(1):1—7

[4] LI Y, HSING T. On Rates of Convergence in Functional Linear Regression[J]. Journal of Multivariate Analysis, 2007,98(9):1782—1840

[5] ZHANG D, LIN X, SOWERS M F. Two-stage Functional Mixed Models for Evaluating the Effect of Longitudinal Covariate Profiles on a Scalar Outcome[J]. Biometrics, 2007,63(2):351—362

[6] SHIN H. Partial Functional Linear Regression [J]. Journal of Statistical Planning and Inference, 2009, 139 (10):3405—3418

[7] SHIN H,LEE M. On Prediction Rate in Partial Functional Linear Regression[J]. Journal of Multivariate Analysis, 2012,103(1):93—106

[8] 王亚飞,杜江,张忠占. 相依误差下部分函数型线性模型的估计[J]. 应用数学学报,2017,40(1):49—65
WANG Y F, DU J, ZHANG Z Z. Partial Functional Linear Models with Dependent Errors [J]. Acta. Mathematica Applicate Sinica,2017,40(1):49—65(in Chinese)

[9] KRAUS D, PANARETOS V M. Dispersion Operators and Resistant Second-order Functional Data Analysis. Biometrika,2012,99(4):813—83

[10] LING X, LIANG L L, VIEU P. Nonparametric Regression Estimation for Functional Stationary Ergodic Data with Missing at Random[J]. Journal of Statistical Planning and Inference, 2015,162(1):75—87

[11] 方连娣,胡凤霞. 核实数据下非线性 EV 模型中经验似然降维推断[J]. 数学杂志,2012,32(1):113—120
FANG L D, HU F X. Empirical Likelihood Dimension Redution Inference in Nonlinear EV Models with Validation Data[J]. Journal of Mathematics, 2012, 32(1):113—120(in Chinese)

[12] FANG L D, WU Y F. Empirical Likelihood Inference of Parameters in Nonlinear EV Models with Censored Data[J]. Journal of Inequalities and Applications, 2015,12(1):379—394

[13] LIANG H. Asymptotic Normality of Parametric Part in Partially Linear Models with Measurement Error in the Nonparametric Part[J]. Journal of Statistical Planning & Inference,2000, 86(1):51—62

Statistical Inference of Partial Function EV Regression Model with Missing Response

FANG Lian-di[1,2]

(1. School of Mathematics and Computer Science, Tongling University,Anhui Tongling 244000, China;
2. School of Electrical and Information Engineering,Jiangsu University,Jiangsu Zhenjiang 212013,China)

Abstract:The purpose of this paper is to study the partial functional linear regression model with a random missing response and measurement errors in explanatory variables. The estimators of unknown parameters and unknown coefficient functions in the model and their asymptotic properties are discussed, respectively. Firstly, the missing data and the data with measurement errors are processed by some data preprocessing method, as well as the model is transformed into a general functional linear regression model. Then the empirical form of the model is given by Karhumen-Loeve expansion (K-L expansion) and principal component analysis. Finally, the least squares estimates of the corresponding unknown variables are obtained by minimizing the objective function with classical multivariate statistical analysis, and the asymptotic normality of the parameter estimators is proved under certain conditions. The convergence rate of the estimator of function and slope function shows that the estimators given are effective estimators, and the statistical inference method of functional data under complete observation is extended to the case of incomplete observation under certain conditions.

Key words:functional data; measurement error; K-L expansion; response missing

责任编辑:罗姗姗

引用本文/Cite this paper:
方连娣. 响应缺失的部分函数型 EV 回归模型的统计推断[J]. 重庆工商大学学报(自然科学版),2020,37(3):42—46
FANG L D. Statistical Inference of Partial Function EV Regression Model with Missing Response[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(3):42—46