系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
29(4) (2009, 4), 490–501

NA 误差下部分线性模型的经验似然推断*

于 卓 熙

(吉林大学数学学院，长春 130012; 长春税务学院应用数学系，长春 130117)

王 德 辉

(吉林大学数学学院，长春 130012)

史 宁 中

(东北师范大学，长春 130021)

摘要 对于部分线性模型 $y_i = \beta x_i + g(t_i) + e_i, 1 \le i \le n$, 这里 (x_i, t_i) 是固定设计点， g 是未知函数， e_i 是负相协 (NA) 随机误差，给出了回归系数的经验似然比统计量，并讨论了似然比统计量的极限分布，可构造参数的经验似然置信区间.

关键词 部分线性模型，负相协随机误差，经验对数似然，置信区间.

MR(2000) 主题分类号 62G10, 62G20, 62M10

1 引 言

考虑下面的部分线性模型

$$y_i = \beta x_i + g(t_i) + e_i, \quad 1 \le i \le n, \tag{1.1}$$

这里 (x_i, t_i) 是非随机设计点， β 是未知参数， $g(\cdot)$ 是定义在闭区间 I 上的未知函数.

模型 (1.1) 中误差项 $\{e_i\}$ 是独立同分布 (i.i.d) 的情形已经被许多学者进行过深入的研究. 这个模型首先由 Engle 等 [1] 提出，又由 Heckman[2], Chen[3], Hamilton 和 Truong[4] 及 Mammen 和 Van de Geer[5] 等学者讨论过. Baek 和 Liang[6] 研究了上述模型在负相协 (NA) 误差下参数 β 的最小二乘估计的相合性和渐近正态性.

在本文中，我们将运用经验似然方法，讨论模型 (1.1) 中误差项 $\{e_i, 1 \le i \le n\}$ 是平稳的 NA 序列的条件下参数的统计推断问题.

称随机变量序列 $\{X_i, 1 \le i \le n\}$ 是 NA 的，如果对 $\{1, 2, \cdots, n\}$ 的任何两个不相交的非空子集 A 和 B 都有

$$\mathrm{Cov}(f_1(X_i, i \in A), f_2(X_j, j \in B)) \le 0,$$

* 国家自然科学基金 (10571073, J10630104), 高等学校博士学科点专项基金 (20070183023), 吉林大学 “985” 工程和长春税务学院科学研究 (2008016) 项目资助.

收稿日期：2007-07-23, 收到修改稿日期：2008-10-27.

其中 f_1 和 f_2 是任何两个使得协方差存在且对每个变元皆非降 (或同时对每个变元皆非增) 的函数. 称 $\{X_i, i \in N\}$ 是 NA 的, 如果对任何自然数 $n \geq 2$, $\{X_i, 1 \leq i \leq n\}$ 是 NA 的.

NA 序列由于其在多元统计分析、渗透性理论、可靠性理论等中的广泛应用而越来越受到人们的关注. NA 序列的定义首先由 Alam 和 Saxena[7] 提出. Joag-Dev 和 Proschan[8], Shao 和 Su[9] 及 Zhang[10] 等都相继进行了研究.

另一方面, 经验似然方法是一类非常重要的构造非参数置信区间和检验的方法, Owen[11] 对此方法的一般性质进行了系统的研究. 许多研究成果表明, 经验似然有类似于参数似然法的优良性, 特别是它的对数形式类似于 Wilk's 理论是趋于卡方分布的. NA 序列的经验似然推断由 Zhang[12] 提出. 目前这种方法还未被应用于 NA 误差下的模型 (1.1).

本文第 2 节构造了 NA 误差下模型 (1.1) 的回归系数的对数经验似然比统计量, 并讨论了统计量的渐近性质. 第 3 节给出数值模拟结果. 定理的证明将在第 4 节给出.

2　方法与主要结果

对模型 (1.1), 如果 β 是已知参数 β_0, 则由 $Ee_i = 0$, 有

$$g(t_i) = E(y_i - x_i\beta_0), \quad 1 \leq i \leq n.$$

所以在给定参数 β 下 $g(\cdot)$ 的一个很自然的估计量是

$$\widehat{g}_n(t, \beta) = \sum_{i=1}^{n} W_{ni}(t)(y_i - x_i\beta),$$

这里 $W_{ni}(\cdot)(1 \leq i \leq n)$ 是一些定义在 I 上的权函数.

为估计 β, 应使下式最小

$$SS(\beta) = \sum_{i=1}^{n}(y_i - x_i\beta - \widehat{g}_n(t_i, \beta))^2. \tag{2.1}$$

(2.1) 式两边对 β 求导可得

$$\sum_{i=1}^{n}\Big[y_i - x_i\beta - \sum_{j=1}^{n} W_{nj}(t_i)(y_j - x_j\beta)\Big]\Big[-x_i + \sum_{j=1}^{n} W_{nj}(t_i)x_j\Big],$$

令 $\beta = \beta_0, m_i = \big[y_i - x_i\beta_0 - \sum\limits_{j=1}^{n} W_{nj}(t_i)(y_j - x_j\beta_0)\big]\big[x_i - \sum\limits_{j=1}^{n} W_{nj}(t_i)x_j\big]$, 我们可以证明当 $n \longrightarrow +\infty$ 时, $Em_i \longrightarrow 0$, 则有下面的表示式

$$-2\mathcal{LR}_n(\beta_0) = 2\sum_{i=1}^{n}\log(1 + \lambda m_i), \tag{2.2}$$

其中 λ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{1 + \lambda m_i} = 0. \tag{2.3}$$

为得到 $-2\mathcal{LR}_n(\beta_0)$ 的渐近性质，我们首先做一些假设.

假设 $\{y_i,x_i,t_i\in I,\ 1\le i\le n\}$ 满足模型 (1.1), $\{e_i,i\ge 1\}$ 是均值为零的平稳 NA 随机误差列， $W_{ni}(\cdot)(1\le i\le n)$ 是定义在 I 上的一些权函数. 令 $\widetilde{x}_i=x_i-\sum\limits_{j=1}^n W_{nj}(t_i)x_j$, $\widetilde{y}_i=y_i-\sum\limits_{j=1}^n W_{nj}(t_i)y_j$.

(C1) i) $\lim\limits_{n\to+\infty}\frac{1}{n}\sum\limits_{i=1}^n\widetilde{x}_i^2=\mathit{\Gamma}_0(0<\mathit{\Gamma}_0<\infty)$, $\max\limits_{1\le i\le n}(\widetilde{x}_i^2)=o(n^{\frac{1}{2}})$.

ii) $\limsup\limits_{n\to\infty}(\frac{1}{\sqrt{n}\log n})\max\limits_{1\le k\le n}\Big|\sum\limits_{i=1}^k\widetilde{x}_{j_i}\Big|<\infty$, 这里 $(j_1,j_2,\cdots,j_n)$ 是 $(1,2,\cdots,n)$ 的任意置换.

(C2) $g(\cdot)$ 定义在 I 上并满足一阶 Lipschitz 条件.

(C3) $\max\limits_{1\le i\le n}\sum\limits_{j=1}^n|W_{nj}(t_i)|=O(1)$, $\max\limits_{1\le i,j\le n}|W_{nj}(t_i)|=O(n^{-\frac{1}{2}}(\log n)^{-1})$.

(C4) i) $\max\limits_{1\le i\le n}\Big|\sum\limits_{j=1}^n W_{nj}(t_i)-1\Big|=O(n^{-\frac{1}{4}})$.

ii) 对某些 $a>0$, $\max\limits_{1\le i\le n}\sum\limits_{j=1}^n|W_{nj}(t_i)|I_{(|t_i-t_j|>an^{-\frac{1}{4}})}=O(n^{-\frac{1}{4}})$.

(C5) 存在 $p>4, E|e_i|^p<+\infty$. 当 $u\to+\infty$ 时， $\sum\limits_{j:|k-j|\ge u}|\mathrm{cov}(e_k,e_j)|\to 0$, 对 $k\ge 1$ 一致成立. $\{e_i\}$ 的谱密度 $f(\omega)$ 满足当 $\omega\in[-\pi,\pi]$ 时， $0<\delta\le f(\omega)<+\infty$.

注 2.1 在一般的条件下，下面的两个权函数就满足假设 (C3)–(C4).

$$W_{ni}^{(1)}(t)=\frac{1}{h_n}\int_{s_{i-1}}^{s_i}K\Big(\frac{t-s}{h_n}\Big)\mathrm{d}s,$$

$$W_{ni}^{(2)}(t)=K\Big(\frac{t-t_i}{h_n}\Big)\Big[\sum_{j=1}^n K\Big(\frac{t-t_j}{h_n}\Big)\Big]^{-1}.$$

这里 $s_i=\frac{t_i+t_{i-1}}{2}$, $i=1,2,\cdots,n-1, s_0=0, s_n=1, K(\cdot)$ 是 Parzen-Rosenblatt 核函数， h_n 是窗宽.

$-2\mathcal{LR}_n(\beta_0)$ 的渐近性质由下面的定理 2.1 给出.

定理 2.1 假设 (C1)–(C5) 成立，则有当 $n\to+\infty$ 时

$$\frac{-2n\mathit{\Gamma}_0\sigma^2\mathcal{LR}_n(\beta_0)}{\sigma_n^2}\longrightarrow_{\mathcal{D}}\chi^2_{(1)},$$

其中 “ $\longrightarrow_{\mathcal{D}}$ ” 是指依分布收敛， $\sigma_n^2=\mathrm{Var}\Big(\sum\limits_{i=1}^n\widetilde{x}_ie_i\Big)$, $Ee_i^2=\sigma^2$.

由定理 2.1 可知， β_0 的一个水平为 $1-\alpha$ 的渐近置信区间为

$$I_\alpha=\Big\{\beta;\frac{-2n\mathit{\Gamma}_0\sigma^2\mathcal{LR}_n(\beta)}{\sigma_n^2}\le u_\alpha\Big\},$$

这里 u_α 是 $\chi^2_{(1)}$ 分布的 $1-\alpha$ 分位数，即 $\Pr(\chi^2_{(1)}\le u_\alpha)=1-\alpha$.

3　模拟计算

这一节我们给出一个数值模拟结果．模拟中的设计点 $x_i, i = 1, 2, \cdots, n$ 取为 $x_i = \Phi^{-1}(\frac{i}{n+1})$, 这里 $\Phi(\cdot)$ 是标准正态分布的分布函数．设计点 t_i，$i = 1, 2, \cdots, n$ 从固定种子 10 的均匀分布 $U[0,1]$ 中产生，以使它们在每次模拟中取值不变．函数 $g(\cdot)$ 取为 $g(t) = t^2$. 核函数 $K(t)$ 是 Gaussian 核且权函数 $W_{ni}(t)$ 取为

$$W_{ni}(t) = K\Big(\frac{t-t_i}{h_n}\Big)\Big[\sum_{j=1}^{n} K(\frac{t-t_j}{h_n})\Big]^{-1}.$$

窗宽 h_n 取为 $n^{-\frac{1}{2}}(\log n)^{-1}$. 误差分布取为标准正态分布，协方差满足

$$Ee_ie_j(i \neq j) = \begin{cases} -0.25, & \text{如果 } |i-j| = 1, \\ -0.15, & \text{如果 } |i-j| = 2, \\ -0.1, & \text{如果 } |i-j| = 3, \\ 0, & \text{否则}. \end{cases}$$

则 $\{e_i\}$ 是 NA 随机变量列．样本容量分别取为 20, 50, 100, 200, 500 和 1000. 置信水平分别为 $1-\alpha = 0.90$ 和 $1-\alpha = 0.95$. 对每组数据重复实现 1000 次，分别计算了经验似然方法的覆盖率，表 1 给出了相关结果．

由表 1, 我们可以看出经验似然方法的优良性．β 取不同值时，覆盖率没有太大的变化．更为有趣的是覆盖率随着样本容量 n 的增大而增大．一般来讲，不是所有情形下都是这样，原因是设计点 (x_i, t_i) 对不同的样本容量 n 取值是不同的．

表 1　β 的覆盖率

		$\beta=-3.5$	$\beta=-1.5$	$\beta=-0.5$	$\beta=0$	$\beta=0.5$	$\beta=1.5$	$\beta=3.5$
$n=20$	$\alpha=0.10$	0.811	0.786	0.787	0.806	0.804	0.803	0.791
	$\alpha=0.05$	0.877	0.875	0.876	0.886	0.879	0.868	0.872
$n=50$	$\alpha=0.10$	0.818	0.796	0.793	0.809	0.806	0.810	0.800
	$\alpha=0.05$	0.879	0.882	0.879	0.888	0.880	0.880	0.879
$n=100$	$\alpha=0.10$	0.819	0.817	0.819	0.828	0.816	0.814	0.812
	$\alpha=0.05$	0.885	0.889	0.892	0.892	0.882	0.892	0.886
$n=200$	$\alpha=0.10$	0.824	0.825	0.837	0.830	0.826	0.817	0.823
	$\alpha=0.05$	0.893	0.891	0.896	0.898	0.898	0.893	0.892
$n=500$	$\alpha=0.10$	0.834	0.835	0.840	0.834	0.829	0.835	0.831
	$\alpha=0.05$	0.903	0.896	0.907	0.901	0.907	0.905	0.897
$n=1000$	$\alpha=0.10$	0.845	0.844	0.846	0.847	0.849	0.849	0.834
	$\alpha=0.05$	0.906	0.900	0.910	0.911	0.909	0.917	0.903

4 定理的证明

在本节，$\{X_i, i \geq 1\}$ 总表示 NA 随机变量序列且 $EX_i = 0$. 为证明定理 2.1, 我们首先给出一系列引理.

引理 4.1 令 $X_1, X_2, \cdots, X_N$ 是平稳的 NA 序列且满足 $E|X_1|^{2+\delta} < \infty$, $\delta > 0$, 则

$$\max_{1 \leq k \leq N} |X_k| = o(N^{\frac{1}{2+\delta}}) \quad \text{a.s.} \tag{4.1}$$

证 见文献 [12].

引理 4.2 假设 $\{a_{ni}, 1 \leq i \leq n, n \geq 1\}$ 是实三角阵满足 $\max\limits_{1 \leq i \leq n} |a_{ni}| = O(n^{-\frac{1}{2}})$ 及 $\sum\limits_{i=1}^{n} a_{ni}^2 = o(n^{-\frac{1}{2}}(\log n)^{-1})$. 如果存在 $p > 4$ 使得 $\sup\limits_i E|X_i|^p < +\infty$, 则有

$$\sum_{i=1}^{n} a_{ni} X_i = o(n^{-\frac{1}{4}}) \quad \text{a.s.} \tag{4.2}$$

证 见文献 [6].

引理 4.3 假设 $\{a_{ni}(\cdot), 1 \leq i \leq n, n \geq 1\}$ 是定义在闭区间 I 上的函数阵并且满足 $\max\limits_{1 \leq i,j \leq n} |a_{ni}(u_j)| = O(n^{-\frac{1}{2}})$ 及 $\max\limits_{1 \leq j \leq n} \sum\limits_{i=1}^{n} |a_{ni}(u_j)| = O(1)$. 如果存在 $p > 2$ 使得 $\sup\limits_i E|X_i|^p < +\infty$, 则

$$\max_{1 \leq j \leq n} \Big| \sum_{i=1}^{n} a_{ni}(u_j) X_i \Big| = o(L(n)) \quad \text{a.s.}, \tag{4.3}$$

$$\max_{1 \leq j \leq n} \Big(\sum_{i=1}^{n} |a_{ni}(u_j) X_i| \Big) = O(1) \qquad \text{a.s.}, \tag{4.4}$$

这里 $L(x) > 0$ 是一个函数，满足 $x \to \infty$ 时，$\sqrt{x} L(x)$ 对于 $x \geq x_0 > 0$ 是非降的.

证 见文献 [6].

引理 4.4 令 $\{a_{ni}, 1 \leq i \leq n, n \geq 1\}$ 是一个实三角阵满足当 $n \to +\infty$ 时，$\sum\limits_{i=1}^{n} a_{ni}^2 = O(1)$ 及 $\max\limits_{1 \leq i \leq n} |a_{ni}| \to 0$. 假设当 $u \to +\infty$ 时，$\sum\limits_{j:|k-j| \geq u} |\mathrm{cov}(X_k, X_j)| \to 0$ 对 $k \geq 1$ 一致成立. 如果 $\mathrm{Var}\big(\sum\limits_{i=1}^{n} a_{ni} X_i\big) \to 1$ 且 X_i 在 L_2 上一致可积，则 $\sum\limits_{i=1}^{n} a_{ni} X_i \longrightarrow_{\mathcal{D}} N(0,1)$.

证 见文献 [13].

引理 4.5 假设定理 2.1 的条件成立，则有

$$\max_{1 \leq i \leq n} |m_i| = o(n^{\frac{1}{2+\delta} + \frac{1}{4}}) \quad \text{a.s.} \tag{4.5}$$

证

$$\begin{aligned}
\max_{1 \leq i \leq n} |m_i| &= \max_{1 \leq i \leq n} \Big| \Big(y_i - x_i\beta_0 - \sum_{j=1}^{n} W_{nj}(t_i)(y_j - x_j\beta_0) \Big) \Big(x_i - \sum_{j=1}^{n} W_{nj}(t_i) x_j \Big) \Big| \\
&= \max_{1 \leq i \leq n} \Big| \widetilde{g}(t_i) + e_i - \sum_{j=1}^{n} W_{nj}(t_i) e_j \Big| |\widetilde{x}_i|
\end{aligned}$$

$$\leq \Big(\max_{1\leq i\leq n}|\widetilde{g}(t_i)|+\max_{1\leq i\leq n}|e_i|+\max_{1\leq i\leq n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\Big)\max_{1\leq i\leq n}|\widetilde{x}_i|,$$

其中 $\widetilde{g}(t_i)=g(t_i)-\sum\limits_{j=1}^{n}W_{nj}(t_i)g(t_j)$. 由引理 4.1 可知 $\max\limits_{1\leq i\leq n}|e_i|=o(n^{\frac{1}{2+\delta}})$ a.s., 这里 $p=2+\delta,\delta>2$.

(C1) i) 表明 $\max\limits_{1\leq i\leq n}|\widetilde{x}_i|=o(n^{\frac{1}{4}})$. 由 (C3) 及引理 4.3, 取 $L(n)=(\log n)^{-1}$, 有

$$\max_{1\leq i\leq n}\left(\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\right)=o((\log n)^{-1})\quad \text{a.s.}$$

由 (C2) 及 (C4) 可得

$$\begin{aligned}
&\max_{1\leq i\leq n}|\widetilde{g}(t_i)|\\
&=\max_{1\leq i\leq n}\Big|g(t_i)-\sum_{j=1}^{n}W_{nj}(t_i)g(t_j)\Big|\\
&\leq \max_{1\leq i\leq n}|g(t_i)|\Big|\sum_{j=1}^{n}W_{nj}(t_i)-1\Big|+\max_{1\leq i\leq n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)\Big||g(t_i)-g(t_j)|I_{(|t_i-t_j|>an^{-\frac{1}{4}})}\\
&\quad+\max_{1\leq i\leq n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)\Big||g(t_i)-g(t_j)|I_{(|t_i-t_j|\leq an^{-\frac{1}{4}})}\\
&=O(n^{-\frac{1}{4}}).
\end{aligned}$$

这表明 $\max\limits_{1\leq i\leq n}|m_i|=o(n^{\frac{1}{2+\delta}+\frac{1}{4}})$ a.s. , 证毕

引理 4.6 假设定理 2.1 的条件成立, 则

$$|\lambda|=o(n^{-\frac{2}{2+\delta}-\frac{3}{4}})\quad \text{a.s.} \tag{4.6}$$

证 记 $\lambda=\rho\psi$, 其中 $\rho\geq 0,|\psi|=1$. 由

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{1+\lambda m_i}=0,$$

可得

$$\begin{aligned}
0&=\frac{1}{n}\psi\sum_{i=1}^{n}\frac{m_i}{1+\rho\psi m_i}=\frac{1}{n}\psi\sum_{i=1}^{n}m_i-\frac{\rho}{n}\sum_{i=1}^{n}\frac{(\psi m_i)^2}{1+\rho\psi m_i}\\
&\leq \frac{1}{n}\psi\sum_{i=1}^{n}m_i-\frac{\rho}{n}\frac{\sum\limits_{i=1}^{n}(\psi m_i)^2}{1+\rho\max\limits_{1\leq i\leq n}|m_i|}\\
&\leq \frac{1}{n}\psi\Big(\sum_{i=1}^{n}m_i\Big)\Big(1+\rho\max_{1\leq i\leq n}|m_i|\Big)-\rho\Big(\frac{1}{n}\sum_{i=1}^{n}m_i^2\Big),
\end{aligned}$$

即

$$\rho\Big(\frac{1}{n}\sum_{i=1}^{n}m_i^2-\psi\Big(\max_{1\le i\le n}|m_i|\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}m_i\Big)\Big)\le\frac{1}{n}\psi\sum_{i=1}^{n}m_i.$$

而

$$\frac{1}{n}\sum_{i=1}^{n}m_i^2\le\frac{1}{n}\sum_{i=1}^{n}\Big(\max_{1\le i\le n}|m_i|\Big)^2=o(n^{\frac{2}{2+\delta}+\frac{1}{2}})\quad \text{a.s.}$$

又

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}m_i&=\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_ie_i-\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)\widetilde{x}_i+\frac{1}{n}\sum_{i=1}^{n}\widetilde{g}(t_i)\widetilde{x}_i\\&=:A_{1n}+A_{2n}+A_{3n}.\end{aligned}$$

由引理 4.2 及条件 (C1) i), 有

$$\max_{1\le i\le n}\frac{|\widetilde{x}_i|}{n}=o(n^{-\frac{3}{4}}),\quad \sum_{i=1}^{n}\frac{|\widetilde{x}_i|^2}{n^2}=O(n^{-1}).$$

所以

$$A_{1n}=o(n^{-\frac{1}{4}})\quad \text{a.s.}$$

显然,

$$A_{2n}=\sum_{j=1}^{n}\Big(\frac{1}{n}\sum_{i=1}^{n}W_{nj}(t_i)\widetilde{x}_i\Big)e_j=:\sum_{j=1}^{n}V_{nj}e_j,$$

由 Abel's 不等式及条件 (C1) ii) 和 (C3), 有

$$\max_{1\le j\le n}|V_{nj}|=\frac{1}{n}\max_{1\le j\le n}\Big|\sum_{i=1}^{n}W_{nj}(t_i)\widetilde{x}_i\Big|\le\frac{C}{n}\max_{1\le i,j\le n}|W_{nj}(t_i)|\cdot\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\widetilde{x}_{j_i}\Big|=O(n^{-1}),$$

其中 $(j_1,j_2,\cdots,j_n)$ 是 $(1,2,\cdots,n)$ 的任意置换.

$$\begin{aligned}\sum_{j=1}^{n}V_{nj}^2&=\sum_{j=1}^{n}\frac{1}{n^2}\Big(\sum_{i=1}^{n}W_{nj}(t_i)\widetilde{x}_i\Big)^2\le\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}W_{nj}^2(t_i)\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\Big)\\&\le\frac{C}{n}\sum_{i=1}^{n}\Big(\max_{1\le j\le n}|W_{nj}(t_i)|\Big)\sum_{j=1}^{n}|W_{nj}(t_i)|=O(n^{-\frac{1}{2}}(\log n)^{-1}),\end{aligned}$$

这里 C 是一个不依赖于 n 的常数，应用引理 4.2, 可得

$$A_{2n}=O(n^{-\frac{1}{4}})\quad \text{a.s.}$$

由 Abel's 不等式及 (C1) ii) 有

$$|A_{3n}|=\frac{1}{n}\Big|\sum_{i=1}^{n}\widetilde{g}(t_i)\widetilde{x}_i\Big|\le\frac{C}{n}\max_{1\le i\le n}|\widetilde{g}(t_i)|\cdot\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\widetilde{x}_{j_i}\Big|=O(n^{-\frac{3}{4}}\log n).$$

所以

$$\frac{1}{n}\sum_{i=1}^{n} m_i = O(n^{-\frac{1}{4}}) \quad \text{a.s.}$$

由以上证明可知

$$\rho = o(n^{-\frac{2}{2+\delta}-\frac{3}{4}}) \quad \text{a.s.} ,$$

这表明

$$|\lambda| = o(n^{-\frac{2}{2+\delta}-\frac{3}{4}}) \quad \text{a.s.}$$

引理 4.7　假设定理 2.1 的条件成立，则有

$$\frac{1}{\sigma_n}\sum_{i=1}^{n} m_i \longrightarrow_{\mathcal{D}} N(0,1), \tag{4.7}$$

其中 $\sigma_n^2 = \mathrm{Var}\Big(\sum\limits_{i=1}^{n} \widetilde{x}_i e_i\Big)$.

证　由 (C5) 及 [14], 可知

$$\begin{aligned}\sigma_n^2 &= \mathrm{Var}\Big(\sum_{i=1}^{n} \widetilde{x}_i e_i\Big) = \int_{-\pi}^{\pi} f(\omega)\Big|\sum_{k=1}^{n} \widetilde{x}_k e^{-ik\omega}\Big|^2 \mathrm{d}\omega \\ &\geq 2\pi\delta \sum_{k=1}^{n} \widetilde{x}_i^2 = O(n).\end{aligned}$$

$$\frac{1}{\sigma_n}\sum_{i=1}^{n} m_i = \frac{1}{\sigma_n}\sum_{i=1}^{n} \widetilde{g}(t_i)\widetilde{x}_i - \frac{1}{\sigma_n}\sum_{i=1}^{n}\Big(\sum_{j=1}^{n} W_{nj}(t_i)e_j\Big)\widetilde{x}_i + \frac{1}{\sigma_n}\sum_{i=1}^{n} \widetilde{x}_i e_i.$$

由 $\max\limits_{1\leq i\leq n} |\widetilde{g}(t_i)| = O(n^{-\frac{1}{4}})$ 及 (C1) ii), 有

$$\frac{1}{\sigma_n}\sum_{i=1}^{n} |\widetilde{g}(t_i)\widetilde{x}_i| \leq \frac{C}{\sqrt{n}}\sum_{i=1}^{n} |\widetilde{g}(t_i)\widetilde{x}_i| \leq \frac{C}{\sqrt{n}}\max_{1\leq i\leq n} |\widetilde{g}(t_i)| \cdot \max_{1\leq k\leq n}\Big|\sum_{i=1}^{k} \widetilde{x}_{j_i}\Big| = O(n^{-\frac{1}{4}}\log n).$$

由 (C1) ii) 并且应用引理 4.3, 取 $L(n) = (\log n)^{-1}$, 可得

$$\begin{aligned}\frac{1}{\sigma_n}\sum_{i=1}^{n}\Big|\Big(\sum_{j=1}^{n} W_{nj}(t_i)e_j\Big)\widetilde{x}_i\Big| &\leq \frac{C}{\sqrt{n}}\sum_{i=1}^{n}\Big|\widetilde{x}_i\sum_{j=1}^{n} W_{nj}(t_i)e_j\Big| \\ &\leq \frac{C}{\sqrt{n}}\Big(\max_{1\leq k\leq n}\Big|\sum_{i=1}^{k} \widetilde{x}_{ji}\Big|\Big)\Big(\max_{1\leq i\leq n}\Big|\sum_{j=1}^{n} W_{nj}(t_i)e_j\Big|\Big) \\ &= \frac{1}{\sqrt{n}} O(n^{\frac{1}{2}}\log n) o((\log n)^{-1}) \\ &= o(1) \quad \text{a.s.}\end{aligned}$$

应用引理 4.4, 由 (C1) 及 (C5) 知 $\max\limits_{1\leq i\leq n} \frac{|\widetilde{x}_i|}{\sigma_n} \to 0$ 且 $\sum\limits_{i=1}^{n} \frac{\widetilde{x}_i^2}{\sigma_n^2} = O(1)$, 可以得到

$$\frac{1}{\sigma_n}\sum_{i=1}^{n} \widetilde{x}_i e_i \longrightarrow_{\mathcal{D}} N(0,1),$$

从而引理 4.7 的结论成立.

引理 4.8 假设定理 2.1 的条件成立, 则有

$$\frac{1}{n}\sum_{i=1}^{n} m_i^2 \longrightarrow \Gamma_0\sigma^2 \quad \text{a.s.} \tag{4.8}$$

证

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} m_i^2 = \frac{1}{n}\Big[&\sum_{i=1}^{n}\widetilde{g}^2(t_i)\widetilde{x}_i^2 + 2\sum_{i=1}^{n}\widetilde{g}(t_i)\widetilde{x}_i^2 e_i - 2\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)\widetilde{g}(t_i)\widetilde{x}_i^2 \\ &-2\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)\widetilde{x}_i^2 e_i + \sum_{i=1}^{n}\widetilde{x}_i^2 e_i^2 + \sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)^2\widetilde{x}_i^2\Big],\end{aligned}$$

由前面证明及 (C1) ii), 可得

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{g}^2(t_i)\widetilde{x}_i^2 \le \Big(\max_{1\le i\le n}|\widetilde{g}(t_i)|\Big)^2\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\Big) = O(n^{-\frac{1}{2}}).$$

而

$$\begin{aligned}\frac{1}{n}\Big|\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)\widetilde{g}(t_i)\widetilde{x}_i^2\Big| &\le \Big(\max_{1\le i\le n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\Big)\Big(\max_{1\le i\le n}|\widetilde{g}(t_i)|\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\Big) \\ &= o(n^{-\frac{1}{4}}(\log n)^{-1}) \qquad \text{a.s.}\end{aligned}$$

又

$$\begin{aligned}\frac{1}{n}\Big|\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)^2\widetilde{x}_i^2\Big| &\le \Big(\max_{1\le i\le n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\Big)^2\Big(\frac{C}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\Big) \\ &= o((\log n)^{-2}) \qquad \text{a.s.}\end{aligned}$$

注意到

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{g}(t_i)\widetilde{x}_i^2 e_i = \sum_{i=1}^{n}\widetilde{g}(t_i)\frac{\widetilde{x}_i^2}{n}e_i,$$

且由 (C1) i), 有

$$\max_{1\le i\le n}\Big|\widetilde{g}(t_i)\frac{\widetilde{x}_i^2}{n}\Big| \le \Big(\max_{1\le i\le n}|\widetilde{g}(t_i)|\Big)\Big(\max_{1\le i\le n}\Big|\frac{\widetilde{x}_i^2}{n}\Big|\Big) = o(n^{-\frac{3}{4}}),$$

$$\sum_{i=1}^{n}\widetilde{g}^2(t_i)\frac{\widetilde{x}_i^4}{n^2} \le \Big(\max_{1\le i\le n}|\widetilde{g}(t_i)|\Big)^2\Big(\max_{1\le i\le n}|\frac{\widetilde{x}_i^2}{n}|\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\Big) = o(n^{-1}),$$

所以, 由引理 4.2 可得

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{g}(t_i)\widetilde{x}_i^2 e_i = o(n^{-\frac{1}{4}}) \qquad \text{a.s.}$$

又知

$$\frac{1}{n}\Big|\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big)\widetilde{x}_i^2e_i\Big| \le \max_{1\le i\le n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\cdot\Big(\frac{1}{n}\sum_{i=1}^{n}|\widetilde{x}_i^2e_i|\Big),$$

由 (C1) i) 且应用引理 4.3, 可得 $\frac{1}{n}\sum\limits_{i=1}^{n}|\widetilde{x}_i^2e_i| = O(1)$ a.s., 所以可以得到

$$\max_{1\le i\le n}\Big|\sum_{j=1}^{n}W_{nj}(t_i)e_j\Big|\cdot\Big(\frac{1}{n}\sum_{i=1}^{n}|\widetilde{x}_i^2e_i|\Big) = o((\log n)^{-1}) \quad \text{a.s.}$$

现在我们来证明

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2e_i^2 \longrightarrow \varGamma_0\sigma^2 \quad \text{a.s.}$$

显然

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2e_i^2 &= \frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2[(e_i^+)^2 - E(e_i^+)^2 + (e_i^-)^2 - E(e_i^-)^2 + Ee_i^2]\\ &= \frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2\{[(e_i^+)^2 - E(e_i^+)^2] + [(e_i^-)^2 - E(e_i^-)^2]\} + \frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2Ee_i^2.\end{aligned}$$

这里 $a^+ = \max(0,a), a^- = \max(0,-a)$. 定义 $[(e_i^+)^2 - E(e_i^+)^2] + [(e_i^-)^2 - E(e_i^-)^2] =: \xi_i$, 则 $\{\xi_i, i\ge 1\}$ 仍然是均值为 0 的 NA 随机变量序列并且满足 $\sup\limits_i E|\xi_i|^{\frac{p}{2}} \le C\sup\limits_i E|e_i|^p < +\infty$, 由引理 4.3 及条件 C(1) i) , 有

$$\frac{1}{n}\Big|\sum_{i=1}^{n}\widetilde{x}_i^2\{[(e_i^+)^2 - E(e_i^+)^2] + [(e_i^-)^2 - E(e_i^-)^2]\}\Big| = o((\log n)^{-1}) \qquad \text{a.s.}$$

所以可以得到

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2Ee_i^2 = \sigma^2\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2 \longrightarrow \varGamma_0\sigma^2,$$

从而引理 4.8 证毕.

定理 2.1 的证明

$$|Em_i| = |\widetilde{g}(t_i)||\widetilde{x}_i| \le \Big(\max_{1\le i\le n}|\widetilde{g}(t_i)|\Big)\Big(\max_{1\le i\le n}|\widetilde{x}_i|\Big) = o(1).$$

所以有

$$Pr(m_{(1)} < 0 < m_{(n)}) \longrightarrow 1,$$

其中 $m_{(i)}$ 代表 m_i 中由小到大的第 i 个次序值. 显然

$$-2\mathcal{LR}_n(\beta_0) = 2\sum_{i=1}^{n}\log(1+\lambda m_i) = 2\sum_{i=1}^{n}\Big(\lambda m_i - \frac{1}{2}\lambda^2m_i^2\Big) + R_{n1},$$

这里 $R_{n1} \le \sum\limits_{i=1}^{n} |\lambda m_i|^3 \le |\lambda|^3 n\big(\max\limits_{1\le i\le n}|m_i|\big)^3 = o(n^{-\frac{3}{2+\delta}-\frac{1}{2}})$ a.s.

令

$$R_{n2} = \frac{1}{n}\sum_{i=1}^{n}\lambda^2 m_i^2 \frac{m_i}{1+\lambda m_i},$$

则有

$$|R_{n2}| \le \frac{1}{n}\sum_{i=1}^{n}|\lambda|\Big(\max_{1\le i\le n}|m_i|\Big)^2 = o(n^{-\frac{1}{4}}) \qquad \text{a.s.}$$

由于

$$\frac{1}{n}\sum_{i=1}^{n} m_i - \frac{1}{n}\sum_{i=1}^{n}\lambda m_i^2 + R_{n2} = \frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{1+\lambda m_i} = 0,$$

可知

$$\lambda = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} m_i}{\frac{1}{n}\sum\limits_{i=1}^{n} m_i^2} + o(1) \ \ \text{a.s.}$$

从而

$$\begin{aligned}
-2\mathcal{LR}_n(\beta_0) &= 2\frac{\frac{1}{n}\sum\limits_{i=1}^{n} m_i}{\frac{1}{n}\sum\limits_{i=1}^{n} m_i^2}\sum_{i=1}^{n} m_i - \left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n} m_i}{\frac{1}{n}\sum\limits_{i=1}^{n} m_i^2}\right)^2\sum_{i=1}^{n} m_i^2 + o(1)\\
&= \frac{2\frac{1}{n}\Big(\sum\limits_{i=1}^{n} m_i\Big)^2}{\frac{1}{n}\sum\limits_{i=1}^{n} m_i^2} - \frac{\frac{1}{n^2}\Big(\sum\limits_{i=1}^{n} m_i\Big)^2}{\frac{1}{n^2}\sum\limits_{i=1}^{n} m_i^2} + o(1)\\
&= \frac{\frac{\sigma_n^2}{n}\Big(\frac{1}{\sigma_n}\sum\limits_{i=1}^{n} m_i\Big)^2}{\frac{1}{n}\sum\limits_{i=1}^{n} m_i^2} + o(1) \ \ \text{a.s.}
\end{aligned}$$

所以，结合引理 4.7 及引理 4.8, 我们得到 $\frac{-2n\Gamma_0\sigma^2\mathcal{LR}_n(\beta_0)}{\sigma_n^2} \longrightarrow_{\mathcal{D}} \chi^2_{(1)}$, 证明完成.

参 考 文 献

[1] Engle R, Granger C, Rice J, Weiss A. Nonparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, 1986, **81**: 310–320.

[2] Heckman N. Spline smoothing in partly linear models. *J. Roy. Statist. Soc. B*, 1986, **48**: 244–248.

[3] Chen H. Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, 1988, **16**: 136–146.

[4] Hamilton S A, Truong Y K. Local linear estimation in partly linear models. *J. Multivariate Anal.*, 1997, **60**: 1–19.

[5] Mammen E, Van de Geer. Penalized quasi-likelihood estimation in partly linear models. *Ann. Statist.,* 1997, **25**: 1014–1035.

[6] Beak J I, Liang H Y. Asymptotics of estimators in semi-parametric model under NA samples. *J. Statist. Plann. Infer.*, 2006, **136**: 3362–3382.

[7] Alam K, Saxena K M L. Positive dependence in multivariate distributions. *Comm. Statist. Theory Methods*, 1981, **A10**: 1183–1196.

[8] Joag-Dev K, Proschan F. Negative association of random variables with applications. *Ann. Statist.*, 1983, **11**: 286–295.

[9] Shao Q M, Su C. The law of the iterated logarithm for negatively associated random variables. *Stochastic. Process. Appl.*, 1999, **83**: 139–148.

[10] Zhang L X. The weak convergence for functions of negatively associated random variables. *J. Multivariate Anal.*, 2001, **78**: 272–298.

[11] Owen A B. Empirical likelihood confidence regions. *Ann. Statist.*, 1990, **18**: 90–120.

[12] Zhang J J. Empirical likelihood for NA series. *Statist. Probab. Lett.*, 2006, **76**: 153–160.

[13] Liang H Y, Beak J I. Convergence of weighted sums for dependent random variables. *J. Korean Math. Soc.,* 2004, **41**(5): 883–894.

[14] Chen Z G. An extension of Lai and Wei's law of the iterated logarithm with application to time series analysis and regression. *Comm. Statist. Theory Methods*, 1990, **32**: 55–69.

EMPIRICAL LIKEHOOD IN PARTIAL LINEAR MODEL UNDER NA SAMPLES

YU Zhuoxi

(*School of Mathematics, Jilin University, Changchun* 130012; *Department of Mathematics, Changchun Taxation College, Changchun* 130117)

WANG Dehui

(*School of Mathematics, Jilin University, Changchun* 130012)

SHI Ningzhong

(*Northeast Normal University, Changchun* 130021)

Abstract This paper is concerned with the typical partial linear model $y_i = \beta x_i + g(t_i) + e_i, 1 \le i \le n$, where (x_i, t_i) are fixed design points, g is an unknown function, and e_i's are negatively associated (NA) random errors. An empirical log-likelihood ratio for the regression coefficient is proposed, the results show that the statistic is asymptotically chi-squared distributed and that the confidence intervals can be constructed accordingly.

Key words Partial linear model, negatively associated random errors, empirical log-likelihood, confidence intervals.