第47卷第3期 中国科学技术大学学报 Vol.47,No.3
2017年3月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Mar. 2017

文章编号:0253-2778(2017)3-0214-07

纵向数据分析中一种压缩经验似然估计方法的大样本性质证明

徐　刚,张　燕,张伟平

(中国科学技术大学管理学院统计与金融系,安徽合肥 230026)

摘要:在一些纵向研究中,协变量往往依赖于观测时间,此时流行的广义估计方程方法在任意的工作相关矩阵下不再保持估计的无偏性和稳健性,若仅使用独立工作相关矩阵则会造成效率低下,而使用不合适的工作相关矩阵可能会错误地包含有偏估计函数,最终导致估计有偏.为此,Leung等提出一种压缩的经验似然估计方法,将大量的无偏估计函数和其他辅助估计函数综合起来,模拟研究表明所提方法相比传统方法是有效的.但他们对其理论性质并未研究.这里对这种压缩经验似然估计量的大样本性质进行了研究,证明了在合适的条件下,压缩经验似然估计量具有相合性和渐近正态性,并给出了经验似然比的渐近分布.

关键词:纵向数据;经验似然;压缩估计;大样本性质

中图分类号:O212.1　　**文献标识码**:A　　doi:10.3969/j.issn.0253-2778.2017.03.003

2010 Mathematics Subject Classification: Primary 62F12; Secondary 62F05

引用格式:徐刚,张燕,张伟平. 纵向数据分析中一种压缩经验似然估计方法的大样本性质证明[J]. 中国科学技术大学学报,2017,47(3):214-220.

XU Gang, ZHANG Yan, ZHANG Weiping. On the asymptotic properties of the shrinkage empirical likelihood estimators for longitudinal data[J]. Journal of University of Science and Technology of China, 2017,47(3):214-220.

On the asymptotic properties of the shrinkage empirical likelihood estimators for longitudinal data

XU Gang, ZHANG Yan, ZHANG Weiping

(*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei* 230026, *China*)

Abstract: When there exist time-dependent covariates in some longitudinal study, it is well-known that the widely used generalized estimating equations approach would not preserve unbiasedness and robustness in an arbitrary working correlation structure. However, incorrect application of the working correlation structure could result in loss of efficiency and biased estimation. To deal with this problem, Leung et al. proposed a shrinkage empirical likelihood approach which combines the unbiased estimating equations and the extracted additional information from the estimating equations that excluded by the independence assumption. Although their simulations have shown the proposed estimators are efficient, the asymptotic properties of the proposed estimators are unknown. Here it is was shown that the proposed estimators are consistent and asymptotically normally distributed under some regular conditions.

Key words: longitudinal data; empirical likelihood; shrinkage estimation; large sample properties

收稿日期:2015-11-20;**修回日期**:2016-04-17

基金项目:国家自然科学基金(11271347, 11671374),中央高校基本科研业务费专项资金(WK2040170010)资助.

作者简介:徐刚,男,1992年生,硕士研究生.研究方向:大样本理论.E-mail: xugang10@mail.ustc.edu.cn

通讯作者:张伟平,博士/副教授.E-mail: zwp@ustc.edu.cn

0 引言

纵向数据经常出现在生物医学、流行病、社会学和经济研究等领域.在纵向研究中,个体的感兴趣指标会依时间顺序多次反复地被观测.因此,同一个体的重复观测之间本质上是相依的.对这种数据,文献中已提出大量的回归模型方法来充分刻画和利用这种相依性[1-2,13]. 其中,广义估计方程(generalized estimating equations, GEE)[3]是应用最广泛的方法之一.广义估计方程方法由边际回归模型和工作相关矩阵构成,当协变量是时间齐次的时候,广义估计方程方法具有众所周知的稳健性,即均值参数估计的相合性不依赖于工作相关矩阵是否正确指定[16].但是,在一些纵向研究中,一些协变量可能会随时间而变化,例如 Alexander 等[4]研究的母亲压力和幼儿患病率关系的实验(MSCM)中,协变量为母亲的压力,因变量为幼儿的患病率,在持续 4 周的追踪调查研究中,母亲的压力随时间而变化,任意一天幼儿的患病率不仅受当天母亲的压力影响,也受过去母亲的压力的影响.Pepe[5]发现此时广义估计方程方法的稳健性不再保持,原因在于其中一些估计函数在任意指定一个工作相关结构下不再是无偏的.他们指出只有在独立相关结构下相合性才成立.显然,独立性假设会造成极大的效率损失,原因在于仅仅使用了少量(无偏)估计函数[6].

选择合适的估计函数属于矩选择问题,文献中已有大量研究[12,14,15,17,21].对纵向数据,Lai 等[7]将时间相依的协变量分为 3 类,并且据此对估计函数进行分类.他们提出了一个检验过程来检验估计函数属于哪种类别,最后使用广义矩方法将通过检验的估计函数(即可认为是无偏的估计函数)综合起来.而 Leung 等[8]提出一种基于压缩经验似然的参数估计方法.这种方法将所有的估计函数分为两类:第一类是确认为无偏的估计函数,其包括了假设独立相依结构下的所有估计函数;第二类是那些可能能够提升参数估计效率的估计函数.这些估计函数中有些是有信息的无偏估计函数,其值对均值参数的变动很敏感,因此在参数估计中包含这类估计函数有助于识别参数真值.还有一些估计函数是无信息的甚至是有偏的,我们显然希望能够在参数估计过程中排除掉这些估计函数.对第二类估计函数需要小心处理,错误地包含无信息或有偏的估计函数会损害估计的性能[9].Leung 等[8]基于经验似然方法,使用压缩估计对第二类估计函数进行自动选择,再和第一类无偏的估计函数综合起来对参数进行估计. 他们通过大量模拟实验和实际数据分析来表明所提估计方法的有效性,但对该方法下估计量的大样本理论性质并未研究.为此,本文对这种估计量的大样本理论性质进行了研究,证明了估计量的相合性和渐近正态性,并且给出了经验似然比统计量的渐近卡方性质及其证明.

1 压缩经验似然估计

记观测的纵向数据为

$$(y_i, X_i), i=1,\cdots,n.$$

式中,$y_i=(y_{i1},\cdots,y_{it})^\tau$ 表示第 i 个观测对象在 T 个时间点的观测,n 为观测对象的总数目.$X_i=(x_{i1},\cdots,x_{iT})^\tau$ 为第 i 个观测对象所对应的协变量矩阵,其中 $x_{it}=(x_{it}^1,\cdots,x_{it}^p)^\tau$ 为观测时间点 t 时观测对象的 p 维协变量.记给定协变量 x_{it} 时,响应变量 y_{it} 的条件期望为

$$E[y_{it} \mid x_{it}] \equiv \mu_{it}(\beta)=\mu(\beta^\tau x_{it}) \tag{1}$$

式中,$\beta=(\beta_1,\cdots,\beta_p)^\tau$ 为 p 维未知参数.记响应变量 y_i 的方差为 $\mathrm{Var}(y_i)=V_i=\phi A_i^{1/2}R_i(\alpha)A_i^{1/2}$,其中 A_i 为正定的对角矩阵,$R_i(\alpha)$为正定相关系数阵,参数 α 未知,ϕ 为一个正的标量参数.ϕA_i 反映了 y_i 的边际方差,相关系数矩阵 R_i 的结构往往是未知的,此时,可以指定一个“工作相关”矩阵,常用的工作相关结构包括独立相关矩阵、对称相关矩阵和 $AR(1)$相关矩阵等. 记逆矩阵 $V_i^{-1}=(\bar{v}_i^{st})$, $s,t=1,\cdots,T$.则广义估计方程方法通过求解下述估计方程来得到 β 的估计值:

$$\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta_j}V_i^{-1}\{y_i-\mu_i\}=\sum_{i=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}\frac{\partial\mu_{is}}{\partial\beta_j}\bar{v}_i^{st}\{y_{it}-\mu_{it}\}=0,\ j=1,\cdots,p \tag{2}$$

式中,$\mu_i=(\mu_{i1},\cdots,\mu_{it})^\tau$.

当协变量为时间齐次时有 $\mu_{is}=\mu_{it}$, $\forall s,t$,因此,条件期望

$$E\frac{\partial\mu_{is}}{\partial\beta_j}\{y_{it}-\mu_{it}\}=0,\ \forall s,t,i,j,$$

即所有估计函数均为无偏估计函数,因此即使相关系数矩阵被错误确定,均值参数 β 的估计依然具有相合性.而当协变量时间相依时,即对某些 s,t,

$\mu_{is} \neq \mu_{it}$，因此 $E \frac{\partial \mu_{is}}{\partial \beta_j}\{y_{it}-\mu_{it}\} \neq 0$.此时，$\beta$ 的估计不再具有相合性. 如果使用工作独立相关矩阵，则会丢弃大量的估计函数，即丢弃所有的 $\partial \mu_{is}/\partial \beta_j\{y_{it}-\mu_{it}\}, s \neq t$，因此估计的效率会有所损失.而如果使用不合适的工作相关矩阵，会导致错误地包含有偏估计函数，最终又会导致估计量有偏[20].从而，选择合适的估计函数是保证估计量相合性和效率的关键所在.

Leung 等[8]提出将估计函数分为两类，一类是独立相关结构下的所有估计函数，这类估计函数总是包含在最后的估计方程中；另一类为剩余的所有估计函数，这里面包含了有信息的无偏估计函数，也包含有偏的估计函数.通过以它们是否为有信息的无偏估计函数的似然度为权值，使用压缩方法选择估计函数，最后通过经验似然估计方法综合所有无偏的估计函数. 模拟研究表明压缩经验似然方法相比于文献[7]的矩估计方法有着相对小的估计偏差和较小的均方误差.

记 $S(\beta)=\{\partial \mu_{\cdot s}/\partial \beta_j(y_{\cdot t}-\mu_{\cdot t}), j=1,\cdots, p; s,t=1,\cdots,T\}$ 为包含所有估计方程的 pT^2 维向量.$S^M(\beta)$ 表示第一类(独立假设下的)所有的无偏估计函数，$S^A(\beta)$ 表示第二类所有的估计函数.即

$$S^M(\beta)=\left\{\frac{\partial \mu_{\cdot t}}{\partial \beta_j}(y_{\cdot t}-\mu_{\cdot t}), t=1,\cdots,T,\ j=1,\cdots,p\right\} \tag{3}$$

$$S^A(\beta)=\left\{\frac{\partial \mu_{\cdot s}}{\partial \beta_j}(y_{\cdot t}-\mu_{\cdot t}), t,s=1,\cdots,T, t\neq s,\ j=1,\cdots,p\right\} \tag{4}$$

引入 $p(T^2-T)$ 维压缩参数 γ，γ 的每一个元素为 $[0,1]$ 区间内的实数.令 $S^{A,\gamma}(\beta)=\gamma^{\tau} S^A(\beta)$，因此第二类所有的估计函数通过压缩参数结合起来，对 S^A 中的无偏有信息的估计函数以较高的权值，而对那些有偏的或无信息的估计函数以较低的权重.使用经验似然估计方法，将第一类估计函数和压缩过的估计函数结合起来，得到如下经验似然目标函数[19]：

$$L(\beta)=\sup\left\{\begin{array}{l}\prod_{i=1}^{n} q_i: q_i \geqslant 0, \sum_{i=1}^{n} q_i=1, \\ \sum_{i=1}^{n} q_i\{S_i^M(\beta), S_i^{A,\gamma}(\beta)\}=0\end{array}\right\} \tag{5}$$

记 $\hat{\beta}^{\gamma}=\operatorname{argmax} L(\beta)$.在实际应用中，需要选择合适的压缩参数 γ.令 S^A 中的任一个估计方程为 $\{S^A{}_k=S_{\cdot st}(\beta_j): k=(j-1)T^2+(s-1)T+t, k=1,\cdots, p(T^2-T)\}$，引入 $p(T^2-T)$ 维向量 c，其元素 c^k 取值为 1 或 0.取值为 1 时，表示对应的第 k 个估计函数 $S^A{}_k$ 可以通过某种检验方法检验为无偏估计函数(例如可以使用文献[7]的检验方法)；反之 c^k 则取值为 0.从而压缩参数 $\gamma=\{\gamma^k, k=1,\cdots,p(T^2-T)\}$ 可以定义如下：

$$\gamma^k= c^k \exp\{-a_0 \mid \sum_{i=1}^{n} S_{ik}^A(\tilde{\beta}) \mid\},\ k=1,\cdots,p(T^2-T) \tag{6}$$

式中，$\tilde{\beta}$ 为独立相关系数矩阵下参数 β 的广义估计方程(2)的解.a_0 为一个正常数，由观测数据决定.其满足以下条件：$\sqrt{n} a_0 \to 0$ 且 $n a_0 \to \infty$ 当 $n \to \infty$，也就是说，$a_0=O(n^{-\delta})$，$1/2<\delta<1$.在实际中，a_0 可以由交叉验证方法来确定.

2　压缩经验似然估计的大样本性质

为研究压缩经验似然估计量 $\hat{\beta}^{\gamma}$ 的大样本性质，我们重点关注估计函数：

$$S_{ik}^A=S_{ist}(y_{it}, x_{it}, x_{is}; \beta_j)=\frac{\partial \mu_{is}}{\partial \beta_j}(y_{it}-\mu_{it}),\ s\neq t,$$

其中，$k=(j-1)T^2+(s-1)T+t$.

定义

$$\Gamma=\begin{Bmatrix}\Gamma^M & 0 \\ 0 & \sum_{k\in\Lambda}\Gamma_k^A\end{Bmatrix} \tag{7}$$

式中，$\Lambda=\{k: \hat{c}^k=1,\ c_0^k=1,\ k=1,\cdots,p(T^2-T)\}$.

$$\Gamma^M=E[\operatorname{Cov}\{S^M(Y,X;\beta_0)\mid X\}+ E\{S^M(Y,X;\beta_0)\mid X\}E\{S^M(Y,X;\beta_0)\mid X\}^{\tau}] \tag{8}$$

$$\Gamma_k^A=E[\operatorname{Cov}\{S_{\cdot k}^A(y_{\cdot t}, x_{\cdot t}, x_{\cdot s}; \beta_{0j})\mid(x_{\cdot t}, x_{\cdot s})\}+ E^2\{S_{\cdot k}^A(y_{\cdot t}, x_{\cdot t}, x_{\cdot s}; \beta_{0j})\mid(x_{\cdot t}, x_{\cdot s})\}] \tag{9}$$

记 X 的概率密度函数记为 $f(x)$.为符号简单起见，以 $g(Y,X;\beta)$ 表示所有估计函数的一般形式，其包括所有的 $S(Y,X;\beta)$，$S^M(Y,X;\beta)$ 和 $S^A(Y,X;\beta)$.记 $m_g(x)=E[g(Y,X;\beta_0)\mid X=x]$，相应地包含

$$m_S(x)=E[S(Y,X;\beta_0)\mid X=x],$$

$$m_{SM}(x)=E[S^M(Y,X;\beta_0)\mid X=x],$$

$$m_{S_{\cdot k}^A}(x)=E[S_{\cdot k}^A(y_{\cdot t}, x_{\cdot t}, x_{\cdot s}; \beta_{0j})\mid$$

$(x_{\cdot t}, x_{\cdot s}) = x]$.

条件①　函数 $f(x)$，$m_g(x)$关于 x 都有 r 阶有界偏导，其中 $r \geqslant 2$，$2r > p$，且对于某个 $c_0 > 0$，$\inf_X p(x) \geqslant c_0$.

条件②　参数真值 β_0 是 $E[g(Y,X;\beta_0)]=0$ 的唯一解，且所有的估计方程 $g(y,x;\beta_0)$关于 x 存在有界 r 阶偏导，$E\|g(Y,X;\beta_0)\|^4 < \infty$.

条件③　$\partial g(y,x;\beta)/\partial\beta$ 和 $\partial^2 g(y,x;\beta)/\partial\beta\partial\beta^\tau$ 在真值β_0 邻域连续.在 β_0邻域，以某个可积函数 $M(x)$ 为界，函数 $\|\partial g(y,x;\beta)/\partial\beta\|$，$\|\partial^2 g(y,x;\beta)/\partial\beta\partial\beta^\tau\|$ 和 $\|g(y,x;\beta)\|^3$ 一致有界.

条件④　矩阵 Γ 正定，$E[\partial g(y,x;\beta)/\partial\beta]$列满秩.

令 c_0，γ_0 对应分别为 c 和 γ 的真实值.若 $S^A_{\cdot k}$ 真实为无偏的，则 $c_0^k=\gamma_0^k=1$；反之 $c_0^k=\gamma_0^k=0$.记 $\hat{\gamma}$ 为 γ_0 的估计，依照文献[8]的方法，若 c 的第 k 个元素通过文献[7]的检验，此时 $S^A_{\cdot k}$ 可被认定为无偏的估计函数，则 $\hat{c}^k$ 取值为 1；反之取值为 0.最后使用交叉验证方法得到合适的 a_0.令

$$\hat{\gamma}^k = \hat{c}^k \exp\{-a_0 \mid \sum_{i=1}^n S^A_{ik}(\tilde{\beta}) \mid\},$$
$$k=1,\cdots,p(T^2-T),$$

从而，我们得到 γ 的估计值.

类似文献[22]的证明，我们可以得到如下结论：

定理 2.1　在条件①～④满足时，当 $n\to\infty$，有 $n^{1/2}(\hat{\beta}^{\hat{\gamma}}-\beta_0)$ 依分布收敛到一个均值为 0，方差为 $V=\{\Omega^\tau\Gamma^{-1}\Omega\}^{-1}$ 取值于 β_0 的正态分布，$\Omega=\left\{E\left[\frac{\partial S^M}{\partial\beta}\right]^\tau, \gamma_0^\tau E\left[\frac{\partial S^A}{\partial\beta}\right]\right\}^\tau$，$\Gamma$ 定义于式(7).

显然极限方差矩阵 V 中有未知多余参数的干扰，在实际中运用上述定理难于进行统计推断.因此下面我们给出经验似然比 $R(\beta)=2l(\beta)-2l(\hat{\beta}^{\hat{\gamma}})$[18-19]的大样本性质，其中 $l(\beta)=-\ln(L(\beta))$，则 $R(\beta_0)=2l(\beta_0)-2l(\hat{\beta}^{\hat{\gamma}})$.下述定理给出了经验似然比 $R(\beta)$的大样本性质.

定理 2.2　满足条件①～④时，当 $n\to\infty$，经验似然比 $R(\beta_0)$依分布收敛到自由度为 p 的卡方分布.

当对参数向量 β 的子集进行统计推断时，可以使用剖面经验似然比.令 $\beta=(\beta_1^\tau,\beta_2^\tau)^\tau$，其中 β_1 是 $q(<p)$维子向量.为了检验原假设 $H_0:\beta_1=\beta_{01}$，记剖面经验似然比为

$$R(\beta_{01},\hat{\beta}_2^{\hat{\gamma}}(\beta_{01}))=2l(\beta_{01},\hat{\beta}_2^{\hat{\gamma}}(\beta_{01}))-2l(\hat{\beta}^{\hat{\gamma}}),$$

其中 $\hat{\beta}_2^{\hat{\gamma}}(\beta_{01})$是在原假设下的估计.则有如下定理：

定理 2.3　条件①～④满足时，当 $n\to\infty$，$R(\beta_{01},\hat{\beta}_2^{\hat{\gamma}}(\beta_{01}))$依分布收敛到自由度为 q 的卡方分布.

为了证明上述定理，先给出以下引理，简单起见，我们记 C 为与 n 无关的正常数的总称.

引理 2.1　假定条件①～④满足，记 $g_i=g_i(y_i,X_i;\beta)$为 $g(Y,X;\beta)$作用在观测对象 i 上的估计方程，当 $n\to\infty$，我们有

$$n^{-1/2}\sum_{i=1}^n g_i(\beta_0) \xrightarrow{d} N(0,\Gamma_g) \tag{10}$$

且

$$\frac{1}{n}\sum_{i=1}^n g_i(\beta_0)g_i(\beta_0)^\tau = \Gamma_g + o_p(1) \tag{11}$$

式中，

$$\Gamma_g = E[\mathrm{Cov}\{g(Y,X;\beta_0)\mid X\} + E\{g(Y,X;\beta_0)\mid X\}E\{g^\tau(Y,X;\beta_0)\mid X\}].$$

证明　由于

$$n^{-1/2}\sum_{i=1}^n g_i(\beta_0) = n^{-1/2}\sum_{i=1}^n \{g_i(\beta_0)-m_g(X_i)\} + n^{-1/2}\sum_{i=1}^n m_g(X_i),$$

易知 $n^{-1/2}\sum_{i=1}^n (g_i(\beta_0)-m_g(X_i)) \xrightarrow{d} E[\sigma_g^2(X)]$，$n^{-1/2}\sum_{i=1}^n m_g(X_i) \xrightarrow{d} \mathrm{Var}\{m_g(X)\}$，由此式(10)得证.由大数定律，易得式(11)的结论，引理得证.

引理 2.2　假定条件①～④满足，当 $n\to\infty$，我们有

$$\mid \hat{\gamma}^k - r_0^k \mid = o_p(1) \tag{12}$$

式中，$k\in\{k:\hat{c}^k=1\}$.

证明　将$\{k:\hat{c}^k=1,\ k=1,\cdots,p(T^2-T)\}$拆分成两部分：$\Lambda=\{k:\hat{c}^k=1,\ c_0^k=1,\ k=1,\cdots,p(T^2-T)\}$和 $\tilde{\Lambda}=\{k:\hat{c}^k=1,\ c_0^k=0,\ k=1,\cdots,p(T^2-T)\}$.对于 $k\in\Lambda$，运用 Taylor 展开，得到

$$\mid\hat{\gamma}^k-\gamma_0^k\mid \left|\exp\left\{-a_0\left|\sum_{i=1}^n S^A_{ik}(\hat{\beta})\right|\right\}-1\right| = a_0\left|\sum_{i=1}^n S^A_{ik}(\hat{\beta})\right| + o_p(1) \tag{13}$$

另外将引理 2.1 的证明运用于 $S^A_{\cdot k}$，我们有

$$n^{-1/2}\sum_{i=1}^n S^A_{ik}(\beta_0) \xrightarrow{d} N(0,\Gamma_k^A) \tag{14}$$

式中，Γ_k^A 的定义见式(9).注意到 $\|\tilde{\beta}-\beta_0\|=$

$o(n^{-1/2})$，利用切比雪夫不等式，我们得到

$$P\left((n^{1/2}a_0)\left|n^{-1/2}\sum_{i=1}^{n}S_{ik}^{A}(\tilde\beta)\right|\geqslant\varepsilon\right)\leqslant\frac{na_0^2\operatorname{Var}\{n^{-1/2}\sum_{i=1}^{n}S_{ik}^{A}(\tilde\beta)\}}{\varepsilon^2}.$$

注意到 $a_0=O(n^{-\delta})$，$1/2<\delta<1$，由此式(12)得证.

对于 $k\in\tilde\Lambda$，

$$|\hat\gamma^k-\gamma_0^k|=|\hat c^k-c_0^k|\exp\left\{-a_0\left|\sum_{i=1}^{n}S_{ik}^{A}(\tilde\beta)\right|\right\}=\exp\left\{-na_0\left|n^{-1}\sum_{i=1}^{n}S_{ik}^{A}(\tilde\beta)\right|\right\}\tag{15}$$

注意到 $\tilde\beta$ 是 β_0 的一个 $\sqrt{n}$ 阶相合估计，$n^{-1}\left|\sum_{i=1}^{n}S_{ik}^{A}(\tilde\beta)\right|=O_p(1)$，并且 $a_0=O(n^{-\delta})$，$1/2<\delta<1$，$na_0\to\infty$，对于 $k\in\tilde\Lambda$，式(12)因此得证.引理 2.2 证毕.

引理 2.3 在条件①～④下，当 $n\to\infty$，

$$n^{-1/2}\sum_{i=1}^{n}\hat\gamma^{\tau}S_i^{A}(\beta_0)\xrightarrow{d}N\left(0,\sum_{k\in\Lambda}\Gamma_k^{A}\right)\tag{16}$$

且

$$\frac{1}{n}\sum_{i=1}^{n}\{\hat\gamma^{\tau}S_i^{A}(\beta_0)\}^2=\sum_{k\in\Lambda}\Gamma_k^{A}+o_p(1)\tag{17}$$

式中，Γ_k^{A} 定义如式(9).

证明 注意到

$$\begin{aligned}n^{-1/2}\sum_{i=1}^{n}\hat\gamma^{\tau}S_i^{A}(\beta_0)&=n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\Lambda}\hat\gamma^{k}S_{ik}^{A}(\beta_0)+n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\tilde\Lambda}\hat\gamma^{k}S_{ik}^{A}(\beta_0)=n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\Lambda}S_{ik}^{A}(\beta_0)+\\&n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\Lambda}(\hat\gamma^{k}-1)S_{ik}^{A}(\beta_0)+n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\tilde\Lambda}\hat\gamma^{k}S_{ik}^{A}(\beta_0)=I_1+I_2+I_3.\end{aligned}$$

对于 $k\in\Lambda$，有 $\max_k S_{ik}^{A}(\beta_0)=O_p(1)$，再次运用引理 2.2，得到 $I_2=o_p(1)$.已知对于 $k\in\tilde\Lambda$，$\gamma^k=o_p(1)$ 且 $n^{-1/2}\sum_{i=1}^{n}S_{ik}^{A}(\beta_0)=O_p(1)$，我们有 $I_3=o_p(1)$.因此由式(14)得证式(16).

现在我们证明式(17)，

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}[\hat\gamma^{\tau}S_i^{A}(\beta_0)]^2&=n^{-1}\sum_{i=1}^{n}\Big[\sum_{k\in\Lambda}S_{ik}^{A}(\beta_0)\Big]^2+o_p(1)=\\n^{-1}\sum_{i=1}^{n}\Big[\sum_{k\in\Lambda}(S_{ik}^{A})^2(\beta_0)&+2\sum_{\substack{k,l\in\Lambda\\k\neq l}}S_{ik}^{A}(\beta_0)S_{il}^{A}(\beta_0)\Big]+o_p(1)\end{aligned}\tag{18}$$

当 $k,l\in\Lambda$，$c_0^k=c_0^l=1$，也就是说，估计方程 S_{ik}^{A} 和 S_{il}^{A} 是真实无偏的，因此

$$\operatorname{Cov}[S_{ik}^{A}(\beta_0),S_{il}^{A}(\beta_0)]=0.$$

事实上，又因为 $\max_{k\in\Lambda}S_{ik}^{A}(\beta_0)=O_p(1)$，有

$$n^{-1}\sum_{i=1}^{n}\sum_{\substack{k,l\in\Lambda\\k\neq l}}S_{ik}^{A}(\beta_0)S_{il}^{A}(\beta_0)=o_p(1)\tag{19}$$

再次将引理 2.1 中证明应用于 $S_{\cdot k}^{A}$，得到

$$n^{-1}\sum_{i=1}^{n}\{S_{ik}^{A}(\beta_0)\}^2=\Gamma_k^{A}+o_p(1),k\in\Lambda\tag{20}$$

由式(18)～(20)，得到式(17)的结论，引理证毕.

引理 2.4 在条件①～④下，当 $n\to\infty$，

$$n^{-1/2}\sum_{i=1}^{n}\{S_i^{M}(\beta_0),\hat\gamma^{\tau}S_i^{A}(\beta_0)\}\xrightarrow{d}N(0,\Gamma)\tag{21}$$

且

$$\frac{1}{n}\sum_{i=1}^{n}\{S_i^{M}(\beta_0),\hat\gamma^{\tau}S_i^{A}(\beta_0)\}\cdot\{S_i^{M}(\beta_0),\hat\gamma^{\tau}S_i^{A}(\beta_0)\}^{\tau}=\Gamma+o_p(1)\tag{22}$$

式中，Γ 的定义见式(7).

证明 运用引理 2.1，令 $g=S^{M}$，易得

$$n^{-1/2}\sum_{i=1}^{n}S_i^{M}(\beta_0)\xrightarrow{d}N(0,\Gamma^{M})\tag{23}$$

且

$$\frac{1}{n}\sum_{i=1}^{n}S_i^{M}(\beta_0)\{S_i^{M}(\beta_0)\}^{\tau}=\Gamma^{M}+o_p(1)\tag{24}$$

式中，Γ^{M} 定义于式(8).已知

$$\operatorname{Cov}(S_{it}(\beta_0),S_{ik}^{A}(\beta_0))=0$$

对所有的方程 $S_{it}\in S_i^{M}$ 和 $S_{ik}^{A}\in S_i^{A}$ 满足，且

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^{n}\hat\gamma^{\tau}S_i^{A}(\beta_0)S_i^{M}(\beta_0)=\\&n^{-1/2}\sum_{i=1}^{n}\sum_{k\in\Lambda}S_{ik}^{A}(\beta_0)S_i^{M}(\beta_0)+o_p(1)=o_p(1)\end{aligned}\tag{25}$$

由此，引理得证.

给定一个 γ 和

$$L(\beta)=\sup\left\{\begin{gathered}\prod_{i=1}^{n}q_i:q_i\geqslant0,\sum_{i=1}^{n}q_i=1,\\\sum_{i=1}^{n}q_i\{S_i^{M}(\beta),S_{i,\gamma}^{A}(\beta)\}=0\end{gathered}\right\}\tag{26}$$

我们记上式 $L(\beta)$ 的对数经验似然为

$$l^{\gamma}(\beta,\lambda)=n^{-1}\sum_{i=1}^{n}\ln\{1+\lambda^{\tau}(\beta)\{S_i^{M}(\beta),S_i^{A,\gamma}(\beta)\}\}\tag{27}$$

基于 β 对式(27)求导并令导数为 0，因此得到如下经验似然等式：

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda^{\tau}(\beta)\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}}\left\{\frac{\partial\lambda^{\tau}(\beta)\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}}{\partial\beta}\right\}^{\tau}\lambda(\beta)=0 \tag{28}$$

求解上式得到式(26)中经验似然方程 $L(\beta)$的极大估计值($\hat{\beta}^{\gamma}$, $\hat{\lambda}^{\gamma}$).

基于引理 2.4 和文献[10,引理 1]的证明,我们能建立下述关于 $L(\beta)$局部最大值存在性的引理.

引理 2.5　假定条件①～④满足,当 $n\to\infty$,运用交叉验证得到合适的 $\hat{\gamma}$,在此基础上,对数经验似然式(27) 在 $\|\beta-\beta_0\|<Cn^{-1/3}$邻域以概率 1 有解 $\hat{\beta}^{\hat{\gamma}}$,且 $L(\beta)$ 在 $\hat{\beta}^{\hat{\gamma}}$ 得到局部最大值.

由以上条件和引理,我们能够完成以下对定理 2.1～2.3 的证明.首先给出定理 2.1 的证明.

定理 2.1 的证明　基于(β,λ),对式(27)中 $l^{\gamma}(\beta,\lambda)$求导,得到

$$Q_{n1}(\beta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}}{1+\lambda^{\tau}\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}},$$

$$Q_{n2}(\beta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda^{\tau}\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}}\left\{\frac{\partial\{S_i^M(\beta),S_i^{A,\gamma}(\beta)\}}{\partial\beta}\right\}^{\tau}\lambda.$$

由合适的 $\hat{\gamma}$ 和引理 2.5,我们得知 $\hat{\beta}^{\hat{\gamma}}$ 和 $\hat{\lambda}^{\hat{\gamma}}=\lambda(\hat{\beta}^{\hat{\gamma}})$满足

$$Q_{n1}(\hat{\beta}^{\hat{\gamma}},\hat{\lambda}^{\hat{\gamma}})=0,\ Q_{n2}(\hat{\beta}^{\hat{\gamma}},\hat{\lambda}^{\hat{\gamma}})=0 \tag{29}$$

在$(\beta_0,\lambda_0)^{\tau}$ 附近对式(29)进行 Taylor 展开,得到

$$n^{1/2}(\hat{\beta}^{\hat{\gamma}}-\beta_0)=V\Omega^{\tau}\Gamma^{-1}n^{1/2}Q_{n1}(\beta_0,\lambda_0)+o_p(1) \tag{30}$$

式中, $V=\{\Omega^{\tau}\Gamma^{-1}\Omega\}^{-1}$ 值基于 β_0, $\Omega=\left\{E\left[\frac{\partial S^M}{\partial\beta}\right]^{\tau},\gamma_0^{\tau}E\left[\frac{\partial S^A}{\partial\beta}\right]\right\}^{\tau}$,$\Gamma$ 的定义见式(7).运用引理 2.4,我们得知

$$n^{1/2}Q_{n1}(\beta_0,\lambda_0)\xrightarrow{d}N(0,\Gamma) \tag{31}$$

定理 2.1 证毕.

定理 2.2 的证明　将对数经验似然式(27)在(β_0,λ_0)进行展开,得到如下经验似然比的展开:

$$R(\beta_0)=n^{1/2}Q_{n1}^{\tau}(\beta_0,\lambda_0)\Gamma^{-1}\Omega V\Omega^{\tau}\Gamma^{-1}n^{1/2}Q_{n1}(\beta_0,\lambda_0)+o_p(1) \tag{32}$$

因此

$$2l(\hat{\beta}^{\hat{\gamma}})=2\sum_{i=1}^{n}\ln\{1+(\hat{\lambda}^{\hat{\gamma}})^{\tau}\{S_i^M(\hat{\beta}^{\hat{\gamma}}),S_i^{A,\hat{\gamma}}(\hat{\beta}^{\hat{\gamma}})\}\}=-nQ_{n1}^{\tau}(\beta_0,\lambda_0)\Gamma^{-1}[I+\Omega V\Omega^{\tau}\Gamma^{-1}]Q_{n1}(\beta_0,\lambda_0)+o_p(1) \tag{33}$$

运用引理 2.4 和式(32),定理 2.2 得证.

定理 2.3 的证明　已知$\tilde{\beta}_2^{\hat{\gamma}}(\beta_{01})$,且 $\tilde{\lambda}_1^{\hat{\gamma}}=\tilde{\lambda}_1^{\hat{\gamma}}\{\tilde{\beta}_2^{\hat{\gamma}}(\beta_{01})\}$满足

$$Q_{n3}\{\beta_{01},\hat{\beta}_2^{\hat{\gamma}},\tilde{\lambda}_1^{\hat{\gamma}}\}=0,Q_{n4}\{\beta_{01},\hat{\beta}_2^{\hat{\gamma}},\tilde{\lambda}_1^{\hat{\gamma}}\}=0,$$

式中,

$$Q_{n3}(\beta_{01},\beta_2,\lambda_1)=\frac{1}{n}\sum_{i=1}^{n}\frac{\{S_i^M(\beta_{01},\beta_2),S_i^{A,\gamma}(\beta_{01},\beta_2)\}}{1+\lambda_1^{\tau}\{S_i^M(\beta_{01},\beta_2),S_i^{A,\gamma}(\beta_{01},\beta_2)\}},$$

$$Q_{n4}(\beta_{01},\beta_2,\lambda_1)=\frac{1}{n}\sum_{i=1}^{n}\frac{\{\partial\{S_i^M(\beta_{01},\beta_2),S_i^{A,\gamma}(\beta_{01},\beta_2)\}/\partial\beta_2\}^{\tau}\lambda_1}{1+\lambda_1^{\tau}\{(S_i^M(\beta_{01},\beta_2),S_i^{A,\gamma}(\beta_{01},\beta_2)\}}.$$

类似于式(33)的证明,有

$$2l(\beta_{01},\hat{\beta}_2^{\hat{\gamma}}(\beta_{01}))=-nQ_{n1}{}^{\tau}(\beta_0,\lambda_0)\Gamma^{-1}[I+\tilde{\Omega}\tilde{V}\tilde{\Omega}^{\tau}\Gamma^{-1}]Q_{n1}(\beta_0,\lambda_0)+o_p(1),$$

其中, $\tilde{\Omega}=\left\{E\left[\frac{\partial S^M}{\partial\beta_2}\right]^{\tau},\gamma_0{}^{\tau}E\left[\frac{\partial S^A}{\partial\beta_2}\right]\right\}^{\tau}$ 和 $\tilde{V}=\{\tilde{\Omega}^{\tau}\Gamma^{-1}\tilde{\Omega}\}^{-1}$值基于 β_0.因此,由引理 2.4 和文献[11]的结论,我们证得

$$R(\beta_{01},\hat{\beta}_2^{\hat{\gamma}}(\beta_{01}))=2l(\beta_{01},\hat{\beta}_2^{\hat{\gamma}})-2l(\hat{\beta}^{\hat{\gamma}})\to\chi_q^2,$$

定理 2.3 得证.

3　结论

经验似然方法的本质优势在于可以灵活综合所有信息并提供似然比统计量,从而进行假设检验和区间估计.在协变量时间相依情形下,本文证明了 Leung 等[8]提出的压缩经验似然估计的相合性和渐近正态性.为便于在实际应用中进行假设检验和区间估计,本文对经验似然比统计量的大样本性质进行了研究,给出了其渐近理论分布,为统计推断提供了理论依据.将现有方法推广到更一般的变系数模型和部分线性模型[23,24],以适应参数随时间变化或者存在非线性函数效应的场合,是我们将来研究的一个方向.

参考文献(References)

[1] DIGGLE P J, HEAGERTY P J, LIANG K Y, et al. Analysis of Longitudinal Data[M]. 2nd ed. New York: Oxford University Press, 2002.

[2] WU H, ZHANG J. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling

Approaches[M]. New York: Wiley, 2006.

[3] LIANG K Y, ZEGER S L. Longitudinal data analysis using generalized linear models[J]. Biometrika, 1986, 73: 13-22.

[4] ALEXANDER C S, MARKOWITZ R. Maternal employment and use of pediatric clinic services[J]. Medical Care, 1986, 24(2): 134-47.

[5] PEPEM S, ANDERSON G L. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data [J]. Communications in Statistics, Simulations and Computation, 1994, 23: 939-951.

[6] FITZMAURICE G M. A caveat concerning independence estimating equations with multivariate binary data[J]. Biometrics, 1995, 51: 309-317.

[7] LAI T L, SMALL D. Marginal regression analysis of longitudinal data with time-dependent covariates: A generalized method-of-moments approach[J]. Journal of the Royal Statistical Society B, 2007, 69: 79-99.

[8] LEUNGD H, SMALL D S, QIN J, et al. Shrinkage empirical likelihood estimator in longitudinal analysis with time-dependent covariates: Application to modeling the health of Filipino children [J]. Biometrics, 2013, 69: 624-632.

[9] NEWEYW K, SMITH R J. Higher order properties of GMM and generalized empirical likelihood estimators [J]. Econometrica, 2004, 72: 219-255.

[10] QIN J, LAWLESS J. Empirical likelihood and general estimating equations[J]. Ann Statist B, 1994, 22: 300-325.

[11] RAO C R. Linear Statistical Inference and Its Applications[M]. New York: Wiley, 1973.

[12] ANDREWS D W K. Consistent moment selection procedures for generalized method of moments estimation[J]. Econometrica, 1999, 67: 543-564.

[13] FITZMAURICEG M, DAVIDIAN M, VERBEKE G, et al. Longitudinal Data Analysis[M]. Boca Raton, Florida: Chapman & Hall/CRC, 2008.

[14] HANSEN L P. Large sample properties of generalized method of moments estimators [J]. Econometrica, 1982, 50:1029-1054.

[15] LAI T, SMALL D, LIU J. Statistical inference in dynamic panel data models[J]. Journal of Statistical Planning and Inference, 2008, 138: 2763-2776.

[16] LIU X, ZHANG W. A moving average Cholesky factor model in joint mean-covariance modeling for longitudinal data[J]. Science in China A, 2013, 56 (11): 2367-2379.

[17] OKUI R. Instrumental variable estimation in the presence of many moment conditions[J]. Journal of Econometrics, 2011, 165: 70-86.

[18] OWEN A B. Empirical likelihood ratio confidence intervals for a single functional[J]. Biometrika, 1988, 75: 237-249.

[19] OWEN A B. Empirical Likelihood[M]. New York: Chapman and Hall/CRC, 2001.

[20] PAN W, CONNETT J E. Selecting the working correlation structure in generalized estimating equations with application to the lung health study[J]. Statistica Sinica, 2002, 23: 475-490.

[21] WANG L, QU A. Consistent model selection and data driven smooth tests for longitudinal data in the estimating equations approach[J]. Journal of the Royal Statistical Society B, 2009, 71: 177-190.

[22] WANG S, QIAN L, CARROLL R J. Generalized empirical likelihood methods for analyzing longitudinal data[J]. Biometrika, 2010, 97(1): 79-93.

[23] XUE L, ZHU L. Empirical likelihood for a varying coefficient model with longitudinal data[J]. Journal of the American Statistical Association, 2007, 102: 642-654.

[24] YOU J, CHEN G, ZHOU Y. Block empirical likelihood for longitudinal partially linear regression models[J]. Can J Statist, 2006, 34: 79-96.