

# 固定效应部分线性单指标面板模型的惩罚经验似然估计 \*

丁 飞 鹏

(江西师范大学数学与信息科学学院, 南昌, 330022)

**摘要:** 结合二次推断函数法、滤子法和经验似然估计法, 为个体内存在相关性的部分线性单指标固定效应面板模型建立了惩罚经验似然估计法。在一些正则条件下, 推导了模型估计量的大样本性质, 证明了所提出的经验似然比渐近于卡方分布。进一步, 用 Monte Carlo 模拟和真实数据分析评价了估计方法在有限样本下的表现。

**关键词:** 经验似然估计; 相关结构; 部分线性单指标面板模型

**中图分类号:** O212.1

**英文引用格式:** DING F P. Penalized empirical likelihood estimation for partially linear single index panel model with fixed effects [J]. Chinese J Appl Probab Statist, 2019, 35(6): 573–593. (in Chinese)

## §1. 引 言

最大似然估计法 (MLE) 是参数推断中重要的统计方法, 广泛应用于众多模型的估计中, Hamilton<sup>[1]</sup> 对该方法及实施操作进行了详细讨论。众所周知, 要获得有效的 MLE 估计量, 必须事先获知现有数据的概率分布簇, 当现有数据的分布类型未知时, 基于该方法的统计推断将可能失效。为克服该缺点, Owen<sup>[2]</sup> 从非参数角度考虑 MLE 方法, 通过运用经验分布函数代替真实分布函数, 提出了经验似然估计法, 关于该方法细节介绍可参见文献 [3]。

经验似然估计法最大的优点在于毋须事先确定枢轴变量, 完全由数据确定置信区间的形状和方向, 现已被广泛应用于非/半参数模型的估计和检验中。Xue 和 Zhu<sup>[4]</sup> 采用经验似然估计法估计了单指标模型, 并证明了对数似然比统计量的渐近分布。与此同时, Zhu 和 Xue<sup>[5]</sup> 发现, 如果没有偏误纠正, 经验似然比不易渐近于极限分布, 因此, 他们对经验似然估计法进行了改进, 提出了偏误纠正的经验似然估计法, 并应用于部分线性单指标模型的估计中。Zhu 等<sup>[6]</sup> 利用偏误纠正经验似然估计法研究了具有多个连接函数的半参数模型, 并得到了一些合理的性质; Xue 和 Xue<sup>[7]</sup> 采用偏误纠正经验似然估计法研究了具有缺失数据的半参数回归模型; Xue<sup>[8]</sup> 采用偏误纠正和可逆选择概率加权方法研究了单指标模型, 该模型的协变量存在缺失值。另一方面, 为了选择合适的模型, 部分学者在经验似然估计中

\*国家自然科学基金项目 (批准号: 71961011) 和江西省教育厅基金项目 (批准号: GJJ160294) 资助。

E-mail: dingfeipeng2008@163.com.

本文 2017 年 5 月 16 日收到, 2018 年 4 月 19 日收到修改稿。

加入惩罚项, 构建惩罚经验似然估计, Otsu<sup>[9]</sup> 采用惩罚经验似然估计法研究了半参数模型, 并证明了所得估计的大样本性质; Fan 等<sup>[10]</sup> 考虑了高维部分线性变系数模型, 采用惩罚经验似然估计法研究了模型的变量选择, 证明了惩罚经验似然估计的 oracle 性质.

近年来, 关于非/半参数面板数据模型的研究进展迅速, 其中利用经验似然估计法研究该模型的文献日益丰富. Xue 和 Zhu<sup>[11]</sup> 及 Tang 和 Zhao<sup>[12]</sup> 将经验似然估计法应用于半参数纵向数据\*模型的估计中, Yang 和 Li<sup>[13]</sup>、Zhou 和 Lin<sup>[14]</sup>、Yang 等<sup>[15]</sup> 及 Zhao 和 Yang<sup>[16]</sup> 应用经验似然估计法研究了半参数变系数纵向数据模型. 值得注意的是, 他们的研究没有考虑个体固定效应, 当真实数据具有个体固定效应时, 忽略个体效应将导致模型估计不满足一致性. 为此, Zhang 等<sup>[17]</sup> 利用经验似然估计法研究了固定效应半参数面板模型, 避免了估计的不一致性; He 等<sup>[18]</sup> 利用经验似然估计法研究了固定效应半参数变系数面板模型.

注意到, 上述关于面板模型的研究均没有考虑个体内的相关性, 忽略这种相关性将可能得到无效估计, 损害统计推断力. 为此, 在研究部分线性纵向数据模型时, 部分学者结合分块经验似然估计法和局部线性估计法, 将个体内的相关性纳入估计方法中, 以提高估计的有效性<sup>[19–23]</sup>. 本文研究内容与前述研究不同之处主要体现在两方面, 一是在建立部分线性单指标面板模型时, 将固定效应和个体内的相关性同时考虑在内, 让所考虑模型更好地拟合真实数据; 二是将 B 样条函数、SCAD 惩罚函数及二次推断估计法 (QIF) 的思想进行有效结合, 为所述模型建立惩罚经验似然估计法. 相比基于核估计理论建立的方法, 该方法不仅能改善估计量的有效性, 还提高了程序运行速度. 此外, 我们证明了所得估计量的大样本性质, 同时用 Monte Carlo 模拟比较了所述方法和广义估计方程法在有限样本下的表现.

## §2. 模型及估计方法

假设个体数为  $N$ , 时间跨度为  $T$ , 则固定效应部分线性单指标面板模型的数学形式为

$$Y_i = X_i\beta + \eta(Z_i, \theta) + \alpha_i \mathbf{1}_T + e_i, \quad (1)$$

其中,  $Y_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$  为因变量;  $X_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$  为  $T \times q$  维线性部分协变矩阵;  $Z_i = (z_{i1}, z_{i2}, \dots, z_{iT})'$  为  $T \times d$  维非线性部分协变矩阵;  $\beta$  为  $q$  维未知参数向量;  $\theta$  为  $d$  维未知参数向量, 并假设满足识别条件  $\|\theta\| = 1$ ;  $\theta$  的第一个元素  $\theta_1 > 0$ ;  $\eta(Z_i, \theta) = (\eta(z'_{i1}\theta), \eta(z'_{i2}\theta), \dots, \eta(z'_{iT}\theta))'$ ,  $\eta(\cdot)$  为未知的一元连接函数;  $\alpha_i$  为第  $i$  个个体未知固定效应, 满足识别条件  $\sum_{i=1}^N \alpha_i = 0$ ;  $e_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$  为误差项, 满足  $E(e_i | X_i, Z_i) = 0$ ,  $\Omega_i = E(e_i e_i')$  为  $e_i$  的正定协方差矩阵, 对  $i \neq j$  有  $E(e_i e_j' | X_i, X_j, Z_i, Z_j) = 0$ ;  $A'$  表示向量或矩阵  $A$  的转置.

\* 本文视面板数据和纵向数据为同一类数据.

在模型估计方面, 由于个体固定效应的存在, 经验似然法不能直接用于模型 (1) 的估计, 必须消除该个体效应才能建立一致有效估计. 本文通过借鉴文献 [24, 25] 的思想, 将滤子法与经验似然法进行有机结合, 以构建模型的估计方法.

首先注意到  $\|\boldsymbol{\theta}\| = 1$ , 意味着  $\boldsymbol{\theta}$  处于单位圆上, 因此,  $\eta(z_{it}\boldsymbol{\theta})$  关于  $\boldsymbol{\theta}$  的导数可能不存在. 为此, 令  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)', \boldsymbol{\theta}^* = (\theta_2, \theta_3, \dots, \theta_d)'$ , 则  $\boldsymbol{\theta}(\boldsymbol{\theta}^*) = (\sqrt{1 - \|\boldsymbol{\theta}^*\|^2}, (\boldsymbol{\theta}^*)')'$ , 真实值  $\boldsymbol{\theta}^*$  满足  $\|\boldsymbol{\theta}^*\| < 1$ ,  $\|\cdot\|$  表示欧几里得模.  $\boldsymbol{\theta}$  关于  $\boldsymbol{\theta}^*$  的 Jacobian 矩阵为

$$J_{\boldsymbol{\theta}^*} = (-\boldsymbol{\theta}^*/\sqrt{1 - \|\boldsymbol{\theta}^*\|^2}, \mathbf{I}_{d-1})',$$

其中,  $\mathbf{I}_{d-1}$  为  $d-1$  阶单位矩阵.

其次, 令  $m$  和  $K_N$  为正整数,  $p = m + K_N$ ,  $a = \min\{z'_{11}\boldsymbol{\theta}_0, z'_{12}\boldsymbol{\theta}_0, \dots, z'_{1T}\boldsymbol{\theta}_0, z'_{21}\boldsymbol{\theta}_0, z'_{22}\boldsymbol{\theta}_0, \dots, z'_{2T}\boldsymbol{\theta}_0, \dots, z'_{N1}\boldsymbol{\theta}_0, z'_{N2}\boldsymbol{\theta}_0, \dots, z'_{NT}\boldsymbol{\theta}_0\}$ ,  $d = \max\{z'_{11}\boldsymbol{\theta}_0, z'_{12}\boldsymbol{\theta}_0, \dots, z'_{1T}\boldsymbol{\theta}_0, z'_{21}\boldsymbol{\theta}_0, z'_{22}\boldsymbol{\theta}_0, \dots, z'_{2T}\boldsymbol{\theta}_0, \dots, z'_{N1}\boldsymbol{\theta}_0, z'_{N2}\boldsymbol{\theta}_0, \dots, z'_{NT}\boldsymbol{\theta}_0\}$ ,  $a = t_0 < t_1 < \dots < t_{K_N} = d$  是区间  $[a, d]$  上的节点序列, 使用这些节点序列可构建  $p$  个  $m-1$  阶的标准化 B 样条基函数 (构造细节参见文献 [26] 或 [27]), 将这些基函数构成样条函数向量  $\mathbf{B}(t) = (B_1(t), B_2(t), \dots, B_p(t))$ , 则一元连接函数  $\eta(t)$  可近似表示为

$$\eta(t) \approx \mathbf{B}'(t)\boldsymbol{\gamma}, \quad (2)$$

其中  $\boldsymbol{\gamma}$  为  $p$  维样条函数系数向量. 将式 (2) 代入式 (1),

$$Y_i = X_i\boldsymbol{\beta} + \tilde{\eta}(Z_i, \boldsymbol{\theta}, \boldsymbol{\gamma}) + \alpha_i\mathbf{1}_T + \mathbf{e}_i, \quad (3)$$

其中,  $\tilde{\eta}(Z_i, \boldsymbol{\theta}, \boldsymbol{\gamma}) = (\mathbf{B}'(z'_{i1}\boldsymbol{\theta})\boldsymbol{\gamma}, \mathbf{B}'(z'_{i2}\boldsymbol{\theta})\boldsymbol{\gamma}, \dots, \mathbf{B}'(z'_{iT}\boldsymbol{\theta})\boldsymbol{\gamma})'$ .

令  $\boldsymbol{\vartheta}_1 = (\boldsymbol{\beta}', \boldsymbol{\theta}^{*'})', \boldsymbol{\vartheta}_2 = \boldsymbol{\gamma}$ , 将 (3) 表示为

$$Y_i^* = X_i^* + \tilde{\eta}^*(Z_i, \boldsymbol{\theta}, \boldsymbol{\gamma}) + \mathbf{e}_i^*,$$

其中,  $Y_i^* = LY_i$ ,  $X_i^* = LX_i$ ,  $\tilde{\eta}^*(\cdot) = L\tilde{\eta}(\cdot)$ ,  $L = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}$  为  $(T-1) \times T$  阶矩阵, 且  $L\mathbf{1}_T = 0$ . 利用 QIF 的思想, 可得两个辅助随机向量:

$$g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) = \begin{pmatrix} u'_{i,1}A_i^{-1/2}M_1A_i^{-1/2}[Y_i^* - u_i(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_{2,0})] \\ \vdots \\ u'_{i,1}A_i^{-1/2}M_sA_i^{-1/2}[Y_i^* - u_i(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_{2,0})] \end{pmatrix}, \quad i = 1, 2, \dots, N, \quad (4)$$

$$g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) = \begin{pmatrix} u'_{i,2}A_i^{-1/2}M_1A_i^{-1/2}[Y_i^* - u_i(\boldsymbol{\vartheta}_{1,0}, \boldsymbol{\vartheta}_2)] \\ \vdots \\ u'_{i,2}A_i^{-1/2}M_sA_i^{-1/2}[Y_i^* - u_i(\boldsymbol{\vartheta}_{1,0}, \boldsymbol{\vartheta}_2)] \end{pmatrix}, \quad i = 1, 2, \dots, N, \quad (5)$$

其中,  $g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})$  表示给定  $\boldsymbol{\vartheta}_2 = \boldsymbol{\vartheta}_{2,0}$  下关于  $\boldsymbol{\vartheta}_1$  的计分函数,  $g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})$  表示给定  $\boldsymbol{\vartheta}_{1,0}$  下关于  $\boldsymbol{\vartheta}_2$  的计分函数; 均值  $u_i = X_i^* \boldsymbol{\beta} + \tilde{\eta}^*(Z_i, \boldsymbol{\theta}, \boldsymbol{\gamma})$ ,  $\dot{u}_{i,1}$  为  $u_i$  关于参数向量  $\boldsymbol{\vartheta}_1 = (\beta', \theta^{*\prime})'$  的导数,  $\dot{u}_{i,2}$  为  $u_i$  关于参数向量  $\boldsymbol{\vartheta}_2 = \boldsymbol{\gamma}$  的导数;  $M_1, M_2, \dots, M_s$  为已知对称的基矩阵;  $\boldsymbol{\vartheta}_{1,0}, \boldsymbol{\vartheta}_{2,0}$  为真实参数向量值. 由假设条件可知  $E[g_{i,1}(\boldsymbol{\vartheta}_{1,0} | \boldsymbol{\vartheta}_{2,0})] = 0$ ,  $E[g_{i,2}(\boldsymbol{\vartheta}_{2,0} | \boldsymbol{\vartheta}_{1,0})] = 0$ . 因此, 由文献 [2] 可得经验似然比为

$$\ell_1(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) = -2 \max \left\{ \sum_{i=1}^N \ln(Np_{i,1}) \mid p_{i,1} \geq 0, \sum_{i=1}^N p_{i,1} = 1, \sum_{i=1}^N p_{i,1} g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) = 0 \right\}, \quad (6)$$

$$\ell_2(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) = -2 \max \left\{ \sum_{i=1}^N \ln(Np_{i,2}) \mid p_{i,2} \geq 0, \sum_{i=1}^N p_{i,2} = 1, \sum_{i=1}^N p_{i,2} g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) = 0 \right\}. \quad (7)$$

由 Lagrange 乘子法, 经过简单计算可得

$$p_{i,1} = \frac{1}{N[1 + \tau'_1 g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})]}, \quad i = 1, 2, \dots, N, \quad (8)$$

$$p_{i,2} = \frac{1}{N[1 + \tau'_2 g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})]}, \quad i = 1, 2, \dots, N. \quad (9)$$

分别将式 (8) 代入式 (6), 式 (9) 代入式 (7),  $\ell_1(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_2)$ 、 $\ell_2(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_1)$  分别可重写为

$$\ell_1(\tau_1, \boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) = 2 \sum_{i=1}^N \ln[1 + \tau'_1 g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})], \quad (10)$$

$$\ell_2(\tau_2, \boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) = 2 \sum_{i=1}^N \ln[1 + \tau'_2 g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})], \quad (11)$$

其中,  $\tau_1$  及  $\tau_2$  为 Lagrange 乘子向量且分别为参数  $\boldsymbol{\vartheta}_1$  和  $\boldsymbol{\vartheta}_2$  的函数. 关于  $\boldsymbol{\vartheta}_1$  最小化  $\ell_1(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})$ , 则参数  $\boldsymbol{\vartheta}_1$  的经验似然估计为

$$\tilde{\boldsymbol{\vartheta}}_1 = \min_{\boldsymbol{\vartheta}_1} \ell_1(\tau_1, \boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}). \quad (12)$$

同理,  $\boldsymbol{\vartheta}_2$  的经验似然估计为

$$\tilde{\boldsymbol{\vartheta}}_2 = \min_{\boldsymbol{\vartheta}_2} \ell_2(\tau_2, \boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}). \quad (13)$$

通常情况下, 用样条函数近似非参数函数将导致过度参数化, 造成过度拟合问题且估计量具有较大的方差. 为此, 借鉴文献 [28] 的做法, 在式 (10) 和式 (11) 中加入惩罚项, 则参数  $\boldsymbol{\vartheta}_1$  和  $\boldsymbol{\vartheta}_2$  的惩罚经验似然估计为

$$\hat{\boldsymbol{\vartheta}}_{1,N} = \arg \min_{\boldsymbol{\vartheta}_1} \left\{ \ell_1(\tau_1, \boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) + \sum_{j=1}^q p_{\lambda_N}(|\beta_j|) + \sum_{l=1}^{d-1} p_{\lambda_N}(|\theta_l^*|) \right\}, \quad (14)$$

$$\hat{\boldsymbol{\vartheta}}_{2,N} = \arg \min_{\boldsymbol{\vartheta}_2} \{ \ell_2(\tau_2, \boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) + p_{\lambda_N}(\|\boldsymbol{\gamma}\|_{K_{\boldsymbol{\theta}}}) \}, \quad (15)$$

其中,  $p_{\lambda_N}(\cdot)$  为依赖于调整参数  $\lambda_N$  的惩罚函数,

$$K_{\boldsymbol{\theta}} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{B}(z'_{it} \boldsymbol{\theta}) \mathbf{B}'(z'_{it} \boldsymbol{\theta}), \quad \|\boldsymbol{\gamma}\|_{K_{\boldsymbol{\theta}}} = (\boldsymbol{\gamma}' K_{\boldsymbol{\theta}} \boldsymbol{\gamma})^{1/2},$$

$|\cdot|$  为绝对值符号. 当进行数值计算时, 采用 Newton-Raphson 迭代算法得出参数  $\boldsymbol{\vartheta}_1$  和  $\boldsymbol{\vartheta}_2$  的估计值.

评论: 本文中采用了 B 样条函数, 该函数具有两个优点: 一是存在局部支撑, 因而在局部近似有良好表现; 二是只需较少节点数便可提供较好的函数近似, 所以数值计算更为有效和稳定. 此外, 通过用样条函数近似未知连接函数, 把非参数回归转化为线性回归, 不仅提高了计算效率, 而且实现了非参数函数的最优收敛率<sup>[29]</sup>.

### §3. 一些实际中的操作问题

#### 3.1 算法的具体实施

令  $\boldsymbol{\vartheta}_1^{(k)}$  表示  $\boldsymbol{\vartheta}_1$  的第  $k$  次迭代值,  $\boldsymbol{\vartheta}_2^{(k)}$  表示  $\boldsymbol{\vartheta}_2$  的第  $k$  次迭代值, 在  $\boldsymbol{\vartheta}_1^{(k)}$  的邻域内, 惩罚函数可近似表示为

$$p_{\lambda_N}(|t|) \approx p_{\lambda_N}(|t^{(k)}|) + \frac{1}{2} \frac{\dot{p}_{\lambda_N}(|t^{(k)}|)}{|t^{(k)}|} [t^2 - (t^{(k)})^2],$$

其中,  $\dot{p}_{\lambda_N}(\cdot)$  表示  $p_{\lambda_N}(\cdot)$  的一阶导数,  $t$  表示一维参数. 根据式 (14) 和 (15) 可知

$$\begin{aligned} N^{-1} \sum_{i=1}^N \frac{g_{i,1}(\boldsymbol{\vartheta}_1^{(k+1)} | \boldsymbol{\vartheta}_2^{(k)})}{1 + \tau'_1 g_{i,1}(\boldsymbol{\vartheta}_1^{(k+1)} | \boldsymbol{\vartheta}_2^{(k)})} &= 0, \\ N^{-1} \sum_{i=1}^N \frac{[\partial g_{i,1}(\boldsymbol{\vartheta}_1^{(k+1)} | \boldsymbol{\vartheta}_2^{(k)}) / \partial \boldsymbol{\vartheta}_1]' \tau_1}{1 + \tau'_1 g_{i,1}(\boldsymbol{\vartheta}_1^{(k+1)} | \boldsymbol{\vartheta}_2^{(k)})} + \sum_{j=1}^q \frac{\dot{p}_{\lambda_N}(|\beta_j^{(k)}|)}{|\beta_j^{(k)}|} \beta_j^{(k+1)} + \sum_{l=1}^{d-1} \frac{\dot{p}_{\lambda_N}(|\theta_l^{(k)}|)}{|\theta_l^{(k)}|} \theta_l^{(k+1)} &= 0. \end{aligned}$$

同理有

$$\begin{aligned} N^{-1} \sum_{i=1}^N \frac{g_{i,2}(\boldsymbol{\vartheta}_2^{(k+1)} | \boldsymbol{\vartheta}_1^{(k)})}{1 + \tau'_2 g_{i,2}(\boldsymbol{\vartheta}_2^{(k+1)} | \boldsymbol{\vartheta}_1^{(k)})} &= 0, \\ N^{-1} \sum_{i=1}^N \frac{[\partial g_{i,2}(\boldsymbol{\vartheta}_2^{(k+1)} | \boldsymbol{\vartheta}_1^{(k)}) / \partial \boldsymbol{\vartheta}_2]' \tau_2}{1 + \tau'_2 g_{i,2}(\boldsymbol{\vartheta}_2^{(k+1)} | \boldsymbol{\vartheta}_1^{(k)})} + \frac{\dot{p}_{\lambda_N}(\|\boldsymbol{\gamma}^{(k)}\|_{K_{\boldsymbol{\theta}}})}{\|\boldsymbol{\gamma}^{(k)}\|_{K_{\boldsymbol{\theta}}}} K_{\boldsymbol{\theta}} \boldsymbol{\gamma}^{(k+1)} &= 0. \end{aligned}$$

经过简单计算可获得

$$\begin{pmatrix} \tau_1^{(k+1)} \\ \boldsymbol{\vartheta}_1^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{\vartheta}_1^{(k)} \end{pmatrix} - S_{1,N}^{-1} \begin{pmatrix} N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)}) \\ \boldsymbol{\Pi}_1 \boldsymbol{\vartheta}_1^{(k)} \end{pmatrix}, \quad (16)$$

$$\begin{pmatrix} \tau_2^{(k+1)} \\ \boldsymbol{\vartheta}_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{\vartheta}_2^{(k)} \end{pmatrix} - S_{2,N}^{-1} \begin{pmatrix} N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)}) \\ \boldsymbol{\Pi}_2 \boldsymbol{\vartheta}_2^{(k)} \end{pmatrix}, \quad (17)$$

其中,

$$\begin{aligned} S_{1,N} &= \begin{pmatrix} N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)}) g'_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)}) & N^{-1} \sum_{i=1}^N \dot{g}_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)}) \\ N^{-1} \sum_{i=1}^N \dot{g}'_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)}) & \boldsymbol{\Pi}_1 \end{pmatrix}, \\ S_{2,N} &= \begin{pmatrix} N^{-1} \sum_{i=1}^N g_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)}) g'_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)}) & N^{-1} \sum_{i=1}^N \dot{g}_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)}) \\ N^{-1} \sum_{i=1}^N \dot{g}'_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)}) & \boldsymbol{\Pi}_2 \end{pmatrix}, \\ \boldsymbol{\Pi}_1 &= \text{diag} \left\{ \frac{\dot{p}_{\lambda_N}(|\vartheta_{1,1}^{(k)}|)}{|\vartheta_{1,1}^{(k)}|}, \frac{\dot{p}_{\lambda_N}(|\vartheta_{1,2}^{(k)}|)}{|\vartheta_{1,2}^{(k)}|}, \dots, \frac{\dot{p}_{\lambda_N}(|\vartheta_{1,q+d-1}^{(k)}|)}{|\vartheta_{1,q+d-1}^{(k)}|} \right\}, \\ \boldsymbol{\Pi}_2 &= \frac{\dot{p}_{\lambda_N}(\|\boldsymbol{\gamma}^{(k)}\|_{K_{\boldsymbol{\theta}}})}{\|\boldsymbol{\gamma}^{(k)}\|_{K_{\boldsymbol{\theta}}}} K_{\boldsymbol{\theta}}, \end{aligned}$$

$\dot{g}_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)})$  表示  $g_{i,1}(\boldsymbol{\vartheta}_1^{(k)} | \boldsymbol{\vartheta}_2^{(k)})$  对参数  $\boldsymbol{\vartheta}_1$  的一阶导数,  $\dot{g}_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)})$  表示  $g_{i,2}(\boldsymbol{\vartheta}_2^{(k)} | \boldsymbol{\vartheta}_1^{(k)})$  对参数  $\boldsymbol{\vartheta}_2$  的一阶导数,  $\vartheta_{1,l}$  为  $\boldsymbol{\vartheta}_1$  的第  $l$  个元素.

综合上述, 惩罚经验似然估计的算法总结如下:

第一步, 取不同最小二乘估计  $\boldsymbol{\vartheta}_1^0$  和  $\boldsymbol{\vartheta}_2^0$  作为  $\boldsymbol{\vartheta}_1$  和  $\boldsymbol{\vartheta}_2$  的初始值;

第二步, 给定当前值  $\boldsymbol{\vartheta}_1^{(k)}$  和  $\boldsymbol{\vartheta}_2^{(k)}$ , 利用式 (16) 和 (17) 得到  $\boldsymbol{\vartheta}_1$  和  $\boldsymbol{\vartheta}_2$  下一步的迭代值;

第三步, 重复步骤第二步, 直至满足收敛条件.

### 3.2 调整参数及节点数的选择

在对估计方法进行具体实施时, 调整参数的作用举足轻重, 模型的表现依赖于该参数选择是否合适. 进一步地, 从第二节的描述可知, 惩罚经验似然估计方法也需要选择合适的节点.

对于节点的选择, 类似于文献 [28, 30], 采用等间距的方式选择节点, 因此, 我们只需选择合适的节点数. Ruppert 等<sup>[31]</sup> 提供了大量的节点数选择方法: 近似选择法、AIC 准则、BIC 准则及广义交叉验证法 (详细可参见文献 [31]). 本文采用文献 [30] 的 BIC 准则选择节点数, 其统计量为

$$\text{BIC}_1(K) = \ell_2(\tilde{\boldsymbol{\vartheta}}_2 | \tilde{\boldsymbol{\vartheta}}_1) + \ln(N) \text{DF}_N / N, \quad (18)$$

其中,  $K$  为节点数,  $\tilde{\boldsymbol{\vartheta}}_1$  和  $\tilde{\boldsymbol{\vartheta}}_2$  由式 (12) 及 (13) 所得,  $\text{DF}_N$  为参数向量  $\tilde{\boldsymbol{\vartheta}}_2$  的维数. 则最优节点数  $K$  为

$$\hat{K} = \arg \min_K \{\text{BIC}_1(K)\}.$$

记  $Q(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \ell_1(\tau_1, \boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_2) + \ell_2(\tau_2, \boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_1) + \sum_{j=1}^{q+d-1} p_{\lambda_N}(|\vartheta_{1,j}|) + p_{\lambda_N}(\|\boldsymbol{\vartheta}_2\|_{K_\theta}), \vartheta_{1,j}$  为  $\boldsymbol{\vartheta}_1$  的第  $j$  个元素. 类似于节点数的选择, 定义

$$\text{BIC}_2(\lambda_N) = Q(\widehat{\boldsymbol{\vartheta}}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{2,N}) + \ln(N) \text{DF}_{\lambda_N}/N, \quad (19)$$

其中,  $\text{DF}_{\lambda_N} = \text{trace}\{[\ddot{Q}(\widehat{\boldsymbol{\vartheta}}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{2,N}) + n\Lambda(\widehat{\boldsymbol{\vartheta}}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{2,N})]^{-1}\ddot{Q}_N(\widehat{\boldsymbol{\vartheta}}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{2,N})\}$ ,  $\ddot{Q}(\cdot)$  为  $Q(\cdot)$  的二阶导数;  $\Lambda(\cdot) = \text{diag}\{\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2\}$ ,  $\boldsymbol{\Pi}_1$  和  $\boldsymbol{\Pi}_2$  的定义见式 (16) 和 (17),  $\widehat{\boldsymbol{\vartheta}}_{1,N}$  和  $\widehat{\boldsymbol{\vartheta}}_{2,N}$  来自于式 (14) 及 (15), 那么最优调整参数值  $\lambda_N$  可通过最小化  $\text{BIC}_2$  获得, 即

$$\widehat{\lambda}_N = \arg \min_{\lambda_N} \text{BIC}_2(\lambda_N).$$

## §4. 演近性质

为了推导出惩罚经验似然法的演近性质, 还需要下列假设条件, 这里用  $\|V\|$  表示向量或矩阵  $V$  的模, 其值等于向量或矩阵  $V$  的最大奇异值.

假设 A1  $T$  为有界正整数, 连接函数  $\eta$  的  $r$  阶导数有界,  $r \geq 2$ ;

假设 A2 参数空间  $\Theta$  为紧致集, 真实参数向量  $\boldsymbol{\vartheta}_0$  为  $\Theta$  的内点;

假设 A3 参数  $\boldsymbol{\vartheta}$  是可识别的, 即存在唯一的  $\boldsymbol{\vartheta}_0$  满足:  $u_i(\boldsymbol{\vartheta}_0) = X_i^* \boldsymbol{\beta}_0 + \tilde{\eta}^*(Z_i, \boldsymbol{\theta}_0, \gamma_0)$ ,  $i = 1, 2, \dots, N$ .

假设 A4 对于随机误差项  $e_i$ , 存在某个  $\zeta > 0$ , 使得  $\sup_i E\|e_i\|^{2+\zeta} < \infty$ , 随机误差项的协方差  $\Omega_i$  满足  $\sup_i \|\Omega_i\| < \infty$ ;

假设 A5 节点序列  $v = \{v_0 \leq v_1 \leq \dots \leq v_{k_N} \leq v_{k_N+1}\}$  为准一致的, 即存在常数  $c$ , 使得  $\max(v_{j+1} - v_j, j = 0, 1, \dots, k_N)/\min(v_{j+1} - v_j, j = 0, 1, \dots, k_N) \leq c$ ,  $k_N = O_p(N^{1/(2r+1)})$ ,  $h = 1/k_N$ ;

假设 A6 所有方差矩阵  $A_i \geq 0$ , 且  $\sup_i \|A_i\| < \infty$ ,  $E(A_{0,i}^{-1/2} M_k A_{0,i}^{-1/2})$  是有界的,  $k = 1, 2, \dots, s$ ,  $A_{0,i}$  为  $A_i$  在  $u = u_i^0$  处的取值,  $i = 1, 2, \dots, N$ ;

假设 A7 存在某个常数  $c$ , 使得方差函数满足  $0 < c \leq \inf_{i,t} V(u_{it}) \leq \sup_{i,t} V(u_{it}) < \infty$ , 且函数  $V$  存在有界二阶导数;

假设 A8 协变量  $\mathbf{X} = (X'_1, X'_2, \dots, X'_N)'$  和  $\mathbf{Z} = (Z'_1, Z'_2, \dots, Z'_N)'$  均为有界紧致的,  $x_{it}$  和  $z'_{it}\theta$  的密度函数均有界且绝对连续,  $i = 1, 2, \dots, N$ ;  $t = 1, 2, \dots, T$ ;

假设 A9 令  $\mathbf{M} = (M'_1, M'_2, \dots, M'_s)'$ , 假设  $\mathbf{M}$  的奇异值有界;

假设 A10 (i)  $N^{-1} \sum_{i=1}^N \mathbb{E}[\dot{g}_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})]$  在  $\boldsymbol{\vartheta}_{1,0}$  的邻域内一致收敛, 用  $\mathbf{G}_{1,0}$  表示其收敛值;  $N^{-1} \sum_{i=1}^N \mathbb{E}[\dot{g}_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})]$  在  $\boldsymbol{\vartheta}_{2,0}$  的邻域内一致收敛的, 用  $\mathbf{G}_{2,0}$  表示其收敛值; (ii)  $N^{-1} \sum_{i=1}^N \mathbb{E}[g_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})g'_{i,1}(\boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0})]$  在  $\boldsymbol{\vartheta}_{1,0}$  的邻域内一致收敛, 用  $\mathbf{C}_{1,0}$  表示其收敛值;  $N^{-1} \sum_{i=1}^N \mathbb{E}[g_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})g'_{i,2}(\boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0})]$  在  $\boldsymbol{\vartheta}_{2,0}$  的邻域内一致收敛, 用  $\mathbf{C}_{2,0}$  表示其收敛值; 当  $N \rightarrow \infty$  时.

假设 A1 是为了让个体数量  $N$  与总样本数  $NT$  保持同阶, 即  $N = O(NT)$ , 并且保证连接函数是平滑的, 假设 A2 至 A9 为一些规则性条件, 很容易验证, 假设 A10 为专门运用于非线性模型的假设.

**定理 1** 在假设条件 A1–A10 下, 令  $\tilde{\eta}(z'_{it}\boldsymbol{\theta}) = \mathbf{B}'(z'_{it}\boldsymbol{\theta})\boldsymbol{\gamma}$ , 则由式 (13) 得到的非参数估计满足:

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\tilde{\eta}(z'_{it}\tilde{\boldsymbol{\theta}}) - \eta(z'_{it}\boldsymbol{\theta}_0)]^2 = O_p(h^{2r}). \quad (20)$$

**定理 2** 在假设条件 A1–A10 下, 如果平滑参数  $\lambda_N = o(N^{-1/2})$ , 则由式 (14) 得到的惩罚经验似然估计  $\widehat{\boldsymbol{\vartheta}}_{1,N}$  为渐近正态且有效的估计, 即

$$\sqrt{N}(\widehat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}) \xrightarrow{L} N(0, \mathbf{V}), \quad (21)$$

其中, “ $\xrightarrow{L}$ ” 表示依分布收敛,  $\mathbf{V} = \{\mathbf{G}'_{1,0} \mathbf{C}_{1,0}^{-1} \mathbf{G}_{1,0}\}^{-1}$ .

**定理 3** 在假设条件 A1–A10 下, 我们有

$$W(\boldsymbol{\vartheta}_{1,0}) = \ell_1(\boldsymbol{\vartheta}_{1,0} | \boldsymbol{\vartheta}_{2,0}) - \ell_1(\widehat{\tau}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{1,N} | \boldsymbol{\vartheta}_{2,0}) \xrightarrow{L} \chi^2_{(q+d-1)}, \quad (22)$$

其中,  $\ell_1(\boldsymbol{\vartheta}_{1,0} | \boldsymbol{\vartheta}_{2,0}) = 2 \sum_{i=1}^N \ln[1 + \tau'_1(\boldsymbol{\vartheta}_{1,0})g_{i,1}(\boldsymbol{\vartheta}_{1,0})]$ ,  $\chi^2_{(q+d-1)}$  表示自由度为  $(q+d-1)$  的卡方分布.

## §5. 数值分析

### 5.1 Monte Carlo 模拟研究

本段将利用 Monte Carlo 模拟评估所述估计方法在有限样本下的表现. 我们用绝对偏度 (A.Bias)、标准误 (S.E.) 和均方误 (MSE) 等指标评估参数估计的表现, 其计算方法分别为  $A.Bias = |\bar{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0|$ ,  $S.E. = \sqrt{M^{-1} \sum_{j=1}^M (\boldsymbol{\vartheta}_j - \bar{\boldsymbol{\vartheta}})^2}$ , 其中,  $\boldsymbol{\vartheta}_j$  为第  $j$  次模拟的估计值,

$\bar{\vartheta} = \sum_{j=1}^M \vartheta_j / M$ ,  $M$  为模拟次数,  $\text{MSE} = M^{-1} \sum_{j=1}^M (\vartheta_j - \vartheta_0)^2$ ; 用  $\text{AMSE}_\eta$  评估非参数部分的表现, 其计算公式为  $\text{AMSE}_\eta = (M \cdot NT)^{-1} \sum_{j=1}^M \sum_{i=1}^N \sum_{t=1}^T [\hat{\eta}(z'_{it} \hat{\theta}_j) - \eta(z'_{it} \theta_0)]^2$ .

**例 4** (估计方法在各相关结构下的表现) 假设个体数为 30、50 及 100, 时间跨度为 5, 数据产生过程为 Cosine 过程:

$$y_{it} = x_{it}\beta + \cos(z_{it,1}\theta_1 + z_{it,2}\theta_2 + z_{it,3}\theta_3) + \alpha_i + \varepsilon_{it}, \quad (23)$$

其中, 协变量  $x_{it}$  独立取值于  $\lambda = 0.5$  的指数分布;  $z_{it1}$ 、 $z_{it2}$  和  $z_{it3}$  均独立取值于均匀分布  $U(-\pi/2, \pi/2)$ ; 真实参数值为  $\theta_0 = 0.8$ ,  $\theta_{10} = 0.5$ ,  $\theta_{20} = 0.3$ ,  $\theta_{30} = \sqrt{0.66}$ ;  $\alpha_i$  为第  $i$  个体的固定效应, 独立取值于区间  $(-1, 2)$ ;  $\varepsilon_{it}$  服从正态分布, 边际方差为 1.

模拟中, 采用立方 B 样条函数的线性组合近似未知连接函数, 利用 SCAD 惩罚函数对参数估计实施惩罚. 此外, 由于独立结构 (WI)、一阶自相关结构 (AR-1) 和等相关结构 (EC) 为误差项在实际应用中最常见的三种相关结构, 因而, 在模拟中, 主要考虑估计方法在这三种结构下的表现, 模拟结果分别呈现在表 1 至表 3 中.

表 1 估计方法在独立结构下的表现

size	Indicator	$\beta$	$\theta_1$	$\theta_2$	$\theta_3$	$\text{AMSE}_\eta$
$N = 30$	mean	0.7987	0.5939	0.2911	0.6538	0.2645
	S.E.	0.0552	0.1851	0.2187	0.2308	
	A.Bias	0.0013	0.0939	0.0089	0.1586	
	MSE	0.0030	0.0342	0.0478	0.0533	
$N = 50$	mean	0.8036	0.5492	0.3015	0.7538	0.1273
	S.E.	0.0348	0.1161	0.1235	0.1025	
	A.Bias	0.0036	0.0492	0.0015	0.0586	
	MSE	0.0012	0.0135	0.0153	0.0105	
$N = 100$	mean	0.7989	0.5404	0.2904	0.7763	0.0969
	S.E.	0.0286	0.0868	0.0914	0.0713	
	A.Bias	0.0011	0.0404	0.0096	0.0361	
	MSE	0.0008	0.0075	0.0083	0.0051	

表 1 为估计方法在独立结构下参数部分和非参数部分的模拟结果. 首先看参数估计的表现, 参数估计值的绝对偏度随着样本量的增加在 0 值附近波动, 说明参数估计值在真实值周围变化; 均方误随着个体数的增加逐渐减小, 意味着估计值偏离真实值的概率随样本量的增加将不断减少, 表明估计方法在独立结构下的表现满足一致性; 从标准误的数值发现, 随着样本量的增加, 标准误逐渐减小, 表明估计的精度不断提高, 说明在独立结构下参数估计的表现较为稳健. 另一方面,  $\text{AMSE}_\eta$  反应了估计连接函数与真实连接函数的接近程度.

度, 从表中数据可以发现, 随着个体数的增加, 估计的连接函数与真实连接函数逐渐接近.

表 2 估计方法在自相关结构下的表现

size	Indicator	$\beta$	$\theta_1$	$\theta_2$	$\theta_3$	AMSE $_{\eta}$
$N = 30$	mean	0.7994	0.4872	0.2927	0.8042	0.1677
	S.E.	0.0267	0.1140	0.0994	0.0856	
	A.Bias	0.0006	0.0128	0.0073	0.0082	
	MSE	0.0007	0.0130	0.0099	0.0073	
$N = 50$	mean	0.7988	0.5074	0.2883	0.8039	0.0764
	S.E.	0.0155	0.0725	0.0739	0.0497	
	A.Bias	0.0012	0.0074	0.0117	0.0085	
	MSE	0.0002	0.0053	0.0055	0.0025	
$N = 100$	mean	0.8007	0.5085	0.3013	0.8037	0.0489
	S.E.	0.0135	0.0428	0.0418	0.0330	
	A.Bias	0.0007	0.0085	0.0013	0.0087	
	MSE	0.0002	0.0018	0.0017	0.0011	

表 2 为估计方法在自相关结构下参数部分和非参数部分的模拟结果. 与独立结构的表现类似, 偏度值在零值附近波动, 均方误随着个体数的增加不断减小, 表明估计方法在自相关结构下的表现满足一致性; 从标准误可知, 参数估计的精度随样本个体的增加不断提高, 说明在自相关结构下参数估计的表现较为稳健; AMSE $_{\eta}$  的数值说明, 随着个体数的增加, 估计的连接函数逐渐接近于真实连接函数.

表 3 估计方法在可交换结构下的表现

size	Indicator	$\beta$	$\theta_1$	$\theta_2$	$\theta_3$	AMSE $_{\eta}$
$N = 30$	mean	0.8006	0.5194	0.2952	0.7889	0.1115
	S.E.	0.0175	0.0895	0.0880	0.0706	
	A.Bias	0.0006	0.0194	0.0048	0.0235	
	MSE	0.0003	0.0080	0.0077	0.0050	
$N = 50$	mean	0.7996	0.5067	0.2978	0.8041	0.0665
	S.E.	0.0175	0.0516	0.0608	0.0402	
	A.Bias	0.0004	0.0067	0.0022	0.0083	
	MSE	0.0003	0.0027	0.0037	0.0016	
$N = 100$	mean	0.7994	0.5090	0.2979	0.8053	0.0462
	S.E.	0.0134	0.0375	0.0390	0.0269	
	A.Bias	0.0006	0.0090	0.0021	0.0071	
	MSE	0.0002	0.0014	0.0015	0.0007	

表3为估计方法在等相关结构下参数部分和非参数部分的模拟结果. 与前两种结构的表现相似, 参数估计值分布在真实周围, 且异常接近, 参数估计在等相关结构下的表现较为稳健.  $AMSE_\eta$  的数值较小, 说明了估计的连接函数接近于真实的连接函数.

**例5** (与广义估计方程法(GEE)的比较) 为更直观地评估估计方法的表现, 考虑数据产生过程:

$$y_{it} = x_{it}\beta + \sin\pi(z_{it,1}\theta_1 + z_{it,2}\theta_2 + z_{it,3}\theta_3 - A)/(C - A) + \alpha_i + \varepsilon_{it}, \quad (24)$$

其中,  $x_{it}$  及  $\alpha_i$  与例4相同,  $z_{it} = (z_{it,1}, z_{it,2}, z_{it,3})'$  独立取值均匀分布  $U(0, 1)$ ,  $\beta = 1$ ,  $\theta_1 = \sqrt{1/2}$ ,  $\theta_2 = \sqrt{1/3}$ ,  $\theta_3 = \sqrt{1/6}$ ,  $A = \sqrt{3}/2 - 1.645/\sqrt{12}$ ,  $C = \sqrt{3}/2 + 1.645/\sqrt{12}$ , 自相关结构和等相关结构所采用的相关系数  $\rho = 0.5$ , 边际方差为1. 模拟结果呈现在表4中.

表4 估计方法与GEE法的比较

相关结构	参数	PEL	GEE
		绝对偏误(标准误)	绝对偏误(标准误)
独立结构	$\beta$	0.0001 (0.0256)	0.0015 (0.0284)
	$\theta_1$	0.0013 (0.0664)	0.0149 (0.1018)
	$\theta_2$	0.0115 (0.0753)	0.0304 (0.1824)
	$\theta_3$	0.0025 (0.0837)	0.0087 (0.1364)
自相关结构	$AMSE_\eta$	0.0938	0.4136
	$\beta$	0.0002 (0.0189)	0.0006 (0.0167)
	$\theta_1$	0.0028 (0.0557)	0.0119 (0.0763)
	$\theta_2$	0.0098 (0.0617)	0.0066 (0.1243)
可交换结构	$\theta_3$	0.0044 (0.0634)	0.0093 (0.1031)
	$AMSE_\eta$	0.0839	0.4317
	$\beta$	0.0006 (0.0158)	0.0009 (0.0164)
	$\theta_1$	0.0010 (0.0433)	0.0066 (0.0500)
	$\theta_2$	0.0072 (0.0584)	0.0018 (0.0633)
	$\theta_3$	0.0004 (0.0536)	0.0018 (0.0795)
	$AMSE_\eta$	0.0979	0.1540

从表4中可以发现, 在独立结构和等相关结构下, 从绝对偏度、参数标准误和非参数部分的  $AMSE_\eta$  值看, PEL 法均优于 GEE 法, 在自相关结构下, 两种估计方法的偏度各有优劣, 但在参数标准误和  $AMSE_\eta$  看, PEL 法要优于 GEE 法. 总之, 与 GEE 法相比, PEL 法的表现更加稳健, 参数估计和非参数估计值更加有效.

## 5.2 真实数据应用

改革开放以来, 我国居民收入持续增加, 人民生活水平不断提高, 但居民消费率一直增

长缓慢。影响居民消费率的因素众多，本文主要研究人口结构与居民消费率的关系。

为此，我们选取的变量主要有：居民消费率 (YC)、老人抚养比 (LR)、儿童抚养比 (CR)、大专以上人口比 (ER)、男性比例 (GR) 和第三产业占 GDP 的比重 (TR)，各变量的涵义及计算方法见表 5，数据来源于 2008 年–2014 年中国统计年鉴。为弄清楚哪些变量

表 5 各变量涵义及计算方法

变量	涵义	计算方法
YC	居民消费率	居民消费占 GDP 的比重
LR	老人抚养比	65 岁以上人口与 14–65 岁人口之比
CR	儿童抚养比	14 岁以下人口与 14–65 岁人口之比
ER	大专以上人口比	具有大专以上学历人口占总人口的比重
GR	男性比例	男性占总人口的比重
TR	第三产业的比重	第三产业占 GDP 的比重

对居民消费率有非线性影响，我们首先考虑以下线性面板回归方程：

$$\begin{aligned} YC_{it} = & \beta_1 LR_{it} + \beta_2 LR_{it}^2 + \beta_3 CR_{it} + \beta_4 CR_{it}^2 + \beta_5 ER_{it} + \beta_6 ER_{it}^2 \\ & + \beta_7 GR_{it} + \beta_8 GR_{it}^2 + \beta_9 TR_{it} + \beta_{10} TR_{it}^2 + \alpha_i + e_{it}, \end{aligned} \quad (25)$$

其中， $\alpha_i$  为个体  $i$  的固定效应， $e_{it}$  为误差项。回归结果呈现在表 6 中，从表 6 可以看到，在 5% 水平下，只有变量 TR 及  $TR^2$  影响不显著，为此，我们去掉变量  $TR^2$ ，再次考虑以下线性面板回归方程：

$$\begin{aligned} YC_{it} = & \beta_1 LR_{it} + \beta_2 LR_{it}^2 + \beta_3 CR_{it} + \beta_4 CR_{it}^2 + \beta_5 ER_{it} + \beta_6 ER_{it}^2 \\ & + \beta_7 GR_{it} + \beta_8 GR_{it}^2 + \beta_9 TR_{it} + \alpha_i + e_{it}. \end{aligned} \quad (26)$$

回归结果呈现在表 6 中，从表 6 可知，在 5% 水平下，所有变量影响均显著，说明变量 TR 对 YC 具有线性影响，其他变量对 YC 具有非线性影响。根据上述分析，我们建立以下固定

表 6 线性面板的估计结果

参数	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$	$\beta_{10}$
模型 (25)	6.424 (***)	-0.228 (***)	-4.747 (***)	0.118 (***)	-2.658 (***)	0.047 (***)	-7.708 (*)	0.036 (*)	0.161	0.006
模型 (26)	6.424 (***)	-0.234 (***)	-4.667 (***)	0.117 (***)	-2.692 (***)	0.049 (***)	-7.677 (*)	0.036 (*)	0.613 (***)	

注：“\*\*\*”、“\*\*”、“\*” 分别表示在 0.1%、1%、5% 下显著，其他表示不显著。

效应部分线性单指标面板模型：

$$YC_{it} = \beta TR_{it} + \eta(\theta_1 LR_{it} + \theta_2 CR_{it} + \theta_3 ER_{it} + \theta_4 GR_{it}) + \alpha_i + e_{it}, \quad (27)$$

其中,  $\eta(\cdot)$  为未知连接函数. 由于不清楚误差项的结构, 因而采用自相关和可交换结构的混合结构, 应用本文估计方法对模型 (27) 进行估计, 估计结果分别呈现在表 7 和图 1 中.

表 7 单指标模型参数部分估计结果

参数	估计值	标准误	t 统计量
$\beta$	0.694	0.104	6.662
$\theta_2$	0.325	0.118	2.753
$\theta_3$	0.326	0.143	2.284
$\theta_4$	0.575	0.013	45.145

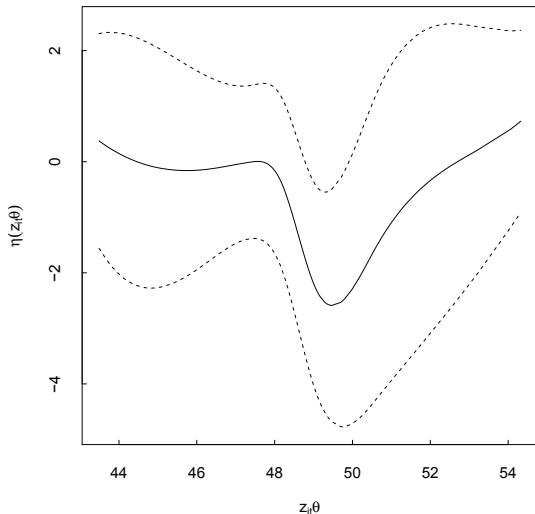


图 1 非线性影响因素对消费率的综合影响

从表 7 可以看到,  $\beta$  的估计值为  $0.694 > 0$ , 根据其 t 统计量可以判断出该估计量为显著的, 说明第三产业比重的上升将导致消费率呈线性上升, 与表 6 所得出结论相同; 在非线性部分参数估计值中, 各系数估计值均显著且为正数, 其中  $\theta_1 = \sqrt{1 - \theta_2^2 - \theta_3^2 - \theta_4^2} = 0.676$ . 图 1 描述了老人抚养比、儿童抚养比、大专以上人口比和男女人口比组成的综合因素对消费率的非线性影响, 图中虚线为连接函数估计量在 95% 水平下的置信区间. 从图中可知, 消费率与单指标  $0.691z_1 + 0.229z_2 + 0.255z_3 + 0.637z_4$  呈开口向上的抛物线型关系. 在表 6 的估计结果中, 老人抚养比对消费率影响呈开口向下的抛物线型影响, 儿童抚养比、大专以上人口比和男女人口比对消费率均呈开口向上的抛物线形影响. 因此, 结合图 1 中的结果表明, 表 6 中的估计结果具有一定的合理性.

## §6. 定理的证明

定理 1–3 的证明, 需要下列结论的支持, 为方便起见, 首先引入一些符号:

记  $g_{i,1}(\boldsymbol{\vartheta}_1) = g_{i,1}(\boldsymbol{\vartheta}_1 \mid \boldsymbol{\vartheta}_{2,0})$ ,  $g_{i,2}(\boldsymbol{\vartheta}_2) = g_{i,2}(\boldsymbol{\vartheta}_2 \mid \boldsymbol{\vartheta}_{1,0})$ ,  $\ell_1(\boldsymbol{\vartheta}_1) = \ell_1(\boldsymbol{\vartheta}_1 \mid \boldsymbol{\vartheta}_{2,0})$ ,  $\ell_2(\boldsymbol{\vartheta}_2) = \ell_2(\boldsymbol{\vartheta}_2 \mid \boldsymbol{\vartheta}_{1,0})$ .

**引理 6** 在假设条件 A1 下, 存在依赖于函数  $\eta(t)$  的参数  $\gamma_0 \in \mathbf{R}^p$ , 使得

$$\sup_{t \in [0,1]} |\eta(t) - \mathbf{B}'(t)\gamma_0| \leq Ch^r. \quad (28)$$

**证明:** 详细证明过程可参见文献 [26] 的推论 6.21.  $\square$

**引理 7** 在定理 1 的假设条件下, 有

$$N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \xrightarrow{L} \mathbf{N}(0, \mathbf{C}_{1,0}), \quad (29)$$

其中,  $\mathbf{C}_{1,0}$  的定义见假设条件 A10.

**证明:** 首先注意到

$$g_{i,1}(\boldsymbol{\vartheta}_{1,0}) = \begin{pmatrix} \dot{u}'_{i,1}\Gamma_{i1}[Y_i^* - X_i^*\boldsymbol{\beta}_0 - \tilde{\eta}^*(Z_i, \boldsymbol{\theta}_0, \gamma_0)] \\ \vdots \\ \dot{u}'_{i,1}\Gamma_{is}[Y_i^* - X_i^*\boldsymbol{\beta}_0 - \tilde{\eta}^*(Z_i, \boldsymbol{\theta}_0, \gamma_0)] \end{pmatrix},$$

其中,  $\Gamma_{ih} = A_i^{-1/2}M_hA_i^{-1/2}$ ,  $i = 1, 2, \dots, N$ ,  $h = 1, 2, \dots, s$ . 令  $\mathbf{P} = L \otimes \mathbf{I}_N$ , 我们有

$$\sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) = \begin{pmatrix} \dot{u}'_1\Gamma_1\mathbf{P}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0 - \tilde{\eta}(\mathbf{Z}, \boldsymbol{\theta}_0, \gamma_0)] \\ \vdots \\ \dot{u}'_1\Gamma_s\mathbf{P}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0 - \tilde{\eta}(\mathbf{Z}, \boldsymbol{\theta}_0, \gamma_0)] \end{pmatrix}, \quad (30)$$

其中,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$ ;  $\mathbf{X}$  和  $\mathbf{Z}$  的表达式与  $\mathbf{Y}$  类似;  $\dot{u}_1 = (\dot{u}'_{1,1}, \dot{u}'_{2,1}, \dots, \dot{u}'_{N,1})'$ ;  $\Gamma_h = \text{diag}(\Gamma_{1h}, \Gamma_{2h}, \dots, \Gamma_{Nh})$ ;  $\mathbf{I}_N$  为  $N$  阶单位矩阵;  $\tilde{\eta}(\mathbf{Z}, \boldsymbol{\theta}, \gamma) = ([\tilde{\eta}(Z_1, \boldsymbol{\theta}, \gamma)]', [\tilde{\eta}(Z_2, \boldsymbol{\theta}, \gamma)]', \dots, [\tilde{\eta}(Z_N, \boldsymbol{\theta}, \gamma)]')'$ . 由假设条件 A8、A9、引理 6 及矩阵  $\mathbf{P}$  的奇异值有界, 则对任意  $k = 1, 2, \dots, s$ , 必存在某个常数  $c^*$ , 使得  $\dot{u}'_1\Gamma_k\mathbf{P}[\eta(\mathbf{Z}, \boldsymbol{\theta}_0) - \tilde{\eta}(\mathbf{Z}, \boldsymbol{\theta}_0, \gamma_0)] \leq c^* \sum_{i=1}^N \dot{u}_{i,1}\Gamma_{i1}[\eta(Z_i, \boldsymbol{\theta}_0) - \mathbf{B}(Z_i, \boldsymbol{\theta}_0, \gamma_0)] = O_p(h^r N^{1/2})$ , 因此,

$$\begin{aligned} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) &= \begin{pmatrix} \dot{u}'_1\Gamma_1\mathbf{P}\mathcal{E} \\ \vdots \\ \dot{u}'_1\Gamma_s\mathbf{P}\mathcal{E} \end{pmatrix} + \begin{pmatrix} \dot{u}'_1\Gamma_1\mathbf{P}[\eta(\mathbf{Z}, \boldsymbol{\theta}_0) - \mathbf{B}(\mathbf{Z}, \boldsymbol{\theta}_0, \gamma_0)] \\ \vdots \\ \dot{u}'_1\Gamma_s\mathbf{P}[\eta(\mathbf{Z}, \boldsymbol{\theta}_0) - \mathbf{B}(\mathbf{Z}, \boldsymbol{\theta}_0, \gamma_0)] \end{pmatrix} \\ &= \begin{pmatrix} \dot{u}'_1\Gamma_1\mathbf{P}\mathcal{E} \\ \vdots \\ \dot{u}'_1\Gamma_s\mathbf{P}\mathcal{E} \end{pmatrix} + O(h^r N^{1/2}), \end{aligned} \quad (31)$$

其中  $\mathcal{E} = (e'_1, e'_2, \dots, e'_N)'$ . 由大数定律知

$$N^{-1} \mathbb{E} \begin{pmatrix} \dot{u}'_1 \Gamma_1 P \mathcal{E} \\ \vdots \\ \dot{u}'_1 \Gamma_s P \mathcal{E} \end{pmatrix} \begin{pmatrix} \dot{u}'_1 \Gamma_1 P \mathcal{E} \\ \vdots \\ \dot{u}'_1 \Gamma_s P \mathcal{E} \end{pmatrix}' \xrightarrow{\text{P}} C_{1,0}.$$

因此,

$$N^{-1} \sum_{i=1}^N g_{i,1}(\vartheta_{1,0}) g'_{i,1}(\vartheta_{1,0}) \xrightarrow{\text{P}} C_{1,0}. \quad (32)$$

取任意  $a \in \mathbf{R}^{s(q+d-1)}$ , 且满足  $a'a = 1$ . 由假设条件 A4 有  $\mathbb{E}[a'g_{i,1}(\vartheta_{1,0})] = 0$ , 又由矩阵  $L$  的奇异值有界可知

$$\begin{aligned} \sup_i \mathbb{E}[\|a'g_{i,1}(\vartheta_{1,0})\|^{2+\delta}] &\leq \|a'\|^{2+\delta} \sup_i \mathbb{E}[\|g_{i,1}(\vartheta_{1,0})\|^{2+\delta}] \\ &\leq \sup_i \mathbb{E}(\|L e_i\|^{2+\delta}) \leq \sup_i \mathbb{E}(\|e_i\|^{2+\delta}) < \infty. \end{aligned}$$

这表明,  $a'g_{i,1}(\vartheta_{1,0})$  满足 Lyapunov 条件, 根据 Lyapunov 中心极限定理知

$$\frac{\sum_{i=1}^N a'g_{i,1}(\vartheta_{1,0})}{\sum_{i=1}^N a'g_{i,1}(\vartheta_{1,0}) g'_{i,1}(\vartheta_{1,0}) a} \xrightarrow{L} N(0, 1). \quad (33)$$

因而可得

$$N^{-1/2} \sum_{i=1}^N g_{i,1}(\vartheta_{1,0}) \xrightarrow{L} N(0, C_{1,0}). \quad \square \quad (34)$$

**引理 8** 在定理 1 的假设条件下, 有

$$\max_{1 \leq i \leq N} \|g_{i,1}(\vartheta_{1,0})\| = o(N^{1/2}). \quad (35)$$

**证明:** 由假设条件 A4 和文献 [32] 立即可得, 证毕.  $\square$

**引理 9** 在定理 1 的假设条件下, 有

$$\tau_1 = O_p(N^{-1/2}). \quad (36)$$

**证明:** 由引理 7 知  $N^{-1} \sum_{i=1}^N g_{i,1}(\vartheta_{1,0}) = O_p(N^{-1/2})$ , 运用类似与文献 [32] 中式 (2.14) 的证明可知式 (36) 成立, 证毕.  $\square$

**定理 1 的证明:** 令  $\Theta_1 = \{(\tau_1, \vartheta_1) | \|\tau_1\| = Ch^r, \|\vartheta_1 - \vartheta_{1,0}\| = Ch^r\}$ ,  $\Theta_2 = \{(\tau_2, \vartheta_2) | \|\tau_2\| = Ch^r, \|\vartheta_2 - \vartheta_{2,0}\| = Ch^r\}$ , 其中,  $C$  为某个充分大的常数;  $Q_1(\tau_1, \vartheta_1) =$

$$\ell_1(\tau_1, \boldsymbol{\vartheta}_1 | \boldsymbol{\vartheta}_{2,0}) + \sum_{j=1}^q p_{\lambda_N}(|\beta_j|) + \sum_{l=1}^{d-1} p_{\lambda_N}(|\theta_l^*|), Q_2(\tau_2, \boldsymbol{\vartheta}_2) = \ell_2(\tau_2, \boldsymbol{\vartheta}_2 | \boldsymbol{\vartheta}_{1,0}) + p_{\lambda_N}(\|\gamma\|_{K_\theta}).$$

我们须证明, 对任意给定  $\varepsilon > 0$ , 使得对  $(\tau_1, \boldsymbol{\vartheta}_1) \in \Theta_1$ ,  $(\tau_2, \boldsymbol{\vartheta}_2) \in \Theta_2$  满足

$$\mathbb{P}\left\{\inf_{(\tau_1, \boldsymbol{\vartheta}_1) \in \Theta_1} Q(\tau_1, \boldsymbol{\vartheta}_1) > Q(0, \boldsymbol{\vartheta}_{1,0})\right\} \geq 1 - \varepsilon. \quad (37)$$

$$\mathbb{P}\left\{\inf_{(\tau_2, \boldsymbol{\vartheta}_2) \in \Theta_2} Q(\tau_2, \boldsymbol{\vartheta}_2) > Q(0, \boldsymbol{\vartheta}_{2,0})\right\} \geq 1 - \varepsilon. \quad (38)$$

这表明, 至少以概率  $1 - \varepsilon$  使得局部最小值落在  $\Theta_1$  和  $\Theta_2$  内. 因而局部最小值满足

$$\|\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{1,0}\| = O(h^r), \quad \|\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_{2,0}\| = O(h^r).$$

为此, 经过简单计算有

$$\begin{aligned} & Q_1(\tau_1, \boldsymbol{\vartheta}_1) - Q_1(0, \boldsymbol{\vartheta}_{1,0}) \\ &= \ell_1(\tau_1, \boldsymbol{\vartheta}_1) - \ell_1(0, \boldsymbol{\vartheta}_{1,0}) + \sum_{j=1}^q [p_{\lambda_N}(|\beta_j|) - p_{\lambda_N}(|\beta_{0,j}|)] + \sum_{l=1}^{d-1} [p_{\lambda_N}(|\theta_l^*|) - p_{\lambda_N}(|\theta_{0,l}^*|)] \\ &= \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3. \end{aligned}$$

首先看第一项  $\mathbf{I}_1$ ,

$$\begin{aligned} \mathbf{I}_1 &= \ell_1(\tau_1, \boldsymbol{\vartheta}_1) - \ell_1(0, \boldsymbol{\vartheta}_{1,0}) \\ &= \tau'_1 \left[ N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + \tau'_1 \left[ N^{-1} \sum_{i=1}^N \dot{g}_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] (\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{1,0}) \\ &\quad + (\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{1,0})' \left[ N^{-1} \sum_{i=1}^N \dot{g}'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] \tau_1 - \tau'_1 \left[ N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] \tau_1 \\ &\quad + o_p(\|\tau'_1\|, \|(\boldsymbol{\vartheta}'_1 - \boldsymbol{\vartheta}'_{1,0})'\|). \end{aligned}$$

由引理 7 的证明知  $N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) = O_p(N^{-1/2})$ ,  $N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) = O_p(1)$ .

又由假设条件 A8、A9 及 A10 有  $N^{-1} \sum_{i=1}^N \dot{g}_{i,1}(\boldsymbol{\vartheta}_{1,0}) = O_p(1)$ , 因此, 可知

$$\mathbf{I}_1 = CO_p(N^{-1/2}h^r) + C^2O_p(h^{2r}). \quad (39)$$

接着由泰勒展式可知

$$\begin{aligned} \mathbf{I}_2 &= \sum_{j=1}^q [p_{\lambda_N}(|\beta_j|) - p_{\lambda_N}(|\beta_{0,j}|)] \\ &= \sum_{j=1}^q \dot{p}_{\lambda_N}(|\beta_{0,j}|) \operatorname{sgn}(\beta_{0,j}) |\beta_j - \beta_{0,j}| + \sum_{j=1}^q \ddot{p}_{\lambda_N}(|\beta_{0,j}|) (\beta_j - \beta_{0,j})^2 [1 + o_p(1)] \\ &\leq C\sqrt{q} a_N h^r + C^2 b_N h^{2r}. \end{aligned}$$

当  $N \rightarrow \infty$  时,  $\lambda_N \rightarrow 0$ , 因此,  $a_N \rightarrow 0$ ,  $b_N \rightarrow 0$ . 因而易证  $\mathbf{I}_1$  主控  $\mathbf{I}_2$ , 同理有  $\mathbf{I}_1$  主控  $\mathbf{I}_3$ . 通过选择充分大的  $C$ , 知式(37)成立, 利用同样的原理可以证明式(38)成立.

利用式(37)和(38)的结论, 我们有

$$\begin{aligned} & (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [\mathbf{B}'(z'_{it}\hat{\boldsymbol{\theta}}_N)\hat{\gamma}_N - \eta(z'_{it}\boldsymbol{\theta}_0)]^2 \\ & \leq (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [\mathbf{B}'(z'_{it}\hat{\boldsymbol{\theta}}_N)\hat{\gamma}_N - \mathbf{B}'(z'_{it}\boldsymbol{\theta}_0)\gamma_0]^2 + O_p(h^{2r}) \\ & \leq (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [\dot{\mathbf{B}}'(z'_{it}\boldsymbol{\theta}_0)\gamma_0(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) + \mathbf{B}'(z'_{it}\boldsymbol{\theta}_0)(\hat{\gamma}_N - \gamma_0) + O(h^r)]^2 + O_p(h^{2r}) \\ & = O_p(h^{2r}). \quad \square \end{aligned}$$

**定理2的证明:** 令  $(\hat{\tau}'_{1,N}, \hat{\boldsymbol{\vartheta}}'_{1,N})' = \min_{\tau_1, \boldsymbol{\vartheta}_1} Q_1(\tau_1, \boldsymbol{\vartheta}_1)$ , 因此有

$$N^{-1} \sum_{i=1}^N \frac{g_{i,1}(\hat{\boldsymbol{\vartheta}}_{1,N})}{1 + \hat{\tau}'_{1,N} g_{i,1}(\hat{\boldsymbol{\vartheta}}_{1,N})} = 0, \quad (40)$$

$$N^{-1} \sum_{i=1}^N \frac{[\partial g_{i,1}(\hat{\boldsymbol{\vartheta}}_{1,N})/\partial \boldsymbol{\vartheta}_1]' \hat{\tau}_{1,N}}{1 + \hat{\tau}'_{1,N} g_{i,1}(\hat{\boldsymbol{\vartheta}}_{1,N})} + \sum_{j=1}^q \dot{p}_{\lambda_N}(|\hat{\beta}_{j,N}|) \operatorname{sgn}(\hat{\beta}_{j,N}) + \sum_{l=1}^{d-1} \dot{p}_{\lambda_N}(|\hat{\theta}_{l,N}^*|) \operatorname{sgn}(\hat{\theta}_{l,N}^*) = 0. \quad (41)$$

注意到  $\dot{p}_{\lambda_N}(|\hat{\beta}_{j,N}|) = \dot{p}_{\lambda_N}(|\beta_{j,0}|) + [\ddot{p}_{\lambda_N}(|\beta_{j,0}|) + o_p(1)](\hat{\beta}_{j,N} - \beta_{j,0})$ , 假设条件A9表明  $\ddot{p}_{\lambda_N}(|\beta_{j,0}|) = o_p(1)$ , 且当  $\lambda_N \rightarrow \infty$  时,  $\dot{p}_{\lambda_N}(|\beta_{j,0}|) = 0$ , 这样立即可得

$$\sum_{j=1}^q \dot{p}_{\lambda_N}(|\hat{\beta}_{j,N}|) \operatorname{sgn}(\hat{\beta}_{j,N}) + \sum_{l=1}^{d-1} \dot{p}_{\lambda_N}(|\hat{\theta}_{l,N}^*|) \operatorname{sgn}(\hat{\theta}_{l,N}^*) = o_p(\|\hat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}\|).$$

将式(40)和式(41)在  $(0', \boldsymbol{\vartheta}'_{1,0})'$  处实施泰勒展式并将上式代入有:

$$\begin{aligned} 0 &= N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) + N^{-1} \sum_{i=1}^N \dot{g}_{i,1}(\boldsymbol{\vartheta}_{1,0})(\hat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}) \\ &\quad - N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0})(\hat{\tau}_{1,N} - 0) + o_p(\|\hat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}\| + \|\hat{\tau}_{1,N}\|), \end{aligned} \quad (42)$$

$$0 = N^{-1} \sum_{i=1}^N \dot{g}'_{i,1}(\boldsymbol{\vartheta}_{1,0})(\hat{\tau}_{1,N} - 0) + o_p(\|\hat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}\| + \|\hat{\tau}_{1,N}\|). \quad (43)$$

令  $S_{22} = N^{-1} \sum_{i=1}^N \dot{g}'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \left[ \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]^{-1} \sum_{i=1}^N \dot{g}_{i,1}(\boldsymbol{\vartheta}_{1,0})$ , 经过简单计算有

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}) = S_{22}^{-1} S_{21} S_{11}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) + o_p(1),$$

其中,  $S_{11} = -N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0})$ ,  $S_{21} = N^{-1} \sum_{i=1}^N \dot{g}'_{i,1}(\boldsymbol{\vartheta}_{1,0})$ . 由引理 7 可得

$$\sqrt{N}(\widehat{\boldsymbol{\vartheta}}_{1,N} - \boldsymbol{\vartheta}_{1,0}) \xrightarrow{L} N(0, \mathbf{V}), \quad (44)$$

其中,  $\mathbf{V} = \{\mathbf{G}'_{1,0} \mathbf{C}_{1,0}^{-1} \mathbf{G}_{1,0}\}^{-1}$ .  $\square$

**定理 3 的证明:** 记  $\ell_1(\boldsymbol{\vartheta}_{1,0}) = 2 \sum_{i=1}^N \ln[1 + \tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})]$ , 其中,  $\tau_{1,0} = \tau_1(\boldsymbol{\vartheta}_{1,0})$ . 运用泰勒展式有

$$\begin{aligned} \ell_1(\boldsymbol{\vartheta}_{1,0}) &= 2 \sum_{i=1}^N \ln[1 + \tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})] \\ &= 2 \sum_{i=1}^N \{\tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0}) - 2^{-1} [\tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})]^2\} + o_p(1). \end{aligned} \quad (45)$$

又因为

$$\begin{aligned} 0 &= N^{-1} \sum_{i=1}^N \frac{g_{i,1}(\boldsymbol{\vartheta}_{1,0})}{1 + \tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})} \\ &= N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) - N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \tau_{1,0} \\ &\quad + N^{-1} \sum_{i=1}^N \frac{g_{i,1}(\boldsymbol{\vartheta}_{1,0}) [\tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})]^2}{1 + \tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})}. \end{aligned}$$

根据引理 7-9 知

$$\tau_{1,0} = \left[ N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]^{-1} \left[ N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + o_p(N^{-1/2}). \quad (46)$$

$$\sum_{i=1}^N [\tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0})]^2 = \sum_{i=1}^N \tau'_{1,0} g_{i,1}(\boldsymbol{\vartheta}_{1,0}) + o_p(1). \quad (47)$$

由式 (46)–(47) 及引理 7 可得

$$\begin{aligned} \ell_1(\boldsymbol{\vartheta}_{1,0}) &= \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]' \left[ N^{-1} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) g'_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]^{-1} \\ &\quad \times \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + o_p(1). \end{aligned}$$

进一步, 根据引理 7 的证明知

$$\begin{aligned} \ell_1(\widehat{\tau}_{1,N}, \widehat{\boldsymbol{\vartheta}}_{1,N}) &= - \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]' S_{11}^{-1} (\mathbf{I} + S_{12} S_{22}^{-1} S_{21} S_{11}^{-1}) \\ &\quad \times \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + o_p(1), \end{aligned}$$

其中,  $\mathbf{I}$  表示单位矩阵. 于是,

$$\begin{aligned} W(\boldsymbol{\vartheta}_{1,0}) &= \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]' (S_{11}^{-1} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1}) \\ &\quad \times \left[ N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + o_p(1) \\ &= \left[ (-S_{11}^{-1/2}) N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right]' (S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2}) \\ &\quad \times \left[ (-S_{11}^{-1/2}) N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0}) \right] + o_p(1). \end{aligned}$$

引理 7 表明  $S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2}$  为对称的幂等矩阵, 其迹为  $(q+d-1)$ ,  $(-S_{11}^{-1/2}) \times N^{-1/2} \sum_{i=1}^N g_{i,1}(\boldsymbol{\vartheta}_{1,0})$  渐近于多元标准正态分布. 因此,  $W(\boldsymbol{\vartheta}_{1,0})$  渐近于  $\chi^2_{(q+d-1)}$ .  $\square$

## §7. 结论总结

本文考虑了一类个体内存在相关性的固定效应部分线性单指标面板模型, 结合惩罚经验似然法与虑子法对模型进行估计. 具体地, 首先用样条函数近似单指标模型中的连接函数, 再运用 QIF 法的思想获得一个关于未知参数的扩展计分函数, 接着利用该扩展计分函数构建关于未知参数的经验似然函数. 由于用样条函数近似未知连接函数会引起过度拟合问题, 为此, 对所构建的经验似然函数实施一个惩罚项, 以克服过度参数化问题. 另一方面, 运用 Lagrange 乘数法估计所构建的惩罚经验似然函数获得未知参数的估计值, 从而获得所有线性部分和非线性部分的估计; 在适当的假设条件下, 证明了所得估计量的一致性和渐近正态性; 进一步, 通过 Monte Carlo 模拟评估了所述方法在有限样本下的表现, 模拟结果显示: (i) 在参数估计方面, 参数估计值在各相关结构下均接近于真实值, 标准误和均方误均随着个体数  $N$  的增大而减少, 说明模型估计的表现满足一致性, 在非参数方面, 随着个体数  $N$  的增大, 估计的函数在各相关结构下均逐渐接近于真实函数; (ii) 本文方法与 GEE 法的模拟比较发现, 在独立结构和等相关结构下, PEL 法的各项指标均优于 GEE 法, 在相关结构下, 除参数估计值的偏度外, PEL 法其他指标均优于 GEE 法; 最后利用所述方法分析了人口结构对消费率的影响, 结果表明, 产业结构对消费率有线性正向影响, 老人抚养比、儿童抚养比、大专以上人口比和男女比例所构成的单指标对消费率呈抛物线型曲线影响.

## 参 考 文 献

- [1] HAMILTON J D. *Time Series Analysis* [M]. New Jersey: Princeton University Press, 1994.

- [2] OWEN A B. Empirical likelihood ratio confidence intervals for a single functional [J]. *Biometrika*, 1988, **75**(2): 237–249.
- [3] OWEN A B. *Empirical Likelihood* [M]. New York: Chapman and Hall/CRC, 2001.
- [4] XUE L G, ZHU L X. Empirical likelihood for single-index models [J]. *J Multivariate Anal*, 2006, **97**(6): 1295–1312.
- [5] ZHUL X, XUE L G. Empirical likelihood confidence regions in a partially linear single-index model [J]. *J R Stat Soc Ser B Stat Methodol*, 2006, **68**(3): 549–570.
- [6] ZHU L X, LIN L, CUI X, et al. Bias-corrected empirical likelihood in a multi-link semiparametric model [J]. *J Multivariate Anal*, 2010, **101**(4): 850–868.
- [7] XUE L G, XUE D. Empirical likelihood for semiparametric regression model with missing response data [J]. *J Multivariate Anal*, 2011, **102**(4): 723–740.
- [8] XUE L G. Estimation and empirical likelihood for single-index models with missing data in the covariates [J]. *Comput Statist Data Anal*, 2013, **68**: 82–97.
- [9] OTSU T. Penalized empirical likelihood estimation of semiparametric models [J]. *J Multivariate Anal*, 2007, **98**(10): 1923–1954.
- [10] FAN G L, LIANG H Y, SHEN Y. Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors [J]. *J Multivariate Anal*, 2016, **147**: 183–201.
- [11] XUE L G, ZHU L X. Empirical likelihood semiparametric regression analysis for longitudinal data [J]. *Biometrika*, 2007, **94**(4): 921–937.
- [12] TANG N S, ZHAO P Y. Empirical likelihood semiparametric nonlinear regression analysis for longitudinal data with responses missing at random [J]. *Ann Inst Statist Math*, 2013, **65**(4): 639–665.
- [13] YANG H, LI T T. Empirical likelihood for semiparametric varying coefficient partially linear models with longitudinal data [J]. *Statist Probab Lett*, 2010, **80**(2): 111–121.
- [14] ZHOU X C, LIN J G. Empirical likelihood inference in mixtures of semiparametric varying coefficient EV models for longitudinal data with nonignorable dropout [J]. *J Korean Statist Soc*, 2013, **42**(2): 215–225.
- [15] YANG Y P, LI G R, PENG H. Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data [J]. *J Multivariate Anal*, 2014, **127**: 1–18.
- [16] ZHAO P X, YANG Y P. Semiparametric empirical likelihood tests in varying coefficient partially linear models with repeated measurements [J]. *Stat Methodol*, 2015, **23**: 73–87.
- [17] ZHANG J H, FENG S Y, LI G R, et al. Empirical likelihood inference for partially linear panel data models with fixed effects [J]. *Econom Lett*, 2011, **113**(2): 165–167.
- [18] HE B Q, HONG X J, FAN G L. Empirical likelihood for semi-varying coefficient models for panel data with fixed effects [J]. *J Korean Statist Soc*, 2016, **45**(3): 395–408.
- [19] YOU J H, CHEN G M, ZHOU Y. Block empirical likelihood for longitudinal partially linear regression models [J]. *Canad J Statist*, 2006, **34**(1): 79–96.
- [20] LI G R, TIAN P, XUE L G. Generalized empirical likelihood inference in semiparametric regression model for longitudinal data [J]. *Acta Math Sin, Engl Ser*, 2008, **24**(12): 2029–2040.
- [21] LI G R, ZHU L X, XUE L G, et al. Empirical likelihood inference in partially linear single-index models for longitudinal data [J]. *J Multivariate Anal*, 2010, **101**(3): 718–732.
- [22] HU S, LIN L. Empirical likelihood analysis of longitudinal data involving within-subject correlation [J]. *Acta Math Appl Sin, Engl Ser*, 2012, **28**(4): 731–744.

- [23] ZHOU X S, ZHOU P X, LIN L. Empirical likelihood for parameters in an additive partially linear errors-in-variables model with longitudinal data [J]. *J Korean Statist Soc*, 2014, **43**(1): 91–103.
- [24] BAI Y, ZHU Z Y, FUNG W K. Partial linear models for longitudinal data based on quadratic inference functions [J]. *Scand J Stat*, 2008, **35**(1): 104–118.
- [25] CHEN J, GAO J T, LI D G. Estimation in partially linear single-index panel data models with fixed effects [J]. *J Bus Econom Statist*, 2013, **31**(3): 315–330.
- [26] SCHUMAKER L L. *Spline Functions: Basic Theory* [M]. New York: Wiley, 1981.
- [27] HE X M, SHI P D. Convergence rate of B-spline estimators of nonparametric conditional quantile functions [J]. *J Nonparametr Stat*, 1994, **3**(3-4): 299–308.
- [28] QU A, LI R Z. Quadratic inference functions for varying-coefficient models with longitudinal data [J]. *Biometrics*, 2006, **62**(2): 379–391.
- [29] HE X M, ZHU Z Y, FUNG W K. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure [J]. *Biometrika*, 2002, **89**(3): 579–590.
- [30] XUE L, QU A, ZHOU J H. Consistent model selection for marginal generalized additive model for correlated data [J]. *J Amer Statist Assoc*, 2010, **105**(492): 1518–1530.
- [31] RUPPERT D, WAND M P, CARROLL R J. *Semiparametric Regression* [M]. London: Cambridge University Press, 2003.
- [32] OWEN A. Empirical likelihood ratio confidence regions [J]. *Ann Statist*, 1990, **18**(1): 90–120.

## Penalized Empirical Likelihood Estimation for Partially Linear Single Index Panel Model with Fixed Effects

DING Feipeng

(School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330022, China)

**Abstract:** This paper constructs a penalized empirical likelihood estimation method via quadratic inference function method, filter method and empirical likelihood estimation method. Under some regular conditions, we derived the large sample properties of estimators and show that the proposed empirical likelihood ratio is asymptotically to chi-square distribution. Furthermore, the infinite sample performance of the proposed method is evaluated by Monte Carlo simulation and real data analysis.

**Keywords:** empirical likelihood estimation; dependence structure; partially linear single index panel model

**2010 Mathematics Subject Classification:** 62G05; 62G08