

高维纵向数据的惩罚 expectile 估计^①

樊梅红, 李婷婷

西南大学 数学与统计学院, 重庆 400715

摘要: 基于期望分位数(expectile)回归理论, 提出高维纵向数据的惩罚 expectile(PGEEE)估计, 在正则条件下, 建立了估计量的 Oracle 性质. 数值模拟和实证结果表明, PGEEE 估计在实现变量选择的同时, 提供了模型回归系数的相合估计, 并且该方法可以有效识别异方差, 刻画数据的异质结构, 挖掘数据中更丰富的信息.

关键词: expectile; 惩罚 expectile 估计方程; Oracle 性质; 异方差

中图分类号: O212.1

文献标志码: A

文章编号: 1000-5471(2023)06-0070-11

Penalized Expectile Estimation for
High-Dimensional Longitudinal Data

FAN Meihong, LI Tingting

School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Abstract: Based on the expectile regression theory, the penalty expectile (PGEEE) estimation of high-dimensional longitudinal data has been proposed, and the Oracle property of the estimator has been established under some regular conditions. Numerical simulation and empirical results demonstrate that the PGEEE estimation provides a consistent estimation of the regression coefficients of the model while realizing the variable selection. Moreover, this method can effectively identify heteroscedasticity, characterize the heterogeneous structure of the data, and mine the richer information in the data.

Key words: expectile regression theory; penalized expectile estimating equation; Oracle property; heteroscedasticity

在统计学中, 有很多方法可以寻找数据间的潜在关系, 刻画数据结构. 文献[1]提出的广义估计方程(generalized estimating equation, GEE)是很常见的一种分析纵向数据的统计方法, 在研究数据内部关系及预测等方面有重要的作用. 此方法的优点是即使相关结构被误判, 所得估计仍然是相合的. 在大数据时代, 高维纵向数据能比时序数据和横截面数据提供更多的信息. 然而数据的高维性使模型变得复杂, 降低了模型的估计精度. 带惩罚项的正则化估计是解决高维数据的常用方法. 文献[2]提出的 SCAD 惩罚和文献[3]提出的 MCP 惩罚是常见的非凸惩罚方法, 具有 Oracle 性质. 文献[4-5]将 GEE 与惩罚函数相结合, 提出了

① 收稿日期: 2022-08-11

基金项目: 国家自然科学基金项目(11701469).

作者简介: 樊梅红, 硕士研究生, 主要从事数理统计方向的研究.

通信作者: 李婷婷, 副教授.

惩罚广义估计方程(penalized generalized estimating equation, PGEE), 模拟研究表明该方法在筛选出重要变量的同时得到模型回归系数的无偏估计.

在实际应用中, 数据往往会呈现异质性. 文献[6]首次提出分位数回归(quantile regression, QR)方法, 可以捕捉整个条件分布的特征. 文献[7]基于独立的数据结构提出纵向数据的线性分位数回归模型. 这不可避免地会损失估计效率. 进一步, 文献[8]考虑纵向数据重复观测样本间的相关性, 建立分位数 GEE 回归模型, 提高了估计效率. 文献[9]对纵向数据的分位数回归模型添加惩罚项, 提出了惩罚分位数回归模型.

QR 方法对应的损失函数具有不可微性, 这给数值计算带来了很大的难度, 尤其对于高维复杂数据来说, 该问题变得更加突出. 受分位数回归的启发, 文献[10]将分位数回归中的非对称绝对值损失函数替换为非对称最小平方损失函数, 提出了期望分位数(expectile)估计量. Expectile 方法不仅继承了 QR 方法可以处理异质性的优点, 且具有连续可微的损失函数, 相较 QR 方法在计算上也有很大的优势. 在独立同分布的截面数据中, 文献[11-12]将 expectile 回归与惩罚函数相结合, 提出带有惩罚项的 expectile 回归模型, 建立了 Oracle 性质, 同时实现了变量选择和异方差识别. 文献[13]将 expectile 应用到纵向数据, 提出了广义 expectile 估计方程(generalized expectile estimating equation, GEEE). 模拟结果显示, GEEE 估计量可以识别出异方差, 在保留分位数优点的同时, 降低了计算难度. 近年来, 作为 QR 方法的替代, expectile 方法受到部分学者的关注, 但在纵向数据变量选择方面的研究还不多见. 本文将截面数据的惩罚 expectile 回归模型扩展到纵向数据, 提出 PGEEE(penalized generalized expectile estimating equation)估计量. 模拟结果和实证分析显示, PGEEE 估计量不仅可以实现高维数据的变量选择, 并且同时为重要变量的回归系数进行估计. 更重要的是, PGEEE 方法可以得到一系列 τ 水平下的变量选择和模型估计结果, 详细地刻画了数据的异质结构, 能够比 GEE 提供更多的信息.

1 纵向数据的 expectile 回归和 PGEEE 模型

1.1 Expectile 回归模型和 PGEEE

定义随机变量 Y 的 τ -expectile 值为

$$\mu_\tau(Y) = \arg \min_{\theta} E(\rho_\tau(Y - \theta)) \quad (1)$$

其中 $\tau \in (0, 1)$, $\rho_\tau(\theta) = |\tau - I(\theta \leq 0)| \cdot \theta^2$ 是非对称平方损失函数, I 是示性函数. 由 τ -expectile 的定义易知, 当 $\tau = 0.5$ 时, $\rho_\tau(\cdot)$ 等价于经典的最小二乘损失函数, 则模型(1) 对应经典的均值回归模型, $\mu_\tau(Y)$ 为随机变量 Y 的数学期望.

假设有纵向样本数据 $(y_{ij}, \mathbf{X}_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, m_i$, 满足如下的 expectile 线性回归模型

$$y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta}_n + \epsilon_{ij} \quad (2)$$

其中: n 是个体的数目, m_i 是第 i 个个体的测量次数, y_{ij} 和 $\mathbf{X}_{ij} = (X_{ij}^1, \dots, X_{ij}^{p_n})^T$ 是第 i 个个体的第 j 次观测值, 总观测次数 $N = \sum_{i=1}^n m_i$. 在本文中, 协变量的维数 p_n 可以是发散的. $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$ 为误差值向量, 满足其 τ -expectile 值 $\mu_\tau(\boldsymbol{\epsilon}_i) = 0$. $\boldsymbol{\beta}_n$ 是 p_n 维未知参数, 值得注意的是, $\boldsymbol{\beta}_n$ 应与 τ 相关, 在不引起混淆的情况下, 本文忽略其下标 τ . 记 $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{im_i})^T$, 则模型(2) 可表示为

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_n + \boldsymbol{\epsilon}_i \quad (3)$$

对 $\boldsymbol{\beta}_n$ 的估计可以通过求解如下目标函数的最小值来获得, 即

$$\min_{\boldsymbol{\beta}_n} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau(y_{ij} - \mathbf{X}_{ij}^T \boldsymbol{\beta}_n) \quad (4)$$

考虑重复观测时个体内的相关性, 文献[13] 在纵向数据协变量数固定的情况下提出了 GEEE 模型, 即通过求解如下估计方程

$$S(\boldsymbol{\beta}_n) = \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{ir}^{-1} \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n)(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) = 0 \quad (5)$$

获得系数 $\boldsymbol{\beta}_n$ 的估计. 其中, $\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) = \text{diag}(\phi_\tau(y_{i1} - \mathbf{X}_{i1}^T \boldsymbol{\beta}_n), \dots, \phi_\tau(y_{im_i} - \mathbf{X}_{im_i}^T \boldsymbol{\beta}_n))$, $\phi_\tau(x) =$

$|\tau - I(x \leq 0)|$, $\hat{\mathbf{V}}_{i\tau} = \mathbf{A}_{i\tau}^{-1} \mathbf{R}_i(\hat{\boldsymbol{\alpha}}_\tau) \mathbf{A}_{i\tau}^{-1}$, $\mathbf{A}_{i\tau} = \text{diag}(\sigma_{i1}, \dots, \sigma_{im_i})$, $\sigma_{ij} = \text{Var}(y_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta}_n)$. $\mathbf{R}_i(\hat{\boldsymbol{\alpha}}_\tau)$ 称为参数为 $\boldsymbol{\alpha}_\tau$ 的工作相关矩阵, 用来代替真实的相关系数矩阵, 具有独立(IND), 自回归(AR(1)), 等相关(CS), 不确定性(UN) 相关结构等. $\hat{\boldsymbol{\alpha}}_\tau$ 是参数 $\boldsymbol{\alpha}_\tau$ 的估计, 本文使用矩估计, 表达式详见算法过程.

进一步地, 本文在协变量维数 p_n 发散的情况下, 提出纵向数据的惩罚非对称最小二乘 PGEEE 估计, 即通过求解如下估计方程

$$\mathbf{Q}(\boldsymbol{\beta}_n) = \mathbf{S}(\boldsymbol{\beta}_n) - n\mathbf{P}'_{\lambda_n}(|\boldsymbol{\beta}_n|) \text{Sign}(\boldsymbol{\beta}_n) = 0 \quad (6)$$

获得系数 $\boldsymbol{\beta}_n$ 的 PGEEE 估计. 其中, $\mathbf{P}'_{\lambda_n}(|\boldsymbol{\beta}_n|) = (p'_{\lambda_n}(|\boldsymbol{\beta}_{n1}|), \dots, p'_{\lambda_n}(|\boldsymbol{\beta}_{np_n}|))^T$, $p_{\lambda_n}(t)$ 是一个含有调节参数 λ_n 的非负惩罚函数, $p'_{\lambda_n}(t)$ 为 $p_{\lambda_n}(t)$ 的导数. $\text{Sign}(\boldsymbol{\beta}_n) = (\text{sign}(\boldsymbol{\beta}_{n1}), \dots, \text{sign}(\boldsymbol{\beta}_{np_n}))^T$, $\text{sign}(t) = I(t > 0) - I(t < 0)$ 为符号函数. $\mathbf{P}'_{\lambda_n}(|\boldsymbol{\beta}_n|) \text{Sign}(\boldsymbol{\beta}_n)$ 定义为对应元素相乘得到的向量. 本文考虑 MCP 和 SCAD 两种惩罚方法. MCP 惩罚函数的数学表达式为

$$p'_{\lambda_n}(t) = \left(\lambda_n - \frac{t}{\gamma} \right) I(|t| \leq \gamma \lambda_n) \quad (7)$$

为简化模型, 参考文献[14], 取 $\gamma = 3$. SCAD 惩罚函数的数学表达式为

$$p'_{\lambda_n}(t) = \lambda_n \left\{ I(|t| \leq \lambda_n) + \frac{(\gamma \lambda_n - t)_+}{(\gamma - 1)\lambda} I(t > \lambda_n) \right\} \quad (8)$$

根据文献[2] 建议取 $\gamma = 3.7$. 此时模型(6) 中需要选择的参数只有 λ_n , 本文使用 BIC 准则来选取, 表达式见算法过程.

1.2 求解算法

Step1: 给定一个 λ_n , 求解(4) 式获得初始值 $\hat{\boldsymbol{\beta}}_n^{(0)}$.

Step2: $\hat{\epsilon}_{i\tau} = y_{i\tau} - \mathbf{X}_{i\tau}^T \hat{\boldsymbol{\beta}}_n^{(k)}$, 计算相关系数.

$$\begin{aligned} \hat{\alpha}_{\tau, \text{CS}} &= \frac{1}{N_1 - p} \sum_{i=1}^n \sum_{t < s}^{m_i} \psi_\tau(\hat{\epsilon}_{i\tau}) \hat{\epsilon}_{i\tau} \psi_\tau(\hat{\epsilon}_{ist}) \hat{\epsilon}_{ist}, \quad N_1 = \frac{1}{2} \sum_{i=1}^n m_i(m_i - 1) \\ \hat{\alpha}_{\tau, \text{AR}(1)} &= \frac{1}{N_2 - p} \sum_{i=1}^n \sum_{t < m_i - 1} \psi_\tau(\hat{\epsilon}_{i\tau}) \hat{\epsilon}_{i\tau} \psi_\tau(\hat{\epsilon}_{i, t+1, \tau}) \hat{\epsilon}_{i, t+1, \tau}, \quad N_2 = \sum_{i=1}^n (m_i - 1) \\ \hat{\alpha}_{i\tau, \text{UN}} &= \frac{1}{N - p} \sum_{i=1}^n \psi_\tau(\hat{\epsilon}_{i\tau}) \hat{\epsilon}_{i\tau} \psi_\tau(\hat{\epsilon}_{ist}) \hat{\epsilon}_{ist} \end{aligned}$$

Step3: 使用如下迭代式计算 $\hat{\boldsymbol{\beta}}_n^{(k+1)}$:

$$\hat{\boldsymbol{\beta}}_n^{(k+1)} = \hat{\boldsymbol{\beta}}_n^{(k)} + \left[\sum_{i=1}^n \mathbf{X}_i^T \hat{\mathbf{V}}_{i\tau}^{-1} \boldsymbol{\Psi}_\tau(y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_n^{(k)}) \mathbf{X}_i + n\mathbf{E}_n(\hat{\boldsymbol{\beta}}_n^{(k)}) \right]^{-1} \cdot [\mathbf{S}(\hat{\boldsymbol{\beta}}_n^{(k)}) - n\mathbf{E}_n(\hat{\boldsymbol{\beta}}_n^{(k)})] \quad (10)$$

其中 $\mathbf{E}_n(\hat{\boldsymbol{\beta}}_n) = \text{diag} \left\{ \frac{p'_{\lambda_n}(|\hat{\beta}_1| +)}{|\hat{\beta}_1|}, \dots, \frac{p'_{\lambda_n}(|\hat{\beta}_{p_n}| +)}{|\hat{\beta}_{p_n}|} \right\}$.

Step4: 重复 Step2-Step3 直至收敛, 并计算 λ_n 对应的 BIC 值, 其表达式为

$$\text{BIC}(\lambda_n) = \log \left(\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau(y_{ij} - \mathbf{X}_{ij}^T \boldsymbol{\beta}_n) \right) + \frac{1}{N} \cdot df \cdot \log N \quad (11)$$

其中, df 表示 λ_n 对应模型所选择的变量个数.

Step5: 重复 Step1-Step4, 选取 BIC 值最小的 λ_n 作为惩罚因子, 对应的 $\hat{\boldsymbol{\beta}}_n$ 为未知参数最优解.

2 Oracle 性质

记 $m = \max\{m_i, i = 1, \dots, n\}$, 参数真值为 $\boldsymbol{\beta}_{n0} = (\boldsymbol{\beta}_{n0(1)}^T, \boldsymbol{\beta}_{n0(2)}^T)^T$. 不失一般性, 设 $\boldsymbol{\beta}_{n0(1)}$ 为 s 维非零元素构成的向量, $\boldsymbol{\beta}_{n0(2)}$ 为 $p_n - s$ 维零向量, $\boldsymbol{\beta}_{n0(2)} = \mathbf{0}$, 则协变量矩阵可记为 $\mathbf{X}_i = (\mathbf{X}_{i(1)}, \mathbf{X}_{i(2)})$, $\mathbf{X}_{i(1)} = (\mathbf{X}_{i1(1)}, \dots, \mathbf{X}_{im_i(1)})$, $\mathbf{X}_{ij(1)}$ 为 $\mathbf{X}_{ij}$ 的前 s 个元素构成的向量. $C, C_1, C_2, \dots$ 代表与 n 无关的正常数, 不同地方可以取不同

的值. 为了证明估计量的性质, 本文需要以下假设条件:

(A1) $\sup_{i,j} \|\mathbf{X}_{ij}\| = O_p(\sqrt{p_n})$; $\{(\mathbf{X}_i, \mathbf{y}_i)\}_{i=1}^n$ 独立, 且 $\text{Var}[\Psi_\tau(\mathbf{y}_i - \mathbf{X}_{i(1)}\boldsymbol{\beta}_{n0(1)})(\mathbf{y}_i - \mathbf{X}_{i(1)}\boldsymbol{\beta}_{n0(1)})^\top] = \boldsymbol{\Sigma}_{i\tau}$; 存在 $\nu > 0, \Delta > 0$, 使 $E|\phi_\tau(\epsilon_{ij\tau})|^{4+\nu} < \Delta, E|\epsilon_{ij\tau}|^{4+\nu} < \Delta$;

(A2) 估计协方差阵 $\hat{\mathbf{V}}_{i\tau}$ 满足 $\|\hat{\mathbf{V}}_{i\tau}^{-1} - \bar{\mathbf{V}}_{i\tau}^{-1}\| = O_p\left(\sqrt{\frac{p_n}{n}}\right)$, 其中 $\bar{\mathbf{V}}_{i\tau}$ 是满足 $C_1\mathbf{I}_{m_i} \leq \bar{\mathbf{V}}_{i\tau} \leq C_2\mathbf{I}_{m_i}$ 的正定阵 ($\bar{\mathbf{V}}_{i\tau}$ 可以不等于真实协方差阵 $\mathbf{V}_{i0\tau}$), $\mathbf{I}_{m_i}$ 是 m_i 阶单位阵. $\|\mathbf{A}\| = [\text{tr}(\mathbf{A}\mathbf{A}^\top)]^{\frac{1}{2}}$ 表示矩阵 $\mathbf{A}$ 的 Frobenius 范数;

(A3) 存在两个正常数 C_1, C_2 , 使得 $C_1 \leq \lambda_{\min}\left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i\right) \leq \lambda_{\max}\left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i\right) \leq C_2, \lambda_{\min}, \lambda_{\max}$ 分别表示矩阵的最小和最大特征根;

(A4) 记 $\mathbf{P}_1 = \mathbf{P}'_{\lambda_n}(|\boldsymbol{\beta}_{n0(1)}|)\text{Sign}(\boldsymbol{\beta}_{n0(1)}), \mathbf{P}_2 = \text{diag}(p''_{\lambda_n}(\beta_{10}), \dots, p''_{\lambda_n}(\beta_{s0})); \mathbf{D}_\tau = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{i(1)}^\top \bar{\mathbf{V}}_{i\tau}^{-1} \boldsymbol{\Sigma}_{i\tau} \bar{\mathbf{V}}_{i\tau}^{-1} \mathbf{X}_{i(1)}, \boldsymbol{\Phi}_\tau = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{i(1)}^\top \bar{\mathbf{V}}_{i\tau}^{-1} E[\Psi_\tau(\mathbf{y}_i - \mathbf{X}_{i(1)}\boldsymbol{\beta}_{n0(1)})] \mathbf{X}_{i(1)}$;

(A5) $\frac{1}{n} p_n^3 \rightarrow 0$;

(A6) $a_n = \max_{1 \leq k \leq p_n} \{p'_{\lambda_n}(|\beta_{j0}|), \beta_{j0} \neq 0\} = O(n^{-\frac{1}{2}}), b_n = \max_{1 \leq k \leq p_n} \{p''_{\lambda_n}(|\beta_{j0}|), \beta_{j0} \neq 0\} \rightarrow 0$;

(A7) $\liminf_{n \rightarrow \infty} \liminf_{\theta \rightarrow 0^+} \frac{p'_{\lambda_n}(\theta)}{\lambda_n} > 0, \frac{1}{\lambda_n} \sqrt{\frac{p_n}{n}} \rightarrow 0$.

定理 1 记方程(6)的解 $\hat{\boldsymbol{\beta}}_n$ 为模型(2)中系数 $\boldsymbol{\beta}_n$ 的 PGEEE 估计. 当(A1)–(A7)满足, 若 $\lambda_n \rightarrow 0$, 且

当 $n \rightarrow \infty$ 时, $\frac{\sqrt{\frac{p_n}{n}}}{\lambda_n} \rightarrow 0$, 那么存在 $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{n(1)}^\top, \hat{\boldsymbol{\beta}}_{n(2)}^\top)^\top$ 满足

(i) (稀疏性) $\hat{\boldsymbol{\beta}}_{n(2)} = \mathbf{0}$;

(ii) (渐近正态性) $\sqrt{n}(\boldsymbol{\Phi}_\tau + \mathbf{P}_2)(\hat{\boldsymbol{\beta}}_{n(1)} - \boldsymbol{\beta}_{n0(1)} + (\boldsymbol{\Phi}_\tau + \mathbf{P}_2)^{-1}\mathbf{P}_1) \rightarrow N(\mathbf{0}, \mathbf{D}_\tau)$.

注 定理 1 表明所提出的方法可以选出正确的模型, 同时实现对重要变量回归系数的参数估计, 称为 Oracle 性质^[2].

定理 1 的证明:

(i) 令 $\alpha_n = \sqrt{\frac{p_n}{n}}$. 第一步证明相合性, 即 $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n0}\| = O_p\left(\sqrt{\frac{p_n}{n}}\right)$. 只需证明对任意的 $\epsilon > 0$, 都存在一个大的常数 D , 使得

$$P(\sup_{\boldsymbol{\beta}_n = \boldsymbol{\beta}_{n0} + \alpha_n \mathbf{u}, \|\mathbf{u}\| = D} (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{Q}(\boldsymbol{\beta}_n) < 0) > 1 - \epsilon \quad (12)$$

成立即可. 根据表达式, 有

$$(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{Q}(\boldsymbol{\beta}_n) = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{S}(\boldsymbol{\beta}_n) - n(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{P}'_{\lambda_n}(|\boldsymbol{\beta}_n|)\text{Sign}(\boldsymbol{\beta}_n) = I + II \quad (13)$$

计算 I 的阶. 令 $\bar{\mathbf{S}}(\boldsymbol{\beta}_n) = \sum_{i=1}^n \mathbf{X}_i^\top \bar{\mathbf{V}}_{i\tau}^{-1} \Psi_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n)(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n)$, 有

$$(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{S}(\boldsymbol{\beta}_n) = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \mathbf{S}(\boldsymbol{\beta}_{n0}) + (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top [\mathbf{S}(\boldsymbol{\beta}_n) - \mathbf{S}(\boldsymbol{\beta}_{n0})] = I_1 + I_2 \quad (14)$$

其中

$$I_1 = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top \bar{\mathbf{S}}(\boldsymbol{\beta}_{n0}) + (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^\top [\mathbf{S}(\boldsymbol{\beta}_{n0}) - \bar{\mathbf{S}}(\boldsymbol{\beta}_{n0})] = I_{11} + I_{12} \quad (15)$$

考虑 I_{11} , 有

$$E\|\bar{\mathbf{S}}(\boldsymbol{\beta}_{n0})\|^2 =$$

$$\begin{aligned}
& \mathbb{E} \left[\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})^T \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \bar{\mathbf{V}}_{ir}^{-1} \mathbf{X}_i \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \right] \leq \\
& \sum_{i=1}^n \lambda_{\max}(\mathbf{X}_i \mathbf{X}_i^T) \lambda_{\max}(\bar{\mathbf{V}}_{ir}^{-2}) \mathbb{E} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})^T \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \leq \\
& C \cdot \text{tr} \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) = C \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_{ij}^T \mathbf{X}_{ij} = O_p(np_n) \quad (16)
\end{aligned}$$

由 Cauchy-Schwarz 不等式知, $|I_{11}| \leq \|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \|\bar{\mathbf{S}}(\boldsymbol{\beta}_{n0})\| = O_p \left(\sqrt{\frac{p_n}{n}} \cdot \sqrt{np_n} \right) \|u\| =$

$\|u\| O_p(p_n)$. 令 $\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1} = (q_{its})_{1 \leq t, s \leq m_i}$, 有

$$\begin{aligned}
& \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) = \\
& \sum_{i=1}^n \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} \mathbf{X}_{it} q_{its} \psi_\tau(\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) (\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) \leq \\
& O_p \left(\sqrt{\frac{p_n}{n}} \right) m \sum_{i=1}^n \mathbf{X}_{it} \psi_\tau(\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) (\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) \quad (17)
\end{aligned}$$

且

$$\begin{aligned}
& \mathbb{E} \left\| \sum_{i=1}^n \mathbf{X}_{it} \psi_\tau(\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) (\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) \right\|^2 = \sum_{i=1}^n \mathbb{E} [\psi_\tau^2(\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0}) (\mathbf{y}_{is} - \mathbf{X}_{is}^T \boldsymbol{\beta}_{n0})^2] \mathbf{X}_{it}^T \mathbf{X}_{it} \leq \\
& C \sum_{i=1}^n \mathbf{X}_{it}^T \mathbf{X}_{it} = O_p(np_n) \quad (18)
\end{aligned}$$

由(18)式知, $|I_{12}| = O_p \left(\sqrt{\frac{p_n}{n}} \cdot \sqrt{\frac{p_n}{n}} \cdot \sqrt{np_n} \right) \|u\| = o_p(p_n) \|u\|$, 因此

$$|I_1| = O_p(p_n) \|u\| \quad (19)$$

将 I_2 分为两部分计算, 有

$$\begin{aligned}
I_2 &= (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \left\{ \sum_{i=1}^n \mathbf{X}_i^T \hat{\mathbf{V}}_{ir}^{-1} [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} = \\
& (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \left\{ \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} + \\
& (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \left\{ \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} = \\
& I_{21} + I_{22} \quad (20)
\end{aligned}$$

记

$$\begin{aligned}
I_{21} &= (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \left\{ \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} + \\
& (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \left\{ \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} = \\
& I_{21}^a + I_{21}^b \quad (21)
\end{aligned}$$

其中由(A3)知

$$\begin{aligned}
I_{21}^a &= -(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i^T \boldsymbol{\beta}_n) \cdot \mathbf{X}_i (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}) \leq \\
& -\lambda_{\min}(\bar{\mathbf{V}}_{ir}^{-1}) \lambda_{\min} \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) \|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\|^2 \leq \\
& -C \cdot p_n \|u\|^2 \quad (22)
\end{aligned}$$

$$I_{21}^b = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})](\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \leqslant$$

$$C(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \quad (23)$$

又因为

$$\mathbb{E} \left\| \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \right\|^2 = \mathbb{E} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})^T \bar{\mathbf{V}}_{ir}^{-1} \mathbf{X}_i \mathbf{X}_i^T \bar{\mathbf{V}}_{ir}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \leqslant$$

$$\sum_{i=1}^n \lambda_{\max}(\bar{\mathbf{V}}_{ir}^{-2}) \lambda_{\max}(\mathbf{X}_i \mathbf{X}_i^T) \mathbb{E} \left\| \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0} \right\|^2 \leqslant \quad (24)$$

$$C \sum_{i=1}^n \text{tr}(\mathbf{X}_i \mathbf{X}_i^T) = O_p(np_n)$$

由(24)式得, $|I_{21}^b| = O_p(\sqrt{np_n} \cdot \alpha_n) \|u\| = O_p(p_n) \|u\|$, 故 $|I_{21}| = -O_p(p_n) \|u\|^2 + O_p(p_n) \|u\|$. 类似地, 计算 I_{22} .

$$I_{22} = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \cdot \left\{ \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n)(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} +$$

$$(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \cdot \left\{ \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n)(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})] \right\} =$$

$$I_{22}^a + I_{22}^b \quad (25)$$

其中

$$I_{22}^a = -(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) \mathbf{X}_i (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}) \leqslant$$

$$-\lambda_{\min}(\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) \lambda_{\min} \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) \left\| \boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0} \right\|^2 \leqslant -C \cdot o_p(p_n) \|u\|^2 \quad (26)$$

$$I_{22}^b = (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) [\boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_n) - \boldsymbol{\Psi}_\tau(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})](\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \leqslant$$

$$C(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})^T \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \quad (27)$$

又

$$\mathbb{E} \left\| \sum_{i=1}^n \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \right\|^2 =$$

$$\mathbb{E} \left[\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0})^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) \mathbf{X}_i \mathbf{X}_i^T (\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0}) \right] \leqslant \quad (28)$$

$$\sum_{i=1}^n \lambda_{\max}(\hat{\mathbf{V}}_{ir}^{-1} - \bar{\mathbf{V}}_{ir}^{-1}) \lambda_{\max}(\mathbf{X}_i \mathbf{X}_i^T) \mathbb{E} \left\| \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{n0} \right\|^2 \leqslant$$

$$O_p \left(\sqrt{\frac{p_n}{n}} \right) \text{tr} \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) = O_p(p_n)$$

由(28)式知, $I_{22}^b = O_p(\sqrt{p_n} \cdot \alpha_n) \|u\| = o_p(p_n) \|u\|$, 则 $I_{22} = -o_p(p_n) \|u\|^2 + o_p(p_n) \|u\|$. 因此, 有

$$I_2 = -O_p(p_n) \|u\|^2 \quad (29)$$

由(19), (29)式可得, (14)式的值由(29)式控制, 小于0. 易知(13)式中的第二项以 $n \alpha_n^2 \|u\| + n b_n \alpha_n^2 \|u\|^2$ 为界, 因此可以找到一个足够大的 D , 使得(13)式的值完全由(29)式决定. (12)式得证.

接着证明稀疏性 $\hat{\boldsymbol{\beta}}_{n(2)} = \mathbf{0}$. 即对于任意 $\left\| \boldsymbol{\beta}_{n(1)} - \boldsymbol{\beta}_{n0(1)} \right\| = O_p \left(\sqrt{\frac{p_n}{n}} \right)$, 存在 $\epsilon_n = C \sqrt{\frac{p_n}{n}}$, $j = s+1, \dots$,

p_n , 使得

$$\begin{aligned} Q(\beta_j) &> 0, -\epsilon_n < \beta_j < 0 \\ Q(\beta_j) &< 0, 0 < \beta_j < \epsilon_n \end{aligned} \quad (30)$$

由相合性证明过程知 $\|\mathbf{S}(\boldsymbol{\beta}_n)\| = O_p(\sqrt{np_n})$, 因此

$$\begin{aligned} Q(\beta_j) &= S(\beta_j) - np'_{\lambda_n}(|\beta_j|) \text{sign}(\beta_j) = \\ &= O_p(\sqrt{np_n}) - np'_{\lambda_n}(|\beta_j|) \text{sign}(\beta_j) = \\ &= n\lambda_n \left[-p'_{\lambda_n}(|\beta_j|) \cdot \frac{\text{sign}(\beta_j)}{\lambda_n} + O_p\left(\frac{1}{\lambda_n} \sqrt{\frac{p_n}{n}}\right) \right] \end{aligned} \quad (31)$$

由(A7)可知, (31)式的符号完全由 β_j 的符号决定. (30)式得证.

(ii) 由相合性证明过程知, $Q((\boldsymbol{\beta}_{n(1)}^T, \mathbf{0}^T)^T)$ 存在 $\sqrt{\frac{n}{p_n}}$ 相合的局部最小值. 令 $\hat{\boldsymbol{\beta}}_{n(1)}$ 为局部最小值, 则其满足如下方程

$$0 = Q((\hat{\boldsymbol{\beta}}_{n(1)}^T, \mathbf{0}^T)^T)$$

即

$$\begin{aligned} 0 &= Q(\hat{\boldsymbol{\beta}}_{n(1)}) = S(\hat{\boldsymbol{\beta}}_{n(1)}) - n \mathbf{P}'_{\lambda_n}(|\hat{\boldsymbol{\beta}}_{n(1)}|) \text{Sign}(\hat{\boldsymbol{\beta}}_{n(1)}) = \\ &= S(\boldsymbol{\beta}_{n0(1)}) + (E[S(\hat{\boldsymbol{\beta}}_{n(1)})] - E[S(\boldsymbol{\beta}_{n0(1)})]) + (S(\hat{\boldsymbol{\beta}}_{n(1)}) - E[S(\hat{\boldsymbol{\beta}}_{n(1)})]) - \\ &= (S(\boldsymbol{\beta}_{n0(1)}) - E[S(\boldsymbol{\beta}_{n0(1)})]) - n \mathbf{P}'_{\lambda_n}(|\hat{\boldsymbol{\beta}}_{n(1)}|) \text{Sign}(\hat{\boldsymbol{\beta}}_{n(1)}) = \\ &= S(\boldsymbol{\beta}_{n0(1)}) + (E[S(\boldsymbol{\beta}_{n0(1)})])' (\hat{\boldsymbol{\beta}}_{n(1)} - \boldsymbol{\beta}_{n0(1)}) - n \mathbf{P}'_{\lambda_n}(|\hat{\boldsymbol{\beta}}_{n(1)}|) \text{Sign}(\hat{\boldsymbol{\beta}}_{n(1)}) + o_p(1) = \\ &= \bar{S}(\boldsymbol{\beta}_{n0(1)}) + (E[\bar{S}(\boldsymbol{\beta}_{n0(1)})])' (\hat{\boldsymbol{\beta}}_{n(1)} - \boldsymbol{\beta}_{n0(1)}) - n \mathbf{P}'_{\lambda_n}(|\hat{\boldsymbol{\beta}}_{n(1)}|) \text{Sign}(\hat{\boldsymbol{\beta}}_{n(1)}) + o_p(1) = \\ &= \bar{S}(\boldsymbol{\beta}_{n0(1)}) + I - II \end{aligned} \quad (32)$$

根据文献[13]可知, $-\frac{1}{n}(E[\bar{S}(\boldsymbol{\beta}_{n0(1)})])' \xrightarrow{p} \boldsymbol{\Phi}_\tau$, $\frac{1}{\sqrt{n}}\bar{S}(\boldsymbol{\beta}_{n0(1)}) \rightarrow N(0, \mathbf{D}_\tau)$. 由条件(A4)可知, $II =$

$n \mathbf{P}_1 + n \mathbf{P}_2(\hat{\boldsymbol{\beta}}_{n(1)} - \boldsymbol{\beta}_{n0(1)})$. 由 Slutsky 定理可得

$$\sqrt{n}(\boldsymbol{\Phi}_\tau + \mathbf{P}_2)(\hat{\boldsymbol{\beta}}_{n(1)} - \boldsymbol{\beta}_{n0(1)}) + (\boldsymbol{\Phi}_\tau + \mathbf{P}_2)^{-1} \mathbf{P}_1 = \frac{1}{\sqrt{n}}\bar{S}(\boldsymbol{\beta}_{n0(1)}) + o_p(1) \rightarrow N(\mathbf{0}, \mathbf{D}_\tau) \quad (33)$$

定理证毕.

3 模拟研究

为了研究所提方法的有限样本性质, 本文比较了不同的惩罚方法及相关结构下所提出方法的效果. 数据来源于以下模型

$$y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta}_n + (1.5 X_{ij}^k) \epsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m_i, k = 2, 9 \quad (34)$$

对于协变量的生成分为两步: 第一步, 生成 $\mathbf{Z}_{ij} = (Z_{ij}^1, \dots, Z_{ij}^{p_n})^T$, 来源于多元正态分布 $N(\mathbf{0}, \mathbf{R}_{p_n})$, $(\mathbf{R}_{p_n})_{a,b} = 0.5^{|a-b|}$, $1 \leq a, b \leq p_n$. 第二步, 令 $X_{ij}^l = Z_{ij}^l$, $l \neq k$, $X_{ij}^k = F(Z_{ij}^k)$, 生成 p_n 维向量 $\mathbf{X}_{ij}^T = (X_{ij}^1, \dots, X_{ij}^{p_n})$. $F(\cdot)$ 表示标准正态分布的分布函数, 主要用于刻画数据中的异方差. $\boldsymbol{\epsilon}_i^T = (\epsilon_{i1}, \dots, \epsilon_{im_i})$ 服从多元正态分布 $N(\mathbf{0}, \mathbf{R}_i)$. 分别考虑如下三种情形:

情形 1 $p_n = 10$, $k = 9$, $m_i = 4$, $n = 50, 100, 200$, $\boldsymbol{\beta}_n = (-3, 5, 0, 0, 4, 0, 0, 2, 0, 0)^T$. $\mathbf{R}_i$ 是参数为 0.9 的等相关结构矩阵.

情形 2 $k = 2$, m_i 服从参数为 (3, 6) 的均匀分布, $\mathbf{R}_i$ 是参数为 0.9 的 AR(1) 结构矩阵. 其余设置和情形 1 一样.

情形 3 $p_n = 30$, $n = 100, 200$. $\boldsymbol{\beta}_n = (-3, 5, 0, 0, 4, 0, 0, 2, 0, 0, \dots, 0)^T$. 其余设置和情形 1 一样.

注 在情形 1 和情形 3 中, $\tau = 0.5$ 时, 真实模型中的自变量为 $X_{ij}^1, X_{ij}^2, X_{ij}^5, X_{ij}^8$; $\tau = 0.9$ 时, 真实模

型中的自变量为 $X_{ij}^1, X_{ij}^2, X_{ij}^5, X_{ij}^8, X_{ij}^9$. 情形 1 与情形 3 的设置可以比较在不同的 τ 下 PGEEE 选择的模型, 同时观察协变量维数增加时的模型估计效果. 情形 2 中, 无论 τ 取何值, 真实模型中的自变量均为 $X_{ij}^1, X_{ij}^2, X_{ij}^5, X_{ij}^8$, 但部分回归系数与 τ 有关. $\tau = 0.5$ 时, $\beta_2 = 5$; $\tau = 0.9$ 时, $\beta_2 = 6.29$. 情形 2 的设置可以比较 PGEEE 在不同 τ 水平下的估计精度.

重复模拟 100 次, 表 1—3 给出了 3 种情形下 MCP 和 SCAD 两种惩罚方法的 PGEEE 估计结果. 使用指标 Size(平均选取变量个数), FN(重要变量被误认为噪音变量的平均个数), FP(噪音变量被选择的平均个数), Prob(X_{ij}^9 被选出的概率), MSE(估计量的均方误差), MAE(回归系数 β_2 估计量的平均绝对误差) 来衡量估计量的优劣. 模拟结果显示:

(i) SCAD 和 MCP 两种惩罚方法并无明显的优劣之分. FN 均为 0, 表示所有重要的变量都被识别, FP 接近 0, 表明噪音变量被选择的可能性很小;

(ii) 在情形 1 和情形 3 中, $\tau = 0.9$ 时, Prob 等于 1, 而 $\tau = 0.5$ 时, Prob 的值接近 0. 这表明所提出的估计量 PGEEE 可以在不同的 τ 水平下, 有效识别出正确的模型, 刻画数据中的异方差结构;

(iii) 在不同的 τ 水平下, 即使选择的变量相同, 参数估计值也可能不同(情形 2). 在此情形下, 估计量的 MSE 和 MAE 随着样本量增大而减小, 表示该方法可以在识别出异方差的同时实现回归参数的一致估计;

(iv) 对比情形 1 和情形 3, 协变量维数 p_n 从 10 增加至 30, 结果显示模型中噪音变量数量增加时, PGEEE 估计表现依然较好, 且估计量 MSE 减小, 表明该方法可以用于分析高维数据, 排除无关变量, 识别出重要变量.

(v) 考虑相关结构时估计量的表现总体上优于独立(IND)的情形. 即使相关结构被误判后, 参数估计效果依然很好, 尤其使用 UN 结构时.

4 实证分析

数据来自 1976 年至 1982 年间对美国经济收入动态的面板研究, 包含了连续 7 年 595 名民众的工资水平, 属于平衡数据, 更多详细信息参考文献[15]. 该研究中, 协变量包括工作经历 E , 工作时间 W , 工作职业 O (蓝领取 1, 否则 0), 工作行业 I (制造业取 1, 否则 0), 居住地 S (居住在南部取 1, 否则 0), 种族 B (黑人取 1, 否则 0), 是否住在都市统计区 A (如果是取 1, 否则 0), 是否结婚 M (结婚取 1, 否则 0), 性别 F (女性取 1, 否则 0), 劳动保障 U (签合同取 1, 否则 0) 及受教育程度 D , 响应变量为对数变换后的工资水平.

表 4 给出了 $\tau = 0.01, 0.5, 0.95$ 下参数的 PGEEE 估计, 其中 $\tau = 0.5$ 对应经典的均值回归估计. 分析结果可知, 不同的惩罚方法和不同的相关结构选出的变量基本一致. 可以看到, 在 3 个水平下均被选择的变量有 O, B, F, D ; 均未被选择的变量有 W . 截距项, B, F 的系数估计随着 τ 不同而变化, 图 1a, b 为不同种族及性别对应的工资随时间变化的箱线图. 男性的工资明显高于女性, 白人的工资明显高于黑人. 在 $\tau = 0.01$ 时, E 被认为是噪音变量, 而在 $\tau = 0.5$ 和 0.95 时被认为是重要变量. 在 $\tau = 0.95$ 时, 除了独立结构下 MCP 估计外, 工作行业 I , 居住地 S , 是否结婚 M , 劳动保障 U 均被剔除在模型外; 而在 $\tau = 0.01$ 和 0.5 时则被认为是重要变量. 图 1c, d, e, f 为这些变量对应的工资分布箱线图. 以变量 S 为例, 可以看到, 在低分位点时, 居住在北部的工资要明显高于南部, 但是在高分位点时, 两者的区别并不明显, 这与 PGEEE 的估计结果相吻合. 由此可见, 该方法比采用普通最小二乘估计($\tau = 0.5$) 挖掘出了更多的信息.

5 结语

本文基于 expectile 提出了高维纵向数据的 PGEEE 估计量, 在实现模型变量选择的同时, 对模型的回归系数进行估计. 在正则条件下本文建立了 PGEEE 估计量的 Oracle 性质. 数值模拟结果显示, MCP 与 SCAD 惩罚及不同的协方差结构在变量选择方面并无明显差异. 相较于独立结构, 考虑相关结构时回归系数的估计效率更高. 多数情况下, 不确定结构(UN) 的 PGEEE 估计量具有较好的估计精度. 最后建立工资数据的 PGEEE 模型, 可以看到在不同的 τ 水平下, 影响工资的因素有所区别, 同一个因素影响程度也可能不同. 这表明 PGEEE 可以有效识别数据中的异质结构, 比经典的惩罚估计方程估计(PGEE) 挖掘出更丰富的信息, 更合理地分析了工资的影响因素.

表 1 情形 1 模拟结果

<i>n</i>		$\tau=0.5$					$\tau=0.9$				
		<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>Prob</i>	<i>MSE</i>	<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>Prob</i>	<i>MSE</i>
50	AR(1). mcp	4.13	0	0.13	0.05	0.019 0	5.53	0	0.53	1	0.104 3
	CS. mcp	4.13	0	0.13	0.07	0.016 9	5.23	0	0.23	1	0.096 4
	IND. mcp	4.29	0	0.29	0.15	0.034 7	5.18	0	0.18	1	0.099 1
	UN. mcp	4.27	0	0.27	0.14	0.027 8	5.08	0	0.08	1	0.086 7
	AR(1). scad	4.17	0	0.17	0.04	0.019 6	5.91	0	0.91	1	0.103 9
	CS. scad	4.15	0	0.15	0.05	0.015 3	5.5	0	0.5	1	0.104 5
	IND. scad	4.25	0	0.25	0.12	0.028 6	5.17	0	0.17	1	0.098 1
	UN. scad	4.2	0	0.2	0.1	0.022 9	5.07	0	0.07	1	0.086 0
100	AR(1). mcp	4.05	0	0.05	0.04	0.009 7	5.11	0	0.11	1	0.046 7
	CS. mcp	4.07	0	0.07	0.04	0.009 4	5.05	0	0.05	1	0.046 0
	IND. mcp	4.28	0	0.28	0.16	0.020 5	5.14	0	0.14	1	0.052 7
	UN. mcp	4.16	0	0.16	0.11	0.012 6	5.01	0	0.01	1	0.044 7
	AR(1). scad	4.04	0	0.04	0.03	0.009 2	5.18	0	0.18	1	0.047 7
	CS. scad	4.09	0	0.09	0.04	0.010 2	5.08	0	0.08	1	0.045 8
	IND. scad	4.21	0	0.21	0.13	0.017 6	5.13	0	0.13	1	0.052 3
	UN. scad	4.09	0	0.09	0.05	0.009 5	5.01	0	0.01	1	0.044 6
200	AR(1). mcp	4.1	0	0.1	0.05	0.005 2	5.01	0	0.01	1	0.020 5
	CS. mcp	4.13	0	0.13	0.08	0.005 9	5	0	0	1	0.019 4
	IND. mcp	4.14	0	0.14	0.11	0.008 6	5.06	0	0.06	1	0.022 4
	UN. mcp	4.07	0	0.07	0.06	0.004 7	5	0	0	1	0.019 6
	AR(1). scad	4.12	0	0.12	0.05	0.004 6	5	0	0	1	0.020 6
	CS. scad	4.12	0	0.12	0.06	0.005 3	5	0	0	1	0.020 1
	IND. scad	4.11	0	0.11	0.08	0.007 0	5.05	0	0.05	1	0.022 2
	UN. scad	4.06	0	0.06	0.05	0.003 6	5	0	0	1	0.019 6

表 2 情形 2 模拟结果

<i>n</i>		$\tau=0.5$					$\tau=0.9$				
		<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>MAE</i>	<i>MSE</i>	<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>MAE</i>	<i>MSE</i>
50	AR(1). mcp	4.02	0	0.02	0.169 5	0.054 9	4.05	0	0.05	0.189 5	0.070 5
	CS. mcp	4.01	0	0.01	0.162 6	0.051 9	4.08	0	0.08	0.197 2	0.071 0
	IND. mcp	4.05	0	0.05	0.166 6	0.061 3	4.19	0	0.19	0.202 1	0.090 1
	UN. mcp	4.01	0	0.01	0.162 4	0.055 8	4.14	0	0.14	0.201 0	0.081 4
	AR(1). scad	4.1	0	0.1	0.171 7	0.056 7	4.05	0	0.05	0.191 1	0.071 9
	CS. scad	4.04	0	0.04	0.162 4	0.052 1	4.04	0	0.04	0.196 6	0.070 1
	IND. scad	4.04	0	0.04	0.166 8	0.061 0	4.1	0	0.1	0.203 2	0.087 1
	UN. scad	4.01	0	0.01	0.162 4	0.055 7	4.11	0	0.11	0.200 9	0.080 2
100	AR(1). mcp	4	0	0	0.136 4	0.031 2	4.04	0	0.04	0.177 9	0.053 4
	CS. mcp	4	0	0	0.132 3	0.030 3	4.05	0	0.05	0.165 1	0.049 8
	IND. mcp	4.01	0	0.01	0.126 9	0.029 8	4.09	0	0.09	0.150 2	0.045 0
	UN. mcp	4	0	0	0.126 7	0.028 2	4.05	0	0.05	0.148 8	0.043 1
	AR(1). scad	4.01	0	0.01	0.136 7	0.031 5	4.04	0	0.04	0.177 8	0.053 4
	CS. scad	4.01	0	0.01	0.132 2	0.030 2	4.03	0	0.03	0.164 5	0.049 5
	IND. scad	4	0	0	0.126 8	0.029 7	4.08	0	0.08	0.150 1	0.044 8
	UN. scad	4	0	0	0.126 7	0.028 2	4.03	0	0.03	0.148 8	0.043 0
200	AR(1). mcp	4	0	0	0.087 8	0.014 2	4.01	0	0.01	0.099 2	0.018 0
	CS. mcp	4	0	0	0.083 4	0.013 5	4.02	0	0.02	0.102 5	0.019 6
	IND. mcp	4.02	0	0.02	0.081 5	0.013 7	4.13	0	0.13	0.099 2	0.020 4
	UN. mcp	4.02	0	0.02	0.082 1	0.013 0	4.11	0	0.11	0.096 5	0.018 8
	AR(1). scad	4.01	0	0.01	0.088 0	0.014 3	4.01	0	0.01	0.099 3	0.018 1
	CS. scad	4	0	0	0.0836	0.013 5	4.02	0	0.02	0.102 6	0.019 6
	IND. scad	4.02	0	0.02	0.081 5	0.013 6	4.09	0	0.09	0.099 5	0.019 9
	UN. scad	4.02	0	0.02	0.082 1	0.013 0	4.05	0	0.05	0.096 7	0.018 5

表 3 情形 3 模拟结果

<i>n</i>		$\tau=0.5$					$\tau=0.9$				
		<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>Porb</i>	<i>MSE</i>	<i>Size</i>	<i>FN</i>	<i>FP</i>	<i>Porb</i>	<i>MSE</i>
100	AR(1). mcp	4.03	0	0.03	0	0.006 0	6.11	0	1.11	1	0.054 0
	CS. mcp	4.03	0	0.03	0.01	0.005 5	5.27	0	0.27	1	0.045 1
	IND. mcp	4.21	0	0.21	0.06	0.013 0	5.17	0	0.17	1	0.046 7
	UN. mcp	4.13	0	0.13	0.04	0.008 7	5.03	0	0.03	1	0.041 5
	AR(1). scad	4.05	0	0.05	0	0.006 4	7.1	0	2.1	1	0.058 2
	CS. scad	4.03	0	0.03	0	0.005 2	5.79	0	0.79	1	0.045 5
	IND. scad	4.2	0	0.2	0.05	0.011 8	5.11	0	0.11	1	0.046 3
	UN. scad	4.09	0	0.09	0.03	0.007 1	5.03	0	0.03	1	0.041 3
200	AR(1). mcp	4	0	0	0	0.002 8	5.05	0	0.05	1	0.019 9
	CS. mcp	4	0	0	0	0.002 3	5.02	0	0.02	1	0.018 6
	IND. mcp	4.07	0	0.07	0.02	0.004 7	5.24	0	0.24	1	0.019 6
	UN. mcp	4.03	0	0.03	0.01	0.003 0	5.02	0	0.02	1	0.016 8
	AR(1). scad	4	0	0	0	0.002 8	5.22	0	0.22	1	0.021 2
	CS. scad	4	0	0	0	0.002 2	5.04	0	0.04	1	0.018 8
	IND. scad	4.07	0	0.07	0.02	0.004 7	5.13	0	0.13	1	0.019 5
	UN. scad	4	0	0	0	0.002 9	5.02	0	0.02	1	0.017 1

表 4 工资数据参数估计结果

		β_0	<i>E</i>	<i>W</i>	<i>O</i>	<i>I</i>	<i>S</i>	<i>B</i>	<i>A</i>	<i>M</i>	<i>F</i>	<i>U</i>	<i>D</i>
0.01	AR(1). mcp	5.050 3	0	0	−0.145 0	0.133 8	−0.068 9	−0.103 6	0.119 1	0.096 6	−0.296 1	0.179 0	0.050 3
	CS. mcp	4.942 7	0.011 6	0	−0.121 1	0.089 4	−0.039 8	−0.184 4	0.066 5	0	−0.365 6	0.233 1	0.060 3
	IND. mcp	5.038 1	0	0	−0.174 2	0.145 8	−0.078 4	−0.105 8	0.113 6	0.114 8	−0.277 9	0.188 1	0.048 8
	UN. mcp	5.042 2	0	0	−0.167 2	0.142 7	−0.079 8	−0.117 8	0.109 4	0.110 5	−0.280 4	0.188 4	0.050 6
	AR(1). scad	5.057 1	0	0	−0.144 8	0.1341	−0.068 6	−0.103 0	0.119 3	0.097 0	−0.296 5	0.179 2	0.050 3
	CS. scad	4.942 1	0.0116	0	−0.121 1	0.089 4	−0.039 9	−0.184 2	0.066 5	0	−0.365 5	0.233 0	0.060 3
	IND. scad	5.038 6	0	0	−0.174 2	0.145 8	−0.078 4	−0.105 8	0.113 6	0.114 8	−0.277 9	0.188 1	0.048 8
	UN. scad	5.042 1	0	0	−0.167 2	0.142 7	−0.079 8	−0.117 7	0.109 4	0.110 6	−0.280 3	0.188 4	0.050 5
0.5	AR(1). mcp	5.360 1	0.021 1	0	−0.049 6	0.016 8	−0.056 4	−0.175 7	0.036 8	−0.050 7	−0.401 3	0.045 2	0.075 2
	CS. mcp	3.253 5	0.086 0	0	−0.028 8	0.013 4	0	−0.264 6	−0.049 7	−0.042 4	−0.173 9	0.040 7	0.136 9
	IND. mcp	5.445 4	0.010 4	0	−0.148 6	0.053 1	−0.053 2	−0.167 1	0.145 4	0.066 2	−0.353 4	0.101 9	0.057 1
	UN. mcp	5.464 6	0.011 4	0	−0.135 6	0.048 1	−0.053 3	−0.168 6	0.138 1	0.049 0	−0.363 8	0.096 3	0.058 7
	AR(1). scad	5.334 2	0.021 2	0	−0.046 3	0.010 4	−0.058 3	−0.167 8	0.019 3	−0.013 8	−0.369 2	0.039 0	0.075 7
	CS. scad	3.251 8	0.086 1	0	−0.028 0	0.012 5	0	−0.264 7	−0.049 0	−0.041 0	−0.173 0	0.039 8	0.136 9
	IND. scad	5.445 6	0.010 4	0	−0.148 6	0.052 4	−0.052 4	−0.167 3	0.145 5	0.066 2	−0.353 5	0.102 1	0.057 1
	UN. scad	5.496 0	0.011 6	0	−0.135 3	0.045 7	−0.051 3	−0.173 5	0.137 0	0	−0.398 9	0.098 1	0.058 9
0.95	AR(1). mcp	6.088 4	0.014 7	0	−0.117 8	0	0	−0.171 4	0.095 7	0	−0.479 9	0	0.065 4
	CS. mcp	5.634 7	0.028 1	0	−0.101 3	0	0	−0.188 4	0	0	−0.430 5	0	0.075 0
	IND. mcp	5.852 2	0.012 2	0	−0.149 0	0.036 5	0	−0.169 9	0.139 4	0.061 5	−0.441 4	0.045 3	0.064 8
	UN. mcp	6.007 2	0.013 3	0	−0.124 2	0	0	−0.168 9	0.128 3	0	−0.499 7	0	0.062 8
	AR(1). scad	6.089 8	0.014 6	0	−0.117 9	0	0	−0.171 3	0.095 8	0	−0.480 1	0	0.065 4
	CS. scad	5.561 7	0.027 4	0	0	0	0	−0.205 8	0	0	−0.412 2	0	0.080 8
	IND. scad	6.018 7	0.012 4	0	−0.134 4	0	0	−0.168 4	0.137 7	0	−0.503 9	0	0.061 1
	UN. scad	5.973 2	0.013 4	0	−0.120 0	0	0	−0.169 1	0.128 0	0	−0.498 1	0	0.063 9

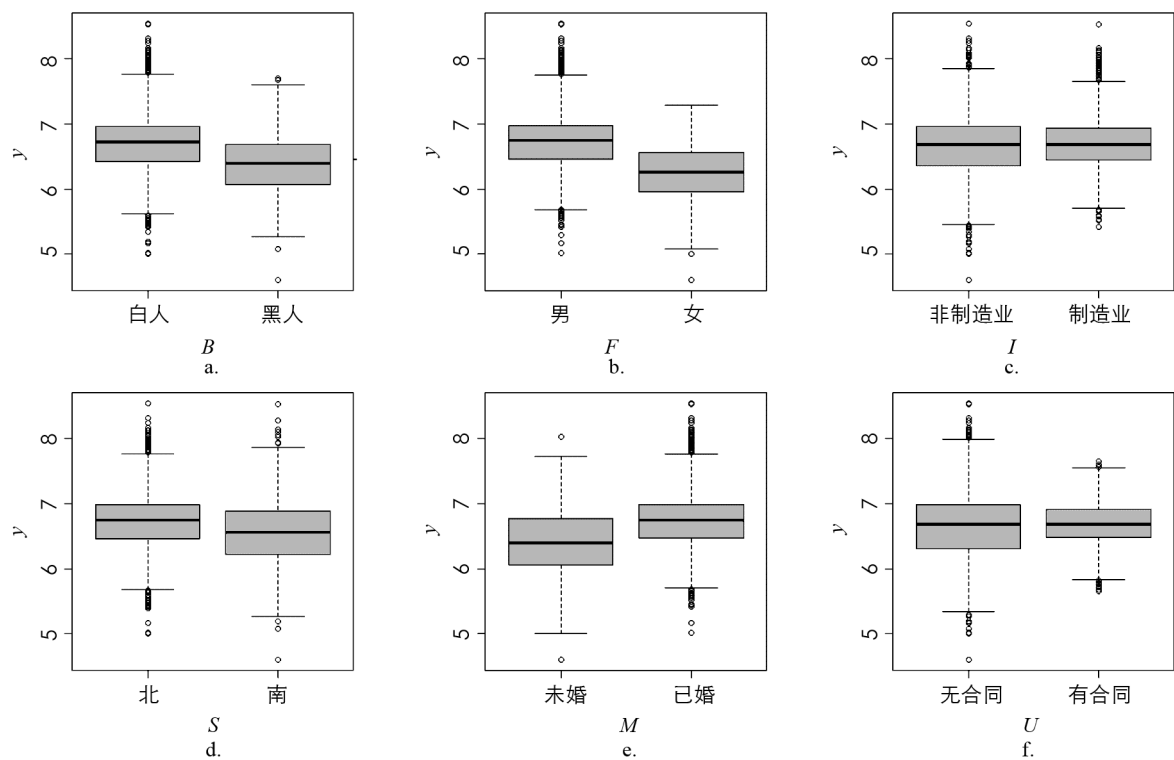


图 1 工资箱线图

参考文献：

[1] LIANG K Y, ZEGER S L. Longitudinal Data Analysis Using Generalized Linear Models [J]. Biometrika, 1986, 73(1): 13-22.

[2] FAN J Q, LI R Z. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties [J]. Journal of the American Statistical Association, 2001, 96(456): 1348-1360.

[3] ZHANG C H. Nearly Unbiased Variable Selection under Minimax Concave Penalty [J]. The Annals of Statistics, 2010, 38(2): 894-942.

[4] WANG L, ZHOU J H, QU A N. Penalized Generalized Estimating Equations for High-Dimensional Longitudinal Data Analysis [J]. Biometrics, 2012, 68(2): 353-360.

[5] DZIAK J, LI R. Penalized Generalized Estimating Equations for Variable Selection with Longitudinal Data [M]. Berlin: Springer, 2020.

[6] KOENKER R, BASSETT G. Regression Quantiles [J]. Econometrica, 1978, 46(1): 33-50.

[7] CHEN L, WEI L J, PARZEN M I. Quantile Regression for Correlated Observations [M]//Proceedings of the Second Seattle Symposium in Biostatistics. New York: Springer New York, 2004: 51-69.

[8] LU X M, FAN Z Z. Weighted Quantile Regression for Longitudinal Data [J]. Computational Statistics, 2015, 30(2): 569-592.

[9] KOENKER R. Quantile Regression for Longitudinal Data [J]. Journal of Multivariate Analysis, 2004, 91(1): 74-89.

[10] NEWEY W K, POWELL J L. Asymmetric Least Squares Estimation and Testing [J]. Econometrica, 1987, 55(4): 819-847.

[11] GU Y, ZOU H. High-Dimensional Generalizations of Asymmetric Least Squares Regression and Their Applications [J]. The Annals of Statistics, 2016, 44(6): 2661-2694.

[12] LIAO L N, PARK C, CHOI H. Penalized Expectile Regression: an Alternative to Penalized Quantile Regression [J]. Annals of the Institute of Statistical Mathematics, 2019, 71(2): 409-438.

[13] BARRY A, OUALKACHA K, CHARPENTIER A. A New GEE Method to Account for Heteroscedasticity Using Asymmetric Least-Square Regressions [J]. Journal of Applied Statistics, 2022, 49(14): 3564-3590.

[14] 周霖, 罗幼喜. 混合效应模型的双 MCP 惩罚分位回归研究 [J]. 华中师范大学学报(自然科学版), 2021, 55(6): 991-999, 1012.

[15] CORNWELL C, RUPERT P. Efficient Estimation with Panel Data: an Empirical Comparison of Instrumental Variables Estimators [J]. Journal of Applied Econometrics, 1988, 3(2): 149-155.