中国科学: 数学　2024 年　第 54 卷　第 4 期: 617～646
SCIENTIA SINICA　Mathematica
《中国科学》杂志社
SCIENCE CHINA PRESS
论　文

高维空间相依数据的 Expectile 回归分析

刘宣[1,2], 马海强[2], 盛志雁[2], 罗良清[2*]

1. 南昌师范学院数学与信息科学学院, 南昌 330032;
2. 江西财经大学统计与数据科学学院, 南昌 330013
E-mail: liuxuanyg@163.com, mhqtc118@163.com, shengzhiyan000303@163.com, llq6429@163.com

收稿日期: 2023-02-10; 接受日期: 2023-06-22; 网络出版日期: 2023-12-08; * 通信作者
国家自然科学基金 (批准号: 12161042)、中国博士后面上项目 (批准号: 2019M662262)、国家社科基金重大项目 (批准号: 21&ZD150)、江西省博士后特别资助项目 (批准号: 2021KY18) 和江西省教育厅科技项目 (批准号: GJJ200522 和 GJJ202603) 资助项目

摘要　空间数据的异质性、空间权重的内生性和解释变量的高维特征会给空间相依数据分析带来重大挑战. 本文基于 Expectile 回归的稳健估计优势和惩罚压缩的有效降维能力, 分别在外生空间和内生空间权重矩阵条件下, 给出高维空间滞后模型未知参数的两步与三步惩罚 Expectile 估计, 并在常规正则条件下证明所提出估计的相合性和变量选择的 Oracle 性质. 数值模拟显示, 两步估计法能有效处理外生空间权重矩阵条件下的稳健统计问题, 同时三步估计法在外生空间和内生空间权重条件下均有优良表现. 最后, 通过分析我国市域空气质量与经济发展的关系, 进一步验证所提出方法的有效性.

关键词　空间滞后模型　Expectile 回归　变量选择　内生性　高维数据

MSC (2020) 主题分类　62F35, 62J99, 91B72

1　引言

Expectile 回归是由 Newey 和 Powell [23] 提出的一种稳健统计分析方法, 主要利用非对称二次函数 (asymmetric least squares, ALS) 替代分位数回归中的检验函数作为新的损失函数, 对线性模型进行分析. 与分位数回归相比, Expectile 回归具有如下优势: 其一, 损失函数可微, 可极大优化大样本理论性质的推导并克服分位数目标函数不可微带来的困难; 其二, 损失函数的可微性有利于未知参数的直接估计, 所得估计量的渐近协方差矩阵将不涉及误差的密度函数估计; 最后, 由于 Expectile 回归与分位数回归在分位点上存在一一对应关系, 从而使它不仅在处理异质数据或厚尾分布数据时具有稳健性, 而且能够全面刻画响应变量的条件分布 (参见文献 [14, 33]).

由于具有以上优势, 因而 Expectile 回归发展迅速, 并引起了学者们的广泛关注. 在截面数据分析方面, Taylor [28]、Xie 等 [32]、Hu 等 [12] 和 Girard 等 [10] 通过构建条件自回归 Expectile 模型来评估资产组合中的在险价值 (value at risk, VaR) 和预期不足 (expected shortfall, ES). 在面板数据分析方面,

英文引用格式: Liu X, Ma H Q, Sheng Z Y, et al. Expectile regression analysis of high-dimensional spatially dependent data (in Chinese). Sci Sin Math, 2024, 54: 617–646, doi: 10.1360/SSM-2023-0031

Barry 等 [2] 将广义矩估计方法用于分析 Expectile 回归, 随后 Barry 等 [3] 研究了固定效应 Expectile 回归模型. 在函数型数据分析方面, Mohammedi 等 [22] 在预测变量为函数、因变量为标量的模型中构建了 Expectile 回归的核估计量并证明了相应的渐近性质; Xiao 等 [31] 提出了部分函数线性 Expectile 回归模型并对其进行统计推断, 其他研究还可参见文献 [9].

伴随信息技术的快速发展, 可获取的数据量和特征变量越来越多, 如何在众多变量中选出重要变量构建统计模型显得尤为重要. Tibshirani [29] 运用惩罚的思想创造性地提出 LASSO (least absolute shrinkage and selection operator) 方法解决高维回归模型的变量选择问题, 由于 LASSO 方法存在不相合问题, 随后出现了改进的惩罚方法, 如 SCAD (smoothly clipped absolute deviation) [7] 和自适应 LASSO [35] 等. 这些方法已成功应用到了多类高维回归模型, 具体可参见文献 [1,17]. 进一步地, 由于高维数据中的异质性和厚尾分布难以避免, 因此人们发展了稳健的变量选择方法, 先后出现了自适应 LASSO 分位数回归分析、带 SCAD 惩罚的分位数回归分析等相关研究, 可参见文献 [8,27,30]. 此外, 在 Expectile 回归框架下, Gu 和 Zou [11]、Liao 等 [18] 及 Ciuperca [4] 采用自适应 LASSO 等惩罚方法研究了高维 Expectile 线性回归模型, 给出了未知参数的大样本性质、相关算法和应用案例.

空间相依数据是人们收集到的常见数据类型之一, 处理此类数据的经典模型是 Cliff 和 Ord [5,6] 提出的空间滞后模型, 该模型主要通过引入空间权重矩阵刻画变量间的空间相依关系, 进而分析空间溢出效应. 因此, 空间权重矩阵的设定在空间回归模型的构建中十分关键. 考虑到理论分析的复杂度以及实际应用的方便性, 基于地理位置或经济关系等方式的外生设定在文献中较为普遍, 同时产生了相关研究方法, 在低维情形下有拟最大似然 [16] 和分位数回归 [21] 等估计方法, 在高维情形下有惩罚拟似然 [20] 和惩罚分位数 [19] 等估计与变量选择方法. 然而, 外生性的设定往往带有较大的主观性, 且当空间权重矩阵涉及社会经济因素时, 外生性设定难以成立 (参见文献 [15]). 近年来, 内生空间权重矩阵开始出现, 主要通过增加额外变量构建某种函数关系来表达内生性, 低维空间回归模型的统计推断得到了进一步发展 (参见文献 [24,25]). 不过, 基于内生空间权重矩阵的高维空间模型的相关研究在文献中尚未可见.

综上可知, 空间相依数据的稳健分析有待进一步发展, 其主要原因表现在: (1) 虽然 Expectile 回归比分位数回归具有一定的比较优势, 但目前大多数文献主要集中于非空间相依数据分析; (2) 高维空间相依数据的稳健分析开始引起学界关注, 但 Expectile 回归分析尚未开始; (3) 由于理论推导的便利, 外生空间权重矩阵下的空间回归分析更为普遍, 但实际数据并非如此, 特别地, 在异质性、内生性存在的条件下设定内生空间权重矩阵的回归分析更为合理. 因此, 本文分别基于外生空间权重矩阵和内生空间权重矩阵的设置, 对高维空间相依数据进行 Expectile 回归分析, 结合惩罚方法给出未知参数的稳健估计, 实现高维解释变量的选择, 并建立相关大样本理论结果, 最后通过数值模拟与实例分析验证所提出方法的有效性.

本文余下内容的安排如下. 第 2 节简要介绍非对称二次函数和惩罚函数. 第 3 节给出空间 Expectile 回归分析, 包括模型设定、假设条件、惩罚方法和大样本理论. 第 4 节考察所提出方法的数值表现, 给出不同参数条件下多种 Monte Carlo 模拟结果, 并进行对比分析. 第 5 节应用所提出方法对实际空间相依数据进行 Expectile 回归分析. 附录给出大样本理论证明过程.

为叙述方便, 以下对文中出现的部分符号进行说明. 记 $\boldsymbol{A}_{i,n}$ 为矩阵 $\boldsymbol{A}_n$ 的第 i 行元素构成的列向量, $\mathrm{vec}_i(\boldsymbol{A}_n)$ 为矩阵 $\boldsymbol{A}_n$ 向量化后的第 i 个元素; $\|\boldsymbol{X}\|_p=[\mathrm{E}(|\boldsymbol{X}|^p)]^{1/p}$ 是随机向量 $\boldsymbol{X}$ 的 L_p 范数, 其中 $|\boldsymbol{X}|$ 表示 $\boldsymbol{X}$ 的 Euclid 范数; $\|\boldsymbol{A}_n\|_\infty=\max_i\sum_j|a_{ij,n}|$, 其中 $a_{ij,n}$ 为 $\boldsymbol{A}_n$ 的第 i 行第 j 列的元素; $l(i)$ 表示个体 i 的位置; $\boldsymbol{B}_s(i)$ 表示中心在 $l(i)$、以 s 为半径的球域; C_a 表示与 a 相关的某常数, $r(\boldsymbol{A}_n)$ 表示矩阵 $\boldsymbol{A}_n$ 的谱半径; $\boldsymbol{l}_n$ 是元素全为 1 的 n 维列向量.

2 非对称二次函数和惩罚函数

2.1 非对称二次函数

对于任意实数 $\tau \in (0,1)$, 非对称二次函数定义为

$$\Psi_\tau(u) = |\tau - I(u<0)|u^2, \quad u \in (-\infty, +\infty).$$

不难发现, $\Psi_\tau(u)$ 可导, 记 $\phi_\tau(u)$ 和 $\varphi_\tau(u)$ 分别是 $\Psi_\tau(u)$ 可导处的一阶导数和二阶导数. 若 $\tau = 0.5$, 则它变成对称函数, 即为二次函数; 若 $\tau \neq 0.5$, 则变为非对称函数, 且在远离 0.5 的地方, 非对称性将更加明显. 因此, τ 的取值控制非对称函数 $\Psi_\tau(u)$ 的不对称程度, 可称为 Expectile 水平.

通常, 以非对称函数 $\Psi_\tau(u)$ 作为损失函数的回归称为 Expectile 回归. 当 Expectile 水平取为 0.5 时, Expectile 回归退化为最小二乘回归. 观察到 $\Psi_\tau(u)$ 只是适当改善了分位数回归损失函数在原点的光滑性, 因此 Expectile 回归与分位数回归一样, 具有全面刻画因变量条件分布的能力 (参见文献 [14,34]), 并且在处理异质类数据时带有较好的稳健性, 同时在数值计算和理论证明方面具有相对较大优势.

2.2 惩罚函数

本文的惩罚函数指在变量选择研究中用来控制模型稀疏性的函数. 在理论与应用上存在 LASSO、自适应 LASSO 和 SCAD 等多种形式, 可参见文献 [26]. 下面介绍性质较好且具有代表性的自适应 LASSO 惩罚函数和 SCAD 惩罚函数.

自适应 LASSO 函数为

$$p_\lambda^{\mathrm{ALAS}}(|\theta|) = \frac{\lambda}{|\hat{\theta}|^r}|\theta|,$$

其中, λ 是调节参数, $\hat{\theta}$ 是真实 θ 的相合估计, r 为某正常数. 此函数是 Zou [35] 在 LASSO 函数的基础上提出的, 它主要通过自适应权重改善惩罚的效果而获得未知参数估计的相合性、稀疏性与渐近正态性等大样本性质, 从而达到有效选出重要变量的目的.

SCAD 惩罚函数为

$$p_\lambda^{\mathrm{SCAD}}(|\theta|) = \begin{cases} \lambda|\theta|, & |\theta| \leqslant \lambda, \\ -\dfrac{1}{2(a-1)}(\theta^2 - 2a\lambda|\theta| + \lambda^2), & \lambda < |\theta| \leqslant a\lambda, \\ \dfrac{a+1}{2}\lambda^2, & |\theta| > a\lambda, \end{cases}$$

其中, λ 是调节参数, a 是大于 2 的某正数. 此函数是 Fan 和 Li [7] 依据 LASSO 函数的惩罚思想, 通过压缩力度和连续性的调整, 从而避免回归模型中 LASSO 惩罚出现的偏差, 获得具有相合性、稀疏性和渐近正态性的参数估计, 同时实现重要变量的选择.

3 空间 Expectile 回归分析

由于大数据易出现空间相依性、异质性和高维特征的叠加影响, 从而形成高维复杂空间相依数据, 并使已有统计分析方法面临适用性和有效性问题. 以下基于 Expectile 方法呈现出稳健、全面、可微等

诸多优势, 同时考虑惩罚方法的优良性、理论的完备性以及估计与算法上的差异, 尝试结合具有代表性的 SCAD 惩罚与自适应 LASSO 惩罚方法, 对高维空间相依数据模型进行 Expectile 回归分析, 包括模型介绍、估计方法、假设条件和理论性质研究.

3.1 外生空间权重矩阵的情形

空间滞后模型是通过引入空间权重矩阵对空间相依数据进行分析的重要模型, 在区域经济学、金融学、环境科学和流行病学等领域有着广泛应用. 在此模型中, 空间权重矩阵主要用于反映变量间的空间相依关系, 常设定为外生的, 即按照某一邻接关系直接给定, 与模型的解释变量和误差项不相关.

定义 3.1 设 $\boldsymbol{y}_n=(y_{1,n},\dots,y_{n,n})^{\mathrm{T}}$ 表示响应变量的观测值, $\boldsymbol{X}_n$ 是 $n\times p$ 维观测值矩阵, 则外生空间权重矩阵条件下的空间滞后模型为

$$\boldsymbol{y}_n=\rho\boldsymbol{W}_n\boldsymbol{y}_n+\boldsymbol{X}_n\boldsymbol{\beta}+\boldsymbol{e}_n, \tag{3.1}$$

其中, $\boldsymbol{W}_n$ 为外生空间权重矩阵, $\boldsymbol{\beta}=(\beta_1,\dots,\beta_p)^{\mathrm{T}}$ 是 p 维回归系数向量, $\boldsymbol{e}_n=(e_{1,n},\dots,e_{n,n})^{\mathrm{T}}$ 为独立同分布的随机误差项, ρ 为空间效应系数. 特别地, 当 $\rho=0$ 时, 模型 (3.1) 退化为普通的线性模型.

在模型 (3.1) 中, 若解释变量维数较高, 而实际重要的变量较少, 则可以考虑对回归系数 $\boldsymbol{\beta}$ 进行惩罚, 结合线性模型的 Expectile 回归分析, 构造如下损失函数:

$$\Phi_n(\boldsymbol{\theta};\tau)=\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-\boldsymbol{X}_{i,n}^{\mathrm{T}}\boldsymbol{\beta}-\rho\boldsymbol{W}_{i,n}^{\mathrm{T}}\boldsymbol{y}_n)+n\sum_{j=1}^{p}p_{\lambda_n}(|\theta_j|),$$

其中 $\boldsymbol{\theta}=(\boldsymbol{\beta}^{\mathrm{T}},\rho)^{\mathrm{T}}=(\theta_1,\dots,\theta_{p+1})^{\mathrm{T}}$. 从而, 未知参数 $\boldsymbol{\theta}$ 的估计量为

$$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\Phi_n(\boldsymbol{\theta};\tau).$$

上述惩罚估计未考虑 $\boldsymbol{W}_n\boldsymbol{y}_n$ 的内生性, 将造成空间效应系数估计不具有无偏性. 为了去掉内生性的影响, 以下引入 $n\times k$ 阶的工具变量矩阵 $\boldsymbol{Q}_n$. 然而, 由于工具变量 $\boldsymbol{Q}_n$ 往往从解释变量 $\boldsymbol{X}_n$ 中选择, 所以高维解释变量的出现会使得可供选择的工具变量维数显著增加. 因此, 面对高维解释变量和工具变量并存的情形, 本文提出如下两步惩罚 Expectile 估计:

第 1 步 选择工具变量的惩罚 Expectile 估计

$$\tilde{\boldsymbol{\kappa}}=\arg\min_{\boldsymbol{\kappa}}\sum_{i=1}^{n}\Psi_\tau(\boldsymbol{W}_{i,n}^{\mathrm{T}}\boldsymbol{y}_n-\boldsymbol{Q}_{i,n}^{\mathrm{T}}\boldsymbol{\kappa})+n\sum_{j=1}^{k}p_{\lambda_n}(|\kappa_j|);$$

第 2 步 去掉内生性后的惩罚 Expectile 估计

$$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-(\boldsymbol{X}_{i,n}^{\mathrm{T}},\boldsymbol{Q}_{i,n}^{\mathrm{T}}\tilde{\boldsymbol{\kappa}})\boldsymbol{\theta})+n\sum_{j=1}^{p}p_{\lambda_n}(|\theta_j|).$$

上述两步均选择 SCAD 惩罚函数或自适应 LASSO 惩罚函数, 它们的优良性质有利于确保在适当的调节参数下可以将不重要变量的系数压缩为 0, 从而同时实现未知参数的估计及重要变量选择的目的. 选择 SCAD 惩罚函数和自适应 LASSO 惩罚函数所得的估计量分别记为 $\hat{\boldsymbol{\theta}}_{\mathrm{SCAD}}$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{ALAS}}$.

令 $\Phi_n^{\mathrm{ex}}(\boldsymbol{\theta};\tau)=\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-\tilde{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\theta})+n\sum_{j=1}^{p}p_{\lambda_n}(|\theta_j|)$, $\tilde{\boldsymbol{X}}_{i,n}=(\boldsymbol{X}_{i,n}^{\mathrm{T}},\boldsymbol{Q}_{i,n}^{\mathrm{T}}\tilde{\boldsymbol{\kappa}})^{\mathrm{T}}$, $\tilde{\boldsymbol{\kappa}}\xrightarrow{\mathrm{P}}\boldsymbol{\kappa}_0$, $\boldsymbol{X}_{i,n}^{\star}=(\boldsymbol{X}_{i,n}^{\mathrm{T}},\boldsymbol{Q}_{i,n}^{\mathrm{T}}\boldsymbol{\kappa}_0)^{\mathrm{T}}$, $\boldsymbol{\theta}_0=(\boldsymbol{\theta}_{10}^{\mathrm{T}},\boldsymbol{0}^{\mathrm{T}})^{\mathrm{T}}$ 是 $\boldsymbol{\theta}$ 的真实值, $\boldsymbol{\theta}_{10}^{\mathrm{T}}$ 是 q 维非零参数向量, 向量 $\tilde{\boldsymbol{X}}_{i,n}=(\tilde{\boldsymbol{X}}_{i,n1}^{\mathrm{T}},\tilde{\boldsymbol{X}}_{i,n2}^{\mathrm{T}})^{\mathrm{T}}$、

$\hat{\boldsymbol{\theta}}_{\mathrm{SCAD}} = (\hat{\boldsymbol{\theta}}_{\mathrm{SCAD},1}^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_{\mathrm{SCAD},2}^{\mathrm{T}})^{\mathrm{T}}$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{ALAS}} = (\hat{\boldsymbol{\theta}}_{\mathrm{ALAS},1}^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_{\mathrm{ALAS},2}^{\mathrm{T}})^{\mathrm{T}}$ 是类似于 $\boldsymbol{\theta}_0$ 的分解, $\tilde{\xi}_{i,n} = y_{i,n} - \tilde{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\theta}_{10}$, $\mu_{\mathrm{ex},\varphi_\tau} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\varphi_\tau(\tilde{\xi}_{i,n})]$, $\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{ex}} = \lim_{n\to\infty}\frac{1}{n}\mathrm{Var}[\sum_{i=1}^{n}\phi_\tau(\tilde{\xi}_{i,n})\tilde{\boldsymbol{X}}_{i,n1}]$, $\boldsymbol{\Sigma}_{\mathrm{ex},11}$ 是 $\boldsymbol{\Sigma}_{\mathrm{ex}}$ (见假设 3.2) 左上角与 $\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{ex}}$ 同阶的方阵.

为了研究上述估计的大样本性质, 下面给出一些常规的正则条件.

假设 3.1 对于任意 $\tau \in (0,1)$, 随机误差项 $\boldsymbol{e}_n$ 是连续型随机向量, 密度函数在 0 的某邻域为正, 存在正数 α, 使得 $\mathrm{E}|e_{i,n}|^{4+\alpha} < \infty$, 每一步估计残差的条件 τ-Expectile 为 0.

假设 3.2 $\mu_{\mathrm{ex},\varphi_\tau}$ 与 $\boldsymbol{\Sigma}_{\varphi_\tau,11}^{\mathrm{ex}}$ 可逆, $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{i,n}^{\star}\boldsymbol{X}_{i,n}^{\star\mathrm{T}}$ 依概率 1 收敛到非奇异矩阵 $\boldsymbol{\Sigma}_{\mathrm{ex}}$.

假设 3.3 对于任意 i、j 和 n, 空间权重矩阵的元素满足

$$w_{ij,n} \geqslant 0, \quad w_{ii,n} = 0, \quad \sum_{i=1}^{n}|w_{ij}| + \sum_{j=1}^{n}|w_{ij}| \leqslant C_w.$$

假设 3.4 (1) $\lim_{n\to\infty}\lambda_n = 0$;

(2) $\lim_{n\to\infty}\sqrt{n}\lambda_n = \infty$;

(3) $\lim_{n\to\infty}\sqrt{n}\lambda_n = 0$;

(4) $\lim_{n\to\infty} n^{(r+1)/2}\,\lambda_n = \infty$.

在经典线性模型的 Expectile 回归分析中需要误差项的 4 阶矩, 叠加空间滞后导致的内生性后需要用到相依序列的中心极限定理, 从而使误差项的矩条件更高, 假设 3.1 对于多数常见分布是合适的, 如正态分布. 每一步估计残差的条件 τ-Expectile 为 0 类似于均值回归中误差项的期望为 0 的假设, 可以通过平移截距项达到这一要求. 假设 3.2 在一定程度上要求工具变量和解释变量观测值不相关, 当空间效应不存在时, 退化为常用条件. 假设 3.3 要求外生空间权重矩阵绝对行和与列和一致有界, 是常见条件. 假设 3.4 是 SCAD 惩罚与自适应 LASSO 惩罚方法的常用条件.

定理 3.1 若假设 3.1–3.3 和 3.4(1) 成立, 则 $\Phi_n^{\mathrm{ex}}(\boldsymbol{\theta};\tau)$ 存在极小值点, 且有

$$\|\hat{\boldsymbol{\theta}}_{\mathrm{SCAD}} - \boldsymbol{\theta}_0\| = O_{\mathrm{P}}(n^{-1/2}).$$

定理 3.2 对于定理 3.1 中 $\boldsymbol{\theta}$ 的 $\sqrt{n}$ 相合估计量 $\hat{\boldsymbol{\theta}}_{\mathrm{SCAD}} = (\hat{\boldsymbol{\theta}}_{\mathrm{SCAD},1}^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_{\mathrm{SCAD},2}^{\mathrm{T}})^{\mathrm{T}}$, 在假设 3.1–3.3、3.4(1) 和 3.4(2) 成立的条件下, 以概率 1 满足

(1) 稀疏性: $\hat{\boldsymbol{\theta}}_{\mathrm{SCAD},2} = \boldsymbol{0}$;

(2) 渐近正态性: $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{SCAD},1} - \boldsymbol{\theta}_{10}) \xrightarrow{L} N(\boldsymbol{0}, \mu_{\mathrm{ex},\varphi_\tau}^{-2}\boldsymbol{\Sigma}_{\mathrm{ex},11}^{-1}\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{ex}}\boldsymbol{\Sigma}_{\mathrm{ex},11}^{-1})$.

定理 3.3 若假设 3.1–3.3、3.4(3) 和 3.4(4) 成立, 则 $\boldsymbol{\theta}$ 的估计量 $\hat{\boldsymbol{\theta}}_{\mathrm{ALAS}} = (\hat{\boldsymbol{\theta}}_{\mathrm{ALAS},1}^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_{\mathrm{ALAS},2}^{\mathrm{T}})^{\mathrm{T}}$, 以概率 1 满足

(1) 稀疏性: $\hat{\boldsymbol{\theta}}_{\mathrm{ALAS},2} = \boldsymbol{0}$;

(2) 渐近正态性: $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{ALAS},1} - \boldsymbol{\theta}_{10}) \xrightarrow{L} N(\boldsymbol{0}, \mu_{\mathrm{ex},\varphi_\tau}^{-2}\boldsymbol{\Sigma}_{\mathrm{ex},11}^{-1}\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{ex}}\boldsymbol{\Sigma}_{\mathrm{ex},11}^{-1})$.

假设 3.1–3.3 表明, 在外生空间权重矩阵情形下的 Expectile 回归分析中, 未知参数的 SCAD 和自适应 LASSO 惩罚估计量具有相合性、稀疏性和渐近正态性, 在大样本条件下能够实现重要参数的有效估计和重要变量的自动选择.

3.2 内生空间权重矩阵的情形

虽然外生空间权重矩阵的设定给空间相依数据的理论分析带来了便利, 但是也给应用工作者带来了不少困惑. 例如, 空间权重矩阵的预先设定形式多样, 该如何选择呢? 错误的选择会对实证分析产生

多大的影响? 事实上, 若空间权重矩阵具有内生性, 则需要重新探究模型的构建和参数估计方法. 基于此, Qu 和 Lee [24] 研究了带内生空间权重矩阵的空间滞后模型, 并给出如下定义.

定义 3.2 [24] 设 $\boldsymbol{y}_n=(y_{1,n},\ldots,y_{n,n})^{\mathrm{T}}$ 为响应变量的观测值, $\boldsymbol{X}_{1n}$ 是 $n\times p_1$ 维观测值矩阵, $\boldsymbol{X}_{2n}$ 是 $n\times p_2$ 维观测值矩阵, $\boldsymbol{Z}_n$ 是 $n\times l$ 维观测值矩阵, 则内生空间权重矩阵条件下的空间滞后模型为

$$\begin{cases}\boldsymbol{y}_n=\rho\boldsymbol{W}_n\boldsymbol{y}_n+\boldsymbol{X}_{1n}\boldsymbol{\beta}+\boldsymbol{e}_n,\\ \boldsymbol{W}_n=\boldsymbol{H}(\boldsymbol{Z}_n,\boldsymbol{D}_n),\\ \boldsymbol{Z}_n=\boldsymbol{X}_{2n}\boldsymbol{\Gamma}+\boldsymbol{V}_n,\end{cases}\tag{3.2}$$

其中, $\boldsymbol{W}_n$ 为内生空间权重矩阵, ρ 是空间效应系数, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{p_1})^{\mathrm{T}}$ 是 p_1 维回归系数向量, $\boldsymbol{e}_n=(e_{1,n},\ldots,e_{n,n})^{\mathrm{T}}$ 为随机误差项, $\boldsymbol{H}(\boldsymbol{Z}_n,\boldsymbol{D}_n)=(h_{ij}(\boldsymbol{Z}_n,d_{ij}))_n$, $h_{ij}(\cdot)$ 是已知的有界函数, $\boldsymbol{D}_n=(d_{ij})_n$, d_{ij} 为个体 i 和 j 所在位置的距离, $\boldsymbol{\Gamma}$ 是 $p_2\times l$ 维系数矩阵, $\boldsymbol{V}_n=(\boldsymbol{v}_{1,n},\ldots,\boldsymbol{v}_{n,n})^{\mathrm{T}}$ 是随机扰动项矩阵, 第 i 行的元素为 $\boldsymbol{v}_{i,n}=(v_{1,in},\ldots,v_{l,in})^{\mathrm{T}}$.

注 3.1 此处的空间权重矩阵 $\boldsymbol{W}_n$ 具有随机性, 依赖于随机变量 $\boldsymbol{Z}_n$ 和位置距离矩阵 $\boldsymbol{D}_n$. 注意到 $\boldsymbol{Z}_n$ 与 $\boldsymbol{V}_n$ 相关, 当 $\boldsymbol{V}_n$ 与 $\boldsymbol{e}_n$ 相关时可导致 $\boldsymbol{W}_n$ 与 $\boldsymbol{e}_n$ 相关, 从而使空间权重矩阵产生内生性.

在模型 (3.2) 中, 不仅面临高维解释变量和工具变量, 而且还面临空间权重矩阵与空间滞后项的双重内生性. 下面借鉴工具变量估计思想 [24], 采用工具变量去掉空间滞后项的内生性, 引入控制变量处理空间权重矩阵的内生性, 结合 Expectile 回归与惩罚方法, 构造出三步惩罚 Expectile 估计:

第 1 步 给出 $\boldsymbol{\Gamma}$ 的惩罚 Expectile 估计

$$\hat{\boldsymbol{\Gamma}}=\arg\min_{\boldsymbol{\Gamma}}\sum_{i=1}^{nl}\Psi_\tau(\mathrm{vec}_i(\boldsymbol{Z}_n-\boldsymbol{X}_{2n}\boldsymbol{\Gamma}))+n\sum_{j=1}^{lp_2}p_{\lambda_n}(|\mathrm{vec}_j(\boldsymbol{\Gamma})|);$$

第 2 步 选择工具变量的惩罚 Expectile 估计

$$\hat{\boldsymbol{\kappa}}=\arg\min_{\boldsymbol{\kappa}}\sum_{i=1}^{n}\Psi_\tau(\boldsymbol{W}_{i,n}^{\mathrm{T}}\boldsymbol{y}_n-\boldsymbol{Q}_{i,n}^{\mathrm{T}}\boldsymbol{\kappa})+n\sum_{j=1}^{k}p_{\lambda_n}(|\kappa_j|);$$

第 3 步 去掉内生性后的惩罚 Expectile 估计

$$\hat{\boldsymbol{\vartheta}}=\arg\min_{\boldsymbol{\vartheta}}\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-(\boldsymbol{X}_{i,1n}^{\mathrm{T}},\boldsymbol{Q}_{i,n}^{\mathrm{T}}\hat{\boldsymbol{\kappa}},(\boldsymbol{Z}_{i,n}-\boldsymbol{X}_{i,2n}^{\mathrm{T}}\hat{\boldsymbol{\Gamma}})^{\mathrm{T}})\boldsymbol{\vartheta})+n\sum_{j=1}^{p_1}p_{\lambda_n}(|\vartheta_j|),$$

其中, $\boldsymbol{\vartheta}=(\boldsymbol{\beta}^{\mathrm{T}},\rho,\boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}=(\vartheta_1,\ldots,\vartheta_{p_1+l+1})^{\mathrm{T}}$, $\boldsymbol{Q}_n$ 为 $n\times k$ 阶的工具变量矩阵, $\boldsymbol{\delta}$ 的定义参见假设 3.5.

第 1 步估计出 $\boldsymbol{\Gamma}$ 后, $\boldsymbol{Z}_n-\boldsymbol{X}_{2n}\hat{\boldsymbol{\Gamma}}$ 可作为控制变量控制空间权重矩阵 $\boldsymbol{W}_n$ 的内生性; 第 2 步估计出的 $\hat{\boldsymbol{\kappa}}$ 用于选择工具变量进而控制 $\boldsymbol{W}_n\boldsymbol{y}_n$ 的内生性; 第 3 步可以看成对去掉内生性后的线性模型进行惩罚 Expectile 回归分析. 上述三步均选择相同的惩罚函数, 在适当的调节参数下可以将非重要变量的系数压缩为 0, 从而同步实现未知参数估计及重要变量选择的目的. 选择 SCAD 惩罚函数和自适应 LASSO 惩罚函数所得的估计量分别记为 $\hat{\boldsymbol{\vartheta}}_{\mathrm{SCAD}}$ 和 $\hat{\boldsymbol{\vartheta}}_{\mathrm{ALAS}}$.

令

$$\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta};\tau)=\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta})+n\sum_{j=1}^{p_1}p_{\lambda_n}(|\vartheta_j|),\quad \hat{\boldsymbol{X}}_{i,n}=(\boldsymbol{X}_{i,1n}^{\mathrm{T}},\boldsymbol{Q}_{i,n}^{\mathrm{T}}\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{Z}}_{i,n}^{\mathrm{T}})^{\mathrm{T}},$$

$$\hat{\boldsymbol{Z}}_{i,n}=(\boldsymbol{Z}_{i,n}-\boldsymbol{X}_{i,2n}^{\mathrm{T}}\hat{\boldsymbol{\Gamma}}),\quad \hat{\boldsymbol{\kappa}}\xrightarrow{\mathrm{P}}\boldsymbol{\kappa}_0,\quad \hat{\boldsymbol{\Gamma}}\xrightarrow{\mathrm{P}}\boldsymbol{\Gamma}_0,$$

$$\boldsymbol{X}^*_{i,n}=(\boldsymbol{X}^{\rm T}_{i,1n},\boldsymbol{Q}^{\rm T}_{i,n}\boldsymbol{\kappa}_0,\boldsymbol{Z}^{*\rm T}_{i,n})^{\rm T},\quad \boldsymbol{Z}^*_{i,n}=(\boldsymbol{Z}_{i,n}-\boldsymbol{X}^{\rm T}_{i,2n}\boldsymbol{\Gamma}_0),$$
$$\mu_{{\rm en},\varphi_\tau}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n{\rm E}[\varphi_\tau(\hat{\xi}_{i,n})],\quad \boldsymbol{\Sigma}^{\rm en}_{\phi_\tau,11}=\lim_{n\to\infty}\frac{1}{n}{\rm Var}\Bigg[\sum_{i=1}^n\phi_\tau(\hat{\xi}_{i,n})\hat{\boldsymbol{X}}_{i,n1}\Bigg],$$

$\boldsymbol{\vartheta}_0=(\boldsymbol{\vartheta}^{\rm T}_{10},\boldsymbol{0}^{\rm T})^{\rm T}$ 是 $\boldsymbol{\vartheta}$ 的真实值, $\boldsymbol{\vartheta}^{\rm T}_{10}$ 是 q 维非零参数向量, $\hat{\boldsymbol{\vartheta}}_{\rm SCAD}=(\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm SCAD,1},\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm SCAD,2})^{\rm T}$、$\hat{\boldsymbol{\vartheta}}_{\rm ALAS}=(\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm ALAS,1},\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm ALAS,2})^{\rm T}$ 和 $\hat{\boldsymbol{X}}_{i,n}=(\hat{\boldsymbol{X}}^{\rm T}_{i,n1},\hat{\boldsymbol{X}}^{\rm T}_{i,n2})^{\rm T}$ 是类似于 $\boldsymbol{\vartheta}_0$ 的分解, $\hat{\xi}_{i,n}=y_{i,n}-\hat{\boldsymbol{X}}^{\rm T}_{i,n1}\boldsymbol{\vartheta}_{10}$, $\boldsymbol{\Sigma}_{\rm en,11}$ 是 $\boldsymbol{\Sigma}_{\rm en}$ (见假设 3.7) 左上角与 $\boldsymbol{\Sigma}^{\rm en}_{\phi_\tau,11}$ 同阶的方阵. 内生空间权重矩阵将导致模型 (3.2) 具有双重内生性, 使其与模型 (3.1) 存在较大差异.

假设 3.5　联合误差 $(e_{i,n},\boldsymbol{v}^{\rm T}_{i,n})^{\rm T}\overset{\rm i.i.d.}{\sim}(0,\boldsymbol{\Sigma}_{ev})$, 其中 $\boldsymbol{\Sigma}_{ev}=\left(\begin{smallmatrix}\sigma_e^2 & \boldsymbol{\sigma}^{\rm T}_{ev}\\ \boldsymbol{\sigma}_{ev} & \boldsymbol{\Sigma}_v\end{smallmatrix}\right)$ 是正定矩阵, $\boldsymbol{\Sigma}_v$ 是 $\boldsymbol{v}_{i,n}$ 的协方差阵, ${\rm E}(e_{i,n}\mid\boldsymbol{v}_{i,n})=\boldsymbol{v}^{\rm T}_{i,n}\boldsymbol{\delta}$, ${\rm Var}(e_{i,n}\mid\boldsymbol{v}_{i,n})=\sigma^2_\xi$, $\boldsymbol{\delta}=\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\sigma}_{ev}$, $\sigma^2_\xi=\sigma^2_e-\boldsymbol{\sigma}^{\rm T}_{ev}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\sigma}_{ev}$. 对于任意 $\tau\in(0,1)$, 随机误差项 $\boldsymbol{e}_n$ 和 $\boldsymbol{v}_{i,n}$ 是连续型随机向量, 密度函数在 0 的某邻域为正, 存在正数 α, 使 ${\rm E}|e_{i,n}|^{4+\alpha}<\infty$, ${\rm E}\|\boldsymbol{v}_{i,n}\|^{4+\alpha}<\infty$, 每一步估计残差的条件 τ-Expectile 为 0.

假设 3.6　格子点集合 $\boldsymbol{D}\subseteq\boldsymbol{R}^{d_0}$ $(d_0\geqslant 1)$ 无穷可数. 位置函数 $l:\{1,2,\ldots,n\}\to\boldsymbol{D}_n\subseteq\boldsymbol{D}$ 是个体 i 到其位置 $l(i)\in\boldsymbol{D}_n\subseteq\boldsymbol{R}^{d_0}$ 的映射. $\boldsymbol{D}$ 中不同元素位置的距离至少为 $d^*>0$, 即 $d_{ij}\geqslant d^*$, 其中 d_{ij} 表示个体 i 和 j 所在位置的距离, 不失一般性, 设 $d^*=1$.

假设 3.7　对于任意 i、j 和 n, 空间权重矩阵的元素满足

$$w_{ij,n}\geqslant 0,\quad w_{ii,n}=0,\quad \sup_n\|\boldsymbol{W}_n\|_\infty=C_w<\infty.$$

未知参数空间为 Euclid 空间中的紧集, 真实参数为紧集中的内点, $\sup(|\rho|C_w)<1$. $\mu_{{\rm en},\varphi_\tau}$ 与 $\boldsymbol{\Sigma}^{\rm en}_{\varphi_\tau,11}$ 可逆, $\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}^*_{i,n}\boldsymbol{X}^{*\rm T}_{i,n}$ 依概率 1 收敛到非奇异矩阵 $\boldsymbol{\Sigma}_{\rm en}$.

假设 3.8　(1) $w_{ij,n}=h_{ij}(\boldsymbol{Z}_{i,n},\boldsymbol{Z}_{j,n},d_{ij})$ $(i\neq j)$, $h_{ij}(\cdot)$ 是 $\boldsymbol{Z}_n$ 的非负、一致有界函数, $0\leqslant w_{ij,n}\leqslant c_1d_{ij}^{-c_3d_0}$ (对某 $c_1\geqslant 0,c_3\geqslant 3$), $\boldsymbol{W}_n$ 至多存在 r 列, 它们的列和超过 C_w, 且不依赖于 n;

(2) $w_{ij,n}=h_{ij}(\boldsymbol{Z}_{i,n},\boldsymbol{Z}_{j,n})\cdot I(d_{ij}\leqslant d_c)$ 或 $h_{ij}(\boldsymbol{Z}_{i,n},\boldsymbol{Z}_{j,n})\cdot I(d_{ij}\leqslant d_c)/\sum_{d_{ij}\leqslant d_c}h_{ik}(\boldsymbol{Z}_{i,n},\boldsymbol{Z}_{k,n})$ $(i\neq j)$, $w_{ij,n}=0$ (若 $d_{ij}>d_c>1$), 其中 $h_{ij}(\cdot)$ 是非负、一致有界函数.

与外生空间权重矩阵情形相比, 内生空间权重矩阵情形下的假设更加复杂. 假设 3.5 多了一些误差项的约束条件, 是依据空间权重矩阵的内生性而设定. 假设 3.6 是个体空间相邻的位置条件. 假设 3.7 和 3.8 是空间权重矩阵元素和结构的要求. 上述假设可参见文献 [24].

定理 3.4　若假设 3.5–3.8 和 3.4(1) 成立, 则 $\Phi^{\rm en}_n(\boldsymbol{\vartheta};\tau)$ 存在极小值点, 且有

$$\|\hat{\boldsymbol{\vartheta}}_{\rm SCAD}-\boldsymbol{\vartheta}_0\|=O_{\rm P}(n^{-1/2}).$$

定理 3.5　定理 3.4 中 $\boldsymbol{\vartheta}$ 的 $\sqrt{n}$ 相合估计量 $\hat{\boldsymbol{\vartheta}}_{\rm SCAD}=(\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm SCAD,1},\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm SCAD,2})^{\rm T}$, 在假设 3.5–3.8、3.4(1) 和 3.4(2) 成立的条件下, 以概率 1 满足

(1) 稀疏性: $\hat{\boldsymbol{\vartheta}}_{\rm SCAD,2}=\boldsymbol{0}$;

(2) 渐近正态性: $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{\rm SCAD,1}-\boldsymbol{\vartheta}_{10})\xrightarrow{L}N(\boldsymbol{0},\mu^{-2}_{{\rm en},\varphi_\tau}\boldsymbol{\Sigma}^{-1}_{\rm en,11}\boldsymbol{\Sigma}^{\rm en}_{\phi_\tau,11}\boldsymbol{\Sigma}^{-1}_{\rm en,11})$.

定理 3.6　若假设 3.5–3.8、3.4(3) 和 3.4(4) 成立, 则 $\boldsymbol{\vartheta}$ 的估计量 $\hat{\boldsymbol{\vartheta}}_{\rm ALAS}=(\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm ALAS,1},\hat{\boldsymbol{\vartheta}}^{\rm T}_{\rm ALAS,2})^{\rm T}$, 以概率 1 满足

(1) 稀疏性: $\hat{\boldsymbol{\vartheta}}_{\rm ALAS,2}=\boldsymbol{0}$;

(2) 渐近正态性: $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{\rm ALAS,1}-\boldsymbol{\vartheta}_{10})\xrightarrow{L}N(\boldsymbol{0},\mu^{-2}_{{\rm en},\varphi_\tau}\boldsymbol{\Sigma}^{-1}_{\rm en,11}\boldsymbol{\Sigma}^{\rm en}_{\phi_\tau,11}\boldsymbol{\Sigma}^{-1}_{\rm en,11})$.

定理 3.4–3.6 表明, 在一般正则条件下, 内生空间权重矩阵与外生空间权重矩阵的高维 Expectile 回归分析结论相似, 均具有优良的大样本性质. 当 $\boldsymbol{\delta}=\boldsymbol{0}$ 时, 空间权重矩阵无内生性, 上述结论与定

理 3.1–3.3 的结论一致. 本节所给的定理表明, 虽然 SCAD 惩罚方法和自适应 LASSO 惩罚方法在调节参数的收敛速度上存在一定差异, 但是都能有效地选出重要变量, 具有优良的 Oracle 性质.

4 数值模拟

本节将通过 Monte Carlo 模拟验证所提出方法的数值表现. 在实施过程中, 分别按照外生和内生空间权重矩阵条件下的惩罚 Expectile 估计步骤, 使用 R 语言中的 Rmosek 包和 lars 包来求解相关优化问题, 并采用 Bayes 信息准则 (Bayesian information criterion, BIC) 选取调节参数 λ 的最优值. 本节给出 SCAD 惩罚函数和自适应 LASSO 惩罚函数下的模拟结果, 并比较它们的异同.

样本生成过程为如下. 解释变量 $\boldsymbol{X}_{1n}=(X_{1,1n},\dots,X_{p,1n})^{\mathrm{T}}$ 按两步来产生, 第一步从标准正态分布中产生 n 个 p 维向量, $\tilde{X}_{k,1n}\sim N(0,1)$, $k=1,\dots,p$; 第二步设置 $X_{4,1n}=\Phi(\tilde{X}_{4,1n})$, $X_{k,1n}=\tilde{X}_{k,1n}$ $(k\neq 4)$, $\Phi(\cdot)$ 是标准正态分布的累积分布函数. 对于响应变量, 其生成机制如下:

$$\boldsymbol{y}_n=(\boldsymbol{I}_n-\rho\boldsymbol{W}_n)^{-1}\{\boldsymbol{X}_{1n}\boldsymbol{\beta}+(X_{4,1n})\boldsymbol{e}_n\},$$

其中, $\boldsymbol{\beta}=(1,1,1,0,\dots,0)^{\mathrm{T}}$ 是 p 维的系数向量, 即前 3 个解释变量是重要变量, 对应的系数为 1, 其余均为非重要变量. 对内生空间权重矩阵 $\boldsymbol{W}_n$, 首先, 生成 n 维向量

$$\boldsymbol{Z}_n=\boldsymbol{X}_{2n}\boldsymbol{\Gamma}+(X_{4,2n})\boldsymbol{V}_n,$$

其中, $\boldsymbol{X}_{2n}=(X_{1,2n},\dots,X_{p,2n})^{\mathrm{T}}=\boldsymbol{X}_{1n}$, $X_{4,2n}=X_{4,1n}$, $\boldsymbol{\Gamma}=\boldsymbol{\beta}=(1,1,1,0,\dots,0)^{\mathrm{T}}$, $\boldsymbol{V}_n=(v_{1,n},\dots,v_{n,n})^{\mathrm{T}}$, $\boldsymbol{e}_n=(e_{1,n},\dots,e_{n,n})^{\mathrm{T}}$ 是误差项, $(v_{i,n},e_{i,n})$ 是来自均值为 $\mathbf{0}$、方差为 $\boldsymbol{\Sigma}$、自由度为 10 的二维 t 分布, $\boldsymbol{\Sigma}=\left(\begin{smallmatrix}1 & \rho_{ev}\\ \rho_{ev} & 1\end{smallmatrix}\right)$, $i=1,\dots,n$. 其次, 通过 Hadamard 乘积构造内生空间权重矩阵 $\boldsymbol{W}_n=\boldsymbol{W}_n^d\circ\boldsymbol{W}_n^e$, 其中, $\boldsymbol{W}_n^d$ 是基于 Rook 邻接关系的矩阵, 即如果第 i 个个体和第 j 个个体是相邻的, 则有 $w_{ij}^d=1$, 否则 $w_{ij}^d=0$, $i,j=1,\dots,n$. $\boldsymbol{W}_n^e$ 是基于经济相似性的矩阵, 即如果 $i\neq j$, 则有

$$w_{ij}^e=\frac{1}{|z_{i,n}-z_{j,n}|+\varepsilon},$$

否则 $w_{ij}^e=0$, 其中, $\varepsilon=10^{-4}, z_{i,n}$ 是向量 $\boldsymbol{Z}_n$ 的第 i 个分量. 最后, 按照空间回归模型的分析惯例, 将空间权重矩阵 $\boldsymbol{W}_n$ 进行行标准化处理.

在模拟中, 使用两种相邻连接矩阵 $\boldsymbol{W}_n^d$, 第一种是基于全国 284 个地级市构造的邻接矩阵, 第二种是基于全国 2,890 个县的空间位置关系构造的邻接矩阵. 因此, 样本量分别为 $n=284$ 和 $n=2{,}890$. 为研究解释变量个数对数值模拟结果的影响, 解释变量的个数 p 分别取为 10 和 100. 为产生不同程度的内生性, 误差相关系数 ρ_{ev} 取为 0 和 0.8. 此外, 模型中的空间效应系数 ρ 选取为 0.4 和 0.8.

同时, 模拟比较两种不同的惩罚估计方法: 一种是外生方法, 即在外生空间权重矩阵条件下或者将内生空间权重矩阵看成外生的情形, 用所提出的两步惩罚 Expectile 方法进行模拟; 另一种是内生方法, 即在内生空间权重矩阵条件下, 用所提出的三步惩罚 Expectile 方法进行模拟. 进一步地, 本文还将给出未知参数的高维 Expectile 回归的 Oracle 估计 (Oracle 方法), 即假设重要变量已知, 然后采用不带惩罚的两步或三步 Expectile 方法对未知参数进行估计. 在每种设置下, 均进行 100 次模拟.

表 1 和 2 分别给出了当 $n=284$, $\rho_{ev}=0.8$ 和 $n=2,890,\rho_{ev}=0.8$ 时, 3 种方法所得参数估计的均值和标准误, 每张表中上半部分对应的是 $p=10$ 的结果, 下半部分对应的是 $p=100$ 的结果. 从表 1 可以看出, 在内生空间权重矩阵条件下, 参数估计的偏差和标准误都非常小, 所提出方法的估计精度

表 1　当 $n=284$、$\rho_{ev}=0.8$ 和 $\rho=0.4$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p=10$	$\hat{\rho}$	0.4907	0.4895	0.4879	0.3993	0.3989	0.4001	0.3996	0.3989	0.3997
		(0.0045)	(0.0041)	(0.0041)	(0.0021)	(0.0020)	(0.0021)	(0.0021)	(0.0020)	(0.0021)
	$\hat{\beta}_1$	0.9623	0.9646	0.9618	1.0026	1.0000	1.0014	1.0029	1.0039	1.0036
		(0.0048)	(0.0043)	(0.0046)	(0.0041)	(0.0036)	(0.0043)	(0.0040)	(0.0037)	(0.0041)
	$\hat{\beta}_2$	0.9585	0.9620	0.9627	1.0023	0.9975	1.0013	1.0000	1.0002	0.9996
		(0.0049)	(0.0046)	(0.0049)	(0.0047)	(0.0045)	(0.0042)	(0.0045)	(0.0040)	(0.0041)
	$\hat{\beta}_3$	0.9631	0.9643	0.9622	1.0040	1.0002	1.0010	1.0037	1.0031	1.0017
		(0.0052)	(0.0048)	(0.0050)	(0.0047)	(0.0042)	(0.0045)	(0.0047)	(0.0042)	(0.0044)
	$\hat{\delta}$				0.7901	0.7849	0.7846	0.8006	0.7944	0.7934
					(0.0059)	(0.0059)	(0.0068)	(0.0055)	(0.0055)	(0.0066)
	$\hat{\Gamma}_1$				1.0003	0.9943	0.9970	1.0017	1.0003	0.9997
					(0.0043)	(0.0038)	(0.0044)	(0.0040)	(0.0038)	(0.0042)
	$\hat{\Gamma}_2$				0.9984	0.9937	0.9992	0.9998	1.0002	1.0007
					(0.0044)	(0.0041)	(0.0042)	(0.0044)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0039	0.9981	1.0000	1.0048	1.0032	1.0028
					(0.0043)	(0.0038)	(0.0041)	(0.0044)	(0.0040)	(0.0041)
$p=100$	$\hat{\rho}$	0.4863	0.4834	0.4822	0.3998	0.3994	0.3999	0.3996	0.3989	0.3997
		(0.0056)	(0.0048)	(0.0049)	(0.0021)	(0.0020)	(0.0021)	(0.0021)	(0.0020)	(0.0021)
	$\hat{\beta}_1$	1.0096	1.0054	1.0057	1.0015	0.9992	1.0015	1.0029	1.0039	1.0036
		(0.0044)	(0.0041)	(0.0047)	(0.0044)	(0.0037)	(0.0043)	(0.0040)	(0.0037)	(0.0041)
	$\hat{\beta}_2$	1.0058	1.0044	1.0069	0.9999	0.9969	0.9998	1.0000	1.0002	0.9996
		(0.0049)	(0.0045)	(0.0050)	(0.0049)	(0.0045)	(0.0044)	(0.0045)	(0.0040)	(0.0041)
	$\hat{\beta}_3$	1.0097	1.0056	1.0059	1.0030	0.9988	1.0007	1.0037	1.0031	1.0017
		(0.0054)	(0.0048)	(0.0051)	(0.0050)	(0.0043)	(0.0046)	(0.0047)	(0.0042)	(0.0044)
	$\hat{\delta}$				0.7930	0.7873	0.7890	0.8006	0.7944	0.7934
					(0.0062)	(0.0059)	(0.0070)	(0.0055)	(0.0055)	(0.0066)
	$\hat{\Gamma}_1$				0.9981	0.9934	0.9958	1.0017	1.0003	0.9997
					(0.0047)	(0.0038)	(0.0043)	(0.0040)	(0.0038)	(0.0042)
	$\hat{\Gamma}_2$				0.9964	0.9926	0.9969	0.9998	1.0002	1.0007
					(0.0045)	(0.0041)	(0.0043)	(0.0044)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0024	0.9972	0.9995	1.0048	1.0032	1.0028
					(0.0046)	(0.0038)	(0.0040)	(0.0044)	(0.0040)	(0.0041)

注: 内生方法指内生空间权重矩阵下的三步惩罚 Expectile 方法. 外生方法指外生空间权重矩阵下的两步惩罚 Expectile 方法. Oracle 方法指已知重要变量情形下的 Expectile 方法. 小括号内的值为标准误, 后续表格类似, 将不再赘述.

较高, 与理想的 Oracle 估计方法具有可比性. 然而, 在内生空间权重矩阵的条件下, 使用外生方法得到的估计偏差和标准误均较大, 特别地, 空间效应系数 ρ 的偏差远大于未知参数 $\boldsymbol{\beta}$ 的偏差, 且外生方法严重高估了 ρ 的值. 对比表 1 上下部分可以看出, 尽管高维情形下 ($p=100$) 3 种方法的估计偏差和标准误比低维情形下 ($p=10$) 大, 但总体上, 高维和低维情形下的估计值表现基本相似. 在大样本

表 2　当 $n = 2{,}890$、$\rho_{ev} = 0.8$ 和 $\rho = 0.4$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$\hat{\rho}$	0.5085	0.5078	0.5066	0.4003	0.4002	0.4004	0.4002	0.4002	0.4003
		(0.0018)	(0.0016)	(0.0016)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0009)	(0.0009)
	$\hat{\beta}_1$	0.9410	0.9421	0.9413	0.9992	0.9998	0.9994	0.9989	0.9995	0.9990
		(0.0019)	(0.0018)	(0.0020)	(0.0021)	(0.0018)	(0.0018)	(0.0020)	(0.0018)	(0.0018)
	$\hat{\beta}_2$	0.9391	0.9407	0.9411	1.0001	0.9995	0.9995	0.9992	0.9988	0.9988
		(0.0018)	(0.0018)	(0.0021)	(0.0017)	(0.0015)	(0.0017)	(0.0017)	(0.0015)	(0.0017)
	$\hat{\beta}_3$	0.9415	0.9432	0.9433	1.0012	1.0022	1.0027	1.0005	1.0015	1.0022
		(0.0018)	(0.0017)	(0.0019)	(0.0018)	(0.0017)	(0.0018)	(0.0018)	(0.0016)	(0.0018)
	$\hat{\delta}$				0.7967	0.7959	0.7982	0.8023	0.8001	0.8030
					(0.0026)	(0.0027)	(0.0031)	(0.0027)	(0.0026)	(0.0029)
	$\hat{\Gamma}_1$				0.7967	0.7959	0.7982	0.8023	0.8001	0.8030
					(0.0021)	(0.0019)	(0.0018)	(0.0021)	(0.0018)	(0.0018)
	$\hat{\Gamma}_2$				0.7967	0.7959	0.7982	0.8023	0.8001	0.8030
					(0.0020)	(0.0017)	(0.0016)	(0.0020)	(0.0017)	(0.0016)
	$\hat{\Gamma}_3$				1.0020	1.0024	1.0030	1.0020	1.0024	1.0030
					(0.0018)	(0.0018)	(0.0020)	(0.0018)	(0.0018)	(0.0020)
$p = 100$	$\hat{\rho}$	0.5082	0.5077	0.5067	0.4005	0.4005	0.4004	0.4002	0.4002	0.4003
		(0.0018)	(0.0016)	(0.0016)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0009)	(0.0009)
	$\hat{\beta}_1$	0.9419	0.9427	0.9417	0.9995	1.0000	1.0000	0.9989	0.9995	0.9990
		(0.0019)	(0.0019)	(0.0020)	(0.0021)	(0.0018)	(0.0018)	(0.0020)	(0.0018)	(0.0018)
	$\hat{\beta}_2$	0.9396	0.9415	0.9415	1.0003	1.0000	1.0003	0.9992	0.9988	0.9988
		(0.0018)	(0.0019)	(0.0022)	(0.0018)	(0.0015)	(0.0017)	(0.0017)	(0.0015)	(0.0017)
	$\hat{\beta}_3$	0.9421	0.9438	0.9437	1.0015	1.0022	1.0031	1.0005	1.0015	1.0022
		(0.0019)	(0.0018)	(0.0019)	(0.0017)	(0.0016)	(0.0018)	(0.0018)	(0.0016)	(0.0018)
	$\hat{\delta}$				0.7945	0.7933	0.7943	0.8023	0.8001	0.8030
					(0.0028)	(0.0026)	(0.0031)	(0.0027)	(0.0026)	(0.0029)
	$\hat{\Gamma}_1$				0.9991	0.9990	0.9980	0.9991	0.9989	0.9980
					(0.0021)	(0.0019)	(0.0018)	(0.0021)	(0.0018)	(0.0018)
	$\hat{\Gamma}_2$				1.0004	1.0007	1.0012	1.0004	1.0007	1.0012
					(0.0020)	(0.0017)	(0.0016)	(0.0020)	(0.0017)	(0.0016)
	$\hat{\Gamma}_3$				1.0020	1.0024	1.0030	1.0020	1.0024	1.0030
					(0.0018)	(0.0018)	(0.0020)	(0.0018)	(0.0018)	(0.0020)

情形下, 从表 2 可以看出, 各种参数估计值的精度都有所提高, 说明增大样本量可以提高估计的精度.

表 3 和 4 分别给出了当 $n = 284, \rho_{ev} = 0.8$ 和 $n = 2{,}890$, $\rho_{ev} = 0.8$ 时使用两种估计方法得到的变量选择结果. 为评估稀疏恢复和变量选择的效果, 本文使用以下指标进行度量: (1) $FZ_{\boldsymbol{\beta}}$ (未知参数 $\boldsymbol{\beta}$ 估计中假零的比例); (2) $FN_{\boldsymbol{\beta}}$ (未知参数 $\boldsymbol{\beta}$ 估计中假的非零比例); (3) p_{β_4} (在最后一步中 $X_{4,1n}$ 被选中的概率); (4) $FZ_{\boldsymbol{\Gamma}}$ (未知参数 $\boldsymbol{\Gamma}$ 估计中假零的比例); (5) $FN_{\boldsymbol{\Gamma}}$ (未知参数 $\boldsymbol{\Gamma}$ 估计中假的非零比例); (6) p_{Γ_4} (在第 1 步中 $X_{4,2n}$ 被选中的概率). 从表 3 可以看出, 在不同的 Expectile 水平下, 外生方法和内生

表 3 当 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.4$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的变量选择结果

		外生方法			内生方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	0.86	0.11	0.93	0.98	0.1	1
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				0.9	0.07	0.89
$p = 100$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	0.43	0.04	0.50	0.73	0.04	0.78
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				0.49	0.03	0.59

表 4 当 $n = 2{,}890$、$\rho_{ev} = 0.8$ 和 $\rho = 0.4$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的变量选择结果

		外生方法			内生方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	1	0.02	1	1	0	1
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				1	0	1
$p = 100$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	1	0.02	1	1	0	1
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				1	0	1

方法都可以完全正确地选出均值 $\boldsymbol{\beta}$ 中的重要变量和非重要变量. 对于影响方差的重要变量 $X_{4,1n}$, 在 $\tau = 0.25$ 和 $\tau = 0.75$ 处, 外生方法选出 $X_{4,1n}$ 的比例达到 80% 以上, 内生方法可以 100% 选出 $X_{4,1n}$, 即在适当的 Expectile 水平下内生方法可有效地用于诊断异方差现象; 而在 $\tau = 0.5$ 处, 两种方法选出 $X_{4,1n}$ 的比例均较低, 约为 10%, 即仅用传统二次平方损失函数难以判断异方差现象; 虽然当 $p = 100$ 时, 在 $\tau = 0.25$ 和 $\tau = 0.75$ 处挑选出重要变量 $X_{4,1n}$ 的比例有下降的趋势, 但是对于均值中的重要变量, 所提出方法依然可以 100% 地挑选出重要变量和非重要变量. 进一步地, 从表 4 可以看出, 当样本量增加到 $n = 2{,}890$ 时, 外生方法和内生方法都可以将均值和方差中的重要变量完美地挑选出来. 不过, 外生方法只执行两步估计, 它无法给出指标 $FZ_{\boldsymbol{\Gamma}}$、$FN_{\boldsymbol{\Gamma}}$ 和 p_{Γ_4} 的值. 总体上, 当空间权重矩阵具有内生性时, 本文所提出的内生方法比外生方法具有更好的数值表现.

为观察空间效应的增强对估计和变量选择的影响, 表 5 和 6 分别给出了在 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.8$ 条件下 SCAD 估计与变量选择的结果. 不难发现, 在内生空间权重矩阵条件下, 当空间效应增强后, 外生方法的估计精度和变量选择效果会降低较多, 而内生方法影响相对较小. 可能的原因是空间效应的增大导致内生空间权重矩阵的内生性更强, 从而降低外生方法的应用效果.

表 7 和 8 分别给出了在 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.4$ 条件下自适应 LASSO 方法的参数估计与

表 5 当 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.8$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$\hat{\rho}$	0.8380	0.8553	0.8399	0.7988	0.7989	0.8001	0.7997	0.7995	0.8002
		(0.0029)	(0.0077)	(0.0025)	(0.0011)	(0.0011)	(0.0012)	(0.0010)	(0.0009)	(0.0010)
	$\hat{\beta}_1$	0.9863	0.9883	0.9839	1.0146	1.0092	1.0050	1.0052	1.0038	1.0035
		(0.0098)	(0.0102)	(0.0081)	(0.0071)	(0.0062)	(0.0075)	(0.0042)	(0.0037)	(0.0043)
	$\hat{\beta}_2$	0.9866	0.9872	0.9905	1.0125	1.0062	1.0109	1.0006	1.0002	0.9999
		(0.0093)	(0.0100)	(0.0101)	(0.0079)	(0.0073)	(0.0081)	(0.0046)	(0.0040)	(0.0044)
	$\hat{\beta}_3$	0.9913	0.9940	0.9811	1.0126	1.0045	1.0095	1.0042	1.0032	1.0016
		(0.0102)	(0.0123)	(0.0087)	(0.0088)	(0.0078)	(0.0085)	(0.0047)	(0.0042)	(0.0043)
	$\hat{\delta}$				0.7319	0.7440	0.7310	0.7979	0.7944	0.7925
					(0.0143)	(0.0125)	(0.0140)	(0.0056)	(0.0055)	(0.0067)
	$\hat{\Gamma}_1$				1.0003	0.9943	0.9970	1.0017	1.0003	0.9997
					(0.0043)	(0.0038)	(0.0044)	(0.0040)	(0.0038)	(0.0042)
	$\hat{\Gamma}_2$				0.9984	0.9937	0.9992	0.9998	1.0002	1.0007
					(0.0044)	(0.0041)	(0.0042)	(0.0044)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0039	0.9981	1.0000	1.0048	1.0032	1.0028
					(0.0043)	(0.0038)	(0.0041)	(0.0044)	(0.0040)	(0.0041)
$p = 100$	$\hat{\rho}$	0.8294	0.8688	0.8362	0.8024	0.8027	0.8021	0.7997	0.7995	0.8002
		(0.0066)	(0.0117)	(0.0106)	(0.0013)	(0.0012)	(0.0011)	(0.0010)	(0.0009)	(0.0010)
	$\hat{\beta}_1$	1.1850	1.1817	1.1816	1.0139	1.0078	1.0095	1.0052	1.0038	1.0035
		(0.0116)	(0.0118)	(0.0111)	(0.0082)	(0.0076)	(0.0087)	(0.0042)	(0.0037)	(0.0043)
	$\hat{\beta}_2$	1.1777	1.1796	1.1754	1.0039	1.0016	1.0079	1.0006	1.0002	0.9999
		(0.0124)	(0.0118)	(0.0142)	(0.0087)	(0.0083)	(0.0090)	(0.0046)	(0.0040)	(0.0044)
	$\hat{\beta}_3$	1.1778	1.1670	1.1751	1.0113	1.0039	1.0037	1.0042	1.0032	1.0016
		(0.0150)	(0.0137)	(0.0150)	(0.0102)	(0.0091)	(0.0093)	(0.0047)	(0.0042)	(0.0043)
	$\hat{\delta}$				0.7160	0.7236	0.7208	0.7979	0.7944	0.7925
					(0.0154)	(0.0137)	(0.0150)	(0.0056)	(0.0055)	(0.0067)
	$\hat{\Gamma}_1$				0.9981	0.9934	0.9958	1.0017	1.0003	0.9997
					(0.0047)	(0.0038)	(0.0043)	(0.0040)	(0.0038)	(0.0042)
	$\hat{\Gamma}_2$				0.9964	0.9926	0.9969	0.9998	1.0002	1.0007
					(0.0045)	(0.0041)	(0.0043)	(0.0044)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0024	0.9972	0.9995	1.0048	1.0032	1.0028
					(0.0046)	(0.0038)	(0.0040)	(0.0044)	(0.0040)	(0.0041)

变量选择结果. 可以看出, 在自适应 LASSO 惩罚函数下, 本文所提出的方法在不同的 Expectile 水平下仍可达到非常精确的估计结果, 并有效地实现变量选择的目的. 不过, 从估计精度与变量选择的效果上看, SCAD 惩罚方法稍优于自适应 LASSO 惩罚方法; 从当前算法的运算时间上看 (为优化本文布局未给出), 自适应 LASSO 惩罚方法明显优于 SCAD 惩罚方法, 平均时间可节省 50% 以上. 因此, 两种方法各有优劣, 在应用上可依据不同条件选择使用.

表 6　当 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.8$ 时, 内生空间权重矩阵条件下 SCAD 惩罚的变量选择结果

		外生方法			内生方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	FZ_β	1	1	1	1	1	1
	FN_β	0.02	0.01	0.01	0	0	0
	p_{β_4}	0.43	0.22	0.40	0.68	0.18	0.62
	FZ_γ				1	1	1
	FN_γ				0	0	0
	p_{γ_4}				0.90	0.07	0.89
$p = 100$	FZ_β	1	1	1	1	1	1
	FN_β	0	0	0	0	0	0
	p_{β_4}	0.11	0.06	0.13	0.30	0.07	0.30
	FZ_γ				1	1	1
	FN_γ				0	0	0
	p_{γ_4}				0.50	0.03	0.59

表 7　当 $n = 284$、$\rho_{ev} = 0.8$ 和 $\rho = 0.4$ 时, 内生空间权重矩阵条件下自适应 LASSO 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$\hat{\rho}$	0.4907	0.4895	0.4879	0.4012	0.3997	0.4013	0.4015	0.3997	0.4018
		(0.0045)	(0.0041)	(0.0041)	(0.0021)	(0.0020)	(0.0022)	(0.0023)	(0.0022)	(0.0024)
	$\hat{\beta}_1$	0.9623	0.9646	0.9618	0.9975	0.9936	0.9945	0.9970	0.9933	0.9946
		(0.0048)	(0.0043)	(0.0046)	(0.0045)	(0.0037)	(0.0043)	(0.0044)	(0.0038)	(0.0044)
	$\hat{\beta}_2$	0.9585	0.9620	0.9627	0.9949	0.9918	0.9933	0.9941	0.9910	0.9927
		(0.0049)	(0.0046)	(0.0049)	(0.0051)	(0.0044)	(0.0046)	(0.0052)	(0.0046)	(0.0047)
	$\hat{\beta}_3$	0.9631	0.9643	0.9622	0.9985	0.9929	0.9936	0.9982	0.9928	0.9937
		(0.0052)	(0.0048)	(0.0050)	(0.0050)	(0.0041)	(0.0045)	(0.0052)	(0.0043)	(0.0048)
	$\hat{\delta}$				0.7820	0.7822	0.7761	0.7825	0.7828	0.7768
					(0.0061)	(0.0059)	(0.0070)	(0.0061)	(0.0059)	(0.0069)
	$\hat{\Gamma}_1$				0.9956	0.9877	0.9901	0.9958	0.9876	0.9908
					(0.0046)	(0.0036)	(0.0041)	(0.0045)	(0.0037)	(0.0041)
	$\hat{\Gamma}_2$				0.9931	0.9877	0.9918	0.9934	0.9877	0.9924
					(0.0047)	(0.0038)	(0.0045)	(0.0046)	(0.0038)	(0.0045)
	$\hat{\Gamma}_3$				0.9989	0.9910	0.9935	0.9992	0.9909	0.9940
					(0.0044)	(0.0034)	(0.0040)	(0.0044)	(0.0034)	(0.0041)

(续表)

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p=100$	$\hat{\rho}$	0.4863	0.4834	0.4822	0.3938	0.4003	0.3892	0.4015	0.3997	0.4018
		(0.0056)	(0.0048)	(0.0049)	(0.0060)	(0.0020)	(0.0072)	(0.0023)	(0.0022)	(0.0024)
	$\hat{\beta}_1$	1.0096	1.0054	1.0057	0.9981	0.9931	0.9845	0.9970	0.9933	0.9946
		(0.0044)	(0.0041)	(0.0047)	(0.0049)	(0.0037)	(0.0111)	(0.0044)	(0.0038)	(0.0044)
	$\hat{\beta}_2$	1.0058	1.0044	1.0069	0.9928	0.9910	0.9826	0.9941	0.9910	0.9927
		(0.0049)	(0.0045)	(0.0050)	(0.0060)	(0.0044)	(0.0110)	(0.0052)	(0.0046)	(0.0047)
	$\hat{\beta}_3$	1.0097	1.0056	1.0059	0.9983	0.9919	0.9821	0.9982	0.9928	0.9937
		(0.0054)	(0.0048)	(0.0051)	(0.0058)	(0.0040)	(0.0110)	(0.0052)	(0.0043)	(0.0048)
	$\hat{\delta}$				0.7876	0.7833	0.7747	0.7825	0.7828	0.7768
					(0.0070)	(0.0059)	(0.0108)	(0.0061)	(0.0059)	(0.0069)
	$\hat{\Gamma}_1$				0.9929	0.9873	0.9873	0.9958	0.9876	0.9908
					(0.0047)	(0.0036)	(0.0041)	(0.0045)	(0.0037)	(0.0041)
	$\hat{\Gamma}_2$				0.9890	0.9871	0.9885	0.9934	0.9877	0.9924
					(0.0047)	(0.0039)	(0.0046)	(0.0046)	(0.0038)	(0.0045)
	$\hat{\Gamma}_3$				0.9964	0.9905	0.9905	0.9992	0.9909	0.9940
					(0.0044)	(0.0034)	(0.0041)	(0.0044)	(0.0034)	(0.0041)

表 8 当 $n=284$、$\rho_{ev}=0.8$ 和 $\rho=0.4$ 时, 内生空间权重矩阵条件下自适应 LASSO 惩罚的变量选择结果

		外生方法			内生方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75
$p=10$	FZ_β	1	1	1	1	1	1
	FN_β	0	0	0	0	0	0
	p_{β_4}	0.86	0.11	0.93	0.79	0	0.83
	FZ_γ				1	1	1
	FN_γ				0	0	0
	p_{γ_4}				0.78	0.01	0.83
$p=100$	FZ_β	1	1	1	1	1	1
	FN_β	0	0	0	0	0	0
	p_{β_4}	0.43	0.04	0.50	0.40	0.01	0.47
	FZ_γ				1	1	1
	FN_γ				0	0	0
	p_{γ_4}				0.41	0.02	0.48

表 9 和 10 给出当 $n=284$, $\rho_{ev}=0$ 时参数估计和变量选择的结果. 在 $\rho_{ev}=0$ 的条件下, 空间权重矩阵的内生性将消失, 这时外生方法和内生方法在理论上应具有相同的效果. 从表 9 可以看出, 在各种情形下, 本文所提出的三步内生方法和两步外生方法的数值估计结果具有高度可比性. 从表 10 可以看出, 内生方法和外生方法在不同 Expectile 水平下的变量选择结果基本相似. 对于均值中的重要变量, 内生方法和外生方法变量选择的结果完全相同. 对于方差中的重要变量, 在 $\tau=0.25$ 和 $\tau=0.75$

表 9　当 $n = 284$、$\rho_{ev} = 0$ 和 $\rho = 0.4$ 时, 外生空间权重矩阵条件下 SCAD 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p=10$	$\hat{\rho}$	0.3991	0.3991	0.3961	0.3966	0.3970	0.3965	0.3949	0.3953	0.3950
		(0.0038)	(0.0034)	(0.0037)	(0.0035)	(0.0033)	(0.0036)	(0.0036)	(0.0035)	(0.0037)
	$\hat{\beta}_1$	1.0105	1.0106	1.0109	1.0091	1.0114	1.0101	1.0054	1.0072	1.0067
		(0.0043)	(0.0038)	(0.0043)	(0.0043)	(0.0039)	(0.0044)	(0.0043)	(0.0040)	(0.0043)
	$\hat{\beta}_2$	1.0040	1.0050	1.0084	1.0036	1.0045	1.0071	1.0005	1.0015	1.0034
		(0.0045)	(0.0042)	(0.0043)	(0.0045)	(0.0041)	(0.0044)	(0.0045)	(0.0041)	(0.0043)
	$\hat{\beta}_3$	1.0076	1.0059	1.0079	1.0064	1.0066	1.0061	1.0034	1.0037	1.0027
		(0.0051)	(0.0044)	(0.0049)	(0.0052)	(0.0047)	(0.0049)	(0.0048)	(0.0044)	(0.0046)
	$\hat{\delta}$				-0.0330	-0.0345	-0.0414	0.0004	-0.0053	-0.0099
					(0.0084)	(0.0075)	(0.0086)	(0.0078)	(0.0072)	(0.0080)
	$\hat{\Gamma}_1$				0.9978	0.9926	0.9958	0.9989	0.9979	0.9977
					(0.0045)	(0.0038)	(0.0044)	(0.0042)	(0.0040)	(0.0042)
	$\hat{\Gamma}_2$				0.9987	0.9949	0.9994	1.0001	1.0005	0.9996
					(0.0046)	(0.0040)	(0.0043)	(0.0045)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0023	0.9986	1.0002	1.0037	1.0030	1.0030
					(0.0042)	(0.0037)	(0.0042)	(0.0041)	(0.0038)	(0.0042)
$p=100$	$\hat{\rho}$	0.4000	0.3990	0.3961	0.3983	0.3970	0.3964	0.3949	0.3953	0.3950
		(0.0048)	(0.0040)	(0.0042)	(0.0037)	(0.0034)	(0.0036)	(0.0036)	(0.0035)	(0.0037)
	$\hat{\beta}_1$	1.0371	1.0362	1.0360	1.0123	1.0135	1.0130	1.0054	1.0072	1.0067
		(0.0043)	(0.0039)	(0.0046)	(0.0046)	(0.0042)	(0.0048)	(0.0043)	(0.0040)	(0.0043)
	$\hat{\beta}_2$	1.0312	1.0307	1.0334	1.0071	1.0086	1.0095	1.0005	1.0015	1.0034
		(0.0047)	(0.0043)	(0.0044)	(0.0048)	(0.0043)	(0.0047)	(0.0045)	(0.0041)	(0.0043)
	$\hat{\beta}_3$	1.0326	1.0310	1.0319	1.0074	1.0081	1.0077	1.0034	1.0037	1.0027
		(0.0053)	(0.0048)	(0.0050)	(0.0053)	(0.0047)	(0.0050)	(0.0048)	(0.0044)	(0.0046)
	$\hat{\delta}$				-0.0586	-0.0540	-0.0675	0.0004	-0.0053	-0.0099
					(0.0088)	(0.0077)	(0.0089)	(0.0078)	(0.0072)	(0.0080)
	$\hat{\Gamma}_1$				0.9973	0.9923	0.9954	0.9989	0.9979	0.9977
					(0.0046)	(0.0040)	(0.0044)	(0.0042)	(0.0040)	(0.0042)
	$\hat{\Gamma}_2$				0.9983	0.9947	0.9973	1.0001	1.0005	0.9996
					(0.0047)	(0.0040)	(0.0046)	(0.0045)	(0.0041)	(0.0043)
	$\hat{\Gamma}_3$				1.0015	0.9983	1.0003	1.0037	1.0030	1.0030
					(0.0043)	(0.0037)	(0.0042)	(0.0041)	(0.0038)	(0.0042)

下, 两种方法都可以相当高的比例将影响方差的重要变量挑选出来, 从而说明数据中存在异方差的现象. 表 11 和 12 给出了当 $n = 2{,}890$、$\rho_{ev} = 0$ 时两种方法的参数估计和变量选择结果. 与表 9 和 10 相比, 由于样本量变大, 因此, 表 11 和 12 中的估计结果与变量选择精度将显著提高, 特别地, 在 $\tau = 0.25$ 和 $\tau = 0.75$ 处, 可以以 100% 的概率挑选出影响方差的重要变量. 因此, 外生空间权重矩阵下, 内生方法仍具有优良表现.

综上可以看出, 本文所提出的外生方法可以较好地处理外生空间权重矩阵下的参数估计与变量选择问题, 同时, 所提出的内生方法不仅适用于内生空间权重矩阵下的情形, 也适用于外生空间权重矩阵下的情形. 当空间权重矩阵外生时, 内生方法和外生方法具有可比性; 当空间权重矩阵内生时, 与外生方法相比, 内生方法估计精度更高, 变量选择结果更精确. 进一步地, 本文所提出的方法不仅可用于处理高维数据, 而且还可用于诊断数据中的异方差现象.

表 10 当 $n = 284$、$\rho_{ev} = 0$ 和 $\rho = 0.4$ 时, 外生空间权重矩阵条件下 SCAD 惩罚的变量选择结果

		外生方法			内生方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	0.82	0.13	0.83	0.85	0.11	0.84
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				0.83	0.06	0.86
$p = 100$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	0.45	0.03	0.46	0.42	0.02	0.40
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				0.46	0	0.56

表 11 当 $n = 2{,}890$、$\rho_{ev} = 0$ 和 $\rho = 0.4$ 时, 外生空间权重矩阵条件下 SCAD 惩罚的估计结果

		外生方法			内生方法			Oracle 方法		
	τ	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$p = 10$	$\hat{\rho}$	0.3985	0.3989	0.3980	0.3987	0.3988	0.3979	0.3985	0.3988	0.3977
		(0.0017)	(0.0016)	(0.0016)	(0.0017)	(0.0016)	(0.0017)	(0.0017)	(0.0016)	(0.0017)
	$\hat{\beta}_1$	0.9994	1.0007	1.0017	1.0000	1.0008	1.0020	0.9997	1.0007	1.0018
		(0.0018)	(0.0017)	(0.0019)	(0.0019)	(0.0018)	(0.0020)	(0.0019)	(0.0017)	(0.0019)
	$\hat{\beta}_2$	0.9986	0.9984	0.9985	0.9993	0.9985	0.9986	0.9989	0.9984	0.9986
		(0.0017)	(0.0016)	(0.0020)	(0.0016)	(0.0017)	(0.0020)	(0.0017)	(0.0016)	(0.0020)
	$\hat{\beta}_3$	1.0000	1.0012	1.0017	1.0007	1.0013	1.0020	1.0002	1.0012	1.0019
		(0.0018)	(0.0017)	(0.0018)	(0.0018)	(0.0017)	(0.0017)	(0.0019)	(0.0017)	(0.0018)
	$\hat{\delta}$				-0.0030	0.0004	-0.0003	0.0007	0.0009	0.0016
					(0.0050)	(0.0039)	(0.0043)	(0.0042)	(0.0039)	(0.0040)
	$\hat{\Gamma}_1$				0.9984	0.9988	0.9985	0.9985	0.9987	0.9985
					(0.0023)	(0.0018)	(0.0017)	(0.0022)	(0.0018)	(0.0017)
	$\hat{\Gamma}_2$				1.0017	1.0018	1.0022	1.0018	1.0018	1.0022
					(0.0020)	(0.0016)	(0.0015)	(0.0020)	(0.0016)	(0.0015)
	$\hat{\Gamma}_3$				1.0020	1.0023	1.0032	1.0020	1.0023	1.0031
					(0.0019)	(0.0019)	(0.0021)	(0.0019)	(0.0019)	(0.0021)

(续表)

	τ	外生方法 0.25	外生方法 0.50	外生方法 0.75	内生方法 0.25	内生方法 0.50	内生方法 0.75	Oracle 方法 0.25	Oracle 方法 0.50	Oracle 方法 0.75
$p=100$	$\hat{\rho}$	0.4039	0.4011	0.4034	0.3990	0.3993	0.3982	0.3985	0.3988	0.3977
		(0.0018)	(0.0016)	(0.0016)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0016)	(0.0017)
	$\hat{\beta}_1$	1.0027	1.0009	1.0046	0.9997	1.0006	1.0017	0.9997	1.0007	1.0018
		(0.0021)	(0.0017)	(0.0023)	(0.0019)	(0.0018)	(0.0020)	(0.0019)	(0.0017)	(0.0019)
	$\hat{\beta}_2$	1.0019	0.9985	1.0017	0.9992	0.9982	0.9985	0.9989	0.9984	0.9986
		(0.0018)	(0.0016)	(0.0025)	(0.0016)	(0.0017)	(0.0020)	(0.0017)	(0.0016)	(0.0020)
	$\hat{\beta}_3$	1.0033	1.0016	1.0051	1.0005	1.0010	1.0018	1.0002	1.0012	1.0019
		(0.0021)	(0.0017)	(0.0021)	(0.0018)	(0.0017)	(0.0017)	(0.0019)	(0.0017)	(0.0018)
	$\hat{\delta}$				−0.0026	0.0002	0.0004	0.0007	0.0009	0.0016
					(0.0051)	(0.0039)	(0.0041)	(0.0042)	(0.0039)	(0.0040)
	$\hat{\Gamma}_1$				0.9985	0.9987	0.9986	0.9985	0.9987	0.9985
					(0.0022)	(0.0018)	(0.0017)	(0.0022)	(0.0018)	(0.0017)
	$\hat{\Gamma}_2$				1.0019	1.0018	1.0022	1.0018	1.0018	1.0022
					(0.0020)	(0.0016)	(0.0015)	(0.0020)	(0.0016)	(0.0015)
	$\hat{\Gamma}_3$				1.0020	1.0024	1.0031	1.0020	1.0023	1.0031
					(0.0019)	(0.0019)	(0.0021)	(0.0019)	(0.0019)	(0.0021)

表 12　当 $n=2{,}890$、$\rho_{ev}=0$ 和 $\rho=0.4$ 时, 外生空间权重矩阵条件下 SCAD 惩罚的变量选择结果

	τ	外生方法 0.25	外生方法 0.50	外生方法 0.75	内生方法 0.25	内生方法 0.50	内生方法 0.75
$p=10$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	1	0	1	1	0	1
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				1	0	1
$p=100$	$FZ_{\boldsymbol{\beta}}$	1	1	1	1	1	1
	$FN_{\boldsymbol{\beta}}$	0	0	0	0	0	0
	p_{β_4}	1	0	1	1	0	1
	$FZ_{\boldsymbol{\Gamma}}$				1	1	1
	$FN_{\boldsymbol{\Gamma}}$				0	0	0
	p_{Γ_4}				1	0	1

5　实例分析

本节研究空气质量与地区经济发展之间的关系, 响应变量选取为废气污染物排放量, 解释变量包括一般公共预算收入、一般公共预算支出、城市建成区面积、人口数、人口密度、工业企业数、人均绿地面积、就业人数、社会消费品零售总额、普通高中数量、普通初中数量、普通小学数量、医疗卫生机

构数量、教育支出、平均工资、外国直接投资、第二产业占地区生产总值的比重和污水处理率等. 本文分析的数据为 2021 年我国 284 个地级及以上城市的数据, 数据的主要来源为《中国城市统计年鉴 (2021 年)》以及国家统计局和地方统计局的官方网站, 极少数缺失数据采用线性插值法和平均趋势法补全.

依据 Newey 和 Powell [23] 提供的异质性检验方法, 并考虑到检验统计量渐近分布估计的困难, 本文基于 10,000 次 Monte Carlo 模拟给出了响应变量分布异质性检验的结果, 其中, 卡方检验统计量的值为 149.80, 在显著性水平 0.05 的条件下临界值为 84.60, 因此可以认为响应变量的分布存在异质性, 故以下采用稳健的 Expectile 回归方法对该数据进行统计分析.

为揭示和探索空间数据的异质性和空间权重矩阵的内生性, 并选取重要的解释变量, 本文采用外生和内生两种高维 Expectile 方法对实际数据进行分析, 分位数 τ 分别取为 0.25、0.50 和 0.75. 在外生方法中, 空间权重矩阵取为 Rook 矩阵; 在内生方法中, 空间权重矩阵与数值模拟中的取法一致, 其中变量 Z_n 取为 2021 年每个城市的 GDP, 同时由于重要的解释变量未知, 且为了避免选择的主观性, 故 X_{2n} 取遍所有的解释变量, 即设定 $X_{2n} = X_{1n}$, 运行结果总结如表 13 所示.

从表 13 可以得到以下结论:

(1) 在内生性识别方面, 根据本文所提出的惩罚 Expectile 估计方法, 通过参数 δ 的估计值来识别, 不难看出, 在不同 Expectile 水平下, $\hat{\delta}$ 均显著非零. 因此, 可以认为空间权重矩阵存在内生性.

表 13　空气质量与地区经济发展关系的 Expectile 回归分析结果

	外生方法			内生方法		
τ	0.25	0.50	0.75	0.25	0.50	0.75
$\hat{\delta}$				10.8843	21.0928	37.0697
$\hat{\rho}$	0.1554	0.1742	0.4736	−0.0894	−0.1493	0.0967
一般公共预算收入	0	0	0	0	0	0
一般公共预算支出	8.611	9.8602	14.1894	5.21	8.3311	14.2507
城市建成区面积	0	0	0	0	0	0
人口数	8.3687	9.1626	17.355	6.0428	10.0122	21.4879
人口密度	0	0	0	0	0	0
工业企业数	0	0	−3.8554	0	0	−1.9102
人均绿地面积	0	0	0	0	0	0
就业人数	−3.2429	−4.1976	−4.8292	−3.8685	−4.6695	−6.4621
社会消费品零售总额	−9.2051	−9.8799	−6.4453	−9.0575	−9.3018	−6.6295
普通高中数量	0	0	0	0	0	0
普通初中数量	−4.959	−4.9668	−4.9497	−4.7901	−5.2822	−6.3262
普通小学数量	0	0	0	0	0	0
医疗卫生机构数量	6.7134	7.5952	0	8.6426	7.845	0
教育支出	−4.357	−5.0645	−10.009	0	−3.7912	−10.8649
平均工资	0	0	0	0	0	0
外国直接投资	0	0	0	0	0	0
第二产业占地区生产总值的比重	3.8912	5.5083	8.0038	3.8079	5.5963	8.0914
污水处理率	0	0	0	0	0	0

(2) 在空间相关性方面, 由于不同城市间的相互影响方向和程度主要由空间效应系数 ρ 决定, 因此我们可依据其符号和大小进行分析. 对于外生方法, 参数 ρ 的估计皆为正数, 这反映出在不同 Expectile 水平下, 相邻城市的影响均具有正向作用. 然而, 对于内生方法, 在不同 Expectile 水平下, 参数 ρ 的估计符号并不相同, 具体而言, 在高 Expectile 水平处, 内生方法的参数估计 $\hat{\rho} > 0$, 这意味着在高污染地区, 相邻城市对本地区的污染排放量具有正向影响, 即相邻城市会加剧本地的空气污染程度; 在中低 Expectile 水平处, 内生方法的参数估计 $\hat{\rho} < 0$, 这意味着在低度污染或中等污染的地区, 相邻城市对本地区的污染排放量具有负向影响, 即相邻城市会减轻本地区的空气污染程度.

(3) 在异质性诊断方面, 当 $\tau = 0.5$ 时, 工业企业数这个变量的系数估计为 0, 即在中 Expectile 水平处, 工业企业数对污染量的均值影响不显著; 然而, 当 $\tau = 0.75$ 时, 工业企业数的系数显著不为 0, 因此, 可以推断出工业企业数主要影响空气污染物的方差, 即数据中存在异质性. 特别地, 当 $\tau = 0.75$ 时, 工业企业数的系数显著不为 0, 且系数的绝对值大于 1, 故说明在高 Expectile 水平处, 工业企业数越多, 空气污染物浓度的波动会越大.

(4) 在变量选取方面, 对不同的 Expectile 水平, 外生和内生两种方法选取出的重要变量集基本相同, 特别需要注意的是, 医疗卫生机构数的影响只表现在中低 Expectile 水平上, 在高 Expectile 水平上, 医疗卫生机构数的影响不显著. 在选取出的重要变量中, 一般公共预算支出、人口数、医疗卫生机构数和第二产业占地区生产总值的比重对响应变量具有正向的影响效应, 即城市一般公共预算支出、人口数和医疗卫生机构数越多, 第二产业占地区生产总值的比重越大, 城市的废气污染物排放量一般也会越多. 这也与我国当前的实际情形基本吻合, 特别地, 对于人口数量多的城市, 为了发展经济、增加就业, 第二产业发展力度一般会较大, 从而造成城市废气污染物排放量增多. 此外, 在选取出的重要变量中, 就业人数、普通中学数量和教育支出对响应变量具有负向的影响效应, 即城市就业人数和普通中学的数量越多, 教育支出越大, 城市的废气污染物排放量会相对减少, 这间接反映出随着公众文明素养的提高, 城市经济发展更注重发展质量.

致谢 感谢编委会和审稿人的认真评阅与宝贵意见, 感谢江西财经大学“共同富裕的统计测度与进程监测”团队支持.

参考文献

1 Aneiros G, Novo S, Vieu P. Variable selection in functional regression models: A review. J Multivariate Anal, 2022, 188: 104871–104884

2 Barry A, Oualkacha K, Charpentier A. Weighted asymmetric least squares regression for longitudinal data using GEE. arXiv:1810.09214, 2018

3 Barry A, Oualkacha K, Charpentier A. Weighted asymmetric least squares regression with fixed-effects. arXiv:2108.04737, 2021

4 Ciuperca G. Variable selection in high-dimensional linear model with possibly asymmetric errors. Comput Statist Data Anal, 2021, 155: 107112

5 Cliff A D, Ord J K. Spatial Autocorrelation. London: Pion, 1973

6 Cliff A D, Ord J K. Spatial Processes: Models and Applications. London: Pion, 1981

7 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statist Assoc, 2001, 96: 1348–1360

8 Fan Y, Tang Y, Zhu Z. Variable selection in censored quantile regression with high dimensional data. Sci China Math, 2018, 61: 641–658

9 Gerlach R, Wang C. Bayesian semi-parametric realized conditional autoregressive expectile models for tail risk forecasting. J Financ Econ, 2022, 20: 105–138

10 Girard S, Stupfler G, Usseglio-Carleve A. Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. Ann Statist, 2021, 49: 3358–3382

11 Gu Y, Zou H. High-dimensional generalizations of asymmetric least squares regression and their applications. Ann Statist, 2016, 44: 2661–2694

12 Hu Z Y, Li Y, Wan C. Evaluating VaR and ES based on the Bayesian GARCH-Expectile model (in Chinese). J Appl Statist Manag, 2020, 39: 467–477 [胡宗义, 李毅, 万闯. 基于贝叶斯 GARCH-Expectile 模型的 VaR 和 ES 风险度量. 数理统计与管理, 2020, 39: 467–477]
13 Jenish N, Prucha I R. On spatial processes and asymptotic inference under near-epoch dependence. J Econometrics, 2012, 170: 178–190
14 Jones M C. Expectiles and M-quantiles are quantiles. Statist Probab Lett, 1994, 20: 149–153
15 Kelejian H H, Piras G. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. Regional Sci Urban Economics, 2014, 46: 140–149
16 Lee L F. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica, 2004, 72: 1899–1925
17 Li R, Liang H. Variable selection in semiparametric regression modeling. Ann Statist, 2008, 36: 261–286
18 Liao L, Park C, Choi H. Penalized expectile regression: An alternative to penalized quantile regression. Ann Inst Statist Math, 2019, 71: 409–438
19 Liu X, Chen J B. Variable selection of the spatial autoregressive quantile model with fixed effects (in Chinese). Acta Math Sinica (Chin Ser), 2023, 66: 405–424 [刘宣, 陈建宝. 固定效应空间自回归分位数模型的变量选择. 数学学报 (中文版), 2023, 66: 405–424]
20 Liu X, Chen J B, Cheng S L. A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. Spatial Stat, 2018, 25: 86–104
21 McMillen D P. Quantile Regression for Spatial Data. Berlin: Springer-Verlag, 2012
22 Mohammedi M, Bouzebda S, Laksaci A. The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. J Multivariate Anal, 2021, 181: 104673
23 Newey W K, Powell J L. Asymmetric least squares estimation and testing. Econometrica, 1987, 55: 819–847
24 Qu X, Lee L F. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. J Econometrics, 2015, 184: 209–232
25 Qu X, Lee L F, Yang C. Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables. J Econometrics, 2021, 221: 180–197
26 Steel M F J. Model averaging and its use in economics. J Economic Literature, 2020, 58: 644–719
27 Tao C Q, Xu Y T. Study on Bayesian adaptive Lasso quantile regression using asymmetric exponential power distribution for panel data (in Chinese). Statist Res, 2022, 39: 128–144 [陶长琪, 徐玉婷. 面板数据贝叶斯自适应 Lasso 分位数回归—基于非对称指数幂分布的研究. 统计研究, 2022, 39: 128–144]
28 Taylor J W. Estimating value at risk and expected shortfall using expectiles. J Financ Econ, 2007, 6: 231–252
29 Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B Stat Methodol, 1996, 58: 267–288
30 Wu Y, Liu Y. Variable selection in quantile regression. Stat Sinica, 2009, 19: 801–817
31 Xiao J, Yu P, Song X, et al. Statistical inference in the partial functional linear expectile regression model. Sci China Math, 2022, 65: 2601–2630
32 Xie S Y, Yao H W, Zhou Y. VaR and ES measurements based on ARCH-Expectile model (in Chinese). Chinese J Manag Sci, 2014, 22: 1–9 [谢尚宇, 姚宏伟, 周勇. 基于 ARCH-Expectile 方法的 VaR 和 ES 尾部风险测量. 中国管理科学, 2014, 22: 1–9]
33 Yao Q, Tong H. Asymmetric least squares regression estimation: A nonparametric approach. J Nonparametr Stat, 1996, 6: 273–292
34 Yu K, Jones M C. Local linear quantile regression. J Amer Statist Assoc, 1998, 93: 228–237
35 Zou H. The adaptive Lasso and its oracle properties. J Amer Statist Assoc, 2006, 101: 1418–1429

附录 A

附录 A.1 相关引理及证明

在模型 (3.2) 下, 当 $\boldsymbol{\delta} = \mathbf{0}$ 时, 内生空间权重矩阵条件下的 Expectile 惩罚损失函数将退化为外生空间权重矩阵情形下的目标函数, 且外生空间权重矩阵在行和与列和一致有界的条件下满足假设 3.8(1) 中内生权重矩阵的假设. 因而其证明思路大同小异. 为简洁起见, 以下只给出内生空间权重矩阵情形下的证明过程.

借鉴近邻相依 (near-epoch dependence, NED) 的概念与相关概率极限定理[13,24], 接下来给出大样本性质证明所需要的若干引理.

记 $\mathscr{F}_{i,n}(s)$ 为由位于球域 $\boldsymbol{B}_i(s)$ 的随机向量 $\boldsymbol{\zeta}_{j,n}(s)$ 产生的 σ 域, $\boldsymbol{X}^{\rm ned}=\{\boldsymbol{X}_{i,n}^{\rm ned}, l(i)\in\boldsymbol{D}_n, n\geqslant 1\}$ 和 $\boldsymbol{\zeta}=\{\boldsymbol{\zeta}_{i,n}, l(i)\in\boldsymbol{D}_n, n\geqslant 1\}$ 为随机域, $\boldsymbol{d}=\{d_{i,n}, l(i)\in\boldsymbol{D}_n, n\geqslant 1\}$ 为有限正常数阵列, $\boldsymbol{D}$ 为 d_0 维的 Euclid 空间.

定义 A.1 设 $\|\boldsymbol{X}_{i,n}^{\rm ned}\|_k<\infty$ $(k\geqslant 1)$, $\boldsymbol{D}_n\subset\boldsymbol{D}$ 且其势为 n. 若对于满足 $\lim_{s\to\infty}g(s)=0$ 的函数 $g(s)$, 有 $\|\boldsymbol{X}_{i,n}^{\rm ned}-{\rm E}(\boldsymbol{X}_{i,n}^{\rm ned}\mid F_{i,n}(s))\|_k\leqslant d_{i,n}g(s)$, 则称随机域 $\boldsymbol{X}^{\rm ned}$ 是关于随机域 $\boldsymbol{\zeta}$ 的 L_k-NED. 不失一般性, 假定 $g(s)$ 是非增的, 并称其为 NED 系数, $d_{i,n}$ 称为 NED 尺度因子. 若对某常数 a 和 b $(0<a<b)$, 有 $g(s)=O(s^{-b})$, 则称 $\boldsymbol{X}^{\rm ned}$ 是在 $-a$ 水平下关于随机域 $\boldsymbol{\zeta}$ 的 NED. 若 $\sup_n\sup_{l(i)\in D_n}d_{i,n}<\infty$, 则称 $\boldsymbol{X}^{\rm ned}$ 一致地 L_k-NED 于随机域 $\boldsymbol{\zeta}$.

令 $\boldsymbol{\zeta}_{i,n}=(e_{in},\boldsymbol{v}_{i,n}^{\rm T})^{\rm T}$, $\boldsymbol{X}_n=(\boldsymbol{X}_{1n},\boldsymbol{X}_{2n})$, $\boldsymbol{\zeta}_{i,1n}^*(\boldsymbol{\vartheta})=\boldsymbol{f}_{1i}(\boldsymbol{\zeta}_{i,n},\boldsymbol{X}_n,\boldsymbol{Q}_n,\boldsymbol{\vartheta})$, $\boldsymbol{\zeta}_{i,1n}^*=\boldsymbol{f}_{1i}(\boldsymbol{\zeta}_{i,n},\boldsymbol{X}_n,\boldsymbol{Q}_n,\boldsymbol{\vartheta}_0)$, $\boldsymbol{\zeta}_{i,2n}^*(\boldsymbol{\vartheta})=\boldsymbol{f}_{2i}(\boldsymbol{\zeta}_{i,n},\boldsymbol{X}_n,\boldsymbol{Q}_n,\boldsymbol{\vartheta})$, $\boldsymbol{\zeta}_{i,2n}^*=\boldsymbol{f}_{2i}(\boldsymbol{\zeta}_{i,n},\boldsymbol{X}_n,\boldsymbol{Q}_n,\boldsymbol{\vartheta}_0)$, $\boldsymbol{\zeta}_{kn}^*=(\boldsymbol{\zeta}_{i,kn}^*,\ldots,\boldsymbol{\zeta}_{i,kn}^*)^{\rm T}$, $k=1,2$, 其中 $\boldsymbol{f}_{1i}$ 和 $\boldsymbol{f}_{2i}$ 分别是当其他分量不变时关于 $\boldsymbol{\vartheta}$ 的多项式函数和非对称二次函数的导函数.

引理 A.1 在假设 3.6 下, 若随机域 $\{\boldsymbol{X}_{i,n}^{\rm ned}, i\in\boldsymbol{D}_n, n\geqslant 1\}$ 是关于独立同分布随机域 $\boldsymbol{\zeta}$ 的 L_1-NED, 且对于某个 $p>1$, $\{\boldsymbol{X}_{i,n}^{\rm ned}\}$ 是一致 L_p 有界的, 则有

$$\frac{1}{n}\sum_{i=1}^n[\boldsymbol{X}_{i,n}^{\rm ned}-{\rm E}(\boldsymbol{X}_{i,n}^{\rm ned})]\xrightarrow{L_1}\boldsymbol{0}.$$

引理 A.2 在假设 3.6 下, 若随机域 $\{\boldsymbol{X}_{i,n}^{\rm ned}, i\in\boldsymbol{D}_n, n\geqslant 1\}$ 是关于独立同分布随机域 $\boldsymbol{\zeta}$ 的 L_2-NED, 且对于某正数 α, $\{\boldsymbol{X}_{i,n}^{\rm ned}\}$ 是一致 $L_{2+\alpha}$ 有界的, $\inf_n\frac{1}{n}{\rm Var}(\sum_{i=1}^n\boldsymbol{X}_{i,n}^{\rm ned})>0$, $\sum_{k=1}^\infty k^{d_0-1}g(k)<\infty$, $\sup_{n,i\in\boldsymbol{D}}d_{i,n}<\infty$, 则有

$$\left(\sum_{i=1}^n[\boldsymbol{X}_{i,n}^{\rm ned}-{\rm E}(\boldsymbol{X}_{i,n}^{\rm ned})]\right)\Big/{\rm Var}\left(\sum_{i=1}^n\boldsymbol{X}_{i,n}^{\rm ned}\right)\xrightarrow{L}N(0,1).$$

引理 A.3 设 $\boldsymbol{A}_n$ 等于 $\boldsymbol{W}_n$ 或 $\boldsymbol{G}_n(\rho)$, $g_{i,kn}(m)=\boldsymbol{e}_{u,in}^{\rm T}\boldsymbol{A}_n^m\boldsymbol{\zeta}_{kn}^*\boldsymbol{a}$, $\boldsymbol{e}_{u,in}$ 是第 i 个元素为 1 其余元素为 0 的 n 维向量, $k=1,2$, m 为非负整数, $\boldsymbol{a}$ 是常数向量. 若假设 3.5–3.8 及 $\sup_{i,n}\|\boldsymbol{\zeta}_{i,kn}^*\|_p<\infty$ 成立, 则有

$$\sup_{i,n}\|g_{i,kn}(m)\|_p\leqslant C_{amp},\quad \sup_{i,n}\|g_{i,kn}(m)-{\rm E}(g_{i,kn}(m)\mid\mathscr{F}_{i,n}(s))\|_p\leqslant C_{amp}g(s),$$

其中 $g(s)$ 满足 $\sum_{k=1}^\infty k^{d_0-1}g(k)<\infty$.

证明 考虑到非对称二次函数的导数受其解释变量的线性变换所控制, 于是在 $\sup_{i,n}\|\boldsymbol{\zeta}_{i,kn}^*\|_p<\infty$ 的条件下可类比文献 [24, 命题 C.1.6–C.2.6] 的推导获得引理 A.3 的结论. □

引理 A.4 在假设 3.5–3.8 及 $\sup_{i,n}\|\boldsymbol{\zeta}_{i,kn}^*\|_4<\infty$ $(k=1,2)$ 成立的条件下, 有

$$\frac{1}{n}{\rm E}|\boldsymbol{a}^{\rm T}\boldsymbol{\zeta}_{1n}^{*{\rm T}}\boldsymbol{M}_n\boldsymbol{\zeta}_{2n}^*\boldsymbol{b}|=O(1),\quad \frac{1}{n}[\boldsymbol{a}^{\rm T}\boldsymbol{\zeta}_{1n}^{*{\rm T}}\boldsymbol{M}_n\boldsymbol{\zeta}_{2n}^*\boldsymbol{b}-{\rm E}(\boldsymbol{a}^{\rm T}\boldsymbol{\zeta}_{1n}^{*{\rm T}}\boldsymbol{M}_n\boldsymbol{\zeta}_{2n}^*\boldsymbol{b})]=o_{\rm P}(1),$$

其中, $\boldsymbol{M}_n=\boldsymbol{M}_{1n}^{\rm T}\boldsymbol{M}_{2n}\boldsymbol{M}_{1n}$, $\boldsymbol{M}_{2n}$ 是 $\boldsymbol{W}_n$ 或 $\boldsymbol{G}_n$ 的非负整数次方, $\boldsymbol{G}_n=\boldsymbol{W}_n(\boldsymbol{I}_n-\rho_0\boldsymbol{W}_n)^{-1}$.

证明 令 $\boldsymbol{a}^{\rm T}\boldsymbol{\zeta}_{1n}^{*{\rm T}}\boldsymbol{M}_n\boldsymbol{\zeta}_{2n}^*\boldsymbol{b}=\sum_{i=1}^n g_{i,n}=\sum_{i=1}^n g_{i,1n}g_{i,2n}=\sum_{i=1}^n(\boldsymbol{e}_{u,in}^{\rm T}\boldsymbol{M}_{1n}\boldsymbol{\zeta}_{i,1n}^*\boldsymbol{a}\cdot\boldsymbol{e}_{u,in}^{\rm T}\boldsymbol{M}_{2n}\boldsymbol{\zeta}_{i,2n}^*\boldsymbol{b})$. 由引理 A.1, 只需要证明 $g_{i,n}$ 是关于 $\boldsymbol{\zeta}$ 的 L_1-NED 且一致 L_p 有界, $p>1$.

应用条件 Jensen 不等式和引理 A.3, 得

$$\|g_{i,1n}-\mathrm{E}(g_{i,1n}\mid \mathscr{F}_{i,n}(s))\|_4 \leqslant \|g_{i,1n}\|_4+\|\mathrm{E}(g_{i,1n}\mid \mathscr{F}_{i,n}(s))\|_4 \leqslant 2\|g_{i,1n}\|_4<\infty.$$

进一步地, 由 Hölder 不等式和引理 A.3 知,

$$\begin{aligned}\|g_{i,n}-\mathrm{E}(g_{i,n}\mid \mathscr{F}_{i,n}(s))\|_2 &\leqslant \|g_{i,1n}g_{i,2n}-\mathrm{E}(g_{i,1n}\mid \mathscr{F}_{i,n}(s))\mathrm{E}(g_{i,2n}\mid \mathscr{F}_{i,n}(s))\|_2\\ &\leqslant \|g_{i,1n}\|_4\|g_{i,2n}-\mathrm{E}(g_{i,2n}\mid \mathscr{F}_{i,n}(s))\|_4+\|g_{i,2n}\|_4\|g_{i,1n}-\mathrm{E}(g_{i,1n}\mid \mathscr{F}_{i,n}(s))\|_4\\ &\leqslant C_{am}g(s).\end{aligned}$$

因此, $g_{i,n}$ 是关于 $\boldsymbol{\zeta}$ 的 L_2-NED. 再由 Lyapunov 不等式, 可知 $g_{i,n}$ 是关于 $\boldsymbol{\zeta}$ 的 L_1-NED. 由

$$\|g_{i,n}-\mathrm{E}(g_{i,n}\mid \mathscr{F}_{i,n}(s))\|_2 \leqslant 2\|g_{i,n}\|_2 \leqslant 2\|g_{i,1n}\|_4\cdot\|g_{i,2n}\|_4<\infty$$

得 $g_{i,n}$ 一致 L_2 有界, 故引理 A.4 成立. □

引理 A.5 在假设 3.5–3.8 及 $\sup_{i,n}\|\zeta^*_{i,kn}\|_4<\infty$ $(k=1,2)$ 成立的条件下, $\frac{1}{n}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}(\boldsymbol{\vartheta})\boldsymbol{G}^{m_1}_n(\rho)^{\mathrm{T}}$ $\boldsymbol{G}^{m_2}_n(\rho)\ \boldsymbol{\zeta}^*_{2n}(\boldsymbol{\vartheta})\boldsymbol{b}$ 等度连续, 且有

$$\sup_{\boldsymbol{\vartheta}\in\Xi}\frac{1}{n}|\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}(\boldsymbol{\vartheta})\boldsymbol{G}^*_n(\rho)\boldsymbol{\zeta}^*_{2n}(\boldsymbol{\vartheta})\boldsymbol{b}-\mathrm{E}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}(\boldsymbol{\vartheta})\boldsymbol{G}^*_n(\rho)\boldsymbol{\zeta}^*_{2n}(\boldsymbol{\vartheta})\boldsymbol{b})|=o_{\mathrm{P}}(1),$$

其中, m_1 和 m_2 为非负整数, $\boldsymbol{G}^*_n(\rho)=[\boldsymbol{G}^{m_1}_n(\rho)]^{\mathrm{T}}\boldsymbol{G}^{m_2}_n(\rho)$.

证明 依引理 A.4 及 $\boldsymbol{\zeta}^*_{kn}(\boldsymbol{\vartheta})$ 的构造形式, 在紧参数空间上只需证 $\frac{1}{n}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}\boldsymbol{G}^{m_1}_n(\rho)^{\mathrm{T}}\boldsymbol{G}^{m_2}_n(\rho)\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}$ 的等度连续性. 令

$$\boldsymbol{B}_n(\rho)=[\boldsymbol{G}^{m_1}_n(\rho)]^{\mathrm{T}}\{m_2\boldsymbol{G}_n(\rho)+[m_1\boldsymbol{G}^{m_1}_n(\rho)]^{\mathrm{T}}\}\boldsymbol{G}^{m_2}_n(\rho),$$

由 $\sup_\rho\|\boldsymbol{G}_n(\rho)\|_1<\infty$ 和 $\sup_\rho\|\boldsymbol{G}_n(\rho)\|_\infty<\infty$, 结合矩阵范数的三角不等式与相容性可知,

$$\sup_\rho\|\boldsymbol{B}^{\mathrm{T}}_n(\rho)\boldsymbol{B}_n(\rho)\|_\infty<\infty.$$

依据 Lagrange 中值定理, 对于参数集上的任意 $\rho_1<\rho_2$, 存在 $\rho^*\in(\rho_1,\rho_2)$, 使得

$$\begin{aligned}&\frac{1}{n}|\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}\boldsymbol{G}^{m_1}_n(\rho_1)^{\mathrm{T}}\boldsymbol{G}^{m_2}_n(\rho_1)\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}\boldsymbol{G}^{m_1}_n(\rho_2)^{\mathrm{T}}\boldsymbol{G}^{m_2}_n(\rho_2)\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}|\\ &\quad=\frac{1}{n}(\rho_2-\rho_1)|\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\zeta}^{*\mathrm{T}}_{1n}\boldsymbol{B}_n(\rho^*)\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}|\\ &\quad\leqslant\frac{1}{n}(\rho_2-\rho_1)\|\boldsymbol{\zeta}^*_{1n}\boldsymbol{a}\|_2\cdot\|\boldsymbol{B}_n(\rho^*)\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}\|_2\\ &\quad\leqslant\frac{1}{n}(\rho_2-\rho_1)\sqrt{r(\boldsymbol{B}^{\mathrm{T}}_n(\rho^*)\boldsymbol{B}_n(\rho^*))}\|\boldsymbol{\zeta}^*_{1n}\boldsymbol{a}\|_2\cdot\|\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}\|_2\\ &\quad\leqslant(\rho_2-\rho_1)\sqrt{\sup_\rho(\|\boldsymbol{B}^{\mathrm{T}}_n(\rho)\boldsymbol{B}_n(\rho)\|_\infty)}\left\|\frac{1}{\sqrt{n}}\boldsymbol{\zeta}^*_{1n}\boldsymbol{a}\right\|_2\cdot\left\|\frac{1}{\sqrt{n}}\boldsymbol{\zeta}^*_{2n}\boldsymbol{b}\right\|_2\\ &\quad=O_{\mathrm{P}}(\rho_2-\rho_1).\end{aligned}$$

从而可知引理 A.5 的结论成立. □

引理 A.6 设 $q_{i,n}(j)=\sum_{k=1}^{n}(\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{\zeta}_{i,1n}^{*\mathrm{T}}\boldsymbol{M}_{jn}(i,k)\boldsymbol{\zeta}_{k,2n}^{*}\boldsymbol{b}_j)$, $\{\boldsymbol{a}_j\}$ 和 $\{\boldsymbol{b}_j\}$ 表示常数列,

$$t_{i,n}=\sum_{j=1}^{m}q_{i,n}(j),\quad \inf_n\frac{1}{n}\mathrm{Var}\left(\sum_{i=1}^{n}t_{i,n}\right)>0,$$

在假设 3.5–3.8 及 $\sup_{i,n}\|\boldsymbol{\zeta}_{i,kn}^{*}\|_{4+\alpha}<\infty$ ($\alpha>0$, $k=1,2$) 成立的条件下, 有

$$\sum_{i=1}^{n}[t_{i,n}-\mathrm{E}(t_{i,n})]\Big/\mathrm{Var}\left(\sum_{i=1}^{n}t_{i,n}\right)\xrightarrow{L}N(0,1).$$

证明 由 $\sup_{i,n}\|\boldsymbol{\zeta}_{i,kn}^{*}\|_{4+\alpha}<\infty$ 和引理 A.4 可知, $q_{i,n}(j)$ ($j=1,\ldots,m$) 是关于独立同分布随机域 $\boldsymbol{\zeta}$ 的 L_2-NED, 结合范数的三角不等式, 有

$$\begin{aligned}\|t_{i,n}\|_{2+\alpha}&=\left\|\sum_{j=1}^{m}q_{i,n}(j)\right\|_{2+\alpha}\leqslant\sum_{j=1}^{m}\|q_{i,n}(j)\|_{2+\alpha}<\infty,\\ \|t_{i,n}-\mathrm{E}(t_{i,n})\|_2&=\left\|\sum_{j=1}^{m}[q_{i,n}(j)-\mathrm{E}(q_{i,n}(j))]\right\|_2\leqslant\sum_{j=1}^{m}\|[q_{i,n}(j)-\mathrm{E}(q_{i,n}(j))]\|_2<C_{mab}g(s).\end{aligned}$$

再依据引理 A.2 可知引理 A.6 的结论成立. □

附录 A.2 定理 3.4 的证明

令 $\hat{\boldsymbol{X}}_n=(\boldsymbol{X}_{1n},\boldsymbol{Q}_n\hat{\boldsymbol{\kappa}},(\boldsymbol{Z}_n-\boldsymbol{X}_{2n}\hat{\boldsymbol{\Gamma}}))$, $\Phi_n(\boldsymbol{\vartheta};\tau)=\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta})+n\sum_{j=1}^{p_1}p_{\lambda_n}(\vartheta_j)$. 只需要证明对于 $\varpi>0$, 存在足够大的常数 C, 使得 $\mathrm{P}\{\inf_{\|\boldsymbol{u}\|=C}\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0+\boldsymbol{u}/\sqrt{n})>\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0)\}\geqslant 1-\varpi$. 考虑

$$\begin{aligned}\Phi_n^{\mathrm{en}}\left(\boldsymbol{\vartheta}_0+\frac{\boldsymbol{u}}{\sqrt{n}}\right)-\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0)=&\sum_{i=1}^{n}\Psi_\tau\left(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}\right)\\&-\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)+\sum_{j=1}^{p_1}n\left[p_{\lambda_n}\left(\left|\vartheta_{j0}+\frac{u_j}{\sqrt{n}}\right|\right)-p_{\lambda_n}(|\vartheta_{j0}|)\right].\end{aligned}$$

由于 $p_{\lambda_n}(0)=0$, 且在 $a>2$ 和 $\theta>0$ 的条件下, 有 $p'_{\lambda_n}(\theta)\geqslant 0$, 于是当 $j=q+1,\ldots,k_1+1$ 时, 有

$$n\left[p_{\lambda_n}\left(\left|\vartheta_{j0}+\frac{u_j}{\sqrt{n}}\right|\right)-p_{\lambda_n}(|\vartheta_{j0}|)\right]=np_{\lambda_n}(|\vartheta_{j0}|)\geqslant 0.$$

因此,

$$\begin{aligned}&\Phi_n^{\mathrm{en}}\left(\boldsymbol{\vartheta}_0+\frac{\boldsymbol{u}}{\sqrt{n}}\right)-\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0)\\&\geqslant\sum_{i=1}^{n}\Psi_\tau\left(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}\right)-\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)+\sum_{j=1}^{q}n\left[p_{\lambda_n}\left(\left|\vartheta_{j0}+\frac{u_j}{\sqrt{n}}\right|\right)-p_{\lambda_n}(|\vartheta_{j0}|)\right]\\&=\sum_{i=1}^{n}\left\{-\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}})_i^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}+\frac{1}{2}\varphi_\tau(\xi_n^*)\left[(\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}})_i^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}\right]^2\right\}\\&\quad+\sum_{j=1}^{q}n\left[p'_{\lambda_n}(|\vartheta_{j0}|)\mathrm{sgn}(\vartheta_{j0})\frac{u_j}{\sqrt{n}}+\frac{1}{2}p''_{\lambda_n}(|\vartheta_{j0}|)\left(\frac{u_j}{\sqrt{n}}\right)^2+o\left(\frac{p''_{\lambda_n}(|\vartheta_{j0}|)}{n}\right)\right]\\&\triangleq\Phi_{n1}+\Phi_{n2}+\Phi_{n3},\end{aligned}$$

其中, ξ_n^* 位于 $y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}$ 与 $y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0$ 之间, $\phi_\tau(\cdot)$ 和 $\varphi_\tau(\cdot)$ 分别是 $\Psi_\tau(\cdot)$ 可导处的一阶和二阶导数, $\Phi_{n1}=-\sum_{i=1}^n\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}+\sum_{j=1}^q n(p'_{\lambda_n}(|\vartheta_{j0}|)\mathrm{sgn}(\vartheta_{j0})\frac{u_j}{\sqrt{n}})$, $\Phi_{n2}=\sum_{i=1}^n\frac{1}{2}\varphi_\tau(\xi_n^*)[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}]^2$, $\Phi_{n3}=\sum_{j=1}^q n[\frac{1}{2}p''_{\lambda_n}(|\vartheta_{j0}|)(\frac{u_j}{\sqrt{n}})^2+o(\frac{p''_{\lambda_n}(|\vartheta_{j0}|)}{n})]$.

依据极值优化条件知 $\Phi_{n1}=o_{\mathrm{P}}(\|\boldsymbol{u}\|)$. 由于 $\varphi_\tau(\cdot)$ 非负有界, 不妨设 $\varphi_\tau(\cdot)\leqslant C_\varphi$, 则

$$\Phi_{n2}=\sum_{i=1}^n\frac{1}{2}\varphi_\tau(\xi_n^*)\left[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{u}}{\sqrt{n}}\right]^2\leqslant\frac{1}{2n}C_h\sum_{i=1}^n[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{u}]^2\leqslant\frac{1}{2}C_h\boldsymbol{u}^{\mathrm{T}}\left[\frac{1}{n}\sum_{i=1}^n\hat{\boldsymbol{X}}_{i,n}\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\right]\boldsymbol{u}.$$

由 $\frac{1}{n}\sum_{i=1}^n\hat{\boldsymbol{X}}_{i,n}\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}=\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}_{i,n}^*\boldsymbol{X}_{i,n}^{*\mathrm{T}}+o_{\mathrm{P}}(1)$ 和假设 3.7 的条件可知 $\Phi_{n2}=O_{\mathrm{P}}(\|\boldsymbol{u}\|^2)$. 对于 SCAD 惩罚函数, 当 $\lambda_n\to0$ 时, 有 $\lim_{n\to\infty}p''_{\lambda_n}(|\vartheta_{j0}|)=\frac{1}{a-1}\lim_{n\to\infty}I(\lambda_n<|\vartheta_{j0}|<a\lambda_n)=0$. 所以,

$$\Phi_{n3}=\sum_{j=1}^q\left[\frac{1}{2}p''_{\lambda_n}(|\vartheta_{j0}|)u_j^2+o(p''_{\lambda_n}(|\vartheta_{j0}|))\right]=O\Big(\max_{1\leqslant j\leqslant q}p''_{\lambda_n}(|\vartheta_{j0}|)\Big)=o(1).$$

综上可知, $\Phi_{n1}+\Phi_{n2}+\Phi_{n3}$ 的大小由 Φ_{n2} 决定, 从而在 n 充分大的条件下, 存在足够大的常数 C, 使得 $\|\boldsymbol{u}\|=C$ 成立时, $\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0+\boldsymbol{u}/\sqrt{n})-\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta}_0)$ 的取值为正, 即存在 $\Phi_n^{\mathrm{en}}(\boldsymbol{\vartheta};\tau)$ 的极小值点 $\hat{\boldsymbol{\vartheta}}_0$ 使得 $\|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|=O_{\mathrm{P}}(n^{-1/2})$ 成立. 定理 3.4 证毕.

附录 A.3 定理 3.5 的证明

(1) 首先给出稀疏性的证明. 令 $g_{\lambda_n}(\boldsymbol{\vartheta}_{10})=(p'_{\lambda_n}(\vartheta_{1,10}),\ldots,p'_{\lambda_n}(\vartheta_{q,10}))^{\mathrm{T}}$. 若目标函数随样本量的增大在 $\boldsymbol{\vartheta}_2=0$ 时达到最小, 即可证明稀疏性. 考虑

$$\begin{aligned}
&\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}},\mathbf{0}^{\mathrm{T}})^{\mathrm{T}})-\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}})\\
&\quad=\sum_{i=1}^n[\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_1)-\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_1-\hat{\boldsymbol{X}}_{i,n2}^{\mathrm{T}}\boldsymbol{\vartheta}_2)]-\sum_{j=q+1}^p np_{\lambda_n}(|\vartheta_j|)\\
&\quad=\sum_{i=1}^n\Big\{\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&\qquad+\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})[\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2+o([\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2)\Big\}\\
&\qquad-\sum_{i=1}^n\Big\{\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}\\
&\qquad+\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2+o([\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2)\Big\}\\
&\qquad-\sum_{j=q+1}^p np_{\lambda_n}(|\vartheta_j|)\\
&\quad\triangleq\Delta_{11}+\Delta_{12}+\Delta_{13}+\Delta_{21}+\Delta_{22}+\Delta_{23}+\Delta_3,
\end{aligned}$$

其中,

$$\Delta_{11}=\sum_{i=1}^n\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10}),\quad\Delta_{12}=\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})[\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2,$$

$$\Delta_{13}=o([\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2),\quad\Delta_{21}=\sum_{i=1}^n[\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}],$$

$$\Delta_{22}=\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2,\quad \Delta_{23}=o([\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2),$$

$$\Delta_3=\sum_{j=q+1}^{p}np_{\lambda_n}(|\vartheta_j|).$$

对于 Δ_{11}, 不难证明 $\lim_{n\to\infty}n^{1/2}\max_{1\leqslant j\leqslant q}p'_{\lambda_n}(|\vartheta_{j,10}|)=0$, 再依据极值优化条件, 有

$$\begin{aligned}
\Delta_{11}&=\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})+ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=\left[\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}+ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})\right](\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=O_{\mathrm{P}}\Big(n^{1/2}q\max_{1\leqslant j\leqslant q}p'_{\lambda_n}(|\vartheta_{10,j}|)\Big)\\
&=o_{\mathrm{P}}(1).
\end{aligned}$$

对于 Δ_{12}, 有

$$\begin{aligned}
\Delta_{12}&=\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})[\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2\\
&\leqslant\frac{1}{2}C_h[\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2\\
&=O_{\mathrm{P}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{l}_q^{\mathrm{T}}\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{l}_q\right)\\
&=O_{\mathrm{P}}(1).
\end{aligned}$$

对于 Δ_{13}, 有

$$\Delta_{13}=o([\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2)=O_{\mathrm{P}}(1).$$

对于 Δ_{21}, 有

$$\begin{aligned}
\Delta_{21}&=\sum_{i=1}^{n}\{\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}\}\\
&=\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}+ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=\left[\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}+ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})\right](\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=-ng_{\lambda_n}^{\mathrm{T}}(\boldsymbol{\vartheta}_{10})(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\
&=O_{\mathrm{P}}\Big(n^{1/2}q\max_{1\leqslant j\leqslant q}p'_{\lambda_n}(|\vartheta_{j,10}|)\Big)\\
&=o_{\mathrm{P}}(1).
\end{aligned}$$

对于 Δ_{22}, 有

$$
\begin{aligned}
\Delta_{22} &= \frac{1}{2}\varphi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2 \\
&\leqslant \frac{1}{2}C_h[\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2 \\
&= O_{\mathrm{P}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{l}_p^{\mathrm{T}}\hat{\boldsymbol{X}}_{i,n}\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{l}_p\right) \\
&= O_{\mathrm{P}}(1).
\end{aligned}
$$

对于 Δ_{23}, 同上, 有

$$
\Delta_{23} = o([\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2) = O_{\mathrm{P}}(1).
$$

对于 Δ_3, 当 $j = q+1, \ldots, p$ 时, 考虑

$$
p_{\lambda_n}(|\vartheta_j|) = \lim_{t \to 0^+} p_{\lambda_n}(t) + \lim_{t \to 0^+} p'_{\lambda_n}(t)|\vartheta_j| + o(|\vartheta_j|) = \lambda_n|\vartheta_j| + o(|\vartheta_j|),
$$

进而,

$$
\Delta_3 = \sum_{j=q+1}^{p} np_{\lambda_n}(|\vartheta_j|) = -n\lambda_n\left(\sum_{j=q+1}^{p}\left(|\vartheta_j| + o\left(\frac{|\vartheta_j|}{\lambda_n}\right)\right)\right) = -n\lambda_n\left(\sum_{j=q+1}^{p}\left(|\vartheta_j| + o\left(\frac{1}{\lambda_n\sqrt{n}}\right)\right)\right).
$$

因此, 当 $n\lambda_n \to \infty$ 且 $\lambda_n\sqrt{n} \to \infty$ 时, $\lim_{n\to\infty}\Delta_3 = -\infty$.

综上可知, 对充分大的 n, 有

$$
\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}) - \Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}) \leqslant 0,
$$

即稀疏性成立.

(2) 接下来考虑渐近正态性.

令 $\boldsymbol{\eta}_1 = (\eta_1, \ldots, \eta_q)^{\mathrm{T}} = \sqrt{n}(\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10}), R_n(\boldsymbol{\eta}_1) = \Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}})$, 则

$$
\begin{aligned}
R_n(\boldsymbol{\eta}_1) &= \sum_{i=1}^{n}\Psi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_1) + \sum_{j=1}^{q} np_{\lambda_n}(|\vartheta_j|) \\
&= \sum_{i=1}^{n}\Psi_\tau\left(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\frac{\boldsymbol{\eta}_1}{\sqrt{n}}\right) + \sum_{j=1}^{q} np_{\lambda_n}\left(\left|\vartheta_{j,0} + \frac{\eta_j}{\sqrt{n}}\right|\right).
\end{aligned}
$$

考虑到 $\hat{\boldsymbol{\eta}}_1 = \sqrt{n}(\hat{\boldsymbol{\vartheta}}_1 - \boldsymbol{\vartheta}_{10})$ 是 $R_n(\boldsymbol{\eta}_1)$ 的局部极小值点, 从而 $\frac{\partial R_n(\boldsymbol{\eta}_1)}{\partial \eta_j}|_{\boldsymbol{\eta}_1 = \hat{\boldsymbol{\eta}}_1} = 0$ $(j = 1, \ldots, q)$, 因此,

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\frac{\partial\Psi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\frac{\boldsymbol{\eta}_1}{\sqrt{n}})}{\partial \eta_j}\bigg|_{\boldsymbol{\eta}_1 = \hat{\boldsymbol{\eta}}_1} + \sum_{j=1}^{q}\frac{\partial np_{\lambda_n}(|\vartheta_{j,0} + \frac{\eta_j}{\sqrt{n}}|)}{\partial \eta_j}\bigg|_{\boldsymbol{\eta}_1 = \hat{\boldsymbol{\eta}}_1} \\
&= \sum_{i=1}^{n}\frac{\partial}{\partial \eta_j}\bigg[\Psi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10}) - \phi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\frac{\boldsymbol{\eta}_1}{\sqrt{n}} \\
&\quad + \frac{1}{2}\varphi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\left(\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\frac{\boldsymbol{\eta}_1}{\sqrt{n}}\right)^2 + o(1)\bigg]\bigg|_{\boldsymbol{\eta}_1 = \hat{\boldsymbol{\eta}}_1}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j=1}^{q} n\frac{\partial}{\partial\eta_j}\left[p_{\lambda_n}(|\vartheta_{j,0}|)+p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{\eta_j}{\sqrt{n}}+\frac{1}{2}p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\eta_j^2}{n}+o\left(\frac{1}{n}\right)\right]\bigg|_{\boldsymbol{\eta}_1=\hat{\boldsymbol{\eta}}_1}\\
=&\sum_{i=1}^{n}\left[-\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\frac{\hat{X}_{ij,n1}}{\sqrt{n}}+\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\frac{\hat{X}_{ij,n1}}{\sqrt{n}}\right]\\
&+\sum_{j=1}^{q} n\left[p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{1}{\sqrt{n}}+p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\hat{\eta}_j}{n}\right]+o(1)\\
=&\sum_{i=1}^{n}\left[-\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\frac{\hat{X}_{ij,n1}}{\sqrt{n}}+\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\frac{\hat{X}_{ij,n1}}{\sqrt{n}}\right]+o(1).
\end{aligned}
$$

进而通过适当变换可得

$$
\begin{aligned}
\hat{\boldsymbol{\eta}}_1=&\left(\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\sum_{i=1}^{n}\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\right)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}\right.\\
&-\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})]\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\hat{\boldsymbol{\eta}}_1\\
&\left.-\sum_{j=1}^{q} n\left[p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{1}{\sqrt{n}}+p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\hat{\eta}_j}{n}\right]+o(1)\right\}\\
\triangleq&\ \Theta_1^{-1}(\Theta_2+\Theta_3+\Theta_4),
\end{aligned}
$$

其中,

$$
\begin{aligned}
\Theta_1&=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\sum_{i=1}^{n}\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}},\\
\Theta_2&=\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})]\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\hat{\boldsymbol{\eta}}_1,\\
\Theta_3&=\sum_{j=1}^{q} n\left[p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{1}{\sqrt{n}}+p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\hat{\eta}_j}{n}\right]+o(1),\\
\Theta_4&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}.
\end{aligned}
$$

对于 Θ_1, 依据假设条件, 有

$$
\Theta_1=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\xrightarrow{\mathrm{P}}\mu_{\mathrm{en},\varphi_\tau}\boldsymbol{\Sigma}_{\mathrm{en},11}.
$$

对于 Θ_2, 由 Markov 大数定理, 得

$$
\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})]\mathrm{E}[\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}]\xrightarrow{\mathrm{P}}0.
$$

而 φ_τ 的有界性和引理 A.1 表明

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})][\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}-\mathrm{E}(\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}})]\right\|
$$

$$
\begin{aligned}
&\leqslant C_{\varphi}\cdot\frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}-\mathrm{E}(\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}})\|\\
&=o_{\mathrm{P}}(1).
\end{aligned}
$$

再依据 $\hat{\boldsymbol{\eta}}_1\xrightarrow{\mathrm{P}}\boldsymbol{\eta}_1$ 及上述结果, 可得

$$
\begin{aligned}
\Theta_2=&\left\{\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})]\mathrm{E}[\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}]\right.\\
&\left.+\frac{1}{n}\sum_{i=1}^{n}[\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})][\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}-\mathrm{E}(\hat{\boldsymbol{X}}_{i,n1}\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}})]\right\}\hat{\boldsymbol{\eta}}_1\\
=&\,o_{\mathrm{P}}(1).
\end{aligned}
$$

对于 Θ_3, 依据 $\hat{\boldsymbol{\eta}}_1\xrightarrow{\mathrm{P}}\boldsymbol{\eta}_1$ 及 $n[p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{1}{\sqrt{n}}+p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\eta_j}{n}]\xrightarrow{\mathrm{P}}0$, 有

$$
\Theta_3=\sum_{j=1}^{q}n\left[p'_{\lambda_n}(|\vartheta_{j,0}|)\mathrm{sgn}(\vartheta_{j,0})\frac{1}{\sqrt{n}}+p''_{\lambda_n}(|\vartheta_{j,0}|)\frac{\eta_j}{n}\right]+\sum_{j=1}^{q}p''_{\lambda_n}(|\vartheta_{j,0}|)(\hat{\eta}_j-\eta_j)+o(1)=o(1).
$$

对于 Θ_4, 由引理 A.3, 有

$$
\Theta_4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}\xrightarrow{L}N(\mathbf{0},\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{en}}).
$$

于是, 再由 Slutsky 定理, 有

$$
\hat{\boldsymbol{\eta}}_1\xrightarrow{L}N(\mathbf{0},\mu_{\mathrm{en},\varphi_\tau}^{-2}\boldsymbol{\Sigma}_{\mathrm{en},11}^{-1}\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{en}}\boldsymbol{\Sigma}_{\mathrm{en},11}^{-1}).
$$

定理 3.5 证毕.

附录 A.4 定理 3.6 的证明

首先证明证渐近正态性. 令 $\boldsymbol{\eta}=(\eta_1,\dots,\eta_p)^{\mathrm{T}}=\sqrt{n}(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)$, 则有

$$
\begin{aligned}
\Re_n(\boldsymbol{\eta})&\triangleq R_n(\boldsymbol{\vartheta})-R_n(\boldsymbol{\vartheta}_0)\\
&=\sum_{i=1}^{n}\Psi_\tau\left(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{\eta}}{\sqrt{n}}\right)-\sum_{i=1}^{n}\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\\
&\quad+\sum_{j=1}^{p}n\lambda_n\frac{1}{|\hat{\vartheta}_j|^r}\left|\vartheta_{j,0}+\frac{\eta_j}{\sqrt{n}}\right|-\sum_{j=1}^{p}n\lambda_n\frac{1}{|\hat{\vartheta}_j|^r}|\vartheta_{j,0}|\\
&=\sum_{i=1}^{n}\left[\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\left(-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{\eta}}{\sqrt{n}}\right)+\frac{\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)}{2}\left(-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{\eta}}{\sqrt{n}}\right)^2+o_{\mathrm{P}}\left(\frac{1}{n}\right)\right]\\
&\quad+n\lambda_n\sum_{j=1}^{p}\frac{1}{|\hat{\vartheta}_j|^r}\left(\left|\vartheta_{j,0}+\frac{\eta_j}{\sqrt{n}}\right|-|\vartheta_{j,0}|\right)\\
&=:\Re_1+\Re_2.
\end{aligned}
$$

分析 $\Re_1$, 类似引理 A.5 的证明可知,

$$
\sum_{i=1}^{n}\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\left(\frac{\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}}{\sqrt{n}}\right)\xrightarrow{L}\boldsymbol{H}N(\mathbf{0},\boldsymbol{\Sigma}_{\phi_\tau}^{\mathrm{en}}),
$$

$$\frac{1}{2}\sum_{i=1}^{n}\frac{\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)}{n}\sum_{i=1}^{n}\frac{\hat{\boldsymbol{X}}_{i,n}\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}}{n}\xrightarrow{\mathrm{P}}\frac{\mu_{\mathrm{en},\varphi_\tau}}{2}\boldsymbol{\Sigma}_{\mathrm{en}},$$
$$\frac{1}{2}\sum_{i=1}^{n}\frac{(\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)-\mathrm{E}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0))\hat{\boldsymbol{X}}_{i,n}\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}}{n}\xrightarrow{\mathrm{P}}\boldsymbol{0}.$$

所以,

$$\begin{aligned}\Re_1&=\sum_{i=1}^{n}\left[\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)\left(-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{\eta}}{\sqrt{n}}\right)+\frac{\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\boldsymbol{\vartheta}_0)}{2}\left(-\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}}\frac{\boldsymbol{\eta}}{\sqrt{n}}\right)^2+o_{\mathrm{P}}\left(\frac{1}{n}\right)\right]\\&\xrightarrow{L}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\eta}+\frac{\mu_{\mathrm{en},\varphi_\tau}}{2}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{en}}\boldsymbol{\eta}.\end{aligned}$$

分析 $\Re_2$, 对于非零参数 $\vartheta_{j,0}$ $(j=1,\ldots,q)$ 的 $\sqrt{n}$ 相合估计 $\hat{\vartheta}_j$, 有

$$\hat{\vartheta}_j\xrightarrow{\mathrm{P}}\vartheta_{j,0},\quad \sqrt{n}\left(\left|\vartheta_{j,0}+\frac{\eta_j}{\sqrt{n}}\right|-|\vartheta_{j,0}|\right)\to\eta_j\mathrm{sgn}(\vartheta_{j,0}).$$

对于零参数 $\vartheta_{j,0}=0$ $(j=q+1,\ldots,p)$ 的 $\sqrt{n}$ 相合估计 $\hat{\vartheta}_j$, 有

$$\sqrt{n}\left(\left|\vartheta_{j,0}+\frac{\eta_j}{\sqrt{n}}\right|-|\vartheta_{j,0}|\right)\to\eta_j,\quad \sqrt{n}\lambda_n\frac{1}{|\hat{\vartheta}_j|^r}=\lambda_n n^{(r+1)/2}\frac{1}{|\sqrt{n}\hat{\vartheta}_j|^r}.$$

进而在假设 3.4(1) 和 3.4(2) 的条件下, 有

$$\Re_2=n\lambda_n\sum_{j=1}^{p}\frac{1}{|\hat{\vartheta}_j|^r}\left(\left|\vartheta_{j,0}+\frac{\eta_j}{\sqrt{n}}\right|-|\vartheta_{j,0}|\right)\xrightarrow{\mathrm{P}}\begin{cases}0, & \eta_{q+1}=\cdots=\eta_p=0,\\ \infty, & \text{其他}.\end{cases}$$

于是,

$$\Re_n(\boldsymbol{\eta})=\Re_1+\Re_2\xrightarrow{L}\Re(\boldsymbol{\eta})=\begin{cases}-\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{\eta}_1+\dfrac{u_{\varphi_\tau}}{2}\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{en},11}\boldsymbol{\eta}_1, & \eta_{q+1}=\cdots=\eta_p=0,\\ \infty, & \text{其他},\end{cases}$$

其中, $\boldsymbol{H}_1$ 和 $\boldsymbol{\eta}_1$ 分别为 $\boldsymbol{H}$ 和 $\boldsymbol{\eta}$ 的前 q 个分量构成的子向量, $\boldsymbol{\Sigma}_{\mathrm{en},11}$ 为 $\boldsymbol{\Sigma}_{\mathrm{en}}$ 的左上角 q 行 q 列元素构成的子矩阵. 再依据上式收敛结果, 可得 $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{\mathrm{ALAS},2}-\boldsymbol{\vartheta}_{20})\xrightarrow{L}\boldsymbol{0}$ 以及渐近正态性

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{\mathrm{ALAS},1}-\boldsymbol{\vartheta}_{10})\xrightarrow{L}(u_{\mathrm{en},\varphi_\tau}\boldsymbol{\Sigma}_{\mathrm{en},11})^{-1}\boldsymbol{H}_1\sim N(\boldsymbol{0},u_{\mathrm{en},\varphi_\tau}^{-2}\boldsymbol{\Sigma}_{\mathrm{en},11}^{-1}\boldsymbol{\Sigma}_{\phi_\tau,11}^{\mathrm{en}}\boldsymbol{\Sigma}_{\mathrm{en},11}^{-1}).$$

故渐近正态性成立.

接下来考虑稀疏性. 对于任意 $\hat{\boldsymbol{\vartheta}}_{\mathrm{ALAS},1}-\boldsymbol{\vartheta}_{10}=O_{\mathrm{P}}(n^{-1/2})$, $0<\|\boldsymbol{\vartheta}_2\|\leqslant Cn^{-1/2}$, 有

$$\begin{aligned}&\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}},\boldsymbol{0}^{\mathrm{T}})^{\mathrm{T}})-\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}},\boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}})\\&\quad=\sum_{i=1}^{n}[\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_1)-\Psi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_1-\hat{\boldsymbol{X}}_{i,n2}^{\mathrm{T}}\boldsymbol{\vartheta}_2)]-\sum_{j=q+1}^{p}np_{\lambda_n}(|\vartheta_j|)\\&\quad=\sum_{i=1}^{n}\{\phi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})\\&\qquad+\frac{1}{2}\varphi_\tau(y_{i,n}-\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}\boldsymbol{\vartheta}_{10})[\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2+o([\hat{\boldsymbol{X}}_{i,n1}^{\mathrm{T}}(\boldsymbol{\vartheta}_1-\boldsymbol{\vartheta}_{10})]^2)\}\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i=1}^{n} \Big\{ \phi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}} \boldsymbol{\vartheta}_0) \hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}} ((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}} \\
&+ \frac{1}{2} \varphi_\tau(y_{i,n} - \hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}} \boldsymbol{\vartheta}_0) [\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}} ((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2 + o([\hat{\boldsymbol{X}}_{i,n}^{\mathrm{T}} ((\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_{10})^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}]^2) \Big\} \\
&- \sum_{j=q+1}^{p} n\lambda_n \frac{1}{|\hat{\vartheta}_j|^r} |\vartheta_j| \\
&\triangleq \Delta_{11} + \Delta_{12} + \Delta_{13} + \Delta_{21} + \Delta_{22} + \Delta_{23} - \Lambda_3.
\end{aligned}
$$

类似定理 3.4 中稀疏性的证明, 可得

$$
\Delta_{11} + \Delta_{12} + \Delta_{13} + \Delta_{21} + \Delta_{22} + \Delta_{23} = O_{\mathrm{P}}(1).
$$

再由初始估计 $\hat{\vartheta}_j$ 的渐近正态性和 $\lim_{n\to\infty} \lambda_n n^{(r+1)/2} = \infty$, 得

$$
\Lambda_3 = \sum_{j=q+1}^{p} n\lambda_n \frac{1}{|\hat{\vartheta}_j|^r} |\vartheta_j| = n^{(1+r)/2} \lambda_n \sqrt{n} \sum_{j=q+1}^{p} \frac{1}{|\sqrt{n}\hat{\vartheta}_j|^r} |\vartheta_j| \to \infty.
$$

因此, 对充分大的 n, 有

$$
\Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{0}^{\mathrm{T}})^{\mathrm{T}}) - \Phi_n^{\mathrm{en}}((\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}) \leqslant 0,
$$

故稀疏性成立.

Expectile regression analysis of high-dimensional spatially dependent data

Xuan Liu, Haiqiang Ma, Zhiyan Sheng & Liangqing Luo

Abstract There are great challenges to the analysis of spatially dependent data with the heterogeneity, the endogeneity of spatial weights, and the high-dimensional characteristics of explanatory variables. Based on the robust estimation advantage of Expectile regression and the effective dimensionality reduction ability of penalty compression, we give the two-step and three-step penalty Expectile estimation of unknown parameters of high-dimensional spatial lag models under the exogenous and endogenous spatial weight matrices respectively, and prove the consistency of the proposed estimation and the Oracle property of variable selection under conventional regularization conditions. The numerical simulation demonstrates that the two-step estimation method can work well with the robust statistical problem under the exogenous spatial weight matrix, and the three-step estimation method has excellent performance under the exogenous and endogenous spatial weights. Finally, the effectiveness of the proposed method is further verified by analyzing the relationship between the urban air quality and the economic development in China.

Keywords spatial lag model, Expectile regression, variable selection, endogeneity, high-dimensional data

MSC(2020) 62F35, 62J99, 91B72

doi: 10.1360/SSM-2023-0031