

文章编号: 1004-4353(2020)01-0015-05

## 空间误差模型的变量选择

刘宣<sup>1</sup>, 马海强<sup>2</sup>, 邓胜伶<sup>1</sup>

( 1. 阳光学院 基础教研部, 福建 福州 350015; 2. 江西财经大学 统计学院, 江西 南昌 330013 )

**摘要:** 基于惩罚拟似然的方法对空间误差模型的变量选择问题进行研究, 得到了惩罚估计量的相合性和渐近正态性. 通过蒙特卡洛模拟, 验证了本文所提惩罚方法选择变量的有效性.

**关键词:** 空间误差模型; 空间权重矩阵; 变量选择; 惩罚似然

**中图分类号:** O212.4

**文献标志码:** A

## Variable selection for the spatial error model

LIU Xuan<sup>1</sup>, MA Haiqiang<sup>2</sup>, DENG Shengling<sup>1</sup>

( 1. Department of Basic Teaching and Research, Yango University, Fuzhou 350015, China;  
2. School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China )

**Abstract:** Based on penalized quasi-likelihood method, we studied the problem of variable selection in spatial error model, and obtained the consistency and asymptotic normality of penalty estimator. Monte Carlo simulation shows the effectiveness of the proposed penalty method for variable selection.

**Keywords:** spatial error model; spatial weight matrix; variable selection; penalized likelihood

空间计量模型是一种描述地理单元间空间相依关系的模型, 因其具有简单、直观和易解释等特点而被广泛应用于经济、管理、能源和交通等领域. 空间计量模型的已有研究主要集中在参数估计和假设检验两个方面<sup>[1-7]</sup>. 近年来, 随着高维数据的出现, 空间计量模型的变量选择问题引起了学者们的关注, 例如: 文献[8]采用自适应 LASSO (least absolute shrinkage and selection operator) 方法对空间自回归模型的变量选择问题进行了研究, 文献[9-10]探讨了空间自回归模型变量选择的大样本渐进性质. 受以上文献的启发, 本文采用 SCAD (smoothly clipped absolute deviation) 惩罚方法对空间误差模型的变量选择问题进行研究, 拟获得该问题的惩罚估计量的理论性质.

### 1 模型及假设

考虑空间误差模型

$$Y_n = X_n \beta + U_n, U_n = \rho W_n U_n + V_n. \quad (1)$$

其中:  $Y_n$  是  $n$  维向量;  $X_n$  是  $n \times k$  阶设计矩阵;  $\beta$  是  $k$  维的回归系数;  $\rho$  是空间自相关系数, 满足  $|\rho| < 1$ ;  $W_n$  是  $n$  阶空间权重矩阵;  $V_n$  是独立同分布的  $n$  维误差向量.

令  $\theta = (\sigma^2, \rho, \beta^T)^T = (\theta_1, \dots, \theta_{k+2})^T$ ,  $\theta_0 = (\sigma_0^2, \rho_0, \beta_0^T)^T = (\theta_{1,0}, \dots, \theta_{k+2,0})^T$  为未知参数  $\theta$  的真实值,  $S_n(\rho) = I_n - \rho W_n$ ,  $L_n(\theta)$  为拟似然函数, 则

收稿日期: 2020-02-08

基金项目: 福建省教育厅中青年教师科研项目(JT180732)

作者简介: 刘宣(1982—), 男, 副教授, 研究方向为概率论与数理统计.

$$\mathbf{I}_n(\boldsymbol{\theta}_0) = -E\left(\frac{1}{n} \frac{\partial^2 \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right), \mathbf{K}_n(\boldsymbol{\theta}_0) = E\left(\frac{1}{n} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right).$$

为证明估计量的大样本性质,本文给出以下正则条件:

条件 1  $\mathbf{V}_n = \{v_i\}_{1 \leq i \leq n}$  独立同分布,  $E(v_1) = 0$ ,  $D(v_1) = \sigma^2$ , 存在  $\eta > 0$ , 使  $E(|v_1|^{4+\eta}) < \infty$ .

条件 2 对任意的  $n$ ,  $\mathbf{X}_n$  的元素一致有界.

条件 3  $\mathbf{W}_n$  的主对角元素为 0, 每一行(列)元素的绝对值之和一致有界.

条件 4  $\mathbf{S}_n(\rho)$  为非奇异,  $\mathbf{S}_n^{-1}(\rho)$  每一行(列)元素的绝对值之和一致有界.

条件 5 存在  $\lim_{n \rightarrow \infty} \mathbf{I}_n(\boldsymbol{\theta}_0)$  和  $\lim_{n \rightarrow \infty} \mathbf{K}_n(\boldsymbol{\theta}_0)$ .

条件 6 设  $\boldsymbol{\Omega}$  为包含真值  $\boldsymbol{\theta}_0$  的开集, 则对所有的  $\boldsymbol{\theta} \in \boldsymbol{\Omega}$  和  $j, k, l$ ,  $\frac{\partial^3 L_n(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l \partial \theta_k}$  存在, 且存在期望有限

的  $M_{jlk}$ , 使得  $\left| \frac{1}{n} \frac{\partial^3 \ln L_n(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l \partial \theta_k} \right| \leq M_{jlk}$ .

## 2 变量选择的方法及其证明

### 2.1 惩罚方法

令  $\mathbf{V}_n(\boldsymbol{\gamma}) = \mathbf{S}_n(\rho) [\mathbf{Y}_n - \mathbf{X}_n \boldsymbol{\beta}]$ ,  $\boldsymbol{\gamma} = (\rho, \boldsymbol{\beta}^T)^T$ ,  $\boldsymbol{\gamma}_0 = (\rho_0, \boldsymbol{\beta}_0^T)^T$ ,  $\mathbf{V}_n = \mathbf{V}_n(\boldsymbol{\gamma}_0)$ . 根据文献[4]的思想, 可得模型(1)的对数拟似然函数为

$$\ln L_n(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |\mathbf{S}_n(\rho)| - \frac{1}{2\sigma^2} \mathbf{V}_n^T(\boldsymbol{\gamma}) \mathbf{V}_n(\boldsymbol{\gamma}), \quad (2)$$

其中  $L_n(\boldsymbol{\theta})$  表示拟似然函数. 基于 SCAD 惩罚函数在经典线性模型下的优良表现<sup>[11]</sup>, 本文采用惩罚对数拟似然函数的方法选择变量, 并构造如下损失函数:

$$Q_n(\boldsymbol{\theta}) = -\ln L_n(\boldsymbol{\theta}) + n \sum_{j=2}^{k+2} p_{\lambda_n}(|\theta_j|). \quad (3)$$

其中  $p_\lambda(\cdot)$  为 SCAD 惩罚函数, 满足  $p'_\lambda(\delta) = \lambda \left\{ I(\delta \leq \lambda) + \frac{(a\lambda - \delta)_+}{(a-1)\lambda} I(\delta > \lambda) \right\}$ , 且  $a > 2$ ,  $\delta > 0$ ,  $p_\lambda(0) = 0$ .

**注 1** 将式(3)所得的估计称为惩罚对数拟似然估计. 为方便起见, 后续简称其为惩罚似然估计.

**注 2** 由于损失函数(3)为非凸函数, 因此需采用数值迭代的算法对其优化. 具体迭代过程为: 首先给出空间回归系数  $\rho$  的初始值(如极大似然估计值); 然后将模型(1)转换成线性模型, 再采用 LQA 算法得到回归系数  $\boldsymbol{\beta}$  的惩罚估计值; 最后把所得估计值代入损失函数将其转化成一元函数的优化问题, 并以此获得  $\rho$  的惩罚估计值. 按上述方法迭代循环到相邻两次估计值的绝对偏差小于给定的精度为止, 所得结果即为所求估计值. 迭代过程中, 调节参数  $\lambda_n$  的选择采用普通 BIC 准则.

### 2.2 大样本性质

令  $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{10}^T, \boldsymbol{\theta}_{20}^T)^T$ ,  $\boldsymbol{\theta}_{10}$  是  $s$  维向量, 包含  $s$  个非零参数,  $\boldsymbol{\theta}_{20}$  为零向量,  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$  为  $\boldsymbol{\theta}$  惩罚似然估计,  $a_n = \max\{p'_{\lambda_n}(|\theta_{j,0}|) : \theta_{j,0} \neq 0, j = 2, \dots, k+2\}$ .

**引理 1** 若满足条件 1—条件 6, 则有

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = O_p(1), -\frac{1}{n} \frac{\partial^2 \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \mathbf{I}_n(\boldsymbol{\theta}_0) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

**证明** 参考文献[4]中定理 3.1 的证明方法, 再根据  $\ln L_n(\boldsymbol{\theta}_0)$  的求导结果及公式  $\mathbf{X} = E(\mathbf{X}) + O_p(\sqrt{\text{var}(\mathbf{X})})$  即可证得引理 1.

**定理 1** 若满足条件 1—条件 6, 且  $\{p''_{\lambda_n}(|\theta_{j,0}|) : \theta_{j,0} \neq 0, j = 2, \dots, k+2\} \rightarrow 0$ , 则当  $n \rightarrow \infty$  时, 损失函数(3)依概率 1 存在极小值点  $\hat{\boldsymbol{\theta}}$ , 使得  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-\frac{1}{2}} + a_n)$ .

**证明** 令  $\alpha_n = n^{-\frac{1}{2}} + a_n$ . 类似文献[11]中估计量相合性的证明,只需证明对任意的  $\epsilon > 0$ , 存在足够大的常数  $C$  使得不等式

$$P\{\inf_{\|u\|=C} Q_n(\theta_0 + \alpha_n u) > Q_n(\theta_0)\} \geq 1 - \epsilon \quad (4)$$

成立即可. 由于  $p_{\lambda_n}(0) = 0$ , 所以

$$\begin{aligned} D_n(u) &= Q_n(\theta_0 + \alpha_n u) - Q_n(\theta_0) = \\ &= -\ln L_n(\theta_0 + \alpha_n u) + \ln L_n(\theta_0) + n \left[ \sum_{j=2}^{k+2} p_{\lambda_n}(|\theta_{j,0} + \alpha_n u_j|) - \sum_{j=2}^{k+2} p_{\lambda_n}(|\theta_{j,0}|) \right] \geq \\ &= -\ln L_n(\theta_0 + \alpha_n u) + \ln L_n(\theta_0) + n \sum_{j=2}^s \{p_{\lambda_n}(|\theta_{j,0} + \alpha_n u_j|) - p_{\lambda_n}(|\theta_{j,0}|)\}. \end{aligned}$$

上述不等式右边按泰勒公式展开,有

$$\begin{aligned} D_n(u) &\geq -\alpha_n \left( \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \right)^T u + \frac{1}{2} u^T I_n(\theta_0) u n \alpha_n^2 \{1 + o_p(1)\} + \\ &\quad \sum_{j=2}^{s+2} [n \alpha_n p'_{\lambda_n}(|\theta_{j,0}|) \operatorname{sgn}(\theta_{j,0}) u_j + n \alpha_n^2 p''_{\lambda_n}(|\theta_{j,0}|) u_j^2 \{1 + o(1)\}]. \end{aligned}$$

再由引理 1 及相关假设条件,可得

$$\begin{aligned} A_1 &= -\alpha_n \left( \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \right)^T u \geq -\left\| \alpha_n \left( \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \right)^T \right\| \cdot \|u\| = -\|u\| \cdot O_p(n \alpha_n^2), \\ A_2 &= \frac{1}{2} u^T I_n(\theta_0) u n \alpha_n^2 \{1 + o_p(1)\} = \|u\|^2 \cdot O_p(n \alpha_n^2), \\ A_3 &= \sum_{j=2}^{s+2} [n \alpha_n p'_{\lambda_n}(|\theta_{j,0}|) \operatorname{sgn}(\theta_{j,0}) u_j + n \alpha_n^2 p''_{\lambda_n}(|\theta_{j,0}|) u_j^2 \{1 + o(1)\}] \leq \\ &\quad \sum_{j=2}^{s+2} [n \alpha_n |p'_{\lambda_n}(|\theta_{j,0}|)| \cdot |u_j| + n \alpha_n^2 |p''_{\lambda_n}(|\theta_{j,0}|)| u_j^2 \{1 + o(1)\}] \leq \sqrt{s} n \alpha_n a_n \|u\| + \\ &\quad n \alpha_n^2 \max\{|p''_{\lambda_n}(|\theta_{j,0}|)| : \theta_{j,0} \neq 0\} \|u\|^2 \{1 + o(1)\} = \|u\| \cdot O_p(n \alpha_n^2) + \|u\|^2 \cdot o_p(n \alpha_n^2). \end{aligned}$$

由以上可知,存在足够大的常数  $C$  使  $A_1$  和  $A_3$  的值不超过  $A_2$  的值,由此可推出结论(4)成立,证毕.

**定理 2** 若条件 1—条件 6 成立,且  $\liminf_{n \rightarrow \infty} \liminf_{\delta \rightarrow 0^+} \frac{p'_{\lambda_n}(\delta)}{\lambda_n} > 0$ ,  $\lim_{n \rightarrow \infty} \lambda_n = 0$ ,  $\lim_{n \rightarrow \infty} \sqrt{n} \lambda_n = \infty$ , 则对满足  $\|\theta_1 - \theta_{10}\| = O_p\left(\frac{1}{\sqrt{n}}\right)$  的  $\theta_1$  和常数  $C$ , 当  $n \rightarrow \infty$  时,依概率 1 有  $Q\left(\begin{smallmatrix} \theta_1 \\ \mathbf{0} \end{smallmatrix}\right) = \min_{\|\theta_2\| \leq \frac{1}{\sqrt{n}} C} Q\left(\begin{smallmatrix} \theta_1 \\ \theta_2 \end{smallmatrix}\right)$ .

**证明** 只需证明当  $n \rightarrow \infty$ ,  $\theta_1$  满足  $\theta_1 - \theta_{10} = O_p\left(\frac{1}{\sqrt{n}}\right)$  时,依概率 1 有  $\frac{\partial Q_n(\theta)}{\partial \theta_j} < 0$  ( $0 < \theta_j < \epsilon_n$ ) 和  $\frac{\partial Q_n(\theta)}{\partial \theta_j} > 0$  ( $-\epsilon_n < \theta_j < 0$ ) 成立,其中  $\epsilon_n = \frac{1}{\sqrt{n}} C$ ,  $j = s+1, \dots, d$ . 根据泰勒公式有

$$\begin{aligned} \frac{\partial Q_n(\theta)}{\partial \theta_j} &= \frac{\partial \ln L_n(\theta)}{\partial \theta_j} - n p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j) = \frac{\partial \ln L_n(\theta_0)}{\partial \theta_j} + \sum_{l=1}^{k+2} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_j \partial \theta_l} (\theta_l - \theta_{l0}) + \\ &\quad \sum_{l=1}^{k+2} \sum_{k=1}^{k+2} \frac{\partial^3 \ln L_n(\theta^*)}{\partial \theta_j \partial \theta_l \partial \theta_k} (\theta_l - \theta_{l0}) (\theta_k - \theta_{k0}) - n p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j), \end{aligned}$$

其中  $\theta^*$  位于  $\theta$  和  $\theta_0$  之间. 根据引理 1 及条件 6,对  $j = s+1, \dots, d$ , 有

$$\begin{aligned} \frac{\partial Q_n(\theta)}{\partial \theta_j} &= n \lambda_n \left\{ \lambda_n^{-1} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \theta_j} + \lambda_n^{-1} \sum_{l=1}^{k+2} \frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_j \partial \theta_l} (\theta_l - \theta_{l0}) + \right. \\ &\quad \left. \lambda_n^{-1} \sum_{l=1}^{k+2} \sum_{k=1}^{k+2} \frac{1}{n} \frac{\partial^3 \ln L_n(\theta^*)}{\partial \theta_j \partial \theta_l \partial \theta_k} (\theta_l - \theta_{l0}) (\theta_k - \theta_{k0}) - \lambda_n^{-1} p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j) \right\} = \\ &= n \lambda_n \left\{ \lambda_n^{-1} O_p\left(\frac{1}{\sqrt{n}}\right) + \lambda_n^{-1} O_p\left(\frac{1}{\sqrt{n}}\right) + \lambda_n^{-1} O_p\left(\frac{1}{n}\right) - \lambda_n^{-1} p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j) \right\} = \end{aligned}$$

$$n\lambda_n \left\{ -\lambda_n^{-1} p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j) + O_p\left(\frac{1}{\sqrt{n}\lambda_n}\right) \right\}.$$

又因为  $\lim_{n \rightarrow \infty} \inf_{\delta \rightarrow 0^+} \liminf \frac{p'_{\lambda_n}(\delta)}{\lambda_n} > 0$ ,  $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}\lambda_n} = 0$ , 所以可知在上述条件下  $\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_j}$  的符号完全由  $\theta_j$  的符号决定, 定理 2 得证.

为方便描述以下定理, 令:

$$\boldsymbol{\Sigma} = \operatorname{diag}\{0, p''_{\lambda_n}(|\theta_{2,0}|), p''_{\lambda_n}(|\theta_{3,0}|), \dots, p''_{\lambda_n}(|\theta_{s,0}|)\},$$

$$\boldsymbol{b} = \{0, p'_{\lambda_n}(|\theta_{2,0}|) \operatorname{sgn}(\theta_{2,0}), p'_{\lambda_n}(|\theta_{3,0}|) \operatorname{sgn}(\theta_{3,0}), \dots, p'_{\lambda_n}(|\theta_{s,0}|) \operatorname{sgn}(\theta_{s,0})\}.$$

**定理 3** 若条件 1—条件 6 成立, 且  $\lim_{n \rightarrow \infty} \inf_{\delta \rightarrow 0^+} \liminf \frac{p'_{\lambda_n}(\delta)}{\lambda_n} > 0$ ,  $\lim_{n \rightarrow \infty} \lambda_n = 0$ ,  $\lim_{n \rightarrow \infty} \sqrt{n}\lambda_n = \infty$ , 则惩罚似然估计  $\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$  满足: (i) 稀疏性, 即  $P(\hat{\boldsymbol{\theta}}_2 = \mathbf{0}) \rightarrow 1$ ; (ii) 渐近正态性, 即  $\sqrt{n}\{(\boldsymbol{I}_{n1}(\boldsymbol{\theta}_{10}) + \boldsymbol{\Sigma})(\hat{\boldsymbol{\theta}}_1 -$

$\boldsymbol{\theta}_{10}) + \boldsymbol{b}\} \xrightarrow{d} N\{\mathbf{0}, \boldsymbol{K}_1(\boldsymbol{\theta}_{10})\}$ , 其中  $\boldsymbol{I}_{n1}(\boldsymbol{\theta}_{10})$  是  $\boldsymbol{I}_n(\boldsymbol{\theta}_0)$  的  $s$  维子矩阵,  $\boldsymbol{K}_1(\boldsymbol{\theta}_{10})$  是  $\boldsymbol{K}(\boldsymbol{\theta}_0)$  的  $s$  维子矩阵,  $\boldsymbol{K}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \boldsymbol{K}_n(\boldsymbol{\theta}_0)$ .

**证明** 由定理 2 知 (i) 成立. 下面证 (ii). 对  $j = 2, \dots, s$ , 有

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta})}{\partial \theta_j} &= -\frac{\partial \ln L_n(\boldsymbol{\theta})}{\partial \theta_j} + n p'_{\lambda_n}(|\theta_j|) \operatorname{sgn}(\theta_j) = \\ &= -\frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j} - \sum_{l=1}^s \left\{ \frac{\partial^2 \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j \partial \theta_l} + o_p(1) \right\} (\hat{\theta}_l - \theta_{l,0}) + \\ &= n [p'_{\lambda_n}(|\theta_{j,0}|) \operatorname{sgn}(\theta_{j,0}) + \{p''_{\lambda_n}(|\theta_{j,0}|) + o_p(1)\} (\hat{\theta}_j - \theta_{j,0})]. \end{aligned}$$

根据取得极值的必要条件, 有

$$\begin{aligned} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j} &= \sum_{l=1}^s \left\{ -\frac{\partial^2 \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j \partial \theta_l} - o_p(1) \right\} (\hat{\theta}_l - \theta_{l,0}) + \\ &= n [p'_{\lambda_n}(|\theta_{j,0}|) \operatorname{sgn}(\theta_{j,0}) + \{p''_{\lambda_n}(|\theta_{j,0}|) + o_p(1)\} (\hat{\theta}_j - \theta_{j,0})], \end{aligned}$$

进而可得

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j} &= \sum_{l=1}^s \left\{ -\frac{\partial^2 \ln L_n(\boldsymbol{\theta}_0)}{\partial \theta_j \partial \theta_l} \right\} \frac{1}{n} \sqrt{n} (\hat{\theta}_l - \theta_{l,0}) + \\ &= \sqrt{n} p''_{\lambda_n}(|\theta_{j,0}|) (\hat{\theta}_j - \theta_{j,0}) + \sqrt{n} [p'_{\lambda_n}(|\theta_{j,0}|) \operatorname{sgn}(\theta_{j,0})] + o_p(1). \end{aligned} \quad (5)$$

令  $\boldsymbol{K}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \boldsymbol{K}_n(\boldsymbol{\theta}_0)$ . 考虑到拟对数似然函数的偏导数可写成扰动项二次型的线性表达式, 再结

合文献[12]提出的二次型中心极限定理, 有  $\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{K}(\boldsymbol{\theta}_0))$ . 因此, 由式(5)可知定理 3 成立.

从定理 1—定理 3 可知, 由最小化损失函数(3)不仅可以得到未知参数的相合估计, 还可以把零系数压缩为零从而可去掉不显著的相应变量, 实现变量选择.

### 3 蒙特卡洛模拟

数据基于模型(1)产生. 其中: 协变量  $\mathbf{X}_n$  来自满足均值为  $\mathbf{0}$ 、协方差阵为  $(0.5^{|i-j|})_{8 \times 8}$  的 8 维正态分布; 回归系数  $\boldsymbol{\beta}$  取  $\{3, 2, 0, 0, 0, 1, 0, 0\}$ ; 空间自回归系数  $\rho$  取  $\{0, 0.3, 0.7\}$ ; 空间权重矩阵为 rook 矩阵 (边相邻的单元设为 1, 否则设为 0);  $\mathbf{V}_n \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ ,  $\sigma = 1, 2$ .

用  $C$  表示 100 次模拟估计中零系数出现的平均数, 用  $I$  表示 100 次模拟估计中非零系数出现的平均数, 用误差平方和的中位数 (MSE) 反映估计的精度, 惩罚函数中的调节参数  $a = 3.7$ ,  $\lambda$  的选取采用 BIC

准则.用“Spatial error model”表示本文方法空间误差模型下模拟的结果,“Linear model”表示忽视空间效应直接利用传统 SCAD 惩罚方法所产生的结果.

由表 1 和表 2 中的  $C$  值和  $I$  值可看出,随着样本量的增加,零系数的正确识别率越来越高,非零系数的错误识别率迅速降为 0.该结果和惩罚估计量的理论性质一致.由表 1 和表 2 中的  $MSE$  值可以看出,惩罚估计量随样本量的增加其精度得到显著改善.由表 1 和表 2 中的“Linear model”栏可以看出,若忽略原本存在的空间效应而直接使用线性模型下相应的变量选择方法会大幅增加误差,因此空间效应不可忽视.

表 1  $\sigma=1$  时变量选择的模拟结果

$\sigma=1$		Spatial error model			Linear model		
$n$	$\rho$	$C$	$I$	$MSE$	$C$	$I$	$MSE$
50	0.6	4.59	0.07	0.162 5	4.61	0.02	0.447 3
	0.0	5.62	0.00	0.132 9	4.60	0.00	0.138 5
150	0.6	4.98	0.00	0.041 2	4.90	0.00	0.200 7
	0.0	5.96	0.00	0.034 7	4.97	0.00	0.034 2

表 2  $\sigma=2$  时变量选择的模拟结果

$\sigma=2$		Spatial error model			Linear model		
$n$	$\rho$	$C$	$I$	$MSE$	$C$	$I$	$MSE$
50	0.6	4.55	0.30	1.459 6	4.67	0.29	3.539 0
	0.0	5.53	0.20	1.177 2	4.63	0.12	1.156 3
150	0.6	4.91	0.00	0.252 7	4.86	0.01	3.011 2
	0.0	5.89	0.00	0.258 0	4.90	0.00	0.251 3

## 参考文献:

- [1] CLIFF A D, ORD J K. Spatial Autocorrelation[M]. London: Pion Ltd, 1973.
- [2] ANSELIN L. Spatial econometrics: methods and models[M]. The Netherlands: Kluwer Academic Publishers, 1988.
- [3] LESAGE J, PACE R. Introduction to spatial econometrics[M]. Boca Raton: CRC Press, 2009.
- [4] LEE L F. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models[J]. Econometrica, 2004,72:1899-1925.
- [5] ANSELIN L. Non-nested tests on the weight structure in spatial autoregressive models: Some Monte Carlo results [J]. Journal of Regional Science, 2006,26(2):267-284.
- [6] LESAGE J P, PARENT O. Bayesian model averaging for spatial econometric models[J]. Geographical Analysis, 2007,39:241-267.
- [7] 李坤明,陈建宝. 非参数空间滞后模型的广义矩估计[J]. 高校应用数学学报: A 辑, 2018,33(2):140-156.
- [8] 郭双,魏传华. 空间自回归模型的变量选择[J]. 中央民族大学学报(自然科学版), 2015,24(3):92-96.
- [9] XIE T, CAO R, DU J. Variable selection for spatial autoregressive models with a diverging number of parameters [J]. Statistical Papers, 2018,1:1-21.
- [10] LIU X, CHEN J, CHENG S. A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model[J]. Spatial Statistics, 2018,25:86-104.
- [11] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American Statistical Association, 2001,96:1348-1360.
- [12] KELEJIAN H H, PRUCHA I R. On the asymptotic distribution of the Moran I test statistic with applications [J]. Journal of Econometrics, 2001,104:219-257.