

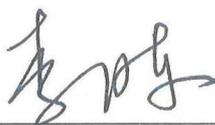
两类空间面板数据模型的变量选择

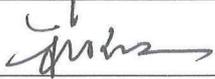
专业名称: 数学

申请人: 唐莹莹

指导老师: 秦永松 教授

论文答辩委员会

主席: 

委员: 





两类空间面板数据模型的变量选择

年级: 2018 级 专业: 数学 研究方向: 概率论与数理统计
研究生: 唐莹莹 导师: 秦永松教授

摘 要

近几十年来,随着科学技术的不断进步,由人们的生活或者发生的经济行为所产生的空间面板数据被大量地采集与记录.这些空间面板数据之中存在着某种相联关系,并非完全独立且具有不可分割的相关性,应运而生的空间面板数据模型便是一种挖掘空间面板数据信息的重要模型之一,因其自身优势及广泛应用已经成为计量经济学的研究热点.

在建立空间面板数据模型时,通过合适的方式选择关键变量会使得空间面板数据模型具备更好地解释能力.从已有文献来看,学者们研究空间面板数据模型的参数估计等方面的文献较多,而关于空间面板数据模型变量选择问题的相关文献较少.在这样的背景下,本文将空间面板数据模型和变量选择的理论知识相结合,以研究线性空间面板自回归模型和空间面板误差模型的变量选择问题.

本文的主要内容分为四章:

第一章为绪论,简要介绍问题的研究背景、空间面板数据模型和变量选择方法的研究概况,以及本文的主要内容与结构.

第二章为预备知识,介绍空间面板数据模型的定义、空间权重矩阵的几种定义方式以及几种常见的变量选择方法.

第三章研究一阶空间面板自回归模型的Adaptive LASSO 变量选择问题,并证明了在一定条件下,该方法具有良好的渐近性质,同时通过蒙特卡洛方法验证该方法的有限样本性能.

第四章研究一阶空间面板误差模型的Adaptive LASSO 变量选择问题,并证明了该方法具有良好的Oracle渐近性质,且蒙特卡洛模拟结果显示了该方法能够有效地识别零系数和非零系数.

关键词: 空间面板自回归模型; 空间面板误差模型; 变量选择; Adaptive LASSO方法

Variable Selection of Two Kinds of Spatial Panel Data Models

Grade: 2018 Major: Math Research field: Probability and Statistics

Graduate: Tang Yingying Supervisor: Qin Yongsong

ABSTRACT

In recent decades, with the continuous progress of science and technology, a large number of spatial panel data generated by people's life or economic behavior have been collected and recorded. Spatial panel data model is one of the important models for mining spatial panel data information.

When building spatial panel data model, it has become a research hotspot of econometrics to select key variables in an appropriate way to make spatial panel data model have better explanatory ability. In this context, this paper combines the spatial panel data model with the theoretical knowledge of variable selection to study the variable selection of linear spatial panel autoregressive model and spatial panel error model.

The main content of this paper is divided into four chapters:

The first chapter is the introduction, which briefly introduces the research background, spatial panel data model and variable selection method, as well as the main content and structure of this paper.

The second chapter introduces the definition of spatial panel data model, several definitions of spatial weight matrix, and several common variable selection methods.

In the third chapter, we study the adaptive lasso variable selection problem of first-order spatial panel autoregressive model, and prove that the method has good asymptotic properties under certain conditions. The finite sample performance of the method is verified by Monte Carlo method.

In Chapter 4, the adaptive lasso variable selection of the first-order spatial panel error model is studied. The results of Monte Carlo simulation show that the method can effectively identify zero and non-zero coefficients.

Key words: Spatial Panel Autoregressive Model; Spatial Panel Error Model; Variable Selection; The Adaptive LASSO Method

目 录

摘要	I
ABSTRACT	II
第一章 绪 论	1
§1.1 研究背景及意义	1
§1.2 研究现状	1
§1.2.1 空间面板数据模型的研究概况	1
§1.2.2 变量选择方法的研究概况	2
§1.3 本文的内容结构	3
第二章 预备知识	4
§2.1 空间面板数据模型	4
§2.2 空间权重矩阵	4
§2.3 降维方法	6
第三章 空间面板自回归模型的 Adaptive LASSO 变量选择	8
§3.1 引言	8
§3.2 主要结果	9
§3.3 数值模拟	10
§3.4 引理及定理的证明	12
第四章 空间面板误差模型的 Adaptive LASSO 变量选择	18
§4.1 引言	18
§4.2 主要结果	19
§4.3 数值模拟	20
§4.4 引理及定理的证明	22
结论及展望	28
参考文献	29
致 谢	

声 明

第一章 绪论

本章将简述问题的研究背景,并概述空间面板数据模型和变量选择方法的研究现状,以及本文的内容和结构.

§1.1 研究背景及意义

在被称为“大数据时代”的21世纪,采集得到的数据量正在以指数的方式不断增长.面对着这些海量数据,如何开展有效的挖掘和合理的转化使用已经成为一个新问题.技术进步被认为是经济增长的最终来源,在地理信息收集技术的加持下发现空间因素对经济问题的影响越来越大,应运而生的空间面板数据模型因其自身优势及广泛应用,引起了学者们的积极关注并成为空间计量经济学的研究热点之一.由人类活动所产生的这些空间面板数据具有更多的信息和事物相关性,从中可以发现潜在的空间和非空间之间的关联模式和检验事件所具有的普遍性,而且在统计分析中空间面板数据模型能得到更加有效的估计结果.

随着数据采集能力的提升,也带来了越变越多的变量(也称特征),但其中可能存在不显著且系数值较小的自变量使得估计性能迅速下降.在进行实际的空间面板数据模型拟合时,如果直接利用繁多的自变量进行建模会面临着过多变量的参与或是混杂变量的掺入的困境,使得模型回归效果不理想或在模型的解读上遇到困难.

如何消除混合因素的干扰,如何将空间面板数据模型中的众多自变量进行精简,通过何种方式去选择关键变量使模型具备更好地解释能力,已经成为计量经济学研究的热点问题.现有的研究主要针对的是空间面板数据模型的参数估计的文献较多,而有关空间面板数据模型变量选择的文献较少.在这样的背景下,本文将空间面板数据模型和变量选择的理论知识相结合,以研究两类空间面板数据模型的变量选择问题,运用合适的方法为纷繁复杂的自变量进行适当降维,为空间面板计量提供便利.

§1.2 研究现状

本节分为二小节;1.2.1小节简要介绍关于空间面板数据模型的研究现状,1.2.2小节介绍有关变量选择方法的研究概况.

§1.2.1 空间面板数据模型的研究概况

空间面板数据模型是空间计量经济学中比较重要的模型之一,相比于时间序列模型和截面数据模型,它能够更加全面的考虑数据的空间依赖性与空间异质性,在以往的统计学方

法中, 往往都是需要样本独立这个基本假设, 而空间数据往往都是具有空间关联性的, 即数据往往都不是独立的, 那么用以往的统计学方法来处理空间数据的问题就会忽略许多数据本身的性质从而产生很大的误差, 为了解决这个问题, 学者们找到新的模型与方法来处理空间数据.

1988年Anlisen出版的《空间计量经济学：模型与方法》^[1]将空间计量经济学理论做出了系统的整理, 为后面学者们对空间计量经济学的研究打下了坚实的基础. 在静态空间面板数据模型研究上, Kapoor等^[2]研究了具有空间自相关误差项的线性回归空间面板数据模型的参数估计, 提出了可估计模型参数的广义矩(GMM)估计和广义最小二乘(FGLS)估计, 并证明了其渐近正态性. Li等^[3]研究了空间面板误差数据模型的经验似然性推断. Baltagi等^[4]对空间面板数据模型的误差项进行了一般性的设定, 发现面板数据具有异质性并在模型中引入个体效应项(固定效应和随机效应)来控制所有不随时间变化且反应空间特质的量. Sun和Malikov^[5]研究了固定效应下空间自回归面板数据模型的估计, 将模型转化为半参数加性模型以消除模型中未观察到的固定效应. 在动态空间面板数据模型研究上, Baltagi等^[6]研究了具有空间相关扰动的动态自回归空间面板数据模型的估计和预测性能, 并通过在新经济地理学中应用了该方法. Lee和Yu^[7]研究了具有时间和个体固定效应的空间动态面板数据模型的极大似然估计和拟极大似然(QML)估计. Jin等^[8]研究了一般空间动态面板数据模型的估计与模型选择问题.

§1.2.2 变量选择方法的研究概况

在建立有效的空间计量模型时, 选择合适的变量也是至关重要的环节之一. 近年来, 惩罚损失函数型变量选择方法受到越来越多的关注. Tibshirani^[9]最早提出了LASSO方法, 并将其应用在线性回归模型的变量选择工作中. 现在, LASSO方法已经变成一种流行的正则化技术而且不断地改进中. Fan和Li^[10]提出了绝对偏差(SCAD)惩罚的变量选择方法, 并建立了估计值的Oracle属性, 以辨别所使用的惩罚函数是否能拥有良好的无偏性、稀疏性和连续性的估计结果. Fan和Peng^[11]概括了Fan和Li^[10]的工作并将其扩展到以允许参数个数随着样本量的增加而增加的情形中. Zou^[12]在Tibshirani^[9]的基础上提出了Adaptive LASSO方法, 改进后的方法能够有效减少模型参数估计的有偏性, 而且具有Oracle性质. Xie等^[13]在研究高维数据部分线性模型的SCAD惩罚变量选择过程中提出了多项式回归样条估计. 王大荣等^[14]做了线性回归模型变量选择方法的综述工作. Zou^[15]结合了二次正则化的强度和自适应加权的套索收缩, 提出了一种自适应弹性网. 自从惩罚损失函数型变量选择方法出现后, 已经被广泛应用到医学、气象、经济、地理等领域之中.

研究空间计量模型变量选择问题的相关文献较少. 在空间计量模型的变量选择问题

中, Wang等^[16]利用了Kullback-Laible信息得出空间AIC, 研究了一种空间AIC的空间自回归模型变量选择方法, 并证明了空间AIC是渐近最优的. Zhang等^[17]从贝叶斯的角度研究了具有自回归扰动的空间自回归模型的变量选择问题, 他们将基于线性模型BIC准则的子集选择方法应用于空间模型, 并证明了在一定条件下BIC准则在空间模型中具有良好的渐近性质. 张元庆和陶志鹏^[18]在广义嵌套空间模型的变量选择中提出了广义空间信息准则. 郭双和魏传华^[19]研究了空间自回归模型的 Adaptive LASSO 变量选择. 王晓瑞^[20]研究了空间自回归模型和部分线性空间自回归模型的变量选择问题. 谢琍等^[21]对比了线性空间自回归模型在不同惩罚函数所得到的参数估计结果. 李春雨^[22]讨论一阶带误差自相关的空间自回归模型和部分线性空间自回归模型的 Adaptive LASSO 变量选择的渐近性质和相合性. Zhu等^[23]采用了带自适应套索的惩罚似然估计对空间线性模型的回归变量和邻域结构进行选择. Liu^[24]提出了以最大似然(ML)估计量用作初步估计量, 结合Wang和Leng^[25]研究的最小二乘近似(LSA)方法来实现对空间自回归模型的惩罚估计. Wang^[26]和王康宁^[27]还研究了空间半参数变系数模型的变量选择问题.

§1.3 本文的内容结构

本文研究两类空间面板数据模型的Adaptive LASSO变量选择, 并证明该方法在空间面板自回归模型和空间面板误差模型的变量选择和参数估计问题上都具有良好的性质.

研究动机: 现有的研究更多的是空间面板数据模型的参数估计和截面空间数据模型的变量选择问题, 而空间面板数据模型变量选择的相关文献较少.

本文的结构分为以下部分:

第一章为绪论, 简要介绍问题的研究背景、空间面板数据模型和变量选择方法的研究概况, 以及本文的内容与结构.

第二章为预备知识, 介绍空间数据模型的定义, 空间权重矩阵的几种定义方式以及常见的变量选择方法.

第三章为一阶空间面板自回归模型的Adaptive LASSO变量选择. 在一阶空间面板自回归模型中, 讨论在Adaptive LASSO惩罚下模型的变量选择和参数估计问题, 证明该方法具有良好的Oracle渐近性质, 并通过蒙特卡洛方法验证该方法的有限样本性能.

第四章为一阶空间面板误差模型的Adaptive LASSO变量选择, 构造 Adaptive LASSO 估计函数且证明该方法具有良好的Oracle渐近性质, 并通过数值模拟验证了该方法能够有效的进行零系数和非零系数的识别.

本文的创新点: 在一阶空间面板自回归模型和一阶空间面板误差模型中, 构造Adaptive LASSO惩罚函数以研究空间面板数据模型的变量选择问题.

第二章 预备知识

本章简要介绍空间面板数据模型的定义, 空间权重矩阵的几种定义方式以及常用的几种变量选择方法.

§2.1 空间面板数据模型

空间面板数据模型是一种穿越空间单元来捕获空间交互和时间的推移所产生的模型, 在各类统计分析中的应用十分广泛, 在地理、经济、生态环境等领域有着广阔的应用前景, 在许多不同的行业也发挥着显著的作用. 一般的空间面板数据模型^[1]:

$$\begin{cases} y_t = X_t\beta + \rho W_n y_t + \mu_t, \\ \mu_t = \lambda M_n \mu_t + \varepsilon_t, t = 1, 2, \dots, T, \end{cases} \quad (2-1)$$

其中, y_t 是 $n \times 1$ 维响应变量, X_t 是 $n \times k$ 维固定设计自变量, β 是 $k \times 1$ 回归系数向量, μ_t 是自相关误差项, W_n 和 M_n 是预先给定的 $n \times n$ 维空间权重矩阵, λ 和 ρ 为空间自相关系数, $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tn})'$ 为 $n \times 1$ 维误差向量, 其分量的均值为 0, 协方差为 σ^2 , n 和 T 分别表示空间单元数和时间单位数.

当 $\lambda = 0$ 时, 模型(2-1)为空间面板自回归模型:

$$y_t = X_t\beta + \rho W_n y_t + \mu_t, \quad (2-2)$$

当 $\rho = 0$ 时, 模型(2-1)为空间面板误差模型:

$$\begin{cases} y_t = X_t\beta + \mu_t, \\ \mu_t = \lambda M_n \mu_t + \varepsilon_t. \end{cases} \quad (2-3)$$

§2.2 空间权重矩阵

在空间计量模型中, 观察结果与空间中的特定位置相关联, 通常涉及包含具有多个空间单元的时间序列观测数据. 每对空间单元之间相互作用的结构可以通过空间权重矩阵表示. 构造一个 $(n \times n)$ 阶的空间计量权重矩阵 W 主要有两种方法, 一种方法是基于外生变量方法来构造的(即用样本在地理分布上的信息来计算空间权重矩阵), 另一种方法是基于数据统

计量方法来构造的. 空间权重矩阵 W 形如下式:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix},$$

其中, 为了排除把自己当作是自己的一个邻居, 令对角线上的元素都为零, 而矩阵的每行上的非零元素定义为“邻域”, 这些非零元素意味着 W 的第 (i, j) 个元素在空间权重矩阵中个体 i 与个体 j 的空间依赖程度. 常见的空间权重矩阵 W 有:

(1)半径距离权重矩阵:

$$w_{ij} = \begin{cases} 1, 0 \leq d_{ij} \leq \bar{d}_i \\ 0, d_{ij} > \bar{d}_i, \end{cases}$$

其中, d_{ij} 表示区域 i 和区域 j 质心的欧氏距离. 距离越远, 空间权重系数越小, 空间相关性越差, \bar{d}_i 是按照研究需求来调整的距离值.

(2) K阶近邻权重矩阵:

$$w_{ij} = \begin{cases} 1, \text{如果区}j \text{是区}i \text{的最近的}k \text{个邻居,} \\ 0, \text{其它.} \end{cases}$$

(3)二进制邻接权重矩阵:

$$w_{ij} = \begin{cases} 1, \text{如果区域}i \text{与区域}j \text{相邻接壤,} \\ 0, \text{如果区域}i \text{与区域}j \text{不相邻接壤.} \end{cases}$$

其中, 若空间个体 i 与空间个体 j 有公共边可为Rook邻接型权重矩阵; 若空间个体 i 与空间个体 j 有公共点可为Bishop邻接型权重; 若空间个体 i 与空间个体 j 有公共点或者公共边可为Queen邻接型权重等. 例如, 给定下面的区域A:

1	2
3	4

图 2-1 区域A

区域A对应的 Rook 加权、 Queen 加权的邻接矩阵如下:

$$W_1 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, W_2 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

在模型中使用的空间权重矩阵通常是经过“行标准化”处理的,如下式:

$$\begin{cases} w_{ij} \geq 0, & \forall 1 \leq i, j \leq n \\ w_{ij} = 0, & \forall 1 \leq i \leq n \\ \sum_{j=1}^n w_{ij} = 1, & \forall 1 \leq i \leq n. \end{cases}$$

例如,邻接矩阵 W_1 和 W_2 进行“行标准化”后的形式如下式:

$$\tilde{W}_1 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \tilde{W}_2 = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

§2.3 降维方法

合适的变量选择方法可以降低模型复杂度,同时使得模型有更好的解释效果,还可以减弱过拟合的程度以得到更高的精度.特征选择的常用方法有过滤法、包装法和嵌入法.过滤法时常应用于预处理阶段,不依赖于任何机器学习算法,它的使用是基于统计检验的得分作为筛选条件(检验特征和响应变量的相关性).包装法则会仅介用特征的一个子集来训练模型,并利用之前模型的结果来判断是否需要增删新的特征,如常见的前向选择法,后向剔除法,迭代剔除法等.嵌入法是在损失函数(即优化目标)加入惩罚项,在遍历求解未知参数过程中考虑到系数的大小,可以通过调整缩减系数(惩罚系数 λ)让影响较小的变量的系数衰减到0,以防止过拟合并提高泛化能力.关于不同惩罚函数的具体讨论可参考Fan和Li^[10]等的研究成果.

考虑模型 $Y = X\beta + \varepsilon$,其中 Y 是 $n \times 1$ 维的响应变量, X 是 $n \times p$ 维的自变量, β 是 $p \times 1$ 维的参数, $\varepsilon \sim (0, \sigma^2 I_n)$.在该模型建立时,可以通过求解 $\|Y - X\beta\|^2$ 函数的最小值,得到参数 β 的最小二乘估计: $\hat{\beta} = (X'X)^{-1}X'Y$.若将变量选择和参数估计相结合,此时模型的目标函数可以变化为:

$$\arg \min_{\beta} \{ \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p p_{\lambda}(|\beta_j|) \},$$

其中, $p_{\lambda}(\cdot)$ 是惩罚函数, $\lambda > 0$ 为可调整的惩罚参数,当 λ 越大时被选入的 β 的个数

就越少. 若 \mathbf{a} 是一个 q 维向量, $\mathbf{a} = (a_1, \dots, a_q)^T$, 设 $|\mathbf{a}|_\infty = \max_{1 \leq k \leq q} |a_k|$, $|\mathbf{a}|_1 = \sum_{k=1}^q |a_k|$ 和 $|\mathbf{a}|_2 = (\sum_{k=1}^q a_k^2)^{1/2}$ 分别为 L_∞ , L_1 和 L_2 . 常用的惩罚函数 $p_\lambda(\cdot)$ 有

1. LASSO 回归惩罚 (L_1 正则化): $p_\lambda(|\theta|) = \lambda|\theta|$;

2. Ridge 回归惩罚 (L_2 正则化): $p_\lambda(|\theta|) = \lambda|\theta|^2$;

3. Adaptive LASSO 回归惩罚: $p_\lambda(|\theta|) = \frac{1}{|\hat{\theta}|} \lambda|\theta|$, 其中 $\hat{\theta}$ 是 θ 的一个初始估计;

4. SCAD 惩罚 (通常令 $a = 3.7$):

$$p_\lambda(|\theta|) = \begin{cases} \lambda|\theta|, & |\theta| \leq \lambda; \\ -\frac{(\theta^2 - 2a\lambda|\theta|)}{2(a-1)}, & \lambda < |\theta| < a\lambda; \\ \frac{(a+1)\lambda^2}{2}, & |\theta| \geq a\lambda. \end{cases}$$

第三章 空间面板自回归模型的 Adaptive LASSO 变量选择

在本章中, 我们研究空间面板自回归模型的 Adaptive LASSO 变量选择问题, 假设有 n 个空间单位和 T 个时间段, 我们考虑以下模型:

$$y_t = \rho W_n y_t + X_t \beta + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (3-1)$$

其中, y_t 是 $n \times 1$ 维响应变量的观测值, X_t 是 $n \times k$ 维固定设计自变量的观测值, β 是 $k \times 1$ 维回归系数向量, W_n 是预先给定的 $n \times n$ 维空间权重矩阵, ρ 为空间自相关系数, $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tn})'$ 为 $n \times 1$ 维误差向量, ε_{ti} 为独立同分布的随机误差项, 且满足 $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2 I_n$, 其中 $0 < \sigma^2 < \infty$, I_n 为单位矩阵.

§3.1 引言

我们继续讨论模型 (3-1). 当 $t = 1, 2, \dots, T$ 时, 模型 (3-1) 可以写成如下矩阵形式:

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_T \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} X_1 \beta \\ X_2 \beta \\ \vdots \\ X_T \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix},$$

其中, $A_t = (I_n - \rho W_n)$, 令 $A_{nT}(\rho) = [I_{nT} - \rho(I_T \otimes W_n)]$ 且设 $A_{nT}(\rho)$ 是非奇异的, 模型(3-1)化简成:

$$A_{nT}(\rho)Y = X\beta + \varepsilon, \quad (3-2)$$

其中, $Y = (y_1', \dots, y_T')'$ 是 $(nT) \times 1$ 维因变量向量, $X = (X_1', \dots, X_T')'$ 是 $(nT) \times k$ 维自变量矩阵, ε 为 $(nT) \times 1$ 维随机扰动项, 需要估计的参数为 $\theta = (\beta', \rho, \sigma^2)' \in R^{k+2}$, 令 $\theta_0 = (\beta_0', \rho_0, \sigma_0^2)'$ 为 θ 的真值.

模型 (3-2)的对数似然函数为:

$$\ln L_{nT}(\theta) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2) + \ln |A_{nT}(\rho)| - \frac{1}{2\sigma^2} \varepsilon' \varepsilon, \quad (3-3)$$

其中, $\varepsilon = A_{nT}(\rho)Y - X\beta$. 根据文献[29] 提出的拟极大似然估计法, 对于给定的 ρ , 关于 β 和 σ^2 的拟极大似然估计为:

$$\hat{\beta} = (X'X)^{-1} X' A_{nT}(\rho)Y,$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{nT} [A_{nT}(\rho)Y - X\hat{\beta}]' [A_{nT}(\rho)Y - X\hat{\beta}] \\ &= \frac{1}{nT} Y' A_{nT}(\rho)' M A_{nT}(\rho) Y,\end{aligned}$$

其中, $M = [I_{nT} - X(X'X)^{-1}X']$, 将 $\hat{\beta}$, $\hat{\sigma}^2$ 代入(3-3)式可得到关于 ρ 的对数似然函数:

$$\tilde{f} = -\frac{nT}{2} (\ln(2\pi) + 1) - \frac{nT}{2} \ln(\hat{\sigma}^2) + \ln |A_{nT}(\rho)|, \quad (3-4)$$

对(3-4)式进行最大化处理后可得到新的 $\hat{\rho}$, 再将 $\hat{\rho}$ 代入模型(3-2)的拟极大似然函数式(3-3)中. 且令 $\bar{Y} = A_{nT}(\hat{\rho})Y$, $\bar{X} = X$, 模型 (3-2)可记为: $\bar{Y} = \bar{X}\beta + \varepsilon$.

为了得到系数的稀疏估计, 考虑 Zou^[12] 提出的 Adaptive LASSO 变量选择方法, 选择一个 $r > 0$, 定义惩罚权重向量为 $\omega_j = 1/|\hat{\beta}_j|^r$, 此时的惩罚目标函数为:

$$\tilde{\beta} = \arg \min_{\beta} \left\{ (\bar{Y} - \bar{X}\beta)' (\bar{Y} - \bar{X}\beta) + \sum_{j=1}^k \lambda_{nT} \omega_j |\beta_j| \right\}, \quad (3-5)$$

其中, $r > 0$ 为调节参数, $\hat{\beta}$ 为 β 的拟极大似然估计, $\lambda_{nT} > 0$ 为可调整的惩罚参数, 当 λ_{nT} 越大时被选入的 β 个数越少, 可通过交叉验证法或广义交叉验证法来选择 λ_{nT} 的值.

给定备选子集 $\mathcal{A} = \{j : \beta_j \neq 0\}$, 不失一般性, 假定 $\mathcal{A} = \{1, \dots, k_0\}$, $\mathcal{A}^c = \{k_0 + 1, \dots, k\}$. 定义 $\tilde{\mathcal{A}} = \{j : \tilde{\beta}_j \neq 0\}$, $\beta = (\beta'_{\mathcal{A}}, \beta'_{\mathcal{A}^c})'$.

§3.2 主要结果

为了方便起见, 引入下述符号:

$$\begin{aligned}G &= (I_T \otimes W_n) A_{nT}(\rho)^{-1}, & G^S &= G + G', & \Psi &= \lim_{n \rightarrow \infty} \frac{1}{nT} X'X \\ \nu &= \lim_{n \rightarrow \infty} \frac{1}{nT} X'(GX\beta_0), & v_1 &= \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(G), & v_2 &= \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(G^S G), \\ \tau &= \lim_{n \rightarrow \infty} \frac{1}{nT} (GX\beta_0)'(GX\beta_0), & \Psi &= \begin{pmatrix} \Psi_{11} & \Psi_{22} \\ \Psi_{33} & \Psi_{44} \end{pmatrix},\end{aligned}$$

其中, Ψ_{11} 是 $k_0 \times k_0$ 维矩阵, Ψ_{22} 是 $k_0 \times (k - k_0)$ 维矩阵, Ψ_{33} 是 $(k - k_0) \times k_0$ 维矩阵, Ψ_{44} 是 $(k - k_0) \times (k - k_0)$ 维矩阵.

为了探讨新估计的渐近正态性, 需作如下假设:

(A1): n 趋于无穷大, T 是有限常数.

(A2): $\{X_t, t = 1, 2, \dots, T\}$ 是固定设计自变量的观测值且一致有界.

(A3): $\{\varepsilon_{ti}, i = 1, 2, \dots, n, t = 1, 2, \dots, T\}$ 是均值为零且方差为 σ^2 的 *i.i.d* 随机变量, 对于某些 $\delta > 0$, 有 $E|\varepsilon_{ti}|^{4+\delta}$ 存在且有界.

(A4): 矩阵 $A_{nT}(\rho)$ 为非奇异矩阵, W_n 具有零对角线且进行“行标准化”处理, $A_{nT}(\rho)^{-1}$ 元

素在 L_1 范数上行和与列和均存在且有界.

(A5): 在一个紧凸集 Δ 上, $A_{nT}^{-1}(\rho)$ 的行和与列和对于 ρ 一致有界, 且 $\Delta \in (-1, 1)$.

(A6): 假设下列极限存在 $\Psi = \lim_{n \rightarrow \infty} \frac{1}{nT} X'X$, $\nu = \lim_{n \rightarrow \infty} \frac{1}{nT} X'(GX\beta_0)$, $v_1 = \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(G)$,
 $v_2 = \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(G^S G)$, $\tau = \lim_{n \rightarrow \infty} \frac{1}{nT} (GX\beta_0)'(GX\beta_0)$.

(A7): 假设极限 $\Sigma = \lim_{n \rightarrow \infty} -\frac{1}{nT} E\left(\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$, $\Omega = \lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta'}\right)$ 存在, 且 Σ 和 Ω 是非奇异的.

注1: 假设(A1)-(A5)是空间数据模型的常见假设, 与Li^[3], Lee^[7]的假设相似.

注2: $E\left(-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$

$$= \begin{pmatrix} \frac{1}{nT\sigma_0^2} X'X & \frac{1}{nT\sigma_0^2} X'(GX\beta_0) & 0 \\ \frac{1}{nT\sigma_0^2} (GX\beta_0)'X & \frac{1}{nT} \text{tr}(G^S G) + \frac{1}{nT\sigma_0^2} (GX\beta_0)'(GX\beta_0) & \frac{1}{nT\sigma_0^2} \text{tr}(G) \\ 0 & \frac{1}{nT\sigma_0^2} \text{tr}(G) & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

可令

$$\Sigma = \begin{pmatrix} \frac{1}{\sigma_0^2} \Psi & \frac{1}{\sigma_0^2} \nu & \mathbf{0} \\ \frac{1}{\sigma_0^2} \nu' & \frac{1}{\sigma_0^2} \tau + v_2 & \frac{1}{\sigma_0^2} v_1 \\ \mathbf{0} & \frac{1}{\sigma_0^2} v_1 & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

定理3.1 假设条件(A1)-(A7)成立, 定义 $\tilde{\mathcal{A}} = \{j : \tilde{\beta}_j \neq 0\}$, 如果当 $\lambda_{nT} \cdot (nT)^{-1/2} \rightarrow 0$ 且 $\lambda_{nT} \cdot (nT)^{(r-1)/2} \rightarrow \infty$, r 为调节参数, Ψ 是非奇异的, 则模型的 Adaptive LASSO 估计满足:

(i) 渐近正态性: $\sqrt{nT}(\tilde{\beta}_{\mathcal{A}} - \beta_{\mathcal{A}}) \xrightarrow{d} \Psi_{11}^{-1} \xi_{\mathcal{A}}$, $\xi \sim N(0, D\Omega D')$,

其中, $\xi_{\mathcal{A}}$ 是 ξ 前 k_0 维子向量, Ψ_{11} 是 $k_0 \times k_0$ 阶矩阵, $D_{k \times (k+2)} = (\nu d_{21} - \sigma_0^2 I_k, \nu d_{22}, \nu d_{23})$, d_{ij} 是矩阵 Σ^{-1} 的第 i 行第 j 列元素, $i, j = 1, 2, 3$.

(ii) 相合性: $\lim_{n \rightarrow \infty} P(\tilde{\mathcal{A}} = \mathcal{A}) = 1$.

§3.3 数值模拟

本节考虑在不同的空间自相关系数情况下, 对蒙特卡罗模拟的结果进行比对分析. 考虑如下模型:

$$y_t = \rho W_n y_t + X_t \beta + \varepsilon_t, \quad t = 1, \dots, 8, \quad (3-6)$$

其中, 令 $X_1 = (x_{11}, x_{12}, \dots, x_{1n})'$, $x_{1i} = \frac{i}{n+1}$, $X_2 = (x_{21}, x_{22}, \dots, x_{2n})'$, $x_{2i} = \frac{i}{n+2}, \dots, X_8 = (x_{81}, x_{82}, \dots, x_{8n})'$, $x_{8i} = \frac{i}{n+8}, 1 \leq i \leq n$, 将 W_n 设置为常用的“Queen”型空间权重矩阵, 假定 $\beta = (0.6, 0, 1, 0, 0, 0.8, 0, 0)$, 惩罚权重系数 $w_j = 1/|\hat{\beta}_j|, j = 1, \dots, 8$. 选取不同的情形进行考察: (1)两种不同样本量 n 为 225 和 400; (2)两种不同 ε 误差项服从的分布分别为

$N(0, 1), t(5)$; (3)三种不同的变量选择方法: LASSO, ALASSO, SCAD; (4)四种不同的自相关系数 ρ 为 $-0.85, 0.2, 0.5, 0.8$.

本节利用“留一法(leave-one-out)”交叉验证来衡量参数估计的精确度(记为 I), 加之对比被正确识别出为零向量的数量(记为 II)以及错误识别零向量的数量(记为 III), 通过重复1000次实验, 对样本量为256, 400组的数据进行模拟. 交叉验证均方误差定义如下:

$$MSE_{CV} = \frac{1}{nT} \sum_{i=1}^{nT} \left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_i^{(-i)} \right\|,$$

其中, $\hat{\mathbf{Y}}_i^{(-i)}$ 表示 $(nT) \times 1$ 维 Y 剔除第 i 行数据后利用剩下的 $(nT - 1)$ 行数据预测得到的第 i 组模拟值.

模型 (3-6) 的模拟结果如表3-1至表3-2所示, 相比于的 ε 为自由度为5的 t 分布时的结果, 当 ε 服从标准正态分布时的变量选择效果更好些. 样本量的增大时, 各假定情况的结果均变得更好. 当空间自回归系数 ρ 较小时, 对因变量的影响越来越小, 系数估计的准确度也随着增加.

表3-1 $\varepsilon \sim N(0, 1)$ 时不同假定的结果对比

ρ		I	II	III	ρ	I	II	III	
$n = 225$									
-0.85	LASSO	0.0068	4.875	0	0.2	LASSO	0.0066	4.970	0
	ALASSO	0.0044	4.986	0		ALASSO	0.0053	4.989	0
	SCAD	0.0042	4.984	0		SCAD	0.0050	4.997	0
0.5	LASSO	0.0059	4.961	0	0.8	LASSO	0.0076	4.963	0
	ALASSO	0.0035	4.983	0		ALASSO	0.0053	4.967	0
	SCAD	0.0032	4.986	0		SCAD	0.0054	4.962	0
$n = 400$									
-0.85	LASSO	0.0063	4.792	0	0.2	LASSO	0.0061	4.983	0
	ALASSO	0.0043	4.984	0		ALASSO	0.0042	4.993	0
	SCAD	0.0041	4.982	0		SCAD	0.0040	4.996	0
0.5	LASSO	0.0051	4.982	0	0.8	LASSO	0.0082	4.989	0
	ALASSO	0.0022	4.992	0		ALASSO	0.0040	4.984	0
	SCAD	0.0020	4.995	0		SCAD	0.0041	4.986	0

表3-2 $\varepsilon \sim t(5)$ 时不同假定的结果对比

ρ		I	II	III	ρ		I	II	III
$n = 225$									
-0.85	LASSO	0.0082	4.967	0	0.2	LASSO	0.0065	4.974	0
	ALASSO	0.0043	4.961	0		ALASSO	0.0055	4.985	0
	SCAD	0.0045	4.964	0		SCAD	0.0054	4.992	0
0.5	LASSO	0.0066	4.947	0	0.8	LASSO	0.0110	4.957	0
	ALASSO	0.0036	4.981	0		ALASSO	0.0069	4.965	0
	SCAD	0.0034	4.988	0		SCAD	0.0069	4.966	0
$n = 400$									
-0.85	LASSO	0.0082	4.966	0	0.2	LASSO	0.0092	4.994	0
	ALASSO	0.0041	4.983	0		ALASSO	0.0050	4.997	0
	SCAD	0.0044	4.984	0		SCAD	0.0051	4.994	0
0.5	LASSO	0.0057	4.982	0	0.8	LASSO	0.0080	4.976	0
	ALASSO	0.0031	4.991	0		ALASSO	0.0042	4.984	0
	SCAD	0.0026	4.991	0		SCAD	0.0043	4.982	0

§3.4 引理及定理的证明

引理3.1 若假设条件(A1)-(A5)满足, $\hat{\theta} = (\hat{\beta}', \hat{\rho}, \hat{\sigma}^2)'$ 为 $\theta_0 = (\beta_0', \rho_0, \sigma_0^2)'$ 的相合估计. 证明: 见文献[29]的定理1.

引理3.2 若假设条件(A1)-(A7)成立, 有下式成立:

$$-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{P} \Sigma, \quad \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Omega),$$

其中, $\Sigma = -\lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial^2 \log L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$, $\Omega = \lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta'}\right)$.

证明: 由文献[29]和[30]可推得,

$$-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{P} \lim_{n \rightarrow \infty} E\left(-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right), \quad (3-7)$$

$$\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} E\left(\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta'}\right)\right), \quad (3-8)$$

结合假设(A7)和式 (3-7) 得,

$$-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{P} \Sigma,$$

在假设(A7)条件下, 由式 (3-8) 得:

$$\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Omega).$$

定理3.1的证明: 令 $\beta = \beta_0 + \mathbf{u}/\sqrt{nT}$, 记

$$\begin{aligned} Q(\mathbf{u}) &= [A_{nT}(\hat{\rho})Y - X(\beta_0 + \mathbf{u}/\sqrt{nT})]' [A_{nT}(\hat{\rho})Y - X(\beta_0 + \mathbf{u}/\sqrt{nT})] \\ &\quad + \lambda_{nT} \sum_{j=1}^k \omega_j \left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right|, \end{aligned}$$

$$Q(\mathbf{0}) = [A_{nT}(\hat{\rho})Y - X(\beta_0)]' [A_{nT}(\hat{\rho})Y - X(\beta_0)] + \lambda_{nT} \sum_{j=1}^k \omega_j \left| \beta_{0j} \right|,$$

记 $\tilde{\mathbf{u}}^{(nT)} = \arg \min_{\mathbf{u}} Q(\mathbf{u})$, 则 $\tilde{\beta} = \beta_0 + \tilde{\mathbf{u}}^{(nT)}/\sqrt{nT}$, 即 $\tilde{\mathbf{u}}^{(nT)} = \sqrt{nT}(\tilde{\beta} - \beta_0)$.

$$\begin{aligned} & (A_{nT}(\hat{\rho}) - A_{nT}(\rho_0))' \\ &= [(I_{nT} - \hat{\rho}(I_T \otimes W_n) - (I_{nT} - \rho_0(I_T \otimes W_n))]' \\ &= -(\hat{\rho} - \rho_0)(I_T \otimes W_n)'. \\ & \quad \frac{-Y'(A_{nT}(\hat{\rho}) - A_{nT}(\rho_0))'X}{\sqrt{nT}} \\ &= \frac{Y'[(\hat{\rho} - \rho_0)(I_T \otimes W_n)]'X}{\sqrt{nT}} \\ &= \sqrt{nT}(\hat{\rho} - \rho_0) \frac{\beta_0' X' (A_{nT}^{-1}(\hat{\rho}))' (I_T \otimes W_n)' X}{nT} \\ & \quad + (\hat{\rho} - \rho_0) \frac{\varepsilon' (A_{nT}^{-1}(\hat{\rho}))' (I_T \otimes W_n)'}{\sqrt{nT}}. \end{aligned}$$

$$\begin{aligned} Q(\mathbf{u}) - Q(\mathbf{0}) &= \frac{\mathbf{u}' X' X \mathbf{u}}{nT} - 2 \frac{\left(A_{nT}(\hat{\rho})Y - X\beta_0 \right)' X}{\sqrt{nT}} \mathbf{u} + \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \\ &= \frac{\mathbf{u}' X' X \mathbf{u}}{nT} + 2 \left\{ \left[\frac{-Y' \left(A_{nT}(\hat{\rho}) - A_{nT}(\rho_0) \right)' X}{\sqrt{nT}} \right]' - \left[\frac{\left(Y' A_{nT}(\rho_0)' - \beta_0' X' \right) X}{\sqrt{nT}} \right]' \right\}' \mathbf{u} \\ & \quad + \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{\mathbf{u}'X'X\mathbf{u}}{nT} + 2 \left\{ \left[\frac{-Y' \left(A_{nT}(\hat{\rho}) - A_{nT}(\rho_0) \right)' X}{\sqrt{nT}} \right]' - \left[\frac{\varepsilon' X}{\sqrt{nT}} \right]' \right\}' \mathbf{u} \\
 &+ \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \\
 &= \frac{\mathbf{u}'X'X\mathbf{u}}{nT} + 2 \left\{ \left[\frac{\sqrt{nT}(\hat{\rho} - \rho_0) \beta_0' X' (A_{nT}^{-1}(\hat{\rho}))' (I_T \otimes M_n)' X}{nT} \right. \right. \\
 &+ \left. \left. (\hat{\rho} - \rho_0) \frac{\varepsilon (A_{nT}^{-1}(\hat{\rho}))' (I_T \otimes M_n)' X}{\sqrt{nT}} \right] - \frac{X' \varepsilon}{\sqrt{nT}} \right\}' \mathbf{u} \\
 &+ \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right).
 \end{aligned} \tag{3-9}$$

由 Slutsky 定理得 $Q(\mathbf{u}) - Q(\mathbf{0}) \xrightarrow{d} V(\mathbf{u})$, 其中

$$V(\mathbf{u}) = \begin{cases} \mathbf{u}'_{\mathcal{A}} \Psi_{11} \mathbf{u}_{\mathcal{A}} - 2\xi'_{\mathcal{A}} \mathbf{u}_{\mathcal{A}}, & u_j = 0, \forall j \notin \mathcal{A} \\ \infty, & \text{其它.} \end{cases}$$

为此, 首先给出 $\sqrt{nT}(\hat{\theta} - \theta_0)$ 和 $\sqrt{nT}(\hat{\rho} - \rho)$ 的表达式:

对 $\partial \ln L_{nT}(\hat{\theta}) / \partial \theta$ 在 $\theta = \theta_0$ 处进行泰勒展开:

$$\frac{\partial \ln L_{nT}(\hat{\theta})}{\partial \theta} = \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0) = 0, \tag{3-10}$$

其中, 由 $\|\hat{\theta} - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$, $\hat{\theta} \xrightarrow{P} \theta_0$, 有 $\theta^* \xrightarrow{P} \theta_0$.

式 (3-10) 等价于

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left[\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} \right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}, \tag{3-11}$$

令 $P = - \left[\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} \right]^{-1} = \begin{pmatrix} * & * & * \\ \eta' & s_1 & s_2 \\ * & * & * \end{pmatrix}$, 其中 (η', s_1, s_2) 表示矩阵 \mathbf{V} 的第二行元素, η'

是 k 维的列向量, s_1 和 s_2 为标量. 由引理 3.2 可知, $P \xrightarrow{P} \Sigma^{-1}$, d_{ij} 是矩阵 Σ^{-1} 的第 i 行第 j 列元素, $i, j = 1, 2, 3$. 其中,

$$\eta' \xrightarrow{P} d_{21}, s_1 \xrightarrow{P} d_{22}, s_2 \xrightarrow{P} d_{23}, \tag{3-12}$$

由式 (3-3) 对 θ 求一阶导数得到:

$$\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_0^2} X' \varepsilon \\ -\text{tr}((I_T \otimes W_n) A_{nT}^{-1}(\rho_0)) + \frac{1}{\sigma_0^2} \varepsilon' (I_T \otimes W_n) Y \\ \frac{1}{2\sigma_0^4} (\varepsilon' \varepsilon - nT\sigma_0^2) \end{pmatrix}, \quad (3-13)$$

由式 (3-11)、式 (3-12) 得到:

$$\sqrt{nT}(\hat{\theta} - \theta_0) = \begin{pmatrix} \sqrt{nT}(\hat{\beta} - \beta_0) \\ \sqrt{nT}(\hat{\rho} - \rho_0) \\ \sqrt{nT}(\hat{\sigma}^2 - \sigma_0^2) \end{pmatrix} = \begin{pmatrix} * & * & * \\ \eta' & s_1 & s_2 \\ * & * & * \end{pmatrix} \cdot \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}, \quad (3-14)$$

所以有,

$$\begin{aligned} \sqrt{nT}(\hat{\rho} - \rho_0) &= \eta' \cdot \frac{1}{\sigma_0^2} X' \varepsilon + s_1 \cdot [-\text{tr}((I_T \otimes W_n) A_{nT}^{-1}(\rho_0)) + \frac{1}{\sigma_0^2} \varepsilon' (I_T \otimes W_n) Y] \\ &+ s_2 \cdot \frac{1}{2\sigma_0^4} (\varepsilon' \varepsilon - nT\sigma_0^2), \end{aligned} \quad (3-15)$$

由式 (3-9)、式 (3-14)、式 (3-15) 和引理3.2得到:

$$\begin{aligned} &\left[-\frac{(A_{nT}(\hat{\rho})Y_n - X\beta_0)' X}{\sqrt{nT}} \right]' \\ &= K \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + (\hat{\rho} - \rho_0) \frac{\varepsilon' (A_{nT}^{-1}(\hat{\rho}))' (I_T \otimes W_n) X}{\sqrt{nT}}, \end{aligned} \quad (3-16)$$

其中,

$$K = \left[\frac{X' (I_T \otimes W_n) A_{nT}^{-1}(\hat{\rho}) X \beta_0}{nT} \eta' - \sigma_0^2 I_k, \frac{X' (I_T \otimes W_n) A_{nT}^{-1}(\hat{\rho}) X \beta_0}{nT} s_1, \frac{X' (I_T \otimes W_n) A_{nT}^{-1}(\hat{\rho}) X \beta_0}{nT} s_2 \right],$$

又由 $\frac{X' (I_T \otimes W_n) A_{nT}^{-1}(\hat{\rho}) X \beta_0}{nT} \xrightarrow{P} \nu$, 以及(3-12)式, 有 $K \xrightarrow{P} D$,

其中 $D = (\nu d_{21} - \sigma_0^2 I_k, \nu d_{22}, \nu d_{23})$,

再结合引理2得:

$$K \frac{1}{\sqrt{nT}} \frac{\partial \ln L(nT)(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, D\Omega^*D'), \quad (3-17)$$

由 $\hat{\rho} - \rho_0 \rightarrow 0$, $\frac{\varepsilon' A_{nT}^{-1}(\hat{\rho})' (I_T \otimes W_n) X}{\sqrt{nT}}$ 依分布收敛到某个正态分布, 所以

$$(\hat{\rho} - \rho_0) \frac{\varepsilon' A_{nT}^{-1}(\hat{\rho})' (I_T \otimes W_n) X}{\sqrt{nT}} \xrightarrow{P} 0, \quad (3-18)$$

再由式 (3-16)、式 (3-17)、式 (3-18) 有：

$$\left[-\frac{\left(A_{nT}(\hat{\rho})Y - X\beta_0 \right)' X}{\sqrt{nT}} \right]' \xrightarrow{d} \xi \sim N(0, D\Omega D'). \quad (3-19)$$

易知, $\frac{1}{nT}XX \xrightarrow{P} \Psi$.

最后考虑 (3-9) 式中的 $\frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} (|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j})$: 如果 $\beta_{0j} \neq 0$ (即 $j \in \mathcal{A}$), 由假设 $\lambda_{nT}(nT)^{-1/2} \rightarrow 0$, $\omega_j = |\hat{\beta}_j|^{-r} \xrightarrow{P} |\beta_j|^{-r}$,

得到：

$$\left| \lambda_{nT} \omega_j \left(|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j} \right) \right| \leq \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} \omega_j \rightarrow 0,$$

如果 $\beta_{0j} = 0$ (即 $j \in \mathcal{A}^c$), 由假设 $\lambda_{nT}(nT)^{(r-1)/2} \rightarrow \infty$, $\omega_j = |\hat{\beta}_j|^{-r}$, $\sqrt{nT} |\hat{\beta}_j|^{-r} = O_p(1)$ 得到：

$$\left| \lambda_{nT} \omega_j \left(|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j} \right) \right| = \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} \omega_j = \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} (nT)^{r/2} (\sqrt{nT} |\hat{\beta}_j|)^{-r} \rightarrow \begin{cases} 0, u_j = 0 \\ \infty, u_j \neq 0. \end{cases}$$

由 Slutsky 定理得到 $Q(\mathbf{u}) - Q(\mathbf{0}) \xrightarrow{d} V(\mathbf{u})$,

$$V(\mathbf{u}) = \begin{cases} \mathbf{u}'_{\mathcal{A}} \Psi_{11} \mathbf{u}_{\mathcal{A}} - 2\xi'_{\mathcal{A}} \mathbf{u}_{\mathcal{A}}, & u_j = 0, \forall j \notin \mathcal{A} \\ \infty, & \text{其它.} \end{cases}$$

又因为 $V(\mathbf{u})$ 是凸函数, 所以存在唯一的最小值 $(\xi'_{\mathcal{A}} \Psi_{11}^{-1}, 0_{1 \times (k-k_0)})'$.

利用 Geyer^[31]、Knight 和 Fu^[32] 的结论得到：

$$\hat{\mathbf{u}}_{\mathcal{A}}^{(nT)} \xrightarrow{d} \Psi_{11}^{-1} \xi_{\mathcal{A}}, \quad \hat{\mathbf{u}}_{\mathcal{A}^c}^{(nT)} \xrightarrow{d} 0. \quad (3-20)$$

得到 $\sqrt{nT}(\hat{\beta}_{\mathcal{A}} - \beta_{\mathcal{A}}) \xrightarrow{d} \Psi_{11}^{-1} \xi_{\mathcal{A}}$, 证毕.

接下来证明相合性: 对 $\forall j \in \mathcal{A}$, (即 $\beta_j \neq 0$), 由渐近正态性有: $\tilde{\beta}_j \xrightarrow{P} \beta_{0j}$. 所以

$$P(\tilde{\beta}_j \neq 0) = P(j \in \tilde{\mathcal{A}}) \rightarrow 1, n \rightarrow \infty,$$

即

$$P(\mathcal{A} \subseteq \tilde{\mathcal{A}}) \rightarrow 1, n \rightarrow \infty.$$

则只需证明: $P(\mathcal{A} \supseteq \tilde{\mathcal{A}}) \rightarrow 1, n \rightarrow \infty$. 考虑事件 $j^* \in \tilde{\mathcal{A}}$, 由 Karush-Kuhn-Tucker(KKT) 最

优化条件^[33]有,

$$-2\bar{X}'_{j^*}(\bar{Y} - \bar{X}\tilde{\beta}) + \lambda_{nT}\omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*}) = 0,$$

上式等价于

$$2\bar{X}'_{j^*}(\bar{Y} - \bar{X}\tilde{\beta}) = \lambda_{nT}\omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*}),$$

又因为

$$2\frac{\bar{X}'_{j^*}(\bar{Y} - \bar{X}\tilde{\beta})}{\sqrt{nT}} = 2\frac{\bar{X}'_{j^*}\bar{X}\sqrt{nT}(\tilde{\beta} - \beta)}{nT} + 2\frac{\bar{X}'_{j^*}\varepsilon}{\sqrt{nT}}, \quad (3-21)$$

其中,

$$\frac{\bar{X}'_{j^*}\bar{X}}{nT} = \frac{X'_{j^*}X}{nT} \xrightarrow{p} (\Psi_{j^*1}, \dots, \Psi_{j^*k}),$$

Ψ_{j^*k} 是 Ψ 的 $j^* \times k$ 阶元素. 通过式(3-19)和定理3.1的(i), 可以得到: $2\frac{\bar{X}'_{j^*}\bar{X}\sqrt{nT}(\tilde{\beta} - \beta)}{nT}$ 依分布收敛到某个正态分布, 以及 $2\frac{\bar{X}'_{j^*}\varepsilon}{\sqrt{nT}} \xrightarrow{d} N(0, 4\|\bar{x}_{j^*}\|^2\sigma^2)$. 因此,

$$P(j^* \in \tilde{\mathcal{A}}) \leq P\left(\left|2\frac{\bar{X}'_{j^*}(\bar{Y} - \bar{X}\tilde{\beta})}{\sqrt{nT}}\right| = \frac{\lambda_{nT}\omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*})}{\sqrt{nT}}\right) \rightarrow 0, \quad n \rightarrow \infty.$$

即 $P(\mathcal{A} \supseteq \tilde{\mathcal{A}}) \rightarrow 1, n \rightarrow \infty$, 综上所述定理3.1的(ii)得证.

第四章 空间面板误差模型的 Adaptive LASSO 变量选择

在本章中, 我们研究空间面板误差模型的 Adaptive LASSO 变量选择问题, 假设有 n 个空间单位和 T 个时间段, 我们考虑以下模型:

$$y_t = X_t\beta + \mu_t, \quad \mu_t = \lambda M_n \mu_t + \varepsilon_t, \quad (4-1)$$

其中, y_t 是 $n \times 1$ 维响应变量的观测值, X_t 是 $n \times k$ 维固定设计自变量的观测值, β 是 $k \times 1$ 维回归系数向量, M_n 是预先给定的 $n \times n$ 维空间权重矩阵, λ 为空间自相关系数, μ_t 是自相关误差项, $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tn})'$ 为 $n \times 1$ 维误差向量, ε_{ti} 为独立同分布的随机误差项, 且满足 $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2 I_n$, 其中 $0 < \sigma^2 < \infty$, I_n 为单位矩阵, $t = 1, 2, \dots, T$.

§4.1 引言

我们继续讨论模型 (4-1). 当 $t = 1, 2, \dots, T$ 时, 模型 (4-1) 可以写成如下矩阵形式:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} X_1\beta \\ X_2\beta \\ \vdots \\ X_T\beta \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix},$$

$$\begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & B_T \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix},$$

其中, $B_t = (I_n - \lambda W_n)$, 令 $B_{nT}(\lambda) = [I_{nT} - \lambda(I_T \otimes M_n)]$ 且设 $B_{nT}(\lambda)$ 是非奇异的, 模型(4-1)化简成:

$$Y = X\beta + B_{nT}^{-1}(\lambda)\varepsilon, \quad (4-2)$$

其中, $Y = (y'_1, \dots, y'_T)'$ 是 $(nT) \times 1$ 维因变量向量, $X = (X'_1, \dots, X'_T)'$ 是 $(nT) \times k$ 维自变量矩阵, ε 为 $(nT) \times 1$ 维随机扰动项, 需要估计的参数为 $\theta = (\beta', \lambda, \sigma^2)' \in R^{k+2}$, 令 $\theta_0 = (\beta'_0, \rho_0, \sigma_0^2)'$ 为 θ 的真值. 模型 (4-2) 的对数似然函数为:

$$\ln L_{nT}(\theta) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2) + \ln |B_{nT}(\lambda)| - \frac{1}{2\sigma^2} \varepsilon' \varepsilon, \quad (4-3)$$

其中, $\varepsilon = B_{nT}(\lambda)(Y - X\beta)$. 根据文献[34] 提出的拟极大似然估计法, 对于给定的 ρ , 关于 β 和 σ^2 的拟极大似然估计为:

$$\hat{\beta} = (X' B_{nT}(\lambda)' B_{nT}(\lambda) X)^{-1} X' B_{nT}(\lambda)' B_{nT}(\lambda) Y,$$

$$\hat{\sigma}^2 = \frac{1}{nT} \|S \cdot Y\|^2,$$

其中, $S = \left\{ I_{nT} - B_{nT}(\lambda) X \left(X' B_{nT}(\lambda)' B_{nT}(\lambda) X \right)^{-1} X' B_{nT}(\lambda)' B_{nT}(\lambda) \right\}^{-1} X' B_{nT}(\lambda)' B_{nT}(\lambda)$. 将 $\hat{\beta}$, $\hat{\sigma}^2$ 代入式(4-3)可得到关于 λ 的对数似然函数:

$$\tilde{f} = -\frac{nT}{2} (\ln(2\pi) + 1) - \frac{nT}{2} \ln(\hat{\sigma}^2) + \ln |B_{nT}(\lambda)|, \quad (4-4)$$

对式(4-4)进行最大化处理后可得到新的 $\hat{\lambda}$, 再将 $\hat{\lambda}$ 代入模型(4-2)的拟极大似然函数式(4-3)中. 且令 $\bar{Y} = B_{nT}(\hat{\lambda})Y$, $\bar{X} = B_{nT}(\hat{\lambda})X$, 模型 (4-2)可记为: $\bar{Y} = \bar{X}\beta + \varepsilon$.

为了得到系数的稀疏估计, 考虑 Zou^[12] 提出的 Adaptive LASSO 变量选择方法, 选择一个 $r > 0$, 定义惩罚权重向量为 $\omega_j = 1/|\hat{\beta}_j|^r$, 此时的惩罚目标函数为:

$$\tilde{\beta} = \arg \min_{\beta} \left\{ (\bar{Y} - \bar{X}\beta)' (\bar{Y} - \bar{X}\beta) + \sum_{j=1}^k \lambda_{nT} \omega_j |\beta_j| \right\}. \quad (4-5)$$

其中, $r > 0$ 为调节参数, $\hat{\beta}$ 为 β 的拟极大似然估计, $\lambda_{nT} > 0$ 为可调整的惩罚参数, 当 λ_{nT} 越大时被选入的 β 个数越少, 可通过交叉验证法或广义交叉验证法来选择 λ_{nT} 的值.

给定备选子集 $\mathcal{B} = \{j : \beta_j \neq 0\}$, 不失一般性, 假定 $\mathcal{B} = \{1, \dots, k_0\}$, $\mathcal{B}^c = \{k_0 + 1, \dots, k\}$. 定义 $\tilde{\mathcal{B}} = \{j : \tilde{\beta}_j \neq 0\}$, $\beta = (\beta'_{\mathcal{B}}, \beta'_{\mathcal{B}^c})'$.

§4.2 主要结果

为了方便起见, 引入下述符号:

$$\begin{aligned} \Phi &= \lim_{n \rightarrow \infty} \frac{1}{nT} X' B_{nT}(\lambda)' B_{nT}(\lambda) X, & H &= (I_T \otimes M_n), \\ \pi_1 &= \lim_{n \rightarrow \infty} \frac{1}{nT} (B_{nT}(\lambda) H X \beta_0)' (B_{nT}(\lambda) H X \beta_0), & \tilde{H} &= H B_{nT}^{-1}(\lambda), \\ \pi_2 &= \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(\tilde{H})^2, & \pi_3 &= \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(\tilde{H}), \\ \tilde{\nu} &= \lim_{n \rightarrow \infty} \frac{1}{nT} X' B_{nT}(\lambda)' B_{nT}(\lambda) H X \beta_0, & \Phi &= \begin{pmatrix} \Phi_{11} & \Phi_{22} \\ \Phi_{33} & \Phi_{44} \end{pmatrix}, \end{aligned}$$

其中, Φ_{11} 是 $k_0 \times k_0$ 维矩阵, Φ_{22} 是 $k_0 \times (k - k_0)$ 维矩阵, Φ_{33} 是 $(k - k_0) \times k_0$ 维矩阵, Φ_{44} 是 $(k - k_0) \times (k - k_0)$ 维矩阵. 为了探讨新估计的渐近正态性, 需作如下假设:

(B1): n 趋于无穷大, T 是有限常数.

(B2): $\{X_t, t = 1, 2, \dots, T\}$ 是固定设计自变量的观测值且一致有界.

(B3): $\{\varepsilon_{ti}, i = 1, 2, \dots, n, t = 1, 2, \dots, T\}$ 是均值为零且方差为 σ^2 的 *i.i.d* 随机变量, 对于某些 $\delta > 0$, 有 $E|\varepsilon_{ti}|^{4+\delta}$ 存在且有界.

(B4): 矩阵 $B_{nT}(\lambda)$ 为非奇异矩阵, M_n 具有零对角线且进行“标准化”处理, $B_{nT}(\lambda)^{-1}$ 元素在 L_1 范数上行和与列和均存在且有界.

(B5): 在一个紧凸集 Δ 上, $B_{nT}^{-1}(\lambda)$ 的行和与列和对于 λ 一致有界, 且 $\Delta \in (-1, 1)$.

(B6): 假设下列极限存在 $\Phi = \lim_{n \rightarrow \infty} \frac{1}{nT} X' B_{nT}(\lambda)' B_{nT}(\lambda) X$, $\pi_2 = \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(\tilde{H})^2$, $\pi_1 = \lim_{n \rightarrow \infty} \frac{1}{nT} (B_{nT}(\lambda) H X \beta_0)' (B_{nT}(\lambda) H X \beta_0)$, $\tilde{\nu} = \lim_{n \rightarrow \infty} \frac{1}{nT} X' B_{nT}(\lambda)' B_{nT}(\lambda) H X \beta_0$, $\pi_3 = \lim_{n \rightarrow \infty} \frac{1}{nT} \text{tr}(\tilde{H})$.

(B7): 假设极限 $\Sigma^* = \lim_{n \rightarrow \infty} -\frac{1}{nT} E\left(\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$, $\Omega^* = \lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta'}\right)$ 存在, 且 Σ^* 和 Ω^* 是非奇异的.

注1: 假设(B1)-(B5)是空间数据模型的常见假设, 与Li^[3], Lee^[7]的假设相似.

注2: $E\left(-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$

$$= \begin{pmatrix} \frac{1}{nT\sigma_0^2} X' B_{nT}(\lambda)' B_{nT}(\lambda) X & \frac{1}{nT\sigma_0^2} X' B_{nT}(\lambda)' B_{nT}(\lambda) H X \beta_0 & 0 \\ * & \frac{1}{nT\sigma_0^2} (B_{nT}(\lambda) H X \beta_0)' (B_{nT}(\lambda) H X \beta_0) + \frac{1}{nT} \text{tr}(\tilde{H})^2 & \frac{1}{nT\sigma_0^2} \text{tr}(\tilde{H}) \\ * & * & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

可令

$$\Sigma^* = \begin{pmatrix} \frac{1}{\sigma_0^2} \Phi & \frac{1}{\sigma_0^2} \tilde{\nu} & \mathbf{0} \\ \frac{1}{\sigma_0^2} \tilde{\nu}' & \frac{1}{\sigma_0^2} \pi_1 + \pi_2 & \frac{1}{\sigma_0^2} \pi_3 \\ \mathbf{0} & \frac{1}{\sigma_0^2} \pi_3 & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

定理4.1 假设条件(B1)-(B7)成立, 定义 $\tilde{\mathcal{B}} = \{j : \tilde{\beta}_j \neq 0\}$, 如果当 $\lambda_{nT} \cdot (nT)^{-1/2} \rightarrow 0$ 且 $\lambda_{nT} \cdot (nT)^{(r-1)/2} \rightarrow \infty$, r 为调节参数, Φ 是非奇异的, 则模型的 Adaptive LASSO 估计满足:

(i) 渐近正态性: $\sqrt{nT}(\tilde{\beta}_{\mathcal{B}} - \beta_{\mathcal{B}}) \xrightarrow{d} \Phi_{11}^{-1} \xi_{\mathcal{B}}$, $\xi \sim N(0, D^* \Omega^* D^{*'})$,

其中, $\xi_{\mathcal{B}}$ 是 ξ 前 k_0 维子向量, Φ_{11} 是 $k_0 \times k_0$ 阶矩阵, $D_{k \times (k+2)}^* = (\tilde{\nu} d_{21}^* - \sigma_0^2 I_K, \tilde{\nu} d_{22}^*, \tilde{\nu} d_{23}^*)$, d_{ij}^* 是矩阵 Σ^{*-1} 的第 i 行第 j 列元素, $i, j = 1, 2, 3$.

(ii) 相合性: $\lim_{n \rightarrow \infty} P(\tilde{\mathcal{B}} = \mathcal{B}) = 1$.

§4.3 数值模拟

本节考虑在不同的空间自相关系数情况下, 对蒙特卡罗模拟的结果进行比对分析. 考虑如下模型:

$$y_t = X_t \beta + \mu_t, \quad \mu_t = \lambda M_n \mu_t + \varepsilon_t, \quad t = 1, \dots, 8, \quad (4-6)$$

其中, 令 $X_1 = (x_{11}, x_{12}, \dots, x_{1n})'$, $x_{1i} = \frac{i}{n+1}$, $X_2 = (x_{21}, x_{22}, \dots, x_{2n})'$, $x_{2i} = \frac{i}{n+2}, \dots, X_8 =$

$(x_{81}, x_{82}, \dots, x_{8n})'$, $x_{8i} = \frac{i}{n+8}, 1 \leq i \leq n$, 将 M_n 设置为常用的“Queen”型空间权重矩阵, 假定 $\beta = (0.6, 0, 1, 0, 0, 0.8, 0, 0)$, 惩罚权重系数 $w_j = 1/|\hat{\beta}_j|, j = 1, \dots, 8$. 选取不同的情形进行考察: (1)两种不同样本量 n 为 225 和 400; (2)两种不同 ε 误差项服从的分布分别为 $N(0, 1), t(5)$; (3)三种不同的变量选择方法: LASSO, ALASSO, SCAD; (4)四种不同的自相关系数值 ρ 为 $-0.85, 0.2, 0.5, 0.8$.

本节利用“留一法(leave-one-out)”交叉验证来衡量参数估计的精确度(记为 I), 加之对比被正确识别出为零向量的数量(记为 II)以及错误识别零向量的数量(记为 III), 通过重复 1000 次实验, 对样本量为 256, 400 组的数据进行模拟. 交叉验证均方误差定义如下: $MSE_{CV} = \frac{1}{nT} \sum_{i=1}^{nT} \|\mathbf{Y}_i - \hat{\mathbf{Y}}_i^{(-i)}\|^2$, 其中 $\hat{\mathbf{Y}}_i^{(-i)}$ 表示 $(nT) \times 1$ 维 Y 剔除第 i 行数据后利用剩下的 $(nT - 1)$ 行数据预测得到的第 i 组模拟值. 模型(4-6)的模拟结果如表 4-1 至表 4-2 所示, 相比于的 ε 为自由度为 5 的 t 分布时的结果, 当 ε 服从标准正态分布时的变量选择效果更好些. 样本量的增大时, 各假定情况的效果变得更佳. 当空间自相关系数 λ 较小时, 系数估计的准确度也随着增加.

表 4-1 $\varepsilon \sim N(0, 1)$ 时不同假定的结果对比

λ		I	II	III	λ		I	II	III
$n = 225$									
-0.85	LASSO	0.0073	4.969	0	0.2	LASSO	0.0054	4.994	0
	ALASSO	0.0054	4.971	0		ALASSO	0.0015	4.995	0
	SCAD	0.0050	4.970	0		SCAD	0.0014	4.998	0
0.5	LASSO	0.0062	4.985	0	0.8	LASSO	0.0043	4.930	0
	ALASSO	0.0031	4.991	0		ALASSO	0.0052	4.985	0
	SCAD	0.0030	4.996	0		SCAD	0.0056	4.986	0
$n = 400$									
-0.85	LASSO	0.0067	4.967	0	0.2	LASSO	0.0047	4.996	0
	ALASSO	0.0043	4.974	0		ALASSO	0.0013	4.998	0
	SCAD	0.0040	4.972	0		SCAD	0.0012	4.996	0
0.5	LASSO	0.0062	4.989	0	0.8	LASSO	0.0073	4.972	0
	ALASSO	0.0016	4.996	0		ALASSO	0.0035	4.984	0
	SCAD	0.0021	4.998	0		SCAD	0.0032	4.988	0

表4-2 $\varepsilon \sim t(5)$ 时不同假定的结果对比

λ		I	II	III	λ		I	II	III
$n = 225$									
-0.85	LASSO	0.0082	4.964	0	0.2	LASSO	0.0087	4.994	0
	Alasso	0.0063	4.974	0		Alasso	0.0052	4.997	0
	SCAD	0.0062	4.972	0		SCAD	0.0052	4.996	0
0.5	LASSO	0.0107	4.981	0	0.8	LASSO	0.0110	4.972	0
	ALASSO	0.0066	4.995	0		ALASSO	0.0069	4.984	0
	SCAD	0.0063	4.998	0		SCAD	0.0068	4.982	0
$n = 400$									
-0.85	LASSO	0.0082	4.975	0	0.2	LASSO	0.0080	4.984	0
	ALASSO	0.0043	4.973	0		Alasso	0.0054	4.994	0
	SCAD	0.0044	4.979	0		SCAD	0.0053	4.998	0
0.5	LASSO	0.0063	4.989	0	0.8	LASSO	0.0074	4.975	0
	ALASSO	0.0032	4.995	0		ALASSO	0.0042	4.987	0
	SCAD	0.0037	4.997	0		SCAD	0.0040	4.985	0

§4.4 引理及定理的证明

引理4.1 若假设条件(B1)-(B5)满足, $\hat{\theta} = (\hat{\beta}', \hat{\lambda}, \hat{\sigma}^2)'$ 为 $\theta_0 = (\beta_0', \lambda_0, \sigma_0^2)'$ 的相合估计. 证明: 见文献[34].

引理4.2 若假设条件(B1)-(B7)成立, 有下式成立:

$$-\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{P} \Sigma^*, \quad \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Omega^*),$$

其中, $\Sigma^* = -\lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial^2 \log L_{nT}(\theta_0)}{\partial \theta \partial \theta'}\right)$, $\Omega^* = \lim_{n \rightarrow \infty} \frac{1}{nT} E\left(\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta'}\right)$.

证明: 见文献[30]的定理1.

定理4.1的证明: 令 $\beta = \beta_0 + \mathbf{u}/\sqrt{nT}$, 记

$$S(\mathbf{u}) = [B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X(\beta_0 + \mathbf{u}/\sqrt{nT})]' [B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X(\beta_0 + \mathbf{u}/\sqrt{nT})] \\ + \lambda_{nT} \sum_{j=1}^k \omega_j |\beta_{0j} + \mathbf{u}_j/\sqrt{nT}|,$$

$$S(\mathbf{0}) = [B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X(\beta_0)]'[B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X(\beta_0)] + \lambda_{nT} \sum_{j=1}^k \omega_j |\beta_{0j}|,$$

记 $\tilde{\mathbf{u}}^{(nT)} = \arg \min_{\mathbf{u}} S(\mathbf{u})$, 则 $\tilde{\beta} = \beta_0 + \tilde{\mathbf{u}}^{(nT)}/\sqrt{nT}$, 即 $\tilde{\mathbf{u}}^{(nT)} = \sqrt{nT}(\tilde{\beta} - \beta_0)$.

$$\begin{aligned} & (B_{nT}(\hat{\lambda}) - B_{nT}(\lambda_0))' \\ &= [(I_{nT} - \hat{\lambda}(I_T \otimes M_n) - I_{nT} + \lambda_0(I_T \otimes M_n))]', \\ &= -(\hat{\lambda} - \lambda_0)(I_T \otimes M_n)'. \end{aligned}$$

$$\begin{aligned} & \frac{-Y'(B_{nT}(\hat{\lambda}) - B_{nT}(\lambda_0))'B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \\ &= \frac{Y'[(\hat{\lambda} - \lambda_0)(I_T \otimes M_n)]'B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \\ &= \sqrt{nT}(\hat{\lambda} - \lambda_0) \frac{Y'(I_T \otimes M_n)'B_{nT}(\hat{\lambda})X}{nT} \\ &= \sqrt{nT}(\hat{\lambda} - \lambda_0) \frac{\beta_0'X'(I_T \otimes M_n)'B_{nT}(\hat{\lambda})X}{nT} + (\hat{\lambda} - \lambda_0) \frac{\varepsilon'B_{nT}^{-1}(\hat{\lambda})'(I_T \otimes M_n)'B_{nT}(\hat{\lambda})}{\sqrt{nT}} \\ &= \sqrt{nT}(\hat{\lambda} - \lambda_0) \frac{(HX\beta_0)'B_{nT}(\hat{\lambda})X}{nT} + (\hat{\lambda} - \rho_0) \frac{\varepsilon'B_{nT}^{-1}(\hat{\lambda})'H'B_{nT}(\hat{\lambda})}{\sqrt{nT}}. \end{aligned}$$

$$\begin{aligned} S(\mathbf{u}) - S(\mathbf{0}) &= \frac{\mathbf{u}'X'B_{nT}(\hat{\lambda})'B_{nT}(\hat{\lambda})X\mathbf{u}}{nT} - 2 \frac{\left(B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X\beta_0\right)'B_{nT}(\lambda_0)X}{\sqrt{nT}}\mathbf{u} \\ &+ \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \\ &= \frac{\mathbf{u}'X'B_{nT}(\hat{\lambda})'B_{nT}(\hat{\lambda})X\mathbf{u}}{nT} + 2 \left\{ \left[\frac{-Y'(B_{nT}(\hat{\lambda}) - B_{nT}(\lambda_0))'B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right]' \right. \\ &- \left. \left[\frac{\left(Y'(B_{nT}(\lambda_0) - \beta_0'X'B_{nT}(\hat{\lambda}))\right)B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right]' \right\} \mathbf{u} \\ &+ \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \\ &= \frac{\mathbf{u}'X'B_{nT}(\hat{\lambda})'B_{nT}(\hat{\lambda})X\mathbf{u}}{nT} + 2 \left\{ \left[\frac{-Y'(B_{nT}(\hat{\lambda}) - B_{nT}(\lambda_0))'B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right]' \right. \\ &- \left. \frac{(\beta_0'X'(B_{nT}(\lambda_0) - B_{nT}(\hat{\lambda})) + \varepsilon')B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right\} \mathbf{u} \\ &+ \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{\mathbf{u}'X'B_{nT}(\hat{\lambda})'B_{nT}(\hat{\lambda})X\mathbf{u}}{nT} + 2\left\{ \left[\sqrt{nT}(\hat{\lambda} - \lambda_0) \left(\frac{\varepsilon'B_{nT}^{-1}(\hat{\lambda})GB_{nT}(\hat{\lambda})}{\sqrt{nT}} \right. \right. \right. \\
 &\quad \left. \left. \left. - \frac{X'B_{nT}(\hat{\lambda})\varepsilon\mathbf{u}}{\sqrt{nT}} \right) \right] \mathbf{u} \right. \\
 &\quad \left. + \frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} \left(\left| \beta_{0j} + \frac{\mathbf{u}_j}{\sqrt{nT}} \right| - \beta_{0j} \right) \right\}.
 \end{aligned} \tag{4-7}$$

由 Slutsky 定理得 $S(\mathbf{u}) - S(\mathbf{0}) \xrightarrow{d} V(\mathbf{u})$, 其中

$$V(\mathbf{u}) = \begin{cases} \mathbf{u}_B' \Phi_{11} \mathbf{u}_B - 2\xi_B' \mathbf{u}_B, & \mathbf{u}_j = 0, \forall j \notin B \\ \infty, & \text{其它.} \end{cases}$$

为此, 首先给出 $\sqrt{nT}(\hat{\theta} - \theta_0)$ 和 $\sqrt{nT}(\hat{\lambda} - \lambda)$ 的表达式:

对 $\partial \ln L_{nT}(\hat{\theta})/\partial \theta$ 在 $\theta = \theta_0$ 处进行泰勒展开:

$$\frac{\partial \ln L_{nT}(\hat{\theta})}{\partial \theta} = \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0) = 0, \tag{4-8}$$

其中, 由 $\|\theta^* - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$, $\hat{\theta} \xrightarrow{P} \theta_0$, 有 $\theta^* \xrightarrow{P} \theta_0$.

式 (4-8) 等价于

$$\sqrt{nT}(\hat{\theta} - \theta_0) = - \left[\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} \right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}, \tag{4-9}$$

令 $\Gamma = - \left[\frac{1}{nT} \frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta \partial \theta'} \right]^{-1} = \begin{pmatrix} * & * & * \\ \varsigma' & p_1 & p_2 \\ * & * & * \end{pmatrix}$, 其中 (ς', p_1, p_2) 表示矩阵 \mathbf{V} 的第二行元素, ς'

是 k 维的列向量, p_1 和 p_2 为标量. 由引理 4.2 可知, $\Gamma \xrightarrow{P} \Sigma^{*-1}$, d_{ij}^* 是矩阵 Σ^{*-1} 的第 i 行第 j 列元素, $i, j = 1, 2, 3$. 其中,

$$\varsigma' \xrightarrow{P} d_{21}^*, p_1 \xrightarrow{P} d_{22}^*, p_2 \xrightarrow{P} d_{23}^*, \tag{4-10}$$

由式 (4-3) 对 θ 求一阶导数得到:

$$\frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_0^2} X' B_{nT}(\hat{\lambda})' \varepsilon \\ -tr((I_T \otimes W_n) B_{nT}^{-1}(\lambda_0)) + \frac{1}{\sigma_0^2} \varepsilon' B_{nT}^{-1}(\lambda_0) \varepsilon \\ \frac{1}{2\sigma_0^4} (\varepsilon' \varepsilon - nT \sigma_0^2) \end{pmatrix}, \tag{4-11}$$

由式 (4-9)、式 (4-10) 得到:

$$\sqrt{nT}(\hat{\theta} - \theta_0) = \begin{pmatrix} \sqrt{nT}(\hat{\beta} - \beta_0) \\ \sqrt{nT}(\hat{\lambda} - \lambda_0) \\ \sqrt{nT}(\hat{\sigma}^2 - \sigma_0^2) \end{pmatrix} = \begin{pmatrix} * & * & * \\ \zeta' & p_1 & p_2 \\ * & * & * \end{pmatrix} \cdot \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}, \quad (4-12)$$

所以有,

$$\begin{aligned} \sqrt{nT}(\hat{\lambda} - \lambda_0) &= \zeta' \cdot \frac{1}{\sigma_0^2} X' B_{nT}(\hat{\lambda})' \varepsilon + p_1 \cdot [-tr((I_T \otimes W_n) B_{nT}^{-1}(\lambda_0)) + \frac{1}{\sigma_0^2} \varepsilon' B_{nT}^{-1}(\lambda_0) \varepsilon] \\ &+ p_2 \cdot \frac{1}{2\sigma_0^4} (\varepsilon' \varepsilon - nT\sigma^2), \end{aligned} \quad (4-13)$$

由式 (4-7)、式 (4-12)、式 (4-13) 和引理4.2得到:

$$\begin{aligned} &\left[-\frac{(B_{nT}(\hat{\lambda})Y_n - B_{nT}(\hat{\lambda})X\beta_0)' B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right]' \\ &= K^* \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + (\hat{\lambda} - \lambda_0) \frac{\varepsilon' (B_{nT}^{-1}(\hat{\lambda}))' H B_{nT}^{-1}(\hat{\lambda}) X}{\sqrt{nT}}, \end{aligned} \quad (4-14)$$

其中,

$$K^* = \left[\frac{X' B_{nT}(\hat{\lambda})' H X \beta_0}{nT} \zeta' - \sigma_0^2 I_K, \frac{X' B_{nT}(\hat{\rho})' H X \beta_0}{nT} p_1, \frac{X' B_{nT}(\hat{\lambda})' H X \beta_0}{nT} p_2 \right],$$

又由 $\frac{X' B_{nT}(\hat{\lambda})' H X \beta_0}{nT} \xrightarrow{P} \tilde{\nu}$, 以及(4-10)式, 可得 $K^* \xrightarrow{P} D^*$,

其中 $D^* = (\tilde{\nu} d_{21}^* - \sigma_0^2 I_k, \tilde{\nu} d_{22}^*, \tilde{\nu} d_{23}^*)$,

再由引理2得:

$$K \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, D^* \Omega^* D^{*'}), \quad (4-15)$$

由 $\hat{\lambda} - \lambda_0 \rightarrow 0$, $\frac{\varepsilon' B_{nT}^{-1}(\hat{\lambda})' H B_{nT}^{-1}(\hat{\lambda}) X}{\sqrt{nT}}$ 依分布收敛到某个正态分布, 所以

$$(\hat{\lambda} - \lambda_0) \frac{\varepsilon' B_{nT}^{-1}(\hat{\lambda})' H B_{nT}^{-1}(\hat{\lambda}) X}{\sqrt{nT}} \xrightarrow{P} 0, \quad (4-16)$$

再由式 (4-14)、式 (4-15)、式 (4-16) 有:

$$\left[-\frac{(B_{nT}(\hat{\lambda})Y - B_{nT}(\hat{\lambda})X\beta_0)' B_{nT}(\hat{\lambda})X}{\sqrt{nT}} \right]' \xrightarrow{d} \xi \sim N(0, D^* \Omega^* D^{*'}). \quad (4-17)$$

易知, $\frac{1}{nT} X' B_{nT}(\hat{\lambda})' B_{nT}(\hat{\lambda}) X \xrightarrow{P} \Phi$.

最后考虑式 (4-7) 中的 $\frac{\lambda_{nT}}{\sqrt{nT}} \sum_{j=1}^k \omega_j \sqrt{nT} (|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j})$: 如果 $\beta_{0j} \neq 0$ (即 $j \in \mathcal{B}$), 由假设 $\lambda_{nT}(nT)^{-1/2} \rightarrow 0$, $\omega_j = |\hat{\beta}_j|^{-r} \xrightarrow{P} |\beta_j|^{-r}$, 得到:

$$\left| \lambda_{nT} \omega_j \left(|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j} \right) \right| \leq \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} \omega_j \rightarrow 0,$$

如果 $\beta_{0j} = 0$ (即 $j \in \mathcal{B}^c$), 由假设 $\lambda_{nT}(nT)^{(r-1)/2} \rightarrow \infty$, $\omega_j = |\hat{\beta}_j|^{-r}, \sqrt{nT} |\hat{\beta}_j|^{-r} = O_p(1)$ 得到:

$$\left| \lambda_{nT} \omega_j \left(|\beta_{0j} + \frac{u_j}{\sqrt{nT}}| - \beta_{0j} \right) \right| = \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} \omega_j = \frac{|u_j|}{\sqrt{nT}} \lambda_{nT} (nT)^{r/2} (\sqrt{nT} |\hat{\beta}_j|)^{-r} \rightarrow \begin{cases} 0, u_j = 0 \\ \infty, u_j \neq 0. \end{cases}$$

由 Slutsky 定理得到 $Q(\mathbf{u}) - Q(\mathbf{0}) \xrightarrow{d} V(\mathbf{u})$,

$$V(\mathbf{u}) = \begin{cases} \mathbf{u}'_{\mathcal{B}} \Phi_{11} \mathbf{u}_{\mathcal{B}} - 2\xi'_{\mathcal{B}} \mathbf{u}_{\mathcal{B}}, & u_j = 0, \forall j \notin \mathcal{B} \\ \infty, & \text{其它.} \end{cases}$$

又因为 $V(\mathbf{u})$ 是凸函数, 所以存在唯一的最小值 $(\xi'_{\mathcal{B}} \Phi_{11}^{-1}, 0_{1 \times (k-k_0)})'$. 利用 Geyer^[31]、Knight 和 Fu^[32] 的结论得到:

$$\hat{\mathbf{u}}_{\mathcal{B}}^{(nT)} \xrightarrow{d} \Phi_{11}^{-1} \xi_{\mathcal{B}}, \quad \hat{\mathbf{u}}_{\mathcal{B}^c}^{(nT)} \xrightarrow{d} 0. \quad (4-18)$$

得到 $\sqrt{nT}(\hat{\beta}_{\mathcal{B}} - \beta_{\mathcal{B}}) \xrightarrow{d} \Phi_{11}^{-1} \xi_{\mathcal{B}}$, 证毕.

接下来证明相合性: 对 $\forall j \in \mathcal{B}$, (即 $\beta_j \neq 0$), 由渐近正态性有: $\tilde{\beta}_j \xrightarrow{P} \beta_{0j}$. 所以

$$P(\tilde{\beta}_j \neq 0) = P(j \in \tilde{\mathcal{B}}) \rightarrow 1, n \rightarrow \infty,$$

即

$$P(\mathcal{B} \subseteq \tilde{\mathcal{B}}) \rightarrow 1, n \rightarrow \infty.$$

则只需证明: $P(\mathcal{B} \supseteq \tilde{\mathcal{B}}) \rightarrow 1, n \rightarrow \infty$. 考虑事件 $j^* \in \tilde{\mathcal{B}}$, 由 Karush-Kuhn-Tucker(KKT) 最优条件^[33] 有,

$$-2\tilde{X}'_{j^*} (\bar{Y} - \bar{X} \tilde{\beta}) + \lambda_{nT} \omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*}) = 0,$$

上式等价于

$$2\tilde{X}'_{j^*} (\bar{Y} - \bar{X} \tilde{\beta}) = \lambda_{nT} \omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*}),$$

又因为

$$2 \frac{\bar{X}'_{j^*} (\bar{Y} - \bar{X} \tilde{\beta})}{\sqrt{nT}} = 2 \frac{\bar{X}'_{j^*} \bar{X} \sqrt{nT} (\tilde{\beta} - \beta)}{nT} + 2 \frac{\bar{X}'_{j^*} \varepsilon}{\sqrt{nT}}, \quad (4-19)$$

其中,

$$\frac{\bar{X}'_{j^*} \bar{X}}{nT} = \frac{X'_{j^*} B'_{nT}(\hat{\lambda}) B_{nT}(\hat{\lambda}) X}{nT} \xrightarrow{p} (\Phi_{j^*1}, \dots, \Phi_{j^*k}),$$

Φ_{j^*k} 是 Φ 的 $j^* \times k$ 阶元素. 通过式 (4-17) 和定理4.1的(i), 可以得到: $2 \frac{\bar{X}'_{j^*} \bar{X} \sqrt{nT} (\tilde{\beta} - \beta)}{nT}$ 依分布收敛到某个正态分布, 以及 $2 \frac{\bar{X}'_{j^*} \varepsilon}{\sqrt{nT}} \xrightarrow{d} N(0, 4 \|\bar{x}_{j^*}\|^2 \sigma^2)$. 因此,

$$P(j^* \in \tilde{\mathcal{B}}) \leq P \left(\left| 2 \frac{\bar{X}'_{j^*} (\bar{Y} - \bar{X} \tilde{\beta})}{\sqrt{nT}} \right| = \frac{\lambda_{nT} \omega_{j^*} \cdot \text{sgn}(\tilde{\beta}_{j^*})}{\sqrt{nT}} \right) \rightarrow 0, \quad n \rightarrow \infty.$$

即 $P(\mathcal{B} \supseteq \tilde{\mathcal{B}}) \rightarrow 1, n \rightarrow \infty$, 综上所述定理4.1的(ii)得证.

结论及展望

本文主要利用Adaptive LASSO方法对两类空间面板数据模型进行变量选择研究,构造空间面板自回归模型和空间面板误差模型对应的Adaptive LASSO惩罚估计函数,并证明了该方法具有良好的渐近性质,这对于今后继续研究空间面板数据模型与变量选择的理论提供了参考.

由于个人能力有限,论文撰写至此.由本文延伸出的问题值得今后进一步探讨:

1. 文中考虑的空间面板数据模型为静态模型,时变系数空间动态面板数据模型的变量选择问题值得后续进行研究,
2. 在数据缺失、半参数模型、部分线性可加等多种复杂情形下的空间面板模型变量问题还有待研究.

参 考 文 献

- [1] Anselin L. Spatial econometrics: methods and models[M]. Kluwer Academic, The Netherlands, 1988.
- [2] Kapoor M, Kelejian H H, Prucha I R. Panel data models with spatially correlated error components[J]. Journal of Econometrics, 2007, 140: 97-130.
- [3] Li Y H, Qin Y S, Li. Empirical likelihood for nonparametric regression models with spatial autoregressive errors[J]. Journal of the Korean Statistical Society, 2020(prepublish). doi: 10.1007/s42952-020-00088-z.
- [4] Baltagi B H, Egger P, Pfaffermayr M. A Generalized Spatial Panel Data Model with Random Effects[J]. Econometric Reviews, 2013, 32(5-6): 650-685.
- [5] Sun Y G, Malikov E. Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects[J]. Journal of Econometrics, 2018, 203(2): 359-378.
- [6] Baltagi B H, Fingleton B, Pirotte A. Estimating and Forecasting with a Dynamic Spatial Panel Data Model[J]. Oxford Bulletin of Economics and Statistics, 2014, 76(1): 112-138.
- [7] Lee L F, Yu J H. Efficient GMM estimation of spatial dynamic panel data models with fixed effects[J]. Journal of Econometrics, 2014, 180(2): 174-197.
- [8] Jin B, Wu Y H, Calyampudi Radhakrishna Rao, Li Hou. Estimation and model selection in general spatial dynamic panel data models[J]. Proceedings of the National Academy of Sciences, 2020, 117(10): 5235-5241.
- [9] Tibshirani, R. Regression Shrinkage and Selection Via the Lasso[J]. Journal of the Royal Statistical Society(Series B): Methodological, 1996: 267-288.
- [10] Fan J Q, Li R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties[J]. Journal of the American Statistical Association, 2001, 96(456): 1348-1360.
- [11] Fan J Q, Peng H. On Nonconcave Penalized Likelihood With Diverging Number of Parameters[J]. The Annals of Statistics, 2004, 32(3): 927-961.

- [12] Zou, Hui. The Adaptive Lasso and Its Oracle Properties[J]. *Publications of the American Statistical Association*, 2006, 101(476): 1418-1429.
- [13] Xie H, Huang J. SCAD-penalized regression in high-dimensional partially linear models[J]. *The Annals of Statistics*, 2009, 37(2): 673-696.
- [14] 王大荣, 张忠占. 线性回归模型中变量选择方法综述[J]. *数理统计与管理*, 2010, 29(04): 615-627.
- [15] Zou H, Zhang H H. On the adaptive elastic-net with a diverging number of parameters[J]. *Annals of Statistics*, 2009, 37(4): 1733-1751.
- [16] 王周伟, 陶志鹏, 张元庆. 基于Spatial AIC准则的空间自回归模型变量选择研究[J]. *数理统计与管理*, 2019, 38(01): 69-80.
- [17] 张元庆, 陶志鹏. 基于贝叶斯法则的空间自相关误差自相关模型变量选择研究[J]. *数理统计与管理*, 2016, 35(05): 826-837.
- [18] 张元庆, 陶志鹏. 广义嵌套空间模型变量选择研究——基于广义空间信息准则[J]. *统计研究*, 2017, 34(09): 100-107.
- [19] 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(03): 92-96.
- [20] 王晓瑞. 空间自回归模型变量选择的理论研究和实证分析[D]. 北京工业大学.
- [21] 谢琍, 唐甜, 王晓瑞. 线性空间自回归模型的不同惩罚函数下参数估计的比较及其实证分析[J]. *数理统计与管理*, 2019, 38(05): 823-835.
- [22] 李春雨. 两类空间计量模型的变量选择研究[D]. 广西师范大学.
- [23] Zhu J, Huang H C, Perla E R. On selection of spatial linear models for lattice data[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2010, 72(3): 389-402.
- [24] Liu T. Model Selection and Adaptive Lasso Estimation of Spatial Models[D]. The Ohio State University. 2017.
- [25] Wang H, Leng C. Unified LASSO Estimation by Least Squares Approximation[J]. *Publications of the American Statistical Association*, 2007, 102(479): 1039-1048.

- [26] Wang K. Variable selection for spatial semivarying coefficient models[J]. *Annals of the Institute of Statistical Mathematics*, 2018, 70(2): 323-351.
- [27] 王康宁, 林路. 空间非参回归的变量选择[J]. *中国科学:数学*, 2016, 46(03): 301-320.
- [28] Moran P A P. The Interpretation of Statistical Maps[J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1948, 10(2): 243-251.
- [29] Lee L F, Jihai Yu. Estimation of spatial autoregressive panel data models with fixed effects[J]. *Journal of Econometrics*, 2009, 154(2): 165-185.
- [30] Su L J, Yang Z L. Instrumental Variable Quantile Estimation of Spatial Autoregressive Models[J]. *Development Economics Working Papers*, East Asian Bureau of Economic Research, 2007:135-137.
- [31] Charles J G. On the Asymptotics of Constrained M -Estimation[J]. *The Annals of Statistics*, 1994, 22(4): 1993-2010.
- [32] Knight K, Fu W J. Asymptotics for lasso-type estimators[J]. *The Annals of Statistics*, 2000, 28(5): 1356-1378.
- [33] 侯定丕编著. 数理经济分析入门[M]. 中国科学技术大学出版社, 2003.
- [34] Li C, Wen L. The Asymptotic Properties of Quasi-maximum Likelihood Estimator for Spatial Error Panel Data Model[C]. *International Conference on Information Science Cloud Computing Companion*. IEEE, 2014.
- [35] Wu Y Q, Sun Y. Shrinkage estimation of the linear model with spatial interaction[J]. *Metrika*, 2017, 80(1): 51-68.
- [36] Baltagi B H, Liu L. Instrumental variable estimation of a spatial autoregressive panel model with random effects[J]. *Economics Letters*, 2011, 111(2): 135-137.
- [37] Lee L F. Asymptotic Distributions of Quasi Maximum Likelihood Estimators for Spatial Autoregressive Models[J]. *Econometrica*, 2004, 72(6): 1899-1925.

致 谢

阳光温热, 日历翻到了夏季, 宣示着毕业季的来临. 在研究生的三年当中, 我收获了知识的果实以及大家无私的帮助. 在离别之际, 我想诚挚地感谢在人生道路上帮助过我的老师和同学.

在此非常感谢我的导师秦永松教授. 秦老师是一位知识渊博, 治学严谨的老师, 能够成为他的学生, 我感到十分地骄傲和自豪. 他对学术研究的热爱和严谨, 让我深感敬佩, 从他身上我学会了对待学术要认真、严谨, 要静下心来去研读文献, 更要动笔去写文章. 生活当中, 老师是一位幽默豁达的长者, 他的谆谆善导, 给我人生上了重要的一课, 让我更加热爱生活 and 珍惜当下, 这三年来, 我一直以他为榜样激励自己学习和钻研学术, 我相信这会是我人生宝贵的财富.

此外, 我还要感谢雷庆祝教授、张军舰教授、张正家老师、晏振老师、廖旭老师等在这三年来对我的帮助和指导; 感谢李春雨师姐对我的支持和开导, 感谢曾庆樊同学、容健荣同学和刘雁同学对我的帮助和鼓励等, 感谢大家鼓励和支持我完成论文的写作. 正是因为有秦老师的帮助和同学的支持我才能坚持了下来, 在此也非常感谢同门的师兄师姐以及师弟师妹们对我学习上的帮助和生活上的关心!

最后, 还要感谢我最爱的家人, 他们力所能及的为我提供了一切可能的条件, 你们的负重前行成为我最坚强的后盾, 才让我在求学路上能够更加安稳地去追逐自己的目标, 谢谢你们给我前进的力量! 感谢评阅论文的各位专家老师, 谢谢你们的建议和指导.

论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下进行的研究工作及取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或其他机构已经发表或撰写过的研究成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

研究生签名： 唐莹莹 日期： _____

论文使用授权声明

本人完全了解广西师范大学有关保留、使用学位论文的规定。本人授权广西师范大学拥有学位论文的部分使用权，即：学校有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文；学校有权向国家有关部门或机构送交学位论文的复印件和电子版。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权广西师范大学学位办办理。

本学位论文属于：

保密

延迟公开 _____ 年（空白处必须填写延迟公开年限，一般不超过2年）

不保密

论文作者签名： 唐莹莹 日期： 2021.06

指导教师签名： 李永 日期： 2021.06

作者联系电话： _____ 电子邮箱： _____