应用Bootstrap方法的空间相关性检验

数理证明与模拟分析

● 龙志和　欧变玲　林光平　著

科学出版社

应用 Bootstrap 方法的空间相关性检验

数理证明与模拟分析

龙志和　欧变玲　林光平　著

科学出版社

北京

内 容 简 介

本书将基础理论研究与实际应用工具相结合，应用一种非参数方法——Bootstrap 方法解决空间经济计量分析中的空间相关性检验难题.

本书首先基于线性回归模型的 OLS 估计残差，利用 Bootstrap 方法构建空间相关性检验统计量，通过数理推导和模拟实验证明 Bootstrap 方法用于空间相关性检验的有效性；其次，基于空间滞后模型的 GMM 或 2SLS 估计残差，提出空间相关性检验统计量(OLL-Moran)的渐近分布和精确分布；最后，基于空间滞后模型的 2SLS 估计残差，利用 Bootstrap 方法构建空间相关性检验统计量，通过数理推导和模拟实验证明 Bootstrap 方法用于空间相关性检验的有效性. 本书在 Gauss 软件中编写一系列 Bootstrap 方法和空间相关性检验程序，丰富了 Gauss 软件工具箱.

本书可供高等院校和科研机构的研究人员，尤其是经济计量、空间经济计量和经济金融领域的研究者使用.

图书在版编目(CIP)数据

应用 Bootstrap 方法的空间相关性检验：数理证明与模拟分析/龙志和，欧变玲，林光平著. —北京：科学出版社，2011

ISBN 978-7-03-031655-4

Ⅰ.①应… Ⅱ.①龙… ②欧… ③林… Ⅲ.①经济计量分析 Ⅳ.①F224.0

中国版本图书馆 CIP 数据核字（2011）第 116494 号

责任编辑：陈玉琢／责任校对：赵桂芬
责任印制：钱玉芬／封面设计：王 浩

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

源海印刷有限责任公司印刷
科学出版社发行 各地新华书店经销
*
2011 年 6 月第 一 版 开本：B5(720×1000)
2011 年 6 月第一次印刷 印张：8 3/4
印数：1—2 000 字数：166 000

定价：28.00 元

(如有印装质量问题，我社负责调换)

目　录

第 1 章　绪　　论

1.1　空间关系：空间相关性和空间异质性

社会经济事物间存在着广泛的、普遍性的联系. 然而, 由于分析技术方面的困难, 在经典经济计量模型中, 基于 Gauss-Markov 假设, 假定变量相互独立, 研究对象间的空间效应被摒弃于模型之外, 所得结论往往与经济管理现实存在较大的差距. 近二十年来, 随着计算机技术的发展, 引入空间效应的空间经济计量分析技术出现了长足发展, 空间经济计量已经成为经济计量领域的一个重要分支和研究热点之一 (Anselin, 2007; Anselin et al, 2000).

空间效应是空间数据的基本性质. 它反映着空间因素的影响, 是空间经济计量学从传统经济计量领域独立出来的根本原因. 空间效应有两种主要形式: 空间相关性① (spatial dependence) 和空间异质性 (Spatial heterogeneity)(Anselin, 1988a). 就实际的空间数据而言, 两类空间效应往往是相关的. 在实际经济管理研究中, 如果忽略研究对象间的空间相关性或空间异质性, 数据分析结果就可能有偏 (Anselin et al, 1988). 故此, 在采用传统经济计量分析方法之前, 首先需要检验研究对象间的空间效应存在与否, 然后才能够决定使用传统经济计量方法, 还是采用空间经济计量方法. 可知, 空间效应检验对正确地进行空间数据分析至关重要.

空间异质性是指反映着空间个体内在特性之间存在差异的空间效应② (Anselin, 1988b), 也有学者将其定义为研究对象空间行为或空间关系上的不稳定性 (Haining, 1990; Bailey et al, 1995). 简单地说, 空间异质性就是以模型函数形式或参数表现出来的结构不稳定性 (Anselin, 1988a, 1988b). 空间经济计量模型设定前, 处理空间异质性十分重要, 也很困难. 研究对象间不稳定的区位分布是产生空间异质性的重要

---

① 文献中通常都没有区分空间相关性、空间依赖性和空间自相关 (spatial autocorrelation). 实际上, 空间相关性涵盖的范围比空间自相关更大. 在具体的空间经济计量分析中, 空间相关性泛指研究对象间存在空间关系 (比如, 空间滞后相关或空间误差相关), 而空间自相关通常特指模型的 OLS 估计残差间存在空间上的自相关关系.

② 空间异质性也是生态学系统的一个普遍特性. 陈玉福和董鸣 (2003) 对空间异质性的生态学研究新进展中提到: Wu (2000) 认为 “空间异质性是指生态学变量在空间上的不均匀性和复杂性, 表现为生态系统的缀块性和环境的梯度变化”, Li et al, (1995)“提出一个定量的、便于描述和应用的空间异质性的概念, 即‘空间异质性定义为所研究的系统特性在空间上的复杂性和/或变异性’”. 然而, 经济计量分析中, 空间异质性可以理解为 “所研究经济体的经济活动在空间上的复杂性或变异性”. 总的来看, 生态学中的空间异质性研究对象与本书不同, 然而其空间异质性的本质特征相似.

原因, 它通常表现为空间 "结构" 上的不一致性. 例如, 发达地区和不发达地区的经济结构, 呈现出空间差异性. 在不存在空间相关性时, 空间异质性的检验方法主要有 Goldfeld-Quandt 检验 (Goldfeld et al, 1965), Breusch-Pagan 检验 (Breusch et al, 1979) 以及 White 检验法 (White, 1980) 等. 当存在空间相关性时, Anselin(1988b) 提出用 Brusch-Pagan 统计量与 LM 统计量之和, 对空间相关性和空间异质性进行联合检验. Brunsdon 等 (1996) 提出使用地理加权回归方法 (geographically weighted regression, 简称 "GWR") 来处理研究对象间的空间差异 (或空间不稳定性), 揭示研究对象间的空间异质性. 在此基础上, Leung Yee 等 (2000) 给出了检验 GWR 估计残差间空间相关性的两个统计量: 基于地理加权回归残差的 Moran's I 和 Geary's c①. 理论上, 空间异质性仍然满足经典统计学和经济计量模型的假定, 可以通过传统经济计量技术来解决空间异质性问题. 但在实际经济分析中, 由于研究对象具有 "结构" 属性, 空间异质性和空间相关性经常共同存在且相互关联, 此时经典统计学和经济计量方法失效 (龙应根等, 2005). 可知, 在空间经济计量分析中, 空间异质性依然有待深入研究.

空间相关性是另一种重要的空间效应, 它是指不同位置的观测值在空间上非独立, 呈现出某种非随机的空间模式 (LeSage, 1999). 当区域位置 i 的观测值依赖于区域位置 $j(j \neq i)$ 的观测值时, 观测变量之间存在空间相关性. 由于空间观测值间存在相关性, 违反了经典统计学和经济计量学有关观测值不相关的假定条件, 传统方法对独立样本的统计推断将不再有效. 粗略来说, 与相同大小的独立样本相比, 存在空间相关性的样本将导致较大的方差估计、假设检验的低显著水平, 以及估计模型较低的拟合度. 简言之, 空间相关性会导致数据信息失真和经济计量分析有偏. 可见, 空间相关性检验是构建空间经济计量模型, 进行空间经济计量分析的关键.

在实际经济分析中, 研究对象的空间变异在一定程度上可以由经验知识进行初步判断和处理, 然而空间相关性的经验判断和处理相对较困难, 需利用专门的检验技术进行识别. 迄今为止, 空间经济计量研究者已经提出了 Moran's I、LM-error、LM-lag 等统计量检验研究对象间是否存在空间相关性. 本书第 1.3 节将详细介绍空间相关性检验方法.

1.2　空间经济计量模型

在经济计量学中, 常见的基本线性回归模型为

$$y = \boldsymbol{X}\beta + \varepsilon, \tag{1-1}$$

① 此处的 Moran's I 和 Geary's c 统计量与基于 OLS 估计残差的传统 Moran's I 和 Geary's c 不同, 详见 (Leung et al, 2000).

式中, y 为因变量, $\boldsymbol{X}$ 为自变量, ε 是模型误差. 空间经济计量模型的基本思想是将研究对象间的相互关系引入式 (1-1), 修正基本线性回归模型. 根据研究对象之间"空间"表现形式的不同, 目前的空间经济计量模型主要包括两种类型: 一种是空间滞后模型 (简称"SAR 模型"), 主要用于研究相邻机构或地区的行为对整个系统内其他机构或地区的行为都有影响的情形. 在实证研究中, 空间滞后模型是经常采用的空间经济计量模型. 比如, Anselin 对哥伦比亚地区犯罪问题的空间经济计量分析 (Anselin, 1988a), 便是空间滞后模型的经典范例.

空间滞后模型的具体表达式为

$$y = \lambda \boldsymbol{W} y + \boldsymbol{X}\beta + \varepsilon, \tag{1-2}$$

其中, $\boldsymbol{W}$ 是空间权重矩阵, 即反映各个机构或地区之间相互关系的网络结构矩阵. 空间权重矩阵的设定方法包括地理权重、经济权重和网络权重等. 通常设定 $\boldsymbol{W}$ 为二进制型的地理空间权重矩阵, 即相邻地区对应权重为 1 , 否则为 0, 且对角线元素为 0; 然后进行行标准化, 即使 $\boldsymbol{W}$ 的各行元素之和为 1 . 因此, $\boldsymbol{W}y$ 为周边因变量的加权平均, 视为空间滞后因变量. λ 是空间自回归系数, 其他变量含义与式 (1-1) 相同.

另一种是空间误差模型, 该模型中机构或地区间的相互关系通过误差项体现. 当研究对象之间的相互作用因所处的相对位置不同而存在差异时, 通常采用这种模型. 具体而言, 误差项的空间相关形式又分成两种基本的表达方式:

其一, 空间误差自相关模型 (简称"SEAR 模型"):

$$y = \boldsymbol{X}\beta + u, \quad u = \rho \boldsymbol{W} u + \varepsilon, \quad \text{或} \quad y = \boldsymbol{X}\beta + (I_N - \rho W)^{-1}\varepsilon; \tag{1-3}$$

其二, 空间误差移动平均模型 (简称"SEMA 模型"):

$$y = \boldsymbol{X}\beta + u, \quad u = \theta \boldsymbol{W}\varepsilon + \varepsilon, \quad \text{或} \quad y = \boldsymbol{X}\beta + (\boldsymbol{I}_N + \theta \boldsymbol{W})\varepsilon, \tag{1-4}$$

其中, ρ 是空间误差自相关系数, θ 是空间误差移动平均系数, $\boldsymbol{W}\varepsilon$ 和 $\boldsymbol{W}u$ 是空间误差滞后项, 空间权重矩阵设定同式 (1-2). 在空间误差模型中, 空间误差自相关模型更常见.

进一步, 在空间滞后模型的误差项中引入空间误差滞后项, 可以得到空间滞后模型的两个扩展模型, 即空间滞后误差自相关模型 (简称"SARAR 模型") 和空间滞后误差移动平均模型 (简称"SARMA 模型")①, 具体表达形式如下:

① 由于此模型是因变量一阶滞后相关和误差项一阶滞后相关, 相关文献中, 分别将"SARAR 模型"和"SARMA 模型"称为"SARAR(1,1) 模型"和"SARMA(1,1) 模型". 鉴于本书考虑的都是一阶情形, 不会引起混淆, 故书中将空间滞后误差自相关模型和空间滞后误差移动平均模型分别简称为"SARAR 模型"和"SARMA 模型".

SARAR 模型为

$$y=\lambda \boldsymbol{W}y+\boldsymbol{X}\beta+u,\quad u=\rho\boldsymbol{W}u+\varepsilon,\quad 或\quad y=\lambda\boldsymbol{W}y+\boldsymbol{X}\beta+(\boldsymbol{I}_N-\rho\boldsymbol{W})^{-1}\varepsilon. \tag{1-5}$$

SARMA 模型为

$$y=\lambda \boldsymbol{X}y+\boldsymbol{X}\beta+u,\quad u=\theta\boldsymbol{W}\varepsilon+\varepsilon,\quad 或\quad y=\lambda\boldsymbol{W}y+\boldsymbol{X}\beta+(\boldsymbol{I}_N+\theta\boldsymbol{W})\varepsilon. \tag{1-6}$$

其中, 各变量含义与式 (1-2)、式 (1-3) 和式 (1-4) 相同.

以上式 (1-2) 至式 (1-6) 是五类最常见的空间经济计量模型.

1.3 空间相关性检验方法

在空间数据分析中, 不论采用何种空间经济计量模型, 需要先对经济变量间是否存在空间相关性进行检验. 空间相关性检验大概可分为两类: 第一, 包括空间误差自相关和空间误差移动平均的空间误差相关检验; 第二, 空间滞后相关检验. 常见的空间相关性检验统计量, 如表 1-1 所示. 此外, 部分统计量既可检验研究对象间的空间误差相关关系又可检验空间滞后相关关系. 比如, 空间相关性 Moran's I 检验没有具体的备择假设模型, 既可以检验空间滞后相关也可以检验空间误差相关 (Anselin 等, 1996). 迄今, Moran's I 检验是最常用的空间相关性检验方法, 第 1.3.1 节将详细介绍 Moran's I 统计量的渐近分布和精确分布.

表 1-1 常见的空间相关性检验统计量

名称	表达式	渐近分布	来源	模型估计
Moran	$e'\boldsymbol{W}e/e'e$	$N(0,1)$	Cliff 和 Ord (1972)	OLS 估计
LM-error	$(e'\boldsymbol{W}e/\hat\sigma^2)^2/T$	$\chi^2(1)$	Burridge (1980)	
稳健 LM-error	$[e'\boldsymbol{W}e/\hat\sigma^2-\boldsymbol{T}\boldsymbol{J}^{-1}(e'\boldsymbol{W}y/\hat\sigma^2)]^2/[\boldsymbol{T}-\boldsymbol{T}^2\boldsymbol{J}^{-1}]$	$\chi^2(1)$	Bera 和 Yoon (1992)	
LR	$N[\ln(\hat\sigma_0^2)-\ln(\hat\sigma_1^2)]+2\ln\lvert\boldsymbol{I}_N-\hat\rho\boldsymbol{W}\rvert$	$\chi^2(1)$	Anselin (1988a)	ML 估计
Wald	$\hat\rho^2[t_2+t_3-(1/N)t_1^2]$	$\chi^2(1)$	Anselin (1988a)	
LM	$(e'\boldsymbol{W}e/\hat\sigma^2)^2\{T-[\mathrm{tr}((\boldsymbol{W}'+\boldsymbol{W})\boldsymbol{W}\boldsymbol{A}^{-1})]^2\mathrm{Var}(\hat\lambda)\}^{-1/2}$	$\chi^2(1)$	Anselin (1988a)	
KP-Moran	$e'\boldsymbol{W}e/\tilde\sigma$	$N(0,1)$	Kelejian 和 Prucha (2001)	2SLS 估计
LM-Lag	$(e'\boldsymbol{W}y/\hat\sigma^2)^2/J$	$\chi^2(1)$	Anselin (1988b)	OLS 估计
稳健 LM-Lag	$(e'\boldsymbol{W}y/\hat\sigma^2-e'\boldsymbol{W}e/\hat\sigma^2)^2/(J-T)$	$\chi^2(1)$	Bera 和 Yoon (1992)	
SARMA	$(e'\boldsymbol{W}y/\hat\sigma^2-e'\boldsymbol{W}e/\hat\sigma^2)^2/(\boldsymbol{J}-\boldsymbol{T})+(e'\boldsymbol{W}e/\hat\sigma^2)^2/T$	$\chi^2(2)$	Anselin (1988b, 1994)	

资料来源: 根据 Anselin 和 Florax (1995) 和 Kelejian 和 Prucha (2001) 整理. 表中各参数含义详见以下说明.

1.3.1 Moran's I 检验

1. Moran's I 检验的渐近分布

空间相关性 Moran's I 检验由 Moran (1948, 1950) 提出, Cliff 和 Ord (1972) 推导了 Moran's I 统计量的矩, 给出 Moran's I 的渐近分布形式, 发展了 Moran's I 检验. 根据式 (1-1) 基本模型, Moran's I 检验的原假设为 OLS 估计残差间不存在空间相关性, 备择假设为 OLS 估计残差间存在空间相关性, 这种空间相关性可能是空间滞后相关, 也可能是空间误差相关, 或二者兼有. Moran's I 统计量的表达式为

$$I = \frac{e'\boldsymbol{W}e}{e'e}, \tag{1-7}$$

其中, $e = y - \boldsymbol{X}\hat{\beta}$ 是线性回归模型式 (1-1) 的 OLS 估计残差; $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$; $\boldsymbol{W}$ 是行标准化的空间权重矩阵. 当模型误差服从正态独立同分布时, 空间相关性检验 Moran's I 统计量渐近服从期望为 $E(I)$ 和方差为 $\mathrm{Var}(I)$ 的正态分布①.

标准化的 Moran's I 统计量:

$$\frac{I - E(I)}{\sqrt{\mathrm{Var}(I)}} \sim N(0,1), \tag{1-8}$$

其中, 期望 $E(I) = \mathrm{tr}(\boldsymbol{MW})/(N-K)$, 矩阵 $M = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, $\boldsymbol{I}_N$ 是 N 阶单位矩阵, N 为样本量, K 为自变量 X 的列数, $\boldsymbol{W}$ 是行标准化的空间权重矩阵, $\mathrm{tr}(\cdot)$ 表示矩阵的迹, 即矩阵对角线元素之和; 方差 $\mathrm{Var}(I)$ 的计算公式为

$$\mathrm{Var}(\boldsymbol{I}) = \frac{\mathrm{tr}(\boldsymbol{MWMW}') + \mathrm{tr}[(\boldsymbol{MW})^2] + [\mathrm{tr}(\boldsymbol{MW})]^2}{(N-K)(N-K+2)} - [E(I)]^2.$$

Burridge (1981) 从 Lagrange 乘数角度给出了 Moran's I 统计量的另一种解释. King (1981) 和 Anselin 等 (1996) 研究了空间相关性检验的有限样本性质. 其中, King 发现, 当模型误差满足正态独立同分布的假定时, 在 $\rho = 0$ 的邻域内, Moran's I 检验是局部最佳不变量 (local best invariant).

Anselin 和 Kelejian (1997) 通过 Monte Carlo 模拟实验研究了回归模型中包含内生变量与采用 2SLS 方法进行模型估计时, Moran's I 检验的有限样本性质②.

① 当空间权重矩阵 $\boldsymbol{W}$ 没有进行行标准化时, Moran's I 统计量的表达式为 $I = Ne'\boldsymbol{W}e/(S_0e'e)$, Moran's I 渐近服从正态分布. 其中, Moran's I 检验渐近分布的期望 $E(I)$ 和方差 $\mathrm{Var}(I)$ 分别为: $E(I) = (N/S_0)\mathrm{tr}(\boldsymbol{MW})/(N-K)$; $\mathrm{Var}(I) = (N/S_0)^2\{\mathrm{tr}(\boldsymbol{MWMW}') + \mathrm{tr}[(\boldsymbol{MW})]^2 + [\mathrm{tr}(\boldsymbol{MW})]^2\}/[(N-K)(N-K+2)] - E(I)^2$, S_0 是空间权重矩阵 $\boldsymbol{W}$ 所有元素之和, 当空间权重矩阵行标准化时, $S_0 = N$, $N/S_0 = 1$. 此外, Tiefelsdorf (2000) 提到空间相关性 Moran's I 检验的渐近分布还与空间权重矩阵有关, 比如, 当空间权重矩阵基于环形单元构造时, Moran's I 检验可能不再服从正态分布而是渐近服从卡方分布.

② Anselin 和 Kelejian (1997) 重点在于通过 Monte Carlo 实验研究空间相关性检验统计量的有限样本性质, Kelejian 和 Prucha (2001) 侧重利用数理推导研究空间相关性检验统计量的大样本性质, 即渐近分布.

Kelejian 和 Prucha (2001) 将基于 OLS 估计残差的 Moran's I 统计量扩展到限制因变量模型和空间滞后模型中, 采用 2SLS 方法进行模型估计, 进而利用 2SLS 估计残差构建出类似的 Moran's I 统计量并提出其渐近分布①. Kelejian 和 Prucha (2001) 推导出检验空间滞后模型的 2SLS 估计残差间空间相关性存在与否的 Moran's I 统计量形式 (见表 1-1 中的 KP-Moran 检验). 空间滞后模型中 Moran's I 检验的原假设模型为空间滞后模型, 备择假设模型为空间滞后误差自相关模型, 或空间滞后误差移动平均模型. 此时, 根据空间滞后模型 (1-2), e 是 2SLS 估计残差. 当模型误差服从正态独立同分布时, 空间滞后模型中 KP-Moran 统计量渐近服从标准正态分布

$$\tilde{\sigma}^2 = \hat{\sigma}^2[\mathrm{tr}[(\boldsymbol{W}+\boldsymbol{W}')\boldsymbol{W}\hat{\sigma}^2 + e'(\boldsymbol{W}+\boldsymbol{W}')\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{W}+\boldsymbol{W}')e],$$

其中, $\hat{\sigma}^2 = e'e/N$, $Z = (\boldsymbol{W}y, \boldsymbol{X})$, $P = \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'$, $\boldsymbol{H}$ 为工具矩阵.

以上是基于 OLS 和 2SLS 估计残差构建的空间相关性检验 Moran's I 统计量的渐近分布. 实际上, 两种情况下的 Moran's I 统计量表达式略有不同. 为保持与常见的 Moran's I (Moran, 1950) 形式一致, 可将 Kelejian 和 Prucha (2001) 中的 Moran's I 检验, 即 KP-Moran 检验 (见表 1-1) 略作变形即可 (欧变玲等, 2009).

2. Moran's I 检验的精确分布

Tiefelsdorf (1995, 2000) 给出了基本模型 (1-1) OLS 估计残差构建的空间相关性检验 Moran's I 统计量的精确分布, 在此进行简单回顾.

因为 e 是线性回归模型的 OLS 估计残差, 所以 $e = \boldsymbol{M}\varepsilon$. 其中, $M = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. 据此可得, 式 (1-7) 等价于

$$I = \frac{\varepsilon'\boldsymbol{MWM}\varepsilon}{\varepsilon'\boldsymbol{M}\varepsilon}, \tag{1-9}$$

其中, 行标准化的空间权重矩阵 W 通常不对称, 为便于研究记 $V = (\boldsymbol{W}+\boldsymbol{W}')/2$. 从而, 上式可作如下等价变换:

$$I = \frac{\varepsilon'\boldsymbol{MVM}\varepsilon}{\varepsilon'\boldsymbol{M}\varepsilon}. \tag{1-10}$$

记随机变量 $\eta = L'\varepsilon$, 并且 $\eta \sim N(0, \sigma^2 I_N)$. 其中, $\boldsymbol{L}$ 是利用实对称矩阵 $\boldsymbol{MVM}$ 和幂等矩阵 $\boldsymbol{M}$ 的特征向量得到的正交矩阵. 进而, Moran's I 统计量可表示为

$$I = \sum_{i=1}^{N} \lambda_i \eta_i^2 \Big/ \sum_{i=1}^{N} \eta_i^2, \tag{1-11}$$

① Kelejian 和 Prucha (2001) P238 中明确指出, 原假设为 $\rho = 0$, 备择假设为 $\rho \neq 0$, 即原假设模型为空间滞后模型, 备择假设模型为空间滞后误差自相关模型 (SARAR 模型). 在表述方面, 本书与 Kelejian 和 Prucha (2001) 略有不同, 但实质相同.

其中, $\lambda_i(i=1,\cdots,N)$ 是矩阵 $\boldsymbol{MVM}$ 的特征值.

Tiefelsdorf (2000) 利用 Imhof 定理 (Imhof, 1961), 给出了基本模型 (1-1) OLS 估计残差的空间相关性检验 Moran's I 统计量的精确分布. 具体来说, 当误差 $\varepsilon \sim N(0,\sigma^2 I_N)$ 时, Moran's I 统计量的分布函数为

$$
\begin{aligned}
P(I \leqslant t) &= P\left(\sum_{i=1}^{N}(\lambda_i - t)\eta_i^2 \leqslant 0\right) \\
&= \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \frac{1}{u}\sin(\theta(u))\rho(u)\mathrm{d}u
\end{aligned}
\tag{1-12}
$$

其中, $\theta(u) = \frac{1}{2}\sum_{i=1}^{N}\arctan[u(\lambda_i - t)]$, $\rho(u) = \prod_{i=1}^{N}[1+u^2(\lambda_i - t)^2]^{-1/4}$.

式 (1-12) 即是基于 OLS 估计残差构建的空间相关性 Moran's I 统计量的精确分布. 尽管利用 Moran's I 统计量的精确分布能更有效地识别 OLS 估计残差间的空间相关性, 但是当样本量较大 ($N > 30$) 时, 这将需要花费大量的计算成本 (Tiefelsdorf, 2000).

利用 Moran's I 统计量的精确分布和渐近分布能够检验研究对象间是否存在空间相关性, 但不能确定空间关系以何种形式存在. LM-error 检验、稳健 LM-error 检验、LM-lag 检验以及稳健 LM-lag 检验等, 则可进一步明确空间关系究竟是空间误差相关还是空间滞后相关.

1.3.2 空间误差相关检验

1. LM-error 检验和稳健 LM-error 检验

LM-error 检验和稳健 LM-error 检验 (Robust LM-error) 都是基于 Lagrange 乘数原理构建, 适用于检验基本模型 (1-1) OLS 估计残差间的空间误差相关. Burridge (1980) 最先提出 LM-error 检验. 此后, Bera 和 Yoon (1992) 又对 LM-error 检验进行改进, 提出稳健 LM-error 检验. LM-error 检验和稳健 LM-error 检验的原假设都是 OLS 估计残差间不存在空间关系, 即 $\rho = 0$ 且 $\theta = 0$; 备择假设是 OLS 估计残差间存在式 (1-3) 或式 (1-4) 所示的空间误差自相关或空间误差移动平均关系, 即 $\rho \neq 0$ 或 $\theta \neq 0$. LM-error 检验和稳健 LM-error 检验的表达式分别如下所示:

$$
\text{LM-error} = (e'\boldsymbol{W}e/\hat{\sigma}^2)^2/\boldsymbol{T}, \tag{1-13}
$$

$$
\text{稳健 LM-error} = [e'\boldsymbol{W}e/\hat{\sigma}^2 - T\boldsymbol{J}^{-1}(e'\boldsymbol{W}y/\hat{\sigma}^2)]^2/[\boldsymbol{T} - \boldsymbol{T}^2\boldsymbol{J}^{-1}], \tag{1-14}
$$

式中, $\hat{\sigma}^2 = e'e/N$, $T = \mathrm{tr}(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}\boldsymbol{W})$, $\boldsymbol{J} = \boldsymbol{T} + (\boldsymbol{W}\boldsymbol{X}\hat{\beta})'\boldsymbol{M}(\boldsymbol{W}X\hat{\beta})/\hat{\sigma}^2$, e 是线性回归模型 (1-1) 的 OLS 估计残差, $e = y - \boldsymbol{X}\hat{\beta}$, $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$. 从而, $\boldsymbol{W}\boldsymbol{X}\hat{\beta}$ 可以看成因变量估计值的空间滞后项.

可见, LM-error 检验和稳健 LM-error 检验都是基于线性回归模型的 OLS 估计残差构建. 它们能够检验研究对象间的空间误差关系存在与否, 但是不能确定这种空间误差关系是空间误差自相关还是空间误差移动平均. 当存在空间滞后型的局部设定偏误时, 稳健 LM-error 检验比 LM-error 检验更可靠① (Anselin et al, 1995).

2. LR 检验和 Wald 检验

与 LM-error 检验类似, Anselin (1988a) 基于线性模型 (1-1) 和空间经济计量模型 (1-3) 的 ML 估计提出了 Wald 检验和 LR 检验 (或似然比检验), 判断残差间是否存在空间自相关. 具体来说, Wald 检验和 LR 检验的原假设 $H_0: \rho = 0$, 备择假设 $H_1: \rho \neq 0$. 为说明 Wald 检验和 LR 检验的构造思路, 先需要简单介绍无约束模型 (空间误差自相关模型) 和有约束模型 (线性回归模型) 的似然函数 (Anselin 1988a). SEAR 模型 ($\rho \neq 0$) 和线性模型 ($\rho = 0$) 的似然函数②分别为

$$\begin{aligned} L = & -\frac{N}{2}\ln(\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|I_N - \rho \boldsymbol{W}| \\ & -\frac{1}{2}\sigma^{-2}(y - X\beta)'(\boldsymbol{I}_N - \rho\boldsymbol{W})'(\boldsymbol{I}_N - \rho\boldsymbol{W})(y - \boldsymbol{X}\beta); \\ L_0 = & -\frac{N}{2}\ln(\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\sigma^{-2}(y - \boldsymbol{X}\beta)'(y - \boldsymbol{X}\beta). \end{aligned}$$

由上可得, LR 检验的表达式为

$$LR = N[\ln(\sigma_0^2) - \ln(\sigma_1^2)] + 2\ln|\boldsymbol{I}_N - \hat{\rho}\boldsymbol{W}|, \tag{1-15}$$

其中, σ_0^2 和 σ_1^2 分别是 SEAR 模型式 (1-3) 和线性回归模型式 (1-1) 的 ML 估计残差的方差估计量, $\hat{\rho}$ 是空间误差自相关系数 ρ 的 ML 估计量. 当模型误差服从正态独立同分布时, LR 检验渐近服从自由度为 1 的卡方分布. 如果约束 $\rho = 0$ 有效, 即增加约束条件 $\rho = 0$ 不会导致似然函数值降低太多, 则似然比 LR 接近 0, 接受原假设; 否则, 拒绝原假设.

Wald 检验的表达式为

$$\text{Wald} = \hat{\rho}^2[t_2 + t_3 - (1/N)t_1^2], \tag{1-16}$$

其中, $t_1 = \text{tr}(\boldsymbol{W}\boldsymbol{B}^{-1})$, $t_2 = \text{tr}(\boldsymbol{W}\boldsymbol{B}^{-1})^2$, $t_3 = \text{tr}[(\boldsymbol{W}\boldsymbol{B}^{-1})'(\boldsymbol{W}\boldsymbol{B}^{-1})]$, $B = \boldsymbol{I}_N - \hat{\rho}\boldsymbol{W}$. 当模型误差服从正态独立同分布时, Wald 检验渐近服从自由度为 1 的卡方分布. 显

① Anselin, Florax (1995) 解释稳健 LM-error 检验时, 原文为: *The adjusted Lagrange Multiplier test of Bera and Yoon (1992) that is robust to local misspecification in the form of a spatial lag term.*

② 详细推导过程参见 Anselin 的经典空间经济计量著作 *Spatial Econometrics: Methods and Models* (1988a) 第六章.

然, 由式 (1-16) 可见, 当 Wald → 0 时, $\hat{\rho} = 0$, 接受模型不存在空间误差自相关的原假设; 否则, 拒绝原假设, 认为模型存在空间误差自相关.

3. LM 检验

以上四种检验的原假设模型都是线性回归模型, Anselin (1988a) 给出了判断空间滞后模型 ML 估计残差间空间相关性存在与否的 LM 检验. LM 检验的原假设模型是空间滞后模型, 即 $\lambda \neq 0, \rho = 0$; 备择假设模型是空间滞后误差自相关模型, 即 $\lambda \neq 0$ 且 $\rho \neq 0$. LM 检验的表达式为

$$LM = (e'\boldsymbol{W}e/\hat{\sigma}^2)^2\{T - [\mathrm{tr}((\boldsymbol{W}' + \boldsymbol{W})\boldsymbol{W}\boldsymbol{A}^{-1})]^2\mathrm{Var}(\hat{\lambda})\}^{-1/2} \tag{1-17}$$

其中, $\hat{\sigma}^2$ 是空间滞后模型 ML 估计残差的方差估计量, $T = \mathrm{tr}(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}\boldsymbol{W})$, $\boldsymbol{A} = \boldsymbol{I}_N - \hat{\lambda}\boldsymbol{W}$, $\hat{\lambda}$ 和 $\mathrm{Var}(\hat{\lambda})$ 分别是空间自回归系数 λ 及其方差的 ML 估计量. 当模型误差服从正态独立同分布时, LM 检验统计量渐近服从自由度为 1 的卡方分布.

式 (1-17) 中的 LM 检验基于空间滞后模型的 ML 估计结果构建, 前述 KP-Moran 检验基于空间滞后模型的 2SLS 估计结果构建, 二者都可用于检验空间滞后模型的估计残差间是否存在空间相关性. 不过, 与 KP-Moran 检验中的 2SLS 估计相比, LM 检验涉及的 ML 估计非常耗时, 不适于大量的 Monte Carlo 模拟实验. LM 检验和 KP-Moran 检验都可检验空间滞后模型中是否存在空间误差相关, Anselin 等 (1996) 提出了另外一种空间相关性检验, 用于检验空间误差模型中是否存在空间滞后相关.

1.3.3 空间滞后相关检验

空间滞后相关检验主要有 LM-lag 检验和稳健 LM-lag 检验两类. 二者的关系类似于 LM-error 检验和稳健 LM-error 检验. Anselin (1988b) 首先提出了 LM-lag 检验来判断研究对象间的空间滞后相关, Bera 和 Yoon (1992) 进一步改进 LM-lag 检验, 提出了稳健 LM-lag 检验. 稳健 LM-lag 检验是指误差间不存在空间相关性的假设条件可能有误时的空间滞后相关检验①. LM-lag 检验和稳健 LM-lag 检验的原假设模型都是线性回归模型 ($\lambda = 0$), 备择假设模型都是空间滞后模型 ($\lambda \neq 0$). LM-lag 检验和稳健 LM-lag 检验的具体形式如下

$$\text{LM-lag} = (e'\boldsymbol{W}y/\hat{\sigma}^2)^2/J, \tag{1-18}$$

$$\text{稳健 LM-lag} = (e'\boldsymbol{W}y/\hat{\sigma}^2 - e'\boldsymbol{W}e/\hat{\sigma}^2)/(\boldsymbol{J} - \boldsymbol{T}), \tag{1-19}$$

① Anselin 和 Florax (1995) 解释稳健 LM-lag 检验时, 原文为: *A test for a spatial lag robust to local misspecification in the form of a spatial moving average error process.*

其中, e 是线性回归模型 (1-1) 的 OLS 估计残差, 其他参数含义与前述相同. 当模型误差满足正态独立同分布的假设时, LM-lag 检验和稳健 LM-lag 检验渐近服从自由度为 1 的卡方分布.

除 Moran's I 检验、空间滞后相关检验和空间误差检验之外, Anselin(1988b, 1994) 提出了基于 OLS 估计残差构建的 SARMA 检验, 如表 1-1 所示. 当模型误差服从正态独立同分布时, SARMA 检验渐近服从自由度为 2 的卡方分布. SARMA 检验能够同时判断 OLS 估计残差间是否存在空间滞后相关和空间误差相关. SARMA 检验的原假设模型是线性回归模型, 备择假设模型是空间滞后误差移动平均模型 ($\lambda \neq 0, \theta \neq 0$). 从表 1-1 可见, SARMA 检验相当于稳健 LM-lag 检验与 LM-error 检验之和, 也是稳健 LM-error 检验和 LM-lag 检验之和.

从目前的空间经济计量实际研究情况看, 上述多种空间相关性检验方法中, Moran's I 检验、LM-error 检验和 LM-lag 检验较常用.

1.4　空间相关性检验方法的异同及面临的问题

从上述各种空间相关性检验论述可以看到彼此间存在一定的差异, 且有着共同的前提条件及局限性. 首先, 不同空间相关性检验方法的备择假设模型不同. 除 Moran's I 检验外, 其他空间相关性检验方法的备择假设模型都是具体的空间经济计量模型. 比如, LM-error 检验的备择假设模型是空间误差自相关模型或空间误差移动平均模型, LM-lag 检验的备择假设模型是空间滞后模型. Moran's I 检验能够识别出任何形式的空间关系, 这使得 Moran's I 检验在空间数据分析中, 更能有效地捕捉到研究对象间的空间相关性.

其次, 空间相关性检验统计量的构造基础不同. 空间误差相关检验中的 LM-error 检验和稳健 LM-error 检验, 空间滞后相关检验中的 LM-lag 检验和稳健 LM-lag 检验, 以及同时检验空间误差相关和空间滞后相关的 SARMA 检验都基于原假设线性回归模型的 OLS 估计构建, 而空间误差相关检验中的 LR 检验和 Wald 检验分别基于原假设线性回归模型和 SEAR 模型的 ML 估计残差构建, LM 检验基于原假设 SAR 模型构建. 与 Moran's I 检验相比, LM-error 检验、稳健 LM-error 检验、LM-lag 检验、稳健 LM-lag 检验, 以及 SARMA 检验虽然也都是利用 OLS 估计残差构建, 但是表达式相对复杂; LR 检验、Wald 检验和 ML 检验都基于 ML 估计残差构建, 计算过程非常耗时, 尤其不适于利用 Monte Carlo 进行模拟实验. 简言之, 与其他空间相关性检验相比, 基于 OLS 和 2SLS 估计残差构建的 Moran's I 检验, 不仅结构简单而且计算速度较快. 故此, 鉴于以上两点, 本书选择最常用的空间相关性 Moran's I 检验为研究对象.

最后, 各种空间相关性检验渐近分布成立的前提条件都是模型误差服从正态

独立同分布[①]. 模型误差满足正态独立同分布和样本量较大这两个假定条件是 Moran's I、LM-error 和 LM-lag 检验等能够有效地判断空间相关性存在与否的必要前提, 二者缺一不可. 然而, 在大量的研究工作中, 以上两个前提条件通常都不满足. 具体来说, 第一, 模型误差通常不服从正态独立同分布. 比如, 存在异方差或分布未知; 第二, 受数据收集的限制, 样本量通常有限. 比如, 对中国大陆省域经济进行空间经济计量分析时, 样本量只有 31 个. 由此可见, 在上述情况下, 常用的空间经济计量检验方法有偏, 不再有效. 实际经济研究中, 忽视模型误差是否服从正态独立同分布与可得样本有限, 是造成许多空间经济计量模型估计结果有偏的原因之一. 也就是说, 在空间经济计量分析中, 模型误差不满足正态独立同分布或样本量有限情况下的空间相关性检验是空间经济计量研究中尚未解决的难题, 亦是本书中心议题.

采用 Bootstrap 模拟技术构造统计量, 是解决空间相关性检验难题可供选择的路径之一 (龙志和等, 2009). 文献中, 国际学界已有学者从数理推导与模拟实验两方面证明, Bootstrap 方法在经济计量模型假设检验中有效 (Davidson R et al., 1999, 2006; Chang, 2004; Hall, 1992, 1994 等). 但是, 在引入空间权重矩阵和空间变量后, 空间经济计量模型变得更为复杂, 采用 Bootstrap 方法构造统计量进行空间相关性检验是否有效则需要严格证明. 文献上, 国内外学者对空间经济计量模型的空间相关性检验中 Bootstrap 方法有效性的理论和应用研究迄今仍是空白. 除本书的相关研究以外, 仅有极少数的空间经济计量研究者 (Warren, 2008a, 2008b, Burridge, 2009 等) 尝试着将 Bootstrap 方法用于空间经济计量模型的空间相关性检验问题.

从国内外对空间经济计量分析的现实需要和理论背景看, 研究有限样本量或模型误差存在异方差或分布未知情况下, 空间相关性检验中 Bootstrap 方法的有效性, 即空间经济计量模型检验的有限样本性质, 具有重要的理论与实际应用价值. 它能够为采用 Bootstrap 方法解决空间经济计量模型检验问题提供理论基础, 开拓一条解决模型误差分布未知或异方差情况下, 空间经济计量模型检验难题的有效路径; 同时, 研究成果将扩展 Bootstrap 方法和空间经济计量模型在现实经济管理的应用范围, 为经济管理应用研究者有效使用 Bootstrap 方法提供便利.

1.5 本书结构

本书研究空间经济计量模型检验的有限样本性质, 内容属于空间经济计量领域中的空间关系检验范畴. 本书研究内容在空间经济计量领域中的位置, 如图 1-1 所示.

① K-R 检验除外, 它不需要正态独立同分布假定, 但 K-R 检验不能有效用于样本量较小情形, 并且求解麻烦.