

叶阿忠 吴继贵 陈生明 等 著

空间计量 经济学

KONGJIAN JILIANG
JINGJIXUE

 厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

叶阿忠 吴继贵 陈生明 等 著

空间计量 经济学

KONGJIAN JILIANG
JINGJIXUE



厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

图书在版编目(CIP)数据

空间计量经济学/叶阿忠等著. —厦门:厦门大学出版社, 2015. 8

ISBN 978-7-5615-5690-0

I. ①空… II. ①叶… III. ①区位经济学-计量经济学-研究 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2015)第 188124 号

官方合作网络销售商:



厦门大学出版社出版发行

(地址:厦门市软件园二期望海路 39 号 邮编:361008)

总编办电话:0592-2182177 传真:0592-2181406

营销中心电话:0592-2184458 传真:0592-2181365

网址:<http://www.xmupress.com>

邮箱:xmup@xmupress.com

沙县四通彩印有限公司印刷

2015 年 8 月第 1 版 2015 年 8 月第 1 次印刷

开本:787 × 1092 1/16 印张:13 插页:2

字数:400 千字 印数:1~3 000 册

定价:28.00 元

本书如有印装质量问题请直接寄承印厂调换

序 言

本书是叶阿忠教授和他的团队多年来从事空间计量经济学理论与应用研究的成果汇集,在我所见过的仅有的几本由国内学者编写的空间计量经济学专门书籍中,本书具有兼顾基础和前沿、教科书和专著的特点。

空间计量经济学的发展虽然已有 30 年的历史,但它仍然是计量经济学领域发展较晚的一个年轻的分支。原因是多方面的。第一,经典的截面数据(包括面板数据)计量经济学模型理论的一个基本假设是:作为样本的截面个体是随机独立抽取的。既然是随机独立抽取,就没有必要考虑它们之间的相关问题,因此,空间相关问题长期被忽视。第二,如果截面个体之间存在相关性,那么这种相关性远比时间序列计量经济学模型中的相关性复杂。例如,人们可以假定时序样本之间的相关性只与时间间隔有关,而与时间点无关,而这个假定对于截面样本显然是不合适的。第三,截面数据计量经济学模型一旦包含描述相关性的空间矩阵,其估计和检验技术就相当复杂,并且对应用软件提出了新的要求。我国的计量经济学教学与研究起步于 20 世纪 80 年代初,从最成熟的经典计量经济学模型开始,逐步深入,直到近 10 年才有学者关注空间计量经济学。所以,在我国,空间计量经济学显得更为年轻,本书的出版也就具有开拓和推动该分支学科发展的意义。

本书总体上可以分为两部分。第一至四章是已经发展的空间计量经济学模型理论方法与应用的系统介绍,包括截面数据空间模型、面板数据空间模型,以及扩展的空间模型。经过作者的系统梳理,辅之以应用软件的介绍和应用实例的演示,形成了完整、清晰的空间计量经济学内容体系,体现了本书的教科书特色。读者通过这部分内容的学习,对空间计量经济学就有了基本的掌握。第五、六两章是作者新的研究成果,将半参数方法应用于各种空间模型,综合了计量经济学的若干前沿,体现了作者对空间计量经济学理论方法的发展和贡献,形成了本书的专著特色。叶阿忠教授是国内较早进入非参数计量经济学领域的专家之一,出版了多本非参数计量经济学专著,由他来承担这部分内容的研究,其学术水平是毋庸置疑的。有兴趣的读者通过这部分内容的学习,可以说已经进入空间计量经济学的前沿了,当然,是前沿之一。

本书中关于描述空间相关性的空间矩阵的构造,以及绝大部分应用实例,都是基于“地理空间”的。这可以从三个方面来解释:一是以地理上的截面个体(例如各个地区)作为样本建立截面数据(或者面板数据)计量经济学模型,一直就是计量经济学模型应用研究的重要领域,关于它的空间相关问题必须着重研究并解决;二是地理上的截面个体样本,例如以我国大陆 31 个省级地区为样本,很难满足“随机独立抽样”的假定;即使采用随机抽样的方法,例如从全国所有县级地区中随机抽取 200 个作为样本,也不能忽略它们之间的相关性。三是地理上的截面个体之间的空间相关性比较直观,描述该相关性具有一

定的依据,即具有可行性。当然,本书中介绍的一些基本理论和方法,同样可以应用于其他截面个体样本,例如家庭、个人、企业等,并没有失去一般性。

本书中包含大量应用实例,为读者学习与掌握理论方法提供了很大的方便。虽然这些实例中的大部分本身就是很有价值的应用研究成果,但是,作为例题,它们更注重的是关于所在章节涉及的理论方法的演示,而没有对建立应用模型的其他方面进行全面的讨论。读者在遇到类似的研究课题时,可以参考,不宜照搬。

本人对空间计量经济学没有专门研究,阅读的文献很少,本不敢为本书作序。浏览了本书书稿,很是受益,写下以上几段话,实为读后感,以飨读者。

李子奈

2015年2月于清华大学经济管理学院

前 言

随着交通的便利和信息网络的高速发展,区域之间经济活动越来越具有空间相关性。空间计量经济学也随着实际研究的需要而迅速发展,本著作涉及很多该领域的前沿研究成果,前四章主要参考了许多国内外学者的研究成果。第五章和第六章是我们自己的研究成果。第五章是教育部人文社会科学基金项目“半参数空间计量经济联立方程模型的理论研究及其应用”(10YJA790227)和教育部高等学校博士点基金项目“半参数空间面板数据联立方程模型理论和应用”(20103514110009)的研究成果。第六章是国家自然科学基金项目“半参数空间向量自回归模型的理论研究及其应用”(71171057)的部分研究成果。

本著作分六章,第一章是空间计量经济学基础,包括空间效应、空间权重和 GeoDa 软件的相关操作;第二章是空间回归模型,包括空间回归模型的动因、空间滞后模型、空间误差模型、空间计量模型的检验和 GeoDa 软件的相关操作;第三章是面板空间回归模型,包括面板数据空间滞后模型、面板数据空间误差模型和 Matlab 的相关操作;第四章是空间回归模型的扩展模型,包括空间变系数回归模型、空间杜宾模型、空间误差修正模型、广义空间回归模型、动态空间回归模型和矩阵指数空间模型;第五章是半参数空间滞后模型,包括半参数横截面空间滞后模型、半参数面板空间滞后模型和半参数动态面板空间滞后模型;第六章是半参数空间向量自回归模型,包括横截面数据半参数空间向量自回归模型、横截面数据半参数空间结构向量自回归模型和面板数据半参数空间向量自回归模型。

该著作的初稿的第一章主要由陈生明和王福军编写,第二章主要由陈君、陈晓玲和王福军编写,第三章主要由邢晓卫和王福军编写,第四章主要由王福军、陈君、陈晓玲和邢晓卫编写。第五章主要由叶阿忠、陈生明、冯烽、郭炬、陈泓、吴继贵、叶娟惠、张枝招、张楠、胡乐琼、陈晓玲、张玺和郑万吉著。第六章主要由叶阿忠、吴继贵、邢晓卫、陈晓玲、张长淮、郑万吉和蒋玲香著。全书最终由叶阿忠统稿完成。

感谢相关国内外学者在该领域的学术贡献!感谢厦门大学出版社经管编辑室吴兴友编辑对此书出版的大力支持!也感谢福建人民出版社黄杰阳的帮助!也感谢陈明英女士的支持!感谢国家自然科学基金委、教育部人文社会科学基金委和教育部高等学校博士点基金委对项目的资助!也特别感谢清华大学博士生导师李子奈教授在我读博士期间对我的培养,使我进入非参数计量经济学理论和应用研究领域。由于我们的学术水平有限,加之时间仓促,书中的错漏和疏忽在所难免,恳请读者批评指正。

叶阿忠

2015年2月于福州大学计量经济研究所

目 录

第 1 章 空间计量经济学基础	1
1.1 空间效应	1
1.1.1 空间相关性	1
1.1.2 空间相关性与检验方法	2
1.2 空间权重	6
1.2.1 空间矩阵的常规设定	6
1.2.2 基于邻近概念的空间权重矩阵	6
1.2.3 其他经济社会空间权重矩阵	7
1.3 GeoDa 软件的相关操作	10
1.3.1 GeoDa 基本操作	10
1.3.2 空间权重矩阵	11
1.3.3 空间相关性	13
第 2 章 空间回归模型	15
2.1 空间回归模型的动因	15
2.1.1 残差与解释变量的正交性问题	15
2.1.2 模型不确定性动因	17
2.1.3 空间回归模型	17
2.2 空间滞后模型	18
2.2.1 IV 估计	18
2.2.2 ML 估计	19
2.2.3 GMM 估计	20
2.3 空间误差模型	21
2.3.1 模型	21
2.3.2 模型的估计	22
2.4 空间回归模型的检验	22
2.4.1 LM 检验	22
2.4.2 残差空间相关性的 Moran's I 检验	24
2.5 实例	25
2.6 GeoDa 软件的相关操作	26

第 3 章 面板空间回归模型	28
3.1 面板数据空间滞后模型	28
3.1.1 固定效应	28
3.1.2 随机效应	29
3.2 面板数据空间误差模型	30
3.2.1 固定效应	30
3.2.2 随机效应	31
3.3 实例	31
3.4 MATLAB 的相关操作	33
第 4 章 空间回归模型的扩展模型	43
4.1 空间变系数回归模型(GWR)	43
4.1.1 地理加权回归估计方法	43
4.1.2 空间变系数的地理加权回归模型	45
4.1.3 实例	47
4.2 空间杜宾模型	49
4.2.1 模型	49
4.2.2 实例	50
4.3 空间误差修正模型	59
4.3.1 模型	60
4.3.2 估计方法	60
4.3.3 实例	60
4.4 广义空间回归模型	64
4.4.1 模型及估计	64
4.4.2 实例	65
4.5 动态空间回归模型	68
4.5.1 模型	68
4.5.2 模型的估计	69
4.5.3 实例	69
4.6 矩阵指数空间回归模型	71
4.6.1 模型	72
4.6.2 实例	73
第 5 章 半参数空间滞后模型	77
5.1 半参数横截面空间滞后模型	77
5.1.1 模型	77
5.1.2 模型的估计及其性质	78

5.1.3 实例	79
5.2 半参数面板空间滞后模型	84
5.2.1 模型	84
5.2.2 工具变量估计	85
5.2.3 两阶段最小二乘估计	97
5.2.4 广义矩估计	104
5.3 半参数动态面板空间滞后模型	114
5.3.1 模型	114
5.3.2 模型的估计	114
5.3.3 实例	116
第 6 章 半参数空间向量自回归模型	126
6.1 横截面数据半参数空间向量自回归模型	127
6.1.1 模型	127
6.1.2 模型估计	128
6.1.3 时空脉冲响应函数	128
6.1.4 实例	129
6.2 横截面数据半参数空间结构向量自回归模型	137
6.2.1 模型	137
6.2.2 模型的估计	138
6.2.3 实例	139
6.3 面板数据半参数空间向量自回归模型	149
6.3.1 模型	149
6.3.2 模型估计	150
6.3.3 实例	152
6.4 面板数据半参数空间结构向量自回归模型	181
6.4.1 模型	181
6.4.2 模型估计	181
6.4.3 实例	183
参考文献	195

第1章 空间计量经济学基础

本章首先介绍空间计量经济学与传统计量经济学的一个重要区别,即空间效应;然后,介绍空间计量经济模型中的空间权重;最后,介绍空间计量经济学基本软件——GeoDa。

1.1 空间效应

空间效应是空间计量经济学的基本特征,它反映着空间因素的影响,是空间计量经济学从传统计量经济领域独立出来的根本原因。空间效应可以分为空间相关性(spatial dependence)和空间异质性(spatial heterogeneity)(Anselin, 1988a)。空间相关性是描述经济变量存在相关性的一种方法,而这一相关性是体现在空间结构上的。当然,空间相关性并不是局限在地理意义上的相关性。例如,该空间结构可以是人与与人之间的关系,也可以是不同政策的辐射能力等。简单地说,空间异质性描述的是不同经济个体之间存在的差异性,是以模型函数形式或参数表现出来的结构特征(Anselin, 1988a, 1988b),并且强调这一差异是由空间分布或者空间结构特点导致的。因为空间异质性可以用传统的计量经济学方法进行处理,例如处理异方差性的方法,所以在本章只关注空间相关性。

1.1.1 空间相关性

Goodchild(1992)指出,几乎所有的空间数据都具有空间依赖(或者称空间自相关)特征,也就是说一个地区空间单元的某种经济地理现象或者某一属性值与邻近地区空间单元上同一现象或属性值是相关的。空间依赖性打破了大多数传统经典统计学和计量经济学中相互独立的基本假设,是对传统方法的继承和发展。

空间依赖是事物和现象在空间上的相互依赖、相互制约、相互影响和相互作用,是事物和现象本身所固有的属性,是地理空间现象和空间过程的本质特征。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)。由于空间观测值间存在相关性,违反了经典统计学和计量经济学有关观测值不相关的假定前提,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传统计量经济分析有偏。可见,空间相关性检验是构建空间经济计量模型和进行空间经济计量分析的关键。

例 1.1.1 2012 年中国各省区国内生产总值为我们提供了一个具有空间特征的经济现象的典型案例。从我国国内生产总值的四分位图(如图 1.1.1)可以看出,我国国内生产总值较大的省份基本集中在东部沿海省份,而内陆省份特别是西部地区(除四川特殊以外)大多数国内生产总值较低。

Quantile: GDP

- [701:1.031e+004] (8)
- [1.141e+004:1.445e+004] (8)
- [1.588e+004:2.225e+004] (7)
- [2.387e+004:5.707e+004] (8)



图 1.1.1 2012 年我国国内生产总值四分位图

1.1.2 空间相关性与检验方法

在实际经济分析中,研究对象的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,空间计量经济学研究者已经提出了 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。

在空间数据分析中,不论采用何种空间计量经济模型,都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分成两类:第一,包括空间误差自相关或空间误差移动平均的误差相关检验,如 LMERR, R-LMERR;第二,空间滞后相关检验,如 LMLAG, R-LMLAG。此外,部分统计量既可以检验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、Geary 检验。迄今为止, Moran's I 检验是最常见的空间相关性检验方法,本节将重点介绍 Moran's I 和 Geary 检验等,至于 LMERR, R-LMERR, LMLAG, R-LMLAG 等检验方法,将结合空间计量的具体模型在第 2 章中介绍。

(1) 全局空间自相关指标

Moran's I 统计量和 Geary 统计量是两个用来度量空间自相关的全局指标。

① Moran's I 统计量

Moran's I 统计量反映的是空间邻接或空间邻近的区域单元属性值的相似程度。如果 Y 是位置(区域)的观察值,则该变量的全局 Moran's I 值用如下公式计算:

$$Moran's I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (1.1.1)$$

其中, $S^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$; $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, Y_i 表示第 i 地区的观测值; n 为地区总数; w_{ij} 为空间权值矩阵。标准化的 Moran's I 统计量为:

$$Z = \frac{Moran's I - E(I)}{\sqrt{VAR(I)}} \quad (1.1.2)$$

其中, $E(I) = -\frac{1}{n-1}$, $VAR(I) = \frac{n^2 w_1 + n w_2 + 3 w_0^2}{w_0^2 (n^2 - 1)} - E^2(I)$, $w_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, $w_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$, $w_2 = \sum_{i=1}^n \sum_{j=1}^n (w_{i.} + w_{.j})^2$, $w_{i.}$ 和 $w_{.j}$ 分别为空间权值矩阵中 i 行和 j 列之和。在不存在空间相关性的原假设下, Z 服从标准正态分布。

Moran's I 统计量的取值一般在 $[-1, 1]$ 之间, 小于 0 表示负相关, 等于 0 表示不相关, 大于 0 表示正相关。越接近 -1 表示单元间的差异越大或分布越不集中; 越接近 1, 则代表单元间的关系越密切, 性质越相似(高值集聚或者低值集聚); 接近 0, 则表示单元间不相关。

以 2012 年国内生产总值为例, 空间权重 W 取常用的一阶邻近矩阵时(1.2 节将详细介绍空间权重), 其 Moran's I 统计量为 0.266584, 呈现较强的空间正相关性。

② Geary 统计量

由于 Moran 统计量不能判断空间数据是高值集聚还是低值集聚, Getis 和 Ord 于 1992 年提出了全局 Geary 统计量。Geary 统计量与 Moran 统计量存在负相关关系。Geary 统计量 C 计算公式如下:

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (1.1.3)$$

式中, C 为 Geary 统计量; 其他变量同上式。Geary 统计量 C 的取值一般在 $[0, 2]$ 之间, 大于 1 表示负相关, 等于 1 表示不相关, 而小于 1 表示正相关。也可以对 Geary 统计量进行标准化:

$$Z(C) = (C - E(C)) / \sqrt{Var(C)} \quad (1.1.4)$$

式中, $E(C)$ 为数学期望, $Var(C)$ 为方差。正的 $Z(C)$ 表示存在高值集聚, 负的 $Z(C)$ 表示低值集聚。

(2) 局部空间自相关指标

局部空间自相关指标包括: 空间联系的局部指标(LISA 集聚图)、 G 统计量、Moran 散点图。

①空间联系的局部指标

LISA 包括局部 Moran 指数(local Moran)和局部 Geary 指数(local Geary)。

局部 Moran 指数被定义为

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_j w_{ij} (x_j - \bar{x}) \tag{1.1.5}$$

正的 I_i 表示该空间单元与邻近单元的属性相似(“高一高”或“低—低”), 负的 I_i 表示该空间单元与邻近单元的属性不相似(“高一低”或“低—高”)。

局部 Geary 指数由 Ord 和 Getis(1992) 提出, 是一种基于距离权重矩阵的局部空间自相关指标, 能探测出高值集聚和低值集聚, 计算公式为

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_k x_k} \tag{1.1.6}$$

在各区域不存在空间相关下, Getis 和 Ord 简化了 G_i^* 的数学期望和方差的表达式:

$$E(G_i^*) = \frac{\sum_j w_{ij}}{n-1} = \frac{W_i}{n-1}, \text{Var}(G_i^*) = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)} \frac{Y_{i2}}{Y_{i1}^2}$$

式中, $Y_{i1} = \frac{\sum_j w_j}{n-1}, Y_{i2} = \frac{\sum_j x_j^2}{n-1} - Y_{i1}^2$ 。

将 G_i^* 标准化, 得到

$$Z_i = \frac{G_i^* - E(G_i^*)}{\sqrt{\text{Var}(G_i^*)}} \tag{1.1.7}$$

此时, 显著的正 Z_i 表示邻近单元的观测值高, 显著的负 Z_i 则表示邻近单元的观测值低。

例 1.1.2 LISA 聚集地图用不同的颜色表示不同的空间自相关类型。以 2012 年国内生产总值为例, 空间权重 W 仍取常用的一阶邻近矩阵。2012 年我国地区国内生产总值 LISA 集聚图见图 1.1.2, 图中有颜色的地区国内生产总值空间关系显著, 而无颜色的

LIST Cluster Map: jiliangke1.I_GDP

- Not Significant (23)
- High-High (3)
- Low-Low (2)
- Low-High (1)
- High-Low (1)
- Neighborless (1)



图 1.1.2 2012 年我国地区国内生产总值 LISA 集聚图

地区国内生产总值集聚关系则不显著。如图中用4种颜色表示:深红色表示高一高,代表国内生产总值高的地区集聚在一起;深绿色表示低—低,表示国内生产总值低的地区集聚在一起;品红色表示高一低,表示本地区国内生产总值高,但是周边地区国内生产总值低;浅绿色表示低—高,表示本地区国内生产总值低,但是周边地区国内生产总值高。并且这四种类型对应于 Moran 散点图中的4个象限。

从图 1.1.2 可以看出山东和江苏呈现显著高一高集聚,这是因为这些地区除安徽外都处于沿海地区,经济都较为发达,存在空间上的相互集聚;新疆、甘肃呈现显著低—低集聚,因为这些地区与相邻地区的经济发展水平相对比较低;湖南呈现显著的低—高空间集聚状态,该地区本身的经济欠发达,但周围地区的经济较为发达,如湖南边上的浙江、江苏的经济较为发达;四川省呈现显著的高—低空间集聚,这是因为它相对于甘肃和青海而言,经济相对要发达,而其他地区的空间集聚关系不显著。

② Moran 散点图

以 (Wz, z) 为坐标点的 Moran 散点图,常用来研究局部的空间特征。它对空间滞后因子 Wz 和 z 数据进行了可视化的二维图示。Moran's I 统计量可看作各地区观测值的乘积和,其取值范围在 -1 到 1 之间,若各区间经济行为为空间正相关,其数值应当较大;负相关则较小。当目标区域数据在空间区位上相似的同时也有相似的属性值时,空间模式整体上就显示出正的空间自相关性;而当在空间上邻接的目标区域数据不同寻常地具有不相似的属性值时,就呈现为负的空间自相关性;零空间自相关性出现在当属性值的分布与区位数据的分布相互独立时。

Moran 散点图中的第 1、3 象限代表观测值的正空间相关性,第 2、4 象限代表观测值的负空间相关性,并且第 1 象限代表了观测值高的区域单元被高值区域所包围(HH);第 2 象限代表了观测值低的区域单元被高值区域所包围(LH);第 3 象限代表了观测值低的区域单元被低值区域所包围(LL);第 4 象限代表了观测值高的区域单元被低值区域所包围(HL)。

例 1.1.3 以 2012 年国内生产总值为例,空间权重 W 仍取常用的一阶邻近矩阵。从 Moran 散点图(如图 1.1.3)上我们可以发现绝大多数省份都处于第 1 或第 3 象限,说明我国国内生产总值呈现较强的空间正相关性。

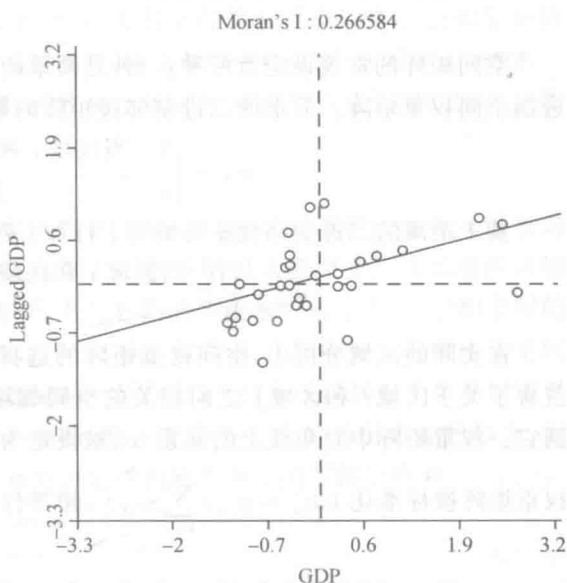


图 1.1.3 2012 年国内生产总值的 Moran 散点图

1.2 空间权重

以区域经济管理研究为例,将空间效应引入经济管理过程的研究,建立空间计量模型进行空间统计分析时,一般要用空间权重矩阵 W 来表达 n 个位置的空间区域的邻近关系,其形式如下:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \quad (1.2.1)$$

式中 w_{ij} 表示区域 i 与 j 的邻近关系。对于空间矩阵的构造,一直是一个有争议的问题。因为无法找到一个完全描述空间相关结构的矩阵,也就是说,理论上讲,不存在最优的空间矩阵。一般讲,空间矩阵的构造必须满足“空间相关性随着‘距离’的增加而减少”的原则。这里的“距离”是广义的,可以是地理上的距离,也可以是经济意义上合作关系的远近,甚至可以是社会学意义上的人际关系的亲疏。

1.2.1 空间矩阵的常规设定

空间矩阵的常规设定有两种:一种是简单的二进制邻接矩阵,另一种是基于距离的二进制空间权重矩阵。简单的二进制邻接矩阵的第 i 行第 j 列元素为

$$w_{ij} = \begin{cases} 1 & \text{当区域 } i \text{ 和区域 } j \text{ 相邻接} \\ 0 & \text{其他} \end{cases} \quad (1.2.2)$$

基于距离的二进制邻接矩阵的第 i 行第 j 列元素为

$$w_{ij} = \begin{cases} 1 & \text{当区域 } i \text{ 和区域 } j \text{ 的距离小于 } d \\ 0 & \text{其他} \end{cases} \quad (1.2.3)$$

在实际的区域分析中,空间权重矩阵的选择设定是外生的,原因是 $(n \times n)$ 阶矩阵 W 包含了关于区域 i 和区域 j 之间相关的空间邻接的外生信息,不需要通过模型来估计得到它。权重矩阵中对角线上的元素 w_{ii} 被设定为 0。为了减少或消除区域间的外在影响,权重矩阵被标准化 ($w_{ij}^* = w_{ij} / \sum_{k=1}^n w_{ik}$),使得行元素之和为 1。

1.2.2 基于邻近概念的空间权重矩阵

基于邻近概念的空间权重矩阵(contiguity based spatial weights)有一阶邻近矩阵和高阶邻近矩阵两种。

一阶邻近矩阵(the first order contiguity matrix)是假定两个地区有共同边界时空间关联才会发生,即当相邻地区 i 和 j 有共同边界时用 1 表示,否则以 0 表示。一般有

Rook 邻近和 Queen 邻近两种计算方法 (Anselin, 2003)。

Rook 邻近用仅有共同边界来定义邻居,而 Queen 邻近则除了共有边界邻区外还包括共同顶点的邻区。由此可见,基于 Queen 邻近的空间矩阵常常与周围地区具有更加紧密的关联结构(拥有更多的邻区)。当然,如果假定区域间公共边界的长度不同(如 10km 和 100km),其空间作用的强度也不一样,则还可以通过将公用边界的长度纳入权重计算过程中,使这种邻近指标更加准确一些。

空间权重矩阵不仅仅局限于第一阶邻近矩阵,也可以计算和使用更高阶的邻近矩阵。Anselin 和 Smirnov(1996)提出了高阶邻近矩阵的算法。二阶邻近矩阵(the second order contiguity matrix)表示了一种空间滞后的邻近矩阵。也就是说,该矩阵表达了邻近的相邻地区的空间信息。当使用时空数据并假设随着时间推移产生空间溢出效应时,这种类型的空间权重矩阵将非常有用,在这种情况下,特定地区的初始效应或随机冲击将不仅会影响其邻近地区,而且随着时间的推移还会影响其邻近地区的相邻地区。当然,这种影响是几何递减的。

可以看出,邻近空间权重矩阵因其对称和计算简单而最为常用,适合于测算地理空间效应的影响。

1.2.3 其他经济社会空间权重矩阵

除了使用真实的地理坐标计算地理距离外,还有包括经济和社会因素的更加复杂的权值矩阵设定方法。比如,根据区域间交通运输流、通讯量、GDP 总额、贸易流动、资本流动、人口迁移、劳动力流动等确定空间权值,计算各个地区任何两个变量之间的距离。例如:

(1) 基于万有引力定律的空间邻接矩阵

近年来,一些学者采用诺贝尔经济学奖获得者 Tinbergen(1962)提出的引力模型研究区域贸易问题,该模型直接把地区间的距离作为解释变量引入模型中。其思想源自物理学中的万有引力定律,即两个物体之间的引力与它们的质量乘积成正比,与它们之间的距离平方成反比。尽管引力模型已经得到了广泛应用,但是基于万有引力定律构造空间邻接矩阵并不多见。我们认为,技术溢出效应是广泛存在的,而不仅仅局限于有共同边界的地区之间,并且两个地区之间的经济实力越强,技术交流与合作的吸引力往往越大,相应的技术溢出效应也越大。为此,基于万有引力定律构建如下空间邻接矩阵 $\mathbf{W}=(w_{ij})$:

$$w_{ij} = \begin{cases} \frac{m_i m_j}{r_{ij}^2}, & i \neq j \\ 0, & i = j \end{cases} \quad (1.2.4)$$

其中, r_{ij} 为地区 i 与地区 j 的地理距离,可由两个地区的经纬度计算; m_i 为地区 i 的经济实力,以样本期内的人均实际 GDP 衡量。为了消除单位选取的影响,邻接矩阵需要标准化使行元素之和为 1。

(2) 基于地理距离标准构造空间权重矩阵

空间邻接标准认为空间单元之间的联系仅仅取决于二者相邻与否,即只要不同空间

单元相邻,则认为它们之间具有相同的影响强度,这在区域创新经济研究中是不符合客观事实的。例如,用空间邻接标准衡量的区域的地理位置,与北京邻接的只有天津、河北两省市,但我们不能认为北京只与津、冀地区发生联系而与其他省区均没有联系,也不能认为北京和在地理区位上与之相近的山东省之间的相互影响和北京与新疆、西藏等相对较远的省份之间的相互影响是等同的(而在邻接权重矩阵中北京、山东和北京、新疆之间的权重都为0)。基于这样的事实,我们通过地理距离标准构造空间权重矩阵,其实,这也符合地理学第一定律(Tobler W.R,1970):任何事物与其他周围事物之间均存在联系,而距离较近的事物总比距离较远的事物联系更为紧密。选用一种常用的空间距离权重矩阵 W_d (Tiiupaas 和 Friso Schlitte,2006)为

$$w_{ij} = \begin{cases} \frac{1}{d^2}, i \neq j \\ 0, i = j \end{cases} \quad (1.2.5)$$

其中 d 为两地区地理中心位置之间的距离。

(3) 社会经济特征空间权重矩阵

以地理区位差异反映出的区域创新的空间联系及其强度仅仅表征了地理邻近特征的影响,是相对粗糙的,区域创新作为一项系统活动,必然受到其他多种非地理邻近因素的综合影响,因此需要从不同角度建构其他类型的空间权重矩阵,以全面客观地揭示区域创新生产的空间影响因素。将区域间的社会经济特征分为经济基础和人力资本两类,分别建立空间权重矩阵。

鉴于不同省区经济水平存在空间相关性的客观现实,许多学者通过建立经济距离空间权重矩阵来对这种关系予以描述,比较有代表性的如林光平等(2006)基于相邻地区间经济发展水平的差异程度越小,其经济上的相互联系强度就越大的假设,建立了基于地区差异的经济距离空间权重矩阵。然而,这一形式的矩阵存在明显不足。该矩阵中各元素所表征的两个空间单元之间的相互影响强度是相同的($w_{ij} = w_{ji}$),而现实情况是经济发展水平较高的地区对经济水平较低地区产生更强的空间影响与辐射作用,如北京对于河北的影响强度显然大于河北对北京的影响强度。由此,我们建立新的经济距离空间权重矩阵,具体为:

$$W = W_d \text{diag}(\bar{Y}_1/\bar{Y}, \bar{Y}_2/\bar{Y}, \dots, \bar{Y}_n/\bar{Y}) \quad (1.2.6)$$

其中 W_d 为地理距离空间权重矩阵, $\bar{Y}_i = \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} Y_{it}$ 为考察期内第 i 省物质资本存量平均值, $\bar{Y} = \frac{1}{n(t_1 - t_0 + 1)} \sum_{i=1}^n \sum_{t=t_0}^{t_1} Y_{it}$ 为考察期内总物质资本存量均值。通过上述矩阵可以发现,当一个地区的物质资本存量占总量比重较大(即 $\bar{Y}_i/\bar{Y} > \bar{Y}_j/\bar{Y}$)时,其对周边地区的影响也越大(即 $w_{ij} > w_{ji}$)。可用省会城市间的地理距离建立 W_d ,用地区物质资本存量表征地区经济发展水平。

人力资本对区域创新活动具有重要的影响。人力资本水平的提高可以增强对知识、技术以及其他信息的获取与运用能力,进而转化为创新产出,促进区域社会经济发展。以卢卡斯为代表的新增长理论认为,地区人力资本对经济发展具有决定性作用(Lucas,