



格致方法·定量研究系列 吴晓刚 主编

空间回归模型

[美] 迈克尔·D.沃德 (Michael D. Ward) 著
克里斯蒂安·格里蒂奇 (Kristian Skrede Gleditsch) 著
宋曦译

- ★ 革新研究理念
- ★ 丰富研究工具
- ★ 最权威、最前沿的定量研究方法指南

格致出版社  上海人民出版社

11

格致方法·定量研究系列 吴晓刚 主编

空间回归模型

[美] 迈克尔·D.沃德(Michael D. Ward) 著
克里斯蒂安·格里蒂奇(Kristian Skrede Gleditsch) 著
宋 曦 译

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

空间回归模型/(美)沃德(Ward, M. D.), (美)格里蒂奇(Gleditsch, K. S.)著;宋曦译. —上海:格致出版社;上海人民出版社, 2012
(格致方法·定量研究系列)
ISBN 978-7-5432-2122-2

I. ①空… II. ①沃… ②格… ③宋… III. ①回归分析-研究 IV. ①0212.1

中国版本图书馆 CIP 数据核字(2012)第 129115 号

责任编辑 王亚丽

格致方法·定量研究系列
空间回归模型

[美] 迈克尔·D. 沃德
克里斯蒂安·格里蒂奇 著
宋曦 译

出版 世纪出版集团 格致出版社
www.ewen.cc www.hibooks.cn
上海人民出版社
(200001 上海福建中路193号24层)



编辑部热线 021-63914988
市场部热线 021-63914081

发行 世纪出版集团发行中心
印刷 浙江临安曙光印务有限公司
开本 920×1168 毫米 1/32
印张 5
字数 96,000
版次 2012 年 7 月第 1 版
印次 2012 年 7 月第 1 次印刷
ISBN 978-7-5432-2122-2/C·74
定价 15.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书,精选了世界著名的 SAGE 出版社定量社会科学研究丛书中的 35 种,翻译成中文,集结成八册,于 2011 年出版。这八册书分别是:《线性回归分析基础》、《高级回归分析》、《广义线性模型》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》、《数据分析方法五种》和《列表数据分析》。这套丛书自出版以来,受到广大读者特别是年轻一代社会科学工作者的欢迎,他们针对丛书的内容和翻译都提出了很多中肯的建议。我们对此表示衷心的感谢。

基于读者的热烈反馈,同时也为了向广大读者提供更多的方便和选择,我们将该丛书以单行本的形式再次出版发行。在此过程中,主编和译者对已出版的书做了必要的修订和校正,还新增加了两个品种。此外,曾东林、许多多、范新光、李忠路协助主编参加了校订。今后我们将继续与 SAGE 出版社合作,陆续推出新的品种。我们希望本丛书单行本的出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总 序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研

究方法的可以在第二年修读另外一个三学期的系列课程,其中头两门课叫“调查数据分析”,第三门叫“研究设计”。除此以外,还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者,提供不同的、更加专业化的课程供学生选修。2001年完成博士学业之后,我又受安德鲁·梅隆基金会资助,在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究,其间旁听谢宇教授为博士生讲授的统计课程,并参与该校社会研究院(Institute for Social Research)定量社会研究方法项目的一些讨论会,受益良多。

2003年,我赴港工作,在香港科技大学社会科学部,教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来,非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课,而第二门课“社会科学中的定量分析”为博士生的必修课(事实上,大部分硕士生修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候,根据社会科学研究生的数理基础比较薄弱的特点,尽量避免复杂的数学公式推导,而用具体的例子,结合语言和图形,帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题,即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”,学完这些课程后,我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章;另一方面,也能在自己的研究中运用这些成熟的

方法论技术。

上述两门课的内容,尽管在线性回归模型的内容上有少量重复,但各有侧重。“社会科学里的统计学”(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始,到多元线性回归模型结束,内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型,以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法,将重点放在因变量为定类数据的分析模型上,包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型,包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生,一直鼓励和支持我将两门课的讲稿结集出版,并帮助我将原来的英文课程讲稿译成了中文。但是,由于种种原因,这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩,每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书,当格致出版社向我提出从这套丛书中精选一批翻译,以飨中文读者时,我非常支持这个想法,因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种

语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程,他们目前大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是:

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾“中央”研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景,书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异,校对也未免挂一漏万,术语的标准译法方面还有很大的改进空间,但所有的参与者都做了最大的努力,在繁忙的学习和研究之余,在不到一年的时间内,完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外,还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林,协助我通读了全稿,在此

我也致以诚挚的谢意。有些作者,如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授,也参与了审校工作。

我们希望本丛书的出版,能为建设国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

序

社会科学定量研究的很多方法都是基于对个体行为的分析。我们通常将个体行为作为结果变量,然后将其表示成由一组解释变量构成的方程,最常见的就是回归方程的形式。社会科学理论主要就是描述这种结果变量和解释变量之间是如何关联的。当然,这种分析不仅停留在个体或者微观层面。有时,研究者也会分析集体层面的数据,比如街道、社区、公司、城市、县、州以及国家。但是分析的逻辑仍然相同。我们试图构造因变量和某些自变量(诸如个人、街坊、社区、公司、城市、县、州和国家)之间的关联,不管这种关联是否为因果关系。当我们这样做的时候,其实已经暗示了观测值的地理或者空间位置并不起作用。尽管研究者也会经常用虚拟变量将处于同一个地区的观测值合成一组,但这样做并没有考虑除了空间以外的其他相似性。比如,数据分析者常用虚拟变量来划分个体是来自美国南部还是其他地区。这样做是为了控制一些特有的文化特征,而不是出于对回归中空间依赖关系的考虑。《空间回归模型》一书将不遗余力地解决线性回归分析中空间依赖关系的相关问题。

确切地说,我们将结果变量 y_i 对解释变量的向量 x_i 做 OLS 回归:

$$y_i = x_i\beta + \epsilon_i$$

这里 β 包括被估参数的一组向量, ϵ_i 是服从独立同分布假设的随机误差。在经典线性回归中,假设分布为正态分布。当存在空间(或其他)依赖关系时, ϵ_i 将不再相互独立,并且导致 β 被低估,从而影响假设检验的正确性。

尽管对空间问题的考虑可以追溯到早期在地图制作和调查方面的尝试,但是现代空间回归是直到最近几十年随着统计知识和计算能力的提高才出现的。本书作者向读者介绍了两种应用最广泛的空间回归模型:空间性定距因变量和空间性误差模型,此外还补充了空间分析中的疑难问题。尽管读者的分析单位可能与书中作者的分析单位并不相同,然而书中大量直观的例子仍然能够为读者提供有益的思考。

廖福挺

前言

空间观念能够为社会科学研究作出巨大贡献。本书试图填补这一领域的空缺,为社会科学家完备地介绍如何将空间依赖性的分析纳入回归框架,让更多对社会科学中空间应用问题感兴趣的读者能够读懂此书。尽管当前已经有非常详尽的空间统计学调查,但是它们中的大部分都过于深奥,并且都假设读者已经具有丰富的高级统计和概率论知识 (Banerjee、Carlin & Gelfand, 2004; Cressie, 1993; Getis & Boots, 1978; Haining, 2003; Ripley, 1981、1988; Schabenberger & Gotway, 2005)。此外,这些调查中的大部分都是自然科学的主题或者该方面的应用,而并不为社会科学研究者所熟知。我们假定此书读者仅仅了解社会科学研究中广泛应用的经典回归模型,同时他们对数据中可能存在的空间依赖性问题感兴趣。在某些部分,虽然我们会用到矩阵的表示形式,但同时也会用非数学语言对其进行详细解释。我们会用到免费且应用广泛的 R 计算平台 (R Development Core Team, 2004) 来演示如何使用这些方法,同时还提供了一段编码,对例子进行解释。如果读者对 R 的了解达到了达尔加

德一书(Dalgaard, 2002)所介绍的程度,将非常有助于理解本书的内容。尽管其他的程序和方法也可用于分析空间数据,但是本书中将不予以介绍,不过我们会以附录的形式提供一些简要的细节内容。

本书的面世离不开外界的帮助。首先,我们要感谢家庭成员对于我们紧张工作的体谅。同时我们要感谢西班牙巴塞罗那经济分析研究所主任琼·埃斯特班(Joan Esteban),她给予我们热情慷慨的欢迎和巨大的支持,帮助我们完成了拖延很久的第一稿。迈克尔·沃德还要感谢统计与社会科学中心主任埃德里安·拉夫特里在他访问期间提供的帮助。同时,沃德还要感谢华盛顿大学艺术与科学学院前院长和现任迈阿密大学校长大卫·霍奇以及华盛顿大学政治科学系主任史蒂夫·马耶斯基的帮助。克里斯蒂安·格里蒂奇得到了来自国家科学基金会的赞助以及来自加泰罗尼亚政府和加州大学的加斯帕·波多拉的旅行资助。

华盛顿大学、加州大学圣地亚哥分校、埃塞克斯大学等地的前任或现任同仁们引导并激发了我们在依赖性数据上的兴趣。我们同样要感谢约翰·阿尔齐斯特、克里斯·巴基、凯尔·比尔兹利、纳撒尼尔·贝克、罗杰·比万德、曹汛、肖娜·费希尔、詹姆士·勒萨热、林泽民、迈克尔·曼格尔、阿西姆·普拉卡什、安德里亚·鲁杰里、伊德里·萨里希安、迈克尔·锡恩、克里斯多夫·沃德、丹尼尔·沃德、安东·韦斯特维尔德三世和艾里克·韦伯尔斯给予的深刻见解和有益的讨论。迈克尔·D.沃德要感谢他以前在科罗拉多大学行为科学研究所的两位邻座安德鲁·科比和约翰·奥洛林对他在地理方面兴趣的影响。克里斯蒂安·格里蒂奇感谢

基德龙的《国家的世界地图集》(*State of the World Atlas*)一书对他在地理和社会科学方面兴趣的影响。

在 20 世纪,迈克尔·刘易斯—贝克鼓励我们从事这个项目。他是一个有耐心的人。

迈克尔·D.沃德、克里斯蒂安·格里蒂奇

目 录

序	1
前言	1
第 1 章 导论	1
第 1 节 交互作用与社会科学	2
第 2 节 世界各国的民主	6
第 3 节 空间依赖关系介绍	11
第 4 节 将地图作为可视化数据	15
第 5 节 空间依赖性和相关性测量	19
第 6 节 接近性测量	25
第 7 节 估计空间模型	36
第 8 节 小结	43
第 2 章 空间滞后因变量	45
第 1 节 空间滞后因变量的回归	47
第 2 节 估计空间滞后 y 模型	54
第 3 节 空间性间隔 y 模型的最大似然估计:以民主研究为例	57
第 4 节 空间滞后 y 模型的均衡效应	59
第 5 节 意大利投票率的空间依赖关系	66

第 3 章	空间误差模型	83
第 1 节	空间误差模型	85
第 2 节	空间误差模型的最大似然估计	88
第 3 节	以民主和发展研究为例	90
第 4 节	空间滞后 y 和空间误差的比较	93
第 5 节	估计成对贸易往来中的空间性误差	95
第 6 节	小结	103
第 4 章	扩展	105
第 1 节	识别连接性	107
第 2 节	推论与模型评估	113
第 3 节	小结	119
附录		121
注释		125
参考文献		127
译名对照表		135

第 **1** 章

导 论