



格致方法·定量研究系列 吴晓刚 主编

非递归因果模型

[美] 威廉·D.贝里 (William D. Berry) 著
洪岩璧 陈陈 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

21

格致方法·定量研究系列 吴晓刚 主编

非递归因果模型

[美] 威廉·D.贝里(William D.Berry) 著
洪岩璧 陈 陈 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

非递归因果模型/(美)贝里(Berry, W. D.)著;
洪岩璧,陈陈译. —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 978-7-5432-2118-5

I. ①非… II. ①贝… ②洪… ③陈… III. ①递归论-研究 IV. ①0141.3

中国版本图书馆 CIP 数据核字(2012)第 129224 号

责任编辑 顾悦

格致方法·定量研究系列

非递归因果模型

[美]威廉·D. 贝里 著

洪岩璧 陈陈 译

出

版

世纪出版集团
www.ewen.cc

格致出版社

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 4.5
字 数 84,000
版 次 2012 年 7 月第 1 版
印 次 2012 年 7 月第 1 次印刷
ISBN 978-7-5432-2118-5/C·70
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上,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

随着社会科学理论的日益复杂化,我们亟须反映此种复杂性的统计模型来补充先前最为简单的模型。这一趋势适用于回归模型,因为在回归模型中,研究已经逐渐从单方程模型转向多方程模型。有时候,当因果效应被认为是单向的,并允许使用所谓的递归模型时,多方程模型也就相对简单。但大多数时候,单向因果作用的假设是不现实的,因此,研究者转向了非递归模型。

随着上述发展的出现,对于全面而且能深入浅出地介绍和讨论这些更复杂的技术的需求就显得尤为迫切。更重要的是,这类著作应当包含对下列过程的一种真正的理解:模型的使用、模型所依据的假设、模型结果的解释等等。威廉·D. 贝里的《非递归因果模型》无疑正是这样的佳作。

贝里假定读者基本熟悉递归模型,比如已经读过早先出版的赫伯特·阿什(Herbert Asher)的《因果模型》(*Causal Modeling*)一书。为了唤起读者的记忆,同时搭建对本书内容理解的合适桥梁,贝里在第1章中简要回顾了递归模型的基本假设,然后才转入讨论非递归模型如何被应用于估计更

为复杂的方程系统。

本书的其余部分大都聚焦于“辨识”问题。因为辨识方程系统是研究工作中最困难的部分,故而努力把这一过程表述清楚成为本书重要的组成部分。这本介绍性的著作还有一个优点:贝里在解释书中材料时,并不要求读者具备线性代数知识。

在最后一章中,贝里回顾了那些已经被确认可用于估计非递归模型参数的通用技术。纵观全书,作者应用了来自经济学、政治科学和社会学的一系列例子。

理查德·G. 尼米

Nonrecursive Causal Models

Copyright © 1984 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

This simplified Chinese edition for the People's Republic of China is published by arrangement with SAGE Publications, Inc. © SAGE Publications, Inc. & TRUTH & WISDOM PRESS 2012.

本书版权归 SAGE Publications 所有。由 SAGE Publications 授权翻译出版。
上海市版权局著作权合同登记号:图字 09-2009-549

目 录

序	1
第 1 章 导论	1
第 1 节 概念定义和符号标记法	6
第 2 节 递归因果模型:简要回顾	8
第 3 节 非递归模型的前提假设	13
第 2 章 可辨识性问题	17
第 1 节 一个供给与需求的例子	20
第 2 节 可辨识性与不可辨识性	26
第 3 节 适度可辨识性和过度可辨识性	30
第 4 节 为什么某些非递归模型无法被辨识?	32
第 3 章 可辨识性的检验	45
第 1 节 次序条件	47
第 2 节 秩条件	49
第 3 节 次序条件的不充分性:一点说明	62

第 4 章	修改不可辨识模型	65
第 5 章	估计方法	71
第 1 节	间接最小二乘法和适度辨识的方程	75
第 2 节	二阶段最小二乘法	79
第 3 节	二阶段最小二乘法和多元共线性问题	85
第 4 节	职业和教育期望模型中的参数估计	88
第 5 节	抗议事件暴力行为修正模型	93
第 6 章	结论	97
第 1 节	其他方程形式	99
第 2 节	滞后内生变量	100
第 3 节	其他类型的可辨识性限定	101
第 4 节	其他的估计方法	102
第 5 节	不可观测变量	103
附录		105
注释		115
参考文献		119
译名对照表		122

第 **1** 章

导 论

社会科学经验研究中,最常用的策略是建立一个单方程模型,然后利用一个样本数据来估计方程中的系数。在这个单方程模型中,一个变量被定义为因变量,它被假定受一个或多个解释变量(自变量)的影响。比如,某人可以建立一个有关选举中投票行为的单方程模型,它认为个人在某次特定的选举中对互为对手的候选人的比较性评价是因变量,该因变量由两个解释变量:某人的政策或议题立场(X_1)和某人的政党认同感(X_2),再加上一个误差项决定。^[1]

$$\text{对候选人评价} = a + b_1 X_1 + b_2 X_2 + e \quad [1.1]$$

接下来,如果我们从一个样本中获得了这三个变量的一组数据,而且愿意为进行这个模型所必要的回归分析作出一组特定假设(参见 Lewis-Beck, 1980),那么我们就可以用一般最小二乘法(OLS)回归分析来估计方程的系数 a 、 b_1 和 b_2 。这个对于解释变量的系数估计值可以被解释为当其他解释变量的系数估计值为常量的情况下,该变量对因变量(候选人评价)的直接效应。比如,假定政党认同保持不变,那么 b_1 的估计值就代表了政策或议题立场对候选人评价的直接作用。

但是,很多社会科学的理论表明,变量之间的因果关系

过于复杂,几乎不可能借助单方程模型来反映。^[2]比如,假定某模型中的因变量受多个解释变量影响,模型背后的理论可能认为其中一些解释变量是另一些解释变量的原因。例如在投票行为中,我们可以修改单方程模型的方程 1.1,认为个人的政策/议题立场变量很可能影响其政党认同。对这两个解释变量的预期也许基于这样一个信念:隶属于某一政党的人们对政党议题的看法和基本的意识形态立场受到其所认同的政党领导人和候选人的引导,至少部分如此。把这个修改加到先前的模型中去,就产生了图 1.1 所示的因果关系模型。^[3]这一模型认为,政党认同和政策/议题立场是决定对候选人评价的两个因素,但同时政策/议题立场又直接影响个体的政党认同。就此而言,政策/议题立场不仅直接(如先前的模型方程 1.1 所示)而且间接(通过它对政党认同的作用)影响对候选人的评价。图 1.1 所示的关于候选人选择的理论无法仅仅用一个方程进行模型化,而是需要一个多方程模型。

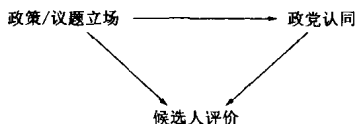


图 1.1 表示变量之间因果关系的“因果模式”：以投票行为为例

的确,多方程(或者说因果)模型被频繁应用于社会科学研究中,并且硕果累累。很多社会科学家想必对多方程模型的一个子类,即常用的递归因果模型非常熟悉。在过去的 20 多年中,该模型被社会学家和政治科学家频繁使用(参见 Land, 1969; Duncan et al., 1971)。所谓递归,就是一个模