



格致方法·定量研究系列 吴晓刚 主编

# 用面板数据做因果分析

[美] 史蒂芬·E.芬克尔 (Steven E.Finkel) 著  
李丁译

- ★ 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

8

格致方法·定量研究系列 吴晓刚

# 用面板数据做因果分析

[美] 史蒂芬·E.芬克尔(Steven E.Finkel) 著  
李 丁 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

用面板数据做因果分析/(美)芬克尔  
(Finkel, S. E.)著;李丁译. —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 978-7-5432-2116-1

I. ①用… II. ①芬… ②李… III. ①经济统计-统计  
数据-经济计量分析 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2012)第 130179 号

责任编辑 顾悦

---

格致方法·定量研究系列  
用面板数据做因果分析  
[美]史蒂芬·E. 芬克尔 著  
李丁 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 4.75  
字数 92,000  
版次 2012年7月第1版  
印次 2012年7月第1次印刷  
ISBN 978-7-5432-2116-1/C·68  
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,他们陆

---

## 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,以给中文读者时,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

---

面板数据(或译“跟踪调查数据”)是指在不同时点对相同样本个体进行重复观察所得的数据。舆情调查就是一个典型的例子,在这种调查中,被抽中的个体将被调查访问两次(比如,年份 1 和年份 2),每一次都询问相同的问题。与常规的截面设计相比,跟踪调查设计有很多值得称赞的地方,特别是它允许更强的因果推论,因为它将因果过程的时间维度明确考虑在内。来看一个假设的例子,假定政治学教授 Alice Green 研究了某次法国选民调查(在时间  $t$  上进行的)中的左—右派意识形态认同( $I$ )与投票意向( $V$ )之间的关系。通过使用这一截面数据,她将  $V_t$  对  $I_t$  做回归(再加上控制变量,统一用  $Z_t$  表示)后,发现意识形态认同与投票意向高度相关:

$$V_t = a + bI_t + cZ_t + e_t \quad [1]$$

不过,Green 教授怀疑这种关系在理论上是错误的,至少部分是虚假的。她认为,首先,可能因为“习惯的力量”,上一次投票意向  $V_{t-1}$  会影响当前的投票意向  $V_t$ 。另外,以往的意

意识形态认同  $I_{t-1}$  会影响当下的意识形态认同  $I_t$ 。并且,她认为意识形态认同  $I_t$  本身就受到投票意向  $V_t$  的影响。总之,她认为模式如下:

$$V_t = a + bV_{t-1} + cI_t + dZ_t + e_t \quad [2]$$

$$I_t = a' + b'I_{t-1} + cV_t + c'Z_t + u_t \quad [3]$$

方程 2 和方程 3 所刻画模型与方程 1 截然不同。可惜,Green 教授凭借一个时点  $t$  的截面样本是无法估计出第二个模型的。不过,如果该调查是某个已经在时间  $t-1$  上完成了第一轮调查的跟踪调查的第二轮,那她应该能估计出第二个模型(假定她还采取了必要的步骤以确保模型可识别)。这就是面板研究设计所承诺的理论功效及统计力量。

在本书中,Finkel 博士阐明了面板数据带来的因果推论机会,以及如何克服各种分析障碍的方法。对于后者,他举例指出了对像方程 2 和方程 3 这样的模型进行一般最小二乘回归估计的不足。由于互为因果问题,两阶段最小二乘法或 LISREL 程序成为必要,对这两种方法,Finkel 在书中都进行了认真的介绍。另外,与自相关有关的某些假定必须满足,以使模型得到识别。除了这些问题外,测量误差对面板数据的影响尤为严重。正如 Finkel 所证明的,当这一因素被考虑时,(模型的)结论可能大不相同。为了演示如何处理测量误差,他从简单的单指标、二期模型一直讲到更复杂的模型,如三指标、三期模型。

此外还有一点也很重要,本书让读者明白,只有满足了一些假定,才能进行有意义的参数估计。另外,为了方便研究实践,书中提供了大量的真实数据应用案例。其中,有两

个数据集用得恰到好处,一个是 1987 年至 1989 年前联邦德国抗议行为的跟踪调查,另一个是 1980 年美国大选跟踪调查。总之,在假定读者已掌握回归分析及 LISREL 用法基础上,Finkel 博士为我们提供了清晰易读而又十分先进的面板分析技术。

迈克尔·S.刘易斯-贝克

# 目 录

---

序	1
第 1 章 导论	1
第 2 章 用面板数据对变化进行建模	7
第 1 节 变化得分模型与滞后内生变量的角色	10
第 2 节 稳定得分模型的估计	19
第 3 节 其他滞后设定	22
第 4 节 面板模型估计中的问题	33
第 3 章 交互因果模型	35
第 1 节 交叉滞后作用模型	38
第 2 节 同步作用模型	47
第 3 节 交叉滞后及同步作用模型	53
第 4 章 测量误差模型	63
第 1 节 基本概念	65
第 2 节 单指标模型	70
第 3 节 多指标模型	80

<b>第 5 章 虚假相关及自相关的扰动项</b>	93
第 1 节 公因子模型	97
第 2 节 不可测量变量模型	101
<b>第 6 章 关于在面板分析中进行因果推论的结语</b>	115
<b>附录</b>	119
<b>注释</b>	122
<b>参考文献</b>	125
<b>译名对照表</b>	130

第 **1** 章

导 论

面板数据在社会科学中通常被用来对个体变化及社会变迁理论进行检验,这种数据由来自相同个体或单位在几个时点上的信息构成。和静止的截面分析相比,面板分析最重要的特征在于,它将变化本身明确地包含在设计中,从而使个体(或其他层次的单位)在一系列变量上的变化得到直接测量。面板数据应该和另外两种形式的纵贯数据区别开(Menard, 1991):“重复截面”或“趋势”数据,这种数据由在不同的时点上向多个不同单位收集的相同变量信息组成;另外一种是“时间序列”数据,这种数据所含的观察结果来自同一个个体在不同时点上的多个变量(Ostrom, 1978)。面板数据的不同之处在于,它包含了多个观察单位在相同变量上的多次重复测量数据。<sup>[1]</sup>

本书对适用于面板数据分析的各种模型进行了概览,特别关注了跟踪调查优于截面研究设计的主要领域:变量间因果关系的分析。众所周知,变量 $X$ 与 $Y$ 之间若要存在因果关系,则必须满足如下条件(Menard, 1991:17):(1) $X$ 和 $Y$ 必须共变,在非实验研究中表现为两变量间的相关系数不为0;(2)从时间上看, $X$ 必须在 $Y$ 之前;(3)这种相关不是“虚假”的,或者说,并不是由 $X$ 和 $Y$ 与第三个变量或其他一系列变



# 原书缺页