



格致方法·定量研究系列

吴晓刚 主编

固定效应回归模型

[美] 保尔·D. 埃里森 (Paul D. Allison) 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

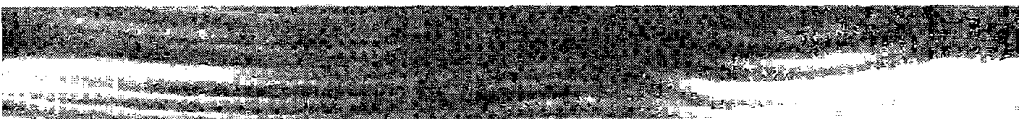
格致出版社  上海人民出版社

7

格致方法·定量研究系列 吴晓刚 主编

固定效应回归模型

[美] 保尔·D.埃里森(Paul D.Allison) 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

固定效应回归模型/(美)埃里森(Allison, P. D.)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 978-7-5432-2115-4

I. ①固… II. ①埃… ②李… 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 127658 号

责任编辑 王亚丽

格致方法·定量研究系列

固定效应回归模型

[美]保尔·D.埃里森 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)



格致出版

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 6
字数 116,000
版次 2012年7月第1版
印次 2012年7月第1次印刷
ISBN 978-7-5432-2115-4/C·67
定价 18.00 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

在最近一次会议上,我听了某研究者分析国家一年度(country-year)数据的报告,理应使用固定效应模型的,他用的却是随机效应模型。而那篇文章却受到了来自不同社会科学背景的学者的热烈欢迎。显然,在诸多社会科学专业里,就如何选用固定效应模型和随机效应模型还存在很多疑惑,很多人甚至还不清楚这些模型有何用处。无疑,埃里森讨论的是这两种模型更重要和一般的方面。本书将很好地满足“社会科学定量研究方法丛书”在这一主题上的需要,尤其是考虑到现在获得跟踪调查数据(panel data)越来越容易的现实^①。

① Panel Data 在经济学文献中通常被翻译为面板数据(如《面板数据计量经济学》[*Panel Data Econometrics*],曼纽尔·阿雷拉诺著,朱平芳、徐伟民译,上海财经大学出版社,2008年10月)或综列数据(如《计量经济学:现代观点》,[美]J. M. 伍德里奇著,费剑平译,中国人民大学出版社,2003年3月)。面板数据分析(Panel Data Analysis)与时间序列分析(Time Series Analysis)及横截面数据回归分析(Regression Analysis with Cross-Sectional Data)构成计量经济学的三大内容。其中,横截面数据是一个时点上收集的不同观察对象的数据,比方说,就一次人口普查来说就是一个截面研究;时间序列数据通常是一个观察单位在不同时点的观察结果构成的数据,如我国1978年以来,每一年的GDP的增长速度数据就构成一个时间序列数据。而Panel Data将这两种数据的特性综合在一起,首先在同一时点上对不同的案例(通常为总体中的(转下页)

上述国家一年度数据代表着这样一种数据类型,在这种数据中,个体案例得到了历时的(多次)观察。跟踪调查(panel survey)之所以近年来非常流行,一个重要的原因是跟踪数据允许研究者把握社会的发展变化,而把握这种变化是真正理解社会机制的必要条件。尽管有些跟踪调查每年都会观测一次,例如英国家户跟踪调查(British Household Panel Study Survey, BHPS)开始于1991年,目前仍在持续进行;其他一些则只有少数几轮调查,例如全美青少年健康跟踪调查(National Longitudinal Study of Adolescent Health in the United States)只在1994年到2002年间进行了三轮。

回归模型中,不管分析单位是个人、单位还是国家,每个案例在不同时点上的残差都将存在一定的相关或互相依赖,这通常是因为不同案例在某些未被观察到的特征上存在差异造成的。此时,回归模型有关误差项相互独立的假定被违背(尽管这个一般规律同样适应于限值因变量回归,但这里我们将讨论限定在线性模型上)。

固定效应模型和随机效应模型都能解决残差相关问题。但固定效应模型做得彻底得多。用埃里森的话说,这些模型“将每个个体作为其自身的控制”。经此处理,它们实际上就控制了所有稳定的、未被观测到的变量,就像这些变量实际得到了观测并被纳入模型一样。就此而言,这些模型所起的

(接上页)一个规模相对较小的子样本)的多个特征进行了观测;其次,对每一个案例在不同时点进行了多次观测;由此所得到的数据就是 Panel Data。面板数据这一翻译尽管因计量经济学而流传甚广,但“面板”的中文字义与英文 panel 含义相差甚远。Panel Survey 翻译成为面板调查显得笨拙,可翻译为小样本重复调查、固定样本长期追踪调查、追踪调查、纵贯调查或者历时调查等等。其中追踪调查或跟踪调查即带有在一段时间内对固定样本进行多次调查的含义。此外,根据艾尔·巴比的观点,Longitudinal Data 为历时研究,它包括趋势研究、队列研究和追踪研究三类。——译者注

作用和实验设计中的随机分配如出一辙。

本书作者在过去 30 年间为社会科学研究方法作出了持续的贡献,涉及诸多重要的主题。他撰写的《事件史分析》(*Event History Analysis*, 1984),至今仍是社会科学领域介绍事件史数据分析著作的榜样和标准。确切地说,在本书中,埃里森介绍了多种形式的固定效应模型——可以用于连续因变量的、分类因变量的、计数因变量的甚至结构方程情境等——并且讨论了如何在固定效应模型及随机效应模型之间作出选择,这一讨论对于本序言开始时提及的那位报告人将大有裨益。

廖福挺

Fixed Effects Regression Models

English language edition published by SAGE Publications of Oaks, London, New Delhi, Singapore and Washington D. C. , © 2008 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 09-2009-547

目 录

序	1
第 1 章 导论	1
第 2 章 线性固定效应模型：基本原理	9
第 1 节 两期数据(固定效应分析)	12
第 2 节 两期数据差分法的扩展	16
第 3 节 每个个体被观察三期及以上的一阶差分方法	19
第 4 节 每个个体被观察两期及以上的虚拟变量法	22
第 5 节 在固定效应法中设置与时间的交互作用	28
第 6 节 与随机效应模型的比较	30
第 7 节 混合(模型)法(A Hybrid Method)	34
第 8 节 总结	37
第 3 章 固定效应 logistic 回归	39
第 1 节 两期数据(固定效应分析)	42
第 2 节 三期及多期数据(固定效应分析)	46

第 3 节	与时间的交互作用	52
第 4 节	混合(模型)法	55
第 5 节	多分类反应变量的(固定效应)方法	58
第 6 节	总结	65
第 4 章	计数变量的固定效应模型	67
第 1 节	每个个体被观察两期的计数数据泊松模型	70
第 2 节	多期数据泊松模型	76
第 3 节	计数数据的固定效应负二项模型	83
第 4 节	混合(模型)法	88
第 5 节	总结	91
第 5 章	事件史数据的固定效应模型	93
第 1 节	Cox 回归	97
第 2 节	带固定效应的 Cox 回归	100
第 3 节	附加说明	106
第 4 节	Cox 回归混合模型法	108
第 5 节	非重复性事件的固定效应事件史法	109
第 6 节	总结	117
第 6 章	固定效应结构方程模型	119
第 1 节	随机效应作为潜变量的模型	121

第 2 节	固定效应作为潜变量的模型	126
第 3 节	固定效应和随机效应的折中	129
第 4 节	带滞后自变量的交互效应	131
第 5 节	总结	136

附录	137
----	-----

注释	159
----	-----

参考文献	162
------	-----

译名对照表	167
-------	-----

第 **1** 章

导 论