



格致方法·定量研究系列 吴晓刚 主编

合并时间序列分析

[美] 洛伊斯·塞耶斯 (Lois W. Sayrs) 著

温方琪 译 范新光 校

- ★ 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

55

格致方法 · 定量研究系列

合并时间序列分析

[美] 洛伊斯·塞耶斯(Lois W. Sayrs) 著
温方琪 译 范新光 校

SAGE Publications, Inc.

格致出版社  上海人民出版社

图书在版编目(CIP)数据

合并时间序列分析/(美)洛伊斯·塞耶斯著;温方琪译. —上海:格致出版社;上海人民出版社, 2016.12

(格致方法·定量研究系列)

ISBN 978-7-5432-1608-2

I. ①合… II. ①洛… ②温… III. ①时间序列分析
IV. ①0211.61

中国版本图书馆 CIP 数据核字(2016)第 303346 号

责任编辑 张苗凤

格致方法·定量研究系列

合并时间序列分析

[美]洛伊斯·塞耶斯 著
温方琪 译 范新光 校

出版 世纪出版股份有限公司 格致出版社
世纪出版集团 上海人民出版社
(200001 上海福建中路 193 号 www.ewen.co)



编辑部热线 021-63914988
市场部热线 021-63914081
www.hibooks.cn

发行 上海世纪出版股份有限公司发行中心

印刷 浙江临安曙光印务有限公司
开本 920×1168 1/32
印张 4.5
字数 74,000
版次 2016 年 12 月第 1 版
印次 2016 年 12 月第 1 次印刷

ISBN 978-7-5432-1608-2/C·163

定价:25.00 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,翻译成中文,起初集结成八册,于 2011 年出版。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热烈欢迎。为了给广大读者提供更多的方便和选择,该丛书经过修订和校正,于 2012 年以单行本的形式再次出版发行,共 37 本。我们衷心感谢广大读者的支持和建议。

随着与 SAGE 出版社合作的进一步深化,我们又从丛书中精选了三十多个品种,译成中文,以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续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

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”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,这两本书拖了多年还没有完成。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,与我的教学理念是相通的。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香港、台湾地区的二十几位

研究生参与了这项工程,他们当时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;哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业,大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映,翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此,当格致出版社和 SAGE 出版社决定在“格致方法·定量研究系列”丛书中推出另外一批新品种时,香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是,香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始,在上海(夏季)和广州南沙(冬季)联合举办《应用社会科学研究方法研修班》,至今已经成功举办三届。研修课程设计体现“化整为零、循序渐进、中文教学、学以致用”的方针,吸引了一大批有志于从事定量社会科学研究博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余,历经近两年的时间,完成了三十多本新书的翻译任务,使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是:东南大学社会学系副教授洪岩璧,香港科技大学社会科学部博士研究生贺光烨、李忠路、王佳、王彦蓉、许多多,硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林,原硕士研究生李兰,密歇根大学社会学系博士研究生王骁,纽约大学社会学系博士研究生温芳琪,牛津大学社会学系研究生周穆之,上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光烨、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之,以及香港科技大学社会科学部硕士研究生陈佳莹,上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版,并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动,我在此致以诚挚的谢意。当然,每本书因本身内容和译者的行文风格有所差异,校对未免挂一漏万,术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议,以便再版时修订。

我们希望本丛书的持续出版,能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾

序^{*}

什么是合并时间序列(pooled time series)? 正如字面上所表达的,时间序列(在一个分析单位下规律出现的具有时间性的观测值)由横截面数据(cross-sections)(在单独时间点上一个分析单位下的观测值)组成的一个数据集。这些分析单位可以是学校、健康组织、商业交易、城市、国家等。为什么需要进行“合并分析”呢? 其中一个原因在于,当下研究者可以获得越来越多的相关横截面数据与时间序列数据。另外一个原因在于,将时间序列数据与横截面数据合并可以显著地扩大样本量,这使之前显得棘手的分析问题变为可能。

举一个简单的例子。布鲁姆(Broom)教授希望使用一个包含 20 个美国城市的数据来解释财产犯罪率的变化情

* 本书最初由前任编辑理查德·尼米(Richard Niemi)与约翰·沙利文(John Sullivan)接收。

况。她提出下面的模型：

$$C = a + b_1U + b_2L + b_3R + e$$

其中 C = 城市财产犯罪率, U = 失业率, R = 区域位置, 所有的变量都包含 15 年中每一年的观测值。假设经典回归的假设都被满足, 那么布鲁姆可对以上等式进行 15 次最小二乘 (OLS) 估计 (每一年的横截面数据一次)。然后, 她可以再运用 20 次 OLS (每一个城市的时间序列一次)。或者, 假设所有的参数 (a 、 b_1 、 b_2 和 b_3) 都在时间与空间上恒定, 她便可以简单地将所有观测值合并进而仅仅只计算一个回归。这个简洁的步骤可以将样本量 N 增加至 300, 同时也可以在很大程度上提高估计的统计有效性。

这种对 OLS 的应用与塞耶斯 (Sayrs) 博士所命名的合并分析的恒定系数模型 (constant coefficients model) 是一致的。此处最大的困难在于恒定参数的假设难以被满足。假设较易满足的一个模型是最小二乘虚拟变量模型 (least squares dummy variable model) [有时它也被称作协方差模型 (covariance model)]。该模型允许截距随时间以及横截面变化。同样地, 这里的虚拟变量不具备实质性意义, 它们极大地减少了自由度以及与此对应的统计解释力。一个可能的替代是误差成分模型 (error components model) [塞耶斯教授也将其称为随机系数模型 (random coefficient model)]。这个模型明确地将横截面上的与时间序列上的

误差都考虑了进去。然而,在滞后因变量(lagged dependent variable)在等式右侧或者是等式嵌套在一个更大的联立方程组(simultaneous-equation system)的情况下,不可以使用这里所需要的加权最小二乘类型的估计(weighted least squares type-estimation)。此外,当严重的时间序列相关存在时,误差假设往往被极大地削弱。为了超越误差成分模型的局限性,塞耶斯教授提出了一个结构方程模型(structural equation model)。她对合并时间序列分析下的估计函数稳健性的评论总结全书。

迈克尔·S.刘易斯—贝克

目 录

序	1
第 1 章 导言	1
第 2 章 合并时间序列模型的理论推导	5
第 1 节 在应用中的合并	6
第 2 节 合并线性回归模型	9
第 3 节 四种合并模型	16
第 4 节 初步诊断与残差分析	19
第 3 章 恒定系数模型	23
第 1 节 估计恒定系数模型	26
第 2 节 纠正自相关	28
第 3 节 异方差性	30
第 4 节 恒定系数模型的局限性	34
第 4 章 LSDV 模型	37
第 1 节 异方差性与单位效应	39
第 2 节 估计 LSDV 模型	43

第 5 章	随机系数模型	49
第 1 节	估计随机系数模型:GLS 方法	52
第 2 节	GLS 模型的一个 ARMA 变异	57
第 3 节	GLS 模型的一个看似不相关回归版本	60
第 4 节	Swamy 随机系数模型	63
第 5 节	Hsiao 随机系数模型	67
第 6 节	转换模型	71
第 7 节	ARCH 模型	75
第 8 节	随机系数模型的总结	79
第 6 章	结构方程模型	81
第 1 节	两步估计	83
第 2 节	最大似然估计	90
第 3 节	LOGIT 与 PROBIT 设定	92
第 4 节	最大似然法的总结	97
第 7 章	稳健性检验:这些估计值有多好?	99
第 1 节	稳健性估计函数	101
第 2 节	异方差性与稳健性	105
第 8 章	合并时间序列分析的总结	113
注释		116
参考文献		119
译名对照表		122

第 **1** 章

导 言

本书讨论使用同时包含横截面与时间序列的数据的回归分析。一个时间序列是一组数列,就一组变量 X_t 与 X_{t+1} 而言,观测值之间的距离是恒定且固定的(Ostrom, 1978)。一个横截面则是一个分析单位在特定时间点上所存在的一组变量($X_1 \cdots X_n$)的全部观测值。当变量在一定时间跨度内在一定数量的不同横截面上被反复观测到时,我们可将这样产生的数据矩阵称为一个合并时间序列(pooled time series)。^[1]

归纳合并时间序列矩阵的特征有许多种方法,但标准的方法是首先归纳横截面上的特征,然后再归纳时间序列上的特征。矩阵的形式被设定,这对于一个横截面内部随时间发生的变化较不同横截面之间产生的变化而言是次要的。

将横截面与时间序列用这样的方式组合在一起的主要好处是,可以捕捉到不同单位在空间上的变化以及同一单位随时间产生的变化。因此,我们可以对结果以及产生

结果的过程进行描述、分析以及假设检验。当时间序列的长度被缩减且/或横截面的样本量较小时,合并数据是应用性研究中一个特别有用的方法。通常来说,单变量时间序列对于常规时间序列技术而言显得太短了。许多时间序列方法要求至少 30 个时间点,而其他的一些方法要求更多。^[2]

本书的主要目的是向已具备有关线性模型、回归方法和单变量时间序列估计基本知识的读者介绍合并时间序列设计。^[3]首先,我们会检验合并时间序列的回归模型,然后再检验各种不同的回归估计技术。本书会穿插对合并时间序列设计进行应用的不同例子,这些例子来自社会科学和行为科学。本书的叙述形式将结合对合并时间序列问题的理论展示与具体应用。为了方便阅读与课堂教学,所有的章节将用大写字母(A、B、C)来表明困难程度。

