

格致方法·定量研究系列

吴晓刚 主编



时间序列分析：回归技术

(第二版)

[美] 小查尔斯·W.奥斯特罗姆 (Charles W. Ostrom, Jr.) 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

61

格致方法·定量研究系列 吴晓刚 主编

时间序列分析:回归技术(第二版)

[美] 小查尔斯·W.奥斯特罗姆 (Charles W. Ostrom, Jr.) 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

时间序列分析:回归技术:第二版/(美)小查尔斯·W.奥斯特罗姆(Charles W.Ostrom)著;温方琪译;范新光校. —上海:格致出版社;上海人民出版社, 2017.6

(格致方法·定量研究系列)

ISBN 978-7-5432-2680-7

I. ①时… II. ①小… ②温… ③范… III. ①时间序列分析 IV. ①0211.61

中国版本图书馆 CIP 数据核字(2017)第 113888 号

责任编辑 裴乾坤

格致方法·定量研究系列

时间序列分析:回归技术(第二版)

[美]小查尔斯·W.奥斯特罗姆 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 上海商务联西印刷有限公司
开本 920×1168 1/32
印张 5.5
字数 110,000
版次 2017年7月第1版
印次 2017年7月第1次印刷

ISBN 978-7-5432-2680-7/C·181

定价:32.00 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,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,北京大学教育学院博士研究生李兰,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

相关数据往往会以时间序列的形式出现,即对单一分析单位在特定时间内获得重复的观测值。现实中有许多这样的例子。例如,一位工业心理学家连续数月监控一个工作小组的生产效率;一个政治学家一个又一个季度地跟踪一位总统的支持率;一个教育政策分析员追踪年度教育预算的数值。这些时间序列是包含变化的,并且可以像所有系统性的社会现象那样被解释。设想有一位都市社会学家珍·多伊提出这样一个假设:在哥谭镇,财产犯罪率(C_t)可以被表达为该城市失业率(U_t)的一个函数。使用过去 26 年的年度观测值,最小二乘(OLS)回归可以产生如下估计:

$$C_t = 2.1 + 3.7U_t + E_t$$

(6.3)

$$R^2 = 0.87 \quad N = 26$$

根据 t 比率(参见括号内数值),失业对犯罪的影响在 0.05 水平上显著。假使经典回归的假设条件都被满足,我们可以认为数据清楚地支持了多伊教授所提出的假设。然而,

时间序列数据往往违反经典回归的一个假设条件——误差项不存在自相关(nonautocorrelation)。比如说,某一观测值在某一时间点的误差项(E_t)有可能与另一观测值在较早一时间点的误差项(E_{t-1})相关。这种自相关会严重影响显著性检验的准确性。在上面的例子中,即使在最小二乘估计中 U_t 有一个很大的 t 比率, U_t 和 C_t 的关系实际上也可能在0.05水平上并不显著。因此,时间序列数据给普通的回归分析造成了不小的困难。

奥斯特罗姆教授向我们展示了诊断自相关的方法。他由简单的一阶自相关过程扩展到高阶、平滑且包含混合误差的过程。他进一步阐释了解决自相关问题的估计步骤。具体来说,他讨论了包括Cochrane-Orcutt法、Prais-Winsten法、Hildreth-Lu法以及Beach-McKinnon法在内的几种非常有用的广义最小二乘法。这些方法都以一种以读者为中心的方式呈现,奥斯特罗姆教授也提供了由计算机软件SPSS、TSP和SAS输出结果的例子。除此之外,本书还讨论了一些重要的特殊主题:Box-Jenkins法与经典回归方法的比较;内生和外生滞后变量;事后和事前预测。在本书第二版中,奥斯特罗姆教授介绍了有关时间序列回归分析的最新发展。更重要的是,他特别强调了有关理解和应用方面的问题。

迈克尔·李维斯-贝克

目 录

序	1
第 1 章 导言	1
第 2 章 时间序列回归分析：无滞后项实例	5
第 1 节 比率目标假设	7
第 2 节 误差项	10
第 3 节 时间序列回归模型	17
第 4 节 无自相关假设	20
第 5 节 违反无自相关假设的后果	26
第 6 节 自相关的常规检验	34
第 7 节 一种替代性的估计方法	39
第 8 节 EGLS 估计(以一阶自相关过程为例)	44
第 9 节 小样本特性	49
第 10 节 比率目标假设再回顾	52
第 11 节 向多元回归扩展	59

第 12 节	结论	61
第 3 章	替代性时间依赖过程	63
第 1 节	替代性过程	66
第 2 节	检验高阶过程	73
第 3 节	过程识别	77
第 4 节	估计	81
第 5 节	结论	89
第 4 章	时间序列回归分析:滞后项实例	91
第 1 节	分布式的滞后模型	93
第 2 节	滞后内生变量	96
第 3 节	用滞后内生变量来检验模型中的自相关	104
第 4 节	估计	106
第 5 节	EGLS 估计	107
第 6 节	改良型比率目标模型	113
第 7 节	对分布式滞后模型的解释	115
第 8 节	结论	119
第 5 章	预测	121
第 1 节	预测误差	125

第 2 节	预测生成	129
第 3 节	修正预测方程	132
第 4 节	预测评估	135
第 5 节	案例:预测美国国防开支	140
第 6 节	结论	146
第 6 章	总结	147
附录		149
注释		151
参考文献		154

第 **1** 章

导 言

这部专著详细介绍了基本回归模型的一个变化形式——所采用的数据为一种特殊时间序列的情况。当一组数据 $X_t (t=1, 2, \dots, T)$ 中 X_t 和 X_{t+1} 之间的间距是固定且恒定的时候,它就被看作一组时间序列(time series)。简单来说,观测值之间的顺序是非常重要的——我们不仅对观察点上的特定数值感兴趣,也对它们出现的顺序感兴趣。例如,关于总统支持率的序列、美国国防开支、世界体系中战争的数量以及失业率都符合时间序列分析的要求。

鉴于数据是按照特定时间顺序排列的,我们可以提出有关变量在过去以及未来是如何表现的问题。时间序列回归分析最大的优势在于,对于我们所感兴趣的变量,它既可以解释其过去的表现又可以预测其未来的表现。因此,一组时间序列的历史承担着两重职责:“首先,它必须告诉我们一个特定机制是如何随时间演化的;其次,它也允许我们利用那个机制预测未来。”(Nelson, 1973:19)正如我们所看到的,这两重职责的实现都必须基于正确地设定一个模型并估计它的参数。例如,对于总统、国会和公众而言,美国国防开支的预算是一个非常重要的决定。因此,充分了解这项决定是如何达成以及该决定会产生哪些附加影响是非常迫切的。为

了让我们在技术上的讨论具有连贯性,这个例子会在这本专著中反复出现。

从本书的最开始,我们就有必要区分时间序列回归分析(time series regression analysis)(Johnston, 1984; Judge et al., 1985; Kmenta, 1986; Pindyck and Rubinfeld, 1981; Wonnacott and Wonnacott, 1979)以及 Box-Jenkins 时间序列分析(time series analysis of the Box-Jenkins)(Box and Jenkins, 1976; Granger and Newbold, 1986; McCleary and Hay, 1980; McDowall, McCleary, Meidinger and Hay, 1980)。时间序列回归分析首先提出一个结构模型,然后再用时间序列数据进行检验。Box-Jenkins 时间序列分析(也被称为 ARIMA 或者转换函数)却是经验导向的,因为这种模型的形式(比如说变量以及滞后结构)取决于数据的拟合情况。

有意思的是,在上面所提到的贯穿整本书进行讨论的例子中,使用这两种不同的方法会得到相反的结论。有关国防开支(或者军备竞赛)文献的核心问题是“美国是否会响应苏联的支出水平”。即究竟美国的国防开支可以基本归结为对苏联的响应,或者是还有其他因素影响可被观测到的时间轨迹? 回归方法(比如说,Marra, 1985; Ostrom, 1978; Ostrom and Marra, 1986)支持两个超级大国军备竞赛的假说。但 Box-Jenkins 的方法或者其变化形式(比如说,Freeman, 1983; Majeski and Jones, 1981)却显示出军备竞赛并不存在。相反,美国当前的国防开支是由美国过去的国防开支所决定的。

与其让我现在就决定说哪一种时间序列分析的方法是

更加可取的，不如先做如下观察评论。这本书主要讨论的是如何采用回归（结构方程）方法来分析时间序列数据。在这里，建模者首先设定好因果结构，然后再来分析数据是否能够在经验上支持这种设定。尽管我们并不涉及 Box-Jenkins 时间序列模型对外生变量的解释，本书却会介绍 Box-Jenkins 时间序列方法作为一种有用的替代方法是如何模拟明显的错误过程。通过这种方法，本书尝试部分地统一这两种分析时间序列数据方法所存在的分歧。

本书对时间序列回归分析的介绍可以分为 4 个部分。首先，通过强调传统回归所存在的问题，我们提出一个正确的估计步骤来讨论时间序列回归模型的一般形式。其次，我们特别关注模型中那些由时间依赖过程（time-dependent process）而非一阶自相关过程（first-order autoregressive process）产生的错误。第三，我们会关注当因变量和自变量的滞后项也作为解释变量被放入方程时对原步骤的修正。最后，我们会讨论使用时间序列回归分析来生成和评估预测的方法。