

HZ BOOKS
华章教育

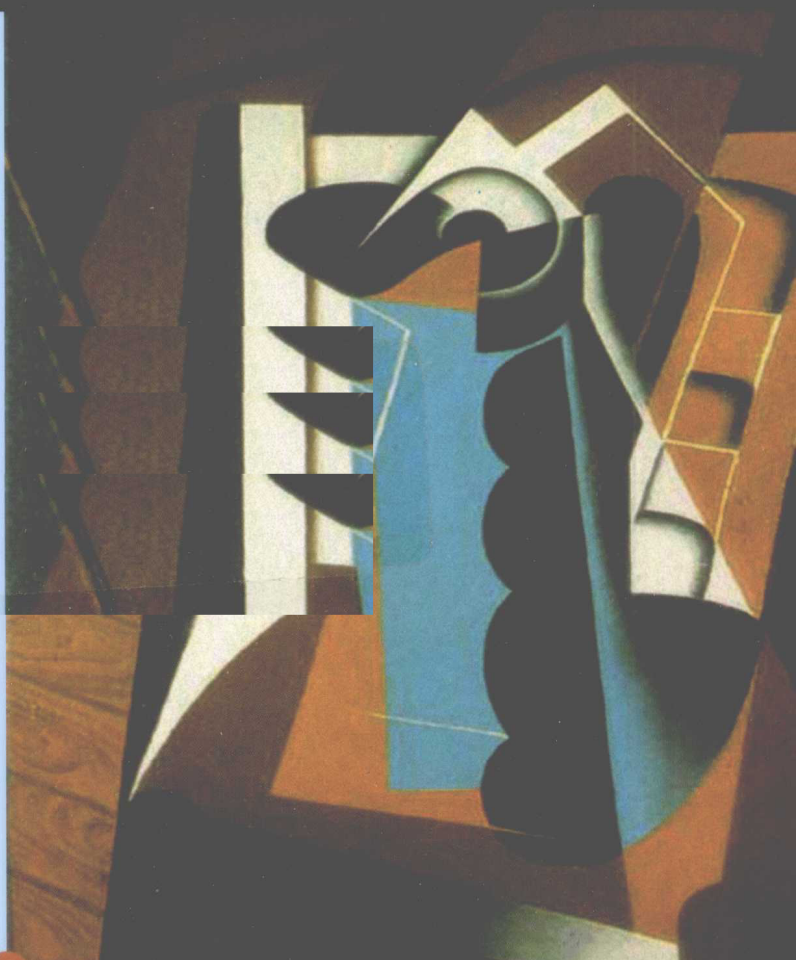
WILEY

经济教材译丛

(第 4 版)

面板数据计量经济分析

Econometric Analysis of Panel Data (4th Edition)



(美) 巴蒂 H. 巴尔塔基 (Badi H. Baltagi) 著

白仲林 等译
张晓峒 主审



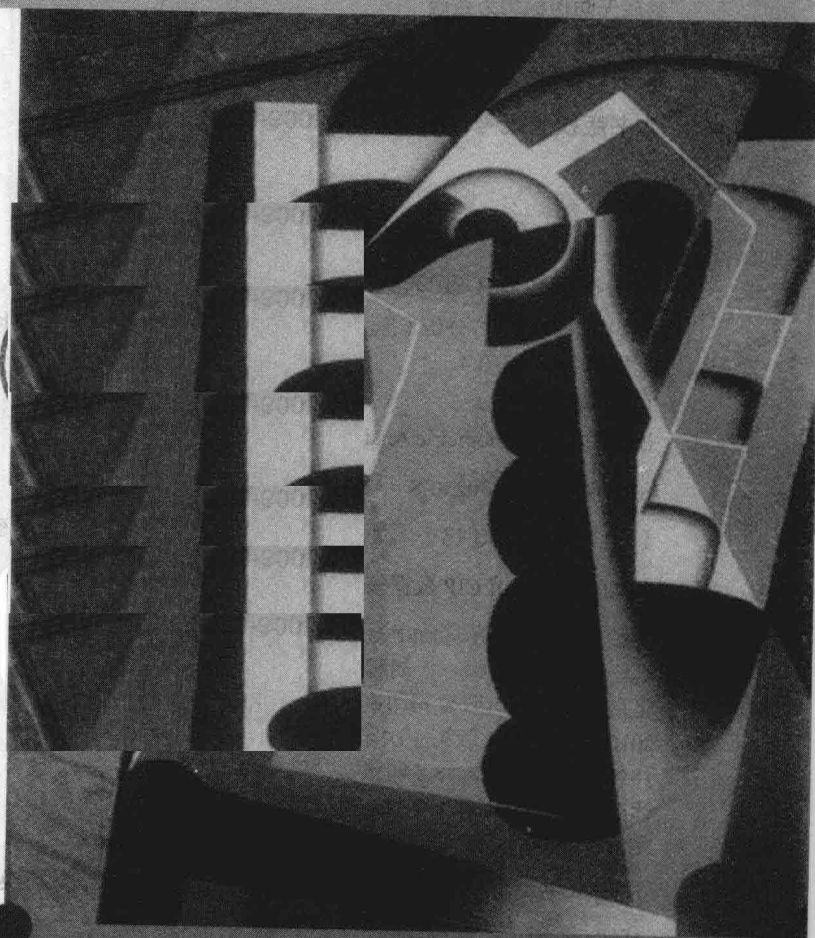
机械工业出版社
China Machine Press

经济教材译丛

(第4版)

面板数据计量经济分析

Econometric Analysis of Panel Data (4th Edition)



(美) 巴蒂 H. 巴尔塔基 (Badi H. Baltagi) 著

白仲林 等译
张晓峒 主审



机械工业出版社
China Machine Press

面板数据计量经济分析已经成为计量经济学研究的重要分支之一，本书系统介绍了面板数据模型的理论方法和应用，其内容包括静态、动态面板数据模型的设定、估计、检验和应用。尤其是对于非经典（非平稳）面板数据的计量经济分析方法的系统介绍是本书的特色之一。其次，本书还集中讨论了受限因变量面板数据模型、非平衡面板数据模型和面板数据联立方程模型的技术方法，指出了面板数据计量经济分析的发展方向。

本书适合高等院校经济学类本科生、研究生使用。

Badi H. Baltagi: *Econometric Analysis of Panel Data*, 4e.

Copyright © 2008 by John Wiley & Sons, Inc.

This translation published under license. Simplified Chinese Translation Copyright © 2010 by China Machine Press.

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

All rights reserved.

本书中文简体字版由 John Wiley & Sons 公司授权机械工业出版社在全球独家出版发行。未经出版者书面许可，不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 John Wiley & Sons 公司防伪标签，无标签者不得销售。

封底无防伪标均为盗版

版权所有，侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2009-6566

图书在版编目（CIP）数据

面板数据计量经济分析（原书第4版）/（美）巴尔塔基（Baltagi, B.）著；白仲林等译. —北京：机械工业出版社，2010.4

（经济教材译丛）

书名原文：Econometric Analysis of Panel Data

ISBN 978-7-111-30230-8

I. 面… II. ①巴… ②白… III. 经济统计—统计数据—经济计量分析—教材 IV. F224.0

中国版本图书馆 CIP 数据核字（2010）第 054081 号

机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码 100037）

责任编辑：刘夏风 胡智辉 版式设计：刘永青

北京京北印刷有限公司印刷

2010 年 5 月第 1 版第 1 次印刷

184mm × 260mm · 18.5 印张

标准书号：ISBN 978-7-111-30230-8

定价：48.00 元

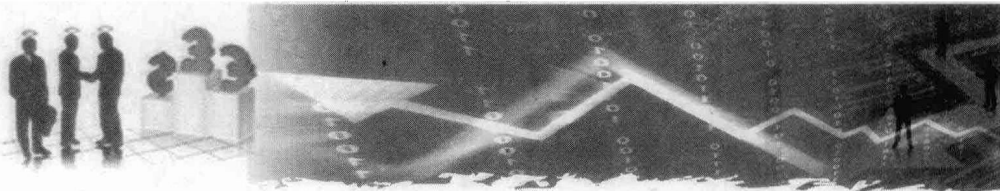
凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：（010）88379210；88361066

购书热线：（010）68326294；88379649；68995259

投稿热线：（010）88379007

读者信箱：hzjg@hzbook.com



译者序

2004年我托朋友从澳大利亚带回来一本《面板数据计量经济分析》(第2版),经过仔细阅读发现该书基础理论知识完整、经验应用内容充实,更值得一提的是系统地概述了面板数据计量经济学研究的前沿。

2006年秋季,在南开大学数量经济研究所张晓峒教授的指示下,开设面板数据专题讨论班。我与南开大学数量经济研究所的博士、硕士们一起研读巴尔塔基教授的这部学术著作。参加讨论的有2008届硕士和2009届博士。在为期一年的讨论学习基础上,王健博士(第11章)、攸频博士(第9章)、栾惠德博士(第4章)、赵娜博士(第3章)、魏学辉博士(第1、2和第8章)、郑妍妍博士(第7章)、刘雪艳博士(第7章)、邓露博士(第6章)和我(第5、10和第12章)合作完成了该书第2版的译稿。但是,在2008年准备联系出版译著时,巴尔塔基教授出版了本书的第4版。于是,在第2版译稿的基础上,我们又翻译了第4版,这次李树生博士和段鹏博士也参加了第4版的翻译。之后,我对全部译稿进行了统一梳理和编纂,张晓峒教授仔细审阅了全部译稿。

回顾从第2版、第3版到第4版的学习和翻译过程,我们深感巴尔塔基教授的这部专著具有深入浅出、系统全面和循循善诱的特色。本书不仅是面板数据计量分析课程的一部优秀教材,也是面板数据计量经济学理论及其应用研究者的一部适用工具书。

另外,在这期间,我们得到了国家自然科学基金委的大力支持,本学术专著的翻译也是国家自然科学基金资助项目(项目批准号:70771072、70571039)的阶段成果之一。

白仲林

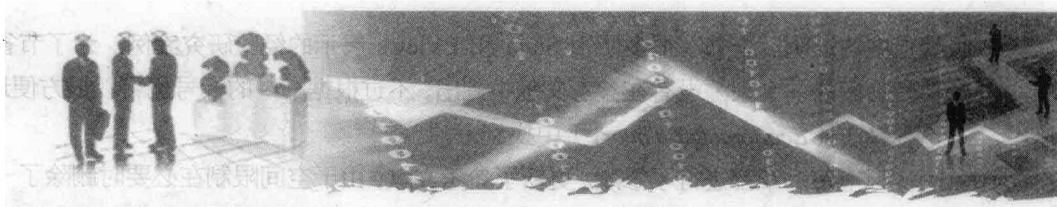
于天津财经大学振财里



作者简介

巴蒂 H.巴尔塔基

自 1979 年在宾夕法尼亚大学获得经济学博士学位以来,巴蒂 H.巴尔塔基先后在美国休斯敦大学和得克萨斯 A&M 大学任教。曾出版了《面板数据计量经济分析》和《计量经济学》等学术专著,编辑出版了《理论计量经济精粹》、《面板数据计量经济学新进展》(卷 I 和卷 II)、《非平稳面板数据、面板协整和动态面板》和《面板数据计量经济学:理论贡献和经验应用》等 100 多部著作以及担任权威经济学和统计学杂志的主编或副主编。巴尔塔基教授是得克萨斯 A&M 大学人文学科 George Summey Jr. 教授主席团负责人,并获得学术研究特别成就奖。他是《体验经济学》(*Empirical Economics*)的主编、《计量经济学杂志》(*Journal of Econometrics*)编委、《计量经济评论》(*Econometric Reviews*)编委和《应用计量经济学杂志》编辑。巴尔塔基教授也是《对经济分析的贡献》丛书的主编,曾获得《计量经济学杂志》特别会员(fellow)和经济计量理论的 Multa & Plura Scripsit 奖。



前 言

本书适合研究生面板计量经济课程使用。学习本书，读者应具有良好的数理统计学知识，同时计量经济学水平应达到 Greene(2003)的程度。另外，矩阵表述对于本书主题也是非常必要的。

本书的主要特点在于它涵盖了当今最新的面板计量技术，特别是在序列相关、空间相关、异方差、似无关回归、联立方程、动态模型、非平衡面板、限值因变量以及非平稳面板等领域。我尽力使本书内容简单易懂，因此在阐述不同领域不同符号表述的基本理论时使用了统一的表述符号。本书还引用并总结了一些使用面板计量技术进行经验研究的文献，这样读者就可以将计量经济方法和应用联系起来。本书从单方程估计方法开始，然后逐渐过渡到联立方程估计方法，这种安排与标准计量经济教材一样，因此对研究生来说，应该非常熟悉。

本书对面板计量只是做了基本概括，而不是像百科全书似的包罗万象。事实上，面板领域的研究非常广泛，本书并未涉及所有主题。30多年前，巴黎举行了第一次面板数据会议，会后 Mazodier 和 Trognon(1978)编辑出版了两册 *Annales de l'INSEE*。此后，又召开过 14 次有关面板数据的国际会议，最近三次分别是 2005 年的丹麦哥本哈根会议、2006 年的英国剑桥会议以及 2007 年的中国厦门会议。在第 3 版之后，《应用计量经济学》杂志 (*Journal of Applied Econometrics*) 出版了一期由 Baltagi 和 Pesaran (2007) 编辑的特刊，题目是《面板数据模型中的异方差与截面相关的理论及应用》(*Heterogeneity and Cross Section Dependence in Panel Data Models: Theory and Applications*)。Baltagi (2006) 以 2004 年 6 月在得克萨斯 State 学院举行的第 11 届国际面板数据会议论文为基础，编辑出版了《面板数据计量经济学：理论贡献和经验应用》一书 (*Panel Data Econometrics: Theoretical Contributions and Empirical Applications*)。

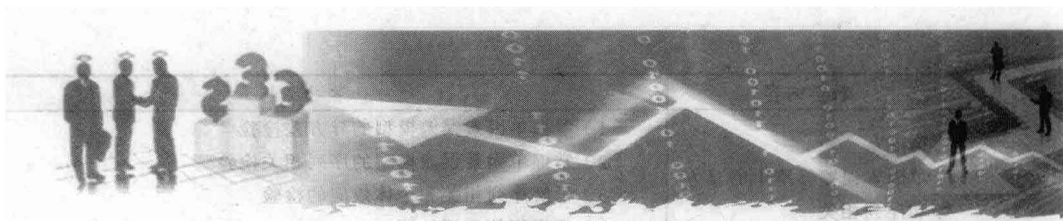
本书的写作与修订得益于我在锡拉库扎大学和得克萨斯 A&M 大学讲授的面板数据课程以及在很多地区所讲授的短期面板数据课程，包括华盛顿特区的国际货币基金组织(IMF)(2004~2007)；西班牙巴塞罗那的 Pompeu Fabra 大学 (2007)；葡萄牙的 Coimbra 大学 (2007)；西班牙塞维利亚的未来技术研究所 (2006)；希腊萨洛尼卡应用健康经济领域的玛丽·居里训练营 (2006)；华盛顿特区的美洲发展银行 (2005)；慕尼黑的经济研究中心 (2005)；乌得勒支大学的荷兰经济网(NAKE) (2005)；柏林的德国经济研究所(DIW)(2004)；辛辛那提大学 (2004)；波哥大罗萨里奥大学 (2003)；圣地亚哥加利福尼亚大学 (2002)；首尔国立大学 (2002)；因斯布鲁克大学 (2002)；罗马 Tor Vergata 大学 (2002)；曼海姆大学 (2002)；慕尼黑大学 (2002)；维也纳高级研究所 (2001)；法兰克福的欧洲中央银行 (2001)；巴黎第二大学 (2000)；Halle 经济研究所 (1997)；亚利桑那大学 (1996)；澳大利亚 Cairns 的国际亚太经济建模会议 (1996)。

本书第 4 版同第 3 版一样，继续引入更多面板数据文献中的经验研究实例，并且删除了第 1

版附录中的证明。全书有很多使用 Stata 和 EViews 展示的经验研究案例。为了节省空间，问题以及参考答案的引文来源没有在参考文献中列出，不过根据书中的标号也可以很方便地在《计量经济理论》杂志 (*Econometric Theory*) 中找到来源。

实际上每一章都进行了修改。更新了参考文献，由于空间限制在必要时删除了一些旧的参考文献。第 8 章动态面板数据模型，第 10 章专题，第 11 章限值因变量面板数据模型以及第 12 章非平稳面板数据都做了重大更新。现在第 1 章将面板数据库按宏观和微观进行了区分。第 2 章和第 3 章有 4 个使用 Stata 和 EViews 进行经验研究的例子。第 4 章展示了标准软件包中已有的检验步骤，同时指出了这些软件包中还没有的一些诊断和检验统计量。第 5 章说明了如何处理异方差和序列相关，以及如何在面板框架下进行检验。第 7 章使用犯罪案例说明了如何处理联立方程中的内生性问题，这里使用了 Stata 进行说明。同时还在使用 PSID 数据估计收入方程中说明了 Hausman 和 Taylor(1981)提出的方法。第 8 章使用最新出版的论文更新了动态面板数据相关内容，并且使用两个经验案例展示了估计方法。第一个案例估计了美国各州对香烟需求的动态变化；第二个案例研究了不同国家的民主和教育的关系。第 9 章通过估计一个享乐主义者住房供给方程表明了非平衡面板数据方法。第 10 章新增了一节涉及计数面板数据问题，并且给出使用 Stata 进行估计的实例，10.5 节关于空间面板的内容也进行了更新。第 11 章使用最新的论文更新了限值因变量面板数据模型的内容，并且给出了应用案例，即估计挪威的护士劳动供给。第 12 章尽力包括最新的有关非平稳面板的相关文献。事实上，在本书第 3 版出版以后，出现了一些新的理论结果和有影响力的文献，现在这些都在本章中进行了总结。特别是按照面板协整的相关文献研究了第一代和第二代面板单位根检验，同时使用 EViews 展示了对购买力平价的经验研究。

非常感谢我的合著者允许我自由引用我们共同的研究成果。我要特别感谢 Jan Askildsen, Georges Bresson, Young-Jae Chang, Peter Egger, Jim Griffin, Tor Helge Holmas, Chihwa Kao, Walter Krämer, Dan Levin, Dong Li, Qi Li, Michael Pfaffermayr, Nat Pinnoi, Alain Piroette, Dan Rich, Seuck Heun Song 以及 Ping Wu。我的很多同事都对这本书的内容有过直接或间接的影响，他们包括 Luc Anselin, George Battese, Anil Bera, Richard Blundell, Trevor Breusch, Chris Cornwell, Bill Griffiths, Cheng Hsiao, Max King, Kajal Lahiri, G.S. Maddala, Roberto Mariano, László Mátyás, Chiara Osbat, M. Hashem Pesaran, Peter C.B. Phillips, Peter Schmidt, Patrick Sevestre, Robin Sickles, Marno Verbeek, Tom Wansbeek 和 Arnold Zellner。Clint Cummins 使用 TSP 给出了本书中的例子的基准结果。David Drukker 提供了 Stata 中 Hausman 和 Taylor 程序以及第 7 章中 EC2SLS 的帮助，还包括第 9 章中 Baltagi 和 Wu 的 LBI 检验。Glenn Sueyoshi 提供了第 12 章中 EViews 面板单位根检验的帮助。同时还要感谢 Wiley 出版社的 Steve Hardman 和 Emma Cooper，感谢他们在编辑上给予的高效、专业的帮助；最后还要感谢我的妻子 Phyllis，她的鼓励和支持给予了我完成本书的力量。当然，本书的所有错误和遗漏都由作者承担。



教学建议

教学目的

本课程的教学目的在于从计量经济学的角度,介绍面板数据和面板数据建模方法的专业知识,并力求结合各类学科,培养既懂本专业,又具备建立面板数据计量模型知识的复合型人才。要求学生掌握面板数据计量分析的基本知识,了解采用面板数据建模的基本原理,并能够就面板数据计量分析的知识运用到经济学、管理学等社会科学研究的实践中。课程内容主要包括四大部分:第一部分为静态面板数据计量模型与分析,主要介绍面板数据的概念、静态面板数据计量模型、模型参数的估计方法和模型设定检验;第二部分介绍动态面板数据计量模型,侧重于 GMM 估计方法的使用及其相关检验;第三部分介绍限值因变量面板数据计量模型、非平衡面板数据建模方法和面板数据计量分析的一些新近专题;第四部分介绍非经典面板数据计量分析方法,包括面板数据单位根检验和协整检验理论。

前期需要掌握的知识

学生课前应掌握“线性代数”、“西方经济学”、“计量经济学基础”。

课时分布建议

教学内容	学习要点	课时安排
第 1 章 导论	(1) 掌握面板数据的概念、特征及其优势 (2) 了解一些经典的面板数据集 (3) 掌握面板数据的局限性 (4) 了解面板数据的测量误差和调查面板样本数据的非随机流失	2
第 2 章 单因素误差回归模型	(1) 掌握单因素误差回归模型的形式 (2) 掌握固定效应模型、LSDV 估计、组内估计和固定效应的检验 (3) 了解单因素误差回归模型的设定检验 (4) 掌握随机效应模型、GLS 估计和组间估计 (5) 了解固定效应和随机效应、ML 估计和模型预测 (6) 掌握案例分析技巧	6
第 3 章 双因素误差回归模型	(1) 掌握双因素误差回归模型的形式 (2) 掌握固定效应模型及其估计、固定效应的检验 (3) 掌握随机效应模型、GLS 估计和组间估计 (4) 了解双因素误差标准差的估计 (5) 了解双因素误差回归模型的 ML 估计和模型预测	4
第 4 章 面板数据模型的假设检验	(1) 掌握面板数据的混合估计检验 (2) 掌握个体效应和时间效应的检验 (3) 掌握 Hausman 设定检验 (4) 掌握案例分析技巧	4

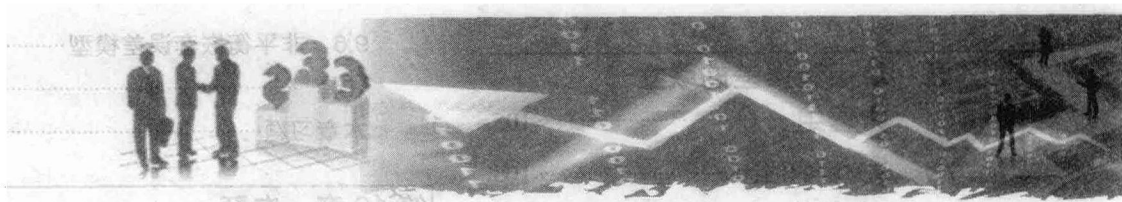
(续)

教学内容	学习要点	课时安排
第 5 章 单因素误差模型中的异方差和序列相关	(1) 掌握单因素误差模型的异方差检验 (2) 掌握单因素误差模型的序列相关检验 (3) 了解序列相关和个体效应的检验 (4) 掌握案例分析技巧	4
第 6 章 因素误差的似不相关回归模型	(1) 掌握单因素误差的似不相关回归模型 (2) 了解双因素误差的似不相关回归模型 (3) 掌握案例分析技巧	2
第 7 章 单因素误差的联立方程组	(1) 了解单方程模型和组内二阶段最小二乘估计 (W2SLS) (2) 了解误差分解二阶段最小二乘估计量 (EC2SLS)	2
第 8 章 动态面板数据模型	(1) 理解动态面板数据模型和 Nickell 偏倚 (2) 掌握面板数据自回归模型的 Arellano 和 Bond 估计 (3) 掌握动态面板数据模型的 Arellano 和 Bover 估计 (4) 掌握动态面板数据模型的 Ahn 和 Schmidt 矩条件 (5) 掌握动态面板数据模型的 Blundell 和 Bond 系统 GMM 估计 (6) 了解 Keane 和 Runkle 估计	6
第 9 章 非平衡面板数据模型	(1) 掌握非平衡的单因素误差模型及其组内估计 (2) 了解单因素误差方差的 ANOVA 方法 (3) 了解非平衡的单因素误差模型的极大似然估计 (4) 掌握案例分析技巧	4
第 10 章 专题	(1) 了解测量误差 (2) 掌握轮换面板数据模型及其估计 (3) 掌握伪面板数据构建方法 (4) 了解合并截面时间序列的其他方法 (5) 了解空间面板数据、异质面板数据 (6) 掌握计数面板数据模型及其估计	4
第 11 章 限值因变量与面板数据	(1) 掌握面板数据 logit 和 probit 模型 (2) 了解面板数据限值因变量模型的模拟估计 (3) 掌握动态面板限值因变量模型 (4) 了解面板数据的选择偏倚 (5) 掌握删截和截尾面板数据模型 (6) 掌握案例分析技巧	6
第 12 章 非平稳面板数据	(1) 掌握假定截面独立的面板单位根检验 (2) 了解截面相关的面板单位根检验 (3) 理解面板数据中的虚假回归 (4) 掌握基于残差的 DF 和 ADF 检验 (5) 掌握 Pedroni 检验 (6) 了解基于似然的面板数据协整检验 (7) 掌握案例分析技巧	6 6
课时总计		30 ~ 51

说明

(1) 在课时安排上, 对于数量经济学或统计学专业必修课建议按周 3 开设, 共 17 周或 51 个学时; 对于经济学、管理学的学科可以按周 3 安排, 共 10 周或 30 学时。建议根据专业特征选择相关内容, 比如管理类专业学生可以偏重微观面板数据的建模方法, 省略一些面板数据分析的扩展内容; 数量经济学或统计学专业的学生可以选择一部分内容让学生自学。

(2) 上机实践、案例讨论等活动可以在课程中穿插进行。



目 录

译者序
作者简介
前言
教学建议

第 1 章 导论..... 1

- 1.1 面板数据：一些例子..... 1
- 1.2 为什么使用面板数据，它们的
 优点和局限性..... 5
- 注释..... 9

第 2 章 单因素误差回归模型..... 10

- 2.1 介绍..... 10
- 2.2 固定效应模型..... 11
- 2.3 随机效应模型..... 14
- 2.4 极大似然估计..... 19
- 2.5 预测..... 20
- 2.6 案例..... 21
- 2.7 精选的应用案例..... 27
- 2.8 计算的注意事项..... 27
- 注释..... 27
- 本章习题..... 28

第 3 章 双因素误差回归模型..... 30

- 3.1 简介..... 30
- 3.2 固定效应模型..... 30
- 3.3 随机效应模型..... 32
- 3.4 极大似然估计..... 36

- 3.5 预测..... 38
- 3.6 案例..... 40
- 3.7 应用精选..... 43
- 注释..... 44
- 本章习题..... 44

第 4 章 面板数据的假设检验..... 49

- 4.1 面板数据的混合估计检验..... 49
- 4.2 对个体效应和时间效应的检验..... 54
- 4.3 Hausman 设定检验..... 63
- 4.4 进一步阅读..... 71
- 注释..... 71
- 本章习题..... 71

第 5 章 单因素误差模型中的异方差和序列相关..... 74

- 5.1 异方差..... 74
- 5.2 序列相关..... 78
- 注释..... 97
- 本章习题..... 98

第 6 章 因素误差的似不相关回归模型..... 100

- 6.1 单因素误差模型..... 100
- 6.2 双因素误差模型..... 101
- 6.3 应用与扩展..... 102
- 本章习题..... 103

第 7 章 单因素误差的联立方程组 105	9.6 非平衡嵌套误差模型..... 171
7.1 单方程的估计..... 105	注释..... 174
7.2 经验案例：北卡罗来纳州的犯罪..... 108	本章习题..... 175
7.3 系统估计..... 113	第 10 章 专题 177
7.4 Hausman 和 Taylor 估计量..... 116	10.1 测量误差和面板数据..... 177
7.5 实证分析：PSID 数据的收入方程..... 119	10.2 轮换面板..... 180
7.6 进一步阅读以及扩展..... 122	10.3 伪面板..... 182
注释..... 124	10.4 合并截面时间序列的其他方法..... 184
本章习题..... 124	10.5 空间面板数据..... 185
第 8 章 动态面板数据模型 128	10.6 混合模型的短期和长期估计..... 188
8.1 简介..... 128	10.7 异质面板数据..... 189
8.2 Arellano 和 Bond 估计量..... 129	10.8 计数面板数据..... 193
8.3 Arellano 和 Bover 估计量..... 135	注释..... 199
8.4 Ahn 和 Schmidt 矩条件..... 137	本章习题..... 199
8.5 Blundell 和 Bond 系统 GMM 估计量..... 139	第 11 章 限值因变量与面板数据 202
8.6 Keane 和 Runkle 估计量..... 141	11.1 固定的和随机的 logit 和 probit 模型..... 202
8.7 最新进展..... 142	11.2 面板数据限值因变量模型的模拟估计..... 208
8.8 经验研究案例..... 147	11.3 动态面板限值因变量模型..... 210
8.9 进一步阅读..... 152	11.4 面板数据的选择偏倚..... 213
注释..... 154	11.5 删截和截尾的面板数据模型..... 217
本章习题..... 155	11.6 实证应用..... 221
第 9 章 非平衡面板数据模型 157	11.7 经验案例：护士劳动力的供给..... 222
9.1 介绍..... 157	11.8 进一步的阅读..... 224
9.2 非平衡的单因素误差模型..... 157	注释..... 226
9.3 经验案例：享乐主义者的住房..... 162	本章习题..... 227
9.4 非平衡双因素误差模型..... 166	第 12 章 非平稳面板 229
9.5 非平衡面板数据个体效应和时间效应的检验..... 168	12.1 引言..... 229
	12.2 假定截面独立的面板单位根检验..... 230

12.3 截面相关的面板单位根检验···	237	12.8 进一步阅读·····	253
12.4 面板数据中的虚假回归·····	240	注释·····	256
12.5 面板数据的协整检验·····	244	本章习题·····	256
12.6 面板协整模型的估计与推断···	249		
12.7 经验案例：购买力平价·····	251	参考文献 ·····	259



第 1 章 导 论

家庭调查得到的面板数据越来越多,这可以说是过去三十年应用社会研究领域所取得的最重要的进展。

—Fitzgerald, Gottschalk, Moffitt (1998, p.252)

1.1 面板数据：一些例子

在本书中,“面板数据”一词指的是一部分家庭、国家或企业等在一段时期内的观测值所构成的集合。这样的数据可以通过在一段时期内对一些家庭或个体进行跟踪调查来获得。个体调查数据通常称为微观面板(micro panels),所搜集数据的特点是个体数 N 较大(通常是几百或几千个),而时期数 T 较短(最少是两年,最长不超过 10 年或 20 年)。与微观面板相反,宏观面板(macro panels)通常涉及一段时期内不同国家的数据。这类数据一般具有适度规模的个体 N (从 7 到 100 或 200 不等,如七国集团、OECD、欧盟、发达国家或发展中国家),时期数 T 一般在 20~60 年之间。微观和宏观面板要求使用不同的计量方法,例如,微观面板必须研究 T 固定而 N 较大时的渐近特性,而宏观面板的渐近特性则是指 T 和 N 都较大时的情况。还有,对于宏观面板,当时间序列较长时需要考虑数据的非平稳问题,如单位根、结构突变以及协整等,参见第 12 章;而微观面板不需要处理非平稳问题,特别是每个家庭或个体的时期数 T 较短时。另外,在处理宏观面板时必须考虑国家之间的相关性,而在微观面板中,如果个体是随机抽样产生,则个体之间不大可能存在相关性,因此不需要考虑此问题。

1.1.1 微观面板的例子

美国微观面板最著名的两个例子是密西根大学社会科学研究所搜集的研究收入动态变化的面板数据集(PSID)(<http://psidonline.isr.umich.edu>),以及劳动统计局发起的全国跟踪调查(NLS)所得到的面板数据集(<http://www.bls.gov/nls/home.htm>)。

1968 年开始时 PSID 只有 4 800 户家庭,到 2001 年已经增至 7 000 多户家庭。截至 2003 年,PSID 数据库中已经搜集了 65 000 多个体长达 36 年的信息。1968~1996 年每年进行一次调查,1997 年时改为两年一次,同时对样本进行了调整:减少了核心样本数,引入了一些 1968 年后的移民家庭样本包括其成人子女。所调查的数据主要集中在经济和人口统计领域,包括收入、贫困状态、食品或住房方面的公共资助、其他金融事件(如纳税额、家庭间资助)、家庭结构和人口数量、劳动市场工作、家务时间、住房状况、家庭迁移、社会经济背景和健康状况。其他补充的调查主题包括

住房和周边环境特征、努力动机、儿童保育、儿童抚养和儿童发展、工作培训和获得工作情况、退休计划、健康状况、直系亲属、财富、教育状况、军事战斗经历、风险承受能力、移民历史以及移民时间。

NLS 通过调查记录了几组男性和女性在不同时间点的劳动力市场行为以及其他重大生活事件的相关信息:

(1) NLSY97, 该组包括从全国抽取的近 9 000 个青少年, 他们 1997 年时年龄都在 12 ~ 17 岁之间。

(2) NLSY79, 该组包括从全国抽取的 12 686 个年轻男性和女性, 他们 1979 年时年龄在 14 ~ 22 岁之间。直到 1994 年对这组人一直是一年调查一次, 最近改为两年调查一次。

(3) NLSY79 儿童和成人组, 这组是由 NLSY79 中女性所生育的子女组成。

(4) 成熟女性和年轻女性组, 该组包括 5 083 名 1967 年时年龄在 30 ~ 44 岁之间的女性, 以及 1968 年时 5 159 名年龄在 14 ~ 24 岁之间的女性。对这组人一直是两年调查一次。

(5) 老年和青年男性组, 该组包括 5 020 名 1966 年时年龄在 45 ~ 59 岁之间的男性, 以及 5 225 名 1966 年时年龄在 14 ~ 24 岁之间的男性。对这两部分人的调查在 1981 年时停止了。

NLS 数据集中的变量主要包括学习经历和工作经历, 婚姻和子女状况, 培训支出, 儿童保育支出以及毒品和酒的使用量等。大量的研究用到了 NLS 和 PSID 数据库。NLS 网站还专门为使用了 NLS 数据库的 3 000 多篇文章、专著以及工作论文提供搜索服务。据 Brown, Duncan 和 Stafford (1996) 估计, 大约有 900 篇使用了 PSID 数据库的期刊论文和书籍出版。PSID 的应用范围较广, 包括跨期劳动力供给模型; 经济周期变化中的工资与就业; 失业、工作变换和劳动力流动; 消费、收入和资产负债表的动态变化; 家庭的延续行为; 贫困、福利和收入的动态变化; 经济状况的代际传递; 经济或人口事件的原因。

由当前人口调查 (current population survey, CPS) 项目所得到数据也可以建立面板数据集。CPS 是由劳动统计局的人口普查处对全国近 50 000 个家庭进行的月度调查项目 (www.census.gov/cps)。该调查项目已经进行了 50 多年, 它是了解美国劳动力特征的主要信息来源。与 NLS 和 PSID 相比, CPS 中包含的变量少, 时期数也短, 而且没有跟踪样本中的迁移人口。但是, CPS 涵盖的样本范围大, 因此对所有的群体都具有代表性。CPS 提供对就业、失业、收入、工作时间以及其他一些指标的估计, 从中还可以了解年龄、性别、种族、婚姻状况以及受教育程度等人口特征, 另外还有职业、行业以及劳动力类型等特征。

另外一个发展中国家家庭调查数据的重要来源是 1980 年建立的世界银行生活标准测量研究 (LSMS) 数据集 (www.worldbank.org/LSMS)。自从 1985 年以来, LSMS 已经在阿尔巴尼亚、越南等近 20 个发展中国家进行了调查。这些调查大都是 2 000 ~ 5 000 个家庭的小样本, 在有的国家调查了一次, 在有的国家调查了多次。在另外一些国家可能是 2 ~ 4 年的面板。调查问卷有 3 种类型: 家庭、社区和价格问卷, 有时也加入学校或健康设施问卷表。LSMS 主要集中在有关贫困本质的规律性上。像 LSMS 这种重复调查所得的数据, 尽管不是真正意义上的面板数据, 但也可以用来构造伪面板数据, 参见第 10 章。

尽管美国的面板数据库在 20 世纪 60 年代已经出现, 但直到 20 世纪 80 年代欧洲才建立了自己的面板数据。1989 年, 《欧洲经济评论》(European Economic Review) 专门有一期刊发了使用德国社会经济面板数据、瑞典家庭市场和非市场行为研究数据以及荷兰家庭面板数据的一些文章。德国社会经济面板数据库 (german socio-economic panel, GSOEP) 的第一次数据搜集是柏林德国经济研究所 (DIW) 在 1984 年进行的, 其中包括了 5 921 个联邦德国家庭 (www.diw.de/soep), 总共 12 290 个被调查者。所搜集的变量包括标准的人口统计变量, 还有工资、收入、福利支出、对生活

各方面的满意程度、希望和担忧、政治参与度等。1990 年由于德国统一，GSOEP 中又增加了 2179 个家庭中的 4453 个成年调查对象。在 GSOEP 中，样本流失率相对较低。Wagner、Burkhauser 和 Behringer (1993) 报告说，在 GSOEP 的八次调查中，有 54.9% 的初始调查对象的数据一次都没有缺失。使用面板数据进行国家研究的详细目录在以下网站上给出：<http://psidonline.isr.umich.edu/Guide/PanelStudies.aspx>。这些包括比利时社会经济面板数据 (BSP) (www.ufsia.ac.be/CSB/sep_nl.htm)，1985 年 BSP 中包括 6471 个比利时家庭样本，1988 年时 3800 个，1992 年时 3800 个 (其中有 900 个是新增的家庭样本)，1997 年时 4632 个 (其中包括 2375 个新增家庭样本)。英国家庭面板调查 (british household panel survey, BHPS) 是 1991 年开始的对英国私人家庭进行的一项年度调查，由艾塞克斯大学社会经济研究所进行 (www.iser.essex.ac.uk/ulsc/bhps/)。该面板由代表全国的 5500 个家庭，10300 个个体组成，他们是从英国 250 个区域抽取的。1999 年又从苏格兰和威尔士新增了 1500 个家庭样本，2001 年从北爱尔兰新增了 2000 个家庭样本。所搜集的数据包括反映人口和家庭特征的变量、家庭组成情况、劳动力市场状况、健康、教育、住房、消费以及收入、社会和政治特征等变量。瑞士家庭面板 (the swiss household panel, SHP) 第一次调查开始于 1999 年，共调查了 5074 个家庭 7799 个个体 (www.swisspanel.ch)。卢森堡社会经济面板数据 (luxembourg panel socio-economique “liewen zu lëtzebuerg”, PSELL I) 选取了具有代表性的 2012 个家庭共 6110 个个体。1994 年 PSELL II 将家庭数和个体数分别增至 2978 和 8232。瑞典市场和而非市场行为研究 (swedish panel study market and nonmarket activities, HUS) 于 1984 年，1986 年，1988 年，1991 年，1993 年，1996 年和 1998 年进行了多次数据搜集 (<http://www.nek.uu.se/faculty/klevmark/hus.htm>)，包括 2619 个个体的数据，数据涉及保育、住房、市场工作、收入和财富、税收改革 (1993 年)、改善环境的支付意愿 (1996 年)、地方税、公共服务以及非法经济行为 (1998 年) 等。

欧共体统计办公室设计并建立了欧共体家庭面板数据集 (ECHP)，参见 Peracchi (2002)。第一次数据搜集起始于 1994 年，包括了除奥地利、芬兰和瑞典之外所有现在欧盟的成员国。奥地利在 1995 年加入，芬兰在 1996 年加入，瑞典的数据直接取自瑞典生活状况调查 (swedish living conditions survey)。发起此项目是为了获得各成员国在有关方面的可比信息：包括收入、工作和就业、贫穷和社会排斥、住房、健康以及很多其他方面反映私人家庭和个人生活条件的社会指标。ECHP 上有链接可以和各国的家庭面板数据集联系 (如比利时和荷兰)，还可以和具有相似内容的面板数据集联系，如 GSOEP, PSELL 和 BHPS。该调查起止时间是 1994~2001 年 (<http://epunet.essex.ac.uk/echp.php>)。

其他一些面板数据集包括：加拿大统计局 (www.statcan.ca) 搜集的加拿大劳动收入动态变化调查数据 (SLID)，其中包括了全国 10 个省份将近 37000 个家庭样本。时间跨度是 1993~2000 年。日本消费者面板调查 (JPSC) 由家庭经济研究所 (www.kakeiken.or.jp) 于 1994 年开始建立，它包括全国范围抽取的 1500 个 1993 年时年龄在 24~34 岁之间的妇女样本 (A 组)；1997 年增加了 500 个年龄在 24~27 岁之间的妇女 (B 组)。所搜集的信息包括家庭构成、劳动力市场行为、收入、消费、储蓄、资产、负债、住房、耐用消费品、家庭管理、使用时间和满意程度等。1992 年北卡罗来纳大学的卡罗来纳人口研究中心建立了俄罗斯纵向跟踪调查数据集 (RLMS) (www.cpc.unc.edu/projects/rlms/home.html)。RLMS 是一项全国性的家庭调查，其目的是衡量俄罗斯改革对经济福利的影响。所搜集的数据包括个人健康和饮食摄入量、支出以及服务利用情况，还包括地区价格和基础设施等反映地区发展情况的数据。韩国劳动力和收入面板研究 (KLIPS) 于 1998~2001 年在 7 个省会城市及 8 个省的非省会城市中调查了 5000 个家庭及其成员 (<http://www.kli.re.kr/klips>)。澳大利亚家庭、收入和劳动力动态研究 (HILDA) 是墨尔本应用经济和社会研究所于 2001 年开始的一项家庭调查 (<http://www.melbourneinstitute.com/hilda>)。它包括了

澳大利亚全国 488 个临近地区的 7 682 个家庭, 19 914 个个体。印度尼西亚家庭生活调查数据 (<http://www.rand.org/FLS/IFLS>) 可用年份为 1993 ~ 1994 年, 1997 ~ 1998 年以及 2000 年。这里的样本包括全国 26 个省份中 13 个省份的 30 000 个个体, 他们代表了印度尼西亚近 83% 的人口。1993 年调查时走访了 7 224 个家庭, 2000 年时超过了 7 700 个。该面板数据集中包括的数据并不全面, 只是精选了一些容易进行经济研究的数据集。

1.1.2 宏观面板的例子

与微观面板数据集相对照, 也有一些“宏观面板数据集”供经济学家利用。这包括: ① 宾夕法尼亚世界表 (PWT), 其网址为 <http://pwt.econ.upenn.edu>。PWT 提供 1950 ~ 2004 年 188 个国家按国际价格计算的购买力平价及国民收入数据。此外, 欧盟 (EU) 或 OECD 还提供各成员国详细的购买力和实际产出估计; 世界银行 (World Bank) 还依据 GDP 水平对大多数 PWT 国家的当前价格进行了估计。② 世界银行也是宏观面板数据的一大来源, 包括世界发展指标集 (world development indicators, WDI) (www.worldbank.org/data), 2007 年的 WDI 中包括了 152 个人口在 100 万以上的经济体的 900 多个发展指标, 其中包括中国。③ 国际货币基金组织 (IMF) (www.imf.org) 也提供了一些宏观面板数据集, 如世界经济展望数据库 (world economic outlook databases), 它提供诸如 GDP 增长、通货膨胀、失业、收支平衡表、出口、进口、外债、资本流动、商品价格等时间序列数据; 还有国际金融统计数据库 (international financial statistics), 它提供 1948 年至今 200 多个国家的近 32 000 个时间序列, 其中有汇率、基金账户以及一些主要的全球性和国家性的经济指标; 贸易方向统计年鉴 (*direction of trade statistics yearbook*) 提供 186 个国家 7 年的贸易数据以及 156 个国家最近 6 个季度的季度贸易数据, 还有相关收支平衡、国际投资位置以及主要商品价格指数等数据; 国际货币基金组织还提供其成员国的国际储备、外汇流动性及金融稳定指标等数据。④ 联合国在 http://unstats.un.org/unsd/economic_main.htm 上提供了很多宏观面板数据, 其中包括国民账户、贸易和工业统计数据等。⑤ 经济合作与发展组织的数据在 www.oecd.org 上可以获得。⑥ 欧洲中央银行 (ECB) 在 www.ecb.int 上提供了欧盟各成员国的有关数据。⑦ 中央情报局的世界实情 (central intelligence agency's world factbook) 可从 www.cia.gov/library/publications/the-world-factbook 获得。提供单个国家一段时期内宏观数据的机构还有很多, 这里仅提及其中的一小部分, 我们可以将这些数据混合起来用在面板数据研究中。

1.1.3 一些基本的参考书

实际上, 所有的计量经济学的研究生教材都用一章或一个主要部分讲述面板数据计量经济学。关于这一主题向大家推荐的阅读材料是 Hsiao (2003) 的计量经济协会专著, 以及《计量经济学手册》中的两章: Chamberlain (1984) 的第 22 章以及 Arellano 和 Honoré (2001) 的第 53 章。Maddala (1993) 编辑了两卷有关这一主题的经典论文集, Baltagi (2002) 对其进行了更新, 又出版了两卷论文集, 涵盖了 1992 ~ 2002 年有关这一主题的主要论文。其他关于本主题的书有 Arellano (2003), Wooldridge (2002), 以及 Mátyás 和 Sevestre (1996) 编辑出版的包含 33 章内容的《面板计量经济学手册》第 2 版; Hsiao 等 (1999) 编辑出版的纪念 G.S. Maddala 的书; Krishnakumar 和 Ronchetti (2000) 编辑出版的纪念 Pietro Balestra 的书; Nerlove (2002) 出版的从历史的角度研究面板数据的书; Baltagi (2006d) 为经济分析系列编辑的一本书, 内容包括了本书未涉及的实证应用和理论贡献。关于非平稳面板数据模型的综述性文献可以参考 Baltagi 和 Kao (2000), Choi (2006) 以及 Breitung 和 Pesaran (2007)。一些期刊也出版了有关面板数据的专刊, 包括 Sevestre (1999) 编辑的两卷《经济统计年刊》(*Annales d'Économie et de Statistique*); Banerjee (1999) 编辑的一期《牛

津经济和统计会刊》(*Oxford Bulletin of Economics and Statistics*); Maasoumi 和 Heshmati (2000) 编辑的两期《计量经济评论》(*Econometric Reviews*), 19 卷中的第 3 期和第 4 期; Baltagi, Fomby 和 Hill (2000) 编辑的一期《计量经济学前沿》(*Advances in Econometrics*); Baltagi (2004) 编辑的一期《经验经济学》(*Empirical Economics*), 以及 Baltagi 和 Pesaran (2007) 编辑的一期《应用计量经济学》(*Journal of Applied Econometrics*)。

本书的目的是简单介绍面板数据分析中的一些基本问题, 适合于具有通常统计学和计量经济学背景的经济学家和社会科学家。面板数据分析方法已经用在了政治学领域, 参见 Beck 和 Katz (1995); 在社会学领域的应用参见 England 等 (1988); 在金融学领域的应用参见 Boehmer 和 Megginson (1990); 在市场营销学领域的应用参见 Erdem (1996) 以及 Keane (1997)。将本书内容集中在基本问题上似乎与现在快速增长的文献研究不相适应, 但考虑到本书的篇幅限制, 这种做法又是不可避免的。本书未涵盖的一些主题包括: 久期模型和风险函数 (参见 Heckman 和 Singer, 1985; Horowitz 和 Lee, 2004), 使用面板数据分析前沿生产函数的文献 (参见 Schmidt 和 Sickles, 1984; Battese 和 Coelli, 1988; Kumbhakar 和 Lovell, 2000; Koop 和 Steel, 2001), 以及关于时变参数、随机系数和贝叶斯模型的文献 (参见 Swamy 和 Tavlak, 2001; Hsiao, 2003), 还有程序运算方面的文献等 (参见 Heckman, Ichimura 和 Todd, 1998; Abbring 和 van den Berg, 2004)。



1.2 为什么使用面板数据, 它们的优点和局限性

Hsiao (2003) 列出了使用面板数据的一些优点。

(1) 可以控制个体异质性。面板数据表明个体、企业、州或国家之间都存在异质性。时间序列和横截面分析没有控制这种异质性, 因而其结果很可能是有偏的, 例子请参见 Moulton (1986, 1987)。现在我们用—个经验研究实例来说明这一点。Baltagi 和 Levin (1992) 研究了 1963 ~ 1988 年美国 46 个州的香烟需求问题。模型将香烟消费量设定为滞后消费量、价格以及收入的函数。这些变量都随着州和时间的不同而变化, 但也有很多其他影响香烟消费量的变量可能并不随着州的不同或时间的不同而变化。我们将这两类变量分别记为 Z_i 和 W_i 。属于 Z_i 的变量如宗教和教育, 对于宗教变量, 我们不可能得到每年每个州某一宗教 (如摩门教) 人数占总人口的百分比, 而且我们一般认为不同年份的百分比也不会有太大变化。同样, 完成高中或大学学业的人数占总人口的百分比也是如此。属于变量 W_i 的例子如电视和广播中的广告, 这些广告是全国性的, 不会随着州的不同而变化。另外, 这类变量中还有一些很难测量或者很难得到, 因此 Z_i 或 W_i 中的很多变量不能包括在消费方程中。遗漏这些变量就会导致估计结果的偏倚。面板数据能够控制这些不随州和时间变化的变量, 而纯时间序列和横截面分析就无法做到。事实上, 从数据中可以看出, 犹他州的人均香烟消费量还不到美国平均水平的一半。这是因为这里几乎是一个摩门教徒州, 而摩门教是禁止吸烟的。在横截面回归中可以加入虚拟变量控制犹他州的这种特征, 这事实上是在回归中剔除了该州的观测数据。使用面板数据我们就可以先对数据进行差分以消除所有 Z_i 类型的变量, 这样就可以有效地控制所有州的个体特征。这种方法不受 Z_i 是否可观测的影响。另外, 还可以使用代表犹他州的虚拟变量来控制所有不同于犹他州的个体特征, 这样就不用剔除犹他州的观测数据。

Hajivassiliou (1987) 给出了另外一个例子, 他使用 1970 ~ 1982 年 79 个发展中国家的面板数据研究了外债偿付问题。这些发展中国家在殖民历史、金融机构、宗教信仰和政治体制等方面有所不同, 所有这些反映国家特征的特定变量都会影响它们在借债或拖欠方面的态度, 同时也会影响债权国对待它们的方式。如果不考虑这些国家之间的异质性就会出现严重的设定错误。

Deaton (1995) 给出了农业经济中的一个例子。在农业经济中的一个问题是: 小农场是否比大