

经
济
科
学
译
库

An Introduction to
Modern
Econometrics
Using Stata

Stata

用Stata 学计量经济学

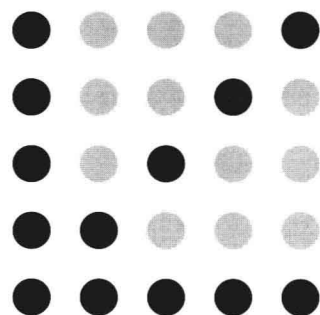
克里斯托弗·F·鲍姆 / 著
Christopher F. Baum

王忠玉 / 译



经济科学译库

用Stata 学计量经济学



克里斯托弗·F·鲍姆 / 著
Christopher F. Baum

王忠玉 / 译

An Introduction to
Modern
Econometrics
Using Stata



中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

用 Stata 学计量经济学/ (美) 鲍姆著; 王忠玉译. —北京: 中国人民大学出版社, 2012. 9
(经济科学译库)
ISBN 978-7-300-16293-5

I. ①用… II. ①鲍…②王… III. ①计量经济学-应用软件 IV. ①F224. 0-39

中国版本图书馆 CIP 数据核字 (2012) 第 221187 号

经济科学译库

用 Stata 学计量经济学

[美] 克里斯托弗·F·鲍姆/著

王忠玉/译

Yong Stata Xue Jiliang Jingjixue

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京东君印刷有限公司

规 格 185 mm×260 mm 16 开本

印 张 20.75 插页 2

字 数 455 000

邮政编码 100080

010-62511398 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2012 年 12 月第 1 版

印 次 2012 年 12 月第 1 次印刷

定 价 65.00 元

版权所有 侵权必究

印装差错 负责调换



中文版序

这本书是为应用研究者设计的，旨在发展他们如何应用 Stata 管理数据以及进行实证分析所需的坚实基础。这里既没有将此书设计成计量经济学教科书，也没有将此书制作成软件手册，而是将此书写作成研究者探索完成可靠的且可重复的实证研究的指南。

这本书一直深受世界各地使用 Stata 的研究者的青睐。我希望，随着中国优秀高校数量日益增多，更多渴望成为应用经济学者或金融专家的大学生群体因为本书的实用性而受益。我在波士顿学院讲授多年博士生计量经济学课程，结识了许多优秀的来自中国大陆的学生。无一例外，他们的数学和统计学背景一直都很优秀。然而，应用计量经济学的研究实践不仅要求掌握做分析应具有的基础知识，而且要求一边做实证一边学习。我希望中译本对那些想要在学术界、研究机构或私人企业领域中成为有效实证研究者的大学生来说，能培训他们更多的技能。

我愿意特别提及我和中国博士生刘博研所进行的卓有成效的互动研究。他来自北京航空航天大学，受到了中国国家自然科学基金和教育部博士点基金项目的慷慨资助。作为访问学者，刘博士对波士顿学院进行为期 1 个学年的访问，我们开展了合作研究。这期间，我们完成了 3 篇学术论文，其中 2 篇已发表。中国对一些“最优秀和最聪明的”博士研究生提供高级培训支持计划，这是一项具有前

瞻性的中国人力资本投资。就某种小的措施而言，我希望本书有助于促进同样的目标。

感谢哈尔滨工业大学的王忠玉教授承担了具有挑战性的中译本的翻译工作。

克里斯托弗·F·鲍姆
美国马萨诸塞州，布莱顿
2012年2月

序 言

本书是为经济和金融应用研究者撰写的一个简明指南，目的在于学习基本的经济计量方法并利用 Stata 对经济中典型的数据集进行分析。读者应具备应用统计学知识，即熟悉线性回归模型（普通最小二乘法或 OLS），并用代数形式表述它们，也就是相当于本科统计学或经济计量学课程水准。^① 此书还会用到某些多元微积分（偏导数）和线性代数的知识。

作者假定读者了解 Stata 的窗口界面，同时掌握数据输入、数据转换以及描述统计学的基本知识。如果需要回顾这些知识，建议读者查阅《Stata 入门手册》（*Getting Started with Stata*）。与此同时，建议那些较熟悉 Stata 的读者略去第 4 章以前的内容，而直接从第 4 章开始学习经济计量学。

对于任何研究项目来说，都要投入大量精力准备作为经济计量模型一个部分的特定数据。尽管本书主要关注点在于做经济计量应用研究，但是我们必须考虑到许多研究者所面临的下列重要挑战：将他们的原始数据资料转换成经济计量模型所要求的形式，甚至为项目提供所需的适当表格及图形。因此，第 2 章关注数据管理，以及 Stata 中为确保准确又有效地完成合适转换所用到的几种工具。如果你精通 Stata 这些方面的用法，就应略读这一章的内容，或者也可以再次学习第 2 章，以便重新理解 Stata 的用法。然后，第 3 章致力于讨论对经济与金融数

^① 在本科阶段，两本优秀的教科书是伍德里奇（Wooldridge, 2006）与斯托克和沃森（Stock and Watson, 2006）的书。

据的组织, 以及为改造几种数据结构形式(横截面、时间序列、混合形式、面板或纵向数据等)所需要的 Stata 命令。倘若你渴望从线性回归经济计量学开始, 则可略读这一章, 不过要注意, 将来会涉及这方面的主要内容。

第 4 章开始讨论本书的经济计量内容, 并阐述经济计量分析中最为广泛运用的工具: 用于连续变量的多元线性回归模型。这一章还讨论如何解释并阐明回归估计, 同时讨论假设检验的逻辑、线性与非线性约束。本章最后一节则考虑残差、预测值以及边际效应。

回归模型依赖于某些假设, 而这些假设经常与现实数据集合相违背。第 5 章讨论在存在设定误差的条件下, 极其重要的误差零条件均值假设可能怎样被违背的问题。这一章还探讨了检查设定误差的统计方法及图形法。第 6 章探讨可能被诸如独立同分布 (i. i. d.) 误差假设等违背的其他假设, 并且阐述广义线性回归模型。另外, 还解释如何诊断并修正两种背离独立同分布误差假设的最重要的情况, 即异方差性与序列相关。

第 7 章利用指示变量或虚拟变量探讨既包括数量因素又包括定性因素的线性回归模型、带有交互作用效应的模型以及结构变化的模型。

在应用经济学中, 许多回归模型都违背了误差零条件均值假设, 原因在于此类模型联立决定响应变量与一个或多个回归元, 或因为回归元出现测量误差。无论怎样, 这类原因将使 OLS 方法不再得出无偏且一致估计值, 因而你将要使用工具变量 (IV) 方法。第 8 章阐述工具变量估计量及其广义矩方法形式, 以及对决定是否运用工具变量法的检验方法。

第 9 章讨论将模型用于面板数据或纵向数据, 这些数据既有横截面维度, 又有时间序列维度。回归模型的扩展形式使得利用面板数据的丰富信息成为可能, 以此说明既有面板单元, 又有时间维度方面的异方差性。

绝大多数的经济计量应用都要对分类变量与受限因变量进行建模。像对购买决定的二值结果, 或者诸如消费额度的约束响应进行建模, 这里将是否购买与在购买的条件下消费多少的决策结合起来。一般地讲, 线性回归方法不适合此类结果的建模, 因此, 第 10 章阐述 Stata 中的几种受限因变量估计量。

附录讨论了将外部数据转换成 Stata 数据的方法, 同时解释了 Stata 基本编程方法。尽管你可以不需要任何编程工作就运用 Stata, 可是你学习 Stata 如何编程有助于节省大量的时间及精力。而且, 你还应学习通过利用 do 文件生成可复制的结果。对于 do 文件, 你能编辑、备份以及再次运行它。按照 Stata 指南的要求去做会使你编写的 do 文件更短小且更易维护和修改。



译者序

自 20 世纪 90 年代以来,我国引进和编写了大量的经济计量学(计量经济学)教科书。这些教科书在普及和推广经济计量理论及应用方面起到了十分重要的作用。随着计算机技术飞速发展以及各种计算机软件的开发和利用,经济计量理论取得了很大的进步,出现了广义线性模型、广义矩方法、非参数估计技术、离散选择模型等,以前的一些教科书尚未包括这些新颖的估计方法。此外,以往绝大多数教科书存在一种缺陷(当然,确实有某些教科书不断更新版本、添加新的理论及方法,这些教科书除外),即只注重理论方面的阐述,却极少给出实施一项经济研究一般程序的应用指南。另外,虽然已经有许多计算机软件提供有关经济计量分析的详细内容,但缺少建模解析内容。经济研究人员和大学生在进行经济实证研究时,搜集并分析原始资料,经常白白浪费宝贵的时间和精力,甚至重复其他人做过的工作。

近年来,经济计量分析方法发展迅速,不论是应用广度还是理论深度都得益于高性能计算机的飞速发展和经济计量软件的开发及使用。在方法背后,实证研究需要大量的经济理论作为根基与灵魂。在这种背景下,对于经济研究人员和大学生来说,如何有效实施高质量的实证研究分析,就成为一个需要认真思考和严肃对待的问题。

我们通过多年来的研究生经济计量学教学发现,虽然各高校研究生院要求经济类研究生学习所谓中高级经济计量学,但学生们在做经济实证分析时往往感到束手无策。即便经过思考设定了模型,也苦于寻找合适的软件或编程求解。实际

上,经济研究人员和大学生在进入学习和研究中高级经济计量学之时,就应该采取学习理论和实践应用二者并重的方式钻研实证分析。经济研究人员和大学生对经济计量建模全过程要有一个清晰的认识和理解,做到知晓经济计量方法、明晰实证分析过程,在探究经济问题时做到有章可循。这样要求终极目标是用规范的实证研究方法诠释新的观点,或者阐述对经济现象规律的发现。具体执行学习理论和实践应用二者并重的过程,无疑要经过多次的反复探索甚至遭遇挫折和失败,如图 1 所示。

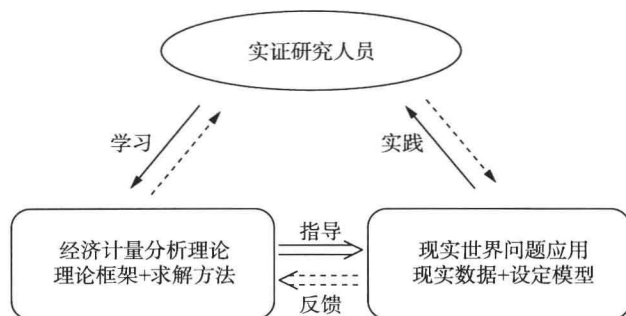


图 1 学习理论和实践应用二者并重

不论学习经济计量理论及方法,还是探索实践应用,都要提升自己对经济问题的计量建模水平认知层次,除扩大特定的前沿专业知识之外,应围绕经济计量建模方法论框架展开。在高级经济计量学理论中,经常出现高等数理统计方法、数据生成过程(DGP)、经济计量方法和模型分析等几个紧密联系的术语,实际上,可利用经济计量建模方法论框架来阐明这些术语之间的关联性。高等数理统计方法是经济计量学理论研究的数理推演工具;经济计量方法是经济理论与统计观测交互作用融合而形成的一种重要的实证研究方法;模型分析构成了理论与数据相互联系的技术桥梁;而数据生成过程则是对数据表象的一种深层次的认识,这些数据揭示了经济规律的某种客观属性。

通常,经济模型有两大类:一种是理论模型,另一种是实证模型。理论模型是从经济理论中直接导出;实证模型则是从理论模型衍生出来,是要用实际数据来估计。一般地,大多数实证模型是以回归模型形式表示,对模型所涉及的变量均要给予明确定义,并对自变量和因变量之间的关系作出详细说明,此外也要对模型的主要系数或由这些系数所导出的弹性可能数值的大小及符号给予一定的预期。

建立实证模型时,如何利用经济理论?一般说来不同研究者有着不尽相同的观点,可能会产生一些争论,甚至出现截然不同的观点。就这个问题而言,目前存在两种极端立场。一种观点认为,理论包含着唯一纯粹真理,因而应成为模型基础。持有该观点的研究者声称,所有残差都应该得到理论的解释,而不给随机性、不确定性或系统的外生冲击留有一席之地。这种建模方法也被称为结构方法,认为数据不可能完全显示自己是怎样产生的。结构方法源于考尔斯委员会方法。

实际上,考尔斯委员会方法是不同流派经济计量方法论中最经典、影响最深远的一种方法论。1989 年诺贝尔经济学奖获得者、计量经济学家哈维尔莫

(Haavelmo) 证明了, 要使方程组可识别, 充分必要条件是什么。识别方程组意指确定可以生成数据的因果(及概率)结构。识别问题源自下面的事实: 数据统计性质通常不足以揭示结构。这是归纳问题的一种形式(相关关系不一定是因果关系)。由考尔斯委员会提出的解决方法是: 运用经济理论, 首先设定具有因果关系的先验结构, 然后将统计方法用于测度因果关系的大小, 如果可能, 还运用统计方法检验源自理论的约束。

支持结构方法的研究者认为, 如果说经济研究的目标是数据生成过程, 只有在研究者模型的协助下才能了解数据产生结构, 尽管研究者模型可能是错误的。从科学研究方法看, 结构方法非常接近于物理学研究方法。

支持结构方法的经济研究人员注重模型, 强调估计模型的原始参数。所谓原始参数指的是那些在偏好和技术方程中的参数。这些参数不会因为政策干涉而变化。相反, 应用简化方法所估计的参数多数不是原始参数, 因而无法用来进行预测, 尤其无法预测从来没实施过的政策会有什么影响。

另一种观点认为, 只依据经济现象所呈现出的规律性和关联性, 建立基于对数据观测的“非理论的”模型。这种建模方法称为简化方法。简化方法认为, 实证研究应该让“数据自己说话”, 认为经济理论模型是由研究人员的意志决定的, 将研究人员的认识和看法施加到数据上而得到的结论, 只有在模型正确的情况下才会正确。由于研究人员不可能知道什么模型是正确的, 他们的主要研究工具很简单: 使用各种各样的回归分析。

实际上, 实证研究不仅仅是收集数据和分析数据, 更重要的一部分工作则是实证研究方法的研究, 因为任何一篇以数据为主的论文研究结果的有效性都取决于所用的研究方法。

结构方法和简化方法的区别在于, 它们对经济理论在实证研究中所起的作用有不同的理解。前者首先依据先验的经济理论设立结构模型, 然后由数据估计模型所含参数值。不难发现, 对结构参数估计值的信赖程度完全取决于对理论模型的信赖程度, 一旦理论模型不正确, 便会导致估计的结构参数大相径庭。因此, 这种方法对先验理论模型具有极强的依赖性。后者将描述数据固有特性作为建模的主要准则, 揭示隐藏于数据背后的经济规律, 因而该方法对经济理论的依赖较少。

对经济计量学方法论的认识, 可用一个“经济理论——经济数据——建模方法——实证(经验)模型”的全程建模框架来刻画及描述, 如图 2 所示。

现代经济计量学从其研究和探索的内容来看, 可分为微观经济计量学和宏观经济计量学两大分支, 这方面存在一些高阶教科书和参考书。在国外非常受欢迎的书, 关于微观经济计量学的有: Colin Cameron 和 Pravin Trivedi 的 *Microeconomics: Method and Application* (2005 年) 和 Jeffrey Wooldridge 的 *Econometric Analysis of Cross Section and Panel Data* (2002 年第 1 版, 2010 年第 2 版) 等; 关于宏观经济计量学的包括: Chetan Dave 和 David N. DeJong 的 *Structural Macroeconometrics* (2007 年) 和 Fabio Canova 的 *Methods for Applied Macroeconomic Research* (2007 年) 等。目前这四本书都有中译本。而更一般的高阶经济计量学教科书, 还包括 Greene 的 *Econometric Analysis* (2007 年第 6 版), 该书

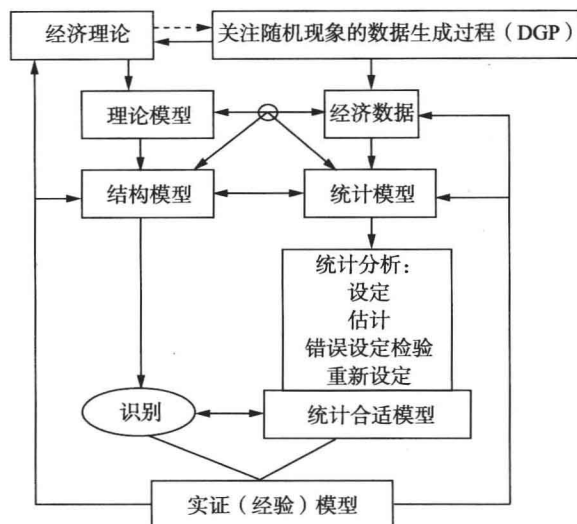


图2 经济计量建模方法论框架

也有中译本，另外该书在2011年出版了英文第7版。此外，Fumio Hayashi的《Econometrics》（2000年）也是一本不可多得的高阶经济计量学教科书，该书同样有中译本。

由美国波士顿学院鲍姆教授编写的这本《用Stata学计量经济学》，凝聚了他多年应用Stata研究经济问题的科研及教学成果。该书不仅在理论内容的选取上充分体现了以应用为导向的特点，而且精心选取大量例子与Stata编程完美结合，既突出经济计量理论的知识性，又强调易于应用的可操作性。作者提供了一种全新的融合数据分析、建模设定与如何应用Stata做经济计量分析的新尝试，为学习者提供了大量来自现实经济问题的应用实例及Stata编程。或许用一个副标题——“如何用Stata有效实施高质量的实证研究”更能恰如其分地揭示本书的特点。全书按照完成一项实证分析和研究的顺序，首先从数据管理以及研究人员每天不得不处理的工作质量管理问题开始，然后逐一阐述大多数经济研究使用的一系列经济计量工具。经济研究人员和大学生通过钻研本书可以掌握做实证研究的真知灼见和技巧，否则单凭自己摸索探究可能要积累多年才会达到如此程度。另外，作者指出在具体应用Stata时常犯的一些错误，比如数据多对多合并易出现的问题等。

本书的宗旨是，为想要做高质量实证研究的学习者提供一种如何用Stata完成高效和稳健的编程的样板。因此该书是面向应用的、不可多得的、极具指南特性的本科生和研究生经济计量学教科书和参考书。

译者在哈尔滨工业大学经济与管理学院的《Stata软件及在经济实证研究中的应用》讲座中曾运用鲍姆这本书的部分内容，取得了良好的预期效果。为此，特别感谢哈尔滨工业大学经济与管理学院院长于勃教授以及副院长梁大鹏教授对译者从事数量经济学教学及科研的关怀、支持。另外，在译者参加的教育部人文社会科学重点研究基地重大项目（项目号08JJD790153）的科研工作，以及译者

主持的黑龙江省哲学社会科学研究规划项目（批准号 11D080）中，本书发挥了十分重要的指导作用。

在翻译本书时，译者曾得到吉林大学商学院赵振全教授的帮助和支持；还得到哈尔滨工业大学王雪松副教授和博士研究生孙薇、刘仕煜、王伟同学的支持，有几位同学试译了部分章节，如孙薇翻译了第 2 章和附录 A 和附录 B；另外哈尔滨医科大学附属第二医院泌尿外科赵柏翻译第 3 章、第 4 章、第 5 章，其余章由王忠玉翻译并做最后的统稿校译工作。徐凤艳、王天元在翻译整理时也做了大量工作，比如翻译校译和文字录入。衷心感谢上述多位教授、同学以及家人的支持及帮助。

这里要特别感谢经济计量学家、美国波特兰州立大学（Portland State University）林光平（Kuan-Pin Lin）教授在本书稿翻译前和翻译期间对译者给予的大力支持及有益指导，使得译文逻辑更加严谨、专业。林教授多年来在促进和提升中国高校经济计量学教学、科研水平方面做了大量的基础性工作，许多人受益匪浅。

感谢中国人民大学出版社经济理论出版中心的马学亮先生一直关心和支持引进国外优秀经济类图书工作。

最后，原书中的一些印刷错误发布在 Stata 出版社网页上，译者据此逐一作了改正。另外，译者也发现了个别印刷错误，并和鲍姆教授联系，得到了他的确认。鲍姆教授对译者翻译时遇到的问题给予了热情的解答和指导，而且为本书撰写了序，译者对此深表感谢！

译者虽竭尽全力、深入研究和认真翻译，但仍可能存在不足和纰漏之处，恳请专家和读者指正。联系方式为：wangzhy@hit.edu.cn。

王忠玉
哈尔滨工业大学经济与管理学院
2012 年 2 月 16 日

记号与印刷体说明

作者将这本书设计成适合于边做边学的方式来学习，因此希望读者学习本书时坐在计算机前面，以便尝试利用书中给出的命令来复制作者的结果。然后，你可以对命令加以推广，以便适应自己所要解决的问题。

一般地说，本书用打字机字体 `command` 意指 Stata 命令、句法以及变量。命令后面的提示“点”(.)，表明为重做本书结果而在点后（正文）你要键入的内容。

为了讲解清楚本书所述内容，在数学记号上要遵循一些习惯：

- 矩阵用黑体、大写字母表示，比如 **X**。
- 向量用黑体、小写字母表示，例如 **x**。
- 标量 (scalars)* 用标准字体小写字母表示，比如 x 。
- 数据向量 ($\mathbf{x}_i$) 用 $1 \times k$ 行向量表示，将它们看成是来自数据矩阵的行。
- 系数向量 (β) 用 $k \times 1$ 列向量表示。

用 N 而不是 n 表示样本量，这样做使得本书要做一个例外处理，即用 N 刻画样本量。类似地，用 T 表示时间序列观测值个数， M 表示聚类 (clusters) 个数， L 表示最大滞后长度。我同样遵循普遍习惯，即杨-博克斯统计量用 Q 表示，类似地，萨尔甘 (Sargan) 之差检验 (difference-in-Sargan test) 用 C 表示。

为了使记号简单，我不使用不同字体来区分随机变量及其实现值。当人们对因变量 y 进行建模时， y 就是随机变量。 y 的观测值是该随机变量的实现值，同时将 y 的第 i 个观测值表示为 y_i ，而

* 标量又称为纯量。——译者注

所有观测值表示为 y 。同样地，回归元 x 表示总体的随机变量，并将此随机变量向量第 i 个观测值记为 x_i ， x_i 是数据矩阵 X 的第 i 行。

这本书不能代替 Stata 的手册的内容，而只是起一个补充作用，因而作者经常用 [R]、[P] 等意指 Stata 的手册。例如，[R] **xi** 意指《Stata 基础参考手册》(*Stata Base Reference Manual*) 关于 x_i 的词条，而 [P] **syntax** 意指《Stata 编程参考手册》(*Stata Programming Reference Manual*) 关于 **syntax** 的词条。

An Introduction to Modern Econometrics Using Stata by Christopher F. Baum

Copyright © 2006 by StataCorp LP

This work is copyright of Stata Press, a division of StataCorp LP. This translation is published with the permission of Stata Press. Stata is a registered trademark of StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845, USA.

Simplified Chinese version © 2012 by China Renmin University Press

All Rights Reserved.

目 录

中文版序	I
序言	III
译者序	V
记号与印刷体说明	X
第 1 章 引论	1
1.1 Stata 特色概述	2
1.2 安装必要的软件	4
1.3 安装支持素材	5
第 2 章 Stata 研究经济和金融数据基础	6
2.1 基础知识	6
2.1.1 use 命令	6
2.1.2 变量类型	8
2.1.3 _n 与 _N	9
2.1.4 generate 与 replace	9
2.1.5 sort 与 gsort	10
2.1.6 if exp 与 in range	10
2.1.7 利用带指示变量的 if exp	12
2.1.8 在统计命令中使用 if exp 与 by varlist	14

2.1.9	Labels 与 notes	16
2.1.10	varlist	18
2.1.11	drop 与 keep	18
2.1.12	rename 与 renvars	19
2.1.13	save 命令	19
2.1.14	insheet 与 infile	19
2.2	常见数据转换方法	20
2.2.1	cond() 函数	20
2.2.2	对离散型与连续型变量重新编码	21
2.2.3	处理缺失数据	22
	mvdecode 与 mvencode	23
2.2.4	字符串与数值间的相互转换	23
2.2.5	设置日期	25
2.2.6	使用有关 generate 与 replace 函数	26
2.2.7	egen 命令	27
	官方的 egen 函数	28
	用户团体提供的 egen 函数	28
2.2.8	用 by-分组计算	30
2.2.9	局部宏	33
2.2.10	变量循环 forvalues 与 foreach	33
2.2.11	标量与矩阵	35
2.2.12	命令语法与返回值	36
	习 题	37
第 3 章	经济数据的组织和整理	39
3.1	横截面数据与标识符变量	39
3.2	时间序列数据	40
3.2.1	时间序列算符	41
3.3	混合横截面时间序列数据	42
3.4	面板数据	42
3.4.1	面板数据运算	43
3.5	处理面板数据的工具	45
3.5.1	非平衡面板与数据筛选	46
3.5.2	面板数据的其他变换	48
3.5.3	移动窗口描述统计量及相关性	49
3.6	横截面与时间序列数据集组合	50
3.7	用 append 创建长格式数据集	51
3.7.1	利用 merge 添加汇总特征	52
3.7.2	多对多的 merges 的危险	53
3.8	reshape 命令	53
3.8.1	xpose 命令	56