



PANEL DATA MODEL AND
ITS APPLICATION IN ECONOMICS

面板数据模型
及其在经济分析中的应用



王志刚 / 编著



经济科学出版社
Economic Science Press



PANEL DATA MODEL AND
ITS APPLICATION IN ECONOMICS

面板数据模型
及其在经济分析中的应用



王志刚 / 编著



经济科学出版社
Economic Science Press

图书在版编目 (CIP) 数据

面板数据模型及其在经济分析中的应用 / 王志刚著.
—北京: 经济科学出版社, 2008. 9
ISBN 978 - 7 - 5058 - 7515 - 9

I. 面… II. 王… III. 经济统计 - 统计数据 - 经济计量
分析 IV. F224. 0

中国版本图书馆 CIP 数据核字 (2008) 第 133300 号

责任编辑: 金 梅 赵 蕾
责任校对: 杨晓莹
版式设计: 代小卫
技术编辑: 董永亭

面板数据模型及其在经济分析中的应用

王志刚 编著

经济科学出版社出版、发行 新华书店经销

社址: 北京市海淀区阜成路甲 28 号 邮编: 100142

总编室电话: 88191217 发行部电话: 88191540

网址: [www. esp. com. cn](http://www.esp.com.cn)

电子邮件: [esp@ esp. com. cn](mailto:esp@esp.com.cn)

北京欣舒印务有限公司印刷

华丰装订厂装订

787 × 1092 16 开 13.5 印张 280000 字

2008 年 9 月第 1 版 2008 年 9 月第 1 次印刷

ISBN 978 - 7 - 5058 - 7515 - 9/F · 6766 定价: 26.00 元

(图书出现印装问题, 本社负责调换)

(版权所有 翻印必究)

序

王志刚博士在北京大学接受了系统、全面的现代经济学训练，着力于在应用计量经济学方面做深入的研究，他的新著《面板数据模型及其在经济分析中的应用》是一本国内不多见、较高水平的应用计量经济学著作，涉及面板数据模型的诸多方面。本书是他在以往多次讲义的基础上所形成的研究成果，凝结着多年的辛劳与汗水。

现代计量经济学正经历飞速的发展，横截面数据分析，以及运用极为广泛的时间序列分析，已经成为经济研究人员所注重的研究工具，毫无疑问，采用这些先进的分析方法必将提升中国经济学研究的质量，丰富中国经济学理论，也为政府决策提供了更科学、更有力的支持。面板数据模型早在 20 世纪 60 年代就开始出现，随后在经济学、社会学、政治学、金融学领域被广泛应用。面板数据模型则介于宏观和微观之间，而更侧重于微观，它对一系列个体（国家、地区、省、行业、企业、家庭、个人等）的时间序列进行追踪调查，是横截面数据和实践序列数据的综合体，是现代经济学数据分析的一个关键点也是难点所在，它已成为当前国际计量经济学界研究的热点领域之一。

如果对比数据类型的特点就会发现，横截面数据主要侧重于微观个体的行为方面，时间序列数据侧重于宏观经济方面，而中国大部分的实证研究主要集中在宏观经济研究上，部分原因在于数据的可得性问题。我们知道，对于一个问题微观分析越细致、越深入，就越能够发现一些具有重要价值的信息。面板数据模型具有横截面数据和时间序列数据所不具备的两大独特优势：第一，它可以控制个体的异质性（heterogeneity），而这些没有出现在时间序列分析中。第二，它可以进行动态分析，而这些没有出现在横截面数据分析中。也就是说，面板模型综合了横截面模型的微观分析之长和时间序列模型的宏观分析之长，具有方法优势。

面板数据模型扩展到几乎所有的计量经济学分析中，包括结构方程、动态分析、时间序列模型（单位根检验，向量自回归（VAR），误差修正模型（VECM）等）、离散选择、选择性偏差、各种诊断检验与自助法、稳健估计等，所有标准的计量分析工具对面板数据都需要重新进行分析。这些特点就决定了面板模型的复杂之处。以自相关为例，它不仅要检测时间序列纬度上的自相关，还要检测某一时刻横截面个体间的自相关，这些问题不一而足，都使得面板模型分析更具挑战性。

随着面板数据的不断增多，各大高校和科研机构都建立了各种类型的面板数据库，这也推动了面板模型的应用。在中国，尽管微观数据的获取并不容易，但是面板数据的获取，至少在区域层面（例如省、地市、县）或产业层面（例如行业、企业）是可行的。我们欣喜地看到在诸多的学术期刊中面板数据应用得到了快速发展。就一般的计量模型使用而言，好的分析方法必须适合我们分析的问题，不能为方法而方法，那样无非是数字游戏，并不能发挥有效作用。滥用方法、无视方法的使用前提假设以及未对模型估计中所出现的计量问题加以修正都是不足取的，有时还是有害的。作为研究人员，掌握一些先进的分析方法是必不可少的，但要系统掌握一种好的分析工具需要花费相当多的时间的，所以选择合适的教材十分重要。此外，要完全掌握面板数据模型，必须注重动手实践，一个好的计量软件必不可少，Stata 软件较适合于经济研究人员，它最早是由一批劳动经济学家开发而成。我们知道，现代劳动经济学除了其经济学背景之外，它还是计量经济学发展最快的领域之一，它开发了许多新的功能强大的实证分析工具，例如最近流行的实验（Experiment）方法等，而别的软件更侧重统计背景。Stata 简便易用、功能强大，几乎经典计量教材中的模型都在菜单中有所体现。那些最新的分析方法，不一定出现在教科书中，但只要被国际权威杂志接受，就会由相关技术人员以最短的时间把它开发成 Stata 命令模块，而且很多是免费下载安装，它的开放性使得其更新能力更强，既能满足教学的需要，也能满足跟踪国际前沿分析的研究需要。

王志刚博士的《面板数据模型及其在经济分析中的应用》一书分别就静态模型、动态模型、单位根和协整分析、受限因变量、变系数模型和随机前沿模型六大领域进行了全面探讨，有详有略，侧重介绍了前三个领域。和同类出版物相比，本书有以下 3 个突出特点：（1）经典例子多。结合最新的国际期刊论文，例如政治经济学代表人物麻省理工学院的阿塞莫格鲁（Acemoglu）教授 2003 年的经典之作《财富的逆转》（Reversal of Fortune）一文，卢卡斯（Lucas）的理性预期模型，经典的跨国经济增长

收敛性检验等，通过提炼这些经典论文的分析要点，可以帮助读者更加深入地思考和熟练掌握各种面板数据分析方法。(2) 实用性强。结合国外比较流行的计量软件 Stata 进行了深入浅出的实战演练，它可以帮助许多经济研究人员，用较短的时间全面系统地掌握面板数据模型。(3) 覆盖面广。它涉及了面板数据模型的许多领域。当然，书中也难免有不尽完美之处，这有待作者进一步的研究来提升和完善。

对于这一新的分析方法，无论对于政策研究人员或高校科研人员而言，都存在着相当大的需求空间，但是，相关出版物在我国还很有限。王志刚博士的努力，我相信会有助于推广并普及这一有力的实证分析工具，丰富中国经济问题的实证研究方法，提高政策研究的水平。



财政部财政科学研究所所长、
研究员、博士生导师

2008年7月

目 录

第1章 引言	(1)
第1节 面板数据发展历程	(1)
一、面板数据分类	(1)
二、面板模型的优势和不足	(3)
第2节 面板数据在社会科学中的应用	(6)
第2章 静态面板模型	(9)
第1节 遗漏变量问题	(9)
一、遗漏变量	(9)
二、个体效应模型	(10)
第2节 随机和固定效应模型	(11)
一、基本概念	(11)
二、等相关模型	(12)
三、固定效应还是随机效应模型	(12)
四、严格外生假定	(13)
第3节 各类面板估计量	(14)
一、混合回归估计量	(14)
二、组间估计量	(15)
三、组内估计量或固定效应估计量	(15)
四、一次差分估计量	(16)
五、随机效应估计量——可行广义最小二乘法	(16)
六、双因子效应模型	(18)
第4节 面板模型的稳健标准差	(19)
一、面板稳健标准差	(19)
二、面板自助标准差	(21)
第5节 各类检验	(21)
一、混合回归模型对随机效应模型	(21)
二、混合回归模型对固定效应模型	(22)

三、随机效应对固定效应：豪斯曼检验	(22)
四、各种自相关检验	(23)
五、群组间的异方差	(24)
第6节 非平衡面板数据	(25)
第7节 预测	(25)
第8节 统计量的特征比较	(26)
第9节 软件命令和应用实例	(27)
一、面板模型的软件命令	(27)
二、应用实例	(28)
第3章 动态面板数据模型	(40)
第1节 工具变量法和广义矩估计	(40)
一、工具变量法	(40)
二、二步法工具变量回归	(41)
三、广义矩估计	(42)
四、各种检验	(42)
第2节 工具变量的选取	(44)
一、同期外生假定	(45)
二、弱外生假定	(45)
三、强外生假定	(45)
四、过多的工具变量	(46)
五、相关的软件命令	(46)
六、应用实例	(47)
第3节 严格外生性条件的违反和初始条件	(54)
一、基本假定	(54)
二、最简单的动态模型	(56)
三、各种检验	(57)
第4节 动态面板模型的带虚拟变量固定效应估计	(57)
一、动态模型	(57)
二、软件命令	(58)
第5节 面板模型工具变量法和广义矩估计	(59)
一、基本模型	(59)
二、一阶差分估计	(59)
三、广义矩估计	(59)
四、各种检验	(63)
第6节 动态面板数据的蒙特卡罗分析	(64)
一、模型设定	(64)

二、软件命令	(65)
第 7 节 极大似然估计	(66)
第 8 节 进一步讨论	(67)
第 9 节 二次差分估计量	(68)
一、一次差分	(69)
二、二次差分	(69)
第 10 节 应用实例	(71)
一、生产函数估计	(71)
二、增长的收敛性验证	(73)
三、就业方程估计	(74)
四、卢卡斯理性预期模型的广义矩估计	(78)
第 11 节 软件命令操作	(79)
一、差分方程估计的软件命令	(79)
二、系统广义矩估计	(87)
三、各种命令对比分析	(93)
第 4 章 面板数据的单根检验和协整分析	(95)
第 1 节 面板数据单位根检验方法	(95)
一、数据生成过程	(96)
二、莱文 - 林检验	(96)
三、伊姆 - 皮萨然 - 申检验	(98)
四、哈德里检验	(99)
第 2 节 面板协整检验	(101)
一、考检验	(101)
二、基于残差的拉格朗日乘数检验	(103)
三、佩德罗尼检验	(103)
四、似然比检验	(104)
第 3 节 面板的误差修正模型	(104)
一、误差修正检验	(104)
二、面板检验统计量	(105)
三、误差修正检验统计量的渐近分布	(106)
四、存在截面相关时的检验统计量	(106)
第 4 节 应用实例	(107)
一、面板单位根检验	(107)
二、面板协整检验和误差修正模型	(110)
第 5 节 软件命令	(112)

第5章 面板受限因变量模型	(113)
第1节 二元选择模型	(113)
第2节 静态面板离散选择模型	(114)
一、二元选择模型	(114)
二、固定效应的逻辑模型	(114)
三、随机效应的概率单位模型	(115)
四、面板随机效应托比模型	(116)
第3节 动态面板离散选择模型	(116)
第4节 选择性样本和面板数据磨损问题	(118)
一、选择性样本	(118)
二、赫克曼二步估计的概率单位模型	(118)
三、样本选择和面板模型	(120)
第5节 应用实例	(121)
第6章 变系数模型	(124)
第1节 介绍	(124)
第2节 随横截面个体而变化的系数	(125)
一、固定系数模型	(125)
二、随机系数模型	(125)
第3节 系数随着时间和横截面单位同时发生变化	(128)
一、固定系数模型	(128)
二、随机系数模型	(129)
第4节 软件命令和应用实例	(130)
一、命令介绍	(130)
二、投资需求函数	(131)
第7章 随机前沿估计	(133)
第1节 随机前沿生产函数	(133)
一、随机前沿生产函数和技术效率	(133)
二、复合残差项的分布	(134)
第2节 异方差情形下的效率估计	(138)
一、随机扰动项具有异方差	(138)
二、无效率项具有异方差	(139)
三、随机扰动项和无效率项都具有异方差	(139)
第3节 面板数据的随机前沿生产函数	(140)
一、基本模型	(140)

二、随时间变化的技术无效率项	(140)
三、进一步讨论	(142)
四、非单调生产效率模型	(142)
第4节 软件命令和应用实例	(144)
一、命令介绍	(144)
二、中国生产效率估计 (1978 ~ 2003 年)	(149)
第8章 Stata 软件基本操作	(153)
第1节 Stata 简介和网上资源	(153)
一、软件简介	(153)
二、网络资源	(153)
三、使用界面	(154)
第2节 软件的基本操作	(156)
一、数据管理	(156)
二、命令的格式与基本命令	(158)
第3节 基本的回归分析	(161)
一、回归命令	(161)
二、预测的语法格式	(162)
三、回归诊断	(163)
四、应用实例	(163)
第4节 简单的编程	(168)
附录 Stata10.0 所有手册和命令一览	(172)
参考文献	(194)
英文 - 中文人名索引	(199)
后记	(203)

第1章

引 言

第1节 面板数据发展历程

微观计量经济学是计量经济学发展最快的领域之一，它涵盖的内容包括离散选择模型、面板数据模型、社会试验方法、计数数据（Count Data）模型、久期（Duration）模型、分层（Stratified）与聚类（Clustered）样本模型、托比（Tobit）模型、样本选择（Sample Selection）模型等，其中面板数据模型的应用范围相当广泛，涵盖了诸多计量模型，这也是我们本书讨论的核心内容，下面我们来简单回顾其发展历程。

1986年，萧政（Hsiao Cheng）出版了第一本《面板数据分析》，同时期有29项研究涉及这一个关键词：“面板数据（Panel Data）或纵列数据（Longitudinal Data）”。根据社会科学文献索引指数（SSCI），截至2003年，有580篇研究，到2004年达到了687篇。自从贝尔斯特拉（Balestra）和纳洛夫（Nerlove）于1966年发表了一篇研讨会议论文以来，对面板数据这一新的计量分析工具的应用取得了快速的发展。国际上已经成立了专门研究Panel Data的协会，每两年举办一次全球学术交流大会，与会者都是从事Panel Data的经济学家、计量经济学家、统计学家和社会学家，大会强调除了在计量经济学领域外，要扩展其应用领域，以发现新的分析方法和研究领域，特别强调在社会学、经济学包括金融学，以及医学中的应用。2007年7月在中国厦门大学举办了第14届国际面板数据大会，大会的主题发言代表了当前面板数据研究所取得的令人振奋的最新进展^①，再次表明了这一分析工具的重要地位和国际影响力，也必将对中国面板数据模型的应用起到积极的推动作用。

一、面板数据分类

面板数据综合了时间序列数据和横截面数据，成为现代计量经济学的重要分析领

^① 会议论文如有需要，可向作者索取，这些属于高级面板模型内容，本书以初中级面板模型为主。

域。面板数据的定义是对同一组个体连续追踪多个时期所得到的数据。这些个体包括：工人、家庭、企业、产业、地区和国家。

一类是较为宏观的面板数据。例如我们把一组国家的时间序列数据混合在一起，就得到了一个国家面板数据，近来应用这些国家面板数据库表现在跨国的增长和收敛模型研究中，这些数据提供了外生的制度或政策变化，这有助于确定我们所感兴趣的经济变量。

另一类面板数据包括家庭或企业的数据库，这些数据基于抽样、普查或是行政管理记录，或是公司的财务统计，这些是“微观的面板数据”。典型的例子是，包含大量个体的短时期数据，包括后面提到的密歇根大学的，欧洲社区家庭收入调查，以及其他一些欧美国家的家庭支出调查数据。这一类起源于横截面数据的计量分析，应用于劳动经济学，需求分析，成本分析和生产函数分析。它的特征体现在较大的横截面纬度，较小的时间段。家庭或企业数据可以克服加总数据中对个人行为原则的误设问题。而且，面板数据通过政策变化前后的不同时期比较，可以提供确定经济结构参数所需的外生性变化。同一组个体的不同时期变化，可以很好地刻画动态行为和控制观测不到的异质性问题。

面板数据的分析重点是对异质性（Heterogeneity）的处理，而且扩展到几乎所有的计量经济学分析中，包括结构方程、动态分析、时间序列模型、离散选择、样本选择性问题 and 单位根，所有标准的计量分析工具对面板数据都需要重新进行分析。

美国有两大类著名的面板数据库，一个是俄亥俄州立大学人力资源中心的全国劳动力市场经历抽样调查（NLS，www.bls.gov/nlshome.htm），另一个是密歇根大学的收入动态调查数据库（PSID，www.isr.umich.edu/src/psid/index.html）。NLS 数据起于 1960 年代中期，它包括五个基础劳动数据库：1966 年 45 ~ 59 岁，1966 年 14 ~ 24 岁的年轻人，1967 年 30 ~ 44 岁的妇女，1968 年 14 ~ 24 岁的年轻妇女，以及 1979 年 14 ~ 21 岁的所有性别的年轻人。1986 年，NLS 扩大到 1979 年那些参与调查的妇女所生的小孩。调查的变量有好几千，主要是强调劳动力市场的供给方。PSID 数据库起自 1968 年，每一年都搜集全国 6 000 户代表性家庭和 15 000 个体的经济信息。包括超过 5 000 个变量，例如，就业、收入、人力资本变量，以及住宅信息、旅游和人口流动等。这些面板数据库用于研究美国，此外还用于分析经济和社会政策对人们福利的影响、贫困的实质和原因。除了这两大类最著名的数据库外，还有很多国家的面板数据库^①。美国北卡莱罗纳大学有一个中国健康与营养调查数据（China Health and Nutrition Survey, CHNS）库^②，其中有面板数据库，而且有很多相关研究成果可供下载。

面板数据格式有两类，我们给出常用的格式，就是先按照个体进行排序，然后再对每一个体按照时间顺序进行排序，结果如表 1-1 所示：

① <http://www.bris.ac.uk/Depts/Economics/Growth/datasets.htm>，其中有一些跨国面板数据。

② <http://www.cpc.unc.edu/projects/china>，先注册再免费下载。

表 1-1 面板数据的常用格式

个体标识 (ID)	年代 (year)	变量 1	变量 2	...	变量 n
Id1	1960
Id1	1961
...
Id1	2006
Id2	1960
Id2	1961
...
Idm	1960
Idm	1961
...
Idm	2006

注：我们还可以先按照时间排序，然后再对每一年的个体进行排序，那样不会影响面板回归分析结果。

通常我们所指的面板模型大部分是横截面纬度 (N) 远大于时间纬度 (T)。时间序列横截面数据 (Time Series Cross-Sectional (TSCS) data) 往往指的是时间纬度高于横截面纬度的数据，即 $T > N$ ，或 $T \approx N$ ，甚至 $N > T$ ，但是 T 具有相当长时期。

二、面板模型的优势和不足

萧政 (Cheng Hsiao, 1985, 1986, 2005); 克莱夫马肯 (Klevmarcken, 1989) 和索伦 (Solon, 1989) 指出面板数据分析具有下面的优点:

1. 控制个体的异质性

面板数据表明个体、企业、州或国家是存在异质性的。单纯的时间序列分析和横截面分析没有控制异质性，面临着估计有偏性的风险 (莫尔顿, 1986, 1987)。我们可以列举一个实证研究的例子来说明这个问题，巴尔塔吉和莱文 (Baltagi & Levin, 1992) 考察了美国 46 个州 1963 ~ 1988 年的香烟需求问题。香烟的消费是上一期消费的，香烟价格和收入的函数。这些变量随着不同的州和时间而发生变化。但是，有许多影响香烟消费的变量可能是不随时间或地点而变化的。我们分别称为 z_i 和 w_i ，前者例如教育程度和宗教信仰，我们不能期待每一个州的摩门教徒比例随着时间而发生极大变化，同样也不能期待某一州内受高中或大学教育程度人口比例会随着时间发生极大变化；后者例如，电视和广播的广告费用，这个广告是全国性的不随着地区不同而不同，但是随着时间而变化。但是我们有时候不能控制所有的这些变量，遗漏变量会导致估计的偏误。而面板数据可以很好地控制这些不随地点或时间变化的变量，这是时间序列数据和横截面数据所做不到的。无论 z_i 能否观察到 (主要是一些地理特征或制度特征)，我们可以通过差分消去其影响。

2. 面板数据可以提供更多的信息，更多的变化，较少的共线性、更多的自由度和效率

时间序列数据常常会带来变量间的共线性，例如前面提到的香烟需求，价格和收

入的时间序列数据常常存在着多重共线性。但是在面板数据中，这种共线性就会小得多，因为横截面数据为价格和收入信息加入了许多变异性。实际上，数据的变异可以分为不同规模和特征的州之间的变异，以及州内部的变异。前面的变异通常占据相当大比例。此外，较多的信息会产生更为可靠的参数估计。

3. 面板数据较好地研究了动态调整过程

看起来相当稳定的横截面纬度的分布掩盖了许多变化。面板数据可以很好地研究失业的持续性（duration）、工作转换（switching）、居民以及收入的流动性（income mobility）。面板数据同样适合研究失业和贫困的持续性问题，他们可以分析经济政策变化所导致的失业或贫困的调整速度。例如，在测度失业时，横截面数据可以估计在某一时刻有多少比例的人群失业。混合截面数据可以表明这种比例的变化。只有面板数据可以估计出在某一时期失业人群中到下一期继续保持失业状态的人群比例。阿申费尔特和索伦（Ashenfelter & Solon, 1982）强调了一些重要的政策问题，包括确定家庭是否经历贫困、失业，以及依赖福利生活的暂时性或是周期性，这些问题都需要面板数据来解决。面板在估计跨期关系，例如生命周期以及跨期模型中也很实用。实际上，面板可以将某一时刻的个体经历和行为与另一时刻的个体经历和行为相连。例如，阿申费尔特（1978）研究了联邦培训项目对这些参与者未来收入的影响，这可以通过对未参与组的对比分析进行估计。

4. 面板数据可以较好地确定和估计一些由纯粹横截面或纯粹时间序列无法确定的经济关系

本·普拉斯（Ben - Porath, 1973）给出了一个例子，假设我们有一个妇女的横截面数据，她们的年劳动参与率平均为 50%。这可能是因为：或是每一个妇女在某些给定年份只有 50% 的就业机会；或是 50% 的妇女一直在工作，其余 50% 一直未就业。前一种情况具有高工作转换率，而后者没有工作转换。只有面板数据才能区分这两种类别。

另外一个例子是研究是否工会会员资格会提高或减少工人工资，这个问题通过观察工人从一个工会会员到非工会会员之间的工作转变就可以得到答案。控制个体的特征不变，我们可以较好地回答工会会员资格如何以及多大程度上影响工人的工资（弗里曼（Freeman），1984）。

5. 与单纯的横截面数据或时间序列数据相比，面板数据模型允许我们构造和验证更为复杂的行为模型

例如面板数据模型可以很好地研究技术效率（technical efficiency）模型（巴尔塔吉和格里芬（Griffin），1988b；考恩威尔（Cornwell）、施密特（Schmidt）和西可（Sickels），1990；昆伯卡（Kumbhakar），1991，1992；巴尔塔吉等，1995；库普（Koop）和斯蒂尔（Steel），2001）。而且在分布滞后模型中，面板数据要求的约束条件较少（萧政，1986）。

6. 面板数据通常是搜集一些微观个体数据，包括个人、企业和家庭

许多变量可以在微观层面上得到更精确的测度，可以消除加总数据所带来的偏差（布伦德尔（Blundell），1988；克莱夫马肯，1989；布伦德尔、梅格尔（Meghir），1990）。

此外，我们还要看到面板数据的一些局限，表现在下面几个方面：

(1) 涉及和数据收集问题。

卡斯普什克 (Kasprzyk, 1989) 对面板数据调查问卷涉及和数据搜集与整理作了详细的分析。这包括覆盖面问题 (对关注群体的不完备记录)，未响应 (Nonresponse) (由于缺乏和被调研人的合作或是因为会见者的问题)，回忆 (回答者不能准确地记忆)，采访频率、采访问隔，参考时期，使用约束或样本的时间偏差 (贝勒 (Bailar), 1989)。

(2) 测量误差的扭曲。

由于问题设计不清楚，记忆错误，有意的扭曲，不合适的被调查者，对回答者的误记录等，这些都会导致测量误差 (卡腾 (Kalton)、卡斯普什克、麦克米伦 (McMillen), 1989)。

(3) 选择性问题。

这包括：首先，自选择 (Self-Selection) 问题。人们选择不去工作是因为保留工资高于市场的工资报价。此时，我们只能观察到他们的个体特征，但是观察不到工资。既然，他们的工资有缺失，这个样本就是非完全截断的 (censored)。但是，如果我们不能到这些人的所有数据，这就是一个完全截断 (truncated) 样本。例如新泽西州的负所得税试验，我们仅仅关注贫困，那些收入水平高于贫困线 1.5 倍的人群就从样本中丢弃掉了。从这个完全截断样本中得到的推断会导致出现一些偏差 (豪斯曼和怀斯 (Hausman & Wise, 1979))

其次，非响应 (Nonresponse) 问题。由于拒绝参与，没有人在家，没有跟踪的样本单位，以及其他原因都会造成非响应问题。当一个或多个问题没有回答，或是不能提供一个有用的相应时，这会导致部分响应。但抽样家庭没有任何信息反馈时，就会出现完全的非响应问题。除了缺失数据所造成的效率损失外，非响应问题还会引起对总体参数估计的严重偏差。赫威茨和马斯金 (1998) 表明这种问题的严重性和非响应的数量成比例。

最后，磨损 (Attrition) 问题。当然，非响应问题也会出现在横截面数据中，但是在面板中造成的后果更严重，因为随后的跟踪调查仍然会出现非响应问题。回答者可能会去世、流动、或是发现回答成本太高而不愿回答。比约克伦 (Björklund, 1989) 和里德 (Ridder, 1990, 1992) 讨论了数据磨损的后果。磨损的程度大小取决于所研究的面板数据 (卡腾、卡斯普什克、麦克米伦, 1989)。贝克特等 (Beckett et al., 1988) 研究了 14 年后 PSID 数据库中那些代表性样本，作者发现 1968 年样本中仅有 40% 的在 1988 年继续存在。然而他们发现，尽管有进出的动态变化，这些数据仍然具有代表性。一种变通的方法是把一定比例的回答者放入每一年的数据中来减少磨损。

(4) 短的时间纬度。

典型的面板数据覆盖的时间段较短，这意味着渐近性主要针对个体数量的无穷大。增加时间长度并不是不花费成本的。随着时间长度的增加，也增加了数据磨损的机会以及有限因变量模型的计算成本。

第2节 面板数据在社会科学中的应用

在社会科学研究中，面板数据很快就替代了以往的横截面研究，这一趋势可以追溯到1960年代，在随后的二十几年中快速发展。面板研究现在在社会科学的研究方法中占据了显著地位。从那些研究大规模社会单位的变化，到福利和贫困（胡珀和斯蒂芬斯（Huber & Stephens），2000；莫勒（Moller）等，2003），甚至到政治暴力（维拉里尔（Villareal），2002）的研究都用到了这种方法。在个体水平的研究也广泛采用这种方法，例如教育，职业选择（巴丁和英格兰（Buding & England），2001），家庭（莫里森和里图劳（Morrison & Ritualo），2000）。在犯罪和行为偏差的研究，就有长期的纵列数据研究传统，在个体层面的分析仍将继续，这一点毫无疑问（黑根和福斯特（Hagan & Foster），2001；奥斯古德（Osgood）等，1996）。在企业层面和组织层面的使用也在不断上升（巴伦（Baron）等，2001；布恩（Boone）等，2002）。

面板数据方法之所以这么受欢迎，主要是它解决了一个数量研究中的核心问题，那就是因果效应的估计。例如沃德佛格等（Waldfoegel et al.，1997）估计了多生一个小孩对母亲工资的影响，彻林等（Cherlin et al.，1998）估计了父母离婚对孩子心理健康的影响。阿塞莫格鲁等（2005）利用跨国的面板数据对教育和民主的因果分析，以及对收入和民主的因果分析，都是非常经典的实证文献。在这些领域，面板模型具有特别的优势。

面板模型在解决因果推断中所涉及的根本问题是处理观测不到的变量，这也是面板数据模型的主要贡献。在非随机试验中所碰到的两种类型的观测不到变量（或隐变量）对因果关系参数进行的确定和估计来说产生了阻碍：（1）一种是时间不变的个体效应^①，它代表个体单位永久性特征；（2）第二种随着时间变化的隐变量，它代表暂时性的影响个体的异质性力量（例如，扰动）。面板数据可以有效地处理各种隐变量问题。

面板数据对因果关系的估计方法在计量领域已经发展了30多年，主要的突破在1970年代，然后在1980~1990年代得到了进一步发展。但是，社会学并没有完全吸收计量经济学的经验教训。相反，社会学家缓慢地使用着面板数据方法。关键点就是时间效应被完全忽略了，盲目地使用检验和估计技术，这需要引起更多的注意。

关于面板数据的经济计量分析教科书有萧政（1986）、巴尔塔吉（2001）、李（Lee，2002）、阿里拉诺（Arellano，2003），这些都是高级版的。伍德里奇（Wooldridge，2002，2003）中、高级教材对这一问题作了很好的阐述。哈拉比（Halaby，2004）对社会学研究中的面板数据模型进行了深入而细致的探讨。

根据他们的整理，我们对1990~2003年间社会科学研究领域所进行的面板数据分析特征列表1-2如下：

^① 在处理时间可变而不随单位变化的隐变量时，面板数据同样有效。