

西安交通大学人文社会科学学术著作出版基金资助

# 函数性数据的分析方法 与经济应用

**Methods of Functional Data Analysis  
and Its Economic Applications**

严明义 著



中国财政经济出版社

西安交通大学人文社会科学学术著作出版基金资助

# 函数性数据的分析方法 与经济应用

Methods of Functional Data Analysis  
and Its Economic Applications

严明义 著

中国财政经济出版社

## 图书在版编目 (CIP) 数据

函数性数据的分析方法与经济应用 / 严明义著. —北京: 中国财政经济出版社, 2014. 9

ISBN 978 - 7 - 5095 - 5639 - 9

I. ①函… II. ①严… III. ①数理统计 - 研究 IV. ①0212

中国版本图书馆 CIP 数据核字 (2014) 第 197400 号

责任编辑: 周桂元

责任校对: 黄亚青

封面设计: 孙俪铭

版式设计: 董生平

中国财政经济出版社 出版

URL: <http://www.cfeph.cn>

E-mail: [jiaoyu@cfeph.cn](mailto:jiaoyu@cfeph.cn)

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100142

营销中心电话: 010 - 88190406 编辑部门电话: 010 - 88190683

北京中兴印刷有限公司印刷 各地新华书店经销

787 × 1092 毫米 16 开 12.75 印张 166 000 字

2014 年 11 月第 1 版 2014 年 11 月北京第 1 次印刷

定价: 26.00 元

ISBN 978 - 7 - 5095 - 5639 - 9/F · 4557

(图书出现印装问题, 本社负责调换)

本社质量投诉电话: 88190744

打击盗版举报热线: 010 - 88190492、QQ: 634579818

# 前 言

与函数性数据分析结缘是 2002 年我访问加拿大滑铁卢大学统计与保险精算系期间, 当时 Jim Ramsay 和 Bernard Silverman 的新作《实用函数性数据分析: 方法和案例研究》将我带入了统计数据分析的一个新奇领域。

“函数性数据”(Functional Data, FD) 的概念由 Ramsay 于 1982 年首次提出, 它是指数据的潜在生成过程为光滑的函数过程时对其进行观测所得的记录。例如, 考古学家挖掘的骨块形状、犯罪者的指纹和足迹、人们书写的笔迹等, 这些样本观察本身就几何图像形式呈现; 另有一些领域的样本资料, 虽然本身并未呈现函数图像的形式, 但其产生过程随时间连续变化的特性也使收集到的观测数据呈现函数性特征, 如人体增高数据、不同地区的多期温度数据、不同时间截面多个行业的就业数据、网上拍卖中竞买者出价数据、不同股票的价格数据等。函数性数据分析(Functional Data Analysis, FDA) 以函数性数据为研究对象, 其基本思想是将观察到的数据看成一个整体进行分析。自 1997 年 Ramsay 和 Silverman 的著作《函数性数据分析》出版以来, 函数性数据的分析方法及其应用受到了国际统计学界及其他相关领域学者的重视, 发表了大量有学术价值的研究文献。

我国在函数性数据方面的研究相对迟缓, 目前尚缺相关中文研究论著。为使我国研究者尽快吸收和掌握函数性数据分析的前沿研究成果, 同时为经济学、管理学、医学、心理学等专业学生提供一本兼顾理论与应用的参考书籍, 我萌发了写作这本书的兴趣。本书主要就函数性数据的相平面图分析方法、多元函数性主成分分析、函数性 Logistic 回归模型、函数性线性回归模型的比



较优势、函数性数据的微分方程和聚类分析方法等进行了研究，对函数性数据分析方法的计算机实现进行了说明，并使用有关方法对网上拍卖等经济问题进行了分析。本书作为国家社会科学基金项目“经济函数性数据分析方法与应用研究”（07XTJ001）的研究成果，部分内容已经公开发表。

本书的最终完成是许多热心人支持与帮助的结果，在本书出版之际，我首先感谢西安财经学院原党委书记刘延年教授、西安交通大学经济与金融学院院长冯根福教授、中国人民大学统计学院何晓群教授，是他们渊博的学识、敏锐的洞察力和无私的支持与帮助，使我坚定了进一步深入研究函数性数据的信心！感谢我的博士生导师，西安交通大学经济与金融学院邱长溶教授的悉心指导和鼓励，使我有力量克服方法应用的计算机编程与网上拍卖数据收集方面的困难，最终将函数性数据的分析方法应用于网上拍卖等经济问题的分析中。感谢加拿大 McGill 大学 Ramsay 教授，他对我在函数性数据分析方法方面的工作给予了许多指导；感谢美国 Maryland 大学史密斯商学院 Galit Shmueli 副教授，她解释了我在网上拍卖数据分析方面的许多疑惑。另外，在本书的编写过程中，我的硕士研究生杜鹏、纪礼文、赵霓、魏婧、蒲涇涇、党敏等做了不少工作。中国财政经济出版社周桂元主任在审稿、出版环节付出了辛勤的汗水，在此向他们一并表示感谢！

作为国家社会科学基金项目的研究成果，本书是我多年来有关函数性数据与经济应用研究的一个总结，同时也仅仅是该方面研究的开始，书中所述肯定存在诸多不足，欢迎广大读者批评指正！

最后，感谢西安交通大学人文社会科学学术著作出版基金的资助！

2014年9月于西安交通大学

# 目 录

## 第一章 绪 论

|                      |    |
|----------------------|----|
| 1.1 选题背景和意义          | 2  |
| 1.1.1 选题背景 / 2       |    |
| 1.1.2 研究意义 / 2       |    |
| 1.2 国内外研究现状          | 4  |
| 1.2.1 国外研究现状 / 4     |    |
| 1.2.2 国内研究现状 / 9     |    |
| 1.2.3 国内外研究文献评析 / 10 |    |
| 1.3 主要研究内容、基本思路和方法   | 12 |
| 1.3.1 主要研究内容 / 12    |    |
| 1.3.2 研究的基本思路 / 14   |    |
| 1.3.3 研究方法 / 14      |    |
| 1.4 研究的重点难点、主要观点与创新点 | 15 |
| 1.4.1 研究的重点和难点 / 15  |    |
| 1.4.2 主要观点和创新点 / 16  |    |

## 第二章 函数性数据及其特征：以网上拍卖为例

|             |    |
|-------------|----|
| 2.1 拍卖与网上拍卖 | 20 |
|-------------|----|

|       |                     |    |
|-------|---------------------|----|
| 2.1.1 | 拍卖及其机制 / 20         |    |
| 2.1.2 | 网上拍卖及其与传统拍卖的差异 / 22 |    |
| 2.2   | 网上拍卖中竞买者的出价记录       | 25 |
| 2.2.1 | 出价记录 / 25           |    |
| 2.2.2 | eBay 网站上的拍卖品数量 / 27 |    |
| 2.3   | 网上拍卖中竞买者出价数据的特征     | 29 |
| 2.3.1 | 数据容量 / 29           |    |
| 2.3.2 | 数据的混合性 / 30         |    |
| 2.3.3 | 数据不等间隔分布 / 31       |    |
| 2.3.4 | 数据密度不均匀 / 33        |    |
| 2.4   | 网上拍卖数据的函数性数据分析方法    | 34 |

### 第三章 函数性数据的修匀

|       |                       |    |
|-------|-----------------------|----|
| 3.1   | 引言                    | 38 |
| 3.1.1 | 函数性数据的结构与分析假设 / 38    |    |
| 3.1.2 | 观察误差和抽样 / 39          |    |
| 3.1.3 | 函数性数据分析的假设条件 / 40     |    |
| 3.2   | 数据修匀的基函数方法            | 40 |
| 3.2.1 | 利用基函数表示函数 / 41        |    |
| 3.2.2 | 描述周期性数据的傅立叶基函数系统 / 42 |    |
| 3.2.3 | 样条函数与 B-样条基函数 / 43    |    |
| 3.3   | 数据修匀的粗糙惩罚方法           | 44 |
| 3.3.1 | 粗糙惩罚修匀的思想 / 45        |    |
| 3.3.2 | 粗糙惩罚调整修匀程度的机理 / 46    |    |

### 第四章 函数性数据的描述性分析

|     |             |    |
|-----|-------------|----|
| 4.1 | 函数性数据的概括统计量 | 48 |
|-----|-------------|----|

|          |    |
|----------|----|
| 4.2 实例分析 | 49 |
|----------|----|

## 第五章 函数性数据的相平面图分析法

|                                  |    |
|----------------------------------|----|
| 5.1 相平面图                         | 56 |
| 5.2 基于相平面图的消费价格指数季节变动诊断          | 57 |
| 5.2.1 样本数据 / 57                  |    |
| 5.2.2 中国消费价格指数的季节变动特征 / 57       |    |
| 5.2.3 相平面图对历史事件信息的呈现 / 61        |    |
| 5.3 基于相平面图的竞买者出价水平动态演变模式研究       | 63 |
| 5.3.1 竞买者出价数据的函数性特征 / 64         |    |
| 5.3.2 竞买者出价水平动态演变的一般模式 / 65      |    |
| 5.3.3 不同特征拍卖中竞买者出价水平相平面图的比较 / 66 |    |
| 5.3.4 中国拍卖网站上竞买者出价水平的动态演变特征 / 70 |    |

## 第六章 函数性主成分分析 (FPCA)

|                              |    |
|------------------------------|----|
| 6.1 函数性主成分的数学模型              | 75 |
| 6.2 函数性主成分的求解方法              | 76 |
| 6.2.1 函数性主成分的特征方程 / 76       |    |
| 6.2.2 权函数的获取方法 / 77          |    |
| 6.2.3 函数性主成分权函数的修匀处理 / 80    |    |
| 6.3 多元函数性主成分分析               | 81 |
| 6.3.1 多元函数性主成分 / 81          |    |
| 6.3.2 获取权函数的方法 / 83          |    |
| 6.4 房屋销售价格指数与居民消费价格指数的联合变异分析 | 85 |
| 6.4.1 描述性分析 / 86             |    |
| 6.4.2 联合变异分析 / 88            |    |





## 第七章 函数性方差分析与函数性 Logistic 回归

|       |                                |     |
|-------|--------------------------------|-----|
| 7.1   | 函数性方差分析                        | 96  |
| 7.2   | 函数性方差分析的经济应用                   | 98  |
| 7.2.1 | 第三产业行业就业人数差异分析 / 98            |     |
| 7.2.2 | 竞买者出价行为的拍卖品类别差异分析 / 100        |     |
| 7.3   | 函数性 Logistic 回归模型              | 101 |
| 7.3.1 | 函数性广义线性模型 / 101                |     |
| 7.3.2 | 函数性 Logistic 模型——单函数性自变量 / 102 |     |
| 7.3.3 | 函数性 Logistic 模型——多函数性自变量 / 108 |     |
| 7.4   | 基于函数性 Logistic 回归的商品房价影响因素分析   | 109 |
| 7.4.1 | 指标选取及数据来源 / 109                |     |
| 7.4.2 | 中国商品房价涨跌影响因素分析 / 110           |     |

## 第八章 函数性线性回归模型的比较优势

|       |                      |     |
|-------|----------------------|-----|
| 8.1   | 模型参数函数展示信息的优势        | 118 |
| 8.1.1 | 变量选取与数据收集 / 118      |     |
| 8.1.2 | 模型构建与参数函数估计 / 119    |     |
| 8.1.3 | 参数估计结果与分析 / 120      |     |
| 8.2   | 函数性线性回归模型的预测优势       | 125 |
| 8.2.1 | 指标选择与数据获取 / 126      |     |
| 8.2.2 | 基于函数性线性回归模型的分析 / 127 |     |
| 8.2.3 | 基于其他模型的分析 / 130      |     |

## 第九章 函数性数据的微分方程分析方法

|     |               |     |
|-----|---------------|-----|
| 9.1 | 函数性数据与微分方程的关系 | 134 |
|-----|---------------|-----|

|       |                                 |     |
|-------|---------------------------------|-----|
| 9.2   | 微分方程系数函数的估计                     | 136 |
| 9.2.1 | 逐点最小化法 / 136                    |     |
| 9.2.2 | 共发函数性线性模型方法 / 137               |     |
| 9.3   | 利用微分方程分析经济增长及其与投资率的关系           | 139 |
| 9.3.1 | 中国经济增长分析 / 140                  |     |
| 9.3.2 | 中国 31 个省、自治区、直辖市经济增长的差异分析 / 142 |     |
| 9.3.3 | 中国、美国及日本经济增长的比较分析 / 143         |     |
| 9.3.4 | 投资率与经济增长的关系 / 144               |     |

## 第十章 函数性数据的聚类方法与案例分析

|        |                         |     |
|--------|-------------------------|-----|
| 10.1   | 一般函数性数据的相异性度量方法         | 149 |
| 10.1.1 | 基于直接距离的函数性数据相异性度量 / 149 |     |
| 10.1.2 | 基于降维的函数性数据相异性度量方法 / 151 |     |
| 10.1.3 | 基于函数曲线极值点的相异性度量 / 154   |     |
| 10.2   | 特殊函数性数据的相异性度量方法         | 156 |
| 10.2.1 | 稀疏不规则数据的相异性度量 / 156     |     |
| 10.2.2 | 空间相关函数性数据的相异性度量 / 158   |     |
| 10.3   | 函数性数据相异性度量方法的优劣分析       | 159 |
| 10.4   | 网上拍卖竞买者出价行为的函数性聚类分析     | 161 |
| 10.4.1 | 研究背景 / 161              |     |
| 10.4.2 | 样本选择与数据的收集整理 / 162      |     |
| 10.4.3 | 函数性数据聚类方法选择 / 163       |     |
| 10.4.4 | 竞买者出价行为的函数性聚类分析 / 166   |     |

## 第十一章 函数性数据分析方法的计算机实现

|        |                      |     |
|--------|----------------------|-----|
| 11.1   | 现有分析函数性数据的程序语言       | 172 |
| 11.1.1 | 函数性数据分析程序应用的语言 / 172 |     |



|        |                                  |     |
|--------|----------------------------------|-----|
| 11.1.2 | 面向对象编程中的两个基本概念 / 173             |     |
| 11.2   | 基于 MATLAB 的函数性数据分析方法的计算机实现       | 174 |
| 11.2.1 | Windows 系统中 MATLAB 函数的安装步骤 / 174 |     |
| 11.2.2 | 分析实例安装 / 175                     |     |
| 11.2.3 | 函数性数据的分析过程 / 175                 |     |
| 附录 1   | 第三产业九类行业的主成分分析结果                 | 177 |
| 附录 2   | 竞买者出价水平导数的主成分分析结果                | 179 |
| 参考文献   |                                  | 183 |

---

# 第一章

## 绪 论

---

本章首先说明了该书研究的选题背景和意义，其次对国内外研究文献进行了梳理和评析。再其次，给出了该书的主要研究内容，研究的基本思路和方法。最后，阐述了研究的重点难点、主要观点与创新点。



## 1.1

### 选题背景和意义

#### 1.1.1 选题背景

随着现代数据收集及贮存技术的迅速发展,研究者收集数据的方法和获取的资料形式正在发生较大的变化。许多领域收集的样本资料本身就以函数图像的形式呈现,如考古学家挖掘的骨块形状、医学检测中患者的心电图及头部影像、人们书写的笔迹、犯罪者的指纹和足迹等。有些领域的样本资料,虽然本身并未呈现函数图像的形式,但其随着时间变化也使收集的观测数据显示出函数性特征,如人体增高数据、不同地区的多期温度与降雨量数据、不同时间截面多个行业的就业数据、网上拍卖中竞买者出价数据、不同股票的价格数据等。为什么我们会观测到具有这种特征的数据?因为这些样本数据的产生过程是一个函数过程。函数性数据分析以函数性数据为研究对象,其基本思想是将观察到的数据看成一个整体。Ramsey (1982) 在其论文中将具有这种特征的数据首次命名为函数性数据 (functional data, FD)。由于函数性数据不同于传统统计方法所分析的时间序列数据 (time series data) 和横截面数据 (cross-sectional data), 也不同于计量经济学中讨论的面板数据 (panel data), 因此, 传统的数据分析方法在具体应用中受到了挑战。

#### 1.1.2 研究意义

由于实际中获得经济数据的方式和技术日新月异、多种多样, 越来越

多的研究者获得的样本观察资料呈现函数性特征。因此,对函数性数据的分析方法及其在经济领域的应用进行研究,具有重要的理论价值和应用价值。

### (1) 理论价值

国外学者对函数性数据的分析方法进行了较为深入地研究,取得了大量富有价值的研究成果。他们在离散数据的函数化处理、粗糙惩罚法、函数注册或列齐(套准)、函数性数据的聚类分析与典型相关分析、函数性数据的假设检验、函数性数据的微分方程方法等方面做了许多研究,尤其在函数性数据的主成分分析、函数性线性回归模型两方面的研究更多、更深、更广,取得的成果也最多。但是,从目前的研究现状来看,有关函数性数据的聚类分析方法缺乏系统的研究文献,对分析函数性数据时微分方程(算子)的设定、多元函数性主成分分析的模型和权函数的估计方法讨论地较少,对函数性 Logistic 回归模型和应用的分析相对薄弱,对函数性线性回归模型的比较优势缺乏分析。另外,国内鲜有文献就函数性数据分析方法的计算机实现进行说明。因此,对这些问题进行研究,能够充实函数性数据的分析方法和应用,具有明显的理论价值。

### (2) 应用价值

在国外学者关于函数性数据分析方法的研究中,所涉及的数据主要来源于生物、医学以及工程领域,极少考虑经济领域的函数性数据。与自然科学中收集的数据相比,经济数据具有收集过程难以控制的特点,致使数据误差较大。调研获得的微观数据,往往具有较大的误差,如家庭可支配收入、消费等。另外,虽然有些金融数据,如股票价格和汇率等,具有较小的测量误差,但有些金融数据是高频数据或超高频数据,从而数据包含的“噪音”会更多一些。因此,不能将函数性数据的分析方法直接、简单地应用于经济函数性数据的分析中,而应在应用中充分考虑经济函数性数据的具体特征,从经济指标的选取、数据的收集、结果的经济解释等方面予以综合考虑,对方法不应只是简单地应用,而应将方法所得结果是否与实际经济预期一致,是否能够利用经济理论对其进行合理解释等作为判断方法使用效果的一条主要标准。



我国在这方面的研究起步较晚，真正引起统计学界重视大约始于 2007 年。虽然我国学者也做了一些工作，但一方面缺乏对真正不同于传统数据形式的经济函数性数据的分析，未能真正显示函数性数据分析方法在经济研究中的优越性；另一方面，对结果的经济分析不够深入。另外，函数性数据的分析方法必须借助于计算机实现，而现有的 SPSS、EViews、SAS 等统计软件不能直接用于函数性数据的分析，因此，对研究人员使用函数性数据的分析方法带来了极大地不便，对于初学者更是困难重重，经常花费大量的时间用于方法的计算机实现，限制了该方法的应用推广。所以，对这些问题进行研究，能够丰富函数性数据分析方法的应用领域，对利用此方法解决实际经济问题大有裨益，因而本研究具有广阔的应用前景和重要的应用价值。

## 1.2

### 国内外研究现状

#### 1.2.1 国外研究现状

函数性数据的概念由加拿大统计学家，McGill 大学心理学教授，加拿大统计学会前主席 Jim Ramsay 于 1982 年首次提出。1991 年，Ramsay 和 Dalzell 在其发表的论文“Some Tools for Functional Data Analysis (with Discussion)”中提出了对函数性数据进行分析的一些方法。六年后，Ramsay 和 Bristol 大学统计学教授 Bernard Silverman（曾获统计学会会长联合委员会颁发的会长奖和两枚皇家统计学会 Guy 奖章），将对函数性数据进行统计分析的已有理论和方法总结在《函数性数据分析》一书中（Ramsay 和 Silverman, 1997），但这本书偏重方法的理论介绍和数学推导，不利于统计基础薄弱者使用。经过五年的努力，Ramsay 和 Silverman 研究了一些函数性

数据案例,并将其具体的分析过程编入他们于2002年出版的专著中。2005年,Ramsay和Silverman对1997年版的《函数性数据分析》进行了修订,出版了《函数性数据分析》第二版,书中吸收了国际上最新研究成果,特别扩充了样条修匀和函数性线性模型方面的内容,对处理函数性数据的微分方程方法进行了改写,我国科学出版社于2006年将其作为国外数学名著引入国内出版。经过三十多年的发展,关于函数性数据的分析取得了许多成果,特别是继Ramsay和Silverman的著作出版后,国际上掀起了一股函数性数据研究热,取得了许多有价值的研究成果,并在经济、生物、医学以及气象等领域被广泛应用。

经过我们对相关研究文献的梳理,国外学者关于函数性数据的研究可以概括为以下七个方面:

### (1) 离散观察数据的函数化处理

在函数性数据分析中,首要的工作是将观测到的离散数据转化为一个函数。关于插值和修匀,国外学者做了大量深入的研究,例如,Schumaker (1981)、Wahba (1990)、Hastie和Tibshirani (1990)就样条函数进行了深入的讨论;Green和Silverman (1994)就样条函数在数学和统计学中的应用进行了研究;Simonoff (1996)和Eubank (1999)就修匀和非参数回归进行了讨论;Silverman (1999, 2000)以及Silverman和Vassilicos (1999)就函数性数据分析中的小波基进行了研究。虽然,关于样条函数不少文献进行了讨论,但系统且深入地论述见de Boor (2001)。Ramsay和Silverman (2006)就B-样条基函数进行了易于掌握的讨论,其针对入门者的知识介绍见函数性数据分析网站 [www.functionaldata.org](http://www.functionaldata.org)。

### (2) 函数性数据的注册 (registration) 或特征列齐 (alignment)

注册问题研究中的一个重要概念是时间变形函数 (time warping function), Sakoe和Chiba (1978)运用动态规划的方法估计了时间变形函数;Kneip和Gasser (1992)、Gasser和Kneip (1995)深入研究了标志注册 (landmark registration),他们将标志称为结构性特征 (structural feature),将位置称为结构点 (structural point); Ramsay (1996)、Ramsay和Li (1996)利用注册技术发展了一类变形函数的拟合;Gervini和Gasser





(2004) 进一步发展了注册技术, 用以解决不能使用标志方法带来的问题, 并考虑了一些重要的理论问题。

### (3) 函数性数据分析中的假设检验问题

Robinson (1989) 提出了半参数检验方法, 这类检验的优势在于可以将这些检验方法扩展到在原假设下函数形式不确定的情况, 因此它们具有一般性, 然而半参数假设检验方法依赖于一系列严格的假设, 如时间序列的平稳性等, 因此该类检验存在一些问题, 如参数估计局限性、维数问题及样本容量选取等。为解决这个问题, Bewley 和 Griffiths (2001)、Montgomery 和 Bradlow (1999) 提出了贝叶斯假设检验方法, 贝叶斯假设检验方法允许加入先验信息, 但当样本容量较小时该方法在估计模型的置信区间时存在一些不足。

在对函数性数据进行分析时, Fan 和 Li (1998) 对曲线数据的显著性检验问题进行了研究。检验均值函数的一个难点在于估计协方差的形式, Bealey 和 Griffiths (2001) 提出了一种方法, 但假设误差项是独立同分布的。由于实际中, 各次观察的误差项之间通常存在一定的相关性, 因此需要将误差项设为较一般的形式; Games 和 Sood (2006) 对这一情况进行了研究, 他们先考虑了从同一均值函数观测到数条曲线情形下的检验方法, 然后将该方法扩展到更复杂、更一般的情形。在具体分析时, 他们对观测曲线和估计的均值函数间的残差序列使用 bootstrap 方法给出了检验过程, 并用其分析了技术进步曲线; Zhang 和 Chen (2007) 基于函数性线性模型和核估计对函数性数据的统计推断问题进行了研究, 提出了一种检验方法。

### (4) 函数性数据的变异性分析

为了对函数性数据的变异性进行分析, 国外学者就函数性主成分分析 (FPCA) 和函数性典型相关分析 (FCCA) 进行了研究。

关于函数性数据的主成分分析, 国外早有文献关注, 取得的成果也颇丰。Dauxois, Pouuse, 和 Romain (1982) 研究了向量随机函数主成分的渐进理论, 提出了函数性主成分分析方法, 这一方法也成为后来很多函数性数据分析方法发展的基础; Ramsay 和 Dalzell (1991) 讨论了使用 FPCA 分