

学校编码: 10384

分类号 _____ 密级 _____

学号: 14220051402869

UDC _____

廈門大學

博 士 学 位 论 文

函数性数据分析方法及应用研究

A Study on the Methods of Functional Data Analysis and
Their Application

靳刘蕊

指导教师姓名: 朱建平教授

专业名称: 统计学

论文提交日期: 2007年3月

论文答辩日期: 2008年 月

学位授予日期: 2008年 月

答辩委员会主席: _____

评 阅 人: _____

2008年3月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

兹提交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（ ），在 年解密后适用本授权书。

2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名： 日期： 年 月 日

导师签名： 日期： 年 月 日

厦门大学博硕士学位论文摘要库

摘要

函数性数据分析是把观测数据 $y_i = (y_{i1}, \dots, y_{iT_i})'$ 看成一个整体, 表示为光滑曲线或连续函数 $u_i(t)$, $i = 1, 2, \dots, N$, 然后从函数视角对其进行分析。与传统的分析方法相比, 函数性数据分析具有其自身的优越性, 它依赖较少的假设条件和较弱的结构约束, 不要求不同观测对象的数据观测点和观测次数相同, 不但可以实现对无限维度数据(即这里所谓的函数性数据)的模式挖掘, 而且还可用于对非函数性数据进行分析。函数性数据分析的一个重要的特性是可以求出导数曲线或微分曲线, 通过对它们的分析能够挖掘出更多的重要信息, 例如对一阶或高阶导数曲线的分析可探索曲线之间的差异和曲线内部动态变化模式等。

面向函数性数据的挖掘已引起国内外学术界的关注, 并取得了丰硕的研究成果, 在增长分析、气象学、生物力学、经济学、医学等许多领域具有广阔的应用前景。国内外关于函数性数据分析的研究还处于起步阶段, 还有许多问题需要进一步研究; 从应用领域来看, 极少用于对经济数据的分析研究。论文以“函数性数据分析方法及其应用研究”为题, 从思想、理论和技术方法等方面对函数性数据分析方法进行了系统概括, 主要包括函数性数据的预处理、线性回归分析、主成份分析、聚类分析、典型相关分析和判别分析等, 并在此基础上进行方法及其在经济数据分析中的应用研究。

论文的创新之处主要体现在以下几个方面:

(1) 提出基于回归系数函数的函数线性模型。与三种基本的函数线性模型相比, 这种方法提高了模型的可估计性, 不仅可以从动态角度对经济现象进行传统的分析预测, 而且可以从连续的视角揭示经济现象发展变化的速度、加速度等更深层次的动态规律性, 体现出了函数性数据分析的实用性。

(2) 提出一种基于里程碑的、同时根据序列数值和序列形状模式进行相异性测度的聚类方法。这种方法克服了基于序列数值模式测度相似性的聚类方法不考虑轨迹形状、而基于序列形状模式测度相似性的聚类方法忽略了序列数值所代表的相似信息和趋势信息的缺陷, 在高频金融时间序列的聚类分析中具有现实意义。

(3) 对函数性数据分析方法在经济数据分析中的应用进行研究和拓展。将

函数性回归分析、主成份分析、聚类分析和典型相关分析分别用于研究城镇居民消费倾向的发展变化趋势、国家财政主要支出项目的主要变化模式、构成上证 50 指数的股票的高频金融时间序列聚类问题和我国 29 个地区城镇居民家庭人均可支配收入与消费性支出的共变模式等。

(4) 在将函数性数据分析方法运用于经济数据分析的过程中, 挖掘出一些用传统分析方法不易识别的经济规律。例如, 城镇居民家庭人均消费的回归分析结果显示, 边际消费倾向的变化在时间上稍微滞后于基本消费的变化, 表明对于收入的变化反应, 边际消费倾向变化的灵敏性要弱于基本消费变化的灵敏性; 城镇居民家庭人均可支配收入与消费性支出的典型相关分析显示, 在可支配收入水平较高的地区, 消费性支出的变化不完全是收入变化的结果, 消费水平的提高可能会促使收入的增加, 因此扩大消费品的种类和提高消费品的质量可能促使该地区的内需增加; 在可支配收入处于中等和较低水平的地区, 消费性支出的变化主要受收入变化的影响, 因此提高居民的收入水平能促使消费的增加, 从而促使该地区的内需增加等。

关键词: 函数性数据分析; 平滑; 正则化

ABSTRACT

In functional data analysis, data are analyzed from a functional angle with viewing the data as an entirety, which is shown as a smooth curve or continuous functions. Compared with the traditional analyzing methods, functional data analysis has its own advantages, it relies relatively on less assumed conditions and looser structural limit, does not require that different observed objects have the same observing point and times, not only can be used to analyze infinite dimensional data, but also can be used to analyze non-functional data. An important characteristic of functional data analysis is that derivative curves or differential curves can be found and used to mine more important information about the data.

The research in this field has just began and there are still many problems needing us to pay more attention and efforts. What's more, this analysis method is seldom used in economic data analysis. This paper introduces systematically the method for functional data analysis and then conducts a research into the method and its application in economic data analysis.

The thesis is entitled *A Study on the Methods of Functional Data Analysis and Their Application*, probes systematically into the methods of functional data analysis, including functional linear models, functional principle components analysis, functional cluster analysis, functional canonical correlation and discriminant analysis, and then apply the methods to analyze economic data.

The creative points of this thesis can be viewed in the following aspects:

(1) The thesis puts forward the functional linear model based on regression parameter functions. Compared with three basic functional linear models, the method improves the computability of the functional linear models, which can be used to not only analyze and forecast economic phenomenon from a dynamic angle in a traditional way, but also discover deeper information behind dynamic laws such as the speed and the acceleration of the developing and changing of economic phenomenon in continuous point of view, which presents the practical applicability of functional data analysis.

(2) The thesis puts forward a new cluster method which measuring similarity based on milestones and taking both sequence values and shapes into account. The method overcomes the disadvantage of the cluster methods based on sequence values ignoring the trajectory shapes and of the cluster methods based on sequence shapes neglecting sequence values which involve information about similarities and trends between curves. The cluster method has practical significance in the cluster analysis of high-frequency financial time series.

(3) The thesis studies on the application of functional data analysis in economic data analysis. The writer applies functional regression into investigating the developing and changing trends of China's urban expenditure tendency; functional principle components analysis into analyzing the main varying modes of the government expenditure by main item; functional cluster analysis into investigating the clustering problem of high-frequency financial time series; functional canonical correlation analysis into analyzing covarying modes between per capita annual income and living expenditure of urban households.

(4) The thesis finds some economic laws, which are not easily identifiable using traditional analysis methods, in the process of applying functional data analysis into economic data. For example, the variation of marginal per capita annual expenditure lags that of income of urban households; the increase of per capita annual living expenditure may spur the increase of income in the regions where per capita annual income of urban households is higher, the increase of income may spur the increase of living expenditure in the regions where per capita annual income of urban households is middling or lower etc.

Key Words: Functional Data Analysis; Smoothing; Regularizing

目录

第 1 章 绪论	1
1.1 选题背景和研究意义.....	1
1.2 函数性数据的特征、分析目的和步骤	2
1.3 国内外相关研究现状.....	6
1.4 论文的研究内容和创新.....	11
第 2 章 函数性数据的预处理	14
2.1 函数性数据的函数表示问题	14
2.2 基函数法.....	16
2.3 局部加权平滑法.....	23
2.4 粗糙惩罚法.....	26
2.5 函数性数据的套准和显示	32
第 3 章 函数性线性模型及其应用	38
3.1 仅响应变量为函数性变量的函数线性模型	38
3.2 仅自变量为函数性变量的函数线性模型	43
3.3 响应变量和自变量均为函数性变量的函数线性模型	47
3.4 基于回归系数函数的函数线性模型	52
3.5 我国城镇居民消费函数的动态分析	56
第 4 章 函数性主成份分析及其应用	60
4.1 一元函数性主成份分析.....	60
4.2 多元函数性主成份分析.....	68
4.3 正则化的函数性主成份分析	70
4.4 混合数据的主成份分析.....	76
4.5 我国财政主要支出项目的主成份分析	80
第 5 章 函数性聚类分析及其应用	86
5.1 基于曲线数值模式的聚类方法	86
5.2 基于轨迹形状模式的非参数聚类法	88

5.3 同时测度序列数值模式和形状模式的函数聚类法	92
5.4 金融时间序列的函数性聚类分析	98
第 6 章 函数性典型相关和判别分析及其应用	105
6.1 经典的典型相关分析.....	105
6.2 函数性典型相关分析.....	107
6.3 函数性典型相关分析算法	111
6.4 惩罚最优得分和判别分析	113
6.5 我国城镇居民家庭人均可支配收入与消费性支出的典型相关分析	115
第 7 章 研究总结与展望	124
7.1 研究总结.....	124
7.2 论文不足与研究展望.....	125
参考文献.....	127
后记.....	140

Contents

Chapter 1 Introduction.....	1
1.1 The background and significance of the study	1
1.2 The characteristics, goals and process of functional data analysis	2
1.3 Status quo of research on functional analysis abroad and at home.....	6
1.4 The contents and innovations of the thesis	11
Chapter 2 Pre-processing of functional data	14
2.1 Problems of representing functional data as functions	14
2.2 Basis function methods	16
2.3 Local weighting smoothing methods	23
2.4 Roughness penalty approach.....	26
2.5 The registration and display of functional data.....	32
Chapter 3 Functional linear models and their application	38
3.1 Linear models in which only responses are functional.....	38
3.2 Linear models in which only covariates are functional	43
3.3 Linear models in which both responses and covariates are functional	47
3.4 Linear models based on regression parameter functions	52
3.5 Dynamic analysis of China's urban consumption.....	56
Chapter 4 Functional principle components analysis and their application	60
4.1 Univariate functional principle components analysis	60
4.2 Multivariate functional principle components analysis	68
4.3 Regularized functional principle components analysis.....	70
4.4 Functional principle components analysis of mixed data	76
4.5 Principle components analysis of China's financial outlay	80
Chapter 5 Functional cluster analysis and their application	86
5.1 Functional cluster methods based on curve values	86
5.2 Functional nonparametric cluster methods based on curve shapes	88
5.3 Functional cluster methods based on curve values and shapes.....	92
5.4 Functional cluster analysis for financial time series	98

Chapter 6 Functional canonical correlation and discriminant analysis and their application	105
6.1 Classical canonical correlation analysis.....	105
6.2 Functional canonical correlation analysis.....	107
6.3 Algorithms of functional canonical correlation	111
6.4 Functional penalized optimal scoring and discriminant analysis	113
6.5 Canonical correlation analysis of per capital annual disposable income and living expenditure of China's urban households.....	115
Chapter 7 Conclusions and Further Research.....	124
7.1 Conclusions.....	124
7.2 Deficiencies and Further Research	125
References.....	127
Postscript	140

第1章 绪论

1.1 选题背景和研究意义

在传统的统计数据分析中，数据一般是时间序列数据或横截面数据。随着信息技术的发展，人们获取和存储数据的能力得到了极大的提高，大量的形式多样的复杂类型的数据不断涌现。在现实生活中，人们需要处理越来越多的具有函数特征的数据（简称为函数性数据）。例如，某地区或多个地区气象站测量的多年的日降雨量数据和气温数据；某条或多条高速公路上传感器网络每天收集并存储的车流量数据；多个地区的月度或年度国内生产总值和外贸交易额数据；证券市场产生的多只股票的分时或日均成交价、收盘价、涨跌幅、交易量、交易额等数据；银行的各分支机构或银行客户多年的日或月交易量、交易额、交易类型等交易数据等。

传统的数据分析中，我们把这些数据看成离散的且有限的，是单个观测值的顺序排列。计量经济学中称单个观测对象的这种数据为时间序列数据，多个观测对象的这种数据为面板数据（Panel Data）或时空数据（Longitudinal Data）。对于时间序列数据的研究早已有丰硕的成果和比较成熟的理论体系。自 Mundlak（1961）、Balestra 和 Nerlove（1966）把面板数据引入到经济计量中之后，近二十多年以来它成为经济计量学的一个研究热点，面板数据能够缓解数据样本容量不足的问题，并具有控制个体行为差异、识别难以度量因素的影响等诸多优点。但采用计量经济模型对时间序列数据或面板数据分析依赖很多的假设条件，有较强的结构约束。此外，传统的数据分析方法忽视了这些数据的部分函数特征，且一般要求数据的观测点和观测次数相同，而在实际中收集到的数据很多时候并不满足这个条件。

函数性数据分析（FDA）是从函数角度对数据进行分析，把函数性数据看作一个整体，表示为光滑曲线或连续函数 $u_i(t)$ ， $i=1, \dots, N$ ，其中， N 表示观测数据的个数； t 可以表示时间，也可以表示时间以外的其它变量。对于所有的观测

对象 $i=1, \dots, N$ ，不要求它们的观测点 t_{ij} 和观测次数 T_i 相同。

与传统的分析方法相比，函数性数据分析具有其自身的优越性，主要体现在五个方面：第一，函数性数据分析方法可对无限维度空间的数据（即这里所谓的函数性数据），例如连续性数据或流式数据等进行分析，实现数据模式的挖掘；第二，相较于传统的数据分析方法，函数性数据分析方法依赖较少的假设条件和较弱的结构约束；第三，不要求不同观测对象的数据观测点和观测次数相同；第四，通过对导数曲线或微分曲线的分析能够挖掘出数据更多的重要信息，例如通过对一阶或高阶导数曲线的分析探索曲线之间的差异和曲线内部动态变化模式等；第五，可以对非函数性数据采用函数性数据分析方法进行分析。

面向函数性数据的挖掘已引起国内外学术界的关注，并取得了很多的研究成果，在增长分析、气象学、生物力学、经济学、医学等许多领域具有广阔的应用前景。但关于函数性数据分析的研究仍处于起步阶段，许多问题还需要做进一步的研究；此外，从目前函数性数据分析方法应用的领域来看，极少涉及对经济数据的分析研究。因此，对函数性数据分析方法及其在经济数据分析中的应用的探索和研究具有重要的理论意义和现实意义。

1.2 函数性数据的特征、分析目的和步骤

1.2.1 函数性数据的基本特征

函数性数据分析是从函数角度对数据进行分析，把函数性数据看成一个整体，表示为光滑曲线或连续函数 $u_i(t)$ ， $i=1, \dots, N$ ，其中 N 表示观测对象（曲线）的个数， t 可以表示时间，也可以表示时间以外的其它变量。术语“函数”是指在数据的内在结构，而非数据的外在形式。在实际操作中，我们观测和记录的函数性数据通常是离散的。当 t 表示时间时，在 T 个时刻上对一个观测对象进行观测得到的 T 个数据点 $y(t_j)$ ($j=1, 2, \dots, T$) 构成一个序列，传统分析中我们称其为时间序列；对 N 个观测对象进行观测得到的数据资料 $y_i(t_j)$

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库