

国内图书分类号: O212.4

国际图书分类号: 519.2

学校代码: 10213

密级: 公开

理学硕士学位论文

函数型数据分析方法在 GDP 研究中的应用

硕 士 研 究 生: 刘宝宇

导 师: 方 茹教授

申 请 学 位: 理学硕士

学 科: 概率论与数理统计

所 在 单 位: 数学学院

答 辩 日 期: 2020 年 6 月

授予学位单位: 哈尔滨工业大学

Classified Index: O212.4

U.D.C: 519.2

Dissertation for the Master Degree in Science

APPLICATION OF FUNCTIONAL DATA ANALYSIS METHOD IN GDP RESEARCH

| | |
|---------------------------------------|---|
| Candidate: | Liu Bao Yu |
| Supervisor: | Prof. Fang Ru |
| Academic Degree Applied for: | Master of Science |
| Speciality: | Probability Theory and Mathematical Statistics |
| Affiliation: | School of Mathematics |
| Date of Defence: | June, 2020 |
| Degree-Conferring-Institution: | Harbin Institute of Technology |

摘 要

计算机技术的迅速发展使得人们收集到的数据量越来越大，经典的多元统计分析方法在处理这种数据时有一定的限制，函数型数据分析方法因此被提出。国内生产总值 (GDP) 是衡量一个国家或地区经济状况和发展水平的重要指标。本文主要研究函数型数据分析方法的理论，并对我国 GDP 分省数据作实证分析。

首先，介绍了函数型数据预处理的几种方法并分别对我国 GDP 分省年度数据和分省季度数据作实证分析。研究表明对于具有循环特征的数据和不具有循环特征的数据，B 样条基都具有相对较好的拟合效果。此外对基于 B 样条基的最小二乘拟合方法进行了改进，使它可以应用到稀疏函数型数据的处理中。

其次，对多元统计主成分分析方法和函数型主成分分析方法进行了探讨，这两种方法的实证分析结果都表明影响我国各省 GDP 均值波动的主要因素有两方面，其中经济大省的拉动作用是主要影响因素。

最后，探讨了系统聚类分析和函数型聚类分析两种聚类分析方法，并且提出了计算函数型变量之间距离的曲线夹角余弦法。这两种方法的实证分析结果表明系统聚类分析方法和函数型聚类分析方法虽然从不同的角度处理数据，但是得到的结果有某些共性。结果显示我国各省份的 GDP 总量以及发展潜力可以按省份分成不同的类别，经济相对落后的省份发展空间反而大，国家应加大对这些省份的政策支持。

关键词：函数型数据；基函数拟合；函数型主成分分析；函数型聚类分析；GDP

Abstract

With the rapid development of computer technology, people have collected more and more data, which makes the classical multivariate statistical analysis methods be limited. Therefore, functional data analysis methods have been proposed. Gross Domestic Product (GDP) is an important indicator to measure the economic status and development level of a country or region. This thesis mainly studies the theory of functional data analysis method and makes an empirical analysis of China's provincial GDP data.

First of all, it introduces several methods of functional data preprocessing and makes empirical analysis on China's provincial annual GDP data and provincial quarterly data respectively. Research shows that B-spline basis has relatively good fitting effect for data with and without cyclic characteristics. In addition, the least square fitting method based on B-spline basis is improved for sparse functional data processing.

Secondly, multivariate statistical principal component analysis method and functional principal component analysis method are discussed. The empirical analysis results of these two methods show that there are two main influencing factors for the fluctuation of GDP mean value in China's provinces, the pulling effect of large economic provinces is the main influencing factor.

Finally, the systematic clustering analysis method and the functional clustering analysis method are discussed, and the cosine method of curve for calculating the distance between functional variables is proposed. The empirical analysis results of the two methods show that although the systematic clustering analysis method and the functional clustering analysis method process data from different angles, the results obtained have some similarities. The results show that the total GDP and development potential of each province in China can be divided into different categories by province. On the contrary, provinces with relatively backward economy have more room for development, and the state should increase its policy support to these provinces.

Keywords: functional data, basis function fitting, functional principal component analysis, functional clustering analysis, GDP

目 录

| | |
|----------------------------|----|
| 摘 要..... | I |
| ABSTRACT..... | II |
| 第 1 章 绪 论..... | 1 |
| 1.1 课题来源..... | 1 |
| 1.2 课题研究背景..... | 1 |
| 1.3 国内外在该方向的研究现状..... | 2 |
| 1.4 主要研究内容..... | 4 |
| 第 2 章 预备知识..... | 5 |
| 2.1 函数型数据..... | 5 |
| 2.2 函数型数据的统计量..... | 5 |
| 2.3 函数型数据分析的工具..... | 6 |
| 2.4 本章小结..... | 8 |
| 第 3 章 函数型数据预处理..... | 9 |
| 3.1 平滑与噪声..... | 9 |
| 3.2 基函数的选择..... | 10 |
| 3.3 函数型数据平滑方法..... | 11 |
| 3.3.1 基于 B 样条基的最小二乘拟合..... | 11 |
| 3.3.2 基于单项式基的最小二乘拟合..... | 13 |
| 3.4 GDP 数据平滑处理..... | 13 |
| 3.5 B 样条基最小二乘拟合方法改进..... | 18 |
| 3.5.1 三次样条插值..... | 19 |
| 3.5.2 GDP 季度数据处理..... | 23 |
| 3.5.3 基于傅里叶基的最小二乘拟合..... | 26 |
| 3.6 本章小结..... | 28 |
| 第 4 章 主成分分析..... | 29 |
| 4.1 多元统计主成分分析..... | 29 |
| 4.2 函数型主成分分析..... | 32 |
| 4.2.1 函数型主成分分析定义..... | 32 |
| 4.2.2 主成分分析和特征值分析..... | 35 |
| 4.2.3 函数型主成分分析推广..... | 35 |
| 4.3 GDP 数据主成分分析..... | 36 |

| | |
|----------------------------|----|
| 4.4 本章小结..... | 41 |
| 第 5 章 聚类分析..... | 42 |
| 5.1 系统聚类分析..... | 42 |
| 5.2 函数型聚类分析..... | 43 |
| 5.3 GDP 数据聚类分析..... | 45 |
| 5.4 本章小结..... | 58 |
| 结 论..... | 59 |
| 参考文献..... | 61 |
| 附 录..... | 65 |
| 哈尔滨工业大学学位论文原创性声明和使用权限..... | 89 |
| 致 谢..... | 90 |

第 1 章 绪 论

1.1 课题来源

近年来,随着计算机和通信技术的迅速发展,数据在计算机中的储存形式也逐渐发生着变化,现代的数据收集技术所收集的信息越来越多地以曲线和图像的形式存在。高频数据得益于电子计算机的快速处理能力而被广泛应用,经典的数据分析方法被应用到这种数据的处理中,甚至还被应用到超高频数据的分析中。在此背景下,由于高频数据维数太大,从而导致了“维数灾难”问题。经典的多元统计分析方法适合处理一般维数的数据,无法处理高维数据。为了解决这个问题,函数型数据分析方法被提出,函数型数据分析方法在处理高频数据时具有一些无可比拟的优势。目前,函数型主成分分析、函数型回归分析等函数型数据分析方法都成为了统计学研究中的热点问题。

1.2 课题研究背景

函数型数据分析 (Functional Data Analysis, FDA) 起源于 20 世纪 50 年代,最初主要用在心理学研究领域。后来,随着研究范围的扩展,函数型数据分析方法更多地被应用于生物统计、海洋、医学等领域。近年来,股票交易、期权期货和保险业等经济领域业务发展迅速,这些领域所产生的数据往往都是以秒甚至毫秒为时间单位记录的高频交易数据,所以函数型数据分析方法也被越来越多地应用到经济学研究中。这种方法为数据处理和数据分析提供了一种新的研究视角和思想,对于处理高维数据问题有明显的优势,因而受到了越来越多学者的青睐。除此之外,与经典的多元统计分析方法相比,函数型数据分析方法具有很多优势,它是对离散数据经过拟合所得到的连续函数进行分析,对观测数据直接进行建模,从数据本身出发,克服了模型必须依赖事先假定分布的局限性。函数型主成分分析的降维方法不仅可以对多个维度的信息进行综合压缩提取,同时可以克服高维数据分析的“维数诅咒”等问题。函数型聚类分析方法不仅可以根据曲线的数值特征研究分类,而且可以从曲线的导数特征研究曲线的变化速率。

国内生产总值 (Gross Domestic Product, GDP) 作为度量一个国家或者地区经济发展情况的主要影响因素,一直都是宏观经济学研究中所要重点关注的对象。GDP 总量代表了一个国家或者地区的经济实力,而从 GDP 的波动变化曲线可以看出国家或地区的经济周期状态和经济运行情况。在宏观经济的相关课题研究中,GDP 被许多学者越来越多地关注。此外,GDP 的研究分析结果对于国家或者地方

各级政府机构制定相关经济决策也有一定的参考价值。因此,研究不同地区 GDP 的区别、探究 GDP 的影响因素以及按发展潜力对 GDP 进行聚类等课题不仅具有深刻的理论意义,而且具有重要的实证研究意义。将函数型数据分析方法应用于我国 GDP 数据的分析中,一方面可以拓展函数型数据分析方法的应用范围,另一方面,相关的研究报告可以作为某些政府职能部门作出相关决策的资料来源。

1.3 国内外在该方向的研究现状

加拿大学者 Ramsay^[1]在 1982 年明确指出了什么是函数型数据,同时也提出了精确区分函数型数据和其它类型数据的方法,函数型数据的相关研究由此开始。Ramsay 指出随着数据收集技术的提高,现代收集系统所收集的数据具有动态性的特征,这种具有动态性特征的数据与传统的静态数据有很大的区别,如果继续使用经典的数据分析方法来处理这种数据,将会造成许多严重的问题,如估计出现较大的偏差,丢失数据所包含的必要信息等。因此,必须对经典的数据分析方法进行扩展,使它可以处理这种具有动态特征的函数型数据。此后,Ramsay 和 Dalzell^[2]于 1991 年给出了函数型数据分析的概念,并提出了函数型数据分析的某些方法,如主成分分析和回归分析。此外,他把这些方法应用到加拿大温度与降水量关系的实证研究中,得到了许多有价值的结果,开启了函数型数据分析方法实证应用的先河。此后,自从 Ramsay 和 Silverman^[3-5],Ferraty 和 Vieu^[6]等学者发表了多本相关著作以来,函数型数据分析方法被应用到越来越多的学科中,受到了许多研究者的关注。具体地,国外对于函数型数据的研究可以概括为以下几个阶段。20 世纪 90 年代以前是第一个阶段,此阶段采用的主要方法是先把函数进行离散化处理,再对离散的数据进行分析。这种处理方法比较简单易用,但是此方法仍然是基于离散的数据进行研究,没有体现出函数的特征,文献[7,8]介绍了这种方法的具体应用。1990 年到 2000 年是第二个阶段,在这段时间里,有关经典函数型数据分析方法的研究有了极大的进展,如函数型主成分分析、函数型回归分析等方法的研究都取得了重要的成果,文献[9]对第二阶段的相关研究进行了综述。2000 年至 2008 年是第三个阶段,此阶段的研究成果包括两个方面,一方面对于稀疏函数型数据^[10]的研究取得了突破性进展,另一方面关于非参数函数型数据^[11]的研究也取得了重要的理论成果。第四个阶段是 2008 年以后,此阶段的研究取得的成果较多,函数型数据分析的范围也逐渐扩大,主要包括三个方面,第一个方面是函数型主成分及函数型线性模型的最优收敛速度^[12,13]、第二个方面是半参数函数型数据^[14-16]、第三个方面是多维函数型数据如二维图像数据^[17,18]及二元甚至多元函数型数据^[19-24]。

目前,函数型数据分析方法在医学、地质学、生物统计学、气象学及海洋研

究等学科领域中都得到了广泛的应用。因为金融市场中的数据大部分符合函数型数据的特征,越来越多的学者开始关注该方法在金融领域中的应用。近年来,国外许多学者开始关注函数型数据分析方法与金融领域相关问题的结合,并取得了一些有意义的成果,但是相关研究仍然处于初始阶段,无论是在理论推导还是实证应用方面都需要做进一步的探究。在经济学的相关研究中,Ramsay 和 Ramsey^[25]将函数型聚类分析方法应用于非耐用消费品指数的研究中。Wang^[26]将函数型主成分分析方法应用于网上商品拍卖价格的分析中,扩展了函数型数据分析方法的应用范围。此外,函数型数据分析方法在金融学的理论研究与应用方面也有一些研究成果,Anderson 和 Newbold^[27]最先把函数型自回归分析方法应用于资产收益建模的研究中,构建出了比较稳定的数学模型。研究结果显示,与传统的统计方法相比,函数型数据分析方法更有利于金融数据的分析。Müller, Stadtmüller 和 Yao^[28,29]首次提出了函数型方差分析过程的概念,并对具体金融数据的方差分析过程进行了阐述,提出了函数型方差分析的新观点。Eddie^[30]在传统的函数型主成分分析方法的基础上提出了基于向量自回归 (VAR) 方法的函数型主成分分析,并且采用原油期货合约的价格进行了实证分析研究,研究结果表明函数型主成分分析方法在处理高频交易数据时具有明显的优势。Ingrassia 和 Costanzo^[31]将函数型数据分析方法应用于金融时间序列数据的分析中,以 MBI300 指数为样本进行了验证,实证结果表明函数型聚类分析方法在处理金融时间序列数据时具有简单、高效的特点。

中国学者对于函数型数据分析方法的研究时间比较短,在理论研究上取得的成果相对较少,目前的研究主要集中在方法的实证应用方面,对方法的扩展及模型的改进研究相对较少。张崇岐等^[32]讨论了把离散函数型数据进行平滑的方法,对比了多种基函数的平滑效果,此外还对平滑过程中需要特别注意的细节进行了详细的阐述与说明。严明义^[33]最先基于统计学的意义从宏观角度介绍了函数型数据分析方法的核心思想,他为函数型数据分析方法在我国的引入与发展做出了巨大的贡献。近年来,关于函数型数据在经济和金融领域的应用,国内也有相关的研究文献。在经济学研究方面,靳刘蕊^[34]主要利用函数型主成分分析和函数型聚类分析方法研究了国内的某些经济数据。剡亮亮^[35]从泛函特征的角度介绍了函数型数据的主微分分析方法,并利用此方法对全国银行间同业拆借利率进行了研究,这是一种比较新颖的研究方式。关于金融学的研究,也有相关的研究文献,毛娟^[36]利用函数型数据分析方法研究了美国股票市场 S&P500 指数在 2003 年 9 月 23 日至 2004 年 6 月 2 日期间的指数隐含波动率。岳敏和朱建平^[37]主要研究了影响股票收益率波动的原因,在研究中主要利用了函数型数据主成分分析方法,结果表明影响股票收益率波动的主要因素有两个,也即利用主成分分析方法所得到的前两个

主成分。王诚^[38]则使用 2013 年 A 股股票为样本数据,进行了同样的研究工作。郭均鹏、孙钦堂和李汶华^[39]研究了 Shibor 收益率,在研究过程中创新性地提出了区间主成分分析方法。胡梦荻^[40]、陈丽琼^[41]、张蕾蕾^[42]利用函数型数据分析方法对我国的各种金融数据进行了理论和实证研究,包括股票和期权等具体金融数据。

综上所述,目前国内外对于函数型数据分析方法的研究均取得了一定的成果,但相比国外的相关研究,国内在此方面仍具有较大的发展空间。从国内外研究现状来看,下面几点需要更进一步深入思考与探究。首先,从函数型数据分析方法引入金融研究的总体框架来看,当前研究多针对金融市场价格变化的某一个方面,或是针对某一种具体的模型进行研究,而将研究对象和模型紧密结合,进行系统性研究的文献很少,尤其是进行多种途径综合分析的实证研究不多。其次,对于具体的函数型数据分析过程,准确认识函数型数据和一般数据的共性与个性以及其统计特征,是进行金融市场分析的基础,但当前的研究大多是直接将函数型数据的一般分析方法套用到所研究的数据上,缺乏对其适用性和实际含义的合理解释。最后,无论是在理论方面还是应用方面,国内的研究都存在一定程度的滞后,有关算法的改进、方法拓展及创新性结论的研究成果较少。

1.4 主要研究内容

本文的主要研究内容如下:

首先介绍预备知识,主要包括函数型数据的描述性统计知识和函数分析的一些技术,如函数型数据的均值、方差、协方差及相关函数,内积和矩阵分解的基本概念。

然后介绍了函数型数据预处理的主要理论和方法,包括平滑与噪声,几种基函数的定义、优点以及缺点,并且利用基于各种基函数的最小二乘拟合方法对我国 GDP 分省数据进行平滑处理,比较不同基函数的拟合效果。

接下来介绍了主成分分析的有关理论,重点探讨了函数型主成分分析方法,包括函数型主成分分析的定义和函数型主成分特征值分析。此外分别利用经典的多元统计主成分分析和函数型主成分分析方法对我国 GDP 分省数据进行实证分析,比较两种分析方法的异同。

最后基于我国各省 GDP 数据的拟合函数及其导函数分别对系统聚类分析和函数型聚类分析两种方法进行了理论和实证研究,分析研究结果可以得到某些有价值的结论。

第 2 章 预备知识

2.1 函数型数据

在通常的研究中，数据一般分为时间序列数据、面板数据和横截面数据。目前，随着信息收集技术的变革，许多研究领域收集到的数据往往以曲线或者图像的形式存在，这种类型的数据通常是时间序列数据与横截面数据的“非标准”结合体，即数据并非是按照固定的时间点或观测个体记录。此时，经典的数据分析方法无法对这种“非标准”类型的数据进行分析，而当用隐含的曲线将这些离散的记录数据连接起来时，这些曲线就可以作为独立的观测值，进而可以对独立的拟合曲线进行分析。从这种观测视角得到的数据具有函数型数据的特征。

如果数据采样点的间隔比较紧密，那么在此情况下，测量得到的离散数据内在会表现出某些函数的特征，这种内在特征的外在表现形式往往就是可以把离散的测量数据拟合成连续的函数，这种以函数为表现形式的数据称为函数型数据。其中，“函数型”代表的是数据内部存在的某种联系或关系，并不是表现在外面的形式。

根据观测变量的个数，函数型数据又具体地可以分为一元函数型数据和多元函数型数据，这里的元也就是观测指标。只对一个观测指标进行分析称作一元函数型数据分析，同理多元函数型数据分析是指观测多个指标的变化情况并进行分析。除此之外，根据记录数据的疏密状况，函数型数据又可以分为稠密函数型数据和稀疏函数型数据。

2.2 函数型数据的统计量

定义 2.1 设 $x_1(t), x_2(t), \dots, x_p(t)$ 是 P 个函数个体，则称

$$\bar{x}(t) = \frac{1}{P} \sum_{j=1}^P x_j(t) \quad (2-1)$$

为函数型数据的均值函数，它是对函数型观测数据逐个求和的算术平均值。

定义 2.2 设 $x_1(t), x_2(t), \dots, x_p(t)$ 是 P 个函数个体，则称

$$\text{Var}_x(t) = \frac{1}{P-1} \sum_{j=1}^P [x_j(t) - \bar{x}(t)]^2 \quad (2-2)$$

为函数型数据的方差函数，函数型数据的标准差函数是方差函数的平方根。

定义 2.3 设 $x_1(t), x_2(t), \dots, x_p(t)$ 是 P 个函数个体，则称

$$\text{Cov}_x(t, s) = \frac{1}{P-1} \sum_{j=1}^P \{x_j(t) - \bar{x}(t)\} \{x_j(s) - \bar{x}(s)\} \quad (2-3)$$

为函数型数据的协方差函数。

定义 2.4 称

$$\text{corr}_x(t, s) = \frac{\text{Cov}_x(t, s)}{\sqrt{\text{Var}_x(t) \text{Var}_x(s)}} \quad (2-4)$$

为函数型数据的相关函数。

定义 2.5 设 $x_1(t), x_2(t), \dots, x_p(t)$ 是 P 个函数个体, 则称

$$\text{Cov}_{x,y}(t, s) = \frac{1}{P-1} \sum_{j=1}^P \{x_j(t) - \bar{x}(t)\} \{y_j(s) - \bar{y}(s)\} \quad (2-5)$$

为函数型数据的交叉协方差函数。

定义 2.6 称

$$\text{corr}_{x,y}(t, s) = \frac{\text{Cov}_{x,y}(t, s)}{\sqrt{\text{Var}_x(t) \text{Var}_y(s)}} \quad (2-6)$$

为函数型数据的交叉相关函数。

2.3 函数型数据分析的工具

定义 2.7 若函数 $x(t)$ 与 $y(t)$ 的乘积在区间 $[a, b]$ 上可积, 那么它们内积定义为

$$\langle x, y \rangle = \int_a^b x(t)y(t)dt \quad (2-7)$$

内积具有如下基本性质

- (1) 对称性: 对于所有的函数 $x(t)$ 和 $y(t)$, $\langle x, y \rangle = \langle y, x \rangle$;
- (2) 正定性: 对于所有的函数 $x(t)$, $\langle x, x \rangle \geq 0$, 其中 $\langle x, x \rangle = 0$ 当且仅当 $x = 0$;
- (3) 线性性: 设 a 与 b 是任意的实数, $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$ 对于所有的函数 $x(t), y(t), z(t)$ 成立。

此外, 若 $\langle x, y \rangle = 0$, 称函数 $x(t)$ 与 $y(t)$ 正交。

定义 2.8 若函数 $x(t)$ 在区间 $[a, b]$ 上可积, 称

$$\|x\| = \sqrt{\int_a^b x(t)^2 dt} \quad (2-8)$$

为 $x(t)$ 的范数, 记为 $\|x\|$. 其中

$$\|x\| \geq 0$$

范数具有如下基本性质

- (1) $\|x\| \geq 0$, 并且 $\|x\| = 0$ 当且仅当 $x = 0$;
- (2) 对于所有的实数 a , $\|ax\| = |a|\|x\|$;
- (3) $\|x + y\| \leq \|x\| + \|y\|$.

定义 2.9 若函数 $x(t)$ 和 $y(t)$ 在区间 $[a, b]$ 上可积, 称

$$d_{xy} = \|x - y\| = \sqrt{\int_a^b (x(t) - y(t))^2 dt} \quad (2-9)$$

为它们之间的欧式距离。

定义 2.10 假设 Z 是 $m \times n$ 维的矩阵, 乘积表达式

$$Z = UDV'$$

称为矩阵 Z 的奇异值分解。

其中, $m \times q$ 维的矩阵 U 满足 $U'U = I_q$, I_q 是 q 阶单位阵; $q \times q$ 维的矩阵 D 是对角线元素严格大于 0 的对角阵; $n \times q$ 维的矩阵 V 满足 $VV' = I_q$. 此外, 矩阵 D 的对角线元素 d_1, d_2, \dots, d_q 称为矩阵 Z 的奇异值, 且满足 $d_1 \geq d_2 \geq \dots \geq d_q > 0$, 矩阵 Z 的秩也为 q .

定义 2.11 假设 A 是 $m \times n$ 维的矩阵, B 是 $p \times q$ 维的矩阵, A 和 B 的克罗内克积 $A \otimes B$ 定义为

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

它是 $mn \times pq$ 维矩阵, 在多元统计分析中, 克罗内克积又称为矩阵的直积。

克罗内克积具有如下简单的性质

$$\begin{aligned} A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\ (A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \\ (A + B) \otimes C &= (A \otimes C) + (B \otimes C) \\ A \otimes (B + C) &= (A \otimes B) + (A \otimes C) \\ \text{trace}(A \otimes B) &= (\text{trace}A)(\text{trace}B) \end{aligned}$$

此外, 若 A 与 B 均为非奇异矩阵, 有

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

2.4 本章小结

本章主要介绍了函数型数据分析相关的背景知识。首先介绍了函数型数据的基本概念，然后介绍了函数型数据的统计量，包括均值函数、方差函数、相关函数、协方差以及交叉协方差函数和交叉相关函数的定义，这些概念都可以看作经典多元统计分析中相关概念的推广。最后介绍了函数型数据分析的一些方法工具，这些处理方法在后文中都会经常用到。

第3章 函数型数据预处理

将离散的函数型数据转化为平滑的曲线是函数型数据分析的首要步骤。只有先把离散数据转化为平滑效果较好，误差在有效范围之内的函数才可以进行函数型主成分分析、函数型聚类分析等研究，而且平滑效果的好坏直接关系到主成分分析及聚类分析的精确度。此外，平滑曲线的展示更有利于挖掘数据的潜在特征，首先通过对离散数据平滑，可以直观看出离散数据所适用的模型，其次对于平滑曲线，可以很容易地求出它的各阶导数，从导数的角度来考察数据个体之间的差异与动态变化规律是离散数据所不具备的。将离散数据合理地转化为平滑曲线是函数型数据分析过程中的重要环节。

3.1 平滑与噪声

若函数具有一阶或二阶导数，那么函数具有平滑性。符号 $d^m x/dt^m$ 代表函数 x 的 m 阶导数， $d^m x(t)/dt^m$ 是 x 的 m 阶导数在参数 t 处的值。此外，平滑函数 x 的相邻函数值 y_j 和 y_{j+1} 在某种程度上必然链接在一起，并且彼此不会相差太大。如果不满足这种特征，那么把数据当做函数型数据处理与当作传统多元数据处理就没有任何的区别。

观测得到的原始数据往往是带有测量误差的数据，此外噪声也常对数据造成一定的影响，在这些因素的影响下，实际得到的数据通常并不具有平滑性特征。噪声具有无形和不可预测的特征，这使得函数值及其导数值的估计成为了一个挑战。虽然原则上可以对具有噪声的数据建模，但在实际中往往选择忽略噪声。一般来说，我们关心的是函数型数据的集合或样本，而不仅仅是一个单个的函数。设函数 x_j 在点 t_{ij} 处的观测值为 y_{ij} ， $i=1, \dots, n_j$ 。每个函数的参数值 t_{ij} 可能是相同的，但也可能因记录而异，观测数据收集的时间间隔 τ 也可能因记录而异。通常，对每条记录的离散观测值进行函数构造的过程是相互独立的，然而在信噪比较低、数据采样稀疏或数据量较少的情况下，使用相邻或相似曲线中的信息来获得更稳定的曲线估计是至关重要的。此外，从原始的离散数据很难直接观察到函数是否具有平滑性特征。

对于具有噪声的函数型数据，文献[43]建议采用“信号加噪声”模型来进行建模，模型表达式为

$$y_i = x(t_i) + \varepsilon_i \quad (3-1)$$

式中 ε_i 代表噪声、误差、扰动或其他外部影响因素对原始数据造成的影响。将原始数据表示为函数的任务之一是在建模之前尽可能有效地过滤掉噪声或者在估计函数过程中先保留噪声，模型建立后再来分析结果的平滑性。此外采用向量符号能使表达式更简洁，向量表达式为 $\mathbf{y} = \mathbf{x}(\mathbf{t}) + \boldsymbol{\varepsilon}$ 。其中 $\mathbf{y}, \mathbf{x}(\mathbf{t}), \mathbf{t}$ 和 $\boldsymbol{\varepsilon}$ 都是长度为 n 的列向量，用符号 Σ_{ε} 代表残差的方差—协方差矩阵。此外，不难发现观察值向量 \mathbf{y} 的方差—协方差矩阵等于残差值向量 $\boldsymbol{\varepsilon}$ 的方差—协方差矩阵，这是因为 $x(t_i)$ 被认为是方差为 0 的固定效应。

3.2 基函数的选择

对基函数进行组合处理是构造函数的一种主要方式。基函数非常适合存储函数型数据的信息，具有较高的灵活性，适用于处理将大量离散点转化为平滑函数的问题。同时，这种方法对数据的要求较低，通常情况下适用于处理各种类型的数据，具有较广的适用范围。基函数系统是指由一系列已知的函数 ϕ_k 构成的函数集 Φ ，系统中的函数相互独立，且任意一个函数都可以通过系统内函数的加权求和或者线性组合表示出来。用 K 个已知的基函数 $\phi_k, k=1, \dots, K$ 的线性组合来表示函数 $x(t)$ 的定义如下

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) \quad (3-2)$$

式 (3-2) 的矩阵符号表达式为

$$\mathbf{x} = \mathbf{c}'\Phi = \Phi'\mathbf{c} \quad (3-3)$$

其中， \mathbf{c} 是系数 c_k 组成的长度为 K 的向量， Φ 是由基函数 ϕ_k 组成的函数向量。实际上，基函数的展开方法代表了像 \mathbf{c} 这样的有限维向量框架内潜在的无限维函数信息。在基函数 $\phi_k(t)$ 确定的条件下，系数向量 $\mathbf{c} = (c_1, \dots, c_k)'$ 的唯一取值就确定了 $x(t)$ 。其中，插值函数的平滑度由基函数的个数 K 决定，因此 K 不是值固定的参数，是根据数据本身的特征确定大小的。

最常见的基函数系统是用来构造幂级数的单项式集合，表示如下

$$1, t, t^2, \dots, t^k, \dots$$

傅里叶基函数 (Fourier base function) 也是一种常用的基函数，主要用来描述周期性数据，表示如下

$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \sin(3\omega t), \cos(3\omega t), \dots, \sin(k\omega t), \cos(k\omega t), \dots$$

此外，还有其它类型的基函数系统，不同的基函数系统有不同的优点。另外，同一种基函数对于不同数据的插值效果也会有很大差别，目前还没有一个对于所有

的数据集都有良好平滑效果的基函数。此外，由（3-2）可计算函数 $x(t)$ 的导数

$$\frac{dx(t)}{dt} = \sum_{k=1}^K c_k \frac{d\phi_k(t)}{dt} = \mathbf{c}'\Phi'(t) \quad (3-4)$$

特别要注意上式中 $\Phi'(t)$ 代表 $\Phi(t)$ 的导数，不是转置符号。有时，一组对离散数据点插值很有效的基函数，它的导函数的估计值却与实际有极大的偏差。这是因为观测值的精确表示可能会迫使 $x(t)$ 产生小的高频率振荡，这会对其导数的估计产生可怕的后果。因此，基函数的选择对于函数导数的估计极其重要，选择基函数的标准不仅仅要考虑函数近似值，还要考察高阶导数值是否合理。

在实践中发现，可以通过选取不同种类的基函数和选择合适的基函数个数等手段来实现复杂数据的拟合。使用基函数时要考虑如下几个因素，即数据拟合是否精确、运算量是否合理、导数条件是否满足。在函数型数据分析中，傅里叶基函数和 B 样条基函数是被使用最多的基函数。此外，还有其它的基函数系统，例如小波基函数、指数函数基函数、幂函数基函数、多项式基函数、折线函数基函数、阶梯函数基函数及主成分基函数等。

3.3 函数型数据平滑方法

3.3.1 基于 B 样条基的最小二乘拟合

样条是工程学科中的一个名词，工程师为了把一些已知的点拟合成一条平滑的曲线，通常使用木条把相近的点连接在一起，再逐步延伸连接起所有的点，形成一条光滑的样条曲线。样条曲线具有很多优良的性质，利用数学方法对曲线进行建模，得到样条函数。样条函数广泛地应用于数据拟合、数值的微积分等领域，在工业生产中也占据着十分重要的地位^[44,45]。样条函数是一类分段的、具有光滑性且在连接点处也具有平滑性的函数，主要由两个因素决定，即分段多项式函数的阶数和分段的节点个数。对于非周期函数型数据或参数，样条函数通常是一个不错的选择。样条函数结合了多项式的快速计算能力和较高的灵活性，通常只需少量的基函数即可实现拟合或插值。

B 样条基 (B-spline base) 是法国雷诺汽车公司工程师 Bezier^[46]为了改善样条插值的缺点而提出的，B 样条函数又称为贝塞尔曲线 (Bezier)。B 样条基通常用于描述没有强烈循环周期变化的数据，是函数型数据预处理中使用最为广泛的基函数系统，文献[47-49]对函数型数据的平滑预处理采用的都是 B 样条基。下面简要叙述 B 样条基的定义和计算过程。为了方便地将数学运算过程转化为计算机程序流程，本节只考虑 B 样条基函数的多项式形式，对于 B 样条基的定义采用了递推

显示方式。设

$$\{t_1, t_2, \dots, t_i, t_{i+1}, \dots\}$$

是单调不减的实数序列, 即 $t_i \leq t_{i+1}$, $i \in \mathbb{Z}$. 令第 i 个 k 次 B 样条基函数 $B_{i,k}(t)$ 为

$$B_{i,k}(t) = \frac{t - t_i}{t_{i+k} - t_i} B_{i,k-1}(t) + \frac{t_{i+k+1} - t}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(t) \quad (3-5)$$

式 (3-5) 中 $k \neq 0$. 当 $k = 0$ 时, 规定

$$B_{i,0}(t) = \begin{cases} 1, & t_i \leq t \leq t_{i+1} \\ 0, & \text{其他} \end{cases} \quad (3-6)$$

其中, $[t_i, t_{i+1}]$ 是 B 样条的第 i 个节点区间, 在递推计算中规定 $0/0 = 0$. 式 (3-5) 说明 $B_{i,k}(t)$ 是 2 个 $k-1$ 次 B 样条基函数的线性组合。

此外, B 样条基还有如下一些性质^[46]

(1) 非负性和局部支撑性:

$$B_{i,k}(t) \begin{cases} > 0, & t_i \leq t \leq t_{i+k+1} \\ = 0, & \text{其他} \end{cases} \quad (3-7)$$

(2) 单位分解性质: B 样条函数在任意选定的节点区间 $x \in [t_i, t_{i+1}]$ 内呈完备分布状态, 且满足

$$\sum_i B_{i,k}(t) = 1 \quad (3-8)$$

事实上利用式 (3-8), 任意阶次的 B 样条函数在区间 $x \in [t_i, t_{i+1}]$ 上的单位分解性质可表示为

$$\sum_{j=0}^k B_{i-k+j,k}(t) = 1 \quad (3-9)$$

(3) 高阶连续性: 以不出现重节点为原则, 在 B 样条基函数取值的区间选定节点时, 即 $t_i \leq t_{i+1}$, 那么 $B_{i,k}(t)$ 具有 $(k-1)$ 阶连续性。当存在重节点, 且重复度为 k_1 时, $B_{i,k}(t)$ 具有 $(k-k_1-1)$ 阶连续性。

在得到了 B 样条基后, 需要计算系数向量, 绝大部分的文献^[2-6]采用普通最小二乘法来估计系数向量。近年来, 文献[11]提出使用带惩罚项的最小二乘法代替普通最小二乘法, 但是由于普通最小二乘法使用广泛且理论成熟, 本文采用普通最小二乘法来估计系数向量。我们的目标是使用模型 $y_i = x(t_i) + \varepsilon_i$ 来拟合离散的观测值 $y_i, i = 1, \dots, n$. 可以利用最小二乘准则得到式 (3-2) 中基函数的展开系数 c_k 的估计值。

$$SMSSE(\mathbf{y}|\mathbf{c}) = \sum_{i=1}^n \left[y_i - \sum_{k=1}^K c_k \phi_k(t_i) \right]^2 \quad (3-10)$$

式 (3-10) 的矩阵形式表达式为

$$SMSSE(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \Phi\mathbf{c})'(\mathbf{y} - \Phi\mathbf{c}) \quad (3-11)$$

上式的等号右侧也通常表示为 $\|\mathbf{y} - \Phi\mathbf{c}\|^2$ 。对 (3-11) 式求关于 \mathbf{c} 的导数，得下式

$$2\Phi'\Phi\mathbf{c} - 2\Phi'\mathbf{y} = 0 \quad (3-12)$$

参数向量 \mathbf{c} 的最小二乘解 $\hat{\mathbf{c}}$ 为

$$\hat{\mathbf{c}} = (\Phi'\Phi)^{-1}\Phi'\mathbf{y} \quad (3-13)$$

向量 \mathbf{y} 的拟合值 $\hat{\mathbf{y}}$ 为

$$\hat{\mathbf{y}} = \Phi\hat{\mathbf{c}} = \Phi(\Phi'\Phi)^{-1}\Phi'\mathbf{y} \quad (3-14)$$

3.3.2 基于单项式基的最小二乘拟合

单项式 $\phi_k(t) = (t - \omega)^k, k = 0, \dots, K$ 构成的基函数也是一种经典的基函数系统，其中 ω 代表移位参数，它的值通常取 0。此外，随着度数的增加，单项式值的相关性越来越高，因此在拟合过程中必须注意避免计算中的舍入误差。下面叙述基于单项式基的最小二乘拟合方法详细计算过程。

假设对于给定的一组数据 $(t_i, y_i), i = 1, 2, \dots, n$ ，寻求作 K 次多项式 ($K \ll n$)

$$x = \sum_{k=0}^K c_k t^k \quad (3-15)$$

使总误差

$$SMSSE(\mathbf{y}|\mathbf{c}) = \sum_{i=1}^n \left(y_i - \sum_{k=0}^K c_k t_i^k \right)^2 \quad (3-16)$$

为最小。展开系数 c_k 估计值的计算方法同式 (3-12) 与 (3-13)，此处不再赘述。

3.4 GDP 数据平滑处理

本节分别利用基于 B 样条基的最小二乘拟合方法和基于单项式基的最小二乘拟合方法对我国 31 个省份 1978-2018 共 41 年的 GDP 分省年度数据进行平滑处理，下面叙述具体的计算过程

(1) 基于 B 样条基的最小二乘拟合方法

下面介绍利用 MATLAB 软件对我国 31 个省份 GDP 数据进行处理的具体方法。

我国 31 个省份 1978-2018 共 41 年的 GDP 年度数据参见附录 1 中的表 1；

调用附录 2 的 `bbasis = create_bspline_basis([1978, 2018], 39, 4)` 和 `plot(bbasis)` 函

数作出区间[1978, 2018]上的 39 个四阶 B 样条基函数图像如下图 3-1 所示;

计算 B 样条基函数在 1978-2018 共 41 个参数值点的函数值

```
bbasismat = eval_basis([1978: 2018]', bbasis);
```

计算最小二乘回归拟合系数 $\text{gdpcoef} = \text{bbasismat} \backslash \text{GDP_1978_2018_year}$;

创建函数型数据对象 $\text{fdoj} = \text{fd}(\text{gdpcoef}, \text{bbasis})$;

调用 $\text{plot}(\text{fdoj})$ 函数作出函数型数据对象 fdoj 的图像如图 3-2、3-3 所示;

分别计算 31 个省份 GDP 的 B 样条基函数拟合曲线残差平方和见表 3-1.

```
y2cMap = (bbasismat' * bbasismat) \ bbasismat';
```

```
yhat = bbasismat * y2cMap * GDP_1978_2018_year;
```

```
SSE = sum((GDP_1978_2018_year - yhat) .* (GDP_1978_2018_year - yhat));
```

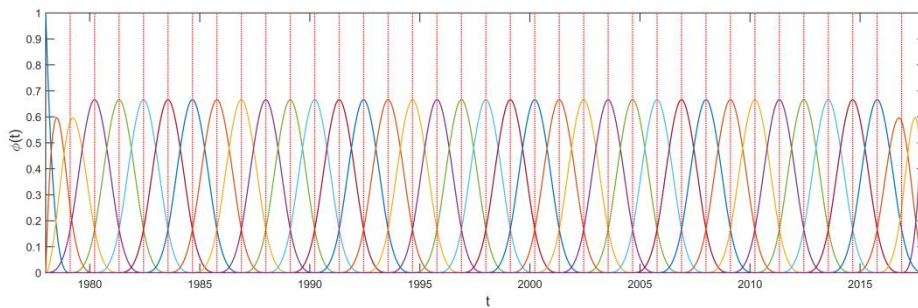


图 3-1 四阶 B 样条基函数

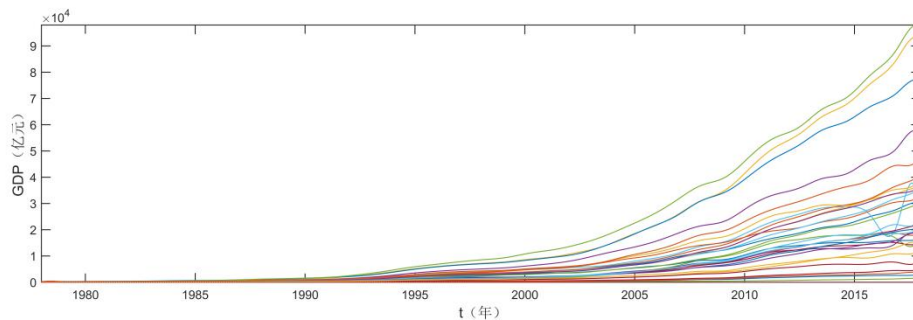
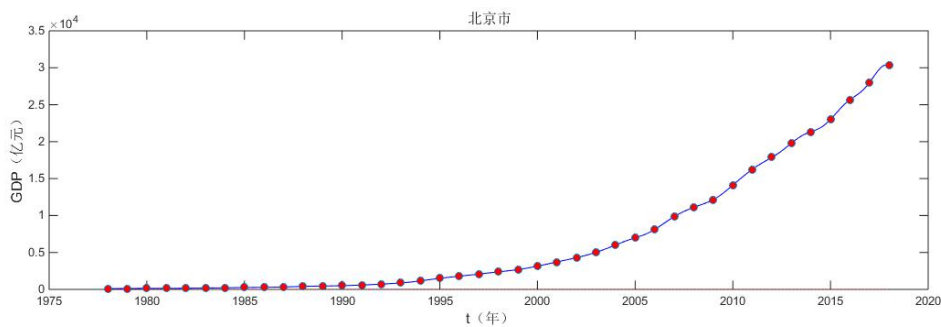


图 3-2 基于 B 样条基的我国 31 个省 GDP 数据拟合曲线



a)

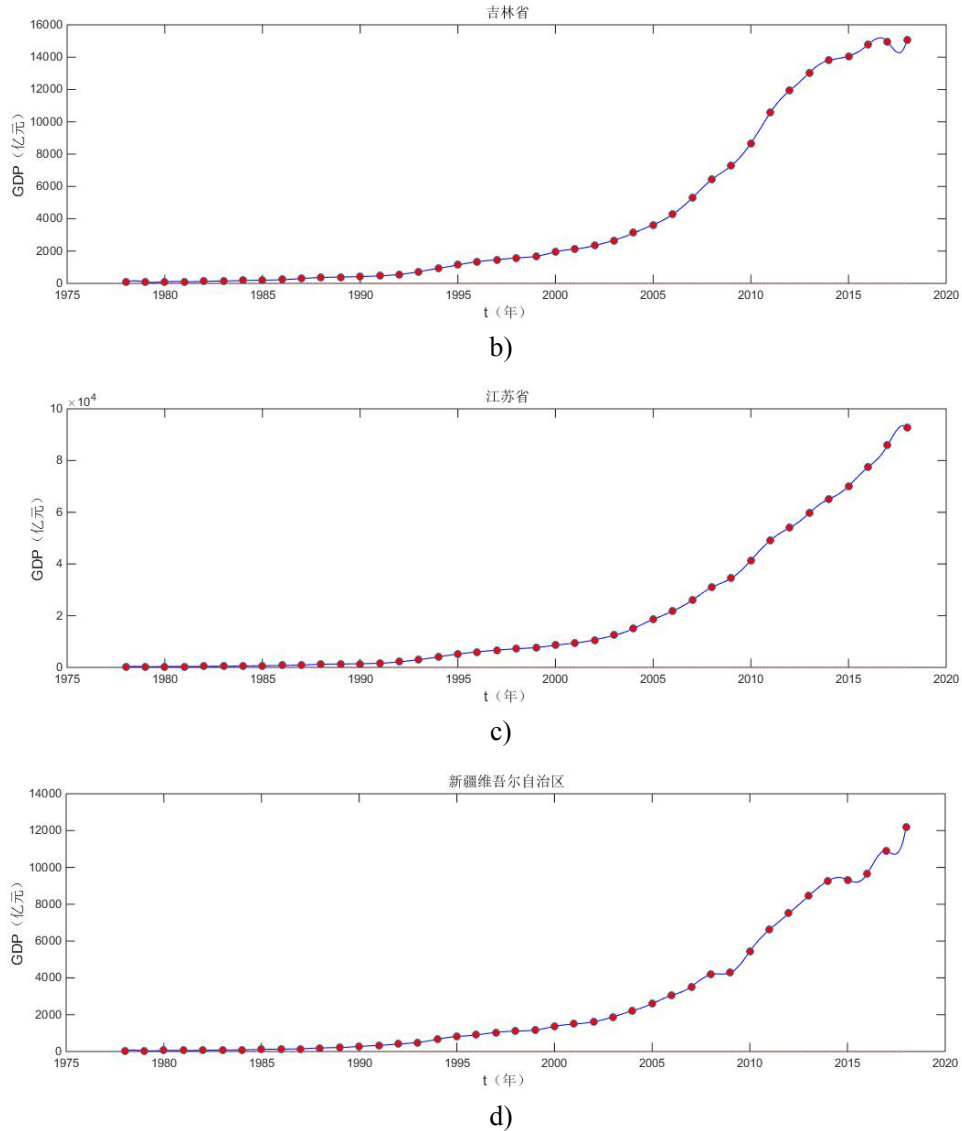


图 3-3 基于 B 样条基的我国部分省市分省 GDP 数据拟合曲线

图 3-3 作出了我国 4 个省份的 GDP 数据拟合曲线，从图像上可以发现利用基于 B 样条基的最小二乘拟合方法得到的平滑曲线对原始数据点逼近效果不错。此外，B 样条基函数对于其它省份 GDP 数据的拟合效果也很好，此处不再作出所有拟合曲线的图像。从图 3-2 和 3-3 可以看出我国各省份的 GDP 在 1978-2018 年间整体都呈现递增趋势，但是某些省份的 GDP 总量远远超过平均水平，而一些省份的 GDP 总量却远远低于平均水平。下面通过计算每条拟合曲线的残差平方和从定量数值上进一步分析拟合效果。

表 3-1 B 样条基拟合曲线残差平方和 (1×10^4)

| | | | | | | | | |
|--------|---------|----------|---------|--------|--------|--------|--------|--------|
| 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
| 0.1648 | 0.2686 | 0.2426 | 0.0041 | 0.0067 | 0.7722 | 0.0571 | 0.0394 | 0.1351 |
| 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
| 2.1440 | 0.1761 | 0.4816 | 0.0718 | 0.3081 | 0.8828 | 0.0151 | 0.0232 | 0.2243 |
| 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
| 1.1743 | 0.0809 | 0.0139 | 0.0486 | 0.0009 | 0.0172 | 0.3476 | 0.3498 | 0.0543 |
| 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 | | | | | |
| 0.0027 | 0.0013 | 0.0022 | 0.2978 | | | | | |

从表 3-1 可以发现一个有趣的现象,即采用同样数量的基函数拟合各省的 GDP 数据,结果显示江苏省、广东省和山东省的拟合残差平方和相对较高,而西藏自治区、青海省和新疆维吾尔自治区等经济发展相对缓慢的地区拟合效果较好。第五章会进一步解释造成这种现象的原因。

(2) 基于单项式基的最小二乘拟合方法

下面介绍利用 MATLAB 软件对我国 31 个省份 GDP 数据进行处理的具体操作过程。

调用附录 2 的 `mbasis = create_monomial_basis([1978, 2018], 4)` 函数定义区间 $[1978, 2018]$ 上的 4 个单项式基函数,生成基对象 `mbasis`;

计算单项式基函数在 1978-2018 共 41 个参数值点的函数值
`mbasismat = eval_basis([1978: 2018]', mbasis);`

计算最小二乘回归拟合系数 `gdpcoef = mbasismat \ GDP_1978_2018_year;`

创建函数型数据对象 `fdobj = fd(gdpcoef, mbasis);`

调用 `plot(fdobj)` 函数作出函数型数据对象 `fdobj` 的图像如图 3-4、3-5 所示;

分别计算每条拟合曲线的残差平方和,31 个省份的单项式基函数拟合曲线残差平方和见表 3-2.

`y2cMap = (mbasismat' * mbasismat) \ mbasismat';`

`yhat = mbasismat * y2cMap * GDP_1978_2018_year;`

`SSE = sum((GDP_1978_2018_year - yhat) .* (GDP_1978_2018_year - yhat));`

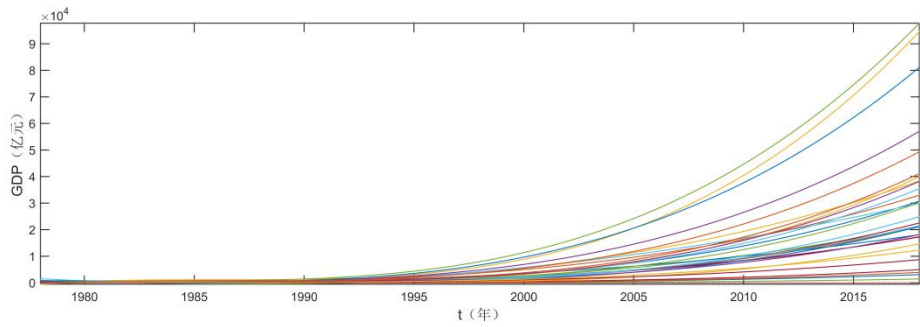
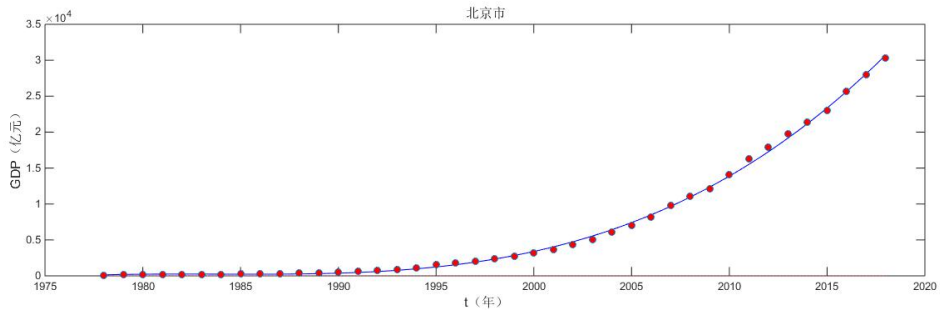
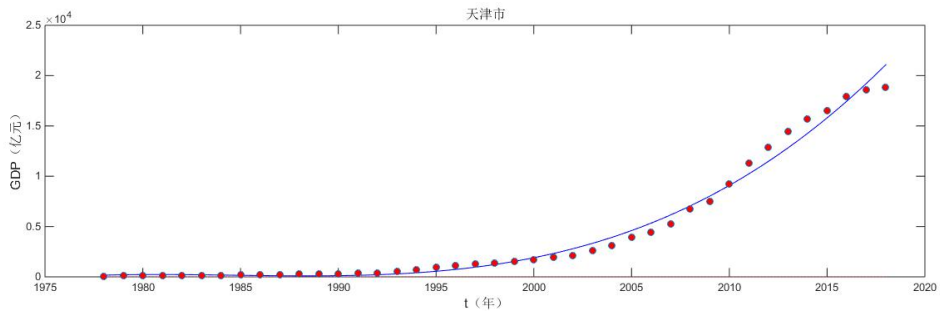


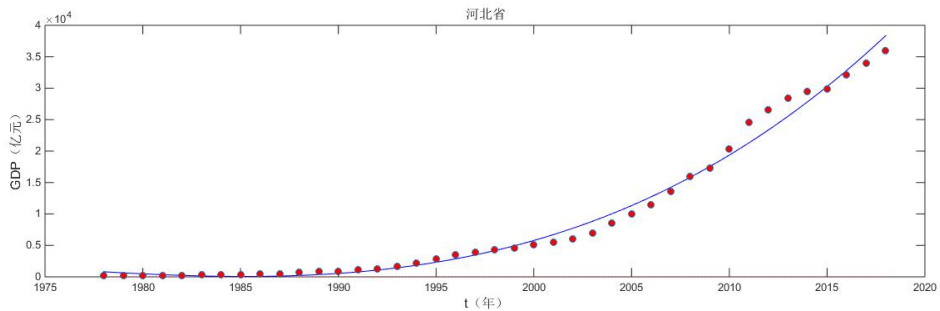
图 3-4 基于单项式基的我国 31 个省 GDP 数据拟合曲线



a)



b)



c)

图 3-5 基于单项式基的我国部分省市分省 GDP 数据拟合曲线

从图 3-4 可以看出基于单项式基的最小二乘拟合方法得到的拟合曲线的形状类似于幂函数曲线的形状。图 3-5 是基于单项式基的我国部分省市分省 GDP 数据拟合曲线，包括北京市、天津市和河北省，从图像可以发现单项式基函数灵活度

较差，导致许多点偏离曲线，拟合效果不好。下面通过计算每条拟合曲线的残差平方和从定量数值上分析拟合效果。

表 3-2 单项式基拟合曲线残差平方和 (1×10^8)

| | | | | | | | | |
|---------|---------|----------|---------|--------|--------|--------|--------|--------|
| 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
| 0.0327 | 0.1939 | 0.5920 | 0.1758 | 0.8999 | 2.3358 | 0.2348 | 0.2228 | 0.0599 |
| 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
| 0.8539 | 0.3119 | 0.1436 | 0.1241 | 0.0772 | 1.2998 | 0.3282 | 0.2416 | 0.3133 |
| 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
| 0.5379 | 0.0305 | 0.0045 | 0.0722 | 0.0002 | 0.0335 | 0.0538 | 0.2675 | 0.1520 |
| 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 | | | | | |
| 0.00039 | 0.0047 | 0.0379 | 0.1487 | | | | | |

从拟合曲线以及残差平方和很容易观察到 B 样条基的拟合效果远远优于单项式基的拟合效果，这是因为单项式基有一些固有的缺陷。首先，正如傅里叶级数展开一样，如果不使用较大的 K 值，拟合多项式就不能表现出非常局部的特征。此外，单项式基在数据区间的中心拟合效果比较好，但在区间尾部的拟合效果不尽人意，因此用它来做推断或预测效果很差。再者，虽然多项式展开式的导函数计算简单，但由于高阶多项式的快速局部振荡特征，使得它作为实际导数的估计值也存在一些偏差。

3.5 B 样条基最小二乘拟合方法改进

本章对我国各省份的 GDP 数据进行平滑处理，GDP 作为衡量国家和地方政府经济发展状况的主要因素，由有关部门专门统一核算，数据准确度相对较高。本节将利用我国 31 个省份 2000-2018 共 19 年的 GDP 季度数据进行平滑处理。但是，我国从 2005 年才开始统计 GDP 的季度数据，2000-2004 年的 GDP 数据是按年份统计的年度数据，这样就使得数据具有稀疏性特征，拟合函数的精确度会大大降低。所以，本节对基于 B 样条基的最小二乘拟合方法进行了改进，首先采用插值的方法对 2000-2004 年的 GDP 年度数据进行插值，插值节点是季度，这样就可以估计出各省份每个季度 GDP 的值，从而大大增加了数据点，将稀疏数据转化成了稠密数据，采用这样的方法可以扩展函数型数据预处理方法的适用范围。下面介绍实践中应用最多的一种插值方法—三次样条插值。

3.5.1 三次样条插值

实践中常有这样的问题，我们已经得到函数 $y = x(t)$ 在自变量点 t_0, t_1, \dots, t_n 处的函数值 y_0, y_1, \dots, y_n ，但是函数的精确表示形式无法得到，需要构造一个简单函数 $\hat{x}(t)$ 作为 $y = x(t)$ 的近似表达式，或者 $y = x(t)$ 虽然有解析式，但是解析式较复杂，在实践中无法应用，此时可以利用一个形式简单且便于求值的函数 $\hat{x}(t)$ 去近似代替 $x(t)$ ，使得

$$\hat{x}(t_0) = y_0, \hat{x}(t_1) = y_1, \dots, \hat{x}(t_n) = y_n \quad (3-17)$$

这类问题称为插值问题。 t_0, t_1, \dots, t_n 称为插值基点， $\hat{x}(t)$ 称为插值函数，式 (3-17) 称为插值条件。

给定 $[a, b]$ 上的一串分点

$$a = t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n = b \quad (3-18)$$

和 $x(t)$ 在分点上的函数值 $x(t_i) = y_i$ 及微商值 $x'(t_i) = m_i$ ($i = 0, 1, 2, \dots, n$)，作一个分段三次埃尔米特插值函数 $H(t)$ ，并且

- (1) $H(t_i) = y_i, H'(t_i) = m_i, i = 0, 1, 2, \dots, n$;
- (2) 在每个小区间 $[t_i, t_{i+1}]$ 上是三次多项式。

作出各个点上的插值基函数

$$h_0(t) = \begin{cases} \left(1 + 2 \frac{t-t_0}{t_1-t_0}\right) \left(\frac{t-t_1}{t_0-t_1}\right)^2, & t_0 \leq t \leq t_1 \\ 0, & t_1 < t \leq t_n \end{cases}$$

$$h_i(t) = \begin{cases} \left(1 + 2 \frac{t-t_i}{t_{i-1}-t_i}\right) \left(\frac{t-t_{i-1}}{t_i-t_{i-1}}\right)^2, & t_{i-1} \leq t \leq t_i \\ \left(1 + 2 \frac{t-t_i}{t_{i+1}-t_i}\right) \left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right)^2, & t_i \leq t \leq t_{i+1} \quad i = 1, 2, \dots, n-1 \\ 0, & \text{其他} \end{cases}$$

$$h_n(t) = \begin{cases} \left(1 + 2 \frac{t-t_n}{t_{n-1}-t_n}\right) \left(\frac{t-t_{n-1}}{t_n-t_{n-1}}\right)^2, & t_{n-1} \leq t \leq t_n \\ 0, & t_0 < t \leq t_{n-1} \end{cases}$$

$$\begin{aligned}
 H_0(t) &= \begin{cases} (t-t_0)\left(\frac{t-t_1}{t_0-t_1}\right)^2, & t_0 \leq t \leq t_1 \\ 0, & t_1 < t \leq t_n \end{cases} \\
 H_i(t) &= \begin{cases} (t-t_i)\left(\frac{t-t_{i-1}}{t_i-t_{i-1}}\right)^2, & t_{i-1} \leq t \leq t_i \\ (t-t_i)\left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right)^2, & t_i < t \leq t_{i+1} \\ 0, & \text{其他} \end{cases} \quad i=1, 2, \dots, n-1 \\
 H_n(t) &= \begin{cases} (t-t_n)\left(\frac{t-t_{n-1}}{t_n-t_{n-1}}\right)^2, & t_{n-1} \leq t \leq t_n \\ 0, & t_0 < t \leq t_{n-1} \end{cases}
 \end{aligned}$$

由上面基函数的表达式，可以构造分段三次埃尔米特插值函数

$$H(x) = \sum_{i=0}^n \{y_i h_i(x) + m_i H_i(x)\} \quad (3-19)$$

下面介绍三次样条插值，前面介绍的分段三次埃尔米特插值函数的构造前提是需要同时知道插值节点上的函数值和微商值，但是在实际问题中，通过取离散点的方法往往很容易得到函数值，而插值点上的微商值就不容易得到。能否在只给出插值点上函数值的情况下构造一个整体上具有二阶连续微商的插值函数呢？三次样条插值方法很好地解决了这一问题。

同分段三次埃尔米特插值函数的构造方法类似，给定 $[a, b]$ 区间上 $n+1$ 个节点，给定点上的函数值 $x(t_i) = y_i$ ($i=0, 1, \dots, n$)，现在要构造一个三次样条插值函数 $s(t)$ ，要求满足条件

- (1) $s(t_i) = y_i, i=0, 1, 2, \dots, n$;
- (2) 在每个小区间 $[t_i, t_{i+1}]$ 上是三次多项式;
- (3) $s(t) \in C_{[a, b]}^2$.

假设在区间 $[a, b]$ 上三次样条函数 $s(t)$ 存在，并用 m_i 来表示 $s(t)$ 在点 t_i 的微商值。由于曲线通过点 (t_i, y_i) ($i=0, 1, 2, \dots, n$)，并且在每一个小区间 $[t_i, t_{i+1}]$ 上满足条件

$$\begin{aligned}
 s(t_i) &= y_i, s(t_{i+1}) = y_{i+1} \\
 s'(t_i) &= m_i, s'(t_{i+1}) = m_{i+1}
 \end{aligned}$$

故可写出小区间 $[t_i, t_{i+1}]$ 上的三次样条函数 $s(t)$ 的计算公式

$$s(t) = \left(1 + 2 \frac{t-t_i}{t_{i+1}-t_i}\right) \left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right)^2 y_i + \left(1 + 2 \frac{t-t_{i+1}}{t_i-t_{i+1}}\right) \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2 y_{i+1} \\ + (t-t_i) \left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right) m_i + (t-t_{i+1}) \left(\frac{t-t_i}{t_{i+1}-t_i}\right) m_{i+1} \quad (3-20)$$

问题是在样点 t_i ($i = 0, 1, \dots, n$) 上的微商值 m_i 并不知道, 因此如果要用公式 (3-20), 就必须设法先求出 m_0, m_1, \dots, m_n . 为了求出 m_0, m_1, \dots, m_n , 利用函数 $s(t)$ 在样点 t_i 上二阶微商连续的性质, 将式 (3-20) 对 t 求微商并令 $h_i = t_{i+1} - t_i$, 不难得到

$$s''(t) = \left(\frac{6}{h_i^2} - \frac{12}{h_i^3}(t_{i+1}-t)\right) y_i + \left(\frac{6}{h_i^2} - \frac{12}{h_i^3}(t-t_i)\right) y_{i+1} \\ + \left(\frac{2}{h_i} - \frac{6}{h_i^2}(t_{i+1}-t)\right) m_i - \left(\frac{2}{h_i} - \frac{6}{h_i^2}(t-t_i)\right) m_{i+1}$$

因为 $s(t)$ 是在区间 $[t_i, t_{i+1}]$ 上的三次多项式, 因此可以得到在区间 $[t_i, t_{i+1}]$ 上点 t_i 的右微商, 记为

$$s''(t_i^+) = -\frac{6}{h_i^2} y_i + \frac{6}{h_i^2} y_{i+1} - \frac{4}{h_i} m_i - \frac{2}{h_i} m_{i+1} \quad (3-21)$$

同样, 可以给出区间 $[t_{i-1}, t_i]$ 上点 t_i 的左微商, 记为

$$s''(t_i^-) = \frac{6}{h_{i-1}^2} y_{i-1} - \frac{6}{h_{i-1}^2} y_i + \frac{2}{h_{i-1}} m_{i-1} + \frac{4}{h_{i-1}} m_i \quad (3-22)$$

由于 $s(t)$ 二阶微商连续, 因此 $s''(t_i^+) = s''(t_i^-)$, 即

$$\frac{6}{h_{i-1}^2} y_{i-1} - \frac{6}{h_{i-1}^2} y_i + \frac{2}{h_{i-1}} m_{i-1} + \frac{4}{h_{i-1}} m_i = -\frac{6}{h_i^2} y_i + \frac{6}{h_i^2} y_{i+1} - \frac{4}{h_i} m_i - \frac{2}{h_i} m_{i+1}$$

将上式整理, 并令

$$\begin{cases} \alpha_i = \frac{h_{i-1}}{h_{i-1} + h_i} \\ \beta_i = 3 \left(\frac{1-\alpha_i}{h_{i-1}} (y_i - y_{i-1}) + \frac{\alpha_i}{h_i} (y_{i+1} - y_i) \right) \end{cases} \quad (3-23)$$

再对每个内点建立方程, 则得到方程组

$$(1-\alpha_i)m_{i-1} + 2m_i + \alpha_i m_{i+1} = \beta_i, \quad i = 1, 2, \dots, n-1 \quad (3-24)$$

这是关于 $n+1$ 个未知量 m_0, m_1, \dots, m_n 的 $n-1$ 个线性方程组, 方程组有无穷多个解, 但实际问题只能选取特定的一个解。这需要根据具体情况, 补充两个附加条件,

它们称为边界条件，常见的边界条件有

(1) 曲线在两端点 t_0, t_n 处的切线斜率已知，即 $s'(t_0) = m_0$ 及 $s'(t_n) = m_n$ 已知，方程组 (3-24) 就成为了具有 $n-1$ 个未知数 m_1, m_2, \dots, m_{n-1} 的 $n-1$ 个方程的方程组，因此它有唯一解。

(2) 函数 $y = x(t)$ 在 t_0, t_n 处二阶微商为零，即 $y_0'' = y_n'' = 0$ 或 $s_0''(t_0) = s_n''(t_n) = 0$ ，从方程 (3-21) 和 (3-22) 可得

$$\begin{cases} 2m_0 + m_1 = \frac{3}{h_0}(y_1 - y_0) \\ m_{n-1} + 2m_n = \frac{3}{h_{n-1}}(y_n - y_{n-1}) \end{cases} \quad (3-25)$$

将式 (3-24) 和 (3-25) 联立，也可唯一地解出未知数 m_0, m_1, \dots, m_n 。

(3) 函数 $y = x(t)$ 是周期函数，基本周期为 $b - a = t_n - t_0$ ，这时 $y_0 = y_n$ ，相应地要求样条函数 $s(t)$ 也是周期函数，在端点上满足条件 $s'(t_0) = s'(t_n)$ 和 $s''(t_0) = s''(t_n)$ ，于是就有

$$\begin{cases} m_0 = m_n \\ \frac{3}{h_0^2}(y_1 - y_0) - \frac{1}{h}(2m_0 + m_1) = \frac{3}{h_{n-1}^2}(y_n - y_{n-1}) + \frac{1}{h_{n-1}}(m_{n-1} + 2m_n) \end{cases} \quad (3-26)$$

把式 (3-24) 与 (3-26) 联立，也可唯一地解出未知数 m_0, m_1, \dots, m_n 。方程组 (3-24) 增补任何一种边界条件，都可以求解。

下面简单叙述一下特殊的三对角方程组的求解过程。以在两端点取第二种边界条件为例，为了符号统一起见，将式 (3-25) 改写成

$$\begin{cases} 2m_0 + \alpha_0 m_1 = \beta_0 \\ (1 - \alpha_n)m_{n-1} + 2m_n = \beta_n \end{cases}$$

其中 $\alpha_0 = 1, \alpha_n = 0, \beta_0 = \frac{3}{h_0}(y_1 - y_0), \beta_n = \frac{3}{h_{n-1}}(y_n - y_{n-1})$ 。与式 (3-24) 联立，方程组成为

$$\begin{cases} 2m_0 + \alpha_0 m_1 & = \beta_0 \\ (1 - \alpha_1)m_0 + 2m_1 + \alpha_1 m_2 & = \beta_1 \\ (1 - \alpha_2)m_1 + 2m_2 + \alpha_2 m_3 & = \beta_2 \\ \dots\dots\dots \\ (1 - \alpha_{n-1})m_{n-2} + 2m_{n-1} + \alpha_{n-1} m_n & = \beta_{n-1} \\ (1 - \alpha_n)m_{n-1} + 2m_n & = \beta_n \end{cases} \quad (3-27)$$

从式 (3-27) 的第一式, 可以得到

$$m_0 = -\frac{\alpha_0}{2}m_1 + \frac{\beta_0}{2} \triangleq a_0m_1 + b_0$$

其中 $a_0 = -\frac{\alpha_0}{2}$, $b_0 = \frac{\beta_0}{2}$. 将 m_0 的这一表达式代入 (3-27) 的第二式, 并加以整理得

$$m_1 = -\frac{\alpha_1}{2+(1-\alpha_1)a_0}m_2 + \frac{\beta_1-(1-\alpha_1)b_0}{2+(1-\alpha_1)a_0} \triangleq a_1m_2 + b_1$$

其中 $a_1 = -\frac{\alpha_1}{2+(1-\alpha_1)a_0}$, $b_1 = \frac{\beta_1-(1-\alpha_1)b_0}{2+(1-\alpha_1)a_0}$. 依次类推, 可以得到递推关系

$$m_i = a_im_{i+1} + b_i, \quad i = 0, 1, \dots, n \quad (3-28)$$

其中

$$a_i = \frac{-\alpha_i}{2+(1-\alpha_i)a_{i-1}}, b_i = \frac{\beta_i-(1-\alpha_i)b_{i-1}}{2+(1-\alpha_i)a_{i-1}}, \quad i = 1, 2, \dots, n \quad (3-29)$$

令 $m_{n+1} = 0$, 并将式 (3-29) 中 $i = n$ 的公式代入 (3-28) 中的最后一个方程得

$$m_n = \frac{\beta_n-(1-\alpha_n)b_{n-1}}{2+(1-\alpha_n)a_{n-1}} = b_n \quad (3-30)$$

这样就可以用式 (3-28) 逐个解出 m_i .

综上所述, 解方程组 (3-27) 有两步

(1) 从 $a_0 = -\frac{\alpha_0}{2}$, $b_0 = \frac{\beta_0}{2}$ 出发, 利用公式 (3-29) 计算出 a_i, b_i ($i = 1, 2, \dots, n$);

(2) 令 $m_{n+1} = 0$, 由公式 (3-28) 计算 m_i ($i = n, n-1, \dots, 0$).

最后, 把计算三次样条的步骤归纳如下

(1) 根据给定的点 (t_i, y_i) 及相应的边界条件计算方程组 (3-27) 及相应边界条件的系数 α_i 和 β_i ;

(2) 在给定的边界条件下解方程组, 计算 m_0, m_1, \dots, m_n ;

(3) 把求得的 m_i 代入公式 (3-20) 求出小区间 $[t_i, t_{i+1}]$ 上的样条函数 $s(t)$;

(4) 计算区间 $[a, b]$ 上的样条插值函数 $s(t)$.

3.5.2 GDP 季度数据处理

本节利用改进的基于 B 样条基的最小二乘拟合方法对我国 2000-2018 年共 19 年的 GDP 季度数据进行平滑处理。在估计出各省份每个季度的 GDP 值后, 再构造 B 样条基及利用最小二乘拟合方法求基函数的系数。计算过程与 3.3.1 介绍的方

法基本一致，下面叙述具体的计算过程。

(1) 对 2000-2004 年的 GDP 数据进行三次样条插值。

2000-2004 年我国 31 个省份的 GDP 年度数据参见附录 1 中的表 1，利用 MATLAB 三次样条插值函数 `spline()` 求得的 GDP 季度数据插值结果见附录 1 中的表 2。其中，为了方便对数据的处理，时间维度不以年份来表示，采用从 1 递增的自然数来表示，即 2000 年的第一季度是 1，2000 年的第二季度是 2，依次类推。

(2) 调用附录 2 的 `bbasis = create_bspline_basis([1, 76], 50, 4)` 函数生成基对象 `bbasis`;

(3) 计算 B 样条基函数在 1-76 共 76 个参数值点的函数值

`bbasismat = eval_basis([1: 76]', bbasis);`

(4) 计算最小二乘回归拟合系数

`gdpcoef = bbasismat \ GDP_2000_2018_quarter;`

(5) 创建函数型数据对象 `fdobj = fd(gdpcoef, bbasis);`

(6) 调用 `plot(fdobj)` 函数作出函数型数据对象 `fdobj` 的图像如图 3-6 和 3-7 所示;

(7) 分别计算 31 个省份 GDP 季度数据的 B 样条基函数拟合曲线残差平方和见表 3-3.

`y2cMap = (bbasismat' * bbasismat) \ bbasismat';`

`yhat = bbasismat * y2cMap * GDP_2000_2018_quarter;`

`SSE = sum((GDP_2000_2018_quarter-yhat) .* (GDP_2000_2018_quarter-yhat));`

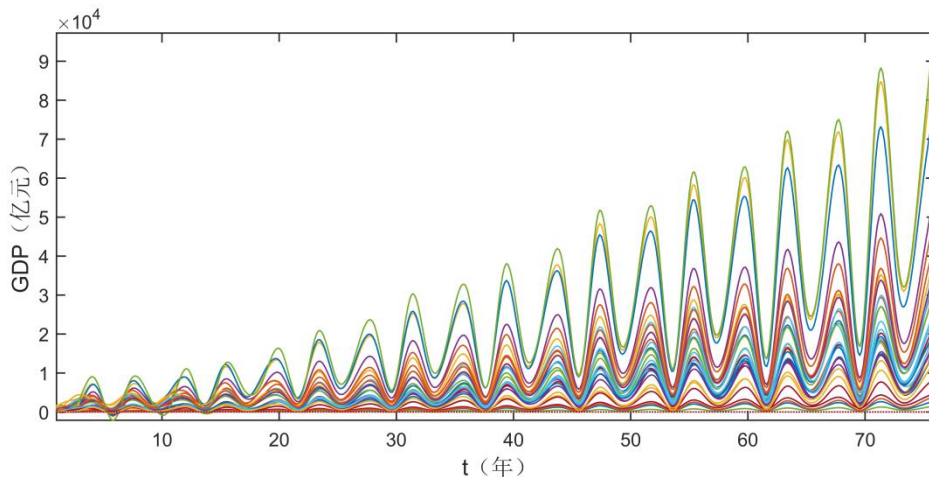


图 3-6 基于 B 样条基的我国 31 个省 GDP 季度数据拟合曲线

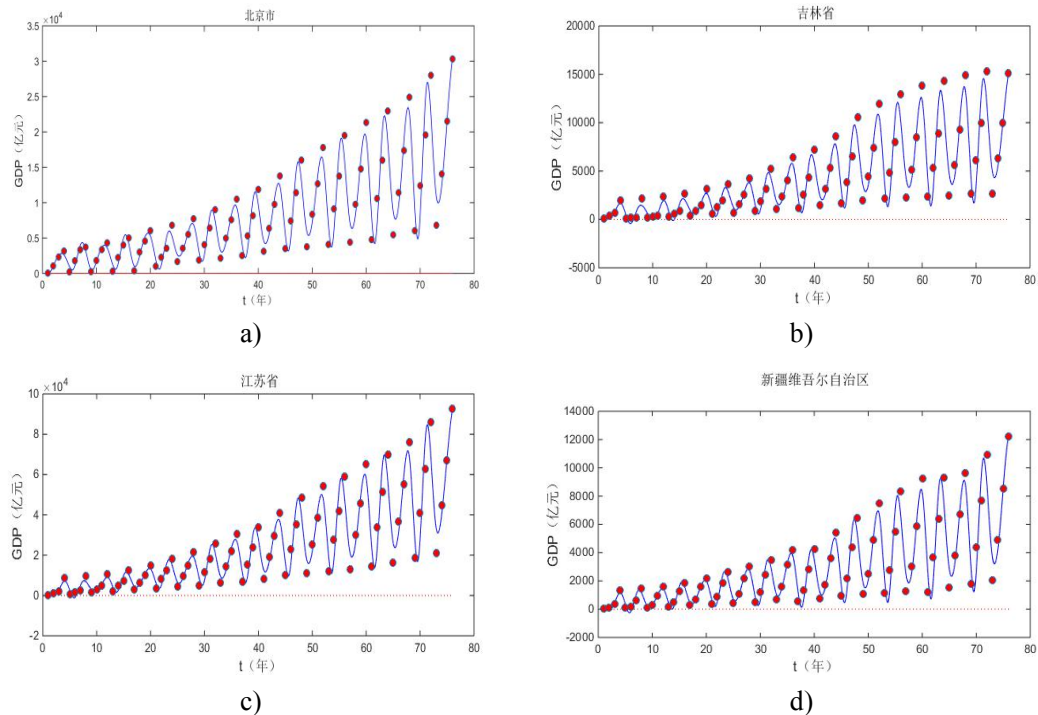


图 3-7 基于 B 样条基的我国部分省市分省 GDP 季度数据拟合曲线

图 3-7 是我国部分省市的 GDP 季度数据拟合曲线，观察图像可以发现 B 样条基函数整体上将 GDP 季度数据的循环周期变化特征表现出来了，但对于每年的第一个季度和最后一个季度的数据点拟合效果不是很好，会出现曲线超过数据点或者达不到数据点的情况，这种现象一般称为过拟合和欠拟合。

表 3-3 改进 B 样条基拟合曲线残差平方和 (1×10^9)

| 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|--------|---------|----------|---------|--------|--------|--------|--------|--------|
| 0.3433 | 0.1568 | 0.6526 | 0.1291 | 0.2065 | 0.4917 | 0.1518 | 0.1918 | 0.4286 |
| 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
| 3.1277 | 1.2956 | 0.3088 | 0.4868 | 0.1718 | 2.4869 | 0.8171 | 0.5529 | 0.5258 |
| 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
| 3.5802 | 0.0292 | 0.0085 | 0.1424 | 0.0006 | 0.0688 | 0.1240 | 0.5097 | 0.2191 |
| 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 | | | | | |
| 0.0037 | 0.0053 | 0.0571 | 0.2051 | | | | | |

将表 3-3 与表 3-1 对比，发现 B 样条基对于 GDP 的年度数据拟合效果要优于 GDP 的季度数据，这很大程度是由于季度数据的波动程度大，数据增加或减少速度快，对拟合函数的灵活度要求很高。因此，本文在进行之后的主成分分析和聚

类分析方法时选择 GDP 的年度数据平滑曲线。

3.5.3 基于傅里叶基的最小二乘拟合

由傅里叶级数定义的基展开如下

$$x(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \dots \quad (3-31)$$

其中展开式的基为 $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin r\omega t$, $\phi_{2r}(t) = \cos r\omega t$. 基是周期性的, 且参数 ω 决定的周期是 $2\pi/\omega$. 如果自变量 t_i 在区间 T 上是等间隔的并且周期恰好等于区间的长度 T , 那么基是正交的, 并且矩阵 $\Phi'\Phi$ 是对角矩阵。此外, 对基函数除以一个合适的常数可以使矩阵 $\Phi'\Phi$ 成为恒等的矩阵, 其中, 若 $i = 0$, 除以 \sqrt{n} , 若 $i \neq 0$, 除以 $\sqrt{n/2}$.

当 n 是 2 的幂次方且自变量等间隔时, 快速傅里叶变换方法可以非常有效地求解拟合函数的系数, 通过 $O(n \log n)$ 步操作可以求得所有的系数 c_k 以及平滑函数在 n 个点 t_i 的值 $x(t_i)$. 正是因为具有这样的特征, 在拟合自变量为较长的时间序列且具有循环特征的数据时, 傅里叶基往往成为首要的选择。基于傅里叶基的导数估计很简单, 因为

$$\begin{aligned} \frac{d(\sin r\omega t)}{dt} &= r\omega \cos r\omega t \\ \frac{d(\cos r\omega t)}{dt} &= -r\omega \sin r\omega t \end{aligned}$$

这意味着 $\frac{dx}{dt}$ 的傅里叶展开具有系数

$$(0, \omega c_1, -\omega c_2, 2\omega c_3, -2\omega c_4, \dots)$$

$\frac{d^2x}{dt^2}$ 的傅里叶展开具有系数

$$(0, -\omega^2 c_1, -\omega^2 c_2, -4\omega^2 c_3, -4\omega^2 c_4, \dots)$$

类似地, 通过对每个系数乘以一个 $r\omega$ 的幂次方可以求出高阶导数的傅里叶展开。在使用中, 也要注意傅里叶基的局限性, 傅立叶级数对于极其稳定的函数特别有用, 也就是没有强局部特征的函数。此外, 傅立叶级数对于基本均匀光滑的函数有很好的展开效果, 但是对于一些已知的或者函数本身及其低阶导数不连续的数据的拟合效果较差。

下面采用基于傅里叶基的最小二乘拟合方法对我国 31 个省份的 GDP 季度数据进行平滑预处理, 具体的 MATLAB 操作命令如下

(1) 调用附录 2 中的 `fbasis = create_fourier_basis([1, 76], 3, 4)` 和 `plot(fbasis)` 函数作出区间[1, 76]上的 3 个周期为四的傅里叶基函数图像如下图 3-8 所示;

(2) 计算傅里叶基函数在 1-76 共 76 个参数值点的函数值

`fbasismat = eval_basis([1: 76]', fbasis);`

(3) 计算最小二乘回归拟合系数

`gdpcoef = fbasismat \ GDP_2000_2018_quarter;`

(4) 创建函数型数据对象 `fdobj = fd(gdpcoef, fbasis);`

(5) 调用 `plot(fdobj)` 函数作出函数型数据对象 `fdobj` 的图像如图 3-9、3-10 所示, 因为傅里叶基对数据拟合效果较差, 故不再作出全部省份拟合曲线图像。

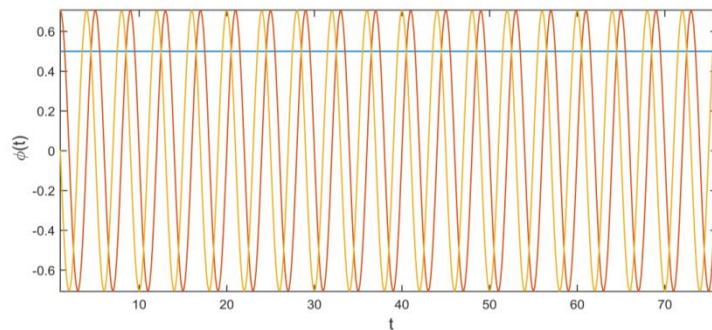


图 3-8 周期为四的傅里叶基函数

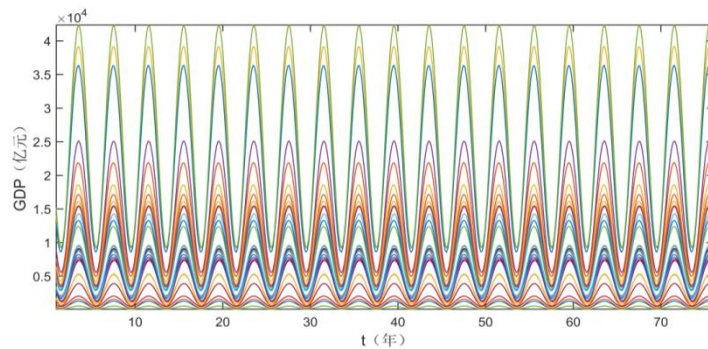


图 3-9 基于傅里叶基的我国 31 个省份 GDP 季度数据拟合曲线

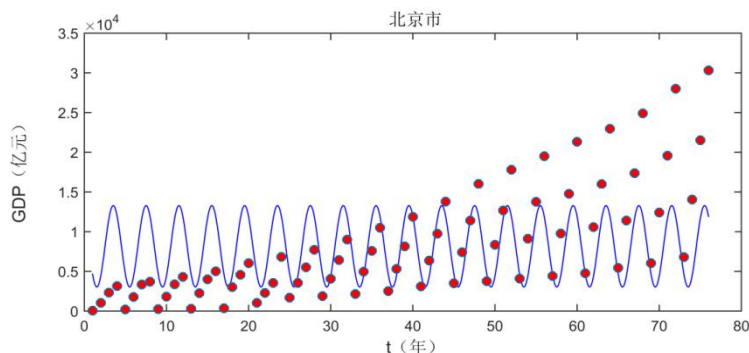


图 3-10 基于傅里叶基的北京市 GDP 季度数据拟合曲线

虽然本节采用了具有按年度周期循环特征的 GDP 季度数据来进行平滑预处理,但是从拟合曲线以及残差平方和很容易发现 B 样条基拟合效果远远优于傅里叶基拟合效果。同时观察图 3-10 也可以发现,傅里叶基函数只适用于仅仅具有周期循环特征的数据,它无法展示数据序列具有的长期递增或者递减趋势。

3.6 本章小结

本章是函数型数据分析的基础,是进行主成分分析等建模方法的前提。本章主要介绍了函数型数据预处理的几种方法,首先介绍了平滑与噪声的概念,然后介绍了几种基函数以及每一种基函数的适用范围、优势和不足,最后介绍了对我国 31 个省份 GDP 数据平滑处理的几种方法以及针对现有方法的缺陷提出的改进办法。本章通过各种基函数对 GDP 数据拟合效果的对比,发现不论是对于具有循环特征的数据还是非循环特征的数据,B 样条基都具有相对较好的拟合效果。

第 4 章 主成分分析

主成分分析最早是由统计学之父 K. Pearson 于 1901 年研究正交回归分析时所提出的^[50]。主成分分析主要通过降维的思想对多变量进行信息提取，用较少的变量来代替这些信息，其实质是把高维空间的统计转化到低维空间中，从而使得问题更容易处理。多变量问题普遍存在于实际生活中，如果变量间相关性比较大，就会使得所观测到的数据在一定程度上反映的信息有所重叠。此外，在低维空间研究问题比在高维空间研究问题简单方便。因此，在主成分分析研究中，往往希望对多变量问题的指标进行压缩处理，得到几个综合变量，要求综合变量可以代表原始数据绝大部分的信息且综合变量所代表的信息无重叠。利用这种降维的思想，产生了主成分分析统计方法，主成分分析又称主分量分析或主轴分析。

4.1 多元统计主成分分析

设 $X = (X_1, \dots, X_p)'$ 是 p 维随机向量，均值 $E(X) = \mu$ ，协方差阵 $D(X) = \Sigma$ 。考虑它的线性变换

$$\begin{cases} Z_1 = a_1'X = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ Z_2 = a_2'X = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots \\ Z_p = a_p'X = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases} \quad (4-1)$$

易见

$$\text{Var}(Z_i) = a_i' \Sigma a_i, (i = 1, 2, \dots, p) \quad (4-2)$$

$$\text{Cov}(Z_i, Z_j) = a_i' \Sigma a_j, (i, j = 1, 2, \dots, p) \quad (4-3)$$

假如我们希望用 Z_1 来代替原来的 p 个变量 X_1, \dots, X_p ，这就要求 Z_1 尽可能多地反映原来 p 个变量的信息，这里所说的“信息”采用 Z_1 的方差来表达。 $\text{Var}(Z_1)$ 越大，表示 Z_1 包含的信息越多。由 (4-2) 和 (4-3) 式可以看出，对 a_1 必须有某种限制，否则可使 $\text{Var}(Z_1) \rightarrow \infty$ ，常用的限制是 $a_1'a_1 = 1$ 。若存在满足以上约束的 a_1 ，使 $\text{Var}(Z_1)$ 达最大， Z_1 就称为第一主成分。如果第一主成分不足以代表原来 p 个变量的绝大部分信息，考虑 X 的第二个线性组合 Z_2 ，为了有效地代表原始变量的信息， Z_1 反映的信息不希望在 Z_2 中出现，用统计语言来讲，就是要求

$$\text{Cov}(Z_2, Z_1) = a_2' \Sigma a_1 = 0 \quad (4-4)$$

于是求 Z_2 就是在约束 $a_2'a_2=1$ 和 (4-4) 式下, 求 a_2 使 $\text{Var}(Z_2)$ 达到最大, 所求之 Z_2 称为第二主成分, 类似地可求得第三主成分, 第四主成分等等。

定义 4.1 设 $X = (X_1, X_2, \dots, X_p)'$ 为 p 维随机向量, 称 $Z_i = a_i'X$ 为 X 的第 i 主成分 ($i=1, 2, \dots, p$), 如果

- (1) $a_i'a_i = 1$ ($i=1, 2, \dots, p$)
- (2) 当 $i > 1$ 时, $a_i' \sum_{j=1}^{i-1} a_j = 0$ ($j=1, \dots, i-1$)
- (3) $\text{Var}(Z_i) = \max_{a'a=1, a' \sum_{j=1}^{i-1} a_j=0} \text{Var}(a'X)$

从代数学观点看, 主成分就是 p 个原始变量的一些特殊的线性组合。而从几何上看, 这些线性组合正是把由 X_1, X_2, \dots, X_p 构成的坐标系经旋转而产生的新坐标系, 新坐标轴使之通过样本变差最大的方向。此外, 对于 $a_1 = (a_{11}, a_{21}, \dots, a_{p1})$ 的求解是条件极值问题, 用拉格朗日乘子法求解, 令

$$\varphi(a_1) = \text{Var}(a_1'X) - \lambda(a_1'a_1 - 1) = a_1'\Sigma a_1 - \lambda(a_1'a_1 - 1)$$

考虑

$$\begin{cases} \frac{d\varphi}{da_1} = 2(\Sigma - \lambda I)a_1 = 0 \\ \frac{d\varphi}{d\lambda} = a_1'a_1 - 1 = 0 \end{cases} \quad (4-5)$$

因 $a_1 \neq 0$, 故 $|\Sigma - \lambda I| = 0$, 求解方程组 (4-5), 其实就是求 Σ 的特征值和特征向量问题。设 $\lambda = \lambda_1$ 是 Σ 的最大特征值, 则相应的单位特征向量 a_1 即为所求。一般地, 求 X 的第 i 主成分可通过求 Σ 的第 i 大特征值所对应的单位特征向量得到。

定理 4.1 设 X 是 p 维随机向量, 且其协方差阵 $D(X) = \Sigma$, Σ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

a_1, a_2, \dots, a_p 是对应的单位正交特征向量, 则 X 的第 i 主成分为

$$Z_i = a_i'X \quad (i=1, 2, \dots, p)$$

证明: 首先显然 Σ 是对称阵, 所以对任意非零向量 a , 有

$$\lambda_p \leq \frac{a'\Sigma a}{a'a} \leq \lambda_1$$

且最大值在 $a = a_1$ 时取得。所以在 $a_1'a_1 = 1$ 约束下

$$\text{Var}(Z_1) = \text{Var}(a_1'X) = a_1'\Sigma a_1 = \lambda_1$$

取得极大值。由定义 4.1, $Z_1 = a_1'X$ 是 X 的第一主成分。

对 $r = 2, 3, \dots, p$, 记 $\mathcal{L}_r = \mathcal{L}(a_r, \dots, a_p)$ 为 a_r, \dots, a_p 张成的向量空间, 则有

$$\max_{a \neq 0, a \in \mathcal{L}_r} \frac{a' \Sigma a}{a' a} = \lambda_r$$

且最大值在 $a = a_r$ 时取得。所以, 在 $a'_1 a_1 = 1$ 条件下, a_r 满足

$$a'_r \Sigma a_j = a'_r \lambda_j a_j = \lambda_j a'_r a_j = 0 \quad (j = 1, \dots, r-1)$$

且使得

$$\text{Var}(Z_r) = \text{Var}(a'_r X) = a'_r \Sigma a_r = \lambda_r$$

达极大值。由定义 4.1, $Z_r = a'_r X$ 为 X 的第 r 主成分。 \square

主成分分析是为了简化数据处理, 因此在实际问题应用中通常不使用 p 个主成分, 而选用 $m (m < p)$ 个主成分。 m 取多大, 这是一个很实际的问题。为此, 我们引进贡献率的概念。

定义 4.2 我们称 $\lambda_k / \sum_{i=1}^p \lambda_i$ 为主成分 Z_k 的贡献率, 称 $\sum_{k=1}^m \lambda_k / \sum_{i=1}^p \lambda_i$ 为主成分 $Z_1, \dots, Z_m (m < p)$ 的累计贡献率。通常取 m 使累计贡献率达到 70% 以上。

在实际问题中, 一般总体的协方差阵 Σ 未知, 需要通过样本来估计。设来自总体 X 的样本为 $X_{(t)} = (x_{t1}, \dots, x_{tp}) (t = 1, \dots, n)$, 记样本数据阵为

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} X'_{(1)} \\ X'_{(2)} \\ \vdots \\ X'_{(n)} \end{bmatrix}$$

那么, 样本协方差阵 S 为

$$S = \frac{1}{n-1} \sum_{t=1}^n (X_{(t)} - \bar{X})(X_{(t)} - \bar{X})' \triangleq (s_{ij})_{p \times p}$$

其中

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_{(t)} = (\bar{x}_1, \dots, \bar{x}_p)'$$

$$s_{ij} = \frac{1}{n-1} \sum_{t=1}^n (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)$$

样本相关阵 R 为

$$R = (r_{ij})_{p \times p}$$

其中

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} (i, j = 1, 2, \dots, p)$$

在实际运算中，往往利用样本协方差阵 S 作为 Σ 的估计或用 R 作为总体相关阵的估计。

4.2 函数型主成分分析

函数型数据通常是同一个观测指标的多个变量在自变量定义区间连续取值的点，若将时间变量看作多元数据中代表变化的因素，则函数型数据面临更严重的“维数灾难”。基于这一特点，将主成分分析的技术引入到函数型数据分析中，即函数型数据主成分分析，以充分地对数据进行降维及信息提取。由于各种原因，函数型主成分分析是进行函数型数据分析时需要首要考虑的关键技术。

通常，函数型数据的主成分分析是函数型数据分析中的关键技术之一。这是因为在对数据进行平滑预处理之后，我们往往希望进一步探索该数据，以了解函数所表现出来的典型特征。在研究中，我们期望某些特征是存在的，例如对于温度曲线，希望它具有按正弦函数周期循环的性质，这样才符合实际情况。主成分分析方法主要用来挖掘数据中存在的潜在特征，因此在许多关于函数型数据分析的文献中，主成分分析方法被首先考虑。

函数型数据主成分分析的思想来源于多元统计主成分分析，但是在变量的形式上与多元统计主成分分析有很大的区别。就像多元统计主成分分析中的相应矩阵一样，往往很难对函数型主成分分析中的方差—协方差矩阵和相关函数进行合理的解释，并且通常不能直接定义出可以完全表示观测数据可变性的一种结构。当然，对于古典的多元统计主成分分析方法，也无法做到这一点。主成分分析提供了一种观察协方差结构的方法，它能提供更多的数据所隐含的补充信息。

4.2.1 函数型主成分分析定义

函数型主成分主要是通过多个函数的组合得到的，相应于多元统计主成分分析中的变量值 x_{ij} ，函数型主成分分析中对应的是函数值 $x(s)$ 。多元统计主成分分析中的第 k 个主成分向量定义形式为

$$Z_k = a'_k X = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p = \sum_j a_{jk} X_j$$

当 a_{jk} 和 X_j 变成函数 $\xi_k(s)$ 和 $x(s)$ 时，多元统计主成分分析中的内积求和表达式应该替换为函数的内积表达式(2-7)所定义的形式，在这里同样假设函数 $\xi_k(s)$ 和 $x(s)$ 以及它们的乘积均在区间 $[a, b]$ 上可积，得到如下的结果，表示形式为

$$f_k = \int_a^b \xi_k(s)x(s)ds \quad (4-6)$$

此时, 权重向量的分量值 a_{jk} 变为连续函数值 $\xi_k(s)$. $\xi_k(s)$ 所对应的主成分得分分量值为

$$f_{kj} = \int_a^b \xi_k(s)x_j(s)ds \quad (4-7)$$

函数型主成分分析的第一步是要计算第一主成分的权重函数 $\xi_1(s)$. 具体求解方法是在 $\int_a^b \xi_1(s)^2 ds = 1$ 的条件约束下, 求出 $\xi_1(s)$, 使

$$\frac{1}{P} \sum_{j=1}^P f_{1j}^2 = \frac{1}{P} \sum_{j=1}^P \left(\int_a^b \xi_1(s)x_j(s)ds \right)^2 \quad (4-8)$$

取得最大值. 通常, 也使用符号 $\|\xi_1\|^2$ 来代替符号 $\int_a^b \xi_1(s)^2 ds$. 对于主成分权重函数的具体计算过程将在下面详细讨论. 通过前面计算过程可以得到第一主成分, 在接下来的计算中, 权重函数 $\xi_m(s)$ 需要满足正交条件约束, 即

$$\int_a^b \xi_n(s)\xi_m(s)ds = 0, \quad n < m \quad (4-9)$$

在计算中, 每一个权重函数都要包含剩余曲线中最重要的变化信息, 并且必须保证与之前步骤已经表示出的变化信息相互正交, 也即与之前定义的权重函数正交. 这样可以保证不同主成分所表示的曲线信息互相不重叠.

下面介绍通过基函数的展开来计算函数型主成分的原理, 前面我们得到了 P 个指标函数的集合 $x_j (j=1, \dots, P)$, 假设这些曲线已经进行了中心化, 即把每条曲线减去均值曲线. 对比多元统计主成分分析方法, 函数型主成分分析的权重函数主要通过指标函数的协方差函数得到. 指标函数的协方差函数定义为

$$v(s, t) = \frac{1}{P} \sum_{j=1}^P x_j(s)x_j(t) \quad (4-10)$$

与多元统计主成分分析类似, 函数型主成分分析的每个权重函数满足如下特征方程

$$\int_a^b v(s, t)\xi(t)dt = \rho\xi(s) \quad (4-11)$$

其中, ρ 是特征值, $\xi(s)$ 是与 ρ 相对应的特征函数. 又因为指标函数 x_j 是基函数的线性组合, 也即

$$x_j(t) = \sum_{k=1}^K c_{jk} \phi_k(t)$$

于是可以得到

$$v(s, t) = \frac{1}{P} \sum_{j=1}^P \left(\sum_{k=1}^K c_{jk} \phi_k(s) \right) \left(\sum_{k=1}^K c_{jk} \phi_k(t) \right) = \frac{1}{P} \phi(s)' C' C \phi(t) \quad (4-12)$$

其中

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1K} \\ c_{21} & c_{22} & \cdots & c_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pK} \end{pmatrix}, \quad \Phi(s) = \begin{pmatrix} \phi_1(s) \\ \phi_2(s) \\ \vdots \\ \phi_K(s) \end{pmatrix}, \quad \Phi(t) = \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_K(t) \end{pmatrix}$$

假设 $\xi(s)$ 的基函数展开表达式为

$$\xi(s) = \sum_{k=1}^K b_k \phi_k(s) \quad (4-13)$$

定义基函数的 K 阶对称内积阵为

$$W = \int_a^b \Phi(t) \Phi'(t) dt \quad (4-14)$$

其元素是

$$w_{k_1, k_2} = \int_a^b \phi_{k_1}(t) \phi_{k_2}(t) dt$$

对于某些基函数, W 很容易求得。例如, 如果采用正交傅里叶基函数来拟合原始数据, 那么 $W = I$, 即是 K 阶单位阵。在采用其它基函数的情况下, 也可以采用数值积分的方法求出矩阵 W 。

下面推导本节的一个重要的结论公式, 将式 (4-12) 和 (4-13) 代入式 (4-11) 中得

$$\int_a^b \frac{1}{P} \Phi(s)' C' C \Phi(t) \left(\sum_{k=1}^K b_k \phi_k(t) \right) dt = \rho \sum_{k=1}^K b_k \phi_k(s)$$

将上式中的 $\sum_{k=1}^K b_k \phi_k(t)$ 写成向量形式为

$$\sum_{k=1}^K b_k \phi_k(t) = \Phi(t)' \mathbf{b}$$

其中

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{pmatrix}$$

所以, 得到下面公式

$$\int_a^b \frac{1}{P} \Phi(s)' C' C \Phi(t) \Phi(t)' b dt = \Phi(s)' \frac{1}{P} C' C W b = \rho \Phi(s)' b \quad (4-15)$$

由于此方程对所有的 s 都成立, 故式 (4-15) 的简化形式为

$$\frac{1}{P} C' C W b = \rho b \quad (4-16)$$

由 $\|\xi\|^2 = 1$ 可以推出

$$\|\xi\|^2 = \int_a^b \xi(s)^2 ds = \int_a^b b' \Phi(s) \Phi(s) b ds = b' W b = 1$$

并且由上式类似可得, 如果两个权重函数 $\xi_m(s)$ 和 $\xi_n(s)$ 正交, 那么当且仅当系数向量满足 $b_m' W b_n = 1$. 为了求解主成分, 定义 $u = W^{1/2} b$ 用来求解与式 (4-16) 等价的对称特征值问题

$$\frac{1}{P} W^{1/2} C' C W^{1/2} u = \rho u \quad (4-17)$$

通过计算 $b = W^{-1/2} u$ 可以求得权重函数 $\xi(s)$ 的展开式系数向量 b , 在得到系数向量 b 后, 由式 (4-13) 可求得每个主成分的权重函数 $\xi(s)$.

4.2.2 主成分分析和特征值分析

本节讨论函数型主成分分析的另一个重要特征, 即方差协方差函数的特征值分析。首先协方差函数 $v(s, t)$ 的定义与式 (4-10) 一致, 这里需要注意, 通常有些文献可能会更喜欢使用 $P-1$ 来代替 (4-10) 式中的 P , 不过在本节我们所讨论的问题中, 这两种定义方法没有本质区别。

对于每一个合适的特征值 ρ , 都有一个权重函数满足方程 (4-11), 式 (4-11) 的左侧是权重函数 ξ 的一个积分变换 V , 定义为

$$V\xi = \int_a^b v(\cdot, t) \xi(t) dt \quad (4-18)$$

这样的积分变换称为协方差算子, 所以特征方程也可表示为

$$V\xi = \rho\xi \quad (4-19)$$

此时的 ξ 是一个函数, 而不是特征向量。总之, 函数型数据的主成分分析主要是求解一组相互正交的归一化权重函数, 可以表示为有关协方差算子的特征值和特征函数求解问题。

4.2.3 函数型主成分分析推广

前面介绍了一元函数型数据的主成分分析原理, 然而在实际问题中, 二元及

多元数据的主成分分析应用更为普遍。但是目前函数型数据主成分分析方法的理论研究尚在起步阶段，对于多元函数型数据的主成分分析研究缺乏相关资料。J.O.Ramsay 在文献^[3]中提出关于多元函数型数据的主成分分析原理类似一元函数型主成分分析。下面叙述二元函数型主成分分析的基本原理，多元函数型主成分分析的原理可以仿照二元函数型主成分分析相关理论得到。假设二元函数型变量分别为 X 和 Y ，其中 $X = (x_1, x_2, \dots, x_p)$ ，即包含 P 个拟合曲线， $Y = (y_1, y_2, \dots, y_p)$ ，同样包含 P 个拟合曲线。 X 和 Y 的均值函数分别为 X_{mn} 和 Y_{mn} 。定义

$$v_{XX}(s, t) = \frac{1}{P} \sum_{j=1}^P x_j(s) x_j(t) \quad (4-20)$$

是函数 X_j 的协方差函数，定义

$$v_{YY}(s, t) = \frac{1}{P} \sum_{j=1}^P y_j(s) y_j(t) \quad (4-21)$$

是函数 Y_j 的协方差函数。此外，定义

$$v_{XY}(s, t) = \frac{1}{P} \sum_{j=1}^P x_j(s) y_j(t) \quad (4-22)$$

为交叉协方差函数，并且如下等式成立。

$$v_{XY}(s, t) = v_{YX}(t, s) \quad (4-23)$$

仿照式 (4-7) 可以得到二元函数型数据第 k 个主成分得分函数为

$$f_{kj} = \int_a^b \xi^X(s) x_j(s) ds + \int_a^b \xi^Y(s) y_j(s) ds \quad (4-24)$$

仿照式 (4-11) 可以得到二元函数型数据主成分的特征方程为

$$\int_a^b v_{XX}(s, t) \xi^X(t) dt + \int_a^b v_{XY}(s, t) \xi^Y(t) dt = \rho \xi^X(s) \quad (4-25)$$

$$\int_a^b v_{YY}(s, t) \xi^Y(t) dt + \int_a^b v_{YX}(s, t) \xi^X(t) dt = \rho \xi^Y(s) \quad (4-26)$$

求解上述特征方程，可以得到权重函数 $\xi^X(s)$ 和 $\xi^Y(s)$ 。

4.3 GDP 数据主成分分析

前面分别介绍了多元统计主成分分析和函数型主成分分析的原理，本节将这两种主成分分析方法应用到我国 1978-2018 年的分省 GDP 年度数据中，分析两种不同方法所得到的结果。其中 GDP 数据参见附录 1 中的表 1。

(1) GDP 数据的多元统计主成分分析

利用 SPSS 软件对附录 1 中的表 1 分省 GDP 年度数据进行多元统计主成分分析，输出如下结果，其中具体的软件操作过程可以参考文献[51]，此处不再叙述操作流程。

表 4-1 总方差分解表

| 成分 | 初始特征值 | | | 提取载荷平方和 | | |
|----|--------|--------|---------|---------|--------|--------|
| | 总计 | 方差百分比 | 累积百分比 | 总计 | 方差百分比 | 累积百分比 |
| 1 | 25.499 | 82.257 | 82.257 | 25.499 | 82.257 | 82.257 |
| 2 | 5.287 | 17.055 | 99.313 | 5.287 | 17.055 | 99.313 |
| 3 | 0.189 | 0.610 | 99.923 | | | |
| 4 | 0.010 | 0.032 | 99.955 | | | |
| 5 | 0.008 | 0.026 | 99.981 | | | |
| 6 | 0.003 | 0.010 | 99.990 | | | |
| 7 | 0.001 | 0.003 | 99.994 | | | |
| 8 | 0.001 | 0.003 | 99.997 | | | |
| 9 | 0.001 | 0.003 | 100.000 | | | |
| 10 | 0.000 | 0.000 | 100.000 | | | |
| 11 | 0.000 | 0.000 | 100.000 | | | |
| 12 | 0.000 | 0.000 | 100.000 | | | |
| 13 | 0.000 | 0.000 | 100.000 | | | |
| 14 | 0.000 | 0.000 | 100.000 | | | |
| 15 | 0.000 | 0.000 | 100.000 | | | |
| 16 | 0.000 | 0.000 | 100.000 | | | |
| 17 | 0.000 | 0.000 | 100.000 | | | |
| 18 | 0.000 | 0.000 | 100.000 | | | |
| 19 | 0.000 | 0.000 | 100.000 | | | |
| 20 | 0.000 | 0.000 | 100.000 | | | |
| 21 | 0.000 | 0.000 | 100.000 | | | |
| 22 | 0.000 | 0.000 | 100.000 | | | |
| 23 | 0.000 | 0.000 | 100.000 | | | |
| 24 | 0.000 | 0.000 | 100.000 | | | |
| 25 | 0.000 | 0.000 | 100.000 | | | |
| 26 | 0.000 | 0.000 | 100.000 | | | |
| 27 | 0.000 | 0.000 | 100.000 | | | |
| 28 | 0.000 | 0.000 | 100.000 | | | |
| 29 | 0.000 | 0.000 | 100.000 | | | |
| 30 | 0.000 | 0.000 | 100.000 | | | |
| 31 | 0.000 | 0.000 | 100.000 | | | |

提取方法：主成分分析

表 4-1 所示为总方差分解表，其中“成分”中列出的是各成分的序号，“初始

特征值”是协方差矩阵的特征值，这些值用来确定哪些成分应当保留，共有三项。其中，“总计”中列出了各成分的特征值，第一成分特征值为 25.499，第二成分特征值为 5.287，第三成分特征值为 0.189。“方差百分比”列出的是各成分所解释的方差占总方差的百分比，也就是各因子特征值占特征值总和的百分比。“累积百分比”自上而下列出的是各因子方差占总方差百分比的累积百分比，前两个主成分的特征值之和占总特征值之和的 99.313%，也即前两个主成分解释了原始 31 个变量的 99.313%的变异。“提取载荷平方和”是未经旋转的因子载荷的平方和，它给出的是每个成分的特征值所解释的方差占总方差的百分比和累积百分比。从结果可以看出，使用前两个主成分已经解释了总方差的 99.313%，因此，使用这两个主成分在很大程度上减少了原始数据的复杂性，仅丢失了 0.687%的信息。

表 4-2 旋转后因子载荷矩阵表

| | 成分 | |
|----------|-------|-------|
| | 1 | 2 |
| 贵州省 | 0.732 | 0.554 |
| 西藏自治区 | 0.711 | 0.579 |
| 重庆市 | 0.786 | 0.616 |
| 湖北省 | 0.775 | 0.631 |
| 福建省 | 0.771 | 0.631 |
| 安徽省 | 0.767 | 0.638 |
| 云南省 | 0.764 | 0.642 |
| 四川省 | 0.761 | 0.646 |
| 海南省 | 0.760 | 0.646 |
| 陕西省 | 0.759 | 0.649 |
| 江西省 | 0.757 | 0.651 |
| 湖南省 | 0.754 | 0.655 |
| 江苏省 | 0.853 | 0.653 |
| 宁夏回族自治区 | 0.744 | 0.667 |
| 北京市 | 0.738 | 0.664 |
| 广东省 | 0.934 | 0.667 |
| 新疆维吾尔自治区 | 0.732 | 0.676 |
| 广西壮族自治区 | 0.732 | 0.679 |
| 河南省 | 0.731 | 0.675 |
| 青海省 | 0.725 | 0.688 |
| 天津市 | 0.724 | 0.688 |
| 浙江省 | 0.821 | 0.682 |
| 山东省 | 0.911 | 0.697 |
| 上海市 | 0.806 | 0.689 |

表 4-2 (续表)

| | 成分 | |
|--------|-------|-------|
| | 1 | 2 |
| 辽宁省 | 0.553 | 0.631 |
| 内蒙古自治区 | 0.611 | 0.690 |
| 黑龙江省 | 0.624 | 0.672 |
| 吉林省 | 0.671 | 0.640 |
| 河北省 | 0.668 | 0.636 |
| 山西省 | 0.672 | 0.629 |
| 甘肃省 | 0.699 | 0.612 |

提取方法：主成分分析

旋转方法：Kaiser 标准化最大方差法

旋转在 4 次迭代后已收敛

表 4-2 是旋转后的因子载荷矩阵，表下方是有关因子提取与旋转方法的说明，即使用了主成分法提取因子，使用最大方差法旋转，经过 4 次迭代达到收敛。从表中可以看出，第一个主成分和“江苏省”、“广东省”、“山东省”、“浙江省”、“上海市”这几个地区有相对较大的相关系数，第二个主成分和 31 个地区的相关系数差别很小，都在 0.6-0.7 之间。根据这些变量的原始含义可以对两个主成分命名，第一个主成分主要概括了经济发达省份的 GDP 与我国 GDP 整体平均水平的差异，第二个主成分主要概括了我国 GDP 发展的整体变化趋势。

(2) GDP 数据的函数型主成分分析

分别调用附录 2 中的 `plot (mean (fdobj))`、`pcastr = pca_fd (fdobj, 2)` 和 `plot_pca_fd (pcastr)` 函数作出 GDP 均值函数和前两个主成分特征函数图像如图 4-1，4-2，4-3 和 4-4 所示。调用 `eigvals = pcastr.values` 函数得到前两个主成分的特征值分别为 2.8385 和 0.0131。调用 `varprop = pcastr.varprop` 函数得到前两个主成分的分差贡献率分别为 89.6%和 7.1%，前两个主成分的分差累计贡献率已经达到 96.7%，足以代表大部分的信息，因此选择前两个主成分。

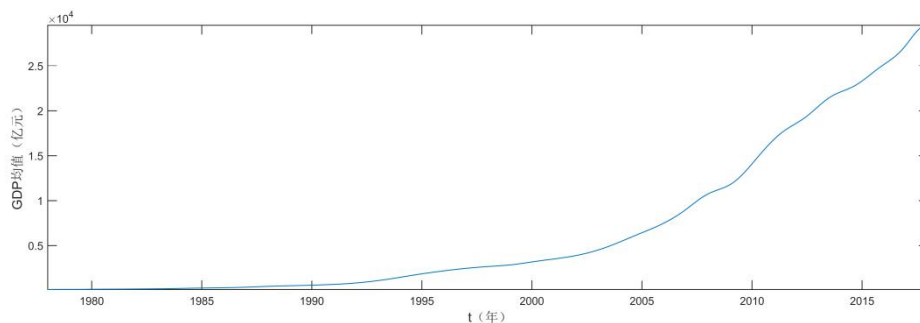


图 4-1 GDP 均值曲线

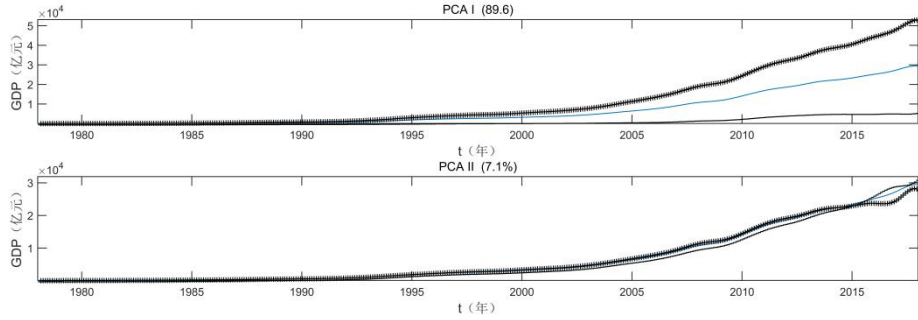


图 4-2 前两个主成分特征函数曲线

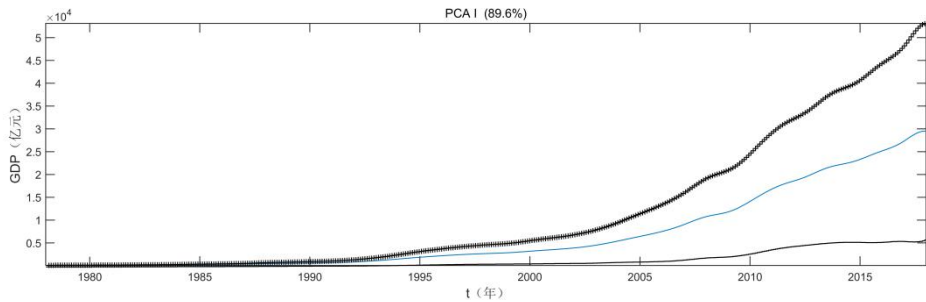


图 4-3 第一主成分特征函数曲线

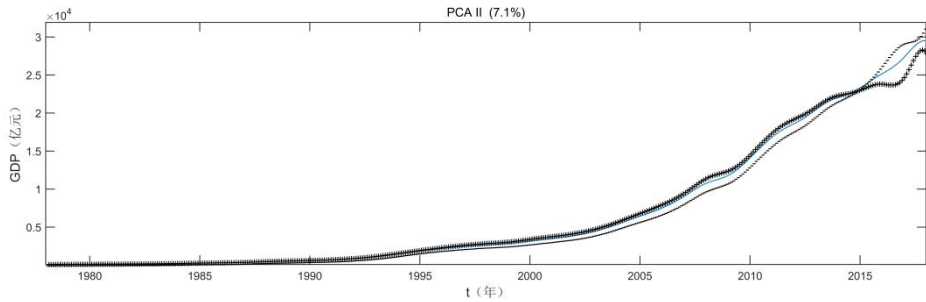


图 4-4 第二主成分特征函数曲线

图 4-3 所示为第一主成分特征函数曲线图像，其中位于中间的线条是我国 31 个省份 1978-2018 年 GDP 的平均值函数曲线，位于均值曲线上边，用“+”符号标识的曲线是由均值曲线加上第一个特征函数（第一主成分）的 2 倍所得到的，位于均值曲线下边，用“-”符号标识的曲线是由均值曲线减去第一个特征函数的 2 倍得到的。图像很好地展示了第一主成分相对于平均值的波动情况。从图像可以看到从 1995 年开始，第一特征函数随着时间增长呈现递增趋势，进而可以得到从 1995 年开始我国 31 个省份之间的 GDP 水平呈现越来越大的差异，这可能是由于某些经济政策或者地区发展不平衡的因素造成的。图 4-4 所示为第二主成分特征函数曲线图像，同样包括均值函数曲线、均值函数加 2 倍的第二特征函数曲线和均值函数减 2 倍的第二特征函数曲线。从图像可以看到这三条曲线在整个区间基本

重合，因此可以推断第二主成分主要代表 1978-2018 年间我国各省市 GDP 发展的平均变化水平。

分析图 4-3 和图 4-4 所得到的结果与前面表 4-2 所得到的结果有很大的相似性，从 1992 年邓小平南巡讲话后，我国改革开放进入新阶段，经济发展迈入新征程。东部某些省市如“江苏省”、“广东省”、“浙江省”、“山东省”等凭借着国家政策的大力支持以及得天独厚的地理位置条件高速发展经济，带动 GDP 发展水平远远超过全国平均水平。经济发达省份的带动是全国 GDP 均值波动的主要影响因素。

4.4 本章小结

本章主要介绍了两种主成分分析方法，即经典的多元统计主成分分析方法和函数型主成分分析方法。重点对函数型主成分分析方法的定义及具体计算细节进行了探讨。在最后分别把这两种方法应用到我国 1978-2018 年的 GDP 分省年度数据中，分析结果显示虽然多元统计主成分分析方法和函数型主成分分析方法从不同的角度处理数据，但是得到的结果具有某些共性。

第 5 章 聚类分析

分类学是一种经典的学科方法，是人类了解世界的基础科学。聚类分析是一种专门研究对事物进行分类的方法。目前，随着计算机技术的高速发展，聚类分析也从传统的工农业生产等领域逐渐被广泛地应用于数据挖掘和人工智能等新兴领域。聚类分析的主要思想是根据事物本身的特性研究个体分类，分类后的结果要达到同一类中的个体有较高相似性，而不同类中的个体相差很大。根据分类对象的不同，聚类分析又分为样品聚类和变量聚类。样品聚类一般也称为 Q 型聚类，是一种对样品或观测进行分类的方法。在统计学中，变量聚类一般也称为 R 型聚类，是一种对变量进行分类的方法。变量聚类和样品聚类在实践中都被广泛应用。

5.1 系统聚类分析

聚类分析又称群分析，它是研究对样品或变量进行分类的一种多元统计方法，所谓的“类”，通俗地说就是相似元素的集合。聚类分析的功能是建立一种分类方法，它将一批样品或变量按照它们在性质上的亲疏、相似程度进行分类。系统聚类法是目前在实际应用中使用较为普遍的一种方法，它是将类由多变到少的一种方法。在聚类过程中，开始时每个对象自成一类，然后每次将最相似的两类合并，合并后重新计算新类与其它类的距离或相近性测度。持续这一过程直到所有对象归为一类为止。本章对变量进行聚类，具体做法如下。

设有 P 个变量，对每个变量测得 n 个观测值，得到观测数据 x_{ij} ，其中 $i=1, \dots, n$ ， $j=1, \dots, P$ 。 P 个 n 元变量记为 $X_j (j=1, \dots, P)$ 。聚类过程的基本步骤如下

(1) 标准化变换

对原始数据进行标准化变换

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5-1)$$

其中

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

变换后的数据，每个变量的样本均值为 0，标准差为 1，而且标准化变换后的数据

$\{x_{ij}^*\}$ 与变量的量纲无关。

(2) 计算 P 个变量两两之间的距离, 得变量间的距离矩阵 $D^{(0)}$. 其中, 采用欧式距离度量变量 X_i 与 X_j 之间的距离

$$d_{ij} = \sqrt{\sum_{t=1}^n |x_{ti} - x_{tj}|^2} \quad (5-2)$$

其中 $i, j = 1, 2, \dots, P$.

(3) 初始时 P 个变量各自构成一类, 此时类的个数 $K = P$, 第 j 类 $G_j = \{X_j\}$ ($j = 1, \dots, P$). 此时类间的距离就是变量间的距离, 也即 $D^{(1)} = D^{(0)}$. 然后对变量 X_j ($j = 2, \dots, P$) 执行并类过程的步骤 (4) 和 (5). 其中类 G_p 与 G_q 之间的距离度量采用最短距离法

$$D_{pq} = \min_{i \in G_p, j \in G_q} d_{ij}$$

上式中, $i \in G_p$ 表示 $X_i \in G_p$, $j \in G_q$ 表示 $X_j \in G_q$. 此种定义方式把类与类间的距离定义为两类中相距最近的样品之间的距离。

(4) 对步骤 (3) 得到的距离矩阵 $D^{(j-1)}$, 合并类间距离最小的两类为一新类, 此时类的总个数 K 减少 1 类, 即

$$K = P - j + 1$$

(5) 计算新类与其他类的距离, 得新的距离矩阵 $D^{(j)}$. 若合并后类的总个数 K 仍大于 1, 重复步骤 (4) 和 (5), 直到类的总个数为 1 时转到步骤 (6).

(6) 画谱系聚类图。

(7) 决定分类的个数及各类的成员。

5.2 函数型聚类分析

无论是在自然科学还是社会科学的研究中, 聚类分析都是一种重要的研究方法, 同时这种方法也是一个重要的研究工具。因为存在维数灾难问题, 经典的聚类分析方法一般用于处理有限维、等间隔取样的离散数据。目前, 随着信息化技术的高速发展和数据采集方式的变化, 越来越多的数据呈现出无穷维、随机间隔、连续动态的特征。经典的聚类分析方法对于这种数据无能为力。

本文第三章介绍了利用基函数和最小二乘法把离散的数据拟合成平滑函数的方法, 本节基于这些曲线的相似度进行聚类。第 i 条曲线 $x_i(t)$ 与第 j 条曲线 $x_j(t)$ 之间的相似度采用式 (2-10) 所定义的曲线距离来度量, 也即欧氏距离

$$d_{ij} = \|x_i - x_j\| = \sqrt{\int_a^b (x_i(t) - x_j(t))^2 dt} \quad (5-3)$$

将 $x_i(t)$ 与 $x_j(t)$ 用相同的 K 维样条基函数 $\Phi(t)$ 展开, 结合式 (4-14) 得

$$\begin{aligned} d_{ij}^2 &= \int_a^b (c_i' \Phi(t) - c_j' \Phi(t))^2 dt \\ &= \int_a^b ((c_i' - c_j') \Phi(t))^2 dt \\ &= (c_i' - c_j') \int_a^b \Phi(t) \Phi'(t) dt (c_i' - c_j')' \\ &= (c_i' - c_j') W (c_i' - c_j')' \end{aligned} \quad (5-4)$$

其中

$$\Phi(t) = \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_K(t) \end{pmatrix}, \quad c_i = \begin{pmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{iK} \end{pmatrix}, \quad c_j = \begin{pmatrix} c_{j1} \\ c_{j2} \\ \vdots \\ c_{jK} \end{pmatrix}$$

在得到了 P 条拟合函数两两之间的距离后, 根据距离最小原则对曲线进行聚类。

文献[37,40,41]对于曲线之间的距离度量采用的都是 (5-3) 式所定义的形式。本节提出一种新的距离度量方法, 即定义第 i 条曲线 $x_i(t)$ 与第 j 条曲线 $x_j(t)$ 之间的距离为

$$\cos \theta = \frac{\int_a^b x_i(t) x_j(t) dt}{\sqrt{\left(\int_a^b x_i(t)^2 dt \right) \left(\int_a^b x_j(t)^2 dt \right)}} \quad (5-5)$$

式 (5-5) 也可以称作曲线 $x_i(t)$ 与曲线 $x_j(t)$ 的夹角余弦。

仿照式 (5-4), 将 $x_i(t)$ 与 $x_j(t)$ 用相同的 K 维样条基函数 $\Phi(t)$ 展开, 结合式 (4-14) 得

$$\begin{aligned} \cos \theta &= \frac{\int_a^b x_i(t) x_j(t) dt}{\sqrt{\left(\int_a^b x_i(t)^2 dt \right) \left(\int_a^b x_j(t)^2 dt \right)}} \\ &= \frac{\int_a^b c_i' \Phi \Phi' c_j dt}{\sqrt{\int_a^b c_i' \Phi \Phi' c_i dt} \cdot \sqrt{\int_a^b c_j' \Phi \Phi' c_j dt}} \\ &= \frac{c_i' W c_j}{\sqrt{c_i' W c_i} \cdot \sqrt{c_j' W c_j}} \end{aligned} \quad (5-6)$$

在得到了 P 个拟合函数两两之间的距离后, 由距离最小原则对曲线进行聚类。

5.3 GDP 数据聚类分析

前面分别介绍了系统聚类分析和函数型聚类分析的原理，本节将这两种聚类分析方法应用到我国 1978-2018 年的 GDP 分省年度数据中，分析两种不同方法所得到的结果。其中 GDP 数据参见附录 1 中的表 1。

(1) GDP 数据的系统聚类分析

利用 SPSS 软件对附录 1 中的表 1 所示的 GDP 分省年度数据进行系统聚类分析，输出如下结果，其中具体的软件操作过程可以参考文献[51]，此处不再叙述操作流程。

表 5-1 变量的距离矩阵

| | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|--------|-------|-------|-------|-------|--------|-------|-------|-------|-------|
| 北京市 | 0.000 | 0.668 | 0.751 | 0.762 | 1.394 | 1.808 | 0.833 | 1.347 | 0.628 |
| 天津市 | 0.668 | 0.000 | 0.946 | 0.971 | 1.026 | 1.690 | 0.621 | 1.500 | 1.146 |
| 河北省 | 0.751 | 0.946 | 0.000 | 0.399 | 1.075 | 1.177 | 0.516 | 0.653 | 0.561 |
| 山西省 | 0.762 | 0.971 | 0.399 | 0.000 | 1.118 | 1.293 | 0.685 | 0.925 | 0.727 |
| 内蒙古自治区 | 1.394 | 1.026 | 1.075 | 1.118 | 0.000 | 1.108 | 0.735 | 1.380 | 1.551 |
| 辽宁省 | 1.808 | 1.690 | 1.177 | 1.293 | 1.108 | 0.000 | 1.138 | 0.945 | 1.639 |
| 吉林省 | 0.833 | 0.621 | 0.516 | 0.685 | 0.735 | 1.138 | 0.000 | 0.946 | 0.951 |
| 黑龙江省 | 1.347 | 1.500 | 0.653 | 0.925 | 1.380 | 0.945 | 0.946 | 0.000 | 0.938 |
| 上海市 | 0.628 | 1.146 | 0.561 | 0.727 | 1.551 | 1.639 | 0.951 | 0.938 | 0.000 |
| 江苏省 | 0.327 | 0.511 | 0.932 | 0.893 | 1.388 | 1.908 | 0.876 | 1.536 | 0.907 |
| 浙江省 | 0.216 | 0.717 | 0.566 | 0.584 | 1.311 | 1.657 | 0.732 | 1.168 | 0.501 |
| 安徽省 | 0.526 | 0.573 | 1.066 | 1.025 | 1.490 | 1.985 | 0.970 | 1.632 | 1.061 |
| 福建省 | 0.458 | 0.676 | 1.079 | 1.050 | 1.582 | 2.047 | 1.044 | 1.648 | 0.991 |
| 江西省 | 0.463 | 0.446 | 1.000 | 0.964 | 1.366 | 1.916 | 0.876 | 1.587 | 1.029 |
| 山东省 | 0.338 | 0.556 | 0.514 | 0.593 | 1.115 | 1.524 | 0.533 | 1.122 | 0.618 |
| 河南省 | 0.227 | 0.569 | 0.669 | 0.660 | 1.267 | 1.696 | 0.705 | 1.278 | 0.671 |
| 湖北省 | 0.626 | 0.565 | 1.192 | 1.154 | 1.515 | 2.076 | 1.045 | 1.762 | 1.190 |
| 湖南省 | 0.531 | 0.368 | 1.006 | 0.996 | 1.318 | 1.875 | 0.824 | 1.575 | 1.070 |
| 广东省 | 0.143 | 0.744 | 0.701 | 0.706 | 1.416 | 1.784 | 0.845 | 1.289 | 0.544 |
| 甘肃省 | 0.599 | 0.635 | 0.445 | 0.582 | 1.044 | 1.307 | 0.397 | 0.951 | 0.734 |
| 海南省 | 0.488 | 0.562 | 0.964 | 0.974 | 1.459 | 1.884 | 0.876 | 1.497 | 0.957 |
| 重庆市 | 0.740 | 0.641 | 1.336 | 1.320 | 1.639 | 2.227 | 1.171 | 1.896 | 1.304 |
| 西藏自治区 | 0.834 | 0.989 | 1.539 | 1.485 | 1.949 | 2.481 | 1.476 | 2.105 | 1.381 |
| 贵州省 | 1.201 | 1.167 | 1.870 | 1.809 | 2.131 | 2.739 | 1.724 | 2.431 | 1.764 |
| 云南省 | 0.486 | 0.617 | 0.983 | 1.001 | 1.511 | 1.908 | 0.919 | 1.508 | 0.946 |
| 四川省 | 0.465 | 0.544 | 0.990 | 0.941 | 1.436 | 1.920 | 0.908 | 1.565 | 0.998 |
| 陕西省 | 0.714 | 0.515 | 1.200 | 1.095 | 1.382 | 2.003 | 1.013 | 1.794 | 1.292 |

表 5-1（续表）

| | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|----------|-------|-------|-------|-------|--------|-------|-------|-------|-------|
| 青海省 | 0.615 | 0.220 | 0.869 | 0.884 | 1.032 | 1.624 | 0.583 | 1.419 | 1.073 |
| 宁夏回族自治区 | 0.647 | 0.337 | 1.056 | 0.985 | 1.208 | 1.866 | 0.830 | 1.650 | 1.196 |
| 新疆维吾尔自治区 | 0.445 | 0.578 | 0.736 | 0.710 | 1.289 | 1.622 | 0.708 | 1.279 | 0.827 |
| 广西壮族自治区 | 0.508 | 0.348 | 0.798 | 0.853 | 1.140 | 1.657 | 0.604 | 1.347 | 0.935 |

| | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 北京市 | 0.327 | 0.216 | 0.526 | 0.458 | 0.463 | 0.338 | 0.227 | 0.626 | 0.531 |
| 天津市 | 0.511 | 0.717 | 0.573 | 0.676 | 0.446 | 0.556 | 0.569 | 0.565 | 0.368 |
| 河北省 | 0.932 | 0.566 | 1.066 | 1.079 | 1.000 | 0.514 | 0.669 | 1.192 | 1.006 |
| 山西省 | 0.893 | 0.584 | 1.025 | 1.050 | 0.964 | 0.593 | 0.660 | 1.154 | 0.996 |
| 内蒙古自治区 | 1.388 | 1.311 | 1.490 | 1.582 | 1.366 | 1.115 | 1.267 | 1.515 | 1.318 |
| 辽宁省 | 1.908 | 1.657 | 1.985 | 2.047 | 1.916 | 1.524 | 1.696 | 2.076 | 1.875 |
| 吉林省 | 0.876 | 0.732 | 0.970 | 1.044 | 0.876 | 0.533 | 0.705 | 1.045 | 0.824 |
| 黑龙江省 | 1.536 | 1.168 | 1.632 | 1.648 | 1.587 | 1.122 | 1.278 | 1.762 | 1.575 |
| 上海市 | 0.907 | 0.501 | 1.061 | 0.991 | 1.029 | 0.618 | 0.671 | 1.190 | 1.070 |
| 江苏省 | 0.000 | 0.446 | 0.267 | 0.253 | 0.178 | 0.467 | 0.283 | 0.335 | 0.264 |
| 浙江省 | 0.446 | 0.000 | 0.628 | 0.586 | 0.566 | 0.256 | 0.218 | 0.750 | 0.622 |
| 安徽省 | 0.267 | 0.628 | 0.000 | 0.190 | 0.168 | 0.655 | 0.475 | 0.173 | 0.223 |
| 福建省 | 0.253 | 0.586 | 0.190 | 0.000 | 0.265 | 0.657 | 0.460 | 0.271 | 0.339 |
| 江西省 | 0.178 | 0.566 | 0.168 | 0.265 | 0.000 | 0.556 | 0.397 | 0.210 | 0.132 |
| 山东省 | 0.467 | 0.256 | 0.655 | 0.657 | 0.556 | 0.000 | 0.231 | 0.750 | 0.568 |
| 河南省 | 0.283 | 0.218 | 0.475 | 0.460 | 0.397 | 0.231 | 0.000 | 0.583 | 0.443 |
| 湖北省 | 0.335 | 0.750 | 0.173 | 0.271 | 0.210 | 0.750 | 0.583 | 0.000 | 0.228 |
| 湖南省 | 0.264 | 0.622 | 0.223 | 0.339 | 0.132 | 0.568 | 0.443 | 0.228 | 0.000 |
| 广东省 | 0.383 | 0.157 | 0.579 | 0.501 | 0.528 | 0.335 | 0.217 | 0.691 | 0.596 |
| 甘肃省 | 0.680 | 0.493 | 0.747 | 0.803 | 0.687 | 0.377 | 0.485 | 0.861 | 0.672 |
| 海南省 | 0.305 | 0.562 | 0.210 | 0.264 | 0.255 | 0.576 | 0.428 | 0.318 | 0.260 |
| 重庆市 | 0.470 | 0.884 | 0.355 | 0.392 | 0.381 | 0.881 | 0.729 | 0.232 | 0.374 |
| 西藏自治区 | 0.653 | 1.014 | 0.570 | 0.504 | 0.634 | 1.084 | 0.900 | 0.502 | 0.695 |
| 贵州省 | 0.958 | 1.372 | 0.839 | 0.827 | 0.896 | 1.400 | 1.227 | 0.715 | 0.917 |
| 云南省 | 0.338 | 0.576 | 0.242 | 0.259 | 0.304 | 0.593 | 0.445 | 0.350 | 0.319 |
| 四川省 | 0.207 | 0.556 | 0.104 | 0.207 | 0.146 | 0.577 | 0.393 | 0.237 | 0.213 |
| 陕西省 | 0.430 | 0.813 | 0.356 | 0.476 | 0.318 | 0.779 | 0.638 | 0.306 | 0.335 |
| 青海省 | 0.469 | 0.643 | 0.498 | 0.613 | 0.383 | 0.499 | 0.493 | 0.519 | 0.319 |
| 宁夏回族自治区 | 0.385 | 0.717 | 0.383 | 0.508 | 0.282 | 0.646 | 0.543 | 0.360 | 0.250 |
| 新疆维吾尔自治区 | 0.413 | 0.432 | 0.435 | 0.483 | 0.412 | 0.432 | 0.347 | 0.564 | 0.438 |
| 广西壮族自治区 | 0.382 | 0.522 | 0.424 | 0.506 | 0.327 | 0.411 | 0.377 | 0.478 | 0.277 |

表 5-1（续表）

| | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|----------|-------|---------|----------|---------|-------|-------|-------|-------|-------|
| 北京市 | 0.143 | 0.599 | 0.488 | 0.740 | 0.834 | 1.201 | 0.486 | 0.465 | 0.714 |
| 天津市 | 0.744 | 0.635 | 0.562 | 0.641 | 0.989 | 1.167 | 0.617 | 0.544 | 0.515 |
| 河北省 | 0.701 | 0.445 | 0.964 | 1.336 | 1.539 | 1.870 | 0.983 | 0.990 | 1.200 |
| 山西省 | 0.706 | 0.582 | 0.974 | 1.320 | 1.485 | 1.809 | 1.001 | 0.941 | 1.095 |
| 内蒙古自治区 | 1.416 | 1.044 | 1.459 | 1.639 | 1.949 | 2.131 | 1.511 | 1.436 | 1.382 |
| 辽宁省 | 1.784 | 1.307 | 1.884 | 2.227 | 2.481 | 2.739 | 1.908 | 1.920 | 2.003 |
| 吉林省 | 0.845 | 0.397 | 0.876 | 1.171 | 1.476 | 1.724 | 0.919 | 0.908 | 1.013 |
| 黑龙江省 | 1.289 | 0.951 | 1.497 | 1.896 | 2.105 | 2.431 | 1.508 | 1.565 | 1.794 |
| 上海市 | 0.544 | 0.734 | 0.957 | 1.304 | 1.381 | 1.764 | 0.946 | 0.998 | 1.292 |
| 江苏省 | 0.383 | 0.680 | 0.305 | 0.470 | 0.653 | 0.958 | 0.338 | 0.207 | 0.430 |
| 浙江省 | 0.157 | 0.493 | 0.562 | 0.884 | 1.014 | 1.372 | 0.576 | 0.556 | 0.813 |
| 安徽省 | 0.579 | 0.747 | 0.210 | 0.355 | 0.570 | 0.839 | 0.242 | 0.104 | 0.356 |
| 福建省 | 0.501 | 0.803 | 0.264 | 0.392 | 0.504 | 0.827 | 0.259 | 0.207 | 0.476 |
| 江西省 | 0.528 | 0.687 | 0.255 | 0.381 | 0.634 | 0.896 | 0.304 | 0.146 | 0.318 |
| 山东省 | 0.335 | 0.377 | 0.576 | 0.881 | 1.084 | 1.400 | 0.593 | 0.577 | 0.779 |
| 河南省 | 0.217 | 0.485 | 0.428 | 0.729 | 0.900 | 1.227 | 0.445 | 0.393 | 0.638 |
| 湖北省 | 0.691 | 0.861 | 0.318 | 0.232 | 0.502 | 0.715 | 0.350 | 0.237 | 0.306 |
| 湖南省 | 0.596 | 0.672 | 0.260 | 0.374 | 0.695 | 0.917 | 0.319 | 0.213 | 0.335 |
| 广东省 | 0.000 | 0.601 | 0.525 | 0.813 | 0.902 | 1.269 | 0.525 | 0.511 | 0.780 |
| 甘肃省 | 0.601 | 0.000 | 0.644 | 1.017 | 1.245 | 1.545 | 0.658 | 0.684 | 0.870 |
| 海南省 | 0.525 | 0.644 | 0.000 | 0.449 | 0.684 | 0.957 | 0.159 | 0.197 | 0.497 |
| 重庆市 | 0.813 | 1.017 | 0.449 | 0.000 | 0.446 | 0.582 | 0.468 | 0.419 | 0.447 |
| 西藏自治区 | 0.902 | 1.245 | 0.684 | 0.446 | 0.000 | 0.431 | 0.653 | 0.626 | 0.665 |
| 贵州省 | 1.269 | 1.545 | 0.957 | 0.582 | 0.431 | 0.000 | 0.940 | 0.905 | 0.826 |
| 云南省 | 0.525 | 0.658 | 0.159 | 0.468 | 0.653 | 0.940 | 0.000 | 0.232 | 0.534 |
| 四川省 | 0.511 | 0.684 | 0.197 | 0.419 | 0.626 | 0.905 | 0.232 | 0.000 | 0.359 |
| 陕西省 | 0.780 | 0.870 | 0.497 | 0.447 | 0.665 | 0.826 | 0.534 | 0.359 | 0.000 |
| 青海省 | 0.678 | 0.511 | 0.472 | 0.657 | 0.969 | 1.179 | 0.527 | 0.461 | 0.483 |
| 宁夏回族自治区 | 0.712 | 0.752 | 0.469 | 0.492 | 0.803 | 0.979 | 0.538 | 0.365 | 0.251 |
| 新疆维吾尔自治区 | 1.462 | 0.401 | 0.359 | 0.734 | 0.907 | 1.213 | 0.367 | 0.370 | 0.582 |
| 广西壮族自治区 | 0.550 | 0.461 | 0.352 | 0.622 | 0.923 | 1.167 | 0.410 | 0.382 | 0.545 |
| | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 | | | | | |
| 北京市 | 0.615 | 0.647 | 0.445 | 0.508 | | | | | |
| 天津市 | 0.220 | 0.337 | 0.578 | 0.348 | | | | | |
| 河北省 | 0.869 | 1.056 | 0.736 | 0.798 | | | | | |
| 山西省 | 0.884 | 0.985 | 0.710 | 0.853 | | | | | |
| 内蒙古自治区 | 1.032 | 1.208 | 1.289 | 1.140 | | | | | |
| 辽宁省 | 1.624 | 1.866 | 1.622 | 1.657 | | | | | |

表 5-1（续表）

| | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|----------|-------|---------|----------|---------|
| 吉林省 | 0.583 | 0.830 | 0.708 | 0.604 |
| 黑龙江省 | 1.419 | 1.650 | 1.279 | 1.347 |
| 上海市 | 1.073 | 1.196 | 0.827 | 0.935 |
| 江苏省 | 0.469 | 0.385 | 0.413 | 0.382 |
| 浙江省 | 0.643 | 0.717 | 0.432 | 0.522 |
| 安徽省 | 0.498 | 0.383 | 0.435 | 0.424 |
| 福建省 | 0.613 | 0.508 | 0.483 | 0.506 |
| 江西省 | 0.383 | 0.282 | 0.412 | 0.327 |
| 山东省 | 0.499 | 0.646 | 0.432 | 0.411 |
| 河南省 | 0.493 | 0.543 | 0.347 | 0.377 |
| 湖北省 | 0.519 | 0.360 | 0.564 | 0.478 |
| 湖南省 | 0.319 | 0.250 | 0.438 | 0.277 |
| 广东省 | 0.678 | 0.712 | 1.462 | 0.550 |
| 甘肃省 | 0.511 | 0.752 | 0.401 | 0.461 |
| 海南省 | 0.472 | 0.469 | 0.359 | 0.352 |
| 重庆市 | 0.657 | 0.492 | 0.734 | 0.622 |
| 西藏自治区 | 0.969 | 0.803 | 0.907 | 0.923 |
| 贵州省 | 1.179 | 0.979 | 1.213 | 1.167 |
| 云南省 | 0.527 | 0.538 | 0.367 | 0.410 |
| 四川省 | 0.461 | 0.365 | 0.370 | 0.382 |
| 陕西省 | 0.483 | 0.251 | 0.582 | 0.545 |
| 青海省 | 0.000 | 0.325 | 0.443 | 0.218 |
| 宁夏回族自治区 | 0.325 | 0.000 | 0.552 | 0.398 |
| 新疆维吾尔自治区 | 0.443 | 0.552 | 0.000 | 0.381 |
| 广西壮族自治区 | 0.418 | 0.398 | 0.381 | 0.000 |

表 5-1 是测量 31 个省市 GDP 数据之间欧式距离的矩阵表，从表可以看出某些省市的 GDP 数据之间的距离很小，而某些省市的 GDP 数据之间的距离很大。比如表中的青海省和宁夏回族自治区之间的欧式距离为 0.325，说明这两个省份的 GDP 数据具有较强的相似性。而广东省和新疆维吾尔自治区之间的欧式距离为 1.462，说明这两个省份的 GDP 数据之间的相似性较弱。从表中得到的这些信息为进一步的聚类分析提供了某些依据。

表 5-2 聚类过程表

| 阶段 | 组合的集群 | | 系数 | 首次出现阶段集群 | | |
|----|-------|------|-------|----------|------|-------|
| | 集群 1 | 集群 2 | | 集群 1 | 集群 2 | 下一个阶段 |
| 1 | 12 | 23 | 0.104 | 0 | 0 | 4 |
| 2 | 14 | 18 | 0.132 | 0 | 0 | 4 |

表 5-2（续表）

| 阶段 | 组合的集群 | | 系数 | 首次出现阶段集群 | | 下一个阶段 |
|----|-------|------|-------|----------|------|-------|
| | 集群 1 | 集群 2 | | 集群 1 | 集群 2 | |
| 3 | 1 | 19 | 0.143 | 0 | 0 | 5 |
| 4 | 12 | 14 | 0.146 | 1 | 2 | 7 |
| 5 | 1 | 11 | 0.157 | 3 | 0 | 11 |
| 6 | 21 | 25 | 0.159 | 0 | 0 | 10 |
| 7 | 12 | 17 | 0.173 | 4 | 0 | 8 |
| 8 | 10 | 12 | 0.178 | 0 | 7 | 9 |
| 9 | 10 | 13 | 0.190 | 8 | 0 | 10 |
| 10 | 10 | 21 | 0.197 | 9 | 6 | 15 |
| 11 | 1 | 16 | 0.217 | 5 | 0 | 14 |
| 12 | 20 | 29 | 0.218 | 0 | 0 | 13 |
| 13 | 2 | 20 | 0.220 | 0 | 12 | 18 |
| 14 | 1 | 15 | 0.231 | 11 | 0 | 19 |
| 15 | 10 | 22 | 0.232 | 10 | 0 | 16 |
| 16 | 10 | 30 | 0.250 | 15 | 0 | 17 |
| 17 | 10 | 27 | 0.251 | 16 | 0 | 18 |
| 18 | 2 | 10 | 0.277 | 13 | 17 | 19 |
| 19 | 1 | 2 | 0.283 | 14 | 18 | 20 |
| 20 | 1 | 31 | 0.347 | 19 | 0 | 21 |
| 21 | 1 | 28 | 0.377 | 20 | 0 | 22 |
| 22 | 1 | 7 | 0.397 | 21 | 0 | 25 |
| 23 | 3 | 4 | 0.399 | 0 | 0 | 25 |
| 24 | 24 | 26 | 0.431 | 0 | 0 | 26 |
| 25 | 1 | 3 | 0.445 | 22 | 23 | 26 |
| 26 | 1 | 24 | 0.446 | 25 | 24 | 27 |
| 27 | 1 | 9 | 0.501 | 26 | 0 | 28 |
| 28 | 1 | 8 | 0.653 | 27 | 0 | 29 |
| 29 | 1 | 5 | 0.735 | 28 | 0 | 30 |
| 30 | 1 | 6 | 0.945 | 29 | 0 | 0 |

表 5-2 是 GDP 的聚类过程表，观察表中的系数栏的数据，可以发现这些数据是由小到大递增的。这些数据就是每一步聚类时计算出的两类之间最小的距离，按照距离最小的先聚为一类的原则所得到的聚类结果。

因为本节对我国 31 个省份按照 GDP 发展水平进行聚类，变量个数相对较多，从 5-1 和表 5-2 很难直观看出应该如何对变量进行分类，不太容易确定分成几类更合适。下面展示对变量聚类得到的谱系聚类图和冰柱图，从图上可以很直观清晰地看出聚类的全过程。

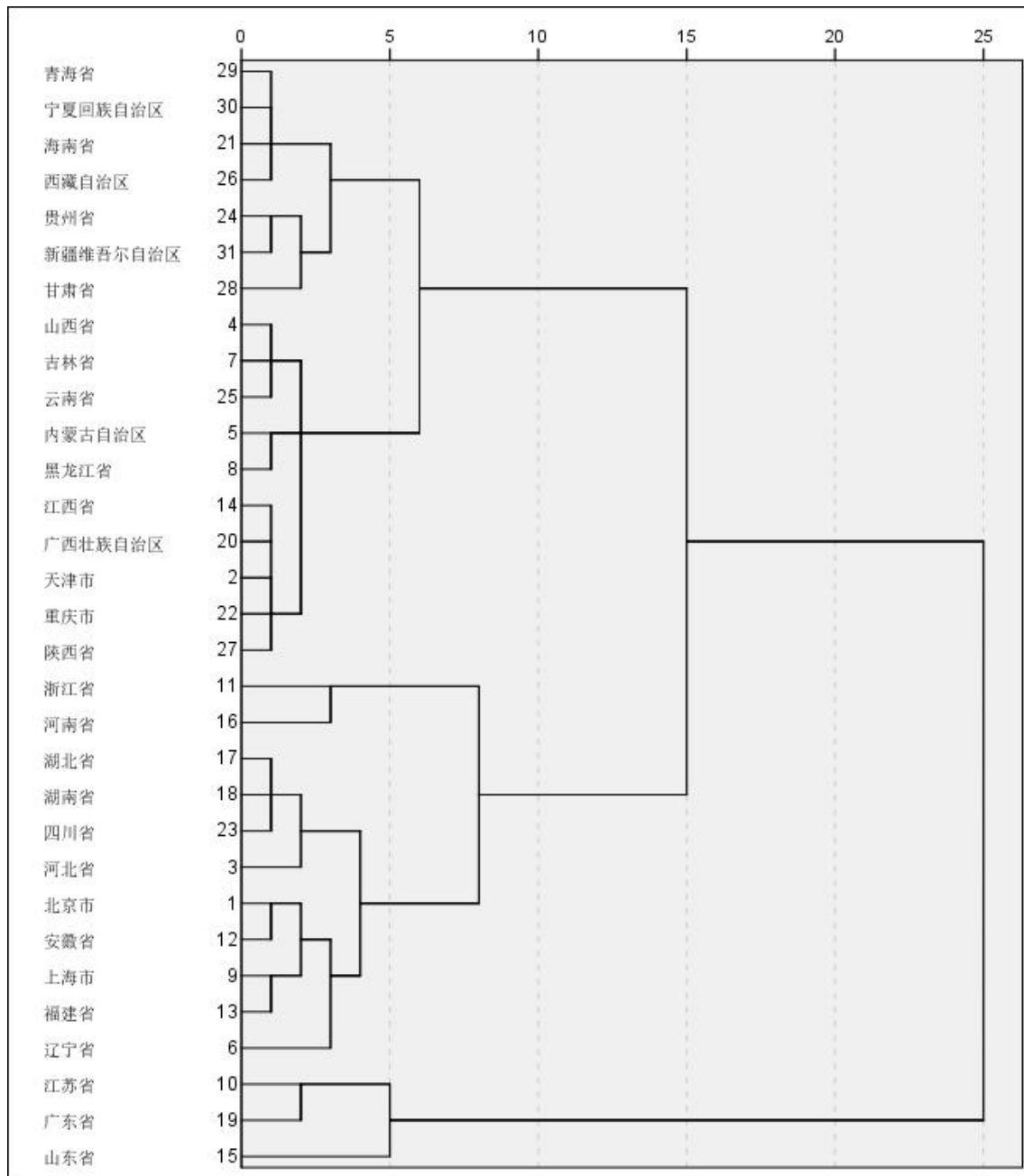


图 5-1 谱系聚类图

如果以横坐标值为 20 划分，可以分为两类；如果按横坐标值为 10 划分，就分为三类；当以横坐标值为 5 划分，可以分为五类。结合我国各省的经济发展水平，发现当分为五类时，分类结果比较合理。

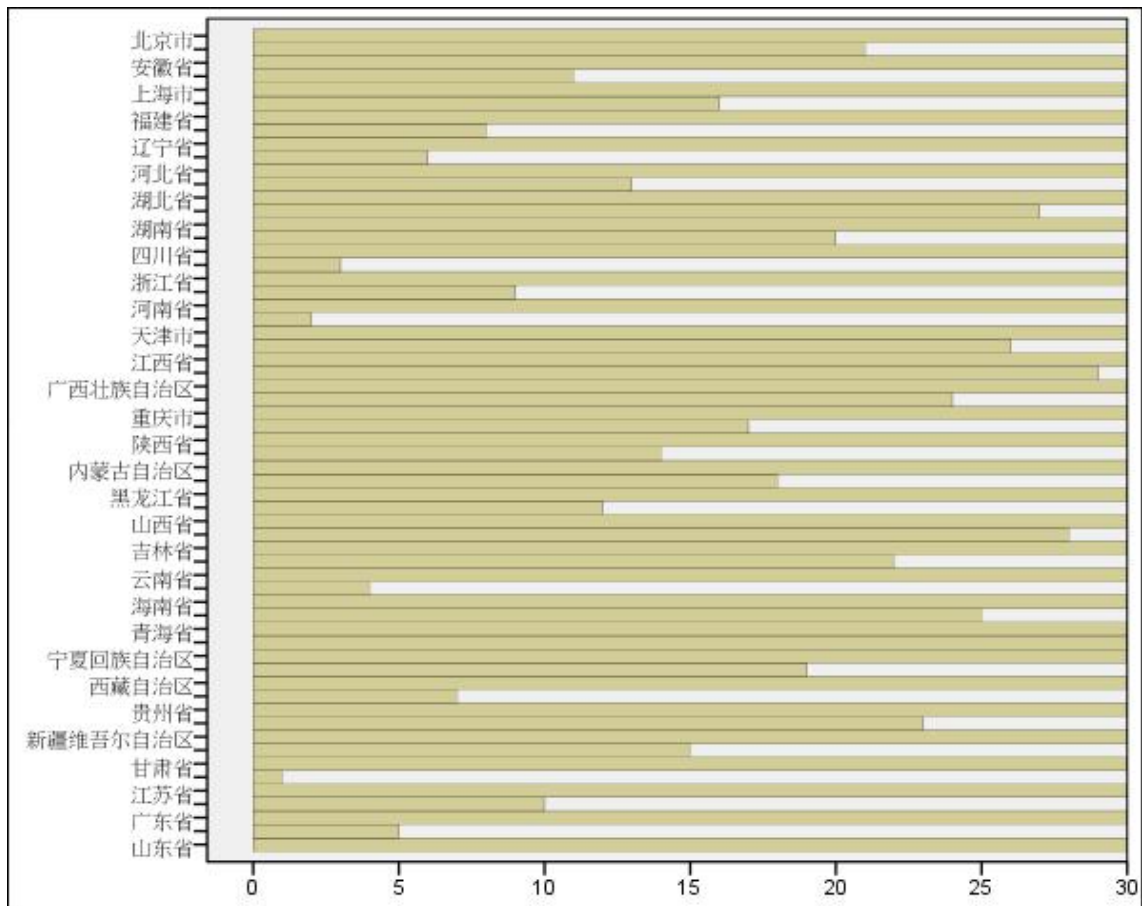


图 5-2 冰柱图

图 5-2 所示是聚类全过程的冰柱图，可以看出当分成五类时所得到的分类结果与图 5-1 的结果是一致的。下面利用 SPSS 软件自动输出当变量分成五类时所得到的分类结果。

表 5-3 分类结果表

| 个案 | 5 个集群 | 个案 | 5 个集群 | 个案 | 5 个集群 |
|-----|-------|---------|-------|----------|-------|
| 江苏省 | 1 | 湖北省 | 3 | 云南省 | 4 |
| 山东省 | 1 | 湖南省 | 3 | 陕西省 | 4 |
| 广东省 | 1 | 四川省 | 3 | 海南省 | 5 |
| 浙江省 | 2 | 天津市 | 4 | 贵州省 | 5 |
| 河南省 | 2 | 山西省 | 4 | 西藏自治区 | 5 |
| 北京市 | 3 | 内蒙古自治区 | 4 | 甘肃省 | 5 |
| 河北省 | 3 | 吉林省 | 4 | 青海省 | 5 |
| 辽宁省 | 3 | 黑龙江省 | 4 | 宁夏回族自治区 | 5 |
| 上海市 | 3 | 江西省 | 4 | 新疆维吾尔自治区 | 5 |
| 安徽省 | 3 | 广西壮族自治区 | 4 | | |
| 福建省 | 3 | 重庆市 | 4 | | |

从表 5-3 可知,对我国 31 个省市按 GDP 相似度进行分类,大致可以分成五类。其中,江苏省、广东省和山东省为第一类;浙江省和河南省为第二类;北京市、河北省、辽宁省、上海市、安徽省、福建省、湖北省、湖南省和四川省为第三类;天津市、山西省、内蒙古自治区、吉林省、黑龙江省、江西省、广西壮族自治区、重庆市、云南省和陕西省为第四类;海南省、贵州省、西藏自治区、甘肃省、青海省、宁夏回族自治区和新疆维吾尔自治区为第五类。

(2) GDP 数据的函数型聚类分析

函数型聚类分析方法利用数学工具研究函数及其导数的特征,可以挖掘出传统统计方法难以挖掘出的信息。相对于直接对离散数据进行建模,函数型聚类分析方法具有很多优势。由于函数型聚类分析方法直接对曲线进行处理,因此可以从曲线的数值特征和曲线的变化特征对函数进行聚类。下面先从数值特征对曲线进行聚类。

调用附录 2 中的 `fdobj = cluster(GDP_1978_2018_year)` 和 `plot(fdobj)` 函数得到的分类结果如表 5-4 及图 5-3 所示。

表 5-4 GDP 聚类结果表

| 类别 | 省份 |
|-----|---|
| 第一类 | 江苏省、山东省、广东省、浙江省 |
| 第二类 | 河北省、北京市、辽宁省、上海市、安徽省、福建省、湖北省、湖南省、四川省、河南省 |
| 第三类 | 天津市、山西省、内蒙古自治区、吉林省、黑龙江省、江西省、广西壮族自治区、重庆市、云南省、陕西省 |
| 第四类 | 海南省、贵州省、西藏自治区、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区 |

从表 5-4 可知,对我国 31 个省市按 GDP 相似度进行分类,可以分成四类。其中,江苏省、山东省、广东省和浙江省为第一类;北京市、河北省、辽宁省、上海市、安徽省、福建省、湖北省、湖南省、四川省和河南省为第二类;天津市、山西省、内蒙古自治区、吉林省、黑龙江省、江西省、广西壮族自治区、重庆市、云南省和陕西省为第三类;海南省、贵州省、西藏自治区、甘肃省、青海省、宁夏回族自治区和新疆维吾尔自治区为第四类。与表 5-3 对比可得,除了河南省和浙江省的分类不一致外,对 GDP 数据进行函数型聚类得到的结果与利用传统的系统聚类法聚类所得到的结果基本一致。观察河南省和浙江省的 GDP 数据发现,浙江省的 GDP 与第一类中的江苏省等省份的 GDP 相似程度较高,同理河南省的 GDP 曲线更接近第二类中的河北省等省份 GDP 均值函数曲线,因此把它划分到第二类也是合理的。

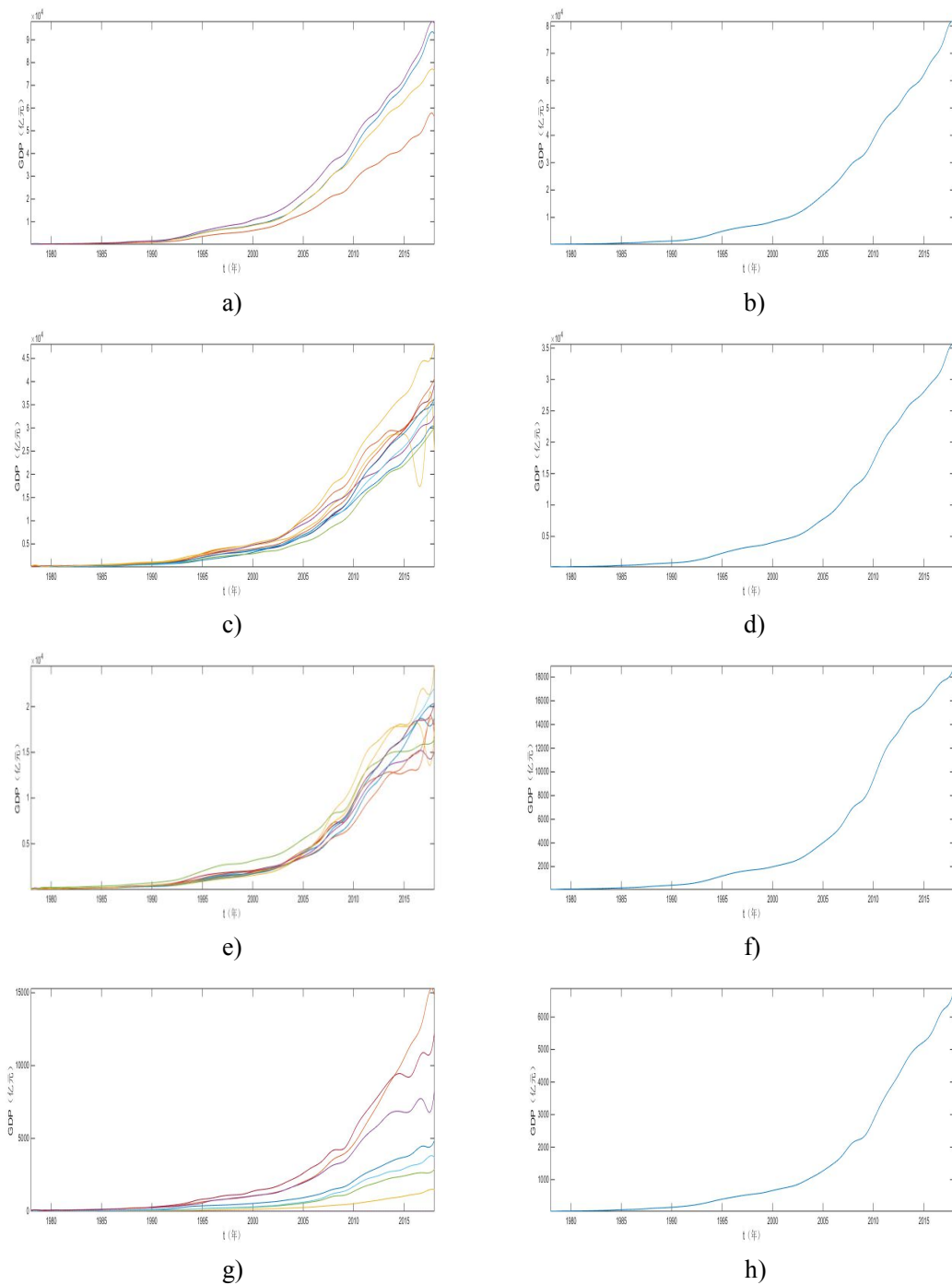


图 5-3 各省分类 GDP 及其均值曲线

图 5-3 展示了聚类后按类别划分的各省份 GDP 平滑曲线以及每类的 GDP 均值曲线。其中子图 a)和 b)代表第一类省份，c)和 d)代表第二类省份，e)和 f)代表第三类省份，g)和 h)代表第四类省份。从图像可以看出同一类中各省份 GDP 总量大小相近。

前面根据我国不同省份 GDP 的数值特征对曲线进行了聚类。由于地理位置、

政策制度等因素，我国各省份的经济发展存在很大差异。经济发展的不平衡造成了资源利用效率低下、教育质量不平衡和就业难等制约我国发展的问题。地区协调发展成为各类政府工作报告中的热词。因此，研究不同地区 GDP 的发展潜力，根据 GDP 发展潜力对地区进行分类，可以为国家制定相关政策提供依据。下面从曲线变化特征方面对曲线进行聚类。

总体上，我国 1978-2018 年的各省份 GDP 数据呈递增趋势，但是不同省份的 GDP 增长快慢差异很大。为了更清晰展示 31 个省各年 GDP 的差异水平，下面先作出数据的描述性统计表和箱线图来观察这种差异。

表 5-5 GDP 年度数据描述统计表 (亿元)

| 年份 | 最小值 | 25%分位数 | 中位数 | 平均值 | 75%分位数 | 最大值 |
|------|--------|---------|---------|---------|---------|----------|
| 1978 | 8.90 | 64.73 | 87.00 | 112.21 | 174.80 | 272.81 |
| 1979 | 11.11 | 67.51 | 104.15 | 127.27 | 190.09 | 298.55 |
| 1980 | 10.90 | 73.90 | 108.76 | 141.94 | 221.00 | 319.80 |
| 1981 | 12.11 | 77.91 | 121.26 | 155.07 | 228.30 | 350.02 |
| 1982 | 14.51 | 93.22 | 133.96 | 172.53 | 251.45 | 395.38 |
| 1983 | 13.02 | 105.88 | 150.14 | 193.75 | 283.21 | 459.83 |
| 1984 | 13.68 | 128.20 | 174.39 | 230.51 | 332.22 | 581.56 |
| 1985 | 17.76 | 163.83 | 207.26 | 278.23 | 421.15 | 680.46 |
| 1986 | 16.95 | 181.58 | 227.23 | 312.11 | 458.23 | 744.94 |
| 1987 | 17.71 | 206.73 | 279.24 | 370.59 | 530.86 | 922.33 |
| 1988 | 20.25 | 259.71 | 368.67 | 468.41 | 659.69 | 1208.85 |
| 1989 | 21.86 | 283.49 | 391.65 | 532.48 | 744.98 | 1381.39 |
| 1990 | 24.45 | 310.95 | 451.67 | 596.73 | 890.95 | 1559.03 |
| 1991 | 30.53 | 342.65 | 518.59 | 685.58 | 1016.31 | 1893.30 |
| 1992 | 33.29 | 411.04 | 646.60 | 839.88 | 1177.27 | 2447.54 |
| 1993 | 37.42 | 537.81 | 871.70 | 1105.64 | 1519.23 | 3469.28 |
| 1994 | 45.99 | 695.06 | 1145.31 | 1465.25 | 2001.41 | 4619.02 |
| 1995 | 56.11 | 857.06 | 1497.56 | 1859.42 | 2499.43 | 5933.05 |
| 1996 | 64.98 | 1023.09 | 1697.90 | 2190.06 | 2957.55 | 6834.97 |
| 1997 | 77.24 | 1153.51 | 1817.25 | 2467.43 | 3438.79 | 7774.53 |
| 1998 | 91.50 | 1262.54 | 1911.30 | 2668.44 | 3801.09 | 8530.88 |
| 1999 | 105.98 | 1379.31 | 1971.41 | 2851.23 | 4171.69 | 9250.68 |
| 2000 | 117.80 | 1539.12 | 2080.04 | 3183.63 | 4669.06 | 10741.25 |
| 2001 | 139.16 | 1713.81 | 2279.34 | 3508.89 | 5033.08 | 12039.25 |
| 2002 | 162.04 | 1940.94 | 2523.73 | 3897.40 | 5458.22 | 13502.42 |
| 2003 | 185.09 | 2388.38 | 2855.23 | 4501.20 | 6002.54 | 15844.64 |
| 2004 | 220.34 | 3034.58 | 3571.37 | 5416.86 | 6672.00 | 18864.62 |
| 2005 | 248.80 | 3462.73 | 4230.53 | 6426.71 | 8047.26 | 22557.37 |
| 2006 | 290.76 | 3907.23 | 4944.25 | 7510.86 | 9304.52 | 26587.76 |

表 5-5 (续表)

| 年份 | 最小值 | 25%分位数 | 中位数 | 平均值 | 75%分位数 | 最大值 |
|------|---------|----------|----------|----------|----------|----------|
| 2007 | 341.43 | 4676.13 | 6423.18 | 9023.80 | 11164.30 | 31777.01 |
| 2008 | 394.85 | 5692.12 | 8314.37 | 10752.06 | 13668.58 | 36796.71 |
| 2009 | 441.36 | 6169.75 | 8587.00 | 11783.99 | 15046.45 | 39482.56 |
| 2010 | 507.46 | 7224.18 | 10368.60 | 14098.13 | 17185.48 | 46013.06 |
| 2011 | 605.83 | 8893.12 | 12582.00 | 16820.68 | 21026.68 | 53210.28 |
| 2012 | 701.03 | 10309.47 | 14453.68 | 18598.45 | 23872.80 | 57067.92 |
| 2013 | 815.67 | 11832.31 | 16205.45 | 20462.75 | 26392.07 | 62474.79 |
| 2014 | 920.83 | 12761.49 | 17689.94 | 22075.79 | 28536.66 | 67809.85 |
| 2015 | 1026.39 | 12766.49 | 17831.51 | 23315.09 | 29550.19 | 72812.55 |
| 2016 | 1151.41 | 13050.41 | 18499.00 | 25163.55 | 32070.45 | 80854.91 |
| 2017 | 1310.92 | 14944.53 | 20006.31 | 27327.10 | 34016.32 | 89705.23 |
| 2018 | 1477.63 | 15074.62 | 21984.78 | 29506.69 | 36425.78 | 97277.77 |

表 5-5 是我国 31 个省市 1978-2018 年 GDP 数据描述统计表，主要展示了各年各省 GDP 数据的最值、均值和分位数。

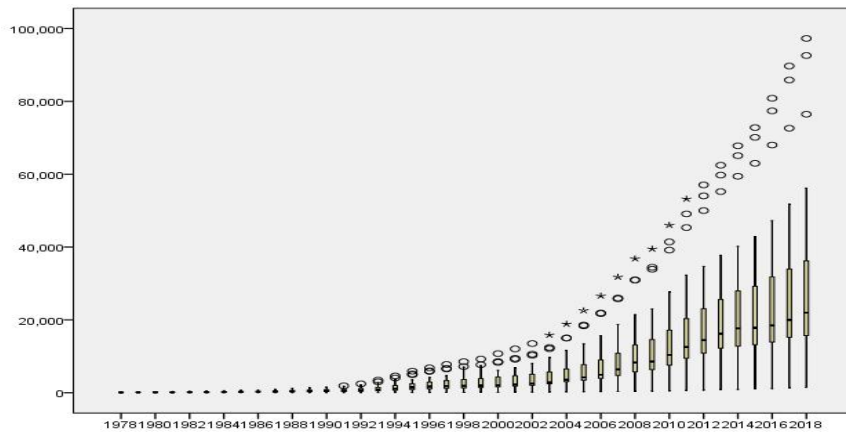


图 5-4 各省 GDP 箱线图

从图 5-4 可以看出，各省的 GDP 差距逐年扩大，图中的离群点和极值点是江苏省、广东省和山东省，表明这几个省份的 GDP 尤为突出，远远超过平均水平。表 5-5 和图 5-4 都显示出我国不同省份在 GDP 发展速度上存在显著的差异，并且这种差异存在鲜明的层次等级，接下来利用函数型聚类分析方法探究这种分层差异。

调用附录 1 中的 $Lfd = \text{deriv_fd}(fdoj)$ 和 $\text{plot}(Lfd)$ 函数得到的分类结果如表 5-6 及图 5-5、图 5-6 所示。

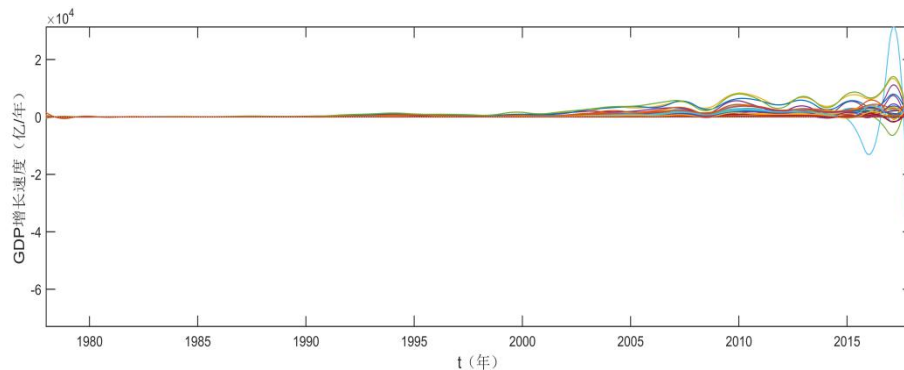


图 5-5 各省 GDP 增长速度

图 5-5 是各省 GDP 的增长速度曲线,从图像可以看出各省 GDP 速度曲线都呈现波动式增长。1995-2007 年各省 GDP 增速显著加快,不过期间也有一些小波动。2007-2008 年 GDP 增速明显放缓,不过在 2008 以后增速又出现了反弹,在 2010 年达到一个高点。从 2010-2018 年 GDP 增速总体呈现波动式变化,值得注意的是从 2015 年开始某些省份 GDP 增速出现负值。这可能是由于地方政策或者整个经济发展背景引起的经济增速减缓现象。

表 5-6 GDP 增速聚类结果表

| 发展潜力 | 省份 |
|------|---|
| 第一梯队 | 江苏省、山东省、广东省 |
| 第二梯队 | 河北省、辽宁省、福建省、湖北省、湖南省、四川省、浙江省、河南省 |
| 第三梯队 | 北京市、天津市、山西省、内蒙古自治区、吉林省、黑龙江省、上海市、安徽省、江西省、广西壮族自治区、重庆市、云南省、陕西省 |
| 第四梯队 | 海南省、贵州省、西藏自治区、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区 |

从表 5-6 可知,江苏省、山东省、广东省的 GDP 发展潜力最大,这几个省份的经济最为发达,发展速度也是令人瞩目;河北省、辽宁省、福建省、湖北省、湖南省、四川省、浙江省和河南省八个省份的发展潜力次之;位于发展潜力第三梯队的地区较多,有北京市、天津市、山西省、内蒙古自治区、吉林省、黑龙江省、上海市、安徽省、江西省、广西壮族自治区、重庆市、云南省和陕西省共 13 个省份,这些省份中大多是传统工业大省,经济基础相对雄厚,进一步上升空间仍然很大;海南省、贵州省、西藏自治区、甘肃省、青海省、宁夏回族自治区和新疆维吾尔自治区的发展潜力相对最低,这几个省份大多位于我国西北地区,地理位置造成发展缓慢,是我国重点扶持的省份。

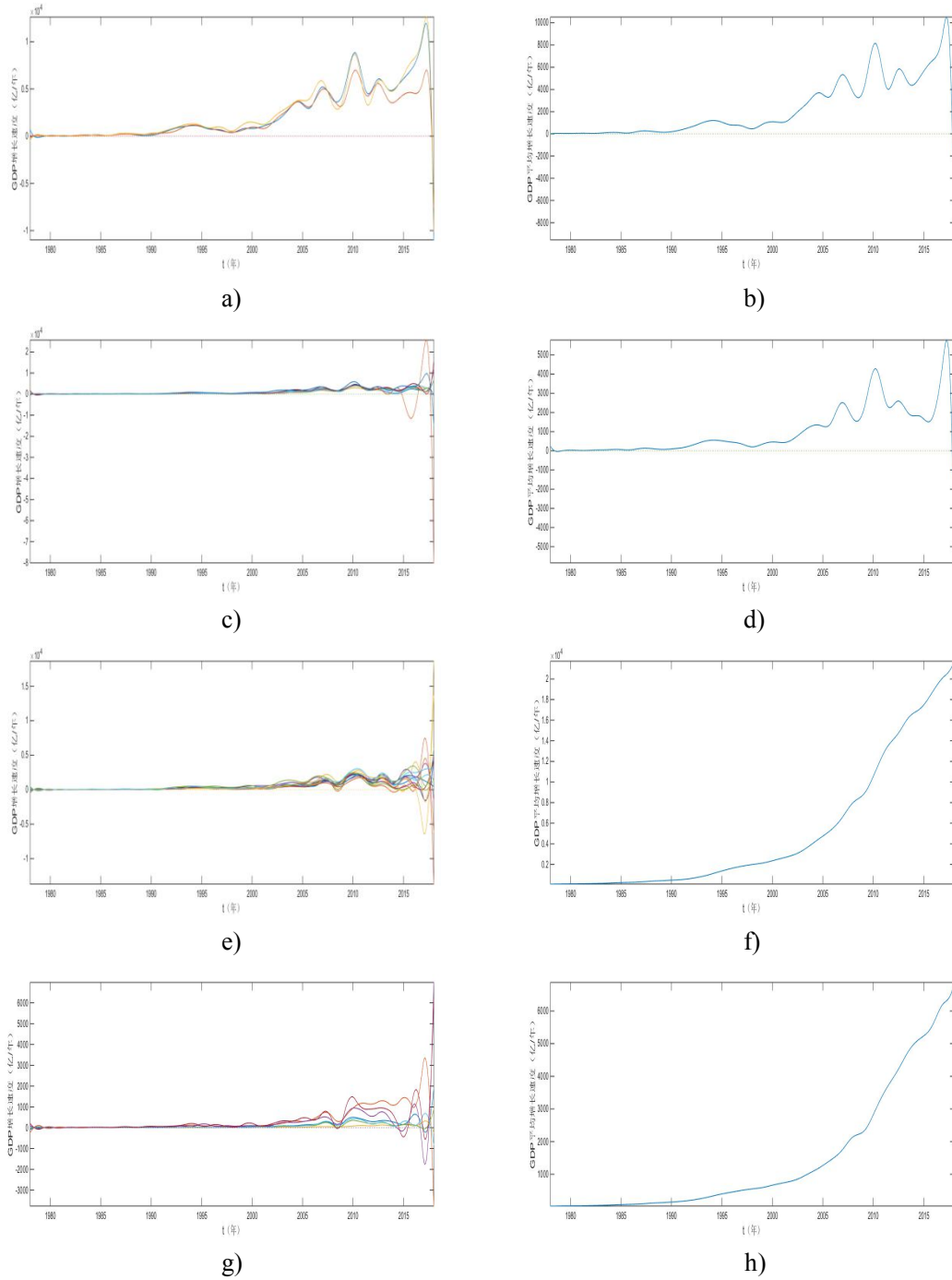


图 5-6 各省分类 GDP 增长速度及其均值曲线

图 5-6 展示了聚类后按类别分的各省份 GDP 增长速度曲线以及每类的 GDP 增长速度均值曲线。其中子图 a)和 b)代表第一梯队，c)和 d)代表第二梯队，e)和 f)代表第三梯队，g)和 h)代表第四梯队。从图像可以看出同一类中各省份 GDP 增长速度较接近。仔细观察图 b)和图 d)可以发现，第一梯队和第二梯队的 GDP 增速均值曲线随时间呈现波动式形状，在 2017-2018 年间曲线递减到负值，这表明前两梯

队的省份 GDP 已经逐渐趋于稳定水平,经济增长速度逐渐放缓。观察图 f)和图 h)可以发现,第三梯队和第四梯队的 GDP 速度均值曲线随时间一直增加,这表明虽然目前这两个梯队省份的经济相对落后,不过 GDP 的发展空间较大,在国家相关政策的大力支持下可以逐渐赶上经济发达省份。

5.4 本章小结

本章主要介绍了两种聚类分析方法,即经典的系统聚类分析方法和函数型聚类分析方法。重点对函数型聚类分析方法的理论进行了研究,并且提出了计算函数型变量之间距离的曲线夹角余弦方法。此外,本章分别把这两种方法应用到我国 1978-2018 年的 GDP 分省年度数据中,分析结果显示虽然系统聚类分析方法和函数型聚类分析方法从不同的角度处理数据,但是得到的结果有某些共性。在应用函数型聚类分析方法处理 GDP 数据时,本文创新性地对曲线的导函数进行了聚类分析,这样就可以从曲线变化特征的角度探究我国各省份 GDP 的发展潜力,这是传统的多元统计聚类分析方法所不能完成的,展示出了函数型数据分析方法一定程度上的优越性。

本章研究结果表明,我国各省份的 GDP 总量以及发展潜力可以按省份分成不同的类别,探究各省份 GDP 发展潜力有助于寻找不同省份发展存在差异的深层原因。国家应适当对经济发展潜力不足的省份给与某些政策倾斜,推进经济发展潜力不足的省份向前发展。同时,要保持经济发展速度高的省份继续发展,带动经济发展潜力不足的省份,发挥领头羊作用,最终实现不同地区协调发展。

结 论

本文主要利用函数型数据分析方法分析我国各省的 GDP 数据，并与经典的多元统计分析方法作对比。

本文的主要结论如下：

(1) B 样条基对我国各省 GDP 数据的拟合效果远远优于单项式基和傅里叶基的拟合效果，傅里叶基函数只适用于仅具有周期循环特征的数据，无法表现时间序列数据的整体递增或递减趋势。此外，B 样条基函数对于某些变化幅度大的数据会出现过拟合或欠拟合现象。

(2) 虽然多元统计主成分分析方法和函数型主成分分析方法从不同的角度处理 GDP 数据，但是得到的结果具有某些共性。对于多元统计主成分分析方法，前两个主成分的方差累计贡献率为 99.313%，而对于函数型主成分分析方法，前两个主成分的方差累计贡献率为 96.7%，且第一个主成分都代表经济发达省份的 GDP 与各省 GDP 平均水平的差异，第二个主成分都代表我国 GDP 的整体变化趋势。因此，经济发达省份的带动作用为 GDP 均值波动的主要影响因素。

(3) 从 GDP 的数值特征出发，利用函数型聚类分析方法可以把我国 31 个省划分成四类。除了河南省和浙江省，与经典的系统聚类法所得到的结果一致。从 GDP 平滑曲线的变化特征角度对曲线进行聚类，发现不同省份在 GDP 发展速度上存在显著的差异，并且这种差异存在鲜明的层次等级。第三和第四等级省份的 GDP 发展空间较大，在国家相关政策的支持下可以赶上第一和第二等级中发展速度较快的省份。

(4) 与经典的多元统计分析方法相比，函数型数据分析方法具有一定的优势，首先函数型数据分析方法在损失少量精确度的条件下，可以对小样本数据进行分析。其次，函数型数据分析方法通过对曲线求导可以探究曲线的变化速率。

本文的主要创新点如下：

(1) 对基于 B 样条基的最小二乘拟合方法提出了改进方法，扩展了函数型数据预处理方法的适用范围。

(2) 建立了度量曲线之间距离的曲线夹角余弦法。

(3) 通过对 GDP 平滑曲线求导数，从曲线的变化特征方面对曲线进行聚类，扩展了函数型聚类分析方法的研究视角。

本文的后续工作可以考虑如下几个方面：

(1) 研究 B 样条基拟合时存在的欠拟合和过拟合现象，提出解决办法。

(2) 进一步利用函数型回归分析、函数型典型相关分析等其它函数型数据分析方法分析我国 GDP 数据。

参考文献

- [1] Ramsay J O. When the Data are Functions[J]. *Psychometrika*, 1982, 47(4): 379-396.
- [2] Ramsay J O, Dalzell C J. Some Tools for Functional Data Analysis[J]. *Journal of the Royal Statistical Society*, 1991, 53(3): 539-572.
- [3] Ramsay J O, Silverman B W. *Functional Data Analysis*[M]. 2nd Edition. New York: Springer-Verlag, 2005: 110-200.
- [4] Ramsay J O, Silverman B W. *Applied Functional Data Analysis: Methods and Case Studies*[M]. New York: Springer-Verlag, 2002: 30-218.
- [5] Ramsay J O, Silverman B W. *Functional Data Analysis*[J]. *International Encyclopedia of the Social and Behavioral Sciences*, 2001, 50(2): 5822-5828.
- [6] Ferraty F, Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice*[M]. New York: Springer, 2006: 112-118.
- [7] Ramsay J O. *The Data Analysis of Vector-valued Functions*[M]. New York: Nova Science Publishers, 1989: 233-245.
- [8] Rao C R. Some Statistical Methods for Comparison of Growth Curves[J]. *Biometrics*, 1958, 14(1): 1-17.
- [9] Ramsay J O. Principal Differential Analysis: Data Reduction by Differential Operators[J]. *Journal of the Royal Statistical Society*, 1996, 58(3): 495-508.
- [10] Müller H G, Yao F. Functional Additive Models[J]. *Journal of American Statistical Association*, 2008, 103 (484): 1534-1544.
- [11] Ferraty F, Vieu P. The Functional Nonparametric Model and Application to Spectrometric Data[J]. *Computational Statistics*, 2002, 17(4): 545-564.
- [12] Yu P, Du J, Zhang Z. Single-index Partially Functional Linear Regression Model[J]. *Statistical Papers*, 2018, 10(11): 1-17.
- [13] Huang J, Chen S. Function-on-Partially Linear Functional Additive Models[J]. *Journal of Applied Mathematics and Physics*, 2020, 08(1): 1-9.
- [14] Yang G, Wang Q, Ma Y, et al. Locally Efficient Estimation in Generalized Partially Linear Model with Measurement Error in Nonlinear Function[J]. *Test*, 2019, 11(2): 5-20.
- [15] Abrams S, Janssen P, Swanepoel J, et al. Nonparametric Estimation of the Cross Ratio Function[J]. *Annals of the Institute of Statistical Mathematics*, 2019, 02(50): 111-116.
- [16] Xie T, Cao R, Du J. Variable Selection for Spatial Autoregressive Models with a Diverging Number of Parameters[J]. *Statistical Papers*, 2018, 15(3): 1-21.

- [17] Martino A, Ghigletti A, Ieva F, et al. A K-means Procedure Based on a Mahalanobis Type Distance for Clustering Multivariate Functional Data[J]. *Statistical Methods and Applications*, 2019, 28(2): 301-322.
- [18] Picheny V, Servien R, Villa-Vialaneix N. Interpretable Sparse SIR for Functional Data[J]. *Stats and Computing*, 2019, 29(2): 255-267.
- [19] Smaga U, Krzyko M. Robust Estimation in Canonical Correlation Analysis for Multivariate Functional Data[J]. *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Stats*, 2019, 48(2): 521-535.
- [20] Bruckner A, Schmerbauch A, Ruess L, et al. Foliar Roundup Application has Minor Effects on the Compositional and Functional Diversity of Soil Microorganisms in a Short-term Greenhouse Experiment[J]. *Ecotoxicology and Environmental Safety*, 2019, 174(7): 506-513.
- [21] Peraro T. FiniteFlow: Multivariate Functional Reconstruction Using Finite Fields and Dataflow Graphs[J]. *Journal of High Energy Physics*, 2019, 2019(7): 2-12.
- [22] Khattree R, Bahuguna M. An Alternative Data Analytic Approach to Measure the Univariate and Multivariate Skewness[J]. *International Journal of Data Science & Analytics*, 2019, 7(1): 1-16.
- [23] Zhu H T, Fan J Q, Kong L L. Spatially Varying Coefficient Model for Neuroimaging Data with Jump Discontinuities[J]. *Journal of American Statistical Association*, 2014, 109(507): 1084-1098.
- [24] Chiou J M, Müller H G. A Pairwise Interaction Model for Multivariate Functional and Longitudinal Data[J]. *Biometrika*, 2016, 103(2): 377-396.
- [25] Ramsay J O, Ramsey J B. Functional Data Analysis of the Dynamics of the Monthly Index of Non-durable Goods Production[J]. *Journal of Econometrics*, 2001, 11(107): 327-344.
- [26] Wang S. Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis[J]. *Journal of Business and Economic Statistics*, 2008, 20(24): 144-161.
- [27] Andersson J, Newbold P. Modeling the Distribution of Financial Returns by Functional Data Analysis[J]. *Institutionen För Informationsvetenskap, Research report*, 2002, 12(4): 1403-7572.
- [28] Müller H G, Stadtmüller U. On Variance Function Estimation with Quadratic Form[J]. *Journal of Statistical Planning and Inference*, 1993, 35(2): 213-231.
- [29] Yao F, Müller H G, Wang J L. Function Data Analysis for Sparse Longitudinal Data [J]. *Journal of American Statistical Association*, 2005, 100(470): 577-590.
- [30] Eddie K H. Consistent Functional PCA for Financial Time Series[J]. *Financial Engineering and Applications*, 2007, 101(51): 103-108.

- [31] Ingrassia S, Costanzo G D. Functional Principal Component Analysis of Financial Time Series[J]. Journal of Statistical Planning and Inference, 2009, 11(7): 2388-2398.
- [32] 张崇岐, 赵娜, 孔丹. 函数数据分析新进展[J]. 广州大学学报(自然科学版), 2006, 12(03): 1-4.
- [33] 严明义. 函数性数据的统计分析: 思想、方法和应用[J]. 统计研究, 2007, 20(02): 87-94.
- [34] 靳刘蕊. 函数性数据分析方法及应用研究[D]. 厦门: 厦门大学博士学位论文, 2008: 40-50.
- [35] 剡亮亮. 基于函数性视角的经济数据分析—以主微分分析方法为例[J]. 统计与信息论坛, 2013, 28(01): 40-46.
- [36] 毛娟. 隐含波动率的函数型数据分析[D]. 武汉: 武汉理工大学硕士学位论文, 2008: 10-21.
- [37] 岳敏, 朱建平. 基于函数型主成分的中国股市波动研究[J]. 统计与信息论坛, 2009, 24(03): 52-56.
- [38] 王诚. 基于函数型数据分析的证券投资收益研究[D]. 北京: 首都经济贸易大学硕士学位论文, 2014: 20-55.
- [39] 郭均鹏, 孙钦堂, 李汶华. Shibor 市场中各期限利率波动模式分析—基于 FPCA 方法[J]. 系统工程, 2012, 30(12): 84-88.
- [40] 胡梦荻. 基于函数型数据分析的波动率研究[D]. 厦门: 厦门大学硕士学位论文, 2014: 24-31.
- [41] 陈丽琼. 函数型数据分析方法在股票价格预测上的应用[D]. 厦门: 厦门大学硕士学位论文, 2014: 32-34.
- [42] 张蕾蕾. 基于函数型数据分析的期货价格曲线形态识别[D]. 广州: 华南理工大学硕士学位论文, 2014: 12-15.
- [43] Zhang W, Kang P, Fang X, et al. Joint Sparse Representation and Locality Preserving Projection for Feature Extraction[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(7): 1731-1745.
- [44] Guo S, Yang N, Zhang Z, et al. Space Object Deconvolution From Wavefront Sensing by Wavefront Phase Modelling Based on Simplex Splines Function[J]. Infrared & Laser Engineering, 2019, 48(1): 11-55.
- [45] Jiale G, Jie F. Explore of Application of NURBS Spline Function in Realization of Nonlinear Compensation in Control System[J]. Chemical and Pharmaceutical Engineering, 2018, 15(2): 21-36.
- [46] Bezier, P E. Example of an Existing System in the Motor Industry: The Unisurf Sy-

- stem[J]. Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences, 1971, 321(1545): 207-218.
- [47] Lamnii A, Lamnii M, Oumellal F. Computation of Hermite Interpolation in Terms of B-spline Basis Using Polar Forms[J]. Mathematics & Computers in Simulation, 2017, 134(124): 17-27.
- [48] Zeinali M, Shahmorad S, Mirnia K. Piecewise Cubic Interpolation of Fuzzy Data Based on B-spline Basis Functions[J]. Iranian Journal of Fuzzy Systems, 2016, 13(1): 211-221.
- [49] Parand K, Bajalan N. A Numerical Approach Based on B-spline Basis Functions to Solve Boundary Layer Flow Model of A Non-Newtonian Fluid[J]. Journal of the Brazilian Society of Mechanical Sciences & Engineering, 2018, 40(10): 10-21.
- [50] Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space [J]. Philosophical Magazine, 1901, 2(6): 559-572.
- [51] 张文彤, 董伟. SPSS 统计分析高级教程[M]. 北京: 高等教育出版社, 2013: 150-170.

附 录

附录 1 相关数据记录

表 1 1978-2018 地区生产总值分省年度数据

(亿元)

| 时间 | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|--------|----------|---------|----------|---------|---------|----------|---------|---------|----------|
| 1978 年 | 108.8 | 82.65 | 183.06 | 87.99 | 58.04 | 229.2 | 81.98 | 174.8 | 272.81 |
| 1979 年 | 120.1 | 93.01 | 203.22 | 106.43 | 64.14 | 245 | 91.12 | 187.2 | 286.43 |
| 1980 年 | 139.1 | 103.53 | 219.24 | 108.76 | 68.4 | 281 | 98.59 | 221 | 311.89 |
| 1981 年 | 139.2 | 107.96 | 222.54 | 121.71 | 77.91 | 288.6 | 111.16 | 228.3 | 324.76 |
| 1982 年 | 154.9 | 114.11 | 251.45 | 139.22 | 93.22 | 315.1 | 121.67 | 248.4 | 337.07 |
| 1983 年 | 183.1 | 123.42 | 283.21 | 155.06 | 105.88 | 364 | 150.14 | 276.9 | 351.81 |
| 1984 年 | 216.6 | 147.53 | 332.22 | 197.42 | 128.2 | 438.2 | 174.39 | 318.3 | 390.85 |
| 1985 年 | 257.1 | 175.78 | 396.75 | 218.99 | 163.83 | 518.6 | 200.44 | 355 | 466.75 |
| 1986 年 | 284.9 | 194.74 | 436.65 | 235.11 | 181.58 | 605.3 | 227.15 | 400.8 | 490.83 |
| 1987 年 | 326.8 | 220.12 | 521.92 | 257.23 | 212.27 | 719.1 | 297.49 | 454.6 | 545.46 |
| 1988 年 | 410.2 | 259.71 | 701.33 | 316.69 | 270.81 | 881 | 368.67 | 552 | 648.3 |
| 1989 年 | 456 | 283.49 | 822.83 | 376.26 | 292.69 | 1003.8 | 391.65 | 630.6 | 696.54 |
| 1990 年 | 500.8 | 310.95 | 896.33 | 429.27 | 319.31 | 1062.7 | 425.28 | 715.2 | 781.66 |
| 1991 年 | 598.9 | 342.65 | 1072.07 | 468.51 | 359.66 | 1200.1 | 463.47 | 822.3 | 893.77 |
| 1992 年 | 709.1 | 411.04 | 1278.5 | 570.12 | 421.68 | 1473 | 558.06 | 959.7 | 1114.32 |
| 1993 年 | 886.21 | 538.94 | 1690.84 | 680.41 | 537.81 | 2010.82 | 718.58 | 1198.3 | 1519.23 |
| 1994 年 | 1145.31 | 732.89 | 2187.49 | 826.66 | 695.06 | 2461.78 | 937.73 | 1604.9 | 1990.86 |
| 1995 年 | 1507.69 | 931.97 | 2849.52 | 1076.03 | 857.06 | 2793.37 | 1137.23 | 1991.4 | 2499.43 |
| 1996 年 | 1789.2 | 1121.93 | 3452.97 | 1292.11 | 1023.09 | 3157.69 | 1346.79 | 2370.5 | 2957.55 |
| 1997 年 | 2077.09 | 1264.63 | 3953.78 | 1476 | 1153.51 | 3582.46 | 1464.34 | 2667.5 | 3438.79 |
| 1998 年 | 2377.18 | 1374.6 | 4256.01 | 1611.08 | 1262.54 | 3881.73 | 1577.05 | 2774.4 | 3801.09 |
| 1999 年 | 2678.82 | 1500.95 | 4514.19 | 1667.1 | 1379.31 | 4171.69 | 1672.96 | 2866.3 | 4188.73 |
| 2000 年 | 3161.66 | 1701.88 | 5043.96 | 1845.72 | 1539.12 | 4669.06 | 1951.51 | 3151.4 | 4771.17 |
| 2001 年 | 3707.96 | 1919.09 | 5516.76 | 2029.53 | 1713.81 | 5033.08 | 2120.35 | 3390.1 | 5210.12 |
| 2002 年 | 4315 | 2150.76 | 6018.28 | 2324.8 | 1940.94 | 5458.22 | 2348.54 | 3637.2 | 5741.03 |
| 2003 年 | 5007.21 | 2578.03 | 6921.29 | 2855.23 | 2388.38 | 6002.54 | 2662.08 | 4057.4 | 6694.23 |
| 2004 年 | 6033.21 | 3110.97 | 8477.63 | 3571.37 | 3041.07 | 6672 | 3122.01 | 4750.6 | 8072.83 |
| 2005 年 | 6969.52 | 3905.64 | 10012.11 | 4230.53 | 3905.03 | 8047.26 | 3620.27 | 5513.7 | 9247.66 |
| 2006 年 | 8117.78 | 4462.74 | 11467.6 | 4878.61 | 4944.25 | 9304.52 | 4275.12 | 6211.8 | 10572.24 |
| 2007 年 | 9846.81 | 5252.76 | 13607.32 | 6024.45 | 6423.18 | 11164.3 | 5284.69 | 7104 | 12494.01 |
| 2008 年 | 11115 | 6719.01 | 16011.97 | 7315.4 | 8496.2 | 13668.58 | 6426.1 | 8314.37 | 14069.86 |
| 2009 年 | 12153.03 | 7521.85 | 17235.48 | 7358.31 | 9740.25 | 15212.49 | 7278.75 | 8587 | 15046.45 |
| 2010 年 | 14113.58 | 9224.46 | 20394.26 | 9200.86 | 11672 | 18457.27 | 8667.58 | 10368.6 | 17165.98 |

表 1 (续表)

| 时间 | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 2011 年 | 16251.93 | 11307.28 | 24515.76 | 11237.55 | 14359.88 | 22226.7 | 10568.83 | 12582 | 19195.69 |
| 2012 年 | 17879.4 | 12893.88 | 26575.01 | 12112.83 | 15880.58 | 24846.43 | 11939.24 | 13691.58 | 20181.72 |
| 2013 年 | 19800.81 | 14442.01 | 28442.95 | 12665.25 | 16916.5 | 27213.22 | 13046.4 | 14454.91 | 21818.15 |
| 2014 年 | 21330.83 | 15726.93 | 29421.15 | 12761.49 | 17770.19 | 28626.58 | 13803.14 | 15039.38 | 23567.7 |
| 2015 年 | 23014.59 | 16538.19 | 29806.11 | 12766.49 | 17831.51 | 28669.02 | 14063.13 | 15083.67 | 25123.45 |
| 2016 年 | 25669.13 | 17885.39 | 32070.45 | 13050.41 | 18128.1 | 22246.9 | 14776.8 | 15386.09 | 28178.65 |
| 2017 年 | 28014.94 | 18549.19 | 34016.32 | 15528.42 | 16096.21 | 23409.24 | 14944.53 | 15902.68 | 30632.99 |
| 2018 年 | 30319.98 | 18809.64 | 36010.27 | 16818.11 | 17289.22 | 25315.35 | 15074.62 | 16361.62 | 32679.87 |

| 时间 | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|--------|----------|----------|---------|---------|---------|----------|----------|---------|---------|
| 1978 年 | 249.24 | 123.72 | 113.96 | 66.37 | 87 | 225.45 | 162.92 | 151 | 146.99 |
| 1979 年 | 298.55 | 157.75 | 127.31 | 74.11 | 104.15 | 251.6 | 190.09 | 188.46 | 178.01 |
| 1980 年 | 319.8 | 179.92 | 140.88 | 87.06 | 111.15 | 292.13 | 229.16 | 199.38 | 191.72 |
| 1981 年 | 350.02 | 204.86 | 170.51 | 105.62 | 121.26 | 346.57 | 249.69 | 219.75 | 209.68 |
| 1982 年 | 390.17 | 234.01 | 187.02 | 117.81 | 133.96 | 395.38 | 263.3 | 241.55 | 232.52 |
| 1983 年 | 437.65 | 257.09 | 215.68 | 127.76 | 144.13 | 459.83 | 327.95 | 262.58 | 257.43 |
| 1984 年 | 518.85 | 323.25 | 265.74 | 157.06 | 169.11 | 581.56 | 370.04 | 328.22 | 287.29 |
| 1985 年 | 651.82 | 429.16 | 331.24 | 200.48 | 207.26 | 680.46 | 451.74 | 396.26 | 349.95 |
| 1986 年 | 744.94 | 502.47 | 382.76 | 222.54 | 227.23 | 742.05 | 502.91 | 442.04 | 397.68 |
| 1987 年 | 922.33 | 606.99 | 442.35 | 279.24 | 261.83 | 892.29 | 609.6 | 517.77 | 469.44 |
| 1988 年 | 1208.85 | 770.25 | 546.94 | 383.21 | 321.36 | 1117.66 | 749.09 | 626.52 | 584.07 |
| 1989 年 | 1321.85 | 849.44 | 616.25 | 458.4 | 366.87 | 1293.94 | 850.71 | 717.08 | 640.8 |
| 1990 年 | 1416.5 | 904.69 | 658.02 | 522.28 | 419.54 | 1511.19 | 934.65 | 824.38 | 744.44 |
| 1991 年 | 1601.38 | 1089.33 | 663.6 | 619.87 | 465.1 | 1810.54 | 1045.73 | 913.38 | 833.3 |
| 1992 年 | 2136.02 | 1375.7 | 801.16 | 784.68 | 559.52 | 2196.53 | 1279.75 | 1088.39 | 986.98 |
| 1993 年 | 2998.16 | 1925.91 | 1037.14 | 1114.2 | 723.04 | 2770.37 | 1660.18 | 1325.83 | 1244.71 |
| 1994 年 | 4057.39 | 2689.28 | 1320.43 | 1644.39 | 948.16 | 3844.5 | 2216.83 | 1700.92 | 1650.02 |
| 1995 年 | 5155.25 | 3557.55 | 1810.66 | 2094.9 | 1169.73 | 4953.35 | 2988.37 | 2109.38 | 2132.13 |
| 1996 年 | 6004.21 | 4188.53 | 2093.3 | 2484.25 | 1409.74 | 5883.8 | 3634.69 | 2499.77 | 2540.13 |
| 1997 年 | 6680.34 | 4686.11 | 2347.32 | 2870.9 | 1605.77 | 6537.07 | 4041.09 | 2856.47 | 2849.27 |
| 1998 年 | 7199.95 | 5052.62 | 2542.96 | 3159.91 | 1719.87 | 7021.35 | 4308.24 | 3114.02 | 3025.53 |
| 1999 年 | 7697.82 | 5443.92 | 2712.34 | 3414.19 | 1853.65 | 7493.84 | 4517.94 | 3229.29 | 3214.54 |
| 2000 年 | 8553.69 | 6141.03 | 2902.09 | 3764.54 | 2003.07 | 8337.47 | 5052.99 | 3545.39 | 3551.49 |
| 2001 年 | 9456.84 | 6898.34 | 3246.71 | 4072.85 | 2175.68 | 9195.04 | 5533.01 | 3880.53 | 3831.9 |
| 2002 年 | 10606.85 | 8003.67 | 3519.72 | 4467.55 | 2450.48 | 10275.5 | 6035.48 | 4212.82 | 4151.54 |
| 2003 年 | 12442.87 | 9705.02 | 3923.11 | 4983.67 | 2807.41 | 12078.15 | 6867.7 | 4757.45 | 4659.99 |
| 2004 年 | 15003.6 | 11648.7 | 4759.3 | 5763.35 | 3456.7 | 15021.84 | 8553.79 | 5633.24 | 5641.94 |
| 2005 年 | 18598.69 | 13417.68 | 5350.17 | 6554.69 | 4056.76 | 18366.87 | 10587.42 | 6590.19 | 6596.1 |

表 1 (续表)

| 时间 | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 2006 年 | 21742.05 | 15718.47 | 6112.5 | 7583.85 | 4820.53 | 21900.19 | 12362.79 | 7617.47 | 7688.67 |
| 2007 年 | 26018.48 | 18753.73 | 7360.92 | 9248.53 | 5800.25 | 25776.91 | 15012.46 | 9333.4 | 9439.6 |
| 2008 年 | 30981.98 | 21462.69 | 8851.66 | 10823.01 | 6971.05 | 30933.28 | 18018.53 | 11328.92 | 11555 |
| 2009 年 | 34457.3 | 22990.35 | 10062.82 | 12236.53 | 7655.18 | 33896.65 | 19480.46 | 12961.1 | 13059.69 |
| 2010 年 | 41425.48 | 27722.31 | 12359.33 | 14737.12 | 9451.26 | 39169.92 | 23092.36 | 15967.61 | 16037.96 |
| 2011 年 | 49110.27 | 32318.85 | 15300.65 | 17560.18 | 11702.82 | 45361.85 | 26931.03 | 19632.26 | 19669.56 |
| 2012 年 | 54058.22 | 34665.33 | 17212.05 | 19701.78 | 12948.88 | 50013.24 | 29599.31 | 22250.45 | 22154.23 |
| 2013 年 | 59753.37 | 37756.59 | 19229.34 | 21868.49 | 14410.19 | 55230.32 | 32191.3 | 24791.83 | 24621.67 |
| 2014 年 | 65088.32 | 40173.03 | 20848.75 | 24055.76 | 15714.63 | 59426.59 | 34938.24 | 27379.22 | 27037.32 |
| 2015 年 | 70116.38 | 42886.49 | 22005.63 | 25979.82 | 16723.78 | 63002.33 | 37002.16 | 29550.19 | 28902.21 |
| 2016 年 | 77388.28 | 47251.36 | 24407.62 | 28810.58 | 18499 | 68024.49 | 40471.79 | 32665.38 | 31551.37 |
| 2017 年 | 85869.76 | 51768.26 | 27018 | 32182.09 | 20006.31 | 72634.15 | 44552.83 | 35478.09 | 33902.96 |
| 2018 年 | 92595.4 | 56197.15 | 30006.82 | 35804.04 | 21984.78 | 76469.67 | 48055.86 | 39366.55 | 36425.78 |

| 时间 | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|--------|----------|---------|--------|---------|--------|---------|---------|----------|---------|
| 1978 年 | 185.85 | 64.73 | 16.4 | 71.7 | 8.9 | 46.62 | 69.05 | 184.61 | 81.07 |
| 1979 年 | 209.34 | 67.51 | 17.45 | 80.98 | 11.11 | 55.28 | 76.83 | 205.76 | 94.52 |
| 1980 年 | 249.65 | 73.9 | 19.33 | 90.68 | 10.9 | 60.26 | 84.27 | 229.31 | 94.91 |
| 1981 年 | 290.36 | 70.89 | 22.23 | 97.2 | 12.11 | 67.89 | 94.13 | 242.32 | 102.09 |
| 1982 年 | 339.92 | 76.88 | 28.86 | 108.08 | 14.51 | 79.39 | 110.12 | 275.23 | 111.95 |
| 1983 年 | 368.75 | 91.5 | 31.12 | 120.01 | 13.02 | 87.38 | 120.07 | 311 | 123.39 |
| 1984 年 | 458.74 | 103.17 | 37.18 | 141.64 | 13.68 | 108.27 | 139.58 | 358.0552 | 149.35 |
| 1985 年 | 577.38 | 123.39 | 43.26 | 164.32 | 17.76 | 123.92 | 164.96 | 421.15 | 180.87 |
| 1986 年 | 667.53 | 140.74 | 48.03 | 184.6 | 16.95 | 139.57 | 182.28 | 458.226 | 208.31 |
| 1987 年 | 846.69 | 159.52 | 57.28 | 206.73 | 17.71 | 165.5 | 229.03 | 530.86 | 244.96 |
| 1988 年 | 1155.37 | 191.84 | 77 | 261.27 | 20.25 | 211.79 | 301.09 | 659.69 | 314.48 |
| 1989 年 | 1381.39 | 216.84 | 91.32 | 303.75 | 21.86 | 235.84 | 363.05 | 744.9788 | 358.37 |
| 1990 年 | 1559.03 | 242.81 | 102.42 | 327.75 | 24.45 | 260.14 | 451.67 | 890.9464 | 404.3 |
| 1991 年 | 1893.30 | 271.39 | 120.52 | 374.18 | 30.53 | 295.9 | 517.41 | 1016.31 | 468.37 |
| 1992 年 | 2447.54 | 317.79 | 184.92 | 461.32 | 33.29 | 339.91 | 618.69 | 1177.27 | 531.63 |
| 1993 年 | 3469.28 | 372.24 | 260.41 | 608.53 | 37.42 | 417.69 | 783.27 | 1486.08 | 678.2 |
| 1994 年 | 4619.02 | 453.61 | 331.98 | 833.6 | 45.99 | 524.46 | 983.78 | 2001.41 | 839.03 |
| 1995 年 | 5933.05 | 557.76 | 363.25 | 1123.06 | 56.11 | 636.21 | 1222.15 | 2443.21 | 1036.85 |
| 1996 年 | 6834.97 | 722.52 | 389.68 | 1315.12 | 64.98 | 723.18 | 1517.69 | 2871.65 | 1215.84 |
| 1997 年 | 7774.53 | 793.57 | 411.16 | 1509.75 | 77.24 | 805.79 | 1676.17 | 3241.47 | 1363.6 |
| 1998 年 | 8530.88 | 887.67 | 442.13 | 1602.38 | 91.5 | 858.39 | 1831.33 | 3474.09 | 1458.4 |
| 1999 年 | 9250.68 | 956.32 | 476.67 | 1663.2 | 105.98 | 937.5 | 1899.82 | 3649.12 | 1592.64 |
| 2000 年 | 10741.25 | 1052.88 | 526.82 | 1791 | 117.8 | 1029.92 | 2011.19 | 3928.2 | 1804 |

表 1（续表）

| 时间 | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|--------|----------|---------|---------|----------|---------|----------|----------|----------|----------|
| 2001 年 | 12039.25 | 1125.37 | 579.17 | 1976.86 | 139.16 | 1133.27 | 2138.31 | 4293.49 | 2010.62 |
| 2002 年 | 13502.42 | 1232.03 | 642.73 | 2232.86 | 162.04 | 1243.43 | 2312.82 | 4725.01 | 2253.39 |
| 2003 年 | 15844.64 | 1399.83 | 713.96 | 2555.72 | 185.09 | 1426.34 | 2556.02 | 5333.09 | 2587.72 |
| 2004 年 | 18864.62 | 1688.49 | 819.66 | 3034.58 | 220.34 | 1677.8 | 3081.91 | 6379.63 | 3175.58 |
| 2005 年 | 22557.37 | 1933.98 | 918.75 | 3467.72 | 248.8 | 2005.42 | 3462.73 | 7385.1 | 3933.72 |
| 2006 年 | 26587.76 | 2277.35 | 1065.67 | 3907.23 | 290.76 | 2338.98 | 3988.14 | 8690.24 | 4743.61 |
| 2007 年 | 31777.01 | 2703.98 | 1254.17 | 4676.13 | 341.43 | 2884.11 | 4772.52 | 10562.39 | 5757.29 |
| 2008 年 | 36796.71 | 3166.82 | 1503.06 | 5793.66 | 394.85 | 3561.56 | 5692.12 | 12601.23 | 7314.58 |
| 2009 年 | 39482.56 | 3387.56 | 1654.21 | 6530.01 | 441.36 | 3912.68 | 6169.75 | 14151.28 | 8169.8 |
| 2010 年 | 46013.06 | 4120.75 | 2064.5 | 7925.58 | 507.46 | 4602.16 | 7224.18 | 17185.48 | 10123.48 |
| 2011 年 | 53210.28 | 5020.37 | 2522.66 | 10011.37 | 605.83 | 5701.84 | 8893.12 | 21026.68 | 12512.3 |
| 2012 年 | 57067.92 | 5650.2 | 2855.54 | 11409.6 | 701.03 | 6852.2 | 10309.47 | 23872.8 | 14453.68 |
| 2013 年 | 62474.79 | 6330.69 | 3177.56 | 12783.26 | 815.67 | 8086.86 | 11832.31 | 26392.07 | 16205.45 |
| 2014 年 | 67809.85 | 6836.82 | 3500.72 | 14262.6 | 920.83 | 9266.39 | 12814.59 | 28536.66 | 17689.94 |
| 2015 年 | 72812.55 | 6790.32 | 3702.76 | 15717.27 | 1026.39 | 10502.56 | 13619.17 | 30053.1 | 18021.86 |
| 2016 年 | 80854.91 | 7200.37 | 4053.2 | 17740.59 | 1151.41 | 11776.73 | 14788.42 | 32934.54 | 19399.59 |
| 2017 年 | 89705.23 | 7459.9 | 4462.54 | 19424.73 | 1310.92 | 13540.83 | 16376.34 | 36980.22 | 21898.81 |
| 2018 年 | 97277.77 | 8246.07 | 4832.05 | 20363.19 | 1477.63 | 14806.45 | 17881.12 | 40678.13 | 24438.32 |

| 时间 | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|--------|--------|---------|----------|---------|
| 1978 年 | 15.54 | 13 | 39.07 | 75.85 |
| 1979 年 | 15.19 | 14.36 | 45.63 | 84.59 |
| 1980 年 | 17.79 | 15.96 | 53.24 | 97.33 |
| 1981 年 | 17.49 | 17.42 | 59.41 | 113.46 |
| 1982 年 | 19.95 | 18.22 | 65.24 | 129.15 |
| 1983 年 | 22.45 | 20.79 | 78.55 | 134.6 |
| 1984 年 | 26.42 | 24.78 | 89.75 | 150.27 |
| 1985 年 | 33.01 | 30.27 | 112.24 | 180.97 |
| 1986 年 | 38.44 | 34.54 | 129.04 | 205.46 |
| 1987 年 | 43.38 | 39.63 | 148.5 | 241.56 |
| 1988 年 | 54.96 | 50.29 | 192.72 | 313.28 |
| 1989 年 | 60.37 | 59.21 | 217.29 | 383.44 |
| 1990 年 | 69.94 | 64.84 | 274.01 | 449.06 |
| 1991 年 | 75.1 | 71.78 | 335.91 | 518.59 |
| 1992 年 | 87.52 | 83.14 | 402.31 | 646.6 |
| 1993 年 | 109.68 | 104.49 | 495.25 | 871.7 |
| 1994 年 | 138.4 | 136.26 | 662.32 | 1198.29 |
| 1995 年 | 167.8 | 175.19 | 814.85 | 1497.56 |

表 1（续表）

| 时间 | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|--------|---------|---------|----------|----------|
| 1996 年 | 184.17 | 202.9 | 900.93 | 1697.9 |
| 1997 年 | 202.79 | 224.59 | 1039.85 | 1817.25 |
| 1998 年 | 220.92 | 245.44 | 1106.95 | 1911.3 |
| 1999 年 | 239.38 | 264.58 | 1163.17 | 1971.41 |
| 2000 年 | 263.68 | 295.02 | 1363.56 | 2080.04 |
| 2001 年 | 300.13 | 337.44 | 1491.6 | 2279.34 |
| 2002 年 | 340.65 | 377.16 | 1612.65 | 2523.73 |
| 2003 年 | 390.2 | 445.36 | 1886.35 | 2821.11 |
| 2004 年 | 466.1 | 537.11 | 2209.09 | 3433.5 |
| 2005 年 | 543.32 | 612.61 | 2604.19 | 3984.1 |
| 2006 年 | 648.5 | 725.9 | 3045.26 | 4746.16 |
| 2007 年 | 797.35 | 919.11 | 3523.16 | 5823.41 |
| 2008 年 | 1018.62 | 1203.92 | 4183.21 | 7021 |
| 2009 年 | 1081.27 | 1353.31 | 4277.05 | 7759.16 |
| 2010 年 | 1350.43 | 1689.65 | 5437.47 | 9569.85 |
| 2011 年 | 1670.44 | 2102.21 | 6610.05 | 11720.87 |
| 2012 年 | 1893.54 | 2341.29 | 7505.31 | 13035.1 |
| 2013 年 | 2122.06 | 2577.57 | 8443.84 | 14449.9 |
| 2014 年 | 2303.32 | 2752.1 | 9273.46 | 15672.89 |
| 2015 年 | 2417.05 | 2911.77 | 9324.8 | 16803.12 |
| 2016 年 | 2572.49 | 3168.59 | 9649.7 | 18317.64 |
| 2017 年 | 2624.83 | 3443.56 | 10881.96 | 18523.26 |
| 2018 年 | 2865.23 | 3705.18 | 12199.08 | 20352.51 |

表 2 2000-2018 地区生产总值分省季度数据

（亿元）

| 时间 | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 50.2 | 70.8 | 1228.31 | 60.3 | 100.78 | 474.67 | 86.21 | 461.46 | 100.58 |
| 2 | 1043.4 | 264.56 | 2866.01 | 126.79 | 756.06 | 1124.7 | 375.12 | 1232.62 | 125.81 |
| 3 | 2313.66 | 304.56 | 3349.95 | 1033.57 | 1338.88 | 2064.35 | 723.77 | 2819.6 | 925.82 |
| 4 | 3161.66 | 1701.88 | 5043.96 | 1845.72 | 1539.12 | 4669.06 | 1951.51 | 3151.4 | 4771.17 |
| 5 | 202.36 | 84.39 | 1370.97 | 100.32 | 155.36 | 681.24 | 100.36 | 584.05 | 200.12 |
| 6 | 1779.83 | 256.13 | 3246.16 | 286.95 | 1167.37 | 1549.26 | 154.16 | 1456.35 | 525.48 |
| 7 | 3342.52 | 311.88 | 4178.52 | 307.9 | 1452.48 | 2693.79 | 176.09 | 3024.5 | 1121.68 |
| 8 | 3707.96 | 1919.09 | 5516.76 | 2029.53 | 1713.81 | 5033.08 | 2120.35 | 3390.1 | 5210.12 |
| 9 | 251.62 | 239.58 | 1513.63 | 114.64 | 200.12 | 887.81 | 149.6 | 706.64 | 316.24 |
| 10 | 1816.26 | 352.3 | 3626.31 | 247.11 | 478.68 | 1973.82 | 266.8 | 1680.08 | 1376.77 |
| 11 | 3371.38 | 628.32 | 5007.09 | 417.77 | 566.08 | 3323.23 | 371.59 | 3229.4 | 2369.18 |
| 12 | 4315 | 2150.76 | 6018.28 | 2324.8 | 1940.94 | 5458.22 | 2348.54 | 3637.2 | 5741.03 |
| 13 | 291.51 | 394.77 | 1656.29 | 314.38 | 220.3 | 1094.38 | 290.5 | 829.23 | 811.36 |

表 2 (续表)

| 时间 | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 14 | 2252.69 | 760.73 | 4006.46 | 492.73 | 10.01 | 2398.38 | 587.76 | 1903.81 | 2228.06 |
| 15 | 4000.24 | 1244.76 | 5835.66 | 1143.44 | 320.32 | 3952.67 | 919.27 | 3434.3 | 3616.68 |
| 16 | 5007.21 | 2578.03 | 6921.29 | 2855.23 | 2388.38 | 6002.54 | 2662.08 | 4057.4 | 6694.23 |
| 17 | 366.18 | 549.96 | 1798.95 | 514.12 | 285.39 | 1300.95 | 431.4 | 951.82 | 1306.48 |
| 18 | 3010.88 | 1169.16 | 4386.61 | 1032.57 | 598.7 | 2822.94 | 908.72 | 2127.54 | 3079.35 |
| 19 | 4570.9 | 1861.2 | 6664.23 | 1869.11 | 1206.72 | 4582.11 | 1466.95 | 3639.2 | 4864.18 |
| 20 | 6033.21 | 3110.97 | 8477.63 | 3571.37 | 3041.07 | 6672 | 3122.01 | 4750.6 | 8072.83 |
| 21 | 1023.87 | 705.15 | 1941.61 | 713.86 | 535.66 | 1507.52 | 572.3 | 1074.41 | 1801.6 |
| 22 | 2274.45 | 1577.59 | 4766.76 | 1572.41 | 1187.39 | 3247.5 | 1229.68 | 2351.27 | 3930.64 |
| 23 | 3542.04 | 2477.64 | 7492.8 | 2594.78 | 2093.12 | 5211.55 | 2014.63 | 3844.1 | 6111.68 |
| 24 | 6814.53 | 3663.86 | 10116.56 | 4121.18 | 3822.77 | 8005.01 | 3614.92 | 5510 | 9143.95 |
| 25 | 1681.56 | 860.34 | 2084.27 | 913.6 | 785.93 | 1714.09 | 713.2 | 1197 | 2296.72 |
| 26 | 3538.02 | 1986.02 | 5146.91 | 2112.25 | 1776.08 | 3672.06 | 1550.64 | 2575 | 4781.93 |
| 27 | 5513.18 | 3094.08 | 8321.37 | 3320.45 | 2979.52 | 5840.99 | 2562.31 | 4049 | 7359.18 |
| 28 | 7720.33 | 4337.73 | 11613.7 | 4746.5 | 4790 | 9257.05 | 4249.23 | 6216.8 | 10296.97 |
| 29 | 1882.78 | 989.72 | 2450.1 | 1098.84 | 967.91 | 2016.53 | 854.91 | 1353 | 2630.5 |
| 30 | 4064.26 | 2312.26 | 6037.28 | 2511.56 | 2320.71 | 4419.92 | 1881.01 | 2927 | 5561.91 |
| 31 | 6419.78 | 3601.3 | 9879.99 | 3973.95 | 3852.19 | 7029.03 | 3139.91 | 4547 | 8590.97 |
| 32 | 9006.23 | 5018.28 | 13863.5 | 5696.15 | 6018.81 | 11021.73 | 5226.08 | 7077.2 | 12001.16 |
| 33 | 2148.34 | 1230.63 | 2963.95 | 1365.27 | 1269.76 | 2488.92 | 1064.34 | 1557 | 3040.76 |
| 34 | 4972.8 | 2951.88 | 7375.81 | 3254.35 | 3069.3 | 5578.05 | 2384.09 | 3414 | 6530.73 |
| 35 | 7586.27 | 4623.93 | 11822.48 | 5105.73 | 5169.35 | 8814.68 | 4008.48 | 5351 | 9929.52 |
| 36 | 10488.03 | 6354.38 | 16188.61 | 6938.73 | 7761.8 | 13461.57 | 6424.06 | 8310 | 13698.15 |
| 37 | 2511.88 | 1416.93 | 3023.46 | 1309.21 | 1471.07 | 2642.23 | 1155.21 | 1460 | 3150.47 |
| 38 | 5308.43 | 3154.85 | 7513.56 | 3105.99 | 3512.84 | 5920.62 | 2572.32 | 3231 | 6612 |
| 39 | 8161.44 | 4997.02 | 12212.87 | 4819.07 | 5895.51 | 9472.96 | 4347.45 | 4955 | 10214.82 |
| 40 | 11865.93 | 7500.8 | 17026.6 | 7365.74 | 9725.78 | 15065.57 | 7203.18 | 8288 | 14900.93 |
| 41 | 3116.99 | 1842.4 | 3665.2 | 1807.73 | 1981.72 | 3554.91 | 1443.59 | 1860 | 3810.63 |
| 42 | 6372.62 | 4106.48 | 9083.52 | 4143.38 | 4690.29 | 8205.52 | 3180.1 | 4202 | 7980.16 |
| 43 | 9754.36 | 6448.59 | 14641.23 | 6262.61 | 7759.11 | 13055.94 | 5281.59 | 6450.46 | 12109.52 |
| 44 | 13777.94 | 9108.83 | 20197.09 | 9088.06 | 11655 | 18278.29 | 8577.06 | 10235 | 16872.42 |
| 45 | 3508.97 | 2256.51 | 4396.11 | 2200.87 | 2391.71 | 4246.19 | 1707.97 | 2180 | 4327.63 |
| 46 | 7418.07 | 5098.65 | 10992.43 | 5134.34 | 5804.83 | 9948.45 | 3890.44 | 5067 | 9164.1 |
| 47 | 11404.26 | 8006.26 | 17821.94 | 7906.51 | 9571.89 | 15709.16 | 6526.69 | 7836.27 | 13725.64 |
| 48 | 16011.43 | 11190.99 | 24228.18 | 11100.18 | 14246.11 | 22025.92 | 10530.71 | 12503.83 | 19195.69 |
| 49 | 3761.52 | 2582.6 | 4919.97 | 2494.07 | 2704.34 | 4811.74 | 1972.46 | 2461.3 | 4593.85 |
| 50 | 8348.59 | 5864.94 | 12219.21 | 5785.87 | 6673.15 | 11228.08 | 4437.97 | 5416.69 | 9552.24 |
| 51 | 12678.14 | 9188.45 | 19581.85 | 8712.89 | 10803.26 | 17643.69 | 7382.56 | 8391.78 | 14374.23 |
| 52 | 17801.02 | 12885.18 | 26575.01 | 12112.81 | 15988.34 | 24801.3 | 11937.82 | 13691.57 | 20101.33 |
| 53 | 4101.23 | 2915.85 | 5312.15 | 2610.08 | 2943.92 | 5296.16 | 2166.02 | 2608.67 | 4937.5 |

表 2 (续表)

| 时间 | 北京市 | 天津市 | 河北省 | 山西省 | 内蒙古自治区 | 辽宁省 | 吉林省 | 黑龙江省 | 上海市 |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 54 | 9112.79 | 6579.01 | 13154.58 | 6016.58 | 7087.55 | 12334.73 | 4808.01 | 5545.1 | 10168.52 |
| 55 | 13766.18 | 10223.04 | 20947.32 | 9028.99 | 11338 | 19263.93 | 8014.61 | 8748.81 | 15474.13 |
| 56 | 19500.6 | 14370.2 | 28301.4 | 12602.2 | 16832.4 | 27077.7 | 12981.5 | 14382.9 | 21602.1 |
| 57 | 4412.95 | 3190.3 | 5426.83 | 2631.67 | 3161.38 | 5663.64 | 2305.1 | 2644.96 | 5313.07 |
| 58 | 9769.32 | 7185.46 | 13640 | 6097.8 | 7338.66 | 13207.48 | 5141.7 | 5672.27 | 10952.64 |
| 59 | 14774.06 | 11101.31 | 21709.99 | 9109.9 | 11709.03 | 20361.68 | 8449.09 | 8987.03 | 16607.08 |
| 60 | 21330.83 | 15722.47 | 29421.15 | 12759.44 | 17769.51 | 28626.58 | 13803.81 | 15039.38 | 23560.94 |
| 61 | 4773.46 | 3709.03 | 6029.46 | 2577.24 | 3274.5 | 5719.1 | 2391.06 | 2561.01 | 5815.79 |
| 62 | 10578.28 | 7884.04 | 13798.13 | 5814.18 | 7424.9 | 13004.97 | 5370.21 | 5435.25 | 11887 |
| 63 | 16002.35 | 12321.02 | 21280.18 | 9003.26 | 12092.76 | 20404.6 | 8896.9 | 8790.9 | 17866.24 |
| 64 | 22968.59 | 16538.19 | 29806.11 | 12802.58 | 18032.79 | 28743.39 | 14274.11 | 15083.67 | 24964.99 |
| 65 | 5451.89 | 4039.44 | 6487.44 | 2465.99 | 3495.54 | 5647.1 | 2484.68 | 2601.71 | 6225.39 |
| 66 | 11413.83 | 8500.91 | 14467.69 | 5714 | 7579.98 | 12812.57 | 5604.85 | 5630.29 | 12956.99 |
| 67 | 17367.77 | 13339.44 | 22496.12 | 8945.55 | 12690.14 | 19952.81 | 9298.11 | 9231.56 | 19529.67 |
| 68 | 24899.26 | 17885.39 | 31827.86 | 12928.34 | 18632.57 | 22037.88 | 14886.23 | 15386.09 | 27466.15 |
| 69 | 6040.48 | 4667.18 | 7512.43 | 2809.42 | 3742.2 | 4574.72 | 2683.43 | 2797.95 | 6922.84 |
| 70 | 12406.79 | 9386.87 | 16404.87 | 6577.91 | 8465.71 | 10296.82 | 6124.22 | 6108.01 | 13908.57 |
| 71 | 19569.76 | 13449.5 | 25603.47 | 10460.38 | 13740.21 | 16639.4 | 9970.58 | 9901.81 | 21617.52 |
| 72 | 28000.35 | 18595.38 | 35964 | 14973.51 | 16103.17 | 23942 | 15288.94 | 16199.88 | 30133.86 |
| 73 | 6801.47 | 4959.42 | 7564.02 | 3122.07 | 3290.47 | 5125.03 | 2659.29 | 3067.19 | 7863.4 |
| 74 | 14051.23 | 9927.6 | 16600.54 | 7482.75 | 7776.67 | 11383.28 | 6299.07 | 6240.62 | 15558.15 |
| 75 | 21511.11 | 14658.36 | 25226.33 | 11640.33 | 12309.17 | 18012.38 | 9957.68 | 9859.52 | 23656.69 |
| 76 | 30319.98 | 18809.64 | 36010.27 | 16818.11 | 17289.22 | 25315.35 | 15074.62 | 16361.62 | 32679.87 |

| 时间 | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|----|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 114.94 | 120.36 | 393.58 | 776.5 | 139.08 | 100.36 | 711.58 | 1674.9 | 462.77 |
| 2 | 979 | 947.05 | 978.24 | 1809.66 | 165.29 | 990.33 | 1642.83 | 2731.84 | 974.92 |
| 3 | 2239.15 | 988.5 | 1415.73 | 3253.03 | 286.84 | 1249.15 | 2641.96 | 4350.91 | 1229.1 |
| 4 | 8553.69 | 6141.03 | 2902.09 | 3764.54 | 2003.07 | 8337.47 | 5052.99 | 3545.39 | 3551.49 |
| 5 | 619.36 | 452.93 | 530.14 | 871.97 | 259.89 | 719.37 | 989.47 | 1651.36 | 640.13 |
| 6 | 1579.37 | 976.44 | 1296.22 | 2033.09 | 444.38 | 1625.79 | 2286.84 | 2824.59 | 1346.8 |
| 7 | 2302.95 | 3092.4 | 1918.61 | 3532.26 | 747.33 | 3675.08 | 3703.66 | 4487.1 | 1882.64 |
| 8 | 9456.84 | 6898.34 | 3246.71 | 4072.85 | 2175.68 | 9195.04 | 5533.01 | 3880.53 | 3831.9 |
| 9 | 1353.66 | 1026.22 | 666.7 | 967.44 | 380.7 | 1441.57 | 1267.36 | 1627.82 | 817.49 |
| 10 | 3237.74 | 3005.83 | 1614.2 | 2256.52 | 723.47 | 3341.91 | 2930.85 | 2917.34 | 1718.68 |
| 11 | 4845.05 | 3073.3 | 2421.49 | 3811.49 | 1207.82 | 6101.01 | 4765.36 | 4623.29 | 2536.18 |
| 12 | 10606.85 | 8003.67 | 3519.72 | 4467.55 | 2450.48 | 10275.5 | 6035.48 | 4212.82 | 4151.54 |
| 13 | 2087.96 | 1599.51 | 803.26 | 1062.91 | 501.51 | 2163.77 | 1545.25 | 1604.28 | 994.85 |
| 14 | 4896.11 | 4035.22 | 1932.18 | 2479.95 | 1002.56 | 5058.03 | 3574.86 | 3010.09 | 2090.56 |
| 15 | 7387.15 | 5054.2 | 2924.37 | 4090.72 | 1668.31 | 8526.94 | 5827.06 | 4759.48 | 3189.72 |

表 2 (续表)

| 时间 | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 16 | 12442.87 | 9705.02 | 3923.11 | 4983.67 | 2807.41 | 12078.15 | 6867.7 | 4757.45 | 4659.99 |
| 17 | 2822.26 | 2172.8 | 939.82 | 1158.38 | 622.32 | 2885.97 | 1823.14 | 1580.74 | 1172.21 |
| 18 | 6554.48 | 5064.61 | 2250.16 | 2703.38 | 1281.65 | 6774.15 | 4218.87 | 3102.84 | 2462.44 |
| 19 | 9929.25 | 7035.1 | 3427.25 | 4369.95 | 2128.8 | 10952.87 | 6888.76 | 4895.67 | 3843.26 |
| 20 | 15003.6 | 11648.7 | 4759.3 | 5763.35 | 3456.7 | 15021.84 | 8553.79 | 5633.24 | 5641.94 |
| 21 | 3556.56 | 2746.09 | 1076.38 | 1253.85 | 743.13 | 3608.17 | 2101.03 | 1557.2 | 1349.57 |
| 22 | 8212.85 | 6094 | 2568.14 | 2926.81 | 1560.74 | 8490.27 | 4862.88 | 3195.59 | 2834.32 |
| 23 | 12471.35 | 9016 | 3930.13 | 4649.18 | 2589.29 | 13378.8 | 7950.46 | 5031.86 | 4496.8 |
| 24 | 18272.12 | 13365 | 5375.84 | 6560.07 | 4056.16 | 18468.3 | 10535.2 | 6484.5 | 6473.61 |
| 25 | 4290.86 | 3319.38 | 1212.94 | 1349.32 | 863.94 | 4330.37 | 2378.92 | 1533.66 | 1526.93 |
| 26 | 9871.22 | 7123.39 | 2886.12 | 3150.24 | 1839.83 | 10206.39 | 5506.89 | 3288.34 | 3206.2 |
| 27 | 15013.45 | 10996.9 | 4433.01 | 4928.41 | 3049.78 | 15804.73 | 9012.16 | 5168.05 | 5150.34 |
| 28 | 21548.36 | 15648.93 | 6141.91 | 7501.63 | 4618.77 | 21846.71 | 12464.09 | 7497.17 | 7493.17 |
| 29 | 5054.5 | 3899.01 | 1418.72 | 1567.11 | 993.33 | 5122.08 | 2872.92 | 1738.46 | 1814.43 |
| 30 | 11752.1 | 8344.4 | 3404.58 | 3752.03 | 2244.26 | 12072.42 | 6817.02 | 3953.61 | 3930.28 |
| 31 | 18010.36 | 12984.08 | 5203.8 | 5941.03 | 3700.72 | 18562.79 | 11094.78 | 6272.22 | 6191.82 |
| 32 | 25560.1 | 18638.36 | 7345.69 | 9160.14 | 5469.25 | 25887.67 | 15058.07 | 9150.01 | 9145 |
| 33 | 6116.83 | 4601.97 | 1725.05 | 1865.23 | 1212.59 | 6185.6 | 3658.01 | 2157.72 | 2170.85 |
| 34 | 14239.82 | 9888.75 | 4180.74 | 4528.32 | 2779.45 | 14736.53 | 8593.53 | 4971.92 | 4911.36 |
| 35 | 21840.63 | 15301.92 | 6295.99 | 7121.95 | 4560.94 | 22506.94 | 13902.26 | 7913.49 | 7694.8 |
| 36 | 30312.61 | 21486.92 | 8874.17 | 10823.11 | 6480.33 | 31072.06 | 18407.78 | 11330.38 | 11156.64 |
| 37 | 6653.17 | 4632.13 | 1885.31 | 1965.06 | 1281.12 | 6469.32 | 4067.44 | 2423.15 | 2409.34 |
| 38 | 15530.49 | 10043.54 | 4512.9 | 4747.44 | 2907.81 | 15688.53 | 8962.8 | 5378.05 | 5350.21 |
| 39 | 23915.19 | 15828.08 | 7200.52 | 7584.29 | 4800.88 | 23959.71 | 14561.5 | 8656.84 | 8382.07 |
| 40 | 34061.19 | 22832.43 | 10052.86 | 11949.53 | 7589.22 | 33805.3 | 19367.28 | 12831.52 | 12930.69 |
| 41 | 8203.53 | 5363.31 | 2303.28 | 2339.71 | 1754.67 | 7672.18 | 4752.01 | 2934.37 | 2956.22 |
| 42 | 19114.1 | 11761.51 | 5537.15 | 5727.6 | 4001.19 | 18645.92 | 10461.25 | 6896.65 | 6805.39 |
| 43 | 29351.53 | 18765.14 | 8820.79 | 9085.47 | 6533.19 | 28261.86 | 16937.95 | 11093.96 | 10583.27 |
| 44 | 40903.34 | 27226.75 | 12263.36 | 14357.12 | 9435.01 | 39416.2 | 22942.68 | 15806.09 | 15902.12 |
| 45 | 9902.13 | 6311.24 | 2843.23 | 2930.73 | 2160.68 | 9049.63 | 5615.44 | 3592.36 | 3646.6 |
| 46 | 22918.26 | 14653.64 | 6883.31 | 7042.53 | 4932.26 | 21880.83 | 12404.68 | 8571.47 | 8762.66 |
| 47 | 35113.06 | 22627.41 | 11078.1 | 11030.05 | 8086.75 | 33030.95 | 20370.43 | 13578.42 | 13625.29 |
| 48 | 48604.26 | 32000.1 | 15110.31 | 17410.21 | 11583.8 | 45429.21 | 27232.04 | 19594.19 | 19635.19 |
| 49 | 10881.46 | 6724.64 | 3289.35 | 3311.67 | 2455.54 | 10089.06 | 6428.7 | 4233.88 | 4218.98 |
| 50 | 25382.81 | 15790.41 | 7781.94 | 7982.63 | 5403.63 | 24118.06 | 13530.55 | 9885.4 | 9909.42 |
| 51 | 38796.36 | 24214.83 | 12516.3 | 12470.57 | 9004.19 | 36235.21 | 22171.49 | 15390 | 15339.15 |
| 52 | 54058.22 | 34606.3 | 17212.05 | 19701.78 | 12948.48 | 50013.24 | 29810.14 | 22250.16 | 22154.23 |
| 53 | 11881.25 | 7261.54 | 3652.36 | 3713.44 | 2712.28 | 11076.48 | 6993.68 | 4699.47 | 4658.82 |
| 54 | 27603.98 | 16953.92 | 8591.28 | 8855.26 | 5901.63 | 25958.15 | 14556.63 | 10949.3 | 10921.81 |
| 55 | 41934.3 | 26195.03 | 13613.86 | 13887.63 | 9874.28 | 39601.68 | 23516.02 | 17099.64 | 16913.58 |

表 2（续表）

| 时间 | 江苏省 | 浙江省 | 安徽省 | 福建省 | 江西省 | 山东省 | 河南省 | 湖北省 | 湖南省 |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 56 | 59161.8 | 37568.5 | 19038.9 | 21759.6 | 14338.5 | 54684.3 | 32155.9 | 24668.5 | 24501.7 |
| 57 | 12892.85 | 7768.46 | 3971.15 | 4037.73 | 2948.38 | 11994.96 | 7456.88 | 5137.26 | 5042.72 |
| 58 | 30115 | 17978.21 | 9350.54 | 9733.61 | 6464.95 | 28080.26 | 15778.62 | 12061.99 | 11975.1 |
| 59 | 45642.77 | 27822.19 | 14709.5 | 15212.75 | 10797.91 | 42814.23 | 25445.43 | 18874.15 | 18531.24 |
| 60 | 65088.32 | 40153.5 | 20848.75 | 24055.76 | 15708.59 | 59426.59 | 34939.38 | 27367.04 | 27048.46 |
| 61 | 14620.67 | 8342.38 | 4238.27 | 4355.87 | 3406.16 | 12931.54 | 7720.21 | 5487.93 | 5392.25 |
| 62 | 33926.9 | 19280.69 | 9976.57 | 10468.34 | 7203.46 | 29731.68 | 16736.55 | 13104.78 | 12800.44 |
| 63 | 51202.37 | 29684.23 | 15628.23 | 16398.67 | 11568.57 | 45712.59 | 26927.01 | 20423.41 | 20250.5 |
| 64 | 70116.38 | 42886.49 | 22005.6 | 25979.82 | 16723.78 | 63002.33 | 37010.25 | 29550.19 | 29047.21 |
| 65 | 16509.04 | 9356.2 | 4647.26 | 5784.8 | 3766.91 | 14914.77 | 8284.26 | 6456.23 | 6318.99 |
| 66 | 36531.73 | 20762.33 | 11028.49 | 11815.51 | 7827.37 | 31688.29 | 17954.9 | 14114.76 | 13613.48 |
| 67 | 55281.46 | 32234.33 | 17131.95 | 18287.01 | 12587.3 | 48703.82 | 28840.57 | 22198.4 | 21771.08 |
| 68 | 76086.17 | 46484.98 | 24117.87 | 28519.15 | 18364.41 | 67008.19 | 40160.01 | 32297.91 | 31244.68 |
| 69 | 18822.6 | 10551.93 | 5826.82 | 6535.19 | 4318.62 | 16653.29 | 9392.22 | 7254.06 | 7051.11 |
| 70 | 40821.22 | 23383.21 | 12645.41 | 13289.77 | 8961.19 | 35017.39 | 20307.72 | 15871.39 | 15275.51 |
| 71 | 62604.46 | 36958.03 | 19811.35 | 20674.99 | 14725.07 | 54866.8 | 32896.69 | 25076.03 | 24492.29 |
| 72 | 85900.94 | 51768.26 | 27518.67 | 32298.28 | 20818.5 | 72678.18 | 44988.16 | 36522.95 | 34590.56 |
| 73 | 21093.31 | 11691.48 | 6601.41 | 7307.22 | 5048.28 | 18900.62 | 10611 | 8188.84 | 7777.65 |
| 74 | 44863.52 | 25674.27 | 14263.97 | 14840.93 | 10124.53 | 39658.06 | 22244.51 | 17958.17 | 16405 |
| 75 | 67039.28 | 39795.62 | 21632.94 | 23311.8 | 15592.53 | 59607.54 | 35537.4 | 27634.35 | 25321.58 |
| 76 | 92595.4 | 56197.15 | 30006.82 | 35804.04 | 21984.78 | 76469.67 | 48055.86 | 39366.55 | 36425.78 |

| 时间 | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|----|----------|---------|--------|---------|--------|---------|---------|---------|---------|
| 1 | 145.23 | 90.51 | 17.11 | 353.3 | 0.95 | 9.31 | 89.9 | 73.46 | 36.23 |
| 2 | 669.84 | 368.06 | 43.41 | 574.72 | 2.59 | 68.52 | 159.9 | 756.82 | 648.14 |
| 3 | 2878.4 | 719.21 | 124.44 | 1080.14 | 10.54 | 55.63 | 397.7 | 1353.6 | 1054.81 |
| 4 | 10741.25 | 1052.88 | 526.82 | 1791 | 117.8 | 1029.92 | 2011.19 | 3928.2 | 1804 |
| 5 | 378.29 | 94.02 | 26.14 | 410.62 | 0.91 | 57.14 | 22.6 | 349.68 | 120.62 |
| 6 | 1155.4 | 457.69 | 45.75 | 723.62 | 21.05 | 80.16 | 88.23 | 1251.77 | 630.43 |
| 7 | 2706.4 | 810.9 | 23.59 | 1286.89 | 43.79 | 295.84 | 113.41 | 2163.14 | 1393.18 |
| 8 | 12039.25 | 1125.37 | 579.17 | 1976.86 | 139.16 | 1133.27 | 2138.31 | 4293.49 | 2010.62 |
| 9 | 738.65 | 100.47 | 20.83 | 467.94 | 11.77 | 123.59 | 315.3 | 625.9 | 205.01 |
| 10 | 1359.04 | 52.68 | 134.91 | 872.52 | 39.51 | 258.84 | 183.44 | 1746.72 | 187.28 |
| 11 | 1465.6 | 297.41 | 171.62 | 1493.64 | 77.04 | 536.05 | 624.52 | 2972.68 | 268.45 |
| 12 | 13502.42 | 1232.03 | 642.73 | 2232.86 | 162.04 | 1243.43 | 2312.82 | 4725.01 | 2253.39 |
| 13 | 1855.59 | 106.96 | 69.8 | 525.26 | 22.63 | 190.04 | 428 | 902.12 | 210.6 |
| 14 | 3873.48 | 263.05 | 224.07 | 1021.42 | 57.97 | 437.52 | 555.11 | 2241.67 | 604.99 |
| 15 | 5637.6 | 605.72 | 319.65 | 1700.39 | 110.29 | 776.26 | 1135.63 | 3782.22 | 930.08 |
| 16 | 15844.64 | 1399.83 | 713.96 | 2555.72 | 185.09 | 1426.34 | 2556.02 | 5333.09 | 2587.72 |
| 17 | 2972.53 | 212.45 | 118.77 | 582.58 | 33.49 | 256.49 | 540.7 | 1178.34 | 426.21 |

表 2（续表）

| 时间 | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|----|----------|---------|---------|----------|--------|---------|---------|----------|----------|
| 18 | 6387.92 | 473.42 | 313.23 | 1170.32 | 76.43 | 616.2 | 926.78 | 2736.62 | 1022.7 |
| 19 | 9809.6 | 914.03 | 467.68 | 1907.14 | 143.54 | 1016.47 | 1646.74 | 4591.76 | 1591.71 |
| 20 | 18864.62 | 1688.49 | 819.66 | 3034.58 | 220.34 | 1677.8 | 3081.91 | 6379.63 | 3175.58 |
| 21 | 4089.47 | 317.94 | 167.74 | 639.9 | 44.35 | 322.94 | 653.4 | 1454.56 | 641.82 |
| 22 | 8902.36 | 683.79 | 402.39 | 1319.22 | 94.89 | 794.88 | 1298.45 | 3231.57 | 1440.41 |
| 23 | 13981.6 | 1222.34 | 615.71 | 2113.89 | 176.79 | 1256.68 | 2157.85 | 5401.3 | 2253.34 |
| 24 | 21701.27 | 1928.14 | 893.14 | 3069.1 | 250.6 | 1942 | 3472.34 | 7385.11 | 3674.75 |
| 25 | 5206.41 | 423.43 | 216.71 | 697.22 | 55.21 | 389.39 | 766.1 | 1730.78 | 857.43 |
| 26 | 11416.8 | 894.16 | 491.55 | 1468.12 | 113.35 | 973.56 | 1670.12 | 3726.52 | 1858.12 |
| 27 | 18153.6 | 1530.65 | 763.74 | 2320.64 | 210.04 | 1496.89 | 2668.96 | 6210.84 | 2914.97 |
| 28 | 25968.55 | 2275 | 1052.43 | 3486.2 | 290.05 | 2267.43 | 4001.87 | 8637.81 | 4383.91 |
| 29 | 6125.85 | 497.46 | 254.87 | 783.77 | 64.1 | 458.28 | 872.39 | 2032.63 | 988.66 |
| 30 | 13544.46 | 1049.5 | 575.37 | 1677.67 | 140.61 | 1165.4 | 2009.6 | 4408.12 | 2252.81 |
| 31 | 21489.58 | 1792.38 | 894.89 | 2757.91 | 246.64 | 1810.27 | 3201.21 | 7465.1 | 3556.23 |
| 32 | 30673.71 | 2699.2 | 1229.64 | 4111.82 | 342.19 | 2710.28 | 4721.77 | 10505.3 | 5369.85 |
| 33 | 7194.35 | 605.13 | 308.04 | 963.77 | 76.03 | 499.22 | 1056.53 | 2555.49 | 1247.02 |
| 34 | 16018.01 | 1270.62 | 707.66 | 2104.76 | 161.81 | 1392.4 | 2451.5 | 5388.53 | 2898.49 |
| 35 | 25299.68 | 2210.14 | 1062.75 | 3419.56 | 277.32 | 2210.27 | 3885.28 | 9057.65 | 4603.21 |
| 36 | 35696.46 | 3176.11 | 1459.23 | 5096.66 | 395.91 | 3333.4 | 5700.1 | 12506.25 | 6851.32 |
| 37 | 7687.73 | 589.13 | 332.41 | 1032.2 | 79.54 | 578.22 | 1219.44 | 2815.04 | 1417.62 |
| 38 | 16538.09 | 1264.13 | 754.15 | 2320.47 | 177.72 | 1496.27 | 2510.75 | 6015.02 | 3187.02 |
| 39 | 26418.08 | 2287.54 | 1129.15 | 3793.38 | 306.3 | 2361.98 | 4090.85 | 10057.67 | 5097.06 |
| 40 | 39081.59 | 3382.35 | 1646.6 | 6528.72 | 441.36 | 3893.51 | 6168.23 | 14151.28 | 8186.65 |
| 41 | 9128.82 | 731.66 | 479.67 | 1698.85 | 86.73 | 700.92 | 1490.06 | 3451.83 | 1936.87 |
| 42 | 19668.55 | 1565.64 | 1018.19 | 3634.61 | 202.98 | 1855.63 | 3037.61 | 7465.93 | 4262.57 |
| 43 | 31542.28 | 2816.73 | 1500.42 | 5631.67 | 354.05 | 2960.6 | 4805.21 | 12327.58 | 6743.42 |
| 44 | 45472.83 | 4119.46 | 2052.12 | 7894.24 | 507.46 | 4593.97 | 7220.14 | 16898.59 | 10021.53 |
| 45 | 10549.74 | 883.07 | 575.53 | 2061.11 | 115.43 | 849.75 | 1739.08 | 4257.59 | 2343.43 |
| 46 | 23421.12 | 1940.65 | 1222.28 | 4450.37 | 250.84 | 2278.06 | 3603.13 | 9370.57 | 5217.29 |
| 47 | 36953.15 | 3471.4 | 1818.15 | 7003.73 | 435.4 | 3632.96 | 5772.68 | 15468.28 | 8234.13 |
| 48 | 52673.59 | 5000.47 | 2515.29 | 10011.13 | 605.83 | 5701.84 | 8750.95 | 21026.68 | 12391.3 |
| 49 | 11510.8 | 1012.95 | 652.6 | 2523.87 | 132.74 | 1071.25 | 1991.61 | 4943.28 | 2755.72 |
| 50 | 26200.92 | 2152.47 | 1370.86 | 5307.19 | 287.21 | 2790.88 | 4165.74 | 10603.91 | 6116 |
| 51 | 40685.76 | 3804.28 | 2026.38 | 8158 | 498.72 | 4400.08 | 6700.3 | 17536.51 | 9568.34 |
| 52 | 57067.92 | 5650.2 | 2855.26 | 11459 | 695.58 | 6802.2 | 10309.8 | 23849.8 | 14451.18 |
| 53 | 12612.88 | 1065.38 | 734.62 | 2725.9 | 152.87 | 1231.86 | 2260.42 | 5436.95 | 3069.41 |
| 54 | 28465.92 | 2349.57 | 1516.69 | 5840.51 | 329.59 | 3249.85 | 4640.59 | 11655.05 | 6777.73 |
| 55 | 44471.53 | 4159.55 | 2224.57 | 8637.1 | 575.73 | 5110.66 | 7545.74 | 19138.94 | 10579.43 |
| 56 | 62164 | 6268 | 3146.5 | 12656.7 | 807.7 | 8006.8 | 11720.9 | 26260.8 | 16045.2 |
| 57 | 13636.91 | 1098.27 | 815.05 | 2982.71 | 172.33 | 1421.15 | 2454.73 | 5878.54 | 3339.16 |

表 2（续表）

| 时间 | 广东省 | 甘肃省 | 海南省 | 重庆市 | 西藏自治区 | 贵州省 | 云南省 | 四川省 | 陕西省 |
|----|----------|---------|---------|----------|---------|----------|----------|----------|----------|
| 58 | 30879.09 | 2491.45 | 1675.57 | 6440.51 | 369.34 | 3780.31 | 5010.33 | 12697.39 | 7422.65 |
| 59 | 48130.36 | 4430.89 | 2453.5 | 9500.83 | 661.56 | 6169.53 | 8212.42 | 20681.54 | 11567.42 |
| 60 | 67792.24 | 6835.27 | 3500.72 | 14265.4 | 920.83 | 9251.01 | 12814.59 | 28536.66 | 17689.94 |
| 61 | 14948.57 | 1129.68 | 840.05 | 3306.69 | 200.13 | 1659.39 | 2613.67 | 6234.51 | 3471.91 |
| 62 | 34526.64 | 2612.88 | 1808.57 | 7237.9 | 418.98 | 4351.07 | 5417.5 | 13300.06 | 7898.36 |
| 63 | 52522.38 | 4630.46 | 2643.53 | 11252.67 | 742.26 | 7195.58 | 8897.37 | 22120.37 | 12334.66 |
| 64 | 72812.55 | 6790.32 | 3702.76 | 15719.72 | 1026.39 | 10502.56 | 13717.88 | 30103.1 | 18171.86 |
| 65 | 17272.24 | 1235.97 | 944.57 | 3772.73 | 238.54 | 2102.97 | 2764.86 | 6703.7 | 3849.22 |
| 66 | 37357.59 | 2720.99 | 2008.95 | 8000.82 | 471.96 | 4936.61 | 5806.57 | 14222.33 | 8207.85 |
| 67 | 57061.17 | 4769.45 | 2880.88 | 12505.05 | 829.2 | 8135.38 | 9536.94 | 23793.6 | 12879.97 |
| 68 | 79512.05 | 7152.04 | 4044.51 | 17558.76 | 1150.07 | 11734.43 | 14869.95 | 32680.5 | 19165.39 |
| 69 | 19438.05 | 1388.75 | 1056.39 | 4306.74 | 271.35 | 2504.83 | 3115.66 | 7552 | 4147.22 |
| 70 | 41957.84 | 2993.02 | 2198.23 | 9143.64 | 555.41 | 5732.35 | 6447.81 | 16080.33 | 9581.1 |
| 71 | 64815.38 | 5641.46 | 3213.68 | 14309.18 | 939.51 | 9499.53 | 10646.65 | 27297.19 | 15202.11 |
| 72 | 89879.23 | 7677 | 4462.54 | 19500.27 | 1310.63 | 13540.83 | 16531.34 | 36980.22 | 21898.81 |
| 73 | 21705.29 | 1575.79 | 1185.25 | 4661.12 | 302.22 | 2904.79 | 3393.06 | 8590.21 | 4989.14 |
| 74 | 46341.93 | 3497.62 | 2434.88 | 9821.09 | 615.86 | 6632.86 | 7008.86 | 18327.01 | 10702.55 |
| 75 | 70635.22 | 6043.71 | 3546.92 | 14773.3 | 1044.03 | 10401.39 | 11619.76 | 30853.5 | 16867.92 |
| 76 | 97277.77 | 8246.07 | 4832.05 | 20363.19 | 1477.63 | 14806.45 | 17881.12 | 40678.13 | 24438.32 |

| 时间 | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|----|--------|---------|----------|---------|
| 1 | 10.46 | 14.65 | 70.36 | 99.72 |
| 2 | 20.47 | 73.48 | 131.25 | 388.9 |
| 3 | 56.92 | 249.91 | 344.75 | 445.58 |
| 4 | 263.68 | 295.02 | 1363.56 | 2080.04 |
| 5 | 18.85 | 20.31 | 80.03 | 109.55 |
| 6 | 62.9 | 96.59 | 149.38 | 111.33 |
| 7 | 123.17 | 226.19 | 650.94 | 139.7 |
| 8 | 300.13 | 337.44 | 1491.6 | 2279.34 |
| 9 | 38.16 | 25.97 | 84.3 | 184.62 |
| 10 | 105.33 | 129.7 | 330.01 | 411.56 |
| 11 | 189.42 | 222.47 | 957.13 | 724.98 |
| 12 | 340.65 | 377.16 | 1612.65 | 2523.73 |
| 13 | 57.47 | 28.37 | 176.63 | 379.79 |
| 14 | 147.76 | 57.19 | 510.64 | 811.79 |
| 15 | 255.67 | 121.25 | 1263.32 | 1310.26 |
| 16 | 390.2 | 445.36 | 1886.35 | 2821.11 |
| 17 | 76.78 | 62.71 | 268.96 | 574.96 |
| 18 | 190.19 | 134.08 | 691.27 | 1212.02 |
| 19 | 321.92 | 244.97 | 1569.51 | 1895.54 |

表 2（续表）

| 时间 | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|----|---------|---------|----------|----------|
| 20 | 466.1 | 537.11 | 2209.09 | 3433.5 |
| 21 | 96.09 | 97.05 | 361.29 | 770.13 |
| 22 | 232.62 | 210.97 | 871.9 | 1612.25 |
| 23 | 388.17 | 368.69 | 1875.7 | 2480.82 |
| 24 | 543.2 | 599.4 | 2609.6 | 4063.25 |
| 25 | 115.4 | 131.39 | 453.62 | 965.3 |
| 26 | 275.05 | 287.86 | 1052.53 | 2012.48 |
| 27 | 454.42 | 492.41 | 2181.89 | 3066.1 |
| 28 | 641.05 | 706.98 | 3018.98 | 4801.98 |
| 29 | 134.75 | 151.42 | 519.66 | 1160.37 |
| 30 | 319.97 | 336.57 | 1179.71 | 2386.26 |
| 31 | 532.98 | 574.63 | 2438.54 | 3717.97 |
| 32 | 760.96 | 834.16 | 3494.42 | 5885.88 |
| 33 | 165.03 | 186.58 | 679.76 | 1414 |
| 34 | 417 | 446.99 | 1590.85 | 3018.37 |
| 35 | 683.65 | 761.23 | 3125.07 | 4701.51 |
| 36 | 961.53 | 1098.51 | 4203.41 | 7171.58 |
| 37 | 180.69 | 204.53 | 547.96 | 1552.38 |
| 38 | 429.42 | 475.77 | 1321.45 | 3225.75 |
| 39 | 703.24 | 835.55 | 2824.93 | 5074.18 |
| 40 | 1081.27 | 1334.56 | 4273.58 | 7700.36 |
| 41 | 239.76 | 281.97 | 744.46 | 1845.81 |
| 42 | 566.65 | 655.99 | 1718.06 | 3833.21 |
| 43 | 929.3 | 1158.3 | 3582.76 | 6030.17 |
| 44 | 1350.43 | 1643.41 | 5418.81 | 9502.39 |
| 45 | 288.14 | 334.99 | 920.5 | 2247.86 |
| 46 | 689.09 | 811.53 | 2187.52 | 4720.1 |
| 47 | 1130.56 | 1442.14 | 4394.08 | 7438.04 |
| 48 | 1634.72 | 2060.79 | 6474.54 | 11714.35 |
| 49 | 322 | 385.86 | 1048.95 | 2546.72 |
| 50 | 797.59 | 930.52 | 2493.91 | 5329.54 |
| 51 | 1298.93 | 1605.77 | 4920.69 | 8296.01 |
| 52 | 1884.54 | 2326.64 | 7466.32 | 13031.04 |
| 53 | 356.81 | 415.69 | 1157.15 | 2779.39 |
| 54 | 888.61 | 1008.16 | 2771.06 | 5810.18 |
| 55 | 1432.46 | 1757.76 | 5451.35 | 9007.22 |
| 56 | 2101.1 | 2565.1 | 8360.2 | 14378 |
| 57 | 388.94 | 442.64 | 1262.25 | 3014.36 |
| 58 | 966.47 | 1071.94 | 3007.18 | 6327.87 |
| 59 | 1541.28 | 1880.27 | 5884.48 | 9792.2 |

表 2（续表）

| 时间 | 青海省 | 宁夏回族自治区 | 新疆维吾尔自治区 | 广西壮族自治区 |
|----|---------|---------|----------|----------|
| 60 | 2301.12 | 2752.1 | 9264.1 | 15672.97 |
| 61 | 425.78 | 460.64 | 1209.75 | 3226.02 |
| 62 | 1011.28 | 1128.03 | 3649.48 | 6769.45 |
| 63 | 1633.66 | 1987.32 | 6411.4 | 10430.75 |
| 64 | 2417.05 | 2911.77 | 9324.8 | 16803.12 |
| 65 | 456.51 | 508.06 | 1546.16 | 3480.32 |
| 66 | 1068.98 | 1203.72 | 3819.98 | 7311.64 |
| 67 | 1741.38 | 2120.31 | 6717.79 | 11345.53 |
| 68 | 2572.49 | 3150.06 | 9617.23 | 18245.07 |
| 69 | 524.01 | 588.12 | 1791.85 | 3909.53 |
| 70 | 1204.3 | 1397.61 | 4362.98 | 8179.81 |
| 71 | 1921.75 | 2449.57 | 7656.35 | 12757.93 |
| 72 | 2642.8 | 3453.93 | 10920.09 | 20396.25 |
| 73 | 532.58 | 673.71 | 2020 | 4271.31 |
| 74 | 1196.99 | 1603.58 | 4892.47 | 8762.63 |
| 75 | 1926.54 | 2784.66 | 8502.39 | 12863.07 |
| 76 | 2865.23 | 3705.18 | 12199.08 | 20352.51 |

注：数据来自《中国统计年鉴》，中国统计出版社

附录 2 MATLAB 源程序

```

function basisobj = create_bspline_basis(rangeval, nbasis, norder)
% 输入参数
% RANGEVAL ... 自变量取值的范围
% NBASIS ... 基函数的个数
% NORDER ... b样条基的阶数(order = degree + 1)
% 返回参数
% BASISOBJ ... 一个函数型数据的基对象
% BREAKS 是 NBASIS - NORDER + 2 个等间隔分割点向量
nbreaks = nbasis - norder + 2;
breaks = linspace(rangeval(1), rangeval(2), nbreaks);
% PARAMS 是 breaks 的内部点,删除了 breaks 的两个边界点
params = breaks(2:(nbreaks - 1));
% 构造基对象
type = 'bspline';
basisobj = basis(type, rangeval, nbasis, params);

function basisobj = basis(basistype, rangeval, nbasis, params)
% 输入参数
% BASISTYPE ... 说明所选用基类型的字符串,例如:'bspline'
% RANGEVAL ... 自变量取值的范围
% NBASIS ... 基函数的个数
% PARAMS ... breaks 的内部点,也即删除了 breaks 的两个边界点后的序列
% 返回参数
% BASISOBJ ... 一个函数型数据的基对象
basisobj.type = basistype;
basisobj.rangeval = rangeval;
basisobj.nbasis = nbasis;
basisobj.params = params;
basisobj = class(basisobj, 'basis'); %类变量 = class(结构变量, '类名')

function plot(basisobj)
% 绘制基对象图象
typex = basisobj.type;
nbasisx = basisobj.nbasis;
% 设置数值的精细网格
nx = max([10*nbasisx+1, 501]);
% 计算基的值
rangex = basisobj.rangeval;
x = linspace(rangex(1), rangex(2), nx)'; %转置,列向量

```

```

basismat = full(eval_basis(x, basisobj)); %把稀疏矩阵转换成全矩阵
% 绘制基值的图像
phdl = plot(x, basismat, '-');
set(phdl, 'LineWidth', 1); %set(句柄, 属性名 1, 属性值 1, 属性名 2, 属性值 2, ...)
% 设置绘图范围
minval = 0;
maxval = 1;
knots = getbasispar(basisobj);
hold on
for k=1:length(knots)
    lhdl = plot([knots(k), knots(k)], [0,1]);
    set(lhdl, 'LineWidth', 1, 'LineStyle', ':', 'color', 'r');
end
hold off
xlabel('\fontsize{13} t')
ylabel('\fontsize{13} \phi(t)')
titstr = ['\fontsize{16} ', typeex, ' basis', ', no. basis fns = ',
        num2str(nbasisx)];

norderx = nbasisx - length(knots);
titstr = [titstr, ', order = ', num2str(norderx)];
title(titstr);
axis([rangex(1), rangex(2), minval, maxval])

function [evalarray, basisobj] = eval_basis(evalarg, basisobj, Lfdobj)
% 输入参数
% EVALARG ... 计算基函数在这些向量元素处的值
% BASISOBJ ... 基对象
% LFD OBJ ... 一个线性微分算子
% 返回参数
% EVALARRAY... 函数在 evalarg 参数向量处所对应的值
% 设立默认的 LFD OBJ 值为 0
if nargin < 3
    Lfdobj = int2Lfd(0);
end
% 检查 LFD OBJ
Lfdobj = int2Lfd(Lfdobj);
% 定义所需导数的最高阶
nderiv = getnderiv(Lfdobj);
% 获取权重系数函数
wfdcell = getwfdcell(Lfdobj);
% 调用 getbasismatrix 函数来计算基
if nargin == 2
    [evalarray, basisobj] = getbasismatrix(evalarg, basisobj, nderiv);

```

```

else
    evalarray = getbasismatrix(evalarg, basisobj, nderiv);
end
nbasis      = size(evalarray,2);
onerow      = ones(1,nbasis);
if nderiv > 0 && ~isinteger(Lfdobj)
    for j = 1:nderiv
        if ~isempty(wfdcell{j})
            wfdParj = wfdcell{j};
            wfdj     = getfd(wfdParj);
            wcoef    = getcoef(wfdj);
            if ~all(all(wcoef == 0.0))
                wjvec = eval_fd(evalarg, wfdj);
                temp  = getbasismatrix(evalarg, basisobj, j-1);
                evalarray = evalarray + (wjvec*onerow).*temp;
            end
        end
    end
end
end

function fdobj = fd(coef, basisobj, fdnames)
defaultfdnames{1} = 'arguments';
defaultfdnames{2} = 'replications';
defaultfdnames{3} = 'functions';
if nargin == 0
    fdobj.coef      = zeros(1,1);
    fdobj.basisobj  = create_constant_basis([0,1]);
    fdobj.fdnames   = defaultfdnames;
    fdobj = class(fdobj, 'fd');
    return;
end
if nargin < 3, fdnames = defaultfdnames; end
if nargin < 2
    error('Less than two arguments.');
```

```

        error('Coefficient matrix is empty.');
```

end

```

if ~strcmp(class(basisobj), 'basis')
    error('Argument BASISOBJ is not of basis class.');
```

end

```

nbasis = getnbasis(basisobj);
if ~isempty(coef) && ~strcmp(getbasistype(basisobj), 'fdVar')
    if coefd(1) ~= nbasis
        error(['First dimension of coefficient array is ', ...
            'not equal to number of basis functions.']);
    end
end
end
if ~iscell(fdnames)
    error('FDNAMES not a cell object.');
```

end

```

if length(fdnames) > 3
    error('FDNAMES has length greater than 3.');
```

end

```

if length(fdnames) == 2
    fdnames{3} = 'functions';
end
if length(fdnames) == 1
    fdnames{2} = 'replications';
    fdnames{3} = 'functions';
end
fdobj.coef      = coef;
fdobj.basisobj  = basisobj;
fdobj.fdnames   = fdnames;
fdobj = class(fdobj, 'fd');
```

function basisobj = create_monomial_basis(rangeval, nbasis, ...
exponents, argtrans, dropind)

```

if nargin < 1, rangeval = [0,1]; end
if length(rangeval) == 1
    if rangeval <= 0
        error('RANGEVAL a single value that is not positive.');
```

end

```

    rangeval = [0,rangeval];
end
if rangechk(rangeval) ~= 1
    error('RANGEVAL is not a legitimate range.');
```

end

```

if nargin < 2, nbasis = 2; end
if nargin < 3 || isempty(exponents)
    exponents = 0:(nbasis-1);
```



```

end
if nargin < 4 || isempty(argtrans)
    argtrans = [0,1];
end
if nargin < 5, dropind = []; end
for ibasis=1:nbasis
    if exponents(ibasis) - round(exponents(ibasis)) ~= 0
        error('An exponent is not an integer.');
```

end

```

    if exponents(ibasis) < 0
        error('An exponent is negative.');
```

end

```

end
if min(diff(sort(exponents))) <= 0
    error('There are duplicate exponents.');
```

end

```

if length(dropind) > 0
    if length(dropind) >= nbasis
        error('Too many index values in DROPIND.');
```

end

```

    dropind = sort(dropind);
    if length(dropind) > 1
        if any(diff(dropind)) == 0
            error('Multiple index values in DROPIND.');
```

end

```

    end
    for i=1:length(dropind);
        if dropind(i) < 1 || dropind(i) > nbasis
            error('An index value is out of range.');
```

end

```

    end
end
parStruct.exponents = exponents;
parStruct.argtrans = argtrans;
type = 'monom';
params = parStruct;
quadvals = [];
values = {};
basisvalues = {};
basisobj = basis(type, rangeval, nbasis, params, ...
    dropind, quadvals, values, basisvalues);

function basisobj = create_fourier_basis(rangeval, nbasis, period, ...
    dropind)
if nargin < 1, rangeval = [0,1]; end

```

```

if length(rangeval) == 1
    if rangeval <= 0
        error('RANGEVAL a single value that is not positive.');
```

end

```

    rangeval = [0,rangeval];
end
if rangechk(rangeval) ~= 1
    error('RANGEVAL is not a legitimate range.');
```

end

```

if nargin < 2, nbasis = 3; end
if nargin < 4, dropind = []; end
width = rangeval(2) - rangeval(1);
if nargin < 3
    period = width;
end
if (period <= 0)
    error ('Period must be positive for a Fourier basis');
```

end

```

if (2*floor(nbasis/2) == nbasis)
    nbasis = nbasis + 1;
end
if length(dropind) > 0
    if length(dropind) >= nbasis
        error('Too many index values in DROPIND.');
```

end

```

    dropind = sort(dropind);
    if length(dropind) > 1
        if any(diff(dropind)) == 0
            error('Multiple index values in DROPIND.');
```

end

```

    end
    for i=1:length(dropind);
        if dropind(i) < 1 || dropind(i) > nbasis
            error('An index value is out of range.');
```

end

```

    end
end
type      = 'fourier';
params    = period;
quadvals  = [];
values    = {};
basisvalues = {};
basisobj = basis(type, rangeval, nbasis, params, ...
    dropind, quadvals, values, basisvalues);
```

```

function pcastr = pca_fd(fdobj, nharm, harmfdPar, centerfns)
if ~isa_fd(fdobj)
    error ('First argument is not a functional data object.');
```

```

end
fdbasis = getbasis(fdobj);
if nargin < 4
    centerfns = 1;
end
if nargin < 3
    Lfdobj = int2Lfd(2);
    lambda = 0;
    harmfdPar = fdPar(fdbasis, Lfdobj, lambda);
else
    if ~isa_fdPar(harmfdPar)
        if isa_fd(harmfdPar) || isa_basis(harmfdPar)
            harmfdPar = fdPar(harmfdPar);
        end
    end
end
if nargin < 2
    nharm = 2;
end
meanfd = mean(fdobj);
if nharm == 0
    pcastr.harmfd = [];
    pcastr.values = [];
    pcastr.harmscr = [];
    pcastr.varprop = [];
    pcastr.fdhatfd = [];
    pcastr.meanfd = meanfd;
    return
end
if centerfns ~= 0
    fdobj = center(fdobj);
end
harmbasis = getbasis(getfd(harmfdPar));
nhbasis = getnbasis(harmbasis);
Lfdobj = getLfd(harmfdPar);
Lfdobj = int2Lfd(Lfdobj);
lambda = getlambda(harmfdPar);
coef = getcoef(fdobj);
coefd = size(coef);
nbasis = coefd(1);
nrep = coefd(2);
ndim = length(coefd);

```

```

if nrep < 2
    error('PCA not possible without replications');
end
if ndim == 3
    nvar = coefd(3);
    ctemp = zeros(nvar*nbasis,nrep);
    for j = 1:nvar
        index = (1:nbasis) + (j-1)*nbasis;
        ctemp(index,:) = coef(:, :, j);
    end
else
    nvar = 1;
    ctemp = coef;
end
Lmat = eval_penalty(harmbasis, int2Lfd(0));
if lambda > 0
    Rmat = eval_penalty(harmbasis, Lfdobj);
    Lmat = Lmat + lambda .* Rmat;
end
Lmat = (Lmat + Lmat')/2;
Mmat = chol(Lmat);
Mmatinv = inv(Mmat);
Wmat = ctemp*ctemp'./nrep;
Jmat = inprod_basis(harmbasis, fdbasis);
MIJW = Mmatinv'*Jmat;
if nvar == 1
    Cmat = MIJW*Wmat*MIJW';
else
    Cmat = zeros(nvar*nhbasis);
    for i = 1:nvar
        indexi = (1:nbasis) + (i-1)*nbasis;
        for j = 1:nvar
            indexj = (1:nbasis) + (j-1)*nbasis;
            Cmat(indexi,indexj) = MIJW*Wmat(indexi,indexj)*MIJW';
        end
    end
end
Cmat = (Cmat + Cmat')./2;
[eigvecs, eigvals] = eig(Cmat);
[eigvals, indsrt] = sort(diag(eigvals));
eigvecs = eigvecs(:,indsrt);
neig = nvar*nhbasis;
indx = neig + 1 - (1:nharm);
eigvals = eigvals(neig + 1 - (1:neig));
eigvecs = eigvecs(:,indx);

```

```

sumvecs = sum(eigvecs);
eigvecs(:,sumvecs < 0) = -eigvecs(:,sumvecs < 0);
varprop = eigvals(1:nharm)./sum(eigvals);
harmnames = getnames(fdobj);
harmlabels = ['I  '; 'II '; 'III'; 'IV  '; 'V  '; ...
              'VI  '; 'VII '; 'VIII'; 'IX  '; 'X   '];
if nharm <= 10
    harmnames2 = cell(1,2);
    harmnames2{1} = 'Harmonics';
    harmnames2{2} = harmlabels(1:nharm,:);
    harmnames{2} = harmnames2;
else
    harmnames{2} = 'Harmonics';
end
if iscell(harmnames{3})
    harmnames3 = harmnames{3};
    harmnames3{1} = ['Harmonics for ',harmnames3{1}];
    harmnames{3} = harmnames3;
else
    if ischar(harmnames{3}) && size(harmnames{3},1) == 1
        harmnames{3} = ['Harmonics for ',harmnames{3}];
    else
        harmnames{3} = 'Harmonics';
    end
end
if nvar == 1
    harmcoef = Mmat\eigvecs;
else
    harmcoef = zeros(nbasis,nharm,nvar);
    for j = 1:nvar
        index = (1:nbasis) + (j-1)*nbasis;
        eigvecs_j = eigvecs(index,:);
        harmcoef(:,j) = Mmat\eigvecs_j;
    end
end
harmfd = fd(harmcoef, harmbasis, harmnames);
if nvar == 1
    harmscr = inprod(fdobj, harmfd);
else
    harmscr = zeros(nrep, nharm, nvar);
    coefarray = getcoef(fdobj);
    harmcoefarray = getcoef(harmfd);
    for j=1:nvar
        coef_j = squeeze(coefarray(:,j));
        harmcoef_j = squeeze(harmcoefarray(:,j));
    end
end

```

```

        fdobjj    = fd(coefj, fdbasis);
        harmfdj   = fd(harmcoefj, fdbasis);
        harmscr(:, :, j) = inprod(fdobjj, harmfdj);
    end
end
if nvar == 1
    fdhatcoef = harmcoef*harmscr';
else
    fdhatcoef = zeros(nbasis, nrep, nvar);
    for j=1:nvar
        fdhatcoef(:, :, j) = harmcoef(:, :, j)*harmscr(:, :, j)';
    end
end
fdhatfd = fd(fdhatcoef, harmbasis);
pcastr.harmfd = harmfd;
pcastr.values = eigvals;
pcastr.harmscr = harmscr;
pcastr.varprop = varprop;
pcastr.fdhatfd = fdhatfd;
pcastr.meanfd = meanfd;

function cl = cluster(amat, normalize)
if nargin < 2, normalize = 0; end
amatd = size(amat);
n = amatd(1);
k = amatd(2);
rotm = eye(k);
onek = ones(1, k);
if k == 1
    return
end
if normalize
    hvec = sum(amat' .^2)';
    amat = amat./ (sqrt(hvec)*onek);
end

function derivfd = deriv_fd(fdobj, Lfdobj)
if nargin < 2
    Lfdobj = int2Lfd(1);
else
    Lfdobj = int2Lfd(Lfdobj);
end
basisobj = getbasis(fdobj);
nbasis = getnbasis(basisobj);
rangeval = getbasisrange(basisobj);

```

```

nfine    = max([201, 10*nbasis+1]);
evalarg  = linspace(rangeval(1), rangeval(2), nfine)';
Lfdmat   = eval_fd(evalarg, fdobj, Lfdobj);
Lfdcoef  = project_basis(Lfdmat, evalarg, basisobj);
Dfdnames = getnames(fdobj);
if iscell(Dfdnames{3})
    Dfdnames{3}{1} = ['L-', Dfdnames{3}{1}];
else
    if ischar(Dfdnames{3}) && size(Dfdnames{3},1) == 1
        Dfdnames{3} = ['L-', Dfdnames{3}];
    else
        Dfdnames{3} = 'L-function';
    end
end
derivfd = fd(Lfdcoef, basisobj, Dfdnames);

```

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《函数型数据分析方法在 GDP 研究中的应用》是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：刘宝宇 日期：2020 年 6 月 6 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：刘宝宇 日期：2020 年 6 月 6 日

导师签名：方茹 日期：2020 年 6 月 6 日

致 谢

时光飞逝，岁月如梭，两年硕士生活即将结束。在论文完成之际首先感谢我的导师方茹教授对我的精心指导，在此向方茹老师表示无限的敬意与感激。

感谢各位评审老师，从开题答辩到终期答辩，各位老师一直在耐心地指出我论文中的格式规范问题以及方法内容等的不足，才使得我不断改正错误，顺利完成论文。

感谢硕士期间授予我知识的各位老师，每位老师有不同的讲课风格，但却是同样的认真负责。

感谢硕士期间认识的所有同学，你们每个人身上都有值得我学习的地方，是我学习的榜样。

感谢我的父母，是你们给予了我生命和最无私的爱。你们对于我的教育让我受用终生。

感谢母校哈尔滨工业大学给我提供的良好学习环境，母校规格严格、功夫到家的校训将会一直影响着我。

谨以此文献给我的学生时光，毕业快乐。