

Empirical Research on Local Fiscal Expenditure Structure in China Based on Factor Analysis and Cluster Analysis

Mengdi Zhai

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 540532394@qq.com

Received: Aug. 10th, 2016; accepted: Aug. 24th, 2016; published: Aug. 31st, 2016

Copyright © 2016 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, based on the fiscal expenditure of 31 provinces in China in 2014, the methods of factor analysis and cluster analysis were used to compare the expenditure structure of our country.

Keywords

Local Fiscal Expenditure, Factor Analysis, Cluster Analysis

基于因子分析与聚类分析的中国地方财政支出结构的实证研究

翟梦迪

云南财经大学, 云南 昆明
Email: 540532394@qq.com

收稿日期: 2016年8月10日; 录用日期: 2016年8月24日; 发布日期: 2016年8月31日

摘要

本文以2014年我国31个省份的财政支出等为基础, 利用因子分析与聚类分析的方法对我国各地支出结

构进行比较分析。

关键词

地方财政支出, 因子分析, 聚类分析

1. 引言

当下时代, 财政支出的结构从一定程度上反应了社会的经济情况, 也在一定程度上影响社会经济的整体水平, 所以对财政支出结构进行分析具有现实意义。

国内很多学者也对此进行了研究, 李建强分析了财政支出结构对居民消费的影响, 认为财政支出结构对居民消费具有挤入效应[1], 傅勇等认为中国的财政分权对财政支出结构产生了扭曲作用, 直接导致文化教育等支出不足[2], 张建迎将财政支出分为经济服务、社会服务、政府服务和其他职能, 以聚类分析法对 2003 年的 31 个省份的财政支出结构进行了聚类分析, 其结果将财政支出结构分为 5 种类型[3]。江克忠的研究表明财政分权是行政管理类支出不断增长的一个重要诱因[4]。张明喜以地方财政支出结构和经济发展水平为主要指标, 采用聚类分析法将全国划分为四大经济区域类型[5]。但是, 这些方法只能对我国各省(自治区、直辖市)进行笼统的分类, 无法探究每一类下具体省(自治区、直辖市)的财政支出结果特点。

综上, 为了解我国财各省(自治区、直辖市)财政支出结构上的异同, 本文基于我国 2014 年 31 个省(自治区、直辖市)财政支出的相关数据, 采用因子分析与聚类分析法对此进行了研究。首先, 利用因子分析寻找出公共因子, 再利用已有的公共因子作为特征对数据进行聚类, 进而将我国 31 个省(自治区、直辖市)归类, 并分别讨论, 进而对我国各个地区的财政支出结构进行分析。

2. 理论基础

2.1. 因子分析的基本思想

因子分析的基本思想是: 根据相关性大小把变量分组, 使得组内的变量相关性较高, 但不同组的变量相关性较低, 则每组变量可以代表一个基本结构, 称为因子, 它反映已经观测到的相关性。

2.2. 正交因子模型

设 p 维随机向量 $X = (x_1, x_2, \dots, x_p)^T$ 的期望为 $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$, 协方差矩阵为 Σ , 假定 X 线性地依赖于少数几个不可观测的随机变量 f_1, f_2, \dots, f_m ($m < p$) 和 p 个附加的方差源 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, 一般称 f_1, f_2, \dots, f_m 为公因子, 称 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ 为特殊因子或误差。那么, 因子模型为:

$$\begin{aligned} x_1 &= \mu_1 + a_{11}f_1 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 &= \mu_2 + a_{21}f_1 + \dots + a_{2m}f_m + \varepsilon_2 \\ &\vdots \\ x_p &= \mu_p + a_{p1}f_1 + \dots + a_{pm}f_m + \varepsilon_p \end{aligned}$$

引入矩阵符号, 记

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix}, \quad F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

那么因子模型可以写为:

$$X = \mu + AF + \varepsilon$$

式中, a_{ij} 称为第 i 个变量在第 j 个因子上的载荷, 矩阵 A 称为载荷矩阵。

2.3. 关于因子模型的假定

我们假定:

$$\begin{aligned} E(F) &= 0, \quad Cov(F) = I \\ E(\varepsilon) &= 0, \quad Cov(\varepsilon) = \Phi = \text{diag}(\phi_1, \phi_2, \dots, \phi_p) \\ Cov(F, \varepsilon) &= 0 \end{aligned}$$

如果模型满足假定, 则称该模型为正交因子模型。

2.4. 聚类分析

所谓聚类分析, 就是将研究对象(若干个个体的集合)按照某种标准分成若干类。

Ward 法聚类, 又称离差平方和法, 基于方差分析思想构建的分类方法, 如果分类正确, 同类样本之间的离差平方和应该较小, 类与类的离差平方和应该较大[6]。

3. 实证分析

为减小我国各个省(自治区、直辖市)在地理环境、人口环境和资源环境上的差异对结果所造成的影响, 在这里利用人均财政支出作为各省(自治区、直辖市)财政支出的直观指标, 即用各个省(自治区、直辖市)的总财政支出除以当年该省(自治区、直辖市)人口总数, 本文所用数据来自于《中国统计年鉴 2015》。

2007 年我国实施财政收入分类改革后, 将财政支出分为一般公共服务、医疗卫生、国防、教育、环境保护等 21 个不同类别, 但是, 通过数据分析发现, 国防和金融支出项目在总支出中所占比例不大, 而外交、援助其他地区和国债还本付息住处在一些省份并没有发生。故本文的分析基于剔除上述五个类别 16 个其他财政支出类别。即一般公共服务支出、公共安全支出、教育支出、科学技术支出、文化体育与传媒支出、社会保障和就业支出、医疗卫生支出、环境保护支出、城乡社区事务支出、农林水事务支出、交通运输支出、资源勘探电力信息等事务支出、商业服务业等事务支出、国土资源气象等事务支出、住房保障支出、粮油物资储备支出。

可以发现的是, 各省的财政支出之间是有相互依存、相互依赖的关系的, 但是若仅从财政支出入手很难发现其中相似之处, 所以本文接下来利用因子分析提取公共因子, 进而对其实现降维处理。

第一步, 采用 R 软件对样本数据进行因子分析, 首先计算样本数据的相关系数矩阵, 观察各变量之间的相关性。其中 $x_1 \sim x_{16}$ 分别代表保留下来的 16 个财政支出类别, 即一般公共服务支出、公共安全支出、教育支出、科学技术支出、文化体育与传媒支出、社会保障和就业支出、医疗卫生支出、环境保护支出、城乡社区事务支出、农林水事务支出、交通运输支出、资源勘探电力信息等事务支出、商业服务业等事务支出、国土资源气象等事务支出、住房保障支出、粮油物资储备支出。利用 R 软件分析, 输出结果如下表 1。

3.1. 数据整理

根据表 1 中的调查数据, 计算得出表 2 中从样本数据各变量的相关系数上可以看出, 许多变量之间存在较强的相关性, 可以考虑提取公因子。为了消除各变量之间的相关性, 下面分别采用 R 软件中主成分法对数据进行因子分析, 提取因子。结果如下表 2、表 3。

Table 1. The output table
表 1. 输出结果表

	x1	x2	x3	x4	x5	x6	x7	x8
x1	1	0.8668	0.74979	0.01264	0.8442	0.579	0.68771	0.6169
x2	0.8668	1	0.94372	0.43439	0.9445	0.7255	0.88144	0.7708
x3	0.74979	0.94372	1	0.57292	0.9032	0.6915	0.91052	0.7414
x4	0.01264	0.43439	0.57292	1	0.4172	0.3715	0.52485	0.3741
x5	0.8442	0.94452	0.90319	0.41722	1	0.7791	0.91694	0.8802
x6	0.57904	0.72554	0.6915	0.37145	0.7791	1	0.78281	0.8269
x7	0.68771	0.88144	0.91052	0.52485	0.9169	0.7828	1	0.8783
x8	0.61693	0.77081	0.74138	0.37407	0.8802	0.8269	0.87828	1
x9	0.23132	0.53671	0.68815	0.66518	0.4611	0.6021	0.54929	0.4439
x10	0.97581	0.80751	0.66881	-0.07394	0.7875	0.5205	0.60237	0.5584
x11	0.9202	0.79586	0.69603	-0.05418	0.835	0.6428	0.73888	0.7334
x12	0.61623	0.77835	0.82063	0.57359	0.7048	0.6576	0.72663	0.5357
x13	0.56688	0.76924	0.75569	0.50708	0.6988	0.6272	0.73685	0.6024
x14	0.18741	0.18814	0.15009	-0.09625	0.2627	0.3349	0.1905	0.3029
x15	0.84298	0.72739	0.59807	-0.13839	0.7271	0.6043	0.64816	0.6344
x16	0.07217	0.06732	-0.02736	-0.16149	0.1171	0.3378	-0.01128	0.2518
	x9	x10	x11	x12	x13	x14	x15	x16
x1	0.23132	0.97581	0.9202	0.61623	0.566878	0.187412	0.843	0.07217
x2	0.53671	0.80751	0.79586	0.77835	0.769236	0.188142	0.7274	0.06732
x3	0.68815	0.66881	0.69603	0.82063	0.755687	0.150092	0.5981	-0.02736
x4	0.66518	-0.07394	-0.05418	0.57359	0.50708	-0.096254	-0.1384	-0.16149
x5	0.46109	0.78748	0.83503	0.70481	0.698839	0.26271	0.7271	0.1171
x6	0.60208	0.52046	0.64285	0.65759	0.627197	0.334851	0.6043	0.33781
x7	0.54929	0.60237	0.73888	0.72663	0.736852	0.190504	0.6482	-0.01128
x8	0.44392	0.55838	0.73335	0.53574	0.602387	0.302862	0.6344	0.25183
x9	1	0.16719	0.17837	0.70862	0.587705	0.074149	0.1073	0.09687
x10	0.16719	1	0.88981	0.56333	0.531115	0.208703	0.844	0.09517
x11	0.17837	0.88981	1	0.59239	0.590154	0.258509	0.8852	0.12304
x12	0.70862	0.56333	0.59239	1	0.839472	0.040817	0.4921	0.01669
x13	0.58771	0.53111	0.59015	0.83947	1	0.007491	0.6012	-0.01753
x14	0.07415	0.2087	0.25851	0.04082	0.007491	1	0.2145	0.18365
x15	0.10731	0.84401	0.8852	0.49214	0.601233	0.214513	1	0.17856
x16	0.09687	0.09517	0.12304	0.01669	-0.017534	0.183647	0.1786	1

Table 2. The cumulative contribution rate**表 2.** 累计贡献率

	Vars	Vars. Prop	Vars. Cum
Factor 1	9.602	0.60012	60.01
Factor 2	2.377	0.14859	74.87
Factor 3	1.354	0.08461	83.33

Table 3. The factor score**表 3.** 因子得分

	Factor 1	Factor 2	Factor 3
x1	0.8571	-0.39922	-0.219609
x2	0.9646	0.019039	-0.093708
x3	0.9312	0.228107	-0.100147
x4	0.4045	0.851919	0.006284
x5	0.9632	-0.040565	0.019284
x6	0.8278	0.075792	0.397028
x7	0.9244	0.155666	-0.00446
x8	0.8503	-0.007246	0.284207
x9	0.5709	0.644204	0.209402
x10	0.8022	-0.473682	-0.216644
x11	0.8572	-0.445561	-0.100605
x12	0.8213	0.334383	-0.122725
x13	0.8078	0.265772	-0.163552
x14	0.2404	-0.275683	0.578293
x15	0.7818	-0.50407	-0.093133
x16	0.1224	-0.240502	0.749733

3.2. 总体总值

从上述主成分法得出的因子分析结果可以看出，主成分法的累计贡献率达到了 83.33%，涵盖了大部分的信息量。而且各个变量在因子上的得分也不尽相同，接下来为了更好地解释因子的含义，我们基于主成分法采用方差最大化作因子正交旋转。R 输出结果如下表 4、表 5。

从上述因子正交旋转的结果可以看出，方差累计贡献率达到了 83.33%，三个因子涵盖了绝大部分信息。表中列示了旋转后的因子载荷矩阵，因子 1 在一般公共服务、医疗卫生、农林水事务、公共安全、交通运输、文化体育与传媒变量上具有较高的载荷，而这些变量大多属于基础服务类支出，因而可以命名为“基础服务类因子”；因子 2 在教育、科技、城乡社区事务、商业服务业等事务等方面载荷很高，这些变量对社会更好发展有十分重要的作用，所以将这一类统一命名为“高级服务类因子”；因子 3 在变量住房保障支出、社保和就业、环保、国土气象和粮油物资事务载荷量很高，这些变量是为了保障社会底层公民的生活质量，所以可以命名为“保障因子”。

Table 4. The cumulative contribution rate
表 4. 累计贡献率

	Vars	Vars. Prop	Vars. Cum
Factor 1	6.707	41.92	41.92
Factor 2	4.923	30.77	72.69
Factor 3	1.703	10.64	83.33

Table 5. The factor score
表 5. 因子得分

	Factor 1	Factor 2	Factor 3
x1	0.95443	0.16764	0.05683
x2	0.77333	0.57862	0.08213
x3	0.63586	0.72435	0.01332
x4	-0.14698	0.91924	-0.15108
x5	0.76995	0.54318	0.20482
x6	0.48524	0.49794	0.60566
x7	0.67349	0.64047	0.12214
x8	0.48239	0.53248	0.62562
x9	0.03361	0.87602	0.12751
x10	0.95091	0.07607	0.0698
x11	0.94318	0.14431	0.18189
x12	0.49918	0.74094	-0.05684
x13	0.53834	0.67373	-0.0793
x14	0.16043	-0.01079	0.6651
x15	0.19104	0.0537	0.91354
x16	-0.00352	-0.03364	0.7961

采用 R 软件运用回归估计法计算出 2014 年各省(自治区、直辖市)在“基础服务类因子”、“高级服务类因子”、“保障因子”上的得分,三个公因子分别从不同侧面反映了不同省(自治区,直辖市)在财政支出结构上的差异, R 软件输出结果如下表 6。

从上述因子得分可以看出:

1) 在基础服务类因子 F1 上得分最高的四个省(自治区、直辖市)依次是西藏、宁夏、新疆、贵州,这四个省的得分远高于其他省,可见,这四地的财政支出偏重于基础服务,即保障公民日常生活水平。

2) 北京、天津、上海三个省(自治区、直辖市)在因子 F2 上的得分较高,说明在财政支出中,这三个省(自治区、直辖市)更偏重于高级服务支出,即在教育、科技和城乡社区事务的财政支出较多。这并不难理解,北京和上海作为我国经济发展的代表城市,其经济水平远超国内平均水平,基础设施完善,更着眼于社会人幸福感的提高,而天津由于其地理位置邻近北京,受到的联动效应较大,所以在此因子上得分也比较高。

3) 在保障因子 F3 上,内蒙古、山西和青海得分较高,说明这三个省(自治区、直辖市)在政府财政支

Table 6. The factor score of each province
表 6. 各省因子得分

	Factor 1	Factor 2	Factor 3
北京	-0.80088	2.91542	0.14221
天津	-0.85786	2.58556	-0.27338
河北	-0.39414	-0.80253	-0.43757
山西	-0.62127	-0.69904	2.09698
内蒙古	-0.36571	0.17935	3.4962
辽宁	-0.43081	0.19392	0.46685
吉林	-0.30958	-0.27833	1.44342
黑龙江	-0.44801	-0.53681	1.26978
上海	-0.59914	2.82714	-0.83917
江苏	-0.5879	0.26809	-0.76345
浙江	-0.53426	0.07575	-1.15962
安徽	-0.40375	-0.44214	-0.33454
福建	-0.93913	-0.25792	-0.89082
江西	-0.13108	-0.55101	-0.62833
山东	-0.38353	-0.48726	-0.64621
河南	-0.36393	-0.85246	-0.55708
湖南	-0.31324	-0.43961	-0.22607
湖北	-0.28444	-0.70413	-0.33251
广东	-0.1606	-0.3341	-0.91128
广西	-0.26745	-0.59961	-0.61723
海南	0.1967	-0.18259	-0.22554
重庆	-0.31826	0.1481	-0.0553
四川	-0.2104	-0.58715	-0.32546
贵州	0.41254	-0.80612	-0.85017
云南	-0.34988	-0.4648	-0.07106
西藏	4.91212	-0.3706	-0.03367
陕西	0.0652	-0.40384	-0.07966
甘肃	0.09519	-0.71761	-0.08434
青海	1.33341	-0.62761	1.489
宁夏	0.52146	0.22038	-0.14214
新疆	0.4386	-0.16486	0.08015

出中更偏重于保障支出，即偏重于住房保障，社保就业等等方面的财政支出，证明这些省(自治区、直辖市)还没有很好的社会物质基础，仍在积累阶段，并且大部分财政支出均用于保障公民基本生活。

为将有相似特征的各个省整合起来，进而得到我国各地地方财政支出结构的具体特征，接下来利用通过因子分析得到的三个因子作为自变量，聚类方法选择 Ward 法和 k 均值法，根据聚类结果，可以将

Table 9. The category of the fiscal expenditure structure in each province
表 9. 各省财政支出结构所属类别

类型	地区
类型 1	北京、上海、天津、浙江
第一小类	广东、江苏、福建、山东、辽宁
类型 2	第二小类 重庆、山西、河北、海南、安徽、陕西、江西、河南、湖北、湖南、广西、云南、四川、吉林、黑龙江、贵州、甘肃
	第三小类 内蒙古、新疆、宁夏、青海
类型 3	西藏

由表可以看出,由北京、上海、天津、浙江四省(直辖市)构成的类型 1 集中了我国目前发展水平较好的地区,这些地区公民生活质量普遍比较高,并且在高级服务类的因子上的分较高,可以很好的归为一类,说明这些地方注重教育、科技等事业的投入,社会经济基础较为扎实。类型 2 中的 17 省(自治区、直辖市)在各个因子上表现较为均衡,没有明显的薄弱方面,代表的大多省(自治区、直辖市)的发展情况,也可以反映我国目前的整体情况,即在一般公共服务和医疗卫生等方面投入较多。类型 3 中西藏属于我国需要重点财政资助的省份,在社保和住房保障等方面支出较多,社会经济基础薄弱,各方面均无法达到国内平均水平,也无法很好的与其他省市归为一类。

4. 结论

1) 本文对我国各省地方财政支出结构进行了分析,并且利用因子分析和聚类分析的方法进行了分类。研究发现剔除在各省表现不突出的五类财政支出之后,利用因子分析可以很好的将财政支出信息用“基础服务类支出”、“高级服务类支出”、“保障支出”三个因子表现出来。这表明,我国各省(自治区、直辖市)的财政支出结构的相似之处便是均用于“基础服务类支出”、“高级服务类支出”和“保障支出”三个方向。

2) 通过进行聚类可以发现,各省的财政支出结构之间存在着一定的依赖性,并可以归纳成为三类地区,即生活水平高等级地区,生活水平均衡地区和需要重点关注地区。北京、上海、天津、浙江四省财政支出结构的特点为在高级服务类支出上比重较大;西藏财政支出结构的特点是主要关注保障服务类支出;而其余省(自治区、直辖市)便是财政重点支出基础社会服务。所以,利用聚类分析的方法,清楚的将我国 31 个省(自治区、直辖市)按照财政支出节结构的特点分为三大类。

3) 通过分析可知,尤其是通过类别 2 可知,多数省份之间的财政支出结构并没有存在很大的差异,只有在利用转移支付所占比例进行再次划分时,细微差异才得以体现。这可以归因于中央政府的积极调控,加大转移支付力度,使得各省(自治区、直辖市)财政的平稳有序运行。尽管中央的转移支付制度还存在局限,但这已成为保障我国公民生活水平的有力办法。

参考文献 (References)

- [1] 李建强. 我国财政支出结构与居民消费异质性动态关系[J]. 山西财经大学学报, 2012(1): 9-21.
- [2] 傅勇, 张晏. 中国式分权与财政支出结构偏向: 为增长而竞争的代价[J]. 管理世界, 2007(3): 4-12, 22.
- [3] 张建迎. 中国地区财政支出结构的聚类分析[J]. 科技信息, 2006(3): 168-169.
- [4] 江克忠. 财政分权与地方政府行政管理支出——基于中国省级面板数据的实证研究[J]. 公共管理学报, 2011, 8(3): 44-52.

-
- [5] 张明喜. 地方财政支出结构与地方经济发展的实证研究——基于聚类分析的新视角[J]. 财经问题研究, 2008(1): 80-86.
- [6] 费宇, 陈贻娟. 多元统计分析[M]. 北京: 中国人民大学出版社, 2014.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>