

天津市工业主要行业发展的比较研究<sup>①</sup>李亚力<sup>1</sup>, 刘嘉焜<sup>1</sup>, 李虹霖<sup>1</sup>, 杨长俊<sup>2</sup>, 于得水<sup>2</sup>

(1. 天津大学, 天津 300072; 2. 天津市计委培训中心, 天津 300040)

**摘要:**对天津市工业中十六个主要行业1997年的数据进行了主成分分析、聚类分析和因子分析, 提出用第一主成分作为行业规模的度量, 第二主成分作为行业效益的度量, 进而从主成分分析和因子分析得出了各项经济指标反映行业经济效益、行业规模的情况和两个公共因子的含义, 还从聚类分析中得出了各行业发展层次的分布情况。

**关键词:**主成分分析; 聚类分析; 因子分析

**中图分类号:** F403.7

**文献标识码:** A

**文章编号:** 1007-9807(2000)03-0082-07

## 0 引言

世纪之交, 世界各国都在以新的发展姿态迎接新世纪的到来。由于当代科学技术的飞速进步和生产力的迅速提高, 使得世界出现了一体化的趋势。加快经济增长, 调整产业结构, 以增强国家的综合国力, 已成为各国共同的战略发展目标。中国经济正面临一个大好时机, 同时也面临严峻的挑战。加快转变经济增长方式, 促进国民经济持续、快速、健康发展的任务已摆在我们面前。产业结构的调整、优化和升级是转变经济增长方式的重要内容, 是经济增长的重要因素, 因此, 产业发展在一定意义上决定着国民经济的发展。

我国是一个幅员辽阔的发展中大国。各省自然条件、资源禀赋相差甚大, 各地区的工业化水平、产业结构特点和社会经济条件极不平衡, 经济具有相当大的地域差异性。然而迄今国家仍没有制定出相应完整的区域产业政策体系, 许多地区大都简单地按国家产业结构统一的发展序列要求作为本地区产业发展的标准, 缺乏区域产业发展指导政策。1994年4月12日国务院颁布的《90年代国家产业政策纲要》中提出“振兴机械电子、石化、汽车和建筑业, 使之成为国民经济的支柱产业”的战略目标, 这一产业导向具有原则性, 但对地区的

作用不够直接和具体, 加之部分地区对支柱产业的概念认识模糊, 并且受利益驱使, 导致许多地区支柱产业选择的趋同。据统计, 全国30个省区在制定发展规划中将汽车、机械电子、石化和建筑业作为地区支柱产业的分别为22、25、24和19个, 一旦付诸实施, 必然会造成新的结构趋同化<sup>[1]</sup>。由此可见, 产业结构的合理性在我国的社会主义市场经济发展中的地位愈加重要, 产业结构调整优化和升级是转变经济增长方式的重要内容, 是经济增长的重要因素。由于产业的调整是非常复杂的, 单用一个指标不能准确地度量一个产业经济发展的效益, 对产业的考察应该是多方面, 全方位的。于是我们需要用多个指标来综合评价产业发展的效益, 从而准确地度量产业发展的水平, 制定出合理的产业结构政策。

产业结构的变化是涉及国家资源配置的一个大问题, 产业的发展与技术的进步意味着资源配置的不断优化。当今世界已进入信息化社会, 对中国这样一个步入社会主义市场经济的发展中国家来说, 加速产业的发展与升级, 将是立于世界之林的关键, 也是实现经济增长方式转变的重要内容。制定经济发展战略, 准确地选定优势产业, 明确产业发展重点, 分析各产业经济效益, 在一定意义上决定着国民经济的发展。由此, 人们应对建造评价

① 收稿日期: 1999-08-20; 修订日期: 2000-01-04。  
作者简介: 李亚力(1953-), 男, 河北省人, 博士生。

产业效益的指标作一些探索和研究,并给选定优势产业提供一定的理论依据。

在分析工业各行业经济效益时,由于反映一个特定行业的经济效益的指标很多,各指标之间还有一定的相关性,缺乏有效的方法进行比较。本文将视具体问题确定适当的数量指标集,采用主成分分析方法,并且还将利用聚类分析和因子分析方法进行进一步的分析,比较出各行业之间的现状和发展趋势。这将为制定天津市整体经济发展战略提供一定策略依据。

## 1 三种分析方法简介

主成分分析是多元统计分析中应用广泛的一种方法,是由 Pearson 于1901年首先引入, Rao (1964), Gnanadesikan (1977), Morrison (1976) 和 Bibby (1979) 等人进一步发展而形成的重要的多元统计方法。主成分分析是研究如何通过少数几个主成分(即原始变量的线性组合)来解释多变量的方差—协方差结构。具体地说,是导出少数几个主分量它们尽可能多地保留了原始变量的信息,且彼此间又不相关,其思想就是从简化方差,协方差的结构来考虑降维,即在一定的约束条件下,把代表各原始变量的各坐标轴构成的坐标系进行一个旋转产生的新坐标系中的若干个代表了具有最大变异方向的新坐标轴,(也可以说是对原始变量作了一次特殊的正交变换)而得到一组具有某种良好的方差性质的新变量(得到的一组新变量彼此互不相关且在各自的特征方向上有最大方差)再从中选取前几个变量来代替原变量,主成分分析的主要目的就是用来减少变量个数,然后用于标图,回归,聚类,等等。也就是说,主成分分析往往不是最终目的,而是达到某种目的的一种手段,因此它常常被用在某个项目研究的中间环节。

聚类分析的目的是把分类对象按一定规则分成组或类,这些组或类不是事先给定的,或依经验确定的,而是根据数据特征而定的。在一个给定的类里的这些对象在某种意义上倾向于彼此相似,而在不同类里的这些对象倾向于不相似。聚类分析也能够用来概括数据而不仅是寻找自然的或“实在的”分类。

因子分析的目的是用有限个不可观测的潜在

变量来解释原变量间的相关性或协方差关系。这些不可观测的潜在变量(或隐变量)称为公因子(common factor)。公因子一般不能表示为原始变量的线性组合。在公因子分析里,为了使得每个变量有不相关的随机误差或特殊方差,总是假定这些隐变量是线性无关的。因子分析常常伴有因子的旋转,旋转的目的是为了更明确地解释分量或因子的意义,在数学上就是对因子进行的非奇异线性变换。

由于篇幅所限,具体分析理论方法见文[2]。

## 2 各行业发展状态的主成分分析

在选取反映各行业效益的指标时,应力求较完善、全面、真实地反映行业规模和效益的状况,体现各行业的基本特征,并且指标的选取还应遵循以下原则:

- (1) 综合性。选择独立性较强、能综合反映指标群特征的指标,尽量避免选择意义相近的指标;
- (2) 可比性。力求操作指标与统计指标的口径相同,便于采集和横向、纵向比较;
- (3) 实用性和易操作性。

为此,选取以下指标: $Z_1$ :工业总产值(万元); $Z_2$ :销售总收入(万元); $Z_3$ :利税总额(万元); $Z_4$ :固定资产合计(万元); $Z_5$ :资本金合计(万元); $Z_6$ :年平均职工人数(人); $Z_7$ :能源消耗(折合为万吨煤)。我们认为,这七项指标对于反映行业规模及效益具有一定的代表性,并且计算简洁,实用性强,易于操作。为了使主成分更加有效,行业个数要大于指标个数,且越大越好,但在实际计算中也不可能取太多行业,根据国家工业普查的统一标准,且考虑到指标数据的采集方便性,我们选取了天津市工业中十六个主要行业进行分析。数据取自《天津市统计年鉴》<sup>[3]</sup>。具体数据见表1。符号解释如下:

- Ahy1:石油和天然气开采业
- Bhy2:食品加工及制造和饮料制造业
- Chy3:纺织业
- Dhy4:造纸及纸制品业
- Ehy5:石油加工及炼焦业
- Fhy6:化学原料及化学制品制造业
- Ghy7:医药制造业

Hhy8:非金属矿物制品业

Ihy9:黑色金属冶炼及压延加工业

Jhy10:金属制品业

Khy11:普通机械、专用设备制造业

Lhy12:交通运输设备制造业

Mhy13:电气机械及器材制造业

Nhy14:电子及通信设备制造业

Ohy15:仪器仪表及文化办公用品机械制造业

Phy16:电力、蒸气、热水的生产和供应业

表1 原始数据

	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	$Z_7$
Ahy1	736 925	727 965	83 848	568 318	358 591	63 004	132.01
Bhy2	1 547 782	1 042 432	157 311	658 868	519 878	77 272	37.23
Chy3	781 472	725 825	86	565 123	271 001	150 890	54.17
Dhy4	185 121	180 625	11 256	293 243	70 469	30 264	15.39
Ehy5	849 914	890 790	78 008	432 557	201 461	25 262	86.78
Fhy6	1 359 855	1 313 592	101 163	1 246 489	728 997	112 274	251.35
Ghy7	387 945	353 677	110 106	130 798	155 481	25 405	14.34
Hhy8	360 389	305 155	11 999	362 903	209 870	73 129	4.84
Ihy9	998 348	999 093	38 860	1 795 253	1 112 502	72 323	230.16
Jhy10	626 973	588 596	6 468	313 497	261 779	100 039	23.93
Khy11	1 113 153	929 329	96 308	520 169	620 842	182 374	34.21
Lhy12	1 741 995	1 858 132	173 078	802 016	677 145	122 586	32.76
Mhy13	678 239	628 982	4 015	382 572	488 248	83 162	17.51
Nhy14	3 047 065	2 477 499	297 845	722 328	619 309	81 065	17.24
Ohy15	185 332	171 545	5 020	97 736	107 313	38 748	3.79
Phy16	239 799	684 519	55 459	754 226	398 909	18 510	31.70

主成分分析方法的具体步骤如下:

(1) 对原始数据进行标准化变换并求相关系数矩阵  $R$ ;(2) 求相关系数矩阵  $R$  的特征根  $\lambda_j, j = 1, 2, \dots$ ;(3) 根据  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ , 确定特征向量  $e_1, e_2, \dots, e_n$ ;(4) 计算特征根  $\lambda_j$  的信息贡献率、累计贡献率, 根据累计贡献率  $\geq 80\%$  的一般原则, 确定主成分个数  $K$ ;(5) 计算综合评价, 计算每个主成分得分值  $Z_1, Z_2, \dots, Z_K$ , 称  $Z = a_1 Z_1 + a_2 Z_2 + \dots + a_K Z_K$  为主成分综合得分值, 也称为综合得分评价, 其中  $a_i, i = 1, 2, \dots, K$ , 为第  $i$  个主成分  $Z_i$  的信息贡献率。用 SAS 软件包首先对变量进行主成分分析<sup>[4]</sup>, 得到样本相关阵的特征值与特征向量, 见表 2、表 3。

表2 样本相关阵的特征值

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	3.900 16	2.108 34	0.557 166	0.557 17
PRIN2	1.791 82	0.887 32	0.255 974	0.813 14
PRIN3	0.904 50	0.658 07	0.129 215	0.942 35
PRIN4	0.246 43	0.156 70	0.035 204	0.977 56
PRIN5	0.089 73	0.049 35	0.012 818	0.990 38
PRIN6	0.040 38	0.013 40	0.005 769	0.996 15
PRIN7	0.026 98	.	0.003 855	1.000 00

表3 样本相关阵的特征向量

	PRIN1	PRIN2	
Z1	0.442 609	-0.336 388	工业总产值(万元)
Z2	0.458 709	-0.277 181	销售总收入(万元)
Z3	0.363 455	-0.448 975	利税总额(万元)
Z4	0.387 899	0.444 349	固定资产合计(万元)
Z5	0.441 123	0.275 221	资本金合计(万元)
Z6	0.233 004	0.004 313	年平均职工人数(人)
Z7	0.249 476	0.578 985	能源消耗(折合为万吨煤)

由此可知, prin1和 prin2的累计贡献率已达81%,故选两个主成分就可以了. 各行业第一、二主成分 prin1、prin2的计算公式(公式中的  $Z_i$  是原始数据标准化后的数值)为

$$\text{prin1} = 0.442\ 609\ Z_1 + 0.458\ 709\ Z_2 + 0.363\ 455\ Z_3 + 0.387\ 899\ Z_4 + 0.441\ 123\ Z_5 + 0.233\ 004\ Z_6 + 0.249\ 476\ Z_7$$

$$\text{prin2} = -0.336\ 388\ Z_1 - 0.277\ 181\ Z_2 - 0.448\ 975\ Z_3 + 0.444\ 349\ Z_4 + 0.275\ 221\ Z_5 + 0.004\ 313\ Z_6 + 0.578\ 985\ Z_7$$

两个主成分已代替了原来七个变量的大部分,现在解释这两个主成分的经济含义. 第一主成分(prin1)在各个变量上的系数都是正的,而且数值上相差不大,数值都在0.2-0.5之间,在各个变量上的系数代表此变量对各行业规模的相对贡献率,因此可以认为 prin1代表各行业的规模大小,即 prin1的数值越大,行业规模越大. 我们把十六个行业的第一主成分值按大小排序,得到行业得分表4.

表4 各行业得分

	PRIN1	PRIN2
Nhy14	3.746 12	-2.970 47
Ihy9	2.657 45	3.304 35
Fhy6	2.552 48	1.865 88
Lhy12	2.358 52	-1.139 68
Bhy2	0.968 11	-0.867 17
Khy11	0.893 35	-0.338 16
Ahy1	-0.180 55	0.522 30

续表4

	PRIN1	PRIN2
Chy3	-0.492 90	0.290 60
Ehy5	-0.724 69	-0.216 75
Mhy13	-0.889 93	0.097 12
Phy16	-0.956 82	0.388 94
Jhy10	-1.256 02	-0.119 51
Hhy8	-1.673 19	0.334 48
Ghy7	-1.847 59	-0.853 82
Dhy4	-2.501 80	-0.040 53
Ohy15	-2.652 52	-0.257 57

此结果显示:电子及通信设备制造业(prin1 = 3.746 12)明显居于首位,黑色金属冶炼及压延加工业(2.657 45),化学原料及化学制品制造业(2.552 48),交通运输设备制造业(2.358 52)次之,与实际比较相符. 第二主成分在变量  $Z_1, Z_2, Z_3$  上的系数是负值,其余为正值. 在负值中,  $Z_1$  是工业总产值,  $Z_2$  是销售总收入,  $Z_3$  是利税总额,这些可视为产出;在正值中,  $Z_4$  是固定资产合计,  $Z_5$  是资本金合计,  $Z_6$  是年平均职工人数,  $Z_7$  是能源消耗,这些可视为投入,故可把 prin2 视为代表行业效益. 其具体含义是:prin2 越大,效益越差;prin2 越小,效益越好.

将第一主成分 prin1 作为横轴,将第二主成分 prin2 作为纵轴,每个行业就有了两个坐标,因而可以在平面上标出一个点,把这十六个行业全部标出就可得到主成分坐标图(图1). 由于 prin1 代表规模,prin2 代表效益,且 prin1 越大越好,prin2 越小越好,故处于坐标系越靠右下越好,即规模越大,效益越好. 越靠右上越差,即规模很大,但效益很差. 从散点图上可以看出, Nhy14 即电子及通信设备制造业最好,也就是说,规模最大,效益也最好,其次为 Lhy12 为交通运输设备制造业,最差的行业 Ihy9 为黑色金属冶炼及加工业,规模很大,仅次于电子及通信设备制造业,但效益却是最差的. 进一步由图1可直观地把十六个行业分成四类: Nhy14 为一类, Fhy6 为一类, Ihy9 为一类,其余为一类.

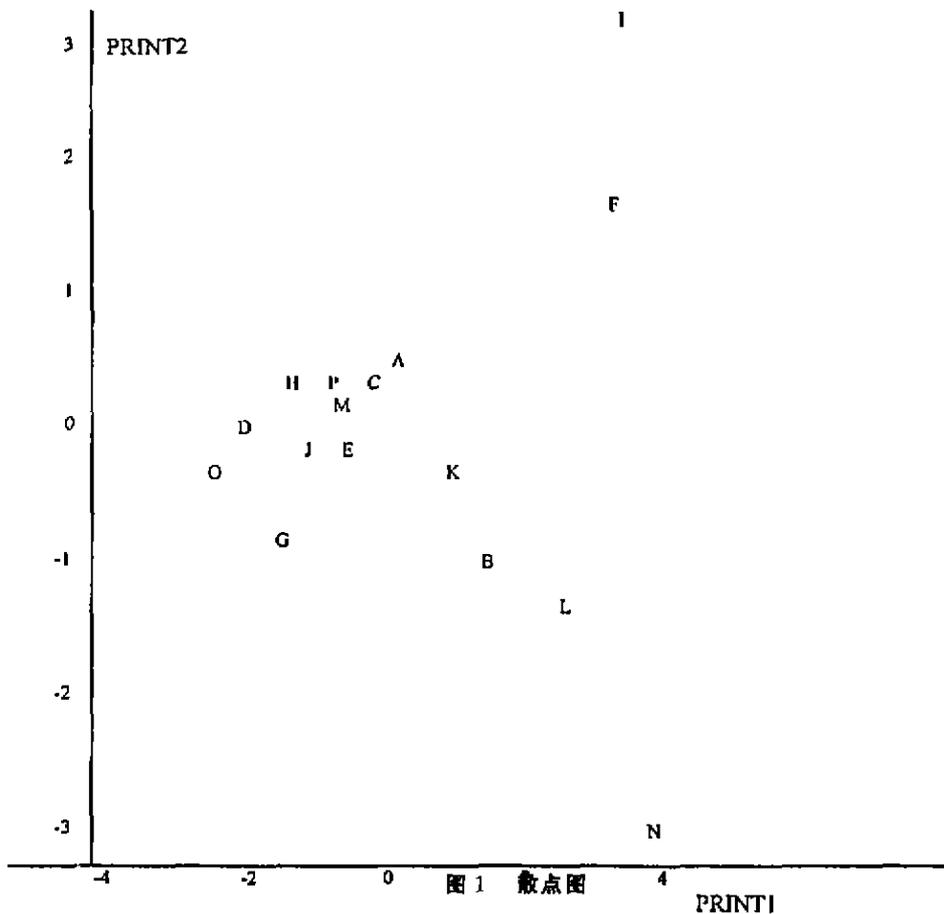


图1 散点图

### 3 各行业发展状态的聚类分析

用 SAS 软件包,用第一主成分和第二主成分来作聚类分析<sup>[5]</sup>.从结果中可以看出,它们可以分成四类:

(1)Nhy14即电子及通信设备制造业单独为一类,记为C1;

(2)Fhy6即化学原料及化学制品制造业单独为一类,记为C2;

(3)lhy9即黑色金属冶炼及压延加工业单独为一类,记为C3;

(4)把其余的分为一类,记为C4;

C1类是最好的一类,即规模最大,效益最好,属于集约型行业,应成为工业中的支柱产业. C4是比较好的一类,有待于进一步发展. C2类和C3类则为比较差的行业,是高投入,低产出,高消耗,低效益的行业,属于粗放型行业.若要分成三类,

则把C2和C3合并成一类即可.

分类结果比较符合目前天津工业的实际情况.电子及通信设备制造业中科技成分含量很高,也属于附加值较高的行业,相对于其他行业来说有一定的优势,所以效益比较好,也应成为工业中的支柱产业.而化学原料及化学制品制造业与黑色金属冶炼及压延加工业对于天津市来说则比较差,落后于国内其它地区,有待于加大科技成分含量,或改变生产经营方针以改变其集约型增长方式.因篇幅所限,聚类分析运行结果略.

### 4 各行业发展状态的因子分析

为使主成分有更鲜明的实际背景,各原始指标与主成分之间有着更合理的解释.作最大方差正交旋转,使各因子仍保持互不相关,且因子负荷最大限度地向两极分化,运行结果见表5,表6.

从主成分分析输出的相关阵中可知有两个大的特征值,3.900 2和1.791 8,它们一起解释了标准方差的81.31%。于是在大多数场合下,前两个公共因子提供了原始数据的足够信息。其中第一个公共因子对所有七个变量都具有大的正负荷,它与 $Z_2$ (销售总收入)相关特别大(0.905 90)。第二公共因子中, $Z_4$ (0.594 80)和 $Z_7$ (0.775 02)与 $Z_1$ (-0.450 29)和 $Z_3$ (-0.600 99)形成鲜明对照,而在 $Z_6$ (0.005 77)上的负荷非常小,最终公共因子方差估计表明,除 $Z_6$ 外,所有变量很好地被两个公共因子所解释,因为最终公共因子方差估计在0.843 398(来自 $Z_7$ )到0.966 810(来自 $Z_1$ )的范围内。

表5 旋转前因子得分

	FACTOR1	FACTOR2
Z1	0.874 10	-0.450 29
Z2	0.905 90	-0.371 03
Z3	0.717 78	-0.600 99
Z4	0.766 05	0.594 80
Z5	0.871 17	0.368 41
Z6	0.460 16	0.005 77
Z7	0.492 68	0.775 02

表6 旋转后因子得分

	FACTOR1	FACTOR2
Z1	0.982 80	0.030 31
Z2	0.972 15	0.115 04
Z3	0.919 22	-0.177 31
Z4	0.381 31	0.891 76
Z5	0.583 05	0.744 78
Z6	0.399 58	0.228 29
Z7	0.054 83	0.916 73
PRIN1	0.874 44	0.485 13
PRIN2	-0.485 13	0.874 44

### 参考文献:

- [1] 国家计委产业经济与技术经济研究所. 中国产业发展报告[R]. 北京: 中国经济出版社, 1997
- [2] Muirhead R J. Aspects of multivariate statistical theory[M]. New York: John Wiley & Sons, 1984
- [3] 天津市统计局. 天津市统计年鉴[R]. 天津: 统计出版社, 1998
- [4] Dillon W R, Goldstein M. Multivariate analysis, methods and applications[M]. New York: Wiley, 1984
- [5] Everitt B. Cluster analysis (second edition). New York: Halsted, 1980
- [6] Mardia K V, Kent J T, Bibby J M. Multivariate analysis[M]. London: Academic Press, 1979
- [7] Johnson R A, Wichern D W. Applied multivariate statistical analysis[M]. New Jersey: Prentice-Hall, Englewood Cliffs, 1982

采用最大方差旋转后,可以看出,旋转后的公共因子有着更鲜明的实际意义<sup>[6-8]</sup>。其中第一公共因子中系数绝对值大的主要是 $Z_1$ (工业总产值), $Z_2$ (销售总收入)和 $Z_3$ (利税总额),故可以认为,这是代表行业的产出,也可以称之为效益;第二公共因子主要反映了 $Z_4$ (固定资产合计), $Z_5$ (资本金合计), $Z_7$ (能源消耗),反映了对行业的投入,也可称为消耗,如果以第一公共因子(Factor1)和第二公共因子(Factor2)为坐标轴画图,显然可看出,坐标轴将穿过 $Z_1(A)$ , $Z_2(B)$ 和 $Z_7(G)$ ,由于 $Z_1$ 为工业总产值, $Z_2$ 为销售总收入, $Z_7$ 为能源消耗,这进一步说明了两个公共因子的含义。第一公共因子代表效益,第二公共因子代表消耗。

## 5 结论

由上面的结果解释中可看出,不仅利用主成分分析给出了各行业规模和效益的排名,电子及通信设备制造业最好,黑色金属冶炼及压延加工业最差。且从主成分分析和因子分析中能看出各项经济指标可在多大程度上反映行业经济效益、发展潜力的状况,还可从聚类分析中得到各行业发展层次的分布情况,与实际比较吻合。

应当指出,评价经济问题是比较复杂的,本文选取的指标由于资料的限制,也未必都能确切地反映各行业的实际情况,但是所得的结论表明文中使用的方法对制定地区整体经济发展战略是有实际意义的。

[8] 方开泰. 实用多元统计分析[M]. 上海: 华东师范大学出版社, 1992

## Research on comparisons of main vocations in the industry of Tianjin

*LI Ya-li<sup>1</sup>, LIU Jia-kun<sup>1</sup>, LI Hong-lin<sup>1</sup>, YANG Chang-jun<sup>2</sup>, YU De-shui<sup>2</sup>*

1. Tianjin University, Tianjin 300072, China;

2. Tianjin Planning Commission Training Center, Tianjin 300040, China

**Abstract:** In this paper, the data of 16 vocations in the industry of Tianjin are dealt with by principal component analysis, cluster analysis and factor analysis. As a result, the first principal component is regarded as the measure of the scale vocation, and also the second principal component is considered as the measure of efficiency. Additionally, the order of them is presented in the paper. The economic efficiency, the scale of vocation and the connotation of common factor implied by the variables are obtained by principal analysis and factor analysis. The hierarchy of the vocation development is acquired by cluster analysis.

**Key words:** principal component analysis; cluster analysis; factor analysis

.....  
(上接第38页)

## Governance to entrepreneurial distribution behavior; role of the entrepreneurial incentive mechanism

*LI Yuan, SUN Kai*

Management School, Xi'an Jiaotong University, Xi'an 710049

**Abstract:** This article divides the entrepreneurial incentive mechanism to obvious incentive and implicit incentive based on the basic function of the entrepreneurial incentive mechanism. Then, the article deeply studies the implicit incentive to managerial entrepreneur and the governance to the entrepreneurial distribution behavior under the circumstances both playing the role of the social restrict mechanism and not playing the role of the social restrict mechanism.

**Keywords:** entrepreneur; incentive; behavior; governance

\*\*\*\*\*  
(上接第74页)

## Research on the existence of nonlinearly cointegrated relationship

*SUN Qing-hua, ZHANG Xi-bin, ZHANG Shi-ying*

School of Management, Tianjin University, Tianjin 300072, China

**Abstract:** In this paper we develop a new estimating and testing procedure on the cointegrated relationship among nonlinear long memory time series by the means of neural network technique. Monte Carlo simulation is also presented to illustrate the application value of the new method. Furthermore, based on the ergodic theory of Markov process we analyze the stationary distribution of network residual series, and prove the efficient conditions under which the unique asymptotic stationary distribution exists.

**Key words:** nonlinear cointegration; neural network algorithm; stationary distribution