Research on Comprehensive Evaluation of Harmonious Society in East China*

Kepei Men, Liangyu Jiang and Jing Liu
Nanjing University of Information Science and Technology, Nanjing 210044, China
Email: menkp@yahoo.com.cn

Abstract *The harmonious society is the general goal of social development. We select 5 subsystems and 35 evaluation index to establish an system for evaluating harmonious society, with the method of Principle Component-Cluster Analysis, according to 2007 Statistical Yearbook of China and some other statistical data. Meanwhile, we use TOPSIS method as an assistant certification to analyze the harmonious society development of east China. The result shows that the six provinces in east China and Shanghai city could be divided into 3 parts. The first part includes Shanghai. The second part includes Zhejiang, Jiangsu and Shandong. And the third part includes Fujian, Jiangxi and Anhui. The result is identical to practice developing of east China.*

Keywords *East China Harmonious Society Comprehensive evaluation Principle Component-Cluster Analysis (PCCA)*

1. Introduction

Realizing social harmony and building a nice society is the social ideal that human being assiduously seeks. It is also the social ideal of all the Marxist Party including the Communist Party of China. To construct a socialist harmonious society is a great task that put forward in whole under the new situation of initiating the socialist road with the specific practice in China. Mr Hu Jintao proposed to realize harmony zone on the 5th anniversary forum of the foundation of Shanghai cooperation organization. East China is the most active economic growth zone. The six provinces and Shanghai city of it account for less than 1/4 of the whole country, the population is 28.74% of the whole nation and the land area only accounts for 8.13%. But it has made great contribution for China, because of its wonderful economy.

According to the statistical bulletins of nation and east China in 2007, we knew that the resident income of the six provinces and Shanghai city was kept growing, and the living standard was improved further. The per capital dominating income of urban residents had exceeded 11 thousand yuan. Among them, Shanghai, Jiangsu, Zhejiang, Fujian and Shandong were over 14 thousand yuan, which were higher than the national average 13.786 thousand yuan. The GDP of this area in 2006 and 2007 were reached 8826.513 billion yuan and 10406.21 billion yuan respectively, which accounted for 41.86% and 42.20% of China's GDP. Therefore, study on the problem about the construction of the harmonious society in east China is strongly representative and has a profound significance. In this paper, we focus on the construction idea of socialist harmonious society, to put forward a harmonious society index system with great operative, for the purpose of providing reference for the construction of harmonious society.

2. The introduction to research methods

2.1 The Method of Principal Component and Clustering Analysis[1-4]

Suppose there are n samples, every sample has p indices, we get a original data matrix

* This work is supported by the Key Projects of National Statistics Research Program (2008LZ022).

$$X = (x_{ij})_{n\times p},\ \ (i = 1,2,\cdots,n;\ \ j = 1,2,\cdots p)$$

and we make linear combination (comprehensive index) with p vectors $X_1, X_2, \cdots, X_p$, that is

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots a_{pi}X_p,\quad i = 1,2,\cdots, p$$

Then we limit the combination coefficients $a_i = (a_{1i}, a_{2i}, \cdots, a_{pi})$ with

$$a_{1i}^2 + a_{2i}^2 + \cdots + a_{pi}^2 = 1,\qquad i = 1,2,\cdots, p$$

in which, a_i is a unit vector and $a_i a_i = 1$.

The comprehensive index is determined by the following rules:

(1) F_i and F_j ($i \neq j, i, j = 1,2,\cdots, p$) are uncorrelated, that is $Cov(F_i, F_j) = 0$.

(2) F_1 is the biggest variance in all the linear combinations of $X_1, X_2, \cdots, X_p$, namely,

$$Var(F) = \max_{c'c=1} Var\left(\sum_{i=1}^{p} c_i X_i\right),$$

in which $c' = (c_1, c_{2,} \cdots, c_p)$.

As to the rule (2), we will get the similar conclusions. F_2, uncorrelated with F_1, is the biggest variance in all the linear combinations of $X_1, X_2, \cdots, X_p$. F_p, uncorrelated with $F_1, F_2, \cdots F_{p-1}$, is the biggest variance in all the linear combinations of $X_1, X_2, \cdots, X_p$, and so on.

The comprehensive vectors $F_1, F_2, \cdots, F_p$, satisfied above requirements, are the principal components. The information extracted from the information content of original indices decrease successively. We use variance to measure information extracted by every principal component, and the contribution of principal component variance is equal to corresponding eigenvalue λ_i to correlation matrix of the original data. $a_i = (a_{1i}, a_{2i}, \cdots, a_{pi})$, the combination coefficient of every principal component, is the eigenvector t_i to corresponding eigenvalue λ_i. The contribution rate of variance is $\alpha_i = \lambda_i / \sum_{j=1}^{p} \lambda_j$. The more α_i, the more information corresponding principal component explains. When the variance contribution rate of one principal component is very small, we think the information it provides is little, then we may delete it. In general case, if the cumulative variance contribution rate of first q principal components reaches 85%, we just consider the first q principal components, then explain properties of random vector X with them. Other principal components are the random errors caused by incorrect observation.

Research on comprehensive evaluation has made process[5-7]. Papers [1-3] point the method popular in recent year is to structure a comprehensive evaluation function of principal components for ranking, which is based on the contribution rate of variance α_i, but it is wrong.

When the contribution rate of variance of the first principal component F_1 is quite high (over 85%), we may think this principal component can almost reflect the information provided by the original variables. Then we may rank and evaluate according to the scores of the first principal component.

To the ranking problems of multiple index system, when the contribution rate of variance of the first principal component is not over 85%, that is, the original data information expressed by the first principal component is not enough, it has one-sidedness, if we still rank and evaluate only by the scores of the first principal component. At this time, we combine principal

component analysis with cluster analysis, which is named “principal component-cluster analysis (PCCA)”.

Table 1 The index system

	Subsystem	Index	Unit	***Directivity***
Harmonious society	Material civilization	1	Yuan/person	positive
		2	yuan	positive
		3	yuan	positive
		4	%	negative
		5	%	negative
		6	Square meter/person	positive
		7	%	positive
		8	Part/10 thousand	positive
		9	agriculture=1	medium
		10	%	positive
	Political civilization	11	Person/10 thousand	positive
		12	female=100	medium
		13	%	negative
		14	Part/10 thousand	negative
	Spiritual civilization	15	%	positive
		16	Person/100 thousand	positive
		17	%	positive
		18	%	positive
		19	Kind/10 thousand	positive
	Social civilization	20	Person/10 thousand	positive
		21	person/100 thousand	negative
		22	old	positive
		23	%	positive
		24	%	positive
		25	Part/10 thousand	positive
		26	%	negative
		27	%	positive
	Ecological civilization	28	ton/10 thousand yuan	negative
		29	%	positive
		20	%	positive
		31	%	positive
		32	%	positive
		33	%	positive
		34	10 thousand stere	positive
		35	%	positive

Note: 1 means Per capital GDP; 2 means per capital disposable income of urban residents; 3 means per capital net income of rural residents; 4 means the Engle's coefficient of urban residents; 5 means the Engle's coefficient of rural residents; 6 means housing areas of rural residents; 7 means the proportion that the tertiary industry accounts for GDP; 8 means numbers of patents application; 9 means consumption level of urban and rural residents; 10 means the proportion that R&D accounts for GDP; 11 means lawyer’s numbers that per ten thousand people have; 12 means sex ratio of senior middle school graduates; 13 means registered unemployment rate of urban residents; 14 means numbers of criminal case; 15 means the proportion that education and culture entertainment account for total consumption expenditure; 16 means average students number of colleges and universities; 17 means excellent rate of product quality; 18 means the proportion that education operating expenses account for fiscal expenditure; 19 means book publishing kinds; 20 means doctor’s numbers that per ten thousand people have; 21 means death toll from traffic accidents; 22 means average life expectancy; 23 means mobile subscription; 24 means gross enrollment ratio of higher education; 25 means numbers of urban community service facilities; 26 means gross divorce rate; 27 means the proportion that numbers of participate in

medical insurance account for total amount; 28 means energy consumption in every unit; 29 means forest coverage; 30 means industrial wastewater treatment level; 31 means per capita park greenery area; 32 means water subscription; 33 means gas subscription; 34 means daily treatment ability of domestic sewage; 35 means harmless treatment rate of domestic waste. [11]

"Principal component-cluster analysis" is such a method. First we do the principal component to get some principal components, with which we do cluster analysis to samples, then we rank and classify the samples by the scores of the first principal component. Specific ways are the following:

(1) select first r principal components by cumulative contribution rate and calculate the scores of the principal components

$$F_l = a_{1l}X_1 + a_{2l}X_2 + \cdots a_{pl}X_p, \quad (l = 1, 2, \cdots, r)$$

(2) do the systematic cluster analysis to the selected new matrix ($F_1, F_2, \cdots, F_r$);

(3) calculate the scores' average value of the first principal component to determine ranking of every class;

(4) determine ranking of every sample in the class to get comprehensive evaluation result, according to every sample's score.

2.2 The Method of TOPSIS

Main steps of TOPSIS:

(1) select p evaluation index to n evaluation units for comprehensive evaluation, then get a evaluation matrix $X = (x_{ij})_{n\times p}$, in which x_{ij} is observation data in unit i .

(2) make original data being dimensionless, then get a normalized evaluation matrix $Z = (z_{ij})_{n\times p}$, in which

$$z_{ij} = x_{ij} / \sqrt{\sum_{k=1}^{n} x_{kj}^2} \quad (i = 1, 2, \cdots, n;\ \ j = 1, 2, \cdots, p) \cdot$$

(3) determine the positive ideal solution Z^+ and negative ideal solution Z^- of matrix Z , then get

$$Z^+ = (z_1^+, z_2^+, \cdots, z_q^+), \qquad Z^- = (z_1^-, z_2^-, \cdots, z_q^-),$$

in which $z_j^+ = \max\{z_{1j}, z_{2j}, \cdots, z_{nj}\}$, $z_j^- = \min\{z_{1j}, z_{2j}, \cdots, z_{nj}\}$, $(j = 1, 2, \cdots, q)$.

(4) calculate the distances between every unit and positive ideal solution, and the distances between every unit and negative ideal solution:

$$D_i^+ = \sqrt{\sum_{j=1}^{q} (z_{ij} - z_j^+)^2}\ , \qquad D_i^- = \sqrt{\sum_{j=1}^{q} (z_{ij} - z_j^-)^2} \qquad (i = 1, 2, \cdots, n)$$

(5) calculate relative approach degrees between every evaluation unit and the optimal solution:

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \times 100\%, \quad (i = 1, 2, \cdots, n)$$

(6) rank according to the relative approach degree. The more it closes to 100, the more C_i . It indicates that unit i is more close to the optimal level. Otherwise, unit i is less close to the optimal level.

3. Establishment of Index System and Empirical Analysis

To build a harmonious society is a great social system engineering. From the horizontal point of view, it is closely associated with society, politics, economy and culture. From the vertical point of view, it involves in macro-view, medium-view and micro-view. The concept of harmonious society is extensive, comprehensive and systematic. It includes every aspect of

economy, politics, society, culture, environment and people's life. Therefore, the index system of harmonious society should embody the comprehensive and systematic connections among index, and reflect progress condition of construction in many aspects. In this paper, we will divide it into 5 subsystems to make it have good operability, that is, material civilization system, political civilization system, spiritual civilization system, social civilization system and ecological civilization system, (Details in Table 1) [8-11] .

3.1 Calculation with the Method of Principal Component and Cluster Analysis

First we do pretreatment to the original data. Then we make principal component analysis to the harmonious index of six provinces and Shanghai city with SAS program. Finally, we get the following eigenvalues of the correlation matrix.

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	18.6926827	12.4745482	0.5341	0.5341
2	6.2181345	2.0407184	0.1777	0.7117
3	4.1774161	1.4626788	0.1194	0.8311
4	2.7147372	0.8307516	0.0776	0.9087
5	1.8839856	0.5709418	0.0538	0.9625
6	1.3130439	1.3130439	0.0375	1.0000

……

In which, PRIN1, PRIN2, PRIN3, PRIN4 denote the first four principal components. The second list is eigenvalues of sample correlation matrix, the fourth list is contribution proportions of variance, and the fifth list is cumulative contribution proportions. As cumulative contribution proportion of the first four eigenvalues reaches 90.87%, which is higher than 85%, we only need to select the first four principal components to summarize the all data, then get the scores of the principle components (see Table 2).

Table 2 The scores of the principle components analysis

Cities / Scores	The first Principal component	The second principal component	The third principal component	The fourth principal component
Shanghai	8.52652	-2.49745	-0.61749	0.58853
Jiangsu	0.68228	1.08935	1.92012	-1.20837
Zhejiang	1.99413	4.24675	-0.908082	-0.56337
Anhui	-3.85481	-2.3237	-2.88072	-1.98804
Fujian	-1.67628	1.85291	-0.83024	1.03237
Jiangxi	-3.62903	-0.86869	-0.02778	2.92133
Shandong	-2.04282	-1.49017	3.34413	-0.78246

The contribution proportion of first principal component's variance is 53.41%, which is the largest of all linear combinations, and the information of the first principal component is the biggest. We get samples' ranking the first time by calculating the scores of the first principal component. If we rank only according to the scores of the first principal component, as contribution proportion is not over 85%, the information will not be big enough and the result will have one-sidedness. Thus, we do the cluster analysis to the first four principal components' score matrix with SAS program, then we get the cluster graph (see Figure 1).

From Figure 1, we know that the six provinces and Shanghai city can be divided into three classes, that is, {Shanghai}; {Jiangsu, Shandong, Zhejiang}; {Fujian, Jiangxi, Anhui}.

Then we rank according to the scores of first principal component. That is, {Shanghai}; {Jiangsu, Shandong, Zhejiang}; {Fujian, Jiangxi, Anhui}.

Finally, we get the ranking by the scores of the first principal component in every class. That is, {Shanghai, Zhejiang, Jiangsu, Shandong, Fujian, Jiangxi, Anhui}.

From Table 3, we get the rankings by principal component-cluster analysis and the scores of the first principal component. We find that they are almost the same, the only different is ranking between Shandong and Fujian. However, 20 index in Shandong are better than Fujian, Shandong should be in front of Fujian. Therefore, it is more reliable to get the ranking by principal

component-cluster analysis.

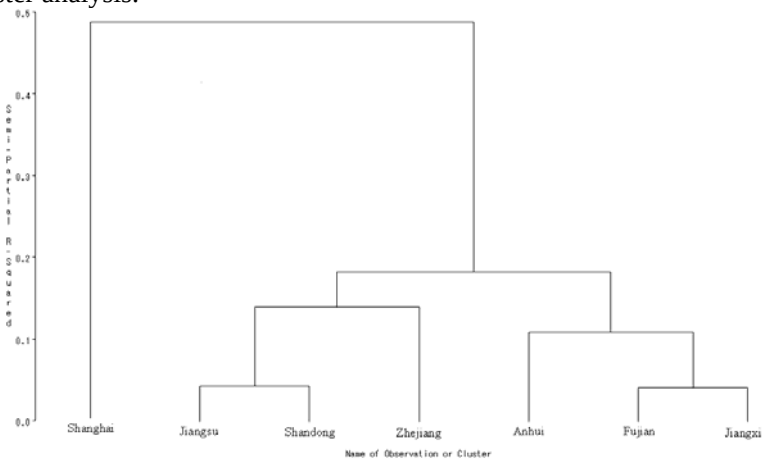

Figure 1 Cluster graph

Table 3 The compositors of the principal component cluster and the first principal component

	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong
Ranking by principal component-cluster analysis	1	3	2	7	5	6	4
Ranking by the scores of the first principal component	1	3	2	7	4	6	5

3.2 Calculation with the Method of TOPSIS

We calculate relative approach degree C_i of every subsystem by MATLAB program. The more it closes to 100, the more C_i. It indicates that unit i is more close to the optimal level. Otherwise, unit i is less close to the optimal level. Details in Table 4.

The construction of material civilization, political civilization, spiritual civilization, social civilization and ecological civilization are closely linked with each other in the process of building a socialist harmonious society and a well-off society. They are very important, so we should treat them equally. We get the final scores of the six provinces and Shanghai city after doing the equal weight. Then we rank by the final scores, which is listed in Table 4. The result is the same as the ranking by principal component-cluster analysis, which verifies rationality of the method of principal component-cluster analysis.

Table 4 The development level of six provinces in east China and Shanghai city

	Material civilization	Political civilization	Spiritual civilization	Social civilization	Ecological civilization	Score	***Ranking***
Shanghai	86.6131	46.87	80.45	71.89	32.28	63.62062	1
Jiangsu	38.4696	32.37	19.50	32.84	57.44	36.12392	3
Zhejiang	53.6959	23.66	20.32	47.55	52.83	39.61118	2
Anhui	12.3374	31.65	9.48	17.73	24.29	19.09748	7
Fujian	21.9262	28.24	23.78	22.12	49.08	29.02924	5
Jiangxi	8.7141	35.27	27.04	27.24	41.41	27.93482	6
Shandong	24.8297	60.62	14.78	26.26	44.15	34.12794	4

4. Conclusions and Discussions

2006 was the beginning of eleventh five-year plan. Under the leadership of provincial and municipal governments in east China, we had acquired remarkable achievement in the field of social economic development, by insisting in carrying out macro-control policy and guiding with scientific development view. According to the methods of principal component-cluster analysis

and TOPSIS, we divide six provinces and Shanghai city into 3 classes on the level of social harmonious development.

(1) Shanghai city is the first class. The score of Shanghai is far higher than any other province. It shows that Shanghai city is superior to other places in the construction of harmonious society. Since reform and opening up to the outside world, Shanghai city has got great achievement as the first metropolitan. The index of per capital GDP, per capital disposable income of urban residents and per capital net income of rural residents are in the first place.

(2) Zhejiang, Jiangsu and Shandong are in the second class, which are high in harmony. The scores of Zhejiang province and Jiangsu province are high in material civilization system, ecological civilization system and social civilization system, but low in spiritual civilization system. These places have strong economic strength, good security systems and harmonious living condition. The scores of Shandong province are low in material civilization system, spiritual civilization system and social civilization system, but high in political civilization system. Shandong province is one of the best provinces in public security.

(3) Fujian, Jiangxi and Anhui are in the third class. It has significant gap to others. The relatively backward social economy restricts social development, leading to the development of society unbalanced.

From above conclusions, we know that Shanghai, Zhejiang, Jiangsu and Shandong are in the leading positions of the whole nation. In the future, we should not only focus on the quantity of economic operation, but also focus on the quality. To other relative undeveloped regions in east China, we should speed up the development to shorten the difference. Six provinces and Shanghai city in east China will obtain new development in the new century through cooperation and exchange and will create a more bright future.

References

[1] Xueming Wang. Applied Multivariate Analysis (2rd edition). Shanghai: Shanghai University of Finance and Economics Press, 2004: 1-358. (in Chinese)

[2] Xueming Wang. Query to the applied method of principal component. *Statistics and Decision*, 2007, 8: 31-32. (in Chinese)

[3] Jingya Xu, Wang Yuanzheng. Improvement to the applied method of principal component. *Mathematics in practice and theory*, 2006, 6: 68-75. (in Chinese)

[4] Daoyuan Zhu, Chengou Wu, Weiliang Qin. Applied Multivariate Analysis and SAS software. Nanjing: Southeast University Press, 1998: 1-408. (in Chinese)

[5] Chuncheng Wu, Shengbao Yao, Chaoyuan Yue. The Model of the Evaluation of the Comprehensive Progress for Large Project. *Advances in System Science and Applications*, 2006, 6(3): 410-415.

[6] Pu Gong, Qiang Si, Jianling Meng. The Evaluation Method of Multi-stage Compound Real Option on Human Capital Valude. *Advances in Systems Science and Applications*, 2006, 6(1): 101-106.

[7] Wei Wang, Zhuangzhi Liu. Comprehensive Evaluation of Project Risk Based on Fuzzy-network Analytical Methods. *Advances in Systems Science and Applications*, 2008, 8(2): 270-275.

[8] State Statistic Bureau Studying Team. Research on the index system of harmonious society. *Statistical Research*, 2006, 5: 23-29. (in Chinese)

[9] Mei Song, Qi Xin. The construction of harmonious society index system. *Beijing Social Science*, 2006, 1: 62-66. (in Chinese)

[10] Jianguo Ouyang. Research on comprehensive evaluation index system of harmonious society. *Zhejiang Social Science*, 2006, 2: 16-22. (in Chinese)

[11] PRC National Statistics Bureau. China Statistical Yearbook-2007. Beijing: China Statistics Press, 2007: 1-1028. (in Chinese)