Spatial geographic information variance analysis in Anhui Province

Jie Zhu [1,2,a] , Zhanqiang Chang [1,2,b*] ,Xiaomeng Liu[1,2], Wen Yu[1,2],Wei Wang[1,2]

[1]College of Resource, Environment & Tourism, Capital Normal University, Beijing 100048, China

[2]Key Lab. of 3D Information Acquisition of Education Ministry of China, Beijing 100048, China

[a]18519196454@163.com,[b]changkkll@tom.com

*Corresponding author

Abstract. Regional industry competition is a momentous aspect of regional geographic information engineering. It is quite vital to investigate industrial competitiveness in both theoretical and practical fields. In this paper, the principle and procedure of principal component analysis (PCA) were respectively introduced in detail. The empirical study and comparative analysis of the regional economic competitiveness was conducted by setting some economic indicators. At the same time, SPSS software was applied to extract available geographic information elements which indirectly reflected the enhancement of industrial competitiveness. By analyzing and comparing the economic development of the 16 cities during 2013, we sought out the primary geographic elements affecting the growth trend, and then calculated the overall score of each city in Anhui Province according to comprehensive evaluation model. The results of spatial analysis indicated that the overall score was consistent with each city's economic strength. The analysis in this paper certainly makes sense to provide theoretical reference and practical value for overall economic strength and sustainable regional construction.

Introduction

Regional economists [1] believe each country (region) elects the division of labor and specialized production for formatting the national (regional) industry through its comparative advantage over other regions .These advantages, thus determine the economic competitiveness of a region. Michael Porter [2], professor of Harvard Business College, insisted that each country has the ability to possess strong competitiveness in only one or several industries, while the other industries are lack of competitiveness. In other words, the difference of geographic information leads to the different regional advantage industries. Consequently, the significance of industrial competitiveness is seeking a region with the comparative industrial advantage to create a competitive advantage in industry and then obtaining the competitive benefits.

Anhui, an extraordinary province in Central China, should also analysis its comprehensive competitiveness for the rapid growth in economy. Based on the major breakthrough of the original industrial structure, it aims at promoting and rationalizing the evolution of industrial structure to a higher level.

Method

Principal component analysis (PCA) [3] is a classical method in multivariate statistical analysis, which is the most suitable method for the analysis and calculation when there are some correlations between variables. The basic idea is that the linear transformation is used to gather indicators by abnegating a small part of the information and replacing the original index with several synthetic indexes. The comprehensive index is highly unrelated to each other, so that the high dimensional index data can be simplified in the best way. The most classical approach is using the variance of the first comprehensive index (F1) to express information, that is, the greater the Var（F_i）, the more information it contains. Hence, the maximum variance should be selected among all the linear combinations, called the first principal component (F1). If F1 is not enough to represent the original

index information, then consider to select the second linear combinations (F2).

Assuming that there exist p variables, each of it has sample size n. The original variable Xi is converted into a set of new features, called the principal components (F). The principal component characterized by orthogonality is a linear combination of the original variables, which can be expressed by polynomial:

$$\begin{cases} F_1 = a_{11}ZX_1 + a_{12}ZX_2 + \cdots + a_{1p}ZX_p \\ F_2 = a_{21}ZX_1 + a_{22}ZX_2 + \cdots + a_{2p}ZX_p \\ \qquad\qquad \cdots \\ F_m = a_{m1}ZX_1 + a_{m2}ZX_2 + \cdots + a_{mp}ZX_p \end{cases} \qquad (1)$$

where $a_{i1}, a_{i2}, \cdots, a_{ip}(i = 1,2,\cdots,m)$ are eigenvectors corresponding to eigenvalues, $ZX_1, ZX_2, \cdots, ZX_p$ are the values of the original variable after normalization. In practical application, the dimension of the index is always different, so the original data should be normalized.

The principal components $(F_1, F_2 \cdots F_m)$ are independent with each other and the proportion in the total variance is taking on descending order. It is always picking up the most significant main components to simplify and settle the dispute. F can be expressed as

$$F = \sum W_i F_i \qquad (2)$$

where F and W_i are the comprehensive score and the variance contribution rate corresponding to eigenvalues respectively.

Spatial geographic information variance analysis

The city [4] was selected as the research unit. It is likely that make a fairly objective evaluation of the industrial structure competitiveness of 16 cities in Anhui province.

Normalization.

$$X'_{ij} = (X_{ij} - \min_i(X_{ij}))/(\max_i(X_{ij}) - \min_i(X_{ij})) \qquad (3)$$

where $\min_i(X_{ij})$ and $\max_i(X_{ij})$ are respectively the minimum of a column and the maximum of each column. X_{ij} is original statistical data, X'_{ij} is the data after normalization.

Correlation coefficient matrix.

$$R = \sum(X_i - \overline{X_i})(X_j - \overline{X_j})/\left[\sum(X_i - \overline{X_i})^2 \sum(X_j - \overline{X_j})^2\right]^{1/2} \qquad (4)$$

where $\overline{X_i}$ and $\overline{X_j}$ are respectively the average of X_i and X_j.

Table 1 the correlation coefficient matrix

index	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.000	-0.073	0.855	0.797	-0.269	0.604	0.605	0.824	0.663	-0.465
X2	-0.073	1.000	-0.119	-0.169	0.089	0.156	-0.180	-0.103	-0.159	0.037
X3	0.855	-0.119	1.000	0.722	-0.083	0.662	0.742	0.703	0.689	-0.422
X4	0.797	-0.169	0.722	1.000	-0.667	0.291	0.771	0.652	0.596	-0.537
X5	-0.269	0.089	-0.083	-0.667	1.000	0.203	-0.296	-0.075	0.015	0.263
X6	0.604	0.156	0.662	0.291	0.203	1.000	0.432	0.446	0.171	0.178
X7	0.605	-0.180	0.742	0.771	-0.296	0.432	1.000	0.636	0.642	-0.479
X8	0.824	-0.103	0.703	0.652	-0.075	0.446	0.636	1.000	0.706	-0.436
X9	0.663	-0.159	0.689	0.596	0.015	0.171	0.642	0.706	1.000	-0.675
X10	-0.465	0.037	-0.422	-0.537	0.263	0.178	-0.479	-0.436	-0.675	1.000

The processing of PCA.

We have calculated eigenvalues and cumulative contribution rate [5,6]. Since the cumulative contribution rate of the first four principal components up to 89.619% (> 85%), they have been able to fully reflect the comprehensive level of industrial development.

Table 2 Significant parameters

component	initial eigenvalues			extraction sums of squared loadings		
	total	% of variance	Cumulative %	total	% of variance	Cumulative %
1	5.247	52.468	52.468	5.247	52.468	52.468
2	1.643	16.431	68.899	1.643	16.431	68.899
3	1.098	10.980	79.879	1.098	10.980	79.879
4	0.974	9.740	89.619	0.974	9.740	89.619
5	0.432	4.318	93.937			
6	0.299	2.988	96.925			
7	0.188	1.877	98.802			
8	0.081	0.805	99.607			
9	0.032	0.320	99.927			
10	0.007	0.073	100.000			

Table 3 Component Matrix

index	component1	component2	component3	component4
X1	0.917	0.151	0.121	0.018
X2	-0.156	0.326	0.277	0.883
X3	0.901	0.269	0.007	-0.074
X4	0.886	-0.303	0.306	-0.024
X5	-0.303	0.700	-0.637	-0.013
X6	0.493	0.779	0.314	-0.108
X7	0.844	-0.046	0.057	-0.095
X8	0.847	0.162	-0.134	0.009
X9	0.804	-0.055	-0.491	0.115
X10	-0.611	0.470	0.383	-0.392

We have selected the load of the first three principal components to make the 3D scatter plot (Fig.1) based on Table3.

From Table 3 & Fig.1[7,8], the six factors, X_1, X_3, X_4, X_7, X_8 and X_9, have a larger load in the first principal component factor as well as can be called economic factor, which mainly reflect the overall strength, technological progress, institutional innovation, national quality, the level of consumer demand and so on; Loads in the second main component are X_5 and X_6; load in the third main component is X_{10}, can be called cultural level; larger load in the fourth main component, the X_2, can be called development potential. As a matter of fact, the contribution rate of the first four principal components were, respectively, 52.468%, 16.431%, 10.980% and 9.740%, the cumulative contribution rate of 89.619%(> 85%). Its contribution to the information ability has gradually weakened.

It can be summed up that the overall strength, the level of national quality and consumer demand for the industrial structure transfer capability play a decisive role in the process of urban industrial competitiveness. While the contribution of cultural level and the development potential factor is relatively less.

Comprehensive evaluation of industrial competitiveness in different cities.

In view of the eigenvalues of the four principal components are 5.247, 1.643, 1.098, 0.974 and the information contribution rates are, respectively, 52.468%, 16.431%, 10.980%, 9.740%, a comprehensive evaluation model will be built. The comprehensive score [9,10] can be exactly

expressed as

$$F=0.52468F_1+0.16341F_2+0.1098F_3+0.0974F_4 \tag{5}$$

Table 4 Urban industrial competitiveness score and ranking in 2013

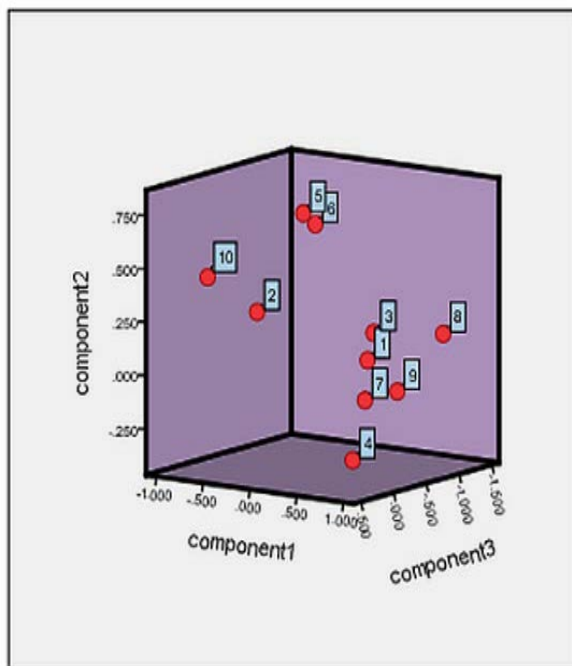| city	F1	F2	F3	F4	F	ranking
Hefei	1.00092	3.21146	0.75942	-0.41175	1.096	1
Huaibei	0.49396	-1.32522	1.04577	1.45783	0.298	5
Bozhou	-1.33053	-0.06611	-0.288	0.242	-0.717	16
Suzhou	-1.12795	0.07177	0.038	0.171	-0.559	12
Bengbu	-0.3043	-0.14637	0.000	1.307	-0.056	9
Fuyang	-1.28277	-0.06118	0.127	0.027	-0.667	15
Huainan	0.8753	-1.29075	-0.976	-1.247	0.018	7
Chuzhou	-0.4253	-0.15692	1.193	0.159	-0.102	10
Luan	-1.14306	0.19659	0.654	-0.702	-0.564	13
Maanshan	1.31796	-0.51649	-0.378	-1.499	0.419	4
Wuhu	0.77759	-0.04007	1.115	0.607	0.583	3
Xuancheng	-0.00427	0.06924	-0.174	0.395	0.029	6
Tongling	1.9489	-0.50509	0.344	0.447	1.021	2
Chizhou	-0.1958	0.23843	-0.96813	0.46387	-0.125	11
Anqing	-0.72534	-0.35429	0.31071	-2.23245	-0.622	14
Huanshan	0.12469	0.67501	-2.80298	0.81603	-0.052	8

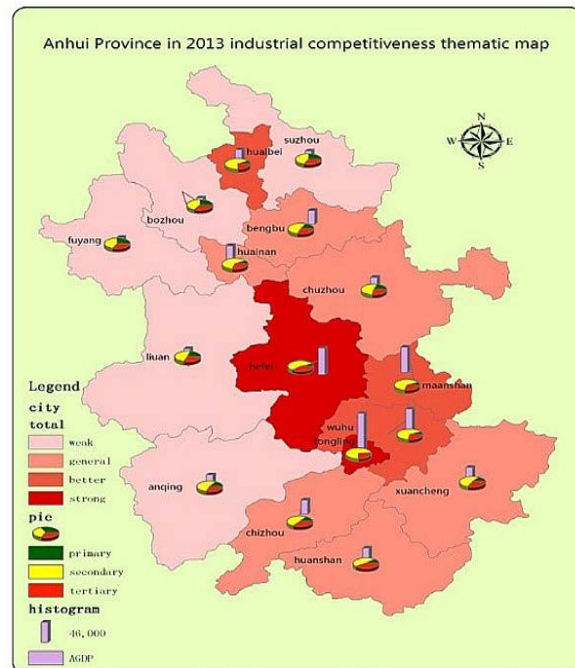
Fig.1 3D scatter diagram of components   Fig.2 Thematic map

Table 4 & Fig.2 demonstrated that there are evident regional differences in Anhui's industrial structure competitiveness. Composite score greater than zero were only seven cities, followed by Hefei, Tongling, Wuhu, Ma'anshan, Huaibei, Xuancheng and Huainan. Mainly located at Southern Anhui, these seven cities, with the industrial competitiveness apparently higher than the provincial average, have already belonging to relatively economically developed areas, while the rest are relatively low and principally gathered in North Anhui. In addition, comprehensive score is consistent with the per capita GDP ranking and total ranking of all prefecture-level cities, indicating the level of competitiveness of the industrial structure and regional economic development is positively correlated. Finally, there exist obvious differences in the contribution of each principal

component in different regions. As the decisive first principal components, the effect distribution is basically consistent with the total score (F), in other words, high scores are mainly distributed in the regions with strong industrial competitiveness, such as Hefei, Ma'anshan, Tongling, Wuhu, while low score mainly concentrated in the weak areas, such as Bozhou, Suzhou, Fuyang, Chuzhou.

Conclusions

The results of spatial geographic information variance analysis, explicitly demonstrated that Hefei, Tongling, Wuhu, Ma'anshan are in the province's top four, and the corresponding industrial output is on the largest contribution of the national economy. More specifically, Hefei has a more developed tertiary industries and more reasonable industrial structure due to its superior investment environment. As well-known industrial city, Ma'anshan and Tongling, it is the developed secondary industrial that contribute to their competitiveness in the forefront. Wuhu, located along the Yangtze River, has attracted immense foreign investment due to its excellent water transportation resource. Bozhou, Fuyang, Anqing are located, respectively, in the first, second, third from bottom. From my perspective, the main causes of its industrial structure less competitive are chiefly as follows. For one thing, their natural elements, such as wide area, large population and poor geographical location, unfortunately result in low productivity in primary industry. For another, it is their terrible investment environment that leads to secondary and tertiary industry of great weak. These two aspects together expose their weakness in urban industrial competitiveness.

In addition, comprehensive score is consistent with per capita GDP ranking, indicating the level of industrial competitiveness and resident income is positively correlated.

Finally, the contribution of each principal component exist obvious differences in different cities. As the decisive first principal components, the effect distribution is basically consistent with the total score (F), while others are lower in turn.

To sum up, the superiority of urban geographical position determines the potential of urban industrial development to a great extent, which further determines the economic competitiveness and influence of the city. It is one of the great significance to study the spatial geographic information engineering to find out the reliable rule of the regional differences.

There is no denying that this paper has only made some fundamental analysis to the industrial competitiveness in view of lacking of further precise data, to some extent, it leads to the existence of certain one-sidedness in consequence. Hence, it is particularly necessary to carry out the more in-depth theoretical analysis for more accurate results.

Acknowledgements

References

[1] G.H. Cui, Q.Q. Wei, Z.G. Chen, Regional Analysis and Planning, Higher Education Press, Beijing (1999).

[2] P.Michael, the Competitive Advantage of Nations, China press, Beijing (1997).

[3] J.H. Xu, Mathematical Methods in Contemporary Geography, Higher Education Press, Beijing (2002).

[4] Anhui Statistical Bureau, Statistical Yearbook of Anhui, China Statistics Press, Beijing: (2014).

[5] Y.G. Zhang, L.M. Wang, in: An empirical analysis of the competitiveness of cultural industry in Shandong Province: Based on principal component and location entropy. Technology and innovation management, Vol.33(2012).

[6] W. Xue, Statistical analysis and application of SPSS, Renmin University of China press, Beijing (2001).

[7] R.J. Yu, G.M. Wu, in: To Evaluate on Industrial Competitiveness of Region Based on Principal Component Analysis .Value Engineering, Vol.7(2008).

[8] S.J. Luo, J.J. Xie, Z.T. Lv, in: the time and space evolution of urban competitiveness in Anhui province .Environment and Natural Resources Research, Vol.4(2008).

[9] Z.G. Liu, Z.G. Gao, in: Evaluation of regional industrial competitiveness in Xinjiang based on principal component analysis. Journal of Xinjiang Finance & Economics Institute, Vol.1(2007) .

[10]Z.G. Gao, Y.L. Han, in: Principal Component Analysis Method in Regional Economy Research. Arid Land Geography, Vol.24(2001).