Asian Agricultural Research 2015, 7(8): 53 – 56

Comprehensive Evaluation of Economic Indicators in Guangxi Economic Zone Based on Parallel Analysis

Guangdong ZENG * , **Chunxiang ZHAO**

College of Science, Engineering University of CAPF, Xi'an 710086, China

Abstract When making comprehensive evaluation on economy of an area, it is always hard to choose evaluation indicator system and composite indicators reasonably. This article uses parallel analysis to choose the number of principal component and classify indicators by R clustering and load rotation matrix to assess the economy of various cities in the economic zone and analyze economic status of each city, to provide a scientific basis for the government functional department to make correct decisions.

Key words Principal component analysis, Parallel analysis, R clustering

1 Introduction

The assessment and prediction of economic development trend in economic zone is of important practical significance to government departments. When we conduct a comprehensive evaluation of economy in one region, it is difficult to choose evaluation indicator system and integrate the indicators. This requires a sound analysis method to determine the important comprehensive indicators. In this paper, according to Guangxi Statistical Yearbook (2013), we select 8 key indicators (land area X_1; permanent population at the end of year X_2; regional GDP X_3; social fixed asset investment X_4; public budget revenue X_5; public budget expenditure X_6; total retail sales of social consumer goods X_7; import and export X_8) to reflect economic conditions. The specific indicators are shown in Table 1.

Table 1 The major economic indicators of Guangxi's economic zones in 2013

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Nanning City	22112	679.08	2503.18	2585.18	229.72	376.51	1255.59	414678
Beihai City	3337	157.20	630.09	725.36	41.13	98.73	146.51	207820
Qinzhou City	10895	313.33	691.32	652.59	33.58	122.67	237.56	376656
Fangchenggang City	6222	88.69	443.99	550.39	35.55	74.73	71.30	489826
Yulin City	12838	558.12	1102.08	1004.26	65.57	192.65	422.83	58807
Chongzuo City	17351	201.97	530.51	532.15	39.49	130.99	84.37	713458
Baise City	36202	351.81	755.24	1000.07	56.58	214.85	156.67	50866
Hechi City	33476	341.55	492.71	277.84	22.17	175.95	176.98	52444
Liuzhou City	18617	382.45	1820.61	1683.13	113.55	221.17	661.84	311234
Guilin City	27809	483.94	1485.02	1462.40	106.01	261.33	536.35	97487
Wuzhou City	12572	292.94	832.58	858.09	73.83	159.21	257.21	121038
Wuzhou City	10602	418.68	679.18	552.24	26.57	126.24	284.05	23144
Hezhou City	11855	198.73	394.21	591.58	19.21	97.69	106.39	15589
Laibin City	13411	213.51	514.29	561.80	32.20	119.82	109.53	14321

2 Basic principles of parallel analysis and principal component analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The information in principal component refers to the variability of the variable. The greater the variability of the variables, the greater the amount of information. In the principal component analysis, standard deviation or variance is used to represent variability. Assuming $X = (X_1, \cdots, X_p)^T$ is p-dimensional random variable. The composite indicator of $X[Y_1, \cdots, Y_k(k \leqslant p)]$ is determined as follows: (i) calculating the eigenvalue of the covariance matrix of X (Σ), denoted as: $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k > 0, \lambda_{k+1} = \cdots = \lambda_p = 0$; (ii) calculating the unit eigenvector $\gamma_i (i = 1, 2, \cdots, k)$ that λ_i corresponds to, and requiring orthogonality; (iii) getting principal component i: $Y_i = \gamma_i X$, $i = 1, 2, \cdots, K$. Obviously, for any two principal components of X, it is easy to verify:

Received: May 21, 2015 Accepted: July 9, 2015

* Corresponding author. E-mail: 1140596135@qq.com

$$\begin{cases} Cov(Y_i, Y_j) = 0, i \neq j, \ j = 1, 2, \cdots, k \\ D(Y_i) = \lambda_i, \ j = 1, 2, \cdots, k \end{cases} \quad (1)$$

The weighted sum of principal components is calculated, and the new composite indicators are established. The weight of the principal components is determined based on the contribution rate of variance. $Y_1, Y_2, \cdots, Y_k$ are k principal components calculated, and their characteristic root is $\lambda_1, \lambda_2, \cdots, \lambda_p$, then the weighting coefficient of principal component i is: [4]

$$w_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \quad (2)$$

Denoting $W = (w_1, w_2, \cdots, w_k)^T$, and establishing the composite indicator as $Z = w_1 Y_1 + w_2 Y_2 + \cdots + w_k Y_k$.

Parallel analysis is the eigenvalue method based on judging the number of principal components, and it is used to determine the number of evaluation indicators[1-2]. Firstly, a random data matrix is established, with the same size as the initial matrix, and the eigenvalues of random data matrix are extracted. Secondly, based on the comparison between eigenvalue of actual data and average eigenvalue of random data matrix, if it is greater than the average eigenvalue, then this principal component can be retained, because if the eigenvalue of original data is smaller than the average eigenvalue of the simulated random matrix, it indicates that the original eigenvalue explains few variances, and it can be ignored.

3 Principal component analysis of economic indicators[4]

According to the data provided in Table 1, we use the psych package in R software to draw scree plot[3], and the average eigenvalue (dashed line) derived by 100 random data matrices and $Y = 1$ horizontal line (eigenvalue greater than 1) can be shown in Fig. 1.

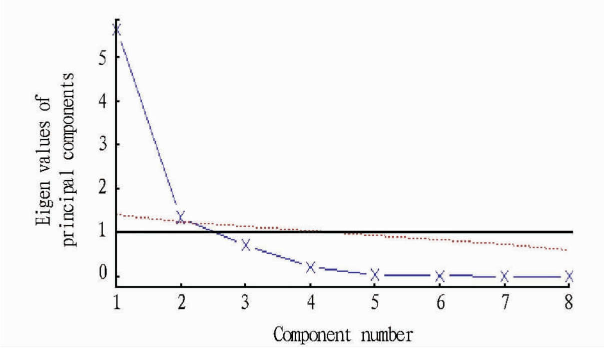

Fig.1 Scree plot of parallel analysis

As can be seen from Fig. 1, the blue line (the lines of eigenvalues connected by the original matrix) and the red line (the average eigenvalues obtained by the parallel analysis random matrix) roughly cross under the second eigenvalue. The eigenvalues of the two principal components prior to selection are also greater than 1, and the variance contribution rate of the first two eigenvalues reaches 87.3%, which verifies the effectiveness of parallel analysis in determining the number of principal component. Using principal component analysis[5], we extract the principal component of economic indicators, determine the weight of each principal component according to the variance contribution rate of each principal component, and establish the composite scores as comprehensive economic evaluation indicators. Using R software, the corresponding results obtained based on Table 1 are shown in Table 2 and 3.

Table 2 Correlation coefficient matrix

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1.000	0.349	0.252	0.245	0.257	0.580	0.215	-0.140
X_2	0.349	1.000	0.781	0.710	0.694	0.842	0.818	-0.302
X_3	0.252	0.781	1.000	0.977	0.961	0.908	0.977	0.031
X_4	0.245	0.710	0.977	1.000	0.976	0.895	0.950	0.044
X_5	0.257	0.694	0.961	0.976	1.000	0.907	0.953	0.133
X_6	0.580	0.842	0.908	0.895	0.907	1.000	0.900	-0.065
X_7	0.215	0.818	0.977	0.950	0.953	0.900	1.000	-0.003
X_8	-0.140	-0.302	0.031	0.044	0.133	-0.065	-0.003	1.000

Table 3 Eigenvalues and eigenvectors

Eigenvalues	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7	Vector 8
5.628	-0.176	0.571	0.725	-0.115	-0.170	-0.093	-0.116	-0.230
1.353	-0.365	0.244	-0.188	0.800	0.042	0.277	-0.056	-0.224
0.723	-0.411	-0.118	-0.098	-0.120	-0.645	0.011	0.610	-0.064
0.222	-0.405	-0.141	-0.084	-0.380	-0.151	0.577	-0.555	0.042
0.036	-0.406	-0.171	-0.019	-0.274	0.605	-0.100	0.220	-0.553
0.027	-0.410	0.152	0.156	-0.006	0.378	0.057	0.255	0.757
0.008	-0.410	-0.111	-0.164	0.077	-0.143	-0.752	-0.435	0.110
0.001	-0.043	-0.718	0.608	0.327	-0.026	0.058	-0.034	0.018

The two principal components are expressed as follows:

$$Y_1 = -0.17575X_1 + 0.570979X_2 + 0.725351X_3 - 0.11525X_4 - 0.1696X_5 - 0.0931X_6 - 0.11599X_7 - 0.22978X_8$$

$$Y_2 = -0.36486X_1 + 0.244459X_2 - 0.18824X_3 + 0.799997X_4 + 0.04201X_5 + 0.276858X_6 - 0.05574X_7 - 0.22398X_8$$

Table 4 Loading matrix of principal component

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8
X_1	-0.176	0.571	0.725	-0.115	-0.170	0.093	-0.116	0.230
X_2	-0.365	0.244	-0.188	0.800	0.042	-0.277	-0.056	0.224
X_3	-0.411	-0.118	-0.098	-0.120	-0.645	-0.011	0.610	0.064
X_4	-0.405	-0.141	-0.084	-0.380	-0.151	-0.577	-0.555	-0.042
X_5	-0.406	-0.171	-0.019	-0.274	0.605	0.100	0.220	0.553
X_6	-0.410	0.152	0.156	-0.006	0.378	-0.057	0.255	-0.757
X_7	-0.410	-0.111	-0.164	0.077	-0.143	0.752	-0.435	-0.110
X_8	-0.043	-0.718	0.608	0.327	-0.026	-0.058	-0.034	-0.018

It can be found that the information hidden in the indicators in the principal component loading matrix is not clear enough, so it is necessary to conduct principal component rotation. The orthogonal rotation in principal component rotation is to conduct column denoising of loading matrix, so that several limited variables can explain principal component, that is, the principal components remain irrelevant. The loading matrix treated is RC, as shown in Table 5.

Under the rotation of principal component, the cumulative variance explanation of the first two principal components after rotation does not change, and the explanation degree of each variance changes. The principal component scores are obtained by the linear combination of the original indicators, and the weight is just the load factor:

$$C_j = a_{j1}x_1 + a_{j2}x_2 + x_{j3}x_3 + \cdots + a_{jp}x_p \text{, where } j = 1, 2, \cdots, m \quad (3)$$

Composite score is the comprehensive evaluation function obtained by linear combination of variance contribution rate of principal components:

$$C = \frac{\lambda_1 C_1 + \lambda_2 C_2 + \cdots + \lambda_m C_m}{\lambda_1 + \lambda_2 + \cdots + \lambda_m} = \sum_{i=1}^{m} \omega_i C_i \quad (4)$$

Table 5 Rotated loading matrix

	RC_1	RC_2	h_1	h_2
x_1	0.32	0.71	0.61	0.385
x_2	0.82	0.40	0.83	0.170
x_3	0.99	0.00	0.97	0.028
x_4	0.97	-0.03	0.95	0.050
x_5	0.98	-0.07	0.97	0.031
x_6	0.94	0.31	0.98	0.024
x_7	0.98	0.00	0.96	0.039
x_8	0.22	-0.81	0.71	0.292

Table 6 Composite score

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8
Nanning City	-6.755	-1.014	-0.056	-0.006	0.261	0.155	-0.096	-0.008
Beihai City	1.787	-0.988	-0.808	-0.470	0.046	-0.029	0.046	-0.080
Qinzhou City	1.066	-0.822	0.079	0.589	-0.127	-0.017	-0.077	-0.020
Fangchenggang City	2.352	-1.809	0.363	-0.253	0.043	0.042	-0.033	0.065
Yulin City	-0.857	0.569	-1.072	0.884	0.020	-0.238	0.056	0.000
Chongzuo City	1.443	-1.643	1.826	0.509	0.068	-0.096	0.043	-0.025
Baise City	-0.266	1.928	1.197	-0.414	0.055	-0.309	-0.100	0.025
Hechi City	0.926	1.992	1.074	0.272	-0.012	0.409	0.050	-0.012
Liuzhou City	-2.404	-0.706	0.013	-0.459	-0.554	0.028	0.062	0.025
Guilin City	-2.371	1.008	0.217	-0.173	-0.073	-0.126	0.087	-0.034
Wuzhou City	0.373	-0.017	-0.512	-0.298	0.283	-0.003	0.160	0.062
Wuzhou City	0.950	0.573	-1.093	0.692	-0.092	0.087	-0.054	0.036
Hezhou City	2.051	0.411	-0.675	-0.427	-0.012	0.033	-0.196	-0.012
Laibin City	1.706	0.517	-0.553	-0.445	0.093	0.065	0.053	-0.023

For preliminary classification of indicators and further analysis, it is necessary to introduce clustering analysis. R clustering in clustering analysis is a method for clustering of variables. Based on R clustering, we perform the clustering on eight indicators (variables). As can be seen from Fig. 2, indicators can be divided into two categories: X_1 and X_8 as a category; X_2, X_3, X_4, X_5, X_6 and X_7 as a category. According to the loading matrix by principal component rotation, it can be found that principal component has a large load on regional GDP (X_3), social fixed asset investment (X_4), public budget revenue (X_5), public budget expenditure (X_6), and total retail sales of social consumer goods (X_7). This factor mainly reflects the financial and capital flow. Principal component has a large load on land area (X_1), permanent population at the end of year (X_2) and import and export (X_8). This factor mainly includes hardware and import and export. The classification derived from two principal components is very similar to the classification derived from R clustering, so based on composite scores, we can evaluate the economic situation of 14 cities in Guangxi.

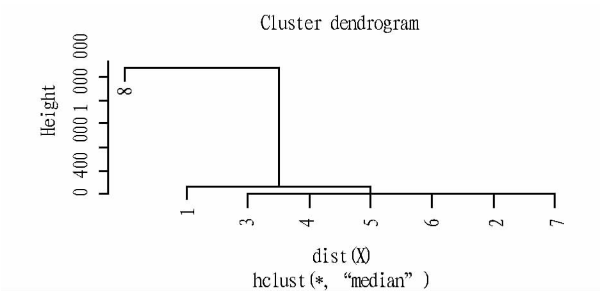

Fig. 2 R clustering

4 Comprehensive evaluation[5]

The cities with high scores on principal component of liquidity and financial balance (Y_1) include Nanning, Liuzhou, Guilin and Fangchenggang. The absolute value of Nanning is higher than that of other cities, indicating that the regional GDP, social fixed asset investment, public budget revenue and total retail sales of social consumer goods in Nanning are much higher than in other cities, because Nanning is a provincial capital, a core city of Beibu Gulf Economic Zone, and also a financial and trading center. Similarly, Liuzhou is an industrial city and an old industrial area in Guangxi. Guilin is a tourist city and Fangchenggang is an important port city. These cities have an advantage in finance. The cities with high scores on principal component of hardware, import and export include Hechi, Baise, Fangchenggang and Chongzuo. Baise and Hechi are ranked first in terms of prefecture-level city area, and Fangchenggang is ranked first in terms of import and export. From the composite score, the top three are still Nanning, Liuzhou and Guilin. Although Hechi and Baise have high scores on the second principal component, Y_1 which represents economic strength still affects ranking.

References

[1] KONG M, BIAN R, ZHANG HC. The application of parallel analysis in exploratory factor analysis[J]. Psychological Science, 2007, 30(4): 924 – 925. (in Chinese).
[2] MU SK, GU HG. The comparison of factor retaining methods in exploratory factor analysis[J]. Exploration of Psychology, 2011, 31(5), 477 – 480. (in Chinese).
[3] Robert I. Kabacoff. R in action: Data analysis and graphics with R[M]. Beijing: Posts & Telecom Press, 2013.
[4] JIANG Y. The economic development evaluation of the boroughs of Tianjin [J]. Application of Statistics and Management, 2002, 21(1): 4 – 9. (in Chinese).
[5] WANG BH. Multivariate statistical analysis and modeling for R language [M]. Guangzhou: Jinan University Press, 2014. (in Chinese).

(From page 52)

[17] ZHAO SJ. Study on the countermeasure and suggestion on developing eco – tourism in Jilin City[D]. Chinese Academy of Agricultural Sciences, 2007. (in Chinese).
[18] WANG SY. A brief analysis on the negative effects of tourism rebate on the industry[J]. Charming China, 2010(17): 26. (in Chinese).
[19] LIU Q, LIU HT, LI R. A brief analysis on the development of tourism in China and its prospect[J]. Group Economics Research, 2007(5Z): 115. (in Chinese).
[20] QU WM. Problems in tourist economy in China and the countermeasures [J]. Economic Research Guide, 2008(6): 157 – 158. (in Chinese).
[21] ZHANG M. The present situation and countermeasures of ecotourism in Linzhi Prefecture, Tibet[J]. Special Zone Economy, 2013(8): 99 – 101. (in Chinese).
[22] ZHU XQ. On situation and development of tourism infrastructure in Nyingchi[J]. Journal of Xi'an University of Arts and Science (Social Sciences Edition), 2013, 16(5): 65 – 68. (in Chinese).
[23] LI LF. Five key problems promoting the development of tourism development[J]. Management Forum, 2009(11): 29 – 41. (in Chinese).
[24] LIU YH, ZONG G, YANG T. Study on the influence of Tibet tourism development on the income of urban and rural residents[J]. China Tibetology, 2014(1): 25 – 30. (in Chinese).
[25] LIU YH. The contribution of Tibet tourist industry and its regional difference[J]. China Tibetology, 2012(4): 78 – 84. (in Chinese).
[26] HU B, QIAO J. An empirical analysis on influence factors and the judgement of regional difference of rural income[J]. Social Science Research, 2005(5): 76 – 80. (in Chinese).