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

<table>
<thead>
<tr>
<th>City</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanning City</td>
<td>22112</td>
<td>679.08</td>
<td>2503.18</td>
<td>2585.18</td>
<td>229.72</td>
<td>376.51</td>
<td>1255.59</td>
<td>414678</td>
</tr>
<tr>
<td>Beihai City</td>
<td>3337</td>
<td>157.20</td>
<td>630.09</td>
<td>725.36</td>
<td>41.13</td>
<td>98.73</td>
<td>146.51</td>
<td>207820</td>
</tr>
<tr>
<td>Qinzhou City</td>
<td>10895</td>
<td>313.33</td>
<td>691.32</td>
<td>652.59</td>
<td>33.58</td>
<td>122.67</td>
<td>237.56</td>
<td>376656</td>
</tr>
<tr>
<td>Fangchenggang City</td>
<td>6222</td>
<td>88.69</td>
<td>443.99</td>
<td>550.39</td>
<td>35.55</td>
<td>74.73</td>
<td>71.30</td>
<td>498926</td>
</tr>
<tr>
<td>Yulin City</td>
<td>12838</td>
<td>558.12</td>
<td>1102.08</td>
<td>1004.26</td>
<td>65.57</td>
<td>192.65</td>
<td>422.83</td>
<td>58807</td>
</tr>
<tr>
<td>Chongzuo City</td>
<td>17351</td>
<td>201.97</td>
<td>530.51</td>
<td>532.15</td>
<td>39.49</td>
<td>130.99</td>
<td>84.37</td>
<td>713458</td>
</tr>
<tr>
<td>Baize City</td>
<td>36202</td>
<td>351.81</td>
<td>755.24</td>
<td>1000.07</td>
<td>56.58</td>
<td>214.85</td>
<td>156.67</td>
<td>58066</td>
</tr>
<tr>
<td>Hechi City</td>
<td>33476</td>
<td>341.55</td>
<td>492.71</td>
<td>277.84</td>
<td>22.17</td>
<td>175.95</td>
<td>176.98</td>
<td>52444</td>
</tr>
<tr>
<td>Liuzhou City</td>
<td>18617</td>
<td>382.45</td>
<td>1820.61</td>
<td>1683.13</td>
<td>113.55</td>
<td>221.17</td>
<td>661.84</td>
<td>311234</td>
</tr>
<tr>
<td>Guilin City</td>
<td>27809</td>
<td>483.94</td>
<td>1485.02</td>
<td>1462.40</td>
<td>106.01</td>
<td>261.33</td>
<td>536.35</td>
<td>97487</td>
</tr>
<tr>
<td>Wuzhou City</td>
<td>12572</td>
<td>292.94</td>
<td>832.58</td>
<td>858.09</td>
<td>73.83</td>
<td>159.21</td>
<td>257.21</td>
<td>121038</td>
</tr>
<tr>
<td>Wuzhou City</td>
<td>10602</td>
<td>418.68</td>
<td>679.18</td>
<td>552.24</td>
<td>26.57</td>
<td>126.24</td>
<td>284.05</td>
<td>23144</td>
</tr>
<tr>
<td>Hezhou City</td>
<td>11855</td>
<td>198.73</td>
<td>394.21</td>
<td>591.58</td>
<td>19.21</td>
<td>97.69</td>
<td>106.39</td>
<td>15589</td>
</tr>
<tr>
<td>Laihin City</td>
<td>13411</td>
<td>213.51</td>
<td>514.29</td>
<td>561.80</td>
<td>32.20</td>
<td>119.82</td>
<td>109.53</td>
<td>14321</td>
</tr>
</tbody>
</table>

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 $X = (X_1, \ldots, X_p)^T$ is $p$-dimensional random variable. The composite indicator of $X\{Y_1, \ldots, Y_k (k \leq p)\}$ is determined as follows; (i) calculating the eigenvalue of the covariance matrix of $X\{\Sigma\}$, denoted as: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0, \lambda_{k+1} = \ldots = \lambda_p = 0$; (ii) calculating the unit eigenvector $\gamma_i (i = 1, 2, \ldots, k)$ that $\lambda_i$ corresponds to, and requiring orthogonality; (iii) getting principal component $i; Y_i = \gamma_i X, \ i = 1, 2, \ldots, K$. Obviously, for any two principal components of $X$, it is easy to verify.
\[
\begin{align*}
\text{Cov}(Y_j, Y_k) &= 0, i \neq j, j = 1, 2, \ldots, k \\
D(Y_j) &= \lambda_i, j = 1, 2, \ldots, k
\end{align*}
\] (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. \(Y_1, Y_2, \ldots, Y_k\) are \(k\) principal components calculated, and their characteristic root is \(\lambda_1, \lambda_2, \ldots, \lambda_k\), then the weighting coefficient of principal component \(i\) is:

\[
w_i = \frac{\lambda_i}{\sum_j \lambda_j}
\] (2)

Denoting \(W = (w_1, w_2, \ldots, w_k)^T\), and establishing the composite indicator as \(Z = w_1Y_1 + w_2Y_2 + \ldots + w_kY_k\).

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\(^{[3]}\)

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.

![Scree plot of parallel analysis](image)

**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>
<thead>
<tr>
<th>Table 2</th>
<th>Correlation coefficient matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>(X_2)</td>
</tr>
<tr>
<td>1.000</td>
<td>0.349</td>
</tr>
<tr>
<td>0.349</td>
<td>1.000</td>
</tr>
<tr>
<td>0.252</td>
<td>0.781</td>
</tr>
<tr>
<td>0.245</td>
<td>0.710</td>
</tr>
<tr>
<td>0.257</td>
<td>0.694</td>
</tr>
<tr>
<td>0.580</td>
<td>0.842</td>
</tr>
<tr>
<td>0.215</td>
<td>0.818</td>
</tr>
<tr>
<td>-0.140</td>
<td>-0.302</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Eigenvalues and eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalues</td>
<td>Vector 1</td>
</tr>
<tr>
<td>5.628</td>
<td>-0.176</td>
</tr>
<tr>
<td>1.353</td>
<td>-0.365</td>
</tr>
<tr>
<td>0.723</td>
<td>-0.411</td>
</tr>
<tr>
<td>0.222</td>
<td>-0.405</td>
</tr>
<tr>
<td>0.036</td>
<td>-0.406</td>
</tr>
<tr>
<td>0.027</td>
<td>-0.410</td>
</tr>
<tr>
<td>0.008</td>
<td>-0.410</td>
</tr>
<tr>
<td>0.001</td>
<td>-0.043</td>
</tr>
</tbody>
</table>
The two principal components are expressed as follows:
\[ Y_i = -0.17575X_1 + 0.570979X_2 + 0.725351X_3 - 0.11525X_4 - 0.1696X_5 - 0.0931X_6 - 0.11559X_7 - 0.22978X_8 \]

Table 4 Loading matrix of principal component

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

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,

Table 5 Rotated loading matrix

<table>
<thead>
<tr>
<th>RC</th>
<th>RC</th>
<th>h1</th>
<th>h2</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0.32</td>
<td>0.71</td>
<td>0.61</td>
</tr>
<tr>
<td>x2</td>
<td>0.82</td>
<td>0.40</td>
<td>0.83</td>
</tr>
<tr>
<td>x3</td>
<td>2.99</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>x4</td>
<td>2.97</td>
<td>-0.33</td>
<td>0.95</td>
</tr>
<tr>
<td>x5</td>
<td>2.98</td>
<td>-0.07</td>
<td>0.97</td>
</tr>
<tr>
<td>x6</td>
<td>2.94</td>
<td>0.31</td>
<td>0.98</td>
</tr>
<tr>
<td>x7</td>
<td>2.98</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>x8</td>
<td>2.22</td>
<td>-0.81</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 6 Composite score

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

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

\[ C = \sum_{i=1}^{p} \omega_i C_i \]

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 \( X_6 \) as a category; \( X_2, X_3, X_4, X_5, X_7 \) and \( X_8 \) 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 \( (X_1) \), social fixed asset invest-
ment \( (X_i) \), public budget revenue \( (X_4) \), public budget expenditure \( (X_3) \), and total retail sales of social consumer goods \( (X_5) \). 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 \( (X_6) \). 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.

![Cluster dendrogram](image)

**Fig. 2 R clustering**

4 Comprehensive evaluation

The cities with high scores on principal component of liquidity and financial balance \( (Y_i) \) 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, Baize, 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 Baize have high scores on the second principal component, \( Y_i \) which represents economic strength still affects ranking.

### References


### About KIT

The Royal Tropical Institute (KIT) in Amsterdam is an independent centre of knowledge and expertise in the areas of international and intercultural cooperation, operating at the interface between theory and practice and between policy and implementation. The Institute contributes to sustainable development, poverty alleviation and cultural preservation and exchange.