Factor Analysis and Clustering-Based Empirical Study on Regional Construction Industry Development in China

Ya-yun Dang and Xing Bi

Abstract The purpose of the study is to provide suggestions for policy-makers and industry practitioners aiming at improving the construction industry development. This paper constructs a complete evaluation index system of regional construction industry development, and then applies factor analysis and clustering to analyze and evaluate the development level of construction industry of 31 regions in China by using SPSS 18.0. These 31 regions are categorized into five clusters by four extracted factors, namely total factor, efficiency and technology factor, per capita factor and profitability factor. The results show that significant differences exist in development level of construction industry among different regions.

Keywords Clustering • Development • Factor analysis • Regional construction industry

1 Introduction

Construction industry is one of the pillar industries in China. Evaluation on construction industry development can reflect the differences of regional construction industry and can guide the market to allocate resources efficiently, thereby improving the overall competitiveness of China's construction industry. There are two types of evaluation methods (Wang Jia-yuan and Yuan Hong-ping 2007), namely subjective method and objective method. The former determines the weight of each evaluation index by experts' subjective judgment according to their own knowledge and experience, such as analytic hierarchy process, fuzzy comprehensive evaluation, etc.; the latter determines the weight according to the objective relationship between

---

Y. Dang (✉) • X. Bi
College of Management and Economics, Tianjin University, Tianjin, China
e-mail: yayunparty@gmail.com; bistar@126.com

E. Qi et al. (eds.), *Proceedings of 20th International Conference on Industrial Engineering and Engineering Management*, DOI 10.1007/978-3-642-40072-8_44,

the indexes, such as DEA (data envelopment analysis), principal component analysis, etc. (Xue et al. 2008; Taewoo Youa and Hongmin Zib 2007; Tsolas 2011; Ruan Lian-fa and Zhang Yue-wei 2009; Deng Rong-hui and Xia Qing-dong 2006; Kang Xue-zeng and Meng Gang 2008). However, there are some deficiencies of these methods: due to the limitations of experts' knowledge and experience, differences exist between expert's weights and actual situation, which influences evaluation results; information overlap or high correlation between indexes makes the results not tally with actual situation.

Some scholars apply cluster and factor analysis to the research of the sustainable development, growth levels of construction industry, having achieved valuable results (Wang Lei et al. 2006; Wang Xue-qing et al. 2011; Kale and Arditi 2002; Wang Wen-xiong and Li Qi-ming 2008; Zhou Jian-hua and Yuan Hong-ping 2007). However, the indexes selected are incomplete. Therefore, on the basis of widely collecting and sorting the existing evaluation index, this paper proposes an evaluation index system, which can reflect the construction industry development and is also suitable for factor analysis and clustering. By adopting the data of China Statistical Yearbook 2011, the paper evaluates the construction industry development of 31 regions in China and categorizes them based on regional similarity through cluster analysis.

2 Methodology

2.1 Factor Analysis

The purpose of factor analysis is to describe the covariance relationships among observed and correlated variables in terms of a few underlying but unobserved random quantity variables called factors. In other words, it is possible that variations in three or four observed variables mainly reflect the variations in fewer unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modeled as linear combinations of the potential factors, plus "error" terms and the factor model is motivated by the hypothesis that variables can be grouped by their correlations (DeCoster 1998; Factor Analysis 2013).

2.2 Hierarchical Clustering

Cluster analysis is a task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis (Du Gang 2003).

Table 1 Evaluation index system

First-level	Second-level
Industry scale	X1 Number of enterprises (unit)
	X2 Number of employed persons (person)
	X3 Total assets (10,000 yuan)
	X4 Gross output value (10,000 yuan)
	X5 Value added of the construction industry (10,000 yuan)
Per capita level	X6 Per capita GDP (10,000 yuan/person)
	X7 Per capita profit (10,000 yuan/person)
Profitability	X8 Rate of return on common stockholders' equity (%)
Performance	X9 Total tax (10,000 yuan)
	X10 The proportion of employment (%)
	X11 Overall labor productivity in terms of gross output value (yuan/person)
Technology	X12 Total number of machinery and equipment owned (set)
	X13 Total power of machinery and equipment owned (10,000 kw)
	X14 Net value of machinery and equipment owned (10,000 yuan)
	X15 Value of machines per laborer (yuan/person)
	X16 Power of machines per laborer (kw/person)

Hierarchical clustering is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will be formed, which can be represented by a dendrogram.

3 Evaluation Index System

According to the principle of purposefulness, scientificalness, integrity, and operability and base on relevant researches, the paper designs an evaluation index system (see Table 1) to reflect construction industry development. The index system includes two hierarchies: first-level indexes and second-level indexes. First-level indexes contain six indexes, which are six aspects of construction industry development; Second-level indexes consist of 16 basic indexes.

4 Factor Analysis & Results

Based on the evaluation index system, the paper analyzes and evaluates construction industry development of 31 provinces in China by using SPSS 18.0 with data collected from China Statistical Yearbook (2011).

Table 2 KMO test results and Bartlett test results

Kaiser-Meyer-Olkin measure of sampling adequacy		0.721
Bartlett's Test of Sphericity	Approx. Chi-Square	817.557
	df	120
	Sig.	.000

Table 3 Characteristic value and contribution rate of common factors

Common factor	F_1	F_2	F_3	F_4
Characteristic value	8.525	2.424	2.155	1.095
Contribution rate/%	53.281	15.148	13.471	6.845
Accumulative contribution rate/%	53.281	68.429	81.901	88.746

4.1 KMO Test and Bartlett's Test of Sphericity

KMO Test measures whether the samplings are enough for factor analysis and whether the partial correlation coefficient between the variables is too small. Bartlett's Test of Sphericity tests whether correlation coefficient matrix is a unit matrix. If it is a unit matrix, it is not suitable for adopting factor model (Table 2).

Kaiser gave the KMO Test standard about whether it is suitable for factor analysis: KMO > 0.9, quite suitable; 0.9 > KMO > 0.8, suitable; 0.8 > KMO > 0.7, generally suitable; 0.7 > KMO > 0.6, not quite suitable; KMO < 0.5, not suitable. SPSS results show that the variables have passed the KMO Test passes. And Bartlett's Test of Sphericity = 817.557; significance = .000, which means that the variables have passed Bartlett's Test of Sphericity. So the variables that the paper selects are suitable for factor analysis.

4.2 Factor Analysis Process and Results

In the process of factor analysis, the paper extracts four common factors by principal components method. Then by using Quartimax method to rotate the factor load matrix, we can obtain the factors' scree plot (see Fig. 1), characteristic value and contribution rate (see Table 3), and rotated component matrix (see Table 4).

Table 3 shows that the accumulative contribution rate of four extracted common factors is 88.745 %, which is bigger than 85 %, i.e., the extraction of common factor is effective. The original 16 indexes can be integrated into four common factors: F_1, F_2, F_3 and F_4. According to the principle of factor analysis, the four common factors have no correlation with each other, but each common factor is highly correlated with its own contained original variables.

Table 4 shows the correlation coefficient between common factors and their own contained original variables. The first common factor F_1 has a large load in Number of Enterprises (X_1), Number of Employed Persons (X_2), Total Assets (X_3), Gross

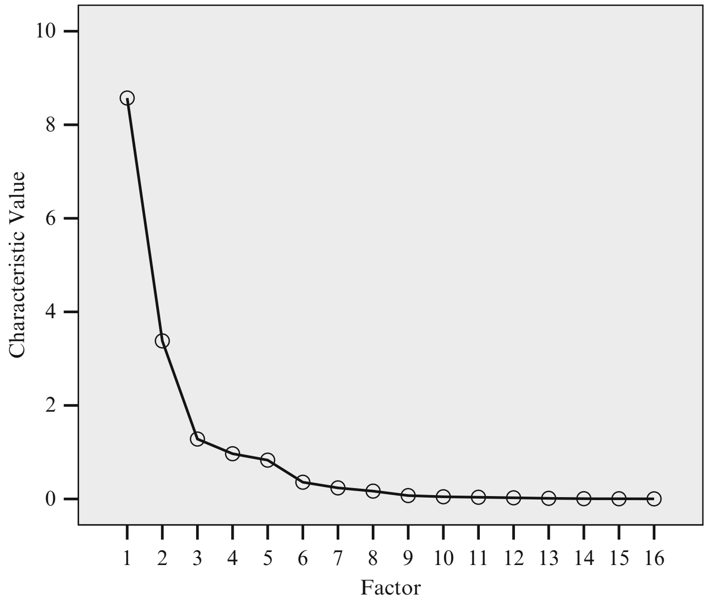

Fig. 1 Scree plot

Table 4 Rotated component matrix

Variable	Common factor F_1	F_2	F_3	F_4
X_1	0.939	−0.071	0.034	0.076
X_2	0.946	−0.211	−0.168	0.031
X_3	0.822	0.031	0.484	−0.153
X_4	0.984	−0.057	0.094	−0.033
X_5	0.977	−0.130	−0.051	0.059
X_6	−0.064	0.486	0.798	−0.230
X_7	−0.013	0.250	0.908	0.226
X_8	0.278	−0.189	0.018	0.902
X_9	0.978	−0.067	0.102	0.001
X_{10}	0.791	−0.021	0.172	−0.082
X_{11}	0.174	0.597	0.378	−0.303
X_{12}	0.772	0.234	−0.368	0.061
X_{13}	0.939	0.070	−0.186	0.156
X_{14}	0.973	0.125	−0.076	0.077
X_{15}	−0.116	0.882	0.242	−0.075
X_{16}	−0.136	0.896	0.031	0.059

Output Value (X_4), Value Added of the Construction Industry (X_5), Total Tax (X_9), The Proportion of Employment (X_{10}), Total Number of Machinery and Equipment Owned (X_{12}), Total Power of Machinery and Equipment Owned (X_{13}) and Net Value of Machinery and Equipment Owned (X_{14}). These ten indexes reflect the scale, economic and social benefits, equipment and assets of regional construction industry, so F_1 can be denominated Total Factor.

The second common factor has a large load in Overall Labor Productivity In Terms of Gross Output Value (X_{11}), Value of Machines per Laborer (X_{15}) and Power of Machines per Laborer (X_{16}). These three indexes reflect the labor productivity and technological level, so F_2 can be denominated Productivity and Technology Factor.

The third common factor has a large load in Per capita GDP (X_6) and Per capita Profit (X_7), both of which reflect the Per capita level. So F_3 can be denominated Per capita Factor.

The fourth common factor has a large load in Rate of Return on Common Stockholders' Equity (X_8), which reflects profitability of construction industry in different regions. So F_4 can be Profitability Factor.

As a result, it is suitable to use Total Factor (F_1), Productivity and Technology factor (F_2), Per capita Factor (F_3) and Profitability Factor (F_4) to represent the original variables and evaluate regional construction industry development.

By using SPSS 18.0, it is easy to obtain the scores and rankings of each common factor of 31 regions. Set contribution rates of each common factor as weight and conduct linear weighted summation to obtain comprehensive scores and rankings (see Table 5). The calculation formula of comprehensive scores is as follows:

$$F = 0.5328 \times F_1 + 0.1515 \times F_2 + 0.1347 \times F_3 + 0.0685 \times F_4 \tag{1}$$

5 Clustering & Results

Take the Total factor, Productivity and Technology Factor, Per capita Factor and Profitability Factor as independent variables for cluster analysis and adopt method of between-groups linkage and measure of squared Euclidean distance to conduct hierarchical cluster analysis to generate Dendrogram (see Fig. 2).

6 Discussion

From comprehensive score and ranking Table 5 and clustering, 31 regions can be categorized into five clusters.

The first cluster includes Beijing and Shanghai. The respective comprehensive rankings of these two regions are 3rd and 7th, with Total Factor 10th and 9th, Productivity and Technology Factor 9th and 16th, Per capita Factor 1st and 2nd, but the respective rankings of Profitability Factor are 30th and 27th, which have an obvious gap with former factors. Therefore, it can be categorized as: upper-middle scale, medium productivity and technology, high per capita and low profitability.

Table 5 Factor analysis results and clustering result of 31 regions

Region	F_1	Ranking	F_2	Ranking	F_3	Ranking	F_4	Ranking	F	Comprehensive ranking	Category
Jiangsu	3.38454	1	0.30451	8	−0.61225	25	0.59196	11	1.80744	1	2
Zhejiang	2.53706	2	−0.93329	26	−0.00682	11	−0.83078	26	1.15261	2	2
Beijing	0.29112	10	0.28406	9	4.05507	1	−1.40912	30	0.64798	3	1
Shanghai	0.36707	9	−0.05795	16	1.94079	2	−0.90358	27	0.38642	7	1
Tianjin	−0.03938	15	3.00921	1	0.71542	5	−0.5748	21	0.49197	6	3
Inner Mongolia	−0.78362	25	−0.46372	22	1.01052	3	2.94336	1	−0.15032	16	4
Tibet	−1.12967	31	0.03901	14	0.97046	4	2.00191	2	−0.32833	22	4
Liaoning	0.56047	6	−0.08053	17	0.21425	8	0.6265	10	0.35813	9	5
Shandong	1.143	3	−0.29276	18	−0.38631	21	1.03014	3	0.58307	4	5
Hubei	0.54115	7	1.30249	5	−0.12495	12	0.76541	5	0.52118	5	5
Guangdong	0.68726	4	0.05858	13	0.17083	10	−0.35474	17	0.37379	8	5
Henan	0.67475	5	0.163	11	−0.81284	29	0.44696	12	0.30528	10	5
Hebei	0.21563	11	1.44212	3	−1.0044	30	0.73982	7	0.24868	11	5
Hunan	0.01576	14	0.09016	12	−0.58511	24	0.78936	4	−0.00277	12	5
Shaanxi	−0.12515	16	0.16529	10	−0.1617	13	−0.65898	22	−0.10849	13	5
Shanxi	−0.30732	18	0.58469	7	−0.3729	20	−0.52528	19	−0.16132	18	5
Xinjiang	−0.83682	27	1.36558	4	−0.27757	15	−0.11028	16	−0.2839	21	5
Qinghai	−0.69874	23	1.79	2	−1.31179	31	−0.95352	29	−0.34302	23	5
Ningxia	−1.03269	29	0.59314	6	−0.28114	16	−0.40043	18	−0.52562	27	5
Heilongjiang	−0.43758	19	0.03453	15	0.17537	9	0.64709	9	−0.16003	17	5
Chongqing	−0.15587	17	−1.31639	29	0.32491	7	0.76306	6	−0.18652	19	5
Jilin	−0.76165	24	−0.60312	25	0.50495	6	0.71074	8	−0.38055	24	5
Sichuan	0.39812	8	−1.53913	31	−0.32963	19	−0.67853	23	−0.11187	14	5
Anhui	0.02348	13	−0.4327	20	−0.52445	23	−0.03516	15	−0.12609	15	5
Fujian	0.09295	12	−1.13986	28	−0.44119	22	−0.80451	16	−0.23762	20	5
Jiangxi	−0.5646	21	−0.95296	27	−0.1685	14	−0.00593	14	−0.4683	25	5
Yunnan	−0.54731	20	−0.58027	24	−0.29882	18	−0.78121	23	−0.4732	26	5
Gansu	−0.82707	26	−0.43852	21	−0.63827	26	0.28323	13	−0.5737	28	5
Guangxi	−0.69726	22	−0.35711	19	−0.64669	27	−0.94365	28	−0.57725	29	5
Guizhou	−0.90287	28	−0.57797	23	−0.80489	28	−1.82726	31	−0.80202	30	5
Hainan	−1.08479	30	−1.4601	30	−0.29235	17	−0.54179	20	−0.87562	31	5

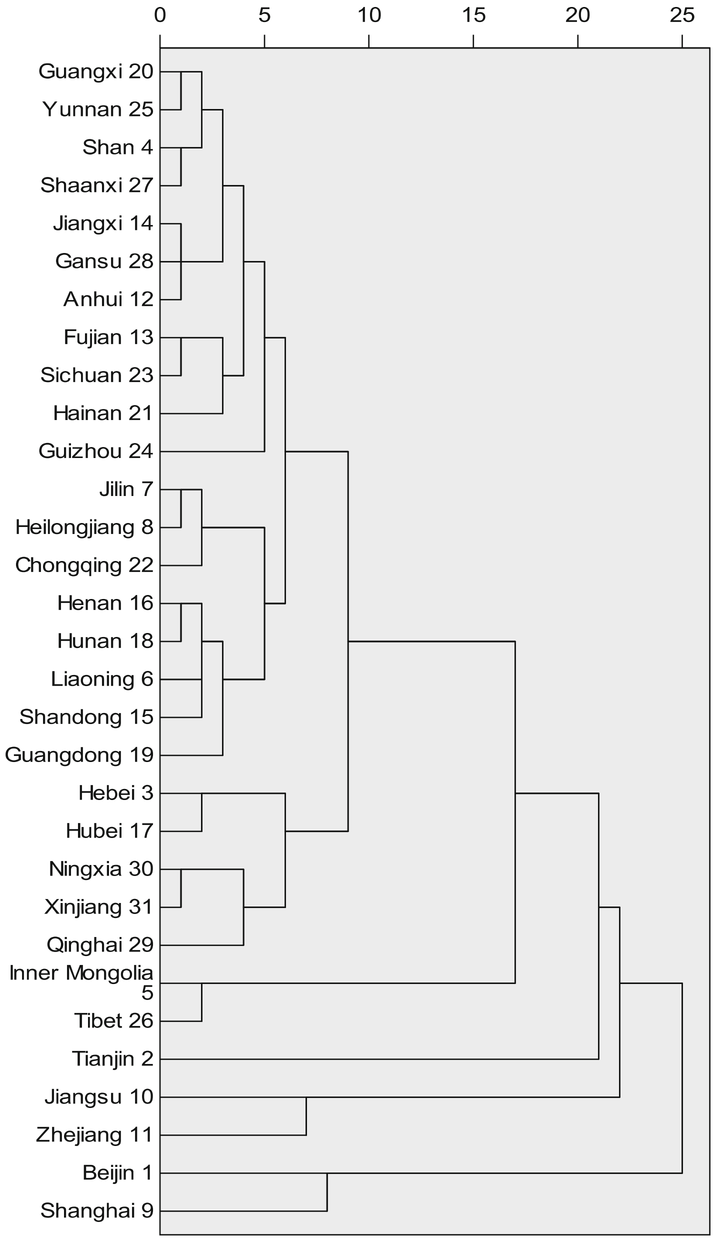

Fig. 2 Cluster genealogy chart

The second cluster includes Jiangsu and Zhejiang. The respective comprehensive rankings of these two regions are 1st and 2nd, and Total Factor also 1st and 2nd. But the other three common factors all rank low. It can be concluded that Total Factor has a huge impact on comprehensive ranking. The scores of Total Factor are 3.38454 and 2.53706 respectively, and much higher than Shandong' score of 1.143, which ranks 3rd. These two regions can be categorized as: large scale, medium-lower productivity and technology, medium-lower per capita and profitability.

Tianjin is a special region and can be categorized by itself. Its score of Total Factor ranks 15th, but Productivity and Technology Factor scores 3.00921 and ranks 1st, much higher than the 2nd ranking score of 1.79 of Qinghai; its score of Per capita Factor ranks 5th, but Profitability Factor 21st. The comprehensive ranking is 6th, which shows that Productivity and Technology Factor improves the comprehensive score a lot and the construction industry of Tianjin is developing towards high productivity and technology. It can be categorized as: middle scale, high productivity and technology, high per capita and medium-lower profitability.

The fourth cluster includes Inner Mongolia and Tibet, the comprehensive ranking of which are 16th and 22nd. The scores of Per capita Factor ranks 3rd and 4th, and Profitability Factor 1st and 2nd. But the score of Total Factor and Productivity and Technology Factor are ranking low. It indicates that the construction industry of Inner Mongolia and Tibet is small-scale, low-productive and low-technological, but due to their small population both of the per capita level and profitability are high. It can be categorized as: small scale, low productivity and technology, high per capita and profitability.

The fifth cluster includes the rest 24 regions. The comprehensive scores of these 24 regions span from 0.58307(Shandong) to −0.87562 (Hainan), and they represent the basic development situation of China's construction industry. The score of each common factor in these regions is not high, which shows that the overall development of China's construction industry is not good and it is still in primary stage no matter from which point of view of the scale, productivity and technology, per capita or profitability. This cluster can be classified as: middle scale, medium productivity and technology, medium per capita and profitability.

7 Conclusion

The study constructs an index system, and then applies factor analysis and cluster analysis to conduct an empirical study on construction industry development of 31 regions in China by using SPSS 18.0. All regions are categorized into five clusters by four extracted factors: total factor, efficiency and technology factor, per capita factor and profitability factor. And the results show that significant differences exist in development level of construction industry among different regions. The purpose of the paper is to help policy-makers and industry practitioners find their own positions, and improve competitiveness.

References

DeCoster J (1998) Overview of factor analysis. Retrieved 2 April 2013, from http://www.stat-help.com/notes.html

Deng Rong-hui, Xia Qing-dong (2006) Competitive performance of construction companies based on analytic Hierarch process (in Chinese). J Chongqing Jianzhu Univ 28(6):124–127

Du Gang (2003) Foundations of managerial mathematics. Tianjin University Press, Tianjin

Factor Analysis, Retrieved 3 April 2013, from http://www2.chass.ncsu.edu/garson/pa765/factor.htm

Kale S, Arditi D (2002) Competitive positioning in United States construction industry. J Constr Eng Manag 128(3):238–247

Kang Xue-zeng, Meng Gang (2008) Dynamic analysis on competitive strength for the building industry in our country (in Chinese). Theory Mon 7:150–153

Ruan Lian-fa, Zhang Yue-wei (2009) Research on Zhejiang construction industry development based on DEA analysis, Technoecon Manag Res 6:22–24

Taewoo Youa, Hongmin Zib (2007) The economic crisis and efficiency change: evidence from the Korean construction industry. Appl Econ 39(14):1833–1842

Tsolas IE (2011) Modeling profitability and effectiveness of Greek listed construction firms: an integrated DEA and ratio analysis. Constr Manag Econ 29(8):795–807

Wang Jia-yuan, Yuan Hong-ping (2007) Comprehensive assessment of construction industry based on factor analysis method (in Chinese). J Shenzhen Univ Sci Eng 24(4):373–378

Wang Lei, Wang Xia, Wang Zhi-ling (2006) The evaluation on growth of construction industry based on factor analysis (in Chinese). Sci Technol Prog 23(7):82–84

Wang Wen-xiong, Li Qi-ming (2008) Research on Chinese construction industry competition based on factor analysis and cluster analysis (in Chinese). Appl Stat Manag 27(2):330–337

Wang Xue-qing, Zhang Shuai, Liu Bing-sheng (2011) Index selection and empirical study using the ANP evaluation model for the regional construction industry competitiveness (in Chinese). J Eng Manag 25(2):121–126

Xue X, Shen Q, Wang Y, Lu J (2008) Measuring the productivity of the construction industry in China by using DEA-based Malmquist productivity indices. J Constr Eng Manag 134(1):64–71

Zhou Jian-hua, Yuan Hong-ping (2007) The construction industry economic evaluation based on factor and cluster analysis (in Chinese). Constr Econ 12:9–14