ISSN : 0974 - 7435 Volume 10 Issue 23

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(23), 2014 [14057-14063]

Assessment of enterprise innovation based on the integrated PCA: A case study of the Chinese listed equipment manufacturing enterprises

Zhi-Yuan Li*
School of Management, Harbin University of Science and Technology (HRBUST), Harbin, (CHINA)
E-mail: andysee2008@sohu.com

ABSTRACT

The innovation capacity, as resource of economic competitiveness, plays a significant role in social and economic development and has become a priority issue. Taking 30 Chinese listed equipment manufacturing enterprises as a sample, this study analyzes their innovation capability based on Principal Component Analysis, Kernel Principal Component Analysis and Clustering Analysis. The assessment results of the proposed method, theory and model coincide with the real conditions of Chinese Listed equipment manufacturing Enterprises, demonstrating that the proposed model is accurate, reliable and objective.

KEYWORDS

Assessment; Enterprise innovation; Integrated PCA; China; Equipment manufacturing enterprises.

INTRODUCTION

With the economic globalization, enterprises are facing increasingly fierce international competition. The innovation capability of enterprises has become a significant factor in the survival and development in the enterprise competition. Chinese listed equipment enterprises, with relatively good assets and relatively transparent financial systems, have become an important force in the promotion of the economic growth. Therefore, an objective assessment and systematic analysis of the innovative capacity of listed equipment enterprises can not only make a clear understanding of their overall innovation strength, but also help listed equipment enterprises to maintain and enhance core competitiveness.

LITERATURE REVIEW

In recent years, many approaches are applied in the assessment of the innovation capability of enterprises. Hollenstein[1] used cluster analysis to identify the iinnovation modes in the Swiss service sector based on firm-level data. Jiang[2] used the evidential reasoning to assess the independent innovation capability. Based on Burgelmans operation level innovation ability audit framework as evaluation system, Liu and Zeng[3] adopted AHP vague integrated evaluation method, and study the enterprises' innovation ability. As can be seen in the literature review, among the approaches to assess the enterprise innovation capability, most of them are single assessment methods. Undoubtedly, a single method to assess the innovation capability certainly has one-sidedness. To overcome this shortcoming, the Principal Component Analysis (PCA), the Kernel Principal Component Analysis (KPCA) and clustering analysis (CA) are integrated in this study to analyze and assess the innovation capability of the listed equipment enterprises, providing reference for the innovation capability assessment of listed equipment enterprises in China and the world as well.

METHODOLOGIES

Principal component analysis (PCA)

Principal Component Analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables[4].

Suppose there are n samples and p variables ($n \succ p$), then we can get the initial matrix

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad (1)$$

where $X = (X_1, X_2, ..., X_p)$, $x_i = (x_{1i}, x_{2i}, ..., x_{ni})^T$, $i = 1, 2, ..., p$.

The linear transformation of X can form the new integrated variables Y, i.e.

$$\begin{cases} Y_1 = u_{11}X_1 + u_{12}X_2 + ... + u_{1P}X_P \\ Y_2 = u_{21}X_1 + u_{22}X_2 + ... + u_{2p}X_p \\ ...... \\ Y_p = u_{p1}X_1 + u_{p2}X_2 + ... + u_{pp}X_p \end{cases}, \quad (2)$$

where $u_{k1}^2 + u_{k2}^2 + ... + u_{kp}^2 = 1$, $k = 1, 2, ..., p$.

Because there are numerous kinds of transformation, in order to obtain the best results, u_{ij} is determined by the following principles.

(1) Y_i and Y_j ($i \neq j$, $i, j = 1, 2, ..., p$)are not relevant, i.e. $Cov(Y_i, Y_j) = 0$.

(2) Y_1 has the largest variance among all the linear combinations of $X_1, X_2, ..., X_P$, satisfying Equation 2, i.e.

$$Var(F_1) = \max_{c'c} Var\left|\sum_{i=1}^{p} c_i X_i\right|, \quad (3)$$

Where $c' = (c_1, c_2, ..., c_p)$.

Y_2 has the largest variance among all the linear combinations that are irrelevant to $X_1, X_2, ..., X_P$. Y_p has the largest variance among all the linear combinations $X_1, X_2, ..., X_P$ that are irrelevant to Y_1, Y_2 … Y_{p-1}. Variable $Y_1, Y_2, ..., Y_p$ satisfying the above principles are the principal components, where the proportion of each variable in the total variance is in the descending order.

Kernel principal component analysis (KPCA)

Essentially, KPCA is a multivariate statistical method that applies the kernel method to PCA. Under the principle of the minimum loss of the initial data, KPCA maps a number of variables into the high-dimensional space through a nonlinear mapping chosen in advance. And then several comprehensive variable indicators are obtained through PCA that can reflect the features of the original problems. And a comprehensive analysis of the system features takes place.

First, map the data of the original space into the feature space F by nonlinear transformationΦ: $R^N \to F$, $x \to X$. Assuming that the feature space to meet $\sum_{i=1}^{l} \Phi(x_i) = \mathbf{0}$, the nonlinear PCA can be seen the linear PCA process of matrix $\overline{K} = \frac{\mathbf{1}}{l}\sum_{j=1}^{l} \Phi(x_j)\Phi(x_j)^T$ in F. Obviously all the eigenvalues λ (λ ≥ 0) and the eigenvector V of $\overline{K}$ meet $\lambda V = \overline{K}V$. Define matrix K, wherein $K_{ij} = [\Phi(x_i)\cdot\Phi(x_j)]$ and get $l\lambda\alpha = k\alpha$. Solve the equation to obtain eigenvalues $\lambda_1, \lambda_2, ..., \lambda_l$ and the corresponding eigenvectors $\alpha_1, \alpha_2, ..., \alpha_l$. Let $V^k = \sum_{i=1}^{l} \alpha_i^k \Phi(x_i)$ and for the selection of the principal components, simply calculate the mapping of the eigenvector V^k of a test point $\Phi(x)$ in F,

$$[V^k \cdot \Phi(x)] = \sum_{i=1}^{l} x_i^k[\Phi(x_i)\cdot\Phi(x)] = \sum_{i=1}^{l} x_i^k K(x_i, x)] \tag{4}$$

At this point the assessment function of KPCA is

$$F(x) = \sum_{k=\mathbf{1}}^{r}\sum_{i=\mathbf{1}}^{l} \omega_k \alpha_i^k K(x_i, x), \tag{5}$$

wherein, r satisfies $\frac{\sum_{i=1}^{r}\lambda_i}{\sum_{j=1}^{l}\lambda_j} \ge \mathbf{85\%}$ and ω_k is the contribution of the corresponding k-th principal components.

If $\sum_{i=1}^{l} \Phi(x_i) \ne \mathbf{0}$, then K can be expressed as K *,

$$K^* = K - AK - KA + AKA, \tag{6}$$

where $A_{ij} = \frac{\mathbf{1}}{l}$.

Principal component analysis-clustering analysis (PCA-CA)

In PCA, if the variance contribution of the first principal component is not big enough, there may be one-sidedness in the analysis. PCA-CA classifies the principal components according to their closeness taken from PCA, thereby obtaining a new comprehensive assessment method as follows.

(1) Extraction of principal components

First calculate the covariance matrix

$$S = (s_{ij})_{p\times p}, \quad s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j), \quad i, j = 1, 2, ..., p. \tag{7}$$

Second, calculate eigenvalues $\lambda_1 \ge \lambda_2 \ge ...\lambda_p \succ 0$ and the corresponding eigenvectors of the covariance matrix.

Finally, extract the principal components. Since the purpose of PCA is to reduce the number of variables, in general, $m \prec p$ principal components will be extracted, whose cumulative contribution is more than 85%, namely

$$\frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{p}\lambda_p} \geq 85\% \quad . \tag{8}$$

(2) Calculation of the principal component scores

$$Y_k = u_{1k}X_1 + u_{2k}X_2 + \ldots + u_{pk}X_p \text{ , } k = 1,2,\ldots,\gamma \text{ .} \tag{9}$$

(3) Cluster the selected matrix ($Y_1, Y_2, \ldots, Y_\gamma$), calculate the average score of the first principal component of each cluster, and determine the rank of the clusters.

(4) Based on the score of the first principal component in the sample, determine the rank of samples in each cluster, and get the comprehensive assessment.

4 Empirical study of the innovation capacity of the Chinese listed equipment enterprises

8 indicators are selected for the index system of the innovation capacity assessment model as Proportion of the technical personnel in the R & D personnel (x_1),Ratio of the staff with tertiary and higher education (x_2),Equipment ageing coefficient (Equipment NAV / equipment assets at cost) (x_3),Equipment rate of fixed assets (net fixed assets / total number of employees) (x_4), Main business growth rate (%) (x_5), Net asset growth (%)(x_6),Total asset growth (%) (x_7), and Intangible assets ratio (%)(x_8).

Data processing

Prior to data processing, the data are normalized processing, and the index values are adjusted to[0-1] based on the following formula,

$$x_{ij}^{'} = \frac{x_{ij} - x_{i\min}}{x_{i\max} - x_{i\min}} \text{ ,} \tag{10}$$

Where $x_{ij}^{'}$ denotes the normalized value of the each indicator, $x_{i\min}$ and $x_{i\max}$ respectively denote the minimum and maximum values.

After the data preprocessing, SPSS and MATLAB are used to analyze the data to get the eigenvalues and the corresponding eigenvectors of PCA and KPCA as shown in TABLE 1 and TABLE 2.

TABLE 1 : Eigenvalues and the variance contribution of PCA and KPCA

No.	PCA			KPCA		
	Eigenvalue	Contribution (%)	Cumulative contribution (%)	Eigenvalue	Contribution (%)	Cumulative contribution (%)
1	2.623	32.782	32.782	5.835	72.935	72.935
2	1.943	24.291	57.073	1.371	17.140	90.075
3	1.128	14.102	71.174	0.499	6.238	96.312
4	0.996	12.446	83.620	0.203	2.541	98.854
5	0.526	6.572	90.193	0.058	0.725	99.579
6	0.442	5.526	95.719	0.034	0.421	100.000
7	0.342	4.281	100.000	1.594	1.992	100.000
8	-2.191	-2.739	100.000	-3.270	-4.088	100.000

TABLE 2 : Eigenvalues of PCA and KPCA

	PCA					KPCA	
	Y_1	Y_2	Y_3	Y_4	Y_5	F_1	F_2
1	0.900	0.028	-0.233	0.355	-0.089	0.951	0.161
2	0.900	0.028	-0.233	0.355	-0.089	0.951	0.161
3	0.628	-0.431	0.348	-0.199	0.361	0.595	0.785
4	0.587	-0.343	0.524	-0.318	-0.095	0.834	0.457
5	0.160	0.633	0.609	0.122	-0.173	0.966	-0.208
6	0.250	0.759	-0.172	-0.132	0.512	0.816	-0.425
7	0.132	0.811	0.141	-0.183	-0.170	0.897	-0.378
8	-0.397	-0.060	0.451	0.732	0.222	0.757	-0.355

As shown in TABLE 1, the cumulative contribution of the first 5 principal components reaches 90.193%. Therefore, the first 5 principal components can represent the initial 8 variables. The score of each principal component can be obtained with the following formulas,

$$Y_1=0.9x_1+0.9x_2+0.628x_3+0.587x_4+0.16x_5+0.25x_6+0.132x_7-0.397x_8, \quad (11)$$

$$Y_2=0.028x_1+0.028x_2-0.431x_3-0.343x_4+0.633x_5+0.759x_6+0.811x_7-0.06x_8, \quad (12)$$

$$Y_3=-0.233x_1-0.233x_2+0.348x_3+0.524x_4+0.609x_5-0.172x_6+0.141x_7+0.451x_8, \quad (13)$$

$$Y_4=0.355x_1+0.355x_2-0.199x_3-0.318x_4+0.122x_5-0.132x_6-0.183x_7+0.732x_8, \quad (14)$$

$$Y_5=-0.089x_1-0.089x_2+0.361x_3-0.095x_4-0.173x_5+0.512x_6-0.17x_7+0.222x_8. \quad (15)$$

As shown in TABLE 2, the eigenvalues contributions of the first 2 principal components reach 90.075%. Therefore, the first 2 principal components can represent the initial 8 variables. The linear combination is as follows,

$$F_1=0.951x_1+0.951x_2+0.595x_3+0.834x_4+0.966x_5+0.816x_6+0.897x_7+0.757x_8, \quad (16)$$

$$F_2=0.161x_1+0.161x_2+0.785x_3+0.457x_4-0.208x_5-0.425x_6-0.378x_7-0.355x_8. \quad (17)$$

The economic significance of the principal components is determined by the comprehensive sense of the several indicators who have bigger weights in a linear combination. Therefore, in the factor analysis of PCA, the weighted sum of Y_1、Y_2、Y_3、$Y_4 and Y_5$ is the final assessment score of the innovation capacity, where the weight is the variance contribution rate of each principal component based on Formula (18); in the factor analysis of PCA, the weighted sum of Y_1、Y_2、Y_3、$Y_4 and Y_5$ is the final assessment score of the innovation capacity, where the weight is the variance contribution rate of each principal component based on Formula (19).

$$Y=\alpha_1 Y_1+\alpha_2 Y_2+\alpha_3 Y_3+\alpha_4 Y_4+\alpha_5 Y_5, \quad (18)$$

$$F=\beta_1 F_1+\beta_2 F_2 \quad (19)$$

Finally, SPSS is used for the cluster analysis of the score matrix of the principal components. 30 listed equipment manufacturing enterprises are clustered into four categories; the 4 categories are ranked based on the average score of the first principal component; then the score of the first principal component is used to rank the listed equipment enterprises within a category in the descending order. TABLE 3 shows the results of comparison of various methods for ranking, the final score and the final rank.

TABLE 3 : Comprehensive assessment score and ranking of the innovation capability of the 30 listed equipment enterprises

Stock code	PCA		KPCA		PCA-CA		Final score	Final rank
	Score	Rank	Score	Rank	Score	Rank		
601808	10.43892	2	4.629845	1	9.879347	2	8.045788	1
000400	11.88861	1	2.798352	5	10.1765	1	7.437679	2
300024	9.99543	3	2.87696	3	8.865423	3	6.587642	3
002608	9.382773	4	2.817737	4	7.000865	4	6.148302	4
601369	6.684346	6	3.489273	2	5.289652	7	4.808843	5
002179	6.732487	5	2.43341	6	5.457796	5	4.741622	6
002013	6.08332	9	2.184349	7	5.438815	6	4.3876	7
000039	6.38462	7	1.944738	9	5.135689	8	4.300504	8
600583	6.177651	8	2.03478	8	5.069832	9	4.22882	9
601299	5.80498	10	1.7845747	11	4.489458	11	3.836888	10
600495	5.704878	11	1.633567	12	4.576532	10	3.756723	11
600893	4.687038	12	1.884894	10	2.615964	14	2.978845	12
601766	4.097482	15	1.604656	14	2.778643	12	2.822282	13
601989	4.190305	14	1.449948	16	2.681257	13	2.633543	14
600458	4.193099	13	1.614736	13	1.958944	17	2.47168	15
600765	3.472845	16	1.27267	20	2.394697	15	2.329961	16
000738	3.320001	18	1.430072	17	2.089662	16	2.1833337	17
002073	3.330863	17	1.36546	18	1.824598	19	2.113378	18
600391	2.776805	20	1.093209	25	1.818763	20	1.850002	19
600685	3.013336	19	1.588393	15	1.096877	24	1.825808	20
002270	2.489062	22	1.295643	19	1.322654	22	1.726764	21
300008	2.265388	24	1.101967	24	1.546824	21	1.648675	22
000410	2.769263	21	1.178363	22	1.0109054	25	1.59003	23
300161	2.104689	26	0.621552	28	1.9300988	18	1.490567	24
600150	2.479968	23	1.263017	21	0.7299433	26	1.448829	25
600320	2.215972	25	1.113169	23	0.414956	28	1.222874	26
000837	1.530444	27	0.799405	26	1.108735	23	1.140831	27
600268	1.069093	29	0.787722	27	0.59882	27	0.823384	28
600072	1.105679	28	0.57389	30	0.2940509	29	0.668557	29
002248	1.021881	30	0.600315	29	0.10802	30	0.571248	30

Empirical analysis

In both PCA and KPCA, the first principal component mainly comes from the Proportion of the technical personnel in the R & D personnel (x_1) and the Ratio of the staff with tertiary and higher education (x_2) and the second principal component depends on the enterprise's technology investments. In rank of the innovation capacity of the Chinese listed equipment enterprises, the top 10 listed equipment enterprises invest more in human resources, particularly in the technical and management than other listed equipment enterprises. Therefore, x_1 and x_2 reflect both the investment of innovative human resources and the enterprise's technical qualifications and level.

CONCLUSIONS

In this study, we assess and study the innovation capability of the listed equipment enterprises based on the integrated PCA, i.e. the integration of PCA, KPCA and CA. The empirical results show that: (1) the integrated approach can utilize the advantages different assessment methods, overcome the shortcomings of a single method, well reflect the information and improve the assessment reliability; (2) the integrated approach has excellent performance, in that it can not only effectively reduce the dimension of the assessment indicators but also effectively deal with the non-linear effects between indicators, providing an objective and scientific basis for the innovation capacity of the Chinese listed equipment enterprises; (3) the integrated comprehensive assessment method is an improved method to overcome the one-sidedness and limitations of a single method and to solve the result differences in various assessment methods.

ACKNOWLEDGEMENT

This work is funded by the Research & Planning Program in Social Sciences of Heilongjiang Province (14E073).

REFERENCES

[1] H.Hollenstein; Innovation modes in the swiss service sector: A cluster analysis based on firm-level data, *Research Policy,* **32(5)**, 845-863 **(2003)**.

[2] B.Jiang, J.Q.Pan, J.Fang; Research on evaluation of enterprises' independent innovation ability based on evidence reasoning, *Journal of Systems & Management*, **(8)**, 416-420 **(2009)**.

[3] H.Y.Liu, Y.L.Zeng; The vague AHP evaluation based on innovation ability of enterprises operation level, *Science & Technology Progress and Policy*, **24(2)**, 116-120 **(2007)**.

[4] Z.Y.Li, B.Jacob; Integrated assessment on profitability of the Chinese aviation industry, *Journal of Intelligent Systems*, **23(1)**, 49–58 **(2013)**.