Analysis on Regional Differences and Influencing Factors of China's High Tech Industry Development - Based on Cross-Sectional Data of 30 Provinces

Yang Liu [a], Jiahang Zhang [b], Rui Zhang [c*], Zhi Li [d]
[a]e-mail: yang_liu202110@163.com, Email:[b]E-mail:zhangjiahang9380@163.com

*Corresponding author: [c]abm201205@163.com, Email: [d]email: lizhi20012022@163.com

School of economics, Yunnan University of Finance and economics, Wuhua, Kunming, Yunnan, China

Abstract—Based on the panel data of 30 provinces in China, this paper first uses ward cluster analysis to divide the development level of high-tech industries of 30 provinces into 7 categories according to the selected 9 indicators. The results show that the development level of high-tech industries in Guangdong and Jiangsu is relatively high, while that of Hainan, Qinghai and Ningxia is relatively low. There are significant differences in the development of high-tech industries in eastern, central and Western China. Then, the nine indicators and the main business income of high-tech industry are linearly regressed. Combined with the results of correlation analysis, the dimension of the selected indicators is reduced by principal component analysis, and the original regression equation is modified. The results show that the level of technology marketization is the primary factor affecting the development of high-tech industry. Finally, the paper points out the significance of high-tech cluster in China.

Keywords-Ward clustering; Principal Component Analysis; Regression Equation; Technical Marketization Level

1. Introduction

1.1 Research background

High tech industry is characterized by intensive knowledge and technology, low consumption of resources and energy, and high added value. It is the "vanguard" of modern industrial system. Since the 1980s, China's high-tech industry has been developing vigorously, gradually penetrating into all fields of social life and becoming the main pillar of high-quality economic development. However, the development level of China's high-tech industry is relatively low, the development factors are not sufficient, and the regional development differences are obvious.

1.2 Research review

Levie & muzyka [1] believes that policies play an important role in the development of high-tech industries. The government should formulate different supporting policies according to the characteristics of enterprise development. Sternberg [2] emphasized the important role of the government in the development of high-tech industries, and discussed the impact on the development of high-tech industries from the aspects of government policies, market demand,

BDEIM 2022, December 02-03, Zhengzhou, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328729

infrastructure and regional R & D level. Stevenson & Lundstrom [3] analyzed the impact of business support, entrepreneurship training and other factors on high-tech industry. Aydalot et al[4] described that R & D and technological innovation are the main factors for the development of high-tech industries. Romer [5] regards technological progress as an important reason for economic growth, and technological transformation and innovation can significantly promote economic growth. Dilling-Hansen [6] found that the ownership structure of enterprises has a significant impact on the R & D performance of high-tech industries. Allen [7] thinks that the financial market is conducive to promoting the development of high-tech industries, financial market development can promote technological innovation. Blomstrom [8] analysis of the spillover effect of foreign direct investment will promote the technological progress and technical efficiency of domestic enterprises.

1.3 Research objectives and work

At present, there is a lack of research on the regional differences and influencing factors of the development of high-tech industries, while there are few theories using Ward clustering and principal component analysis to analyze the development level of high-tech industries. Based on the high-dimensional data, this paper uses the high-level cluster analysis method to construct the high-level indicators of industry development, some suggestions are put forward for the development of high-tech industry.

2. Materials and Methods

2.1 Data collection and processing

In order to explore the factors that influence the development of China's high-tech industries in the past 30 years, this paper sorts out the important factors that influence the development of China's industries in the past 30 years. Because the units of each index are different, the influence of dimension can be eliminated by data standardization.

2.2 Cluster analysis

In this paper, ward clustering is first used. The basic process of ward clustering is to find out the average point of each index, calculate the square of the distance, and then sum. Based on the nine indicators of the data, this paper divides 30 provinces into 7 categories and divides the development level of high-tech industries in each province.

2.3 Linear regression

Taking the main business income of high-tech industry in each province as the dependent variable, the OLS regression equation is constructed, and the other nine indicators are analyzed by linear regression analysis. The correlation between other indicators and the development of high-tech industry is analyzed, and the rationality and significance of this model are tested.

2.4 Principal component analysis

On the basis of linear regression, by reducing the dimension, eliminating redundant indicators, the new indicators are recombined into new indicators with higher discrimination. Most of the information of the original indicators is retained in the new indicators, and the principal components and the main business income of high-tech industries are analyzed and tested by re-

gression. Then, the regression equation of principal component is reduced, compared with the original linear equation, the advantages and disadvantages of the two are compared, and the model with the best regression effect is selected. Finally, the relationship between each index and the development of high-tech industry is analyzed, focusing on individual indicators.

2.5 Result analysis

This paper analyzes the results of regression from two aspects of practice and theory, and points out the significance and value of the development of high-tech industry.

3. Results & Discussion

3.1 Data processing

Because the units of the selected indicators are not consistent, it is necessary to standardize the panel data to eliminate the influence of data dimension.

$$y_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}} \tag{1}$$

3.2 Index introduction

Table 1 Indicator symbol and explain

Index	Explain	Symbol
Labor endowment	The number of employed people in cities and towns of each province.	L
Capital conditions	The number of new fixed assets in provinces.	K
Full time equivalent of R&D personnel	The number of people engaged in high-tech industries according to their working hours represents the scientific research level of each province.	R
Technology marketization level	It is measured by the sales revenue of the new product developed.	T
Gross domestic product by region	The sum of industrial added value, which represents the economic development level of each province.	G
Total import and export volume by region	It is determined by the high-tech industry export delivery value.	TMX
Amount of investment	Investment amount of high-tech industries in various provinces.	I
Value of industrial output	Gross industrial output value refers to the final achievement of industrial production activities expressed by industrial enterprises in monetary form during the reporting period.	IP
National financial education funds	The amount of each province's expenditure on developing education at all levels.	NF
Main business income of high-tech industry	It refers to the sum of the income of high-tech products or services and their technical income generated by enterprises through technological innovation and R&D activities.	Y

3.3 Ward clustering

Ward clustering is a common method to measure the distance between clusters in hierarchical clustering aggregation method. Its steps are to first calculate the ESS of each cluster, and then calculate the total ESS.

$$D_{st} = \frac{n_s + n_p}{n_s + n_t} D_{sp} + \frac{n_s + n_q}{n_s + n_t} D_{sq} - \frac{n_s}{n_s + n_t} D_{pq} \tag{2}$$

According to the selected first nine indicators, the development status of high-tech industries in 30 provinces is clustered, and the classification number is 7.

The clustering results of category 7 are as follows:

Table 2 Classification of provinces by clustering

Category	Province
first kind	Hainan, Qinghai, Ningxia
second kind	Beijing, Shanghai
The third category	Guangxi, Shaanxi, Tianjin, Liaoning, Guizhou, Yunnan, Gansu, Xinjiang, Jilin, Heilongjiang, Shanxi, Inner Mongolia
The fourth category	Jiangsu
The fifth category	Guangdong
The sixth category	Chongqing, Anhui, Jiangxi, Fujian, Sichuan, Hebei, Hubei, Hunan
The seventh category	Zhejiang, Shandong, Henan

According to the classification results, the development level of high-tech industries in Guangdong and Jiangsu is relatively high, and there is a significant gap with other provinces, while the development level of high-tech industries in Hainan, Qinghai and Ningxia is relatively low, and the development of high-tech industries in other provinces is mostly at a medium level. There are problems of unbalanced and uncoordinated development of high-tech industries in eastern, central and western China.

3.4 Regression analysis

Then it discusses the factors that influence the development of high-tech industries in China. This paper takes the main business income of high-tech industries in each province as the dependent variable, and analyzes the relationship between the other nine indicators and the dependent variable.

Table 3 The result of linear regression

variable	Estimate	Std.Error	t value	Pr(>|t|)	
Intercept	0.008	0.008	0.982	0.338	
L	-0.209	0.078	-2.684	0.014	*
K	0.161	0.103	1.572	0.132	

R	0.072	0.127	0.568	0.576	
T	0.747	0.105	7.088	0.000	***
G	0.163	0.186	0.874	0.393	
TMX	0.369	0.072	5.124	0..000	***
I	0.158	0.091	1.749	0.096	
IP	-0.308	0.130	-2.366	0.028	*
NF	0.011	0.076	0.138	0.892	
R^2	0.991				
P	0.000				

3.5 OLS linear regression model is established as follows

$$Y = 0.008 - 0.209L + 0.161K + 0.072R + 0.747T + 0.163G + 0.369TMX + 0.158I - 0.308IP + 0.011NF \quad (3)$$

From the regression results, the value of P is extremely small, and this model is significant, with R^2 being 0.991, and the fitting degree of the model is high. However, at the significant level of 0.05, K, R, G, I, and NF don't pass the significance test. Therefore, there are some defects in regression analysis directly through the existing indicators. On the basis of the original variables, this paper carries out principal component analysis by reducing the dimension of indicators and removing the influence of redundant information.

3.6 Principal component analysis

Judge the correlation between each index and Y, and make the heat diagram of all index correlations as follows:

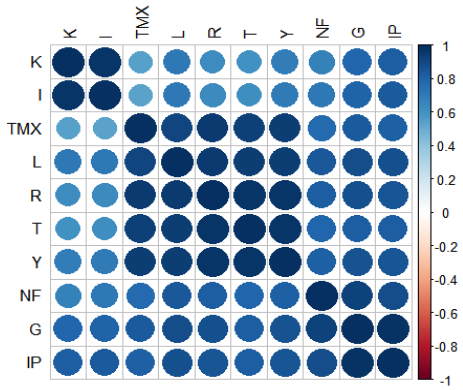

Figure 1 Correlation heat diagram

According to the results of correlation analysis, Y has a strong correlation with TMX L, R and T. Therefore, the principal component can be used to reduce the dimension. The regression results of principal components are as follows:

Table 4 Regression results of principal components

variable	Standard deviation	Proportion of Variance	Cumulative Proportion
Comp.1	2.737	0.833	0.833
Comp.2	0.979	0.107	0.939
Comp.3	0.535	0.032	0.971
Comp.4	0.341	0.013	0.984
Comp.5	0.235	0.006	0.990
Comp.6	0.216	0.005	0.995
Comp.7	0.161	0.003	0.998
Comp.8	0.120	0.002	1.000
Comp.9	0.060	0.000	1.000

The cumulative contribution rate of the first two principal components reaches 93.906%, which shows that these two components retain a lot of information of the original indicators. Therefore, it is only necessary to select the first two principal components and analyze the relationship between these two principal components and Y.

The relationship between components and variables is as follows:

Table 5 Relationship between principal components and indicators

variable	Comp.1	Comp.2
L	0.350	0.159
K	0.298	-0.555
R	0.344	0.299
T	0.336	0.320
G	0.354	
TMX	0.328	0.391
I	0.302	-0.549
IP	0.352	-0.123
NF	0.332	

Assuming that the first three principal components are w_1 and w_2 respectively, then

$$w_1 = 0.350L + 0.298K + 0.344R + 0.336T + 0.354G + 0.328TMX + 0.302I + 0.352IP + 0.332NF \quad (4)$$

$$w_1 = 0.159L - 0.555K + 0.299R + 0.320T + 0.391TMX - 0.549I - 0.123IP \quad (5)$$

Then, the sample data of the first two principal components are calculated, and the regression analysis is made between these two principal components and Y.

Table 6 Regression analysis of principal components

variable	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	0.120	0.010	12.580	0.000	***
w_1	0.071	0.004	20.360	0.000	***
w_2	0.043	0.010	4.390	0.000	***
R^2	0.937				
P	0.000				

In the new regression equation, the coefficients of two variables, w_1 and w_2, are highly significant, and the regression effect is ideal. R^2 is 0.937, and the fitting degree of the equation is high. The regression equation of the principal component:

$$Y = 0.120 + 0.071w_1 + 0.043w_2 \tag{6}$$

Using the relationship between the principal components and the original indexes, the original two principal components are changed into the original variables, and the regression equation of the variables can be obtained as follows:

$$\begin{aligned} Y = &-0.053 + 0.153L - 0.013K + 0.185R + 0.196T \\ &+ 0.089G + 0.191TMX - 0.009I + 0.077IP + 0.110NF \end{aligned} \tag{7}$$

The principal component regression equation is changed into the original variable regression equation. K, I and IP have no significant influence on the equation, but R, T and TMX have significant influence on the equation. T is the primary factor affecting the development of high-tech industries. Compared with the original regression equation, the principal component regression equation is obviously more effective.

The high level of technology marketization, a large number of high-tech talents and open free trade have promoted the development of high-tech industries in the eastern region, while the low level of technology marketization, brain drain and poor geographical location in the western region have led to the low level of development of high-tech industries.

4. Conclusion

Based on the panel data, this paper first analyzes the regional differences of high-tech industry development in China by cluster analysis, and then makes regression analysis on L, K, R, T, G, TMX, I, and IP.

4.1 Draw the following conclusions

Firstly, due to various factors, the development level of high-tech industries in various provinces is uneven, and there is a big difference between the development of high-tech industries in the east and the west. At present, some provinces are not dominated by high-tech industries or have not updated and followed up the development of high-tech industries in time, which makes the allocation of market resources tend to other industries.

Secondly, the development of high-tech industry is closely related to many indicators. This paper makes a regression between the selected indicators and the high-tech industry, and finds that most of them show a positive relationship, which means that the improvement of these indicators is helpful to the improvement of the development level of the high-tech industry.

Therefore, all provinces should increase the investment of relevant indicators, so as to promote the overall development of high-tech industries.

Third, the development of high-tech industry depends on the development of frontier science and technology. R&D activists provide technical support for the development of high-tech industries, and the level of technology marketization has a positive impact on the main business income of high-tech industries, which indicates that the level of technology marketization provides financial guarantee and strong market demand for high-tech industries.

4.2 Research significance

Defining the regional differences in the development of high-tech industries is conducive to the rational decision-making of the government, rational allocation of resources, enhancement of regional innovation and development, gradual narrowing of the gap in the development of high-tech industries in various regions, and coordinated development.

Understanding the important indicators affecting high-tech industries will help enterprises to prescribe the right medicine, innovate and develop new technologies and products, and open up new markets, thus driving consumption, improving consumer satisfaction and creating a good market atmosphere.

It is helpful for the state to increase investment in the field of education, continuously cultivate high-tech talents, and encourage R&D innovation, so as to better serve the high-tech industry.

It is beneficial to study the impact of effective indicators on high-tech industries, innovate related theories, improve the system and provide theoretical guidance for the development of high-tech industries.

4.3 Shortcomings

In this paper, the selection of indicators is not comprehensive enough, which may cause the accuracy of the results of the model to decrease. At the same time, the factors influencing the development of high-tech industries are not static. With the development of high-tech industries, more influencing factors should be taken into account in the index system.

References

[1] Levie, J., & Muzyka, D. (1996). New venture growth in France: do local government policies make a difference?. Frontiers of Entrepreneurship Research, 8, 46-69.

[2] Sternberg, R. (1996). Reasons for the genesis of high-tech regions—theoretical explanation and empirical evidence. Geoforum, 27(2), 205-223.

[3] Stevenson, L., & Lundström, A. (2001, June). Towards a framework for the development of entrepreneurship policy and practice. In Proceedings of the Annual Conference Interntionl Council Small Business, Washington, DC.

[4] Aydolat, P., & Keeble, D. (1998). High technology industry and innovative environments: the European experience. GREMI.

[5] Romer, P. M. (1990). Endogenous Technological Change; Journal of Political Economy, October.

[6] Dilling-Hansen, M., Madsen, E. S., & Smith, V. (2003). Efficiency, R&D and ownership–some empirical evidence. International Journal of Production Economics, 83(1), 85-94.
[7] Allen, F. (1993). Stock Markets and Resource Allocation in Capital Markets and Financial Intermediation. Capital Markets and Financial Intermediation.
[8] Blomström, M., & Persson, H. (1983). Foreign investment and spillover efficiency in an underdeveloped economy: Evidence from the Mexican manufacturing industry. World development, 11(6), 493-501.