ISSN:2790-1661 Volume-5-(2023)

Empirical research on the influencing factors and improvement path of big data industry construction based on factor analysis method

Zhuolun Li

Faculty of Shipping Economics and Management,Dalian Maritime University,Dalian,China

E-mail:zhuolunli@dlmu.edu.cn

Abstract. With the continuous development of the information technology era, the big data industry is becoming one of the important criteria for measuring the comprehensive strength of the region. To better understand and explore the development status of the big data industry, it is essential to analyze the influencing factors of the big data industry, and then determine effective countermeasures that can promote the long-term development of the big data industry according to the influencing factors. This paper is committed to studying the influencing factors of the big data industry. By selecting 16 indicators, including policies and regulations, management institutions, technological innovation,talents distribution and per capita GDP in 31 provinces and municipalities directly under the Central Government in 2021, the factor analysis method is used to reduce 16 indicators into four factors affecting the government promotion, industrial ecosystem, industry platform and industrial resources affecting the big data industry, and establishes a multivariate linear regression model that affects the scale of the big data industry. This study found that there is a positive correlation between the industrial scale of the big data industry and the industrial ecosystem, government promotion and industrial resources. Combined with the specific situation and empirical analysis results of the development of provinces and cities across the country, this paper puts forward relevant suggestions in order to provide some help for the development of the big data industry in 31 provinces and cities across the country, hoping to provide a certain theoretical basis for the relevant departments in charge of the big data industry to formulate corresponding policies.

Keywords: Big data industry； Factor analysis； Multiple linear regression models； Industrial ecosystem; Industrial resources.

1. Introduction

With the advent of the era of big data, the popularity of intelligent terminals and mobile networks not only meets people's living needs but also makes the social operation process show distinct data characteristics. Big data refers to large and complex structured and unstructured data from different sources. The concept of Big Data originated in the United States. As early as 1980, the famous futurist Alventole called big data "the colorful movement of the third wave" in his book The Third Wave. In 1997, David Eisleworth and Michael Cox of NASA Ames Research Center first used the concept of big data when studying the visualization of data. In 1998, an article entitled "Visualization of Data Science" was published in the American journal Nature, so big data officially appeared in public journals as a topic term. Chen Yue, a Chinese scholar, believes that "using big data in human resources management at a time when big data applications are popular can help enterprises better manage employees and improve their overall management ability" [1]. Wang Shuguang and others proposed that "the deep combination of big data technology and government governance is the result of choosing to improve government governance capacity and the result of the era of big data" [2].

The big data industry originated in developed countries. The United States, the United Kingdom, Australia, and other countries have successively promoted big data as a national strategy to promote the development of the big data industry. Lv Ping believes that the big data industry has become an important engine to accelerate the qualitative, efficiency and dynamic transformation of economic and social development, and is one of the keys to promoting the Fourth Global Industrial

Revolution [3]. Fu Wenyu and other scholars used the hierarchical analysis method and the principle of fuzzy comprehensive evaluation to put forward a performance evaluation method for financial support for the development of the big data industry [4]. Zhang Mu et al. used the dynamic panel data model, the intermediate effect test method and the threshold effect model to empirically test the impact of financial technology on the development of the big data industry and its mechanism [5]. Han Xianfeng and others used the DEA method to analyze the technological efficiency of China's big data industry and its influencing factors [6]. Zhou Ying et al. analyzed the factors affecting the big data industry using Delphi and AHP methods [7]. Shen Junxin and others used the fuzzy set qualitative comparative analysis (fsQCA) method to explore the impact of multiple factors on the development capacity of the big data industry [8]. At present, there are views on how to effectively guide and develop the big data industry, such as "government promotion theory" and "platform support theory". At present, there is no literature based on factor analysis and multivariate linear regression method to study the influencing factors of the big data industry from the government, platform, industry, society, and other dimensions, so the research in this article is innovative.

Combined with the actual situation of the development of the big data industry in 31 provinces and municipalities directly under the Central Government, this paper explores factors that can mainly affect the big data industry, and puts forward relevant suggestions based on the experimental data results. This paper fills some theoretical gaps in the research of the influencing factors of the big data industry, provides a theoretical basis for further in-depth research and development of the big data industry in the future, and also provides theoretical support for the big data development and construction policies of 31 provincial and municipal governments in China from different angles, further helps to promote the development of the digital era, and provides reference for the further addition and standardization of industry platforms. It looks forward to providing some help for the sustainable development of China's big data industry and contributing to the great progress of China's big data industry.

2. Basic Theory

2.1 Factor analysis process

The cross-section data is N samples, and each sample has p indicators. 1， 2， 3，…， p are special factors，F (F1, F2 ,⋯,Fm) (m < p) is an unobservable random vector. It is assumed that the random observation variable X satisfies the following model:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \cdots \\ X_P = a_{p1}F_2 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_3 \end{cases} \quad (1)$$

2.2 Multiple linear regression analysis

Using the new variable F obtained from factor analysis, a regression model is established with Scale of the industrial scale of the big data industry. The sample capacity is n, k<n, Scale is a dependent variable, F is all independent variables, which is an N* (k+1) matrix composed of N factor values, and β is the k+1-dimensional coefficient vector; is the n-dimensional perturbation term column vector. The model is:

$$\text{Scale} = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_k F_k + \mu \quad (2)$$

3. Model building and empirical analysis

3.1 Selection of evaluation indicators and data sources

According to the Big Data White Paper released by the Chinese Academy of Information and Communications Technology, in 2021, China's big data industry developed rapidly and the amount of data increased rapidly. The overall scale of the big data industry reached 1.3 trillion yuan, with a compound growth rate of more than 30%. Now it has gradually entered the stage of high-quality development. Therefore,the actual situation data of the big data industry in 2021 is of great significance for studying the construction and development of China's future big data industry. Therefore, this article chooses data from 2021 to describe the development status and trend of the big data industry. The combination of factor analysis and regression analysis of cross-section data can effectively and accurately overcome the problem of multivariate colinearity, and a more concise and clear regression model can be obtained. This paper uses SPSS 26.0 software to analyze the data, reduces the dimension of 16 variable indicators to 4, and then uses SPSS 26.0 software for regression analysis. By sorting out and summarizing the existing literature, combining the data availability of the big data industry in 31 provinces and cities across the country, and following the principles of data systematism, availability, operability and representativeness, this paper chooses to evaluate the influencing factors of the big data industry from the following four perspectives:

First, from the perspective of the government. In the process of the growth of the big data industry, the government assumes the role of guiding, cultivating and promoting it. Wang Qian and others proposed that the role of the government in guiding, supporting, popularizing and standardizing has become the key to achieving the high-quality development of the big data industry [9]. Ma Liping and other scholars believes that the construction of a big data sharing platform can provide strong support for the efficient governance of the government, and the construction of a big data sharing platform is the path choice to improve the government's governance capabilities. Building a big data sharing platform is a path choice to improve the government's governance capacity [10]. Industrial policy is a tool to achieve national economic goals. Luo Tao believes that industrial policy is an important factor leading to the development gap between the big data industry between China and the United States [11]. Jia Chuanling and others believe that "from the development experience of various places, the big data management institutions, as a functional department of the government, mainly plays the role of data integration, providing a technical basis for optimizing government services and improving administrative efficiency; at the same time, it manages and guides the development of the data economy" [12]. Tae Hee Kwon and others used the ANP model and also studied policy formulation and policy priorities to activate the big data industry [13].

By investigating the big data industry and summarizing the existing research and methods, it is determined that the first dimension is measured from the perspective of the government through three indicators: big data management institutions, big data industry policies and regulations, and open sharing resources of big data. The big data management institutions said that the establishment of institutions to supervise the big data industry within the government reflects the importance that local governments attach to big data. Big data industry policies and regulations refer to the policies promulgated, promulgated, formulated and revised by the government to promote and develop the big data industry and measure the promotion and guidance direction of local governments to the big data industry. Big data open sharing resources refer to the open data sharing resources published by provincial and municipal governments, reflecting the strength of local governments to promote data sharing and build an industrial ecosystem.

The second is the platform angle. With the competition of global Internet giants to lay out the big data industry and the in-depth promotion of the implementation of the "big data +" ecosystem strategy, China's big data industry has shown a good trend of vigorous development, agglomeration and platform innovation. The industrial platform is the core of the big data industry and an

important carrier for data-driven industrial upgrading. With the strong support of the government, the industrial platform has become an important force in the development of domestic big data. Wang Wei and others believe that strengthening the construction of a big data trading platform will help data sharing, give full play to the value of data, and improve the efficiency of data sharing [14]. At the same time, data centers are an important new type of infrastructure in the era of big data, which is an important guarantee for the digital transformation and upgrading of society and regional high-quality development. Wang Xinfeng and other scholars believe that "the data center will be the traffic center of the Internet in the future, and the collaborative trend of cloud network numbers is obvious" [15].

Therefore, it is determined that the second dimension is measured from the perspective of the platform and measured through two indicators: trading platform and data platform. Among them, the trading platform is a platform to realize massive data trading, reflecting the support for data trading links. The trading platforms that have been operated in China include Guiyang Big Data Exchange, Beijing Big Data Trading Support Platform,etc. The industrial platform is an important condition and guarantee for the development of the big data industry. Its quantity reflects the support for digital technology, industrial technology innovation, and the digital economy.

Third, the industry perspective. Big data industry refers to industrial clusters and industrial parks of big data, covering big data industrial parks composed of big data technology product research and development and big data industry service system. In the fields related to the big data industry, many scholars have carried out in-depth research. Wu Xue put forward big data industry cluster refers to relying on big data industry chain upstream and downstream data resources, IT infrastructure, etc., through the way of professional division of labor and collaboration, on the basis of big data talent, technology, capital collaborative innovation, value exchange of a dynamic interaction, network new industrial organization form [16]. After empirical analysis of 40 domestic listed companies, Fu Wei et al. proposed that "strengthening the investment in innovation factors is an important aspect of improving the total factor productivity of the data industry"[17]. Xia Dawen et al. put forward that talent is the main factor affecting the development of big data industry[18]. Big data industry belongs to the contemporary high-tech industry, and its development needs more support of high-end talents. However, at present, there is a serious lack of big data talents in China. In view of the problem of talent shortage, Liu Wenhua proposed that the main way to solve the lack of big data talent is to reconstruct the talent training system, optimize the teaching content, and reform the teaching method [19]. Zhang Taiping proposed that only by strengthening the foundation of big data industry can we accelerate the development of digital economy [20]. Because the consolidation of the big data industry foundation can not only help to strengthen the innovation ability of enterprises, but also drive the development of emerging industries. In the big data industry, the leading enterprises of big data are in a dominant position and are responsible for leading the development of the regional big data industry. Enterprises with investment value are the backbone of the big data industry, reflecting the development potential of the regional big data industry in the future. The total amount of big data enterprises represents the overall scale of the big data industry in the region, and is gradually becoming one of the important indicators to measure the regional economy. The number of newly built enterprises reflects whether the development of big data industry in each region can continuously inject vitality. The larger the number of new big data enterprises, the larger the scale of the local big data industry, and the greater the contribution to the overall big data economy.

From this point of view, the third dimension in the industry point of view, through the big data leading enterprises, investment value, big data, new big data enterprises, big data technology innovation degree, big data talent distribution, set up data science and big data technology big data professional university and industry environment indicators to measure. If the regional innovation atmosphere, strong information industry foundation, big data talent, active data technology research and development, leading enterprises can fully play a leading role, backbone enterprises can grow under the good industry environment, with these favorable development conditions, on large

probability can create more big data enterprise, and for the development of big data industry provide a steady stream of nutrients, promote its rapid growth for industrial highlands.

Fourth, from a social perspective. An industry wants to have a considerable development cannot do without a good social environment training. GDP is the core indicator of national economic accounting and an important indicator to measure the economic situation and development level of a country or region. It can fully represent the specific socio-economic development status of various provinces and cities in China.The development of the big data industry and the development of the Internet are also complementary. On the one hand, the development of the Internet has provided more data, information and resources for the development of big data; on the other hand, the development of big data has provided more support, services and applications for the development of the Internet. The development of the Internet cannot be separated from the construction of solid infrastructure, and the development of the big data industry is also closely related to the basic Internet resources in various regions, while the IPV 4 occupancy ratio of provinces reflects the development of provinces in terms of the basic social network resources. The research shows that the overall cultural level of the society has an important influence on the development of enterprises. The higher the overall cultural level of the society, the better the development prospect of the enterprise; the contrary, the worse the development prospect of the enterprise. Therefore, the cultural and education level in various regions also greatly affects the development of the big data industry.To sum up, from a social perspective, the fourth dimension chooses to measure it with three indicators: cultural education level, GDP and IPV4 ratio, which represents the overall cultural level of society, social and economic situation and the current situation of social basic network resources. The specific measurement indicators from the four perspectives of government, platform, industry and society are shown in Table 1.

This article selects the big data of the big data industry in 31 provinces and cities across the country in 2021. The data mainly comes from the 2021 China Big Data Regional Development Level Assessment Report, the 2021 China Big Data Industry Statistics Yearbook, 2021 White Paper on the Development of China's Big Data Industry and other authoritative national statistics.

Table 1 Measurement indicators and meanings

Identification	Metrics	Meaning
Y	Industry scale	The scale of output or operation of the big data industry
X1	policies and regulations	Number of policies and regulations issued by provincial and municipal governments on the big data industry
X2	Management institutions	The number of big data management institutions set up by the provincial and municipal governments
X3	Data sources	The number of data resources opened by provincial and municipal governments
X4	Trading platform	Number of big data trading centers and exchanges
X5	Data platform	Number of large urban centers, government centers and supercomputing centers
X6	Technological innovation	Current status of innovation input and output of big data technology
X7	Talents distribution	The number of big data talents
X8	Industrial environment	The basic situation of the big data industry
X9	Subject construction	Number of universities offering big data-related majors
X10	Cultural education	Number of people with high school education in provinces and municipalities
X11	GDP per capita	GDP per capita of provinces and cities
X12	IPv4 ratio	Provincial and municipal IPv4 share ratio

X13	Leading Enterprises	Top 50 domestic big data enterprises
X14	Total number of enterprises	The total number of big data enterprises in provinces and cities
X15	The backbone enterprise	Top 100 enterprises with domestic investment value
X16	New enterprise	Number of new big data enterprises in China

3.2 Factor analysis

Before factor analysis of each variable, it is necessary to check whether there is a strong correlation between the original variables. In this paper, 16 indicator data were tested by KMO and Bartlett sphere using SPSS26.0 statistical software. The original data is standardized. Table 2 shows that the KMO test value is 0.783 and greater than 0.5, indicating that the information overlap between variables is high. The approximate chi-square value of the Bartley sphericality test is 619.863, and the significance probability is 0.000 and less than 0.05, indicating that there is a correlation between various variables. It is appropriate to measure the factor analysis method, indicating that the variables and data selected in this paper are suitable for factor analysis.

Table 2 KMO and Bartlett test

Test indicators		Statistical value
KMO sampling tangent quantity.		.783
Bartlett spherical test	Approximate chi square	619.863
	Degree of freedom	120
	Significance	.000

In this paper, factor analysis is used to analyze the correlation matrix, and the sample data is dimensionally reduced. Four common factors are extracted according to the principle that the cumulative variance contribution rate exceeds 85%. According to the slope of Figure 1, it is appropriate to select four factors. It can be seen from Table 3 that the cumulative contribution rate after rotation reaches 85.827%, indicating that the four extracted factors can interpret the original variable by 85.827%, and the factor analysis method can fully explain the original variable. Among them, the variance contribution rates of common factors F1, F2, F3 and F4 are 54.261%, 20.972%, 6.342% and 4.253%, respectively, indicating that the common factor F1 has the greatest impact on the overall data.

Table 3 Explanation of total variance

Extract the sum of load squares			The sum of rotating load squares		
Total Amount	Deviation Percentage	Accumulate Percentage	Total Amount	Deviation Percentage	Accumulate Percentage
8.682	54.261	54.261	6.307	39.422	39.422
3.355	20.972	75.233	3.260	20.374	59.796
1.015	6.342	81.575	2.633	16.455	76.251
.680	4.253	85.827	1.532	9.576	85.827

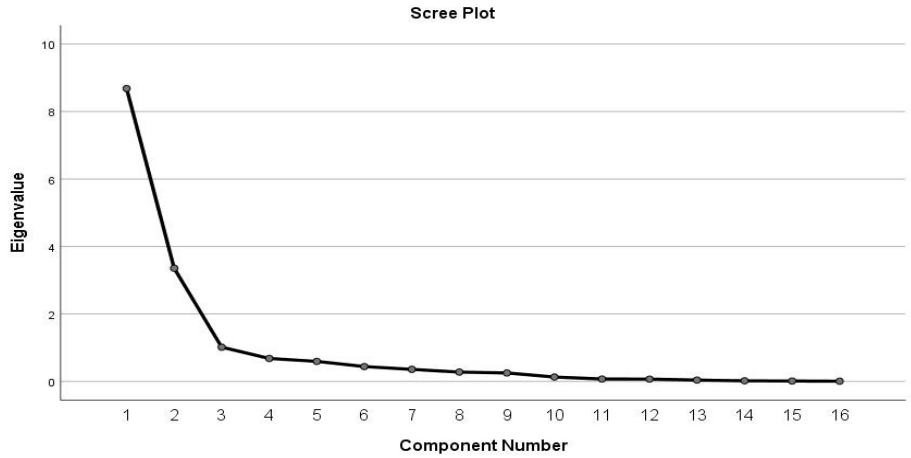

Figure 1 Gravel diagram

To analyze the rotating component matrix more clearly and conveniently and observe the attribution of various problem factors, this paper sets the operation of prohibiting the display of small coefficients and sets the absolute value of prohibiting the display of small coefficients to 0.5. According to the results of Table 4, it is found that the F1 factor has a great impact on X4, X6, X7, X8, X11, X12, X13, X14 and X15. According to the common characteristics of these indicators, we named them the industrial ecosystem factors. The F2 factor has a great impact on X1, X2, X9 and X10. According to the commonality, it is named as a government-driven factor. The F3 factor has a great impact on X3 and X16. According to the commonality, it is named as an industrial resource factor. The F4 factor has a great impact on X5. According to the characteristics of this index, we will use the industry platform factor obtained from factor analysis for regression analysis.

Table 4 Rotating component matrix

	Component part			
	1	2	3	4
policies and regulations		.551		
Management institutions		.694		
Data sources			.571	.550
Trading platform	.626			
Data platform				-.833
Technological innovation	.750			
Talents distribution	.774			
Industrial environment	.720			
Subject construction		.785		
Cultural education		.908		
GDP per capita	.723		.557	
IPv4 ratio	.964			
Leading Enterprises	.978			
Total number of enterprises	.700		.628	
The backbone enterprise	.983			
New enterprise			.789	

From the common factor deviation we can see the commonality of the index, and the information extracted from each original variable is given. From Table 5, we can see that the extraction degree of raw data by all indicators is more than 60%, and the extraction value of most indicators is greater than 80%, which shows that the common factors extracted by factor analysis have a strong ability to interpret each index.

Table 5 Common factor deviation

	initial	extract

Leading Enterprises	1.000	.968
The backbone enterprise	1.000	.977
Total number of enterprises	1.000	.940
New enterprise	1.000	.825
Policies and Regulations	1.000	.646
Industrial environment	1.000	.927
Data sources	1.000	.785
Management institutions	1.000	.633
Subject construction	1.000	.857
Data platform	1.000	.846
Trading platform	1.000	.715
Technological innovation	1.000	.965
Talents distribution	1.000	.957
GDP per capita	1.000	.860
IPv4 Ratio	1.000	.949
Cultural education	1.000	.883

3.3 Regression model establishment

Taking the four-factor values of the factor analysis results as the interpretation variable, Scale of the big data industry as the explained variable, SPSS 26.0 software is used for multiple regression analysis to discuss the relationship between the scale of the big data industry and the F1 industrial ecosystem (Ecosphere), F2 government promotion, F3 industrial resources, and Platform.

$$\text{Scale} = \beta_0 + \beta_1 \text{Ecosphere} + \beta_2 \text{Promotion} + \beta_3 \text{Resources} + \beta_4 \text{Platform} + \mu \quad (3)$$

Where μ is the error term and β0 is the constant term. Make a hypothesis:

H1：The industrial ecosystem has a positive impact on the scale of the industry.

H2：Government promotion has a positive impact on the scale of the industry.

H3：Industrial resources have a positive impact on the industrial scale.

H4：The industry platform has a positive impact on the scale of the industry.

Table 6 shows that R = 0.953 indicates a high correlation. The adjusted R square is 0.895, indicating that the four independent variables explain the 89.5% variation of the industrial scale of the big data industry, and the model fit reaches 89.5%.

Table 6 Model Abstract

					Change statistics				
Model	R	R Square	Adjust R Square	Errors in standard estimates	R Square Variation	F Variation	Degree of Freedom 1	Degree of Freedom 2	F Significance Variation
1	.953	.909	.895	7.45412	.909	64.629	4	26	.000

The data in Table 7 shows that F=64.629, P=0.000<0.001 in this case, which means that at least one independent variable explains the variation of some dependent variables, which makes the regression variation larger and the residual variation smaller, which means that the established multiple linear regression model is appropriate.

Table 7 ANOVA

Model		Sum of Square	Degree of Freedom	Mean Square	F	Significance
1	Regression	14364.209	4	3591.052	64.629	.000
	Residual	14444.663	26	55.564		
	Total Amount	15808.872	30			

Table 8 shows that the industrial ecosystem (b=16.277, β= 0.709, p<0.001), government promotion (b=4.470, β= 0.195, p<0.01), industrial resources (b=13.923, β= 0.067, p<0.001) will affect the industrial scale of the big data industry, because the regression coefficients of the industrial ecosystem, government promotion and industrial resources are all positive, so the industrial ecosystem, government promotion and industrial resources are all positively affecting the industrial scale of the big data industry. The industry platform (b=0.212, β= 0.009, p=0.877) Because it is not significant, there is no evidence to support that the industry platform will affect the industrial scale of the big data industry. Because of the factor analysis conducted in this paper, the VIF values of the four factors are all equal to 1, indicating that there is no collinearity between the factors.

Table 8 Coefficient

		Unstandardized coefficient		Standardization coefficient			Collinearity statistics	
Model		B	Standard error	Beta	t	Significance	Tolerance	VIF
1	(constant)	18.068	1.339		13.496	.000		
	Industrial ecosystem	16.277	1.361	.709	11.960	.000	1.000	1.000
	Government promotion	4.470	1.361	.195	3.284	.003	1.000	1.000
	Industrial resources	13.923	1.361	.067	10.230	.000	1.000	1.000
	Industry platform	.212	1.361	.009	.156	.877	1.000	1.000

It can be seen from the normal P-P diagram in Figure 2 of the model test that the points of the multiple linear regression model basically meet the distribution on and around the 45 degree line, indicating that the error term has good normality and the regression of the equation is good.

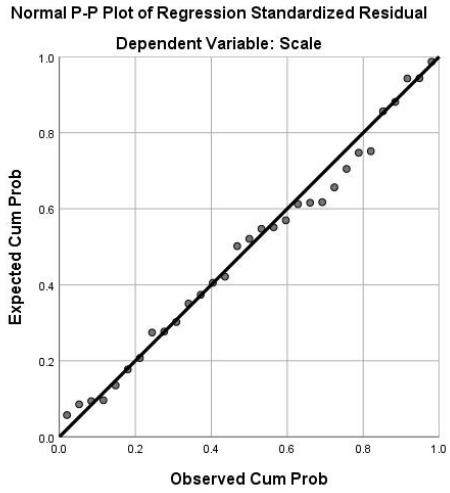

Figure 2 Normal P-P diagram

It can be seen from the scatter diagram in Figure 3 that the scatter points of this model are basically distributed on both sides of the 0 line and have no regularity, which means that the error terms are independent of each other and there is no heteroscedasticity.

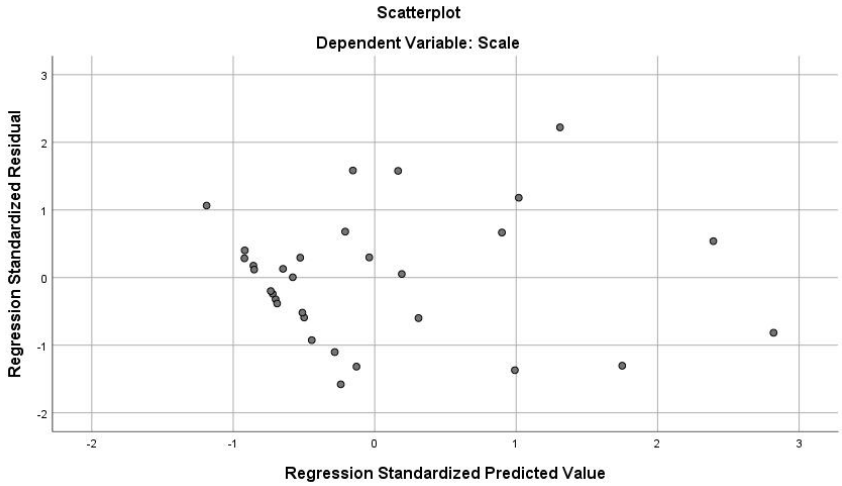

Figure 3 Scatter Chart

The measurement results of multiple regression analysis show that:

Assuming that H1, H2 and H3 are tenable, H4 is not tenable. The industrial scale of the big data industry is positively correlated with the industrial ecosystem, government promotion and industrial resources, of which the industrial ecosystem has the largest impact on the big data industry, and the industrial resources are the second important.

Assume that H1 is an industrial ecosystem that has a positive impact on industrial scale. First of all, an excellent industrial ecosystem can provide more big data talents for the big data industry, help big data enterprises occupy a more favorable position in the national competition, and effectively improve the comprehensive competitiveness of enterprises. Secondly, the industrial ecosystem can provide more and richer basic resources, which can effectively help enterprises integrate resources and achieve synergy. Thirdly, the industrial ecosystem can promote cooperation and exchange between upstream and downstream enterprises of big data, bring more cooperation opportunities for big data enterprises, and greatly promote the further establishment and development of the industrial ecosystem of big data industry. Finally, the dynamic industrial ecosystem can continuously innovate, provide new solutions for big data enterprises, and effectively promote the development of big data industry. In short, the industrial ecosystem has a great impact on the development of the big data industry. It can not only improve the competitiveness of enterprises, but also promote resource integration, collaborative innovation and other aspects, thus promoting the better development of the big data industry.

It is assumed that H2 government promotion has a positive impact on industrial scale. The positive impact of government promotion on industrial scale is helpful. First, the government can provide financial funds to support the development of big data and related industries, including technology research and development, talent training, data center construction, etc. Secondly, the government can provide policy support, such as tax incentives, land supply and other policies, to encourage enterprises to invest and develop big data and related industries. Third, the government can provide legal protection to ensure that issues such as data security and privacy protection are properly resolved and protect the legitimate rights and interests of enterprises. Fourth, government promotion helps to optimize the allocation of industrial resources and promote the optimal allocation of talents, funds, technology and other resources. Finally, the government promotes the establishment of a sound industrial ecosystem, including the construction of industrial chain, supply chain, innovation chain and other aspects. To sum up, the government's promotion has a positive impact on the industrial scale, mainly because it will not only affect the industrial structure, resource allocation and technological progress, but also promote industrial development and

innovation, and help optimize the allocation of industrial resources, promote the construction of industrial ecosystem, so as to achieve the long-term development of the big data industry.

It is assumed that H3 industrial resources have a positive impact on industrial scale. First of all, industrial resources can provide talents, funds, data and other resources needed by the industrial chain, and improve the industrial chain of the big data industry to improve the efficiency and efficiency of the big data industry. Secondly, industrial resources can provide technical support, improve the technical level of the big data industry, and promote the development of the big data industry. Finally, industrial resources can provide a good innovation environment, help big data enterprises develop new technologies and products, and promote innovation in the big data industry. In short, industrial resources have a great impact on big data and related industries. It can effectively promote the development of big data industry by influencing big data technology, industrial chain, innovation and other aspects.

Assuming that H4 is an industry platform has a positive impact on the industrial scale, because the empirical results of H4 are not significant, there is no evidence to prove that the industrial platform has an impact on the industrial scale of the big data industry. The industry platform factor only contains data platform indicators, so it is inferred that the data platform will not affect the development of the big data industry. The main reasons are the following two points. First of all, some regions lack data platforms. Although some cities have data platforms, they also have the problem of repeated positioning and fighting separately, which makes it difficult to form comprehensive advantages. Therefore, it is difficult to meet the long-term development needs of the big data industry. Secondly, the direction of data platforms in some cities is right, but the asymmetry of data supply and demand makes it difficult for data transactions to meet the effective needs of society. The data transaction rate and turnover are not high, resulting in unsatisfactory operation effect. Therefore, the data platform will not provide obvious help to the development of the big data industry. However, if we make good use of the opportunities brought by the data platform, we can provide more opportunities and resources for enterprises to promote the development of the big data industry.

3.4 Policy recommendations

Based on the above experimental data, the following suggestions are made in this paper. First, create an industrial ecosystem conducive to the development of the big data industry. It is necessary to increase the research, development and application of core big data technologies, such as using artificial intelligence, machine learning, block chain and other technologies to strengthen the research and development and application of key big data technologies, and promote the wide application of big data in various fields. At the same time, it is necessary to cultivate and introduce big data professionals and industry talents. To cultivate and introduce talents related to big data, we must first carry out curriculum construction and talent training related to big data technology in colleges and universities, strengthen cooperation with well-known universities, research institutions and enterprises, introduce high-level talents, and establish a talent training system that combines industry, university and research. In addition, it is necessary to strengthen the infrastructure construction of the big data industry to lay a solid foundation for the further development of the big data industry, thus providing favorable preconditions for future take-off.

Secondly, give more support and help to the big data industry in terms of industrial resources. It is necessary to increase investment in big data enterprises, and the government should actively guide and encourage the development of the big data industry, provide policy support in terms of capital, taxation and other aspects, and set up a big data industry investment fund to provide financial support for high-growth big data enterprises, and pay attention to cultivating advantageous enterprises. There should be corresponding resource support for new enterprises to help new enterprises survive smoothly, so as to promote the industrial scale and industrial cluster development of the big data industry.

Finally, the government should scientifically formulate policies and set up big data management institutions to effectively promote the development of the big data industry. First, government departments should clarify the plan for the development of the big data industry, formulate policies scientifically, clarify the direction of industrial development, give full play to the role of big data, and establish a perfect standard system, product system, data security and other relevant regulations and policies. In addition, it is necessary to establish and improve the big data supervision system, strengthen industry supervision, establish a special big data Management institutions, and carry out unified planning and management to support the development of the big data industry.

4. Conclusion

Using the combination of factor analysis and regression analysis, this paper analyzes the influencing factors of the big data industry from the perspective of government, industry, platform and society, reduces 16 indicators, such as management institutions, policies and regulations to 4-dimensional factors, connects factors with the industrial scale of the big data industry, and establishes a multivariate linear regression model. The empirical analysis is also provided to show that the industrial scale of the big data industry is positively correlated with three variables such as industrial ecosystem, government promotion and industrial resources. The importance is promoted by the industrial ecosystem, industrial resources and government from large to small. The selection of indicators and data in this paper is comprehensive, and accurate and does not redundantly cover the information that has an impact on the development of the big data industry. At the same time, the selection of factor analysis eliminates the problem of multiple collinearity caused by too many variables.

This paper fills some theoretical gaps in the research of influencing factors of the big data industry. It also provides theoretical support for the development and construction policies of 31 provinces and cities in China from different perspectives and provides a reference for the further addition and standardization of industry platforms. The disadvantage of this paper is that the research index data used for factor analysis in this paper are limited. Subsequent research will continue to search for more dimensions of research index data in the big data industry and further empirical research on the main influencing factors of the big data industry based on richer research data so as to provide a more effective basis for the future development of the big data industry in 31 provinces and cities in China.

Reference

[1] Chen Yue. Thinking about the human resources management model of enterprises in the era of big data [J]. Sichuan Architecture, 2022,42(06):291-293.

[2] Wang Shuguang, Song Jiayao. Research on Innovation in Government Governance in the Age of Big Data [J]. Business economy, 2023(01):152-155.

[3] Lu Ping. How to break through the bottleneck of the development of the big data industry [J]. China's industry and informatization, 2022(12):1-5.

[4] Fu W, Zhao J, Wang C, et al. Research on the Financial Support Performance Evaluation of Big Data Industry Development[J]. Mathematical Problems in Engineering, 2022.

[5] Zhang M, Cao C, Lv Z. Financial Technology, Big Data Enterprise Financing Constraints and Big Data Industry Development: Empirical Analysis Based on Mediating Effect and Threshold Effect[J]. Journal of Risk Analysis and Crisis Response, 2021, 11(4).

[6] Han Xianfeng, Hui Ning. Analysis of the technological efficiency of China's big data industry and its influencing factors [J]. Science and Technology Management Research, 2016, 36 (14):107-112.

[7] Zhou Ying, Liu Yue. Research on influencing factors for the development of the big data industry [J]. Modern Intelligence, 2017, 37(08):129-134.

[8] Shen Junxin, Li Shuang, Zhang Jingyang. Research on the Influencing Factors of the Development Capability of the Big Data Industry: Based on fsQCA Method [J]. Science and Technology Management Research, 2019, 39 (07): 140-147.

[9] Wang Qian, Li Tianyun, Du Yu. Mechanism analysis of the role of the government in the development of the big data industry [J]. Journal of University of Electronic Science and Technology (Social Science Edition), 2019, 21(04):49-58.

[10] Ma Liping, Zhang Lingjie. Use a big data sharing platform to help high-quality development of government governance [J]. Jiangnan Forum, 2021(03):7-9.

[11] Luo Tao. American experience in the big data industry and China's countermeasures [J]. High-tech and industrialization, 2013(05):54-57.

[12] Jia Chuanling, Wu Qi. Take advantage of big data management institutions to accelerate the construction of smart cities [J]. Mass, 2019(07):44-45.

[13] Kwon T H, Kwak J H, Kim K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea[J]. Technological Forecasting and Social Change, 2015, 96: 144-152.

[14] Wang Wei, Zhang Mengjun, Wang Jing. Research and analysis of big data trading platforms at home and abroad [J]. Intelligence Magazine, 2019,38(02):181-186+194.

[15] Wang Xinfeng, Zhang Yunhua. Data center industry development and capital layout [J]. Communication Enterprise Management, 2022(12):51-53.

[16] Wu Xue, Wu Junmin. Research on the cultivation of big data industrial clusters from the perspective of the industrial ecosystem: Take Suzhou as an example [J]. Journal of Changzhou University (Social Science Edition), 2015, 16(01):56-62.

[17] Fu Wei, tea is booming. Empirical analysis of the contribution rate of innovation factors in China's data industry [J]. Statistics and decision-making, 2019,35(11):129-131.

[18] Xia Dawen, Zhang Zili. Explore the big data talent training model in the DT era [J]. Journal of Southwest Normal University (Natural Science Edition), 2016, 41(09):191-196.

[19] Liu W. Research on the Practical Dilemma and Solution of Big Data Talents Cultivation[C]//2022 3rd International Conference on Big Data and Social Sciences (ICBDSS 2022). Atlantis Press, 2022: 776-783.

[20] Zhang Taiping. Strengthen the foundation of the big data industry and accelerate the development of the digital economy [J]. Northern Economy, 2018 (Z1):26-28.