https://dx.doi.org/10.22161/ijels.51.20 ISSN: 2456-7620

Using Big Data Analysis to Assist the choice of Leading Industries in County Areas

Mimi Ning[1], Guanglei Cui[2]

[1]Department of Economic, Shandong University, China
Email: ningm@buffalostate.edu
[2]Department of Scientific Research and Training, Shandong Big Data Center, China
Email: 379798223@qq.com

Abstract— *Big data plays an increasingly important role in government decision-making. Based on statistical data, this paper studies how to choose the leading industry at the county level by using a factor analysis method, and offers some suggestions for the local government to determine its leading industry. To avoid duplication of leading industries among counties, local governments should determine their leading industries scientifically and effectively and formulate industrial planning based on their resource endowments.*
Keywords— ***Big Data Analysis, Factor Analysis Method, Leading Industries.***

I. INTRODUCTION

With the continuous development and application of big data, more and more countries have raised data management to the strategic level, and big data is also entering the scope of public management. With the advent of the era of big data, the data resources produced and owned by the government are increasingly rich. When the government makes decisions, how to measure the effectiveness, pertinence, and operability of its decisions? This paper holds that the conclusions drawn from the collection, collation, and analysis of big data are of scientific and auxiliary significance. It is the future trend to improve the government management level by collecting big data, analyzing big data and applying big data, but collecting data and analyzing data does not mean that the problem is completely solved, and assisting the government to complete the decision-making is the ultimate goal.

In order to make use of big data to assist government decision-making, big data analysis and mining need to be carried out, and big data analysis and mining need to be in-depth analyzed according to specific objectives. The data itself contains a large amount of information. The big data mining tool is to show the information contained in the data itself and analyze the relationship between things according to our needs. The most basic role of big data mining is to let big data itself tell its story. The factor analysis method used in this paper is one of the big data mining tools, mainly through the mining of the information contained in the data itself, and then find out the most important influencing factors and correlation relations, and comprehensively summarize the regular quantitative characteristics, to provide guidance and reference for decision-making. Although this method is rigorous and scientific, as a quantitative analysis method, it cannot cover all factors of industrial development, so it needs to be qualitatively modified in practical application.

In this paper, factor analysis is used to assist the government in choosing the leading industry, which shows the application of big data mining method in the actual government decision-making. It has strong operability and replicability and has a high reference value in the field of government decision-making big data research.

II. DETERMINANTS OF LEADING INDUSTRY SELECTION

At present, when some counties are looking for their leading industries, they are easy to fall into the situation of blind choice or following the trend. These are limited to the availability of data, which is caused by human judgment factors. Taking Guangrao County of Shandong Province as an example, this paper uses the factor analysis method of big data to guide its selection of leading industries, which can objectively reflect the characteristics of Guangrao County's industries to a large extent, and this method of leading industry selection is also adaptable in other counties.

The county economy is complex and diverse. The determination of the selection standard of the leading industry needs to take into account both the generality and particularity, as well as certain pertinence and wide adaptability. Therefore, based on the existing county economic research, following the general law of the development of leading industries, combined with the

https://dx.doi.org/10.22161/ijels.51.20 ISSN: 2456-7620

actual development of the county economy, the factors to be considered in the selection of leading industries are determined as follows:

2.1 Comparative Advantage. The theory of comparative advantage originated from the classical economic theory of "comparative advantage", and then developed through "resource endowment", it has become a more rigorous theoretical system of logical reasoning and can be better used to select the leading industry. The difference between comparative advantage and absolute advantage mainly highlights the comparative advantage of the county industry. For example, in a county with backward economic development, there may not be an industry with absolute advantage or good economic benefits at the provincial level, but it does not mean that such a county cannot select its leading industry. In this case, based on the theory of comparative advantage, the leading industries in the county are those industries that have comparative advantage compared with other industries.

2.2 Economic Benefits. Economic efficiency is the final comprehensive index to measure all economic activities. The choice of leading industry should pay attention to the economic efficiency of the industry. Only with good economic efficiency can the leading industry play a better role in the regional economy. The market potential, technological innovation, and economic scale only provide conditions for a certain industry to achieve good benefits. Whether the industry can achieve benefits must rely on the internal management of the industrial organization. Therefore, the choice of county leading industries should also take economic benefits as an indispensable benchmark.

2.3 Industrial Scale. A suitable industrial scale is an important aspect that the government and the state need to consider when making industrial policies. The purpose of leading industry selection is not only for its own development and expansion, but also to promote the development of the whole national economy through its development. The driving role of leading industry to the regional economy depends on the inter-industry correlation effect, which depends on many factors. The economic factor is the economic scale of the industrial sector. If an industrial sector is too small, it cannot play a role in promoting the rapid development of other industrial sectors. Therefore, economic scale is the basis and guarantee for leading industries to play a role. Therefore, the economic scale should also be a basis for the selection of the leading industries in the county.

2.4 Market Potential. Market potential is an important indicator to determine the market development prospect of an industry and its influence on the economy. Market potential provides a prerequisite for the formation and development of leading industries. If an industry wants to become a leading industry, it must have a strong market expansion ability, so that it has the possibility of continuous development and expansion, and also can play a role in promoting the development of other industrial sectors. Therefore, county leading industries should be selected according to market potential.

III. USING DATA ANALYSIS METHOD TO LEADING INDUSTRY SELECTION

In the empirical analysis, generally, there is no case that all index values of one industry are higher than those of other industries. According to different indexes, industries will have a different order. Therefore, when analyzing the importance of industries, we need to comprehensively deal with the indexes. In this paper, the factor analysis method combined with the principal component analysis process of extracting common factors is suitable for the selection of county leading industries.

Factor analysis is a data analysis method that selects the most important influencing factors from multiple variables. Through the classification of observation variables, the variables with high correlation, that is, closely related variables, are classified into the same category, while the variables with different categories have low correlation. The method of factor analysis is easy to evaluate the evaluated object and classify the evaluation indexes. It has a clear practical significance and is easy to link with the objective economic phenomenon so that the quantitative analysis and qualitative analysis can be better combined. Based on the selection system of the county leading industry, this paper uses factor analysis to select the county leading industry.

3.1 Data cleaning. Calculate the correlation coefficient matrix, and get the correlation of indicators. Data cleaning is the first step of data processing, which is the basic processing of the overall data. Data cleaning in factor analysis is to calculate the correlation coefficient matrix. The purpose of data cleaning in factor analysis is to get the correlation of all indicators, that is, the relationship between indicators.

3.2 Data analysis. Calculate the variance of common factors and mine the information content of common factors. Data analysis is carried out based on data cleaning. Therefore, this paper measures the common factor variance based on the calculation of the correlation coefficient matrix. Through data analysis, we can find that one to four columns of common factor variance describe the characterization of the original variable population by the initial solution. Table1, the first column is the sequence

https://dx.doi.org/10.22161/ijels.51.20 *ISSN: 2456-7620*

number of initial solutions; the second column is the characteristic root of the principal component (or common factor), which is an indicator to measure the importance of the principal component. For example, 5.866 in the first row indicates that the first principal component characterizes 5.866 in the total variance of the original variable, which characterizes the largest variance, and the following characteristic roots decrease in turn, indicating that the ability of the principal component to describe the original variable decreases in turn; the third and fourth columns respectively represent the variance contribution rate and cumulative variance contribution rate of each principal component. It can be seen from table1 that the cumulative variance contribution rate of the first two common factors is 88.12%, indicating that the two common factors reflect 88.12% of the information of the original variable, and the number of main factors can be determined.

Table.1: Common factor variance

composition	Initial eigenvalue		
	Total	variance （%）	accumulate （%）
1	5.866	58.662	58.662
2	2.946	29.459	88.121
3	0.58	5.795	93.916
4	0.278	2.776	96.692
5	0.197	1.975	98.667
6	0.103	1.027	99.693
7	0.026	0.265	99.958
8	0.002	0.024	99.982
9	0.002	0.018	100
10	0	0	100

3.3 Data mining. The main purpose of data mining is to excavate the information behind the data at a deeper level, to make up for the missing information in data analysis. Therefore, on the basis of data analysis, we can further mine the data. According to the cumulative variance contribution rate, we can basically determine the two common factors. In order to further mine the data information and determine whether the number of factors is correct, after calculating the common factor variance, we use the gravel map to verify. A gravel map is used to determine the number of factors, which is often determined by some criteria in practical application. Because there is no precise quantitative method to determine the number of factors, some criteria are often used in practical application. There are several commonly used: 1. Eigenvalue criterion. That is to say, the principal component whose eigenvalue is greater than or equal to 1 is taken as the initial factor, and the principal component whose eigenvalue is less than 1 is discarded; 2. The cumulative variance contribution rate of factors, generally the number of factors selected should meet the cumulative variance contribution rate of more than 80%. For the data samples in this paper, the first two points represent the number of factors, so two common factors are selected.

3.4 Build the model. Calculate the component matrix and establish the measurement model. After data cleaning, data analysis, and data mining, the measurement model is established. The ultimate purpose of establishing the measurement model is to simulate and optimize the overall data and get the coefficient estimation results. In the factor analysis method, the weight determined according to the variation degree of the correlation among the indexes is objective. SPSS software is used to calculate the matrix coefficient of data components (see Table2).

Table.2: Composition matrix

	composition1	composition2
Location quotient	0.89	0.276
Comparison of capital profit and tax rate	-0.466	0.758
Asset profit tax rate	-0.446	0.78
Sales profit margin	-0.417	0.826
Profit margin of output value	-0.424	0.832
Output value scale	0.94	0.253
Fixed assets scale	0.885	0.192
Scale of profits and taxes	0.952	0.284
Employment scale	0.922	0.258
Market share	0.938	0.252

Finally, according to the sum of the component matrix coefficients and the common factor variance calculated in the data analysis, we establish the county leading industry selection model as follows:

$$C_1=0.890X_1-0.466X_2-0.446X_3-0.417X_4-0.424X_5+0.940X_6+0.885X_7+0.952X_8+0.922X_9+0.938X_{10} \quad (1)$$

$$C_2=0.276X_1+0.758X_2+0.780X_3+0.826X_4+0.832X_5+0.253X_6+0.192X_7+0.284X_8+0.258X_9+0.252X_{10} \quad (2)$$

$$F_i=58.662\%*C_{i1}+29.459\%*C_{i2} \ (i=1、2、\ldots10) \quad (3)$$

In which, C_1 and C_2 represent common factors, X_1-X_{10} represents 10 original variables such as location quotient and market share, and F_i represents the comprehensive evaluation value of an industrial sector, i represents industries.

According to the above analysis results, common factors have higher loads on many variables. The first common factor C_1 has a high load number in six variables, which basically reflects the scale index, location quotient and market potential of the industry; the second common factor C_2 basically reflects the industrial efficiency and comparative capital profit margin of the industry.

https://dx.doi.org/10.22161/ijels.51.20 *ISSN: 2456-7620*

3.5 The results. Through four steps of data cleaning, analysis, mining and modeling, the most important information contained in these indicators is extracted. At the same time, the factor analysis method is used to score all industries in the county, according to the ranking of the comprehensive score value to provide the basis for the selection of leading industries. In order to achieve this goal, according to the analysis principle of factor analysis method, factor analysis method in SPSS statistical analysis software package is used, common factors are extracted by principal component analysis method, correlation coefficient matrix, factor load matrix, etc. are calculated, and comprehensive evaluation value is finally obtained, according to which, the selection results of leading industries are obtained (see Table3).

Table.3: Comprehensive evaluation of industries

Industry	Score	ranking
Rubber and plastic products industry	4.28	1
Chemical raw materials and chemical products manufacturing industry	2.33	2
textile industry	1.93	3
Paper and paper products industry	1.43	4
Metal products industry	1.24	5
Petroleum processing, coking and nuclear fuel processing industry	0.65	6
Printing and recording media reproduction industry	0.53	7
Agricultural and sideline food processing industry	0.42	8
Automobile manufacturing industry	0.22	9
Production and supply of power and heat	0.18	10
Nonmetal mining and processing industry	0.15	11
Food manufacturing	0.13	12
Nonmetallic mineral products industry	0.10	13
Wine, beverage and refined tea manufacturing	0.08	14
Wood processing and wood bamboo, rattan, palm and grass products industry	0.07	15
General equipment manufacturing	0.07	16
Chemical fiber manufacturing	-0.02	17
Ferrous metal smelting and rolling industry	-0.05	18
Pharmaceutical manufacturing	-0.10	19
Special equipment manufacturing industry	-0.14	20
Non ferrous metal smelting and rolling industry	-0.19	21
Leather, fur, feather and their products and footwear industry	-0.21	22
Electrical machinery and equipment manufacturing industry	-0.41	23

IV. CONCLUSION

By comparing the analysis results with the actual selection, it is found that the leading industries selected by the factor analysis method are consistent with the actual selection, and the degree of compliance is more than 80%. The six leading industries determined by Guangrao County are consistent with those through the factor analysis, and the comprehensive evaluation scores are all in the forefront. It is scientific and effective to choose the leading industry by factor analysis. Only the automobile manufacturing industry ranked lower, ranking 9th, which was different from the leading industry established by the local county government. This is mainly because the automobile industry has a strong radiation driving role. The automobile industry is located at the upper end of the industrial chain, which can effectively drive the development of the lower industry of the industrial chain, such as the machinery manufacturing industry. Based on this consideration, it has a certain foresight for the local government to determine it as the leading industry.

REFERENCES

[1] Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). The economics of privacy". Journal of Economic Literature 54.2, pp. 442(92).
[2] Bundeskartellamt (2019). Case Summary: Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing. Germany: Bundeskartellamt.
[3] Liang Yan, Wang Qing. (2011) Selection of leading industries in Yangzhou Based on principal component analysis [J]. Forum on industry and technology (3).
[4] Prufer, Jens and Christoph Schottmuller (2017). Competing with big data". Working Paper.
[5] Sun Jiting, Meng Qingwu. (2012). Study on the selection of marine leading industries in the blue economic zone of Shandong Peninsula [J] China fishery economy (3).
[6] Wang Guangfeng, Wu Hongxia, sun Fengqin. (2015). Research on the selection of regional leading industries in the context of low carbon economy [J]. Business economy research (1).