A Study on the Level of e-Commerce Development in China based on factor analysis and cluster analysis

Na Long

Beijing Technology and Business University, Beijing 100000, China.

Abstract: The rapid development of e-commerce is of great significance in promoting the development of the real economy, accelerating the transformation of the industrial development mode and enhancing the strength of the regional economy. This paper first summarizes the relevant previous studies, measures the level of regional e-commerce development through three dimensions: the level of economic development, the level of regional informatization and the level of logistics and commerce, and establishes an index evaluation system; then, through factor analysis, the measurement indexes of e-commerce development level are downscaled, public factors are selected, and the public factors are named to calculate the score of each region's e-commerce development level; finally, according to the The cluster analysis was conducted based on the e-commerce development of 31 provinces and cities in China, and policy suggestions were put forward to promote the development of e-commerce in China.

Keywords: Electronic Commerce; Data Processing; Economic Development

1. Introduction and literature review

According to data from the National Bureau of Statistics, national online retail sales reached RMB 1,176.01 billion in 2020, up 10.9% year-on-year. The role of e-tailing in promoting consumption, stabilising foreign trade, expanding employment and protecting people's livelihood has been increasing, and e-commerce is playing an increasingly large role in economic life.

In recent years, there have been many scholars who have conducted more in-depth studies on e-commerce development. He Shaoli[1] constructs an interactive relationship model between the level of e-commerce development and the quality of economic development, proving the correlation between e-commerce development and economic development; Xie Hui[2] et al. believe that e-commerce development is closely related to the construction of modern information technology; Yang Xiaoguang[3] et al. analysed the role of the universal information network in promoting e-commerce activities. Therefore, on this basis, this paper studies the level of e-commerce development in China, evaluates the ability of e-commerce development in each region, and makes suggestions for the development of e-commerce in each region of China.

2. Source of data and description of variables

In recent years, many scholars have conducted research on the development level of e-commerce. Through the collation of relevant literature, the indicators that affect the development level of e-commerce mainly include the following aspects: First, the level of regional economic development. E-commerce has an obvious role in promoting economic growth, while the level of economic development will also in turn affect the development of e-commerce. Second, the level of information development. The level of regional informatization can have a very important impact on the development of e-commerce, while e-commerce is the booster of informatization construction. The third is the level of development of regional trade and logistics. As the artery of modern economic operation, logistics connects different economic sectors, and the level of development of logistics in a country directly reflects the modernization of its economy.[4] The level of logistics development directly reflects the modernisation of a country's economy. The level of logistics development determines whether online products can reach the customer, how long it takes to reach the customer and

the corresponding logistics costs paid by the customer, all of which can affect the customer's enthusiasm for online shopping. In summary, the level of economic development, the level of regional informatization and the level of logistics and commerce have a very important impact on the development of e-commerce, so this paper chooses these three as the first-level evaluation indicators. Meanwhile, on this basis, in order to comprehensively, objectively and reasonably reflect the level of e-commerce development in China, this paper selects nine evaluation factors that can comprehensively reflect the level of e-commerce development to establish a relevant statistical index system. The selected evaluation indicators are shown in Table 1.

The data used in this paper for each province and city is from the China Statistical Yearbook 2021, so the analysis is of the e-commerce development of each province and city in 2020.

Table 1 Indicators for evaluating the level of e-commerce development

Level 1 evaluation indicators	Secondary evaluation indicators	Indicator Code
Level of economic development	Gross regional product (billion yuan)	X1
	Disposable income per capita (yuan)	X2
	Fiscal revenue (billion yuan)	X3
Level of regional informatization	Number of computers in use (units)	X4
	Internet broadband access ports (million)	X5
	Length of fibre optic cable lines (km)	X6
Levels of logistics and commerce	Online retail sales (billion yuan)	X7
	Employment in the postal industry (persons)	X8
	Cargo volume (million tonnes)	X9

3. Data processing and analysis

This paper uses SPSS 23.0 to analyse and process the data. Firstly, descriptive statistics were conducted for the selected indicators. From the results of the descriptive statistics, it can be seen that the mean values of the number of computers in use and the length of fibre optic cable lines are large, 1863964.226 units and 1599315.097 km respectively. At the same time, the standard deviation of the regional GDP, the length of fibre-optic cable lines, the volume of freight transport and the number of computers in use is extremely large, indicating that the level of economic development, the level of logistics and commerce and the degree of informatization vary greatly between regions. Therefore, in this paper, based on the standardisation of the original data, the public factors were solved based on the correlation array and factor analysis was conducted to reduce the dimensionality of the indicators measuring the level of e-commerce development, followed by determining the main factors and calculating the component score coefficient matrix based on the cumulative contribution rate; then cluster analysis was conducted based on the evaluation results; based on the analysis of the data results, policies related to e-commerce development were discussed. The results of the descriptive statistics are shown in Table 2.

Table 2 Descriptive statistics results

	Minimum value	Maximum value	Average	Standard deviation
X1	1902.74	110760.94	32658.55	26661.81
X2	20335.10	72232.40	32086.38	12661.02
X3	220.99	12923.85	3230.43	2824.28
X4	40868.00	9914338.00	1863964.23	2124335.15
X5	219.00	8653.20	3051.77	2099.21
X6	33442.00	3990069.00	1599315.10	1050423.81
X7	116.20	25782.20	3750.12	5810.59
X8	2591.00	100660.00	29065.52	22461.91
X9	4091.00	374503.00	149776.81	102139.20

3.1 Factor analysis applicability test

This paper uses 2 tests, KMO test and Bartlett test, to determine the correlation between the variables.

The results of the KMO test showed that the KMO test for the data selected for this paper was 0.809, which was greater than the critical value of 0.6, indicating that these data were highly correlated and well suited for factor analysis. Meanwhile, the p-value of Bartlett's spherical test was 0.000, indicating that the original hypothesis of the covariance array being a unit array was rejected at the significance level of 0.01, making it suitable for factor analysis.

Table 3 Results of KMO test and Bartlett's test

Test name	Results
KMO values	.809
Bartlett's test	.000

3.2 Determine the number of factors and name them/

Principal component analysis was used as a method of constructing the factor variables and the factor loading matrix was rotated using the maximum variance method to obtain the results as shown in Table 4. It is obvious through Table 4 that before rotation, the variance contribution of the first 2 factors are 73.14% and 17.49% respectively, and the total variance explained by the first 2 factors is 90.63% cumulatively, which can basically extract the information contained in the sample. Therefore, the level of e-commerce development of each province and city in China can be comprehensively evaluated by the first 2 public factors, which are recorded as F1 and F2 respectively.

Table 4 Table of total variance explained

Ingredients	Extraction of sum of squares of loads			Sum of squared rotating loads		
	Total	Variance %	Cumulative %	Total	Variance %	Cumulative %
1	6.58	73.14	73.14	4.12	45.77	45.77
2	1.57	17.49	90.63	4.04	44.86	90.63

The paper then proceeds to obtain the table of factor rotation matrices shown in Table 5 by the maximum variance orthogonal rotation method.

Table 5 Factor rotation load matrix

	Ingredients	
	1	2
Gross regional product (billion yuan)	.806	.547
Disposable income per capita (yuan)	-.106	.909
Fiscal revenue (billion yuan)	.634	.756
Number of computers in use (units)	.505	.844
Internet broadband access ports (million)	.905	.369
Length of fibre optic cable lines (km)	.930	.103
Online retail sales (billion yuan)	.445	.843
Employment in the postal sector (persons)	.352	.863
Cargo volume (million tonnes)	.892	.163

From the rotated factor loading matrix, it can be seen that the first public factor has large loading values on X1, X5, X6 and X9, thus naming the first public factor as the e-commerce scale factor F1. A higher score on this factor means that the province or city has a high level of economic development, a high degree of information technology and a large scale of e-commerce business. The second public factor is named the current situation factor of commerce development due to its large loadings on X2, X3, X4, X7 and X8. The higher the score on this numerator, the more importance the region attaches to commerce development.

3.3 Calculating the overall score

Based on the factor scores and the corresponding eigenvalues, the composite factor score for each region can be calculated, and

the other composite factor is F, which is calculated by the following formula:

$$F=(4.12F1+4.02F2)/8.16$$

The 31 provinces, municipalities and autonomous regions were ranked according to the results of the calculations according to 2 public factor scores and the overall score.

3.4 Analysis of results

By constructing a factor evaluation model, this paper has derived the ranking of e-commerce development level of 31 provinces, municipalities and autonomous regions in China in 2020. According to the comprehensive score ranking, it can be seen that the regions ranked in the top 10 are Guangdong, Jiangsu, Zhejiang, Shandong, Shanghai, Beijing, Sichuan, Henan, Anhui and Hebei. They are divided into 3 categories based on their comprehensive scores, pioneer provinces in e-commerce development, medium provinces and potential provinces, the results of which are shown in Table 6.

Table 6 Regional distribution structure of e-commerce development level in China

Category	Region
Pioneer provinces	Guangdong, Jiangsu, Zhejiang, Shandong
Middle provinces	Shanghai, Beijing, Sichuan, Henan, Anhui, Hebei, Fujian, Hunan, Hubei
Potential provinces	Liaoning, Guangxi, Shaanxi, Chongqing, Yunnan, Shanxi, Jiangxi, Inner Mongolia, Heilongjiang, Tianjin, Guizhou, Xinjiang, Jilin, Gansu, Hainan, Ningxia, Qinghai, Tibet

And the analysis is carried out from the F1 score ranking. Provinces with scores greater than 1 include Guangdong, Jiangsu, Shandong, Sichuan, Zhejiang and Anhui, indicating that these six provinces have higher regional GDP, more internet access ports, longer fibre optic cable lines and larger freight volumes, and their e-commerce has a larger market and wider trade. Among them, Guangdong, Jiangsu, Zhejiang and Shandong also ranked in the top four in the overall score. Meanwhile, Beijing has the lowest F1 score of -1.82, while ranking 6th in the composite score, probably due to its low F1 score as a result of the pressure on Beijing's urban areas, tight urban land use and low intra-city freight volumes. Similarly, Shanghai's F1 score is -1.22, ranking 29th, again indicating its tight land resources and low intra-city freight volumes. Fujian, on the other hand, has an F1 score of -0.06 and is ranked 16th, probably due to its coastal nature, with less goods being transported by sea and lower freight volumes by land.

Analysing the F2 score rankings, Beijing, Guangdong, Shanghai and Zhejiang are the top four provinces and cities, indicating that these four regions have higher per capita disposable income, higher fiscal revenue, more computers in use, and higher online retail sales and postal employment, suggesting that the trade and commerce economies in these airports are developing well and the economic situation is better. Jiangsu, Tianjin, Fujian and Shandong follow in the rankings. Among them, Jiangsu's F2 score ranks fifth, with F1 and F ranked second, indicating that Jiangsu lacks in F2, probably due to the low number of people employed in the postal industry in Jiangsu Province; while Shandong's ranking eighth may be due to its larger population and lower per capita disposable income; Chongqing's score of F_2 is much higher than the score of F_1 and the overall score, probably due to the development of the online economy and the government's fiscal revenue The increase in

Analyzed from the overall score, Guangdong, Jiangsu, Zhejiang and Shandong ranked the top four, belonging to the pioneer provinces of e-commerce development, they are also the top 4 economic powerhouse provinces in terms of GDP in 2020, once again illustrating the mutually reinforcing relationship between the level of economic development and the level of e-commerce development. In addition, these four provinces are located in the eastern coastal region, in general, is also the eastern and southern e-commerce development level is higher, the western and northern e-provinces business development still has some room for upward mobility.

4. Cluster analysis

Using the Wald method based on squared Euclidean distances, the 31 provinces and cities were classified into three categories and the classification results obtained are shown in Table 7.

Table 7 Cluster analysis results

Category	Region
High level provinces	Guangdong, Jiangsu, Zhejiang, Shandong, Sichuan
Medium level provinces	Shanghai, Beijing
Potential provinces	Liaoning, Guangxi, Shaanxi, Chongqing, Yunnan, Shanxi, Jiangxi, Inner Mongolia, Heilongjiang, Tianjin, Guizhou, Xinjiang, Jilin, Gansu, Hainan, Ningxia, Qinghai, Tibet, Henan, Anhui, Hebei, Fujian, Hunan, Hubei

The results of the cluster analysis are roughly the same as those of the factor analysis evaluation system, with the difference that Sichuan has changed from a medium province to a high level development province, and the rest of the medium provinces have been classified as potential provinces, except for Beijing. This is because some of the indicators selected do not directly reflect the level of e-commerce development, but more reflect the strength of the economy, so with the exception of Guangdong, Jiangsu, Zhejiang, Shandong Sichuan, Shanghai and Beijing, most provinces and cities are classified as potential provinces.

A one-way ANOVA was conducted on the means of the variables for each group based on the subgroups and the results are shown in Table 8.

Table 8 One-way ANOVA results

Variables	F	Significance
Gross regional product (billion yuan)	25.441	.000
Disposable income per capita (yuan)	52.299	.000
Fiscal revenue (billion yuan)	34.402	.000
Number of computers in use (units)	30.650	.000
Internet broadband access ports (million)	32.607	.000
Length of fibre optic cable lines (km)	20.930	.000

Continued from Table 8

Variables	F	Significance
Online retail sales (billion yuan)	22.500	.000
Employment in the postal industry (persons)	21.031	.000
Cargo volume (million tonnes)	7.430	.003

As can be seen from Table 8, the means of the nine variables are significantly different at the 0.05 level of significance in all three subgroups and the groupings are appropriate.

The results of the sub-groups illustrate the similarity of the provinces on the three main indicators used to evaluate the level of e-commerce development. For the high level provinces, they have a high level of economic development and a high level of information technology and trade logistics; for the medium level provinces, Beijing and Shanghai, both of them have similar score rankings in F1, F2 and F. Both of them have a high level of economic development, but due to the small amount of land resources and policy reasons, they mainly develop trade in the urban areas and have a low freight volume; the potential provinces, due to their geographical location or economic development level, still There is still more room for development.

5. Policy recommendations

First of all, each province and city can identify their own shortcomings and improve them according to the factor evaluation system. The factor evaluation system of e-commerce development level selects three primary indicators, namely, the level of economic development, the degree of informatization and the level of development of trade and logistics, and sets up three specific secondary indicators under each primary indicator, so that each province and city can use each secondary indicator as a guide according to their own situation, and strive to improve their own e-commerce development level, so as to drive residents' consumption and economic growth.

Secondly, observing the large geographical differences between the levels of e-commerce development in China, a one-to-one support relationship can be established between the various pioneering provinces and cities on the eastern coast and the relatively less developed inland provinces and cities, such as relying on the vast land or production raw material resources of the inland provinces to sell through the vast market in the east, thus driving the development of e-commerce and the improvement of the economic level.

Finally, improve the construction of logistics infrastructure and promote the development of e-commerce with logistics development. Based on the impact of the development of trade logistics on the level of e-commerce development, each region should integrate the existing logistics infrastructure, improve the degree of information technology of the infrastructure, and promote the infrastructure more in line with the needs of the times[5] For example, in the organisation of transport, the main hubs will be more efficient. For example, in the transport organisation, the main hub cities should be quickly connected to each other to achieve fast transport between provinces and cities and improve the efficiency of transport and distribution. The rational planning and conservation of land, the formation of industrial clusters, the improvement of logistics service systems, and the strengthening of the radiation function of logistics centres, in order to enhance the level of logistics development, while promoting the development of e-commerce, to achieve the interaction and coordinated development between the two.

References

[1] He SL. Study on the interaction between the level of e-commerce development and the quality of economic development in China[D]. Guizhou Normal University,2019.

[2] Xie H, Ren H. Analysis and research on the current situation of e-commerce development in China and its influencing factors[J]. Mall Modernization, 2021, (13): 25-27.

[3] Yang XG, An analysis of economic trade under network information technology [J]. Global Market, 2018, 26(4): 13.

[4] Cheng LL. Research on the mutual influence of e-commerce and logistics service quality [D]. Jiangxi University of Finance and Economics, 2019.

[5] Wang SS. An empirical analysis of China's e-commerce development level based on factor analysis[J]. Journal of Xinyang Normal College (Philosophy and Social Science Edition), 2019, 39(03): 34-38.

About the author:Long Na (2004.09-), female, Han nationality, Hunan Province, undergraduate, Beijing Technology and Business University, Time series analysis.