An Empirical Study on the Influencing Factors of Big Data Industry in Guizhou

Yuting Fu(✉)

Dalian Polytechnic University, Dalian, China
974431118@qq.com

Abstract. The Party and the State attach great importance to the development of the big data industry, and from 2014 onwards, the development of big data was incorporated into the work deployment at the national level. Subsequently, programmatic documents for the promotion and development of big data industry as well as relevant industrial development plans have been formulated and released in the name of the State Council and national ministries, elevating big data to a national industrial development strategy. This paper aims to study the influencing factors of the development of big data industry in Guizhou Province, and to identify the factors that really affect the development of big data industry in Guizhou Province by building a model and testing the heteroskedasticity and autocorrelation of the influencing factors. The data query for the influencing factors was obtained from the public data provided by the National Bureau of Statistics. By testing the validity of the model, the factors influencing the development of the big data industry in Guizhou province are analysed and suggestions for the future development of the big data industry in Guizhou province are given based on the influencing factors.

Keywords: big data industry · influencing factors · empirical analysis

1 Introduction

The development of the big data industry, entering a new era, has received great attention from the Party and the State. At the 19th Party Congress held in 2017, clear requirements and special arrangements were made to develop the big data industry, especially to promote big data technology to empower traditional industries and the real economy, and to achieve comprehensive and deep integration of big data and other technologies with the real economy [1]. The National 13th Five-Year Plan and 14th Five-Year Plan also regard the big data industry as a strategic and leading industry for China's economic and social development in the new era, and attach great importance to the development of the big data industry as a national industrial development strategy and make relevant action deployment [2]. By reviewing the available literature and research reports on the big data industry in Guizhou, this thesis compares and analyses the relevant practices and initiatives in the development of the big data industry in Guizhou Province and other

© The Author(s) 2023
V. Gaikar et al. (Eds.): ECIT 2023, AHE 18, pp. 212–219, 2023.
https://doi.org/10.2991/978-94-6463-210-1_27

regions, analyses and studies the relevant factors affecting the development of the big data industry in Guizhou Province, endeavors to grasp the general laws that promote the benign development of the big data industry, and through the analysis of these influencing factors and the laws of industrial development Through the analysis of these influencing factors and the grasp of the laws of industrial development, we will take them as the entry point and put forward some rationalized suggestions for the next step to enhance and develop Guizhou's big data industry, so as to promote Guizhou's big data industry to actively cope with the challenges, maintain its advantages and avoid its disadvantages, so as to achieve better and faster development in the new era and lead and drive the transformation and upgrading of Guizhou's economic and social development [3].

2 Building a Multiple Linear Regression Model

2.1 Identification of Explanatory and Explained Variables

The sample data of indicators of the model are shown in Table 1. Above, and the variables of the model are standardized by using spss, [4] and the results are shown in 1. The standard deviation of GDP is 3122.82846, the standard deviation of industrial output value of high-tech industry in Guizhou Province is 394.06663, and the standard deviation of the number of enterprises related to big data industry in Guizhou Province is 3275.49679. From the above, we can see that the standard deviation statistics between variables are still quite different, so it is necessary to standardize the selected variables. From Table 2, we can see that the kurtosis and skewness distributions of the industrial output value of Guizhou's big data industry, Guizhou's GDP and the industrial output value of Guizhou's high-tech industry are all close to zero, so we can judge that these indicators do not deviate too much from the normal distribution.

Table 1. Influencing Factors for the Development of Big Data Industry in Guizhou Province

Year/Sh adow Impact Factors	Total Industrial Output Value of Big Data in Guizhou Province (in billions of dollars)	GDP of Guizhou Province (Single) (Position: $ billion)	Industrial production of high-tech industries in Guizhou Province Value (in billions)	Big Data Industry Related Enterprises in Guizhou Province Number (in units)
2013	61.13	8006.8	518.96	1401
2014	130.41	9266.39	638.31	1907
2015	309.57	10541	889.72	3042
2016	510.47	11792.25	1209.54	4851
2017	766.09	13540.83	1603.95	8080
2018	612.54	14806.45	1164.95	8179
2019	632.76	16769.34	1380.87	9344

Data source: Peking University Digital Inclusion Index.

Table 2. Descriptive statistics of factors influencing the development of big data industry in Guizhou Province

Descriptive Statistics								
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Total industrial output value of big data	7	61.13	766.094	3185e2	268.99906	-.340	.794	-1.605	1.587
GDP	7	8006.80	16769.34	1.2103E4	3122.82846	.221	.794	-1.062	1.587
High-tech industrial output Value	7	518.96	1603.951	0580e3	394.06663	-131	.794	-1.208	1.587
Number of related businesses	7	1401.00	9344.005	2577e3	3275.49679	.064	.794	-2.211	1.587
Valid N (listwise)	7								

1) Regression analysis.

Table 3. Regression analysis results

Model		Sum of Squares	df	Mean Square	F	Sig
	Regression	427483.301	3			
	Residual	6679.675	3			
	Total	434162.977	6	142494.434		
1	Regression	426904.480	2	2226.558	63.998	0.003a
2	Residual	7258.496	4	213452.240	117.629	0.000b
3	Total	434162.977	6	1814.624	147.481	0.000c
	Regression	419926.365	1	419926.365		
	Residual	14236.612	5	2847.322		
	Total	434162.977	6			

Table 4. ANOVA table of factors influencing the development of big data industry in Guizhou Province

ANOVAb				
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	419926.365	1	419926.365	147.481	.000a
Residual	14236.612	5	2847.322		
Total	434162.977	6			

a. Predictors: (Constant), Industrial output value of high technology industries

b. Dependent Variable: Big data industrial output

2.2 Significance Tests

1) F-test of model regression coefficients.

As shown the Table 3 an ANOVA table of the factors influencing the development of big data industry in Guizhou province [5]. The data in this table shows that the mean variance of the regression is 419926.365, the remaining mean variance is 2847.322, the observed value of the F-test statistic is 147.481, and the corresponding probability P-value is 0.000, which is less than 0.05, so it is concluded that the explanatory variable, the total industrial output value of the big data industry in Guizhou province (Y), and the explanatory variable, the high-tech industry in Guizhou province industrial output value (X2) has a linear relationship.

2) t-test of model regression coefficients.

2.3 Residual Analysis of the Model

As shown the Tables 2, 3, 4, 5 and 6 the residual statistics of the linear regression model of the factors influencing the development of the big data industry in Guizhou Province. The following table shows the statistics: the minimum value of the predicted value is 69.946, the maximum value is 798.3411, the median value is 4.3185, the standard deviation is 264.55194, the number of cases is 7, and the standardised predicted value, the standard deviation of the predicted value, the adjusted residual of the predicted value, the residual values, standard residuals, etc. Table 2, 3, 4, 5 and6 Residual statistics of the linear regression model of factors influencing the development of big data industry in Guizhou Province.

As shown the Table 7 the scatter plot of the linear regression equation of the influencing factors of the development of big data industry in Guizhou Province [6]. From the Table, it can be seen that all the scatter points are basically distributed around the

Table 5. Regression coefficients for the analysis of factors influencing the development of big data industry in Guizhou Province

Coefficientsa						
	dUnstandardiz	Standardized Coefficients	t	Collinearity Sig.	Statistics	
Model						
Coefficients						
	StdB Error	Beta	Tolerance	VIF		
(Constant)	61.869		-4.501	.006		
-278.452 1						
Industrial output value of high technology industries.671	.055	.983	12.144	.000	1.000	1.000
a. Dependent Variable: Big Data Industry						
Total output						

Table 6. Goodness-of-fit tests of the model regression coefficients

Model Summaryb					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.983a	.967	.961	53.36031	2.528
a. Predictors: (Constant), Industrial output value of high technology industries					
b. Dependent Variable: Big data industrial output					

diagonal line, and this phenomenon shows that the predicted values of the linear regression equation are very close to the observed values, indicating that the linear regression model established for the influencing factors of the development of big data industry in Guizhou Province has a good prediction effect.

Table 7. Residual statistical results

Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	69.9460	798.3411	4.3185E2	264.55194	7
Std. Predicted Value	-1.368	1.385	.000	1.000	7
Standard Error of Predicted Value	21.016	36.297	27.860	6.600	7
Adjusted Predicted Value	77.2991826.11524.3712E2	267.89790	7		
Residual	-3.22511E11	.08916E2	.0000048.71107		
Std. Residual	-.6042.041	.000	.913		
Stud. Residual	-.8252.221	-.0411.019	7		
Deleted Residual	-6.00252E11.28914E2	5.2707061.23802	7		
Stud. Deleted Residual	-.79316.9372.0886.551	7			
Mahal. distance	.0741.919	.857	.800	7	
Cook's Distance	.004	.453	.123	.177	7
Centered Leverage Value	.012	.320	.143	.133	7

a. Dependent Variable: Big Data Gross Industrial Product

3 Conclusion

Through the empirical analysis in Chapter 4, it can be concluded that among the factors affecting the development of big data industry in Guizhou Province, the most important influencing factor is the level of development of high-tech industry in Guizhou Province [7]. Secondly, the development level of Guizhou province's economy and the number of enterprises related to Guizhou province's big data industry, which are the two influencing factors.

All of these factors also have a greater or lesser impact on the development of the big data industry in Guizhou Province.

Based on the above conclusions, combined with the understanding of the past development of big data in Guizhou and the practices of other regions in developing big data industry, the following recommendations are made to accelerate the development of big data industry in Guizhou Province in the next step [8].

References

1. Outline of Planning for the Development and Application of Big Data Industry in Guizhou Province (20142020)
2. HIGUCHI T. Data scientist: a key factor in innovation driven by big data [J]. Journal of Information Processing & Management, 2013,56(1):2-11.
3. Neal Caren and Aaron Panofsk y, TQCA: A Technique for Adding Temporality to Qualitative Comparative Analysis, Sociolog y Methods & Research, Vol. 34 , No.2, 2005,pp.147 -172.
4. Ministry of Industry and Information Technology of the People's Republic of China . Development plan of big data industry (2016–2020) [Z]. Beijing : Ministry of Industry and Information Technology , 2016.
5. ChaoLe Men, Ma Guanghui, Lu Haijuan. Analysis of the characteristics of China's big data industry and policy recommendations [J]. Intelligence Theory and Practice, 2016,39(10): 5-10.
6. Cui Xiaohui , Wu New Year . Big data applications to promote big data industry on the ground [J]. Science and Technology Management Research ,2016,36(2):203–207. [11] Delia. Research on the development of big data industry in China [J]. Science and technology progress and countermeasures ,2014(4):56–60.
7. Zou Guangsheng. Reflections on promoting cross-strait big data industry cooperation [J]. Strait Science, 2020, 000(001):54-56.
8. Pang Lujing, Zhang Mu. A study on the coupled and coordinated development of China's provincial big data industry and financial agglomeration[J]. Science and Technology Management Research, 2020(15).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.