Research on the Impact of Big Data Development on Green Economy—An Empirical Study Based on Panel Data from Beijing-Tianjin-Hebei Region

GuanHui Wang *[,1,a] XiaoYu Zhou[1,b] XiaoDi Fan[1,c]

[1]School of International Business,Tianjin Foreign Studies University, Tianjin, 300204, China

Email: [a]newhop@126.com ,[b]zxyhebut@163.com
[c]1625085934@qq.com

Abstract. This study utilizes panel data from large-scale enterprises in the Beijing-Tianjin-Hebei region from 2013 to 2021. R software is used to perform Winsorize processing on the data. By constructing a index system for the devel-opment of big data, the relationship between big data development and the green economy in the region is analyzed. The research findings indicate that the development of big data plays a positive role in promoting the green economy. Firstly, big data technologies provide abundant data resources for enterprises and decision-makers, driving green innovation and technological development. Secondly, big data technologies help optimize resource utilization efficiency, achieving sustainable development in the green economy. This study provides references and insights for the development of the green economy in the Beijing-Tianjin-Hebei region and other areas, emphasizing the significant role of big data in driving the transformation of the green economy.

Keywords: big data;green economy; Beijing-Tianjin-Hebei;R software;

1 Introduction

As a core region of China's economic development, the Beijing-Tianjin-Hebei area holds significant importance in the transformation towards a green economy (GE). The GE, as a significant topic in response to the 2008 financial crisis and the 2012 United Nations Conference on Sustainable Development, has received widespread attention from the international community. It is a sustainable economic model aimed at achieving a win-win situation of economic growth and environmental protection. The concept of the GE is highly controversial. Merino et al. (2020) selected 150 definitions of the GE and green growth through extensive literature review, and through network analysis, they argue that the core of the GE lies in achieving a potential balance and synergy between the economy and the environment, while also addressing social issues, which undoubtedly supports the research conclusions of Rennings and Jaffe[1][2][3]. In terms of measuring the efficiency of the GE, Li Ruzi et al. (2019) used the DEA model to measure the efficiency level of the GE in practical operations[4] ; Li Qingshui et al. (2021)

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_21

considered factors such as economic structure and economic development level to construct a multi-level index system for evaluating the level of the GE[5] ;However, the GE still faces many challenges, primarily the problem of information asymmetry. Enterprises face information asymmetry with demand-side when making R&D in-vestments, which leads to the inability to translate R&D achievements into market value, ultimately causing enterprises to reduce green R&D investment (Sinnandavar 2018)[6].

The development and application of big data technology can alleviate the aforementioned problem of information asymmetry to some extent and promote the improvement of enterprises' level of green innovation. Through big data technology can help enterprises analyze and predict market trends more effectively, providing more accurate guidance and decision support for green innovation. J Wang (2022), Internet development significantly contributes to green economy growth in China's prefecture-level cities[7]. Ai Yongfang (2021), Big data development promotes green innovation in enterprises by facilitating market demand matching, alleviating financing constraints, and enhancing government subsidies[8]. Wu Yingning (2022) found regional disparities in big data development, but no significant regional variations in green economic efficiency in mainland China[9].

2 Research Design

2.1 Sample Selection and Data Sources

The Beijing-Tianjin-Hebei Big Data Comprehensive Experimental Zone was offi-cially established in 2016 with the aim of using data flow to guide the flow of technology, materials, funds, and talents, supporting cross-regional public ser-vices, social governance, and industrial transfers, and promoting the integrated development of the Beijing-Tianjin-Hebei region. After five years of construction, the experimental zone has achieved significant results. This study selects panel data from the Beijing-Tianjin-Hebei region from 2013 to 2021. The data on green patents of enterprises are manually collected from the National Patent Database. The data on enterprises and regional indicators in the Beijing-Tianjin-Hebei region are obtained from the Wind database. The policy reports related to big data are manually collected from the official websites of the local governments in the three regions. To handle extreme values and skewed distributions, this study uses the R software to perform Winsorization and logarithmic transformation on the relevant variables.

2.2 Variable Selection and Indicator Construction

Dependent Variable: Green innovation level. Due to the lack of a dedicated da-tabase for GE-related data, this study adopts a keyword screening method to collect relevant data from the National Patent Database.

Independent Variable: Big data development index. Since 2016, several research institutions in China have successively released research results on the development of big data indices. However, the indices compiled by these institutions lack continuity, and there are differences in the calculation methods used, making it difficult to apply the index data available in the market to empirical research. Therefore, this study borrows the compilation ideas of existing big data development indices and compiles the Beijing-Tianjin-Hebei big data development indices from 2013 to 2021 based on data availability, as shown in Table 1.

Table 1. Index System for Big Data Development in the Beijing-Tianjin-Hebei Region

Primary Indicator	Secondary Indicator	Representative Indicator	Indicator Definition
Institutional Development	Policy Intensity	Number of Policies Related to Big Data Released	Reflects the government's enthusiasm in promoting big data development
Enterprise Development	Technological Investment	Proportion of R&D Expenditure by Large Industrial Enterprises to GDP	Reflects the research and development investment in big data development
Enterprise Development	Related Enterprise Scale	Proportion of Enterprises Engaged in E-commerce Transactions	Reflects the level of attention given by enterprises to big data
Civil Development	Related Industry Scale	Proportion of Software and Information Technology Service Industry Revenue to GDP	Reflects the development level of related industries in big data
Civil Development	Terminal Accessibility	Number of Computer Users per 100 People	Reflects the convenience of personal data collection
Civil Development	Consumer Capacity	Proportion of Household Expenditure on Transportation and Communication to Total Consumption Expenditure	Reflects the public's capacity for information consumption

Drawing on the construction of the big data evaluation index system by Ai Yongfang (2021) and Joubert[1] (2023) mentioned earlier and expanding it appropriately, this study selects six indicators from the perspectives of institutional development, enter-prise development, and civil use development[10]. These indicators are policy intensity, technological investment, related enterprise scale, related industry scale, terminal penetration, and consumption capacity. The entropy method is then used to weight the indicators, with "lnbigdata" representing the level of big data development in the Bei-jing-Tianjin-Hebei region.

Control Variables: In addition to the main independent and dependent variables, this study includes control variables to account for other factors that may influence the research results. These control variables include the GDP of the Beijing-Tianjin-Hebei region, the growth rate of enterprises (grow), the level of financial leverage (lev), and the innovation capability of enterprises (nip).

2.3 Model Construction

To investigate the impact of big data development on the GE in the Bei-jing-Tianjin-Hebei region, this study constructs the following regression model:

$$lnegil_{ij}=C+\beta_1 lnbigdata_{ij}+\beta_2 gdp_{ij}+\beta_3 grow_{ij}+\beta_4 lev_{ij}+\beta_5 nip_{ij}+\varepsilon_{ij} \quad (1)$$

[1] Joubert A, Murawski M, Bick M. Measuring the big data readiness of developing countries–index development and its application to Africa[J]. Information Systems Frontiers, 2023, 25(1): 327-350.

where i and j represent provinces and years, respectively, and εij denotes the unobserved individual effects. If $\beta 1 > 0$, it indicates that big data development can promote the growth of the GE in the region.

3 Empirical analysis

3.1 Basic Regression

Using the R software for econometric estimation, the parameter estimates of the empirical model are presented in Table2. In this analysis, we first examine the impact of big data development index (lnbigdata) on green innovation level (lnegil) in Model 1. Then, we add control variables in Models 2-9. Only the models with significant p-values at the 1% level are shown in Table 2. It can be observed that in Models 1-9, with lnegil as the dependent variable and lnbigdata as the explanatory variable, the regression coefficients of lnbigdata are all positive and significant at the 1% level.

Table 2. Big Data Development Affects Green Economy Regression Results

Var	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6	Mod7	Mod8	Mod9
lnbigdata	0.466***	0.444***	0.348***	0.340***	0.414***	0.292***	0.286***	0.265***	0.263***
	(10.37)	(9.05)	(4.93)	(4.73)	(6.84)	(3.9)	(3.77)	(4.26)	(4.13)
gdp					0.001	-0.01*	-.001*	-.003***	-.003***
					(0.86)	(-1.77)	(-1.68)	(-3.05)	(-2.91)
grow								0.004***	0.004***
								(3.32)	(3.14)
lev		0.009		0.007	0.008		0.060		0.020
		(-1.08)		(-0.83)	(-0.98)		(-0.72)		(-0.32)
nip			0.006**	7.214***		0.001**	0.001**	0.002***	0.002***
			(2.05)	(12.38)		(2.57)	(2.41)	(3.86)	(3.64)
C	6.392***	6.998***	6.789***	7.214***	6.963***	7.362***	7.697***	7.237***	7.364***
	(29.38)	(11.59)	(24.13)	(12.38)	(11.44)	(17.51)	(12.25)	(20.8)	(13.8)
F	10.87	10.07	13.89	12.48	2.56	5.5	5.57	4	7.74

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$. t-values are in parentheses

3.2 Robustness Test and Endogeneity Discussion

To ensure the accuracy and reliability of the research results, this study conducted a robustness test to verify the robustness of the regression results. The approach used was the Least Absolute Deviation (LAD) estimation, which helps to reduce the influence of outliers or extreme observations on the estimation of regression coefficients, thereby improving the robustness of the regression results. The specific results are shown in Table 3.After re-estimating the model, it was found that the regression coefficient of lnbigdata remained positive and significant at the 1% level. This further supports the findings of the baseline regression results.

Table 3. Robustness check and Endogeneity issue

Var	Robustness check	Endogeneity issue
lnbigdata	0.366***	0.635***
	(12.3)	(11.6)
gdp	-0.001***	-0.005***
	(-7.09)	(-3.89)
grow	0.002***	0.005**
	(12.72)	(2.72)
lev	-0.009**	-0.04
	(-2.3)	(-0.46)
nip	0.001***	0.003***
	(3.42)	(6.77)
C	7.190***	8.759***
	(32.38)	(15.81)

4 Conclusion

This study analyzed the impact of big data development on the green economy (GE) in the Beijing-Tianjin-Hebei region from 2013 to 2021. The findings indicate a positive relationship between big data development and the GE. The utilization of big data technologies has facilitated environmental innovation, improved re-source utilization efficiency, and enhanced waste management, thereby promoting the development of the GE.:

A. big data technologies provide businesses and organizations with more comprehensive and accurate environmental data and information. Through anal-ysis and mining of big data, patterns and trends of environmental issues can be revealed, helping businesses and decision-makers to better understand environ-mental challenges and opportunities.

B. big data technologies enable fine-grained and intelligent monitoring and optimization of resource acquisition, utilization, and management. Real-time monitoring and analysis of data can accurately predict resource demand, optimize resource allocation, and achieve efficient resource utilization.

Based on the conclusions of this study, the following implications can be drawn: Big data technologies offer valuable insights and information for businesses and decision-makers, facilitating the development of green technologies and solutions. By optimizing resource allocation and enhancing environmental management, big data contributes to efficient resource utilization and sustainable economic growth. Collaboration and data sharing among stakeholders are vital for maximizing the benefits of big data in promoting a greener and more prosperous future.

Acknowledgement

The research is supported by 2021 Tianjin Educational Science Planning Project A Study of the Collaborative Development of Higher Education Clusters in the Beijing–Tianjin–Hebei (BTH) Region (No.FIE210052)

References

1. Rennings ,K.Towards a Theory and Policy of Eco-lnnovation-Neoclassical and(Co-) Evolutionary Perspectives[J].ZEW Discussion Papers,1998,35(1):23-32.
2. Jaffe, A. B, Palmer ,K. Environmental Regulation and lnnovation; A Panel Data Study[J].Review of Economics & Statistics,1997,79(4):610-619.
3. Merino-Saum A, Clement J, Wyss R, et al. Unpacking the Green Economy concept: A quantitative analysis of 140 definitions[J]. Journal of cleaner production, 2020, 242: 118339.
4. Li, R., & Pan, K. (2019). Spatial-temporal changes in urban green economy efficiency and influencing factors in Jiangxi Province. Ecological Economy, 35(5), 100-104.
5. Li, Qingshui, Li Dengfeng, Li Hui, et al. Multi-index evaluation of regional green economic development level based on prospect theory. Operations Research and Management Science, 2021, 30(6): 118-123.
6. Sinnandavar C M, Wong W P, Soh K L. Dynamics of supply environment and information system: Integration, green economy and performance[J]. Transportation Research Part D: Transport and Environment, 2018, 62: 536-550.
7. Wang J, Wang W, Ran Q, et al. Analysis of the mechanism of the impact of internet development on green economic growth: evidence from 269 prefecture cities in China[J]. Environmental Science and Pollution Research, 2022: 1-15.
8. Ai, Y., & Kong, T. (2021). Can Regional Big Data Development Promote Green Innovation in Enterprises? Journal of Zhongnan University of Economics and Law, (06), 116-126+160. DOI: 10.19639/j.cnki.issn1003-5230.2021.0072.
9. Wu Yingning, Wang Zhanglin. Research on the Coordination Degree of Big Data Development and Green Economic Efficiency: Based on Provincial Panel Data[J]. Journal of Changchun University, 2022, 32(09): 21-30.
10. Joubert A, Murawski M, Bick M. Measuring the big data readiness of developing countries–index development and its application to Africa[J]. Information Systems Frontiers, 2023, 25(1): 327-350.