Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 21, No. 1, 2026

Research on the Impact Effect of Data Factor Agglomeration Empowering Industrial Chain Resilience

Jili Tong[1, *]

[1] School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China
[*] Corresponding author: Jili Tong (Email: 623742044@qq.com)

Abstract: This study examines how data factor agglomeration empowers industrial chain resilience, using the establishment of national big data comprehensive pilot zones as a quasi-natural experiment. Based on panel data of 266 Chinese cities from 2009 to 2022, we apply the multi-period difference-in-differences (DID) and double machine learning (DDML) methods. Results show that data factor agglomeration significantly improves industrial chain resilience, mainly through technological innovation, digital infrastructure construction, and industrial structure rationalization. The positive effects are more pronounced in eastern coastal regions, areas with high market, and developed digital economies. These findings provide empirical support for big data policy optimization and industrial chain upgrading.

Keywords: Big data policy, data factor agglomeration, difference-in-differences method, industrial chain resilience.

1. Introduction

This paper evaluates the impact of big data policies on industrial chain resilience using multi-period difference-in-differences (DID) method, based on the unbalanced panel data of 266 prefecture-level and above cities in China from 2009 to 2022. It also conducts mechanism analysis from three aspects: technological innovation, digital infrastructure construction, and rationalization of industrial structure. Compared with previous studies, the marginal contributions of this paper are as follows:

First, most existing studies have explored the economic effects of data factor agglomeration, such as new quality productive forces, economic growth, industrial transformation, green development, enterprise innovation, and labor income share [6,7]. However, few have focused on the impact of data factor agglomeration on industrial chain resilience [2,4]. This paper investigates the effect of data factor agglomeration on industrial chain resilience from both theoretical and empirical perspectives, which not only enriches the research scope of factors influencing industrial chain resilience but also expands the research on the economic consequences of big data policies [12].

Second, existing studies have discussed the formation and influencing factors of industrial chain resilience, but they mostly focus on qualitative analysis and pay insufficient attention to the causal relationship between data factor agglomeration and industrial chain resilience [3]. Given that data factor agglomeration has become a key factor in enhancing industrial chain resilience, this paper takes big data policies as the research entry point to answer the question of "how data factor agglomeration promotes the improvement of industrial chain resilience", further clarifies the relationship between the two, and provides new evidence for understanding the laws of industrial chain resilience in the new era [16].

2. Theoretical Analysis

2.1. Theoretical Analysis

Against the backdrop of digital economic transformation, the economic growth model is shifting from extensive to intensive growth. Distinct from traditional production factors such as labor and capital, data factors exert a pivotal impact on high-quality economic development [8,13]. Currently, data factors have emerged as a core determinant of economic and social advancement, as well as a new driving force facilitating manufacturing upgrading and industrial competitiveness improvement.

The application of new-generation data factors continuously elevates the digitalization, informatization and intellectualization of industrial development. It can directly strengthen industrial chain resilience, and further enhance resilience by boosting digital technological innovation, optimizing digital infrastructure and rationalizing industrial structure [4,6].

The construction of comprehensive big data pilot zones serves as a crucial strategy for China to gather data factors and advance the exploitation and utilization of data resources, which plays an essential role in industrial chain renewal and modernization [9,10]. This section first analyzes the direct effect of data factor agglomeration on industrial chain resilience. It then adopts digital technological innovation, digital infrastructure construction and industrial structure rationalization as influencing mechanisms to explore the indirect impact of data factor agglomeration on industrial chain resilience.

Data Factor Agglomeration and Industrial Chain Resilience.

Data factor agglomeration refers to the concentration and integration of data resources within specific regions, whose agglomeration effect profoundly influences technological innovation, regional coordination, environmental governance and economic development [15]. In economics, resilience denotes the capacity of an economy to cope with internal and external disturbances, withstand shocks and achieve sustainable development via structural and growth pattern adjustment. Industrial chain resilience means the capability of all industrial chain links to maintain stable system operation, avoid chain rupture and resist impacts amid internal and external risks and challenges [14].

This paper elaborates the direct effects of data factor agglomeration on industrial chain resilience from three

perspectives: collaborative optimization, improved resource allocation efficiency and enhanced risk early-warning capacity.

On one hand, data factor agglomeration optimizes industrial collaborative operation mechanisms and effectively eliminates information silos in research and development, production and circulation. Real-time sharing of capacity utilization data enables rapid response and coordinated resource allocation amid supply chain disruptions, preventing partial shocks from spreading across the entire industrial chain. On the other hand, it improves industrial resource utilization efficiency. Enterprises establish intelligent decision-making models to dynamically optimize regional inventory management and production scheduling, cutting redundant inventory and delivery cycles, lowering resource idle rate and accelerating delivery efficiency [15]. Such data-driven decision-making reduces resource misallocation risks and stabilizes industrial chain operation. Furthermore, data factor agglomeration improves the industrial chain risk prevention and control system. Real-time monitoring and intelligent prediction systems help enterprises identify potential risks in advance and formulate prompt countermeasures, greatly strengthening the industrial chain's buffering capacity against external shocks [16].

Based on the above theoretical analysis, the research hypothesis is proposed as follows:

H1: Data factor agglomeration contributes to the enhancement of industrial chain resilience.

3. Research Design

3.1. Variable Definition

Variables are presented in Table 1.

Table 1. Definitions and Measurements

Variable	Name	Symbol
Explained	Industrial chain resilience	indusres
Explanatory	Data factor agglomeration	did
Control	Economic development level	eco
	Industrialization level	industry
	Population scale	peop
	Technological development level	tech
	Government intervention degree	gov

3.2. Model Specification

Given that the national big data pilot zones are established in batches, this paper adopts a multi-period Difference-in-Differences model as the benchmark panel regression model, which is specified as follows:

$$indusres_{it}=\beta_0+\beta_1 did_{it}+\gamma Controls_{it}+\mu_{city}+\lambda_{year}+\varepsilon_{it} \quad (1)$$

where $indusres_{i,t}$ denotes the industrial chain resilience of city in year, and $did_{i,t}$ represents data factor agglomeration, which is assigned a value of 1 in and after the year when the national big data pilot zone is established, and 0 otherwise. Controls refers to a set of city-level control variables.City and year denote city fixed effects and year fixed effects, respectively. ε_{it} is the random disturbance term, and robust standard errors are adopted in all regression analyses.ped in manually. Do not use Word's References feature or numbered list. In the reference list, provide up to three authors' names; if more than three authors, use "et al." Place a space between an authors' initials. Papers that have not been published should be cited as "unpublished". Papers that have been submitted or accepted for publication should be cited as "submitted for publication". Please give affiliations and addresses for personal communications.Use sentence case for the words in a paper title.

3.3. Data Sources and Descriptive Statistics

According to the descriptive statistical results reported in Table 2, the overall mean value of industrial chain resilience in China is at a moderately low level, leaving considerable room for further improvement. The standard deviation of industrial chain resilience is relatively small, whereas a distinct disparity exists between its maximum and minimum values [15].This suggests that while the industrial chain resilience of most cities clusters around the average level, a small number of cities have witnessed a significant enhancement in industrial chain resilience benefited from the implementation of the national comprehensive big data pilot zone policy and effective industrial transformation [10].

Table 2. Descriptive Statistical Results

Variable	N	Mean	SD	Min	Max
indusres	3182	0.149	0.052	0.121	0.929
did	3182	0.125	0.331	0	1
eco	3182	10.68	0.626	9.205	12.09
industry	3182	15.73	1.026	13.32	18.14
peop	3182	5.874	0.695	3.807	7.245
tech	3182	0.016	0.015	0.002	0.082
gov	3182	0.193	0.098	0.071	0.626
indusres	3182	0.149	0.052	0.121	0.929

4. Empirical Results and Discussion

4.1. Benchmark Regression Analysis

Table 3 reports the empirical estimation results of the impact of national comprehensive big data pilot zone establishment on industrial chain resilience. Columns (1), (2), and (3) in Table 3 correspond to the regression results without control variables and fixed effects, with fixed effects but without control variables, and with both control variables and fixed effects, respectively. The results show that the coefficients of DID are significantly positive at the 1% statistical level across all specifications, indicating that data factor agglomeration can substantially improve industrial chain resilience. Accordingly, the results from Columns (1) to (3) provide preliminary empirical evidence for Hypothesis H1 that data factor agglomeration is conducive to the enhancement of industrial chain resilience [1].

Table 3. Descriptive Statistical Results

	(1)	(2)	(3)
VARIABLES	indusres	indusres	indusres
did	0.031***	0.016***	0.011***
Control	NO	NO	YES
City FE	NO	YES	YES
Year FE	NO	YES	YES
Constant	0.146***	0.147***	-0.707***
	-151.51	-270.48	(-3.56)
Observations	3,182	3,182	3,182
R^2	0.039	0.711	0.764

4.2. Parallel trend test results

A fundamental prerequisite for applying the Difference-in-Differences (DID) method is that the treatment group and the control group share a consistent trend of outcome variables in the absence of policy intervention. To verify the validity of the DID model setting in this study, a parallel trend test is conducted. Referring to Beck et al. [1], this paper adopts the event study approach and constructs the following dynamic model for hypothesis testing:

$$indusres_{it}=\alpha_0+\sum_{k=-m}^{n}\theta_k Policy_{it}^{k}+\gamma Controls_{it}+\mu_{city}+\lambda_{year}+\varepsilon_{it} \quad (2)$$

Given that the sample period of this study spans from 2009 to 2022, incorporating all year dummy variables into the model may cause multicollinearity. Therefore, this paper selects six pre-policy periods and four post-policy periods for the parallel trend test, with the sixth period before policy implementation set as the base period. Meanwhile, policy time aggregation is performed: all periods with t<-6 are merged into t=-6, and all periods with t>4 are merged into t=4.

Figure 1 illustrates the dynamic policy effects with a 95% confidence interval. The results indicate that the confidence intervals include zero in all periods before the implementation of the big data pilot zone policy, whereas the confidence intervals exclude zero starting from the second year after policy implementation. This finding suggests that the estimated coefficients are statistically insignificant in the pre-policy periods but become significantly positive in the post-policy periods. There is no significant difference in the development trend of industrial chain resilience between pilot and non-pilot cities before policy implementation, while a significant divergence emerges two years after the policy shock. Overall, the parallel trend assumption is satisfied [1,15].

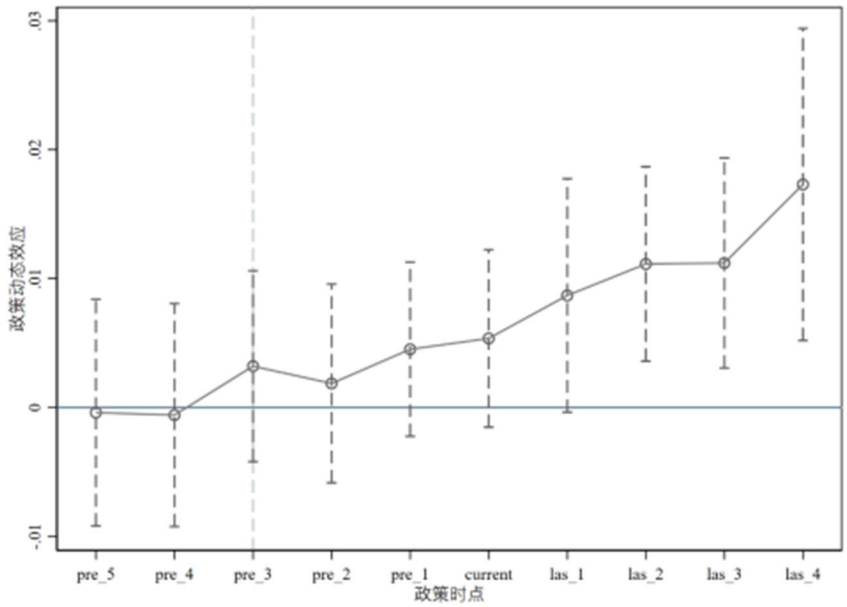

Figure 1. Dynamic effects of big data pilot zone policy

4.3. Placebo Test

To eliminate the interference of urban individual characteristics on policy effect estimation, this study incorporates a series of city-level control variables into the baseline model and controls for both city fixed effects and year fixed effects. Nevertheless, unobserved omitted variables may still lead to biased estimation results. To further exclude the confounding effects of unobservable factors on the DID estimation, this paper adopts the mixed placebo test proposed by Chen. By randomly assigning policy implementation time and fictitious treatment groups simultaneously, false interaction terms are generated to replace the core explanatory variable in the baseline model, and the simulation process is repeated 500 times. The corresponding results are presented in Figure 2 and Figure 3.

Further analysis of statistical significance demonstrates that the kernel density estimation and histograms of the fictitious treatment effects in Figure 2 and Figure 3 peak near zero with a compact and symmetric distribution, showing no obvious deviation from zero. The null hypothesis of zero placebo effect is verified, indicating that the fictitious policy shocks produce no significant impact under fixed structural constraints such as group scale and time scope. The two-tailed p-values are 0.0100 for Figure 2 and 0.0000 for Figure 3, confirming the significance of the genuine treatment effect. This evidence suggests that the causal findings are not dependent on specific group settings or policy time selections and reflect a general and robust causal relationship[1,15].

In summary, the placebo test results reveal that the benchmark estimation results cannot be replicated by random policy interventions. It is confirmed that the core policy effect passes the placebo test and is barely affected by unobservable confounding factors, thereby verifying the high robustness of the causal inference in this study.

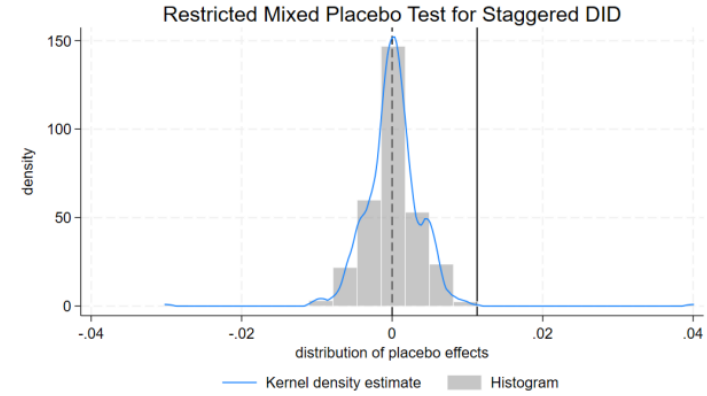

Figure 2. Restricted Mixed Placebo Test for Staggered DID

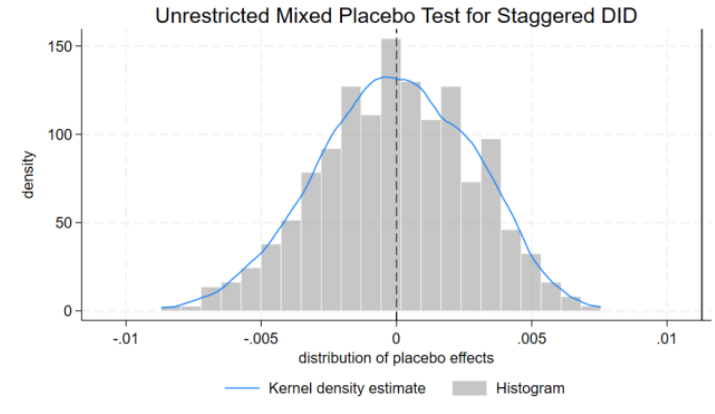

Figure 3. Unrestricted Mixed Placebo Test for Staggered DID

4.4. Instrumental Variable Method

The DID method mitigates endogeneity via comparative analysis between treatment and control groups, which premised on the random selection of big data pilot zones. In reality, regions with advanced big data industry and great application potential are more likely to be approved as pilot sites. The endogeneity caused by reverse causality may undermine the reliability of research conclusions.

This paper adopts the instrumental variable method to conduct robustness checks and alleviate estimation bias. Following Wang Lan (2019), the one-period lagged term of the core explanatory variable is selected as the instrumental variable, and the two-stage least squares (2SLS) model is constructed for estimation.

The regression results are presented in Table 5. The F statistic exceeds 10, ruling out the weak instrumental variable problem. The LM statistic rejects the null hypothesis of under-identification at the 10% significance level, and the Wald statistic passes the critical value test. Both tests verify the validity and exogeneity of the instrumental variable[1].

The first-stage regression shows a significantly positive coefficient of the instrumental variable, implying a positive correlation with industrial chain resilience. The second-stage

result also presents a significantly positive coefficient. The core conclusion still holds after eliminating endogeneity interference.

Table 3. Results of Instrumental Variable Method

	(1)	(2)
	first	second
VARIABLES	did	indusres
did		0.016*
		-1.85
IV	0.821***	
	-86.17	
Control	YES	YES
CityFE	YES	YES
YearFE	YES	YES
Observations	2,384	2,384
R-squared		0.337

4.5. Double Machine Learning

Against the background of the big data era, the double machine learning (DDML) method proposed by Chernozhukov et al. (2018) can effectively overcome the limitations of traditional causal inference approaches[9]. This method employs Neyman orthogonal moment cross-fitting estimation. When applied to the difference-in-differences (DID) framework, it mitigates functional form misspecification bias and addresses the curse of dimensionality caused by numerous control variables in linear regression[7].Even when the functional form of covariates is unknown, DDML yields unbiased estimates of the treatment effect. This study uses the DDML model for robustness checks, with the following specification:

$$indusres_{i,t}=a_0 did_{i,t}+g(X_{i,t})+U_{i,t}, \quad E(U_{i,t}|did_{i,t},X_{i,t})=0 \qquad (3)$$

where Xi,t denotes the set of high-dimensional control variables, and all other variables are consistent with the baseline regression.

The empirical procedure is as follows:

First, an initial random forest model is constructed with a sample splitting ratio of 1:5, and the results are reported in Column (1) of Table 5.Next, the machine learning algorithm is adjusted from random forest to LASSO regression and gradient boosting, with the corresponding results presented in Columns (2) and (3) of Table 5.All results indicate that the DDML approach does not alter the baseline regression conclusion, i.e., the implementation of the big data policy significantly promotes the development of industrial chains[1,15].

Table 4. Robustness Test Based on Double Machine Learning

	(1)	(2)	(3)
VARIABLES	indusres	indusres	indusres
did	0.006***	0.011***	0.007***
	-0.002	-0.003	-0.002
control	YES	YES	YES
CityFE	YES	YES	YES
YearFE	YES	YES	YES
Observations	3,182	3,182	3,182

5. Conclusion

The establishment of big data pilot zones attracts the agglomeration of data factors to the pilot regions, which serves as an important measure to enhance regional industrial chain resilience. Based on the unbalanced panel data of 266 prefecture-level and above cities in China from 2009 to 2022, this paper takes the establishment of national big data comprehensive pilot zones as an exogenous shock, and systematically investigates the impact and mechanism of data factor agglomeration on industrial chain resilience by using methods such as multi-period difference-in-differences (DID) and double/debiased machine learning (DDML). The results show that:The establishment of big data comprehensive pilot zones has significantly strengthened the industrial chain resilience of the pilot areas;Data factor agglomeration enhances industrial chain resilience through three mechanisms: promoting digital technology innovation, improving digital infrastructure construction, and optimizing the rationalization of industrial structure;The policy effects are more prominent in the developed eastern coastal areas, regions with high digital economy development and strong marketization, while no significant effects are found in the central and western regions, areas with weak digital foundation or low marketization, indicating that the release of policy dividends is closely related to regional resource endowments. The conclusions of this study are of great significance for in-depth understanding and evaluation of the policy effects of China's big data pilot zones, and provide important implications for formulating reasonable big data development strategies from the perspective of data factor agglomeration to promote the improvement of industrial chain resilience.

Acknowledgment

We sincerely thank all reviewers for their constructive comments and valuable suggestions, which greatly improved the quality of this paper. We are grateful to the colleagues and researchers who provided assistance and discussions during the research process. This work was supported by the research funding of the authors' institutions.

References

[1] Beck, T., Levine, R., & Levkov, A. (2010). Big bad banks? The winners and losers from bank deregulation in the United States. *The Journal of Finance*, *65*(5), 1637-1667.

[2] Zhu, Y. G., Zhang, W. F., Wang, D., et al. (2023). Evaluation of the resilience of China's copper resource industrial chain and supply chain. *Resources Science*, *45*(9), 1761-1777.

[3] Zhang, B., & Yang, L. (2024). The impact of international industrial transfer on China's industrial chain resilience—From the perspective of industrial chain vertical correlation. *Journal of International Trade*, (8), 1-18. https://doi.org/10.13510/j.cnki.jit.2024.08.001

[4] Chao, X. J., Lian, Y. M., Yuan, R. J., et al. (2024). Digital infrastructure construction and industrial chain resilience—An empirical analysis based on industrial chain recovery capacity data. *The Journal of Quantitative & Technical Economics*, *41*(11), 112-131. https://doi.org/10.13653/j.cnki.jqte.20240812.001

[5] Streimikiene, D., et al. (2023). Review and assessment of import diversification methods and measures in the primary economic sector. *Acta Montanistica Slovaca*, *28*(1), 83-97. https://doi.org/10.46544/AMS.v28i1.08

[6] Zheng, W., & Luo, R. F. (2024). Can data factor agglomeration promote the modernization of industrial chains?—Dual perspectives of digital finance development and digital talent agglomeration. *Industrial Economics Research*, (6), 43-55+69. https://doi.org/10.13269/j.cnki.ier.2024.06.001

[7] Ahrens, A., Hansen, C. B., Schaffer, M. E., & Wiemann, T. (2024). ddml: Double/debiased machine learning in Stata. *The Stata Journal*, *24*(1), 3-45.

[8] Wang, Q., & Fu, X. D. (2021). Research on the mechanism of data factors empowering economic growth. *Shanghai Journal of Economics*, (4), 55-66. https://doi.org/10.3969/j.issn.1005-1309.2021.04.006

[9] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C. B., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1-C68.

[10] Wei, J., & Zhang, X. (2023). The role of big data in promoting green development: Based on the quasi-natural experiment of big data pilot zones. *International Journal of Environmental Research and Public Health*, *20*(5), 4097. https://doi.org/10.3390/ijerph20054097

[11] Qi, P., Sun, D., Xu, C., et al. (2023). Can data elements promote the high-quality development of China's economy? *Sustainability*, *15*(9), 7102. https://doi.org/10.3390/su15097102

[12] Jianing, P., Fangyi, J., & Yimeng, Z. (2022). An analysis of the impact of the digital economy on high-quality economic development in China—A study based on the effects of supply and demand. *Sustainability*, *14*(24), 16991. https://doi.org/10.3390/su142416991

[13] Shi, D. (2022). Evolution of industrial development trends under the background of digital economy. *China Industrial Economics*, (11), 26-42. https://doi.org/10.3969/j.issn.1006-480X.2022.11.006

[14] Duan, H. (2020). Stress test and countermeasures of COVID-19 on China's industrial chain resilience. *China Industry & Information Technology*, (3), 94-96. https://doi.org/10.19609/j.cnki.cn10-1299/f.2020.03.015

[15] Liu, C. M., Chen, L., & Wei, X. M. (2023). Research on the impact of data factor agglomeration on technological innovation—A quasi-natural experiment based on national big data comprehensive pilot zones. *Journal of Shanghai University of Finance and Economics*, *25*(5), 107-121. https://doi.org/10.16538/j.cnki.jsufe.2023.05.008

[16] Xie, H. Q., Zhang, F., & Wu, X. D. (2024). Can virtual agglomeration improve industrial chain resilience?—Empirical test based on urban panel data in China. *Modern Finance and Economics (Journal of Tianjin University of Finance and Economics)*, *44*(8), 3-17. https://doi.org/10.19559/j.cnki.12-1387.2024.08.001