The Impact of Public Data on Firm Intelligent Development

Wenhui Yan[1,a,*], Wanru Wang[2,b]

[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]National Food and Strategic Reserves Administration, Beijing, China

*[a]18734864643@163.com, [b]569764894@qq.com

Abstract. Utilizing panel data from listed companies between 2009 and 2022 and employing a difference-in-differences model, we examine the impact of public data openness on firm intelligent development. Our findings reveal that the policy of public data significantly promotes firm intelligent development. Public data openness has a more significant effect on the intelligent development of non-state-owned and small- to medium-sized firms. The conclusions of this study provide policy insights into harnessing the benefits of public data for intelligent development, thereby promoting high-level growth for Chinese firms.

Keywords: Public data; firm intelligent development; Quasi-natural experiment

1 Introduction

With the advancement of information technology, big data has become a significant driving force for social progress and economic growth. The Chinese government places great emphasis on the openness and sharing of data resources and actively promotes the opening of public data (PDO). PDO refers to the practice where government departments make the data they collect and generate available to the public for free use by businesses and individuals (Abhishek, 2020)[1]. This approach provides firms with a wealth of data resources, fostering favorable conditions for intelligent development. An increasing number of scholars have also turned their attention to the economic effects of PDO. They argue that PDO facilitates the integration of data as a new production factor with traditional production, thereby reducing transaction costs (Avi and Catherine, 2019)[2] and accelerating digital transformation (Chen and Zhang, 2024)[3]. However, existing research has overlooked the impact of PDO on firm intelligent development. This study fills this gap by investigating the effect of PDO on the intelligent development of firms.

PDO provides essential support for enterprise intelligent development. On one hand, companies can leverage publicly available data to effectively enhance their innovation capabilities (Long et al., 2021)[4] and optimize resource allocation, significantly strengthening market competitiveness and intelligent capabilities. On the other hand, the openness and sharing of public data facilitate deep collaboration between gov-

S. M. Chung et al. (eds.), *Proceedings of 2024 4th International Conference on Public Management and Big Data Analysis (PMBDA 2024)*, Advances in Economics, Business and Management Research 318,
https://doi.org/10.2991/978-94-6463-656-7_16

ernment and firms, fostering digital transformation (Chen and Zhang, 2024)[3], thereby further advancing enterprise intelligent development. This study aims to explore the impact of PDO on firm intelligent development, providing theoretical insights for policymakers and enterprise development.

Compared to existing research, the innovations in our study include the following aspects. (1) Previous studies have primarily examined factors influencing firm intelligent development from the perspectives of technological innovation (Gillenwater and Comlon, 1995)[5], informatization strategy development (Wang and He, 2022)[6], yet have overlooked the impact of PDO on firm intelligent development. Our research expands this field by investigating the influence of PDO on firm intelligent development in China, thus broadening the research perspective on firm intelligence. (2) Existing literature has found that PDO benefits business innovation (Thorhildur et al., 2014)[7], reduces firm information acquisition costs (Hughes-Cromwick and Coronado, 2019)[8], and supports the high-quality development of manufacturing firms (Wu, 2024)[9]. This study contributes to the literature by examining whether PDO influences firms' intelligent development from the perspective of public data itself, thereby expanding research on the economic outcomes of PDO policies to some extent.

2 Institutional Background and Theoretical Analysis

2.1 Institutional Background

China is currently in the digital economy era, with data emerging as a critical production factor that plays a key role in economic and social development. Building a unified national data market requires breaking down data silos, promoting data flow and sharing, and unlocking the value of data as a resource. China's practice of PDO began relatively late, with the first pilot public data open platforms launched in 2012 in three cities, including Shanghai and Beijing. Since then, PDO policies have been gradually implemented across various regions each year. By the end of 2023, 226 provincial and municipal governments in China had established public data open platforms.

2.2 Theoretical Analysis

The diversity, reliability, and usability of government-opened data enable firms to use this data more confidently and conveniently, lowering the barriers to accessing and utilizing public data and allowing more businesses to leverage it for intelligent transformation. The timeliness and frequency of data updates ensure that public data is refreshed regularly, allowing firms to access the latest market and societal trends (Balci, 2021)[11], which is crucial for real-time decision-making and intelligent systems. From a managerial perspective, the comprehensive and timely market and societal information provided by PDO enables firm leaders to make data-driven decisions and devise effective, scientifically grounded strategies for intelligent transformation. Lastly, by analyzing public data, companies can identify bottlenecks and deficiencies in their business processes, allowing for process optimization and increased operational efficiency, which are key aspects of intelligent development. Based on the above

analysis, this study proposes the following hypothesis: public data facilitates firm intelligent development.

3 Research Design

3.1 Data Sources

This study focuses on A-share listed companies on the Shanghai and Shenzhen stock exchanges in China, covering the period from 2009 to 2022. Data on firm intelligent development is obtained from annual reports of listed companies, while information on PDO is sourced from the China Local Public Data Open Report (Cities). Control variables for the companies come from the China Stock Market & Accounting Research Database (CSMAR) database. To ensure data quality, we exclude the following sample: (1) financial sectors; (2) ST, *ST, and PT. We perform winsorization on all continuous variables to mitigate the influence of outliers.

3.2 Data Variables

Dependent variable: firm intelligent development (AI). We extract keywords related to "artificial intelligence" from the annual reports. Following the methodology of Tomas et al. (2013)[12], we measure the level of firm intelligent development using the logarithm of the keyword count plus one.

Independent variable: public data open ($Open$). $Open_{it}$ indicates a dummy variable representing whether a public data open platform was launched in city i in year t. If a public data platform was launched in city i in year t, $Open_{it}$ is assigned a value of 1; otherwise, it is 0.

Control variables: firm size, debt-to-asset ratio, operating profit margin, return on assets, asset turnover ratio, revenue growth rate, fixed asset ratio, proportion of independent directors on the board, administrative expense ratio, audit opinion, shareholding ratio of the top five shareholders, whether the chairman and CEO are the same person, and chairman's shareholding ratio.

3.3 Research Model

To examine the impact of PDO on firm intelligent development, this study constructs the following regression model, referencing Elsawy and Shehata (2023)[13]:

$$AI_{i,t} = \alpha + \beta Open_{i,t} + \gamma Control\_Var_{i,t} + FirmFE + YearFE + \varepsilon_{i,t} \quad (1)$$

Where $AI_{i,t}$ indicates the level of intelligent development of firm i in year t; $Open_{i,t}$ represents a dummy variable for public data. $Control\_Var_{i,t}$ represents the set of control variables influencing corporate intelligent development; $FirmFE$ denotes firm fixed effects, $YearFE$ denotes year fixed effects, and $\varepsilon_{i,t}$ represents the random disturbance term.

4 Empirical Results

4.1 Baseline Estimation

Table 1 reports the regression results on the impact of PDO on corporate intelligent development. These regression results indicate that the coefficient of *Open* remains significantly positive across all specifications, suggesting that PDO significantly promotes corporate intelligent development. Additionally, based on the results in Column (4), the coefficient of *Open* is 0.105, which implies that, all else being equal, PDO enhances corporate intelligent development by 10.5%.

Table 1. Regression Results on the Impact of Public Data on Corporate Intelligent Development

Variables	(1)	(2)	(3)	(4)
Open	0.788***	0.126***	0.608***	0.105***
	(30.542)	(6.376)	(24.574)	(5.492)
Control	No	No	Yes	Yes
Firm and Year FEs	No	Yes	No	Yes
Observations	31903	31903	31903	31903
R-squared	0.106	0.768	0.207	0.776

Notes: Robust t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

4.2 Parallel Trend Test

Following the study by Li et al. (2024)[14], we use an event study approach to examine whether the treatment and control groups exhibit the same trend prior to the policy implementation. The results in Figure 1 show that the pre-policy coefficients are all insignificant, indicating no significant differences in corporate AI development between the treatment and control groups before PDO.

4.3 Robustness Test

To ensure the robustness of the regression results, we conduct a placebo test by randomly constructing pseudo-policy variables (Figure 2). The p-values of most estimated coefficients are greater than 0.1, and the coefficients are largely clustered around zero, demonstrating a significant difference from the baseline regression coefficient (0.105). These results indicate that, after conducting the placebo test, the conclusion that PDO positively impacts corporate intelligent development still holds, confirming the robustness of the core finding.

To mitigate self-selection bias in the regression, we follow the approach of Li et al. (2024)[14] and use the propensity score matching (PSM) method for robustness testing (Column (1) of Table 2). The regression coefficient for *Open* remains significantly positive at the 1% level, indicating that PDO has a significant impact on corporate intelligent development.

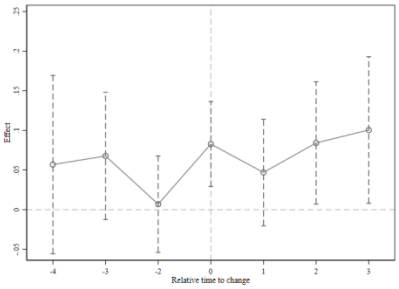

Fig. 1. Parallel Trend Test

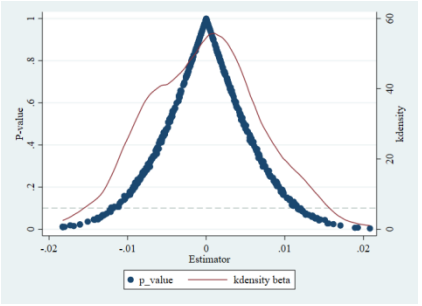

Fig. 2. Placebo Test

To further mitigate the impact of heterogeneous treatment effects, we reference the method of De Chaisemartin and D'Haultfoeuille (2020)[10]. The counts of negative and positive weights are 3,239 and 11,053, respectively, with sums of -0.2196 and 1.2196 for negative and positive weights, respectively. This indicates that the proportion of negative weights is much smaller than that of positive weights, suggesting that the influence of heterogeneous treatment effects is negligible.

Following the approach of Pang et al. (2024)[15], this study measures corporate AI development by the frequency of precise terms reflecting intelligent development appearing in corporate annual reports (Column (2) of Table 2). The coefficient for *Open* is significantly positive at the 1% level, consistent with the baseline regression results, confirming the robustness of this study's findings.

The presence of unmeasurable factors associated with industry and city variations may lead to biased estimates. To address this, we incorporate industry and city fixed effects into the baseline model (Column (3) of Table 2). The results show our conclusion remains valid.

Table 2. Robustness Tests

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Open	0.112***	0.098***	0.094***	0.090***	0.096***	0.092***	0.078***
	(4.583)	(5.223)	(4.913)	(4.083)	(4.216)	(4.071)	(3.481)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18667	31903	31806	23238	23238	23238	23238
R-squared	0.766	0.772	0.783	0.773	0.773	0.774	0.774

During the sample period, the Chinese government implemented a series of policy initiatives, such as the National Big Data Comprehensive Reform Pilot and Information Benefiting the People Pilot, which may affect our regression results. Therefore, we control for the impact of these policies on the study results (Columns (4)-(7) of Table 2). After excluding other policy interventions, *Open* remain significantly positive, indicating that our conclusions are robust.

4.4 Heterogeneity Test

Different operational mechanisms may lead to variations in the impact of PDO on corporate intelligent development (Columns (1) and (2) of Table 3). These results indicate that the *Open* coefficient is significantly positive regardless of firm ownership, with a relatively larger impact on AI observed in non-state-owned firms. Non-state-owned firms, being more market-sensitive, are more likely to adopt and apply new technologies, thus facilitating AI transformation.

Firms of different sizes exhibit distinct characteristics in the process of AI development (Columns (3) and (4) of Table 3). The results indicate that the coefficient for *Open* is significantly positive for both large and SME groups, but the coefficient for *Open* is slightly higher for large firms compared to SMEf. This is likely because public data platforms lower the barriers for SMEf to access technology and resources, providing new opportunities that drive their AI development.

Table 3. Heterogeneity Test

Variables	(1) *Soe=1*	(2) *Soe=0*	(3) *Large*	(4) *SMEf*
Open	0.086***	0.107***	0.083***	0.099***
	(2.835)	(4.341)	(3.239)	(3.560)
Control	Yes	Yes	Yes	Yes
Firm and Year FEs	Yes	Yes	Yes	Yes
Observations	11302	20336	15815	15798
R-squared	0.745	0.790	0.796	0.811
Empirical p-value	0.226***		0.048**	

5 Conclusions and Policy Recommendations

The following conclusions are drawn: First, PDO significantly promotes corporate intelligent development. Second, this effect remains robust after a series of tests, including parallel trend and placebo tests. Third, the impact of PDO policy exhibits significant heterogeneity, with a more pronounced effect on the intelligent development of non-state-owned and small- to medium-sized firms.

The policy recommendations are as follows: First, gradually expand the scope of public data openness by incorporating data from more government departments across various fields and promote intelligent development. Second, the government should provide differentiated support based on characteristics, helping various types of firms better utilize public data to drive intelligent development.

References

1. Nagaraj, A. (2021) The private impact of public data: Landsat satellite maps increased gold discoveries and encouraged entry. Management Science, 68: 564-582. https://doi.org/10.1287/mnsc.2020.3878.

2. Goldfarb, A. and Tucker, C., (2019) Digital Economics. Journal of Economic Literature, 57(1): 3-43. https://doi.org/10.1257/jel.20171452.
3. Chen, K. and Zhang, S. (2024) How does open public data impact enterprise digital transformation? Economic Analysis and Policy, 83: 178-190. https://doi.org/10.1016/j.eap.2024.06.007.
4. Lam, L., Nguyen. P., Le, N., Tran, K. (2021) The relation among organizational culture, knowledge management, and innovation capability: its implication for open innovation. Journal of Open Innovation: Technology, Market, and Complexity, 7(1): 66. https://doi.org/10.3390/joitmc7010066.
5. Gillenwater, E. L., Conlon, S., Hwang, C. (1995) Distributed manufacturing support systems: The integration of distributed group support systems with manufacturing support systems. Omega, 23(6): 653-665. https://doi.org/10.1016/0305-0483(95)00037-2.
6. Wang, X. and He, M. (2022) The impact of information strategy cognition and construction behavior on the intelligent transformation of manufacturing firms. Science & Technology Progress and Policy, 39(3): 87-95. https://doi.org/10.1016/j.eng.2020.07.017.
7. Jetzek, T., Avital, M., Bjorn, N. (2014) Data-driven innovation through open government data. Journal of Theoretical and Applied Electronic Commerce Research, 9(2): 15-16. https://doi.org/10.4067/S0718-18762014000200008.
8. Hughes-Cromwick, E. and Coronado, J. (2019) The value of US government data to US business decisions. Journal of Economic Perspectives, 33 (1): 131–46. https://doi.org/10.1257/jep.33.1.131.
9. Wu, T. (2024) How can open public data promote efficient and equitable green production? Evidence from eco-efficiency in China. Environmental Science and Pollution Research, 31: 35173-35193. https://doi.org/10.1007/s11356-024-33464-x.
10. De Chaisemartin, C. and D'Haultfœuille, X. (2020) Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9): 2964–96. https://doi.org/10.1257/aer.20181169.
11. Balci, G. (2021) Digitalization in container shipping: Do perception and satisfaction regarding digital products in a non-technology industry affect overall customer loyalty?. Technological Forecasting and Social Change, 172: 12-16. https://doi.org/10.1016/j.techfore.2021.121016.
12. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. Dean, J. (2013) Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems, 3111-3119. https://doi.org/10.48550/arXiv.2108.12419.
13. Elsawy, E., Shehata, A. (2023) Open government data initiatives in the Maghreb countries: An empirical analysis. IFLA journal, 49(1): 61-73. https://doi.org/10.1177/03400352221104847.
14. Li, X., Liu, Z., Ye, Y. (2024) Public data and corporate employment: Evidence from the launch of Chinese public data platform. Economic Analysis and Policy, 84: 124-144. https://doi.org/10.1016/j.eap.2024.08.023.
15. Pang, S., Li, Z., Wang, Y. (2024) Digital technology and domestic value-added ratio in export: Evidence from China's pilot zones for integrating informatization and industrialization. Economic Analysis and Policy, 84424-439. https://doi.org/10.1016/j.eap.2024.08.030.