

Big Data Policies and Corporate Earnings Management: A Quasi-Natural Experiment Based on National Big Data Comprehensive Pilot Zones

Feiyan Gao

*School of International Business, Henan University of Economics and Law, Zhengzhou, China
3153271661@qq.com*

Abstract. Taking the establishment of national big data comprehensive pilot zones as an exogenous policy shock, this paper examines whether big data policies can constrain corporate earnings management. Based on panel data of China's A-share listed companies from 2000 to 2024, the paper constructs a difference-in-differences model. The estimation results show that after the establishment of pilot zones, the degree of corporate earnings management decreases by about 7.5% on average, and the inhibitory effect of the policy is significant. Group tests further show that non-high-tech enterprises, small-scale enterprises and enterprises with weaker external supervision have more prominent policy responses. This indicates that the improvement of information environment and enhancement of regulatory capacity can form complementary constraints under the condition of relatively weak governance foundations. Relevant findings provide empirical evidence for the institutional optimization of national big data comprehensive pilot zones and the practice of digital governance.

Keywords: Earnings Management, National Big Data Comprehensive Pilot Zone, Difference-in-Differences Method

1. Introduction

Against the background of the digital economy, data resources have become an important element connecting production, transactions and governance. In August 2015, the State Council issued the *Outline for Promoting Big Data Development* (Guo Fa [2015] No. 50), and subsequently launched the construction of national big data comprehensive pilot zones. In 2016, eight regions including Guizhou and Beijing-Tianjin-Hebei were successively incorporated into the pilot zone layout, focusing on exploring the realization paths of data opening and sharing, resource integration and industrial innovation. The pilot zones provide a clear policy shock and make it possible to identify the firm-level consequences of big data development.

Regarding the digital economy and big data policies, existing literature mostly discusses regional efficiency and enterprise innovation, while corporate governance behavior has not yet become the main observation object. Research on earnings management has formed two types of explanations: internal governance and external institutional environment. However, whether government-led big

data policies constitute a new external constraint still has limited relevant evidence. Yan et al. [1] found governance effects from the perspective of digital finance, but digital finance is mainly driven by market forces, which is different from the government-led characteristics of pilot zone policies. Therefore, the direction and mechanism of big data policies affecting earnings management still need to be empirically tested.

To answer this question, this paper regards the establishment of national big data comprehensive pilot zones as a quasi-natural experiment, and uses a difference-in-differences model to identify the impact of policies on corporate earnings management. The results show that the pilot zone policy significantly reduces earnings management; grouping results show that this effect is more prominent in non-high-tech enterprises, small-scale enterprises and enterprises with weaker external supervision, presenting a complementary effect on weak links in governance. Compared with existing literature, the incremental contributions of this paper are mainly reflected in three aspects: First, it extends the analysis of the consequences of big data policies to earnings management, a corporate governance behavior. Second, starting from government-led digital governance policies, it supplements research on the external influencing factors of earnings management. Third, it provides micro-level policy evidence for the improvement of the pilot zone system and the advancement of digital governance.

The rest of the paper is arranged as follows: the second part discusses relevant literature and puts forward research hypotheses; the third part explains the sample, variables and econometric model; the fourth part reports baseline regression, robustness checks and heterogeneity results; the fifth part summarizes research findings and puts forward policy suggestions.

2. Literature review and research hypothesis

2.1. Literature review

2.1.1. Economic consequences of big data

From the perspective of regional-level evidence, the digital economy can improve total factor productivity [2, 3]. Data accumulation and the reduction of information frictions contribute to more efficient allocation of resources among different uses [4], and the organizational forms and operation modes of enterprises will also adjust with the application of digital technologies [5]. These findings reveal the macro efficiency effect of the digital economy, but pay relatively limited attention to internal governance decisions of enterprises.

Research on pilot zones further shifts the perspective to the firm level. Yang and Zhu [6] point out that pilot zones can improve enterprise innovation efficiency by promoting digital transformation and alleviating financing constraints; Lin [7] finds that digital policies have an enabling effect on enterprises' new quality productivity; Xin et al. [8] find that pilot zone policies promote the diversified configuration of enterprise supply chains. Existing micro-level studies have thus accumulated evidence on innovation and supply chains, but there is still a lack of direct tests on whether and how big data policies affect corporate earnings management.

2.1.2. Influencing factors of earnings management

The motives for earnings management can be summarized as responding to capital market expectations, meeting financing needs and serving the private interests of management [9, 10]. In terms of internal factors, Bo and Wu [11] find that state-owned enterprises have more positive

earnings management, and institutional investors can exert governance effects; Fang and Jin [12] confirm that high-quality internal control can inhibit earnings management; Su and Lin [13] point out that there is a complex relationship between equity incentives and earnings management. In terms of external factors, Chen and Chen [14] find that policy uncertainty will exacerbate earnings management; Huang et al. [15] prove that economic policy uncertainty increases the degree of real earnings management by deteriorating the information environment. In terms of digital governance, Jiang and Zhang [16] examine the dual impact of digital transformation on real earnings management; Du et al. [17] discuss the relationship between digital finance and accounting information transparency; Yan et al. [1] find from the perspective of digital finance that it can inhibit corporate earnings management.

Overall, research on external governance mainly focuses on policy uncertainty and traditional financial development, and the role of big data policies as a new governance mechanism has not been fully discussed. Yan et al. [1] analyze the governance effect of earnings management from the perspective of digital finance, but digital finance is driven by market forces, and its mechanism and causal identification environment are different from those of government-led big data pilot zones. This paper uses the establishment of pilot zones as a quasi-natural experiment to supplement empirical evidence on the relationship between big data governance and corporate earnings management from a policy-driven perspective.

2.2. Research hypothesis

Information asymmetry provides conditions for management to adjust accounting information [10]. When management has more internal information and external supervision is insufficient, there are both motives and operational space for manipulating accruals. The establishment of big data pilot zones may narrow this space from two dimensions: information environment and regulatory capacity.

At the information environment level, pilot zones promote the sharing, opening and integrated use of data resources, and improve the efficiency of information production, transmission and processing. With more transparent corporate financial and operational information, investors and regulators can more easily identify abnormal activities, and the reputation costs and regulatory risks faced by management in earnings manipulation increase accordingly. At the regulatory capacity level, the construction of pilot zones requires the improvement of data security standards, sharing rules and management systems. These institutional arrangements help enhance local governments' ability to obtain information and implement supervision, thereby reducing the room for enterprises to manipulate earnings.

H1: The establishment of national big data comprehensive pilot zones significantly inhibits corporate earnings management behavior.

Policy effects may not be the same across all enterprises. Non-high-tech enterprises usually face more serious information asymmetry, and the marginal benefits brought by the improvement of information environment may be greater; small-scale enterprises have relatively limited internal governance resources and are more likely to obtain external constraints from the construction of pilot zones; enterprises with weaker external supervision have insufficient original supervision mechanisms, and the complementarity of pilot zone policies may be more obvious. Therefore, this paper expects that the inhibitory effect of big data pilot zones on earnings management is more prominent in these weak links of governance.

3. Research design

3.1. Model specification

The key to identifying the causal effect of big data comprehensive pilot zone policies is to eliminate the confusion between pilot selection and corporate governance outcomes. Pilot cities are not randomly selected; their establishment may be related to economic development, industrial structure and institutional environment, and these factors also affect corporate earnings management. For this reason, this paper adopts the difference-in-differences method to compare the changes of the treatment group and the control group before and after the policy, so as to extract the net effect of the policy shock and establish an econometric model accordingly.

$$AbsDA_{it} = \beta_0 + \beta_1 did_{it} + \beta_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (1)$$

In the econometric expression, i denotes the firm index, t denotes the year index, and $AbsDA_{it}$ is the degree of corporate earnings management. The core explanatory variable did_{it} is the product of the treatment group dummy variable and the policy implementation time dummy variable, which is defined as follows:

$$did_{it} = Treat_i \times Post_t \quad (2)$$

$Treat_i$ is used to distinguish whether an enterprise is located in a pilot city: 1 if it belongs to a pilot city, 0 otherwise; $Post_t$ is used to identify the policy implementation stage: 1 after the policy takes effect, 0 in other years. Therefore, the interaction term did_{it} equals 1 only when an enterprise simultaneously satisfies "located in a pilot city" and "the policy has been implemented", and 0 in all other cases.

In the model, X_{it} is a set of firm-level control variables; μ_i absorbs firm-specific time-invariant heterogeneity; λ_t controls for common macro shocks in each year; and ε_{it} is the random disturbance term. This paper takes β_1 as the parameter of interest, which measures the average treatment effect of the big data pilot zone policy on earnings management. To make statistical inferences robust to within-firm error correlation, standard errors are clustered at the firm level.

3.2. Variable definitions

3.2.1. Dependent variable

This paper takes the degree of earnings management ($AbsDA$) as the dependent variable. Based on the modified Jones model proposed by Dechow et al. [18], discretionary accruals are first estimated, and then their absolute values are taken. Two alternative measures are also used in robustness analysis: one is the modified Jones model with performance factors ($AbsDA1$), and the other is the Lu Jianqiao model ($AbsDA2$).

3.2.2. Explanatory variable

The policy dummy variable did is the core explanatory variable, formed by the interaction of $Treat$ and $Post$, that is, $did_{it} = Treat_i \times Post_t$. In accordance with the pilot arrangement specified in the State Council's *Outline for Promoting Big Data Development* (Guo Fa [2015] No. 50), did takes the value of 1 only when the city where the enterprise is located has entered the pilot and the year is

after the policy implementation, and 0 for other observations. Its estimated coefficient β_1 depicts the average treatment effect of the pilot zone policy on corporate earnings management.

3.2.3. Control variables

To control for other factors that may affect corporate earnings management, referring to the studies of Wei [9] and Su et al. [13], this paper introduces the following control variables: firm size (*Size*), leverage (*Lev*), return on assets (*ROA*), board size (*Board*), proportion of independent directors (*Indep*), listing age (*ListAge*), nature of property rights (*SOE*), management shareholding ratio (*Mshare*), high-tech enterprise (*HighTech*).

The descriptive analysis of main variables is shown in Table 1:

Table 1. Descriptive statistics

VARIABLES	Measurement	mean	sd	min	p50	max	N
did	Policy dummy variable	0.308	0.462	0.000	0.000	1.000	43,600
AbsDA	Degree of earnings management, calculated with reference to Dechow et al. (1995)	0.053	0.050	0.001	0.038	0.246	43,600
SOE	Whether it is a state-owned enterprise	0.340	0.474	0.000	0.000	1.000	43,600
Size	Natural logarithm of total assets of the enterprise	22.233	1.277	19.959	22.041	26.063	43,600
Lev	Asset-liability ratio	0.423	0.203	0.059	0.416	0.882	43,600
ROA	Return on total assets	0.033	0.061	-0.217	0.034	0.182	43,600
Board	Natural logarithm of the number of board members	2.113	0.197	1.609	2.197	2.639	43,600
Indep	Proportion of independent directors	0.377	0.053	0.333	0.364	0.571	43,600
ListAge	Listing age	2.167	0.788	0.693	2.303	3.401	43,600
Mshare	Management shareholding ratio	0.136	0.193	0.000	0.010	0.675	43,600
HighTech	Whether it is a high-tech industry	0.615	0.486	0.000	1.000	1.000	43,600

4. Empirical results and analysis

4.1. Regression results analysis

Table 2 presents the baseline estimation of the big data comprehensive pilot zone policy. The coefficient of the core variable *did* is -0.004, which is significantly negative at the 5% level, indicating that corporate earnings management is inhibited after the establishment of pilot zones.

The absolute change of this estimated value is 0.004; compared with the sample mean of AbsDA (0.053), the average reduction corresponding to the policy is about 7.5%. Therefore, the result has not only statistical significance but also certain economic significance.

The signs of control variables also provide supplementary information: enterprises with higher leverage (*Lev*) are more likely to have a higher degree of earnings management; firm size (*Size*), return on assets (*ROA*), listing age (*ListAge*) and state-owned property rights (*SOE*) are associated with a lower degree of earnings management.

Table 2. Model regression results

VARIABLES	AbsDA
did	-0.004** (-2.40)
Size	-0.002** (-2.22)
Lev	0.035*** (10.07)
Board	0.001 (0.19)
ListAge	-0.003** (-2.57)
Indep	0.001 (0.14)
SOE	-0.006*** (-2.90)
ROA	-0.122*** (-12.32)
Mshare	-0.006 (-1.33)
HighTech	-0.002 (-1.06)
Constant	0.093*** (4.74)
Observations	43,278
Adjusted R-squared	0.170
FIRM	YES
YEAR	YES

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

4.2. Robustness checks

To test whether the baseline conclusion depends on specific variable or sample settings, this paper carries out robustness analysis from three directions: replacing the dependent variable, adjusting the

control variable set, and shortening the sample time window. The corresponding results are summarized in Table 3.

Table 3. Robustness checks

VARIABLES	AbsDA1	AbsDA2	ControlChange	NewWindow
did	-0.003** (-2.37)	-0.003** (-2.12)	-0.003** (-2.35)	-0.007*** (-2.95)
Size	-0.002*** (-2.81)	-0.002** (-2.55)	-0.002** (-2.49)	0.003* (1.92)
Lev	0.037*** (12.29)	0.035*** (9.86)	0.034*** (9.83)	0.034*** (6.46)
Board	-0.001 (-0.27)	-0.000 (-0.10)		0.004 (0.74)
ListAge	-0.002* (-1.72)	-0.002* (-1.78)		-0.005** (-2.31)
Indep	0.003 (0.33)	-0.001 (-0.15)	0.001 (0.08)	0.005 (0.33)
SOE	-0.002 (-1.48)	-0.006*** (-2.93)	-0.006*** (-2.90)	-0.006** (-2.10)
ROA	0.054*** (8.74)	-0.147*** (-13.72)	-0.121*** (-12.35)	-0.204*** (-15.81)
Mshare	-0.005 (-1.23)	-0.005 (-1.09)		-0.004 (-0.64)
HighTech	-0.001 (-0.39)	-0.002 (-0.90)		-0.008* (-1.65)
Constant	0.082*** (4.72)	0.101*** (5.02)	0.090*** (4.98)	-0.003 (-0.10)
Observations	43,277	43,139	43,278	23,611
Adjusted R-squared	0.152	0.173	0.170	0.226

Table 3. (continued)

FIRM	YES	YES	YES	YES
YEAR	YES	YES	YES	YES

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

4.2.1. Replacement of dependent variable

The baseline model uses the absolute value of discretionary accruals (AbsDA) obtained from the modified Jones model to measure earnings management. Considering that estimation methods may affect indicators, this paper further adopts two alternative calibers: adding ROA to the modified Jones model to control for performance factors, and the absolute value of the obtained residuals is recorded as AbsDA1; on this basis, the Lu Jianqiao model which incorporates intangible assets and other long-term assets is adopted, and the absolute value of the obtained residuals is recorded as AbsDA2. Under the two alternative indicators, the *did* coefficients are both significantly negative at the 5% level, and the baseline conclusion remains unchanged.

4.2.2. Adjustment of control variables

To eliminate the interference caused by differences in control variable sets, this paper adjusts the control variables and re-estimates the model. In the new regression, the *did* coefficient is still significantly negative at the 5% level, and the direction is consistent with the baseline result. This indicates that the estimated effect of the pilot zone policy on earnings management does not depend on a specific set of control variables.

4.2.3. Shortening of time window

The baseline sample covers 2000–2024. To mitigate the interference caused by data structure changes and macro-environmental differences over a long observation period, this paper changes the window to 2015–2022 for re-estimation. This interval covers before and after policy implementation, while avoiding sample heterogeneity that may exist in earlier and later years. The results show that the *did* coefficient is -0.007, which is significantly negative at the 1% level, and the conclusion remains robust to sample period adjustment.

4.3. Heterogeneity analysis

The governance effect of pilot zones may vary with firm characteristics. To observe this difference, this paper conducts grouping and interaction term tests from three dimensions: industry attribute, firm size and external supervision, and the results are listed in Table 4.

Table 4. Heterogeneity analysis

VARIABLES	HighTech	Large	regulated
1.did	-0.008*** (-3.82)	-0.001 (-0.39)	-0.002 (-1.13)
1.HighTech	-0.004*		

Table 4. (continued)

	(-1.73)		
1.did#1.HighTech	0.007***		
	(3.25)		
1.Large		-0.000	
		(-0.26)	
1.did#1.Large		-0.004**	
		(-2.10)	
1.regulated			0.006**
			(2.42)
1.did#1.regulated			-0.005**
			(-2.21)
Size	-0.002**		-0.002**
	(-2.26)		(-2.39)
Lev	0.035***	0.033***	0.035***
	(10.11)	(9.81)	(10.04)
Board	0.000	-0.000	0.001
	(0.13)	(-0.04)	(0.23)
ListAge	-0.004***	-0.004***	-0.004***
	(-2.95)	(-2.91)	(-2.76)
Indep	0.001	0.001	0.002
	(0.10)	(0.07)	(0.19)
SOE	-0.006***	-0.006***	-0.006***
	(-2.85)	(-2.85)	(-2.97)
ROA	-0.122***	-0.124***	-0.122***
	(-12.32)	(-12.59)	(-12.36)
Mshare	-0.005	-0.006	-0.005
	(-1.20)	(-1.27)	(-1.27)
Constant	0.096***	0.056***	0.095***
	(4.92)	(5.76)	(4.83)
Observations	43,278	43,278	43,278
Adjusted R-squared	0.170	0.170	0.170
FIRM	YES	YES	YES
YEAR	YES	YES	YES

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

4.3.1. Industry attribute

High-tech enterprises usually have more mature digital foundations and information management capabilities, so the new governance role of pilot zone policies may be limited. This paper adds the high-tech enterprise dummy variable (HighTech) and its interaction term with *did*. The results show

that the *did* coefficient corresponding to non-high-tech enterprises is -0.008, which is significantly negative at the 1% level; the coefficient of the interaction term *did*×*HighTech* is 0.007, which is also significantly positive at the 1% level. This means that compared with non-high-tech enterprises, the inhibitory effect of the policy on earnings management of high-tech enterprises is weakened, and pilot zones are more like complementary governance for enterprises with weaker digital foundations.

4.3.2. Firm size

The grouping results of firm size show that the interaction term of *did* and *Large* is significantly negative, indicating that compared with small-scale enterprises, the policy has a stronger inhibitory effect on earnings management of large-scale enterprises. Large-scale enterprises have wider disclosure scope and more complex operation chains, and the strengthened data supervision and external supervision after policy implementation may more easily act on such enterprises. Therefore, the construction of pilot zones further compresses the room for earnings manipulation of large-scale enterprises by improving information transparency and regulatory efficiency.

4.3.3. External supervision

The heterogeneity test of external supervision shows that the *did* coefficient is negative, and the *did*×*regulated* interaction term is also significantly negative, indicating that regulated enterprises are subject to stronger inhibition than non-regulated enterprises. Regulated enterprises are originally in a stricter supervision environment, and pilot zone policies can form synergy with existing systems, enabling regulatory authorities to obtain and process enterprise information more efficiently, thereby strengthening the constraint on earnings management behavior.

5. Conclusions and policy implications

Taking the establishment of national big data comprehensive pilot zones as a quasi-natural experiment, this paper uses the difference-in-differences method to test its impact on corporate earnings management. The baseline results show that the construction of pilot zones significantly reduces the degree of earnings management, and the conclusion still holds after replacing indicators, adjusting control variables and shortening the time window. Heterogeneity tests show that non-high-tech enterprises, small-scale enterprises and enterprises with weaker external supervision have more obvious policy responses, indicating that big data policies have a complementary effect on weak links in governance.

The research conclusions correspond to three policy implications. First, the construction of pilot zones should continue to be promoted, and data sharing and information transparency should be incorporated into the capital market governance framework to exert a restraining effect on corporate earnings manipulation. Second, policy implementation needs to focus on small-scale enterprises and groups with insufficient external supervision, and improve the marginal effect of policies through differentiated disclosure requirements and data supervision. Third, it is necessary to promote the coordinated operation of big data governance and traditional audit supervision, so that pilot zone policies and existing governance mechanisms complement each other, and jointly improve the quality of capital market information and the efficiency of resource allocation.

References

- [1] Yan, W. C., Wu, J. J., & Deng, L. (2026). Research on the governance effect of digital finance: From the perspective of earnings management. *Journal of Management Sciences in China*, 29(4), 40-65.
- [2] Qiu, Z. X., & Zhou, Y. H. (2021). Digital economy development and regional total factor productivity: An analysis based on national big data comprehensive pilot zones. *Journal of Finance and Economics*, 47(7), 4-17.
- [3] Shen, K. R., & Yan, J. M. (2024). Digital technology and enterprise total factor productivity: Influence effect and mechanism. *Collected Essays on Finance and Economics*(12), 3-15.
- [4] Li, Z. R., & Xu, Z. W. (2024). Data, production and macroeconomy: A general equilibrium theory with enterprise digitalization. *Management World*, 40(9), 80-101.
- [5] Qi, Y. D., & Xiao, X. (2020). Enterprise management reform in the era of digital economy. *Management World*, 36(6), 135-152+250.
- [6] Yang, J., & Zhu, Q. S. (2025). Impact of the establishment of national big data comprehensive pilot zones on enterprise innovation efficiency. *Statistics & Decision*, 41(18), 158-164.
- [7] Lin, X. S. (2025). Research on the enabling effect of digital China construction on enterprises' new quality productivity: Quantitative analysis based on multi-period difference-in-differences model. *Journal of Southwest University (Social Sciences Edition)*, 51(1), 148-164+304.
- [8] Xin, D. L., Yi, Y. Z., & Liu, Z. S. (2026). Excellence through multiplication: Big data development empowers enterprise supply chain diversification configuration——A quasi-natural experiment based on national big data comprehensive pilot zones. *Collected Essays on Finance and Economics*, 1-13. Retrieved June 8, 2026, from [URL not provided]
- [9] Wei, M. H. (2000). Basic theory of earnings management and its research review. *Accounting Research*, (9), 37-42.
- [10] Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 13(4), 365-383.
- [11] Bo, X. H., & Wu, L. S. (2009). Governance effects of state ownership and institutional investors: From the perspective of earnings management. *Economic Research Journal*, 44(2), 81-91+160.
- [12] Fang, H. X., & Jin, Y. N. (2011). Can high-quality internal control inhibit earnings management?——Empirical research based on voluntary internal control attestation reports. *Accounting Research*, (8), 53-60+96.
- [13] Su, D. W., & Lin, D. P. (2010). Equity incentives, earnings management and corporate governance. *Economic Research Journal*, 45(11), 88-100.
- [14] Chen, D. Q., & Chen, Y. S. (2018). Policy uncertainty and earnings management of listed companies. *Economic Research Journal*, 53(6), 97-111.
- [15] Huang, Q. C., Wen, Y. C., & Chen, Q. H. (2022). The impact of economic policy uncertainty on enterprises' real earnings management. *Securities Market Herald*(5), 69-79.
- [16] Jiang, Q. W., & Zhang, M. (2025). Digital transformation: "Shackles" or "invisibility cloak" of real earnings management? *The Chinese Certified Public Accountant*(7), 55-60.
- [17] Du, X. Q., Hong, J., Tao, H. X., et al. (2024). Is digital finance always beneficial to accounting information transparency? *Accounting Research*(10), 18-32.
- [18] Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *The Accounting Review*, 70(2), 193-225.