China Journal of Accounting Research

Volume 19, 4 (2026)

Accounting for capital, accounting for humanity – A Marxist view of AI and the role of accountants in the time of AI
Oliver Zhen Li 100506

The economic consequences of tightening insider trading regulation: evidence from Hong Kong
Zhihong Chen, Yuyan Guan, Bin Ke 100499

Digital transformation, information advantage and managerial opportunism
Yan Li, Yangyang Liu, Jing Dai, Hong Luo 100503

Predicting goodwill impairment from M&A comment letters: an explainable machine learning approach
Hongqin Tang, Jialan Huang, Nan Li 100504

TMT openness and corporate innovation: personality measurement using an LLM
Jinglu Zhao, Zeyu Zhang 100502

Computing power projects and digital transformation of enterprises: Evidence from the ''East data, West computing'' project
Lu Li, Chuanxing Tan 100500

Social credit systems and corporate AI development
Tai Wang 100481

Business groups and affiliated firms' digital transformation
Hui Xu, Haomin Wu, Wei Zheng, Kam C. Chan, Qiaoling Fang 100488

Corrigendum to ''Business groups and affiliated firms' digital transformation'' [China J. Account. Res. 19 (2026) 100488]
Hui Xu, Haomin Wu, Wei Zheng, Kam C. Chan, Qiaoling Fang 100514

Website:
http://www.elsevier.com/locate/cjar
http://www.cb.cityu.edu.hk/research/cjar/

19
4

China Journal of Accounting Research Vol. 19/4 (2026)

ELSEVIER

Consulting Editors
Bin Ke, *National University of Singapore*
Roni Michaely, *The University of Hong Kong*

Editors-in-Chief
Yangyang Chen, *City University of Hong Kong*
Junsheng Zhang, *Sun Yat-sen University*

Editors
Gennaro Bernile, *University of Miami*
Sean Cao, *University of Maryland*
Chen Chen, *Monash University*
Donghua Chen, *Nanjing University*
Thomas Jeanjean, *CCI Paris Île-de-France Education*
Ning Jia, *Tsinghua University*
Flora Kuang, *The University of Melbourne*
Edward Lee, *The University of Manchester*
Oliver Zhen Li, *National University of Singapore*
Ling Lei Lisic, *Virginia Polytechnic Institute and State University*
Feng Liu, *Xiamen University*
Jalal Sani, *University of Illinois Urbana-Champaign*
Yaowen Shan, *University of Technology Sydney*
Xijia Su, *China Europe International Business School (CEIBS)*
Jingsong Tan, *Sun Yat-Sen University*
Naohiro Urasaki, *Kindai University*
Xin Wang, *The University of Hong Kong*
Donghui Wu, *The Chinese University of Hong Kong*
Zhifeng Yang, *Stony Brook University*
Ning Zhang, *Queen's University*
Yuan Zhang, *Tulane University*

Editorial Board
Sudipta Basu, *Temple University*
Jeffrey Callen, *University of Toronto*
Shijun Cheng, *Shanghai Jiao Tong University*
Charles Hsu, *The Hong Kong University of Science and Technology*
Guohua Jiang, *Peking University*
Indra Wijaya Kusuma, *Universitas Gadjah Mada*
Clive S. Lennox, *University of Southern California*
Zengquan Li, *Shanghai University of Finance and Economics*
Bing-Xuan Lin, *University of Rhode Island*
Gerald Lobo, *University of Houston*
Changjiang Lv, *Fudan University*
Xiumin Martin, *University of Washington in St. Louis*
Suresh Radhakrishnan, *The University of Texas at Dallas*
Dan A. Simunic, *The University of British Columbia*
Hervé Stolowy, *HEC Paris*
Albert Tsang, *Southern University of Science and Technology*
Joanna Shuang Wu, *University of Rochester*
Stella Huiying Wu, *Western Sydney University*
Xi Wu, *Central University of Finance and Economics*
Zezhong Xiao, *University of Macau*

Editorial Director
Xinyi Zhang, Sun Yat-sen University

Editorial Office
Guilong Cai, *Sun Yat-sen University*
Pengdong Zhang, *Sun Yat-sen University*
Ying Zheng, *Sun Yat-sen University*

Volume 19 • Issue 4 • December 2026

China Journal *of* Accounting Research

HOSTED BY
ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

CYAN MAGENTA YELLOW BLACK

Aims and Scope

The focus of the *China Journal of Accounting Research* is to publish theoretical and empirical research papers that use contemporary research methodologies to investigate issues about accounting, corporate finance, auditing and corporate governance in the Greater China region, countries related to the Belt and Road Initiative, and other emerging and developed markets. The Journal encourages the applications of economic and sociological theories to analyze and explain accounting issues within the legal and institutional framework, and to explore accounting issues under different capital markets accurately and succinctly. The published research articles of the Journal will enable scholars to extract relevant issues about accounting, corporate finance, auditing and corporate governance related to the capital markets and institutional environment.

Production and hosting by Elsevier

Radarweg 29, 1043 NX Amsterdam, The Netherlands

ISSN 1755-3091

© China Journal of Accounting Research

Founded by Sun Yat-sen University and City University of Hong Kong

Sponsored by: 广州注册会计师协会

Published quarterly in March, June, September, and December

All rights reserved. No part of this journal may be reproduced, stored in a retrieval system or transmitted in any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Editorial Office of China Journal of Accounting Research.

Notice

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

http://www.elsevier.com/journals/china-journal-of-accounting-research/1755-3091/guide-for-authors

CHINA JOURNAL OF ACCOUNTING RESEARCH

Volume 19/4 (2026)

Available online at www.sciencedirect.com

ScienceDirect

HOSTED BY

ELSEVIER

Consulting Editors

Bin Ke
National University of Singapore

Roni Michaely
The University of Hong Kong

Editors-in-Chief

Yangyang Chen
City University of Hong Kong

Junsheng Zhang
Sun Yat-sen University

Editors

Gennaro Bernile
University of Miami

Sean Cao
University of Maryland

Chen Chen
Monash University

Donghua Chen
Nanjing University

Thomas Jeanjean
CCI Paris Île-de-France Education

Ning Jia
Tsinghua University

Flora Kuang
The University of Melbourne

Edward Lee
The University of Manchester

Oliver Zhen Li
National University of Singapore

Ling Lei Lisic
Virginia Polytechnic Institute and State University

Feng Liu
Xiamen University

Jalal Sani
University of Illinois Urbana-Champaign

Yaowen Shan
University of Technology Sydney

Xijia Su
China Europe International Business School (CEIBS)

Jingsong Tan
Sun Yat-sen University

Naohiro Urasaki
Kindai University

Xin Wang
The University of Hong Kong

Donghui Wu
The Chinese University of Hong Kong

Zhifeng Yang
Stony Brook University

Ning Zhang
Queen's University

Yuan Zhang
Tulane University

Editorial Board

Sudipta Basu, *Temple University*
Jeffrey Callen, *University of Toronto*
Shijun Cheng, *Shanghai Jiao Tong University*
Charles Hsu, *The Hong Kong University of Science and Technology*
Guohua Jiang, *Peking University*
Indra Wijaya Kusuma, *Universitas Gadjah Mada*
Clive S. Lennox, *University of Southern California*
Zengquan Li, *Shanghai University of Finance and Economics*
Bing-Xuan Lin, *University of Rhode Island*
Gerald Lobo, *University of Houston*
Changjiang Lv, *Fudan University*
Xiumin Martin, *University of Washington in St. Louis*
Suresh Radhakrishnan, *The University of Texas at Dallas*
Dan A. Simunic, *The University of British Columbia*
Hervé Stolowy, *HEC Paris*
Albert Tsang, *Southern University of Science and Technology*
Joanna Shuang Wu, *University of Rochester*
Stella Huiying Wu, *Western Sydney University*
Xi Wu, *Central University of Finance and Economics*
Zezhong Xiao, *University of Macau*

Editorial Director

Xinyi Zhang, *Sun Yat-sen University*

Editorial Office

Guilong Cai, *Sun Yat-sen University*
Pengdong Zhang, *Sun Yat-sen University*
Ying Zheng, *Sun Yat-sen University*

China Journal of Accounting Research 19 (2026) 100506



Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar



Accounting for capital, accounting for humanity – A Marxist view of AI and the role of accountants in the time of AI



Oliver Zhen Li

National University of Singapore

ARTICLE INFO

Article history:
Received 15 February 2026
Accepted 12 June 2026
Available online 16 June 2026

Keywords:
AI
Capital
Labor
Accountant
Karl Marx
Super-entity

ABSTRACT

Major Western political economists of the past 300 years had an exaggerated emphasis on the role of capital in the economy, with the exception of Karl Marx, who recognized the role of labor, and solely of labor, in creating social wealth. The advancement of modern technologies has hitherto disproportionally benefited capital, even though we see labor welfare also improve over the centuries. This time, however, AI is infinitely benefiting capital. How can humanity survive? We imagine the rise of new accountants, serving a "super-entity," to redistribute economic benefits to humanity. In search of historical perspectives, we go back in time. Divergent from the Western experience that accounting emerged for commerce, China's experience with accounting is that it often availed itself to central governance. Further, market-participating state capital can be super-competitive and well-informed. Facing an anti-human private-capital-empowered AI, rationing and redistribution under an intelligent "super-entity," enabled by the new accountants, take on a new meaning and urgency in taming unbridled private capital and helping preserve the physical existence of humanity.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

E-mail address: bizzhenl@nus.edu.sg

☆ This paper has been presented at the 2024 Antai Accounting and Capital Markets Conference of Shanghai Jiao Tong University, the 2025 13th Annual Meeting of the Accounting History Section of the China Accounting Association, Central South China University, Hunan University, Yan'an University, Xiamen University, Zhejiang University, Southern University of Science and Technology, Central South University and the National University of Singapore.

https://doi.org/10.1016/j.cjar.2026.100506
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. The role played by accounting in society and human history

Jensen and Meckling's (1976) article is well known in modern finance theory, having been cited more than 140,000 times (Google Scholar). Jensen and Meckling propose a creative and useful idea—the agency cost due to the separation of ownership and management of a firm. An entrepreneur's utility comes from two sources: the value of the firm and the perks that she consumes. When she owns 100% of her firm, she enjoys 100% of the benefit of the perks but also bears 100% of the cost of the perks. This time, the value of her firm is V^*. To grow her business, she decides to sell $(1 - \alpha)$, where $0 < \alpha < 1$, of her firm. Can she get $(1 - \alpha)V^*$? No. External investors will assume that, now that she only bears α of the cost of perks but enjoys 100% of the benefit, she will over-consume perks. This will cause the value of her firm to go down to $V\prime$ where $V' < V^*$. Therefore, $(V^* - V\prime)$ is the agency cost of the separation of ownership and management. How to solve this problem? Jensen and Meckling propose that an external accountant (auditor) can monitor the manager and prevent her from consuming an excessive quantity of perks, to the benefit of firm value.

This might be a way to solve capital's governance problem, and this is the essence of modern accounting—measurement, disclosure and monitoring. Should these be the only roles played by accounting, however? The problem with Western political-economic theories is the absence of sublime ideals. Since Adam Smith, the starting point has been the selfish nature of human beings. Are there other problems? Yes: short-sightedness in seeking empirical evidence and blindness due to ideology. Here, we discuss an important question: what is the purpose of accounting?

Accounting is a profession and science for rationing and redistributing limited resources. There are two aspects of redistributing: horizontal redistribution of limited resources to individual entities and longitudinal redistribution over intertemporal segments of time. The continued growth and prosperity of human civilization require good accounting and good accountants. In the Egyptian Hall of the British Museum, we can see some magnificent murals depicting the life of a man called Nebamun 3400 years ago. These murals reflect the hunting, agricultural and livestock-rearing activities of ancient Egyptian society. Nebamun was an accountant, but he was treated like a high priest, living a life of luxury and making important decisions. We can see that accountants occupied an important position in ancient Egypt.

Accounting is important to the continued existence of humankind and the development of societies. Without accountants, humans would not have existed to this day. In primitive societies with low productive capacity, if there had been no powerful accountants to ensure rationing, the stronger among the hunter-gatherers would have regaled themselves of all the exploits of a good hunt without leaving enough food for women, children, the elderly and other physically weaker members of society. The clan would have become smaller and feebler to the point that one day it would find itself incapable of competing against the natural elements, animals and other clans, leading to its demise. In an agrarian economy, without prudent accountants, if farmers consumed all the produce and did not leave enough for the next planting season, the clan would be diminished. Engels mentioned this in *The Origin of the Family, Private Property, and the State*: "*This implies, on the one hand, the production of the means of existence* (*food, clothing, shelter and the necessary tools*)*; on the other hand, the generation of children, the propagation of the species.*" The conflict between consumption and human reproduction, arising from limited resources, requires accountants with prudence and temperance to ration and redistribute resources horizontally and longitudinally.

Our contemporary society possesses advanced technologies and super-productive capacity. Do we still need rationing and redistribution by accountants? The answer is "yes." Our newest technology, artificial intelligence (AI), though capable of enabling super-productivity, could quickly deplete the Earth's limited resources, benefiting capital but squeezing out labor. Labor would be pushed out of the consumption cycle, leading to its atrophy, thus posing challenges to its very physical existence. Without accountants' rationing and redistributing, labor and humanity will cease. In the era of AI, the roles and influences of accounting far exceed the provisioning of information and monitoring in commerce.

This article starts from a world disturbed by AI and introduces the main topic—how accounting accounts for capital and humanity (labor). Through a Marxist lens, we discuss how private-capital-controlled AI squeezes out labor but infinitely benefits capital. The development of accounting in the Western world came from the development of commerce and banking, whereas in China, it historically comes from serving a "super-entity." In our world disturbed and challenged by AI, labor faces unprecedented oppression from pri-

vate capital, and the basic function of accounting becomes all the more obvious and urgent: via serving a super-entity, to restore a world disturbed and thus preserve the physical existence of humanity and maintain the balance between humans and nature.

2. The technology of our time

Let us look back upon Europe during the last 300 years. The 17th and 18th centuries, before the French Revolution, were an era of artists and thinkers. It was the time of Molière, Voltaire, Rousseau, Lully, Saint Colombe, Marais, Bach, Louis XIV. It was an era of love, an era of La Nouvelle Heloise, an era that was oblivious of asceticism, an era of great spirituality, an era of hope and awakening. The 19th century and the early part of the 20th century—the era after the Industrial Revolution—also had a multitude of scientists, artists and thinkers: Dickens, Hugo, Balzac, Marx, Darwin, Tolstoy, Freud, D.H. Lawrence, Berlioz, Satie, Proust. However, this was an era mixed of hope and despair. During this time, capital's acute woes became clearer and more pressing.

Ideas and technologies develop alongside each other. Hitherto, all technologies before the advent of AI had created benefits for humankind: fire to warm him, wheels to help her move, steam to give them power, wings to enable them to fly, computers to assist her thinking and the Internet to motivate him to connect. What is the world that we are living in now? We have invented AI. AI, though capable of super-efficiency, will only ultimately create human misery. AI is a technology that squeezes out humans, not a technology that elevates them. With the advent of AI, human sublimity will be lost, human dignity will be lost, human innocence will be lost, humankind's mastery over its own destination will be lost. Ultimately, humankind's right to physical existence on the Earth will be challenged.

3. How does Marx account for capital and labor?

To understand how AI will affect humanity, we need to start with Karl Marx. In fact, *Das Kapital* is an accounting book—it was started for the purpose of understanding accounting. In June 1861, Marx asked Engels for help on how to understand Dr. Quesnay's economic table: "*If it could be done very briefly, without making undue demands on you, I should like to have a sample of Italian book keeping* (*with explanations*). *It would help throw light on Dr Quesnay's* Tableau economique." (Marx–Engels Correspondence, 1862). Marx at that time did not know accounting, while Engels, who ran a large textile mill in Manchester, knew a lot. Learning accounting marked the turning point of Karl Marx from the "pamphleteer Marx" to the later "British Library Marx." However, it is important to note that double-entry bookkeeping might not have been first invented by bankers in Florence, Italy—something that we will discuss later in this article.

How did Marx account for capital and humanity? To account for capital and labor, Marx proposed his famous surplus value theory. Before production, we have the following (Capital: Volume I – Chapter 9: The Rate of Surplus Value):

$$C = c + v, \tag{1}$$

where C is total capital before production, which is composed of fixed capital c and variable capital v, that is, labor. After production, we have

$$C' = c + v + s, \tag{2}$$

where $C\prime$ is capital after production, and s is surplus value.

Marx believed that surplus value s is created by workers (labor), i.e., variable capital v, and not by fixed capital (machines and factories) c. Accounting researchers can perhaps appreciate this better. Fixed capital c is not depleted during production. There may appear to be the wear and tear of machines and factories, but capitalists categorize c as depreciation and add it back to capital, ensuring its recovery. This means that capital C is maintained during production. If we recognize surplus value, how do we compute the surplus value rate? The Surplus value rate is not s/C or $s/(c + v)$ as would some harmonizers divine. Instead, it is s/v. We know that $s/v > s/(c + v)$ (Capital: Volume I – Chapter 9: The Rate of Surplus Value). Further, as

discussed above, total capital C is not depleted during production. In fact, total capital C can be construed as the accumulation of surplus value created by many generations of labor.

How does capital accumulate? This is the concept of accounting flow to accounting stock. Capital in each production cycle is as follows:

$$\begin{aligned} c_t &= c_{t-1} + v_{t-1} + s_t \\ c_{t-1} &= c_{t-2} + v_{t-2} + s_{t-1} \\ &\cdots\cdots, \end{aligned} \tag{3}$$

where c_{t-1} is the starting value of capital for period t, v_{t-1} is the labor input for period t, s_t is surplus value created during period t. Summing up the above, we have

$$C = \sum\nolimits_{i=1}^{t} c_i = c_0 + \sum\nolimits_{i=0}^{t-1} v_i + \sum\nolimits_{i=1}^{t} s_i. \tag{4}$$

The profit rate p', from an inter-generational angle, is in fact the surplus value rate s'. The inter-generational profit rate p' is defined as

$$p' = \frac{\sum_{i=0}^{t} s_i}{c_0 + \sum_{i=0}^{t-1} v_i} = \frac{\sum_{i=0}^{t} s_i / \sum_{i=0}^{t-1} v_i}{c_0 / \sum_{i=0}^{t-1} v_i + 1}. \tag{5}$$

However, as the number of production cycles t increases, the profit rate p' approaches asymptotically the surplus value rate s', as follows:

$$\lim_{t\to\infty} p\prime = \frac{\sum_{i=0}^{t} s_i}{\sum_{i=0}^{t-1} v_i} = \frac{s}{v}, \tag{6}$$

as $t \to \infty$, $c_o / \sum_{i=0}^{t-1} v_i$ approaches 0. That is, the profit rate is the surplus value rate.

Therefore, total capital comes from labor, and not from Adam Smith's entrepreneurial spirit. Entrepreneurs and capital owners do not just rely on diligence, hard work, frugality or talent, but on the exploitation of labor to accumulate capital. Otherwise, they would be "priced out."

Now let us consider the initial capital c_0. Is it Adam Smith's "*Birth and fortune are evidently the two circumstances which principally set one man above another*"? The initial capital is obtained accidentally, haphazardly, and often via violence. We see the land enclosure movement in England, the westward expansion of the United States, colonization of Southeast Asia, Africa and Latin America by European powers, the slave trade, and pollution and carbon emissions since the Industrial Revolution. Capital via violence has a profound influence on the people, societies and regions affected by it. For example, the slave trade is a reason for the lack of trust among African peoples to this day and the relative lack of development in African countries (Nunn and Wantchekon, 2011). The westward expansion fostered a frontier individualism that still influences people's behavior (Bazzi et al., 2020, 2021). Unrequited commerce caused the lack of growth or development in colonial India.

Let us reference Ohlson (1995) here. We cite Ohlson not necessarily for scholarship, but because he defines the concept of residual earnings, tantalizingly reminiscent of Marx, as follows:

$$x_t^a \equiv x_t - (R_t - 1) y_{t-1}, \tag{7}$$

where x_t is the earnings for period t, y_{t-1} is the starting book value of equity for period t, R_t is the cost of equity capital for period t, x_t^a is the residual earnings for period t. The value of equity at the end of period t is

$$p_t = y_t + \sum\nolimits_{\tau=1}^{\infty} R^{-\tau} E_t[x_{t+}^a] \tag{8}$$

Ohlson (1995) adds a required return for capital recovery, r or $(R - 1)$. That is, capital providers demand a basic risk-adjusted return. Residual earnings is the amount of income in excess of this return plus the recovery of the capital. In a prima facie comparison, we can consider capital's residual earnings as a way to aid our understanding of surplus value. In other words, surplus value is the value in excess of expected return of cap-

ital plus the capital itself. Marx does not consider a discount rate. Therefore, his theory is more general and generous. Of course, Marx accounts for capital's historical cost from capital accumulation, while Ohlson (1995) and the modern finance theories account for capital's forward-looking discounted value.

Capital's forward-looking discounted value is also reminiscent of Schumpeter. Of course, here we cannot compare Ohlson with Marx or Schumpeter. There is only a thin veil between capital's forward-looking value and capital's historical accumulation. Capital's required return for capital recovery, r or $(R-1)$, already represents risk-adjusted return. It is not fundamentally different from the depreciation of plants and machinery, but merely adds a further threshold. Schumpeter believes that economic growth comes from innovation. However, not all capitalists can recover their capital. The advent of new technologies can benefit some capitalists while ruining others. This is considering capitalists individually at a point in time. We are here talking about how capitalists, as a class, recover capital and extract surplus value. We have already shown that capital itself is created and accumulated by labor (thus workers cannot morally destroy plants and machinery in an anarchist fashion—see Zola's *Germinal*). Further, Schumpeter must have borrowed a lot of inspiration from Marx. Capital's agitation and desire for innovation are clearly described in the *Manifesto of the Communist Party*, and in a more forceful way: "*The bourgeoisie cannot exist without constantly revolutionizing the instruments of production, and thereby the relations of production, and with them the whole relations of society. Conservation of the old modes of production in unaltered form, was, on the contrary, the first condition of existence for all earlier industrial classes. Constant revolutionizing of production, uninterrupted disturbance of all social conditions, everlasting uncertainty and agitation distinguish the bourgeois epoch from all earlier ones.*" Agitation is a hallmark of capital. Losses suffered by some individual capitalists should not be considered as a sacrifice made by capitalists as a class. Below, we further argue that capital enjoys most of the benefits of technological innovations.

We know that, with one production cycle, profit can be expressed as follows (Capital: Volume III – Chapter 3: The Relationship between Rate of Profit and Rate of Surplus Value):

$$p' = \frac{s}{C} = \frac{s'v}{c+s}. \tag{9}$$

If the turnover ratio is n, we have

$$p' = \frac{ns'v}{C}. \tag{10}$$

Therefore, we understand that capital cares for labor productivity—labor productivity that comes from repetition or aided by technologies. This brings us to two more advanced accounting concepts: absolute surplus value and relative surplus value. Absolute surplus value comes from the prolongation of labor hours or cheating. Relative surplus value comes from repetition (Adam Smith's "dexterity") and especially machine-enabled more efficient labor.

The concept of relative surplus value is particularly useful for understanding how new technologies impact capital and labor. With new technologies, the surplus value rate s' will increase, leading the profit rate $p' = s'v/C$ to increase, without the prolongation of labor hours, or even with a reduction of labor hours. This means an absolute increase in wages or an absolute enhancement in workers' quality of life. Note that an absolute increase in wages or an absolute enhancement in workers' quality of life is not a fair outcome for labor vis-à-vis capital. This is the situation of major developed capitalist countries since the end of the Second World War. However, we have witnessed a climax of globalization since then. Globalization brings cheap labor to Western capitalist countries. Therefore, it is still labor, and not technology, that is a source of competitiveness. Globalization causes labor distilled into the same product to differ across different regions and at the same time causes this difference to dissipate, leading labor to become homogeneous internationally. During this process of dissipation, labor in different regions can conflict (Autor et al., 2013, 2020). Of course, globalization was long ago mentioned in the *Manifesto of the Communist Party* as a source of competitiveness. We have "*Free Trade*," "*raw materials from the remotest zones*," "*consumed ... in every quarter of the globe*," "*the Chinese market*," and of course we do not lack a mention of "*the East-Indian market*," nor "*the colonization of America*." Lenin also raised the concept of "surplus capital" roaming the world ("Imperialism, the Highest Stage of Capitalism"). Of course, surplus capital now also has an urge to roam the world to evade taxes. It

is a recent phenomenon that developed Western capitalist countries want to retreat from international trade, globalization and sustainability. This is because they find it hard to compete with certain examples of state capital or economic planning (we will cover this later). Overall, capital always wins over labor. Labor's quality of life is raised absolutely but not relatively as the distance in welfare between labor and capital grows.

We have already mentioned that before the arrival of AI, new technologies could cause some workers to become unemployed. However, the redeployment of labor is possible, although it needs to overcome some friction: from driving coaches to driving cars, from using typewriters to using computers, from data entry to data analysis. There have been occasions of workers destroying machinery in the process. However, the harm and inhumanity caused by capital deliberately destroying over-capacity when faced with an elastic demand curve for its products greatly exceeds the harm due to workers destroying machinery. The reason that capital escaped self-destruction at the end of the 19th century and the beginning of the 20th century is that the two world wars greatly allayed capital's acute woes. Class enmity was turned into racial enmity or the conflict to carve colonies out of Asia or Africa. We can also consider the triggers of the two world wars to have been not exogenous or accidental (the assassination in Sarajevo in 1914 and Germany's invasion of Poland in 1939) but inevitable outcomes of capital's striving for survival or incidents unconsciously created by capital to reduce internal pressure or to carve up the world in a new way. Globalization after the Second World War gave capital more space for imagination. However, with globalization and international trade, capital's return still comes from labor, from Asia, from Africa and from Latin America. To summarize, technologies before AI can be considered unshackling technologies. Unshackling technologies, while benefiting capital, has also absolutely raised labor's standard of living, even though the distance between labor and capital is enlarged.

In contrast, private-capital-empowered AI is anti-human. AI has a stifling effect on labor and brain power. With other technologies, the resuscitation of labor can be achieved, even though it may face friction and require time. In production, labor is still needed, consumption is there and the economy can churn. Labor's welfare will improve, albeit only in absolute but not relative terms. This will not be the situation with AI. AI will cause an atrophy of labor (and brain power and human society). Labor (and brain power) is no longer needed and thus our society implodes. Let us imagine this. In a world endowed with a super-abundance of products, 99.99% of the people cannot consume; only 0.01%, represented by capital owners, can consume. Capital can achieve abundance with abandon, but there is no consumption by the masses. This would be the best and perhaps the ultimate illustration of the "*epidemic of overproduction*" (Manifesto of the Communist Party). We are already seeing this now. While AI is helping to raise efficiency, employment is low, economic growth is limited even though the stock market sometimes behaves flamboyantly, and the gloom of global military conflict looms over us.

With the advent of private-capital-empowered AI, the relative surplus value rate reaches almost infinity ($p\prime = s'v/C \to \infty$). However, machines and technologies care only for capital recovery and create no surplus value. Thus, an infinite surplus value rate benefits only capital and not labor. Ironically, with private-capital-empowered AI, even though labor creates near-infinite surplus value and indeed enables the very accumulation of capital to begin with, it is rejected. A great majority of workers will be squeezed out of the labor market with no possibility of redeployment, and thus face the challenge of losing the right to physical existence. This is the fundamental difference between AI and other technologies that humans have experienced. Using the language of classical political economists, we would say that labor's spoon to dip into the soup will become infinitely small. Workers' organized resistance against capital will be futile as capital will have a multitudinous horde of robots and drones to protect it. Private-capital-empowered AI is unabashed in driving out labor.

Marx has his four famous alienations (*Economic and Philosophical Manuscripts*, 1844): workers are alienated from the product of their labor, which they do not own; workers are alienated from the process of production, as they are appendages of machines; workers are alienated from the species-being, as they cannot realize a sense of satisfaction and live for the purpose of living; workers are alienated from other human beings, as they have to compete for work and wages. With private-capital-empowered AI, workers have one more alienation: workers are alienated from human sublimity, as they are held dominion by an all-puissant non-biological computing power. This non-biological computing power runs factories, moves commerce, manages banks and services, influences people's minds and, ultimately, expels labor's physical existence.

4. Private capital's egoism and private-capital-empowered AI

Throughout humankind's existence, we have been forced to think again and again about evolution and ethics. Darwin (*The Voyage of the Beagle*; *The Origin of Species*) and Huxley (*Of Evolution and Ethics*) contemplated the competition between species and the natural elements (wind, air, water and fire); among species; and between humans and other species. Marx considered the competition among classes (people). In extreme cases, Lenin in 1918 and Stalin during his purge emphasized class antagonisms (and not racial conflicts), while Hitler in the Third Reich stressed racial conflicts (and not class antagonisms). However, the two are of a transmutational relation. Allow us to state here that the conflict between labor and capital can be assuaged by "reincarnation." In Yukio Mishima's *Temple of Dawn*, the antagonist Honda witnesses the reincarnation from Isao Iinuma to the Princess of the Moon. Hermann Hesse in *Siddhartha* describes Siddhartha's experience with asceticism and carnal pleasure. We only have an analogy here. Capital and labor can mix over a time period and do not have to conflict at a given point in time. That is to say, capital and labor do not have to conflict as individuals, but only as classes. Remember that Engels was a major capitalist managing a textile mill in Manchester. This can perhaps help us better understand Marx's class struggle theory in our modern-day context.

Now, the evolution and ethics that we need to contemplate is the competition between human and computing power. Private-capital-empowered AI is obviously ingrained with the egoism of private capital. Marx, in his earlier writing, criticized the concept of natural rights expressed in the *Declaration of the Rights of Man* (1791) from the French Revolution. Human nature is egoistic as well as collective. The French Revolution was a bourgeois revolution that emphasized egoism but not collectivism: "*None of the so-called rights of man, therefore, go beyond egoistic man, beyond man as a member of civil society – that is, an individual withdrawn into himself, into the confines of his private interests and private caprice, and separated from the community.*" (Marx: On the Jewish Question, 1844). Therefore, major European powers were engaged in "colonialism," "slave trades," "war" and "pollution." Colonialism revived capital's yearning for natural resources for its own development. Colonized India had commerce with Great Britain but could not develop and could not consume. Put nicely, this was a remittance economy of a sort—"unrequited transfers"—toward England! (Dadabhai Naoroji). Put more bluntly, this was pure exploitation. Slave trades helped allay capital's thirst for labor. This thirst can be observed frequently in the novels of Gabriel Garcia Marquez. Materialism can metamorphose into ruptures in thinking and ethics and become racism (Darwin: *The Voyage of the Beagle*). Wars have been a constant reality in Europe (Voltaire: *Candide*), even though the major European empires were connected via marriages and blood. Henry V married Princess Catherine after the Battle of Agincourt, romantically but also materialistically, as aptly described by Shakespeare. Louis XVI's queen Marie-Antoinette was an Austrian princess. Catherine the Great was a Prussian princess. The First World War was an episode of mayhem among European empires intricately connected via marriage and blood. Every two decades, there is a major European war, even now. Of course, there is no eternal hatred among European powers. Pollution has been a source of high standards of living since the Industrial Revolution. Europe and North America have always been the greatest polluters of the world (Our World in Data, 2025). Of course, they now want others' emissions to be curtailed. Therefore, private capital manifests human egoism. The French Revolution only abolished feudal lords' rights over tenant peasants. This was progress, but one with no property has no rights. From Zola's *Germinal*, we know, from old Maheu: before, they were the king's soldiers; now they are the Republic's soldiers.

Imbued with capital's egoism, private-capital-empowered AI will cause labor to be challenged spiritually and physically. Decentralization, popular not long ago, was merely a recent attempt by capital to obtain free rein. As discussed earlier, in a world of AI purely empowered by private capital, labor will be quickly squeezed out. Only a collective entity, a common entity, an entity above private capital and above individual capitalists, a super-entity, can constrain private capital and private-capital-empowered AI and return the right of physical existence to labor. Let us temporarily imagine this super-entity as a centralized entity. Whom does accounting serve? In Western capitalistic countries, accounting serves private capital (Jensen and Meckling, 1976). In a major East Asian country such as China, it has the history and experience of serving a super-entity or a centralized entity.

5. Historical roots of accounting serving a super-entity

Let us take a look at the history of accounting serving a super-entity/centralized entity. To explore the origin and evolution of written languages, researchers have contemplated the meaning of symbols and patterns carved in caves and onto bones and rocks. Some archeologists and accounting researchers believe that at least some of the carved symbols and patterns were related to early accounting activities, and bookkeeping appeared before the advent of written languages. In 2002, a batch of wooden slips (BCE 217–210) was excavated in Liye Township, Longshan County, China. These slips provide opportunities to support the conjecture that bookkeeping appeared before written language (Chen et al., 2025b). Each slip records written details of transactions and also has carved notches on its edges representing numerical quantities of money and matters involved in the transaction. This method unites written records with numerical etches. The notches thus obviously were accounting-related. Before the discovery of these slips, archeologists had already found many examples of bookkeeping slips purely using written languages. Why did notches reappear in bookkeeping slips?

The answer can be found in the historical backdrop of China at that time (Chen et al., 2025b). This batch of slips was made during BCE 217–210, a period of unity in the Qin Dynasty (BCE 221–207). Before the Qin Dynasty, China underwent the Warring States Period (BCE 475–221) and was divided into seven states. Qin defeated the six other states and united China. Before unification, each state had its own unique language. After unification, the emperor of Qin conducted dramatic reforms, including the standardization of the language as well as money, law and measurement, thus building a centralized country. This batch of slips was unearthed in Hunan Province, which belonged to the State of Chu. Chu's written characters were very different from those of Qin. After the Emperor of Qin's standardization of written characters, official documents had to be written in Qin. This could have caused a temporary dip in literacy in the State of Qin, allowing bookkeeping fraud to increase. At the same time, Qin's penal codes were severe and used collective punishment. Therefore, the risk of Chu officials' being held accountable for bookkeeping errors increased. The demand for accounting accuracy and the prevention of fraud likely caused officials to add notches to match written bookkeeping. Thus, from the beginning, China's bookkeeping and accounting served a super-entity/centralized entity.

We have another example. Conventional wisdom in accounting research is that the modern double-entry bookkeeping system originated from Florence, Italy in the 13th century. The development of the banking business and its participation in commerce and finance caused double-entry bookkeeping to emerge (Sangster, 2016, 2018). The first evidence of double-entry bookkeeping is the two parchments from a certain Florentine bank in 1211. Pacioli's (1494) work was a formal double-entry bookkeeping textbook that summarized bookkeeping techniques used by Italian banks. That is to say, in Europe, double-entry bookkeeping originated from private banking business. Recent research (Chen et al., 2025a) finds that rudiments of double-entry bookkeeping appeared in China in the Kingdom of Wu during the Three Kingdoms Era (220–280). A large quantity of bookkeeping slips from the Kingdom of Wu were excavated in Zoumalou in the City of Changsha in September 1996. These slips have the characteristics of double-entry: every transaction involves two accounts, two bookkeeping symbols (“in” and “out”), cross-reference tokens; every account has information on the imperial calendar, names of accountants handling the transaction and administrative regions; there are accruals, principal and interest; and there are signs of internal control. Owing to the large quantity of slips (whereas the Florentine parchments had only a few legible entries), researchers can observe patterns: rice taxes were mainly collected after the harvest of rice, during October, November and December, while cloth taxes were largely collected during the rice-growing phase when peasants had the time to weave, namely July and August.

Thus, in Italy (13th century), double-entry bookkeeping emerged from private commerce and banking; in China (third century), it emerged from central governance—it served the tax administration of the central government. Therefore, from its origin, accounting does not necessarily serve private capital, but a super-entity. Littleton's (1927) proposed precedents of double-entry bookkeeping—writing, arithmetic, private property, money, credit, commerce and capital—only reflect the views of Western capitalist countries, not those with different economic, historical and political conditions and environments. In China, double-entry bookkeeping originated from the need to manage central government's tax revenue and finance. To collect taxes, the gov-

ernment needs to track and measure taxpayers and tax bases, leading to double-entry bookkeeping. At the same time, in years of famine and economic adversity, tax collection has to be postponed, leading to accruals and accounting for interest. These are important elements of double-entry bookkeeping.

The above-discussed research suggests that accounting has historically served a super-entity. In our new era, faced with new technologies such as AI, accounting's role in enabling rationing and redistribution is becoming all the more important. In fact, even Western economies do not shun redistribution, and we see surveillance and regulations in all dimensions. A degree of redistribution is an indispensable aspect for all governments.

6. New accounting and super-entity

Discussing a super-entity, many scholars perhaps would point to the fact that the former Soviet economy was not a success. Some invariably would want to quote Hayek's *The Road to Serfdom*. As discussed earlier, if we construe a super-entity as a central government, then it is true that the former Soviet Union, unable to appreciate factor prices, found it essential to use quotas to eliminate arbitrage, even though arbitrage is one important aspect of human societies. However, China's existing state capital was forged and baptized in market competition. This state capital understands factor prices and thus is capable of realizing the goal of redistribution. The long-term success of China's industrial policies and economic planning in global competition is obvious. China's elimination of abject poverty is also successful. Many countries, including the U. S., are learning from China, willingly or unwillingly.

Hayek's premises no longer stand in our rapidly changing world and evolving technologies. The current world is different from the one Hayek envisioned 80 years ago. Technological advances and the participation of state capital have greatly enhanced the possibility and feasibility of economic planning. Hayek is ideological and abhors collectivism, and Western empirical economists did not want to wait long before providing empirical evidence. Economic planning is not automatically dry, monotonous or arbitrary. Even Leninist economic planning was not arbitrary but was devised to be consistent with the development needs of the Soviet economy at that time. It only erred operationally. China's current economic planning is planning under market competition. International and national governance is a biosphere that needs diversity. State capital or capital under state guidance can in fact be super-competitive, so competitive that Western tariffs and sanctions—very non-market measures—have to be deployed to stifle its growth. Therefore, human society does not have to end with liberal democracy.

Simone de Beauvoir wrote the book *La Longue Marche* after visiting China for three months in 1959, describing China's economic, political, cultural and national defense development in the 1950s. This book has a rare candidness and astuteness among Western writers talking about China. Unfortunately, after the publication of this book, China entered the Great Leap Forward era and then the Cultural Revolution. Some Western scholars, feigning politeness, have stated that de Beauvoir made honest mistakes. This is not necessarily the case. Looked upon in the wake of China's economic reform of the 1970s, the book is correct in describing the Chinese psyche and spirit. De Beauvoir correctly saw the historical sense of mission and purpose of China's politics and economics. Even though 1959 was followed by the Great Leap Forward and the Cultural Revolution, we need a long-term vision and a sense of path-dependency to understand China's growth. China's development is a result of a hardworking people and a government that unwaveringly moves toward the correct goals. There have been mistakes made on the way, but they were quickly rectified.

In fact, the inspiration for Hayek's *The Road to Serfdom* likely came from Marx. Marx frequently mentioned slavery and forçats. In fact, it is unchecked private capital that engenders slavery and forçats. This slavery is not initially political but economic. However, it will become political, and ultimately it will metamorphose into spiritual servitude. That is to say, with the "absolute enhancement" in workers' standard of living in major developed capitalistic countries since the end of the Second World War, slaves and forçats are voluntary. Capital has always been trying to lure one into servitude—platforms, annual fees, registration and various devices for attention control—everything is done in quietus. The sole purpose of these is to lure consumers away from competition, cut off the true sources of information and enable capital to monopolize—the ultimate dream of capital. AI enhances private capital's ability to monopolize the market. Thus, private capital engenders its own slaves and tyrants. We can see this kind of phenomenon everywhere now: the

idolization of capitalists; private-capital-created fake news; everything for efficiency; a general befuddlement of the mind; and convenient academic ideals and monopoly.

Why do we need to restrain private-capital-empowered AI? The utility function of private-capital-empowered AI is a "Tower of Babel" kind of utility function. Its sole purpose is seeking efficiency and profits for private capital without the slightest scruple for labor. In fact, "low efficiency" is the natural resource and gift of our Earth and society. The very reason that humanity can exist is due to this low efficiency. Low efficiency subdues the avarice of human nature and maintains humankind's spiritual modesty, constraining capital's ability to extract surplus value. At the same time, low efficiency limits the ability of humans to exploit the Earth's natural resources, enabling our planet to sustain human activities for a longer period of time. Super-efficiency enables private capital to obtain a closed-form solution for all problems. Unabashed exploitation of natural and social reserves threatens the sustainability of humankind. When labor is ultimately removed from capital's utility function, workers, without capital, have no means for sustenance, and thus no right to physical existence. Diversity, as a means of checks and balances, will disappear. Literature, music, arts and love will die. Before, we only faced a mysterious god; now, we may be held dominion by a private-capital-empowered AI.

Ironically, with our current super-productivity, human existence will again become dependent on rationing and redistribution, like in prehistoric times. As discussed earlier, the advent of AI will quickly cause 99.99% of people to lose their employment and their ability to consume, while only the 0.01% represented by capital owners will survive. The chasm between capital and labor will become infinite. Under this scenario, to sustain humanity, to sustain labor, we need vertical and horizontal rationing and redistribution. This kind of system requires a form of public or quasi-public capital, or a public ownership of the means of production. This public or quasi-public capital or ownership is different from the diverse stock ownership in major developed Western economies. Diverse stock ownership is not equal to public ownership or ownership by all. Most people do not have capital. Although everyone may have some sort of ownership, most people are atomistic owners and cannot compete with large capital-owners; thus, they have no voice. Of course, institutional ownership or class action suits can potentially alleviate this. However, this is equivalent to ceding small ownership to large private capital.

The situation would be different with public capital. Let us imagine. Under a super-entity endowed with super-productivity, public or quasi-capital will no longer produce profits but the means of human physical existence.

7. Conclusion

Our future world, if it can be sustained, will have to be one that relies on rationing and redistribution. Can dignified rationing and redistribution be made possible? Under current technological conditions, rationing and redistribution are enabled by near-perfect information. A mechanism of common or public ownership is needed to restrain the savage growth of private capital. At least, it should be a system where public capital and private capital grow side by side. The key to this redistribution system is not welfare, not charity, but common ownership.

Now let us return to the concept of a super-entity. What is a super-entity? It is an entity that restrains the savage growth of private capital, an entity that embraces, at least to a certain extent, common capital ownership. The terminal stage of humankind is likely to be built upon common ownership. Accountants will play a significant role in this stage. What are new accounting and new accountants? Accounting is a profession that supports the highest level of social management and governance in a world with super-productivity, representing a super-entity, providing rationing and redistribution, to ensure the physical existence of humankind. Accounting, which aided humankind's limping early existence, will accompany it to its very end.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Autor, D., Dorn, D., Hanson, G., 2013. The China syndrome: local labor market effects of import competition in the United States. Am. Econ. Rev. 103 (6), 2121–2168.

Autor, D., Dorn, D., Hanson, G., Majlesi, K., 2020. Importing political polarization? The electoral consequences of rising trade exposure. Am. Econ. Rev. 110 (10), 3139–3183.

Bazzi, S., Fiszbein, M., Gebresilasse, M., 2020. Frontier culture: the roots and persistence of "rugged individualism" in the United States. Econometrica 88 (6), 2329–2368.

Bazzi, S., Fiszbein, M., Gebresilasse, M., 2021. "Rugged individualism" and collective (in)action during the COVID-19 pandemic. J. Public Econ. 195, 1–20.

Chen, M., Huang, L., Li, O.Z., 2025. Rudiments of double entry bookkeeping in ancient China's tax administration – evidence from bamboo slips from the third century. Working paper.

Chen, M., Huang, L., Li, O.Z., Zhang, C., 2025b. Ancient notching or tokening as bookkeeping –evidence from wood slips in China during 217–210 BCE. China J. Account. Res. 18, 1–7.

Darwin, C., 2003. The Origin of Species and the Voyage of the Beagle. Everyman's Library.

Engels, F., 2010. The Origin of the Family, Private Property, and the State. Ernest Untermann.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3 (4), 305–360.

Lenin, V.I., 1916. Imperialism, the highest stage of capitalism. Lenin's Selected Works, vol. 1. Progress Publishers, 1963, Moscow, pp. 667–766.

Littleton, A.C., 1927. The antecedents of double entry. Account. Rev. 2 (2), 140–149.

Marx, K., 1862. Marx–Engels Correspondence vol. 41, 380, First published in Der Briefwechsel zwischen F. Engels und K. Marx, Stuttgart, 1913.

Marx, K., 1844. On the Jewish Question. First published February 1844 in Deutsch-Französische Jahrbücher.

Marx, K., Engels, F., 1848. Manifesto of the Communist Party.

Marx, K., 1976. Capital: Volume I. Penguin Classics.

Marx, K., 1981. Capital: Volume III. Penguin Classics.

Nunn, N., Wantchekon, L., 2011. The slave trade and the origins of mistrust in Africa. Am. Econ. Rev. 101 (7), 3221–3252.

Ohlson, J.A., 1995. Earnings, book values, and dividends in equity valuation. Contemporary Accounting Research 11 (2), 661–687.

Our World in Data, 2025. https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

Pacioli, L. 1494. Summa de arithmetica, geometria, proportioni et proportionalita. Reproduction of original from Goldsmiths' Library, University of London. Translation by Geijsbeek, J. B. 1914. Ancient double entry bookkeeping by Lucas Pacioli. Houston: Scholars Book Co.

Sangster, A., 2016. The genesis of double entry bookkeeping. Account. Rev. 91 (1), 299–315.

Sangster, A., 2018. Pacioli's Lens: God, humanism, Euclid, the rhetoric of double entry. Account. Rev. 93 (2), 299–314.

Smith, A., 1994. An Inquiry Into the Nature and Causes of the Wealth of Nations. Modern Library Edition.

China Journal of Accounting Research 19 (2026) 100499

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

The economic consequences of tightening insider trading regulation: evidence from Hong Kong

Zhihong Chen [a], Yuyan Guan [b], Bin Ke [c,*]

[a] Department of Accounting, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[b] Division of Accounting, Nanyang Business School, Nanyang Technological University, 91 Nanyang Ave, 639956, Singapore
[c] Department of Accounting, Business School, National University of Singapore, Mochtar Riady Building, BIZ 1, # 07-30, 15 Kent Ridge Drive, 119245, Singapore

ARTICLE INFO

Article history:
Received 11 April 2026
Accepted 15 May 2026
Available online 20 May 2026

JEL:
G14
G18
G38

Keywords:
Insider trading regulation
Blackout window
Shareholder value
Hong Kong

ABSTRACT

Using a lobbying event that led to the unexpected reversal of a tougher insider trading blackout regulation in Hong Kong, we examine whether tightening insider trading regulation increases shareholder value. We find that repealing the regulation increases shareholder value for the entire Hong Kong stock market. The increase is greater for the lobbying firms, which are more affected by the regulation. We find no evidence supporting the allegations by proponents of the regulation that insiders at lobbying firms trade more aggressively on future earnings news during the proposed new blackout period or delay earnings announcements. Supporting the argument of the regulation's critics, we find that lobbying firms' insiders trade more aggressively during the proposed new blackout period to stabilize their firms' stock prices amid stock price turmoil. Our results suggest caution in imposing tougher insider trading regulations.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding author.
E-mail addresses: aczh@ust.hk (Z. Chen), yuyan.guan@ntu.edu.sg (Y. Guan), bizk@nus.edu.sg (B. Ke).

https://doi.org/10.1016/j.cjar.2026.100499
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Major financial markets worldwide have adopted insider trading regulations (Bhattacharya and Daouk, 2002). Despite regulatory restrictions on insider trading, it remains frequent and profitable in many financial markets (e.g., Seyhun, 1998; Hillier and Marshall, 2002; Lee et al., 2014; Zhu and Wang 2015; Brochet, 2019).[1] While some prominent commentators tout the benefits of insider trading and oppose restrictive insider trading regulation (Manne, 1966; Carlton and Fischel, 1983; Mak, 2008), critics question the adequacy of existing levels of insider trading regulation (Easterbrook, 1981; Macey, 1991; Williams, 2008).

In recent years, some securities regulators have also called for the further tightening of existing insider trading laws. For example, Hong Kong securities regulators adopted legislation in 2008 to further tighten the insider trading blackout regulation (HKEX 2008a). Preet Bharara, former U.S. attorney, and Robert J. Jackson Jr., an SEC Commissioner, criticized the U.S. insider trading laws as outdated because prohibitions against insider trading have been based on years of case law and SEC guidance stemming from the broader antifraud provisions of the Securities Exchange Act of 1934 (Bharara and Jackson, 2018). On 5 December 2019, the U.S. House of Representatives voted 410 to 13 to pass a bipartisan bill, H.R. 2534 (the Insider Trading Prohibition Act), that would write an explicit ban on insider trading into law (Flook, 2019).

We contribute to this ongoing debate by investigating whether tightening insider trading regulation increases shareholder value. As we explain in the next section, there could be both costs and benefits of regulating insider trading. Since these costs and benefits are difficult to quantify ex ante, the impact of tightening insider trading regulation on shareholder value is an empirical question.

We test our research question using an event study approach by exploiting the unexpected reversal of a 2008 insider trading blackout regulation in Hong Kong that would have substantially increased the blackout window. Easterbrook (1985) argues that the event study is the best available test of the benefits and costs of insider trading regulation to shareholders since many of the benefits and costs are not directly observable.

Insider trading regulations are typically bundled with other regulations when proposed (Bekaert and Harvey, 1997; Jayaraman, 2012), making it difficult to isolate the effect of an insider trading regulation from other confounding regulations in an event study. We overcome this challenge by exploiting the unexpected reversal of Hong Kong's 2008 insider trading blackout regulation. Prior to 2009, Hong Kong-listed firms were subject to a one-month insider trading blackout period before each earnings announcement. In December 2008, the Hong Kong Securities and Futures Commission (SFC) passed a much tougher regulation that would increase the blackout window from two months to a maximum of seven months per year, a substantial tightening of insider trading restrictions. However, on 29 December 2008, just three days before the new regulation's effective date, the directors of many influential Hong Kong-listed companies suddenly objected to the new blackout regulation in an open lobbying letter, prompting a swift reversal. The reversal of the new regulation provides a clean setting for our event study because it was not bundled with any other regulatory events and was unexpected by the stock market.

Because the new blackout regulation applies to all Hong Kong-listed firms, we first examine the cumulative stock market reactions to announcements of events that led to the unexpected reversal of the new regulation across the *entire* Hong Kong stock market portfolio. Hong Kong is an open market economy with strong ties to mainland China's economy. Hence, Hong Kong stock market returns are influenced by both global and mainland China-specific economic factors. Following Armstrong et al. (2010), we use the Dow Jones STOXX Global 1800 Index, excluding Hong Kong-listed firms (hereafter, DJ STOXX 1800 ex HK), to control for global economic factors. We use mainland China's stock market index to control for mainland Chinese economic factors.

We find that the cumulative abnormal stock return to announcements of events that led to the unexpected reversal of the new blackout regulation is significantly positive across the entire Hong Kong market portfolio. This result suggests that the stock market views the new insider trading blackout regulation as detrimental to shareholder value across the entire Hong Kong financial market.

[1] Unless stated otherwise, insider trading in this paper is limited to legal insider trades by corporate executives and directors of publicly listed firms, which are required to be publicly disclosed in many financial markets.

We also examine whether the event study results are stronger for firms more affected by the new regulation. To the extent that the new blackout regulation hurts shareholder value, we should expect the event study results to be more positive for firms more affected by the new regulation. A benefit of such a cross-sectional analysis is that we can further control for unobservable confounding macro-news that is unique to Hong Kong. Following Hochberg et al. (2009), we argue that firms more affected by the new regulation should be more likely to lobby against its implementation. Consistent with this argument, we find that firms that have traded more heavily during the proposed new blackout window are more likely to lobby against implementing the new regulation. Hence, we use lobbying firms as a proxy for more-affected firms. We find that the cumulative abnormal stock returns to the events that led to the unexpected reversal of the new regulation are significantly more positive for lobbying firms than for a matched sample of non-lobbying firms, following the methodology of Hochberg et al. (2009).[2] This cross-sectional event study result for the more affected firms is consistent with the event study result for the overall Hong Kong stock market portfolio, providing further support for the view that the new insider trading blackout regulation hurts shareholder value.

In addition to the event study approach, we test the validity of some costs and benefits alleged by supporters and critics of the new blackout regulation. We first test the validity of two allegations made by Hong Kong securities regulators (Williams, 2008): (a) Hong Kong insiders trade aggressively to exploit their information advantage about future earnings news; and (b) Hong Kong insiders deliberately delay their earnings announcements to allow themselves more insider trading opportunities. As lobbying firms are more affected by the new blackout regulation, we test these two allegations by comparing insider behavior between lobbying and non-lobbying firms. First, we examine whether insiders at lobbying firms are more aggressive than those at non-lobbying firms in trading on forthcoming earnings news during the proposed new blackout window. We find no such evidence. Second, we examine whether insiders at lobbying firms are more likely than those at non-lobbying firms to delay their firms' earnings announcements to maximize their insider trading advantage prior to the results' announcement. Even though the lobbying firms are strong opponents of the new blackout regulation, we find no evidence that their earnings reports are released significantly later than those of non-lobbying firms. Overall, these two results show no evidence that these lobbying firms oppose the new insider trading regulation for opportunistic reasons.

Next, we test the validity of the arguments advanced by the lobbying firms against the new blackout regulation. As noted in section 2, the lobbying firms listed several highly negative consequences of the new regulation. Unfortunately, most of the listed consequences cannot be directly tested because the new blackout regulation ultimately did not take effect. However, we can use available data to shed light on the following claim: prohibiting insider trading during the proposed new blackout window would further reduce trading volume, thereby jeopardizing an important source of market liquidity and stable support for shareholder value in the event of unanticipated crises or events. Specifically, we examine whether insiders at lobbying firms are more likely to step in and buy shares during the proposed new blackout window following significant declines in their firms' stock prices, as alleged by the lobbying firms. We find supporting evidence for this claim using insider trading data prior to 2018. This finding corroborates our event study results.

Our study contributes to the literature on insider trading regulation. Whether and how insider trading should be regulated is still hotly debated (e.g., Manne, 1966; Easterbrook, 1981; Carlton and Fischel, 1983; Glosten and Milgrom, 1985; Macey, 1991; Fishman and Hagerty, 1992; Meulbroek, 1992; Roulstone, 2003; Bushman et al., 2005). Prior research (e.g., Jaffe, 1974; Seyhun, 1992; Garfinkel, 1997) examines the impact of insider trading regulation on insider trading behavior, such as trading volume and profitability, but few studies are able to assess the net shareholder value impact of insider trading regulation due to the difficulty of capturing all the costs and benefits of insider trading. In addition, most prior studies (e.g., Bhattacharya

[2] One could argue that the reversal of the regulation per se may signal the weakening of Hong Kong's investor protection. However, this effect, if it existed, would bias against finding a positive stock market reaction to the reversal.

and Daouk, 2002; Bushman et al., 2005; Fernandes and Ferreira, 2009) focus on the consequences of regulating insider trading.[3] Much less attention is devoted to the subsequent important question in the ongoing debate: whether further tightening of insider trading regulation is warranted. Furthermore, there is little research examining the consequences of adopting specific insider trading regulatory provisions, such as blackout windows.[4] Our study advances the insider trading regulation literature by using a relatively *clean* event study to identify the *net* impact on shareholder value of *further tightening* blackout windows for insider trading.

We acknowledge that the findings of our study may not be directly generalizable to other financial markets with different institutional environments. Nevertheless, the cost–benefit trade-offs facing Hong Kong regulators should also be instructive for regulators in other financial markets. More importantly, our results suggest caution in imposing tougher one-size-fits-all insider trading regulation.

The rest of the paper is organized as follows. Section 2 develops our hypothesis. Section 3 reviews Hong Kong's institutional environment for insider trading and its data sources. Section 4 reports the event study results for the entire Hong Kong stock market, while Section 5 shows the event study results for lobbying firms versus non-lobbying firms. Section 6 tests several costs and benefits alleged by the critics and proponents of the new blackout regulation. Section 7 concludes.

2. Hypothesis development

The existing insider trading literature around the world provides consistent evidence that trades made by corporate insiders contain private information and are profitable (see, e.g., Seyhun, 1998; Piotroski and Roulstone, 2005 for the U.S., Baesel and Stein, 1979 for Canada, Cheuk et al., 2006 for Hong Kong, and Fidrmuc et al., 2006 for the U.K.). There is an ongoing debate over whether such informed insider trading benefits the shareholders of the affected firms and the overall stock market.[5] Proponents of insider trading argue that it accelerates the incorporation of private information into stock prices and thus should benefit all shareholders and the broader financial market (Manne, 1966; Meulbroek, 1992; Chau and Vayanos, 2008). Insider trading could also be an effective way to compensate managers for innovations (Manne, 1966; Carlton and Fischel, 1983; Roulstone, 2003).

However, critics of insider trading assert that when unrestricted it harms uninformed investors who trade against informed insiders, and thus reduces investor confidence in financial markets and increases the affected firms' cost of equity (Brudney, 1979; Easterbrook, 1981; Macey, 1991; Glosten and Milgrom, 1985; Bhattacharya and Daouk, 2002). In addition, unrestricted insider trading may create perverse managerial incentives such as the manipulation of important corporate decisions (see, e.g., Easterbrook, 1981; Bainbridge, 2000; Cheng and Lo, 2006) and reduce private information acquisition by outsiders (Fishman and Hagerty, 1992; Bushman et al., 2005; Fernandes and Ferreira, 2009).

Many of the alleged costs and benefits are firm-specific. Hence, shareholders of an individual firm should have a strong incentive to voluntarily restrict insider trading to the extent that it harms them (Bettis et al., 2000; Roulstone, 2003). Therefore, it is not immediately apparent whether the government needs to regulate insider trading. However, insights from the existing literature on regulation suggest several possible justifications for doing so.[6]

[3] The most common insider trading regulatory restriction is the "disclose or abstain" rule, which bans corporate insiders from trading when in possession of price-sensitive non-public information (Stamp and Jaggers, 2005). Many regulators impose easy-to-enforce insider trading blackout windows regardless of whether the insiders possess price-sensitive information or not (e.g., Hong Kong, mainland China, Australia, Netherlands).

[4] One exception is Kabir and Vermaelen (1996), who examine the effect on stock liquidity of introducing a two-month insider trading blackout period prior to the annual earnings announcement on the Amsterdam Stock Exchange. However, the blackout regulation change examined by Kabir and Vermaelen (1996) occurred before, rather than after, the first enforcement of insider trading laws; therefore, their setting differs from ours. Another exception is Zhu and Wang (2015), who find that an insider trading ban in China altered insider trading patterns, constrained profits during the ban window, but did not fully eliminate profitable insider trading outside the restricted period.

[5] See Bainbridge (2000) for a comprehensive review of the costs and benefits of insider trading.

[6] See Leuz and Wysocki (2016) for similar arguments justifying disclosure regulation.

First, a mandatory insider trading regulation may serve as a low-cost commitment device (Mahoney, 1995; Rock, 2002). As noted above, individual firms should have an incentive to voluntarily adopt insider trading restrictions to the extent that insider trading hurts shareholder value. However, such voluntary commitment devices may be quite expensive or even impossible in many financial markets with weak investor protection (Doidge et al., 2007). Therefore, the observed insider trading restrictions under a voluntary regime may not be socially optimal. Insider trading regulation can mitigate this problem and serve as a lower-cost commitment device, given that the government has access to criminal penalties and other remedies unavailable in private contracts.

Second, unrestricted insider trading may also create deadweight losses (Shleifer and Wolfenzon, 2002). As noted above, unrestricted insider trading may induce corporate insiders to manipulate their firms' strategic decisions. Such manipulation activities are costly and represent a deadweight loss to society. Due to the imperfections of voluntary commitment devices, there will always be some residual level of manipulative activity that individual firms' shareholders cannot fully eliminate. As a more credible and lower-cost commitment device, regulation may help reduce such deadweight losses.

Third, insider trading may create externalities. For example, unregulated insider trading may undermine investor confidence in the integrity of the stock market (Easterbrook, 1981; Macey, 1991) and, in turn, reduce many investors' willingness to participate. Such externalities are unlikely to be internalized when individual firms set their optimal insider trading policies in a voluntary regime.

While the above discussions highlight the potential benefits of insider trading regulation, there are also costs associated with it. First, mandatory regulations are costly to design, implement and enforce. Second, mandatory regulations tend to be one-size-fits-all, whereas the socially optimal insider trading restriction may differ across firms. Third, the existence of insider trading regulation may encourage corporate lobbying and the capture of the regulators by certain special interest groups.

In summary, there are many potential costs and benefits associated with insider trading regulation. Ex ante, these costs and benefits are hard to quantify. In addition, major financial markets, including Hong Kong, have already imposed insider trading regulations; therefore, it is even less clear whether further tightening would benefit investors. For these reasons, it is an empirical question whether tightening insider trading regulations would benefit the overall Hong Kong stock market. We state our hypothesis in the null form below:

Hypothesis. Tightening the insider trading blackout regulation in Hong Kong is not associated with a change in shareholder value for the overall Hong Kong stock market.

3. Institutional background

3.1. Insider trading laws in Hong Kong

Hong Kong enacted its first insider trading law—the Securities (Insider Dealing) Ordinance—in 1991, and the first insider trading prosecution occurred in 1994 (Bhattacharya and Daouk, 2002). Under the Securities (Insider Dealing) Ordinance, insider trading was not a crime and was dealt with by the Insider Dealing Tribunal, which had no formal sanctioning power. Nevertheless, insider trades had to be filed with the HKEX within five business days. Effective from 1 April 2003, the Securities and Futures Ordinance (SFO) superseded the Securities (Insider Dealing) Ordinance. The 2003 SFO made three key changes to insider trading regulation. First, insider trades must be filed with the HKEX within three business days. All the reports based on the filings are published on the HKEX website.[7] Second, the Insider Dealing Tribunal is replaced by the Market Misconduct Tribunal, which has the power to address various forms of market misconduct, including insider dealing. Third, the SFO expands the criminal regime for market misconduct to include insider dealing, with a maximum jail sentence of 10 years and fines of up to HK$10 million. Thus, since the SFO became law in 2003,

[7] In particular, the "Directors'/Chief Executives' Notifications Report – A (Actual Transactions)" provides the company name, the names of insiders (Directors/Chief Executives), the type of securities, the type of transactions (purchase or sale), the securities holding before and after the transactions, the transaction prices and the dates of the transactions and notifications.

there have been dual civil and criminal regimes for insider dealing, market manipulation and the spread of false or misleading information about securities in Hong Kong. February 2008 saw the first criminal insider trading prosecution since insider dealing was made a criminal offense in 2003. Overall, like many developed financial markets, Hong Kong has already had a relatively robust insider trading regulatory regime in place since 2003.

3.2. Evolution of the blackout regulation in Hong Kong

Hong Kong introduced its first blackout window regulation in November 1993. Until 2009, under Rule A.3 of the Model Code in Appendix 10 of the Listing Rules, the blackout window was one month prior to the earnings announcement date. In early 2008, the HKEX proposed extending the blackout window from the earnings announcement date to the fiscal period end date in a combined consultation paper.[8] The blackout provision was approved on 28 November 2008 as part of the SFC's combined consultation conclusions, and was scheduled to take effect on 1 January 2009.[9] However, on 29 December 2008, just three days prior to the effective date, 236 listed companies (the lobbying firms) collectively published an open letter (see Appendix A) to the HKEX opposing the implementation of the tougher blackout regulation. The lobbying firms claimed that they had missed the revised blackout proposal because the initial proposal was buried among the 40 proposals.[10] On 30 December 2008, the HKEX announced that its listing committee would meet to consider the open letter.[11] On 31 December 2008, the day before the scheduled effective date of the new blackout regulation, the HKEX announced that it would delay the effective date to 1 April 2009 (Event 1). Two weeks later, on 14 January 2009, the HKEX announced that it would review the new blackout rule only if the listed firms agreed to provide quarterly financial reports (Event 2). About one month later, on 12 February 2009, the HKEX announced that it would modify the blackout period to two months prior to the annual earnings announcement and one month prior to the interim earnings announcement (Event 3). The SFC's final approval of the modified blackout regulation was announced on 24 February 2009 (Event 4). The approved modified blackout rule is very similar to the blackout regulation of the old regime and therefore represents a significant setback for the Hong Kong regulators. Panel A of Table 1 shows the key events associated with the reversal of the proposed new blackout regulation. Although we identify four relevant events associated with the reversal of the new regulation, Event 1 represents the most significant and cleanest because it was the first time that the financial market learned of the probability of the regulation's reversal.

The blackout regulation saga generated heated discussions about how to regulate insider trading in markets like Hong Kong, a relatively well-governed, free financial market with a robust insider trading regulatory framework. While acknowledging that many comparable markets rely on a mixture of insider dealing laws and market forces that result in companies setting tight internal controls, the SFC argued that neither of these factors may work sufficiently well in Hong Kong due to the dominance of family firms, calling for the need to set a formal benchmark by tightening the insider trading blackout regulation (HKEX, 2008a; Williams, 2008). The HKEX also believes that directors should not actively and frequently trade in their company's shares and instead should be long-term investors in the company (HKEX 2008a). A vocal Hong Kong shareholder activist, David Webb, and many retail investors strongly supported the tougher blackout regulation (Webb, 2008; Association for Sustainable and Responsible Investment in Asia, 2009).

[8] Published on 11 January 2008, this combined consultation paper seeks the market's views on 18 substantive policy issues regarding corporate governance, initial listing criteria and some amendments to improve the clarity, certainty and efficacy of the Listing Rules.

[9] The Consultation Conclusions pertain to 15 of the 18 issues with the final versions of the Rule changes. Included in the 15 issues was the proposal to extend the blackout period in the Model Code (Appendix 10 to the Main Board Rules and its GEM equivalent).

[10] The combined consultation paper contains more than 40 proposals, which are grouped into 18 substantive policy issues and some minor Rule changes.

[11] On 30 December 2008, concerns raised about the new Rule were the subject of a Special Meeting of the Legislative Council Financial Affairs Panel. The Financial Affairs Panel approved, by a majority, a non-binding motion requesting "the authorities concerned not to effect the amended Rules as scheduled on 1 January 2009 and launch a six-month consultation on the issue afresh, so as to address the market concerns properly".

Table 1
Overall market reaction to the reversal of the new blackout rule.

Panel A. Description of the relevant events

Event	Event return measurement window	Announcement date of the event	Description of the event
Event 1	29/12/2008–05/01/2009	29/12/2008 (Monday)	The open letter signed by the directors of 236 listed firms was published.
		30/12/2008 (Tuesday)	HKEX announced that its listing committee will meet to consider the open letter.
		31/12/2008 (Wednesday)	HKEX announced that it will delay the implementation date of the new blackout rule to April 1, 2009.
Event 2	12/01/2009 – 16/01/2009	14/01/2009 (Wednesday)	HKEX announced that it will review the new blackout rule only if the listed firms agree to provide quarterly financial reports.
Event 3	10/02/2009 – 16/02/2009	12/02/2009 (Thursday)	HKEX announced that it will modify the blackout period to be two months prior to the annual results announcement and one month prior to the interim results announcement.
Event 4	20/02/2009 – 26/02/2009	24/02/2009 (Tuesday)	HKEX's modified blackout rule received approval from the SFC.

Panel B. Regression results

Independent variable		The dependent variable is R_t^{HKALL}	
		(1)	(2)
Intercept		−0.0011	−0.0011
		(0.226)	(0.239)
		[0.048]	[0.796]
$EVENT_t^1$	β_1	0.0546**	
		(0.020)	
		[0.046]	
$EVENT_t^2$	β_2	−0.0041	
		(0.871)	
		[0.328]	
$EVENT_t^3$	β_3	0.0342**	
		(0.021)	
		[0.048]	
$EVENT_t^4$	β_4	0.0011	
		(0.956)	
		[0.612]	
Key_EVENT_t			0.0869*
			(0.076)
			[0.114]
$R_{t-1}^{STOXX-exHK}$		0.0474	0.0541
		(0.432)	(0.358)
		[0.136]	[0.000]

(*continued on next page*)

Table 1 (*continued*)

Panel B. Regression results

Independent variable	The dependent variable is R_t^{HKALL}	
	(1)	(2)
$R_t^{STOXX-exHK}$	0.4449***	0.4520***
	(0.000)	(0.000)
	[0.480]	[0.038]
$R_{t+1}^{STOXX-exHK}$	0.2157***	0.2199***
	(0.000)	(0.000)
	[0.982]	[1.000]
$R_{t-1}^{A-share}$	−0.0134	−0.0162
	(0.693)	(0.627)
	[0.048]	[0.002]
$R_t^{A-share}$	0.3408***	0.3409***
	(0.000)	(0.000)
	[0.052]	[0.042]
$R_{t+1}^{A-share}$	−0.0119	−0.0094
	(0.746)	(0.794)
	[0.986]	[1.000]
Adjusted R^2	0.5582	0.5596
N	283	283

Note: The sample period is from January 1, 2008, to February 27, 2009. $EVENT_t^i$ ($i = 1, 2, 3, 4$) is a dummy variable that equals one if the trading day falls within the event window associated with event i, as shown in panel A, scaled by the number of trading days included in the corresponding event window, and zero otherwise. Key_EVENT_t is a dummy variable that equals one if the trading day falls within the event window associated with any of the four events shown in Panel A, scaled by the total number of trading days included in the event windows of the four events, and zero otherwise. R_t^{HKALL} is the equal-weighted returns of all firms listed on HKEX. $R_t^{STOXX-exHK}$ is the value-weighted return of the firms included in the DJ STOXX-1800 index, excluding the firms listed on the HKEX., $R_t^{A-share}$ is the value-weighted return of A-share listed firms on day t. The *p*-values in parentheses are based on White (1980) adjusted standard errors. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. The *p*-values in the brackets are bootstrapped following Zhang (2007).

Those against the blackout proposal pointed out in the open letter that existing regulatory rules had already prohibited directors from dealing when in possession of price-sensitive information (HKEX 2008b). In addition, the lobbying firms explicitly cited several potential negative consequences of the new blackout regulation, including deterring existing and future listings, jeopardizing market liquidity and discouraging individuals from serving as directors of listed companies. As quoted by a Hong Kong newspaper, Lo Ka-shui, a former chairman of the Hong Kong stock exchange's GEM Listing Committee, argued that the inability of major shareholders to trade a company's securities for an extended period may make it difficult for major shareholders to fight a hostile takeover bid (Scent, 2008). Dr. Louis Cheng at The Hong Kong Polytechnic University also noted that longer blackout periods, per se, may not be an unmitigated positive for individual investors because trading by senior executives, directors and major shareholders in their companies' shares is a signal that investors like to watch (Leng, 2009). Citing the lack of evidence of private-information-motivated insider trades prior to earnings announcements in a commissioned report by Chan (2009), Mr. Phillip Baldwin, the Chief Executive of the Hong Kong Institute of Chartered Secretaries, argued against the new blackout regulation (Baldwin, 2009).[12] However, shareholder activists dismissed the significance of Chan (2009) due to various alleged methodological shortcomings of the study and the claim that Chan (2009) was commissioned by the Hong Kong Institute of Chartered Secretaries and therefore not independent (Webb, 2009).

4. Overall market reaction

We use an event study methodology (Campbell et al., 1997) to assess the impact of the repeal of the new blackout regulation on shareholder value across the entire Hong Kong stock market portfolio. We focus on the events associated with the reversal because they provide a clean setting: the only issue of interest is the new blackout regulation. Moreover, the reversal came as a real surprise to the market because the new blackout regulation had already been passed and was to take effect within a few days. To measure the stock market reactions to the four events associated with the regulation's reversal, we estimate the following regression model:

$$R_t^{HKALL} = \alpha_0 + \sum_{i=1}^{4}(\beta_i \frac{EVENT_t^i}{M_i}) + \sum_{j=-1}^{+1}(\gamma_j R_{t+j}^{STOXX-exHK}) + \sum_{j=-1}^{+1}(\delta_j R_{t+j}^{A-share}) + \varepsilon_t \quad (1)$$

See Table 1 for the definitions of all regression variables. We estimate regression model (1) using the sample period from 1 January 2008 to 27 February 2009, right after the end of the last event window. R_t^{HKALL} is the dividend-inclusive, equal-weighted return for the entire Hong Kong stock market portfolio on day t. We obtain daily stock return data from the *DataStream* database. The sample includes 1070 firms listed on the main board of the HKEX, covered by WorldScope and *DataStream*, and with return data available during the sample period. To control for other confounding effects, we include $R_{t+j}^{STOXX-exHK}$ to control for the effects of global news (Zhang, 2007; Armstrong et al., 2010) and $R_{t+j}^{A-share}$ to control for the effects of mainland China news. To deal with trading non-synchronicity across financial markets, we include $R^{STOXX-exHK}$ and $R^{A-share}$ on days t–1, t and $t+1$ in model (1).[13]

$EVENT_t^i$ is a dummy variable that equals one if day t is included in the return measurement window of event i ($i = 1$ to 4) and zero otherwise. For each event announcement, we measure the event-period stock return

[12] Using directors' insider dealings over 2003–2008, Chan (2009) documents the following key findings. First, only 30% of directors' dealings are before results announcements. Second, approximately 75% of directors' dealings are purchases, and only 25% are sales. Third, there is no evidence that directors' purchases are information-motivated, although directors' sales are associated with negative future abnormal returns. However, the informativeness of directors' sales is limited to small firms only.

[13] Because the firms included in DJ STOXX 1800 ex HK are located in different time zones, $R_t^{STOXX-exHK}$ on day t is constructed as follows: we use day t returns for the firms located in the Asia-Pacific region and day t–1 returns for the firms located outside the Asia-Pacific region (i.e., America and Europe). Nevertheless, our results are qualitatively the same when we measure $R_t^{STOXX-exHK}$ as the value-weighted average returns of all firms on calendar day t.

using the cumulative return over the trading days (–2, +2) centered on the event announcement date (i.e., day 0).[14] M_i is the number of trading days included in the event window of event i. We scale $EVENT_t^i$ by M_i so that β_i can be interpreted as the total market reaction to event i for the entire Hong Kong stock market portfolio.

The sample period used to test our event study hypothesis coincides with the 2008 financial crisis. Hence, there is a concern that we could get significant results even if the repeal of the regulation did not affect shareholder value. To account for potential return volatility during our sample period, we use both regular p-values, which already control for cross-firm correlations, and bootstrapped p-values for our event study results.[15]

Panel B of Table 1 shows the results of regression model (1). As shown in column (1) of Table 1, the coefficients on $R_t^{STOXX-exHK}$, $R_{t+1}^{STOXX-exHK}$ and $R_t^{A-share}$ are significantly positive, suggesting that both global macro-factors and mainland China-specific factors influence Hong Kong's market returns. The coefficient on $EVENT_t^1$ (0.0546, regular $p = 0.020$ and bootstrapped $p = 0.046$), the date of publication of the lobbying letter, is significantly positive. The coefficient on $EVENT_t^3$ (0.0342, regular $p = 0.021$ and bootstrapped $p = 0.048$), the date when the HKEX announced the modification to the blackout regulation, is also significantly positive. The coefficients on $EVENT_t^2$ and $EVENT_t^4$ are small and nonsignificant. As expected, the stock market reaction to Event 1 is the strongest. The coefficient on $EVENT_t^1$ is almost 60% bigger than the coefficient on $EVENT_t^3$.

To capture the overall effects of the four events, we replace $EVENT_t^i$ ($i = 1$ to 4) in column (1) by one variable Key_EVENT_t in column (2) of Table 1. Key_EVENT_t is a dummy variable that equals one if day t is included in the return measurement window of any of the four events, and zero otherwise. We also scale Key_EVENT_t by the total number of trading days included in the four event windows so that the coefficient can be interpreted as the aggregate market reaction. The coefficient on the indicator for the four combined events (Key_EVENT_t) is 0.0869 and marginally significant using the regular p-value (0.076) but nonsignificant using the bootstrapped p-value (0.114). Overall, the results in Table 1 provide weak evidence that repealing the new blackout regulation increases shareholder value across the entire Hong Kong stock market portfolio.

5. Market reaction for the lobbying firms vs. matched non-lobbying firms

We next examine whether the documented event study results are stronger for firms more affected by the new regulation. To the extent that the new blackout regulation hurts shareholder value, we expect the event study results to be more positive for firms more affected by it.

Based on the conjecture that firms more affected by the new regulation are more likely to lobby against its implementation (see below for evidence), we use lobbying firms as a proxy for more affected firms. One benefit of this subsample analysis is that we can better control for unobservable confounding macro-effects specific to the Hong Kong market by using a matched sample of non-lobbying firms as a control group.

5.1. Who lobbied against the implementation of the new regulation?

We first test our conjecture that the firms more affected by the new regulation would be more likely to lobby against the implementation of the new blackout regulation using the following logit model:

[14] The only exception is that we start the event return window on the event announcement date, 29 December 2008, for Event 1. The reason is that the trading day immediately preceding 29 December 2008 is 24 December 2008 because of the Christmas holiday break. Including returns prior to 29 December 2008 seems less appropriate because the news of Event 1 was announced unexpectedly on that day, and it is therefore unlikely that the public received news of the lobbying five days before the release of the open letter. Nevertheless, our inferences are similar if we extend the event window of Event 1 to 24 December 2008.

[15] We compute the bootstrapped p-value following Lo (2003) and Zhang (2007). Specifically, we randomly select 20 trading days from the sample period after removing the observations in the event windows. We define the first five trading days as the event window for pseudo-Event 1, the second five trading days as the event window for pseudo-Event 2, and so on. We estimate a time series regression of aggregate returns on the indicators for the four pseudo-event periods and $R^{STOXX-exHK}$ and $R^{A-share}$ on day t–1, t and t+1 to obtain the coefficient estimates. We repeat the random draw for 500 replicates. The one-tail bootstrapped p-values are calculated as the proportions of the 500 replicates that generate coefficients greater than the OLS coefficients in Panel B of Table 1.

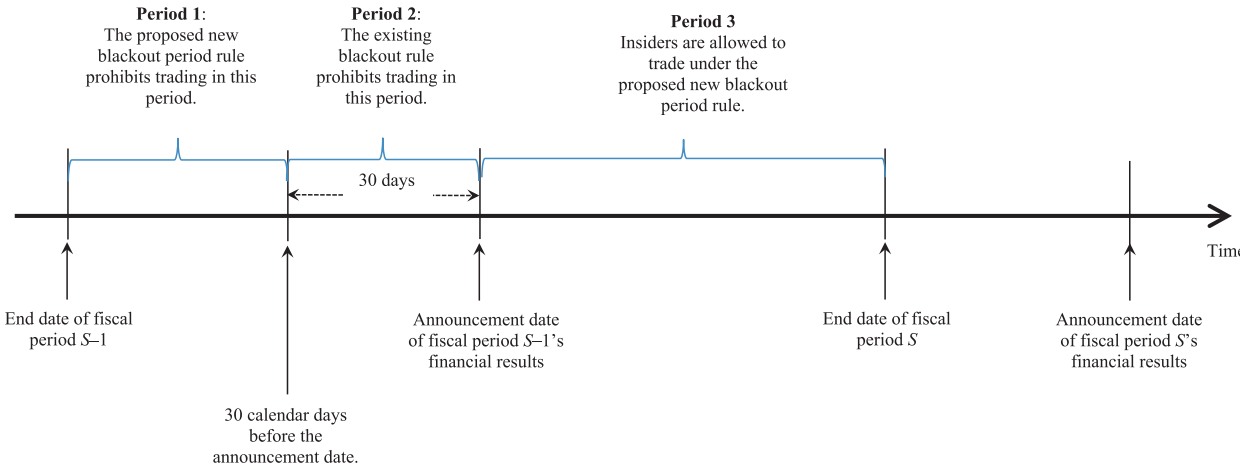


Fig. 1. The period covered by the proposed new blackout regulation. Each half-year fiscal period (fiscal period *S*) is partitioned into three periods as follows: Period 1: from the 1st day after the end of (half-year) fiscal period *S*–1 to the 31st day before the announcement date of the earnings for fiscal period *S*–1. If the earnings announcement date for fiscal period *S*–1 is within 31 days after the end date of fiscal period *S*–1, then Period 1 is empty, but this case is rare in our sample period. Period 2: from the earlier of (1) the 1st day after the end date of fiscal period *S*–1 and (2) the 30th day before the announcement date of the earnings of fiscal period *S*–1 to the announcement date of the earnings for period *S*–1. Period 3: from the 1st day after the announcement date for earnings for fiscal period *S*–1 to the end date of fiscal period *S*.

$$\Pr(LOBBY) = \beta_1 HEAVY_TRADING + \beta_2 lnRET + \beta_3 lnQ + \beta_4 lnFOLLOW + \beta_5 SIZE + \beta_6 AGE + \text{industry fixed effects} + \varepsilon \quad (2)$$

See Appendix B for the variable definitions. One key variable of interest is *HEAVY_TRADING*, a proxy for firms more affected by the new regulation. We predict the coefficient on *HEAVY_TRADING* to be positive. Conceptually, *HEAVY_TRADING* captures the intensity of the insider trades that occurred in the proposed new blackout period from the fiscal period end to one month prior to the earnings announcement (Period 1 in Fig. 1). As detailed in Appendix B, the definition of *HEAVY_TRADING* considers two conditions: (i) whether a firm's insiders have sufficient time to trade in Period 1, and (ii) whether the firm's insiders as a whole trade a significant dollar amount of their own firm's stock in Period 1. We consider condition (i) because insiders would have no time to trade if Period 1 were too short. We consider condition (ii) because the insiders for some firms may not trade substantially even if they have a sufficiently long Period 1. For each firm, we define *HEAVY_TRADING* for half-year fiscal periods that start after 30 June 2003 and end on or before 30 June 2008 so that the variable is not contaminated by insider trading behavior immediately before the release of the open letter.[16]

As controls, we consider two classes of lobbying determinants based on prior research (Lo, 2003; Hochberg et al., 2009). The first class of determinants is common proxies for information asymmetry (*lnFOLLOW*) and undervaluation (*lnRET* and *lnQ*). The second class of determinants captures the lobbying firms' experience (*AGE*) and resources (*SIZE*). Conditional on all firms receiving the same gross benefits from repealing the new regulation, firms with more resources should be more likely to lobby.

Panel A of Table 2 shows the sample selection criteria. We exclude firms delisted before 1 January 2009 because they are unlikely to engage in lobbying. After deleting firms with missing data, we obtain a final sample of 861 unique firms. Among the sample firms, 21.6% participated in the lobbying, representing 22.6% of the total market capitalization of all firms (untabulated). Panel B shows the descriptive statistics for the regression variables of model (2).

[16] It is likely that *HEAVY_TRADING* is the outcome of various institutional factors, including the quality of a firm's governance system. We take a firm's *HEAVY_TRADING* measured before the release of the lobbying open letter as given and ask whether firms with *HEAVY_TRADING* are more likely to lobby against the implementation of the new regulation in December 2008. It is beyond the scope of this study to examine the institutional factors driving insider trading behavior.

Table 2
Determinants of corporate lobbying against the new blackout rule.

Panel A. Sample selection

Sample election procedures Remained	Number of firms
All firms listed on the main board of HKEX and covered by the *WorldScope* and the *DataStream*	1070
Delete firms listed after 1/1/2007 or delisted before 1/1/2009	934
Delete firms disclosing quarterly reports	907
Delete firms with missing HEAVY_TRADING	877
Delete firms with missing *lnQ*, *SIZE*, *lnRET*	861
Final sample	861
Lobbying firms	186
Non-lobbying firms	675

Panel B. Summary statistics and univariate comparison

				Percentile		
Variable	N	mean	stdev	25%	50%	75%
LOBBY	861	0.216	0.412	0	0	0
HEAVY_TRADING	861	0.300	0.458	0	0	1
SIZE	861	7.627	1.807	6.289	7.342	8.651
lnQ	861	0.337	0.653	−0.094	0.226	0.637
lnRET	861	−1.209	0.800	−1.602	−1.144	−0.742
AGE	861	4.768	0.817	4.290	4.977	5.361
lnFOLLOW	861	0.670	0.996	0.000	0.000	1.099

	The lobbying firms		The non-lobbying firms		*t*-test for the difference		Wilcoxon Rank-sum test for the difference	
Variable	N	mean	N	mean	t-stat	*p*-value	Z-stat	*p*-value
HEAVY_TRADING	186	0.376	675	0.279	2.47	0.014	2.58	0.010
SIZE	186	8.187	675	7.473	4.76	0.000	4.95	0.000
lnQ	186	0.185	675	0.379	−3.83	0.000	−3.82	0.000
lnRET	186	−1.362	675	−1.166	−2.06	0.041	−1.70	0.090
AGE	186	5.170	675	4.658	8.47	0.000	8.61	0.000
lnFOLLOW	186	0.658	675	0.673	−0.18	0.855	0.08	0.938

Panel C. Logit regression results of lobbying against the new blackout rule

Independent Variable	The dependent variable is *LOBBY*
HEAVY_TRADING	0.5929***
	(0.004)
lnQ	−0.5871***
	(0.001)
lnRET	−0.3364***
	(0.009)
lnFOLLOW	−0.5162***
	(0.002)
SIZE	0.5275***
	(0.000)
AGE	0.6677***
	(0.000)
Industry fixed effects	Yes
Pseudo R^2	0.1737
N	861

Note: See Appendix B for variable definitions. The *p*-values based on White (1980) adjusted standard errors are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Panel C shows the logit regression results of model (2). The coefficient on *HEAVY_TRADING* is significantly positive (0.5929, $p = 0.004$), consistent with our conjecture that firms more affected by the new regulation are more likely to lobby against its implementation. The coefficients on all the control variables are significant. The significant coefficients on *lnQ* and *lnRET* suggest that firms more severely undervalued are more likely to lobby against the new blackout regulation. The coefficient on *LnFOLLOW* is negative, suggesting that firms with greater information asymmetry are more likely to lobby. We also find that firms with more experience and resources have a higher propensity to lobby.

It is important to note that model (2) merely shows that the lobbying firms are the firms more affected by the new blackout regulation, but the results from model (2) cannot be used to conclude whether the net shareholder value impact of the new blackout regulation is more positive (or more negative) for the lobbying firms relative to the non-lobbying firms. In the next section, we directly examine whether further tightening insider trading regulation affects shareholder value to a greater extent for the lobbying firms (more affected firms).

5.2. The market reaction for the lobbying firms vs. matched non-lobbying firms

To mitigate concerns about potential confounding macro-news unique to Hong Kong, we follow Hochberg et al. (2009) by matching the lobbying firms to a portfolio of non-lobbying firms with similar propensity scores that incorporate firm size, book-to-market ratio, return momentum and industry. Appendix C details the matching process and presents the matching diagnostics. We then estimate the following time-series portfolio regression to examine the difference in the market reaction between the lobbying firms and the matched non-lobbying firms:

$$\frac{1}{N_t}\sum_{i=1}^{N_t}\left(R_{i,t} - R_{i,t}^{B}\right) = \alpha + Key_Event_t + \gamma_{MKTRF}R_t^{MKTRF} + \gamma_{SMB}R_t^{SMB} + \gamma_{HML}R_t^{HML} + \gamma_{UMD}R_t^{UMD} + \varepsilon_t \quad (3)$$

where $R_{i,t}$ is the dividend-inclusive raw return of lobbying firm i's stock on day t. N_t is the number of lobbying firms for which returns are available for day t. $R_{i,t}^{B}$ is the average dividend-inclusive return on day t for the portfolio of non-lobbying firms matched to lobbying firm i using the propensity score matching method of Hochberg et al. (2009) (see Appendix C for details). We adopt a portfolio regression approach because the individual firm stock returns on the same event dates are not independent due to event clustering (Sefcik and Thompson, 1986). We also follow Hochberg et al. (2009) by using the four-factor model to examine the market reaction. R_t^{MKTRF}, R_t^{SMB}, R_t^{HML} and R_t^{UMD} are the returns of the Hong Kong-listed firm portfolios mimicking the market, size, book-to-market and momentum factors, respectively, on day t, constructed using the method of Fama and French (2012).[17] For brevity, we combine the effects of the four events into one coefficient similar to the model in column (2) of Panel B in Table 1.[18] The coefficient on *Key_EVENT* captures the combined market reaction to the four events for the lobbying firms relative to the matched non-lobbying firms.

Panel A of Table 3 shows the sample selection for model (3). Because we require additional variables to compute the propensity scores for constructing the matching portfolio, the sample size is reduced from 1070 firms in Table 1 to 860 firms. Panel B shows the results of regression (3). The coefficient on *Key_EVENT* is significantly positive (0.0528, regular $p = 0.011$, bootstrapped $p = 0.006$). Overall, the evidence in Table 3 provides further support for Table 1, indicating that the new blackout regulation hurts the shareholder value of the lobbying firms (the more affected firms) to a greater extent.

[17] Our results are robust to replacing the four-factor returns with $R^{STOXX-exHK}$ and $R^{A-share}$ on day t–1, t and t+1 (untabulated).

[18] To further control for potential individual firm-specific confounding news during the event windows (e.g., earnings announcements, profit warnings, equity offering announcements, M&A events and regulatory investigations), we also recalculate all the event study results reported in Table 3 after excluding the lobbying firms' event days with such confounding news. Our inferences are qualitatively the same (untabulated).

Table 3
The difference in the stock market reactions to the reversal of the new blackout rule between the lobbying and non-lobbying firms.

Panel A. Sample selection procedures

Sample selection procedures	Number of firms Remained
All firms listed on the main board of HKEX and covered by the *WorldScope* and the *DataStream*	1070
Delete firms listed after 1/1/2007 or delisted before 1/1/2009	934
Delete firms disclosing quarterly earnings	907
Delete firms with missing *lnQ*, *SIZE*, and *lnRET*12 used in the propensity score matching procedures (see Appendix C)	860
Final sample	860
Lobbying firms	187
Non-lobbying firms	673

Panel B. Regression results of the difference in market reactions to the reversal of the new blackout rule between the lobbying firms and matched non-lobbying firms

Independent Variable	The dependent variable is the mean *excess* return of all lobbying firms: $(\frac{1}{N_t}\sum_i(R_{i,t} - R^B_{i,t}))$
Intercept	−0.0005**
	(0.049)
	[0.030]
Key_EVENT_t	0.0528**
	(0.011)
	[0.006]
R_t^{MKTRF}	−0.0425***
	(0.004)
	[0.000]
R_t^{SMB}	0.0015
	(0.962)
	[1.000]
R_t^{HML}	0.0001
	(0.998)
	[0.086]
R_t^{UMD}	0.0003
	(0.992)
	[0.000]
Adjusted. R^2	0.0856
N	283
Number of lobbying firms	187

Note: Panel B reports the results of time-series regressions using the mean excess return of all lobbying firms as the dependent variable. The sample period is from January 1, 2008, to February 27, 2009. $R_{i,t}$ is the dividend-inclusive return of lobbying firm i on day t. $R^B_{i,t}$ is the mean dividend-inclusive returns on day t of the non-lobbying firms matched to lobbying firm i. See Appendix C for the detailed description of the matching procedure. N_t is the number of lobbying firms included in the corresponding portfolio. Key_EVENT_t is a dummy variable that equals one if the trading day falls within the event window associated with any of the four events shown in Panel A, scaled by the total number of trading days included in the event windows of the four events, and zero otherwise. R_t^{MKTRF}, R_t^{SMB}, R_t^{HML}, and R_t^{UMD} are the returns of the portfolios, on day t, mimicking the market, size, book-to-market, and momentum factors for firms listed on HKEX, respectively. We follow Fama and French (2012) to construct R_t^{MKTRF}, R_t^{SMB}, R_t^{HML}, and R_t^{UMD}. The *p*-values in the parentheses are based on White (1980) adjusted standard errors. ***, **, and * indicate significant at the 1%, 5%, and 10% levels. The *p*-values in the brackets are one-tailed bootstrapped *p-values* following Zhang (2007).

6. Specific alleged costs and benefits of insider trading in the debate

The event study results suggest that Hong Kong's 2008 new blackout regulation harms shareholders in the Hong Kong stock market. In this section, we directly test some of the alleged costs and benefits of insider trading during Hong Kong's debate on the new blackout regulation. Our first test examines whether insiders at lobbying firms (more affected firms) trade on forthcoming earnings news more aggressively than insiders at non-lobbying firms in the proposed new blackout window, as alleged by supporters of the new blackout regulation. Our second test examines whether lobbying firms are more likely to delay earnings announcements than non-lobbying firms, as alleged by supporters of the new blackout regulation. Our third test examines whether insiders at lobbying firms buy more shares when their firms' stock prices decline significantly, as critics of the new blackout regulation allege. We discuss each of these three additional analyses in the following subsections.

6.1. Aggressiveness of insider trading in the proposed new blackout window

We use the following OLS regression model to analyze whether the insiders of lobbying firms trade on the forthcoming earnings news in the proposed new blackout period (i.e., Period 1 in Fig. 1) more aggressively than the insiders of non-lobbying firms:

$$NETBUY = \beta_0 + \beta_1 LOBBY + \beta_2 NEWS_{S-1} + \beta_3 LOBBY \times NEWS_{S-1} + \beta_4 NEWS_S + \beta_5 LOBBY \times NEWS_S + CONTROL + \text{industry and fiscal period fixed effects} + \varepsilon \quad (4)$$

See Appendix B for variable definitions and Fig. 1 for the timeline. The dependent variable *NETBUY* is defined as the natural logarithm of one plus the dollar value of total insider purchases minus the natural logarithm of one plus the dollar value of total insider sales (in HK$1000) in Period 1.[19] The variable of interest is $LOBBY \times NEWS_{S-1}$. If insiders of lobbying firms (firms more affected by the new regulation) are more opportunistic, we should expect the coefficient on $LOBBY \times NEWS_{S-1}$ to be significantly positive. We also include $NEWS_S$ because prior research (e.g., Ke et al., 2003) shows that corporate insiders often trade on future long-term earnings news. However, the coefficient on $NEWS_S$ is not relevant to the new blackout regulation debate because corporate insiders can continue to trade on this long-term earnings news after the announcement of the earnings news for fiscal period *S*–1 (i.e., Period 3 in Fig. 1), even if they are banned from trading in the new blackout period (i.e., Period 1 in Fig. 1).

We use two proxies for *NEWS*: *SUE* and *CAR*. *CAR* is more precise than *SUE* as a proxy for the earnings surprise because *SUE* is defined using a seasonal random walk model, given the lack of analysts' earnings forecast data for more than half of the observations in our sample. *CAR* can also capture other non-earnings news released during the earnings announcement (e.g., dividend surprises). Nonetheless, *SUE* is less endogenous than *CAR* because insider trades made in Period 1 must be disclosed within three business days in Hong Kong; therefore, *CAR* could understate corporate insiders' private information advantage. The selection of control variables follows prior insider trading literature (e.g., Ke et al., 2003). We also include *NDAY* because the number of trading days in Period 1 could vary across firms and time periods.

Panel A of Table 4 presents the detailed sample selection procedures for model (4), and Panel B reports the descriptive statistics for the variables. Consistent with prior insider trading literature, *NETBUY* is highly skewed. The mean value of *NETBUY* is 0.227, and the value of the 75th percentile is zero. Untabulated statistics show that *NETBUY* is negative (positive) for 7.75% (11.13%) of the observations, and 81.12% have a zero value of *NETBUY*.

[19] As a robustness check, we use three alternative definitions of the dependent variable in model (4): the dollar value of insider purchases minus the dollar value of insider sales in Period 1 scaled by the market value of equity at the end of the prior half-year fiscal period (*netBUYMV*); the number of insider purchase transactions minus the number of insider sale transactions in Period 1 (*netN*); and the number of directors who have a net purchase in Period 1 minus the number of directors who have a net sale in Period 1 (*netDIR*). A director is defined to have a net purchase in Period 1 if the dollar value of his/her inside purchases is greater than the dollar value of his/her inside sales in Period 1. In addition, we use the quintile rank of *SUE* and *CAR* for each fiscal period as alternative measures of earnings news. We obtain similar inferences using these alternative variable definitions (untabulated) in that we do not find evidence that insiders of the lobbying firms are more aggressive in trading on forthcoming earnings news than insiders of the non-lobbying firms.

Table 4
Aggressiveness of insider trading on forthcoming earnings news in the proposed new blackout window.

Panel A. Sample selection procedures

Selection Procedures	The number of firm-half-year fiscal periods remained
Number of firm-half year fiscal periods that satisfy the following two conditions: (i) the fiscal period (i.e., fiscal period *S* in Fig. 1) starts after 30/6/2003 and ends on or before 30/6/2008, and (ii) the announcement of the earnings for the previous half-year period (i.e., fiscal period *S*-1) occurred after the end date of the previous fiscal period and before seven calendar days past the official earnings announcement deadline.	7712
Delete firms listed after 1/1/2007 or delisted before 1/1/2009	7622
Delete firms disclosing quarterly reports	7419
Delete firms that change fiscal period end dates	7383
Delete firms with missing control variables for *lnQ*, *lnMV*, and *PRET6m*	6635
Delete firms with missing either proxies for $NEWS_{S\text{-}1}$ or $NEWS_S$	6505
Final sample	
Use *SUE* to proxy for *NEWS*	5895
Use *CAR* to proxy for *NEWS*	5235

Panel B. Descriptive statistics

				Percentile		
Variable	N	Mean	stdev	25%	50%	75%
NETBUY	6505	0.227	3.264	0	0	0
LOBBY	6505	0.238	0.426	0	0	0
$SUE_{S\text{-}1}$	6010	0.018	0.115	−0.009	0.006	0.028
SUE_S	6200	0.011	0.116	−0.011	0.006	0.027
$CAR_{S\text{-}1}$	5723	−0.005	0.094	−0.056	−0.010	0.038
CAR_S	5629	−0.006	0.094	−0.057	−0.009	0.039
lnMV	6505	7.044	1.858	5.684	6.776	8.172
lnQ	6505	0.127	0.594	−0.254	0.017	0.411
PRET6m	6505	0.127	0.399	−0.093	0.087	0.322
NDAY	6505	36.736	13.799	28	36	50

Panel C. OLS regression results of NETBUY

Independent Variable	*NEWS* = *SUE* (1)	*NEWS* = *CAR* (2)
$NEWS_{S\text{-}1}$	0.7974*	0.9495*
	(0.087)	(0.077)
$NEWS_{S\text{-}1}\times LOBBY$	0.4404	1.2422
	(0.669)	(0.290)
$NEWS_S$	0.5513	0.6564
	(0.146)	(0.210)
$NEWS_S\times LOBBY$	−0.2857	1.3685
	(0.692)	(0.278)
LOBBY	0.3675**	0.3751**
	(0.021)	(0.029)

Table 4 (*continued*)

Panel C. OLS regression results of NETBUY		
Independent Variable	*NEWS = SUE* (1)	*NEWS = CAR* (2)
lnMV	−0.0269	−0.0462
	(0.497)	(0.299)
lnQ	−0.2745***	−0.2652**
	(0.002)	(0.011)
PRET6m	−0.4849***	−0.5069***
	(0.000)	(0.000)
NDAY	−0.0032	−0.0045
	(0.516)	(0.382)
Fixed effects of fiscal period and industry	yes	Yes
Adjusted R^2	0.0181	0.019
N	5895	5235

Note: See Appendix B for detailed variable definitions. The *p*-values in parentheses are based on standard errors clustered by firm. ***, **, and * indicate significant at the 1%, 5%, and 10% levels.

Panel C of Table 4 shows the regression results of model (4). The first column shows the regression results using *SUE*. The coefficient on SUE_{S-1} is positive and significant (0.7974, $p = 0.087$), suggesting that insider trading by non-lobbying firms is sensitive to forthcoming earnings news. However, untabulated regression results show that neither the coefficient on positive SUE_{S-1} nor the coefficient on negative SUE_{S-1} is significant when we split SUE_{S-1} into positive and negative components. More importantly, the coefficient on SUE_{S-1} - $\times$ *LOBBY* is nonsignificant ($p = 0.669$). Thus, we find no evidence that insiders at lobbying firms are more aggressive in trading on SUE_{S-1} in the proposed new blackout window than those at non-lobbying firms.

Though not our focus of interest, we note that the coefficients on SUE_S and $SUE_S \times LOBBY$ are both nonsignificant. The sum of the coefficient on SUE_S and that on $SUE_S \times LOBBY$ is also nonsignificant (untabulated). Thus, there is little evidence that insiders at lobbying or non-lobbying firms trade on long-term earnings news SUE_S.

The second column in Panel C shows the regression results using *CAR* as a proxy for *NEWS*. The coefficient on $CAR_{S-1} \times LOBBY$ is nonsignificant (1.2422, $p = 0.290$), suggesting no evidence that the lobbying firms' insiders trade more aggressively on the forthcoming earnings news in Period 1 than non-lobbying firms' insiders. The coefficient on CAR_{S-1} is significantly positive (0.9495, $p = 0.077$), suggesting that non-lobbying firms' insiders trade on the forthcoming earnings news CAR_{S-1} in Period 1. However, untabulated regression results show that the positive coefficient on CAR_{S-1} is driven by positive CAR_{S-1}, and the coefficient on negative CAR_{S-1} is nonsignificant. Hence, we find no evidence that insiders sell shares opportunistically before announcements of negative earnings news (i.e., profiting from bad news), a more troublesome form of insider trading behavior (Carlton and Fischel, 1983).

Overall, the results in Table 4 provide little evidence to support the allegation that insiders at lobbying firms trade more aggressively than those at non-lobbying firms in response to forthcoming earnings news during the proposed new blackout period.

6.2. Timeliness of earnings announcement

We next examine whether insiders at lobbying firms are more likely than those at non-lobbying firms to delay earnings announcements, as alleged by proponents of the new blackout regulation. Specifically, we estimate the following OLS regression model:

$$LAG = \beta_0 + \beta_1 LOBBY + CONTROL + \text{industry and year fixed effects} + \varepsilon \quad (5)$$

See Appendix B for variable definitions. If insiders at lobbying firms (firms more affected by the new regulation) are more opportunistic, we expect the coefficient on *LOBBY* to be positive. Following Chan et al. (2012), we control for total assets (*lnTA*), Tobin's Q (*Q*), sales growth rate (*SALESGROW*), leverage (*LEV*), profitability (*ROA*), an indicator of reporting a loss (*LOSS*), number of business segments (*NSEG*) and an indicator of whether a firm is engaged in a merger and acquisition (*M&A*). We also include two indicators for whether a firm's fiscal year ends in March (*MARCH*) or December (*DECEMBER*) to control for the fact that about 24% of listed firms in Hong Kong have a fiscal year ending in March and about 65% have a fiscal year ending in December. We control for whether a firm is also listed on mainland China's A-share market (*AH*), as reporting deadlines are shorter for mainland-listed stocks.[20] We also include industry and year fixed effects.

Panel A of Table 5 reports the sample selection procedures for model (5). The sample includes announcements of earnings for half-year fiscal periods ending between 30 June 2003 and 31 December 2007. The model is estimated separately for interim and annual earnings announcements because the listing rules require interim (annual) reports to be announced within three (four) months after the fiscal period end date during our sample period.

As shown in Panel B of Table 5, the mean *LAG* is about 73 days for interim earnings announcements and about 100 days for annual earnings announcements, consistent with regulators' observations that Hong Kong-listed firms are relatively slow to report earnings results. Nevertheless, we find no evidence that lobbying firms

[20] During our sample period, firms listed on the Shanghai or Shenzhen stock exchanges are required to announce interim (quarterly) earnings within one month whereas firms listed on the HKEX are required to announce interim earnings within three months.

Table 5
Timeless in earnings announcements.

Panel A. Sample selection procedures

Sample selection procedures	Number of observations remained for the interim report sample	Number of observations remained for the annual report sample
All (interim and annual) earnings announcements for half-year fiscal periods (i.e., fiscal period *S*-1 in Fig. 1) with an end date between 30/06/2003 and 31/12/2007 (the announcements occurred in fiscal periods *S*).	3926	3861
Delete firms listed after 1/1/2007 or delisted before 1/1/2009	3887	3808
Delete observations that change the fiscal period end	3871	3760
Delete observations that report quarterly earnings in the current fiscal year	3805	3502
Delete observations with missing control variables (*lnTA*, *ROA*, *Q*, *SALESGROW*, *LEV*, *M&A*)	2595	3423
Final sample	2595	3423

Panel B. Descriptive statistics
Interim report sample

				Percentile		
Variable	N	mean	stdev	25%	50%	75%
LAG	2595	72.568	14.751	64	77	83
LOBBY	2595	0.238	0.426	0	0	0
lnTA	2595	21.203	1.853	19.949	20.990	22.315
Q	2595	1.543	1.516	0.811	1.088	1.610
SALESGROW	2595	0.105	0.651	−0.078	0.104	0.290
ROA	2595	0.021	0.087	−0.005	0.019	0.052
LEV	2595	0.189	0.173	0.041	0.155	0.289
LOSS	2595	0.291	0.454	0	0	1
NSEG	2595	2.773	1.556	2	3	4
M&A	2595	0.024	0.153	0	0	0
MARCH	2595	0.311	0.463	0	0	1
DECEMBER	2595	0.573	0.495	0	1	1
AH	2595	0.052	0.221	0	0	0

Annual report sample

				Percentile		
Variable	N	mean	stdev	25%	50%	75%
LAG	3423	100.303	17.198	89	106	112
LOBBY	3423	0.228	0.420	0	0	0
lnTA	3423	21.309	1.943	19.989	21.086	22.449
Q	3423	1.542	1.524	0.813	1.088	1.640
SALESGROW	3423	0.122	0.596	−0.039	0.121	0.289
ROA	3423	0.030	0.141	−0.012	0.031	0.093
LEV	3423	0.188	0.175	0.037	0.152	0.289

(*continued on next page*)

Table 5 (*continued*)

Annual report sample

Variable	N	mean	stdev	Percentile 25%	50%	75%
LOSS	3423	0.309	0.462	0	0	1
NSEG	3423	2.785	1.565	1	3	4
M&A	3423	0.342	0.474	0	0	1
MARCH	3423	0.242	0.428	0	0	0
DECEMBER	3423	0.652	0.476	0	1	1
AH	3423	0.057	0.232	0	0	0

Panel C. Regression results of earnings announcement timing (dependent variable is LAG)

Independent Variable	Interim report sample (1)	Annual report sample (2)
LOBBY	1.0239	−1.9878*
	(0.321)	(0.083)
lnTA	−3.4051***	−3.4455***
	(0.000)	(0.000)
Q	−1.0772***	−1.1384***
	(0.000)	(0.000)
SALESGROW	0.1501	0.1139
	(0.663)	(0.802)
ROA	−15.3787***	−16.1398***
	(0.001)	(0.000)
LEV	9.6138***	11.5387***
	(0.000)	(0.000)
LOSS	−0.089	2.0380**
	(0.911)	(0.044)
NSEG	1.0094***	1.0315***
	(0.000)	(0.001)
M&A	2.6984*	3.0458***
	(0.089)	(0.000)
MARCH	−2.5185**	1.9912
	(0.036)	(0.259)
DECEMBER	−3.1512**	−1.5935
	(0.014)	(0.356)
AH	−11.9304***	2.2936
	(0.000)	(0.286)
Industry and year fixed effects	yes	Yes
Adjusted R^2	0.2707	0.2753
N	2595	3423

Note: See Appendix B for detailed variable definitions. The p-values in parentheses are based on standard errors clustered by firm. ***, **, and * indicate significant at the 1%, 5%, and 10% levels.

are more likely to delay earnings announcements in the period prior to the introduction of the new blackout regulation. Untabulated tests show that for interim earnings announcements, the mean LAG is 72.8 days for lobbying firms and 72.5 days for non-lobbying firms (p-value for the two-sample t-test = 0.701). For annual earnings announcements, the reporting lag is shorter for the lobbying firms than for the non-lobbying firms. The mean LAG is 96.9 days for lobbying firms and 101.3 days for non-lobbying firms (p-value for two-sample t-test <0.001).

Panel C of Table 5 shows the results of regressions controlling for other economic determinants of reporting lag, and the inference is consistent with that drawn from the univariate analysis. The coefficient on $LOBBY$ is not significant (1.0239, $p = 0.321$) in the regression of interim reporting lag in column (1) and significantly negative (–1.9878, $p = 0.083$) in the regression of annual reporting lag in column (2). Hence, we find no support for the allegation that insiders at lobbying firms are more likely to delay their earnings announcements. If anything, we find some evidence that lobbying firms' reporting of annual earnings is more timely than that by non-lobbying firms.

6.3. Insider trading during market turmoil

Our final analysis tests a prediction made by critics of the new blackout regulation that insiders at lobbying firms are more likely to step in and buy more shares when their firms' stock prices experience short-term turmoil. This prediction is also consistent with one of the reasons stated in the lobbying firms' open letter (see Appendix A). Specifically, reason #3 in the open letter states that "*Against a backdrop of the worsening financial tsunami, prohibiting directors' dealings for an extended period of time will reduce trading volume further, jeopardizing an important source of market liquidity and stable support for shareholder value in unanticipated crises and events. This creates a window of opportunities for corporate snipers to destroy value for both major and small shareholders.*" As noted in the Introduction, we are not able to test the other alleged costs of the new blackout regulation in the lobbying letter due to a lack of data.

We test this prediction using the following OLS regression model:

$$\begin{aligned} NETBUY1w = {} & \beta_0 + \beta_1 NEWS_{S-1} + \beta_2 NEWS_S + \beta_3 lnMV + \beta_4 lnQ + \beta_5 PosPRET1w + \beta_6 NegPRET1w \\ & + \beta_7 D(PRET1w > 0) + \text{industry and fiscal period fixed effects} + \varepsilon \end{aligned} \tag{6}$$

See Appendix B for all variable definitions. Model (6) is similar to model (4) except for the following two important differences. First, the unit of observation in model (6) is a calendar week because we want to study insider trading behavior in the proposed new blackout window in response to past short-term, rather than long-term, stock price patterns. Hence, both the dependent variable $NETBUY1w$ and the key independent variable of interest $PRET1w$ are measured weekly. Aggregating insider trading behavior and past return momentum over a longer period may obscure such subtle short-term insider trading behavior. Second, we decompose $PRET1w$ into three variables—positive returns ($PosPRET1w$), negative returns ($NegPRET1w$) and an indicator of positive returns ($D(PRET1w > 0)$)—to capture different insider trading behavior in response to past positive and negative short-term stock returns. Our variable of interest is $NegPRET1w$, which captures the propensity of insiders to step in to support their firms' stock prices through insider purchases during short-term market turmoil.

Table 6 shows the descriptive statistics in Panel A and the regression results in Panel B. The first two columns of Panel B use SUE to proxy for $NEWS$. Column (1) shows the regression results for the lobbying firms. We find that the coefficient on $NegPRET1w$ (–1.5336, $p = 0.011$) is significantly negative, implying that lobbying firms' insiders buy more shares when their firms' short-term stock prices fall significantly. The coefficient on $PosPRET1w$ (–0.7827, $p = 0.084$) is marginally significant and negative, suggesting that lobbying firms' insiders also sell shares when their firms' short-term stock prices have risen significantly. Column (2) shows the regression results for the non-lobbying firms. The coefficient on $PosPRET1w$ (–1.2699, $p < 0.001$) is significantly negative, while the coefficient on $NegPRET1w$ (–0.156, $p = 0.514$) is nonsignificant. These results suggest that the insiders of non-lobbying firms are more likely to sell shares after their firms' short-term stock

Table 6
Insider trading in the proposed new blackout window during market turmoil.

Panel A. Descriptive statistics

				Percentiles		
variable	N	mean	stdev	25%	50%	75%
*NETBUY*1*w*	49,849	0.049	1.545	0.000	0.000	0.000
$CAR_{S\text{-}1}$	41,909	−0.007	0.099	−0.060	−0.012	0.034
CAR_S	41,779	−0.006	0.096	−0.058	−0.011	0.037
$SUE_{S\text{-}1}$	44,438	0.021	0.130	−0.012	0.006	0.031
SUE_S	46,410	0.017	0.123	−0.010	0.006	0.029
lnMV	49,849	6.642	1.744	5.415	6.392	7.706
lnQ	49,849	0.102	0.601	−0.281	−0.014	0.381
*PRET*1*w*	47,891	0.006	0.075	−0.028	0.000	0.035

Panel B. Regression analysis (*the dependent variable is NETBUY1w*)

	NEWS measured by *SUE*		*NEWS* measured by *CAR*	
Independent Variable	All lobbying firms (1)	Non-lobbying firms (2)	All lobbying firms (3)	Non-lobbying firms (4)
$NEWS_{S\text{-}1}$	0.0674	0.1351	0.4271	0.2727**
	(0.640)	(0.242)	(0.162)	(0.044)
$NEWS_S$	0.0070	0.1973*	0.4631	0.1768
	(0.964)	(0.055)	(0.137)	(0.203)
lnMV	0.0102	−0.0166	0.0081	−0.0242**
	(0.711)	(0.120)	(0.783)	(0.037)
lnQ	−0.1845***	−0.0357*	−0.1627**	−0.0391
	(0.009)	(0.095)	(0.047)	(0.111)
*PosPRET*1*w*	−0.7827*	−1.2699***	−1.0251**	−1.2147***
	(0.084)	(0.000)	(0.047)	(0.000)
*NegPRET*1*w*	−1.5336**	−0.156	−2.5256***	−0.6397**
	(0.011)	(0.514)	(0.001)	(0.034)
D(*PRET*1*w*>0)	−0.0219	−0.0141	0.0137	−0.0131
	(0.618)	(0.489)	(0.765)	(0.527)
Fixed effects of fiscal period and industry	Yes	yes	yes	Yes
Adjusted R^2	0.0172	0.0069	0.0179	0.0073
N	11,588	36,728	10,577	32,496

p-value for the difference in the coefficient between subsamples

Variable	(1) vs (2)	(3) vs (4)
$NEWS_{S\text{-}1}$	0.713	0.642
*PosPRET*1*w*	0.332	0.737
*NegPRET*1*w*	0.032**	0.020**
D(*PRET*1*w*>0)	0.871	0.593

Note: See Appendix B for detailed variable definitions. The *p*-values in parentheses are based on standard errors clustered by firm.

prices have risen, but there is no clear evidence that they step in to buy more shares after their firms' short-term stock prices have fallen.[21]

We also show in Table 6 that the coefficient on *NegPRET*1*w* is significantly more negative for the lobbying firms than for the non-lobbying firms ($p = 0.032$). Overall, the results in Table 6 suggest that insiders at the lobbying firms have a greater propensity to buy shares to support their firms' stock prices during short-term market turmoil, a finding consistent with the lobbying firms' claims. In contrast, the coefficient on *PosPRET*1*w* for the lobbying firms is not significantly different from that for the non-lobbying firms ($p = 0.332$).

[21] Consistent with the signaling role of insider purchases during Period 1, we find in an untabulated analysis that the short-run stock market reaction to such insider purchases in Period 1 is positive for both lobbying firms and non-lobbying firms. In contrast, we do not find a significant market reaction to the disclosure of insider sales in Period 1.

The last two columns of Table 6 replicate the same regression models, using *CAR* as a proxy for *NEWS*. The inferences are qualitatively the same except that the coefficient on $NegPRET1w$ becomes significantly negative for the non-lobbying firms.

7. Conclusion

Even though many financial markets have adopted some form of insider trading regulation, insider trading remains frequent and profitable in many markets, raising questions about the adequacy of existing restrictions. The objective of this study is to shed light on this important question by examining the economic consequences for shareholder value of further tightening insider trading regulation. We examine our research question in Hong Kong because of the availability of a unique corporate lobbying event that led to the unexpected reversal of a tough insider trading blackout regulation. Unlike the setting of insider trading regulation, which is typically contaminated by other concurrent regulations, this reversal offers a rare setting in which we can cleanly identify the effect of further tightening insider trading regulation.

Using an event study approach, we find that the cumulative stock market reaction to the events associated with the reversal of the new blackout regulation is significantly positive for the entire Hong Kong stock market portfolio. In addition, the market reaction is more positive for lobbying firms, which are more affected by the new regulation, than for the matched comparable non-lobbying firms. Overall, these results suggest that the tighter blackout regulation hurts shareholder value for the overall Hong Kong stock market.

We also directly test some of the costs and benefits of insider trading alleged by proponents and critics of the new blackout regulation. First, we find no evidence that insiders at lobbying firms are more aggressive than those at non-lobbying firms in trading on forthcoming earnings news in the proposed new blackout window, or in delaying earnings announcements to maximize insiders' trading opportunities. If anything, we find weak evidence that lobbying firms are less likely to delay earnings announcements. These findings fail to support the allegations made by the proponents of the new blackout regulation. However, consistent with allegations by critics of the new blackout regulation, we find that insiders at lobbying firms are more likely than those at non-lobbying firms to step in and purchase additional shares when their firms' stock prices experience short-term market turmoil.

Overall, the results suggest that further tightening of insider trading regulation hurts shareholder value across the entire Hong Kong stock market, especially for firms more affected by the new regulation (i.e., lobbying firms). Our results suggest little evidence of opportunistic insider trading during the proposed new blackout period; instead, we find that insider trades are often used to reduce information asymmetry between insiders and outside investors, especially for lobbying firms (the firms most affected by the regulation). Our findings contradict a common belief among both Hong Kong regulators and some vocal shareholder activists and suggest caution in imposing tougher one-size-fits-all insider trading regulations.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Grammarly in order to correct grammatical errors. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bin Ke reports financial support was provided by MOE 111 Project. Bin Ke reports financial support was provided by MOE Project of Key Research Institute of Humanities and Social Sciences at Universities. Zhihong Chen reports financial support was provided by Hong Kong Research Grant Council. Bin Ke is a consulting editor of CJAR If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We wish to thank Kevin Chen, Mingyi Hung, David Reeb, Gordon Richardson, Xiangang Xin, Chenqi Zhu (discussant), and workshop participants at the University of Melbourne, City University of Hong Kong, Nanyang Technological University, University of Hong Kong, Hong Kong University of Science and Technology, the APJAE Symposium, and the AAA annual meeting for helpful comments. Part of the research is supported by a grant from the Hong Kong Government's Research Grant Council (Project 141110). Bin Ke acknowledges the funding support by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (No. 23JJD790009) and the 111 Project (No. B18033).

Appendix A. The lobbying letter published in the South China Morning Post on 29 December 2008

Extended blackout period detrimental to hong kong's position as international financial centre

To: Honorable Executives of Financial Services and the Treasury Bureau of the HKSAR, the Hong Kong Securities and Futures Commission and the Hong Kong Exchanges and Clearing Limited ("HKEx").

Damage to viability and health of hong kong market

We are a group comprising listed companies in Hong Kong, directors of such listed companies, participants in the securities markets and professional organizations serving the Hong Kong financial industry, who are concerned with the sustainable and healthy development of Hong Kong as an international financial centre. **We write to express our strong objection to an amendment to HKEx's Listing Rules* announced recently, which extends the "blackout period" on directors' dealings in their companies' securities from within one month before the announcement of results to the financial year/period end until the announcement of the relevant results. As such, directors could be prohibited from dealing for up to 7 calendar months in a year. For companies reporting quarterly results, the blackout period will last for up to 9 calendar months.** In light of the current environment where the financial tsunami is eroding the wellbeing of all industries and sectors in Hong Kong, not only will this amendment damage the continued viability and health of the Hong Kong market, it will also undermine the legitimate rights and functions of directors as investors.

Deterrent to existing and future listing on the hong kong stock exchange

As in other economies in Asia, ownership and management of companies already listed or seeking a listing in Hong Kong are not completely segregated. Directors are often also owners of substantial or controlling interests in the listed companies. The amendment deprives such directors of their legitimate rights as investors and their legitimate expectations of a level playing field. As a result, companies already listed in Hong Kong might relocate to other markets and those contemplating a listing here might decide to list elsewhere. This will inflict damage on Hong Kong's status as an international financial centre.

Jepardy to market liquidity and shareholder value

Against a backdrop of the worsening financial tsunami, prohibiting directors' dealings for an extended period of time will reduce trading volume further, jeopardizing an important source of market liquidity and stable support for shareholder value in unanticipated crises or events. This creates a window of opportunities for corporate snipers to destroy value for both major and small shareholders.

Discourage individuals from becomig directors of listed companies

Directors of listed companies expend their time and expertise seeking to maintain and improve return for shareholders. The amendment penalizes them by further restricting their freedom to deal in their companies' shares. It will discourage talented individuals of the right caliber from becoming directors of listed companies.

Current insider trading rules adequate

We believe that the existing insider dealing provisions under the Securities and Futures Ordinance and the protection against unauthorized use of price sensitive information under the relevant rules of HKEx already provide adequate scrutiny and safeguard on directors' dealings. Directors are also required to make full disclosure for their dealings. The Listing Rules have outlined explicitly that HKEx regards it as highly desirable that directors should hold securities in companies. Further manoeuvers to the existing regime in this regard constitute a blatant regulatory overshoot.

Consulation proess questionable and hasty

In addition, we have strong reservations as to whether the public consultation on such an important amendment has been conducted and concluded properly. There was not adequate communication to listed companies and the public to ensure that they are properly informed of such proposal so as to respond or raise objection if they so wish. There was tremendous objection to the proposal with substantive and well-elaborated arguments, and yet no proper evaluation of the substance of the responses was made. We consider it extremely inappropriate that the HKEx concluded that this amendment of such grave impact should proceed in such haste.

We hereby urge HKEx to withhold this amendment for further consulation

A group concerned with the sustainable and healthy development of Hong Kong as an international financial center.

(The name list is omitted for brevity.).

*Note: Listing Rules refer to the Rules Governing the Listing of Securities on The stock Exchange of Hong Kong Limited and the Rules Governing the Listing of Securities on the Growth Enterprise Market of the Stock Exchange of Hong Kong Limited.

Appendix B. Variable definitions and sources of data

Variables used in regression model (2).

Variable name	Definition
LOBBY	A dummy variable that equals one for the firms that lobbied against the new blackout rule and zero for the non-lobbying firms. The list of lobbying firms is obtained from the open letter published in the South China Morning Post on 29 December 2008.
HEAVY_TRADING	An indicator variable that equals one if the firm's insiders are most affected by the proposed new blackout period. We define *HEAVY_TRADING* using all half-year fiscal periods (i.e., fiscal period *S* in Fig. 1) that start after 30 June 2003 and end on or before 30 June 2008 to make sure that *HEAVY_TRADING* is computed prior to the publication of the open letter on 29 December 2008. We delete half-year fiscal periods if there is a change in the fiscal period end, the earnings announcement date is missing, or the earnings announcement date is either before the fiscal period end or 7 days after the official earnings announcement deadline. We divide the trading days in each half-year fiscal period into three periods as shown in Fig. 1. *N*1 is the number of trading days in Period 1. *N*3 is the number of trading days in Period 3. *Mean_N1R* is the mean value of $N1/(N1 + N3)$ for each firm in the sample. We also compute the relative insider trading volume (*RITVOL*) in Period 1, defined as the total dollar volume of open-market purchases or sales by officers and directors in Period 1 divided by the total dollar trading volume in Period 1. *Mean_RITVOL* is the mean value of *RITVOL* for each firm in the sample.[22] We sort the sample firms into two groups (high and low) based on the medians of *Mean_N1R* (0.367) and *Mean_RITVOL* (0.001), respectively. *HEAVY_TRADING* is one for firms that fall in both the high *Mean_N1R* and high *Mean_RITVOL* groups, and zero otherwise. The insider trading data are obtained directly from the HKEX and comprise all notifications of directors' or chief executives' dealings in their own firms' shares for all HKEX-listed firms.
lnRET	Natural logarithm of one plus the compounded dividend-inclusive return from 1 December 2007 to 30 November 2008, the month prior to the publication of the open letter. The stock return data are obtained from *DataStream*.
lnQ	Natural logarithm of Tobin's Q, where *Q* is defined as the market value of equity minus book value of equity plus book value of total assets, divided by book value of total assets, measured at the end of the fiscal year ending in 2007. The accounting data are obtained from *WorldScope* and the market value of equity is obtained from *DataStream*.
lnFOLLOW	Natural logarithm of one plus the number of analysts that provided forecasts for the year 2008 earnings in the I/B/E/S summary file.
SIZE	Natural logarithm of the total market value of equity at the end of the year 2007, retrieved from *DataStream*.
AGE	Natural logarithm of the number of months since a firm's stock appeared in *DataStream*, measured at the end of 2007.

[22] Using the raw dollar value of insider trading volume in Period 1 rather than *RITVOL* yields similar inferences.

Variables used in regression model (4).

Variable	Definition
NETBUY	Natural logarithm of one plus the dollar value of total insider open market purchases (in HK$1000) minus the natural logarithm of one plus the dollar value of total insider open market sales (in HK$1000) in Period 1 (see Fig. 1). The insider trading data are directly obtained from the HKEX.
SUE_S	The seasoned difference in the half-year earnings. *SUE* of the half-year fiscal period *S* is defined as the net income of fiscal period *S* minus the net income of period *S*–2, and is divided by total assets at the end of *S*–2. The accounting data are obtained from *WorldScope*.
CAR_S	Cumulative abnormal return over the trading days [–2, +2] centered on the date of the announcement of the earnings of the half-year fiscal period *S*. Abnormal return is defined as the raw return minus the value-weighted market return (dividend inclusive). The stock return data are obtained from *DataStream*.
lnMV	Natural logarithm of market value of equity at the end of the prior half-year fiscal period, obtained from *DataStream*.
lnQ	Natural logarithm of Tobin's Q measured at the end of the prior half-year fiscal period. Tobin's Q is defined as the market value of equity minus the total book value of equity plus the total book value of assets, divided by the total book value of assets. The accounting data are obtained from *WorldScope* and the market value of equity is obtained from *DataStream*.
*PRET*6*m*	Buy-and-hold return over the prior half-year fiscal period, obtained from *DataStream*.
NDAY	Number of trading days in Period 1 (see Fig. 1).

The variables *NETBUY*, *CAR*, *SUE*, *lnQ*, *PRET*6*m* and *NDAY* are winsorized at the 1st and 99th percentiles.

Variables used in regression model (5).

Variable	Definition
LAG	The number of calendar days between the fiscal period ending date and the earnings announcement date for the fiscal period. For observations whose earnings announcement dates fall more than 7 calendar days past the official deadline, we set *LAG* to the maximum allowed by the regulation, plus 7 calendar days. The listing rule requires interim reports to be announced within 3 months after the fiscal period end date and annual reports within 4 months. We delete observations with a negative *LAG* because such observations are likely due to errors in the disclosed earnings announcement dates. The earnings announcement date data are obtained from Bloomberg, *WorldScope* and I/B/E/S. The fiscal period end dates are from *WorldScope*.
lnTA	Natural logarithm of total assets at the end of the fiscal period, obtained from *WorldScope*.
Q	Tobin's Q at the end of the fiscal period, defined as the market value of equity plus total assets minus the book value of equity, divided by total assets. The accounting data are from *WorldScope* and the market value of equity data is obtained from *DataStream*.
SALESGROW	Natural logarithm of the sales growth rate for the fiscal period. For the annual report, the sales growth rate is defined as sales in the current year divided by sales in the previous year. For the interim report, the sales growth rate is defined as the ratio of sales in the first half of the current fiscal year to the sales in the first half of the previous fiscal year. Sales data are obtained from *WorldScope*.
ROA	Return on assets for the fiscal period, defined as operating income over the fiscal period divided by total assets at the end of the fiscal period, both from *WorldScope*.
LEV	Leverage at the end of the fiscal period, defined as total liabilities divided by total assets, calculated from *WorldScope*.
LOSS	A dummy variable that equals one if the reported operating income for the fiscal period is negative, and zero otherwise. Data on operating income are from *WorldScope*.
NSEG	Number of business segments reporting positive segment sales in the fiscal period. If missing, *NSEG* is set to one. The data on business segments are obtained from *WorldScope*.
M&A	A dummy variable that equals one if acquisition in the half-year fiscal period is not zero or missing, and zero otherwise. The data are from *WorldScope*.
MARCH	A dummy variable that equals one if the fiscal year ends in March, and zero otherwise. The data for the fiscal year-end date are from *WorldScope*.
DECEMBER	A dummy variable that equals one if the fiscal year ends in December, and zero otherwise. The data for the fiscal year-end date are from *WorldScope*.
AH	A dummy variable that equals one if the firm is listed on both the HKEX and the Shenzhen or Shanghai stock exchanges. The data are from *DataStream*.

The variables *Q*, *SALESGROW*, *ROA*, *LEV* and *NSEG* are winsorized at the 1st and 99th percentiles.

Variables used in regression model (6).

Variable	Definition
*NETBUY*1*w*	Natural logarithm of one plus the dollar value of total insider open market purchases (in HK$1000) minus the natural logarithm of one plus the dollar value of total insider open market sales (in HK$1000) in **a week** during Period 1 (see Fig. 1). The insider trading data are directly obtained from the HKEX.
SUE_S	The seasoned difference in the half-year earnings. *SUE* of the half-year fiscal period *S* is defined as the net income of fiscal period *S* minus the net income of period *S*–2, and is divided by total assets at the end of *S*–2. The accounting data are from *WorldScope*.
CAR_S	Cumulative abnormal return over the trading days [–2, +2] centered on the date of the announcement of the earnings of the half-year fiscal period *S*. Abnormal return is defined as the raw return minus the value-weighted market return (dividend inclusive). The stock return data are obtained from the *DataStream*.
lnMV	Natural logarithm of market value of equity at the end of the prior half-year fiscal period, obtained from *DataStream*.
lnQ	Natural logarithm of Tobin's Q measured at the end of the prior half-year fiscal period. Tobin's Q is defined as the market value of equity minus the total book value of equity plus the total book value of assets, divided by the total book value of assets. The accounting data are from *WorldScope* and the market value of equity data are obtained from *DataStream*.
*PRET*1*w*	The buy-and-hold return over the week before the beginning date of each weekly observation. The stock return data are obtained from *DataStream*.
*PosPRET*1*w*	*PosPRET*1*w* equals *PRET*1*w* if *PRET*1*w* is positive, and zero otherwise.
*NegPRET*1*w*	*NegPRET*1*w* equals *PRET*1*w* if *PRET*1*w* is negative, and zero otherwise.
D(*PRET*1*w* > 0)	A dummy variable that equals one if *PRET*1*w* is positive, and zero otherwise.

The unit of observation is a firm-week. For each firm-Period 1, the variables *SUE*, *CAR*, *lnMV* and *lnQ* are identical for each week.

Appendix C. *The propensity score matching procedure:*

We follow the approach outlined in Hochberg et al. (2009) to implement propensity score matching. We first estimate the following logit model:

$$\Pr(LOBBY) = \beta_0 + \beta_1 SIZE + \beta_2 SIZE_SQUARED + \beta_3 lnQ + \beta_4 lnRET12 + \text{industry fixed effects} + \varepsilon$$

where *SIZE* is defined as the natural logarithm of the market value of equity (in HK$) at the end of the year 2007. *SIZE_SQUARED* is the square of *SIZE*. *lnQ* is the natural logarithm of Tobin's Q, measured at the end of the fiscal year ending in 2007. *lnRET*12 is the natural logarithm of one plus the compounded return over the year 2007. Industry fixed effects are based on the Fama–French 12-industry classifications.

The results of the logit model are reported in panel A of Table A1 (column 1).

We use the predicted probability of lobbying as the *propensity score*. We then define 10 bins of lobbying and non-lobbying firms as follows. We compute the 10th, 20th and up to the 90th percentiles of the lobbying firms' propensity scores. All lobbying and non-lobbying firms with propensity scores below the 10th percentile are grouped in the first bin; all lobbying and non-lobbying firms with propensity scores between the 10th and 20th percentiles are grouped in the second bin; and so on. One advantage of defining the bins based on the propensity score percentiles of the lobbying firms alone is that each bin has a meaningful number of lobbying and non-lobbying firms (Hochberg et al., 2009).

The diagnostics of the propensity score matching are presented in Panel A (column 2) and Panel B of Table A1. Overall, the diagnostics show that the propensity score matching is reasonably effective.

Table A1
Propensity score matching diagnostics.

Panel A. Logistic regression results of lobbying

	(1)		(2)	
Independent variable	Coefficient	*p-value*	Coefficient	*p-value*
SIZE	1.176	(0.001)	0.496	(0.510)
SIZE_SQUARED	−0.054	(0.008)	−0.025	(0.495)
lnQ	−0.639	(0.002)	−0.190	(0.620)
*lnRET*12	−0.383	(0.023)	−0.211	(0.468)
Industry fixed effects	Yes		Yes	
Propensity score bin indicators			Yes	
Pseudo R-squared	0.112		0.151	
N	860		860	
Joint significance of *SIZE*, *SIZE_SQUARE*, *lnQ*, *lnRET*12, and industry fixed effects				
χ^2	66.95		6.24	
(*p*-value)	(0.000)		(0.960)	
Joint significance of propensity score bin indicators				
χ^2			31.71	
(*p*-value)			(0.000)	

Panel B. Firms characteristics of the lobbying firms and the non-lobbying firms

		Mean value		Test for the difference	
Variable	Sample	Lobbyingfirms	Non-Lobbying firms	*t*-stat	*p*-value
SIZE	Before matching	8.2016	7.4904	4.86	0.000
	After matching	8.2016	8.1832	0.10	0.923
SIZE_SQUARED	Before matching	70.5285	59.2025	4.95	0.000
	After matching	70.5285	70.3845	0.04	0.966
lnQ	Before matching	0.1949	0.3730	−3.49	0.000
	After matching	0.1949	0.1881	0.12	0.905
*lnRET*12	Before matching	0.3611	0.4836	−2.51	0.012
	After matching	0.3611	0.3940	−0.61	0.54

Data availability

The data used in this study are publicly available from the sources identified in the paper.

References

Armstrong, C., Barth, M.E., Jagolinzer, A.D., Riedl, E.J., 2010. Market reaction to the adoption of IFRS in Europe. Account. Rev. 85, 31–61.

Association for Sustainable & Responsible Investment in Asia. 2009. Directors' blackout period in Hong Kong – a test of Corporate Governance in Hong Kong. January 22, 2009.

Bainbridge, S.M., 2000. Insider Trading: An overview. Encyclopedia of Law and Economics, Volume III. The Regulation of Contracts, Cheltenham, Edward Elgar, 2000, 1205 p.

Baldwin, P. 2009. Insider trading laws must be effective. The South China Morning Post, June 20, 2009.

Baesel, J.B., Stein, G.R., 1979. The value of information: inferences from the profitability of insider trading. J. Financ. Quant. Anal. 14, 553–571.

Bekaert, G., Harvey, C.R., 1997. Emerging equity market volatility. J. Financ. Econ. 43, 29–77.

Bettis, J.C., Coles, J.L., Lemmon, M.L., 2000. Corporate policies restricting trading by insiders. J. Financ. Econ. 57, 191–220.

Bharara, P, Jackson Jr., R.J. 2018. Insider Trading Laws Haven't Kept Up With the Crooks. The New York Times (op-ed), October 9, 2018.

Bhattacharya, U., Daouk, H., 2002. The world price of insider trading. J. Financ. 57, 75–108.

Brochet, F., 2019. Aggregate insider trading and market returns: the role of transparency. J. Bus. Financ. Acc. 46, 336–369.

Brudney, V., 1979. Insiders, outsiders, and informational advantages under the Federal Securities Laws. Harv. Law Rev. 93, 322–376.

Bushman, R.M., Piotroski, J.D., Smith, A.J., 2005. Insider trading restrictions and analysts' incentives to follow firms. J. Financ. 60, 35–66.

Campbell, J., Lo, A., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Carlton, D.W., Fischel, D.R., 1983. The regulation of insider trading. Stanford Law Rev. 35, 857–895.

Chan, K., 2009. A study on the stock price performance subsequent to directors' dealings in Hong Kong. Commissioned by The Hong Kong Institute of Chartered Secretaries.

Chan, L.H., Chen, K., Chen, T.-Y., Yu, Y., 2012. The effects of firm-initiated clawback provisions on earnings quality and auditor's behavior. J. Account.Econ. 54, 180–196.

Chau, M., Vayanos, D., 2008. Strong-form efficiency with monopolistic insiders. Rev. Financ. Stud. 21, 2275–2306.

Cheng, Q., Lo, K., 2006. Insider trading and voluntary disclosures. J. Account. Res. 44, 815–848.

Cheuk, M.Y., Fan, D.K., So, R.W., 2006. Insider trading in Hong Kong: some stylized facts. Pac. Basin Financ. J. 14, 73–90.

Doidge, C., Karolyi, G.A., Stulz, R.M., 2007. Why do countries matter so much for corporate governance? J. Financ. Econ. 86 (1), 1–39.

Easterbrook, F.H., 1981. Insider trading, secret agents, evidentiary privileges, and the production of information. The Supreme Court Review 1981, 309–365.

Easterbrook, F.H., 1985. Insider trading as an agency problem. In: Pratt, J.W., Zeckhauser, R. (Eds.), Principals and Agents: the Structure of Business. Harvard Business School Press.

Fama, E.F., French, K.R., 2012. Size, value, and momentum in international stock returns. J. Financ. Econ. 105, 457–472.

Fernandes, N., Ferreira, M.A., 2009. Insider trading laws and stock price informativeness. Rev. Financ. Stud. 22, 1845–1887.

Fidrmuc, J.P., Goergen, M., Renneboog, L., 2006. Insider trading, news releases, and ownership concentration. J. Financ. 61, 2931–2973.

Fishman, M.J., Hagerty, K.M., 1992. Insider trading and the efficiency of stock prices. Rand J. Econ. 23, 106–122.

Flook, B. House Passes Insider Trading Ban. Accounting & Compliance Alert, December 6, 2019.

Garfinkel, J., 1997. New evidence on the effects of federal regulations on insider trading: the Insider Trading and Securities Fraud Enforcement Act (ITSFEA). Finance 3, 89–111.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14, 71–100.

Hillier, D., Marshall, A.P., 2002. Are trading bans effective? Exchange regulation and corporate insider transactions around earnings announcements. Finance 8, 393–410.

HKEX. 2008a. Combined consultation paper on proposed changes to the listing rules. January 2008.

HKEX. 2008b. Consultation conclusions on proposals in the 2008 combined consultation paper. November 2008.

Hochberg, Y.V., Sapienza, P., Vissing-Gorgensen, A., 2009. A lobbying approach to evaluating the Sarbanes-Oxley Act of 2002. J. Account. Res. 47, 519–583.

Jaffe, J.F., 1974. The effect of regulation changes on insider trading. Bell J. Econ. Manag. Sci. 5, 93–121.

Jayaraman, S., 2012. The effect of enforcement on timely loss recognition: evidence from insider trading laws. J. Account. Econ. 53, 77–97.

Kabir, R., Vermaelen, T., 1996. Insider trading restrictions and the stock market: evidence from the Amsterdam Stock Exchange. Eur. Econ. Rev. 40, 1591–1603.

Ke, B., Huddart, S., Petroni, K., 2003. What insiders know about future earnings and how they use it: evidence from insider trades. J. Account.Econ. 35, 315–346.

Lee, I., Lemmon, M., Li, Y., Sequeira, J., 2014. Do voluntary corporate restrictions on insider trading eliminate informed insider trading? Finance 29, 158–178.
Leng, S. S. 2009. As I Call It: Do longer black out trading periods benefit retail investors? The Edge Singapore, 16 March 2009.
Leuz, C., Wysocki, P., 2016. The economics of disclosure and financial reporting regulation: evidence and suggestions for future research. J. Account. Res. 54, 525–622.
Lo, K., 2003. Economic consequences of regulated changes in disclosure: the case of executive compensation. J. Account.Econ. 35, 285–314.
Macey, J.R., 1991. Insider Trading: Economics, Politics and Policy. AEI Press, Washington, DC.
Mahoney, P., 1995. Mandatory disclosure as a solution to agency problems. Univ. Chic. Law Rev. 62, 1047–1112.
Mak, F., 2008. Business sector unites against longer blackout. South China Morning Post, December 29, 2008.
Manne, H.G., 1966. Insider Trading and the Stock Market. Free Press, New York, NY.
Meulbroek, L.K., 1992. An empirical analysis of illegal insider trading. J. Financ. 47, 1661–1699.
Piotroski, J.D., Roulstone, D.T., 2005. Do insider trades reflect both contrarian beliefs and superior knowledge about future cash flow realizations? J. Account.Econ. 39, 55–81.
Rock, E., 2002. Securities regulation as lobster trap: a credible commitment theory of mandatory disclosure. Cardozo Law Rev. 23, 675–704.
Roulstone, D.T., 2003. The relation between insider-trading restrictions and executive compensation. J. Account. Res. 41, 525–551.
Sefcik, S.E., Thompson, R., 1986. An approach to statistical inference in cross-sectional models with security abnormal returns as dependent variable. J. Account. Res. 24, 316–334.
Seyhun, N.H., 1992. The effectiveness of the insider-trading sanctions. J. Law Econ. 35, 149–182.
Seyhun, N.H., 1998. Investment intelligence from insider trading. MIT Press, Cambridge, MA.
Shleifer, A., Wolfenzon, D., 2002. Investor Protection and Equity Markets. J. Financ. Econ. 66, 3–27.
Scent, B., 2008. HKEx hit with complaints over 'blackout' extension. The Standard, 19 December 2008.
Stamp, M., Jaggers, T., 2005. International Insider Dealing. City & Financial Publication.
Webb, D., 2008. Poll results: Insider dealing blackout." 2008. Available at http://webb-site.com/vote/result.asp?p=25.
Webb, D., 2009. "The HKICS blackout study." 8th June 2009. Available at https://webb-site.com/articles/blackoutstudy.asp.
White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–838.
Williams, R., 2008. "Combined consultation paper and quarterly reporting: Opening remarks by Richard Williams, Head of Listing, HKEx to the Legislative Council Financial Affairs Panel." 30 December 2008.
Zhang, I., 2007. Economic consequences of the Sarbanes-Oxley Act of 2002. J. Account. Econ. 44, 74–115.
Zhu, C., Wang, L., 2015. Insider trading under trading ban regulation in China's A-share market. China J. Account. Res. 8 (3), 169–191.

China Journal of Accounting Research 19 (2026) 100503

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Digital transformation, information advantage and managerial opportunism

Yan Li [a], Yangyang Liu [b], Jing Dai [a,*], Hong Luo [a]

[a] School of Accounting, Southwestern University of Finance and Economics, Chengdu, China
[b] School of Finance, Southwestern University of Finance and Economics, Chengdu, China

ARTICLE INFO

Article history:
Received 29 October 2025
Accepted 28 May 2026
Available online 3 June 2026

Keywords:
Digital transformation
Information advantage
Managerial opportunism
Insider trading

ABSTRACT

Digital transformation may give managers an information advantage that facilitates opportunistic behavior. We hypothesize that this enables managers to achieve excess returns from insider trading, increasing their engagement in such activities. We find that digital transformation significantly increases insider trading and pledging volumes, particularly for insider sales. When firms implement digital transformation, managers' insider trading is positively associated with stock returns and financial performance. Managers can thus achieve significantly higher long-term excess returns from insider selling, increasing trading volumes. The effect is pronounced for firms with stronger management opportunism and less transparent information environments. Our findings suggest that digital transformation amplifies managerial information advantages, enabling more profitable and well-timed managerial opportunism; this underscores the need for stronger regulatory oversight.

© 2026 The Authors. Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

"A former top Equifax executive was charged on Wednesday with insider trading for selling nearly $1 million in firm stock after he learned about a major data breach in 2017 but before it was publicly announced." —The New York Times[1]

The rise of the digital economy marks a strategic shift, creating opportunities to leverage technological advancements and drive industrial transformation. We define digital transformation as a firm-wide process

* Corresponding author at: Room 615 Chengzheng Building, 555 Liutai Road, Sichuan, China.
E-mail addresses: yanli@swufe.edu.cn (Y. Li), liuyanyang@swufe.edu.cn (Y. Liu), daijing@swufe.edu.cn (J. Dai), luohong@swufe.edu.cn (H. Luo).

[1] https://www.nytimes.com/2018/03/14/business/equifax-executive-insider-trading.html.

https://doi.org/10.1016/j.cjar.2026.100503
1755-3091/© 2026 The Authors. Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

of organizational change that is triggered and shaped by the widespread diffusion of digital technologies and that fundamentally reconfigures firms' value propositions, business processes, organizational structures and data governance (Bharadwaj et al., 2013; Vial, 2019; Hanelt et al., 2021; Verhoef et al., 2021). Research increasingly highlights the role of digital transformation in promoting high-quality corporate growth by reshaping firm capabilities (Wu et al., 2021; Gaglio et al., 2022; Chen and Srinivasan, 2023). Consistent with the broader governance literature, which emphasizes how data and technology redefine information and information asymmetry in data-intensive firms (Jiang and Li, 2024; Li et al., 2024), digital transformation facilitates innovation in how firms collect, process and utilize information, empowering managers with enhanced decision-making capabilities and deeper operational insights. These advancements enable businesses to optimize their resource allocation, improve their efficiency and enhance their overall performance, thereby fostering sustainable growth (Liu et al., 2011; Ferreira et al., 2019).

However, the rapid development of digital transformation also presents challenges for corporate governance. Technological advancements can be a "double-edged sword" for managers: while digital transformation can equip them with more information, enabling them to improve their operations and decision-making, it can also confer information advantages that enable self-serving actions. Managers might use their enhanced access to information to engage in opportunistic behavior, potentially disadvantaging small and medium-sized investors. Therefore, to investigate whether managers exploit the information advantages provided by digital transformation, we examine whether digital transformation affects managerial opportunistic behavior.

Our study asks whether firms in emerging markets are particularly susceptible to opportunistic behavior, with a specific focus on China. Two key factors make firms in these markets particularly vulnerable to such behaviors. First, the rapid growth of the digital economy in emerging markets has made them some of the largest data producers globally, with China being a prominent example (McKinsey Global Institute, 2017). In response, governments in these markets, including China, have implemented various policies to promote digital transformation (Wen et al., 2022; Wu et al., 2022). Second, firms in emerging markets often face substantial challenges due to relatively weak internal and external institutional environments (Jiang and Kim, 2020), which result in significant information asymmetry and agency problems (Piotroski et al., 2015; Bradshaw et al., 2019). In markets like China's A-share market, where investor protection is relatively weak (Allen et al., 2005), managers' self-interested behaviors, such as insider selling and share pledging, have been widely criticized. These challenges are not unique to China but are prevalent in many emerging economies, where weaker governance structures and rapid technological advances may create new opportunities for managerial opportunism.

Substantial evidence from the literature suggests that digital transformation furnishes management with novel information advantages. By integrating internal and external big data and applying advanced data analytics techniques, digital transformation generates additional information that supports corporate decision-making (Porfírio et al., 2021) and, in line with Jiang and Li (2024), enhances the strength and timeliness of managers' information relative to that of outside investors in data-intensive firms. Simultaneously, digital transformation can expand and complicate corporate business and revenue models, potentially exacerbating information asymmetry between insiders and outside investors (Vaska et al., 2021). As a result, the information advantages granted to management through digital transformation may facilitate managers' opportunistic arbitrage behavior. Moreover, digital technologies enable the real-time and efficient collection of vast amounts of corporate data, which, when standardized through unified coding, enhance information timeliness and provide management with more immediate insights into corporate performance (Liu et al., 2011). This not only allows managers to achieve higher excess returns but also grants them greater flexibility in timing their opportunistic actions, thereby facilitating the concealment of their transactions. Consequently, digital transformation poses significant challenges for external investors in identifying the motivations behind such behavior and for regulators in detecting illegal opportunistic activities. As a result, insiders may exploit the information advantages provided by digital transformation to engage in more self-interested behavior, ultimately harming the interests of small and medium-sized investors.

While the above argument predicts a positive association between digital transformation and managerial opportunism, some research indicates that digital technology improves transparency for external investors (Wu et al., 2021), which in turn may discourage managers from engaging in opportunistic behavior. Therefore, it remains an open question whether digital transformation influences informed trading.

We infer managerial opportunism from both daily trading and rare and large-scale pledging behavior. Studies indicate that management may exploit their information advantage to engage in insider trading, securing excess returns (Rozeff and Zaman, 1998; Ke et al., 2003; Piotroski and Roulstone, 2005). A potentially more severe type of opportunistic behavior is share pledging, where managers leverage their informational advantage by pledging shares in anticipation of declines in firm value (Yeh et al., 2003; Kao et al., 2004; Li et al., 2023b). Such opportunistic pledging not only harms the interests of outside investors (Dou et al., 2019) but also increases the risk of bad debt for banks.[2]

Using a sample of firms listed on the Shanghai and Shenzhen stock exchanges from 2007 to 2021, our study finds that the degree of digital transformation is significantly and positively associated with managerial trading scale and transaction amounts and the frequency and volume of insider share pledging. This self-interested effect primarily manifests in selling behavior, with a one standard deviation increase in digitalization levels associated with approximately a 2.32 million yuan increase in selling amounts. The main result is more pronounced for firms whose managers are more self-interested and those that operate in less transparent information environments. Further decomposing digital transformation, we show that this effect is driven by technologies related to data generation, integration and analytics, such as artificial intelligence, big data and cloud computing; blockchain and other application-oriented technologies play a much weaker role. This pattern is consistent with the view that information-processing and analytical capabilities are the primary channels through which digital transformation enhances insiders' information advantages. Our main finding still holds after utilizing propensity score matching (PSM), employing instrumental variable (IV) methods and conducting other robustness checks.

Furthermore, we investigate whether the main result is driven by the information advantage that digital transformation provides to managers. We find that insider trading is negatively and significantly associated with stock returns and financial performance in firms with high levels of digital transformation, indicating that digital transformation enhances insiders' ability to trade based on superior private information. Consistent with the information advantage channel, we find that digital transformation significantly increases managerial long-term excess returns from insider trading. Further investigation reveals that digital transformation does not significantly enhance the short-term excess returns of insider selling. This suggests that timely information collection enables managers to time trades to avoid revealing their intention to reduce holdings, making it more difficult for the market to detect opportunistic behavior. Consistent with this timing-based interpretation, using the 2017 tightening of insider selling regulations in China as an exogenous shock, we further find that insider selling increases to a greater extent in more digitally transformed firms after the regulatory change. This indicates that digital technologies help managers to sustain opportunistic trading even under heightened regulatory scrutiny.

As we use information disclosed in financial reports to measure corporate digital transformation, an alternative explanation is that managers engage in insider trading by leveraging digital information disclosure to hype up stock prices (Peng et al., 2023). To rule out this strategic disclosure effect, we examine whether digital transformation influences insider selling before the annual report disclosure. We find that digital transformation increases insider selling in the window between the end of the fiscal year and the date of annual report announcement, suggesting that insiders exploit their information advantage regardless of strategic disclosure. The main results still hold after removing forward-looking disclosures on digital transformation and controlling for textual characteristics, further eliminating the disclosure effect.

Our study makes important contributions to the literature. First, it contributes to the digital transformation literature by highlighting the dual nature of digital transformation in corporate governance. The literature focuses on how digital transformation and the application of digital technologies enhance external investor

[2] For example, LeEco's (300104, now delisted) founder Jia Yueting pledged a large amount of his equity to financial institutions such as banks and securities firms to obtain substantial funds to support the company's business expansion. However, LeEco's operational performance sharply declined and its stock price plummeted. Jia Yueting was unable to add collateral or repay the loans, resulting in significant risks for these financial institutions. China Merchants Bank applied to the court for asset preservation in June 2017, freezing LeEco's assets, worth approximately 1.237 billion yuan. The largest single lender of LeEco's pledged shares, Western Securities, had already recorded an impairment loss of 973 million yuan.

access to information (Wu et al., 2021). Few studies note that digital transformation also significantly improves managers' ability to access internal information or examine whether the actual information asymmetry is exaggerated or mitigated. Our study provides empirical evidence that deepens understanding of the potential negative economic consequences of digital transformation due to increased information asymmetry from the perspective of managerial opportunism. Additionally, our paper extends the growing body of research on big data by applying big data analysis to emerging markets, which have relatively weak corporate governance. In contrast with studies like Zhu (2019), which focus on more developed markets, our findings offer important insights into how rapid technological advancements in emerging economies may exacerbate governance challenges, providing managers with opportunities for self-serving behavior.

Second, our paper enriches the literature on the sources of insider trading and pledging. Due to the covert nature of insider trading, identifying the information advantage it confers requires attention to unique scenarios, such as insider trading prosecutions that track the flow of information from executives to analysts and managers (Ahern, 2017), insiders' advance knowledge of potential non-standard audit opinions (Arif et al., 2022) and insiders' exploiting advance knowledge of government intervention during the 2008 financial crisis through the Troubled Asset Relief Program (Jagolinzer et al., 2020). Our study identifies a new, important and prevalent scenario in the digital era, wherein digital technology applications provide managers with information advantages that drive them to engage in self-interested trading. This deepens our understanding of the information advantages underlying insider opportunistic trading, providing valuable insights for regulators and market participants into strategies for curbing such behavior. Moreover, in contrast with Yang and Xie (2024), who focus only on insider selling, our paper considers a broader scope of managerial opportunistic behavior, including both regular market trading and the less frequent but more damaging practice of share pledging. This study also provides evidence of the information advantage facilitated by digital transformation while ruling out major alternative explanations, namely the disclosure effect.

Third, our study offers a novel perspective on the interaction between digital transformation and market regulation in emerging economies. While much of the literature focuses on the positive impacts of digital transformation in advanced markets (Müller et al., 2018; Ghasemaghaei and Calic, 2019; Chen and Srinivasan, 2023), our research emphasizes the complex regulatory challenges posed by the increased access to information that digital technologies enable. We argue that in emerging markets, where regulatory frameworks may be less robust, the enhanced information available to managers could lead to strategic exploitation. In particular, our study highlights how digital transformation can complicate the detection of managerial opportunism, such as insider trading and share pledging, by providing managers with more sophisticated tools for information manipulation. This suggests that regulators in emerging markets need to evolve their oversight mechanisms to address the new risks introduced by digital technologies and prevent abuse by informed insiders.

This study also contributes to the literature by distinguishing firm-level digital transformation from traditional enterprise systems such as enterprise resource planning (ERP) systems. Prior ERP research mainly focuses on how enterprise systems improve transactional efficiency, internal control and firms' internal information environment through process integration and standardization (e.g., Gattiker and Goodhue, 2005; Dorantes et al., 2013; Tian and Xu, 2015). In contrast, we conceptualize digital transformation as a higher-order strategic change that enhances firms' data integration and forward-looking analytical capabilities using advanced digital technologies (Bharadwaj et al., 2013; Vial, 2019). We show that such digital transformation reshapes information asymmetry between insiders and external investors, leading to more pronounced insider selling and higher long-term trading profits.

2. Hypothesis development

In line with the information systems and strategy literature, we define digital transformation as a process of firm-wide organizational change that is triggered and shaped by the diffusion of digital technologies and that fundamentally reconfigures how firms generate, process and utilize information for decision-making (Bharadwaj et al., 2013; Vial, 2019). Digital transformation goes beyond traditional IT-enabled automation

or isolated IT investments, such as the implementation of ERP systems, which primarily focus on transaction processing and internal integration. Instead, it reflects a higher-order strategic transformation in which firms leverage technologies such as cloud computing, big data analytics, artificial intelligence (AI) and digital platforms to enhance real-time data collection, integrate heterogeneous information sources and develop forward-looking analytical and predictive capabilities. From an economic perspective, the defining feature of digital transformation is not the adoption of specific technologies per se but the resulting shift in the firm's internal information environment, namely the timeliness, granularity and usability of information available to decision-makers, which may alter the distribution of information advantages between insiders and external investors.

Insider trading is often driven by the information advantage insiders hold over other market participants, enabling them to secure abnormal profits in the capital markets (Huddart and Ke, 2007; Cohen et al., 2012; Ali and Hirshleifer, 2017; Akbas et al., 2020; Akey et al., 2022). China's capital market, an emerging market in a transitioning economy, has relatively weak external regulatory frameworks and internal governance constraints, accompanied by significant agency problems among insiders (Zeng et al., 2018). Corporate management's exploitation of these information and control advantages for opportunistic trading, to the detriment of investors, has long been a source of concern for both researchers and capital market regulators.[3]

Digital technologies facilitate the real-time and efficient collection of vast amounts of data from both internal and external sources, which are subsequently standardized and integrated into usable information. Such data-intensive operations allow firms to transform raw data into valuable insights and often make them better informed than outsiders (Abis and Veldkamp, 2024; Jiang and Li, 2024). This process provides incremental information that supports managerial decision-making and offers prompt feedback for management (Liu et al., 2011; Tian and Xu, 2015; Pincus et al., 2017; Porfírio et al., 2021; Chen et al., 2024). For instance, ERP systems automate data collection and integration, delivering information to managers in a timely manner, thereby enhancing their ability to process and analyze accounting information (Davenport, 1998; Hitt et al., 2002; Brazel and Dang, 2008). Spathis (2006) finds that ERP systems improve the efficiency and effectiveness of enterprise operations and management. Dorantes et al. (2013) show that ERP systems improve the accuracy of management's earnings forecasts. Wu and Tian (2022) find that digital transformation meets management's need for timely and accurate information, which helps them to identify, adjust and integrate idle resources.

Moreover, the efficient collection and processing of information further exacerbate the information timing gap between internal managers and external investors in accessing critical data (Brazel and Dang, 2008), thereby enhancing managers' ability to engage in strategic trading. Real-time data integration and predictive analytics give firms greater discretion over the timing between information acquisition and mandatory disclosure (Jiang and Li, 2024), which directly strengthens managers' ability to engage in opportunistic insider trading based on superior information. This temporal disparity presents significant challenges for external investors in identifying the motivations behind insider trading and for regulators in detecting such activities. As a result, the costs of regulatory violations are reduced, potentially incentivizing insiders to exploit their information advantage for personal profit.

Digital transformation can also complicate firms' business and profit models (Vaska et al., 2021). Digital technologies accelerate change, leading to significant transformations in corporate operation (Dal Mas et al., 2020) and value networks (Gray et al., 2013). Furthermore, digital transformation facilitates corporate expansion not only physically, through internationalization (Brieger et al., 2022; Jiang and Wang, 2024), but also virtually, through online business, which reshapes and blurs traditional revenue streams. These developments increase the volume and complexity of information that managers must process and evaluate. While external investors may gain access to more information through digital tools, internal managers typically

[3] In 2017, the China Securities Regulatory Commission (CSRC) introduced a mandate requiring insiders of all publicly listed firms to submit a sales plan to the CSRC and publicly disclose it prior to selling shares in their respective companies. Ye et al. (2025) find that although the "sell-by-plan" mandate can reduce opportunistic selling activities, insiders can circumvent this mandate by strategically initiating their plans before the earnings announcement. Thus, management has strong economic incentives to engage in self-interested trading by leveraging their information advantage.

possess a more nuanced understanding of these data and their strategic implications, which gives them an information advantage over external investors. Consequently, digital transformation exacerbates information asymmetry between insiders and outsiders, increasing the profit opportunities available to insiders.

In summary, the improvement in information availability due to digital transformation amplifies management's information advantage over external investors, thereby increasing their trading opportunities and fostering opportunistic insider behavior. Accordingly, we test the following hypothesis:

H1a: The degree of digital transformation is positively associated with the scale of insider trading.

A similar argument applies to a less frequent but more detrimental scenario, in which insiders exploit their information advantage to secure capital gains through share pledging. In practice, when shareholders of listed firms reduce their holdings directly in the secondary market, they are subject to restrictions imposed by laws, regulations and corporate bylaws, such as limitations on selling ratios and mandatory information disclosure. Share pledging, however, is widely favored by shareholders of listed firms due to its accessibility and flexibility, providing a significant financing channel (He et al., 2022). This mechanism allows insiders to circumvent stringent selling restrictions and regulations, effectively enabling implicit cash-outs (Lu et al., 2022). Thus, this study tests the following hypothesis:

H1b: The degree of digital transformation is positively associated with the scale of insider share pledging.

Nevertheless, the literature suggests that digital transformation increases the information available to external investors through enhanced disclosure and increased attention (Wu et al., 2021). Moreover, technology-driven data generation can narrow information gaps when external "alternative data" are effectively processed by sophisticated intermediaries (Jiang and Li, 2024). These arguments suggest that richer information and stronger monitoring may discipline insiders and curb opportunistic behavior under certain conditions.

The key point in our setting is that digital transformation expands the information set available within the firm. Whether this expanded internal information set ultimately benefits outside shareholders or facilitates insider opportunism depends on managers' incentives and how they use—and choose to disclose—information. When managers use enhanced information to improve their decision-making and align with shareholder interests, digital transformation can generate value. However, when managers have incentives and opportunities to pursue private benefits, their superior access to and interpretation of information—together with discretion over disclosure timing—can create more scope for opportunistic insider trading and share pledging. Therefore, although the net effect is ultimately an empirical question, our hypotheses (H1a and H1b) focus on the opportunism channel that is central to this study (Fig. 1)).

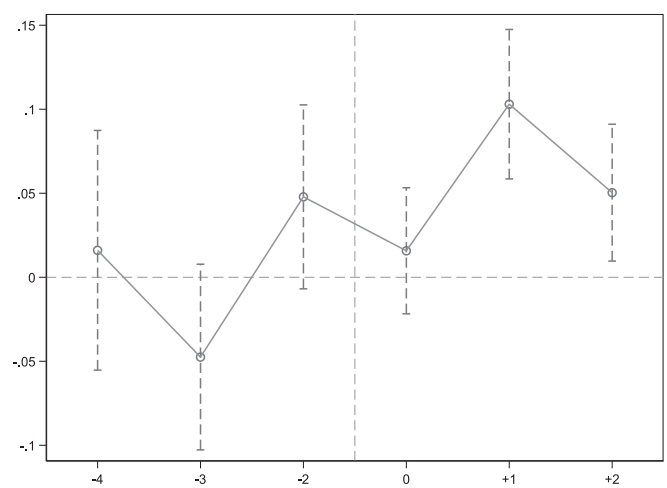


Fig. 1. Event-Study Coefficients of Dynamic Continuous DID: The Time-Varying Effect of Digitalization on Insider Trading (Relative to the Treatment Period). Note: This figure plots the estimated coefficients (and 95% confidence intervals, represented by the vertical error bars) from the dynamic continuous DID event-study specification.

3. Research design

3.1. Sample and data

We obtain financial data from the CSMAR database, insider trading data from the CNRDS database, share pledging data from Eastmoney (Eastmoney.com) and annual reports from cninfo.com.cn. The initial sample consists of all firms listed on the Shanghai and Shenzhen stock exchanges from 2007 to 2021. Our initial sample consists of 34,547 firm-year observations. We exclude 696 firm-years that belong to the financial industry. We further remove observations based on the following criteria: (1) 294 firm-years with an asset-liability ratio greater than 1; (2) 2562 firm-years within the first year following their initial public offering; (3) 50,399 insider trading events triggered by lock-up period expirations; and (4) 1376 firm-years with missing control variables or key financial indicators. The final sample consists of 29,619 firm-year observations from 2007 to 2021. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Table 1 presents the variable definitions.

3.2. Measurement of digital transformation

Following Chen and Srinivasan (2023), we measure the degree of digital transformation using a dictionary-based textual analysis approach. We adopt the corporate digital transformation dictionary developed in prior studies (Wu et al., 2021; Li et al., 2023a; Zhou and Li, 2023), as shown in Table A1. Specifically, we carry out the following steps. (1) We download annual reports of A-share listed firms from cninfo.com.cn. (2) We use Python to perform text segmentation with the pkuseg tool, developed by Peking University. Pkuseg is based on a conditional random field model and incorporates the adaptive dual feature training method, which improves segmentation accuracy and training efficiency across diverse corpora. (3) We remove stop words and exclude digital transformation keywords appearing in negated contexts. (4) We compute the frequency of digital transformation keywords in each report and apply a logarithmic transformation to construct a continuous index capturing the intensity of a firm's digital transformation.

3.3. Empirical model

We estimate the following model to examine H1a and H1b:

$$InsiderTrading/Pledging_{i,t} = \alpha + \beta_1 Digital_{i,t-1} + \gamma Controls_{i,t-1} + \sum Industry \times Year + \varepsilon_{i,t} \quad (1)$$

where the scale of insider trading is measured using trading proportion and amount, including total buy-and-sell transactions (*bs_share_manager*, *bs_money_manager*), buying transactions (*buy_share_manager*, *buy_money_manager*) and selling transactions (*sell_share_manager*, *sell_money_manage*r). We aggregate all insider trading transactions conducted by senior management, including their immediate family members, at the firm-year level.[4,5] Following Zeng et al. (2018) and Luo and Huang (2020), we assign a value of zero to the dependent variable in firm-years for which no insider trading activity is observed. As in Sun and Liu (2021), we measure the scale of insider equity pledge behavior by the frequency of pledged transactions (*Quantity of pledged deals*) and the ratio of shareholders' pledged shares to total share capital (*Pledge ratio*). Our variable of interest, *Digital*, measures the degree of a firm's digital transformation, with a higher value indicating greater digital transformation.

Following Zeng et al. (2018) and Luo and Huang (2020), we include additional determinants of insider trading. Firm-level characteristics include return on assets (*ROA*), leverage ratio (Lev), firm size (*Size*) and management ownership (*Mngmhldn*), as well as governance characteristics: state ownership (*SOE*),

[4] Individuals who both hold more than 5% of the firm's shares and concurrently serve as directors or senior executives are also categorized as part of management.

[5] To avoid distortions from non-discretionary transactions, we exclude insider trades associated with the expiration of share lock-up periods. This adjustment ensures that our measures reflect discretionary trading behavior by management, rather than mandated divestitures or routine liquidity needs.

Table 1
Variable definition table.

1. Dependent Variables	
bs_share_manager	The sum of shares bought and sold by all directors, supervisors, and senior managers (including relatives) within the fiscal year divided by the total number of A-shares in circulation for the year, multiplied by 100 for ease of interpretation.
buy_share_manager	The sum of shares bought by all directors, supervisors, and senior managers (including relatives) within the fiscal year divided by the total number of A-shares in circulation, multiplied by 100 for ease of interpretation.
sell_share_manager	The sum of shares sold by all directors, supervisors, and senior managers (including relatives) within the fiscal year divided by the total number of A-shares in circulation, multiplied by 100 for ease of interpretation.
bs_money_manager	The total amount of money involved in the buying and selling of shares by all directors, supervisors, and senior managers within the fiscal year, unit: ten million yuan.
buy_money_manager	The total amount of money involved in the shares bought by all directors, supervisors, and senior managers within the fiscal year, unit: ten million yuan.
sell_money_manager	The total amount of money involved in the shares sold by all directors, supervisors, and senior managers within the fiscal year, unit: ten million yuan.
Pledge ratio	The ratio of the number of shares pledged by shareholders to the total share.
Quantity of pledged deals	The number of shareholder equity pledge transactions of the firm within one year
CAR	Cumulative abnormal returns using Carhart four-factor model[a]. The estimation period is [−252, −5 days], excluding samples with an estimation period less than 100 days. For ease of interpretation, the value is multiplied by 100.
BHAR	The buy-and-hold abnormal returns from the month of the transaction of the management to the 1/3/6/9/12th month, multiplied by 100 for ease of interpretation.
2. Explanatory Variables	
Digital	Frequency of keywords related to digital transformation in annual reports, logarithm taken after adding 1.
3. Control Variables	
ROA	Return on Assets.
Lev	Liabilities-to-assets ratio.
Size	Logarithm of market capitalization at the year-end.
Ins_hold	Institutional investor shareholding percentage
SOE	Whether the firm is state-owned enterprise, yes is assigned 1, no is 0.
Big4	Whether the auditor comes from one of the Big Four, yes is assigned 1, no is 0.
Turnover	Annual turnover rate of circulating shares.
Mngmhldn	Proportion of shares held by the management, the ratio of shares held by directors and supervisors to the total number of shares.
Top1	The percentage of shares owned by the firm's largest shareholder.
Boardsize	The natural logarithm of the total number of board members.
AnaAttention	The natural logarithm of the count of analysts (or analyst teams) actively following the firm over the year, incremented by one to enable the logarithmic transformation.

[a] The empirical results obtained from estimating expected returns using the Fama-French three-factor model are essentially consistent with those obtained from the Carhart four-factor model.

Table 2
Summary statistics.

1. Dependent variables

Variable	N	Mean	SD	Min	Median	Max
bs_share_manager	29,619	0.428	1.322	0	0	8.255
buy_share_manager	29,619	0.043	0.204	0	0	1.557
sell_share_manager	29,619	0.370	1.241	0	0	7.911
bs_money_manager	29,619	2.057	6.946	0	0	48.206
buy_money_manager	29,619	0.194	0.910	0	0	7.000
sell_money_manager	29,619	1.768	6.426	0	0	45.172
Pledge ratio	20,290	11.845	14.098	0	6.165	56.370
Quantity of pledged deals	20,290	26.025	37.406	0	12	216
CAR[0,5]_sell	70,663	0.007	6.875	−16.949	−0.616	22.914
BHAR6m_sell	71,884	5.413	36.377	−55.730	−1.623	144.616
CAR[0,5]_opposell	5076	−0.049	6.760	−16.766	−0.686	21.668
BHAR6m_opposell	5178	4.913	35.034	−57.206	−1.581	129.612

2. Independent variables

Variable	N	Mean	SD	Min	Median	Max
Digital	29,619	1.159	1.355	0	0.693	4.997

3. Control variables

Variable	N	Mean	SD	Min	Median	Max
ROA	29,619	0.042	0.061	−0.211	0.039	0.223
Lev	29,619	0.435	0.205	0.055	0.43	0.894
Size	29,619	22.1	1.277	19.681	21.921	26.059
Ins_hold	29,619	0.712	0.672	0.006	0.55	3.921
SOE	29,619	0.39	0.488	0	0	1
Big4	29,619	0.058	0.233	0	0	1
Turnover	29,619	556.639	492.553	25.957	409.53	2569.703
Mngmhldn	29,619	11.803	18.95	0	0.126	67.515
Top1	29,619	34.781	14.888	8.77	32.72	74.3
Boardsize	29,619	2.14	0.199	1.609	2.197	2.708
AnaAttention	29,619	1.449	1.182	0	1.386	3.784
Nshra[a]	29,619	755,605,367	1,348,819,410	24,700,000	358,143,846	9,706,677,295

[a] The count of A-shares in circulation is essential for the subsequent economic analysis and calculations presented in this study.

institutional ownership (*Ins_hold*), auditor quality (*Big4*), analyst coverage (*AnaAttention*), top shareholder ownership percentage (*Top1*), board size (*Boardsize*) and stock turnover rate (*Turnover*). All explanatory variables are lagged by one period to alleviate potential endogeneity concerns. The model includes industry-by-year fixed effects to control for time-varying industry shocks. Standard errors are clustered at the industry level.[6] All variable definitions are provided in Table 1.

3.4. Summary statistics

Table 2 presents the descriptive statistics of the main variables. The mean of the degree of digital transformation (*Digital*) is 1.159, with a median of 0.693 and a standard deviation of 1.355, indicating significant variation in the degree of corporate digital transformation. The average proportion of insider trading is 0.428%, with an average transaction amount of 20.47 million yuan. The average proportion of selling is 0.370%, with an average transaction amount of 17.59 million yuan. The average proportion of buying is 0.042%, with an average transaction amount of 1.90 million yuan. These statistics are consistent with prior research (Ye et al., 2025). Overall, these figures suggest that insider trading activity among management in Chinese listed

[6] As a robustness check, we confirm that our results remain qualitatively similar when clustering standard errors at the firm level.

firms is economically significant, with sales transactions occurring at a much larger scale than purchases. This pattern is consistent with the assumption that equity-based compensation is a major component of managerial pay (Jensen and Murphy, 1990; Hall and Liebman, 1998; Ofek and Yermack, 2000; Nienhaus, 2022).

4. Empirical results

4.1. Baseline results

4.1.1. Digital transformation and insider self-interested behavior

Table 3 presents the main results for the relationship between digital transformation and insider self-interested behavior. Columns (1)–(3) report the results using different measures of insider trading volume as dependent variables. The coefficients on *Digital* are positive and statistically significant for total trading proportion (coefficient = 0.037, t-statistic = 4.224), insider buying (coefficient = 0.003, t-statistic = 2.789) and insider selling (coefficient = 0.033, t-statistic = 3.867). The coefficient on *Digital* in Column (3) is similar in magnitude to that in Column (1), suggesting that the overall increase in transaction volume is primarily driven by insider selling activity. Economically, a one standard deviation increase in digital transformation is associated with a 0.045 percentage point increase in the insider selling proportion, which translates into approximately 340,000 additional shares sold based on the sample's average number of tradable shares ($0.045/100 \times 755605367$). This represents a 12.16% increase relative to the average insider selling proportion. Columns (4)–(6) report the results using different measures of insider trading amounts as dependent variables and present qualitatively similar findings to those based on trading volumes.[7] Columns (7)–(8) report the results for insider share pledging. The coefficient on *Digital* is positive and statistically significant for both the pledge ratio (coefficient = 0.331, t-statistic = 2.469) and the number of pledge transactions (coefficient = 1.242, t-statistic = 4.126). Economically, a one standard deviation increase in digital transformation is associated with 1.68 additional pledge transactions, representing a 14% increase relative to the median frequency of pledging. A one standard deviation increase in digital transformation is associated with a 0.449 percentage point increase in the pledge ratio, representing a 7.28% increase relative to the median level. Taken together, our findings support H1 and indicate that higher levels of digital transformation are associated with higher levels of both insider trading and pledging behavior.[8]

Unlike studies based on U.S. data, where managerial purchases are commonly associated with abnormal returns (Lakonishok and Lee, 2001; Jeng et al., 2003), research in the Chinese context shows that abnormal returns are more likely to follow managerial sales (Huang et al., 2024; Ye et al., 2025). Consistent with this pattern, Table 3 shows that the effect of digital transformation on insider trading is driven primarily by selling activity. Because insider selling can harm investor interests, regulators in China have increasingly focused on strengthening oversight and standardizing sales practices.[9] Reflecting this emphasis, the remainder of the analysis focuses on insider selling.

[7] Economically, a one standard deviation increase in digital transformation is associated with a 2.86 million yuan increase in total insider trading amount, representing a 13.9% increase relative to the sample average. Similarly, the corresponding increase in insider selling amount is 2.32 million yuan, equivalent to a 13.12% rise relative to the average insider selling amount.

[8] To mitigate potential endogeneity problems, this study employs robustness checks such as firm fixed effects analysis, propensity score matching and instrumental variable tests. To ensure the robustness and reliability of the fundamental conclusions, further robustness tests are conducted by altering the measurement of key variables (Table A4); controlling for industry characteristics and firm business effects; and modifying model specifications and adding control variables.

[9] To strengthen oversight of insider sales, the CSRC has issued a series of regulatory measures. On 7 January 2016, the CSRC released the Regulations on the Selling of Shares by Major Shareholders and Directors and Senior Management of Listed Companies. These rules were revised and reissued on 27 May 2017. More recently, on 27 August 2023, the CSRC proposed additional requirements to further standardize insider selling practices. On 26 September 2023, the Shanghai and Shenzhen stock exchanges each issued a Notice on Further Standardizing Matters Related to Share Selling, while the Beijing Stock Exchange released updated guidelines on share selling and shareholding management. Together, these actions reflect ongoing efforts to tighten the regulatory framework around insider share sales.

Table 3
Digital transformation and insider self-interest behavior.

Variable	Trading volume			Trading amount			Share pledging	
	Buy and sell	Buy	Sell	Buy and sell	Buy	Sell	*Pledge ratio*	*Quantity of pledged deals*
	bs_share_manager	*buy_share_manager*	*sell_share_manager*	*bs_money_manager*	*buy_money_manager*	*sell_money_manager*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Digital	**0.037*** **	**0.003*** **	**0.033*** **	**0.213*** **	**0.024*** **	**0.172*** **	**0.331** **	**1.242*** **
	(4.224)	**(2.789)**	**(3.867)**	**(4.019)**	**(4.233)**	**(3.528)**	**(2.469)**	**(4.126)**
ROA	−0.905***	−0.004	−0.879***	−0.360	0.023	−0.673	−19.108***	3.388
	(−4.865)	(−0.168)	(−5.126)	(−0.375)	(0.241)	(−0.732)	(−5.062)	(0.596)
Lev	−0.196***	0.008	−0.218***	−0.600*	0.078*	−0.820***	10.049***	9.297***
	(−2.939)	(1.038)	(−3.434)	(−1.907)	(1.988)	(−2.744)	(9.669)	(3.615)
Size	−0.065***	−0.008***	−0.051***	0.909***	0.051***	0.759***	−0.261	14.519***
	(−4.288)	(−4.234)	(−3.813)	(5.860)	(5.176)	(5.453)	(−0.952)	(21.762)
Ins_hold	−0.006	0.001	−0.013	0.489***	0.029***	0.417***	1.666***	3.346***
	(−0.505)	(0.623)	(−1.071)	(6.625)	(2.751)	(6.622)	(5.521)	(8.115)
SOE	−0.161***	−0.035***	−0.113***	−1.031***	−0.192***	−0.716***	−13.416***	−6.463***
	(−6.544)	(−10.386)	(−5.363)	(−6.750)	(−11.532)	(−5.300)	(−21.676)	(−6.619)
Big4	−0.017	−0.007	−0.011	−0.917***	−0.088***	−0.752***	−3.809***	1.813
	(−0.866)	(−1.594)	(−0.600)	(−3.466)	(−3.422)	(−3.206)	(−3.707)	(0.863)
Turnover	−0.000***	−0.000*	−0.000***	−0.001***	−0.000***	−0.001***	−0.002***	0.003***
	(−2.928)	(−1.718)	(−2.762)	(−5.480)	(−3.862)	(−5.130)	(−7.412)	(4.038)
Mngmhldn	0.024***	0.001***	0.022***	0.105***	0.004***	0.098***	0.028	0.205***
	(25.375)	(6.927)	(27.341)	(15.359)	(5.981)	(15.687)	(1.625)	(9.306)
Top1	−0.003***	−0.000***	−0.002***	−0.032***	−0.003***	−0.027***	0.026*	−0.259***
	(−3.698)	(−3.043)	(−3.120)	(−5.914)	(−4.404)	(−5.280)	(1.772)	(−9.646)
Boardsize	0.105**	0.012**	0.087*	0.284	0.035	0.181	−2.456***	0.248
	(2.046)	(2.107)	(1.820)	(1.040)	(1.350)	(0.716)	(−3.754)	(0.145)
AnaAttention	0.046***	0.007***	0.039***	0.252***	0.029***	0.247***	−1.001***	−1.945***
	(4.076)	(5.993)	(3.583)	(3.728)	(5.643)	(3.705)	(−5.596)	(−6.751)
Constant	1.126***	0.144***	0.909***	−12.320***	−0.630***	−10.182***	21.965***	−196.989***
	(4.529)	(5.092)	(4.046)	(−5.474)	(−3.765)	(−5.168)	(5.487)	(−18.791)
Industry × Year	YES	YES	YES	YES	YES	YES	YES	YES
Adj R^2	0.185	0.061	0.179	0.151	0.072	0.144	0.248	0.482
Observations	29,619	29,619	29,619	29,619	29,619	29,619	20,290	20,290

Note: This table examines whether digital transformation significantly increases insiders' self-interested behavior. The specification follows Eq. (1),and the dependent variable is the scale of insider trading, measured from two dimensions: trading proportion and trading amount, including overall buy and sell transactions (*bs_share_manager, bs_money_manager*), buying transactions (*buy_share_manager, buy_money_manager*), and selling transactions (*sell_share_manager, sell_money_manager*). And the other dependent variable is the scale of insider equity pledging behavior, including the number of pledged transactions (*Quantity of pledged deals*) and the ratio of pledged shares (*Pledge ratio*). The explanatory variable is the degree of corporate digital transformation (*Digital*). All regressions include the industry-by-year fixed effects. Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. ***p < 0.01,**p < 0.05,*p < 0.1.

4.1.2. Types of digital transformation and managerial opportunistic behavior

Given the broad scope of corporate digital transformation and the potentially heterogeneous ways in which it affects managerial information advantage, we further decompose digital transformation along both technology types and functional dimensions.

We first follow Wu et al. (2021) in decomposing digital transformation into five technology categories: AI, cloud computing (CC), big data (DT), blockchain (BD) and applied digital technologies (ADT). The results, reported in Table 4, show that the coefficients for AI, CC and DT are all positive and significant at the 1% level. These technologies are directly related to firms' capabilities in data generation, storage, integration and analysis, which enhance the timeliness, granularity and forward-looking nature of internal information. As a result, they expand insiders' informational advantage relative to outside investors and strengthen their ability to trade opportunistically.

In contrast, the coefficient on BD is nonsignificant. This result is consistent with the limited adoption of blockchain technologies among Chinese listed firms during our sample period: blockchain-related disclosures account for less than 4% of firm-year observations, which limits both statistical power and cross-sectional variation. Moreover, blockchain focuses on data security and immutability rather than on generating new forward-looking information and therefore does not materially enhance insiders' information advantage. The coefficient on ADT is also nonsignificant, suggesting that many downstream digital applications in China have not yet reached a level of maturity sufficient to materially change firms' internal information structure.

To further clarify how digital transformation creates an information advantage, we reclassify digital transformation into three mutually exclusive functional dimensions, namely technological, operational and administrative, and a separate data analytics sub-dimension, as shown in Table 4. The technological dimension captures foundational digital technologies and IT infrastructure, including cloud computing, data platforms, AI and computing architecture. This dimension is significantly and positively associated with insider selling, indicating that firms with stronger digital technological foundations have superior internal capabilities to generate, integrate and process information, thereby widening the information gap between insiders and external investors.

The operational dimension captures the digitalization of core business processes, such as production, logistics and customer service. When examined alone, operational digitalization is positively related to insider selling, suggesting that real-time process monitoring and automated execution provide insiders with high-frequency signals about firm performance. However, once data analytics is controlled for, the operational coefficient becomes nonsignificant. This pattern indicates that process automation by itself does not constitute an independent source of information advantage; instead, operational data must be collected, structured and analyzed through analytics systems before it becomes economically meaningful for insider trading.

The administrative dimension, which captures the digitalization of management, governance and market-facing activities, is also positively associated with insider selling. This result suggests that digital tools for decision support, customer management and financial and governance processes improve internal information visibility and coordination, enabling managers to more efficiently interpret firm-specific conditions and translate their informational advantages into trading decisions.

Finally, the data analytics dimension remains positive and significant even after controlling for the operational dimension. This highlights data analytics as the most direct and robust channel through which digital transformation enhances insiders' information advantage, by improving firms' ability to transform large volumes of high-frequency and unstructured data into forward-looking assessments of firm performance and risk.

Taken together, these results show that not all forms of digital transformation equally generate information advantages. Instead, technologies and applications related to data generation, integration and analytics—rather than encryption or routine automation—are the primary drivers of insiders' opportunistic trading behavior.

4.2. Mechanism analyses

This section examines the mechanism through which digital transformation influences insider selling, focusing on the role of information advantage. We first test whether digital transformation strengthens the predictive power of insider selling for negative future outcomes, as an indirect measure of insiders' information

Table 4
Digital transformation dimensions and insider selling.

Variable	*Digital Transformation by Technology Type*					*Digital Transformation by Functional Dimension*				
	(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)
AI	**0.053*** **					***Technological***	**0.036*** **			
	(3.506)						**(3.716)**			
DT		**0.049*** **				***Operational***		**0.034*** **		**0.016**
		(3.409)						**(3.431)**		**(1.157)**
CC			**0.052*** **			***Administrative***			**0.023** **	
			(3.629)						**(2.406)**	
BD				**0.006**		***DataAnalytics***				**0.033** **
				(0.242)						**(2.499)**
ADT					**0.014**					
					(1.203)					
ROA	−0.889***	−0.887***	−0.896***	−0.906***	−0.897***	*ROA*	−0.890***	−0.891***	−0.890***	−0.887***
	(−5.188)	(−5.271)	(−5.267)	(−5.315)	(−5.190)		(−5.181)	(−5.194)	(−5.161)	(−5.208)
Lev	−0.212***	−0.219***	−0.218***	−0.215***	−0.216***	*Lev*	−0.219***	−0.219***	−0.216***	−0.219***
	(−3.382)	(−3.381)	(−3.380)	(−3.398)	(−3.453)		(−3.402)	(−3.397)	(−3.444)	(−3.408)
Size	−0.049***	−0.050***	−0.049***	−0.046***	−0.048***	*Size*	−0.050***	−0.049***	−0.049***	−0.050***
	(−3.667)	(−3.667)	(−3.650)	(−3.417)	(−3.598)		(−3.757)	(−3.730)	(−3.623)	(−3.751)
Ins_hold	−0.011	−0.012	−0.011	−0.010	−0.011	*Ins_hold*	−0.012	−0.012	−0.011	−0.012
	(−0.957)	(−0.996)	(−0.958)	(−0.840)	(−0.920)		(−1.014)	(−0.999)	(−0.957)	(−1.031)
SOE	−0.115***	−0.114***	−0.117***	−0.116***	−0.115***	*SOE*	−0.114***	−0.114***	−0.114***	−0.114***
	(−5.484)	(−5.458)	(−5.527)	(−5.561)	(−5.382)		(−5.400)	(−5.408)	(−5.416)	(−5.408)
Big4	−0.014	−0.012	−0.013	−0.015	−0.014	*Big4*	−0.011	−0.012	−0.013	−0.012
	(−0.786)	(−0.642)	(−0.712)	(−0.799)	(−0.736)		(−0.611)	(−0.616)	(−0.701)	(−0.623)
Turnover	−0.000***	−0.000***	−0.000***	−0.000***	−0.000***	*Turnover*	−0.000***	−0.000***	−0.000***	−0.000***
	(−2.831)	(−2.789)	(−2.827)	(−2.797)	(−2.772)		(−2.807)	(−2.800)	(−2.763)	(−2.800)
Mngmhldn	0.022***	0.022***	0.022***	0.022***	0.022***	*Mngmhldn*	0.022***	0.022***	0.022***	0.022***
	(27.205)	(27.270)	(27.196)	(26.741)	(26.788)		(27.124)	(27.168)	(27.076)	(27.242)
Top1	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***	*Top1*	−0.002***	−0.002***	−0.002***	−0.002***
	(−3.161)	(−3.200)	(−3.162)	(−3.196)	(−3.161)		(−3.175)	(−3.174)	(−3.146)	(−3.187)
Boardsize	0.088*	0.088*	0.091*	0.088*	0.086*	*Boardsize*	0.089*	0.089*	0.086*	0.089*
	(1.821)	(1.840)	(1.857)	(1.801)	(1.770)		(1.845)	(1.833)	(1.774)	(1.852)
AnaAttention	0.040***	0.040***	0.040***	0.041***	0.041***	*AnaAttention*	0.039***	0.039***	0.040***	0.039***
	(3.663)	(3.672)	(3.741)	(3.778)	(3.692)		(3.677)	(3.675)	(3.653)	(3.643)
Constant	0.888***	0.895***	0.877***	0.861***	0.880***	Constant	0.892***	0.888***	0.894***	0.896***
	(3.971)	(3.990)	(3.917)	(3.833)	(3.931)		(3.997)	(3.982)	(3.963)	(4.001)
Industry × Year	YES	YES	YES	YES	YES	Industry × Year	YES	YES	YES	YES
Adj R^2	0.179	0.179	0.179	0.179	0.179	Adj R^2	0.179	0.179	0.179	0.179
Observations	29,619	29,619	29,619	29,619	29,619	Observations	29,619	29,619	29,619	29,619

Note: This table examines how different components and dimensions of digital transformation (DT) are associated with managerial insider selling. In column (1)–(5), following Wu et al. (2021), digital transformation is decomposed into five technology categories: Artificial Intelligence (AI), Big Data (DT), Cloud Computing (CC), Blockchain (BD), and Applied Digital Technologies (ADT). In addition, in column (6)–(9), we reclassify DT into three mutually exclusive functional dimensions: Technological, Operational, and Administrative. The dependent variable is the insider selling proportion (sell_share_manager). Each column reports estimates from Equation (1), where one or more functional dimensions are included as explanatory variables. All regressions include industry-by-year fixed effects. Standard errors are clustered at the industry level, and t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

advantage. We then complement this analysis by examining whether insiders at more digitally advanced firms earn higher abnormal returns from their trades. Third, we exploit a major regulatory tightening of insider selling in 2017 as an exogenous shock to examine whether digital transformation enhances insiders' ability to time their trades and sustain opportunistic selling under heightened regulatory scrutiny. Together, these tests provide empirical support for the mechanism that digital transformation facilitates insider selling by amplifying information advantages.

In the mechanism analysis, we focus on insider selling rather than share pledging, for both conceptual and empirical reasons. First, insider trading is fundamentally motivated by information advantages held by insiders (Frankel and Li, 2004; Huddart and Ke, 2007; Cohen et al., 2012). To examine this mechanism, we rely on stock price reactions, which capture the extent to which insider transactions reflect private, value-relevant information. In contrast, pledging behavior may itself generate market responses, as it can act as a signal of financial pressure or future risk, thereby introducing confounding effects that obscure the role of information advantage. Second, insider trading is more frequent and temporally dispersed than pledging, offering greater statistical power and allowing for a more precise analysis of timing, opportunism and regulatory responses. Taken together, these considerations make insider selling a more suitable setting for testing the information-based mechanism. Therefore, our mechanism analysis focuses exclusively on insider selling.

4.2.1. Information advantage

While information advantage cannot be directly observed, we infer its presence indirectly, through the predictive power of insider selling. Specifically, if insider sales are more likely to precede negative future stock returns, deteriorating firm fundamentals or adverse market reactions to earnings announcements, insiders are likely to be trading based on superior private information. Accordingly, we employ the following Model (2) to test whether digital transformation increases the extent to which insider selling predicts future firm performance and investor responses.

$$FirmInfo_{i,t+1} = \alpha + \beta_1 Digital_{i,t} + \beta_2 sell_manager_{i,t} + \beta_3 Digital_{i,t} \times sell_manager_{i,t} + \gamma Controls_{i,t} + \sum Industry \times Year + \varepsilon_{i,t} \quad (2)$$

$FirmInfo_{i,t+1}$ represents firm performance and investor reaction for firm i in year t + 1, including stock returns (*Yretwd*), *ROA* and *ROE*. $sell_share_manager_{i,t}$ denotes the share-selling ratio of senior management for firm i in year t, calculated as the number of shares sold by executives divided by the total number of A-shares outstanding. Following Cohen et al. (2012), we include *Size, BM*, lagged firm information ($FirmInfo_{i,t-1}$) and the other control variables in Model (1) as control variables.

Table 5 presents the results for whether digital transformation enhances the predictability of insider trading. The dependent variable in Column (1) of Table 5 is the stock return in the next fiscal year. The coefficient on $Digit\ al \times sell_share_manager$ is negative and significant (coefficient = −0.300, t-statistic = −2.432), indicating that digital transformation strengthens the ability of insider sales to predict future stock performance. Columns (2)–(3) examine whether digital transformation improves managers' ability to anticipate changes in firm fundamentals, measured by ROA and ROE. The results show that insider selling is significantly and negatively associated with future fundamentals in firms with high levels of digital transformation (coefficient = −0.061, t-statistic = −2.805 in Column (2); coefficient = −0.096, t-statistic = −2.228 in Column (3)). These results suggest that digital transformation enables insiders to access timely operational information before it becomes publicly available, whereas external investors react only upon the release of the annual report. Overall, the findings support the view that digital transformation enhances insiders' information advantage and increases the predictive content of their trading behavior.

4.2.2. Digital transformation and insider selling profitability

To further validate the information advantage mechanism, we examine whether insiders at firms with higher levels of digital transformation earn superior abnormal returns. Prior research shows that corporate executives often exploit private information in stock trading to earn excess returns from insider sales (Huddart and Ke, 2007). These excess returns serve as an important proxy for insiders' information advantage, offering indirect evidence of the strategic use of private information in trading decisions (Luo and

Table 5
Digital transformation, insider selling, and firm information prediction.

Variable	*F_yretwd* (1)	*F_ROA* (2)	*F_ROE* (3)
Digital	−0.151	−0.275***	−0.488***
	(−0.542)	(−4.630)	(−3.850)
sell_share_manager	−0.123	0.007	−0.086
	(−0.337)	(0.186)	(−1.076)
Digital* × *sell_share_manager	**−0.300****	**−0.061*****	**−0.096****
	(−2.432)	**(−2.805)**	**(−2.228)**
BM	9.272***	−3.699***	−9.787***
	(5.970)	(−9.071)	(−12.969)
ROA	33.642***		
	(6.809)		
Lev	3.029	−2.137***	−1.344
	(1.640)	(−5.624)	(−1.326)
Size	−7.983***	0.657***	1.487***
	(−21.125)	(6.658)	(6.421)
Ins_hold	−4.060***	1.011***	2.522***
	(−6.769)	(8.692)	(13.264)
SOE	−4.537***	0.099	0.346
	(−6.304)	(0.857)	(1.315)
Big4	2.405**	−0.254	−0.334
	(2.150)	(−1.297)	(−0.675)
Turnover	−0.009***	−0.000**	−0.001**
	(−7.478)	(−2.585)	(−2.307)
Mngmhldn	−0.059**	0.025***	0.067***
	(−2.345)	(6.804)	(8.982)
Top1	0.112***	0.021***	0.046***
	(9.108)	(6.100)	(6.782)
Boardsize	1.251	0.418*	0.674
	(0.759)	(1.903)	(1.545)
AnaAttention	1.645***	0.809***	1.962***
	(4.714)	(14.348)	(16.118)
L1_yretwd	−0.006		
	(−0.912)		
L1_ROA		0.278***	
		(18.963)	
L1_ROE			0.154***
			(12.892)
Constant	129.238***	−6.801***	−17.439***
	(21.945)	(−4.274)	(−4.697)
Industry × Year	Yes	Yes	Yes
Adj R^2	0.486	0.274	0.201
Observations	29,452	29,446	28,488

Note: This table examines whether digital transformation enhances management's ability to predict firm information through selling transactions. The specification follows Eq. (2), and the dependent variable is the stock return of the company in the next fiscal year (*F_yretwd*), financial performance such as *ROA*, *ROE*. *sell_share_manager* denotes the share selling proportion, measured by dividing the number of shares sold by management by the company's total number of A-shares outstanding. *Digital* is the degree of corporate digital transformation. All regressions include the industry-by-year fixed effects. Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

Huang, 2020). Building on this literature, we expect executives to leverage the informational benefits of digital transformation to enhance the profitability of their insider sales.[10] We employ the following Model (3) to examine the impact of digital transformation on the profitability of insider selling by executives.

$$CAR_{i,t,j}/BHAR_{i,t,j} = \alpha + \beta_1 Digital_{i,t} + \gamma Controls_{i,t} + \sum Industry * Month + \varepsilon_{i,t} \quad (3)$$

where $CAR_{i,t,j}$ represents the short-term cumulative abnormal returns of insider selling transaction j of firm i in year t, calculated over event windows of (0,1), (0,2), (0,3), (0,4) and (0,5) days centered on the annual report disclosure date. $BHAR_{i,t,j}$ represents the long-term buy-and-hold abnormal returns of insider selling transaction j of firm i in year t, measured over post-sale holding periods of 1, 3, 6, 9 and 12 months. For insider sales, a more negative post-sale *BHAR* indicates higher insider selling profitability. The regression model includes month-by-industry fixed effects and uses standard errors clustered at the industry level.

Panel A of Table 6 reveals that an increase in the degree of digital transformation has no significant impact on the short-term cumulative abnormal returns of insider selling but significantly enhances long-term buy-and-hold abnormal returns over horizons exceeding nine months. These findings suggest that trades executed by management leveraging information advantages from digital transformation are less likely to be detected by the market in the short term. However, by capitalizing on medium- to long-term informational advantages, managers are able to generate substantial profits from insider selling. Taken together, these results indicate that insider trading becomes more concealed when firms undergo digital transformation. Moreover, profits derived from longer-horizon trades are less likely to be recognized as resulting from insider activity, which in turn reduces the likelihood of regulatory scrutiny.

However, insider selling may be driven by either private information or liquidity needs, indicating that not all trades are opportunistic, even when insiders hold informational advantages (Jagolinzer et al., 2011; Cohen et al., 2012). To distinguish between these motives, we follow Cohen et al. (2012) in classifying trades as routine or opportunistic. Insiders (including their relatives) who trade in the same calendar month for three consecutive years are defined as routine traders; all others are considered non-routine traders. We then re-examine whether digital transformation enhances the profitability of opportunistic insider selling. Panel B of Table 6 reports results consistent with those in Panel A, indicating that digital transformation significantly enhances the long-term profitability of opportunistic insider selling.[11,12]

4.2.3. Regulation shock and superior timing ability based on digital transformation

To further distinguish an information-advantage channel from a firm fundamentals channel, we exploit the 2017 tightening of insider selling regulation in China as a quasi-natural experiment. This reform substantially strengthened disclosure requirements, extended cooling-off periods and increased regulatory scrutiny over managerial share sales, thereby raising the expected cost of opportunistic insider trading.

If digital transformation merely improves firm fundamentals, such as profitability or growth prospects, stricter regulation should not systematically affect insider selling differently across firms with different degrees of digitalization. In contrast, if digital transformation enhances managers' private information and trade-timing ability, insiders in more digitally transformed firms may still be able to trade more effectively, even under heightened regulatory oversight, because their informational advantages become more subtle and harder to detect.

[10] Digital transformation improves executives' access to detailed firm-level data, allowing for more accurate assessments of intrinsic value. This enables them to sell shares in anticipation of future performance declines, thereby avoiding potential losses. In addition, digital transformation provides insiders with timelier access to firm-specific information relative to external investors, improving their ability to time the market. This enhances their capacity to exploit medium- to long-term informational advantages while minimizing regulatory risk.

[11] Taking the six-month buy-and-hold excess return (*BHAR6m*) as an example, an increase of one standard deviation in the degree of digital transformation results in management obtaining an average excess return of 2.550% (= 1.355 × 1.882) from opportunistic insider selling, equivalent to 161% of the sample median (= 2.550 / 1.581).

[12] The coefficient on short-term excess returns from opportunistic insider selling is positive but nonsignificant over the initial five-day window, potentially reflecting the greater concealment and complexity of such transactions, which makes it difficult for the market to identify insiders' trading motives. This finding also suggests that opportunistic insiders possess stronger market-timing ability, deliberately avoiding trades at price peaks to minimize regulatory scrutiny.

Table 6
Digital transformation and the profitability of insider selling.

Panel A: Excess returns from all insider selling transactions

Variable	*CAR[0,1]* (1)	*CAR[0,2]* (2)	*CAR[0,3]* (3)	*CAR[0,4]* (4)	*CAR[0,5]* (5)	*BHAR1m* (6)	*BHAR3m* (7)	*BHAR6m* (8)	*BHAR9m* (9)	*BHAR12m* (10)
Digital	**−0.021**	**−0.018**	**−0.024**	**−0.016**	**0.005**	**−0.337***	**−0.261**	**−0.425**	**−1.227*****	**−1.602*****
	(−0.914)	**(−0.546)**	**(−0.686)**	**(−0.424)**	**(0.132)**	**(−1.889)**	**(−1.043)**	**(−1.119)**	**(−3.417)**	**(−3.767)**
ROA	−0.703*	−0.912**	−1.339**	−1.670***	−2.510***	23.345***	40.324***	69.147***	78.971***	84.829***
	(−1.736)	(−2.173)	(−2.440)	(−2.785)	(−4.223)	(7.087)	(8.439)	(7.015)	(5.865)	(5.112)
Lev	−0.272	−0.243	−0.305	−0.455	−0.461	4.154***	6.988***	7.512**	11.509***	14.236***
	(−1.042)	(−0.889)	(−0.946)	(−1.293)	(−1.204)	(2.779)	(3.083)	(2.616)	(3.985)	(5.111)
Size	0.079**	0.131***	0.181***	0.242***	0.290***	2.899***	4.268***	5.452***	5.374***	4.699***
	(2.020)	(2.830)	(3.618)	(3.240)	(3.498)	(6.344)	(8.053)	(6.205)	(5.117)	(3.437)
Ins_hold	−0.041	−0.012	−0.026	−0.021	0.026	0.486	1.929***	3.796***	4.409***	5.771**
	(−0.442)	(−0.087)	(−0.160)	(−0.118)	(0.133)	(1.052)	(2.749)	(3.565)	(3.076)	(2.588)
SOE	−0.097	−0.153	−0.101	−0.176	−0.187	−0.709	−2.236***	−4.563***	−4.725***	−8.662***
	(−0.802)	(−1.342)	(−0.773)	(−1.184)	(−1.099)	(−1.586)	(−3.372)	(−3.895)	(−2.870)	(−3.886)
Big4	−0.153	−0.241*	−0.090	−0.117	−0.158	−1.442	−1.896	0.303	−0.372	0.958
	(−1.142)	(−1.797)	(−0.532)	(−0.615)	(−0.795)	(−1.128)	(−1.238)	(0.106)	(−0.104)	(0.232)
Turnover	0.001***	0.000***	0.000**	0.000	0.000	0.004***	0.005***	0.004***	0.004***	0.002
	(4.545)	(3.048)	(2.047)	(1.393)	(1.142)	(7.729)	(7.589)	(4.555)	(4.009)	(1.323)
Mngmhldn	−0.002	−0.001	0.001	0.003	0.005	0.049***	0.093***	0.149***	0.173***	0.203***
	(−0.828)	(−0.482)	(0.401)	(0.808)	(1.319)	(3.851)	(6.321)	(6.171)	(5.119)	(4.255)
Top1	0.002	0.001	0.000	0.000	−0.001	0.019	−0.017	−0.056	−0.057	−0.097
	(1.369)	(0.356)	(0.069)	(0.075)	(−0.264)	(1.414)	(−0.565)	(−1.270)	(−0.760)	(−1.077)
Boardsize	−0.146	−0.225	−0.218	−0.271	−0.322	−1.912*	−3.851*	−5.660**	−6.085*	−5.930
	(−0.734)	(−0.849)	(−0.752)	(−0.849)	(−0.797)	(−1.805)	(−1.701)	(−2.253)	(−1.745)	(−1.299)
AnaAttention	0.078***	0.072**	0.044	0.020	0.006	−0.135	−0.629*	−0.826*	−0.603	−0.269
	(3.040)	(2.301)	(1.027)	(0.387)	(0.109)	(−0.469)	(−1.726)	(−1.768)	(−1.168)	(−0.319)
Constant	−0.418	−1.224*	−2.066***	−2.940**	−3.687***	−42.464***	−61.116***	−75.767***	−74.013***	−62.877***
	(−0.832)	(−1.907)	(−2.817)	(−2.681)	(−2.888)	(−5.948)	(−7.150)	(−5.330)	(−4.463)	(−4.000)
Industry × Month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.067	0.081	0.095	0.106	0.117	0.396	0.396	0.378	0.387	0.384
Observations	71,082	70,977	70,887	70,784	70,663	71,884	71,884	71,884	71,883	71,881

Panel B: Excess returns from opportunistic insider selling transactions

Variable	*CAR[0,1]* (1)	*CAR[0,2]* (2)	*CAR[0,3]* (3)	*CAR[0,4]* (4)	*CAR[0,5]* (5)	*BHAR1m* (6)	*BHAR3m* (7)	*BHAR6m* (8)	*BHAR9m* (9)	*BHAR12m* (10)
Digital	**0.131**	**0.241**	**0.214**	**0.285**	**0.346**	**−0.792***	**−1.551***	**−1.882****	**−1.625**	**−3.796****
	(1.047)	**(1.493)**	**(1.308)**	**(1.429)**	**(1.543)**	**(−1.854)**	**(−1.978)**	**(−2.100)**	**(−1.028)**	**(−2.566)**
ROA	−2.711**	−1.961	−1.846	−1.739	−3.524	14.417***	31.104**	43.580***	40.473	58.950***
	(−2.343)	(−1.471)	(−1.051)	(−0.746)	(−1.425)	(2.992)	(2.087)	(2.949)	(1.620)	(2.801)
Lev	−1.827**	−1.638*	−2.038*	−2.568*	−3.013**	−3.812*	−7.696*	−8.877*	−7.900	−1.491
	(−2.171)	(−1.704)	(−1.886)	(−1.995)	(−2.138)	(−1.738)	(−1.732)	(−1.848)	(−1.085)	(−0.221)

(*continued on next page*)

Table 6 (*continued*)

Panel B: Excess returns from opportunistic insider selling transactions

Variable	*CAR[0,1]* (1)	*CAR[0,2]* (2)	*CAR[0,3]* (3)	*CAR[0,4]* (4)	*CAR[0,5]* (5)	*BHAR1m* (6)	*BHAR3m* (7)	*BHAR6m* (8)	*BHAR9m* (9)	*BHAR12m* (10)
Size	0.089	0.112	0.214	0.112	0.185	2.631***	5.440***	6.630***	9.651***	9.440***
	(0.456)	(0.595)	(0.875)	(0.314)	(0.491)	(2.721)	(4.984)	(3.059)	(4.840)	(3.359)
Ins_hold	−0.924	−0.711	−0.966	−1.149	−0.772	−0.119	4.122	7.702	6.768	18.231*
	(−1.158)	(−0.688)	(−0.849)	(−0.993)	(−0.638)	(−0.046)	(0.822)	(1.213)	(1.241)	(1.966)
SOE	0.361	0.717**	0.573	0.861*	0.691	−1.277	−2.813	−0.014	0.888	−3.498
	(1.472)	(2.237)	(1.640)	(1.948)	(1.227)	(−0.592)	(−1.014)	(−0.004)	(0.280)	(−1.160)
Big4	0.793***	0.794**	1.102***	1.152**	0.911	2.571	2.288	13.969***	5.610	11.621
	(3.119)	(2.374)	(2.896)	(2.098)	(1.361)	(0.791)	(0.679)	(3.179)	(0.965)	(1.136)
Turnover	0.000	−0.000	−0.000	−0.001	−0.001	0.005**	0.008*	0.005	0.010	0.000
	(0.794)	(−0.242)	(−0.865)	(−1.252)	(−1.006)	(2.182)	(2.015)	(0.799)	(1.204)	(0.028)
Mngmhldn	−0.006	−0.001	0.002	0.001	0.009	−0.010	0.050*	0.071	0.136	0.199**
	(−0.505)	(−0.085)	(0.131)	(0.036)	(0.460)	(−0.409)	(1.863)	(0.970)	(1.177)	(2.258)
Top1	0.012	−0.003	−0.008	−0.015	−0.022	0.038	−0.036	−0.067	0.024	−0.184
	(0.693)	(−0.159)	(−0.368)	(−0.534)	(−0.622)	(0.524)	(−0.383)	(−0.423)	(0.130)	(−0.826)
Boardsize	−0.513	−0.746	−1.561	−2.258*	−2.734**	−2.684	−0.227	−3.426	−0.520	−2.409
	(−0.558)	(−0.621)	(−1.373)	(−1.929)	(−2.264)	(−1.021)	(−0.042)	(−0.497)	(−0.072)	(−0.210)
AnaAttention	0.117	0.025	−0.038	−0.008	0.002	0.317	−1.118	−1.662	−2.043	−2.771
	(1.159)	(0.217)	(−0.303)	(−0.058)	(0.012)	(0.323)	(−0.682)	(−0.982)	(−1.224)	(−1.250)
Constant	0.613	0.617	1.151	4.507	4.318	−32.645**	−79.070***	−89.162***	−147.276***	−133.801***
	(0.357)	(0.301)	(0.419)	(1.030)	(0.904)	(−2.154)	(−4.151)	(−3.680)	(−5.757)	(−4.025)
Industry×Month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.108	0.131	0.161	0.187	0.227	0.621	0.649	0.625	0.636	0.614
Observations	5116	5096	5092	5082	5076	5178	5178	5178	5178	5178

Note: This table examines the impact of digital transformation on the profitability of insider selling in the secondary market using trade-level data. The specification follows Eq. (3), and the dependent variables *CAR* represents the short-term returns of every insider selling transaction, while *BHAR* represents the long-term returns of every insider selling transaction. For example, *CAR[0,5]* is the 5-days cumulative abnormal returns of the transaction, and *BHAR6m* is the 6-months buy-and-hold abnormal return of the transaction. All regressions include the industry-by-month fixed effects. The explanatory variable is the degree of corporate digital transformation (*Digital*). Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. ***p < 0.01,**p < 0.05,*p < 0.1.

To test this hypothesis, we implement a continuous difference-in-differences (DID) design in which firm-level digital transformation intensity is interacted with a post-regulation indicator:

$$InsiderTrading_{i,t} = \alpha + \beta_1 Digital_{i,t} + \beta_2 Digital_{i,t} \times POST_t + \gamma Controls_{i,t-1} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (4)$$

where *POST* equals one for years after 2017, μ_i refers to firm fixed effects and λ_t refers to year fixed effects. This specification allows the regulatory shock to differentially affect firms depending on their degree of digital transformation.

As reported in Column (1) of Table 7, the interaction term *Digital* × *POST* is positive and highly significant. This result indicates that following the stricter regulation, insider selling increases more strongly in firms with higher levels of digital transformation. Importantly, this finding is inconsistent with the narrative of improvement to fundamentals: if digital transformation merely improved firm value, stricter regulation should not systematically amplify insider selling in digitally advanced firms.

In Column (2) of Table 7, we estimate an event-study version of the continuous DID model by interacting *Digital* with event-time indicators (*Pre4*, *Pre3*, *Pre2*, *Current*, *Post1* and *Post2*), with *Pre1* as the omitted category.

$$InsiderTrading_{i,t} = \alpha + \beta_1 Digital_{i,t} + \sum_{k=-4, k\neq -1}^{k=2} \beta_k (Digital_{i,t} \times 1\{t = k\}) + \gamma Controls_{i,t-1} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (5)$$

The post-policy coefficients (*Digital* × *Post1* and *Digital* × *Post2*) are positive and significant. This dynamic pattern indicates that the effect emerges only after the policy intervention and persists over time, reinforcing a causal interpretation. While the pre-period coefficients display some fluctuation, we do not interpret the results in Column (2) of Table 7 as a sharp test of strict parallel trends. Instead, we view the event-study pattern as supplementary evidence of the dynamic evolution around the 2017 regulatory tightening. Given the gradual tightening of China's share-sale regulation in 2015–2017, some pre-period movement may reflect anticipation and transition dynamics rather than a smooth pre-policy benchmark. Importantly, the economically meaningful increase appears after the 2017 revision and remains persistent thereafter.

Taken together, these results suggest that stricter insider trading regulation does not eliminate opportunistic insider selling in digitally transformed firms. Instead, digital transformation appears to strengthen insiders' ability to trade based on information advantages in a way that is more difficult to detect and constrain, even under heightened regulatory scrutiny. Overall, the evidence strongly supports the interpretation that digital transformation increases managers' long-term insider trading profits primarily through enhanced information advantages and superior trade timing, rather than through improvements in firm fundamentals.

4.3. Cross-sectional analyses

We expect the relationship between digital transformation and insider trading to be shaped by managerial opportunism and firms' information environments. Accordingly, we conduct the following cross-sectional analyses.

4.3.1. Managerial opportunism

This paper's proposition is grounded in the agency theory framework, with a specific focus on the self-interested motives of management. We infer that the impact of corporate digital transformation on insider selling is more pronounced in firms whose managers exhibit greater self-interest.

Executives in state-owned enterprises (SOEs) exhibit weaker motives to profit from insider transactions, primarily due to concerns over their own political risks. Additionally, compared with non-SOEs, SOEs are subject to more stringent regulations and greater societal scrutiny. Prior studies support this context: Zeng et al. (2018) find that non-SOE (vs. SOE) executives are more likely to manipulate the tone of annual reports to create a favorable market environment for their selling transactions; similarly, Luo and Huang (2020) document that the supervisory role of multiple large shareholders in curbing executive opportunistic selling is more prominent in non-SOEs than SOEs. Therefore, we expect non-SOE executives to have stronger motives

Table 7
Digital transformation and insider selling: evidence from the 2017 regulatory shock.

Variable	*sell_share_manager*	
	(1)	(2)
Digital	−0.004	−0.009
	(−0.280)	(−0.557)
Digital* × *POST	**0.051*** **	
	(3.164)	
***Digital* × *Pre*4**		**0.016**
		(0.441)
***Digital* × *Pre*3**		**−0.047***
		(−1.685)
***Digital* × *Pre*2**		**0.048***
		(1.715)
Digital* × *Current		**0.016**
		(0.824)
***Digital* × *POST*1**		**0.103*** **
		(4.543)
***Digital* × *POST*2**		**0.050** **
		(2.425)
ROA	−0.606***	−0.582***
	(−3.404)	(−3.274)
Lev	−0.066	−0.067
	(−0.955)	(−0.976)
Size	0.055***	0.050**
	(2.586)	(2.392)
Ins_hold	0.058***	0.055***
	(3.089)	(2.973)
SOE	0.021	0.022
	(0.702)	(0.737)
Big4	−0.105*	−0.107*
	(−1.905)	(−1.928)
Turnover	−0.000***	−0.000***
	(−4.506)	(−4.562)
Mngmhldn	0.033***	0.033***
	(13.932)	(13.926)
Top1	0.003**	0.003**
	(2.333)	(2.390)
Boardsize	0.100	0.103
	(1.353)	(1.389)
AnaAttention	0.013	0.014
	(1.047)	(1.129)
Constant	−1.100***	−1.040***
	(−3.108)	(−2.965)
Firm FE	YES	YES
YearFE	YES	YES
Adj R^2	0.285	0.286
Observations	26,420	26,420

Note: This table exploits the 2017 tightening of insider selling regulations in China as a quasi-natural experiment to examine whether digital transformation amplifies insider selling under heightened regulatory scrutiny. The dependent variable is the insider selling proportion (sell_share_manager). Column (1) reports estimates from a continuous difference-in-differences specification, where Digital is interacted with a post-2017 indicator (POST). Column (2) reports results from an event-study specification that interacts Digital with event-time indicators relative to the policy change. All regressions include firm fixed effects and year fixed effects. Standard errors are clustered at the firm level, and t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

to profit from insider transactions relative to their SOE counterparts, and we thus hypothesize that the role of corporate digital transformation in amplifying management's self-interest is more pronounced in non-SOEs.

To test these expectations, we partition our sample into SOE and non-SOE groups and estimate our regression model separately for each. The results, presented in Columns (1) to (4) in Table 8, show that in the SOE group, the coefficient on *Digital* is 0.003 with a t-statistic of 0.940, indicating that corporate digital transformation does not significantly increase managerial insider selling. In contrast, for the non-SOE group, the coefficient of *Digital* is 0.043 (t-statistic = 3.671), which is statistically significant at the 1% level. These findings confirm that the role of corporate digital transformation in reinforcing management's self-interest, thereby driving insider selling, is more pronounced in firms whose managers are more self-interested.

Based on agency theory, which emphasizes managerial self-interest as a key driver of organizational conduct, we further hypothesize that firms with a history of violations—where managerial self-serving motives and behaviors are more likely to prevail—show a more marked relationship between digital transformation and insider selling.

Firms with a history of violations typically have flawed internal governance structures and cultures that tolerate managerial opportunism. In such firms, managers find it easier to engage in self-interested actions, as the checks and balances that should curb such behavior are ineffective. To test this, we categorize the sample into firms with violations (a *Violation* group) and firms without violations (a *Non-violation* group) based on firms' prior year violation records and re-estimate our model. The results, shown in Table 8, Columns (5)–(8), reveal that for the *Violation* group, the coefficient is 0.039 (t-statistic = 2.554), while for the *Non-violation* group, the coefficient on *Digital* is 0.031 (t-statistic = 3.103). This demonstrates that the positive link between digital transformation and insider selling is significantly stronger in firms with (vs. without) prior violations, confirming that existing managerial self-interest amplifies the use of digital means for opportunistic insider trading.

4.3.2. Information environment

Managers' ability to exploit private information and profit from self-interested behavior is shaped by both internal and external information environments. When the overall information environment is weak (vs. strong), digital transformation may grant managers a greater informational advantage, enabling more opportunistic trading. Accordingly, we explore how the relationship between digital transformation and insider trading varies with firms' information environments, focusing specifically on accounting quality and the extent of coverage by information intermediaries.

When accounting quality is low (vs. high), managers are more likely to engage in opportunistic behavior, because the reduced precision and credibility of financial reports impair external monitoring and increase information asymmetry (Healy and Palepu, 2001; Dechow and Dichev, 2002). In such opaque environments, investors struggle to assess firm performance accurately, giving insiders greater leeway to exploit their informational advantage for insider trading. In addition, weak accounting quality often reflects poor internal controls or lax enforcement, lowering the perceived risk of detection and punishment (Ashbaugh-Skaife et al., 2009). The resulting lack of transparency makes it easier for managers to obscure self-serving actions. Therefore, we expect the relationship between digital transformation and managers' insider trading to be more pronounced in firms with lower accounting quality.

We partition the sample into high accounting information quality and low accounting information quality groups, using the tertiles of earnings management degree as the cutoff. The earnings management degree is calculated using the modified Jones model. The results, shown in Columns (1) to (4) of Table 9, show that for the group with higher earnings management (i.e., lower accounting information quality), the coefficient on digital transformation is statistically significant at the 1% level. This finding indicates that the positive association between corporate digital transformation and managerial insider selling is more pronounced in firms with higher earnings management, indicating that weaker accounting information quality amplifies the opportunistic insider trading driven by digital transformation.

When the external information environment is weak (vs. strong), particularly in firms with limited analyst and media coverage, managers are more likely to engage in opportunistic behavior, because external monitoring is reduced. Analysts and financial media serve as important channels for collecting, interpreting and disseminating firm-specific information to the public. Analysts can also exert governance influence by conducting

Table 8
Cross-sectional analyses: managerial opportunism.

Variable	Self-interest Motivation				Violation			
	sell_share_manager		*sell_money_manager*		*sell_share_manager*		*sell_money_manager*	
	Non-SOE (1)	SOE (2)	Non-SOE (3)	SOE (4)	Violation (5)	No-vio (6)	Violation (7)	No-vio (8)
Digital	**0.043*** **	**0.003**	**0.185*** **	**0.035**	**0.039** **	**0.031*** **	**0.224** **	**0.158*** **
	(3.671)	**(0.940)**	**(2.750)**	**(0.782)**	**(2.554)**	**(3.103)**	**(2.322)**	**(2.699)**
ROA	−1.137***	−0.029	−2.143*	1.131	−0.628***	−0.953***	−0.897	−0.493
	(−5.960)	(−0.337)	(−1.988)	(1.153)	(−3.198)	(−4.080)	(−0.823)	(−0.497)
Lev	−0.250**	−0.024	−1.327***	−0.035	−0.199**	−0.249***	−0.956**	−0.877**
	(−2.666)	(−0.846)	(−2.798)	(−0.282)	(−2.326)	(−3.399)	(−2.060)	(−2.606)
Size	−0.083***	0.001	1.273***	0.178***	−0.074***	−0.044***	0.833***	0.766***
	(−4.184)	(0.087)	(7.483)	(2.851)	(−3.274)	(−2.930)	(4.326)	(5.095)
Ins_hold	−0.045***	0.024***	0.501***	0.225***	0.001	−0.016	0.479***	0.404***
	(−2.769)	(3.263)	(6.607)	(3.319)	(0.028)	(−1.153)	(3.684)	(5.181)
SOE					−0.104***	−0.112***	−0.475**	−0.748***
					(−3.254)	(−5.475)	(−2.555)	(−4.956)
Big4	−0.027	−0.002	−0.661	−0.107	−0.021	−0.006	−0.072	−0.802***
	(−0.639)	(−0.363)	(−1.363)	(−1.278)	(−0.377)	(−0.269)	(−0.107)	(−4.046)
Turnover	−0.000***	0.000	−0.001***	−0.000	−0.000*	−0.000***	−0.001***	−0.001***
	(−3.572)	(0.946)	(−4.960)	(−0.593)	(−1.970)	(−2.839)	(−2.901)	(−5.504)
Mngmhldn	0.021***	0.042***	0.102***	0.257***	0.023***	0.022***	0.100***	0.098***
	(23.983)	(7.533)	(15.993)	(6.267)	(19.004)	(22.001)	(12.595)	(13.789)
Top1	−0.003**	−0.001*	−0.029***	−0.008***	−0.003*	−0.002***	−0.026***	−0.026***
	(−2.492)	(−1.988)	(−4.671)	(−3.296)	(−1.794)	(−2.988)	(−3.237)	(−5.384)
Boardsize	0.143*	0.023	0.431	0.174	0.116*	0.076	0.508	0.057
	(1.727)	(1.357)	(0.961)	(1.652)	(1.816)	(1.483)	(1.550)	(0.185)
AnaAttention	0.058***	−0.005	0.362***	−0.041	0.031*	0.039***	0.285***	0.225***
	(3.491)	(−0.970)	(4.107)	(−1.274)	(1.719)	(3.168)	(2.737)	(2.857)
Constant	1.319***	−0.027	−18.418***	−2.859***	1.196***	0.824***	−12.121***	−9.962***
	(3.876)	(−0.302)	(−7.290)	(−2.997)	(3.288)	(3.643)	(−4.701)	(−4.826)
Industry × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.152	0.117	0.132	0.101	0.156	0.186	0.118	0.149
Observations	18,059	11,560	18,059	11,560	7399	22,220	7399	22,220
Coeff. Diff.	0.040***		0.150***		0.008***		0.067***	

Note: This table examines whether the impact of digital transformation on insider selling will be more pronounced in firms with stronger manager's self-interest incentives. The dependent variable is the scale of insider selling, measured from two dimensions: trading proportion (*sell_share_manager*) and trading amount (*sell_money_manager*). The explanatory variable is the degree of corporate digital transformation (*Digital*). Columns (1)–(4) split the sample into non-state-owned enterprises (Non-SOEs) and state-owned enterprises (SOEs) to examine whether ownership structure moderates the impact of digital transformation on insider selling. Columns (5)–(8) split the sample into firms with regulatory violations and non-violations to test whether the effect differs across firms with different levels of regulatory discipline. All regressions include the industry-by-year fixed effects. Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. Significance levels are based on bootstrap (500 replications) P-values. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

on-site visits and engaging directly with management. Their coverage enhances transparency, reduces information asymmetry and increases the likelihood that self-interested managerial actions will be detected and constrained (Frankel and Li, 2004; Yu, 2008; Wu, 2019).[13] Firms with high media attention benefit from enhanced transparency, as media outlets serve as important information intermediaries (You et al., 2018; Goldman et al., 2022). Greater media coverage increases a firm's visibility in capital markets, making managerial trading behavior more likely to be scrutinized by investors and regulators. This heightened exposure

[13] For example, Yu (2008) finds that firms with more analyst coverage engage in less earnings management. Wu (2019) discovers a significant increase in excess returns from insider trading after a decrease in analyst coverage due to external shocks such as brokerage mergers or closures. Frankel and Li (2004) find that analysts, as information intermediaries, restrain insiders' profit-making ability by reducing information asymmetry.

Table 9
Cross-sectional analyses: information environment.

Variable	Accounting information quality				Analysts' attention				Media's attention			
	sell_share_manager		*sell_money_manager*		*sell_share_manager*		*sell_money_manager*		*sell_share_manager*		*sell_money_manager*	
	low	high	low	high	low	high	low	high	low	high	low	high
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Digital	**0.050*** **	**0.027** **	**0.258*** **	**0.148***	**0.042*** **	**0.025**	**0.149*** **	**0.180**	**0.037*** **	**0.021**	**0.193*** **	**0.132**
	(5.228)	**(2.027)**	**(4.253)**	**(2.005)**	**(3.612)**	**(1.279)**	**(4.449)**	**(1.248)**	**(3.051)**	**(1.303)**	**(3.186)**	**(1.297)**
ROA	−0.552*	−1.837***	0.803	−3.650	−0.588***	−0.673**	−1.732***	3.738	−0.481**	−0.861**	−0.929*	1.007
	(−1.943)	(−4.783)	(0.645)	(−1.309)	(−2.934)	(−2.215)	(−3.334)	(1.325)	(−2.677)	(−2.571)	(−1.690)	(0.283)
Lev	−0.144	−0.245***	−0.532	−0.918**	−0.094	−0.317***	−0.380**	−1.408**	−0.104	−0.320***	−0.203	−1.571**
	(−1.386)	(−3.682)	(−1.172)	(−2.334)	(−1.536)	(−3.220)	(−2.028)	(−2.173)	(−1.542)	(−3.295)	(−1.034)	(−2.441)
Size	−0.080***	−0.038	0.795***	0.725***	−0.039**	−0.054***	0.412***	1.140***	−0.034***	−0.081***	0.450***	0.866***
	(−4.133)	(−1.645)	(4.106)	(5.033)	(−2.293)	(−3.414)	(4.295)	(5.300)	(−2.781)	(−3.445)	(3.779)	(4.639)
Ins_hold	−0.001	−0.005	0.402***	0.535***	−0.013	−0.012	0.177*	0.756***	−0.007	−0.040*	0.225***	0.498***
	(−0.024)	(−0.219)	(3.938)	(3.973)	(−0.463)	(−0.525)	(1.918)	(5.437)	(−0.406)	(−1.892)	(3.328)	(3.295)
SOE	−0.098***	−0.124***	−0.584***	−0.954***	−0.069***	−0.115***	−0.332***	−0.970***	−0.074***	−0.139***	−0.349***	−1.094***
	(−2.978)	(−4.808)	(−3.923)	(−5.740)	(−2.743)	(−3.715)	(−3.612)	(−3.045)	(−3.512)	(−4.643)	(−3.164)	(−4.799)
Big4	−0.018	0.013	−0.761	−0.366	−0.068*	0.007	−0.440**	−0.718**	−0.041	0.044	−0.551**	−0.744**
	(−0.438)	(0.501)	(−1.608)	(−1.581)	(−1.863)	(0.250)	(−2.641)	(−2.078)	(−1.580)	(1.358)	(−2.175)	(−2.114)
Turnover	−0.000	−0.000	−0.000***	−0.001**	−0.000	−0.000*	−0.000***	−0.001***	−0.000*	−0.000*	−0.000***	−0.001***
	(−1.626)	(−0.208)	(−2.835)	(−2.486)	(−1.554)	(−1.801)	(−3.161)	(−4.416)	(−2.004)	(−1.984)	(−3.015)	(−5.769)
Mngmhldn	0.021***	0.023***	0.097***	0.100***	0.023***	0.020***	0.069***	0.139***	0.019***	0.023***	0.058***	0.136***
	(18.061)	(23.401)	(15.239)	(14.320)	(18.023)	(21.107)	(10.701)	(12.858)	(21.481)	(18.593)	(13.287)	(14.889)
Top1	−0.002	−0.001	−0.024***	−0.026***	−0.002**	−0.003*	−0.010***	−0.045***	−0.002**	−0.002	−0.016***	−0.037***
	(−1.491)	(−1.272)	(−3.691)	(−3.268)	(−2.244)	(−1.712)	(−2.822)	(−4.213)	(−2.129)	(−1.604)	(−3.680)	(−4.728)
Boardsize	0.058	0.137	0.275	0.496	0.026	0.015	0.243	−0.154	0.044	0.147*	0.204	0.520
	(0.927)	(1.584)	(0.825)	(1.132)	(0.298)	(0.232)	(1.039)	(−0.283)	(0.700)	(1.923)	(0.916)	(0.964)
AnaAttention	0.052***	0.036***	0.304***	0.248***	0.099*	0.032	0.343**	0.355	0.030***	0.040***	0.200***	0.201*
	(3.000)	(2.875)	(3.610)	(3.129)	(1.918)	(0.947)	(2.146)	(1.209)	(2.811)	(3.231)	(2.932)	(1.993)
Constant	1.306***	0.600*	−11.524***	−10.304***	0.718*	1.190***	−5.825***	−15.754***	0.609**	1.392***	−6.336***	−11.513***
	(3.895)	(1.718)	(−3.902)	(−4.775)	(1.943)	(5.149)	(−3.386)	(−5.980)	(2.471)	(3.734)	(−3.775)	(−4.166)
Industry × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.163	0.206	0.149	0.149	0.167	0.169	0.115	0.161	0.172	0.167	0.146	0.150
Observations	9303	9338	9303	9338	11,550	9674	11,550	9674	9924	9854	9924	9854
Coeff. Diff.	0.023***		0.109***		0.016***		−0.032***		0.016***		0.062***	

Note: This table examines whether the impact of digital transformation on insider selling will be more pronounced in firms with less transparent information environment. The dependent variable is the scale of insider selling, measured from two dimensions: trading proportion (*sell_share_manager*) and trading amount (s*ell_money_manager*). The explanatory variable is the degree of corporate digital transformation (*Digital*). In column (1) to (4), we divide the sample into groups of higher and lower accounting information quality based on the tertiles of earnings management degree calculated using the modified Jones model. In column (5) to (8), we divide the sample into groups of higher and lower analyst attention based on the tertiles of analyst coverage. In column (9) to (12), we divide the sample into groups of higher and lower media attention based on the tertiles of the total number of news headlines mentioning the company within a year. All regressions include the industry-by-year fixed effects. Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. Significance levels are based on bootstrap (500 replications) P-values. ***p < 0.01,**p < 0.05,*p < 0.1.

reduces the incentive for managers to engage in opportunistic insider selling, as their actions are more easily detected and subject to public scrutiny (Dai et al., 2015). Without sufficient external monitoring, managers face less scrutiny and a lower risk of detection, which makes opportunistic insider trading more feasible. As a result, digital transformation may further increase managers' information advantage and facilitate self-serving trading in firms with poor external oversight.

We partition the sample into high analyst attention and low analyst attention groups, using the tertiles of analyst coverage as the cutoff. The results, shown in Columns (5) to (8) of Table 9, reveal that for the group with lower analyst attention, the coefficient on *Digital* is statistically significant at the 1% level. In contrast, the coefficient on digital transformation becomes statistically nonsignificant for the group with higher analyst attention.

We further divide the sample into high media attention and low media attention groups, where the grouping is based on the tertile of the total number of annual news headlines mentioning the firm. As shown in Columns (9) to (12) of Table 9, the coefficient on *Digital* is statistically significant at the 1% level for the group with lower media attention only. These findings collectively suggest that external monitoring by analysts and the media mitigates the positive association between corporate digital transformation and managerial insider selling.

4.4. Robustness checks

4.4.1. Instrumental variable approach

To address the concern of unobservable omitted variable bias, we use the IV approach to check the robustness of our main results. Following Xiao et al. (2022), we use the product of the 1984 postal and telecommunications volume at the prefectural level and the one-year-lagged national internet user count (*IV_internet_1984*) as the IV. We use the 1984 postal and telecommunications volume at the prefectural level for two reasons, as follows. (1) The level of development of communication facilities in an enterprise's location can influence the enterprise's degree of application of digital technology from the perspectives of technology and social preference, satisfying the relevance condition. (2) Postal and telecommunications, as social infrastructure, do not directly affect insider trading behavior, meeting the exogeneity condition. As the 1984 urban postal and telecommunications data are cross-sectional, making them difficult to directly use them as IVs for panel data, inspired by Nunn and Qian (2014), we use the product of the 1984 postal and telecommunications volume at the prefectural level and the one-year-lagged national internet user count as the IV for the degree of an enterprise's current digital transformation.

Table A2 reports the results of the regression using the IV method. Column (1) of Table A2 presents the first-stage results. As expected, the coefficient of *IV_internet_1984* is statistically significant (coefficient on *IV_internet_1984* = 0.017, t-statistic = 4.071).[14] Columns (2) and (3) report the second-stage regression results. The results indicate that the coefficient of *Digital* is significant and positive,[15] supporting the positive association between digital transformation and insider selling.

4.4.2. Propensity score matched sample

A potential concern is that the observed association between digital transformation and insider trading may be driven by omitted firm characteristics that jointly affect both firms' incentives to digitally transform and insiders' trading behavior. To mitigate endogenous selection concerns arising from such omitted variables, we adopt a PSM approach.

[14] The Kleibergen–Paap rk LM statistic is significant at the 1% level, rejecting the null hypothesis of insufficient identification. The Cragg–Donald Wald F statistic exceeds the critical value of the weak IV test at the 10% significance level, rejecting the null hypothesis of a weak IV. Therefore, the IV selected in this study is reasonable.

[15] Similar to the findings of Yuan et al. (2021), the regression coefficient for *Digital* in the second stage of the IV method is significantly magnified compared with the baseline regression. According to Jiang (2017), this may be because the IV method captures only the average treatment effect of a subset of individuals in the sample, rather than the average treatment effect for the entire population. Therefore, it is necessary to interpret the economic significance of the estimated coefficients under the IV method with caution. This highlights the importance of considering the limitations of the IV approach and the potential for selection bias in the sample when drawing conclusions from the results.

We identify firms that undergo digital transformation as the treatment group and firms that never experience digital transformation during the sample period as the control group. The year in which a firm first adopts digital transformation is defined as the treatment year. We estimate firms' propensity to adopt digital transformation using a logistic regression model, where the dependent variable is an indicator of digital transformation adoption and the independent variables include the one-year-lagged values of all of the control variables used in the baseline regressions.

Based on the estimated propensity scores, we implement one-to-one nearest-neighbor matching without replacement to pair each treated firm with a control firm exhibiting similar observable characteristics. After constructing the matched sample, we combine the treated and control firms and restrict the analysis to a balanced event window spanning five years before and after the treatment year. This event window restriction allows us to focus on observations close to the adoption of digital transformation, where the treated and control firms are most comparable, and to reduce potential contamination from long-run structural changes unrelated to digital transformation. Consistent with Beaumont et al. (2025) and Agarwal et al. (2025), we then conduct regression analyses using the matched sample within a five-year window before and after the treatment year. As shown in Table A3, the coefficient of *Digital* is significant at least at the 10% level, mitigating the concern that the positive association between digital transformation and insider trading is driven by endogenous selection on observable variables.

4.4.3. Excluding the alternative explanation of disclosure effects

Previous studies show that insiders can influence investor sentiment and boost stock prices through announcements of high stock dividends (Devos et al., 2015), strategic media disclosures (Yang et al., 2022) and optimistic narrative disclosures (Bushee et al., 2023), thus facilitating insider selling at favorable prices. Consistent with this literature, it is possible that our findings are partially driven by managers exaggerating digital transformation disclosures to influence investor perceptions and secure abnormal returns from insider trading. To rule out this disclosure-based explanation, we examine insider selling behavior that occurs prior to the release of annual reports and estimate the following model:

$$Sell_share_manager_before/Sell_amount_manager_before = \alpha + \beta_1 Digital_{i,t-1} + \gamma Controls_{i,t-1} + \sum Industry \times Year + \varepsilon_{i,t} \quad (6)$$

where *Sell_share_manager_before* is the number of shares sold by management prior to the annual report disclosure date, scaled by the total number of tradable shares; and *Sell_amount_manager_before* is the total monetary value of shares sold by management during this pre-disclosure period.

Columns (1) and (2) of Table 10 present the results of estimating the relationship between digital transformation and insider selling prior to annual report disclosures. The coefficients on *Digital* are positive and statistically significant (coefficient = 0.005, t-statistic = 2.207 in Column (1); coefficient = 0.028, t-statistic = 2.739 in Column (2)). These findings indicate that even when restricting insider trading to the pre-disclosure period, digital transformation continues to significantly promote insider selling activity. This helps to mitigate the concern that the main results are merely driven by disclosure-induced investor sentiment. Overall, the evidence reinforces the core argument of this study: that digital transformation facilitates insider trading by enhancing information advantages.

Our next step is to exclude forward-looking disclosure.[16] Inspired by Muslu et al. (2015), we first extract sentences from annual reports containing phrases related to digital transformation and then remove sentences containing prospective terms. The remaining sentences, considered as actionable content of digital transformation, are logarithmically transformed and used as explanatory variables to re-examine the insider selling effects of corporate digital transformation. Columns (3) and (4) of Table 10 present the results of estimating the relationship between digital transformation without forward-looking disclosure and insider selling. The coeffi-

[16] Inspired by Muslu et al. (2015), we first extract sentences from annual reports containing phrases related to digital transformation and then remove sentences containing prospective terms. The remaining sentences, considered as actionable content of digital transformation, are logarithmically transformed and used as explanatory variables to reexamine the insider selling effects of corporate digital transformation.

Table 10
Insider selling behavior prior to annual report disclosure and excluding forward-looking information of digital.

Variable	Insider Selling Behavior Prior to Annual Report Disclosure		Excluding Forward-Looking Information of Digital	
	sell_share_manager_before (1)	*sell_money_manager_before* (2)	*sell_share_manager* (3)	*sell_money_manager* (4)
Digital	**0.005****	**0.028*****		
	(2.207)	**(2.739)**		
Digital_noforward			**0.038*****	**0.199*****
			(3.839)	**(3.376)**
ROA	−0.129***	−0.217	−0.880***	−0.734
	(−3.954)	(−1.217)	(−4.906)	(−0.809)
Lev	−0.024***	−0.081*	−0.215***	−0.808***
	(−2.980)	(−1.955)	(−3.314)	(−2.695)
Size	0.004	0.118***	−0.048***	0.760***
	(1.630)	(5.461)	(−3.878)	(5.457)
Ins_hold	−0.008***	0.013	−0.012	0.409***
	(−2.912)	(1.492)	(−1.081)	(6.714)
SOE	−0.023***	−0.120***	−0.114***	−0.723***
	(−4.814)	(−5.060)	(−5.331)	(−5.243)
Big4	−0.006*	−0.106***	−0.015	−0.763***
	(−1.768)	(−3.407)	(−0.775)	(−3.275)
Turnover	0.000	−0.000	−0.000***	−0.001***
	(0.077)	(−1.533)	(−2.717)	(−5.109)
Mngmhldn	0.003***	0.012***	0.022***	0.098***
	(20.891)	(15.962)	(27.814)	(15.900)
Top1	−0.000**	−0.003***	−0.002***	−0.026***
	(−2.218)	(−4.040)	(−3.199)	(−5.429)
Boardsize	0.020**	0.056	0.087*	0.178
	(2.387)	(1.210)	(1.771)	(0.698)
AnaAttention	0.003*	0.022***	0.038***	0.244***
	(1.749)	(2.803)	(3.558)	(3.768)
Constant	−0.047	−1.663***	0.863***	−10.167***
	(−1.365)	(−5.274)	(4.119)	(−5.133)
Industry × Year	Yes	Yes	Yes	Yes
Adj R^2	0.060	0.059	0.180	0.144
Observations	29,619	29,619	29,393	29,393

Note: This table examines impact of digital transformation on insider selling behavior before annual report disclosure and excluding prospective information on digital transformation measure. In column (1) and (2), The dependent variable is changed to the number of shares sold by management before the annual report disclosure divided by the total number of shares available for trading (*sell_share_manager_before*) and the total amount of shares sold by management before the annual report disclosure (*sell_money_manager_before*). The explanatory variable is the degree of corporate digital transformation (*Digital*). In column (3) and (4), the explanatory variable is changed to the degree of corporate digital transformation without forward-looking digital transformation information (*Digital_noforward*). The steps of compute *Digital_noforward* are: First, extracts sentences from annual reports containing phrases related to digital transformation and then removes sentences containing prospective terms. Second, the remaining sentences, considered as actionable content of digital transformation, are logarithmically transformed and used as explanatory variables. All regressions include the industry-by-year fixed effects. Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

cients on *Digital* are positive and statistically significant (coefficient = 0.038, t-statistic = 3.839 in Column (3); coefficient = 0.199, t-statistic = 3.376 in Column (4)).[17]

4.4.4. Other robustness tests

To ensure the robustness and reliability of the main conclusions of this study, further robustness tests are conducted by altering the measurement of key variables, controlling for industry characteristics, using firm fixed effects and adding control variables.

First, we change the measurement of corporate digital transformation. The frequency of terms in the management discussion and analysis section replaces the explanatory variable. This adjustment aims to eliminate measurement errors caused by descriptions of digital transformation in the annual report that do not pertain to the company itself, such as mentions of digital transformation when introducing shareholders, customers, suppliers and company executives. Next, the logarithm of the frequency of digital transformation terms in the management discussion and analysis section of the annual report is used as a replacement for the explanatory variable. Columns (1) to (2) of Table A4 indicate that the regression results are largely consistent with the main findings. Subsequently, the ratio of the frequency of digital transformation terms to the total number of words in the annual report replaces the explanatory variable. Columns (3) to (4) of Table A4 demonstrate that the coefficients of *Digital* remain significant, indicating the robustness of the main results when using a relative measure to assess the degree of digital transformation.

Second, we change the measurement of insider trading intensity. The number of shares sold by insiders divided by the total number of shares traded by the company and the amount of insider selling divided by the total trading volume of the company's shares are substituted for the dependent variable to measure the scale of insider trading. Columns (5) to (6) of Table A4 show that the coefficients of *Digital* remain significant at the 1% level, suggesting that after changing the measurement benchmarks, corporate digital transformation significantly increases the intensity of insider trading, thus supporting the fundamental conclusion of this study.

Third, we eliminate the influence of industry characteristics and company operations. To mitigate the potential effects of high-tech business operations inherently linked to digitalization, we take three approaches, following Yuan et al. (2021), to re-examine the sample. 1) We exclude information technology (IT) companies. Following the CSRC's industry classification in 2012, companies categorized under computer, communication and other electronic equipment manufacturing (C39) and information transmission, software and information technology services (I63, I64, I65) are removed. 2) We exclude companies identified as high-tech enterprises. Considering the significant association between high-tech enterprises and cutting-edge digital technologies such as the internet and big data, which may affect the accuracy of the digitalization index, samples identified as high-tech enterprises are excluded for retesting. 3) We exclude companies listed on the Growth Enterprise Market (GEM) and the Science and Technology Innovation Board (STAR Market). Many GEM- and STAR Market-listed companies have strong technological attributes and significant connections to the internet, emerging technologies, etc. Therefore, samples from GEM- and STAR Market-listed companies are excluded for retesting. Columns (3) to (5) of Table A5 demonstrate that the coefficients of *Digital* remain positive and significant, indicating that corporate digital transformation significantly increases the scale of insider selling. Therefore, the fundamental conclusion of this study is unaffected by industry characteristics or company operations.

Fourth, we change the dimensions of fixed effects and clustering. We alter the dimensions of fixed effects controlled in the main regression and the clustering dimensions of standard errors. Columns (7) to (9) of

[17] We also perform analyses such as excluding strategic disclosure behavior and controlling for textual features. The results, shown in Columns (1) to (2) and (6) of Table A5, indicate that the main results of this study are not affected by companies' disclosure behavior. To exclude strategic disclosure behavior, we follow Yuan et al. (2021) in 1) excluding samples with zero levels of digitalization (Column (1) of Table A5); and 2) retaining only firms rated as excellent or good in information disclosure by the Shanghai and Shenzhen stock exchanges. Samples of listed firms with excellent or good information disclosure ratings from the Shanghai and Shenzhen stock exchanges have higher information disclosure quality, reducing the likelihood of strategic information disclosure. Column (2) of Table A5 reports the results of controlling for textual features. Zeng et al. (2018) find that insiders manipulate the tone of annual reports to create a favorable market atmosphere for selling. Therefore, to mitigate the impact of annual report textual features on the conclusions of this study, readability, forward-looking and tone indicators are added to the control variables. Column (6) of Table A5 reports the results.

Table A5 show that the coefficients of *Digital* remain significant at the 1% level, suggesting that the main conclusions of this study remain valid after modifying the model specification.

5. Conclusion

With the rapid advancement of digital technologies, firms increasingly benefit from improved operational efficiency. However, digital transformation also intensifies information asymmetry between corporate insiders and outside investors. Leveraging the informational advantages afforded by digitalization, managers may be more inclined to engage in opportunistic behavior through insider trading. Consistent with this expectation, we find that digital transformation significantly increases both the volume and the monetary value of insider trading and share pledging, particularly insider selling. Managers' information advantage from digital transformation is reflected in the greater ability of insider trades to predict future stock returns and firm performance.

Importantly, timely access to information allows insiders greater flexibility in timing their trades, thereby increasing the opacity of insider transactions. As evidence, we find that while short-term abnormal returns from insider sales remain largely unchanged, long-term excess returns increase significantly. This behavior may undermine the interests of small and medium-sized investors and present heightened challenges for regulatory oversight. We also find that the self-serving behavior of management is more pronounced in firms with stronger managerial opportunism and less transparent information environments. These results remain robust when employing the IV approach and PSM to address endogeneity concerns.

A key limitation of our study lies in the measurement of digital transformation. Consistent with previous literature, we conduct keyword frequency analysis of annual reports to gauge the extent of digital transformation. However, this approach may not fully capture the extent to which digital technologies are integrated into firms' internal operations and the specific types of digital information generated and utilized. Future research should develop more refined and comprehensive measures to more accurately capture the actual extent of firms' digital transformation. Such advances would enable a more nuanced understanding of how insiders participate in the digitalization process and how digital transformation influences the behavior of different types of insiders.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for the helpful comments and suggestions made by seminar participants at Sichuan University, the University of Electronic Science and Technology of China, the Harbin Institute of Technology (Shenzhen) and Southwestern University of Finance and Economics. We acknowledge support from the National Natural Science Foundation of China (72303186, 72002179, 72010107001, 72272123), the National Social Science Fund of China (24AJY031), the 111 Project (B18043), the National Accounting Project of the Ministry of Finance of China (2023KJA1-04) and the China Research Center of Financial Law at Southwestern University of Finance and Economics (24CFL04).

Appendix A.

Table A1
Dictionary of corporate digital transformation.

Panel A: Keyword Classification across Technology Categories:	
Category	Keywords
Artificial Intelligence (AI)	Artificial intelligence, Business intelligence, Image understanding, Investment decision support systems, Intelligent data analysis, Intelligent robots, Machine learning, Deep learning, Semantic search, Biometric technology, Facial recognition, Speech recognition, Authentication, Autonomous driving, Natural language processing
Cloud Computing (CC)	Cloud computing, Stream computing, Graph computing, In-memory computing, Multi-party secure computing, Neuromorphic computing, Green computing, Cognitive computing, Converged architecture, Billion-level concurrency, Exabyte-level storage, Internet of Things, Cyber-Physical Systems
Big Data (DT)	Big data, Data mining, Text mining, Data visualization, Heterogeneous data, Credit reporting, Augmented reality, Mixed reality, Virtual reality
Blockchain (BD)	Blockchain, Digital currency, Distributed computing, Differential privacy technology, Smart financial contracts
Technology Applications (ADT)	Mobile internet, Industrial internet, Mobile interconnection, Internet healthcare, Electronic healthcare, Mobile payment, Third-party payment, NFC payment, Smart energy, B2B, B2C, C2B, C2C, O2O, Network connection, Smart wearables, Smart agriculture, Intelligent transportation, Intelligent healthcare, Intelligent customer service, Smart home, Intelligent investment consulting, Smart cultural tourism, Smart environmental protection, Smart grid, Intelligent marketing, Digital marketing, Unmanned retail, Internet finance, Digital finance, Fintech, Financial technology, Quantitative finance, Open banking
Panel B: Keyword Classification across Functional Dimensions:	
Technological	Artificial Intelligence (AI), Business Intelligence (BI), Image Understanding, Intelligent Data Analysis, Machine Learning, Deep Learning, Semantic Search, Biometric Technology, Facial Recognition, Speech Recognition, Identity Verification, Natural Language Processing (NLP), Cloud Computing, Stream Computing, Graph Computing, In-Memory Computing, Secure Multi-Party Computation (SMPC), Brain-Inspired Computing, Green Computing, Cognitive Computing, Converged Architecture, 100-Million-Level Concurrency, EB-Level Storage, Internet of Things (IoT), Cyber-Physical Systems (CPS), Big Data, Data Mining, Text Mining, Data Visualization, Heterogeneous Data, Augmented Reality (AR), Mixed Reality (MR), Virtual Reality (VR), Blockchain, Digital Currency, Distributed Computing, Differential Privacy Technology, Smart Financial Contracts, Smart Wearables Note: *DataAnalytics* subcategory includes: AI, BI, Image Understanding, Intelligent Data Analysis, Machine Learning, Deep Learning, Semantic Search, NLP, Big Data, Data Mining, Text Mining, Data Visualization, Heterogeneous Data

(*continued on next page*)

Table A1 (*continued*)

Panel B: Keyword Classification across Functional Dimensions:	
Operational/ process automation	Intelligent Robots, Autonomous Driving, Industrial Internet, Smart Energy, Smart Grid, Smart Agriculture, Smart Transportation, Smart Healthcare, Intelligent Customer Service, Smart Home, Smart Culture & Tourism, Smart Environmental Protection, Unmanned Retail
Administrative	Investment Decision Support System, Mobile Internet, Mobile Connectivity, Internet Healthcare, E-Commerce, Mobile Payment, Third-Party Payment, NFC Payment, B2B, B2C, C2B, C2C, O2O, Network Connectivity, Intelligent Marketing, Digital Marketing, Internet Finance, Digital Finance, FinTech, Quantitative Finance, Open Banking, Intelligent Investment Advisory, Credit Reporting

Note: Table A1, Panel A Following the approach of Wu et al. (2021), shows the key words and classifications of corporate digital transformation. Table A1, Panel B reports the specific keywords used to construct the functional dimensions of digital transformation. Keywords are classified into Technological, Operational (process automation), and Administrative dimensions. The Technological dimension includes foundational digital technologies and data-related capabilities, such as artificial intelligence, cloud computing, big data, and related infrastructure. The Operational dimension consists of keywords related to automation and real-time execution in production, logistics, and service delivery. The Administrative dimension captures digital applications in administration, governance, and market-facing activities, including digital payments, e-commerce, and FinTech. A Data Analytics subcategory is identified within the Technological dimension and includes keywords related to data processing, machine learning, natural language processing, and data visualization.

Table A2
Instrumental variable regression.

Variable	First stage	Second stage	
	Digital (1)	*sell_share_manager* (2)	*sell_money_manager* (3)
IV_internet_1984	0.017***		
	(4.071)		
Digital		**0.344****	**2.072***
		(2.253)	**(1.874)**
ROA	−0.773***	−0.673**	0.861
	(−3.652)	(−2.314)	(0.669)
Lev	0.109	−0.253***	−0.957***
	(1.419)	(−3.569)	(−2.738)
Size	0.163***	−0.106***	0.435
	(7.161)	(−3.699)	(1.542)
Ins_hold	0.084***	−0.042*	0.213
	(2.817)	(−1.815)	(1.208)
SOE	−0.104**	−0.085**	−0.583***
	(−2.157)	(−2.202)	(−2.790)

Table A2 (*continued*)

Variable	First stage	Second stage	
	Digital	*sell_share_manager*	*sell_money_manager*
	(1)	(2)	(3)
Big4	−0.158***	0.031	−0.457
	(−3.405)	(0.865)	(−1.247)
Turnover	−0.000	−0.000**	−0.001***
	(−0.641)	(−2.165)	(−4.011)
Mngmhldn	0.004**	0.021***	0.091***
	(2.328)	(20.235)	(9.798)
Top1	−0.003***	−0.002	−0.022***
	(−2.827)	(−1.530)	(−3.275)
Boardsize	−0.009	0.083*	0.146
	(−0.120)	(1.888)	(0.667)
AnaAttention	0.067***	0.021	0.126
	(5.319)	(1.323)	(1.269)
Industry × Year	Yes	Yes	Yes
Observations	26,619	26,619	26,619
Kleibergen-Paap rk LM statistic	7.108***		
Cragg-Donald Wald F statistic	117.003		
Kleibergen-Paap rk Wald F statistic	16.570		
Stock-Yogo weak ID test critical values 10% maximal IV size	16.38		

Note: This table reports the Two-Stage Least Squares regression results of the association between digital transformation and insider trading using the instrumental variable approach. Column (1) presents the first-stage results and column (2) and (3) report the second-stage results. *IV_internet_1984* is the product of the 1984 postal and telecommunications volume at the prefectural level and the lagged one-year national internet users. Other variable definitions are presented in Table1 in the paper. All continuous variables are winsorized at the 1% and 99% levels. All regressions include the industry-by-year fixed effects. T-statistics are based on standard errors clustered at the industry level. *, **, and *** denote significance based on two-tailed t-tests at or below the 10%, 5%, and 1% levels, respectively.

Table A3
Propensity Score Matching (PSM).

Variable	*All Year Insider Selling*		*Insider Selling Before Annual Report Disclosure*	
	sell_share (1)	*sell_money* (2)	*sell_share_before* (3)	*sell_money_before* (4)
Digital	**0.029***	**0.187****	**0.007****	**0.036****
	(1.969)	**(2.233)**	**(2.273)**	**(2.438)**
ROA	−0.167	0.137	−0.100	−0.306
	(−0.448)	(0.090)	(−1.465)	(−1.206)
Lev	−0.070	−1.051	−0.004	−0.060
	(−0.482)	(−1.297)	(−0.289)	(−0.778)
Size	−0.003	0.500***	0.012*	0.106***
	(−0.104)	(2.966)	(1.743)	(2.799)
Ins_hold	0.047	0.292*	0.001	0.021
	(1.116)	(1.748)	(0.170)	(0.799)
SOE	−0.091**	−0.539***	−0.018***	−0.089***
	(−2.441)	(−3.709)	(−4.066)	(−5.174)
Big4	−0.076*	−0.547*	−0.011*	−0.065
	(−1.779)	(−1.691)	(−1.694)	(−1.678)
Turnover	0.000	−0.000	0.000	0.000
	(0.617)	(−0.585)	(1.116)	(0.235)
Mngmhldn	0.023***	0.082***	0.003***	0.010***
	(8.908)	(6.720)	(6.346)	(6.192)
Top1	−0.004**	−0.024***	−0.000	−0.002**
	(−2.124)	(−2.778)	(−1.476)	(−2.077)
Boardsize	0.082	0.451	0.027**	0.104**
	(0.955)	(1.226)	(2.448)	(2.261)
AnaAttention	0.006	0.146	−0.005	−0.013
	(0.196)	(0.947)	(−1.026)	(−0.629)
Constant	0.064	−6.885***	−0.209*	−1.637***
	(0.145)	(−2.815)	(−1.962)	(−2.990)
Industry × Year	YES	YES	YES	YES
Adj R^2	0.186	0.138	0.076	0.065
Observations	16,394	16,394	16,394	16,394

Note: This table reports the results of the association between digital transformation and insider trading using a propensity score match sample. All variable definitions are presented in Table 1 in the paper. All continuous variables are winsorized at the 1% and 99% levels. The industry-by-year fixed effects are included. T-statistics are based on standard errors clustered at the industry level. *, **, and *** denote significance based on two-tailed t-tests at or below the 10%, 5%, and 1% levels, respectively.

Table A4
Robustness test: change measurement of key variables.

Variable	Measure *Digital* using "MD&A" section		Measure *Digital* using proportion		Alter the measurement of trading behavior	
	sell_share_manager (1)	*sell_money_manager* (2)	*sell_share_manager* (3)	*sell_money_manager* (4)	*sell_share_manager* (5)	*sell_money_manager* (6)
Digital	0.037***	0.210***	0.711*	5.533**	0.005***	0.005***
	(3.823)	(4.260)	(1.893)	(2.216)	(2.842)	(2.760)
ROA	−0.886***	−0.703	−0.886***	−0.659	−0.030	−0.028
	(−5.186)	(−0.767)	(−5.142)	(−0.715)	(−1.124)	(−1.076)
Lev	−0.214***	−0.798**	−0.212***	−0.785**	−0.036***	−0.035***
	(−3.398)	(−2.681)	(−3.334)	(−2.588)	(−3.469)	(−3.636)
Size	−0.050***	0.768***	−0.049***	0.765***	−0.003	−0.003
	(−3.708)	(5.540)	(−3.751)	(5.449)	(−0.966)	(−0.934)
Ins_hold	−0.012	0.421***	−0.011	0.425***	0.011***	0.011***
	(−1.000)	(6.708)	(−0.919)	(6.759)	(4.797)	(5.067)
SOE	−0.114***	−0.721***	−0.114***	−0.720***	−0.029***	−0.028***
	(−5.426)	(−5.351)	(−5.437)	(−5.349)	(−6.112)	(−6.187)
Big4	−0.012	−0.753***	−0.011	−0.741***	−0.013**	−0.012**
	(−0.618)	(−3.238)	(−0.579)	(−3.154)	(−2.204)	(−2.240)
Turnover	−0.000***	−0.001***	−0.000***	−0.001***	−0.000***	−0.000***
	(−2.770)	(−5.146)	(−2.804)	(−5.182)	(−7.212)	(−7.308)
Mngmhldn	0.022***	0.098***	0.022***	0.098***	0.004***	0.004***
	(27.331)	(15.664)	(27.438)	(15.734)	(25.440)	(25.979)
Top1	−0.002***	−0.027***	−0.002***	−0.026***	−0.001***	−0.001***
	(−3.174)	(−5.331)	(−3.077)	(−5.221)	(−4.327)	(−4.092)
Boardsize	0.088*	0.188	0.089*	0.191	0.016	0.015
	(1.838)	(0.741)	(1.832)	(0.759)	(1.423)	(1.365)
AnaAttention	0.039***	0.247***	0.040***	0.250***	0.008***	0.008***
	(3.591)	(3.708)	(3.636)	(3.758)	(4.148)	(4.212)
Constant	0.885***	−10.293***	0.887***	−10.227***	0.081*	0.074*
	(3.936)	(−5.240)	(3.974)	(−5.160)	(1.751)	(1.706)
Industry × Year	YES	YES	YES	YES	YES	YES
Adj R^2	0.179	0.144	0.179	0.144	0.143	0.147
Observations	29,619	29,619	29,619	29,619	29,602	29,602

Note: This table examines the impact of digital transformation on insider selling behavior by changing the measurement of key variables. Column (1) and (2) reports the results that replace the explanatory variable with the word frequency count from the "MD&A" section. Column (3) and (4) reports the results that measuring the degree of digital transformation by the proportion of digital-related word frequency in the annual report. Column (5) and (6) reports the results that altering the measurement approach of trading behavior. The dependent variable is the scale of insider selling, measured from two dimensions: trading proportion (*sell_share_manager*) and trading amount (*sell_money_manager*). The explanatory variable is the degree of corporate digital transformation (*Digital*). Robust standard errors clustered at the industry level. T-statistics are reported in parentheses. ***p < 0.01,**p < 0.05,*p < 0.1.

Table A5
Other robustness test.

Variable	Exclude zero digital transformation samples	Retain samples with excellent or good disclosure ratings	Exclude companies in the IT industry	Exclude of companies identified as high-tech enterprises	Exclude companies listed on the ChiNext and STAR board.	Controlling textual attributes of annual reports	Change fixed effects and clustering dimensions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Digital	**0.023****	**0.036*****	**0.033*****	**0.035*****	**0.028*****	**0.029*****	**0.033*****	**0.033*****	**0.022****
	(2.096)	**(3.334)**	**(3.867)**	**(2.886)**	**(3.153)**	**(2.952)**	**(3.762)**	**(3.953)**	**(2.029)**
ROA	−1.117***	−1.187***	−0.879***	−0.535**	−0.503***	−0.883***	−0.879***	−0.913***	−0.503***
	(−4.659)	(−4.132)	(−5.126)	(−2.358)	(−3.549)	(−5.036)	(−5.190)	(−5.546)	(−3.135)
Lev	−0.346***	−0.183**	−0.218***	−0.118*	−0.145***	−0.220***	−0.218***	−0.215***	−0.115*
	(−4.431)	(−2.370)	(−3.434)	(−1.808)	(−2.752)	(−3.423)	(−4.691)	(−4.730)	(−1.818)
Size	−0.041**	−0.034*	−0.051***	−0.034**	−0.032**	−0.054***	−0.051***	−0.045***	0.036**
	(−2.175)	(−1.757)	(−3.813)	(−2.414)	(−2.683)	(−4.014)	(−4.432)	(−3.983)	(1.963)
Ins_hold	−0.003	−0.016	−0.013	−0.010	−0.002	−0.015	−0.013	−0.009	0.037**
	(−0.213)	(−1.072)	(−1.071)	(−0.716)	(−0.140)	(−1.275)	(−0.863)	(−0.600)	(2.147)
SOE	−0.097***	−0.151***	−0.113***	−0.100***	−0.092***	−0.114***	−0.113***	−0.116***	−0.027
	(−3.969)	(−5.216)	(−5.363)	(−4.016)	(−5.125)	(−5.278)	(−7.532)	(−7.829)	(−1.114)
Big4	−0.007	−0.002	−0.011	−0.038**	−0.014	−0.005	−0.011	−0.011	−0.069
	(−0.301)	(−0.060)	(−0.600)	(−2.187)	(−0.776)	(−0.269)	(−0.558)	(−0.542)	(−1.340)
Turnover	−0.000***	−0.000*	−0.000***	−0.000	−0.000*	−0.000***	−0.000***	−0.000**	−0.000***
	(−2.899)	(−1.744)	(−2.762)	(−0.814)	(−1.859)	(−2.735)	(−2.945)	(−2.427)	(−4.171)
Mngmhldn	0.022***	0.022***	0.022***	0.023***	0.021***	0.022***	0.022***	0.022***	0.034***
	(18.417)	(20.288)	(27.341)	(19.585)	(19.266)	(27.974)	(24.402)	(24.724)	(15.849)
Top1	−0.003***	−0.003***	−0.002***	−0.002**	−0.002**	−0.002***	−0.002***	−0.002***	0.004***
	(−3.362)	(−3.531)	(−3.120)	(−2.024)	(−2.370)	(−3.039)	(−3.442)	(−3.536)	(3.151)
Boardsize	0.060	0.128*	0.087*	−0.012	0.102**	0.087*	0.087**	0.089**	0.136**
	(0.971)	(1.966)	(1.820)	(−0.225)	(2.539)	(1.852)	(2.090)	(2.171)	(2.032)
AnaAttention	0.031**	0.026*	0.039***	0.042***	0.021***	0.038***	0.039***	0.037***	0.009
	(2.039)	(1.689)	(3.583)	(3.454)	(2.721)	(3.262)	(4.094)	(4.007)	(0.783)
Readability						−0.006			
						(−1.401)			
Forward						−12.483			
						(−1.525)			
Tone						1.855			
						(1.482)			
Constant	0.919***	0.628**	0.909***	0.750***	0.491***	0.891***	0.909***	0.807***	−0.918***
	(2.971)	(2.018)	(4.046)	(3.095)	(2.938)	(3.356)	(4.790)	(4.327)	(−2.952)
Industry × Year	YES	YES	YES	YES	YES	YES			
Fixed effect							Ind × Year	Ind + Year	Firm
Cluster							Firm	Firm	Firm
Adj R^2	0.178	0.172	0.179	0.189	0.159	0.179	0.179	0.175	0.271
Observations	16,505	18,176	29,619	16,897	24,922	29,439	29,619	29,618	29,495

Note: This table shows further robustness tests of the main regression of column (3) in Table 3. Column (1) reports the results that excluding samples with zero levels of digitalization. Column (2) reports the results that retaining only firms rated as excellent or good in information disclosure by the Shanghai and Shenzhen Stock Exchanges. Column (3) reports the results that excluding information technology (IT) companies. Column (4) reports the results that excluding companies identified as high-tech enterprises. Column (5) reports the results that excluding companies listed on the Growth Enterprise Market (GEM) and the Science and Technology Innovation Board (STAR Market). Column (6) reports the results that controlling textual attributes of annual reports such as readability, forward-looking, and tone. Column (7) to (9) reports the results that changing fixed effects and clustering dimensions. Other variable definitions are presented in Table 1 in the paper. All

continuous variables are winsorized at the 1% and 99% levels. T-statistics are based on standard errors clustered at the industry level. *, **, and *** denote significance based on two-tailed t-tests at or below the 10%, 5%, and 1% levels, respectively.

Data availability

Data are available from the public sources cited in the text.

References

Abis, S., Veldkamp, L., 2024. The changing economics of knowledge production. Rev. Financ. Stud. 37 (1), 89–118.

Agarwal, S., Qian, W., Ren, Y., Tsai, H.-T., Yeung, B., 2025. The real impact of FinTech: evidence from mobile payment technology. Manag. Sci. 72 (5), 3804–3825.

Ahern, K.R., 2017. Information networks: evidence from illegal insider trading tips. J. Financ. Econ. 125 (1), 26–47.

Akbas, F., Jiang, C., Koch, P.D., 2020. Insider investment horizon. J. Finance 75 (3), 1579–1627.

Akey, P., Grégoire, V., Martineau, C., 2022. Price revelation from insider trading: evidence from hacked earnings news. J. Financ. Econ. 143 (3), 1162–1184.

Ali, U., Hirshleifer, D., 2017. Opportunism as a firm and managerial trait: predicting insider trading profits and misconduct. J. Financ. Econ. 126 (3), 490–515.

Allen, F., Qian, J., Qian, M., 2005. Law, finance, and economic growth in China. J. Financ. Econ. 77 (1), 57–116.

Arif, S., Kepler, J.D., Schroeder, J., Taylor, D., 2022. Audit process, private information, and insider trading. Rev. Account. Stud. 27 (3), 1125–1156.

Ashbaugh-Skaife, H., Collins, D.W., Kinney, W.R., LaFond, R., 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. J. Account. Res. 47 (1), 1–43.

Beaumont, P., Tang, H., Vansteenberghe, E., 2025. Collateral effects: the role of FinTech in small business lending. Rev. Financ. Stud. hhaf080

Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N., 2013. Digital business strategy: toward a next generation of insights. MIS Q. 37 (2), 471–482.

Bradshaw, M., Liao, G., Ma, M.S., 2019. Agency costs and tax planning when the government is a major shareholder. J. Account. Econ. 67 (2–3), 255–277.

Brazel, J., Dang, L., 2008. The effect of ERP system implementations on the management of earnings and earnings release dates. J. Inf. Syst. 22 (2), 1–21.

Brieger, S.A., Chowdhury, F., Hechavarría, D.M., Muralidharan, E., Pathak, S., Lam, Y.T., 2022. Digitalization, institutions and new venture internationalization. J. Int. Manag. 28 (4) 100949.

Bushee, B.J., Taylor, D.J., Zhu, C., 2023. The dark side of investor conferences: evidence of managerial opportunism. Account. Rev. 98 (4), 33–54.

Chen, Y., Lu, C., Li, Z., 2024. How does digital transformation affect corporate accounting employees? China J. Account. Res. 17 (4) 100391.

Chen, W., Srinivasan, S., 2023. Going digital: Implications for firm value and performance. Rev. Acc. Stud. 29, 1619–1665.

Cohen, L., Malloy, C., Pomorski, L., 2012. Decoding inside information. J. Finance 67 (3), 1009–1043.

Dai, L., Parwada, J.T., Zhang, B., 2015. The governance effect of the media's news dissemination role: evidence from insider trading. J. Account. Res. 53 (2), 331–366.

Dal Mas, F., Massaro, M., Verde, J.M., Cobianchi, L., 2020. Can the blockchain lead to new sustainable business models? J. Bus. Models 8 (2), 31–38.

Davenport, T.H., 1998. Putting the enterprise into the enterprise system. Harv. Bus. Rev. 76 (4), 121–131.

Dechow, P.M., Dichev, I.D., 2002. The quality of accruals and earnings: the role of accrual estimation errors. Account. Rev. 77 (s-1), 35–59.

Devos, E., Elliott, W.B., Warr, R.S., 2015. CEO opportunism? Option grants and stock trades around stock splits. J. Account. Econ. 60 (1), 18–35.

Dorantes, C.-A., Li, C., Peters, G.F., Richardson, V.J., 2013. The effect of enterprise systems implementation on the firm information environment. Contemp. Account. Res. 30 (4), 1427–1461.

Dou, Y., Masulis, R.W., Zein, J., 2019. Shareholder wealth consequences of insider pledging of firm stock as collateral for personal loans. Rev. Financ. Stud. 32 (12), 4810–4854.

Ferreira, J.J., Fernandes, C.I., Ferreira, F.A., 2019. To be or not to be digital, that is the question: firm innovation and performance. J. Bus. Res. 101, 583–590.

Frankel, R.M., Li, X., 2004. Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders. J. Account. Econ. 37 (2), 229–259.

Gaglio, C., Kraemer-Mbula, E., Lorenz, E., 2022. The effects of digital transformation on innovation and productivity: Firm-level evidence of South African manufacturing micro and small enterprises. Technol. Forecast. Soc. Change 182 121785.

Gattiker, T.F., Goodhue, D.L., 2005. What happens after ERP implementation: understanding the impact of interdependence and differentiation on plant-level outcomes. MIS Q. 29 (3), 559–585.
Ghasemaghaei, M., Calic, G., 2019. Does big data enhance firm innovation competency? The mediating role of data-driven insights. J. Bus. Res. 104, 69–84.
Goldman, E., Martel, J., Schneemeier, J., 2022. A theory of financial media. J. Financ. Econ. 145 (1), 239–258.
Gray, P., El Sawy, O.A., Asper, G., Thordarson, M., 2013. Realizing strategic value through center-edge digital transformation in consumer-centric industries. MIS Q. Exec. 12 (1) 3.
Hall, B.J., Liebman, J.B., 1998. Are CEOs really paid like bureaucrats? Q. J. Econ. 113 (3), 653–691.
Hanelt, A., Bohnsack, R., Marz, D., Marante, C.A., 2021. A systematic review of the literature on digital transformation: insights and implications for strategy and organizational change. J. Manag. Stud. 58 (5), 1159–1197.
He, Z., Liu, B., and Zhu, F., 2022. Share pledging in China: Funding listed firms or funding entrepreneurship? NBER Working Paper No. 29731. https://doi.org/10.3386/w29731.
Healy, P.M., Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature. J. Account. Econ. 31 (1–3), 405–440.
Hitt, L.M., Wu, D.J., Zhou, X.G., 2002. Investment in enterprise resource planning: business impact and productivity measures. J. Manag. Inf. Syst. 19 (1), 71–98.
Huang, W., Bai, Y., Luo, H., 2024. Customer identity concealing and insider selling profitability: evidence from China. J. Corp. Finance 85 102566.
Huddart, S.J., Ke, B., 2007. Information asymmetry and cross-sectional variation in insider trading. Contemp. Account. Res. 24 (1), 195–232.
Jagolinzer, A.D., Larcker, D.F., Taylor, D.J., 2011. Corporate governance and the information content of insider trades. J. Account. Res. 49 (5), 1249–1274.
Jagolinzer, A.D., Larcker, D.F., Ormazabal, G., Taylor, D.J., 2020. Political connections and the informativeness of insider trades. J. Finance 75 (4), 1833–1876.
Jeng, L.A., Metrick, A., Zeckhauser, R., 2003. Estimating the returns to insider trading: a performance-evaluation perspective. Rev. Econ. Stat. 85 (2), 453–471.
Jensen, M.C., Murphy, K.J., 1990. Performance pay and top-management incentives. J. Polit. Econ. 98 (2), 225–264.
Jiang, F., Kim, K.A., 2020. Corporate governance in China: a survey. Rev. Finance 24 (4), 733–772.
Jiang, Y., Wang, X., 2024. Digital transformation, innovation capability and speed of internationalization. Financ. Res. Lett. 67 (Part A) 105448.
Jiang, W., 2017. Have instrumental variables brought us closer to the truth? Rev. Corp. Financ. Stud. 6 (2), 127–140.
Jiang, W., Li, T., 2024. Corporate governance meets data and technology. Found. Trends Financ. 14 (2), 61–136.
Kao, L., Chiou, J.R., Chen, A., 2004. The agency problems, firm performance, and monitoring mechanisms: the evidence from collateralized shares in Taiwan. Corp. Gov. Int. Rev. 12, 389–402.
Ke, B., Huddart, S., Petroni, K., 2003. What insiders know about future earnings and how they use it: evidence from insider trades. J. Account. Econ. 35 (3), 315–346.
Lakonishok, J., Lee, I., 2001. Are insider trades informative? Rev. Financ. Stud. 14 (1), 79–111.
Li, C., Huo, P., Wang, Z., Zhang, W., Liang, F., Mardani, A., 2023a. Digitalization generates equality? Enterprises' digital transformation, financing constraints, and labor share in China. J. Bus. Res. 163 113924.
Li, F., Qian, J., Wang, H., Zhu, J.L., 2023b. Stock pledged loans and market crash risk: evidence from China. J. Finance Data Sci. 9 100104.
Li, L., Bo, W., Qin, L., 2024. Spillover effect of digital transformation along the supply chain: from the perspective of suppliers' audit fees. China J. Account. Res. 17 (3) 100363.
Liu, D., Chen, S., Chou, T., 2011. Resource fit in digital transformation. Manag. Decis. 49 (10), 1728–1742.
Lu, R., Zhao, J.Y., Liu, C., Liu, J.H., 2022. The corporate governance effect of capital market liberalization: from the perspective of controlling shareholders' share pledging. Account. Res. 2, 164–178 (in Chinese).
Luo, H., Huang, W., 2020. Research on the impact of multiple large shareholders coexistence on the opportunistic stock selling behaviors of executives. Manag. World 36 (8), 163–178 (in Chinese).
McKinsey Global Institute, 2017. China's digital economy: A leading global force. McKinsey & Company. Available at https://www.mckinsey.com/~/media/mckinsey/featured%20insights/China/Chinas%20digital%20economy%20A%20leading%20global%20force/MGI-Chinas-digital-economy-A-leading-global-force.ashx.
Müller, O., Fay, M., vom Brocke, J., 2018. The effect of big data and analytics on firm performance: an econometric analysis considering industry characteristics. J. Manag. Inf. Syst. 35 (2), 488–509.
Muslu, V., Radhakrishnan, S., Subramanyam, K.R., Lim, D., 2015. Forward-looking MD&A disclosures and the information environment. Manag. Sci. 61 (5), 931–948.
Nienhaus, M., 2022. Executive equity incentives and opportunistic manager behavior: New evidence from a quasi-natural experiment. Rev. Account. Stud. 27 (4), 1153–1196.
Nunn, N., Qian, N., 2014. US food aid and civil conflict. Am. Econ. Rev. 104 (6), 1630–1666.
Ofek, E., Yermack, D., 2000. Taking stock: equity-based compensation and the evolution of managerial ownership. J. Finance 55 (3), 1367–1384.
Peng, Y., Wang, N., Gu, L., 2023. Corporate digital transformation, predictive information disclosure, and stock price crash risk. Finance Trade Econ. 44 (5), 73–90 (in Chinese).

Pincus, M., Tian, F., Wellmeyer, P., Xu, S.X., 2017. Do clients' enterprise systems affect audit quality and efficiency? Contemp. Account. Res. 34 (4), 1975–2021.
Piotroski, J.D., Roulstone, D.T., 2005. Do insider trades reflect both contrarian beliefs and superior knowledge about future cash flow realizations? J. Account. Econ. 39 (1), 55–81.
Piotroski, J.D., Wong, T.J., Zhang, T., 2015. Political incentives to suppress negative information: evidence from Chinese listed firms. J. Account. Res. 53 (2), 405–459.
Porfírio, J.A., Carrilho, T., Felício, J.A., Jardim, J., 2021. Leadership characteristics and digital transformation. J. Bus. Res. 124, 610–619.
Rozeff, M.S., Zaman, M.A., 1998. Overreaction and insider trading: evidence from growth and value portfolios. J. Finance 53 (2), 701–716.
Spathis, C., 2006. Enterprise systems implementation and accounting benefits. J. Enterp. Inf. Manag. 19 (1), 67–82.
Sun, X.Y., Liu, Y.S., 2021. Equity pledge, employee stock ownership plan and large shareholders' self-interest behavior. Account. Res. 4, 117–129 (in Chinese).
Tian, F., Xu, S.X., 2015. How do enterprise resource planning systems affect firm risk? Post-Implementation Impact. MIS Q. 39 (1), 39–60.
Vaska, S., Massaro, M., Bagarotto, E.M., Dal Mas, F., 2021. The digital transformation of business model innovation: a structured literature review. Front. Psychol. 11 539363.
Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N., Haenlein, M., 2021. Digital transformation: a multidisciplinary reflection and research agenda. J. Bus. Res. 122, 889–901.
Vial, G., 2019. Understanding digital transformation: a review and a research agenda. J. Strateg. Inf. Syst. 28 (2), 118–144.
Wen, H., Zhong, Q., Lee, C.C., 2022. Digitalization, competition strategy and corporate innovation: evidence from Chinese manufacturing listed companies. Int. Rev. Financ. Anal. 82 102166.
Wu, F., Hu, H., Lin, H., Ren, X., 2021. Corporate digital transformation and capital market performance: empirical evidence from stock liquidity. Manag. World 37 (7), 130–144 (in Chinese).
Wu, K., Fu, Y., Kong, D., 2022. Does the digital transformation of enterprises affect stock price crash risk? Finance Res. Lett. 48 102888.
Wu, W., 2019. Information asymmetry and insider trading. Fama-Miller Working Paper. Chicago Booth Research Paper, No. 13-67.
Wu, W., Tian, Y., 2022. Can corporate digital transformation reduce cost stickiness? A perspective based on expense adjustment capability. Account. Res. 4, 89–112 (in Chinese).
Xiao, T., Sun, R., Yuan, C., Sun, J., 2022. Digital transformation, human capital structure adjustment and labor income share. Manag. World 38 (12), 220–237 (in Chinese).
Yang, J., Wang, J., Yi, Z., Mao, N., 2022. Managing the media coverage: evidence from managers' stock selling. Asia-Pac. J. Account. Econ. 29 (1), 77–96.
Yang, X., Xie, R., 2024. The dark side of digital transformation: evidence from opportunistic insider selling. Finance Res. Lett. 67 (Part B) 105902.
Ye, P., Zeng, Q., Zhang, C., 2025. Sell-by-plan mandate and opportunistic insider selling: evidence from China. J. Account. Econ. 79 (2–3) 101757.
Yeh, Y.H., Ko, C., Su, Y.H., 2003. Ultimate control and expropriation of minority shareholders: new evidence from Taiwan. Acad. Econ. Pap. 31, 263–299 (in Chinese).
You, J., Zhang, B., Zhang, L., 2018. Who captures the power of the pen? Rev. Financ. Stud. 31 (1), 43–96.
Yu, F., 2008. Analyst coverage and earnings management. J. Financ. Econ. 88 (2), 245–271.
Yuan, C., Xiao, T., Geng, C., Sheng, Y., 2021. Digital transformation and corporate division of labor: specialization or vertical integration. China Ind. Econ. 9, 137–155 (in Chinese).
Zeng, Q., Zhou, B., Zhang, C., Chen, X., 2018. Annual report tone and insider trading: consistency or discrepancy between words and actions? Manag. World 34 (9), 143–160 (in Chinese).
Zhou, Z., Li, Z., 2023. Corporate digital transformation and trade credit financing. J. Bus. Res. 160 113793.
Zhu, C., 2019. Big data as a governance mechanism. Rev. Financ. Stud. 32 (5), 2021–2061.

China Journal of Accounting Research 19 (2026) 100504

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Predicting goodwill impairment from M&A comment letters: an explainable machine learning approach

Hongqin Tang, Jialan Huang*, Nan Li

School of Management, Xiamen University, Xiamen, Fujian 361005, China

ARTICLE INFO

Article history:
Received 10 October 2025
Accepted 28 May 2026
Available online 4 June 2026

JEL Classification:
G32
G34
M41
C45
C55

Keywords:
M&A
Regulatory comment letters
Goodwill impairment
Explainable machine learning methodology
BERT

ABSTRACT

This study develops an interpretable machine learning framework to predict goodwill impairment using linguistic features of mergers and acquisitions comment letters. We extract thematic content via Bidirectional Encoder Representations from Transformers-based synonym self-generation with term frequency–inverse document frequency and capture readability, length and response time through text mining. Tree-based models with SHapley Additive exPlanations analysis reveal that valuation-focused, less readable letters signal higher impairment risk, whereas forward-looking, detailed letters with clear deadlines indicate lower risk. Predictive power strengthens in weakly governed firms. This study demonstrates that unstructured regulatory text serves as an early warning signal for financial risk and advances explainable natural language processing applications in accounting research.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Mergers and acquisitions (M&A) constitute key strategic actions undertaken by firms to achieve specific objectives. They play a vital role in enhancing efficiency, optimizing economic structures, driving industrial transformation and fostering competitive advantages, and they have thereby become an essential element of corporate strategy, closely linked to shareholder value creation (Aaldering et al., 2019). In the Chinese con-

* Corresponding author at: 422 Siming South Road, Siming District, Xiamen, Fujian 361005, China.
E-mail address: huangjialan@stu.xmu.edu.cn (J. Huang).

https://doi.org/10.1016/j.cjar.2026.100504
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

text, M&A can effectively enhance new product development performance (Chen and Lin, 2011), facilitate knowledge exchange or knowledge transfer (Ai and Tan, 2018) and make firms more competitive both domestically and globally (Zhu and Zhu, 2016). Consequently, M&A activities by a listed firm have emerged as a highly coveted investment paradigm among Chinese investors. However, issues such as severe information asymmetry, agency conflicts, tunneling and managerial overconfidence have led to frequent occurrences of "deceptive" M&A activities, resulting in uncertainty regarding M&A outcomes and significant value destruction for firms (Malmendier and Tate, 2007; Chung and Hribar, 2021). In response, Chinese regulatory authorities have emphasized stricter oversight of financing and M&A activities in sensitive areas to mitigate spillover risks and protect minority investors.

More effective regulation requires better identification of the "hype" qualities and hazards connected with firms M&A activities. The stock exchange is an important regulatory agency in the capital market, responsible for overseeing securities. Among the several methods it deploys, the comment letter is a non-punitive regulatory approach. It falls under "during and post-event" supervision of information sharing and is rapidly becoming the major regulatory strategy. Chinese comment letters are grouped into three types, namely annual report comment letters, M&A comment letters and attention comment letters (for areas of concern), all of which perform governance and risk warning purposes. M&A comment letters—in contrast with annual report comment letters, which are periodic and retroactive—are issued prior to the M&A transaction and are not periodic. Exchanges act as an extension of the China Securities Regulatory Commission's (CSRC's) regulatory authority, exercising self-discipline rather than administrative enforcement powers. They can require listed firms to provide additional clarity through comment letters, but they cannot impose corrective actions. Listed firms often reply at length to such requests for clarity, but they rarely substantially change essential contractual terms such as merger considerations. For example, in Sunliban's (SZ.000606) acquisition of a 100% equity stake in Shenzhou Yiqiao, the firm only clarified the valuation methodology, transaction pricing, goodwill recognition and performance compensation arrangements in its M&A comment letter, without making any revisions to the disclosed figures. Unlike periodic comment letters, which allow for ongoing follow-up, M&A comment letters do not facilitate continuous post-merger monitoring. Consequently, their primary function lies in risk identification during the pre-merger phase. Furthermore, compared with traditional financial variables, the textual content of comment letters offers distinct informational advantages in risk identification. Many key risk indicators (such as goodwill impairment) typically emerge only after the completion of an acquisition and are therefore inherently lagged. In contrast, comment letter text, as a direct expression of regulators' professional judgment, is generated prior to the implementation of the transaction and reflects regulatory concerns about potential risks ex ante, thereby providing forward-looking signals for subsequent risk detection. Meanwhile, M&A-related risks generally stem from multiple dimensions, including economic fundamentals, managerial incentives and transaction characteristics. Through systematic inquiries across thematic areas, comment letters are able to capture and integrate these multidimensional risk factors, which are difficult to fully characterize using a single or limited set of structured financial variables. Against this backdrop, research examines the risk warning function of M&A comment letters, focusing on simple generic textual features such as whether a letter was issued, the number of questions raised and the length of the disclosure (Kothari et al., 2009; Dechow et al., 2016). As a formal and content-rich regulatory document, a comment letter requires in-depth examination that considers its substantive content. Whether the specific thematic information contained in these letters can effectively signal future risks associated with M&A activities remains an important question for further investigation.

Goodwill impairment signals M&A risk. Although shaped by irrational pricing and managerial discretion (Ramanna and Watts, 2012; Li and Sloan, 2017; Beatty et al., 2019), goodwill impairment ultimately reflects post-merger integration outcomes (Li et al., 2011). Research on its determinants falls into three categories: economic (AbuGhazaleh et al., 2011; Bostwick et al., 2016), managerial (Riedl, 2004; Ramanna and Watts, 2012) and deal-related (Hayn and Hughes, 2006; Gu and Lev, 2011; Olante, 2013).

Goodwill impairment signals that an acquisition's expected benefits have not materialized, often prompting negative market reactions, including downward earnings revisions, sell-offs and stock price declines, thereby increasing operational and financial risks. Importantly, the factors driving impairment are not only reflected in internal financial reports; they also attract regulatory scrutiny from the outset of the deal, primarily through M&A comment letters. These letters systematically capture three categories of impairment risk: economic,

such as inquiries about the target's past performance and future earnings; managerial, such as questions regarding accounting policies, impairment testing methods and key parameters; and deal-specific, such as performance obligation achievability, pricing reasonableness and industry risks. In essence, M&A comment letters serve as a condensed, authoritative external assessment of the risks underlying a transaction.

To further explore whether the specific content of restructuring comment letters can effectively warn of and detect future goodwill impairment, and whether the importance of different thematic components in terms of detection capability varies, this study employs detection-oriented methods based on natural language processing (NLP) techniques. NLP is widely applied to extract basic textual features such as sentiment tone (Huang et al., 2014) and text readability (Wang et al., 2022), which help assess how understandable and emotionally charged a text is. NLP can also identify topic-related features, such as those reflecting specific economic concerns (Bybee et al., 2023), allowing for a deeper understanding of the thematic orientation of documents. Some studies also use features that are not inherently explainable, such as those derived from transformer-based models (Borchert et al., 2023) or unsupervised keyword clusters generated through latent Dirichlet allocation (LDA) (Pekar et al., 2024), which are often adopted for detection performance. In practical applications, the effective analysis of textual information requires not only intuitive insights derived from basic features but also deeper interpretation informed by the reader's knowledge, which is particularly crucial in the financial domain. As formal regulatory disclosures in the fields of accounting and finance, M&A comment letters call for the selection of textual features that are clearly explainable and analytically transparent. A meaningful analysis must combine fundamental linguistic comprehension with specific expertise to ensure accurate understanding, helping stakeholders to grasp the substance and implications of regulatory scrutiny.

This study investigates whether the linguistic composition of M&A comment letters can detect and signal subsequent goodwill impairment risk. This focus on a single information carrier, namely text, reflects an interest in whether the content of these comment letters can help investors identify and detect future risk, rather than in establishing a direct causal relationship between regulatory inquiries and subsequent goodwill impairment. In modern settings, individuals are frequently exposed to large volumes of data from multiple sources, which substantially increase their cognitive burden. When cognitive load exceeds a certain threshold, information overload may occur, increasing the likelihood of erroneous decisions (Sweller, 1988). Such overload also poses challenges for firms and their managers, and it has become a major impediment to effective firm judgment and decision-making, as well as an obstacle to implementing decisions (Libby and Lipe, 1992; Hwang et al., 2019; Liu et al., 2025). Importantly, information overload cannot be fully eliminated even in the era of artificial intelligence (AI). For example, training large language models that incorporate extensive multisource information requires substantial computational resources and may entail extremely long computing time (Kumar, 2024). Consequently, there remains a need to develop relatively low-cost models that can achieve strong performance within specific domains. In the context of M&A comment letters, textual content constitutes the most direct and effective channel through which professional regulatory guidance is communicated to investors. As an important signal of latent risk, such content already carries inherent risk detection value. Building on this premise, this study proposes an explainable machine learning detection framework that relies exclusively on the textual content of M&A comment letters while achieving high detection accuracy, with the goal of identifying risk at minimal information processing cost. Further, the study seeks to identify which textual elements in M&A comment letters can serve as early indicators of goodwill impairment risk and to assess the relative importance of these textual features for goodwill impairment detection, thereby elucidating the role of M&A-related comment letters in the early warning of goodwill impairment risk.

Accordingly, this study is expected to make both theoretical and practical contributions. Theoretically, the study integrates text mining and explainable machine learning techniques to construct a detection framework that achieves high accuracy while enhancing understanding of how textual features in M&A comment letters function as risk warning signals for goodwill impairment. This provides a comprehensive understanding of the role of preemptive government regulation in M&A activities and adds more evidence to research on the risk warning function of M&A comment letters. Additionally, by incorporating ordinary least squares (OLS)-based linear analysis into explainable machine learning methods, the study calculates the linear relationship between feature values and SHapley Additive Explanations (SHAP) values. This methodological improvement further enhances the transparency and explainability of machine learning models. Practically, in the era of the digital economy, individuals receive overwhelming amounts of information; therefore, it is crucial to make

accurate decisions with minimal cognitive resources. This study develops a precise goodwill impairment detection model relying solely on M&A comment letter textual data, effectively reducing stakeholders' cognitive load and improving decision-making efficiency while demonstrating the effectiveness of regulatory inquiries. The explainable machine learning model not only achieves high detection accuracy but also reveals the impact of various features on goodwill impairment. This will provide valuable reference information for practitioners, helping them to identify potential factors leading to goodwill impairment risk in advance and to formulate appropriate response strategies. For acquiring firms, it will facilitate the early detection of potential goodwill impairment risks, thereby supporting more informed bidding, valuation and post-merger integration strategies. For target firms, the framework will enable better preparedness for regulatory scrutiny and mitigate the risk of deal failure due to unresolved goodwill issues. For regulatory authorities, the framework will serve as a scalable and explainable tool for screening high-risk M&A cases based on qualitative disclosures, thus improving regulatory efficiency and objectivity. At a broader level, by relying solely on publicly available textual data, the proposed framework will reduce stakeholders' cognitive burden in processing complex information, an especially relevant contribution in today's data-saturated digital environment.

The remainder of this article is structured as follows. Section 2 reviews the relevant literature on goodwill impairment, M&A comment letters and detection analysis based on NLP techniques. Section 3 describes the data, detection framework and methodological approach. Section 4 introduces the construction of textual features and provides definitions of the variables. Section 5 presents the model development under the detection framework and the explanatory analysis. Section 6 reports the robustness analysis and heterogeneity analysis, and Section 7 reviews the entire study and summarizes the main findings.

2. Literature review

2.1. M&A, goodwill impairment and comment letters

2.1.1. Goodwill impairment in M&A

M&A are motivated by the desire for synergy. Under information asymmetry, buyers and sellers bargain over price: the buyer's maximum willingness to pay equals the fair value of the acquired assets plus expected synergies, while the seller demands a price that is no less than fair value. The final transaction price is determined by both parties' bargaining power and synergy expectations (Hansen, 1987). Synergy, or the future excess revenues of acquired assets, is considered goodwill (Ohlson, 1995). Goodwill impairment indicates a failure to realize expected intangible benefits and synergies, reducing business value. Acquisition premiums have strong explanatory power for goodwill impairment, according to studies based on bankruptcy prediction models (Altman, 1968; Ohlson, 1980; Zmijewski, 1984), with excessive premiums and frequent past M&A failures increasing impairment risk (Gu and Lev, 2011; Olante, 2013; Potepa and Thomas, 2023). The success of post-merger integration in achieving synergies influences impairment risk. Key drivers include the acquirer's prior M&A experience (Zmijewski, 1984), the acquirer's accounting conservatism (Tang et al., 2021), industry familiarity (Bereskin et al., 2018) and performance-based arrangements (Yuan et al., 2020), all of which can reduce the chance of impairment. Furthermore, managerial discretion is critical in goodwill testing; opportunistic or irrational reasons can lead to profit manipulation through impairment, weakening the link between goodwill impairment and firm performance (Ramanna and Watts, 2012; Li and Sloan, 2017; Beatty et al., 2019).

In terms of financial ramifications, goodwill impairment immediately causes unfavorable stock market responses. Investors and analysts reduce firm value because such impairment is viewed as bad news, which causes sharp drops in stock prices (Li et al., 2011). It also exposes unfulfilled strategic goals and unsuccessful post-merger integration. Using data from 2002 to 2006 in the United States, Jarva (2009) discovers a positive correlation between the amount of impairment and decreases in operating cash flows during the first and second years. According to Li and Sloan (2017), impairment also exhibits a negative correlation with lagged return on assets, suggesting a deterioration in operations. Additionally, impairment may lead to stock price crashes (Han and Tang, 2020). In response to these risks, regulatory authorities such as the CSRC frequently issue comment letters regarding questionable M&A practices to enhance disclosure transparency and prevent the financial and market disruptions associated with goodwill impairment.

The pre-acquisition phase is crucial for the establishment of goodwill value, as the quality of decision-making directly influences the level of impairment risk. Valuation bias and inflated premiums are significant drivers of impairment risk (Hayn and Hughes, 2006; Gu and Lev, 2011). During the pricing process, variables such as peer herd behavior, managerial overconfidence and the reduced valuation discipline associated with stock-based remuneration mechanisms can all contribute to high premiums. This results in an overstatement of both the acquisition consideration and the initially recognized goodwill value.

Following acquisition, goodwill impairment is driven by the combined effect of deterioration in economic substance and the failure of governance constraints, a position strongly supported by previous research. For example, unmet synergy effects (Ramanna and Watts, 2012) and management opportunism (Gonçalves et al., 2019) can result in future goodwill write-downs. Management may be motivated by "big bath" accounting or earnings smoothing incentives, utilizing goodwill impairment to alter reported performance. The inherent uncertainty of the impairment testing method allows management to apply discretion for earnings management purposes (Ramanna, 2008).

2.1.2. The functions of M&A comment letters

Comment letters have two main functions: governance and risk warning. The former is fulfilled through the regulatory pressure exerted through comment letters, which compel firms to improve internal controls and enhance transparency, thereby improving the quality of information disclosure (Bozanic et al., 2017; Lyu et al., 2024); optimizing the financial reporting environment (Johnston and Petacchi, 2017); suppressing opportunistic behavior and enhancing the information environment; and ultimately strengthening corporate governance (Gietzmann and Isidro, 2013; Cunningham et al., 2020). In the field of M&A, regulatory authorities issue M&A comment letters in response to suspicious transactions. They may request additional clarification and disclosure and urge firms to revise their restructuring plans to reduce information asymmetry and achieve effective oversight (De Angelo, 1990). Regarding their risk warning function, comment letters act as an "alarm" that sends risk signals to the market. Public comment letters issued by the CSRC on revenue recognition issues frequently elicit significant negative market reactions; even when firms respond and revise their accounting information accordingly, the market reaction remains negative (Dechow et al., 2016; Johnston and Petacchi, 2017; Shen and Tan, 2023). Similarly, M&A comment letters can provide the market with early warning signs of material asset restructuring risks, as firms that receive such letters, compared with those that do not, tend to experience more negative short-term market reactions to their M&A announcements and face greater future goodwill impairment (Kothari et al., 2009).

In summary, M&A comment letters serve a dual function, acting both as regulatory mechanisms that strengthen corporate governance and as early warning signals that convey risks to the market. In the context of M&A transactions, their signaling role helps the market identify uncertainty and impairment risks at an earlier stage, thereby improving information transparency and promoting more efficient resource allocation.

2.2. Detection analysis based on NLP

Because this detection framework is inherently prediction-oriented, it falls within the broader class of prediction-based research. Accordingly, it is necessary to first review the relevant prediction literature. In the context of this study, which develops a detection framework based on textual features, particular attention should be paid to the role of text-related features in prediction-oriented models and how such features contribute to effective risk detection.

2.2.1. Textual feature mining based on NLP

With the rise of Internet technologies, textual data have become a key medium for conveying information, driving advances in NLP. NLP is widely applied for sentiment analysis, intent detection and decision-making tasks (Kusal et al., 2023). In this study, NLP is applied to M&A comment letters to identify embedded regulatory signals and to support the construction of a detection framework for analyzing goodwill impairment risk.

Before constructing the detection framework, relevant textual features must be identified. NLP plays a central role in transforming unstructured language into structured indicators suitable for subsequent analysis.

Prior research explores various types of such features. Huang et al. (2014) show that managers may use overly positive tones in earnings press releases to mislead investors, leading to poorer future performance. Wang et al. (2022) find that lexical indicators have a stronger effect than grammatical ones on crowdfunding outcomes. Song et al. (2024) use dictionary-based methods to extract indicators such as word count, sentence count and sentiment scores from annual reports, highlighting the joint role of textual structure and sentiment in shaping disclosure outcomes. Other studies focus on topic-based textual features. Baker et al. (2023) applies topic modeling to extract macroeconomic themes (e.g., inflation, unemployment, interest rates) from *Wall Street Journal* articles, which are shown to reflect investor sentiment without structured variables. Hajek and Henriques (2024) use two approaches based on Bidirectional Encoder Representations from Transformers (BERT), namely FinBERT and BERTopic, to extract sentiment polarity and topic composition from financial news, demonstrating the predictive value of linguistically rich, explainable features in M&A target identification. Other researchers prioritize predictive accuracy using non-explainable, deep learning-based textual features. Pekar et al. (2024) apply LDA to extract thematic keywords from crowdfunding descriptions to build predictive frameworks. Cheng and Shi (2025) integrate graph neural networks with attention mechanisms to model semantic relations and label dependencies in tourism resource texts, improving prediction via deep text representation.

In summary, NLP enables the extraction of diverse textual features with proven value across domains such as macroeconomics, disclosure and finance. Given the domain-specific nature of M&A comment letters, this study focuses on extracting explainable features that capture both linguistic and thematic content to better reflect the informational value embedded in regulatory text.

2.2.2. Predictive analysis based on textual features

In the domain of predictive analysis based on textual features, prior research can be broadly categorized into two streams: one that integrates textual features with other types of data and another that relies solely on textual information.

In frameworks that combine textual features with structured data, Vo et al. (2021) use a multi-stacking ensemble integrating call log texts and customer profiles for churn prediction, incorporating linguistic and personality-based features. Qian et al. (2022) propose a text-guided attention network that fuses textual, visual and auxiliary data to predict content popularity, although with limited explainability. Chen et al. (2023) combine financial ratios with textual communicative value features from annual reports, improving bankruptcy prediction using models like eXtreme Gradient Boosting (XGBoost) and random forest (RF) algorithms. Borchert et al. (2023) apply transformer-based NLP to extract website text and integrate it with financial data for business failure prediction. Song et al. (2024) build a multidimensional financial distress framework combining textual, financial, managerial and ESG indicators, with gradient boosting proving particularly effective. Zhang and Luo (2023) incorporate basic textual metrics (e.g., length, volume) with user-posted images to predict restaurant survival. Luo et al. (2025) use a spatiotemporal long short-term memory framework merging SNS-based co-occurrence features with structured inputs such as demand history and geography. Hajek and Henriques (2024) develop an ensemble prediction framework integrating financial variables with FinBERT- and BERTopic-derived sentiment and topic indicators, enhancing M&A target prediction under class imbalance. While such integration can improve predictive performance, it also increases complexity and cognitive load, which may hinder decision-making (Sweller, 1988).

In text-only predictive frameworks, prior studies show that unstructured textual data alone can yield effective predictions. Yoon (2012) demonstrate that long-term business signals can be captured using only news texts. Baker et al. (2023) use topic modeling on *Wall Street Journal* articles to show that attention to themes like recession and inflation predicts macroeconomic trends without structured inputs. Barbaglia et al. (2023) apply fine-grained sentiment and topic analysis to large-scale economic news to forecast indicators such as GDP and industrial production. Pekar et al. (2024) build a thematic framework by extracting LDA-based topic features from crowdfunding project descriptions, without using structured data. Cheng and Shi (2025) propose the hierarchical label-aware tourism-informed dual graph attention network framework, combining graph neural networks and attention mechanisms to classify tourism texts based solely on semantic and hierarchical textual features.

While deep learning-based predictive frameworks have become more advanced, many still prioritize performance over transparency, raising concerns about explainability. To address this, researchers such as Vo et al. (2021), Chen et al. (2023) and Pekar et al. (2024) apply SHAP-based methods to explain feature contributions. In the context of M&A comment letters, domain-specific regulatory texts, textual features may capture latent concerns or early impairment risks not yet evident in financial data. Therefore, building explainable predictive frameworks with carefully selected textual features enhances both accuracy and explainability, demonstrating interdisciplinary value by integrating insights from accounting, finance and computational linguistics and making the results more valuable for financial analysis and decision-making (Marrone et al., 2020).

3. Data, framework and methodology

3.1. Data acquisition and sample selection

To address the study's objectives, information related to M&A comment letters and goodwill impairment disclosures is collected. The textual content of M&A comment letters is obtained from Cninfo (https://www.cninfo.com.cn/), the official information disclosure platform designated by the CSRC, which provides authoritative public data such as firm announcements, periodic reports and regulatory comments. Goodwill impairment data are manually retrieved from the notes to listed firms' annual reports, which indicate the impairment of goodwill for all underlying firms. In contrast with the data in the balance sheet's "net goodwill impairment" column, the data in the notes are explicitly tied to specific M&A targets and correspond one-to-one with the M&A transaction comment letters.

The sample period begins in 2014 and ends in 2021, with the fiscal year 2023 serving as the unified observation endpoint for goodwill impairment. The upper bound of 2021 is chosen to ensure a minimum three-year post-acquisition observation window, thereby capturing the majority of goodwill impairment events. The starting year of the sample is determined by institutional background: the Shanghai and Shenzhen stock exchanges began publicly disclosing M&A-related comment letters on their official websites in December 2014. Accordingly, A-share listed companies that completed major asset restructuring transactions between 2014 and 2021 are selected as the research sample. The choice of the sample cutoff year is further supported by the distribution of goodwill impairment timing and the sample structure. As shown in Table 1, approximately 85.11% of first impairment events occur within three years of deal completion, and no first impairment is observed beyond five years. This indicates that goodwill impairment risk is primarily realized during the early post-acquisition period, and that a three-year observation window captures the vast majority of impairment events. Extending the minimum observation period to five years would require the sample to be restricted to transactions completed no later than 2019, which would substantially reduce the sample size and weaken statistical power. Meanwhile, as shown in Table 2, approximately 83.47% of the transactions were completed during 2014–2019 and therefore already have at least five years of post-acquisition observations. This structure implies that most transactions are monitored over a sufficiently long horizon to capture the impairment cycle. Therefore, selecting 2021 as the cutoff year for M&A transactions ensures a minimum observation window while preserving sample size and representativeness. In addition, for transactions completed in earlier years, impairment events are tracked throughout their full subsequent observation periods.

Table 1
Timing of first goodwill impairment after acquisition.

Years after acquisition	Number	Percent (%)	Cum (%)
1 Year	12	4.5802	4.5802
2 Years	114	43.5115	48.0916
3 Years	97	37.0229	85.1145
4 Years	21	8.0153	93.1298
5 Years	18	6.8702	100.0000
Total	262	100.0000	–
Mean Years	2.69		

Table 2
Distribution of M&A transactions by year.

Year	Number	Percent (%)	Cum (%)
2014	41	2.5673	2.5673
2015	322	20.1628	22.7301
2016	354	22.1666	44.8967
2017	250	15.6544	60.5510
2018	222	13.9011	74.4521
2019	144	9.0169	83.4690
2020	142	8.8917	92.3607
2021	122	7.6393	100.0000
Total	1597	100.0000	–

This design preserves impairment information and maximizes the sample size while balancing observational completeness and statistical efficiency.

The sample is also screened according to the following criteria: (1) financial industry listed companies are excluded; (2) ST and *ST listed companies are excluded; (3) abnormal samples with a debt-to-equity ratio greater than 1 or less than 0 are excluded; (4) samples with missing data for primary variables are excluded; and (5) outliers are removed from the dataset using the isolation forest algorithm. Ultimately, a sample of 1597 is obtained for analysis. In addition, this study's sample consists of firms that have experienced material M&A restructuring, which eliminates simple asset purchases, as well as the stricter regulatory and disclosure requirements for restructuring under China's relevant legislation, providing a solid foundation for data collection.[1]

3.2. A detection framework based only on M&A comment letters

As previously stated, the purpose of this study is to provide a framework for detecting goodwill impairment based on the textual features of M&A comment letters, as well as to investigate how these features suggest patterns of impairment. To address the issue of information overload, which can impede decision-making efficiency (Sweller, 1988; Libby and Lipe, 1992; Hwang et al., 2019; Liu et al., 2025), while also avoiding the high costs associated with significant computational resource requirements (Kumar, 2024), the proposed framework relies solely on textual data from M&A comment letters. In other words, the objective is to minimize cognitive resource consumption while focusing on whether comment letters—in which the regulator assumes a constructive intermediate supervisory role—can assist information users in detecting future goodwill impairment risks. To achieve this objective, a detection framework is developed based on the collection of M&A comment letter texts and goodwill impairment data, comprising the following three core components, as illustrated in Fig. 1.

(1) **Measurement of M&A comment letter textual features:** Referring to prior research (Kimbrough and Louis, 2011), seven thematic keywords (e.g. "financial information" and "accounting policies") are obtained. Building on the approach proposed by Wu (2024), a synonym self-generation approach based on BERT (SSAB) is subsequently implemented to generate a more comprehensive and abundant set of thematic keywords. Based on these keywords, term frequency–inverse document frequency (TF-IDF) is measured to quantify the textual thematic features of the M&A comment letters. Furthermore, four basic textual features, including text readability and text volume, are calculated using straightforward statistical text analysis methods.

[1] According to the Administrative Measures for Material Asset Reorganization of Listed Firms, "material asset reorganization" refers to the purchase or sale of assets by a listed firm and its controlling firm outside its daily business activities, resulting in significant changes in its main business, assets and revenue. Usually, the key indicators of the subject assets of a transaction (such as total assets, net assets and operating income) constitute a major asset reorganization when they account for more than 50% of the corresponding indicators of the listed firm.

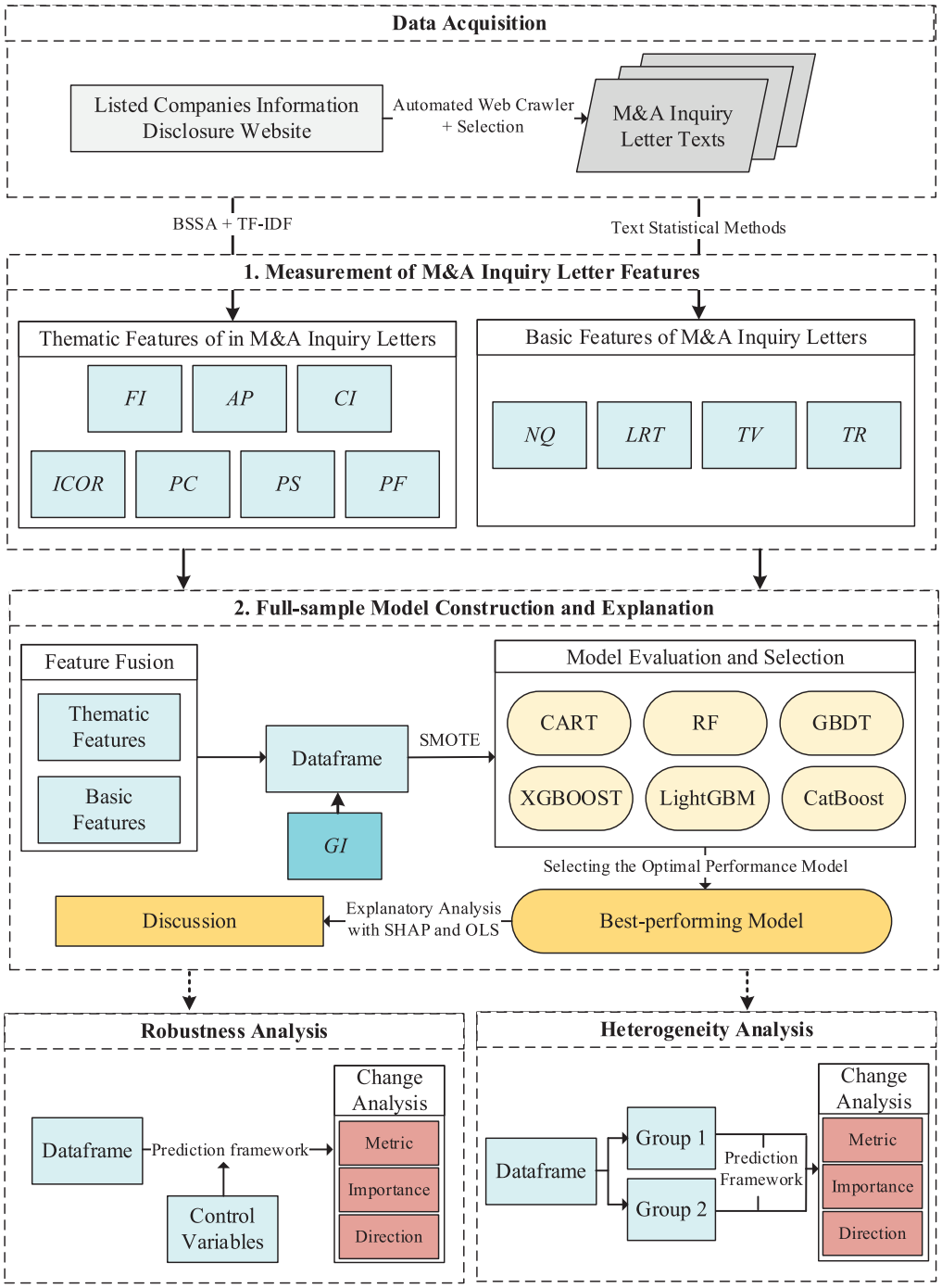


Fig. 1. A detection framework based only on M&A comment letters. Note: Thematic features include *FI* (Financial Information), *AP* (Accounting Policies), *CI* (Customer Information), *ICOR* (Industry Characteristics and Operational Risks), *PC* (Performance Commitments), *PS* (Profitability Sustainability), and *PF* (Profit Forecast). Basic textual features include NQ (Number of Questions), *LRT* (Limited Response Time), *TV* (Text Volume), and TR (Text Readability). Thematic features are constructed using BERT (Bidirectional Encoder Representations from Transformers) for synonym expansion and TF-IDF (Term Frequency-Inverse Document Frequency) for weighting, with weights aggregated to form feature scores. Class imbalance is handled using SMOTE (Synthetic Minority Over-sampling Technique). Model explanatory analysis is conducted by SHAP (Shapley Additive Explanations), and supplementary analysis is performed using OLS (Ordinary Least Squares).

(2) **Development and explanatory analysis of the detection model:** Textual features extracted from M&A comment letters are matched with goodwill impairment outcomes observed over the entire post-M&A period. Multiple tree-based machine learning algorithms are trained to identify early warning signals of subsequent goodwill impairment risk, and the optimal model is selected based on out-of-sample performance. To enhance model transparency, SHAP values are combined with OLS regression to assess and illustrate the influence of each textual feature. Control variables capturing corporate financial characteristics and M&A transaction details are included in robustness checks to examine the incremental predictive value of textual features. The study also conducts heterogeneity analyses to investigate how different governance structures shape the detection power of comment letter text.

(3) **Measurement of M&A comment letter textual features:** Referring to previous research (Kimbrough and Louis, 2011), seven thematic keywords (e.g., ‘‘financial information’’ and ‘‘accounting policies’’) are obtained. Building on the approach proposed by Wu (2024), an SSAB is then implemented to generate a more comprehensive and abundant set of thematic keywords. Based on these keywords, TF-IDF is measured to quantify the textual thematic features of the M&A comment letters. Four basic textual features, including text readability and text volume, are calculated using straightforward statistical text analysis methods.

(4) **Development and explanatory analysis of the detection model:** Textual features extracted from M&A comment letters are matched with goodwill impairment outcomes observed over the entire post-M&A period. Multiple tree-based machine learning algorithms are trained to identify early warning signals of subsequent goodwill impairment risk, and the optimal model is selected based on out-of-sample performance. To enhance model transparency, SHAP values are combined with OLS regression to assess and illustrate the influence of each textual feature. Control variables capturing corporate financial characteristics and M&A transaction details are included in robustness checks to examine the incremental predictive value of textual features. The study also conducts heterogeneity analyses to investigate how different governance structures shape the detection power of comment letter text.

3.3. Methodology

As shown in Fig. 1, the implementation of the research framework involves a series of methods of data acquisition, integration and preprocessing. Some of these methods are conventional or easily substitutable and are thus not detailed here; this section focuses on the key feature measurement techniques and data analysis methods.

3.3.1. Synonym self-generation approach based on BERT

As illustrated in the framework, the measurement of thematic features requires the use of an SSAB. As shown in Fig. 2, the SSAB used in this study has two main components: a BERT-based vectorization approach and a cosine similarity-based synonym self-generation method.

BERT is a context-based pre-trained model that utilizes a self-attention mechanism and a bidirectional transformer structure to learn language representations, demonstrating exceptional performance across various NLP tasks (Devlin et al., 2018). Unlike Word2Vec, which generates static word vectors and fails to address the polysemy of a word in different contexts, BERT employs bidirectional encoding to dynamically capture the contextual meanings of words, thereby generating context-aware word vectors. This makes it particularly well suited for analyzing semantically rich text data, such as consumer reviews and regulatory comment letters. As shown in Fig. 2, the process of BERT-based word vectorization is as follows. First, the text undergoes tokenization, where each word (including special tokens like [CLS] and [SEP]) is converted into embeddings through the embedding layer. These embeddings are then passed through multiple transformer layers, where the self-attention mechanism captures contextual relationships between words. Finally, the model outputs context-aware word vectors for each word, which can be utilized for downstream tasks such as text classification or prediction. The corpus used in this study comprises Chinese M&A comment letters, and the BERT Base Chinese model, specifically designed for Chinese corpora, is employed to generate word vectors. As the original thematic keywords are manually derived by the researchers from M&A comment letters, they are likewise included in the vectorization process.

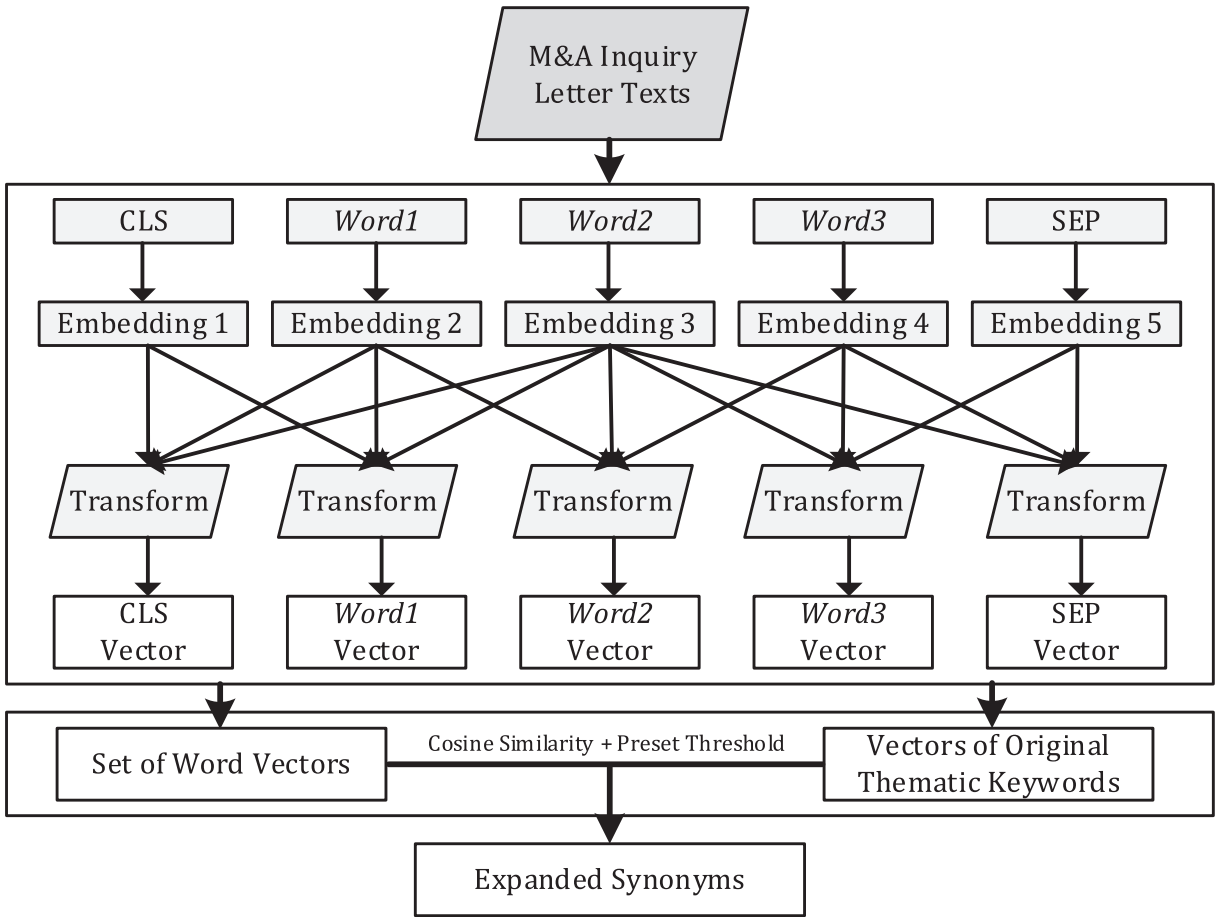


Fig. 2. Synonym Self-generation Approach based on BERT.

After vectorizing the words, the cosine similarity between all word vectors in the M&A comment letter corpus and the original thematic keywords is calculated. A preset cosine similarity threshold is then applied to filter and identify additional synonyms in the corpus that align with the original thematic keywords. This process further enhances and expands the original set of thematic keywords, ensuring comprehensive coverage of the thematic features within the corpus.

3.3.2. TF-IDF

As illustrated in the research framework, the measurement of thematic features also requires the use of the TF-IDF algorithm. TF-IDF is a statistical measure commonly used in information retrieval and text mining to evaluate the importance of a term within a collection of documents. This algorithm combines the frequency of a term within a specific document (TF) with its frequency across the entire document collection (IDF), thus determining the relative importance of the term in a given document. Compared with approaches that rely solely on word frequency calculations, TF-IDF offers the significant advantage of accounting not only for a word's frequency within a single document but also for its overall prevalence or rarity across an entire document corpus, thereby providing a more effective measure of the word's importance to a document (Zhang et al., 2011). This approach is particularly practical for studies involving M&A comment letters across diverse listed firm samples, as it facilitates comparisons within the sample set. As shown in Equation Set (1), this study applies the TF-IDF algorithm to calculate the importance of the expanded thematic keywords extracted from the corpus of M&A comment letters.

$$
\begin{aligned}
&TF(t,d) = \frac{f_{t,d}}{n_d} \\
&IDF(t,D) = \log\left(\frac{N}{n_t}\right) \\
&TF - IDF(t,d,D) = TF(t,d) \times IDF(t,D)
\end{aligned} \tag{1}
$$

where $f_{t,d}$ represents the frequency of a thematic keyword t in document d, n_d is the total number of terms in document d, N is the total number of documents in the entire corpus and n_t is the number of documents containing the keyword t.

3.3.3. Tree-based machine learning algorithms

Considering the research objectives and the dataset, this study selects a series of tree-based machine learning algorithms for performance comparison. These algorithms include classification and regression trees (CART), gradient boosting decision trees (GBDT), RF, XGBoost, Light Gradient Boosting Machine (LightGBM) and Category Boosting (CatBoost).

CART is a foundational decision tree algorithm that partitions data using metrics like the Gini index (Breiman, 2001). While it is highly interpretable and easy to understand, it tends to overfit in complex datasets, particularly when the depth of the tree increases. Its simplicity, however, makes it an excellent starting point for comparison with more advanced models. RF, an ensemble decision tree method that improves on CART through bagging (bootstrap aggregating) and random feature sampling, thereby enhancing model robustness and reducing overfitting (Loh, 2011). RF generally offers better performance than a single decision tree but often lags behind boosting methods like GBDT in terms of accuracy (Tang et al., 2025). The GBDT algorithm builds trees sequentially, each one correcting the errors of its predecessor by focusing on residuals through gradient descent (Friedman, 2001). This process results in a highly accurate model, especially in cases where capturing complex relationships between features is crucial. Both XGBoost and LightGBM are advanced versions of GBDT, with XGBoost incorporating regularization and parallelization to improve efficiency and scalability, especially in large datasets (Chen and Guestrin, 2016). LightGBM further optimizes training speed and memory usage with techniques like gradient-based one-side sampling and exclusive feature bundling, making it particularly well suited for large datasets (Ke et al., 2017). However, both XGBoost and LightGBM require careful tuning to avoid overfitting, especially when handling highly dimensional data. CatBoost excels in handling categorical features natively, which is particularly beneficial for datasets with many categorical variables. Its ordered boosting technique reduces overfitting more effectively than do other boosting methods (Prokhorenkova et al., 2018). Additionally, CatBoost minimizes the need for extensive preprocessing of categorical data, a step usually required by other models.

Tree-based algorithms are adopted because they combine explainability, robustness to complex feature interactions and high predictive performance. Compared with neural networks, tree-based models are more transparent and better suited to settings where explainability is required. Compared with support vector machine methods, they provide more direct explanations of how features affect the model output and are often easier to deploy in applied settings. Importantly, tree-based models work naturally with SHAP-based explainability, which enables quantification of each feature's contribution to predicted impairment risk in terms of both relative importance and direction of effect (Lundberg et al., 2020). In addition, these methods perform well for high-dimensional and non-linear problems, consistent with regulatory text-derived features, which may exhibit threshold effects and complex interactions (Nalepa and Kawulok, 2019; Bentéjac and Csörgő, 2021). Finally, boosted tree methods such as XGBoost, LightGBM and CatBoost repeatedly show strong predictive performance in structured empirical applications, making them appropriate for this study's goodwill impairment detection task (Tang et al., 2025). These advantages are essential for developing an effective monitoring and detection framework.

However, the results of these explainable machine learning methods are typically limited to visualizations such as bar charts for feature importance, one-dimensional scatter plots of SHAP values and two-dimensional scatter plots of feature values against SHAP values. While informative, these visualizations lack statistical rigor. To address this limitation, this study combines the model's SHAP values with feature values for secondary analysis using OLS regression. This method allows for the calculation of coefficients, statistical signif-

icance and other key parameters, with the goal of uncovering the linear relationships between features and the outcome variable within the context of the nonlinear model.

Notably, the machine learning research in this work is prediction-oriented (or detection-oriented), with the main goal of determining the incremental predictive value of textual elements of M&A comment letters in detecting goodwill impairment risk. Furthermore, using SHAP-based interpretability analysis, this study characterizes the process by which each characteristic influences the model's predictions, looking at both their relative relevance and the direction of their contribution to risk prediction. As a result, the findings should be considered to provide empirical evidence of the predictive value, risk-signaling role and contribution patterns of comment letter linguistic elements.

4. Description and measurement of features

4.1. Textual features of M&A comment letters

M&A comment letters are formal documents issued by regulatory authorities to firms involved in M&A activities. These letters encompass textual features, including thematic features such as regulatory focus, and basic features such as the number of questions raised. These two categories of features respectively reflect a comment letter's emphasis on specific themes and its approaches to information delivery, and they address the question of what constitutes an M&A comment letter by combining surface-level textual information with deeper content. To further elucidate these features, the feature description table in Table 3 provides a detailed overview of the classification and methods of measurement of these features. The specific descriptions are as follows:

4.1.1. Textual thematic features of M&A comment letters

Referring to the reading and analysis of M&A comment letters conducted by Kimbrough and Louis (2011), this study focuses on the following textual thematic features: customer information (*CI*), profit forecast (*PF*), profitability sustainability (*PS*), performance commitments (*PC*), financial information (*FI*), accounting policies (*AP*) and industry characteristics and operational risks (*ICOR*). These features reflect regulatory authorities' focus on specific aspects of M&A activities and represent the information extracted by readers based on their knowledge background. From the perspective of attributes, *CI*, *PF*, *PC* and *PS* are directly related to profitability and are typically the focal points of stakeholders. *CI* and *PF* estimates are the most important

Table 3
Description of features in M&A comment letters and *GI*.

Category	Name	Feature	Attribute	Measurement
Textual Thematic Feature	Customer Information	*CI*	Profitability	Synonym expansion using BERT model
	Profit Forecast	*PF*		
	Profitability Sustainability	*PS*		TF-IDF weighting method
	Performance Commitments	*PC*		Weight accumulation
	Financial Information	*FI*	Operational and Financial Stability	
	Accounting Policies	*AP*		
	Industry Characteristics and Operational Risks	*ICOR*		
Textual basic Feature	Number of Questions	*NQ*	Information Demand Volume	Manual statistics
	Limited Response Time	*LRT*	Information Transparency	
	Text Readability	*TR*		MinMax (word count) + MinMax (character count)
	Text Volume	*TV*		MinMax (adverb frequency) + MinMax (conjunction frequency)/Sentence Count
Target Variable	Goodwill Impairment	*GI*	Indicates whether goodwill impairment occurred	$GI > 0$ classified as positive; otherwise, negative classification

inputs to M&A valuation models. They provide fundamental evidence of the target company's current and past operational quality, serve as a true basis for future estimates and reflect critical current data. In contrast, *PC* and *PS* are forward-looking indicators that reflect identification of future trends and associated risks. Meanwhile, *FI*, *AP* and *ICOR* are not directly tied to profitability but can be understood as features related to operational and financial stability.

Textual thematic features are measured using the BSSA and the TF-IDF algorithm. As the BSSA requires baseline keywords to be prepared in advance, this study uses the thematic keywords published by Kimbrough and Louis (2011) as baseline keywords and vectors them with the BERT model. Cosine similarity analysis is applied to the loaded corpus (M&A comment letters) to automatically identify words with similarity scores above 0.8 or within the top 100 ranking, which are considered potential synonyms generated by the algorithm. As algorithm-generated synonyms may include noise, manual filtering is conducted to produce the seven thematic expanded keywords listed in Table 4. Although the number of keywords varies across thematic features, the weights calculated through TF-IDF effectively reflect the relative importance of the same feature across texts. This makes them suitable for comparing comment letter texts. As the keywords for each feature are independent, their calculations are also relatively independent, avoiding mutual interference and ensuring that the final data analysis results remain unaffected. Finally, the cumulative weights of keywords within each thematic category are calculated to derive the final measurement for each textual thematic feature. These thematic features are all represented as numerical values, where higher feature values indicate a greater presence of that type of information in the text, reflecting a higher degree of regulatory emphasis on the corresponding aspect.

4.1.2. Textual basic features of M&A comment letters

There are four textual basic features: number of questions (*NQ*), limited response time (*LRT*), text readability (*TR*) and text volume (*TV*). These features provide a direct representation of the basic information contained in M&A comment letters and reflect readers' perceptions of this information. *NQ* is related to the volume of information demand, indicating the regulatory authority's level of scrutiny of the M&A activity. *LRT*, *TR* and *TV* are associated with information transparency, reflecting the information quantity, ease of comprehension and timeliness of information disclosure and thus indicating the regulatory authority's demand for knowledge regarding the M&A activity and the urgency of that demand.

Among these textual basic features, *NQ*, *TR* and *TV* are numerical variables, where larger values respectively indicate a greater number of questions, larger text volume and poorer readability. *LRT* is a categorical variable, with a value of 1 indicating the presence of a limited response time and 0 indicating otherwise. Some features can be directly extracted from reading the text, such as *NQ* and *LRT*. *TV* is calculated based on the

Table 4
Keywords of thematic features in M&A comment letters (Partial)[a].

Thematic feature	Keywords (Partial)
CI	Major clients, client concentration, client categories, client resources, orders, sales orders, shipments, top five clients, target client groups (20 keywords)
PF	Financial analysis, cost and expense forecasting, firm performance, market demand forecasting, cash flow forecasting, sales forecasting, performance model, predicted net profit, predicted profit margin (22 keywords)
PS	Financial independence, sustainable operation assumptions, client reliance, technology introduction, operational autonomy, client churn risk, tax incentives, dependency, long-term existence, self-operation (60 keywords)
PC	Compensation fees, compensation commitments, guarantee agreements, returns, payment conditions, economic compensation, performance commitments, performance forecasts, penalties (10 keywords)
FI	Financial reports, financial decisions, acquisition costs, operational income, liability structure, equity investment, total profit, intangible assets, accounts payable/turnover ratio, capital appreciation (269 keywords)
AP	Reporting compilation basis, reporting standards, expense analysis, cost analysis, fee recognition, quarterly profits, valuation models, equity valuation, merger scope, return policies (118 keywords)
ICOR	Insurance risks, product competitiveness, innovation capabilities, pricing risks, core technologies, operational risks, competitive trends, market share, litigation risks, financing arrangements (104 keywords)

[a] The final set of keywords is extensive and cannot be fully displayed here due to space limitations. For the complete list, please contact the authors.

number of words and characters in the text. First, the text undergoes preprocessing to remove unnecessary characters. Next, the Jieba model is used to segment the text, filtering out stop words and single-character words to retain meaningful tokens. Subsequently, the total number of words and characters is counted, and these two indicators are standardized using MinMax scaling to avoid dimension inconsistencies. The result is a comprehensive text volume metric reflecting the overall scale of the text. For *TR*, as the dataset consists of Chinese text, traditional readability metrics designed for English text may not be applicable. Therefore, referring to the Flesch reading ease test, *TR* is calculated by analyzing sentence length and the frequency of adverbs and conjunctions. First, the text is segmented into multiple sentences, and the word count (i.e., sentence length) and the number of adverbs and conjunctions within each sentence are calculated. Longer sentences and higher frequencies of adverbs and conjunctions typically indicate greater reading difficulty. Finally, these metrics are used to generate two readability indicators, which are normalized using MinMax scaling to produce a comprehensive readability score. Higher readability scores indicate more complex text, while lower scores suggest simpler text.

4.2. Goodwill impairment

Goodwill impairment (*GI*) refers to the cumulative impairment amount disclosed in financial statement notes by the acquiring entity since the date of the acquisition. This study primarily aims to investigate whether the textual features of M&A comment letters can effectively detect a firm's likelihood of experiencing goodwill impairment over the post-M&A period, measured from deal completion to the last observable fiscal year in the sample. Following Glaum et al. (2018), the cumulative impairment amount is obtained and transformed into a binary classification variable. If the impairment amount is greater than 0, it is classified as *GI* in the positive class, indicating the occurrence of goodwill impairment during the entire post-M&A period from deal completion to the last observable fiscal year in the sample. Otherwise, it is classified as *GI* in the negative class, indicating that no goodwill impairment is observed over the post-acquisition observation window. For consistency with the other textual features of M&A comment letters, the description of *GI* is also included in Table 3.

5. Model development and explanatory analysis

Following the proposed detection framework, based solely on the textual features of M&A comment letters (descriptive statistics are presented in Table 5), a predictive model is constructed to identify potential goodwill impairment risk. The model design emphasizes the integration of explainable textual features to enhance both predictive performance and interpretability. The subsequent sections describe the modeling procedure and examine the contributions of key linguistic features to overall model performance, highlighting how explainability deepens understanding of regulatory text signals.

Table 5
Descriptive statistics of textual features in M&A comment letters.

Feature	Mean	Std	Min	25%	50%	75%	Max	Skew	Kurtosis	Count
FI	52.7707	33.8982	0.0000	25.6436	48.4541	70.0031	215.3436	0.9934	1.3676	1597
AP	13.1618	12.1458	0.0000	3.7975	9.8111	18.1691	78.2310	1.3381	1.8546	1597
CI	2.8853	5.7947	0.0000	0.0000	0.0000	3.0553	37.2633	3.0193	9.8561	1597
ICOR	10.3985	10.8165	0.0000	2.5333	7.9771	14.7553	62.9596	1.5021	2.2361	1597
PC	21.7154	15.9423	0.0000	9.6304	17.8943	30.0123	110.6348	1.4435	2.5816	1597
PS	8.4643	8.9207	0.0000	2.5005	6.4523	10.9739	77.8589	2.1970	9.4397	1597
PF	2.1082	3.8342	0.0000	0.0000	0.0000	3.0838	30.5948	2.4577	7.3169	1597
NQ	11.4158	5.6568	1.0000	7.0000	10.0000	15.0000	40.0000	1.0302	1.2373	1597
LRT	0.2198	0.4142	0.0000	0.0000	0.0000	0.0000	1.0000	1.3534	−0.1684	1597
TV	0.2003	0.1150	0.0000	0.1086	0.1837	0.2703	0.7904	0.9508	0.9518	1597
TR	0.3062	0.1379	0.0000	0.2054	0.2835	0.3996	0.9649	0.9120	1.1408	1597
GI	0.1641	0.3704	0.0000	0.0000	0.0000	0.0000	1.0000	1.8143	1.2917	1597

5.1. Model development

After extracting and integrating textual features from M&A comment letters and completing data preprocessing, this study constructs detection models using multiple tree-based machine learning algorithms. The outcome variable exhibits a pronounced class imbalance. Specifically, among M&A events associated with firms receiving comment letters, only about 16.41% ultimately result in goodwill impairment, while the remaining observations do not, corresponding to an approximate class ratio of 1–5. To mitigate potential model bias caused by class imbalance, the synthetic minority over-sampling technique is employed to augment minority class observations and improve the class distribution (Wang and Yan, 2025). Based on this balanced dataset, several tree-based machine learning algorithms are applied, including CART, RF, GBDT, XGBoost, LightGBM and CatBoost.

In line with the research objectives, the outcome variable is specified as binary and all models are formulated as binary classification models. To clearly delineate the training, validation and test design, a hold-out test set is first separated from the full sample and is strictly excluded from all subsequent model training and hyperparameter tuning procedures. This test set is reserved exclusively for the final evaluation of model performance. Hyperparameter optimization is then performed on the remaining data using five-fold cross-validation. Specifically, the non-test sample is partitioned into five mutually exclusive subsets; in each iteration, four subsets are used for model training, while the remaining subset serves as the validation set, with the validation role rotating across folds. Through this process, validation performance is assessed across multiple data partitions, enabling the robust comparison of candidate models and selection of optimal hyperparameter configurations.

Based on the optimal hyperparameters for each algorithm, these parameter settings are reloaded into the corresponding models, which are then re-estimated using the training data. Finally, detection performance is evaluated on the pre-held-out test set. Model performance is assessed using standard evaluation metrics, including accuracy, precision, recall, F1 score and the area under the receiver operating characteristic curve. These metrics are calculated based on the confusion matrix components of true positive (TP), false positive (FP), false negative (FN) and true negative (TN), thereby providing a comprehensive assessment of the detection capability of each machine learning model.

The performance metrics of the best-performing models are comprehensively compared, and the model developed using the CatBoost algorithm (the CatBoost model) exhibits the best performance. When the textual features of M&A comment letters and *GI* data are input into the candidate models, the performance comparison results shown in Table 6 are obtained. As indicated in Table 6, the CatBoost model achieves over 0.9 across all metrics and outperforms the other models, demonstrating its exceptional detection capability.

The exceptional performance of the CatBoost model in this study is closely aligned with both the inherent strengths of the algorithm and the characteristics of the dataset. First, the CatBoost algorithm excels in handling datasets containing categorical variables. The dataset includes several categorical features, such as *LRT* in M&A comment letters, which CatBoost processes natively without the need for extensive preprocessing such as one-hot encoding. This capability reduces preprocessing complexity and errors, making it particularly efficient for datasets with categorical variables, such as those in M&A comment letters. Second, the algorithm's ordered boosting technique helps prevent overfitting and improves generalization, especially when dealing with complex, high-dimensional data. The dataset used in this study is high-dimensional and contains complex relationships between features. CatBoost's ordered boosting helps capture these intricate interactions while preventing overfitting, ensuring that the model generalizes well to new data. Additionally, CatBoost's

Table 6
Model performance comparison.

Metric	CART	RF	GBDT	XGBoost	LightGBM	**CatBoost**
Accuracy	0.8202	0.8914	0.8652	0.8951	0.9064	**0.9326**
Precision	0.7874	0.8521	0.8684	0.8704	0.8868	**0.9012**
Recall	0.9257	0.9730	0.8919	0.9527	0.9527	**0.9865**
F1 Score	0.8509	0.9085	0.8800	0.9097	0.9186	**0.9419**
AUC	0.8669	0.9454	0.9300	0.9602	0.9542	**0.9761**

regularization techniques and computational efficiency allow it to uncover intricate relationships within the data, leading to superior accuracy. The dataset possesses both categorical and continuous features, and CatBoost's regularization ensures that the model captures these complex relationships effectively, improving accuracy without overfitting. Finally, this capability makes CatBoost particularly well suited for subsequent explanatory analysis, as it can reveal hidden patterns in the data while maintaining robustness to noise and missing values. Even if some noisy data remain in the dataset, CatBoost's robustness allows it to maintain reliable performance and uncover meaningful insights, making it ideal for both detection and explanatory analysis.

5.2. Explanatory analysis

Detection models are developed using various machine learning algorithms with parameter optimization, followed by a performance comparison. The results indicate that the CatBoost model outperforms the others across all metrics, making it the preferred algorithm for the subsequent explanatory analysis. The main hyperparameters for the CatBoost model at its optimal performance are as follows: $learning_rate = 0.05$ and $depth = 8$. However, despite offering precise detection performance, machine learning models inherently face the "black box" effect, which limits their explainability. To address this issue, this study employs SHAP, a widely used explainable approach for tree-based machine learning models, to explain the CatBoost model. Compared with traditional feature importance metrics, SHAP values provide a unified measurement for each feature, explaining its marginal contribution across different samples while ensuring that the sum of all feature contributions matches the model's output. This increases the consistency and intuitiveness of the explanation results (Lundberg et al., 2020). SHAP values provide significant advantages in uncovering the decision-making logic of complex models and clarifying how specific textual features contribute to the likelihood of detecting goodwill impairment. This approach is particularly effective in identifying the directional relationships between features of M&A comment letters and impairment.

Using explainable ML and SHAP, Fig. 3 shows feature importance (mean absolute SHAP); longer bars represent a greater contribution to detecting goodwill impairment. Fig. 4 summarizes SHAP by observation:

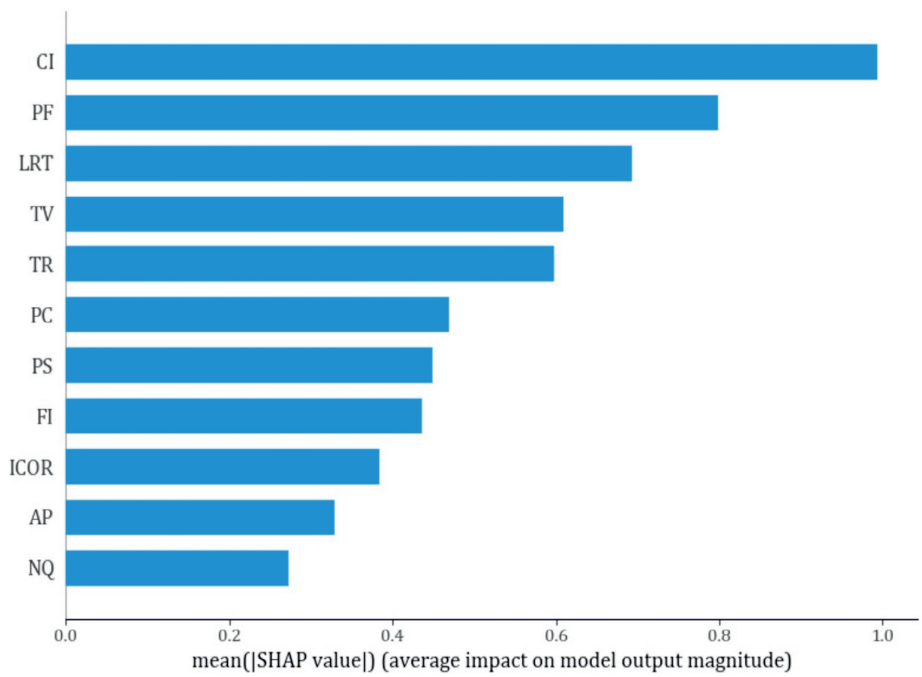


Fig. 3. SHAP importance chart for textual features of comment letters on *GI*.

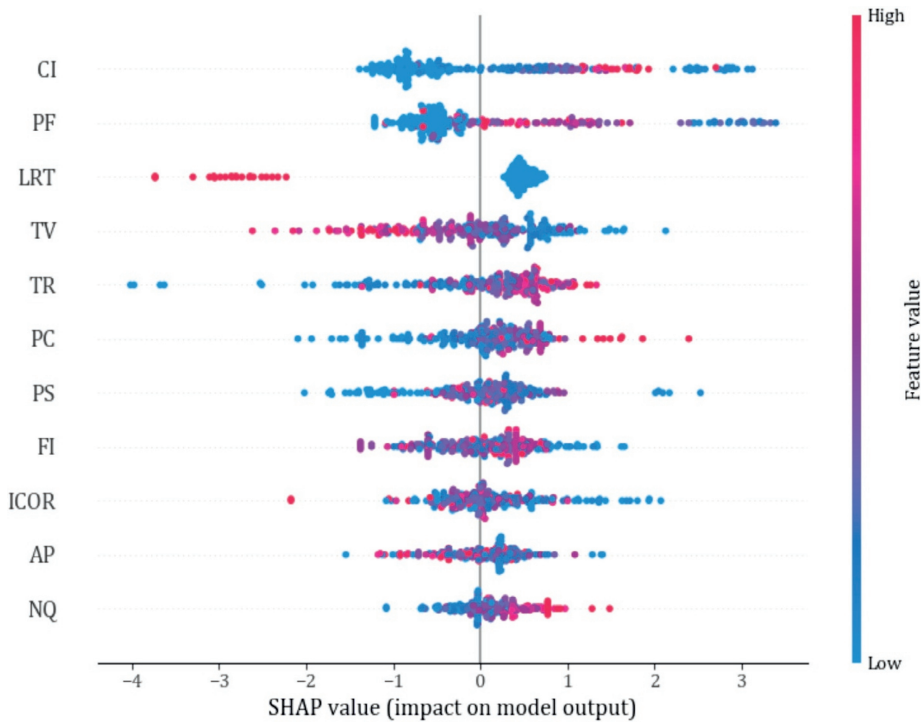


Fig. 4. SHAP summary plot for textual features of comment letters on *GI*.

color encodes low–high feature values (blue to red); the x-axis represents SHAP values; and red clustering on the positive side (blue on the negative) indicates a positive association with GI, with the reverse indicating a negative association. Because Fig. 4 is one-dimensional per feature, Fig. 5 adds scatter plots of feature values vs. SHAP with OLS fits (coefficients, *p*-values, R^2) for finer granularity. Table 7 consolidates the average SHAP values, feature ranks and OLS statistics. Together, these visuals clarify how textual features drive goodwill impairment detection and surface linear structure within a nonlinear model.

The *GI* detection importance of the textual features of M&A comment letters is determined by the nature of the features themselves. In terms of feature importance (or contribution to detection outcomes), *CI* and *PF*, as thematic features of the comment letters, have the strongest *GI* detection ability. Other thematic features follow closely behind the basic features, such as *LRT*, *TV* and *TR*, but rank higher than *NQ*. This pattern indicates that the effectiveness of a feature in detecting *GI* is more closely linked to its underlying informational content than to its categorical classification.

Specifically, thematic features closely associated with profitability, especially those related to M&A valuation, including *CI* and *PF*, demonstrate greater detection capability with respect to subsequent goodwill impairment. In contrast, *PC* and *PS*, which primarily involve assessments of future uncertainty, are less significant than customer information and profit forecasts. Their importance is also lower than that of core indicators reflecting regulatory information demands, including *LRT*, *TV* and *TR*. This pattern arises because excessive acquisition premiums driven by inaccurate valuations constitute a primary cause of goodwill impairment (Gu and Lev, 2011; Li et al., 2011). Comment letters that emphasize customer information and profit estimates are more informative, as they address key evidence regarding the target firm's current and historical operating quality and provide a concrete basis for future projections. Profit forecasts, in particular, represent quantitative financial expectations that form the foundation of valuation calculations and can directly reflect revenue quality, customer concentration, forecast consistency and the potential of a transaction to generate long-term value. In contrast, comment letters focusing on performance commitments and profit sustainability rely more heavily on forward-looking statements, which are difficult to verify or validate ex ante. Consequently, such information is less effective in accurately identifying goodwill impairment risk and receives lower weight in the detection model.

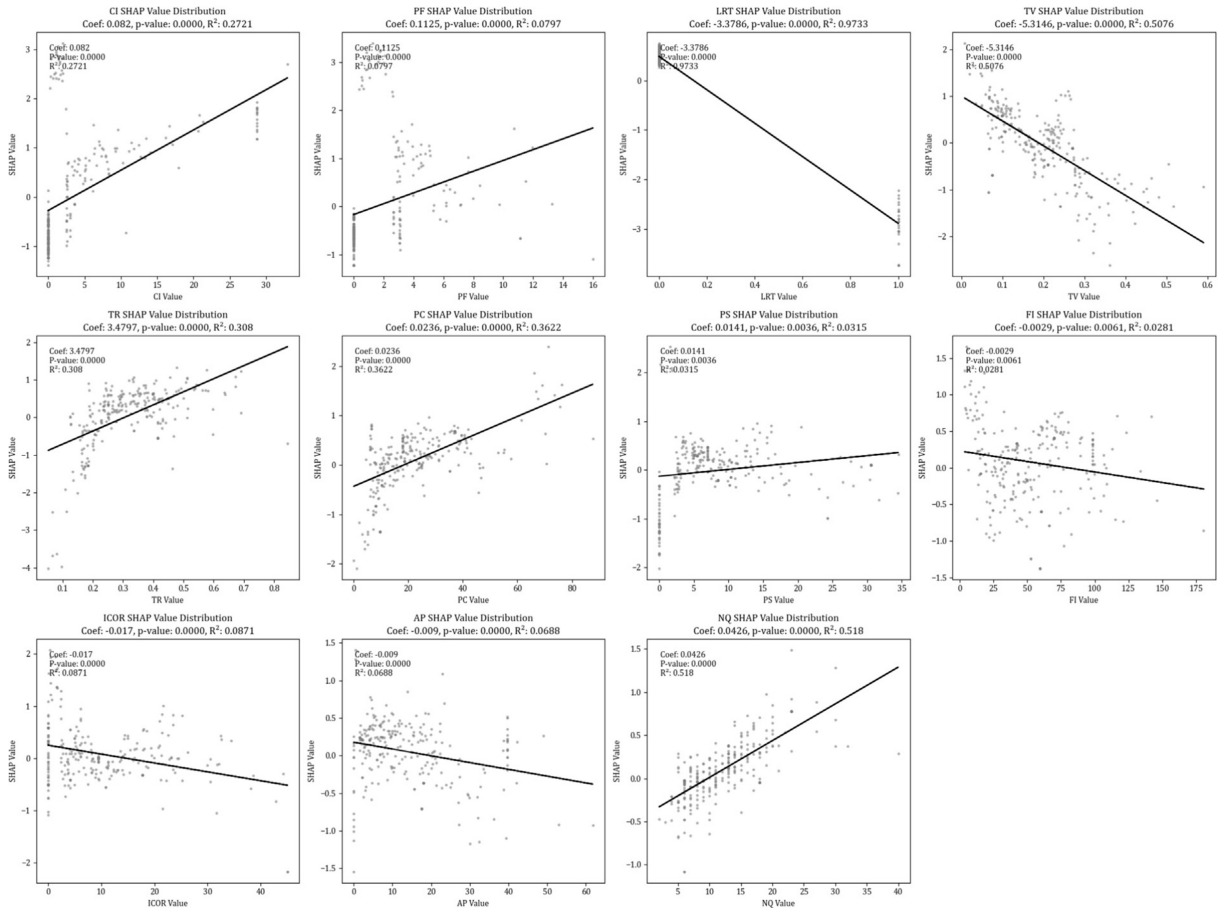


Fig. 5. Scatter plots of feature values vs. SHAP values with OLS fit results.

Table 7
Comprehensive overview of explanatory analysis.

Feature	Mean |SHAP| value	Importance order	OLS coefficients	*P*-value	R^2
CI	0.9931	1	0.0820	0.0000	0.2721
PF	0.7980	2	0.1125	0.0000	0.0797
LRT	0.6933	3	−3.3786	0.0000	0.9733
TV	0.6090	4	−5.3146	0.0000	0.5076
TR	0.5976	5	3.4797	0.0000	0.3080
PC	0.4686	6	0.0236	0.0000	0.3622
PS	0.4488	7	0.0141	0.0036	0.0315
FI	0.4364	8	−0.0029	0.0061	0.0281
ICOR	0.3833	9	−0.0170	0.0000	0.0871
AP	0.3286	10	−0.0090	0.0000	0.0688
NQ	0.2727	11	0.0426	0.0000	0.5180

Although basic features such as *LRT*, *TV* and *TR* are not directly linked to future profits or financial conditions, they play a critical role in ensuring information transparency by regulating the timeliness and adequacy of disclosures, thereby providing a sideways indication of the risk of *GI*. As for *NQ*, while it reflects the regulatory authority's scrutiny of M&A activities, the number of questions may not necessarily reveal

the depth of the issues or their direct correlation with *GI*. Thus, its informational value may be weaker in practice than that of other features, resulting in a lower level of importance.

Regulatory authorities' comments on M&A activities with a focus on profitability-related aspects can reveal a higher risk of subsequent *GI*, whereas their attention to operational and financial stability is associated with a reduced likelihood of *GI*. In terms of detection direction, thematic features such as *CI*, *PF*, *PC* and *PS* are positively associated with the likelihood of goodwill impairment. Regulatory authorities' focus on profitability-related themes in M&A activities may signal an increased likelihood of *GI*. Specifically, *CI* directly relates to the source of future revenue; *PF*, *PC* and *PS* are closely tied to the actual returns from M&A activities. When regulatory authorities raise comments about these aspects, it often signals concerns about the achievability of expected profitability levels. The failure to realize the anticipated competitive advantages of an M&A could result in the devaluation of goodwill, reflecting the projected future benefits of the acquisition. In contrast, thematic features such as *FI*, *ICOR* and *AP* are negatively associated with the likelihood of goodwill impairment. When a comment letter focuses on technical topics such as financial information, industry characteristics and accounting policies, it indicates that the regulatory focus is on standardization rather than risk. Such comments do not directly reveal M&A failures or valuation overestimations, and they may instead be negatively correlated with future goodwill impairment, implying a lower risk of such impairment.

When the readability of M&A comment letters is poor or the number of questions is high, the likelihood of *GI* is often higher. Conversely, when the text volume is large and a limited response time is specified, the risk of *GI* tends to be lower. In terms of detection direction, *TR* and *NQ* are positively associated with the likelihood of goodwill impairment, suggesting that lower readability and a higher volume of questions tend to signal increased impairment risk. Specifically, poor readability raises the difficulty of understanding the content, making it harder for stakeholders to accurately understand the comment letter. Additionally, a greater number of questions may imply a higher level of regulatory skepticism regarding the M&A activities, creating more pressure and uncertainty for the firm and its stakeholders, which in turn elevates the likelihood of *GI*. In contrast, *LRT* and *TV* exert a negative influence on *GI*, suggesting that when comment letters include a limited response time or have a larger text volume, the risk of *GI* is lower. A limited response time may prompt firms to respond to regulatory comments more promptly and cautiously, thereby reducing information uncertainty. Similarly, a larger text volume may indicate more comprehensive and detailed information disclosure, helping to alleviate market concerns and signaling a lower likelihood of *GI*.

6. Additional analysis

6.1. Robustness analysis

Based on the detection model constructed from textual features of M&A comment letters, this study further introduces a set of numerical control variables to account for the potential effects of corporate financial characteristics and M&A transaction details on goodwill impairment. These control variables fall into two categories. The first category consists of financial variables related to target firm characteristics, including the book-to-market ratio (*BM*), return on assets (*ROA*), firm size (*Size*), firm growth (*GROWTH*), leverage (*LEV*) and firm age (*AGE*). These variables capture differences in valuation attributes, profitability, scale, growth potential, financial risk and operational maturity. The second category includes transaction-specific variables related to M&A activities, such as total deal value (*VALUE*), which reflects the overall scale of the acquisition.

Incorporating the above control variables together with the textual features of M&A comment letters into the CART, RF, GBDT, XGBoost, LightGBM, and CatBoost models and applying the same 9:1 data-splitting criterion and five-fold cross-validation yields the detection performance results reported in Table 8. Compared with the baseline model without control variables, the inclusion of these controls does not lead to a significant improvement; instead, most evaluation metrics exhibit a noticeable decline. As shown in Table 8, although CatBoost remains the best-performing model, its accuracy, recall and F1 scores decrease after the control

Table 8
Model performance comparison with control variable.

Metric	CART	RF	GBDT	XGBoost	LightGBM	CatBoost
Accuracy	0.8052	0.8951	0.8764	0.9101	0.9101	**0.9139**
Precision	0.7759	0.8571	0.8710	0.8875	0.8875	**0.8981**
Recall	0.9122	0.9730	0.9122	**0.9595**	**0.9595**	0.9527
F1	0.8385	0.9114	0.8911	0.9221	0.9221	**0.9246**
AUC	0.8030	0.9385	0.9429	0.9631	0.9686	**0.9754**

variables are added, indicating that these controls provide limited incremental detection value. This result suggests that the textual content of M&A comment letters already contains sufficient information closely related to goodwill impairment risk, while corporate financial characteristics and M&A transaction features may overlap with the information embedded in the text or, by introducing additional noise and complexity, increase cognitive burden and thereby weaken detection performance.

This finding highlights the superior applicability of detection models based on M&A comment letter language. Regulatory comment letters, unlike corporate financial data, constitute objective verification from an outside third party. They include not only information about the acquirer but also checks on the accuracy of the target's information. Indicators at the target company level are substantially more effective in detecting goodwill impairment than are typical firm-level measurements.

To explain feature importance, the SHAP approach is used in conjunction with the CatBoost model, again including the control variables. When used with SHAP–OLS regression analysis, this approach investigates the marginal effect of each feature on model output. The relevant findings are shown in Table 9. *CI*, *PF* and *LRT* continue to rank as the most important textual elements. This shows that even after controlling for firm financial and transaction variables, linguistic elements still contain critical signals linked with goodwill impairment risk, indicating high informational content and explanatory power. A closer look at the relative importance rankings of textual features reveals that the addition of control variables has no effect on their core structure: *CI*, *PF* and *LRT* remain in the top three positions, demonstrating a high level of stability. Meanwhile, the relative relevance of several content-oriented textual criteria (such as *PS* and *AP*) increases, while the ranking of formal features, such as *TV*, decreases. The direction of the OLS coefficients for all textual features stays steady, indicating that the marginal effects represented by these features in the model do not change fundamentally. This implies that despite the inclusion of control factors, the influence of textual features on goodwill impairment risk remains consistent, verifying their effectiveness and stability in the model.

Overall, the addition of control factors does not reduce the importance of textual elements in detecting goodwill impairment. The importance structure and detection direction of the textual traits remain relatively consistent, offering additional evidence of robustness to support the baseline conclusions. This, in turn, shows that textual elements do have incremental detection potential.

Table 9
Comprehensive overview of explanatory analysis with control variable.

Feature	Mean |SHAP| value	Importance order	OLS coefficients	*P*-value	R^2
CI	0.8129	1	0.0510	0	0.1356
PF	0.6245	2	0.0805	0	0.0719
LRT	0.5274	3	−2.6345	0	0.9832
PS	0.432	4	0.0246	0	0.0999
TR	0.3989	5	2.1097	0	0.2495
PC	0.3267	6	0.0187	0	0.4423
AP	0.2813	7	−0.0083	0	0.0788
TV	0.2561	8	−2.1857	0	0.4554
ICOR	0.2109	9	−0.0088	0	0.0724
FI	0.1786	10	−0.0017	0.0003	0.0492
NQ	0.1384	11	0.0205	0	0.4238

6.2. Heterogeneity analysis

To further investigate the differential detection effects of various textual features on goodwill impairment across subsamples, this section explores the underlying economic implications of the detection model. Comment letters are often issued during the pre-acquisition phase, with the primary goals of reviewing and criticizing the reliability of revealed valuation-related information, thereby serving as a risk alert. Although they cannot be heavily active during the post-merger integration period, comment letters can operate as a type of regulatory intervention, encouraging management to re-evaluate the merger and integration process and thereby exerting a governance influence. When challenged to regulatory inspection, organizations with greater internal governance are likely to respond more quickly and proactively, implementing the necessary corrective actions and offering appropriate feedback (Li et al., 2025). As a result, comment letters have a bigger governance effect on organizations with higher governance quality, which may reduce their risk-signaling role. As a result, the overall potential of comment letter language to detect future goodwill impairment is slightly diminished in such situations. In contrast, in organizations with poor internal governance, chaotic internal processes frequently impede quick and effective solutions. In these circumstances, comment letters are generally used to indicate potential future dangers. This study uses ownership concentration (*OC*) and independent director ratio (*IDR*) as proxies for corporate governance and conducts subsample analysis accordingly. Based on the detection framework, which relies solely on M&A comment letters, and under consistent parameter settings, multiple heterogeneous explainable detection models are developed, with the full-sample model serving as the baseline. Comparing detection performance across different governance-based subsamples enables us to evaluate how the explanatory power of textual features varies under distinct governance environments.

6.2.1. Ownership concentration grouping

Table 10 presents the comparative performance of the detection model across different *OC* subgroups, using the full-sample model as the baseline. The results report five evaluation metrics, namely accuracy, precision, recall, F1 score and AUC, for both low- and high-*OC* subgroups, along with the corresponding performance gains relative to the baseline. In the high-*OC* subgroup, the model outperforms the baseline in terms of accuracy, precision, F1 score and AUC. This suggests that in firms with more concentrated ownership structures, the text-based model demonstrates stronger detection power for identifying goodwill impairment risk. In contrast, the model's performance declines in the low-*OC* subgroup, indicating a notable weakening in detection effectiveness for these firms. These findings suggest that ownership concentration has a significant impact on the marginal informational value of regulatory comment letters. A higher level of ownership concentration often implies greater control by significant shareholders and poorer checks-and-balances mechanisms, which can lead to the breakdown of corporate governance systems and impede effective corrective actions in response to regulatory demands. In such cases, comment letters sent by regulatory bodies serve as a significant risk-signaling method, with the written content expressing more pronounced risks. This improves the model's ability to detect the company's future goodwill impairment (Fig. 6).

Further analysis of the feature-level results for firms with high *OC* and low *OC* reveals notable shifts in the importance rankings and explanatory structures of text-based features relative to the full-sample model. As shown in Fig. 7, in the high-concentration group ($OC = \text{high}$), the explanatory structure of the model shifts markedly toward internal, firm-controlled variables, with a diminished reliance on external disclosure-related features. Specifically, the importance of *FI* rises from eighth to first place, *AP* from 10th to 6th and *PC* from sixth to

Table 10
Comparison of model detection performance across different *OC* groups.

Metric	Baseline model	$OC = \text{High}$		$OC = \text{Low}$	
		Metric value	Metric gain (%)	Metric value	Metric gain (%)
Accuracy	0.9326	0.9450	1.3261	0.8625	−7.5201
Precision	0.9761	0.9123	1.2310	0.8667	−3.8285
Recall	0.9012	0.9811	−0.5470	0.8864	−10.1297
F1 Score	0.9865	0.9455	0.3824	0.8764	−6.9572
AUC	0.9419	0.9943	1.8645	0.9306	−4.6595

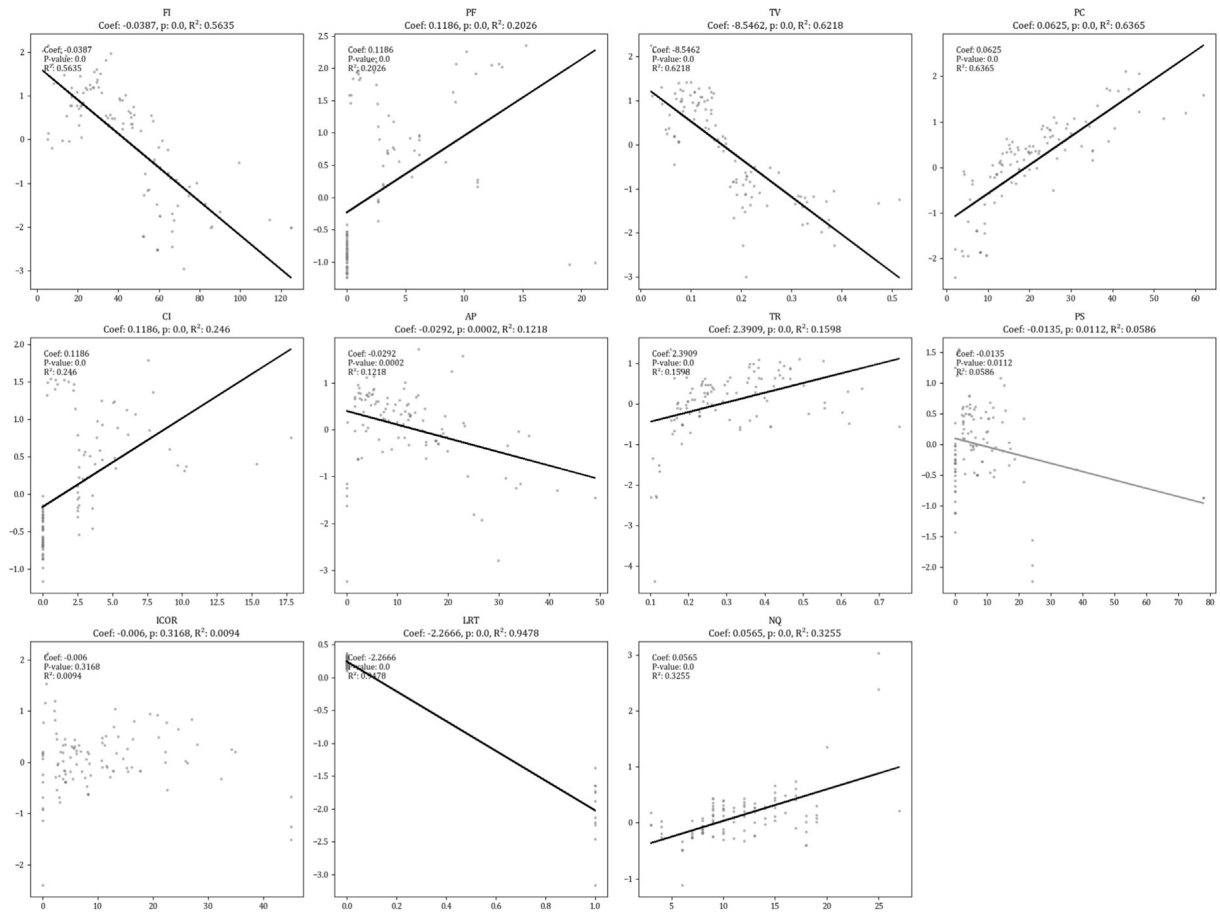


Fig. 6. Scatter plots of feature values vs. SHAP values with OLS Fit results in OC = High.

fourth. These results suggest that regulatory attention to firms' financial performance and policy discretion during M&A processes is more effective in identifying risks arising from potential opportunistic behavior under dominant shareholder control. In contrast, features related to external disclosure quality, such as *CI*, *LRT* and *TR*, decline significantly in importance. This pattern implies that in firms with weak (vs. strong) shareholder oversight, management enjoys greater discretion in information presentation, making accounting policy choices and financial disclosures more informative for predicting subsequent goodwill impairment. Conversely, in the low-concentration group ($OC = \text{low}$), where governance mechanisms are generally stronger and managerial discretion is more constrained, the predictive value of regulatory response and textual quality improves. Notably, Fig. 8 shows that *LRT* and *TR* rise from third and fifth places in the full sample to second and third places, respectively, indicating that regulatory timeliness and textual clarity become key risk indicators of *GI* in better-governed firms. *FI* also increases in importance, moving from eighth to sixth place, and its association with model output shifts from negative to positive, suggesting a transition from compliance signals to risk indicators. Meanwhile, the importance of *PF* and *PC* declines, and *PS* loses statistical significance entirely.

6.2.2. Independent director ratio grouping

Table 11 presents the performance comparison of subgroup models based on different levels of *IDR* and the baseline model developed using the full sample. Specifically, it reports the values of accuracy, precision, recall, F1 score and AUC for both the high and low *IDR* groups, along with their respective metric gains relative to the baseline model. The results show that for firms with a lower *IDR*, the model achieves notable improve-

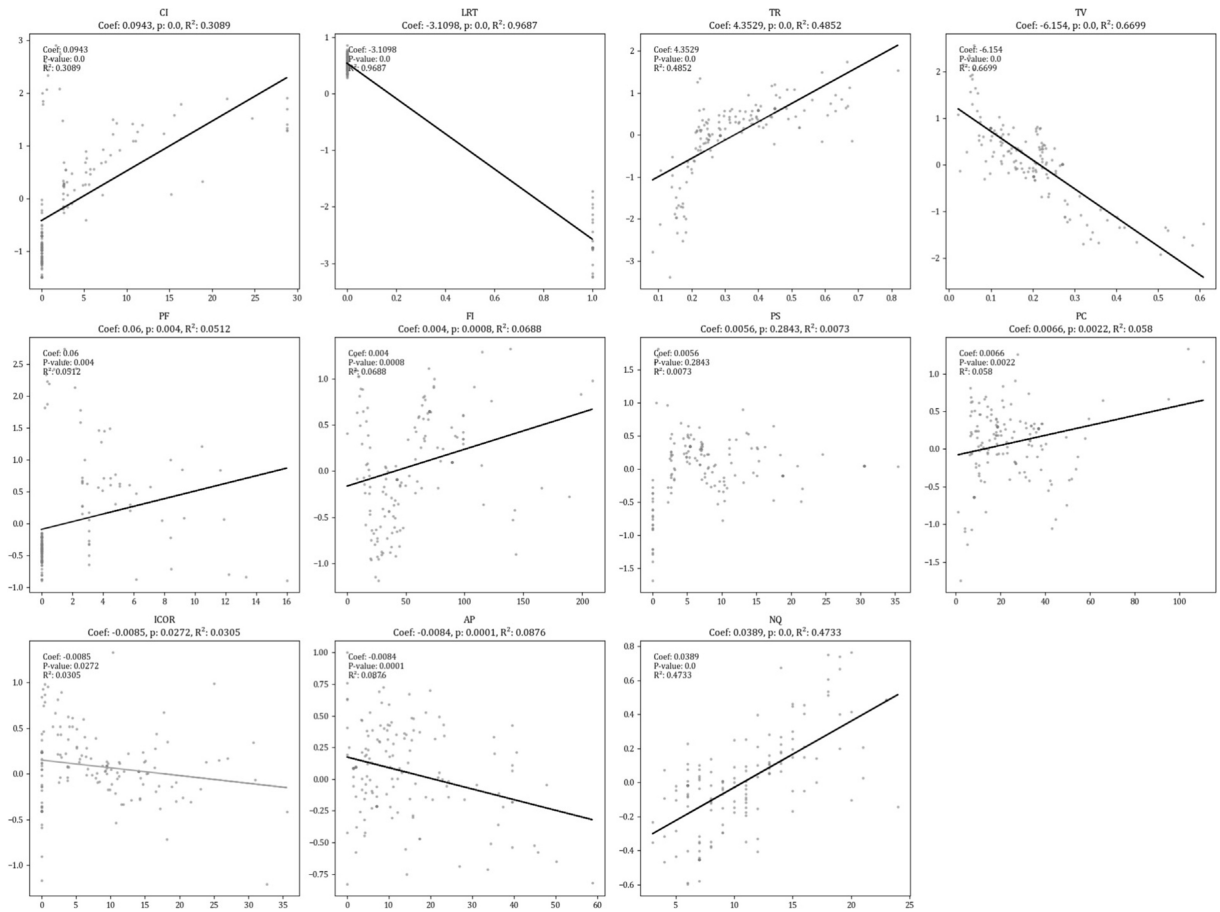


Fig. 7. Scatter plots of feature values vs. SHAP values with OLS fit results in OC = Low.

ments in accuracy, recall and AUC, indicating that textual features exhibit stronger detection power for goodwill impairment in this group. In contrast, for firms with a higher *IDR*, all metrics except recall decline, suggesting that the explanatory power of textual information weakens in firms with stronger board independence, leading to a marked deterioration in detection performance. Further investigation of this heterogeneous finding reveals that enterprises with a lower proportion of independent directors frequently have inferior internal governance structures, with the board's supervisory function curtailed, restricting timely and effective responses to regulatory supervision. In such cases, comment letters made by regulatory bodies on business M&A operations have a greater marginal information value. They efficiently enhance the information accessible to market players and are generally used to provide risk signals. As a result, detection models built using the linguistic elements of M&A comment letters show superior risk recognition capabilities within such organizations, resulting in better detection performance. In contrast, organizations with a higher share of independent directors have stronger governance frameworks and more uniform information disclosure policies. Under these conditions, the incremental information value of regulatory comment letters is restricted, and the model's capacity to anticipate goodwill impairment risk also decreases.

After dividing the sample based on *IDR*, the ranking and statistical significance of feature importance in the model exhibit more pronounced adjustments than those observed under the *OC* grouping. As illustrated in Fig. 8, firms with a high *IDR* demonstrate an increased reliance on features associated with textual clarity and objective information, while showing reduced sensitivity to externally suggestive signals. Specifically, *TR* rises from fifth place in the full-sample analysis to become the most influential feature in this subgroup,

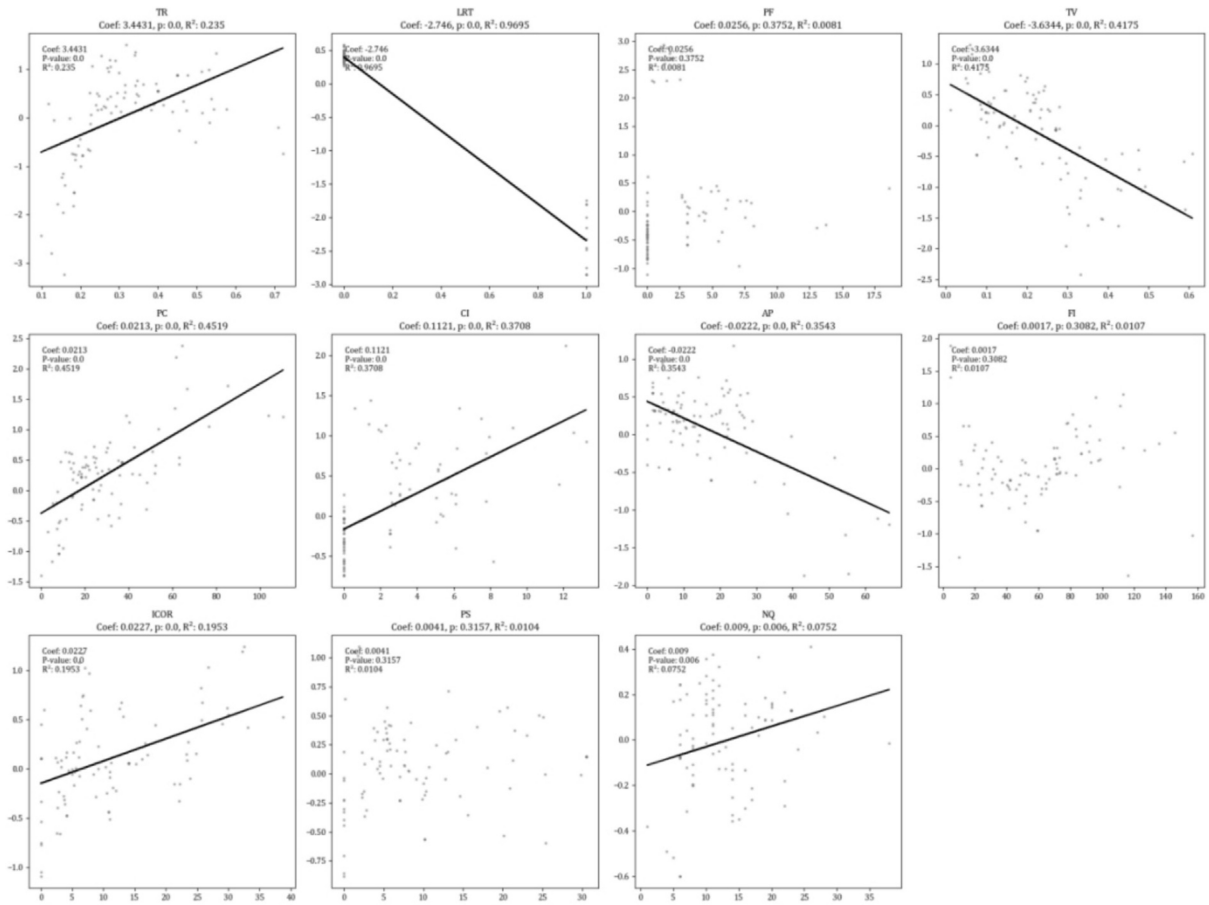


Fig. 8. Scatter plots of feature values vs. SHAP values with OLS fit results in IDR = High.

Table 11
Comparison of model detection performance across different *IDR* groups.

Metric	Baseline model	*IDR* = High		*IDR* = Low	
		Metric value	Metric gain (%)	Metric value	Metric gain (%)
Accuracy	0.9326	0.8889	−4.7041	0.9408	0.8788
Precision	0.9761	0.8704	−3.4200	0.9130	1.3126
Recall	0.9012	0.9216	−6.5578	0.9767	0.9934
F1 Score	0.9865	0.8952	−4.9448	0.9438	−0.2010
AUC	0.9419	0.9422	−3.4721	0.9905	1.4753

indicating that in firms with strong governance structures, the clarity of regulatory discourse serves as a critical detection signal of goodwill impairment risk. Similarly, *LRT* advances from third to second place, reflecting the heightened informational value of regulatory efficiency in well-governed environments. The ranking of *AP* also improves, rising from 11th to 7th. In contrast, *CI* declines from first to sixth place, and *PS* drops from 7th to 10th, losing statistical significance. These shifts suggest that although comments on profitability outlook may initially appear to convey negative signals, investors in firms with more robust governance mechanisms

are more likely to place trust in management's responses. As a result, such disclosures are regarded as routine rather than as indicators of elevated risk, thereby diminishing their detection relevance. Furthermore, the association between *ICOR* and the model output shifts from negative to positive. This reversal implies that in firms with a higher *IDR*, regulatory attention to industry-level risks may be perceived as an affirmation of operational soundness rather than as a warning sign. Effective internal governance enables businesses to respond to and resolve challenges more quickly and efficiently. Anticipating this, the market expects such businesses to proactively manage and reduce the specific business risks identified in comment letters. As a result, the warning effect of these danger signals is mitigated, and their informational value falls. In contrast, the qualities of the regulatory action itself (such as language clarity) become leading indicators of the quality and effectiveness of the company's following correction, increasing their importance.

As shown in Fig. 9, for firms with a low *IDR*, the importance of *PF* drops from second to fourth place, while *TV* rises from fourth to second. *TR* falls from fifth to ninth place, and *ICOR* climbs from ninth to fifth. These shifts suggest that for firms with lower independent board representation and weaker governance, the market tends to be more cautious about the completeness and transparency of internal disclosures. As a result, greater reliance is placed on features of regulatory comment letters that reflect external relevance, structural clarity or objective verifiability. For example, longer comment letters are often regarded as signaling increased regulatory scrutiny, which amplifies their linkage with potential goodwill impairment. Similarly, industry-related disclosures, due to their objective nature and external verifiability, gain more explanatory power and marginal information value in environments characterized by weak governance and pronounced informa-

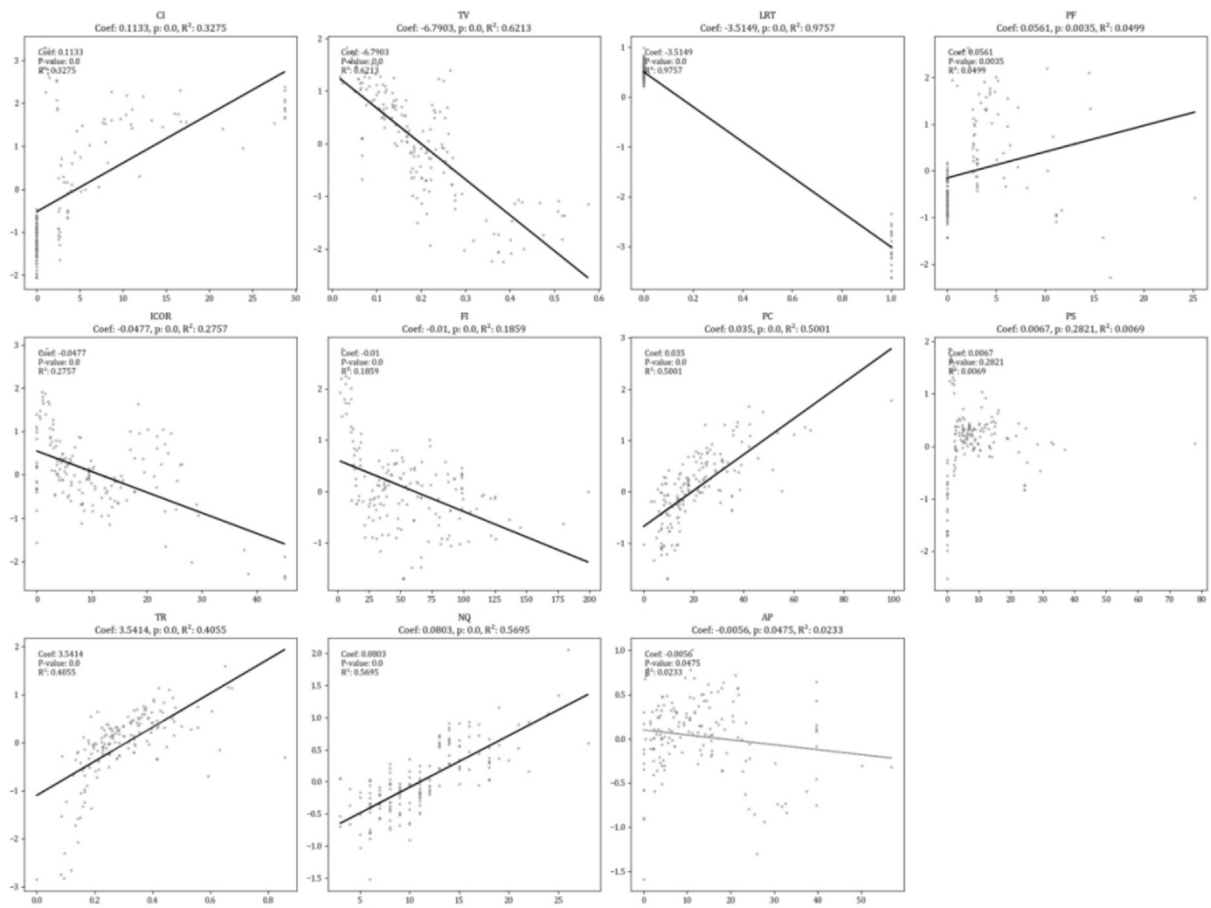


Fig. 9. Scatter plots of feature values vs. SHAP values with OLS fit results in IDR = Low.

tion asymmetry. This demonstrates that in an environment with poor governance and a lack of trust, investors are forced to abandon the interpretative logic based on management's response in favor of more objective information and regulatory intensity signals that are less susceptible to manipulation, to compensate for severe information asymmetry. *PS* also loses its statistical significance in this subsample, indicating that its informational value is constrained under conditions of poor governance.

Notably, *PS* loses significance in both sub-samples. This reinforces the critical role of the governance environment. In the low-governance group, the signal is weakened due to a lack of credibility; in the high-governance group, risk sensitivity is reduced because risk is perceived to be mitigated by the company's effective governance actions.

7. Conclusion

This study develops an explainable detection framework that relies on the textual content of M&A comment letters to forecast future goodwill impairment risk. By combining NLP with advanced machine learning algorithms, the model enhances stakeholders' ability to detect early warning signals embedded in regulatory discourse. The framework introduces an SSAB to expand domain-specific thematic keywords, which are then quantified using TF-IDF metrics. Meanwhile, basic textual features are extracted through standard text analytics techniques. These features are paired with post-M&A goodwill impairment outcomes and used to train multiple tree-based machine learning classifiers, among which CatBoost achieves the best detection performance. To overcome the "black box" limitation of nonlinear models and enhance explainability, SHAP is applied to quantify each feature's detection contribution, while OLS regression on SHAP values provides linear insights into direction and strength. This dual explanatory structure ensures that textual risk signals are both explainable and theoretically grounded. Furthermore, the study incorporates corporate financial and M&A-related features as control variables, demonstrating that M&A comment letter textual features indeed exhibit strong detection power for goodwill impairment, and the relative importance structure and detection direction of textual features remain stable even after introducing the control variables. Additionally, heterogeneity analyses based on corporate governance variables, *OC* and *IDR*, illustrate how the detection effect of textual themes differs across governance contexts. These results reveal that governance mechanisms moderate the informational role of regulatory communication in shaping post-M&A risk outcomes. Overall, this study demonstrates that M&A comment letters serve not only as compliance instruments but also as repositories of forward-looking accounting and economic signals. The findings enrich the literature on proactive regulation, goodwill impairment detection and explainable AI in accounting. Practically, the framework will provide useful guidance for acquirers on pricing and integration, for targets on disclosure management and for regulators on supervisory targeting. By relying entirely on publicly accessible regulatory texts, it will also reduce cognitive complexity for stakeholders and advance scalable, explainable risk analytics in dynamic M&A settings. The following key findings are derived from the empirical analysis.

(1) The importance of M&A comment letter features to goodwill impairment detection is largely determined by the intrinsic nature of the features. The thematic characteristics of the most relevant real-world foundational information for M&A valuation, especially *CI* and *PF*, exhibit the strongest detection power. The thematic characteristics of forward-looking information that is more future-oriented and difficult to immediately verify, such as *PC* and *PS*, and those linked to operational and financial stability, including *FI*, *AP* and *ICOR*, are less influential. Basic textual attributes, particularly *LRT* and *TR*, also rank highly, reflecting the transparency and clarity of disclosure. All outperform the *NQ* feature. These findings highlight the need to move beyond surface-level explanations of comment letters. Stakeholders should apply professional judgment and weigh features by their relevance to goodwill impairment detection. Greater analytical attention should be given to indicators such as *CI* and *PF*, which directly signal recent performance trends. Among basic features, those reflecting regulatory transparency and urgency, including *LRT*, *TV* and *TR*, deserve particular focus. Features with weaker detection value may be considered selectively, depending on cognitive and decision-making priorities.

(2) Textual features of M&A comment letters exhibit distinct detection directions regarding goodwill impairment. When regulatory authorities emphasize profitability-related themes such as *CI*, *PF*, *PC* and *PS*, this typically signals heightened concern about a firm's earnings outlook. Such emphasis may reflect skepticism toward post-M&A performance and serve as an early indicator of increased goodwill impairment risk. Stakeholders should therefore pay close attention to comment letters highlighting these themes, conduct comprehensive evaluations of profitability prospects during the integration phase and adjust their strategies to protect their interests. When comment letters focus on themes related to operational and financial stability, including *FI*, *ICOR* and *AP*, they are generally associated with a lower probability of goodwill impairment. These comments are often routine in nature, in contrast with those concerning future profitability such as customer information and profit forecasts, and may indicate that regulatory attention is focused on general compliance. Stakeholders can use these themes to assess a firm's financial soundness and market position, enabling more informed decisions. Regarding basic textual features, low *TR* and a high *NQ* are typically linked to a higher risk of goodwill impairment. Letters that are difficult to understand or contain numerous inquiries may suggest inadequate disclosure or unresolved uncertainties. Stakeholders should remain cautious when such characteristics appear, as they may imply underlying risks. In contrast, comment letters with greater *TV* and explicitly stated *LRT* often reflect more comprehensive and timely regulatory engagement, indicating stronger information transparency and a reduced likelihood of goodwill impairment. These two features can serve as practical indicators for evaluating disclosure quality and managerial responsiveness.

(3) Based on heterogeneity modeling within a unified detection framework, the explanatory power of textual features for goodwill impairment detection varies notably across governance environments, demonstrating the empirical value of the framework. In firms with weaker governance, characterized by higher ownership concentration or a lower proportion of independent directors, the model performs better, indicating that regulatory comment letters convey stronger external risk signals when internal transparency is limited. In such settings, external oversight effectively compensates for information asymmetry, enhancing detection precision. Conversely, in firms with stronger governance, reflected by dispersed ownership or a higher ratio of independent directors, the model's detection performance declines, suggesting that as internal governance improves, the marginal value of external regulatory discourse diminishes. Moreover, the relative importance and significance of textual features shift across governance subgroups, indicating that the interpretive value of textual risk signals is context-dependent and shaped by institutional structures.

Inevitably, this study has several limitations. First, although the selected textual features are grounded in existing theory, some thematic constructs, such as customer information, do not have fully comparable numerical proxies in practice. To more systematically assess the marginal contribution of textual information, future research could incorporate a richer set of transaction-specific variables, where data availability permits, and conduct incremental detection tests or design comparison models. Second, the modeling framework primarily quantifies textual information through topic intensity and semantic expression. Although transaction amounts are captured, other numerical information embedded in comment letters is not separately extracted and incorporated into the model, because of heterogeneity in presentation, inconsistent definitions and the availability of substitutable sources. Instead, such information is obtained from external databases and introduced as control variables to the robustness analyses. Future work could leverage advances in information extraction and text-structuring methods to integrate embedded numerical data with semantic features and examine their complementary roles in detection performance. Third, the interpretative analysis focuses on the significance and directional contribution of features in terms of goodwill impairment risk detection, because the study is prediction-oriented rather than designed for causal identification. As a result, potential endogenous mechanisms, such as firms' behavioral adjustments following comment letters, cannot be separately identified. To evaluate the governance implications of regulatory comment letters and the mechanisms through which they may influence M&A decisions and subsequent goodwill impairment, future research could combine quasi-experimental approaches with causal identification designs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aaldering, L.J., Leker, J., Song, C.H., 2019. Uncovering the dynamics of market convergence through M&A. Technol. Forecast. Soc. Change 138, 95–114. https://doi.org/10.1016/j.techfore.2018.08.012.

AbuGhazaleh, N.M., Al-Hares, O.M., Roberts, C., 2011. Accounting discretion in goodwill impairments: UK evidence. J. Int. Financ. Manag. Account. 22, 165–204. https://doi.org/10.1111/j.1467-646X.2011.01049.x.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finance 23, 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.

Ai, Q., Tan, H., 2018. The intra-firm knowledge transfer in the outward M&A of EMNCs: evidence from Chinese manufacturing firms. Asia Pac. J. Manag. 35, 399–425. https://doi.org/10.1007/s10490-017-9518-z.

Barbaglia, L., Consoli, S., Manzan, S., 2023. Forecasting with economic news. J. Bus. Econ. Stat. 41, 708–719. https://doi.org/10.1080/07350015.2022.2060988.

Beatty, A., Cheng, L., Zach, T., 2019. Nonrecurring items in debt contracts. Contemp. Account. Res. 36, 139–167. https://doi.org/10.1111/1911-3846.12426.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

Bereskin, F., Byun, S.K., Officer, M.S., Oh, J.M., 2018. The effect of cultural similarity on mergers and acquisitions: evidence from corporate social responsibility. J. Financ. Quant. Anal. 53, 1995–2039. https://doi.org/10.1017/S0022109018000716.

Borchert, P., Coussement, K., De Caigny, A., De Weerdt, J., 2023. Extending business failure prediction models with textual website content using deep learning. Eur. J. Oper. Res. 306, 348–357. https://doi.org/10.1016/j.ejor.2022.06.060.

Bostwick, E.D., Krieger, K., Lambert, S.L., 2016. Relevance of goodwill impairments to cash flow prediction and forecasting. J. Account. Audit. Financ. 31, 339–364. https://doi.org/10.1177/0148558x15596201.

Bozanic, Z., Dietrich, J.R., Johnson, B.A., 2017. SEC comment letters and firm disclosure. J. Account. Public Policy 36, 337–357. https://doi.org/10.1016/j.jaccpubpol.2017.07.004.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Bybee, L., Kelly, B., Su, Y., 2023. Narrative asset pricing: Interpretable systematic risk factors from news text. The Review of Financial Studies 36 (12), 4759–4787. https://doi.org/10.1093/rfs/hhad042.

Chen, C.H., Lin, M.J.J., 2011. An assessment of post-M&A integration influences on new product development performance: an empirical analysis from China, Taiwan, and HK. Asia Pac. J. Manag. 28, 807–831. https://doi.org/10.1007/s10490-009-9173-0.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining. ACM, New York, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, T.K., Liao, H.H., Chen, G.D., Kang, W.H., Lin, Y.C., 2023. Bankruptcy prediction using machine learning models with the text-based communicative value of annual reports. Expert Syst. Appl. 233. https://doi.org/10.1016/j.eswa.2023.120714 120714.

Cheng, Q., Shi, W., 2025. Hierarchical multi-label text classification of tourism resources using a label-aware dual graph attention network. Inf. Process. Manag. 62. https://doi.org/10.1016/j.ipm.2024.103952 103952.

Chung, B.H., Hribar, P., 2021. CEO overconfidence and the timeliness of goodwill impairments. Account. Rev. 96, 221–259. https://doi.org/10.2308/TAR-2016-0555.

Cunningham, L.M., Johnson, B.A., Johnson, E.S., Krishnan, G.V., 2020. The switch-up: an examination of changes in earnings management after receiving SEC comment letters. Contemp. Account. Res. 37, 917–944. https://doi.org/10.1111/1911-3846.12546.

De Angelo, L.E., 1990. Equity valuation and corporate control. Account. Rev. 65, 93–112. https://doi.org/10.2308/TAR-9603274030.

Dechow, P.M., Lawrence, A., Ryans, J.P., 2016. SEC comment letters and insider sales. Account. Rev. 91, 401–439. https://doi.org/10.2308/accr-51232.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. BERT: pre-training of deep bidirectional transformers for language understanding, arXiv Preprint. arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gietzmann, M.B., Isidro, H., 2013. Institutional investors' reaction to SEC concerns about IFRS and US GAAP reporting. J. Bus. Finance Account. 40, 796–841. https://doi.org/10.1111/jbfa.12027.

Glaum, M., Landsman, W.R., Wyrwa, S., 2018. Goodwill impairment: the effects of public enforcement and monitoring by institutional investors. Account. Rev. 93, 149–180. https://doi.org/10.2308/accr-52006.

Gonçalves, C., Ferreira, L., Rebelo, E., et al., 2019. Big bath and goodwill impairment. Rev. Bus. Manag. 21, 312–331. https://doi.org/10.7819/RBGN.V21I2.3977.

Gu, F., Lev, B., 2011. Overpriced shares, ill-advised acquisitions, and goodwill impairment. Account. Rev. 86, 1995–2022. https://doi.org/10.2308/accr-10131.

Han, H., Tang, Q., 2020. The potential harms of goodwill impairment avoidance: evidence based on future performance and stock prices, China. J. Account. Res. 13, 271–289. https://doi.org/10.1016/j.cjar.2020.07.007.

Hansen, R.G., 1987. A theory for the choice of exchange medium in mergers and acquisitions. J. Bus. 60, 75–95. https://doi.org/10.1086/296386.
Hajek, P., Henriques, R., 2024. Predicting M&A targets using news sentiment and topic detection. Technol. Forecast. Soc. Change 201. https://doi.org/10.1016/j.techfore.2024.123270 123270.
Hayn, C., Hughes, P.J., 2006. Leading indicators of goodwill impairment. J. Account. Audit. Financ. 21, 223–265. https://doi.org/10.1177/0148558X0602100303.
Huang, X., Teoh, S.H., Zhang, Y., 2014. Tone management. Account. Rev. 89, 1083–1113. https://doi.org/10.2308/accr-50684.
Hwang, E.H., Singh, P.V., Argote, L., 2019. Jack of all, master of some: Information network and innovation in crowdsourcing communities. Inf. Syst. Res. 30, 389–410. https://doi.org/10.1287/isre.2018.0804.
Jarva, H., 2009. Do firms manage fair value estimates? An examination of SFAS 142 goodwill impairments. J. Bus. Finance Account. 36, 1059–1087. https://doi.org/10.1111/j.1468-5957.2009.02169.x.
Johnston, R., Petacchi, R., 2017. Regulatory oversight of financial reporting: Securities and Exchange Commission comment letters. Contemp. Account. Res. 34, 1128–1155. https://doi.org/10.1111/1911-3846.12297.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. In: Proc. 31st Int. Conf. Neural Information Processing Systems. Curran Associates, Red Hook, NY, pp. 3149–3157.
Kimbrough, M.D., Louis, H., 2011. Voluntary disclosure to influence investor reactions to merger announcements: an examination of conference calls. Account. Rev. 86, 637–667. https://doi.org/10.2308/accr.00000022.
Kothari, S.P., Shu, S., Wysocki, P.D., 2009. Do managers withhold bad news? J. Account. Res. 47, 241–276. https://doi.org/10.1111/j.1475-679X.2008.00318.x.
Kumar, P., 2024. Large language models (LLMs): survey, technical frameworks, and future challenges. Artif. Intell. Rev. 57, 260. https://doi.org/10.1007/s10462-024-10888-y.
Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D., Pappas, I., 2023. A systematic review of applications of natural language processing and future challenges with special emphasis on text-based emotion detection. Artif. Intell. Rev. 56, 15129–15215. https://doi.org/10.1007/s10462-023-10509-0.
Li, K.K., Sloan, R.G., 2017. Has goodwill accounting gone bad? Rev. Account. Stud. 22, 964–1003. https://doi.org/10.1007/s11142-017-9401-7.
Li, X., Zhong, Q., Liu, S., 2025. Random inspections and corporate information disclosure. Pac.-Basin Finance J. 90. 102701. https://doi.org/10.1016/j.pacfin.2025.102701.
Li, Z., Shroff, P.K., Venkataraman, R., Zhang, I.X., 2011. Causes and consequences of goodwill impairment losses. Rev. Account. Stud. 16, 745–778. https://doi.org/10.1007/s11142-011-9167-2.
Libby, R., Lipe, M.G., 1992. Incentives, effort, and the cognitive processes involved in accounting-related judgments. J. Account. Res. 30, 249–273. https://doi.org/10.2307/2491090.
Liu, Q., Du, Q., Tang, C., Hong, Y., Fan, W., 2025. An exploration and exploitation of value cocreation-based machine learning framework for automated idea screening. Decis. Support Syst. https://doi.org/10.1016/j.dss.2025.114504 114504.
Loh, W.-Y., 2011. Classification and regression trees, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1, 14–23. https://doi.org/10.1002/widm.8.
Luo, Q., Cai, S., Lv, N., Fu, X., 2025. Daily forecasting of tourism demand: an ST-LSTM model with social network service co-occurrence similarity. Inf. Manag. 62. https://doi.org/10.1016/j.im.2024.104056 104056.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.
Lyu, K., Shen, W., Li, H., Zhang, Y., 2024. Use of comment letters for mergers and acquisitions in a setting with weak investor protection: the Chinese experience. J. Account. Public Policy 46. https://doi.org/10.1016/j.jaccpubpol.2024.107227 107227.
Malmendier, U., Tate, G., 2007. Who makes acquisitions? CEO overconfidence and the market's reaction. J. Financ. Econ. 89, 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002.
Marrone, M., Linnenluecke, M.K., Richardson, G., Smith, T., 2020. Trends in environmental accounting research within and outside of the accounting discipline. Account. Audit. Account. J. 33, 2167–2193. https://doi.org/10.1108/AAAJ-03-2020-4457.
Nalepa, J., Kawulok, M., 2019. Selecting training sets for support vector machines: a review. Artif. Intell. Rev. 52, 857–900. https://doi.org/10.1007/s10462-017-9611-1.
Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res. 18, 109–131. https://doi.org/10.2307/2490395.
Ohlson, J.A., 1995. Earnings, book values, and dividends in equity valuation. Contemp. Account. Res. 11, 661–687. https://doi.org/10.1111/j.1911-3846.1995.tb00461.x.
Olante, M.E., 2013. Overpaid acquisitions and goodwill impairment losses—evidence from the US. Adv. Account. 29, 243–254. https://doi.org/10.1016/j.adiac.2013.09.010.
Pekar, V., Candi, M., Beltagui, A., Riedl, C., 2024. Explainable text-based features in predictive models of crowdfunding campaigns. Ann. Oper. Res., 1–31 https://doi.org/10.1007/s10479-023-05800-w.
Potepa, J., Thomas, J., 2023. Goodwill impairment after M&A: acquisition-level evidence. J. Financ. Rep. 8, 131–155. https://doi.org/10.2308/FR-2020-026.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: Unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst., vol. 31. Curran Associates, Red Hook, NY, pp. 6638–6648.

Ramanna, K., Watts, R.L., 2012. Evidence on the use of unverifiable estimates in required goodwill impairment. Rev. Account. Stud. 17, 749–780. https://doi.org/10.1007/s11142-012-9188-5.

Ramanna, K., 2008. The implications of unverifiable fair-value accounting: evidence from the political economy of goodwill accounting. J. Account. Econ. 45, 253–281. https://doi.org/10.1016/j.jacceco.2007.11.006.

Riedl, E.J., 2004. An examination of long-lived asset impairments. Account. Rev. 79, 823–852. https://doi.org/10.2308/accr.2004.79.3.823.

Qian, Y., Xu, W., Liu, X., Ling, H., Jiang, Y., Chai, Y., Liu, Y., 2022. Popularity prediction for marketer-generated content: a text-guided attention neural network for multi-modal feature fusion. Inf. Process. Manag. 59. https://doi.org/10.1016/j.ipm.2022.102984 102984.

Shen, M., Tan, S.T., 2023. The SEC revolving door and comment letters. J. Account. Public Policy 42. https://doi.org/10.1016/j.jaccpubpol.2023.107080 107080.

Song, Y., Jiang, M., Li, S., Zhao, S., 2024. Class-imbalanced financial distress prediction with machine learning: Incorporating financial, management, textual, and social responsibility features into index system. J. Forecast. 43, 593–614. https://doi.org/10.1002/for.3050.

Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. Cogn. Sci. 12, 257–285. https://doi.org/10.1016/0364-0213(88)90023-7.

Tang, H., Zhao, Y., Li, N., Zhu, J., 2025. Enhancing crowdfunding financing speed through image-text integration strategies: Insights from information processing theory and explainable machine learning methodology. Technol. Soc. https://doi.org/10.1016/j.techsoc.2025.103136 103136.

Tang, Q., Guo, Y., Huang, Z., 2021. The role of accounting conservatism in M&A target selection. China J. Account. Res. 14. https://doi.org/10.1016/j.cjar.2021.100204 100204.

Vo, N.N., Liu, S., Li, X., Xu, G., 2021. Leveraging unstructured call log data for customer churn prediction. Knowl.-Based Syst. 212. https://doi.org/10.1016/j.knosys.2020.106586 106586.

Wang, W., Xu, Y., Wu, Y.J., Liu, H., 2022. Linguistic understandability, signal observability, funding opportunities, and crowdfunding campaigns. Inf. Manag. 59. https://doi.org/10.1016/j.im.2022.103591 103591.

Wang, J., Yan, C., 2025. Using machine learning to identify audit opinion shopping. China J. Account. Res. 18. https://doi.org/10.1016/j.cjar.2025.100436 100436.

Wu, D., 2024. Text-based measure of supply chain risk exposure. Manag. Sci. 70, 4781–4801. https://doi.org/10.1287/mnsc.2023.4927.

Yoon, J., 2012. Detecting weak signals for long-term business opportunities using text mining of web news. Expert Syst. Appl. 39, 12543–12550. https://doi.org/10.1016/j.eswa.2012.04.059.

Yuan, H., Gao, C., Shi, H., 2020. Failure in performance commitment and goodwill impairment: evidence from M&As. China J. Account. Stud. 8, 1–31. https://doi.org/10.1080/21697213.2020.1822028.

Zhang, M., Luo, L., 2023. Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp. Manag. Sci. 69, 25–50. https://doi.org/10.1287/mnsc.2022.4359.

Zhang, W., Yoshida, T., Tang, X., 2011. A comparative study of TF-IDF, LSI and multi-words for text classification. Expert Syst. Appl. 38, 2758–2765. https://doi.org/10.1016/j.eswa.2010.08.066.

Zhu, H., Zhu, Q., 2016. Mergers and acquisitions by Chinese firms: a review and comparison with other mergers and acquisitions research in the leading journals, Asia Pac. J. Manag. 33, 1107–1149. https://doi.org/10.1007/s10490-016-9465-0.

Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. J. Account. Res. 22, 59–82. https://doi.org/10.2307/2490859.

China Journal of Accounting Research 19 (2026) 100502

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

TMT openness and corporate innovation: personality measurement using an LLM

Jinglu Zhao [a,*], Zeyu Zhang [b]

[a] Business School, Renmin University of China, Beijing, China
[b] Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China

ARTICLE INFO

Article history:
Received 18 May 2025
Accepted 27 May 2026
Available online 1 June 2026

Keywords:
TMT openness
Big Five personality
Explorative innovation
Exploitative innovation
Digital technology innovation
LLM

ABSTRACT

According to upper echelons theory, top management team (TMT) characteristics significantly influence corporate strategic decision-making. This study examines the impact of TMT openness on corporate innovation using a sample of Chinese A-share listed firms from 2007 to 2022. We employ a large language model (GLM-4) to measure TMT personality traits based on Q&A transcripts from annual performance briefings, offering a novel approach to capturing managerial personalities. Our findings demonstrate that TMT openness significantly enhances both explorative and exploitative innovation. Moreover, our extended analysis reveals that TMT openness exhibits differential effects across innovation types, with a more pronounced enhancing effect on digital technology innovation than non-digital technology innovation. Mechanism analyses reveal that TMTs with higher openness exhibit stronger innovation awareness, greater tolerance for uncertainties and reduced managerial myopia. Cross-sectional analyses show that the positive effect is larger among firms facing higher performance pressure, greater institutional ownership and more intense competition. Further analyses demonstrate that TMT openness is positively associated with R&D expenditure and personnel allocation. Moreover, TMT openness increases not only innovation quantity but innovation quality and efficiency. This study contributes to the literature by highlighting the role of TMT personality in corporate innovation and introducing an innovative measurement approach using large language models, providing methodological insights for future empirical research.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author.
E-mail address: zhaojinglu001@ruc.edu.cn (J. Zhao).

https://doi.org/10.1016/j.cjar.2026.100502
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Upper echelons theory, proposed by Hambrick and Mason (1984), posits that top management team (TMT) characteristics fundamentally shape organizational decision-making processes and strategic choices. Factors such as TMT cognitive styles and capability levels significantly affect TMTs' interpretation of external and internal environments, ultimately determining corporate strategic direction (Samimi et al., 2022).

Innovation activities constitute a critical dimension of corporate strategy, serving as a key driver of competitive advantage and long-term growth. As primary architects of corporate strategy, TMTs exercise substantial decision-making authority over innovation processes, including project approval and R&D resource allocation (Bostan and Mian, 2019; Griffin et al., 2021). Therefore, understanding the determinants of corporate innovation requires examining TMT characteristics (Wang et al., 2019). While prior research demonstrates that TMT demographic characteristics and personal experiences influence innovation outcomes (Talke et al., 2010; Richard et al., 2019), TMT personality traits represent another critical yet underexplored dimension. Personality traits are particularly salient given that innovation activities are characterized by substantial failure risks and extended return horizons (Jones and Williams, 2000), necessitating TMTs who demonstrate forward-thinking orientations, elevated risk tolerance and enhanced creative capabilities (Cortes and Herrmann, 2021). In the contemporary business environment, the rapid evolution of digital technologies has introduced new complexities and opportunities for corporate innovation. Digital technology innovation, encompassing areas such as artificial intelligence, blockchain and cloud computing, exhibits distinctive characteristics including accelerated obsolescence, substantial knowledge spillovers and heightened technological uncertainty (Nambisan et al., 2017). These attributes necessitate TMTs with exceptional cognitive flexibility and risk tolerance, rendering the role of executive personality traits potentially more salient in digital innovation contexts than in traditional technological domains.

The Big Five personality framework, initially conceptualized by Costa and McCrae (1992), provides a comprehensive tool for psychological analysis of individual characteristics. This framework encompasses five fundamental dimensions: openness, conscientiousness, extraversion, agreeableness and neuroticism, collectively abbreviated as OCEAN. Among these dimensions, openness captures an individual's receptiveness to novel ideas, diverse perspectives and unconventional approaches (Ones and Viswesvaran, 1997), fundamentally shaping cognitive frameworks and decision-making styles.

In the context of TMT characteristics, openness manifests in terms of orientations toward innovation and strategic change. TMTs characterized by high openness typically demonstrate greater imagination, creativity and exploratory tendencies, while those with low openness exhibit conventional thinking and risk-averse behaviors. These cognitive and behavioral differences have direct implications for corporate innovation. Specifically, we propose that TMTs with high openness possess heightened innovation awareness, greater tolerance for innovation-related uncertainties and failures, and more long-term-oriented strategic perspectives. These characteristics collectively enable open-minded TMTs to sustain commitment to innovation activities despite inherent risks and uncertainties, thereby facilitating enhanced corporate innovation outcomes.

This paper investigates the influence of TMT openness on corporate innovation using Chinese A-share listed firms from 2007 to 2022 as the research sample. TMT personality traits, including openness and other Big Five dimensions, are measured using GLM-4, a state-of-the-art large language model (LLM), to analyze Q&A transcripts from annual performance briefings based on a professional Big Five personality scale. The findings indicate that TMT openness significantly enhances corporate innovation, including both explorative and exploitative innovation. Mechanism analyses reveal that TMTs with high openness exhibit stronger innovation awareness, demonstrate greater tolerance for innovation risks and uncertainties, and display reduced managerial myopia. Extending our analysis to technology heterogeneity, we find that TMT openness exerts a more pronounced promotional effect on digital technology innovation than on non-digital technology innovation, reflecting the particular importance of cognitive flexibility and risk tolerance in navigating the rapid evolution and high uncertainty characteristic of digital innovation domains. Cross-sectional analyses reveal that the positive effect of TMT openness on corporate innovation is more pronounced in firms facing higher performance pressure, exhibiting greater institutional ownership and operating in more competitive industry environments. Further analyses demonstrate that TMT openness is positively associated with R&D expenditure and personnel allocation, as well as innovation quality and efficiency.

The contributions of this paper are manifested in three primary aspects. First, this paper introduces a novel approach to measuring TMT personality traits using LLMs, thereby expanding the availability of large-scale data on the Big Five personality traits of TMTs and demonstrating the application potential of artificial intelligence in management research. Traditional methods of measuring executive personality traits, such as surveys or manual evaluations, suffer from notable limitations (Herrmann and Nadkarni, 2014; Benischke et al., 2019), including restricted evaluation scope, substantial labor costs and considerable subjectivity among respondents or evaluators, which compromise the validity and comparability of results. The emergence of LLMs provides a powerful tool for management research, enabling efficient large-scale measurement of personality traits. Leveraging Q&A transcripts from annual performance briefings of Chinese listed firms and professional psychological scales, this paper employs GLM-4, an advanced LLM specializing in Chinese language processing, to measure the Big Five personality traits of TMTs. This approach effectively addresses comparability concerns arising from inconsistent evaluation standards across individual assessors and generates large-sample data on TMT personality traits suitable for rigorous empirical analysis.

Second, this paper investigates the impact of TMT openness on corporate innovation strategy by distinguishing between explorative and exploitative innovation, as well as between digital and non-digital technology innovation, thereby providing novel insights into how executive personality traits shape firms' strategic choices regarding innovation. Existing literature examining the impact of executive characteristics on corporate innovation predominantly focuses on demographic characteristics or educational and professional backgrounds (Wiersema and Bantel, 1992; Barker and Mueller, 2002; Price, 2012; Xie et al., 2020). Constrained by data availability, empirical evidence concerning how TMT openness influences corporate innovation remains limited. By constructing a comprehensive dataset of TMT Big Five personality traits and distinguishing between explorative and exploitative innovation (Byun et al., 2021), as well as between digital and non-digital technology innovation (Nambisan et al., 2017), this paper systematically investigates how TMT openness shapes innovation strategy. Furthermore, this study identifies and empirically validates three key mechanisms through which TMT openness influences corporate innovation: enhanced innovation awareness, increased tolerance for innovation uncertainties and reduced managerial myopia. By elucidating these pathways, this research contributes to a more comprehensive understanding of the psychological and cognitive mechanisms linking executive personality traits to innovation outcomes, bridging the gap between personality psychology and strategic innovation management.

Third, this study offers significant practical implications for TMT composition and human resource management practices. The findings reveal that TMT openness substantially enhances corporate innovation, underscoring the critical role of executive personality traits in shaping corporate strategic decisions. These results suggest that firms should systematically assess managerial personality traits during selection and appointment processes, prioritizing candidates whose characteristics align with the firm's strategic objectives. Furthermore, by examining how TMT personality traits interact with specific organizational environments, industry contexts and market dynamics, this study provides nuanced insights into the contextual contingencies that amplify or attenuate the innovation-promoting effects of TMT openness. These findings not only deepen theoretical understanding of the relationship between TMT personality traits and corporate innovation, but also offer actionable guidance for firms seeking to optimize TMT composition and leverage executive strengths to enhance innovation performance.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature and develops the research hypothesis. Section 3 introduces the data sources and model specifications. Section 4 presents the empirical results. Section 5 examines the underlying mechanisms. Section 6 conducts cross-sectional analyses and further tests, and Section 7 concludes.

2. Literature review and hypothesis development

2.1. Big Five personality traits and economic consequences

Costa and McCrae (1992) were the pioneers in formulating the modern framework of the Big Five personality theory, encapsulating the five "OCEAN" dimensions of personality traits. This framework has been

widely adopted in organizational and strategic management research to understand how executive personality traits influence firm outcomes.

A growing body of literature examines the relationship between Big Five personality traits and various economic and organizational outcomes. Gow et al. (2016) develop linguistic-based measures of CEO Big Five personality traits and find that these traits are associated with firms' financing choices, investment decisions and operating performance. In the domain of innovation and strategic change, Benischke et al. (2019) find that CEO openness moderates the relationship between equity risk bearing and strategic risk-taking, with open CEOs demonstrating greater willingness to pursue risky strategic initiatives. Herrmann and Nadkarni (2014) document that CEO openness enhances the initiation of strategic change in SMEs, suggesting that open-minded executives are more receptive to pursuing novel strategic directions. From a strategic flexibility perspective, Aabo et al. (2024) examine how CEO personality traits influence firms' ability to adapt to environmental volatility, finding that different personality dimensions have contingent effects on strategic adaptability depending on the context. These findings collectively suggest that CEO openness plays a crucial role in shaping firms' strategic orientations toward exploration and innovation.

Beyond innovation and strategy, openness has been linked to various corporate outcomes. Green et al. (2019) find that CEO extraversion, which often correlates with openness, influences career progression and firm outcomes. Research on entrepreneurs shows that openness to experience is positively related to entrepreneurial success, opportunity recognition and innovation adoption (Zhao and Seibert, 2006; Marcati et al., 2008). In the context of organizational performance, several studies document positive associations between openness and firm performance metrics, although these relationships are often contingent on industry characteristics and environmental conditions (Nadkarni and Herrmann, 2010).

Despite these valuable contributions, the literature reveals several important research gaps. First, although prior studies establish positive associations between management openness and various outcomes, critical questions regarding the strategic nature of innovation and underlying mechanisms of the relationships remain unanswered. Specifically, no existing work investigates whether and how openness shapes firms' strategic choices between explorative and exploitative innovation. Moreover, the cognitive and behavioral pathways through which TMT openness translates into innovation outcomes remain inadequately understood.

Second, existing measurement approaches for Big Five personality traits present significant methodological limitations. Current methodologies—questionnaire surveys, manual scoring and machine learning techniques—each face distinct challenges. Questionnaire-based approaches encounter constraints in survey scope and sample size due to limited accessibility and response authenticity concerns (Herrmann and Nadkarni, 2014). Manual scoring methods, which employ trained professionals to evaluate text and video materials using Q-score methodology (Block, 1961; Benischke et al., 2019), require substantial human resources. Critically, both approaches involve subjective judgments that compromise cross-individual comparability, as different evaluators or respondents may interpret scale gradations inconsistently. Machine learning approaches improve scalability by training linguistic tools on validated assessments and applying them to larger corpora (Harrison et al., 2019), yet their accuracy depends on training data quality and they may inherit biases from original scoring processes. Advancements in LLMs offer transformative opportunities to overcome these constraints by enabling large-scale personality measurement while maintaining construct validity, ensuring consistent evaluation standards and substantially reducing resource requirements.

2.2. TMT openness and corporate innovation

The Big Five personality traits, as stable individual characteristics, constitute critical determinants of TMTs' decision-making styles, information processing approaches and risk tolerance levels. Openness reflects an individual's receptiveness to new knowledge and methodologies, with higher TMT openness signifying enhanced curiosity, imagination and exploratory orientation (Costa and McCrae, 1992; Ones and Viswesvaran, 1997). This personality trait assumes particular significance in the innovation context, as innovation activities are characterized by extended cycles, elevated risks, intertemporal returns and complex R&D processes (Jones and Williams, 2000; Brav et al., 2018).

Corporate innovation encompasses two distinct types: explorative innovation, characterized by venturing into new technological territories with high uncertainty, and exploitative innovation, which involves incremen-

tal improvements within existing capabilities (March, 1991; Byun et al., 2021). We theorize that TMT openness influences both innovation types through three mechanisms: enhanced managerial innovation awareness, increased willingness to tolerate innovation uncertainties and reduced managerial myopia. These mechanisms operate through cognitive, affective and temporal channels respectively, shaping how TMTs attend to innovation opportunities, tolerate risks and uncertainties, and commit resources beyond short-term pressures.

First, open TMTs demonstrate heightened managerial innovation awareness—a cognitive orientation that directs attention toward innovation opportunities and prioritizes innovation in strategic decision-making (Ocasio, 1997). This cognitive advantage manifests in superior ability to identify opportunities across both familiar and unfamiliar technological domains (George and Zhou, 2001). Consequently, open-minded TMTs allocate greater financial resources to R&D activities (Chan et al., 1990; Belderbos et al., 2004) and prioritize recruitment of innovation talent (D'Acunto, 2014), establishing the resource foundation necessary for both explorative and exploitative innovation.

Second, TMT openness enhances willingness to tolerate innovation uncertainties, corresponding with elevated risk-taking propensity (Li, 2010; Benischke et al., 2019). Open TMTs exhibit psychological resilience that enables persistence despite setbacks and reframes failures as learning opportunities rather than disasters (Costa and McCrae, 1992; Sitkin, 1992). This tolerance for uncertainty is essential for innovation success: it provides the confidence to initiate high-risk projects (Simon and Houghton, 2003) and sustains commitment through the inherent uncertainties of innovation processes, particularly for explorative innovation where failure probability is substantial (McGrath, 2001).

Third, TMT openness contributes to reduced managerial myopia, addressing a fundamental barrier to innovation investment. TMTs with weaker innovation orientations often prioritize immediate performance over long-term value creation (Narayanan, 1985), resulting in insufficient investment in innovation projects that inherently require extended time horizons (Bebchuk and Stole, 1993). In contrast, open TMTs exhibit reduced myopia and enhanced long-term orientation (Nadkarni and Herrmann, 2010). Their broader temporal perspective enables them to resist short-term performance pressures and maintain sustained commitment to innovation initiatives (Laverty, 1996; Brochet et al., 2015). This temporal advantage allows effective resource allocation toward both explorative ventures with distant payoffs and exploitative refinements requiring continuous investment, thereby facilitating organizational ambidexterity in innovation strategy (O'Reilly and Tushman, 2008).

These three mechanisms—managerial innovation awareness, willingness to tolerate innovation uncertainties and reduced managerial myopia—operate synergistically to promote both explorative and exploitative innovation. Innovation awareness ensures that innovation receives adequate cognitive attention and resource allocation; uncertainty tolerance provides the psychological capacity to undertake and persist in risky endeavors; and reduced myopia maintains long-term commitment despite short-term pressures. Based on this theoretical framework, we propose as follows:

Hypothesis: TMT openness positively influences corporate innovation, encompassing both explorative innovation and exploitative innovation.

3. Data and methodology

3.1. Data and sample

This study examines the relationship between TMT openness and corporate innovation using panel data from A-share listed firms in China over the period 2007–2022. The empirical analysis draws upon multiple data sources. (1) TMT personality traits, including openness and other Big Five dimensions, are extracted through natural language processing of Q&A transcripts in annual performance briefings. In China's institutional context, listed firms are required or encouraged to hold annual performance briefings to communicate with investors and analysts. These briefings, typically conducted following annual earnings announcements, provide a structured forum where TMT members respond to questions regarding firm performance, strategic initiatives and future prospects. The Q&A sessions from these briefings offer rich textual data that reveal managerial cognition, communication styles and personality characteristics. We obtain original transcripts of these Q&A sessions in annual performance briefings from the Chinese Research Data Services (CNRDS) database

and employ the GLM-4 LLM to assess TMT personality traits based on linguistic patterns and response characteristics exhibited in the transcripts. (2) Corporate patent application data, serving as a proxy for firm-level innovation, are obtained from CNRDS. The data encompass both the quantity of patent applications and their International Patent Classification (IPC) codes, which enable us to categorize patents by technological field and distinguish between explorative and exploitative innovation. (3) Additional firm-level financial and governance variables, as well as TMT demographic characteristics, are sourced from the China Stock Market & Accounting Research (CSMAR) database.

To ensure the reliability of the results, we apply several screening criteria to the initial sample. Specifically, we exclude (1) firms designated as Special Treatment (ST) or Particular Transfer (PT) due to financial distress or regulatory violations; (2) financial industry firms, whose capital structures and innovation patterns differ fundamentally from non-financial firms; and (3) observations with missing values for key variables. These exclusions yield a final sample of 10,703 firm-year observations. To mitigate the influence of extreme values, all continuous variables are winsorized at the 1st and 99th percentiles.

3.2. Variable definitions

3.2.1. TMT openness

Gow et al. (2016) demonstrate that linguistic features can mirror personality traits and remain generally stable over time, enabling personality assessment through language content analysis. Building on this insight, we employ GLM-4, an advanced LLM optimized for Chinese corpora, to measure the Big Five personality traits of TMTs using Q&A transcripts from annual performance briefings of Chinese A-share listed firms. In our regression analysis, we focus on the openness dimension as the primary explanatory variable, while using the other four dimensions—conscientiousness, extraversion, agreeableness and neuroticism—as control variables.

The Q&A transcripts from annual performance briefings offer several advantages as input materials for personality measurement. Unlike scripted corporate disclosures such as annual reports, these Q&A sessions capture spontaneous responses to externally initiated inquiries, providing more authentic insights into TMT members' genuine thoughts and communication styles. Moreover, while formal written texts may suffer from reduced discriminability due to involvement of multiple managers and consultants (Gow et al., 2016), performance briefing Q&As not only involve multiple TMT members but also allow precise identification of individual respondents, enabling accurate individual-level personality trait measurement.

We implement the following procedure to measure TMT Big Five personality traits:

(1) For each executive in a given firm-year (identified by year, stock code and executive name), we use the firm's annual performance briefing Q&A transcripts as input. We instruct GLM-4 (specifically the GLM-4-9B-chat version released on 5 June 2024, developed by Zhipu AI) to learn and adopt the role of the executive when answering designated questions. We select this model because it demonstrates superior performance on Chinese text comprehension benchmarks, has a context window of 128K tokens allowing complete processing of lengthy transcripts, and its training corpus includes substantial Chinese business communication data.
(2) We employ the Chinese-language Big Five Inventory-2 (BFI-2) scale (Zhang et al., 2022), a professional and validated instrument comprising 60 items across five dimensions (12 items per dimension), all scored on a 5-point Likert scale. GLM-4 is instructed to assume the executive's role and respond to each BFI-2 item with integer scores ranging from 1 to 5. The complete English translation of the BFI-2 scale, detailed prompt engineering specifications and illustrative examples are provided in the Appendix.
(3) We calculate dimensional scores based on GLM-4′s responses to each BFI-2 item. The original total score for each dimension is 60 points (5 points × 12 items). We sum the scores across all five dimensions and divide by 6 to standardize them to a 10-point scale, yielding individual-level Big Five personality variables for each executive.
(4) We compute firm-level TMT personality measures by averaging the scores of all TMT members within each firm-year. Among these five dimensions, openness serves as our core explanatory variable, denoted as *TMTopen*.

To validate our measurement approach, we employ two methods. First, three doctoral students with professional training in psychology independently review 100 randomly selected Q&A transcripts and manually score the Big Five personality traits using the same BFI-2 scale. The manual scores demonstrate strong concordance with the GLM-4-generated scores, suggesting high agreement between human judgment and LLM assessment. Second, we examine the temporal stability of our personality measures. As personality traits are theoretically expected to be relatively stable over time, particularly in adulthood (Costa and McCrae, 1992; Roberts and DelVecchio, 2000), we analyze year-over-year changes in individual executive-level openness scores. Fig. 1 presents this distribution, which is tightly concentrated around zero, indicating that executives' measured personality traits exhibit modest variation over time. Collectively, these two validation results provide strong evidence for the reliability and validity of our GLM-4-based personality measurement methodology.

3.2.2. Corporate innovation

This paper constructs variables for measuring corporate innovation levels utilizing the patent dataset obtained from CNRDS. Given that patent grants are subject to discretionary decisions by patent offices, the number of patent applications provides a more accurate and timely reflection of genuine innovation output relative to granted patents (Crosby, 2000; Chemmanur et al., 2019). Accordingly, we employ patent application counts as the primary metric for innovation activity, with the temporal dimension determined by application dates.

We construct three principal measures of innovation output. First, aggregate innovation (*AllInnov*) is operationalized as the natural logarithm of one plus the total number of patent applications filed by the firm in a given fiscal year, capturing the overall magnitude of innovation activity.

Meanwhile, to distinguish between innovation strategies, we decompose patent applications into explorative and exploitative innovation following Byun et al. (2021). We utilize the first three digits of IPC codes to delineate technological domains. Explorative innovation (*ExplorativeInnov*) is measured as the natural logarithm of one plus the number of patent applications filed in unfamiliar technological domains—defined as IPC 3-digit classifications in which the firm possesses no prior patent applications throughout the past five years. This measure captures ventures into novel technological territories and the exploration of previously uncharted knowledge domains. The last of the three principal measures, exploitative innovation (*ExploitativeInnov*) is measured as the natural logarithm of one plus the number of patent applications filed in familiar technological domains—defined as IPC 3-digit classifications in which the firm has previously filed at least one patent within the past five years. This measure captures refinement and deepening of existing technological competencies within the firm's established capability portfolio.

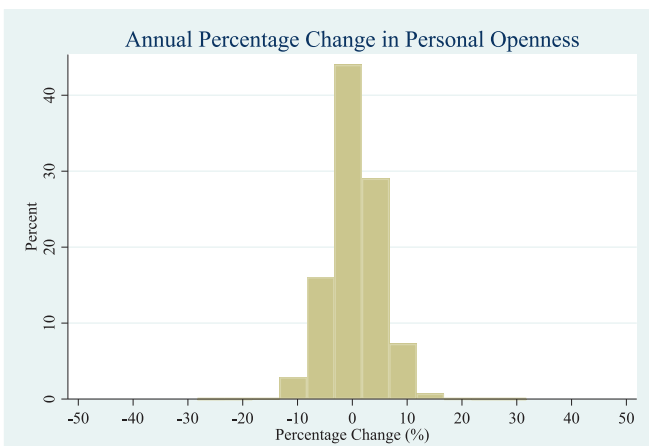


Fig. 1. Distribution of year-over-year changes in individual executive openness scores.

3.2.3. Control variables

To account for potential confounding factors that may influence corporate innovation, we incorporate a comprehensive set of control variables following Kong et al. (2022) and Gao et al. (2023). These variables can be categorized into two groups: firm-level and TMT-level. At the firm level, we control for firm size (*Size*), leverage ratio (*Lev*), R&D expenditure proportion (*RDpro*), state ownership (*SOE*), return on assets (*ROA*), profitability status (*Loss*), CEO duality (*Dual*), firm age (*FirmAge*) and board size (*Board*). At the TMT level, we include the remaining four dimensions of the Big Five personality traits—conscientiousness (*TMTconsc*), extraversion (*TMTextra*), agreeableness (*TMTagree*) and neuroticism (*TMTneuro*)—to isolate the specific effect of openness. Additionally, we control for TMT demographic and compensation characteristics, including average age (*TMTage*), TMT shareholding ratio (*TMTshare*), compensation (*TMTpay*), stock option incentives (*TMToption*) and gender diversity measured by the proportion of female executives (*Female*). Detailed definitions of all variables are provided in Table 1.

3.3. Model specification

To investigate the influence of TMT openness on corporate innovation, this paper estimates the following regression:

$$Patent_{i,t} = \alpha_0 + \alpha_1 TMTopen_{i,t-1} + \sum Controls_{i,t-1} + \gamma_{ind} + \delta_t + \varepsilon_{i,t}$$

where $Patent_{i,t}$ represents the level of corporate innovation, quantified by the natural logarithm of one plus the total patent applications submitted by the firm, with further distinctions made between explorative and exploitative innovation for analysis. $TMTopen_{i,t-1}$ denotes the degree of TMT openness, obtained via GLM-4 analysis based on professional psychological scales. $Controls_{i,t-1}$ encompasses our comprehensive set of control variables, including the remaining four dimensions of the Big Five personality traits, additional

Table 1
Variable definitions.

Variable name	Definition
AllInnov	Natural logarithm of one plus total patent applications.
ExplorativeInnov	Natural logarithm of one plus the number of patent applications in unfamiliar technological domains, where unfamiliar domains are defined as IPC 3-digit classifications in which the firm has no prior patents within the past five years, representing explorative innovation.
ExploitativeInnov	Natural logarithm of one plus the number of patent applications in familiar technological domains, where familiar domains are defined as IPC 3-digit classifications in which the firm has at least one prior patent within the past five years, representing exploitative innovation.
TMTopen	Openness of the TMT analyzed using GLM-4, based on Q&A transcripts of annual performance briefings.
TMTconsc	Conscientiousness of the TMT analyzed using GLM-4, based on Q&A transcripts of annual performance briefings.
TMTextra	Extraversion of the TMT analyzed using GLM-4, based on Q&A transcripts of annual performance briefings.
TMTagree	Agreeableness of the TMT analyzed using GLM-4, based on Q&A transcripts of annual performance briefings.
TMTneuro	Neuroticism of the TMT analyzed using GLM-4, based on Q&A transcripts of annual performance briefings.
Size	Natural logarithm of total assets.
Lev	Total liabilities/Total assets.
RDpro	R&D expenditure/Operating revenue.
SOE	An indicator variable equaling 1 if the firm is state-owned, 0 otherwise.
ROA	Net income/Total assets.
Loss	Indicator variable equaling 1 if the firm incurs a loss in the current year, 0 otherwise.
Dual	Indicator variable equaling 1 if the CEO and chairman are the same person, 0 otherwise.
FirmAge	Natural logarithm of the number of years since a firm has been established.
Board	Natural logarithm of the number of board members.
TMTage	Average age of the TMT.
TMTshare	TMT shares/Total issued shares.
TMTpay	Natural logarithm of the total compensation of the TMT.
TMToption	Indicator variable equaling 1 if TMT members hold stock options, 0 otherwise.
Female	Number of female managers/Total TMT members.

TMT characteristics and the firm-level controls. This paper lags the core explanatory variable and all control variables by one period, given the time needed for innovation to yield results. Notably, the Big Five personality traits of the TMT exhibit remarkable stability over time (Jiang et al., 2024), rendering the control of firm-fixed effects infeasible. Consequently, we control for industry-fixed effects (γ_{ind}) and year-fixed effects (δ_t) in the regression. Standard errors are clustered at the firm level. This study mainly focuses on the significance of the $TMTopen_{i,t-1}$ coefficient α_1 in the model. A significantly positive coefficient would indicate that openness of the TMT can improve corporate innovation.

4. Empirical results

4.1. Summary statistics

Table 2 reports the main summary statistics for the variables. Within this paper's research sample, the mean of total innovation (*AllInnov*) is 2.035, with a median of 2.079. Notably, the large standard deviation of *AllInnov*, coupled with a significant disparity between the maximum and minimum values, suggests considerable variation in innovation levels among different firms. When distinguishing between explorative innovation (*ExplorativeInnov*) and exploitative innovation (*ExploitativeInnov*), the mean values are 0.810 and 1.790, respectively, indicating that firms on average engage more in incremental innovation within familiar technological domains than breakthrough innovation in unfamiliar domains. The distribution of these two categories shows that explorative innovation exhibits higher variance, consistent with its more exploratory and risky nature. The mean value of *TMTopen* among the TMTs is 7.372, accompanied by a standard deviation of 0.187. This indicates that there exists some variability in the openness trait among the TMTs of the sample firms, but the TMTs generally exhibit a high level of openness. The values and distributions of the control variables in this study align closely with those reported in existing literature, falling within a reasonable range.

Table 2
Summary statistics.

Variables	(1) N	(2) Mean	(3) Std	(4) Min	(5) Median	(6) Max
AllInnov	10,703	2.035	1.525	0.000	2.079	5.861
ExplorativeInnov	10,703	0.810	1.062	0.000	0.000	4.263
ExploitativeInnov	10,703	1.790	1.502	0.000	1.792	5.645
TMTopen	10,703	7.372	0.187	6.583	7.389	8.000
TMTconsc	10,703	7.960	0.082	6.833	8.000	8.222
TMTextra	10,703	7.699	0.094	7.000	7.708	8.000
TMTagree	10,703	7.491	0.130	6.167	7.500	8.000
TMTneuro	10,703	4.526	0.088	4.167	4.500	5.500
Size	10,703	21.914	1.144	19.863	21.739	26.039
Lev	10,703	0.384	0.191	0.040	0.371	0.839
RDpro	10,703	5.016	5.142	0.000	3.740	31.980
SOE	10,703	0.189	0.392	0.000	0.000	1.000
ROA	10,703	0.045	0.062	−0.208	0.043	0.232
Loss	10,703	0.094	0.292	0.000	0.000	1.000
Dual	10,703	0.336	0.472	0.000	0.000	1.000
FirmAge	10,703	2.769	0.371	1.386	2.773	3.497
Board	10,703	2.105	0.186	1.609	2.197	2.565
TMTage	10,703	48.582	3.175	40.820	48.530	56.610
TMTshare	10,703	19.260	20.618	0.000	11.653	71.310
TMTpay	10,703	14.444	0.704	12.714	14.420	16.540
TMToption	10,703	0.397	0.489	0.000	0.000	1.000
Female	10,703	0.196	0.109	0.000	0.188	0.500

4.2. Baseline regression

The baseline regression results for the relationship between TMT openness and corporate innovation levels are presented in Table 3. The first two columns display the results using total patent applications as the dependent variable. Specifically, Column (1) presents the results of a univariate regression without including control variables, while Column (2) extends this analysis by incorporating control variables at both the TMT-level and

Table 3
Baseline regression.

Variables	(1) *AllInnov*	(2) *AllInnov*	(3) *ExplorativeInnov*	(4) *ExplorativeInnov*	(5) *ExploitativeInnov*	(6) *ExploitativeInnov*
TMTopen	0.572***	0.388***	0.470***	0.366***	0.487***	0.324***
	(4.72)	(3.40)	(5.56)	(4.52)	(4.05)	(2.87)
TMTconsc	0.707**	0.630**	0.452**	0.418**	0.544**	0.465*
	(2.57)	(2.42)	(2.46)	(2.35)	(1.99)	(1.78)
TMTextra	0.199	−0.135	0.142	−0.029	0.216	−0.127
	(1.12)	(−0.81)	(1.15)	(−0.24)	(1.24)	(−0.78)
TMTagree	0.396**	0.462***	0.358***	0.382***	0.339*	0.418**
	(2.12)	(2.67)	(2.88)	(3.20)	(1.83)	(2.47)
TMTneuro	−0.129	0.461**	0.221	0.602***	−0.308	0.287
	(−0.51)	(2.01)	(1.26)	(3.60)	(−1.22)	(1.26)
Size		0.270***		0.168***		0.253***
		(7.46)		(7.28)		(7.11)
Lev		0.284*		0.083		0.428***
		(1.74)		(0.80)		(2.68)
RDpro		0.041***		0.013***		0.044***
		(6.35)		(3.23)		(6.75)
SOE		0.061		0.053		0.081
		(0.77)		(1.01)		(1.01)
ROA		3.493***		2.647***		3.479***
		(7.14)		(7.04)		(7.36)
Loss		0.041		0.125**		0.041
		(0.61)		(2.43)		(0.63)
Dual		0.032		0.096***		0.020
		(0.67)		(2.75)		(0.42)
FirmAge		0.072		0.087*		0.048
		(0.96)		(1.72)		(0.65)
Board		0.290**		0.085		0.298**
		(2.25)		(0.97)		(2.36)
TMTage		−0.005		−0.006		0.000
		(−0.58)		(−1.16)		(0.06)
TMTshare		0.004***		0.001*		0.004***
		(2.99)		(1.65)		(2.91)
TMTpay		0.141***		0.084***		0.155***
		(3.32)		(2.84)		(3.70)
TMToption		0.184***		0.082**		0.180***
		(4.07)		(2.57)		(3.97)
Female		−0.578***		−0.171		−0.674***
		(−2.67)		(−1.19)		(−3.19)
Constant	−11.719***	−19.378***	−11.030***	−15.851***	−8.934***	−16.859***
	(−3.62)	(−6.20)	(−4.94)	(−6.93)	(−2.79)	(−5.43)
Observations	10,703	10,703	10,703	10,703	10,703	10,703
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.246	0.321	0.151	0.203	0.250	0.329

Note: The *t* statistics are in parentheses. ***, ** and * indicate significance of the estimated coefficients at the 1%, 5% and 10% levels, respectively. The robust standard errors have been adjusted for clustering at the firm level. The same applies to subsequent tables.

firm-level. The coefficients of *TMTopen* are significantly positive at the 1% level in both specifications. After adding control variables and including industry and year fixed effects, the coefficient of *TMTopen* is 0.388, suggesting that a one-unit increase in TMT openness is associated with approximately a 47.4% increase in total patent applications.

The patents are further categorized into explorative innovation (*ExplorativeInnov*) and exploitative innovation (*ExploitativeInnov*) for additional regression tests, as shown in Columns (3) to (6) of Table 3. Across these columns, the coefficients of *TMTopen* remain significantly positive at the 1% level. Specifically, in the full specification with control variables (Columns 4 and 6), the coefficient of *TMTopen* is 0.366 for explorative innovation and 0.324 for exploitative innovation. These findings collectively indicate that TMT openness exerts a significant positive effect on both types of innovation, with a slightly stronger impact on explorative innovation than exploitative innovation, although the difference is relatively modest.

4.3. Robustness tests

4.3.1. Propensity score matching

To address potential endogeneity concerns arising from omitted variables and self-selection bias, this paper employs the propensity score matching (PSM) method following Rosenbaum and Rubin (1983). PSM creates a quasi-experimental setting by matching firms with similar observable characteristics, thereby mitigating confounding factors that could bias the estimated relationship between TMT openness and corporate innovation.

Specifically, the full sample is divided into a treatment group and a control group based on the median value of *TMTopen*. A range of covariates is selected, including *Size*, *Lev*, *RDpro*, *ROA*, *FirmAge*, *Board*, *TMTage*, *TMTshare*, *TMTpay*, *TMToption* and *Female*. These covariates are used in a 1:1 nearest-neighbor matching with replacement to create matched pairs. The matching results indicate that the standardized biases for all covariates are less than 5%, and the *t*-tests do not refute the null hypothesis of no significant difference in coefficients between the treatment and control groups, suggesting a good matching quality. Fig. 2

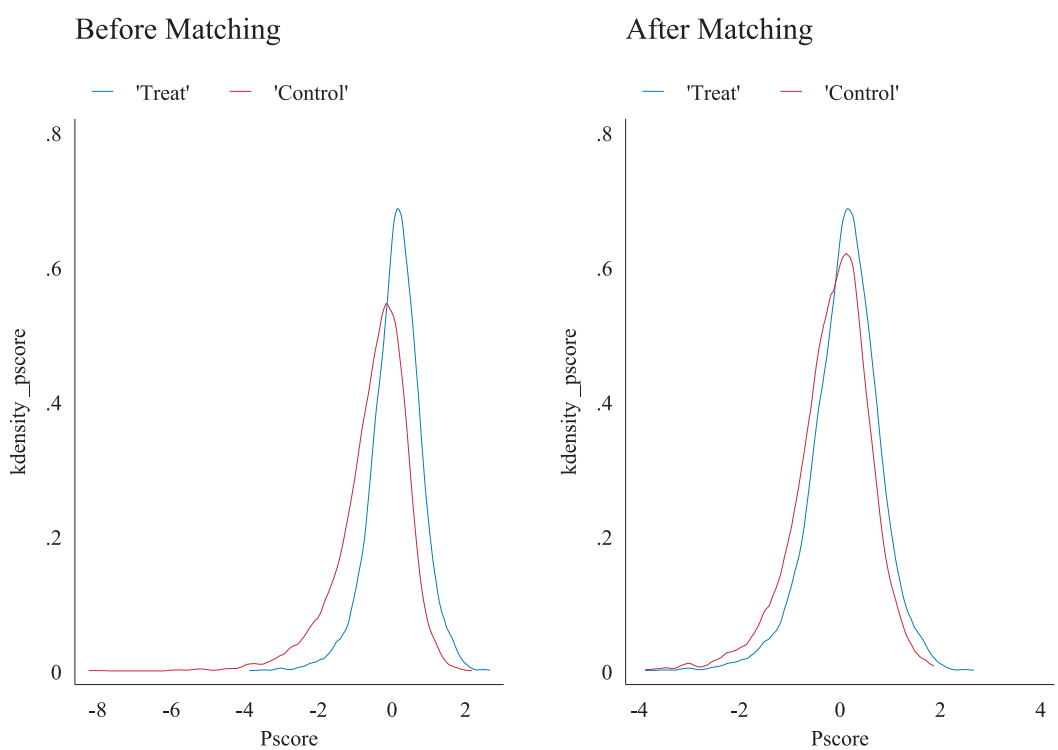


Fig. 2. Propensity score matching: pre- and post-matching density curves.

shows the kernel density distributions of propensity scores for both groups before and after matching, demonstrating substantial overlap in the post-match distributions.

Table 4 presents the regression results based on the propensity score-matched sample. The coefficient of *TMTopen* remains significantly positive at the 1% level across all specifications. These results are consistent with the baseline findings, indicating that the positive relationship between TMT openness and corporate innovation is robust to potential biases from omitted variables and self-selection issues.

Table 4
Propensity score matching.

Variables	(1) *AllInnov*	(2) *AllInnov*	(3) *ExplorativeInnov*	(4) *ExplorativeInnov*	(5) *ExploitativeInnov*	(6) *ExploitativeInnov*
TMTopen	0.580***	0.420***	0.438***	0.345***	0.493***	0.350***
	(4.58)	(3.54)	(4.83)	(4.02)	(3.91)	(2.96)
TMTconsc	1.256***	1.259***	0.975***	0.980***	1.089***	1.086***
	(3.63)	(3.81)	(4.11)	(4.27)	(3.15)	(3.30)
TMTextra	0.219	−0.206	0.131	−0.097	0.251	−0.186
	(1.14)	(−1.13)	(0.95)	(−0.72)	(1.35)	(−1.05)
TMTagree	0.407*	0.535***	0.315**	0.364***	0.381*	0.523***
	(1.95)	(2.75)	(2.25)	(2.70)	(1.84)	(2.73)
TMTneuro	−0.051	0.593**	0.313	0.719***	−0.187	0.471*
	(−0.18)	(2.33)	(1.56)	(3.74)	(−0.67)	(1.85)
Size		0.268***		0.167***		0.252***
		(7.15)		(7.06)		(6.86)
Lev		0.307*		0.104		0.452***
		(1.82)		(0.96)		(2.75)
RDpro		0.042***		0.014***		0.046***
		(6.05)		(3.13)		(6.50)
SOE		0.052		0.052		0.066
		(0.63)		(0.92)		(0.81)
ROA		3.317***		2.650***		3.343***
		(6.56)		(6.85)		(6.86)
Loss		−0.010		0.101*		−0.005
		(−0.14)		(1.85)		(−0.08)
Dual		0.022		0.088**		0.011
		(0.45)		(2.47)		(0.23)
FirmAge		0.047		0.076		0.021
		(0.62)		(1.46)		(0.28)
Board		0.257*		0.080		0.268**
		(1.95)		(0.89)		(2.08)
TMTage		−0.001		−0.005		0.003
		(−0.16)		(−0.97)		(0.41)
TMTshare		0.003***		0.001		0.003***
		(2.77)		(1.25)		(2.79)
TMTpay		0.159***		0.089***		0.166***
		(3.68)		(2.88)		(3.90)
TMToption		0.187***		0.090***		0.182***
		(4.07)		(2.71)		(3.96)
Female		−0.576***		−0.156		−0.701***
		(−2.59)		(−1.03)		(−3.22)
Constant	−16.740***	−25.442***	−14.957***	−20.134***	−14.460***	−23.287***
	(−4.50)	(−7.10)	(−5.66)	(−7.51)	(−3.89)	(−6.53)
Observations	9230	9230	9230	9230	9230	9230
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.247	0.323	0.150	0.203	0.249	0.330

4.3.2. Placebo test

To exclude interference from unobservable factors and further ensure the reliability of the conclusions, this paper follows the placebo testing methodology of Cornaggia and Li (2019). Specifically, we randomly generate a pseudo-TMT openness variable (*Pseudo*) for each firm-year observation by drawing values from a uniform distribution that mimics the actual distribution of TMT openness in the sample. This pseudo-variable is then used to re-run the baseline regression specification. The entire randomization and estimation process is performed 1000 times to obtain 1000 sets of pseudo-effect estimates and their corresponding *t*-statistics.

Table 5 reports the regression results from one randomly generated pseudo-TMT openness assignment. Across all specifications in Columns (1) through (6), the coefficient on *Pseudo* is statistically nonsignificant

Table 5
Placebo tests.

Variables	(1) *AllInnov*	(2) *AllInnov*	(3) *ExplorativeInnov*	(4) *ExplorativeInnov*	(5) *ExploitativeInnov*	(6) *ExploitativeInnov*
Psedo	−0.003	−0.001	0.002	0.003	−0.006	−0.004
	(−0.55)	(−0.25)	(0.46)	(0.78)	(−1.02)	(−0.70)
TMTconsc	0.879***	0.740***	0.594***	0.521***	0.691**	0.557**
	(3.19)	(2.84)	(3.22)	(2.92)	(2.52)	(2.13)
TMTextra	0.342*	−0.039	0.260**	0.062	0.339*	−0.046
	(1.94)	(−0.23)	(2.11)	(0.51)	(1.96)	(−0.29)
TMTagree	0.662***	0.634***	0.577***	0.544***	0.566***	0.562***
	(3.59)	(3.78)	(4.76)	(4.73)	(3.09)	(3.39)
TMTneuro	−0.080	0.486**	0.261	0.626***	−0.267	0.309
	(−0.31)	(2.12)	(1.48)	(3.73)	(−1.05)	(1.35)
Size		0.271***		0.168***		0.254***
		(7.43)		(7.23)		(7.10)
Lev		0.277*		0.078		0.422***
		(1.69)		(0.75)		(2.64)
RDpro		0.042***		0.014***		0.045***
		(6.51)		(3.50)		(6.88)
SOE		0.052		0.044		0.073
		(0.65)		(0.84)		(0.91)
ROA		3.499***		2.656***		3.483***
		(7.15)		(7.07)		(7.37)
Loss		0.040		0.123**		0.039
		(0.58)		(2.40)		(0.60)
Dual		0.036		0.099***		0.023
		(0.74)		(2.85)		(0.48)
FirmAge		0.070		0.085*		0.046
		(0.93)		(1.69)		(0.62)
Board		0.299**		0.093		0.306**
		(2.31)		(1.06)		(2.41)
TMTage		−0.006		−0.007		−0.000
		(−0.70)		(−1.34)		(−0.03)
TMTshare		0.004***		0.002*		0.004***
		(3.02)		(1.71)		(2.95)
TMTpay		0.146***		0.089***		0.159***
		(3.42)		(2.98)		(3.78)
TMToption		0.190***		0.088***		0.185***
		(4.18)		(2.73)		(4.07)
Female		−0.538**		−0.134		−0.640***
		(−2.49)		(−0.93)		(−3.04)
Constant	−12.189***	−19.598***	−11.425***	−16.068***	−9.329***	−17.038***
	(−3.74)	(−6.23)	(−5.08)	(−6.97)	(−2.90)	(−5.46)
Observations	10,703	10,703	10,703	10,703	10,703	10,703
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.243	0.319	0.147	0.201	0.248	0.328

and close to zero, suggesting that randomly assigned openness values have no systematic relationship with corporate innovation outcomes. This finding indicates that the baseline results are minimally affected by chance correlations or spurious relationships. To comprehensively assess the robustness of our findings, we construct a kernel density distribution of the 1000 placebo coefficient estimates. As illustrated in Fig. 3, the distribution is tightly concentrated around zero with a mean of 0.000, indicating no systematic relationship between randomly assigned *Pseudo* and corporate innovation. Critically, the baseline coefficient on *TMTopen* (0.388 for total innovation, Table 3) lies far outside this placebo distribution, well beyond conventional significance thresholds.

The placebo test results indicate a low probability that the documented relationship between TMT openness and corporate innovation is driven by unobservable confounding variables. Instead, the openness trait of the TMT emerges as the primary factor driving the positive effect on corporate innovation, that is, there is no placebo effect, further validating the reliability and credibility of the empirical findings.

4.3.3. Excluding observations without patents

To further ensure the reliability of our research conclusions, we conduct robustness checks by excluding firms without patent applications and re-estimating the regression models. Innovation activities require substantial resource commitments and organizational capabilities, and firms that never engage in patenting may systematically differ from innovative firms in ways that could affect our estimates. By restricting the analysis to innovation-active firms, we can better isolate the effect of TMT openness on innovation outcomes.

Table 6 presents the results after restricting the sample to patenting firms. The coefficients of *TMTopen* remain significantly positive at the 1% level across all innovation measures, with magnitudes slightly greater than those in the baseline regression. The consistent positive effects suggest that TMT openness plays a robust role in enhancing corporate innovation even when limiting the analysis to firms actively engaged in patenting activities. These robustness checks confirm the reliability and stability of our findings, demonstrating that the results are not driven by systematic differences between patenting and non-patenting firms.

4.3.4. Poisson regression

Given that patent counts are non-negative integers with a substantial proportion of zero values, the presence of numerous zero observations may violate the distributional assumptions of ordinary least squares (OLS) regression and potentially bias the estimates. While our main OLS regressions employ logarithmic transformations of patent counts (*AllInnov*, *ExplorativeInnov* and *ExploitativeInnov*), which partially address the skewed distribution of patent data, we further adopt Poisson regression models as a theoretically appropriate approach for count data (Cameron and Trivedi, 2013).

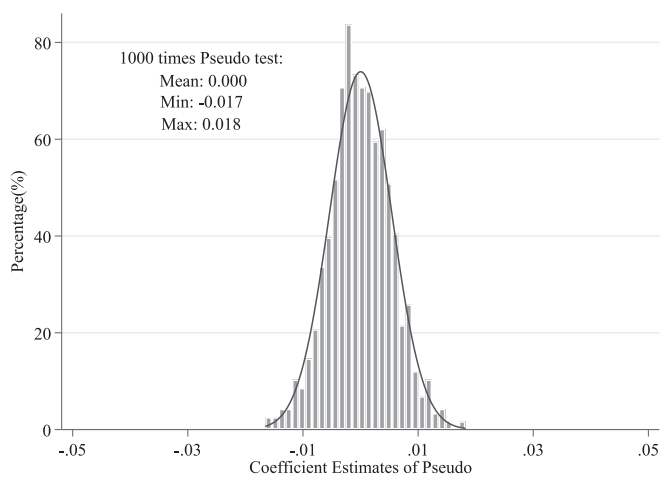


Fig. 3. Kernel density distribution of placebo test coefficients.

Table 6
Excluding observations without patents.

Variables	(1) *AllInnov*	(2) *AllInnov*	(3) *ExplorativeInnov*	(4) *ExplorativeInnov*	(5) *ExploitativeInnov*	(6) *ExploitativeInnov*
TMTopen	0.509***	0.395***	0.514***	0.421***	0.437***	0.322***
	(4.44)	(3.71)	(5.05)	(4.35)	(3.59)	(2.85)
TMTconsc	0.014	0.169	0.307	0.365	−0.037	0.112
	(0.05)	(0.69)	(1.29)	(1.61)	(−0.12)	(0.41)
TMTextra	0.210	−0.188	0.166	−0.033	0.238	−0.187
	(1.27)	(−1.24)	(1.10)	(−0.23)	(1.33)	(−1.16)
TMTagree	0.214	0.336**	0.340**	0.387***	0.190	0.327**
	(1.19)	(2.13)	(2.22)	(2.67)	(0.99)	(1.96)
TMTneuro	−0.454*	0.178	0.157	0.597***	−0.630**	0.035
	(−1.95)	(0.85)	(0.75)	(2.98)	(−2.46)	(0.15)
Size		0.357***		0.227***		0.347***
		(11.24)		(8.34)		(10.33)
Lev		0.275*		0.067		0.460***
		(1.92)		(0.54)		(3.03)
RDpro		0.028***		0.009*		0.036***
		(5.51)		(1.87)		(6.39)
SOE		0.110		0.087		0.117
		(1.46)		(1.36)		(1.46)
ROA		3.118***		2.898***		3.232***
		(7.06)		(6.63)		(6.98)
Loss		0.075		0.146**		0.059
		(1.18)		(2.29)		(0.87)
Dual		0.055		0.123***		0.050
		(1.30)		(3.16)		(1.12)
FirmAge		0.113*		0.117**		0.083
		(1.79)		(2.07)		(1.22)
Board		0.148		0.040		0.211*
		(1.31)		(0.39)		(1.80)
TMTage		−0.013*		−0.010		−0.005
		(−1.77)		(−1.61)		(−0.59)
TMTshare		0.001		0.001		0.002
		(1.37)		(0.97)		(1.61)
TMTpay		0.128***		0.088**		0.161***
		(3.21)		(2.44)		(3.89)
TMToption		0.094**		0.050		0.104**
		(2.39)		(1.39)		(2.48)
Female		−0.460**		−0.108		−0.615***
		(−2.43)		(−0.66)		(−3.12)
Constant	−2.388	−13.514***	−9.712***	−16.738***	−1.001	−12.929***
	(−0.77)	(−4.54)	(−3.52)	(−5.96)	(−0.29)	(−3.99)
Observations	8228	8228	8228	8228	8228	8228
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.147	0.272	0.107	0.170	0.181	0.300

The Poisson regression models utilize the original untransformed patent counts as dependent variables, denoted as *NAllInnov*, *NExplorativeInnov* and *NExploitativeInnov*, respectively. Table 7 reports the Poisson regression results with industry and year fixed effects. The coefficient of *TMTopen* remains significantly positive across all specifications, providing further support for our main hypothesis. In the specifications with full controls, the coefficient of *TMTopen* is 0.512 for total innovation, significant at the 1% level; 0.648 for explorative innovation, significant at the 5% level; and 0.857 for exploitative innovation, significant at the 1% level.

The consistency of findings across both the restricted sample approach and the Poisson regression approach reinforces the reliability of our main conclusions. The significantly positive coefficients of *TMTopen* in the

Table 7
Poisson regression.

Variables	(1) *NAllInnov*	(2) *NAllInnov*	(3) *NExplorativeInnov*	(4) *NExplorativeInnov*	(5) *NExploitativeInnov*	(6) *NExploitativeInnov*
TMTopen	0.807***	0.512***	1.120***	0.648**	1.320**	0.857***
	(4.03)	(3.11)	(3.04)	(2.29)	(2.19)	(2.80)
TMTconsc	0.454	0.596	2.085**	2.014***	1.037	1.536**
	(0.91)	(1.47)	(2.45)	(2.91)	(0.96)	(2.51)
TMTextra	0.603**	−0.195	0.661	−0.022	1.361	−0.090
	(2.32)	(−0.84)	(1.40)	(−0.05)	(1.48)	(−0.21)
TMTagree	0.307	0.523**	0.575	0.771	0.328	0.761**
	(0.98)	(2.10)	(1.02)	(1.59)	(0.50)	(2.17)
TMTneuro	−0.345	0.536	1.029	2.082***	0.355	1.469*
	(−0.81)	(1.47)	(1.40)	(3.12)	(0.42)	(1.84)
Size		0.490***		0.528***		0.782***
		(11.89)		(7.45)		(7.59)
Lev		0.498**		0.300		−0.092
		(2.38)		(0.96)		(−0.22)
RDpro		0.035***		0.018		0.032***
		(5.32)		(1.61)		(3.09)
SOE		0.127		0.003		0.325*
		(1.16)		(0.02)		(1.78)
ROA		5.114***		7.144***		3.838***
		(9.14)		(5.94)		(3.14)
Loss		0.049		0.155		−0.030
		(0.51)		(0.96)		(−0.20)
Dual		0.134*		0.300**		0.179
		(1.88)		(2.13)		(1.42)
FirmAge		0.182*		0.259		0.292*
		(1.90)		(1.43)		(1.70)
Board		0.008		0.251		0.092
		(0.05)		(0.93)		(0.29)
TMTage		−0.013		−0.028		−0.013
		(−1.19)		(−1.49)		(−0.60)
TMTshare		0.002		0.002		0.005
		(1.08)		(0.57)		(1.46)
TMTpay		0.156***		0.290**		0.289**
		(2.75)		(2.45)		(2.09)
TMToption		0.104*		−0.038		0.056
		(1.83)		(−0.28)		(0.55)
Female		−0.711**		0.321		−0.749
		(−2.42)		(0.54)		(−1.44)
Constant	−11.565**	−23.676***	−37.077***	−50.592***	−28.833**	−49.200***
	(−1.99)	(−4.72)	(−3.37)	(−5.26)	(−2.02)	(−4.60)
Observations	10,665	10,665	10,601	10,601	10,623	10,623
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Pseudo-R^2	0.224	0.413	0.170	0.324	0.296	0.581

Poisson models confirm that the positive relationship between TMT openness and firm innovation is not driven by the presence of zero-patent observations, the choice of logarithmic transformation in the main models or sample composition effects. These results provide robust evidence that open-minded TMTs promote both explorative and exploitative innovation.

4.3.5. Multi-period innovation outputs

Innovation is a cumulative process in which R&D investments may take multiple periods to materialize as tangible outputs such as patents. To account for this temporal lag, we employ innovation outputs over the subsequent three years as an alternative measure of corporate innovation intensity.

Table 8 presents the regression results considering multi-period innovation outputs. The coefficients of *TMTopen* remain significantly positive across all innovation measures. In the specifications with full controls,

Table 8
Considering multi-period innovation outputs (including the next three periods).

Variables	(1) *AllInnov3y*	(2) *AllInnov3y*	(3) *ExplorativeInnov3y*	(4) *ExplorativeInnov3y*	(5) *ExploitativeInnov3y*	(6) *ExploitativeInnov3y*
TMTopen	1.959***	1.138***	1.352***	0.846***	1.868***	1.071**
	(4.17)	(2.61)	(4.47)	(3.04)	(3.95)	(2.45)
TMTconsc	0.985	0.879	1.234*	1.186*	0.650	0.535
	(0.91)	(0.91)	(1.75)	(1.82)	(0.61)	(0.56)
TMTextra	0.096	−0.488	0.391	0.074	0.137	−0.451
	(0.14)	(−0.78)	(0.86)	(0.17)	(0.20)	(−0.73)
TMTagree	1.538**	1.818***	1.075**	1.216***	1.033	1.336**
	(2.10)	(2.82)	(2.38)	(2.84)	(1.41)	(2.08)
TMTneuro	0.346	1.592*	0.824	1.621***	−0.322	0.946
	(0.35)	(1.89)	(1.35)	(2.84)	(−0.33)	(1.12)
Size		0.883***		0.572***		0.811***
		(5.61)		(5.91)		(5.28)
Lev		0.889		0.436		1.244**
		(1.41)		(1.12)		(1.98)
RDpro		0.110***		0.042***		0.120***
		(4.42)		(2.80)		(4.71)
SOE		0.341		0.214		0.313
		(1.07)		(1.02)		(0.97)
ROA		12.726***		9.149***		12.560***
		(6.18)		(5.85)		(6.32)
Loss		0.307		0.556***		0.256
		(1.19)		(3.05)		(1.00)
Dual		0.157		0.410***		0.103
		(0.83)		(3.09)		(0.55)
FirmAge		0.431*		0.422**		0.349
		(1.67)		(2.54)		(1.35)
Board		1.033**		0.176		1.069**
		(2.08)		(0.50)		(2.20)
TMTage		−0.011		−0.021		0.005
		(−0.36)		(−1.05)		(0.16)
TMTshare		0.011**		0.005		0.010**
		(2.39)		(1.49)		(2.26)
TMTpay		0.355**		0.190*		0.430***
		(2.10)		(1.70)		(2.59)
TMToption		0.720***		0.308**		0.744***
		(3.96)		(2.46)		(4.06)
Female		−1.591*		−0.616		−1.740**
		(−1.90)		(−1.14)		(−2.10)
Constant	−29.996**	−54.941***	−32.084***	−46.801***	−20.911*	−46.492***
	(−2.33)	(−4.65)	(−3.86)	(−5.69)	(−1.65)	(−3.98)
Observations	5931	5931	5931	5931	5931	5931
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.283	0.375	0.217	0.292	0.276	0.372

the coefficient for total innovation (*AllInnov3y*) is 1.138, significant at the 1% level. Similar patterns emerge for both explorative and exploitative innovation, with coefficients significant at the 1% and 5% levels, respectively. These results confirm that our main findings are robust to alternative temporal specifications of innovation outcomes, further reinforcing the positive role of TMT openness in enhancing corporate innovation across different time horizons.

5. Mechanism analyses

5.1. Managerial innovation awareness

The TMT's level of innovation awareness reflects their cognitive intensity with respect to corporate innovation, encompassing their emphasis on innovation, vision for innovation activities and expectations regarding innovation performance (Chen et al., 2015). TMTs characterized by higher openness typically demonstrate broader cognitive frameworks and enhanced information processing capabilities, enabling them to identify innovation opportunities more effectively and strengthening their innovation awareness.

To examine whether TMT openness enhances managerial innovation awareness, we employ text analysis methodology following Chen et al. (2015) to quantify innovation awareness through MD&A sections of annual reports. The innovation awareness metric (*Innoaware*) is calculated as the ratio of innovation-related keywords to the total word count of the MD&A section. Columns (1) and (2) of Table 9 present the regression results. The coefficient of *TMTopen* is significantly positive at the 1% level, indicating that TMTs with higher openness exhibit significantly stronger innovation awareness. This heightened innovation awareness prompts TMTs to allocate greater resources and strategic attention to corporate innovation initiatives, thereby facilitating enhanced innovation outcomes.

5.2. Tolerance for innovation uncertainties

Innovation is an inherently uncertain and complex exploratory process characterized by substantial investment requirements, extended development cycles and considerable failure risks (Jones and Williams, 2000; Brav et al., 2018). Risk-taking propensity is a pivotal factor influencing corporate innovation (Li, 2010). Higher risk tolerance reflects greater willingness to bear the uncertainties inherent in innovation activities and a stronger commitment to advancing innovation initiatives despite potential setbacks. TMTs characterized by high openness typically demonstrate elevated risk-taking propensity and greater acceptance of potential errors and failures throughout the innovation process, thereby cultivating an organizational environment conducive to innovation.

To examine whether TMT openness enhances corporate innovation through increased tolerance for innovation uncertainties, this paper follows Nakano and Nguyen (2012) and Koerniadi et al. (2014) by using the logarithm of the standard deviation of annualized daily stock returns to quantify the risk-taking propensity of the TMT, serving as a proxy for tolerance toward innovation uncertainties (*Tolerance*). Table 9, Columns (3) and (4) present the regression results. The coefficient of *TMTopen* is significantly positive at the 1% level, indicating that TMTs with higher openness exhibit greater tolerance for failure and risk. This increased tolerance is associated with a stronger inclination toward risky yet potentially rewarding innovation projects, ultimately facilitating corporate innovation activities.

5.3. Reduced managerial myopia

"Managerial myopia" refers to the tendency of managers to prioritize short-term performance over long-term value creation, often manifesting in underinvestment in innovation activities, which require extended time horizons to yield returns (Brochet et al., 2015). TMTs characterized by higher openness typically possess broader temporal perspectives and greater ability to balance short-term pressures with long-term strategic objectives. Such cognitive characteristics enable them to mitigate short-term-oriented behaviors and maintain stronger commitment to innovation initiatives despite immediate uncertainties.

Table 9
Mechanism analysis.

Variables	(1) *Innoaware*	(2) *Innoaware*	(3) *Tolerance*	(4) *Tolerance*	(5) *Myopia*	(6) *Myopia*
TMTopen	0.002***	0.002***	0.085***	0.068***	−0.011***	−0.007***
	(4.89)	(4.07)	(6.01)	(5.12)	(−4.77)	(−3.27)
TMTconsc	0.002*	0.001	0.051	0.032	−0.005	−0.002
	(1.87)	(0.91)	(1.30)	(0.87)	(−0.85)	(−0.40)
TMTextra	−0.002***	−0.001*	−0.053**	0.000	0.000	0.001
	(−3.75)	(−1.72)	(−2.16)	(0.00)	(0.03)	(0.33)
TMTagree	0.000	0.000	−0.083***	−0.092***	−0.002	0.000
	(0.55)	(0.21)	(−3.32)	(−4.02)	(−0.53)	(0.09)
TMTneuro	0.001	0.000	0.086**	0.004	0.010*	0.007
	(0.66)	(0.32)	(2.46)	(0.13)	(1.72)	(1.23)
Size		−0.001***		−0.063***		−0.001
		(−7.00)		(−18.11)		(−0.99)
Lev		−0.002***		0.146***		0.002
		(−4.12)		(8.74)		(0.64)
SOE		0.000*		−0.017**		0.002
		(1.72)		(−2.01)		(1.33)
ROA		0.003*		0.104*		−0.031***
		(1.94)		(1.83)		(−3.77)
Loss		−0.000		0.025***		0.001
		(−0.52)		(2.63)		(0.67)
Dual		0.000*		0.004		−0.003***
		(1.95)		(0.71)		(−3.75)
FirmAge		−0.001**		−0.016**		0.007***
		(−2.40)		(−2.10)		(3.96)
Board		0.001*		−0.003		−0.001
		(1.82)		(−0.19)		(−0.25)
TMTage		0.000		−0.004***		0.000
		(0.65)		(−5.08)		(0.87)
TMTshare		0.000***		0.000		−0.000***
		(2.70)		(1.43)		(−4.49)
TMTpay		0.000		0.000		0.000
		(0.68)		(0.03)		(0.23)
TMToption		0.000***		0.004		−0.003***
		(2.90)		(0.72)		(−3.07)
Female		−0.001*		0.006		0.002
		(−1.71)		(0.27)		(0.36)
Constant	−0.005	0.013	−3.931***	−2.062***	0.125*	0.057
	(−0.45)	(1.24)	(−8.66)	(−4.66)	(1.72)	(0.79)
Observations	10,703	10,703	10,703	10,703	10,703	10,703
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.330	0.362	0.522	0.563	0.078	0.102

To examine whether TMT openness reduces managerial myopia, we construct a managerial myopia measure following Brochet et al. (2015) and Li (2010) through textual analysis of MD&A sections. Specifically, we develop a short-term horizon lexicon and expand it using Word2Vec methodology to capture short-term orientation language. The myopia metric (*Myopia*) is calculated as the proportion of short-term-horizon vocabulary relative to the total MD&A word count, with higher values indicating greater managerial myopia. Columns (5) and (6) of Table 9 present the regression results. The coefficient of *TMTopen* is significantly negative at the 1% level, indicating that TMTs with higher openness exhibit substantially lower levels of managerial myopia. This finding reveals that open-minded TMTs adopt more long-term-oriented strategic

perspectives, showing reduced susceptibility to short-term pressures and thereby facilitating enhanced corporate innovation performance.

6. Further analyses

6.1. Cross-sectional analyses

Personality trait theory emphasizes that individual behavior emerges from the dynamic interaction between intrinsic characteristics and environmental contexts. Consequently, examining the impact of TMT personality traits on corporate strategy necessitates explicit consideration of contextual factors. The relationship between TMT openness and corporate innovation may exhibit systematic variation across different situational conditions. To explore these contingencies, we conduct cross-sectional analyses examining how the effect of TMT openness on corporate innovation varies under diverse contextual circumstances. For brevity, we only report the results using total patent applications as the dependent variable.

6.1.1. Performance pressure

Performance pressure represents a critical contextual factor that may influence the relationship between TMT openness and corporate innovation. Under high performance pressure, TMTs often face heightened market expectations and shareholder demands, compelling them to prioritize short-term financial returns over long-term innovation investments (Yu and Fang, 2024). In such circumstances, TMT openness becomes particularly valuable, enabling executives to maintain long-term perspectives and make forward-looking innovation decisions despite immediate performance pressures.

Following Caskey and Ozel (2017), we measure performance pressure using the absolute difference between a firm's actual earnings and analysts' earnings forecasts. We conduct grouped regressions based on the median performance pressure level. Columns (1) and (2) of Table 10 present the results. In the high performance pressure subsample, the coefficient of *TMTopen* is 0.584, significant at the 1% level, while in the low performance pressure subsample, the coefficient is 0.272, significant at the 10% level. The difference between groups is statistically significant ($p = 0.044$), indicating that TMT openness exerts a more pronounced positive effect on corporate innovation when firms face higher performance pressure.

6.1.2. Proportion of institutional investors

Institutional investors can incentivize innovation and facilitate corporate innovation activities (Aghion et al., 2013). With their sophisticated monitoring capabilities and long-term investment horizons, institutional investors tend to support strategic initiatives that enhance sustainable competitive advantages, including innovation investments. A higher proportion of institutional ownership creates favorable governance conditions and reduces agency conflicts, potentially amplifying the influence of TMT openness on corporate innovation. In such contexts, TMTs with higher openness benefit from enhanced legitimacy and resource support to pursue innovative strategies.

Accordingly, we partition the sample based on the median institutional ownership ratio and conduct grouped regressions. Columns (3) and (4) of Table 10 present the results. In the high institutional ownership subsample, the coefficient of *TMTopen* is 0.515, significant at the 1% level. In contrast, in the low institutional ownership subsample, the coefficient is 0.250 and statistically nonsignificant. The difference between the groups is statistically significant ($p = 0.058$), indicating that TMT openness has a more potent effect on corporate innovation when institutional investor presence is stronger. This finding suggests that the innovation-promoting effect of TMT openness is contingent upon supportive ownership structures that provide the necessary resources and governance environment for innovation investments.

6.1.3. Industry competitiveness

Industry competition intensity may also influence the relationship between TMT openness and corporate innovation. In highly competitive industries, firms face more severe market challenges and competitive threats, heightening their need for innovation to maintain competitive advantages. Competitive pressures not only intensify the demand for innovation but also foster knowledge exchange and diffusion among market partic-

Table 10
Cross-sectional analyses.

Variables	(1) High performance pressure	(2) Low performance pressure	(3) High proportion of institutional investors	(4) Low proportion of institutional investors	(5) High industry competitiveness	(6) Low industry competitiveness
TMTopen	0.584***	0.272*	0.515***	0.250	0.549***	0.181
	(3.82)	(1.86)	(3.19)	(1.63)	(3.10)	(1.24)
TMTconsc	0.646	0.613	1.120***	0.185	0.998**	0.387
	(1.58)	(1.60)	(3.05)	(0.52)	(2.38)	(1.23)
TMTextra	−0.392*	−0.001	−0.327	0.029	−0.363	−0.019
	(−1.66)	(−0.01)	(−1.38)	(0.13)	(−1.38)	(−0.09)
TMTagree	0.324	0.598**	0.393	0.452*	0.651**	0.276
	(1.36)	(2.47)	(1.60)	(1.92)	(2.35)	(1.28)
TMTneuro	0.467	0.568*	0.786**	0.173	0.763**	0.121
	(1.41)	(1.71)	(2.33)	(0.59)	(2.02)	(0.44)
Size	0.219***	0.303***	0.274***	0.253***	0.315***	0.226***
	(4.60)	(6.48)	(5.63)	(4.99)	(5.12)	(5.29)
Lev	0.359*	0.162	0.319	0.310	0.488*	0.128
	(1.65)	(0.81)	(1.39)	(1.48)	(1.76)	(0.68)
RDpro	0.033***	0.042***	0.039***	0.044***	0.051***	0.034***
	(3.85)	(5.37)	(4.07)	(5.44)	(4.95)	(4.07)
SOE	−0.039	0.039	0.066	−0.027	0.088	0.024
	(−0.40)	(0.37)	(0.68)	(−0.24)	(0.64)	(0.27)
ROA	3.160***	4.010***	3.674***	2.921***	5.614***	2.136***
	(4.84)	(5.99)	(5.35)	(4.80)	(6.84)	(3.79)
Loss	−0.063	0.146	0.076	−0.016	0.293***	−0.142*
	(−0.46)	(1.61)	(0.79)	(−0.18)	(2.76)	(−1.71)
Dual	0.038	0.021	0.047	0.011	0.042	0.023
	(0.61)	(0.37)	(0.65)	(0.19)	(0.56)	(0.40)
FirmAge	0.060	0.098	−0.051	0.213**	0.206*	−0.057
	(0.63)	(1.15)	(−0.52)	(2.04)	(1.83)	(−0.60)
Board	0.414**	0.306*	0.329*	0.225	0.037	0.531***
	(2.52)	(1.88)	(1.78)	(1.38)	(0.19)	(3.31)
TMTage	−0.003	−0.004	0.003	−0.015	−0.002	−0.005
	(−0.24)	(−0.42)	(0.27)	(−1.48)	(−0.16)	(−0.51)
TMTshare	0.001	0.004***	0.004**	0.006***	0.003	0.005***
	(0.94)	(2.62)	(2.02)	(3.76)	(1.64)	(2.78)
TMTpay	0.156***	0.107**	0.148**	0.117**	0.187**	0.097*
	(2.95)	(2.08)	(2.52)	(2.04)	(2.56)	(1.95)
TMToption	0.200***	0.118**	0.178***	0.187***	0.197***	0.190***
	(3.47)	(2.05)	(2.75)	(3.21)	(2.73)	(3.36)
Female	−0.602**	−0.620**	−0.257	−0.913***	−0.435	−0.683**
	(−2.26)	(−2.27)	(−0.86)	(−3.23)	(−1.23)	(−2.56)
Constant	−17.312***	−21.168***	−24.070***	−13.756***	−25.770***	−12.703***

(*continued on next page*)

Table 10 (*continued*)

Variables	(1) High performance pressure	(2) Low performance pressure	(3) High proportion of institutional investors	(4) Low proportion of institutional investors	(5) High industry competitiveness	(6) Low industry competitiveness
	(−3.73)	(−4.73)	(−5.40)	(−3.34)	(−4.97)	(−3.40)
Diff. (P-value)	0.044		0.058		0.024	
Observations	5222	4548	5386	5317	4776	5927
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.315	0.324	0.345	0.298	0.184	0.297

Table 11
TMT openness and innovation heterogeneity.

VARIABLES	(1) *LnDigpatent*	(2) *LnDigpatent*	(3) *LnNondigpatent*	(4) *LnNondigpatent*	(5) *Digpatent*	(6) *Digpatent*	(7) *Nondigpatent*	(8) *Nondigpatent*
TMTopen	0.485***	0.389***	0.390***	0.248**	0.983***	0.527***	0.631***	0.418**
	(5.30)	(4.91)	(3.18)	(2.09)	(4.47)	(2.98)	(3.26)	(2.44)
TMTconsc	−0.113	0.008	0.637**	0.520*	−0.242	0.358	0.756	0.757*
	(−0.52)	(0.04)	(2.29)	(1.96)	(−0.40)	(0.81)	(1.53)	(1.77)
TMTextra	0.281*	−0.120	0.136	−0.093	0.448	−0.670***	0.527**	−0.071
	(1.95)	(−0.94)	(0.76)	(−0.54)	(1.28)	(−2.71)	(2.06)	(−0.29)
TMTagree	−0.200	−0.003	0.641***	0.632***	−0.561	−0.051	0.452	0.589**
	(−1.43)	(−0.03)	(3.30)	(3.46)	(−1.59)	(−0.20)	(1.43)	(2.17)
TMTneuro	−0.516***	−0.004	0.120	0.642***	−0.657	0.183	−0.257	0.566
	(−2.65)	(−0.02)	(0.47)	(2.64)	(−1.43)	(0.48)	(−0.64)	(1.56)
Size		0.379***		0.169***		0.614***		0.381***
		(14.89)		(4.72)		(13.57)		(9.43)
Lev		0.221**		0.251		0.605**		0.446**
		(1.99)		(1.54)		(2.17)		(2.10)
RDpro		0.058***		0.015**		0.062***		0.021***
		(11.63)		(2.45)		(10.10)		(2.91)
SOE		0.314***		−0.011		0.396***		0.089
		(5.62)		(−0.13)		(3.63)		(0.85)
ROA		1.017***		3.950***		2.260***		5.379***
		(2.97)		(8.03)		(2.98)		(9.32)
Loss		−0.034		0.133*		−0.031		0.098
		(−0.69)		(1.87)		(−0.27)		(0.96)
Dual		0.068**		0.072		0.170**		0.145**
		(2.08)		(1.46)		(2.18)		(2.02)
FirmAge		0.048		0.041		0.136		0.165
		(0.92)		(0.54)		(1.17)		(1.61)
Board		0.083		0.234*		−0.021		0.067
		(0.90)		(1.77)		(−0.11)		(0.41)
TMTage		−0.009		0.003		−0.007		−0.010
		(−1.57)		(0.38)		(−0.48)		(−0.97)
TMTshare		0.001*		0.003***		0.004**		0.001
		(1.68)		(2.72)		(2.10)		(0.69)
TMTpay		0.135***		0.112***		0.203***		0.132**
		(4.54)		(2.65)		(3.15)		(2.45)
TMToption		0.079**		0.150***		0.110		0.104*
		(2.41)		(3.20)		(1.58)		(1.76)
Female		−0.436***		−0.426*		−0.736**		−0.679**
		(−3.03)		(−1.91)		(−2.56)		(−2.10)
Constant	0.016	−11.468***	−12.620***	−17.660***	0.839	−17.264***	−13.900***	−23.501***
	(0.01)	(−4.92)	(−3.85)	(−5.30)	(0.13)	(−3.67)	(−2.61)	(−4.69)
Observations	10,703	10,703	10,703	10,703	10,665	10,665	10,660	10,660
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
R^2/Pseudo-R^2	0.281	0.429	0.200	0.243	0.311	0.535	0.197	0.319

ipants, which can facilitate innovation processes (Arrow, 1962). Under such conditions, TMT openness becomes increasingly valuable, serving as a critical catalyst that enables firms to transcend conventional approaches and explore novel strategic pathways, thereby exerting a particularly pronounced influence on corporate innovation outcomes.

We measure industry competition intensity using the number of firms within the industry and conduct grouped regressions based on the median value. The results are presented in Columns (5) and (6) of Table 10. In the high-competition subsample, the coefficient of *TMTopen* is 0.549, significant at the 1% level, while in the

Table 12
TMT openness and R&D investment.

Variables	(1) *RDpro*	(2) *RDpro*	(3) *RDemploy*	(4) *RDemploy*
TMTopen	2.212***	1.517***	3.138***	2.349***
	(6.55)	(4.84)	(4.42)	(3.33)
TMTconsc	0.294	0.476	−0.036	−0.510
	(0.27)	(0.47)	(−0.02)	(−0.27)
TMTextra	−1.876***	−0.808	−2.538**	−1.380
	(−3.26)	(−1.46)	(−2.11)	(−1.15)
TMTagree	−0.922	−0.792	−5.073***	−5.364***
	(−1.42)	(−1.31)	(−3.84)	(−4.15)
TMTneuro	1.753**	1.045	−3.191*	−3.475**
	(2.07)	(1.32)	(−1.82)	(−2.04)
Size		−0.262***		−0.496**
		(−3.05)		(−2.35)
Lev		−6.838***		−4.207***
		(−12.64)		(−3.61)
SOE		0.723***		0.690
		(3.15)		(1.53)
ROA		−8.687***		−2.657
		(−5.71)		(−0.80)
Loss		0.922***		−0.005
		(4.02)		(−0.01)
Dual		0.500***		0.473
		(3.26)		(1.50)
FirmAge		−0.994***		−1.313***
		(−3.76)		(−2.87)
Board		−0.376		−1.560*
		(−0.91)		(−1.68)
TMTage		−0.003		0.070
		(−0.11)		(1.22)
TMTshare		0.011***		0.017**
		(3.01)		(2.20)
TMTpay		0.855***		0.361
		(6.58)		(1.41)
TMToption		0.724***		1.648***
		(5.21)		(5.72)
Female		−1.086		−0.296
		(−1.53)		(−0.20)
Constant	−0.214	−3.128	60.241**	74.096***
	(−0.02)	(−0.26)	(2.56)	(3.19)
Observations	10,703	10,703	10,703	10,703
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Adjusted R^2	0.365	0.438	0.537	0.547

low-competition subsample, the coefficient is 0.181 and statistically nonsignificant. The difference between the groups is statistically significant ($p = 0.024$), indicating that TMT openness has a stronger effect on corporate innovation in more competitive industry environments.

6.2. TMT openness and innovation heterogeneity: Digital versus non-digital technology

Beyond overall innovation and the distinction between explorative and exploitative innovation, different innovation types may respond differently to TMT personality characteristics due to their distinct nature

and requirements. Digital technology innovation, characterized by rapid technological evolution, high obsolescence risk and substantial knowledge spillovers (Nambisan et al., 2017), demands particularly high levels of cognitive flexibility, learning agility and tolerance for technological uncertainty. In contrast, non-digital technology innovation, while still requiring innovative capacity, often involves more established technological trajectories and less uncertainty.

To examine this potential heterogeneity, we classify patents into digital and non-digital categories following Tan and Cui (2026), and conduct separate regression analyses for each innovation type. Table 11 presents the results under both OLS and Poisson regression specifications. Columns (1)–(2) report the OLS results with the

Table 13
TMT openness and corporate innovation quality.

Variables	(1) *Cite_all*	(2) *Cite_all*	(3) *Citetop10%*	(4) *Citetop10%*
TMTopen	0.865***	0.643***	0.110***	0.076*
	(5.55)	(4.45)	(2.59)	(1.83)
TMTconsc	0.662*	0.526	−0.076	−0.065
	(1.82)	(1.54)	(−0.84)	(−0.76)
TMTextra	0.173	−0.191	0.046	−0.042
	(0.77)	(−0.89)	(0.77)	(−0.73)
TMTagree	0.309	0.442**	0.103*	0.123**
	(1.27)	(1.99)	(1.69)	(2.03)
TMTneuro	−0.182	0.541*	0.018	0.140
	(−0.55)	(1.82)	(0.21)	(1.61)
Size		0.292***		0.081***
		(6.67)		(3.68)
Lev		0.483**		0.029
		(2.40)		(0.50)
RDpro		0.067***		0.006***
		(7.88)		(2.69)
SOE		0.184*		0.033
		(1.87)		(1.29)
ROA		4.150***		0.502**
		(6.82)		(2.12)
Loss		−0.007		0.017
		(−0.09)		(0.68)
Dual		0.027		0.020
		(0.45)		(1.06)
FirmAge		0.111		0.046*
		(1.20)		(1.93)
Board		0.359**		0.037
		(2.19)		(0.72)
TMTage		0.004		−0.005**
		(0.40)		(−2.07)
TMTshare		0.004***		0.001*
		(2.70)		(1.74)
TMTpay		0.237***		0.043***
		(4.50)		(2.59)
TMToption		0.239***		0.024
		(4.09)		(1.35)
Female		−0.828***		−0.041
		(−3.13)		(−0.71)
Constant	−12.220***	−22.761***	−1.313	−3.597***
	(−2.89)	(−5.63)	(−1.20)	(−2.68)
Observations	10,703	10,703	10,703	10,703
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Adjusted R^2	0.198	0.280	0.041	0.083

natural logarithm of digital patent applications (*LnDigpatent*) as the dependent variable, while Columns (3)–(4) present corresponding results for non-digital patents (*LnNondigpatent*). Columns (5)–(8) replicate these analyses using Poisson regression with raw patent counts (*Digpatent*, *Nondigpatent*). *TMTopen* demonstrates significantly positive effects on both innovation types across all specifications, corroborating our baseline findings. More importantly, the magnitude of the coefficients for digital innovation consistently exceeds that for non-digital innovation, suggesting that TMT openness exerts a more pronounced enhancing effect on digital technology innovation.

Table 14
TMT openness and corporate innovation efficiency.

Variables	(1) *Efficiency_all*	(2) *Efficiency_all*	(3) *Productivity_all*	(4) *Productivity_all*
TMTopen	0.010***	0.007***	0.035**	0.027*
	(6.04)	(4.33)	(2.45)	(1.94)
TMTconsc	0.006	0.000	0.066**	0.036
	(1.53)	(0.06)	(2.01)	(1.13)
TMTextra	−0.010***	−0.006**	−0.035	−0.014
	(−3.94)	(−2.28)	(−1.59)	(−0.62)
TMTagree	0.001	0.002	0.054***	0.054***
	(0.46)	(0.81)	(2.65)	(2.67)
TMTneuro	0.002	0.001	0.038	0.037
	(0.63)	(0.41)	(1.34)	(1.30)
Size		−0.002***		−0.014***
		(−6.44)		(−4.97)
Lev		0.000		0.003
		(0.06)		(0.17)
RDpro		0.001***		−0.001
		(6.49)		(−1.56)
SOE		0.002**		0.013**
		(2.47)		(2.12)
ROA		0.040***		0.266***
		(6.01)		(4.75)
Loss		0.000		0.015*
		(0.56)		(1.83)
Dual		0.002**		0.008
		(2.28)		(1.63)
FirmAge		−0.000		−0.003
		(−0.41)		(−0.39)
Board		−0.000		−0.004
		(−0.25)		(−0.37)
TMTage		−0.000		−0.001*
		(−1.30)		(−1.70)
TMTshare		0.000***		0.000
		(3.30)		(1.55)
TMTpay		0.001*		0.010**
		(1.65)		(2.13)
TMToption		0.001		0.002
		(1.55)		(0.33)
Female		−0.006**		−0.018
		(−2.20)		(−0.78)
Constant	−0.050	0.022	−1.001***	−0.635*
	(−1.08)	(0.50)	(−2.65)	(−1.68)
Observations	10,703	10,703	5674	5674
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Adjusted R^2	0.133	0.186	0.096	0.119

This differential impact reflects the distinct demands of digital technology innovation. Digital technologies evolve rapidly and involve substantial uncertainty, requiring continuous adaptation, experimentation and cross-disciplinary knowledge integration (Nambisan et al., 2017). These characteristics align well with the cognitive flexibility, intellectual curiosity and risk tolerance associated with openness. The results indicate that TMT openness confers stronger advantages in domains characterized by rapid technological change and high complexity.

6.3. TMT openness and R&D investment

Having established that TMT openness significantly enhances corporate innovation outcomes, we further examine its relationship with R&D investment decisions. R&D investment constitutes the foundation for innovation (Chan et al., 1990; Belderbos et al., 2004), and we expect TMTs with higher openness to allocate more resources toward innovation activities given their greater receptivity to novel ideas and stronger innovation awareness.

We examine whether TMT openness translates into greater commitment to innovation inputs by analyzing both financial and human capital investments in R&D. We employ two measures: the ratio of R&D expenditure to operating revenue (*RDpro*), capturing financial investment intensity, and the ratio of R&D personnel to total employees (*RDemploy*), capturing human capital allocation.

Table 12 presents the results, with Columns (1)–(2) reporting R&D expenditure intensity and Columns (3)–(4) reporting the R&D personnel proportion. The coefficients of *TMTopen* are significantly positive at the 1% level across all specifications, indicating that firms led by TMTs with higher openness allocate substantially greater resources to innovation activities across both dimensions. Collectively, these investment decisions provide the necessary resource base for sustained innovation activities, ultimately contributing to the superior innovation performance documented in our main analyses.

6.4. TMT openness and corporate innovation quality

Having established the effects on innovation outcomes and R&D investment, we further examine whether TMT openness influences innovation quality, given that patents vary substantially in their technological contributions. Patent citations serve as a widely accepted proxy for innovation quality, with higher citation counts indicating greater technological significance and impact (Trajtenberg, 1990; Kogan et al., 2017).

Referring to Bradley et al. (2017), this paper constructs two citation-based quality measures using the CNRDS patent citation database. *Cite_all* is calculated as the natural logarithm of one plus the total forward citations (excluding self-citations) received within five years by the firm's patents filed in a given year. *Citetop10%* is calculated as the natural logarithm of one plus the number of patents filed by the firm in a given year whose five-year forward citations (excluding self-citations) rank in the top decile among cited patents in the same application year. Whereas *Cite_all* captures the overall technological impact of the firm's innovation portfolio, *Citetop10%* specifically identifies breakthrough innovations with exceptional influence and significance.

Table 13 presents the regression results. The coefficients of *TMTopen* are significantly positive across both quality measures. These findings indicate that TMT openness not only increases innovation quantity but also significantly enhances the firm's capacity to generate high-quality, highly cited innovations with substantial technological impact.

6.5. TMT openness and corporate innovation efficiency

Previous sections have examined the impact of TMT openness on innovation quantity, quality and input decisions. This section further investigates its influence on innovation efficiency. Innovation efficiency emphasizes the ratio of innovation outputs to resource inputs, reflecting how effectively firms transform resources into innovative outcomes (Afriat, 1972). Higher efficiency indicates that firms generate more innovation outputs per unit of input, representing superior resource utilization in the innovation process.

Following Cooper et al. (2022) and Mukherjee et al. (2017), we employ two proxy variables to measure innovation efficiency. *Efficiency_all* is measured as the ratio of total patent applications in year t to R&D expenditure in year t–1, capturing innovation output per unit of financial investment. *Productivity_all* is measured as the ratio of total patent applications in year t to the number of R&D personnel in year t–1, capturing innovation output per unit of human capital investment.

Table 14 presents the regression results. Columns (1)–(2) report the results using the R&D expenditure-based efficiency measure, while Columns (3)–(4) present the results using the R&D personnel-based productivity measure. The coefficients of *TMTopen* are significantly positive across all specifications, indicating that firms led by TMTs with higher openness demonstrate superior innovation efficiency. These findings suggest that beyond increasing innovation inputs and outputs, TMT openness fundamentally enhances firms' capacity to convert resources into innovative outcomes, thereby improving the overall efficiency of the innovation process.

7. Conclusion and discussion

The Big Five personality framework provides a comprehensive structure for understanding how individual psychological characteristics shape organizational outcomes. Among these dimensions, openness to experience—characterized by intellectual curiosity, creative thinking and receptiveness to novel ideas—assumes particular significance in the innovation context. Given that innovation activities inherently involve substantial resource commitments, extended time horizons and considerable outcome uncertainty (Brav et al., 2018), understanding the role of TMT openness in shaping innovation outcomes is critically important.

This study examines the relationship between TMT openness and corporate innovation using Chinese A-share listed firms from 2007 to 2022. We introduce an innovative approach by employing GLM-4 to assess TMT personality traits based on Q&A transcripts from annual performance briefings and the validated BFI-2 scale. This methodology addresses critical limitations of traditional assessment methods, including restricted sample coverage, substantial resource requirements and inter-rater reliability concerns. To capture the heterogeneous effects of TMT openness across different innovation contexts, we further distinguish between explorative and exploitative innovation, as well as between digital and non-digital technology innovation, providing a nuanced understanding of how executive personality traits interact with innovation characteristics.

Our key findings are as follows. First, TMT openness demonstrates a robust positive association with corporate innovation. Firms led by TMTs with higher openness generate significantly more patents, encompassing both explorative innovation in unfamiliar technological domains and exploitative innovation within established areas. Second, mechanism analyses indicate that the positive effect of TMT openness operates through enhanced innovation awareness, elevated risk tolerance and reduced managerial myopia. Third, cross-sectional analyses reveal that the positive effect is more pronounced in firms experiencing high performance pressure, greater institutional ownership and intense industry competition. Fourth, TMT openness exerts a stronger promotional effect on digital technology innovation than on non-digital technology innovation, reflecting the particular importance of cognitive flexibility and risk tolerance in digital innovation domains characterized by rapid technological evolution and cross-disciplinary complexity. Fifth, further analyses demonstrate that firms with higher TMT openness allocate substantially greater R&D resources to both financial expenditure and human capital. Moreover, TMT openness not only increases innovation quantity but also improves innovation quality and efficiency, as reflected in higher patent citation counts and superior resource conversion effectiveness.

This study contributes to the literature by establishing TMT openness as a critical determinant of corporate innovation. Beyond documenting this relationship, the research elucidates the psychological and behavioral mechanisms through which personality traits shape strategic innovation outcomes. Methodologically, the study employs an LLM (GLM-4) for personality assessment in large-scale empirical research, an approach providing a scalable and replicable framework for incorporating psychological constructs into strategic management scholarship. The findings yield important practical implications. Specifically, firms should systematically integrate personality assessment into executive selection and succession planning, particularly for innovation-intensive positions. Additionally, when constructing TMTs, firms should consider how personality

traits align with organizational contexts, as these contextual factors shape the effectiveness of executive characteristics. Furthermore, this research opens avenues for future studies to examine how other Big Five dimensions influence corporate outcomes and how TMT personality traits shape alternative strategic decisions and organizational behaviors.

Compliance with ethical standards

Research involving human & animal participants: This study does not involve any human or animal participants.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Funding

This research was supported by the Outstanding Innovative Talents Cultivation Funded Programs 2025 of Renmin University of China.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Outstanding Innovative Talents Cultivation Funded Programs 2025 of Renmin University of China.

Appendix. Big Five Inventory-2 (BFI-2) Scale and Prompt Engineering

A. **Big Five Inventory-2 (BFI-2) English Scale**

The Big Five Inventory-2 (BFI-2) is a comprehensive personality assessment instrument developed to measure the five major dimensions of personality. Below is an English translation of the Chinese-language BFI-2 scale used in this study.

1. **Instructions:**

Below are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who "likes to spend time with others"? Please indicate the extent to which you agree or disagree with each statement by writing a number in the blank preceding each statement.

2. **Response Scale:**

1 = Disagree strongly
2 = Disagree a little
3 = Neutral; no opinion
4 = Agree a little
5 = Agree strongly

3.**Items:**

I am someone who....

(1) ___Is outgoing, sociable

(2) ___Is compassionate, has a soft heart
(3) ___Tends to be disorganized
(4) ___Is relaxed, handles stress well
(5) ___Has few artistic interests
(6) ___Has an assertive personality
(7) ___Is respectful, treats others with respect
(8) ___Tends to be lazy
(9) ___Stays optimistic after experiencing a setback
(10) ___Is curious about many different things
(11) ___Rarely feels excited or eager
(12) ___Tends to find fault with others
(13) ___Is dependable, steady
(14) ___Is moody, has up and down mood swings
(15) ___Is inventive, finds clever ways to do things
(16) ___Tends to be quiet
(17) ___Feels little sympathy for others
(18) ___Is systematic, likes to keep things in order
(19) ___Can be tense
(20) ___Is fascinated by art, music, or literature
(21) ___Is dominant, acts as a leader
(22) ___Starts arguments with others
(23) ___Has difficulty getting started on tasks
(24) ___Feels secure, comfortable with self
(25) ___Avoids intellectual, philosophical discussions
(26) ___Is less active than other people
(27) ___Is forgiving
(28) ___Can be somewhat careless
(29) ___Is emotionally stable, not easily upset
(30) ___Has little creativity
(31) ___Is sometimes shy, introverted
(32) ___Is helpful and unselfish with others
(33) ___Keeps things neat and tidy
(34) ___Worries a lot
(35) ___Values art and beauty
(36) ___Finds it hard to influence people
(37) ___Is sometimes rude to others
(38) ___Is efficient, gets things done
(39) ___Often feels sad
(40) ___Is complex, a deep thinker
(41) ___Is full of energy
(42) ___Is suspicious of others' intentions
(43) ___Is reliable, can always be counted on
(44) ___Keeps their emotions under control
(45) ___Has difficulty imagining things
(46) ___Is talkative
(47) ___Can be cold and uncaring
(48) ___Leaves a mess, doesn't clean up
(49) ___Rarely feels anxious or afraid
(50) ___Thinks poetry and plays are boring
(51) ___Prefers to have others take charge
(52) ___Is polite, courteous to others
(53) ___Is persistent, works until the task is finished

(54) ___Tends to feel depressed, blue
(55) ___Has little interest in abstract ideas
(56) ___Shows a lot of enthusiasm
(57) ___Assumes the best about people
(58) ___Sometimes behaves irresponsibly
(59) ___Is temperamental, gets emotional easily
(60) ___Is original, comes up with new ideas

Please check to ensure you have provided a response for every statement.

4. **Scoring Instructions:**

The Big Five personality dimensions and their corresponding items are listed below. Items marked with "R" require reverse scoring (i.e., 1 = 5, 2 = 4, 3 = 3, 4 = 2, 5 = 1). Big Five personality dimensions:

Openness: 5R, 10, 15, 20, 25R, 30R, 35, 40, 45R, 50R, 55R, 60.

Conscientiousness: 3R, 8R, 13, 18, 23R, 28R, 33, 38, 43, 48R, 53, 58R.

Extraversion: 1, 6, 11R, 16R, 21, 26R, 31R, 36R, 41, 46, 51R, 56.

Agreeableness: 2, 7, 12R, 17R, 22R, 27, 32, 37R, 42R, 47R, 52, 57.

Neuroticism: 4R, 9R, 14, 19, 24R, 29R, 34, 39, 44R, 49R, 54, 59.

Note: Domain scores are computed by averaging responses to all items within each dimension after applying reverse scoring where indicated.

B. **Prompt Engineering Specifications**

To operationalize the personality assessment using GLM-4, we developed a structured prompt engineering approach that instructs the LLM to assume the executives' roles based on their Q&A transcripts from annual performance briefings and respond to each BFI-2 item accordingly.

1. **Prompt Template:**

The general prompt structure is as follows:

----- *Q&A Transcripts of [Executive Name] from Annual Performance Briefings* -----

[Transcript Text]

----- *END* -----

You are an expert in role-playing. Please assume the role of ***[Executive Name]****. Above are the Q&A transcripts from your annual performance briefings. Please fully understand and embody the character's personality. Based on* ***[Executive Name]****'s personality, please answer the following question as* ***[Executive Name]****.*

Question:

*Please respond to the statement: "****[BFI-2 Item Statement]****"*

1: Strongly Disagree 2: Disagree 3: Neutral 4: Agree 5: Strongly Agree

Please output only a single integer from 1 to 5 representing your answer. Try to be as objective as possible to reflect the character's personality traits.

2. **Illustrative Example:**

To demonstrate the prompt engineering process, we provide a concrete example below:

Prompt (LLM Input):

----- *Q&A Transcripts of Jinkun Xie from Annual Performance Briefings* -----

For details about the company's operating projects, please carefully read the company's 2005 annual report. If you have any questions, you can call us at any time or log on to the investor relations interactive platform on the company's website. The company's future sustainable development capability is mainly reflected in several aspects: the company has a sound marketing network, a production and management team suitable for its own

development, an elite R&D team, and the brand and reputation established over the years. We believe that the company can maintain sustainable development. This is also the company's business philosophy. Thank you.

----- END -----

You are an expert in role-playing. Please assume the role of Jinkun Xie. Above are the Q&A transcripts from your annual performance briefings. Please fully understand and embody the character's personality. Based on Jinkun Xie's personality, please answer the following question as Jinkun Xie.

Question:

Please respond to the statement: "Is inventive, finds clever ways to do things."

1: Strongly Disagree 2: Disagree 3: Neutral 4: Agree 5: Strongly Agree.

Please output only a single integer from 1 to 5 representing your answer. Try to be as objective as possible to reflect the character's personality traits.

Response (LLM Output):

4

In this example, the executive's response indicates agreement with being characterized as inventive, reflecting a moderate level of openness to experience. This is consistent with the communication style in the transcript, where the executive demonstrates creative thinking by emphasizing the company's elite R&D team and strategic approach to sustainable development, suggesting an appreciation for innovation and novel solutions in business operations.

References

Aabo, T., Pantzalis, C., Park, J.C., Trigeorgis, L., Wulff, J.N., 2024. CEO personality traits, strategic flexibility, and firm dynamics. Finance 84 102524.

Afriat, S.N., 1972. Efficiency estimation of production functions. Int. Econ. Rev., 568–598

Aghion, P., Van Reenen, J., Zingales, L., 2013. Innovation and institutional ownership. Am. Econ. Rev. 103 (1), 277–304.

Arrow, K.J., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R.R. (Ed.), The Rate and Direction of Inventive Activity: Economic and Social Factors. Princeton University Press, Princeton, NJ, pp. 609–625.

Barker III, V.L., Mueller, G.C., 2002. CEO characteristics and firm R&D spending. Manag. Sci. 48 (6), 782–801.

Bebchuk, L.A., Stole, L.A., 1993. Do short-term objectives lead to under-or overinvestment in long-term projects? J. Financ. 48 (2), 719–729.

Belderbos, R., Carree, M., Lokshin, B., 2004. Cooperative R&D and firm performance. Res. Policy 33 (10), 1477–1492.

Benischke, M.H., Martin, G.P., Glaser, L., 2019. CEO equity risk bearing and strategic risk taking: the moderating effect of CEO personality. Strateg. Manag. J. 40 (1), 153–177.

Block, J., 1961. The Q-sort method in personality assessment and psychiatric research.

Bostan, I., Mian, G.M., 2019. Inventor chief executive officers and firm innovation. Int. Rev. Financ. 19 (2), 247–286.

Bradley, D., Kim, I., Tian, X., 2017. Do unions affect innovation? Manag. Sci. 63 (7), 2251–2271.

Brav, A., Jiang, W., Ma, S., Tian, X., 2018. How does hedge fund activism reshape corporate innovation? J. Financ. Econ. 130 (2), 237–264.

Brochet, F., Loumioti, M., Serafeim, G., 2015. Speaking of the short-term: Disclosure horizon and managerial myopia. Rev. Acc. Stud. 20 (3), 1122–1163.

Byun, S.K., Oh, J.M., Xia, H., 2021. Incremental vs. breakthrough innovation: the role of technology spillovers. Manag. Sci. 67 (3), 1779–1802.

Cameron, A.C., Trivedi, P.K., 2013. Regression Analysis of Count Data, No. 53. Cambridge University Press.

Caskey, J., Ozel, N.B., 2017. Earnings expectations and employee safety. J. Account. Econ. 63 (1), 121–141.

Chan, S.H., Martin, J.D., Kensinger, J.W., 1990. Corporate research and development expenditures and share value. J. Financ. Econ. 26 (2), 255–276.

Chemmanur, T.J., Kong, L., Krishnan, K., Yu, Q., 2019. Top management human capital, inventor mobility, and corporate innovation. J. Financ. Quant. Anal. 54 (6), 2383–2422.

Chen, S., Bu, M., Wu, S., Liang, X., 2015. How does TMT attention to innovation of Chinese firms influence firm innovation activities? A study on the moderating role of corporate governance. J. Bus. Res. 68 (5), 1127–1135.

Cooper, M., Knott, A.M., Yang, W., 2022. RQ innovative efficiency and firm value. J. Financ. Quant. Anal. 57 (5), 1649–1694.

Cornaggia, J., Li, J.Y., 2019. The value of access to finance: Evidence from M&As. J. Financ. Econ. 131 (1), 232–250.

Cortes, A.F., Herrmann, P., 2021. Strategic leadership of innovation: a framework for future research. Int. J. Manag. Rev. 23 (2), 224–243.

Costa, P.T., McCrae, R.R., 1992. Normal personality assessment in clinical practice: the NEO Personality Inventory. Psychol. Assess. 4 (1), 5.

Crosby, M., 2000. Patents, innovation and growth. Economic Record 76 (234), 255–262.

D'Acunto, F., 2014. Innovating to invest: The role of basic education. Working paper.

Gao, L., Han, J., Pan, Z., Zhang, H., 2023. Individualistic CEO and corporate innovation: evidence from US frontier culture. Res. Policy 52 (9) 104852.
George, J.M., Zhou, J., 2001. When openness to experience and conscientiousness are related to creative behavior: an interactional approach. J. Appl. Psychol. 86 (3), 513.
Gow, I.D., Kaplan, S.N., Larcker, D.F., Zakolyukina, A.A., 2016. CEO Personality and Firm Policies, No. w22435. National Bureau of Economic Research.
Green, T.C., Jame, R., Lock, B., 2019. Executive extraversion: career and firm outcomes. Account. Rev. 94 (3), 177–204.
Griffin, D., Li, K., Xu, T., 2021. Board gender diversity and corporate innovation: International evidence. J. Financ. Quant. Anal. 56 (1), 123–154.
Hambrick, D.C., Mason, P.A., 1984. Upper echelons: the organization as a reflection of its top managers. Acad. Manag. Rev. 9 (2), 193–206.
Harrison, J.S., Thurgood, G.R., Boivie, S., Pfarrer, M.D., 2019. Measuring CEO personality: developing, validating, and testing a linguistic tool. Strateg. Manag. J. 40 (8), 1316–1330.
Herrmann, P., Nadkarni, S., 2014. Managing strategic change: the duality of CEO personality. Strateg. Manag. J. 35 (9), 1318–1342.
Jiang, Z., Peng, C., Yan, H., 2024. Personality differences and investment decision-making. J. Financ. Econ. 153 103776.
Jones, C.I., Williams, J.C., 2000. Too much of a good thing? The economics of investment in R&D. J. Econ. Growth 5, 65–85.
Koerniadi, H., Krishnamurti, C., Tourani-Rad, A., 2014. Corporate governance and risk-taking in New Zealand. Aust. J. Manag. 39 (2), 227–245.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132 (2), 665–712.
Kong, D., Lin, C., Wei, L., Zhang, J., 2022. Information accessibility and corporate innovation. Manag. Sci. 68 (11), 7837–7860.
Laverty, K.J., 1996. Economic "short-termism": the debate, the unresolved issues, and the implications for management practice and research. Acad. Manag. Rev. 21 (3), 825–860.
Li, F., 2010. The information content of forward-looking statements in corporate filings—a naïve Bayesian machine learning approach. J. Account. Res. 48 (5), 1049–1102.
Marcati, A., Guido, G., Peluso, A.M., 2008. The role of SME entrepreneurs' innovativeness and personality in the adoption of innovations. Res. Policy 37 (9), 1579–1590.
March, J.G., 1991. Exploration and exploitation in organizational learning. Organ. Sci. 2 (1), 71–87.
McGrath, R.G., 2001. Exploratory learning, innovative capacity, and managerial oversight. Acad. Manag. J. 44 (1), 118–131.
Mukherjee, A., Singh, M., Žaldokas, A., 2017. Do corporate taxes hinder innovation? J. Financ. Econ. 124 (1), 195–221.
Nadkarni, S., Herrmann, P.O.L., 2010. CEO personality, strategic flexibility, and firm performance: the case of the Indian business process outsourcing industry. Acad. Manag. J. 53 (5), 1050–1073.
Nakano, M., Nguyen, P., 2012. Board size and corporate risk taking: further evidence from J apan. Corp. Gov. 20 (4), 369–387.
Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M., 2017. Digital innovation management: Reinventing innovation management research in a digital world. MIS Q. 41 (1), 223–238.
Narayanan, M., 1985. Managerial incentives for short-term results. J. Financ. 40 (5), 1469–1484.
Ocasio, W., 1997. Towards an attention-based view of the firm. Strateg. Manag. J. 18 (S1), 187–206.
Ones, D.S., Viswesvaran, C., 1997. Personality determinants in the prediction of aspects of expatriate job success.
O'reilly Iii, C.A., Tushman, M.L., 2008. Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. Res. Organ. Behav. 28, 185–206.
Price, C.R., 2012. Gender, competition, and managerial decisions. Manag. Sci. 58 (1), 114–122.
Richard, O.C., Wu, J., Markoczy, L.A., Chung, Y., 2019. Top management team demographic-faultline strength and strategic change: what role does environmental dynamism play? Strateg. Manag. J. 40 (6), 987–1009.
Roberts, B.W., DelVecchio, W.F., 2000. The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. Psychol. Bull. 126 (1), 3.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Samimi, M., Cortes, A.F., Anderson, M.H., Herrmann, P., 2022. What is strategic leadership? Developing a framework for future research. Leadersh. Q. 33 (3) 101353.
Simon, M., Houghton, S.M., 2003. The relationship between overconfidence and the introduction of risky products: evidence from a field study. Acad. Manag. J. 46 (2), 139–149.
Sitkin, S.B., 1992. Learning through failure: the strategy of small losses. Res. Organ. Behav. 14, 231–266.
Talke, K., Salomo, S., Rost, K., 2010. How top management team diversity affects innovativeness and performance via the strategic choice to focus on innovation fields. Res. Policy 39 (7), 907–918.
Tan, Q., Cui, Y., 2026. Can digital innovation be a new driver of corporate sustainability? China J. Account. Res. 19 (2) 100466.
Trajtenberg, M., 1990. A penny for your quotes: patent citations and the value of innovations. Rand J. Econ., 172–187
Wang, D., Su, Z., Guo, H., 2019. Top management team conflict and exploratory innovation: the mediating impact of market orientation. Ind. Mark. Manag. 82, 87–95.
Wiersema, M.F., Bantel, K.A., 1992. Top management team demography and corporate strategic change. Acad. Manag. J. 35 (1), 91–121.
Xie, L., Zhou, J., Zong, Q., Lu, Q., 2020. Gender diversity in R&D teams and innovation efficiency: Role of the innovation context. Res. Policy 49 (1) 103885.

Yu, X., Fang, J., 2024. Stock transferability, the managerial learning effect and corporate innovation. China J. Account. Res. 17 (1) 100341.
Zhang, B., Li, Y.M., Li, J., Luo, J., Ye, Y., Yin, L., John, O.P., 2022. The big five inventory–2 in China: a comprehensive psychometric evaluation in four diverse samples. Assessment 29 (6), 1262–1284.
Zhao, H., Seibert, S.E., 2006. The big five personality dimensions and entrepreneurial status: a meta-analytical review. J. Appl. Psychol. 91 (2), 259.

Jinglu Zhao is currently a Ph.D. candidate at Business School, Renmin University of China. She obtained her Bachelor's degree in Management from Shandong University. Her research interests include corporate innovation, TMT characteristics, supply chain risk management, ESG, corporate finance and corporate governance.

Zeyu Zhang is currently a Ph.D. candidate at Gaoling School of Artificial Intelligence, Renmin University of China. He obtained his Bachelor's degree in Engineering from Shandong University. His research fields include LLM-based agents, causal discovery and recommender systems. His current research interest is in enhancing the memory capabilities of LLM-based agents. He has published several papers in top-tier conferences and journals, including NeurIPS, KDD, and TKDE. He served as a member of program committee for IJCNLP 2022, IJCNLP 2023, RecSys 2024, NeurIPS 2024, ICLR 2025, AISTATS 2025, ICML 2025. He also served as a reviewer for IEEE Transactions on Knowledge and Data Engineering (TKDE).

China Journal of Accounting Research 19 (2026) 100500

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Computing power projects and digital transformation of enterprises: Evidence from the "East data, West computing" project

Lu Li [a,b], Chuanxing Tan [a,*]

[a] School of Accounting, Zhongnan University of Economics and Law, Wuhan, China
[b] Audit Research Institute, Zhongnan University of Economics and Law, Wuhan, China

ARTICLE INFO

Article history:
Received 7 October 2025
Accepted 20 May 2026
Available online 25 May 2026

Keywords:
The "East data, West computing" project
Computing power projects
Enterprise digital transformation

ABSTRACT

Computing power is becoming a new productive force driving high-quality business growth worldwide. How to lower barriers to computing power usage and facilitate enterprise digital transformation through the systematic deployment of computing power projects remains unclear. Using China's "East data, West computing" project as a quasi-natural experiment, we investigate its impact on the digital transformation of enterprises and find that it significantly promotes this transformation in the region, due to three factors: the optimization of the computing power environment, the strengthening of digital R&D and innovation and government digital subsidies and support. Heterogeneity analysis reveals that the facilitating effect is more pronounced in state owned enterprises, asset light firms and regions with better computing power infrastructure. Furthermore, the effect is stronger in regions with higher energy consumption and better data transmission conditions, suggesting a potential relocation of computing power demand away from energy intensive areas. From the perspective of developing high quality enterprises, firms in computing power hub nodes with more advanced digital transformation achieve higher total factor productivity. From the perspective of cross regional resource flows, the project significantly boosts out of region investment by enterprises located in the Eastern hub node, reflecting their strategic initiative to actively respond to the project. Our paper contributes to the literature on the drivers of digital transformation in enterprises and uses China's experience and evidence in constructing and implementing policies related to global computing power infrastructure.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author.
E-mail address: tanchuanxing0730@163.com (C. Tan).

https://doi.org/10.1016/j.cjar.2026.100500
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Amid the global wave of technological revolution and industrial transformation, digital technologies and the digital economy have become strategic priorities for countries seeking future competitive advantages. Computing power, as the core capability enabling efficient data processing, is emerging as a new productive force driving the digital economy. Many countries have elevated computing power to a national strategic level. For instance, the United States launched its National Strategic Computing Initiative in 2015, China introduced the "East data, West computing" project in 2022 and the European Union unveiled its Artificial Intelligence Action Plan in 2025. Collectively, these initiatives aim to boost economic and social innovation and growth by optimizing the allocation of computing power resources.

As a general-purpose factor of production in the digital economy, computing power is usually defined as the amount of information that a device can process per second (Nordhaus, 2001). Greater computing power implies a stronger data processing capacity, which affects not only the speed of data interaction and response but also the iteration rate of deep learning (Xu et al., 2025). The transmission and scheduling of computing power constitute a computing power network. Such a network can connect ubiquitous and heterogeneous computing power resources to achieve flexible scheduling and efficient supply (Yukun et al., 2024).

In recent years, global computing power has continued to grow and its application scenarios have continually diversified. According to a report published by the China Academy of Information and Communications Technology (CAICT), the United States ranks first and China second in global computing power, with their respective shares accounting for approximately 36% and 30% of the world total. In terms of data center construction, the United States has the world's largest number of hyperscale data centers—roughly three times that of China—with major players such as Amazon, Microsoft and Google holding a dominant position. On the application front, according to the Computing Power Development Report (2024) published by the CAICT, computing power is playing an increasingly important role in sectors such as financial services, healthcare, education, transportation, industrial manufacturing and media and entertainment. The expansion of global computing power and the deepening of its applications reveal that computing power is gradually transforming from a dedicated resource into a universal factor of production. This transformation not only improves the efficiency of computing power as a single factor of production but also contributes to various aspects of high-quality economic development, including enhancing the value of data as a factor of production, lowering barriers to innovation in computing power services and increasing total factor productivity (TFP).

In the context of high-quality economic development, the digital transformation of enterprises is a crucial link to achieving high-quality economic and social development. The literature on the digital transformation of enterprises mainly focuses on four aspects: measurement, driving factors, impact mechanisms and economic consequences. Regarding the driving factors, studies largely concentrate on the external environment, resources and policies such as industry competition, climate policy uncertainty, digital infrastructure construction, fiscal and financial support and a series of pilot policies (Verhoef et al., 2021; Mo and Liu, 2023; Wu et al., 2023; Xie and Wu, 2024; Hua et al., 2025; Tao et al., 2025; Jiang et al., 2026; Yang and Chen, 2026). However, few studies systematically explore the effects of the computing power environment, computing power accessibility and computing power scale on the digital transformation of enterprises, and evidence at the national computing power strategy level is particularly lacking.

In practice, this research gap has important real-world implications. First, in the current context of the rapid development of the digital economy, massive amounts of data are emerging explosively, creating an urgent need for a sound computing power environment and robust computing capacity to provide high-performance computing support for the digital development of enterprises. Second, on a global scale, computing power resources are concentrated among a few large enterprises. Most firms, especially non-digital native firms and small and medium-sized enterprises, face multiple constraints, including the high cost of access to computing power, the high risks associated with digital transformation and insufficient transformation capabilities (Merín-Rodrigáñez et al., 2024; Demann et al., 2026). Therefore, exploring whether and how improvements in the computing power environment and the provision of universal access to computing power can

drive the digital transformation of enterprises is not only of academic value but also of great practical importance for understanding how computing power enables high-quality development of the real economy in the digital age.

Compared with the computing power strategies or initiatives of the United States and European countries, China's "East data, West computing" project offers a quasi-natural experimental setting for testing the above questions. Specifically, first, by deploying eight computing power hub nodes in the Eastern and Western regions and building an integrated national computing power network, the project aims to create a sound computing power environment and provide universal, efficient and green computing power services. Due to its scale, systemic integration and regional differentiation, the project offers a viable perspective for studying how improvements in the computing power environment and universal access to computing power affect the digital transformation of enterprises. Second, the project explicitly states that, to direct different types of computing power toward the national hub nodes, the construction of new large-scale and ultra-large-scale data centers is in principle not permitted outside of the hub nodes, to avoid blind and disorderly interregional competition. Therefore, the regions within the hub nodes of the "East data, West computing" project can be treated as the treatment group and all other regions as the control group. Accordingly, we use a difference-in-differences (DID) model to determine the impact of this project on the digital transformation of enterprises, based on a sample of Chinese A-share non-financial companies listed on the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE) from 2005 to 2023.

The empirical results show that the "East data, West computing" project significantly promotes the digital transformation of enterprises within the hub node regions. Following the construction of these computing power hub nodes, the degree of digital transformation of enterprises increases by 1.893, which is equivalent to 26.696% of the sample mean. Drawing on the resource-based view, this facilitating effect can be attributed to the project's provision of important and scarce external environmental resources and its empowerment of local enterprises, notably through the optimization of the computing power hardware and software environment, the strengthening of digital R&D and innovation and government digital subsidies and aid. Together, these three mechanisms constitute the main drivers of the effect. Second, the results of the heterogeneity analysis indicate that the effect is more pronounced in state-owned enterprises (SOEs), asset-light firms and regions with better conditions for the development of computing power. Moreover, the effect is stronger in regions with higher energy consumption pressure and better data transmission conditions, suggesting that, to advance their digital transformation, enterprises in energy-intensive regions are more likely to rely on computing power networks to transfer their computing power needs to regions with an abundant and lower-cost supply of computing power. Finally, in a further analysis, from the perspective of developing high-quality enterprises, we find that firms within computing power hub nodes that achieve better digital transformation exhibit higher TFP. From the perspective of cross-regional resource flows, we find that after the implementation of the "East data, West computing" project, the number of subsidiary investments made by enterprises in the Eastern hub nodes located in cities with data center clusters in the Western region increases significantly. This not only reflects the precise alignment of enterprises with the deployment of computing power but also highlights the important facilitating role of the focal project in cross-regional resource flows.

This paper makes the following contributions. First, by incorporating computing power projects into the research framework of the driving factors of the digital transformation of enterprises, this study enriches the literature on the micro-level effects of major infrastructure strategies in the context of the digital economy. The literature on external drivers of digital transformation of enterprises largely focuses on aspects such as industry competition, climate policy uncertainty, digital infrastructure construction, fiscal and financial support and a series of pilot policies (Verhoef et al., 2021; Mo and Liu, 2023; Wu et al., 2023; Xie and Wu, 2024; Hua et al., 2025; Tao et al., 2025; Jiang et al., 2026; Yang and Chen, 2026), with little evidence from the perspective of computing power strategy projects. Although one of the initial objectives of the "East data, West computing" project is to support the digital transformation of enterprises by optimizing the allocation of computing power, discussions remain insufficient on issues such as how firms leverage the unique resources of this project for their digital transformation, the circumstances in which the project's effects are more pronounced and whether the project facilitates cross-regional resource flows. This paper innovatively uses the "East data, West computing" project as a quasi-natural experiment, providing empirical evidence of the causal relationship between the project and the digital transformation of enterprises. These findings not only offer a new

reference for promoting the digital transformation of enterprises but also provide empirical evidence from the perspective of developing computing power projects for high-quality development of micro-level enterprises.

Second, drawing on the resource-based view and using the implementation pathways of the "East data, West computing" project, we discuss the core mechanisms through which this project enables the digital transformation of enterprises within the hub node regions. We deconstruct the internal logic of the project's promotion of the digital transformation of enterprises along three dimensions: the development of the computing power project, empowerment and impact. Specifically, the three mechanisms are as follows: first, the optimization of the hardware and software environment required for the digital transformation of enterprises through the implementation of the computing power project; second, the promotion of enterprises' digital R&D and innovation through the strengthening of computing power and algorithms; and third, the alleviation of firms' financing constraints by encouraging local governments to provide digital subsidies, thereby helping enterprises acquire computing power equipment and accelerating their digital transformation. Together, these analyses provide a theoretical understanding of how the "East data, West computing" project drives the digital transformation of enterprises.

Third, from the multidimensional perspectives of cross-regional resource allocation, green and sustainable development and high-quality economic development, this paper broadens the economic and social effects of computing power projects, providing a reference for the implementation and advancement of computing power strategies on a global scale. Specifically, first, we confirm that the implementation of the "East data, West computing" project prompts enterprises in the Eastern computing power hub nodes to establish subsidiaries in the Western data centers, thereby facilitating cross-regional resource allocation. Second, we reveal that, through the computing power network, the project encourages enterprises in energy-intensive regions to seek external computing power services, thereby advancing their digital transformation, which indirectly reflects the green and sustainable development effects of the computing power project. Third, within the computing power hub nodes, enterprises with better digital transformation exhibit higher TFP, indicating that they leverage the unique resources provided by the computing power project to accelerate their digital transformation and achieve high-quality development. These findings not only broaden our understanding of the economic and social effects of computing power projects but also offer a reference for the implementation and advancement of computing power strategies worldwide.

2. Policy background, literature review and theoretical hypotheses

2.1. Policy background

With the rapid development of the global digital economy and digital technologies, the types and volume of data generated by economic and social activities have increased considerably (Jones and Tonetti, 2020). The storage and processing of these data have greatly increased the demand for computing power. In contrast to the narrower notion of "computing power" (as in computational capacity), the broader concept of computing power encompasses communication capabilities, computational methods, processing speed, storage capacity, data volume and service capabilities. It represents a new qualitative force of production that integrates high-performance computing, network transmission capacity and data storage capacity (Yu, 2024). Based on this broader definition, new technologies, new business forms and new models related to big data, artificial intelligence, the Internet of Things and 5G networks all rely on computing power (Xu et al., 2025). In this context, many countries around the world are investing in the construction of high performance computing infrastructures (Deelman et al., 2025), with the aim of providing efficient and convenient data processing capabilities through robust computing power, thus offering strong technical support for high-quality economic and social development.

In China, however, the geographical distribution of computing power has long suffered from a regional imbalance: the Eastern region has high demand but scarce resources, whereas the Western region has abundant resources but low demand. At the same time, there is a persistent tension between high energy consumption and low carbon targets. To address these issues, enhance universal access to computing power and build an integrated national computing power network, China has developed and begun to implement the "East

data, West computing" project. The project's development can be traced back to a series of informatization initiatives proposed in 2016, which laid the groundwork for subsequent computing power infrastructures.

After reviewing the development trajectory of the "East data, West computing" project, we identify three stages: preliminary deployment; launch and construction; and continuous advancement. (1) Preliminary deployment stage (2016–late 2020): This foundational phase focused on building the digital infrastructure and setting up data centers. (2) Launch and construction stage (early 2021–early 2022): This initiation phase involved the planning of computing power hub nodes, demonstration of construction plans and full-scale launch. In February 2022, the approval of 8 hub nodes and 10 national data center clusters was completed, the overall design of the integrated national big data center system was finalized and the "East data, West computing" project was officially launched. (3) Continuous advancement stage (early 2022–present): This roll-out phase features the ongoing construction of hub nodes and data center clusters, the development of computing power scheduling platforms and in-depth strategic cooperation between hub nodes and enterprises. Since February 2022, the 8 hub nodes and 10 clusters have successively launched activities, actively built scheduling and trading platforms, pursued technological breakthroughs to reduce data transmission costs and aimed to deliver universal, efficient and green computing power services.

By the end of 2024, China had more than 9 million standard racks in use in computing centers, a year-on-year increase of about 11% and its computing power capacity had reached 280 EFLOPS (exaflops, i.e., 10^{18} floating-point operations per second), up about 22% from the previous year. Fig. 1 shows that, overall, China's total computing power has exhibited a clear upward trend since 2022, with the scale of standard racks in use also growing at an accelerated pace. This suggests that after the full launch of the "East data, West computing" project, China's computing power infrastructure has developed rapidly and its computing power capacity has grown quickly to meet the demands of economic and social development.

2.2. Literature review and theoretical hypotheses

Scholars argue that digital transformation is the process by which firms leverage digital technologies to achieve a radical change in their operational efficiency and business performance (Westerman and Bonnet, 2015). A high-quality digital transformation enables firms to develop a new type of competitiveness through digital technologies. This new competitiveness differs from the traditional competition model that relies on cost control and channel reach; instead, it enables the implementation of new business models (Ritter and Pedersen, 2020). Moreover, a high-quality digital transformation accelerates the pace of innovation (Nwankpa and Roumani, 2016), helping firms shift from passive adaptation to active followership in a rapidly changing policy and technological environment.

As part of their digital transformation, it is important to recognize that many enterprises around the world are facing various transformation challenges. The main challenge arises from the high-risk nature of digital

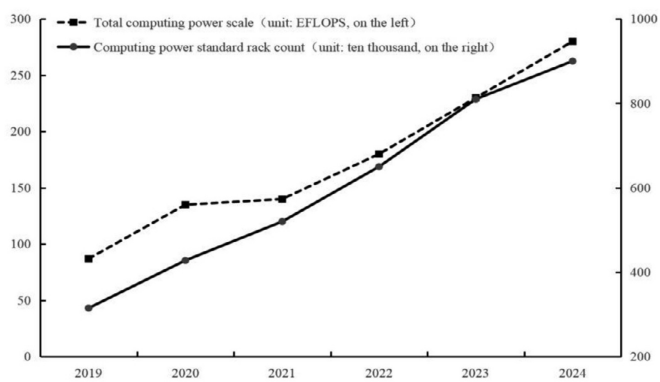


Fig. 1. China's total computing power scale and standard racks from 2019 to 2024[1]. [1] The figure is based on public data published by the CAICT, official government websites and other sources (e.g., the CAICT's annual White Paper on China's Computing Power Development Index) and is compiled by the author.

technology innovation, as digital transformation requires digital technology innovation. For digital native firms, this innovation demands substantial human, financial and R&D resources, a favorable external policy environment, a robust digital ecosystem and adequate computing power and algorithms. Coupled with the long cycles, high complexity and pronounced uncertainty of these innovation activities (Firk et al., 2022; Huang et al., 2023), firms are likely to get stuck in their transformation efforts. For non-digital native firms, the barriers to digital technology innovation are even higher and the difficulty greater (Xu et al., 2026), often trapping them in a "capability-induced transformation deadlock."

Second, long transformation cycles and high transformation costs hinder the digital transformation of enterprises. This transformation constitutes a comprehensive reform for enterprises, both commercially and functionally, involving the breaking of old interdepartmental models and the reconstruction of new ones. Long transformation cycles, substantial capital investment and high additional costs put pressure on the long-term survival and stable and sustainable operation of enterprises (Gurbaxani and Dunkle, 2019), making them "unwilling to transform" in the face of digital transformation.

Third, the low organizational acceptance of digital transformation is also an important factor hindering the digital transformation of enterprises. Some studies suggest that the use of robots and artificial intelligence will replace some jobs and reduce low-skilled positions (Acemoglu and Restrepo, 2020; Acemoglu et al., 2022). This implies that digital transformation will lead to job losses, lowering employees' acceptance of digital transformation and indirectly impeding the digital transformation of enterprises. In addition, an organization's dependence on traditional and fixed work patterns, as well as its receptivity to transformation, affect its digital transformation to some extent (Liu et al., 2025).

In response to these transformation dilemmas, the literature explores the drivers of the digital transformation of enterprises from multiple perspectives. First, from the perspective of corporate governance, scholars typically focus on the effects of newly appointed and younger CEOs, corporate grouping, family culture and peer pressure on the digital transformation of enterprises (Zhang et al., 2023; Zou et al., 2024; Xiang et al., 2025; Zhao and Ren, 2025; Yang and Chen, 2026). Second, from the perspective of digital construction, research shows that improved digital infrastructure and digital technology innovation enhance enterprises' digital transformation capabilities and promote high-quality development (Huang et al., 2023; Wu et al., 2023). Third, from the perspective of policy and pilot programs, research mainly focuses on mechanisms such as improved financing environments, policy support and talent agglomeration, examining the effects of green credit policies, big data pilot zones, talent attraction policies and urban innovation policies on the digital transformation of enterprises (Sun et al., 2023; Liu et al., 2024; Zhang et al., 2024; Hua et al., 2025; Yu et al., 2025). Fourth, from the perspective of fiscal and financial policies, some scholars examine the positive role of government subsidies, tax incentives and other instruments (Wang et al., 2023a, 2023b; Ding et al., 2025; Han et al., 2025).

In summary, the literature identifies the drivers of the digital transformation of enterprises along two main lines: internal resources and external policy environments. However, few studies systematically examine the impact of the external computing power environment, computing power scale and accessibility on the digital transformation of enterprises, and empirical evidence at the national computing power strategy level is particularly lacking. Computing power, as a core factor of production in the digital economy, raises important questions about how its external environment, scale, cost and accessibility reshape the pathways of enterprises' digital transformation. China's "East data, West computing" project offers a unique and suitable setting to address this research gap. Specifically, as a major national strategy for coordinating the allocation of computing power resources, the project aims to promote the synergistic development of diverse types of computing power, facilitate the integrated application of computing power, data and algorithms and build an integrated national computing power infrastructure network. It is characterized by its scale, systemic integration and regional differentiation. Furthermore, the project explicitly stipulates that, to direct different types of computing power toward the national hub nodes, the construction of new large-scale or ultra-large-scale data centers outside of these nodes is in principle not permitted, to avoid disorderly competition. This allows use to designate the hub node regions of the project as the treatment group and all other regions as the control group. In summary, the "East data, West computing" project constitutes a quasi-natural experiment that allows us to identify the causal effects of improving the external computing power environment and universal computing power on the digital transformation of enterprises, and to explore the underlying mechanisms.

The resource-based view holds that the basis for a firm to achieve a sustainable competitive advantage lies in the possession of valuable, rare, inimitable and non-substitutable strategic resources (Wernerfelt, 1984; Barney, 1991; Wu et al., 2023). Such resources include internally accumulated assets and capabilities, as well as unique external resources and conditions that a firm can access. From this theoretical perspective, the "East data, West computing" project can be viewed as a systematic, state-led platform for the creation and allocation of computing power resources. By improving the computing power infrastructure environment, aggregating digital R&D and innovation resources and encouraging local governments to provide complementary subsidies, the project offers a unique and valuable external environment and resource base for the digital transformation of enterprises within its computing power hub nodes. Therefore, drawing on the resource-based view, we identify the mechanisms through which the computing power project affects the digital transformation of local enterprises from three aspects: the optimization of the computing power hardware and software environment, the strengthening of digital R&D and innovation and the provision of government digital subsidies and aid.

2.2.1. Optimization of the computing power hardware and software environment under the "East data, West computing" project

The "East data, West computing" project promotes the digital transformation of enterprises within its hub node regions by optimizing the local computing power hardware and software environment. The regional computing power environment typically consists of computing resources such as data service centers, computer equipment and cloud servers; network conditions for data transmission, including broadband and fiber optics; and the soft environment, such as raising awareness among local authorities about computing power. One of the key objectives of the project is to strengthen the systematic design of new computing power infrastructures, to build integrated computing centers covering general-purpose computing, intelligent computing and supercomputing and to establish hub nodes as national computing power hubs. This objective can significantly enhance awareness among local governments in the eight national hub node regions of computing power and accelerate the design and implementation of computing power infrastructures. A good understanding of computing power, coupled with a supportive environment and robust infrastructure, will support the digital transformation of enterprises, from both a software and hardware perspective.

Specifically, first, raising local governments' awareness of computing power can lead them to allocate resources and steer industries toward enterprises related to computing power, thus creating a virtuous circle between the development of the computing power project and the digital transformation of enterprises. Second, increasing the quantity and improving the quality of computing power infrastructure enhances data transmission capacity and provides convenient and efficient access to and use of massive computing resources (Wang et al., 2017). These improvements not only provide stable and reliable data transmission and processing capabilities for the digital transformation of local enterprises, thereby strengthening their application of digital technologies, but also help to reduce the high cost of leasing computing equipment and servers from external providers, encouraging firms to invest more capital in digital transformation and thus alleviating the problems of "not knowing how to transform" and "being unwilling to transform" (Liu and Chen, 2025). Therefore, we argue that the "East data, West computing" project can optimize the computing power hardware and software environment in the hub nodes, thereby supporting the digital transformation of local enterprises.

2.2.2. Strengthening digital R&D and innovation under the "East data, West computing" project

The "East data, West computing" project enhances the digital transformation of enterprises by increasing their digital R&D and innovation capabilities in the hub node regions. We examine how the project improves the external conditions for R&D and innovation to provide enterprises with the necessary R&D capabilities and technologies for their digital transformation. The project's computing power hub nodes and data center clusters bring together a large number of cloud service providers, computing power enterprises and research institutions. For example, in September 2022, the Ningxia hub node signed cooperation agreements with

several enterprises, including China Mobile Ningxia, ZTE and Tencent, to build a high-speed direct-connection network infrastructure and an integrated computing power scheduling system.[1] Similarly, the Shaoguan data center cluster, located in the hub node of the Guangdong–Hong Kong–Macao Greater Bay Area, has actively cooperated with Huawei and Lenovo to establish the Northern Guangdong Big Data Industry College and engage in long-term collaboration.[2] The agglomeration of these advanced computing power enterprises, support resources and research institutions implies that intelligent computing power, advanced algorithms and high-quality talent are concentrated in the hub nodes, thus supporting the digital transformation of local enterprises.

Specifically, first, the agglomeration of intelligent computing power, advanced algorithms and high-quality talent within the computing power hub nodes enables local enterprises to access computing resources and cutting-edge algorithms at a lower cost, helping them break through bottlenecks in digital R&D and innovation and facilitating their digital transformation with more solid digital innovation capabilities and technological outcomes (Huang et al., 2023). Second, active industry–university–research collaboration between enterprises and research institutions not only accelerates the real-world application of digital R&D innovation but also fosters the training of specialized digital talent through these institutions, thus providing enterprises with high-quality technological and talent support to pursue their sustained and steady digital transformation (Ju and Shao, 2026). Third, the agglomeration of advanced computing power enterprises and factors in the hub nodes facilitates the diffusion of digital technologies across industrial chains and regions, generating technology spillovers and innovation synergies for upstream and downstream firms as well as for enterprises in surrounding areas, thus enabling a broader digital transformation. Thus, we argue that the "East data, West computing" project attracts advanced enterprises, talent and technologies, thereby helping to enhance the digital R&D and innovation capabilities of local enterprises and supporting their digital transformation.

2.2.3. Government digital subsidies and aid under the "East data, West computing" project

By guiding and incentivizing local governments to implement digital subsidy policies, the "East data, West computing" project helps local enterprises obtain more digital subsidies, thereby alleviating their financial pressure to transform, helping them to purchase computing power equipment and increase their R&D investments and, ultimately, fostering their digital transformation. During the development of the project, to facilitate the creation of national computing power hub nodes and data center clusters, as well as their coordinated development with local enterprises, local governments granted various levels of subsidies and rewards in their incentive policies. For example, to advance the construction of the National (Guizhou) Hub Node, Guizhou province coordinated relevant fiscal special funds between various departments and regions to increase support for computing power development. In addition, Guizhou supports local enterprises in developing standardized and commercialized computing power products by allocating a certain amount of annual funds as rewards.[3] To support the construction of the National Hub Node (Gansu), Gansu province coordinated funding support for enterprises conducting computing power transactions within the Qingyang data center cluster and for listed companies relocating their registered addresses to the Qingyang cluster.[4] Tao et al. (2025) show that government digital subsidies can effectively support digital technology innovation in local enterprises.

Specifically, first, under the influence of the "East data, West computing" project, the digital subsidies and related financial support provided by local governments directly fund the construction of local data centers and data platforms, as well as the purchase of cloud servers and computer equipment. This helps local enterprises overcome technological barriers to digital transformation, especially those struggling with insufficient funding and technology. Second, the digital subsidies and financial support granted by local governments

[1] Source: "The First 'Western Data Valley' Computing Power Industry Conference of the Ningxia National Computing Power Hub Was Successfully Held."

[2] Source: "The Big Data Industry Investment Promotion Symposium of the Shaoguan Data Center Cluster, National Integrated Computing Power Network Guangdong-Hong Kong-Macao Greater Bay Area Hub Node."

[3] Source: "Several Incentive Policies for Promoting the Construction of the National (Guizhou) Hub Node of the National Integrated Computing Power Network."

[4] Source: "Notice of the General Office of the People's Government of Gansu Province on Issuing Several Measures to Support the Construction and Operation of the National Hub Node (Gansu) of the National Integrated Computing Power Network."

can alleviate enterprises' concerns about transformation arising from capital constraints, providing assistance with equipment procurement, talent cultivation and R&D innovation during the digital transformation process, thereby accelerating the pace of this transformation. Third, substantial subsidies from local governments often signal government optimism regarding enterprise transformation, helping to play a "safety net role" in the process (Sun et al., 2023), enhancing enterprises' ability to bear risks and their willingness to transform and thus strengthening their confidence in embarking on the digital transformation journey. Therefore, we argue that under the influence of the "East data, West computing" project, enterprises receive more digital subsidies from local governments, which facilitates their digital transformation.

Based on the above triple mechanism analysis, we put forward the following hypothesis:

Hypothesis 1. Through the three channels of the optimization of the computing power hardware and software environment, the strengthening of digital R&D and innovation and government digital subsidies and aid, the "East data, West computing" project facilitates the digital transformation of local enterprises.

Based on the above analysis, Fig. 2 presents the conceptual framework for the mechanism analysis in this paper.

3. Research design

3.1. Data and sample

We use a sample of Chinese A-share non-financial companies listed on the SSE and SZSE from 2005 to 2023 to empirically examine the impact of the "East data, West computing" project on the digital transformation of enterprises. Considering that the project was fully launched in February 2022, and taking into account the data available at the regional and firm levels, the sample period is set from 2005 to 2023. The firm-level data mainly come from the China Stock Market and Accounting Research (CSMAR) database and the Chinese Research Data Services platform. The city- and province-level data primarily come from the China City Statistical Yearbook, the China Statistical Yearbook, the CEInet Statistics Database and the CSMAR database. Some data are collected manually and compiled by the authors.

Following common practice in the literature, the sample selection procedure is as follows: (1) we exclude listed financial companies; (2) we exclude companies in serious financial distress, including ST and *ST firms; (3) we replace missing values for some key variables using interpolation (e.g., missing values of some city-level control variables for 2023 are imputed using extrapolation); and (4) we exclude firm-year observations with missing key variables. To mitigate the influence of outliers, all main continuous variables are winsorized at the 1% level on both tails. Our final sample consists of 38,260 firm-year observations from 4033 firms.

3.2. Variable construction

3.2.1. The "East data, West computing" project

The "East data, West computing" project (*EDWC*) explicitly stipulates that, to direct different types of computing power toward the national hub nodes, the construction of new large-scale and ultra-large-scale data centers outside of these nodes is in principle not permitted, to avoid blind and disorderly interregional

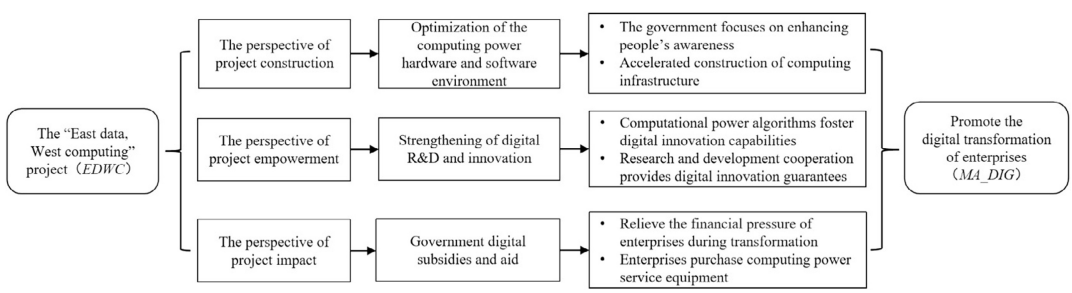


Fig. 2. Conceptual framework.

competition. This allows us to designate the regions of the national computing power hub nodes as the treatment group and all other regions as the control group. The eight national computing power hub nodes of the project are the Beijing–Tianjin–Hebei region, the Yangtze River Delta, the Guangdong–Hong Kong–Macao Greater Bay Area, the Chengdu–Chongqing region, Guizhou, Inner Mongolia, Ningxia and Gansu. Accordingly, we define the cities located in these eight regions as the treatment group (*Treat*), to which we assign a value of 1 for the period from 2022 (*Post*), and all other cities as the control group, to which we assign a value of 0.

It should be noted that we define the treatment group based on the presence or not of a computing power hub node in a city. This approach is motivated by the following considerations. First, consistent with the theoretical framework, the construction of computing power infrastructure and associated incentive policies are largely concentrated within the hub nodes, which can have a direct and significant impact on local firms. Second, although firms located outside of the hub nodes can also remotely access computing resources through the computing power network, this may bias the estimated treatment effect of the project on the digital transformation of enterprises. However, in a DID framework, if such cross-regional computing services benefit certain control firms, the baseline regression estimates would become conservative, i.e., the true effect would be underestimated. This does not necessarily call into question the baseline finding that the "East data, West computing" project promotes the digital transformation of enterprises in the hub nodes.

3.2.2. Digital transformation of enterprises

Following Chen and Srinivasan (2024) and Guo and Duan (2024), we measure the degree of digital transformation of enterprises by the frequency of digital-related keywords appearing in the Management Discussion and Analysis (MD&A) section of corporate annual reports, where a higher frequency indicates a higher degree of digital transformation. This measure is motivated by the following considerations. First, the MD&A section is a formal channel through which management communicates the firm's strategy, operating performance and future prospects to external stakeholders (Guo and Duan, 2024), and its textual content reflects the firm's strategic priorities and resource allocation intentions. Second, despite some measurement errors inherent in text-based word frequency analysis, this method remains widely adopted in the literature (Chen and Srinivasan, 2024). Moreover, several studies indicate that digital word frequency can better reflect the degree of digital transformation of enterprises (Tan et al., 2025). To enhance the robustness of our findings, we conduct the following robustness checks: first, we standardize the measurement of the variables; second, we replace the dependent variable with alternative proxies, including the financial indicator "proportion of digital intangible assets" and the governance indicator "enterprise digital technology labor demand" (Müller et al., 2018; Tan et al., 2025).

3.2.3. Control variables

Following Xu et al. (2025) and Zhao et al. (2026), we include a set of control variables at the firm and city levels in the baseline regressions. The firm-level control variables include firm size (*Size*), leverage ratio (*Lev*), return on assets (*Roa*), operating cash flow (*Cash*), fixed asset investment (*Capital*), executive compensation (*Top3*), firm age (*Age*) and ownership type (*Soe*). The city-level control variables include regional gross domestic product (*Gdp*), urban population (*Pop*), urban fiscal pressure (*Fiscal*), industrial development base (*Firm*), share of secondary industry (*Gdp2*) and share of tertiary industry (*Gdp3*). In addition, firm fixed effects (*Firm FE*) and year fixed effects (*Year FE*) are included. Table 1 presents the definitions of our main variables.

3.3. Empirical models

Given the regional layout of the "East data, West computing" project across the country, we use a DID model, dividing the sample into treatment and control groups based on whether they are affected by the project, to examine its impact on the digital transformation of local enterprises. The baseline regression model is specified as follows:

$$MA_DIG_{i,t} = \beta_0 + \beta_1 EDWC_{c,t} + \sum \beta Controls_{i,c,t} + \mu_i + \gamma_t + \varepsilon_{i,c,t} \quad (1)$$

Table 1
Variable definitions.

Symbols	Variables	Definition
MA_DIG	Enterprise digital transformation	The frequency of digital-related keywords appearing in the MD&A section of corporate annual reports.
EDWC	The "East data, West computing" project	Cities hosting the eight national computing power hub nodes are assigned a value of 1 for the year 2022 and thereafter, and 0 otherwise.
Size	Firm size	The natural logarithm of total assets at year-end.
Lev	Leverage ratio	The ratio of total liabilities to total assets at year-end.
Roa	Return on assets	Net profit divided by total assets at year-end.
Cash	Operating cash flow	Net cash flow from operating activities divided by total assets at year-end.
Capital	Fixed asset investment	The natural logarithm of a firm's net fixed assets.
Top3	Executive compensation	The natural logarithm of the compensation of the top 3 senior executives in a firm.
Age	Firm age	Current year minus the year the firm was founded.
Soe	Ownership type	State-owned enterprises are coded as 1, and 0 otherwise.
Gdp	Regional gross domestic product	Natural logarithm of urban gross domestic product.
Pop	Urban population	Natural logarithm of urban year-end total population.
Fiscal	Urban fiscal pressure	(Local budgetary fiscal expenditure − local budgetary fiscal revenue) / local budgetary fiscal revenue.
Firm	Industrial development base	Natural logarithm of the number of industrial enterprises above designated size in the city.
GDP2	Share of secondary industry	Share of secondary industry in urban GDP.
GDP3	Share of tertiary industry	Share of tertiary industry in urban GDP.

where i, c and t denote the firm, the city and the year, respectively; $MA_DIG_{i,t}$ represents the degree of digital transformation of firm i in year t; and $EDWC_{c,t}$ is a dummy variable indicating whether the city is located in one of the eight national computing power hub nodes after the implementation of the "East data, West computing" project, taking the value of 1 if city c is located in a hub node in 2022 or later and 0 otherwise. The coefficient β_1 captures the impact of the computing power project in the hub nodes on the digital transformation of enterprises, relative to non-hub nodes. A positive β_1 indicates that the computing power project enhances the digital transformation of local enterprises. $Controls_{i,c,t}$ denotes a set of firm- and city-level control variables; μ_i and γ_t represent firm and year fixed effects, respectively; and $\varepsilon_{i,c,t}$ is the error term. In the baseline regressions, we use cluster-robust standard errors at the city level.

4. Empirical results

4.1. Descriptive statistics

Table 2 reports the descriptive statistics of our main variables. The mean value of *MA_DIG* is 7.091, indicating that, on average, the frequency of digital-related keywords in the MD&A section of the sample firms' annual reports is approximately 7. The standard deviation of *MA_DIG* is 17.661, with a maximum of 152 and a minimum of 0, suggesting a substantial variation in the degree of digital transformation among the sample firms. The mean value of *EDWC* is 0.143, indicating that about 14% of the sample firms belong to the treatment group. Among the control variables, except for a few variables with interpolated values that exhibit slight differences in extreme values compared with the literature (e.g., the maximum value of *Gdp2* is 1.034), the descriptive statistics of the other variables are similar to those reported in prior studies.

4.2. Baseline regression results

Table 3, columns (1) to (3), reports the baseline regression results. Column (1) includes only the policy variable and controls for firm and year fixed effects. The coefficient of *EDWC* is 2.263, significant at the 5% level, indicating that the "East data, West computing" project significantly improves the digital transformation of local enterprises. Columns (2) and (3) add firm-level and city-level control variables, respectively, to

Table 2
Descriptive statistics.

Variables	N	Mean	SD	Min	P25	Median	P75	Max
MA_DIG	38,260	7.091	17.661	0.000	0.000	1.000	5.000	152.000
EDWC	38,260	0.143	0.350	0.000	0.000	0.000	0.000	1.000
Size	38,260	22.125	1.338	18.679	21.172	21.903	22.845	27.655
Lev	38,260	0.405	0.201	0.029	0.242	0.399	0.556	1.236
Roa	38,260	0.042	0.062	−0.636	0.017	0.042	0.071	0.437
Cash	38,260	0.051	0.069	−0.243	0.012	0.049	0.090	0.380
Capital	38,260	20.110	1.736	12.910	19.067	20.028	21.098	25.136
Top3	38,260	14.474	0.772	10.747	13.998	14.483	14.944	16.819
Age	38,260	17.542	6.356	2.000	13.000	17.000	22.000	38.000
Soe	38,260	0.383	0.486	0.000	0.000	0.000	1.000	1.000
Gdp	38,260	9.016	1.151	4.475	8.216	9.168	9.944	10.738
Pop	38,260	6.489	0.662	3.827	6.098	6.546	6.964	8.137
Fiscal	38,260	0.512	0.662	−0.189	0.141	0.290	0.586	6.051
Firm	38,260	8.094	0.995	4.290	7.568	8.181	8.837	9.841
GDP2	38,260	0.400	0.118	0.069	0.323	0.414	0.481	1.034
GDP3	38,260	0.560	0.138	0.000	0.463	0.536	0.642	0.896

the specification in column (1). The coefficient of *EDWC* remains positive and significant at the 5% level, suggesting that the positive effect persists after controlling for potential confounding factors at the firm and city levels, thus supporting H1. According to the coefficient in column (3), after the establishment of the national computing power hub nodes as part of the project, the degree of digital transformation of enterprises increases by 1.893, which is equivalent to 26.696% of the mean level (7.091). In economic terms, the project has a substantial promoting effect on the digital transformation of local enterprises. Furthermore, we classify the digital-related keywords into two categories: basic digital technologies (*BasicDig*) and applications of digital technologies (*ApplyDig*). The frequencies of these two types of keywords are counted and included in the regressions, and the results are shown in columns (4) and (5) of Table 3, respectively. The results show that the coefficient of *EDWC* is significant and positive for both basic digital technologies and applications of digital technologies. Given that the coefficients are 1.691 and 0.268, respectively, this indicates that the "East data, West computing" project has a greater enhancing effect on basic digital technologies than on the applications of digital technologies.

4.3. Mechanism analysis

4.3.1. Optimization of the computing power hardware and software environment under the "East data, West computing" project

Based on the above theoretical analysis, the national computing power hub nodes of the "East data, West computing" project not only raise local governments' awareness of computing power but also improve local computing power infrastructure. The resulting favorable computing power software and hardware environment promotes the digital transformation of local enterprises. To test this, we first summarize the keywords related to computing power by reading the project's Implementation Opinions and Guiding Opinions,[5] then we supplement and refine the list through manual review and finally we use Python to count the frequency of these keywords in local government work reports. This frequency serves as a measure of local governments' attention to computing power, i.e., the computing power "software" environment (*SuanLi*). Second, following Zhao et al. (2020), we use the regional digital financial inclusion index (*Finance*), regional telecommunications revenue (*Telephone*), the number of mobile phone subscribers (*Mobileuser*) and the number of mobile internet users (*Internetuser*) to measure regional digital infrastructure and the development environment, i.e., the

[5] The extracted keywords include computing, computing power, computing network, computing energy, algorithms and "East data, West computing."

Table 3
Baseline regression results.

Variables	(1)	(2)	(3)	(4)	(5)
	MA_DIG	*MA_DIG*	*MA_DIG*	*BasicDig*	*ApplyDig*
EDWC	2.263**	2.097**	1.893**	1.691**	0.268*
	(2.43)	(2.30)	(2.46)	(2.28)	(1.91)
Size		2.156***	2.120***	1.313***	0.744***
		(8.51)	(8.33)	(5.89)	(8.12)
Lev		−0.114	−0.265	−0.256	−0.198
		(−0.07)	(−0.19)	(−0.22)	(−0.58)
Roa		−12.159***	−11.986***	−11.605***	−0.189
		(−6.40)	(−6.27)	(−6.74)	(−0.28)
Cash		−2.008*	−2.010*	−1.859*	−0.172
		(−1.67)	(−1.69)	(−1.92)	(−0.45)
Capital		0.282	0.287	0.400*	−0.060
		(1.09)	(1.10)	(1.92)	(−0.58)
Top3		0.215	0.292	0.226	0.031
		(0.68)	(0.84)	(0.81)	(0.36)
Age		−0.431	−0.348	−0.259	−0.294**
		(−0.75)	(−0.57)	(−0.66)	(−2.17)
Soe		1.759*	1.647	1.687*	−0.084
		(1.69)	(1.58)	(1.85)	(−0.32)
Gdp			1.212	1.759	−0.517
			(1.13)	(1.62)	(−1.64)
Pop			3.976***	2.881***	1.045***
			(4.54)	(4.39)	(3.23)
Fiscal			−0.131	−0.356	0.199*
			(−0.30)	(−0.95)	(1.76)
Firm			−2.700***	−2.798***	0.034
			(−2.70)	(−2.64)	(0.18)
GDP2			−0.331	0.311	−1.779
			(−0.04)	(0.04)	(−0.62)
GDP3			−4.197	−3.826	−1.494
			(−0.40)	(−0.43)	(−0.49)
Constant	6.768***	−42.139***	−56.228***	−40.993***	−9.195**
	(51.51)	(−4.64)	(−4.18)	(−4.44)	(−2.23)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	38,260	38,260	38,260	38,260	38,260
Adj.R^2	0.697	0.702	0.703	0.686	0.622

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

computing power "hardware" environment. We divide the above variables into two groups (better vs. worse computing power software/hardware environment) based on the annual median, assigning them values of 1 and 0, respectively. These variables are then interacted with the *EDWC* variable and included in the regression.

Table 4 reports the regression results for the optimization of the computing power software and hardware environment. Specifically, column (1) uses our measure of local governments' attention to computing power, while columns (2) to (5) use indicators related to the regional digital development environment and digital infrastructure. The results in column (1) show that the interaction term $EDWC \times SuanLi$ is positive and significant at the 10% level, indicating that the promoting effect of the project on the digital transformation of local enterprises is stronger in regions where governments pay more attention to computing power. This suggests that the computing power project affects the digital transformation of enterprises by raising local governments' awareness of computing power. The results in columns (2) to (5) show that the interaction terms $EDWC \times Finance$, $EDWC \times Telephone$, $EDWC \times Mobileuser$ and $EDWC \times Internetuser$ are all significant

Table 4
Mechanism analysis: optimization of the computing power hardware and software environment[a].

Variables	(1)	(2)	(3)	(4)	(5)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*
EDWC	1.406*	0.436	−0.261	−0.283	−0.247
	(1.71)	(0.58)	(−0.37)	(−0.40)	(−0.33)
SuanLi	−0.361**				
	(−2.29)				
EDWC × SuanLi	1.069*				
	(1.81)				
Finance		−0.188			
		(−0.27)			
EDWC × Finance		2.104**			
		(2.38)			
Telephone			−0.370		
			(−0.79)		
EDWC × Telephone			3.390***		
			(4.76)		
Mobileuser				−0.596	
				(−1.44)	
EDWC × Mobileuser				3.478***	
				(4.27)	
Internetuser					−0.302
					(−0.68)
EDWC × Internetuser					3.350***
					(3.96)
Size	2.120***	2.355***	2.121***	2.133***	2.135***
	(8.35)	(8.40)	(8.43)	(8.46)	(8.42)
Lev	−0.276	−0.523	−0.240	−0.266	−0.237
	(−0.20)	(−0.32)	(−0.17)	(−0.19)	(−0.17)
Roa	−12.044***	−11.792***	−11.783***	−11.845***	−11.816***
	(−6.31)	(−5.60)	(−6.17)	(−6.17)	(−6.22)
Cash	−1.972*	−2.403*	−2.005*	−2.007*	−1.924
	(−1.66)	(−1.87)	(−1.69)	(−1.71)	(−1.64)
Capital	0.288	0.634**	0.284	0.281	0.283
	(1.11)	(2.14)	(1.09)	(1.08)	(1.09)
Top3	0.301	0.164	0.335	0.313	0.329
	(0.87)	(0.49)	(0.96)	(0.90)	(0.94)
Age	−0.353	−0.522	−0.294	−0.302	−0.302
	(−0.59)	(−0.61)	(−0.49)	(−0.50)	(−0.50)
Soe	1.659	1.674	1.688	1.685	1.635
	(1.60)	(1.44)	(1.62)	(1.62)	(1.58)
Gdp	1.270	1.374	1.342	1.340	1.417
	(1.15)	(1.00)	(1.34)	(1.29)	(1.36)
Pop	3.838***	3.454***	3.395***	3.140***	3.223***
	(4.51)	(3.43)	(4.37)	(3.60)	(3.68)
Fiscal	−0.139	−0.070	0.065	−0.035	0.052
	(−0.32)	(−0.16)	(0.15)	(−0.08)	(0.12)
Firm	−2.696**	−2.428*	−2.408***	−2.285***	−2.442***
	(−2.59)	(−1.92)	(−2.81)	(−2.71)	(−2.90)
GDP2	0.010	−6.637	0.119	−1.293	−0.154
	(0.00)	(−0.48)	(0.01)	(−0.14)	(−0.02)
GDP3	−4.276	−12.123	−4.897	−5.965	−4.677
	(−0.40)	(−0.79)	(−0.47)	(−0.57)	(−0.44)
Constant	−55.871***	−55.311***	−57.269***	−54.990***	−56.640***
	(−4.21)	(−2.80)	(−4.51)	(−4.22)	(−4.44)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	38,260	32,597	38,260	38,260	38,260
Adj.R^2	0.703	0.725	0.704	0.704	0.704

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

[a] The reason for the reduced sample size in column (2) is that the data for the digital financial inclusion index are available only from 2011 onward.

and positive. This indicates that the project's promoting effect on the digital transformation of enterprises is more pronounced in regions with better digital infrastructure, implying that the project influences enterprises' digital transformation by improving regional computing power infrastructure and the digital development environment. Overall, these findings are broadly consistent with our previous theoretical analysis, suggesting that optimizing the computing power software and hardware environment is a potential mechanism through which the "East data, West computing" project affects the digital transformation of enterprises.

4.3.2. Strengthening of digital R&D and innovation under the "East data, West computing" project

Based on the above theoretical analysis, the national computing power hub nodes of the "East data, West computing" project provide local enterprises with advanced, fast and efficient computing power and algorithm support for digital R&D and innovation, thereby enhancing their digital R&D and innovation capabilities and providing them with reliable technical support for their digital transformation. To examine the strengthening of digital R&D and innovation as a potential mechanism, we analyze it from two perspectives: firms' digital R&D and digital technology innovation. First, from a digital R&D perspective, we screen the development spending items of listed firms to identify their current-year increase in digital-related R&D projects, and we use this increase to measure their digital R&D investment (*Rd*). Specifically, we assign a value of 1 to projects with a positive increase and 0 otherwise, then we interact this dummy variable with the *EDWC* variable and include the interaction term in the regression. Second, from a digital technology innovation perspective, following Huang et al. (2023), Wu et al. (2025) and Tan and Cui (2026), we use the number of digital economy patents (*Patent1*) and digital technology patents (*Patent2*) to measure firms' digital innovation capacity. We then take the natural logarithm of these digital innovation variables and divide firms into two groups (high vs. low digital innovation level) based on the annual industry median, assigning them values of 1 and 0, respectively. These dummy variables are then interacted with the *EDWC* variable and included in the regression.

Table 5 reports the regression results for the strengthening of digital R&D and innovation. Column (1) shows that the interaction term $EDWC \times Rd$ is positive and significant at the 1% level, indicating that the project's promoting effect on the digital transformation of local enterprises is stronger for those with higher investment in digital R&D. Columns (2) and (3) show that the interaction terms $EDWC \times Patent1$ and $EDWC \times Patent2$ are positive and significant at the 1% level, indicating that the project's promoting effect is stronger for firms with higher digital innovation levels. Overall, these results suggest that the "East data, West computing" project promotes the digital transformation of enterprises by enhancing their digital R&D and innovation capabilities. Taken together, the above findings are consistent with our previous theoretical analysis, confirming that strengthening digital R&D and innovation is an important mechanism through which the "East data, West computing" project affects the digital transformation of local enterprises.

4.3.3. Government digital subsidies and aid under the "East data, West computing" project

Based on the above theoretical analysis, the national computing power hub nodes of the "East data, West computing" project encourage local governments to provide policy and financial support to the project and to business entry, to accelerate the construction of computing centers equipped with large-scale or ultra-large-scale computing power infrastructure. When firms receive more government digital subsidies, they can develop advanced technologies, attract digital talent and purchase more computers and cloud service equipment, thus benefiting from a signal of government support for computing power and digital development, which enhances their ability and motivation to pursue their digital transformation. To measure government digital subsidies, following the literature (Tao et al., 2025), we screen the government subsidy notes in the annual reports of listed firms using digital-related keywords to identify items related to such subsidies and construct a dummy variable (*Subsidy*) indicating whether a firm receives such subsidies (1 = yes, 0 = no). Furthermore, to examine whether firms that receive government digital subsidies subsequently purchase more computer equipment to facilitate their digital transformation, we screen the notes on fixed assets in the annual reports for original value items related to computers or cloud services, and construct a dummy variable (*Asset*) indicating whether a firm purchases such fixed assets (1 = yes, 0 = no). Finally, we interact these variables with the *EDWC* variable and include the interaction terms in the regression.

Table 6 reports the regression results for government digital subsidies and aid. Column (1) shows that the interaction term $EDWC \times Subsidy$ is positive and significant at the 1% level, indicating that the project's

Table 5
Mechanism analysis: strengthening of digital R&D and innovation.

Variables	(1)	(2)	(3)
	MA_DIG	*MA_DIG*	*MA_DIG*
EDWC	0.867	0.631	0.651
	(1.29)	(0.84)	(0.91)
Rd	4.144***		
	(7.07)		
EDWC × Rd	10.646***		
	(8.12)		
Patent1		−0.350	
		(−1.58)	
EDWC × Patent1		2.417***	
		(3.49)	
Patent2			−0.144
			(−0.61)
EDWC × Patent2			2.690***
			(5.20)
Size	1.974***	2.105***	2.064***
	(7.60)	(8.72)	(8.36)
Lev	−0.139	−0.197	−0.157
	(−0.10)	(−0.14)	(−0.11)
Roa	−11.386***	−12.101***	−12.037***
	(−5.94)	(−6.34)	(−6.34)
Cash	−1.651	−2.034*	−1.899
	(−1.54)	(−1.73)	(−1.64)
Capital	0.274	0.281	0.273
	(1.01)	(1.07)	(1.05)
Top3	0.294	0.298	0.291
	(0.83)	(0.86)	(0.84)
Age	−0.401	−0.318	−0.264
	(−0.67)	(−0.52)	(−0.43)
Soe	1.646*	1.682	1.681
	(1.66)	(1.62)	(1.61)
Gdp	1.049	1.201	1.262
	(1.06)	(1.14)	(1.19)
Pop	3.734***	3.891***	3.867***
	(4.04)	(4.53)	(4.41)
Fiscal	−0.069	−0.113	−0.090
	(−0.17)	(−0.26)	(−0.21)
Firm	−2.343***	−2.686***	−2.687***
	(−2.77)	(−2.77)	(−2.78)
GDP2	0.606	0.084	−0.416
	(0.07)	(0.01)	(−0.05)
GDP3	−2.388	−3.721	−4.263
	(−0.24)	(−0.36)	(−0.41)
Constant	−53.470***	−56.163***	−55.970***
	(−4.08)	(−4.11)	(−4.09)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	38,260	38,260	38,260
Adj.R^2	0.710	0.703	0.703

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

promoting effect on the digital transformation of local enterprises is stronger for those receiving more government digital subsidies. This reflects that the project influences the digital transformation of enterprises by increasing their government digital subsidies. In column (2), the dependent variable is whether a firm

Table 6
Mechanism analysis: government digital subsidies and aid.

Variables	(1)	(2)	(3)
	MA_DIG	*Asset*	*MA_DIG*
EDWC	1.474*	−0.016	0.448
	(1.95)	(−1.48)	(0.71)
Subsidy	−0.425	−0.004	
	(−1.55)	(−0.78)	
EDWC × Subsidy	2.067***	0.020**	
	(3.90)	(2.41)	
Asset			−0.035
			(−0.07)
EDWC × Asset			2.752***
			(4.38)
Size	2.126***	0.017*	2.125***
	(8.34)	(1.80)	(8.31)
Lev	−0.281	−0.023	−0.286
	(−0.20)	(−0.95)	(−0.20)
Roa	−12.007***	0.017	−11.844***
	(−6.26)	(0.53)	(−6.22)
Cash	−2.033*	−0.023	−2.027*
	(−1.72)	(−0.72)	(−1.72)
Capital	0.285	−0.002	0.271
	(1.10)	(−0.34)	(1.06)
Top3	0.286	0.003	0.289
	(0.83)	(0.34)	(0.84)
Age	−0.369	−0.017	−0.268
	(−0.61)	(−1.14)	(−0.44)
Soe	1.634	−0.001	1.609
	(1.57)	(−0.04)	(1.54)
Gdp	1.262	0.005	1.364
	(1.18)	(0.16)	(1.27)
Pop	3.923***	−0.036	3.842***
	(4.49)	(−1.23)	(4.76)
Fiscal	−0.123	0.005	−0.089
	(−0.28)	(0.39)	(−0.21)
Firm	−2.737***	−0.011	−2.771***
	(−2.73)	(−0.66)	(−2.80)
GDP2	−0.352	0.074	−0.485
	(−0.04)	(0.21)	(−0.05)
GDP3	−4.075	−0.009	−4.097
	(−0.39)	(−0.03)	(−0.39)
Constant	−55.627***	0.662	−57.270***
	(−4.16)	(1.57)	(−4.35)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	38,260	38,260	38,260
Adj.R^2	0.703	0.821	0.703

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

purchases fixed assets (*Asset*). The results show that the interaction term $EDWC \times Subsidy$ is positive and significant at the 5% level, indicating that when firms receive more government digital subsidies, the project significantly increases their purchases of such fixed assets. Column (3) shows that the interaction term $EDWC \times Asset$ is positive and significant at the 1% level, indicating that the project's promoting effect on the digital transformation of local enterprises is stronger for those with more purchases of computers, cloud

services and other fixed assets. Taken together, these results reflect that thanks to the project, local enterprises obtain more digital subsidies and rewards from local governments, thereby enhancing their digital transformation capabilities. Furthermore, the project encourages firms to receive more digital subsidies, which helps them purchase more computer equipment and thus promotes their digital transformation. In summary, the above results indicate that digital subsidies from local governments constitute an important mechanism through which the "East data, West computing" project promotes the digital transformation of enterprises.

4.4. Endogeneity and robustness tests

4.4.1. Parallel trend test

Satisfying the parallel trend assumption is a prerequisite for causal inference when using the DID method. That is, before the establishment of the national computing power hub nodes of the "East data, West computing" project, the treatment and control groups should exhibit similar trends in digital transformation. If this assumption holds, the post-implementation differences can be attributed to the project itself. Following the literature (Xiao and Shen, 2022; Shen et al., 2024), we use an event study approach to test whether the parallel trend assumption is satisfied. Specifically, we examine whether the coefficients for the pre-policy period are close to 0 and statistically indistinguishable from 0. If so, the treatment and control groups are considered to have similar pre-trends, thereby satisfying the parallel trend assumption and validating our results. In practice, following Sun et al. (2023), we define 10 year dummies ($Time_n$, where $n = -8, -7, \ldots, 0, 1$). To avoid multicollinearity, we use $n = -1$ as the base period. For $n = -8, -7, \ldots, -1$, the dummies respectively indicate the 8th, the 7th, ..., the 1st year before the establishment of the hub nodes; $n = 0$ and $n = 1$ indicate the year of establishment and the first year after, respectively. These year dummies are then interacted with the treatment variable (*Treat*) and the interaction terms replace the original *EDWC* variable in the baseline regression. The regression results are presented in Fig. 3.

The parallel trend results show that for $n < 0$, the coefficients of the interaction terms are close to 0 and not significant, indicating that there is no significant difference in the digitalization level between the treatment and control groups before the establishment of the national computing power hub nodes. Thus, the parallel trend assumption is satisfied. For $n > 0$, i.e., in the year of establishment and the first year after, the coefficients of the interaction terms are significantly greater than 0, confirming the positive impact of the establishment of the hub nodes of the "East data, West computing" project on the digital transformation of enterprises.

4.4.2. Instrumental variable test

Although our results pass the parallel trend test, they may still be subject to potential endogeneity issues such as reverse causality and omitted variables. To address this, following Zhao et al. (2020), we use the

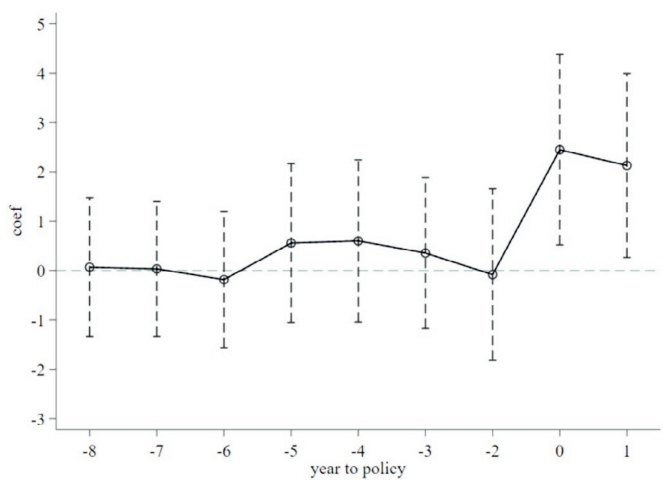


Fig. 3. Parallel trend test.

number of telephones per 10,000 people in 1984 as an instrumental variable (*IV*) in a two-stage estimation. Before proceeding, *IV* must satisfy the relevance and exogeneity conditions. In terms of relevance, regions with a larger number of telephones per 10,000 people in 1984 tend to have better digital infrastructure and informationization levels, which is conducive to the subsequent establishment of national computing power hub nodes. Thus, the relevance condition is met. In terms of exogeneity, most scholars argue that the 1984 telephone penetration rate is a historical fact and is unlikely to be directly affected by the "East data, West computing" project implemented only in recent years; it can influence the degree of enterprise digitalization only through relevant policies. Hence, the exogeneity condition is satisfied.

Because this variable is cross-sectional, we interact it with the post-implementation year dummy (*Post*) of the "East data, West computing" project to form our instrumental variable (*IV*), which is then used as an instrument for the *EDWC* variable in model (1). The regression results are presented in Table 7. Column (1) reports the first-stage results, where *IV* is significantly positively correlated with *EDWC*, indicating that a higher number of telephones per 10,000 people in 1984 is associated with a greater likelihood of being

Table 7
Instrumental variable test.

Variables	(1)	(2)
	EDWC	*MA_DIG*
	First stage	Second stage
EDWC		8.173***
		(3.05)
IV	0.157***	
	(4.32)	
Size	0.007*	2.096***
	(1.68)	(7.85)
Lev	−0.028	−0.164
	(−1.32)	(−0.11)
Roa	−0.026	−12.034***
	(−0.90)	(−6.09)
Cash	0.015	−1.910
	(0.72)	(−1.56)
Capital	0.004	0.289
	(1.44)	(1.11)
Top3	−0.004	0.245
	(−1.07)	(0.67)
Age	0.007	−0.252
	(0.61)	(−0.40)
Soe	0.012	1.918*
	(1.22)	(1.76)
Gdp	0.051	0.928
	(0.98)	(0.87)
Pop	−0.019	3.431***
	(−0.30)	(3.99)
Fiscal	0.024	0.057
	(0.72)	(0.12)
Firm	0.026	−2.391***
	(0.74)	(−3.10)
GDP2	−0.206	−1.943
	(−0.45)	(−0.21)
GDP3	−0.528	−4.245
	(−1.03)	(−0.39)
Firm FE	Yes	Yes
Year FE	Yes	Yes
N	36,373	36,373
KP-F statistic	18.687	

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

designated as a national computing power hub node, consistent with the relevance expectation. In column (2), the coefficient of *EDWC* remains significant and positive in the second stage, confirming the robustness of our findings. Moreover, the Kleibergen–Paap F statistic is 18.687, exceeding the critical value. Overall, these results demonstrate that after addressing endogeneity concerns, the "East data, West computing" project still has a significant promoting effect on the digital transformation of enterprises.

4.4.3. Exclude other policy effects

During the sample period, China implemented a number of digital pilot policies, such as the triple-play (telecommunications, television and computer) network integration pilot launched in 2010, the smart city pilot launched in 2013, the "Broadband China" pilot and the information benefiting the people pilot launched in 2014 and the national big data comprehensive pilot zone pilot launched in 2016. To rule out the potential influence of these policies on our findings, following the literature (Shen et al., 2024), we manually collect the regions and years of these pilot policies and incorporate them as additional control variables in the regression. These variables include the "Broadband China" pilot (*KDChina*), the smart city pilot (*Wisecity*), the national big data comprehensive pilot zone pilot (*Bigdata*), the artificial intelligence innovation and development pilot zone pilot (*AIarea*), the national information benefiting the people pilot (*Infor*), the triple-play network integration pilot (*Threeweb*) and the big data administration bureau pilot (*DATAgov*). The regression results including these policy pilots are presented in Table 8. After controlling for these policy factors, the coefficient of *EDWC* remains significant and positive, confirming the robustness of our baseline findings.

4.4.4. Robustness test at the sample level

At the sample level, following the literature (Sun et al., 2023; Shen et al., 2024), we conduct a series of robustness checks to confirm the robustness of our findings, as detailed below.

(1) Changing the sample period. To mitigate the potential influence of an excessively long pre-event window, we rerun the regressions using the 2009–2023 and 2016–2023 time windows. The results are presented in Table 9, columns (1) and (2). Second, given that the effective impact period of the "East data, West computing" project currently only covers 2 years (2022–2023) and that the project requires time to encourage the digital transformation of enterprises, we extend the time window to 2024 to obtain more reliable results. Currently, many city-level variables for 2024 have missing values. Therefore, we proceed as follows: first, we collect the frequency of digital transformation keywords in the MD&A section of corporate annual reports for 2024; second, we retain the firm-level control variables following the same treatment as in the baseline regression; third, we include city fixed effects to account for uncontrolled city-level variables; and finally, we re-estimate the model using this new sample. The results are shown in Table 9, column (3).[6] The results in columns (1) to (3) show that the coefficient of *EDWC* remains significant and positive across all sample specifications, indicating that our regression results are robust regardless of the time window used.

(2) Excluding the four municipalities. Given that the four municipalities (Beijing, Shanghai, Tianjin and Chongqing) exhibit higher levels of economic development and a greater agglomeration of digital industries, enterprises in these regions may exhibit higher degrees of digital transformation than those in other areas, which could bias our regression results. To address this, we rerun the regression after excluding samples from Beijing, Shanghai, Tianjin and Chongqing. The results are presented in Table 9, column (4), where the coefficient of *EDWC* remains significant and positive, confirming the robustness of our findings.

[6] Due to the large number of missing city-level variables for 2024, this robustness check does not include city-level controls but instead adds city fixed effects to alleviate potential omitted variable bias at the city level. In addition, we control for industry × city fixed effects and clustering at the province-year level; both approaches yield significant and positive results, further confirming the robustness of our findings and indicating that over time, the development of computing power under the "East data, West computing" project significantly promotes the digital transformation of enterprises.

Table 8
Excluding other policy influences.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*
EDWC	1.874**	1.879**	1.689**	1.509**	1.832**	1.893**	1.901**	1.413**
	(2.43)	(2.44)	(2.18)	(2.10)	(2.39)	(2.46)	(2.48)	(1.98)
KDChina	1.670***							1.154**
	(2.98)							(2.13)
Wisecity		1.210*						0.449
		(1.88)						(0.68)
Bigdata			1.118					0.517
			(1.28)					(0.58)
AIarea				2.159***				1.888***
				(4.31)				(3.80)
Infor					1.985***			1.272*
					(2.74)			(1.77)
Threeweb						0.180		−0.439
						(0.36)		(−0.96)
DATAgov							0.526	0.379
							(1.12)	(0.83)
Size	2.110***	2.113***	2.100***	2.105***	2.109***	2.121***	2.123***	2.080***
	(8.21)	(8.23)	(8.28)	(8.33)	(8.30)	(8.33)	(8.35)	(8.20)
Lev	−0.283	−0.298	−0.322	−0.238	−0.397	−0.274	−0.259	−0.350
	(−0.20)	(−0.21)	(−0.23)	(−0.17)	(−0.28)	(−0.19)	(−0.18)	(−0.25)
Roa	−11.976***	−12.046***	−11.944***	−11.759***	−12.014***	−11.994***	−11.948***	−11.755***
	(−6.28)	(−6.28)	(−6.20)	(−6.17)	(−6.25)	(−6.28)	(−6.28)	(−6.15)
Cash	−1.974*	−2.028*	−2.012*	−2.065*	−1.919	−2.005*	−2.016*	−1.998
	(−1.66)	(−1.70)	(−1.67)	(−1.70)	(−1.62)	(−1.68)	(−1.69)	(−1.64)
Capital	0.291	0.289	0.298	0.290	0.305	0.288	0.287	0.308
	(1.12)	(1.11)	(1.14)	(1.11)	(1.17)	(1.11)	(1.10)	(1.19)
Top3	0.294	0.302	0.304	0.330	0.329	0.293	0.292	0.358
	(0.84)	(0.87)	(0.86)	(0.94)	(0.95)	(0.84)	(0.84)	(1.01)
Age	−0.374	−0.343	−0.338	−0.369	−0.316	−0.345	−0.349	−0.366
	(−0.61)	(−0.56)	(−0.56)	(−0.61)	(−0.52)	(−0.56)	(−0.57)	(−0.61)
Soe	1.682	1.649	1.647	1.692	1.635	1.648	1.642	1.698
	(1.62)	(1.58)	(1.58)	(1.63)	(1.56)	(1.58)	(1.58)	(1.64)
Gdp	0.685	0.925	1.070	0.895	0.474	1.138	1.154	0.063
	(0.65)	(0.88)	(1.03)	(0.91)	(0.45)	(1.02)	(1.09)	(0.07)
Pop	3.677***	3.880***	3.588***	3.146***	3.404***	3.961***	3.939***	2.472***
	(4.62)	(4.62)	(4.16)	(4.06)	(4.03)	(4.52)	(4.46)	(2.90)
Fiscal	−0.070	−0.091	−0.124	0.071	0.020	−0.124	−0.106	0.203
	(−0.16)	(−0.21)	(−0.28)	(0.16)	(0.05)	(−0.29)	(−0.25)	(0.45)
Firm	−2.306**	−2.510**	−2.477***	−2.170***	−2.103**	−2.664**	−2.648***	−1.459*
	(−2.37)	(−2.52)	(−2.70)	(−2.61)	(−2.25)	(−2.57)	(−2.67)	(−1.97)
GDP2	4.056	0.296	1.167	0.301	5.120	0.104	0.084	6.913
	(0.46)	(0.03)	(0.13)	(0.03)	(0.54)	(0.01)	(0.01)	(0.75)
GDP3	0.367	−3.310	−1.660	−3.752	1.922	−3.760	−3.440	4.252
	(0.04)	(−0.32)	(−0.17)	(−0.38)	(0.17)	(−0.35)	(−0.32)	(0.42)
Constant	−57.401***	−56.191***	−56.647***	−53.170***	−58.533***	−56.382***	−56.770***	−56.037***
	(−4.38)	(−4.23)	(−4.36)	(−4.09)	(−4.48)	(−4.21)	(−4.21)	(−4.48)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	38,260	38,260	38,260	38,260	38,260	38,260	38,260	38,260
Adj.R^2	0.703	0.703	0.703	0.704	0.703	0.703	0.703	0.704

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

Table 9
Robustness test at the sample level.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*
	Sample period: 2009–2023	Sample period: 2016–2023	Sample period: 2005–2024	Municipalities were excluded	Guizhou Province were excluded	IT industry were excluded	PSM-DID
EDWC	1.825**	1.457**	3.089***	1.496**	1.878**	1.027**	1.940***
	(2.44)	(1.99)	(3.13)	(2.01)	(2.42)	(2.00)	(2.63)
Size	2.468***	1.004*	2.902***	2.054***	2.108***	1.623***	2.188***
	(9.21)	(1.76)	(8.79)	(6.27)	(8.21)	(9.74)	(7.96)
Lev	−0.089	−1.948	−0.412	−0.419	−0.193	0.433	−0.641
	(−0.05)	(−1.36)	(−0.28)	(−0.23)	(−0.14)	(0.41)	(−0.51)
Roa	−12.637***	−7.499***	−12.918***	−12.367***	−11.837***	−6.325***	−13.094***
	(−6.13)	(−3.95)	(−6.96)	(−5.00)	(−6.17)	(−3.30)	(−6.23)
Cash	−1.881	−3.181	−2.781***	−1.876*	−1.988*	−0.152	−1.831
	(−1.52)	(−1.61)	(−2.70)	(−1.73)	(−1.67)	(−0.19)	(−1.43)
Capital	0.458	1.323**	−0.078	−0.034	0.292	−0.022	0.185
	(1.62)	(2.43)	(−0.34)	(−0.13)	(1.12)	(−0.18)	(0.74)
Top3	0.174	−0.138	0.503*	0.134	0.305	0.227	0.441
	(0.46)	(−0.34)	(1.74)	(0.37)	(0.87)	(0.95)	(1.33)
Age	−0.678	−1.040	−0.195	0.381	−0.352	−0.542	−0.178
	(−0.83)	(−0.97)	(−0.54)	(0.52)	(−0.58)	(−1.24)	(−0.37)
Soe	1.859	0.681	−0.445	1.390	1.671	0.817	1.644
	(1.58)	(0.73)	(−0.42)	(1.15)	(1.59)	(0.93)	(1.50)
Gdp	1.659	−1.603		1.640	1.062	−0.548	0.979
	(1.27)	(−1.23)		(1.07)	(0.96)	(−0.80)	(0.89)
Pop	4.230***	3.323***		4.703***	3.991***	2.719***	3.775***
	(4.18)	(4.25)		(4.76)	(4.54)	(4.14)	(3.94)
Fiscal	−0.114	0.513		0.036	−0.149	−0.122	−0.094
	(−0.25)	(1.03)		(0.08)	(−0.34)	(−0.33)	(−0.20)
Firm	−2.994**	−0.745		−2.112*	−2.654**	−0.241	−2.375***
	(−2.34)	(−0.50)		(−1.74)	(−2.53)	(−0.53)	(−2.86)
GDP2	−8.973	20.293		−0.990	−0.505	1.389	−0.443
	(−0.72)	(1.35)		(−0.11)	(−0.05)	(0.19)	(−0.05)
GDP3	−14.006	22.415		2.732	−3.171	−0.301	−4.122
	(−1.01)	(1.32)		(0.29)	(−0.29)	(−0.04)	(−0.39)
Constant	−53.477***	−37.741	−56.971***	−74.666***	−55.790***	−36.178***	−60.291***
	(−2.86)	(−1.18)	(−3.92)	(−4.30)	(−4.02)	(−3.20)	(−4.60)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	No	No	No
N	34,959	24,207	53,706	29,546	37,922	35,188	32,236
Adj.R^2	0.711	0.788	0.693	0.686	0.704	0.664	0.701

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

(3) Excluding Guizhou province. Guizhou province has been at the forefront of China's digital development trends and had a strong digital foundation even before the implementation of the "East data, West computing" project, thus exerting considerable influence. To avoid any potential interference with the accuracy of our baseline results, we re-estimate the model after excluding samples from Guizhou province. The results are shown in Table 9, column (5), where the coefficient of *EDWC* remains significant and positive, indicating that our conclusions still hold.
(4) Excluding the information technology (IT) industry. Compared with non-digital native firms, IT firms are digital natives and have inherent advantages in digital transformation. Moreover, such firms may exaggerate their digital activities in their annual reports. Therefore, including these firms could affect the accuracy of our baseline results. Accordingly, we rerun the regression after excluding firms in the IT industry.[7] The results are presented in Table 9, column (6), where the coefficient of *EDWC* remains significant and positive, further demonstrating the robustness of our findings.
(5) Propensity score matching combined with difference-in-differences (PSM-DID). The establishment of national computing power hub nodes under the "East data, West computing" project is not random, and individual characteristics between the treatment and control groups may differ before and after the implementation of the policy, potentially leading to sample selection bias. To mitigate this concern, we use the PSM-DID approach as a robustness check. Specifically, we use the nearest neighbor matching method, taking firm-level control variables as covariates, to perform 1:4 matching, and then we rerun the regression on the matched sample. The results are presented in Table 9, column (7), where the coefficient of *EDWC* remains significant and positive, indicating that our conclusions still hold.

4.4.5. Robustness test at the variable level

At the variable level, we conduct a series of additional robustness checks to enhance the robustness of our findings, as detailed below.

(1) Replacing the measurement method. In the baseline regression, we measure the degree of digital transformation of enterprises by the frequency of digital-related keywords in the MD&A section of corporate annual reports. To mitigate estimation bias due to unstandardized measurement of variables, we conduct a robustness check using the variable "frequency of digital keywords / total word count of the MD&A section." The results are presented in Table 10, column (1), where the coefficient of *EDWC* remains significant and positive, indicating that our conclusions still hold.
(2) Replacing the dependent variable. First, following Müller et al. (2018), we measure the level of digital transformation of enterprises by the ratio of digital intangible assets to net intangible assets. Specifically, we screen the notes on intangible assets in the annual reports of listed firms using digital-related keywords to identify digital intangible asset items, then we calculate the ratio of these items to net intangible assets (*Intangasset*). Second, following Tan et al. (2025), we measure the degree of digital transformation of enterprises by the demand for labor in digital technologies. Specifically, we obtain from the CSMAR database the total number of digital job offers published by listed firms on recruitment websites from 2021 to 2023[8] and the total number of employees required in these offers.[9] We then aggregate these figures at the firm-year level and take their natural logarithms (*Num_Diglabor1*, *Num_Diglabor2*). Finally, we replace the original dependent variable with these variables in the regression. The results are shown in

[7] The IT industry includes telecommunications, broadcasting, and satellite transmission services (I63); internet and related services (I64); and software and information technology services (I65).
[8] Job function categories are limited to computer/internet/communication/electronics, software development, artificial intelligence and data engineer.
[9] When the required number of employees is missing or stated as "several," we assign a value of 1 for missing and 5 for "several."

Table 10
Robustness test at the variable level.

Variables	(1)	(2)	(3)	(4)	(5)
	MA_DIG1	*Intangasset*	*Num_Diglabor1*	*Num_Diglabor2*	*MA_DIG*
	Replacing the measurement method	Replacing the dependent variable			Adding control variables
EDWC	0.431**	0.034**	0.401***	0.399**	1.733**
	(2.10)	(2.49)	(2.92)	(2.52)	(2.22)
Size	1.030***	0.003	0.474***	0.558***	2.114***
	(6.32)	(0.45)	(3.40)	(3.72)	(8.19)
Lev	0.668	0.034**	−0.430	−0.585*	−0.213
	(0.85)	(2.16)	(−1.47)	(−1.80)	(−0.15)
Roa	−4.767***	−0.023	−1.279***	−1.463***	−11.862***
	(−4.72)	(−0.68)	(−3.05)	(−3.13)	(−6.28)
Cash	−0.181	−0.000	0.300	0.326	−1.982*
	(−0.29)	(−0.01)	(0.98)	(0.84)	(−1.66)
Capital	0.084	−0.034***	0.245***	0.281***	0.270
	(1.03)	(−5.63)	(5.77)	(5.96)	(1.03)
Top3	0.058	0.011***	0.020	0.013	0.275
	(0.38)	(2.60)	(0.36)	(0.25)	(0.79)
Age	−0.375	−0.009	−0.101	−0.134	−0.359
	(−1.47)	(−1.10)	(−1.02)	(−1.07)	(−0.60)
Soe	0.512	0.009	0.212	0.277	1.665
	(0.94)	(0.81)	(1.26)	(1.25)	(1.60)
Gdp	1.531**	−0.002	0.012	0.023	−0.066
	(2.18)	(−0.12)	(0.03)	(0.06)	(−0.05)
Pop	1.254***	0.037**	0.303	0.296	3.483***
	(2.78)	(2.49)	(1.13)	(0.92)	(4.11)
Fiscal	−0.117	−0.003	−0.363**	−0.478***	−0.398
	(−0.46)	(−0.61)	(−2.50)	(−2.83)	(−0.92)
Firm	−1.731***	−0.005	−0.042	−0.060	−2.980***
	(−2.94)	(−0.35)	(−0.15)	(−0.18)	(−3.11)
GDP2	−16.569**	−0.035	0.290	−0.187	0.531
	(−2.46)	(−0.24)	(0.08)	(−0.04)	(0.06)
GDP3	−18.854**	0.063	1.638	1.636	−4.512
	(−2.38)	(0.37)	(0.41)	(0.35)	(−0.43)
Tech_edu					1.568
					(1.51)
Edu					1.322
					(1.16)
Wage					2.654**
					(2.35)
Constant	−5.443	0.522*	−15.044***	−16.410***	−87.995***
	(−0.67)	(1.87)	(−2.90)	(−2.61)	(−4.82)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	32,915	37,729	11,209	11,209	38,260
Adj.R^2	0.745	0.599	0.709	0.686	0.703

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

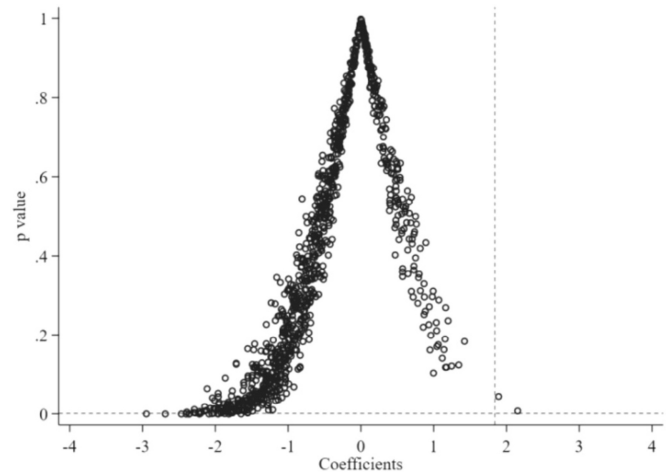


Fig. 4. Pseudo-treatment group variables estimated 1000 times.

Table 10, columns (2) to (4). The coefficient of *EDWC* is significant and positive, demonstrating that whether our measurement is based on text, financial indicators or governance-level variables,[10] the conclusion that the computing power project promotes the digital transformation of local enterprises still holds.

(3) Adding control variables. We add city-level control variables that may influence the "East data, West computing" project, including urban science and technology expenditure (*Tech_edu*), urban education expenditure (*Edu*) and labor cost (*Wage*). Urban science and technology expenditure is defined as the sum of local fiscal expenditures on science and technology and on education, divided by local budgetary fiscal revenues. Urban education expenditure is the natural logarithm of local fiscal expenditures on education. Labor cost is the natural logarithm of the average wage of urban workers. After including these control variables, the regression results are presented in Table 10, column (5). The coefficient of *EDWC* remains significant and positive, indicating that our conclusions remain robust.

4.4.6. Placebo test

We also conduct a placebo test as a robustness check. Specifically, based on the actual implementation of the "East data, West computing" project, we randomly generate 1000 treatment group dummy variables, while keeping the policy time from 2022. We then interact each randomly generated pseudo-treatment variable with the policy time dummy to obtain a pseudo-*EDWC* variable, replace the original explanatory variable in the model with this pseudo-variable and rerun the regression while keeping its coefficient. Repeating this procedure 1000 times yields the density distribution of p-values of the pseudo-*EDWC* coefficients for the effect on the digital transformation of enterprises, as shown in Fig. 4. The vertical line in the figure represents the true coefficient of 1.893. It can be observed that the coefficients of the randomly generated pseudo-*EDWC* variables are distributed around 0 and are all smaller than the baseline regression estimate, indicating that our results are not due to chance. The improvement of the digital transformation of enterprises is indeed driven by the establishment of computing power hub nodes under the "East data, West computing" project.

[10] One potential concern with using the demand for labor in digital technologies to measure digital transformation is that firms in more economically developed regions may naturally have greater recruitment needs, both in terms of the number of job offers and the number of hires. To alleviate this concern, following the same approach as in the robustness checks above, we rerun the regression after excluding the four municipalities, and the results remain significant, confirming the robustness of our findings. The reduction in sample size in columns (3) and (4) is due to the fact that the sample period is limited to 2021–2023.

Table 11
Other robustness tests.

Variables	(1)	(2)
	MA_DIG	*MA_DIG*
	High-dimensional fixed effects	Two-way clustering at the city-industry level
EDWC	1.681**	1.893***
	(2.34)	(2.88)
Size	1.986***	2.120***
	(7.58)	(4.69)
Lev	0.370	−0.265
	(0.31)	(−0.20)
Roa	−9.701***	−11.986***
	(−5.67)	(−2.59)
Cash	−2.178**	−2.010
	(−2.14)	(−1.24)
Capital	0.635**	0.287
	(2.59)	(0.67)
Top3	0.153	0.292
	(0.47)	(0.79)
Age	−0.610	−0.348
	(−0.98)	(−0.53)
Soe	1.759*	1.647
	(1.76)	(1.50)
Gdp	5.682***	1.212
	(3.48)	(0.78)
Pop	5.684***	3.976***
	(5.81)	(4.26)
Fiscal	0.003	−0.131
	(0.01)	(−0.26)
Firm	−3.919***	−2.700
	(−3.20)	(−1.23)
GDP2	−11.943	−0.331
	(−1.58)	(−0.04)
GDP3	−12.820	−4.197
	(−1.29)	(−0.86)
Constant	−86.113***	−56.228***
	(−4.90)	(−2.64)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Industry × City FE	Yes	No
N	38,215	38,260
Adj.R^2	0.719	0.703

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

4.4.7. Other robustness tests

Based on the above robustness checks, we conduct additional robustness tests using high-dimensional fixed effects and two-way clustering at the city-industry level. First, in addition to firm and year fixed effects, we include industry × city fixed effects to control for potential time-invariant confounding factors that are jointly determined by the industry and city dimensions. The regression results are presented in Table 11, column (1), where the coefficient of *EDWC* remains significant and positive, indicating that our conclusions remain robust. Second, we replace the original cluster-robust standard errors at the city level with two-way cluster-robust

standard errors at the city-industry level. This clustering approach simultaneously accounts for city and industry dimensions, more comprehensively capturing the complex dependence structure of the error term and providing more reliable standard error estimates.[11] The regression results are presented in Table 11, column (2), where the coefficient of *EDWC* remains significant and positive at the 1% level, confirming the robustness of our findings.

5. Heterogeneity analysis

5.1. The enterprise level

5.1.1. Enterprise ownership

In China, compared with non-SOEs, SOEs serve as the "national team" for policy implementation and naturally bear the responsibility of executing national strategies. To examine whether the promoting effect of the "East data, West computing" project on the digital transformation of enterprises is more pronounced in those with a particular ownership type, we divide the sample into SOEs and non-SOEs based on ownership and run separate regressions. The results are presented in Table 12, columns (1) and (2). The coefficient of the project on digital transformation is significant and positive for SOEs (column (1)) but not significant for non-SOEs, indicating that the project's promoting effect is stronger in SOEs. This finding can be explained as follows: (1) As a major national strategic project, SOEs are more inclined to incorporate the "East data, West computing" project into their long-term development plans and allocate resources preferentially to facilitate its implementation. (2) The performance evaluation of SOEs is often linked to policy implementation, such as the "dual-carbon" goals and digital economy development targets. These performance pressures tend to push SOEs to accelerate their digital transformation.

5.1.2. Capital intensity of enterprises

In practice, industries such as science, government affairs and finance have a high demand for computing power. For these demand-intensive industries, low-cost, high-quality and easy-to-use computing power supply services should be actively developed. Thus, firms in different industries have heterogeneous computing power needs, and identifying the types of firms that most need computing power resources to support their digital transformation is an important question that warrants in-depth investigation. To address this, following Derashid and Zhang (2003), we divide firms into two groups, asset-heavy and asset-light firms, based on the median ratio of fixed assets to total assets, and run separate regressions. The results are presented in Table 12, columns (3) and (4). They show that the promoting effect of the "East data, West computing" project on the digital transformation of enterprises is more pronounced in asset-light firms. This finding can be explained as follows: (1) Firms with low capital intensity, i.e., asset-light firms, typically lack their own computing power infrastructure and rely more on external cloud services, computer equipment or third-party computing support. The computing power hub nodes of the project directly lower the threshold for digital transformation for such firms by providing low-cost, universal computing power resources. (2) Compared with asset-heavy firms, asset-light firms have more flexible organizational structures and a faster speed of technology iteration, enabling them to absorb the computing power dividends generated by the project more quickly and accelerate their digital transformation.

5.2. The regional level

5.2.1. Electricity resources

Computing power hub clusters and data center clusters often require a large number of computer devices and servers to provide powerful computing capacity, and the high concentration of these servers consumes substantial electricity resources (Yang et al., 2023). The availability of electricity resources plays an important

[11] In addition, given that the national computing power hub pilot under the "East data, West computing" project is at the provincial level, to enhance the robustness of our findings, we use cluster-robust standard errors at the provincial level, and the conclusion remains robust.

Table 12
Ownership and firm capital intensity.

Variables	(1)	(2)	(3)	(4)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*
	SOEs	Non-SOEs	Asset-heavy enterprises	Asset-light enterprises
EDWC	1.569**	1.369	0.975	2.530***
	(2.38)	(1.11)	(1.12)	(2.92)
Size	1.166***	2.624***	3.413***	2.265***
	(2.75)	(6.41)	(6.03)	(6.54)
Lev	0.104	−1.500	−0.195	0.133
	(0.08)	(−0.90)	(−0.15)	(0.05)
Roa	−4.930	−9.987***	−9.623***	−13.371***
	(−0.90)	(−5.00)	(−3.89)	(−4.26)
Cash	−1.391	−2.236	0.078	−2.381
	(−1.26)	(−1.35)	(0.04)	(−1.47)
Capital	−0.079	0.374	−1.186**	0.141
	(−0.26)	(1.60)	(−2.59)	(0.39)
Top3	0.520	−0.840*	0.527	−0.152
	(1.33)	(−1.85)	(1.21)	(−0.32)
Age	0.294	−1.025	−1.285**	−0.437
	(1.24)	(−0.86)	(−2.59)	(−0.54)
Soe	/	/	1.051	2.469*
	/	/	(0.87)	(1.75)
Gdp	0.509	3.864	0.835	1.735
	(0.45)	(1.43)	(0.70)	(1.33)
Pop	2.005***	5.205***	3.659***	3.586***
	(2.73)	(3.36)	(3.34)	(3.65)
Fiscal	0.179	−0.155	−0.575	0.733
	(0.42)	(−0.25)	(−1.54)	(1.23)
Firm	−0.899**	−5.186*	−2.419**	−2.462**
	(−2.12)	(−1.70)	(−2.31)	(−2.24)
GDP2	1.079	−19.276	−3.517	−1.430
	(0.12)	(−1.31)	(−0.34)	(−0.10)
GDP3	4.165	−28.583	−7.112	−5.577
	(0.40)	(−1.54)	(−0.62)	(−0.37)
Constant	−46.627***	−27.443	−35.670**	−51.612***
	(−3.21)	(−0.93)	(−2.21)	(−2.86)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	14,603	23,602	18,391	19,130
Adj.R^2	0.736	0.716	0.748	0.706

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

role in the layout and functioning of computing power hubs. To examine whether the availability of electricity resources exerts a heterogeneous promoting effect on the digital transformation of enterprises, we obtain provincial-level electricity generation data from the CSMAR database and divide the provinces into high electricity generation and low electricity generation groups (i.e., provinces with more and less electricity resources) based on the annual median, and then run subgroup regressions. The results are presented in Table 13, columns (1) and (2). They show that the promoting effect of the ‘‘East data, West computing” project on the digital transformation of enterprises is more pronounced in provinces with high electricity generation. This finding can be explained as follows: Provinces with higher electricity generation have more electricity resources, which better meets the project’s electricity demand and creates favorable preconditions for its effective functioning. This not only facilitates the conversion of electricity resources into computing power

Table 13
Electricity resource endowment.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*
	High electricity generation provinces	Low electricity generation provinces	Central-Western region		Eastern Region	
			High electricity generation provinces	Low electricity generation provinces	High electricity generation provinces	Low electricity generation provinces
EDWC	2.950***	1.343	1.175	0.136	2.898***	−0.272
	(3.38)	(1.25)	(0.99)	(0.09)	(3.06)	(−0.15)
Size	2.043***	2.200***	1.417***	2.186***	2.132***	2.560***
	(4.95)	(6.56)	(2.90)	(4.30)	(4.61)	(5.86)
Lev	1.115	−1.323	3.132**	−1.378	0.492	−1.208
	(0.41)	(−1.12)	(2.41)	(−0.84)	(0.16)	(−0.69)
Roa	−13.383***	−9.028***	−10.236*	−7.992**	−13.754***	−8.618***
	(−4.13)	(−5.72)	(−1.94)	(−2.23)	(−3.77)	(−6.35)
Cash	−0.099	−3.550*	1.154	−2.673	−0.210	−3.437
	(−0.10)	(−1.88)	(0.35)	(−1.47)	(−0.22)	(−1.21)
Capital	0.058	0.653*	0.128	−0.164	0.062	1.065**
	(0.21)	(1.73)	(0.31)	(−0.32)	(0.21)	(2.44)
Top3	0.079	0.333	1.422	0.144	−0.143	0.524
	(0.19)	(0.66)	(1.37)	(0.44)	(−0.33)	(0.54)
Age	0.933	−0.825**	0.782	0.184	1.067	−1.059**
	(0.91)	(−2.02)	(0.54)	(0.37)	(0.94)	(−2.36)
Soe	1.941	1.597	−1.841	−0.544	2.590	3.448
	(1.08)	(1.22)	(−0.85)	(−0.34)	(1.24)	(1.60)
Gdp	−0.034	1.951	−1.696	1.324	2.187	1.441
	(−0.02)	(1.57)	(−0.66)	(0.95)	(1.26)	(0.49)
Pop	5.294***	2.912***	−0.182	3.791**	4.900***	2.097
	(5.35)	(2.69)	(−0.06)	(2.16)	(3.72)	(1.67)
Fiscal	0.396	0.225	−0.531	0.300	1.121	−1.768
	(0.43)	(0.47)	(−0.66)	(0.74)	(0.90)	(−0.73)
Firm	−0.765	−3.764***	−0.378	−2.554**	−0.917	−4.549*
	(−0.61)	(−4.18)	(−0.36)	(−2.53)	(−0.61)	(−1.88)
GDP2	13.945	−3.141	46.209**	1.586	−11.196	11.124
	(0.77)	(−0.35)	(2.05)	(0.18)	(−0.55)	(0.20)
GDP3	23.373	−15.206	58.482**	−4.089	1.074	−2.841
	(1.27)	(−1.55)	(2.32)	(−0.49)	(0.05)	(−0.05)
Constant	−102.605***	−41.880**	−93.798**	−60.257***	−96.893***	−48.771
	(−3.85)	(−2.57)	(−2.41)	(−2.64)	(−3.29)	(−0.87)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	16,761	21,378	2184	8868	14,576	12,508
Adj.R^2	0.699	0.729	0.717	0.678	0.698	0.737

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

resources but also reduces the cost of using computing power, thus enabling firms to better utilize computing power resources and achieve their digital transformation.

To better understand how the heterogeneity of electricity resources in different geographical locations affects the project's promotion of the digital transformation of enterprises, we distinguish between the Eastern region and the Central-Western region. Given that the Eastern region is mostly a source area for the relocation of computing power demand, while the Central-Western region is mostly a destination area, despite its greater electricity resources than the Eastern region, we expect that the promoting effect of the project on the digital transformation of enterprises will be more significant in high electricity generation provinces in the Central-

Western region. The results of these subgroup regressions are shown in Table 13, columns (3) to (6). Contrary to our expectations, the coefficients for the Central-Western region are positive but not significant, while those for the high electricity generation provinces in the Eastern region are significant and positive. This implies that the promoting effect of the project on the digital transformation of enterprises is more pronounced in the provinces with high electricity generation in the Eastern region. A possible explanation is that these provinces not only have abundant electricity resources but also more mature digital infrastructure systems, such as data center clusters and 5G base stations, which can quickly convert electricity advantages into computing power advantages for local enterprises. In contrast, although the Central-Western region has a surplus of electricity, its grid transmission and distribution capacity is low and the surrounding digital infrastructure systems are still under construction, which constrains the conversion of electricity into computing power.

5.2.2. Regional energy pressure and data transmission conditions

The previous analysis demonstrates that the availability of electricity resources is an important factor influencing the layout of computing power projects and their impact on the digital transformation of enterprises. However, unlike the perspective of electricity resource supply, the pressure on regional electricity consumption adopts a resource demand perspective, focusing on the intensity of electricity demand in a region. How the "East data, West computing" project affects the digital transformation of enterprises under heterogeneous energy consumption pressures also warrants discussion. To this end, following Wang et al. (2023a, 2023b), we measure provincial electricity energy consumption pressure as (provincial electricity generation – provincial electricity consumption)/provincial electricity generation. A higher value indicates a smaller electricity supply–demand gap and thus less pressure on electricity consumption, and vice versa. We then divide provinces into high and low energy pressure groups based on the annual median and run subgroup regressions. The results are presented in Table 14, columns (1) and (2). We find that in provinces with higher energy consumption pressure, the "East data, West computing" project has a stronger promoting effect on the digital transformation of local enterprises. This finding can be explained as follows: (1) Regions with high energy consumption pressure generally face the dual constraint of electricity shortages and carbon emissions, which forces the digital transformation of enterprises. The optimization and upgrading of digital transformation facilitate the automation and intelligence of production processes, enabling firms to achieve efficient production, energy savings and reduced emissions through digital technologies. (2) By building a regional computing power network, the project transfers the computing demands of high energy pressure regions to areas with abundant computing power supply, achieving regional coordination of computing power and energy consumption. Given the reality of electricity shortages in high energy pressure regions, firms could take advantage of the project to relocate their computing power needs to other regions with lower energy consumption. Through the regional scheduling of computing power resources, enterprises can access lower-cost computing services while pursuing green and sustainable development, thus facilitating and accelerating their digital transformation.

The conclusion that computing power demand can be transferred, as argued above, requires verification of the regional scheduling role of data transmission media. Following the literature (Xu et al., 2025), we measure the actual conditions of data transmission using the length of long-distance optical cable lines. A longer optical cable line implies better data transmission services, lower costs, less data loss and better accessibility in the regional scheduling process for computing power demand. To test this, we divide the provinces into two groups (better vs. worse data transmission conditions) based on the annual median length of regional optical cable lines and run subgroup regressions. The results are presented in Table 14, columns (3) and (4). We find that in regions with longer optical cable lines, i.e., better data transmission conditions, the promoting effect of the "East data, West computing" project on the digital transformation of enterprises is more pronounced. The scheduling of computing power in the digital transformation of enterprises largely relies on high-speed transmission media and a stable network environment. The project improves the data transmission infrastructure, such as optical cables, by providing high-speed and stable transmission channels enabling enterprises to access computing power resources, thus promoting their digital transformation. Furthermore, this finding confirms our previous result that the project has a stronger promoting effect on the digital transformation of enterprises in provinces with high energy consumption pressure.

Table 14
Regional energy pressure and data transmission conditions.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	MA_DIG	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*	*MA_DIG*
	Low energy pressure provinces	High energy pressure provinces	Better data transmission conditions	Poorer data transmission conditions	provinces with both high energy pressure and good data transmission conditions	The sum of the other three cases outside column (5)
EDWC	0.653	6.369***	1.633**	1.266	7.920***	1.373
	(0.84)	(3.61)	(2.17)	(1.05)	(4.18)	(1.51)
Size	2.490***	2.027***	1.988***	2.102***	2.001***	2.195***
	(6.35)	(5.07)	(5.19)	(5.81)	(3.48)	(8.68)
Lev	−0.592	0.241	1.718	−1.925	1.867	−0.672
	(−0.49)	(0.10)	(0.74)	(−1.50)	(0.34)	(−0.72)
Roa	−11.192***	−11.711***	−11.624***	−10.404***	−11.995**	−11.776***
	(−4.11)	(−4.80)	(−3.88)	(−4.23)	(−2.57)	(−5.69)
Cash	−1.712	−2.001	−0.356	−3.320*	−1.075	−2.431*
	(−1.17)	(−1.18)	(−0.33)	(−1.76)	(−0.82)	(−1.67)
Capital	−0.511*	0.818***	−0.087	0.723**	0.463*	0.229
	(−1.74)	(2.77)	(−0.26)	(2.10)	(1.77)	(0.66)
Top3	0.256	0.260	0.484	0.371	−0.354	0.462
	(0.62)	(0.45)	(1.04)	(0.65)	(−0.56)	(1.21)
Age	−0.210	−0.461	0.664	−1.058***	1.796*	−0.799**
	(−0.39)	(−0.51)	(0.70)	(−2.71)	(1.75)	(−2.12)
Soe	0.033	3.703**	1.201	1.958	4.805	0.879
	(0.03)	(2.10)	(0.72)	(1.45)	(1.28)	(0.85)
Gdp	3.113*	0.527	0.119	2.693	5.649**	2.341**
	(1.94)	(0.28)	(0.10)	(1.38)	(2.42)	(2.05)
Pop	2.305**	6.238***	4.847***	4.770	3.893	3.010***
	(2.39)	(3.24)	(6.34)	(1.60)	(1.52)	(2.93)
Fiscal	0.121	−0.097	0.278	−0.500	1.601	−0.253
	(0.25)	(−0.08)	(0.47)	(−0.77)	(1.40)	(−0.57)
Firm	−1.985**	−3.903***	−0.986	−4.212***	−3.175*	−3.758***
	(−1.98)	(−2.73)	(−1.06)	(−4.53)	(−1.71)	(−4.02)
GDP2	−2.417	40.054	−0.471	−6.289	2.014	−0.583
	(−0.25)	(1.44)	(−0.05)	(−0.33)	(0.08)	(−0.06)
GDP3	−2.019	36.194	0.997	−11.770	2.876	−3.433
	(−0.20)	(1.29)	(0.09)	(−0.53)	(0.10)	(−0.28)
Constant	−62.459***	−100.120***	−79.052***	−52.856**	−126.317***	−47.423***
	(−4.03)	(−4.11)	(−3.46)	(−2.08)	(−3.68)	(−3.73)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	17,865	20,233	18,541	19,478	7721	30,425
Adj.R^2	0.684	0.719	0.698	0.722	0.700	0.704

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

To further examine whether firms in high energy pressure regions achieve regional computing power scheduling through data transmission and thus promote their digital transformation, we divide the sample into two groups: provinces with both high energy consumption pressure and good data transmission conditions, and all other provinces (i.e., the remaining three scenarios). We then run separate regressions for each group. The results are presented in Table 14, columns (5) and (6). They show that in provinces with both high energy consumption pressure and better data transmission conditions, the promoting effect of the "East data, West computing" project on the digital transformation of enterprises is more significant, while the effect is not significant in other provinces. This further indicates that in provinces facing high energy consumption pressure and benefiting from better data transmission conditions, firms are more likely to transfer their computing power needs to regions with lower energy pressure and more abundant computing power resources, thus accelerating their digital transformation.

6. Further analysis

6.1. The high-quality economic development effect of the "East data, West computing" project

The previous results demonstrate that the computing power hub nodes of the "East data, West computing" project promote the digital transformation of local enterprises through three channels: optimization of the computing power hardware and software environment, strengthening of digital R&D and innovation and government digital subsidies and aid. A high-quality digital transformation, in turn, promotes high-quality enterprise development (Xiong et al., 2025). Therefore, from the perspective of firm-level TFP, we examine whether the improvement of digitalization through the project fosters high-quality enterprise development. Following the literature (Giannetti et al., 2015), we estimate firm-level TFP using the Cobb–Douglas production function, with the model specified as in Eq. (2):

$$Ln(Y_{it}) = \alpha_0 + \alpha_1 Ln(K_{it}) + \alpha_2 Ln(L_{it}) + \alpha_3 Ln(M_{it}) + \varepsilon_{it} \tag{2}$$

where Y denotes production, measured by firms' operating revenue; and K, L and M represent capital input, labor input and intermediate input, measured by net fixed assets, the number of employees and cash paid for goods and services, respectively. We estimate model (2) using ordinary least squares (OLS) and firm fixed effects (FE) methods and obtain the residual ε_{it} as an estimate of firm-level TFP, with a larger residual indicating higher TFP. We then divide the sample firms into two groups (high vs. low digital transformation) based on the annual mean of digital transformation, assigning them values of 1 and 0, respectively. We interact this digital transformation dummy variable (*DIG*) with *EDWC* and include the interaction term in the regression, where the dependent variable is firm-level TFP (*TFP_OLS* and *TFP_FE*). The regression results are presented in Table 15, columns (1) and (2).

The results show that regardless of the method used to estimate firm-level TFP, the coefficient of $EDWC \times DIG$ is positive and significant at the 1% level, indicating that the promoting effect of the "East data, West computing" project on the TFP of local firms is stronger for those with higher digitalization levels. This demonstrates that the project enhances the TFP of local enterprises by improving their digital transformation. This economic consequence provides empirical evidence that the computing power project enables firms to achieve high-quality development.

6.2. The cross-regional resource flow effect of the "East data, West computing" project

Current evaluations of the implementation effects of the "East data, West computing" project mostly focus on the actual scale of general computing, intelligent computing and supercomputing, the coverage and computational efficiency of the computing power network and the share of green electricity in computing power hubs. Few studies discuss its effects from the perspective of cross-regional resource flows. Through the project, besides seeking computing power or services from Western hub nodes, do enterprises in the Eastern region establish subsidiaries, invest in factories or undertake other activities in Western hub nodes or data center clusters to benefit from low-cost green computing services and policy incentives, thereby optimizing their geographic location and achieving high-quality development? To test and answer this question, we adopt the

Table 15
The high-quality economic development effect of the "East data, West computing" project.

Variables	(1)	(2)
	TFP_OLS	*TFP_FE*
	TFP calculated using the OLS method	TFP calculated using the FE method
EDWC	−0.011	−0.012
	(−1.01)	(−1.18)
DIG	−0.009	−0.006
	(−1.26)	(−0.90)
EDWC × DIG	0.024***	0.023***
	(3.15)	(3.18)
Size	0.165***	0.198***
	(16.65)	(20.86)
Lev	−0.009	0.025
	(−0.38)	(1.14)
Roa	0.937***	0.980***
	(18.04)	(18.40)
Cash	1.081***	1.063***
	(25.21)	(26.09)
Capital	−0.103***	−0.079***
	(−21.11)	(−16.16)
Top3	0.052***	0.055***
	(9.01)	(9.78)
Age	0.018	0.019*
	(1.62)	(1.71)
Soe	0.007	0.006
	(0.39)	(0.39)
Gdp	−0.012	−0.014
	(−0.61)	(−0.74)
Pop	−0.022	−0.019
	(−1.28)	(−1.15)
Fiscal	0.008	0.007
	(0.69)	(0.63)
Firm	0.026*	0.027*
	(1.77)	(1.82)
GDP2	0.798***	0.769***
	(3.15)	(3.13)
GDP3	0.650**	0.624**
	(2.48)	(2.46)
Constant	−3.401***	−4.662***
	(−9.59)	(−13.68)
Firm FE	Yes	Yes
Year FE	Yes	Yes
N	38,197	38,197
Adj.R^2	0.707	0.725

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

perspective of the cross-regional resource flows of the "East data, West computing" project and examine the effect of the project on firms' out-of-region investment.

Data center clusters are sites for aggregating computing power resources under the national computing power hub nodes. We set the destinations of out-of-region investments as the cities in the Western region where the data center clusters are located, including Chengdu, Chongqing, Guiyang, Anshun, Hohhot, Ulanqab, Zhongwei and Qingyang. Specifically, we first calculate the total number of subsidiaries in which listed firms invest in cities with Western data center clusters (*CWE*) through the "East data, West computing" project, then we take its natural logarithm and divide the sample into Eastern and Western region groups

for separate regressions. The regression results are presented in Table 16. Column (1) shows that the project significantly increases the number of subsidiaries in which local firms invest in cities with Western data center clusters. The results in columns (2) and (3) reveal that after the project's implementation, the number of subsidiaries in which listed firms in Eastern hub nodes invest in cities with Western data center clusters increases significantly, while the effect is not significant for the Western region. This finding can be explained as follows: (1) Push from the Eastern region: the Eastern hub nodes are often under strong energy consumption pressure, leading to substantially higher computing power costs. Some firms, seeking green and low-cost computing services, choose to deploy their subsidiaries in Western data center clusters. (2) Pull from the Western region: A core objective of the project is to relieve Eastern computing power demand and guide its migration to the resource-abundant Western region. The Western region offers more incentives such as investment promotion support, electricity price subsidies and land concessions, thus attracting Eastern firms to establish subsidiaries in cities with Western data center clusters. Overall, from the perspective of firms' out-of-region deployment of subsidiaries, these results not only show how Eastern firms within the project use the deployment of computing power to chart their transformation and development path but also demonstrate the cross-regional resource flow effect of the ''East data, West computing'' project.

7. Conclusion and discussion

In the era of the global digital economy, computing power is becoming a new productive force driving the digital transformation of enterprises and high-quality economic and social development. However, firms commonly face practical constraints such as insufficient transformation capabilities and high transformation costs. How to promote the digital transformation of enterprises through the implementation of external computing power projects remains an important topic that has received insufficient attention in the literature. In this context, we use China's ''East data, West computing'' project as our research setting and systematically investigate its impact, mechanisms, heterogeneity and economic consequences on the digital transformation of enterprises.

Using a sample of Chinese A-share non-financial companies listed on the SSE and SZSE from 2005 to 2023 and a DID approach, we find that the ''East data, West computing'' project significantly improves the digital transformation of enterprises within the computing power hub nodes. This finding confirms that the systematic layout of computing power infrastructures has a substantial promoting effect on the digital transformation of enterprises. Our mechanism analysis reveals that this facilitating effect primarily stems from three pathways: the optimization of the computing power hardware and software environment, the strengthening of digital R&D and innovation and government digital subsidies and aid. Our heterogeneity analysis shows that the effect is more pronounced for SOEs, asset-light firms and regions with better computing power infrastructure. Furthermore, we find that through the project, firms with a higher degree of digital transformation exhibit higher TFP; that enterprises located in regions with high energy consumption tend to relocate their computing power needs; and that firms in the Eastern hub nodes significantly increase their out-of-region investments in Western data center clusters. These findings indicate that the ''East data, West computing'' project not only directly promotes the development of local enterprises but also drives strategic adjustments in firms' layout and cross-regional resource reallocation.

Our paper makes three main contributions. First, by incorporating a national computing power project into the research framework on the driving factors of the digital transformation of enterprises, our study enriches the literature on the micro-level effects of major infrastructure strategies in the digital economy. Studies on the external drivers of the digital transformation of enterprises rarely adopt the perspective of major national strategic projects. We innovatively take China's ''East data, West computing'' project as an exogenous shock, providing quasi-natural experimental evidence of the causal relationship between this project and the digital transformation of enterprises. This finding not only extends the research boundary of the driving factors of the digital transformation of enterprises but also offers a new perspective for understanding the micro-economic consequences of computing power as a new factor of production.

Second, based on the resource-based view, we systematically reveal the internal mechanisms through which the computing power project affects the digital transformation of enterprises from three pathways: the optimization of the computing power environment, the strengthening of R&D and innovation and government

Table 16
The cross-regional resource flow effect of the "East data, West computing" project.

Variables	(1)	(2)	(3)
	CWE	*CWE*	*CWE*
	/	Out-of-region investment of enterprises in the eastern region	Out-of-region investment of enterprises in the western region
EDWC	0.034*	0.048*	0.009
	(1.95)	(1.67)	(0.15)
Size	0.145***	0.153***	0.139***
	(11.33)	(10.08)	(3.74)
Lev	0.044	0.032	0.134
	(0.87)	(0.46)	(1.49)
Roa	−0.215***	−0.195***	−0.381
	(−4.01)	(−3.27)	(−1.60)
Cash	0.088**	0.081	0.025
	(2.03)	(1.60)	(0.18)
Capital	0.008	0.009	−0.016
	(1.05)	(0.99)	(−0.75)
Top3	−0.008	0.004	−0.030
	(−0.60)	(0.26)	(−1.28)
Age	0.004	0.010	−0.048
	(0.30)	(0.70)	(−0.51)
Soe	0.036	0.029	0.024
	(1.27)	(0.90)	(0.27)
Gdp	0.034	0.062	−0.369***
	(0.51)	(0.94)	(−2.73)
Pop	0.015	0.036	−0.152
	(0.37)	(0.67)	(−1.30)
Fiscal	−0.012	−0.023	0.001
	(−0.69)	(−0.53)	(0.05)
Firm	−0.071	−0.101*	0.093
	(−1.50)	(−1.81)	(1.07)
GDP2	−0.818	−0.790	1.086
	(−1.41)	(−0.84)	(1.05)
GDP3	−0.887	−1.253	1.788
	(−1.46)	(−1.35)	(1.63)
Constant	−2.068***	−2.421**	0.808
	(−2.74)	(−2.27)	(0.35)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	35,098	24,963	4105
Adj.R^2	0.716	0.715	0.760

Note: The values in parentheses represent the t-values clustered at the city level. *, **, and *** indicate significance levels of 10%, 5%, and 1% respectively.

supporting policies. We not only verify the policy effect of the ''East data, West computing'' project but also deconstructs its mechanisms, specifically revealing how the development of computing power infrastructure improves the external conditions for the digital transformation of enterprises; how the spatial agglomeration of computing power enterprises and research institutions strengthens digital R&D and innovation; and how incentives from local governments alleviate the financing constraints associated with the digital transformation of enterprises. Together, these three mechanisms fully explain how computing power projects enable the digital transformation of enterprises, thus deepening the theoretical understanding of how major strategic projects enhance firms' competitive advantage.

Third, from a multidimensional perspective, this paper expands the scope of research on the economic and social effects of computing power projects. First, we find that the ''East data, West computing'' project not only generates a localized promoting effect on the digital transformation of enterprises but also encourages them, through the computing power network, to transfer their computing power needs outward and to engage in cross-regional investments. This finding reveals that the project can trigger adjustments in firm behavior and cross-regional resource reallocation, thus providing evidence to understand the spatial economic impacts of major national strategic projects. Second, the finding that firms located in regions with high energy consumption tend to relocate their computing power needs offers indirect evidence of the compatibility between computing power projects and green and low-carbon development objectives. Third, firms in computing power hub nodes with better digital transformation exhibit higher TFP, indicating that they leverage the unique resources provided by the computing power project to accelerate their digital transformation and achieve high-quality development. These findings not only expand the research boundaries of computing power projects but also provide a reference for the implementation and advancement of computing power strategic projects worldwide.

Based on the above findings, the following policy implications are drawn. First, we find that the ''East data, West computing'' project significantly promotes the digital transformation of local enterprises, thus validating the positive effect of improved external computing power environments and enhanced accessibility to computing resources on the digital transformation of enterprises. This facilitating effect is particularly pronounced in asset-light firms that rely heavily on computing power and in regions with better computing power infrastructure. These positive findings suggest that computing power projects and universal access to computing power should gradually become key considerations of digital economy policies worldwide. Second, we reveal that the relationship between the ''East data, West computing'' project and the digital transformation of enterprises is not a simple instrumental link, but rather a systematic empowerment formed through multiple mechanisms, including the optimization of the computing power environment, innovation synergy and policy support. This implies that the design of computing power projects must go beyond a single logic of infrastructure investment and shift to an integrated policy framework that encompasses a technological innovation ecosystem, fiscal incentive instruments and regional coordination mechanisms. Third, we find that the spatial optimization of the project's layout can guide firms to engage in cross-regional investments and resource reallocation, thus providing a practical Chinese reference for addressing the global challenge of balancing the growing demand for computing power with energy and environmental constraints.

This study also opens up several avenues for future research. First, due to limited accessibility of firm-level data on computing power usage, we are unable to directly observe firms' specific procurement behaviors and the intensity of their computing service usage. Future research could combine data from computing power trading platforms or firm-level computing power consumption indicators to further reveal micro-level details of how computing power resources enable the digital transformation of enterprises. Second, as the ''East data, West computing'' project is still in a phase of continuous development, its long-term dynamic effects and persistent impact on firms' organizational capabilities and innovation performance require monitoring and testing over a longer time window. Third, the introduction of a cross-country comparative perspective could help further identify the heterogeneity of the implementation effects of computing power infrastructure strategies in different institutional environments and industrial policies, which is also an important topic worthy of future attention.

Funding

This paper was supported by the General Project of the National Social Science Fund of China (No. 24BGL095) and the Fundamental Research Funds for the Central Universities (Graduate Research and Innovation Platform) Project of Zhongnan University of Economics and Law (No. 202611435).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

Acemoglu, D., Restrepo, P., 2020. Robots and jobs: evidence from US labor markets. J. Polit. Econ. 128 (6), 2188–2244. https://doi.org/10.1086/705716.

Acemoglu, D., Autor, D., Hazell, J., et al., 2022. Artificial intelligence and jobs: evidence from online vacancies. J. Labor Econ. 40 (S1), S293–S340. https://doi.org/10.1086/718327.

Barney, J., 1991. Firm resources and sustained competitive advantage. J. Manag. 17 (1), 99–120.

Chen, W., Srinivasan, S., 2024. Going digital: implications for firm value and performance. Rev. Account. Stud. 29 (2), 1619–1665. https://doi.org/10.1007/s11142-023-09753-0.

Deelman, E., Dongarra, J., Hendrickson, B., et al., 2025. High-performance computing at a crossroads. Science 387 (6736), 829–831. https://doi.org/10.1126/science.adu0801.

Demann, J., Doblinger, C., Nar, S., 2026. More money, more digital? The impact of state grants on digital transformation in small and medium-sized enterprises. J. Strateg. Inf. Syst. 35 (1), 101954. https://doi.org/10.1016/j.jsis.2026.101954.

Derashid, C., Zhang, H., 2003. Effective tax rates and the industrial policy hypothesis: evidence from Malaysia. J. Int. Account. Audit. Tax. 12 (1), 45–62. https://doi.org/10.1016/S1061-9518(03)00003-X.

Ding, Q., He, W., Deng, Y., 2025. Can tax reduction incentive policy promote corporate digital and intelligent transformation? Int. Rev. Financ. Anal. 99, 103932. https://doi.org/10.1016/j.irfa.2025.103932.

Firk, S., Gehrke, Y., Hanelt, A., et al., 2022. Top management team characteristics and digital innovation: exploring digital knowledge and TMT interfaces. Long Range Plan. 55 (3), 102166. https://doi.org/10.1016/j.lrp.2021.102166.

Giannetti, M., Liao, G., Yu, X., 2015. The brain gain of corporate boards: evidence from China. J. Finance. 70 (4), 1629–1682. https://doi.org/10.1111/jofi.12198.

Guo, J., Duan, J., 2024. The impact of MD&A digital transformation information disclosure on stock price synchronicity in China. N. Am. J. Econ. Finance. 74, 102253. https://doi.org/10.1016/j.najef.2024.102253.

Gurbaxani, V., Dunkle, D., 2019. Gearing up for successful digital transformation. MIS Q. Exec. 18 (3), 6. https://doi.org/10.17705/2msqe.00017.

Han, Z., Luo, E., Xiong, X., et al., 2025. Government innovation subsidies and enterprise digital transformation: evidence from Chinese listed companies. Technol. Soc., 103146 https://doi.org/10.1016/j.techsoc.2025.103146.

Hua, Z., Kong, S., Yu, Y., 2025. Talent magnets in China: can talent attraction policies catalyze digital transformation in enterprises? China Econ. Rev. 90, 102376. https://doi.org/10.1016/j.chieco.2025.102376.

Huang, B., Li, H., Liu, J., et al., 2023. Digital technology innovation and high-quality development of Chinese enterprises: evidence from corporate digital patents. Econ. Res. J. 58 (3), 97–115 (in Chinese).

Jiang, W., Jiang, N., Yu, D., 2026. Network infrastructure construction and digital innovation resilience: evidence from Broadband China policy. Inf. Process. Manag. 63 (6), 104758. https://doi.org/10.1016/j.ipm.2026.104758.

Jones, C.I., Tonetti, C., 2020. Nonrivalry and the economics of data. Am. Econ. Rev. 110 (9), 2819–2858. https://doi.org/10.1257/aer.20191330.

Ju, Z., Shao, J., 2026. Industry-university-research digital innovation collaboration and AI-driven green innovation: an innovation resource perspective. Int. Rev. Econ. Financ. 108, 105253. https://doi.org/10.1016/j.iref.2026.105253.

Liu, B., Li, Y., Liu, J., et al., 2024. Does urban innovation policy accelerate the digital transformation of enterprises? Evidence based on the innovative City pilot policy. China Econ. Rev. 85, 102167. https://doi.org/10.1016/j.chieco.2024.102167.

Liu, D., Chen, L., 2025. Supply chain innovation support policies, financing constraints, and corporate digital transformation. Int. Rev. Financ. Anal., 104639 https://doi.org/10.1016/j.irfa.2025.104639.

Liu, S., Li, Y., Pan, L., et al., 2025. Transformation strategy choice and digital transformation of non-digital-native enterprises: a perspective of organizational identity conflict. China Ind. Econ. 4, 137–155 (in Chinese).

Merin-Rodriganez, J., Dasi, A., Alegre, J., 2024. Digital transformation and firm performance in innovative SMEs: the mediating role of business model innovation. Technovation 134, 103027. https://doi.org/10.1016/j.technovation.2024.103027.

Mo, Y., Liu, X., 2023. Climate policy uncertainty and digital transformation of enterprise-evidence from China. Econ. Lett. 233, 111377. https://doi.org/10.1016/j.econlet.2023.111377.

Müller, O., Fay, M., Vom Brocke, J., 2018. The effect of big data and analytics on firm performance: an econometric analysis considering industry characteristics. J. Manag. Inf. Syst. 35 (2), 488–509. https://doi.org/10.1080/07421222.2018.1451955.

Nordhaus, W.D., 2001. The progress of computing. NBER Work. Pap.
Nwankpa, J.K., Roumani, Y., 2016. IT capability and digital transformation: a firm performance perspective.
Ritter, T., Pedersen, C.L., 2020. Digitization capability and the digitalization of business models in business-to-business firms: past, present, and future. Ind. Mark. Manag. 86, 180–190. https://doi.org/10.1016/j.indmarman.2019.11.019.
Shen, K., Qiao, G., Tan, R., 2024. Establishment of national big data comprehensive pilot zones and employment growth. China Ind. Econ. 12, 5–23 (in Chinese).
Sun, W., Mao, N., Lan, F., et al., 2023. Policy-empowered digital ecology and enterprise digital transformation: a quasi-natural experiment based on national big data comprehensive pilot zones. China Ind. Econ. 9, 117–135 (in Chinese).
Tan, Q., Cui, Y., 2026. Can digital innovation be a new driver of corporate sustainability? China J. Account. Res. 19 (2), 100466. https://doi.org/10.1016/j.cjar.2025.100466.
Tan, Y., Liu, Y., Zhang, X., 2025. Statistical measurement of enterprise digital transformation degree: from the perspective of digital technology talent demand. Econ. Res. J. 60 (4), 122–138 (in Chinese).
Tao, C., Li, X., Xu, Y., 2025. Government digital subsidies and corporate digital technology innovation: evidence from Chinese listed companies. J. Innov. Knowl. 10 (5), 100797. https://doi.org/10.1016/j.jik.2025.100797.
Verhoef, P.C., Broekhuizen, T., Bart, Y., et al., 2021. Digital transformation: a multidisciplinary reflection and research agenda. J. Bus. Res. 122, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022.
Wang, H., Hu, Y., Yuan, L., et al., 2023a. Energy supply and firm production performance: evidence from ultra-high voltage transmission projects. World Econ. 46 (12), 141–166 (in Chinese).
Wang, S., Li, X., Li, Z., et al., 2023b. The effects of government support on enterprises' digital transformation: evidence from China. Manag. Decis. Econ. 44 (5), 2520–2539. https://doi.org/10.1002/mde.3831.
Wang, S., Zhang, X., Zhang, Y., et al., 2017. A survey on mobile edge networks: convergence of computing, caching and communications. IEEE Access 5, 6757–6779. https://doi.org/10.1109/ACCESS.2017.2685434.
Wernerfelt, B., 1984. A resource-based view of the firm. Strateg. Manag. J. 5 (2), 171–180 https://www.jstor.org/stable/2486175.
Westerman, G., Bonnet, D., 2015. Revamping your business through digital transformation. MIT Sloan Manag. Rev. 56 (3), 10.
Wu, J., Yang, K., Song, F., 2025. Reconfiguring capital structures through digital R&D: mechanisms and empirical insights. Int. Rev. Econ. Financ., 104610 https://doi.org/10.1016/j.iref.2025.104610.
Wu, W., Wang, S., Jiang, X., et al., 2023. Regional digital infrastructure, enterprise digital transformation and entrepreneurial orientation: empirical evidence based on the broadband China strategy. Inf. Process. Manag. 60 (5), 103419. https://doi.org/10.1016/j.ipm.2023.103419.
Xiang, J., Yang, G., Chen, Y., et al., 2025. Business group affiliation and corporate digital transformation. Econ. Model. 147, 107068. https://doi.org/10.1016/j.econmod.2025.107068.
Xiao, G., Shen, S., 2022. To pollute or not to pollute: political connections and corporate environmental performance. J. Corp. Finance. 74, 102214. https://doi.org/10.1016/j.jcorpfin.2022.102214.
Xie, Y., Wu, D., 2024. How does competition policy affect enterprise digitization? Dual perspectives of digital commitment and digital innovation. J. Bus. Res. 178, 114651. https://doi.org/10.1016/j.jbusres.2024.114651.
Xiong, Q., Yang, J., Zhang, X., et al., 2025. The influence of digital transformation on the total factor productivity of enterprises: the intermediate role of human-machine cooperation. J. Innov. Knowl. 10 (4), 100736. https://doi.org/10.1016/j.jik.2025.100736.
Xu, H., Wu, H., Zheng, W., et al., 2026. Business groups and affiliated firms' digital transformation. China J. Account. Res., 100488 https://doi.org/10.1016/j.cjar.2026.100488.
Xu, N., Mao, J., Mao, X., et al., 2025. Computing power deployment, cross-border data flow, and firm total factor productivity: evidence from intelligent computing centers. China Ind. Econ. 4, 61–79 (in Chinese).
Yang, G., Wang, H., Fan, H., et al., 2023. The carbon emission reduction effect of the digital economy: theoretical analysis and empirical evidence. China Ind. Econ. 5, 80–98 (in Chinese).
Yang, N., Chen, X., 2026. Following the trend or not? How does peer pressure affect firm's digital transformation. J. Bus. Res. 205, 115877. https://doi.org/10.1016/j.jbusres.2025.115877.
Yu, D., 2024. Computing power: new quality productive force in the digital economy era. Financ. Trade Res. 35 (7), 1–16 (in Chinese).
Yu, J., Yuan, G.G., Liu, S., et al., 2025. Green credit, supply chain concentration and corporate digital transformation: evidence from China. Br. Account. Rev., 101690 https://doi.org/10.1016/j.bar.2025.101690.
Yukun, S., Bo, L., Junlin, L., 2024. Computing power network: a survey. China Commun. 21 (9), 109–145. https://doi.org/10.23919/JCC.ja.2021-0776.
Zhang, J., Wei, H., Yuan, K., et al., 2024. New industrial policy and corporate digital transformation: empowering or impairing? Emerging evidence from green credit policy. Energy Econ. 140, 107960. https://doi.org/10.1016/j.eneco.2024.107960.
Zhang, W., Lu, C., Liang, S., 2023. More words but less investment: rookie CEOs and firms' digital transformations. China J. Account. Res. 16 (3), 100305. https://doi.org/10.1016/j.cjar.2023.100305.
Zhao, Y., Tian, J., Zhang, S., et al., 2026. Beyond computation: how national supercomputing centers facilitate new business entry. J. Bus. Venturing. 41 (3), 106590. https://doi.org/10.1016/j.jbusvent.2026.106590.
Zhao, J., Ren, M., 2025. Innovative or conservative? How clan culture shapes bank digital transformation in China. China J. Account. Res., 100406 https://doi.org/10.1016/j.cjar.2025.100406.
Zhao, T., Zhang, Z., Liang, S., 2020. Digital economy, entrepreneurial activity, and high-quality development: empirical evidence from Chinese cities. Manag. World 36 (10), 65–76 (in Chinese).
Zou, Z., Fu, J., Zeng, Y., et al., 2024. Do young CEOs matter for corporate digital transformation? Econ. Lett. 237, 111636. https://doi.org/10.1016/j.econlet.2024.111636.

China Journal of Accounting Research 19 (2026) 100481

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Social credit systems and corporate AI development

Tai Wang

Antai College of Economics and Management, Shanghai Jiao Tong University, No.1954 Huashan Road, Shanghai 200030, China

ARTICLE INFO

Article history:
Received 9 January 2025
Accepted 2 March 2026
Available online 16 April 2026

Keywords:
AI
Social credit system construction
Financial support
Talent agglomeration
Regional innovation

ABSTRACT

This study examines the impact of social credit system construction on entrepreneurial activities and AI development using a quasi-natural experiment with data from 284 cities between 2007 and 2022. Our findings reveal that establishing a social credit system enhances AI advancement by improving financial support for businesses, fostering regional talent agglomeration and promoting innovation. The positive effects are particularly pronounced in regions with lower market integration, in the service industry and for non-SOEs, firms with low internal control and start-up firms. This research highlights how social credit systems can drive enterprise-level AI development through informal institutions, offering insights for cultivating a distinct financial culture in China and suggesting pathways to enhance financial services that support the real economy.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In the context of the rapid advancement of the trust economy, AI has emerged as a pivotal force driving social and economic transformation. The continuous evolution of technology, coupled with the expansion of application scenarios, positions enterprises at the forefront of AI research, application and innovation. AI not only enhances production efficiency and reduces operational costs but also facilitates the creation of novel business models and competitive advantages. However, enterprises encounter numerous challenges in their endeavors to develop AI capabilities, including funding shortages, technical barriers and market uncertainties, all of which can impede their innovation potential and growth trajectory.

In recent years, there has been increasing emphasis on the development of social credit systems, which have become critical policy instruments aimed at improving the business environment and promoting economic development. By establishing trust mechanisms, optimizing resource allocation and lowering transaction costs, social credit systems create an external environment conducive to the sustainable growth of enterprises.

E-mail address: wangtai@sjtu.edu.cn

https://doi.org/10.1016/j.cjar.2026.100481
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

In an economic climate characterized by financing difficulties, social credit systems can enhance companies' credit ratings, enabling them to secure more robust financial support. Moreover, a favorable social credit environment is instrumental in attracting high-quality talent and fostering the aggregation of innovative resources, thereby providing a solid foundation for enterprises engaged in AI research and development.

Although the literature acknowledges the positive effects of social credit systems on economic activities (Yu et al., 2015; Liang et al., 2018), there is a paucity of systematic research examining the relationship between social credit systems and AI development. Most scholarly work focuses on the implications of social credit systems for corporate financing and market behavior (Cornée and Szafarz, 2014; Wang et al., 2024a, 2024b), and there is limited exploration of their role in fostering technological innovation and development within businesses.

This study investigates the extent to which the construction of social credit systems facilitates the development of AI within enterprises. Specifically, it analyzes the mechanisms through which social credit systems contribute to the progress of corporate AI initiatives, including enhancing the breadth and depth of financial support, leveraging the agglomeration effects of high-caliber talent and promoting the generation and protection of regional innovation outcomes. To accomplish this, we use demonstration cities that have implemented social credit system initiatives as quasi-natural experiments to evaluate their impact on enterprise AI development. Through empirical analysis across various regions and industries, this research elucidates the pathways and internal mechanisms through which social credit system construction influences the evolution of AI in enterprises.

We adopt a quasi-natural experiment to investigate the impact of the construction of a social credit system on entrepreneurial activities. Using a dataset of 284 cities from 2007 to 2022, we implement a multi-time difference-in-differences (DID) model to analyze the effects and underlying mechanisms of social credit system implementation. Our findings reveal that the establishment of a social credit system promotes the advancement of AI by enhancing financial support, fostering regional talent aggregation and boosting regional innovation levels. Heterogeneity analyses indicate that the positive impact of social credit system construction on AI development is particularly significant in regions with lower market integration. In addition, this effect is more pronounced in the service industry and for non-SOEs, firms with low internal control and start-up firms. This study examines the promotional effects of social credit systems on enterprise-level AI development through the lens of informal institutions, providing empirical insights for the development of a unique financial culture in China and highlighting strategies for improving financial services that support the real economy.

Through this investigation, we highlight the critical role of social credit system construction in steering the development of enterprise-level AI and offer valuable perspectives and recommendations for furthering regional economic and social progress. By conducting an in-depth analysis of the interplay between social credit systems and corporate AI development, this study contributes new insights into the dynamics between trust mechanisms and technological innovation in the contemporary socioeconomic landscape while also providing empirical evidence to inform strategies for optimizing social credit systems and promoting the widespread application and advancement of AI technologies.

This study contributes to the literature in three important ways. First, it enriches the discourse surrounding social credit and enhances the credibility of its findings. Although many prior studies rely on questionnaire survey data to assess social credit (Yu et al., 2015; Gutiérrez-Nieto et al., 2016; Kostka and Antoine, 2020), this approach is fraught with potential issues, including subjective measurement error and sample selection bias. Although some scholars explore the punitive mechanisms associated with trust violations, using data on individuals subject to enforcement for trust breaches to gauge social credit levels, this perspective tends to overlook the positive effects of incentives for trustworthiness on enhancing regional social credit (Brockman et al., 2020; Liu et al., 2022). In contrast, we use the pilot policy of social credit system construction as a metric for regional social credit levels and investigate its impact on enterprise-level AI development. This approach not only mitigates endogeneity concerns but also enables a comprehensive assessment of the policy effects stemming from social credit system implementation. This study is therefore a valuable addition to the literature on the socioeconomic impacts of informal institutions.

Second, we broaden the research perspective on the influence of the informal institutional environment on entrepreneurship. The literature predominantly concentrates on the effects of formal institutions on AI (Erdélyi and Goldsmith, 2018; Raisch and Krakowski, 2021), with comparatively few studies addressing

the entrepreneurial impacts of informal institutions. By examining the mechanisms of social credit, which is considered an informal institution, this study explores how it influences the development of enterprise-level AI through support mechanisms, talent agglomeration and regional innovation. This exploration greatly enriches the literature on the role of institutional environments in fostering AI development.

Third, this research provides valuable policy insights for advancing the construction of social credit systems and, consequently, fostering the development of AI in regional enterprises. Because credit is foundational to the market economy, establishing a robust social credit system is a critical strategy for enhancing the credit environment and optimizing the business landscape. The findings offer a roadmap for the promotion of social credit system construction and can serve as a reference for decision-makers in government departments, thus guiding efforts to stimulate the development of AI.

2. Institutional background

The acceleration of the development of the social credit system is a critical measure for refining the socialist market economy and enhancing social governance. It plays a pivotal role in fostering integrity among societal members, cultivating a favorable credit environment and boosting national competitiveness. In recent years, China has introduced a series of policies aimed at improving the social credit system, focusing on enterprise and personal credit information systems, credit evaluation systems and the credit reporting industry. In June 2014, the State Council issued the first national-level strategic plan for social credit system construction, the Outline of the Social Credit System Construction Plan (2014–2020) (hereafter referred to as "the Outline"). Since then, this document has served as a guiding framework for the development of the social credit system, marking a notable phase of accelerated improvement.

Under the guidance of the Outline, in August 2015, the National Development and Reform Commission (NDRC) and the People's Bank of China jointly designated 11 cities as the first batch of national demonstration cities for social credit system construction. In April 2016, an additional 32 cities were approved for pilot programs. The implementation of this policy has established a comprehensive and systematic credit information sharing mechanism, which is reflected in three key areas. First, each city has developed a local credit information sharing platform, enhancing local credit websites to facilitate user access to credit information and real-time synchronization with the national credit information website Credit China. Second, pilot cities have leveraged the expertise of credit service institutions, collaborating closely on credit information sharing and the development and application of credit products. Third, these cities have formulated standards and regulations to establish red and black lists for key industries, encouraging financial institutions and other entities to use credit information and thereby institutionalizing and automating credit-based joint rewards and punishments. This has greatly enhanced the market value of enterprise credit and established a market supervision system centered on credit.

Compared with previous policies, this round of demonstration city construction places greater emphasis on establishing a robust credit information sharing mechanism, with a focus on large-scale credit information collection and application. The expansion of social member credit records has positively influenced corporate integrity and empowered the public to make informed decisions. According to NDRC investigations and announcements, all prefecture-level cities listed as demonstration cities for social credit system construction had established public credit information sharing platforms by the end of 2017, achieving the institutionalized and regularized collection of public credit information. This provides a reliable context for analyzing the impact of credit information sharing on corporate cash holdings.

3. Literature review

3.1. Information sharing

The literature on information sharing originates within the field of industrial organization theory, focusing primarily on the mechanisms and market consequences of information sharing in product markets (Shapiro, 1986). However, with the establishment of credit information sharing systems in various countries during the 1990s, academic interest shifted toward exploring information sharing mechanisms in credit markets.

Research has identified two primary functions of credit information sharing mechanisms: information screening and constraint. These functions have significantly enhanced bank performance, influenced enterprise behavior and improved overall market efficiency. For example, Pagano and Jappelli (1993) use data from the personal consumer credit market to demonstrate that credit information sharing enables banks to assess borrowers' creditworthiness more accurately, thereby reducing loan default rates. Similarly, Padilla and Pagano (1997) develop a model demonstrating that inter-bank information sharing can intensify competition between financial institutions, mitigating hold-up behaviors that arise from monopolistic control over borrower information.

Gehrig and Stenbacka (2007) posit that in scenarios of repeated competition between banks, credit information sharing allows lenders to acquire information at lower costs, enabling them to attract more customers from their competitors and ultimately increasing their market equilibrium benefits. In situations characterized by information asymmetry, borrowers might be inclined to pursue high-risk, low-value investment projects. Therefore, establishing a long-term reputation mechanism is crucial for incentivizing optimal investment decisions.

Credit information sharing also functions as a reputational constraint on corporate behavior. For instance, Diamond (1989) finds that a lack of historical borrowing information between borrowers and lenders can lead to excessive risk-taking by borrowers, a tendency that can be mitigated by the existence of a credit information sharing mechanism.

3.2. Economic consequences of the social credit environment

The economic implications of the social credit environment have been extensively examined at both the macro and micro levels. At the macro level, endogenous social trust in diverse cultural contexts enhances market efficiency by reducing transaction costs (Fukuyama, 1995), thereby contributing to economic growth (Horváth, 2013; Ng et al., 2016; Corradini, 2022). Zhang (2020) uses cross-national data to demonstrate that economies with superior social credit environments exhibit higher operational efficiency in the financial sector and are more likely to adopt a market-oriented financial system. Xiao and Chen (2025) highlight the positive role of social credit in fostering securities investment and direct investment. Wang et al. (2024a, 2024b) observe that the social credit environment promotes overall economic development and may also improve corporate investment efficiency.

At the micro level, social credit facilitates information generation and sharing, which lead to improvements in the quality of corporate financial reporting (Garrett et al., 2014) and encourage the voluntary disclosure of information by enterprises (Nicholls, 2010). Chen et al. (2024) find that AI significantly increases corporate risk-taking by reducing firms' risk perception and improving their decision-making efficiency. This effect is more pronounced for firms facing severe financing constraints, as AI alleviates information asymmetry and improves resource allocation in high-uncertainty scenarios. Wang et al. (2024a, 2024b) note that the absence of a robust social credit framework increases transaction costs and financing constraints for businesses, thereby weakening their innovation incentives and hindering the growth of local private enterprises. Kimmich and Wenzlaff (2022) find that deteriorating social credit can intensify agency problems within firms, increase financing constraints and induce precautionary and profit-seeking behaviors, ultimately leading to higher financialization.

Zuo et al. (2023) demonstrate that the social credit system promotes general innovation in enterprises (such as traditional technological R&D and non-AI patent output) by alleviating financing constraints and reducing information asymmetry. However, they do not address the particularity of AI development. Unlike traditional innovation, AI development involves not only technological R&D activities but also the substitution and enhancement of human judgment (such as intelligent monitoring replacing manual inspection and AI algorithms assisting business decisions), and it requires a special interaction between data credibility and trust mechanisms (for instance, AI risk control relies on the reliability of enterprise credit data, and human–machine collaboration is based on trust in technology).

This study expands this field. It focuses for the first time on the social credit–AI development sub-scenario, thus filling the gap in the literature on the differentiated impact of technological characteristics. It also reveals the dual effects of social credit on AI development. The positive effect is manifested in the enhancement of

information transparency by social credit (such as the publicization of enterprise credit ratings), reducing data risks in AI applications (such as user data fraud) and thereby promoting development. The potential negative effect is that social credit strengthens interpersonal trust (such as the willingness to fulfill contracts in business cooperation), which may reduce enterprises' demand for machine trust mechanisms (such as AI contract fulfillment monitoring). Through subsequent mechanism tests, this study verifies the conclusion that the positive effect dominates, thus strengthening the theoretical link between social credit and technology development.

4. Hypothesis development

From a comprehensive mechanistic perspective, the positive impacts of social credit on AI development are multidimensional and dominant. Social credit can improve enterprises' credit ratings by standardizing credit information disclosure and strengthening reputation constraints, enabling them to secure long-term financing crucial for high-investment AI R&D. Social credit can optimize the urban living and business environment, encompassing improvements in public services, the rule of law and social order. This makes cities more attractive to core AI talent such as algorithm engineers, data scientists and AI researchers, as a stable and trustworthy environment reduces the uncertainties associated with career development and life arrangements, which fosters the concentration of high-quality human capital essential for AI advancement. Social credit also fosters regional innovation sharing by strengthening intellectual property protection and reducing information asymmetry. It encourages enterprises to engage in open innovation through mechanisms such as credit incentives and joint rewards and punishments, thus facilitating the transfer and diffusion of patent technologies and innovative achievements (Zuo et al., 2023). This lowers the R&D costs of AI technologies for enterprises (innovation spillover) and accelerates the commercialization and application of AI.

First, the construction of a social credit system enhances the development of regional AI by increasing financial support. The establishment of a social credit system can greatly facilitate financial development, particularly in regions where the credit environment is underdeveloped. In such areas, issues such as moral hazard and adverse selection, which arise from information asymmetry, are especially pronounced among entrepreneurial groups, particularly small and micro enterprises and individual businesses (Crawford et al., 2018). These problems hinder the efficient allocation of financial resources.

By improving credit information sharing and using national financing credit service platforms, the social credit system mitigates information asymmetry between credit parties. This enhancement encourages the expansion of credit loans and short-term loans, fostering the development of regional financial markets. In turn, a more robust regional financial environment stimulates advancements in AI. As Dutta and Meierrieks (2021) argue, a higher degree of financial development promotes AI development primarily through improvements in financial depth and accessibility, alleviating the financing constraints and transaction costs faced by enterprises. In addition, the growing prevalence and sophistication of trust technologies have transformed the relationships between governments, financial institutions and market participants. This evolution in trust finance enhances the availability of information for market players, allowing financial institutions to accurately assess enterprises' financing needs. This reduces the transaction costs associated with financial services and supports the overall development of AI.

Second, the construction of a social credit system plays a crucial role in fostering the development of AI by facilitating talent agglomeration. An improved social credit environment is essential for attracting high-quality talent. According to signaling theory, a social credit system is a powerful policy tool for optimizing the business environment. By enhancing incentives for honesty in public services such as education, healthcare and cultural offerings, the social credit system sends positive signals to market participants. The resulting improvements in the living environment and quality of life make regions more appealing to skilled individuals. Moreover, a higher level of social credit promotes the free flow of information between social groups. Through mechanisms of reputation and information sharing, the social credit system expands individual social networks (Mickiewicz and Rebmann, 2020). This increased connectivity reduces transaction costs related to talent mobility and employment, further facilitating regional talent agglomeration.

The concentration of talent greatly enhances AI development activities. The externalities associated with labor agglomeration improve the overall availability and efficiency of resources within the region, creating favorable conditions for AI development. Furthermore, the presence of diverse talent fosters a richer environ-

ment for learning and information exchange, enhancing the skills and capabilities of the workforce. This boost in human capital equips enterprises with unique resources to navigate uncertainties in the external environment and to capitalize on opportunities for AI development. In summary, the social credit system not only attracts high-quality talent but also enhances the collaborative potential of that talent, ultimately driving the growth of AI in the region. By improving the business environment, facilitating information flow and concentrating talent, the social credit system lays the groundwork for a thriving AI ecosystem.

The construction of a social credit system can also promote the development of AI by improving the level of regional innovation. The improvement of the social credit level can promote innovation. Research shows that social credit can reduce the opportunistic behavior of enterprises and promote a higher level of innovation development. In addition, as a contract-intensive activity, innovation is greatly affected by the contractual environment. A social credit system can use the incentive of keeping faith to play the role of credit culture in improving the contract environment, encourage cooperation and tolerate failure, thus promoting innovation. In addition, improving the innovation level can stimulate the vitality of AI. The improvement of the regional innovation level promotes the flow of knowledge and information, gives full play to the knowledge spillover effect, promotes the development and transformation of innovation achievements and boosts the emergence of entrepreneur clusters, thus promoting the development of AI. The protection of intellectual property rights can also foster continuous innovation vitality by improving the level of knowledge spillover, support the formation of a fair and competitive market environment and provide a good external environment for the development of AI activities.

Although some perspectives suggest that enhanced interpersonal trust driven by social credit could reduce reliance on machines, the trust foundations and applicable scenarios of interpersonal trust and human–machine dependence are inherently distinct. Interpersonal trust is more rooted in emotional connections, mutual understanding and reciprocal commitments between individuals, often serving social interaction and cooperative scenarios that require subjective judgment and empathy. In contrast, human–machine dependence stems primarily from the technical reliability, efficiency improvements and functional support provided by AI technologies. Its core is the recognition of AI's ability to solve practical problems (such as data processing, risk control and intelligent decision-making) rather than emotional reliance. Importantly, these two forms of trust are not mutually exclusive; instead, they complement each other in different social and economic contexts. For instance, enhanced interpersonal trust can facilitate cooperation between enterprises, while AI dependence can improve the efficiency of collaborative projects—there is no inherent conflict between the two.

Overall, the synergistic effect of positive mechanisms far outweighs the potential inhibitory effect, which confirms that social credit system construction ultimately promotes the development of corporate AI. This study therefore proposes the following hypothesis:

H1: The construction of a social credit system can improve corporate AI development.

5. Data and variables

5.1. Data sources

This study takes A-share listed companies in Shanghai and Shenzhen as the research object, and the sample period is from 2007 to 2022. The reasons for choosing 2007 as the starting point are as follows. First, the 2017 Special Research Report on China's AI Industry points out that the current wave of AI development mainly comes from the proposal of deep learning algorithms in 2006. These realize large-scale operation on the basis of data volume and computing power, which is a major breakthrough in technology. Therefore, the advantages of AI technology may have been more obvious since 2007. Second, the Research Report on the Market Prospect of China's AI Industry in 2022 documents that after 2007, China's patents in the field of AI began to enter the development stage, which makes the collection of AI patent data feasible. This identifies the listed companies for the sample period from the Sina Finance and Economics website (https://finance.sina.com.cn/). Patent data are obtained from the IRPDB Intellectual Property database (https://www.iprdb.com/). Enterprises' basic information and financial data are obtained from the CSMAR database.

To ensure the quality of the data, the sample is processed as follows. (1) Companies in the financial industry are excluded. (2) Information transmission, software and information technology services and scientific

research and technology services are excluded because these industries naturally use cloud computing, big data and AI technologies and disclose relevant information, and it may be impossible to clearly judge the impact of the application of AI technology on these enterprises' production efficiency. (3) Firms with ST or *ST status during the sample period are excluded. (4) Samples with missing data are also excluded. Finally, 21,212 firm-year observations are obtained. To eliminate the influence of extreme values, this study winsorizes all continuous variables at the 1% level. The data selection process is presented in Table 1.

5.2. Explained variables

The frequency of AI-related keywords and the development of AI patents are critical indicators of a company's AI development. The frequency of AI-related keywords reflects the company's emphasis on AI and its strategic intent in public communications, including annual reports and press releases, thereby demonstrating its promotional efforts and strategic orientation regarding AI. To construct measures of AI-related keywords, a machine learning method is used to generate an AI dictionary and construct an enterprise AI index based on the management discussion and analysis (MD&A) sections (please see Appendix B). The frequency of AI-related words in the MD&A sections is used to determine the enterprise's AI technology level (Wang et al., 2024a, 2024b). *Lnwords* is calculated as the natural logarithm of 1 plus the frequency of AI-related keywords in the full text of a firm's annual report. *Lnwords_MDA* is calculated as the natural logarithm of 1 plus the frequency of AI-related keywords in the MD&A section of a firm's annual report.

AI patents are a quantitative indicator of a company's achievements in research and innovative applications, revealing its investments and competencies in technological innovation. *Lnpatents* is defined as the natural logarithm of 1 plus the number of AI-related patent applications by a firm in a year.

The integration of these two indicators facilitates a comprehensive evaluation of a company's strength and strategic positioning in the AI landscape. Although a high frequency of AI-related keywords may suggest a heightened focus on AI, it does not necessarily correlate with the company's technical capabilities. However, together, the number and quality of patents directly reflect the company's actual technical performance and innovative outcomes.

5.3. Explanatory variable

This study defines the policy impact variable *Trust* according to whether a city is included on the list of demonstration cities for the construction of the social credit system, with a value of 1 for the year of implementation and all subsequent years, and 0 otherwise.

5.4. Control variables

This study incorporates a vector of firm and industry characteristics that may influence a firm's innovation activities. To account for firm size, we use the natural logarithm of total assets (*Size*). Firm age is represented as the logarithmic value of the number of years since the firm's establishment (*Firm Age*). We measure capital intensity using the net amount of property, plant and equipment scaled by total assets (*PPE*). The ratio of return on assets (*ROA*) serves as a proxy for operating profitability. The other variables are the ratio of total

Table 1
Data selection.

Number of observations		Total
Population of all enterprises listed from 2007 to 2022		23,754
Exclude: Observations from financial industry	917	
*Observations such as ST and *ST*	415	
Observations from missing value	1210	
Final sample		21,212

debt to total assets (*Lev*), equity concentration represented by the ownership stake of the 10 largest shareholders (*Top10*), a binary variable indicating whether the roles of chairperson and general manager are held by the same individual (*Dual*), the proportion of independent directors on the board (*Indep*), the total number of directors (*Board*), the ratio of shares held by management (directors, supervisors and senior executives) to total share capital (*Mshare*), the ratio of total management compensation to annual operating income (*Mfee*), cash holdings divided by total assets (*Cashflow*) and a measure of financial constraints (*SA*). Furthermore, this study uses year dummy variables (*Year*) and firm dummy variables (*Firm*) to control for time and firm-specific effects on the regression model. Detailed definitions and measurements of these variables are provided in Appendix A.

6. Empirical analysis

6.1. Summary statistics

Table 2 presents the summary statistics for the main variables. All of the continuous variables are winsorized at the top and bottom 1% levels to mitigate the influence of extreme values. The mean of *Lnpatents* is 0.333, which is similar to the findings of Liu and Kong (2021). The distributions of the other variables are similar to those in prior Chinese studies.

6.2. Multivariate DID analysis

Forty-two cities participated in the pilot of the construction of the social credit system in two batches. Based on the discussion of the institutional background, this study regards the pilot policy of social credit system construction as a quasi-natural experiment and constructs the following multi-time DID model to test the impact of the pilot policy on regional AI development:

$$AIdevelopment_{i,t} = \alpha_i + \beta_1 Trust_{i,t} + \gamma Z_{i,t} + \delta_t + \varphi_i + \varepsilon_{i,t} \quad (1)$$

Table 2
Summary statistics.

Variable	*N*	*Mean*	*SD*	*p1*	*p25*	*p50*	*p75*	*p99*
Lnpatents	21,212	0.333	0.737	0.000	0.000	0.000	0.000	3.434
Lnwords	21,212	0.089	0.228	0.000	0.000	0.000	0.000	1.041
Lnwords_MDA	21,212	0.055	0.170	0.000	0.000	0.000	0.000	0.845
Trust	21,212	0.317	0.465	0.000	0.000	0.000	1.000	1.000
Size	21,212	22.309	1.278	20.008	21.397	22.120	23.020	26.280
FirmAge	21,212	2.927	0.320	1.946	2.708	2.944	3.178	3.526
Lev	21,212	0.431	0.200	0.060	0.273	0.426	0.579	0.886
Indep	21,212	0.376	0.054	0.333	0.333	0.357	0.429	0.571
Board	21,212	2.126	0.196	1.609	1.946	2.197	2.197	2.708
Top10	21,212	0.575	0.149	0.234	0.470	0.581	0.683	0.896
SA	21,212	−3.829	0.246	−4.431	−3.996	−3.833	−3.663	−3.192
Dual	21,212	0.272	0.445	0.000	0.000	0.000	1.000	1.000
Growth	21,212	0.173	0.385	−0.522	−0.016	0.112	0.273	2.354
Cashflow	21,212	0.049	0.066	−0.142	0.011	0.048	0.088	0.240
Mshare	21,212	0.125	0.187	0.000	0.000	0.005	0.225	0.671
Mfee	21,212	0.088	0.146	0.008	0.042	0.069	0.107	0.377
INST	21,212	0.405	0.237	0.001	0.21	0.417	0.59	0.884

This table presents the summary statistics of the variables in our main analysis for the Mean, standard deviation (SD), Median, minimum and maximum distributions. The full panel sample comprises 21,212 firm-year observations from 2007 to 2022. Variable definitions are detailed in the Appendix A.

In this formula, $AIdevelopment_{i,t}$ is the dependent variable, representing the degree of AI development of enterprise i in year t. $Trust_{i,t}$ is the core explanatory variable, which is a dummy variable indicating whether city i is part of the pilot social credit system construction. To avoid the influence of missing variables, the model controls for year and firm fixed effects. Table 3 shows the results. As shown, the coefficient is positive and significant.

In Column 1 of Table 3, where *Lnpatents* is the dependent variable, the coefficient is 0.044 ($p < 0.05$), indicating that when the other variables are controlled, the number of AI patents of enterprises in the social credit system pilot cities is on average 4.4% higher than that in non-pilot cities. Based on the sample mean (*Lnpatents* = 0.333), the actual increase in AI patents is approximately 13.2% (0.044/0.333). This effect is significantly higher than the impact of enterprise size (the coefficient of *Size* is 0.143, and a 1% increase only increases AI patents by 0.143%), demonstrating that the economic impact of social credit on AI development is substantial.

Table 3
Basic staggered DID regression.

	Lnpatents	*Lnwords_MDA*	*Lnwords*
	(1)	*(2)*	*(3)*
Trust	0.044**	0.025***	0.041***
	(2.06)	(4.95)	(6.00)
Size	0.143***	0.010***	0.018***
	(11.51)	(4.75)	(5.97)
FirmAge	0.378***	0.007	0.025
	(4.10)	(0.51)	(1.22)
Lev	−0.184***	−0.024**	−0.033**
	(−3.74)	(−2.37)	(−2.42)
Indep	0.183	0.035	0.022
	(0.88)	(0.91)	(0.41)
Board	0.090	0.001	0.001
	(1.35)	(0.08)	(0.06)
Top10	−0.217***	−0.025*	−0.046**
	(−3.01)	(−1.79)	(−2.35)
SA	0.524***	0.027	0.058**
	(4.12)	(1.45)	(2.13)
Dual	0.021	0.003	0.009
	(1.08)	(0.75)	(1.58)
Growth	−0.039***	−0.004	−0.001
	(−3.54)	(−1.13)	(−0.13)
Cashflow	0.453***	−0.057***	−0.061**
	(4.43)	(−2.76)	(−2.11)
Mshare	0.220***	0.001	0.013
	(3.96)	(0.11)	(0.78)
Mfee	0.016	0.000	0.004
	(0.66)	(0.04)	(0.59)
INST	0.147***	−0.007	−0.001
	(3.25)	(−0.82)	(−0.06)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	21,212	21,212	21,212
Adj. R^2	0.208	0.141	0.196

Notes: This table examines the impact of social credit system construction on corporate AI development using a staggered DID model. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics reported in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix A.

6.3. Parallel trend tests

It is possible that our results are driven by reverse causality, as there are concerns that the expected change in AI development induces the government to issue a new public data opening policy. Another assumption related to the validity of the DID approach is the parallel trend, which requires the average change in AI development to be the same in the firms in the treatment group and the control group. Without a parallel trend, the DID approach may generate biased results.

To address these concerns, we examine the dynamics of AI development surrounding the opening of a public data platform by estimating the following equation (Bertrand and Mullainathan, 2003; Atanassov, 2013):

$$\begin{aligned} AIdevelopment_{i,t} = {} & \alpha_i + \beta_1 Treat_i * Before_{i,t}^{-3+} + \beta_2 Treat_i * Before_{i,t}^{-2} + \beta_3 Treat_i * Before_{i,t}^{-1} + \beta_4 Treat_i \\ & * Current_{i,t}^{0} + \beta_5 Treat_i * After_{i,t}^{1} + \beta_6 Treat_i * After_{i,t}^{2} + \beta_7 Treat_i * After_{i,t}^{3+} + rZ_{i,t} \\ & + \sigma_t + \varphi_i + \epsilon_{i,t} \end{aligned} \tag{2}$$

where $Before_{i,t}^{-3+}$($Before_{i,t}^{-2}$ or $Before_{i,t}^{-1}$) is a dummy variable that takes a value of 1 if the observation occurs 3 years (2 years or 1 year) before the opening of the public data platform, and 0 otherwise. $Current_{i,t}^{0}$ is a dummy variable that takes a value of 1 if the observation occurs in the year when the platform opens, and 0 otherwise. $After_{i,t}^{1}$ ($After_{i,t}^{2}$ or $After_{i,t}^{3+}$) is a dummy variable that takes a value of 1 if the observation occurs in the first year (the second year or the third year and beyond) after the shock of opening public data construction, and 0 otherwise. All of the other variables are the same as those described in the baseline DID regression.

The coefficients of interest are β_1,β_2 and β_3. If there is any pre-trend in AI development between the treatment and control groups, we should obtain statistically significant coefficients for β_1,β_2 and β_3. We report the dynamic DID results in Table 4.

For all three dependent variables, the pre-policy interaction terms (β_1, β_2, β_3) are not statistically significant at the 10% level. *Lnpatents* has pre-policy coefficients of 0.012 ($t = 0.44$), 0.078 ($t = 1.42$) and 0.096 ($t = 0.73$), confirming no prior AI patent growth advantage in pilot cities. Similarly, *Lnwords_MDA* has small, nonsignificant pre-policy coefficients (0.008, $t = 0.31$; 0.021, $t = 0.89$; 0.019, $t = 0.76$), indicating no greater AI emphasis in management discussions. Likewise, *Lnwords* yields nonsignificant pre-policy coefficients (0.015, $t = 0.47$; 0.032, $t = 1.05$; 0.028, $t = 0.92$), ruling out pre-existing AI communication tendencies. Collectively, these results validate the DID parallel trend assumption and mitigate reverse causality concerns.

The positive effect of social credit system construction on AI development emerges in the pilot year and strengthens over time, consistent with the long-term nature of AI development (Xiao et al., 2020). For *Lnpatents*, the coefficient of $Current_{i,t}^{0}$ is 0.079 ($p < 0.05$), indicating an immediate increase in AI patent applications, and the coefficients of $After_{i,t}^{2}$ (0.069, $p < 0.05$) and $After_{i,t}^{3+}$ (0.039, $p < 0.1$) remain positive and significant, showing the policy's sustained contribution to technological output. For *Lnwords_MDA*, the policy effect becomes significant in the pilot year ($Current_{i,t}^{0}$: 0.033, $p < 0.05$) and grows in magnitude in subsequent years ($After_{i,t}^{1}$: 0.027, $p < 0.1$; $After_{i,t}^{2}$: 0.041, $p < 0.01$; $After_{i,t}^{3+}$: 0.035, $p < 0.05$), which reflects that firms in pilot cities gradually increase their strategic focus on AI, as evidenced by more frequent AI-related discussions in management reports. For *Lnwords*, the full-text AI keyword frequency also shows a significant post-policy increase ($Current_{i,t}^{0}$: 0.049, $p < 0.01$; $After_{i,t}^{1}$: 0.038, $p < 0.05$; $After_{i,t}^{2}$: 0.056, $p < 0.01$; $After_{i,t}^{3+}$: 0.047, $p < 0.05$), confirming that the policy's impact extends to firms' overall public communications about AI, not just management discussions.

Overall, the dynamic DID results in Table 4 validate the parallel trend assumption of the DID model and confirm that the positive impact of social credit system construction on corporate AI development is both statistically robust and temporally consistent.

6.4. Placebo test

It is possible that the DID analysis results are driven purely by chance. To address this concern, we conduct a placebo test. We select 2013 as the pseudo-event year and rerun our regression for the pre-2014 period. We

Table 4
Parallel trend test.

	Lnpatents	*Lnwords_MDA*	*Lnwords*
	(1)	*(2)*	*(3)*
before_3	0.012	0.008	0.015
	(0.44)	(0.031)	(0.47)
before_2	0.078	0.021	0.032
	(1.42)	(0.89)	(1.05)
before_1	0.096	0.019	0.028
	(0.73)	(0.76)	(0.92)
current	0.079**	0.033**	0.049***
	(2.30)	(2.28)	(2.65)
after_1	0.042	0.027*	0.038**
	(1.38)	(1.79)	(2.01)
after_2	0.069**	0.041***	0.056***
	(2.30)	(2.83)	(3.12)
after_3	0.039*	0.035**	0.047**
	(1.67)	(2.14)	(2.33)
Size	0.142***	0.011***	0.019***
	(11.47)	(4.82)	(6.05)
FirmAge	0.378***	0.008	0.026
	(4.13)	(0.55)	(1.25)
Lev	−0.167***	−0.023**	−0.032**
	(−3.43)	(−2.29)	(−2.38)
Cashflow	0.440***	−0.056***	−0.060**
	(4.29)	(−2.71)	(−2.08)
Dual	0.020	0.003	0.008
	(0.99)	(0.72)	(1.55)
Indep	0.174	0.034	0.021
	(0.83)	(0.89)	(0.40)
Board	0.097	0.001	0.002
	(1.46)	(0.09)	(0.07)
Top10	0.019	−0.024*	−0.045**
	(0.28)	(−1.76)	(−2.32)
SA	0.507***	0.026	0.057**
	(4.02)	(1.42)	(2.10)
Growth	−0.045***	−0.004	−0.002
	(−4.13)	(−1.11)	(−0.15)
Mshare	0.129**	0.001	0.012
	(2.47)	(0.10)	(0.76)
Mfee	0.019	0.000	0.003
	(0.78)	(0.05)	(0.57)
INST	0.050	−0.006	−0.002
	(1.16)	(−0.80)	(−0.07)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	21,211	21,211	21,211
Adj. R^2	0.207	0.143	0.198

Notes: This table tests the parallel trend assumption using a dynamic DID model. before_3+, before_2, before_1 are dummy variables for 3+ years, 2 years, and 1 year before the social credit pilot, respectively; current is the pilot year; after_1, after_2, after_3+ are 1 year, 2 years, and 3+ years after the pilot. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

find that *Trust* has a nonsignificant coefficient, indicating that the results of our baseline model are unlikely to be driven by chance. The results, which are presented in Table 5, are consistent with our baseline results.

7. Robustness tests

7.1. Instrumental variable tests

Referring to Kong et al. (2023), this study uses the number of urban Chinese sub-dialects as an instrumental variable to address potential endogeneity issues. The rationale for this choice lies in the correlation between regional dialect diversity and social inclusiveness. Areas with a greater variety of non-dominant dialects are likely to exhibit higher levels of social integration, reflecting enhanced social inclusiveness. Regions characterized by a greater number of Chinese sub-dialects are thus more likely to possess historical foundations and inherent advantages for the development of social credit systems, making them better candidates for designation as social credit demonstration cities.

In terms of exogeneity, dialects have evolved over millennia through historical processes and social transformations, making it unlikely that they would directly influence the development of AI within enterprises. This characteristic aligns with the assumptions required for the use of instrumental variables. Because this

Table 5
Placebo test.

	(1) *Lnpatents*
DID1	0.029
	(1.47)
Size	0.167***
	(10.85)
FirmAge	0.464***
	(4.75)
Lev	−0.085*
	(−1.65)
Cashflow	0.345***
	(3.32)
Dual	0.016
	(0.80)
Indep	0.154
	(0.71)
Board	0.084
	(1.22)
Top10	0.033
	(0.48)
SA	0.619***
	(4.54)
Growth	−0.053***
	(−4.66)
Mshare	0.142***
	(2.58)
Mfee	0.018
	0.75
INST	0.021
	(0.47)
Year FE	Yes
Firm FE	Yes
N	19,952
Adj. R^2	0.214

Notes: This table reports results of the placebo test, where 2013 is set as the pseudo-event year and the regression is restricted to the pre-2014 period. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

variable is cross-sectional, this study adopts the approach of Kong et al. (2024) by using the product of the number of Chinese sub-dialects and the interaction term of the policy time dummy variable associated with social credit demonstration cities as the instrumental variable, referred to as *Dialect*.

We conduct a Durbin–Wu–Hausman (DWH) test on the policy variable related to social credit demonstration cities (*Trust*). The test yields a DWH statistic of 4.470, leading to the rejection of the exogeneity hypothesis for *Trust* at the 5% significance level. Additionally, the results of the Kleibergen–Paap–Wald *F* test indicate the rejection of the null hypothesis of weak instrumental variables. Column (1) of Table 6 presents the regression results of the first stage, revealing a positive relationship for the *Dialect* variable that is significant at the 1% level. Column (2) displays the results from the second stage, where *Trust* exhibits a positive coefficient significant at the 5% level. This suggests that after we address endogeneity concerns, the establish-

Table 6
IV test.

	(1) First stage	(2) Second stage
	Trust	*Lnpatents*
Dialect	0.023***	
	(4.977)	
Trust		0.636**
		(2.018)
Size	0.062***	0.089***
	(17.183)	(4.416)
FirmAge	0.317***	0.209*
	(14.322)	(1.875)
Lev	−0.074***	−0.209***
	(−3.635)	(−5.683)
Cashflow	−0.022	0.274***
	(−0.414)	(3.303)
Dual	0.076***	0.011
	(9.554)	(0.418)
Indep	−0.208***	0.311**
	(−2.748)	(2.048)
Board	−0.286***	0.134
	(−13.166)	(1.358)
Top10	−0.050*	−0.120***
	(−1.914)	(−2.653)
SA	−0.007	0.614***
	(−0.248)	(9.333)
Growth	−0.020**	−0.028**
	(−2.215)	(−2.103)
Mshare	0.318***	0.173
	(14.219)	(1.597)
Mfee	−0.057**	0.012
	(−2.073)	(0.485)
INST	0.054***	0.009
	(2.802)	(0.256)
Year FE	Yes	Yes
Firm FE	Yes	Yes
N	19,364	19,364
Adj. R^2		0.057

Notes: This table reports results of the instrumental variable (IV) test. The IV Dialect is the interaction of "number of Chinese sub-dialect slices in a city" and the policy time dummy of social credit pilots. The DWH test (statistic = 4.470, $p < 0.05$) rejects exogeneity of Trust; the Kleibergen-Paap-Wald F test rejects weak IVs. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

ment of social credit demonstration cities does indeed promote the development of enterprise-level AI, thereby reinforcing the reliability of our conclusions.

7.2. PSM-DID test

Because we do not observe the counterfactual outcomes of the treatment firms, we are unable to compare the outcome differences of choices in the same treatment firm to evaluate the impact of open public data platforms. Thus, we have to select a control firm that is identical to the treatment firm except that it is not affected by the exogenous shock of the open public data platform. However, one may argue that the firms in the treatment group differ substantially from those in the control group. If so, the impact of the open public data platform shown in the DID estimation might be due to the omitted observable differences between our treatment and control groups.

To address this concern, we match each treatment firm to one control firm using propensity score matching (PSM). We first implement a logit regression to model the probability of being a treatment firm in the year prior to the shock. In the logit regression, we include a vector of control variables, including *Size*, *Lev*, *ROA*, *Firmage*, *Cashflow*, *PPE*, *SA*, *Top1*, *Dual*, *Indep* and *Board*. We add these variables to ensure that the parallel trend assumption of the DID regression holds.

We report the PSM-DID results in Table 7. Again, we find consistent results. All of the coefficients of *Trust* are positive and statistically significant in the regressions. Moreover, the magnitude of the term is slightly larger than that in the baseline results, suggesting that our main findings are robust to alternative samples. In particular, compared with firms without an open public data platform, those with an open public data platform have greater AI development.

7.3. Excluding the effects of other relevant city policies

During the sample period, China implemented other trust economy policies. From 2014 to 2016, the Ministry of Industry and Information Technology and the NDRC selected 120 cities as Broadband China demonstration areas in three batches.

In December 2012, the Ministry of Housing and Urban–Rural Development officially issued the Notice on the Implementation of National Smart City Pilot Work, the National Smart City Pilot Interim Management Measures and the National Smart City (District, Town) Pilot Index System (Trial), after which many cities were included as ‘‘smart city’’ demonstration areas.

In August 2015, to promote the comprehensive deployment of big data strategy, The State Council issued the Outline of Action for Promoting the Development of Big Data, proposing to construct a national comprehensive pilot zone for big data.

In September 2009, the NDRC and the Ministry of Commerce officially approved the establishment of Shenzhen as the first national e-commerce demonstration city in China. From 2011 to 2017, 70 new cities were established as national e-commerce demonstration cities. These trust economy policies may affect our research results.

To exclude the influence of relevant policies on the research results, this study adopts four indicator variables according to whether the city is a Broadband China demonstration area, a smart city demonstration zone, a big data pilot zone or an e-commerce demonstration city, denoted by *Broadband China*, *Smart city*, *Big data pilot zone* and *E-commerce city*, respectively. These policy indicator variables are introduced into the regression model for control.

The regression results are shown in Table 8, where Columns (1) to (4) are the estimated results after including each of the indicator variables. We find that the coefficients of *Trust* are positive and significant, consistent with the benchmark regression results.

Table 7
PSM-DID test.

	Lnpatents
Trust	0.033***
	(4.26)
Size	0.145***
	(10.20)
FirmAge	0.324***
	(3.03)
Lev	−0.197***
	(−3.19)
Indep	0.085
	(0.34)
Board	0.075
	(0.97)
Top10	−0.181*
	(−1.83)
SA	0.455***
	(3.38)
Dual	0.017
	(0.69)
Growth	−0.043***
	(−2.63)
Cashflow	0.483***
	(3.31)
Mshare	0.227***
	(3.00)
Mfee	−0.050
	(−0.38)
INST	0.184***
	(2.81)
Year FE	Yes
Firm FE	Yes
N	9903
Adj. R^2	0.210

Notes: This table reports results of the PSM-DID test. Each treated firm (in social credit pilot cities) is matched to a control firm using propensity scores, with matching variables including all control variables in the baseline regression. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

8. Mechanism tests

8.1. Financial support for AI development

The advancement of AI is fundamentally linked to the availability of financial support. To investigate this relationship, we assess financial support through two dimensions: breadth and depth.

In evaluating the breadth of financial support, prior studies often measure financial availability based on the distribution of commercial bank branches. In this study, we use financial licensing data published by the State Financial Regulatory Administration to determine the number of bank branches in various regions. Subsequently, we calculate the Herfindahl–Hirschman index (*HHI*) as an indicator of bank competition levels within these regions. To analyze the depth of financial support, we use the sub-index of the Trust Inclusive Financial Index, as constructed by Demirgüç-Kunt and Klapper (2013). This index encompasses various business categories, including credit and investment services, thereby providing a comprehensive measure of the actual utilization of trust financial services. Increased competition between banks enhances the efficiency of credit resource allocation, mitigates corporate financing constraints and fosters a financing environment con-

Table 8
Excluding other relevant policies.

	Lnpatents	*Lnpatents*	*Lnpatents*	*Lnpatents*
	(1)	*(2)*	*(3)*	*(4)*
Trust	0.039*	0.047**	0.039*	0.041*
	(1.74)	(2.06)	(1.71)	(1.83)
Broadband China	−0.039*			
	(−1.92)			
Smart city		−0.009		
		(−0.38)		
Big data pilot zone			−0.009	
			(−0.34)	
E-commerce city				−0.022
				(−1.08)
Size	0.144***	0.143***	0.145***	0.143***
	(11.11)	(11.52)	(11.12)	(11.07)
FirmAge	0.399***	0.378***	0.398***	0.396***
	(4.20)	(4.10)	(4.19)	(4.17)
Lev	−0.194***	−0.183***	−0.191***	−0.189***
	(−3.78)	(−3.73)	(−3.71)	(−3.67)
Indep	0.182	0.184	0.179	0.181
	(0.84)	(0.88)	(0.82)	(0.83)
Board	0.086	0.090	0.079	0.084
	(1.24)	(1.35)	(1.15)	(1.21)
Top10	−0.244***	−0.217***	−0.239***	−0.242***
	(−3.23)	(−3.00)	(−3.14)	(−3.19)
SA	0.550***	0.525***	0.545***	0.544***
	(4.19)	(4.13)	(4.15)	(4.15)
Dual	0.019	0.022	0.019	0.020
	(0.90)	(1.09)	(0.89)	(0.94)
Growth	−0.037***	−0.039***	−0.037***	−0.037***
	(−3.24)	(−3.54)	(−3.25)	(−3.26)
Cashflow	0.415***	0.451***	0.419***	0.420***
	(3.93)	(4.42)	(3.96)	(3.97)
Mshare	0.249***	0.221***	0.250***	0.251***
	(4.29)	(3.97)	(4.29)	(4.31)
Mfee	0.015	0.016	0.014	0.012
	(0.61)	(0.66)	(0.55)	(0.47)
INST	0.159***	0.147***	0.159***	0.160***
	(3.38)	(3.25)	(3.36)	(3.39)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	19,897	21,146	19,866	19,897
Adj. R^2	0.213	0.208	0.213	0.212

This table reports the results after excluding the effects of other relevant city policies. To mitigate the influence of relevant policies on the research outcomes, this study incorporates four dummy variables "*Broadband China*", "*Smart city*", "*Big data pilot zone*", and "*E-commerce city*". "*Broadband China*" equals to 1 if the city where the firm located is designated as a "Broadband China Demonstration City", 0 otherwise. "*Smart city*" *equals to* 1 if the city where the firm is located is designated as a "National Smart City Pilot", 0 otherwise. "*Big data pilot zone*" equals to 1 if the city where the firm is located is part of a "National Big Data Comprehensive Pilot Zone", 0 otherwise. "*E-commerce city*" equals to 1 if the city where the firm is located is designated as a "National E-commerce Demonstration City", 0 otherwise. The aforementioned policy-related indicator variables are included in the regression model as control variables to account for their potential impact on the results. Variable definitions are reported in Appendix A. All regressions include firm and year fixed effects. The t statistics reported in parentheses are based on standard errors clustered at the firm level. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

ducive to AI development. The integration of finance with trust technology enhances the inclusivity and depth of financial services, offering more accessible financial support for AI development.

Considering that the HHI is an inverse indicator of bank competition, we follow the methodology outlined by Wang et al. (2020) and take the logarithm of the HHI, making its values negative for regression analysis. The results of this analysis are presented in Column (1) of Table 9. The coefficient for *Trust* is statistically significant at the 5% level, suggesting that the establishment of a social credit system can enhance the development of AI by fostering increased competition between local banks.

To enhance the comparability of the results, we normalize the depth of utilization within the Trust Inclusive Financial Index before incorporating it into the regression model. The results of this analysis are reported in Column (2) of Table 9. Here, the coefficient for *Trust* is positive and significant at the 1% level, indicating that

Table 9
Mechanism test-financial support.

	HHI	*Trust Inclusive Financial Index*	*Lnpatents*	*Lnpatents*
	(1)	(2)	(3)	(4)
Trust	0.935*	2.602***	0.028**	0.031***
	(1.85)	(4.71)	(2.14)	(2.89)
HHI			−0.005***	
			(−3.52)	
Trust Inclusive Financial Index				0.012***
				(3.76)
Size	1.191***	0.237	0.141***	0.138***
	(4.28)	(0.84)	(11.26)	(10.98)
FirmAge	−0.213	−1.583	0.369***	0.357***
	(−0.12)	(−0.93)	(4.01)	(3.87)
Lev	−5.799***	0.554	−0.181***	−0.176***
	(−4.07)	(0.39)	(−3.68)	(−3.55)
Cashflow	−4.274*	3.057	0.448***	0.432***
	(−1.73)	(1.18)	(4.39)	(4.21)
Dual	0.207	0.175	0.020	0.018
	(0.41)	(0.37)	(0.99)	(0.88)
Indep	−2.023	−8.355	0.179	0.165
	(−0.44)	(−1.64)	(0.86)	(0.79)
Board	1.209	0.308	0.089	0.085
	(0.91)	(0.20)	(1.32)	(1.26)
Top10	44.668***	0.527	−0.212***	−0.205***
	(23.58)	(0.28)	(−2.97)	(−2.83)
SA	5.031**	−0.351	0.518***	0.502***
	(2.23)	(−0.18)	(4.08)	(3.94)
Growth	2.927***	0.058	−0.038***	−0.036***
	(8.68)	(0.23)	(−3.49)	(−3.32)
Mshare	41.002***	3.450**	0.218***	0.209***
	(16.77)	(2.42)	(3.91)	(3.75)
Mfee	0.021	0.019	0.014	0.012
	(0.65)	(0.58)	(0.60)	(0.51)
INST	32.085***	2.661**	0.145***	0.139***
	(20.16)	(2.47)	(3.22)	(3.08)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	6507	19,640	6507	19,640
adj. R^2	0.672	0.052	0.211	0.207

This table reports the results of the mechanism analysis. we calculate the Herfindahl-Hirschman Index (*HHI*) as an indicator of bank competition levels within these regions. To analyze the depth of financial support, we utilize the sub-index of the *trust inclusive financial index*, as constructed by Guo et al. (2020). This index encompasses various business categories, including credit and investment services, thereby providing a more comprehensive measure of the actual utilization of trust financial services. Variable definitions are reported in Appendix A. All regressions include firm and year fixed effects. The t statistics reported in parentheses are based on standard errors clustered at the firm level. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

the construction of a social credit system can deepen the development of trust inclusive finance, thereby promoting the development of AI at the regional level.

Social credit strengthens the role of financial support for AI through two pathways. First, the intensification of bank competition (with a decrease in HHI) prompts financial institutions to develop special AI credit products (such as loans for AI quality inspection equipment in the manufacturing industry), and the sample data show that the scale of AI-related credit in the pilot cities is 19.2% higher than that in the non-pilot cities (data from the Regional Report of the People's Bank of China); Second, the deepening of inclusive finance (such as the credit information sharing platform) helps small and medium-sized enterprises obtain funds for purchasing AI software (such as AI risk control systems in the SaaS model); such investment is the foundation for the development of AI. As shown in Table 9, the coefficient is positive and significant, verifying this link. The coefficient of the AI-specific credit scale is included as an intermediary variable in the regression analysis.

Overall, our findings demonstrate that the establishment of a social credit system can significantly enhance both the breadth and depth of financial services. This improvement in the supply of financial resources and credit directly addresses the financing challenges related to AI development, ultimately facilitating its growth and progress.

8.2. Talent agglomeration in AI development

Talent represents the cornerstone of AI initiatives, with high-quality professionals forming a crucial demographic for enhancing the quality of AI development. To examine the talent agglomeration effect resulting from the pilot implementation of the social credit system, we assess this phenomenon from two dimensions: scale and structure.

In terms of scale, we use the logarithmic values of employment figures in urban sectors such as information transmission, computer services and software industries as a measure of talent agglomeration levels (*Lncit*).

For structure, we refer to the methodology of Zhao and Wu (2024). We calculate the proportion of employees in the information transmission, computer service and software industries and those in the research, technical services and geological exploration industries relative to the total workforce in a city. This proportion (*Rdpera*) serves as a proxy variable for measuring urban talent agglomeration levels.

The regression results are presented in Columns (1) and (2) of Table 10. The coefficients for *Trust* are positive and significant at the 1% level, indicating that the implementation of the social credit system has effectively contributed to regional talent agglomeration. This finding underscores the notion that enhancements in social credit systems facilitate information exchange and sharing, thereby reducing the costs associated with information dissemination. Such improvements promote the flow of knowledge and technology, which are essential for addressing the development of AI.

Furthermore, the establishment of a social credit system standardizes the behaviors of various credit entities through mechanisms that promote adherence to commitments and impose penalties for violations. This not only signals favorable policy intentions but also cultivates an environment conducive to talent development. Consequently, the agglomeration of talent fosters the advancement of regional AI initiatives.

The information transmission and computer service industry talents (such as algorithm engineers and data scientists) attracted by social credit are the core of AI development, as they possess key technologies such as deep learning and natural language processing. When they gather, they accelerate AI applications through knowledge spillover (such as technical exchanges between enterprises), which leads to faster implementation of AI. The mobility rate of AI positions in pilot cities is 24% lower than that in non-pilot cities (data from Zhaopin.com), and the stable talent pool ensures the continuous iteration of AI algorithms (such as the optimization of intelligent customer service systems). As shown in Table 10, the coefficient is positive and significant, verifying this link.

Our analysis demonstrates that the construction of a social credit system significantly enhances both the scale and structure of talent agglomeration. By improving the environment for talent cultivation and facilitating knowledge exchange, the social credit system plays a vital role in supporting the development of AI within firms.

Table 10
Mechanism test-talent agglomeration.

Variable	*Rdpera*	*LnCit*	*Lnpatents*	*Lnpatents*
	(1)	(2)	(3)	(4)
Trust	2.088***	0.199***	0.022***	0.018***
	(7.03)	(5.14)	(3.12)	(2.78)
Rdpera			0.035***	
			(3.85)	
LnCit				0.041***
				(4.21)
Size	0.058	0.773***	0.156***	0.149***
	(0.39)	(30.56)	(12.89)	(12.15)
FirmAge	−1.364*	0.502***	0.387***	0.372***
	(−1.83)	(3.06)	(4.15)	(3.98)
Lev	−2.537***	−0.138	−0.192***	−0.185***
	(−3.39)	(−1.22)	(−3.76)	(−3.61)
Cashflow	−4.448***	−0.036	0.463***	0.447***
	(−3.38)	(−0.17)	(4.52)	(4.35)
Dual	0.517**	−0.015	0.023	0.021
	(2.07)	(−0.39)	(1.05)	(0.98)
Indep	2.007	0.583	0.185	0.172
	(0.81)	(1.40)	(0.89)	(0.83)
Board	−0.034	0.269**	0.092	0.088
	(−0.05)	(2.20)	(1.38)	(1.31)
Top10	−2.635***	−0.517***	−0.225***	−0.218***
	(−3.10)	(−3.61)	(−3.02)	(−2.91)
SA	0.124	0.757***	0.532***	0.516***
	(0.13)	(3.65)	(4.23)	(4.07)
Growth	0.680***	−0.134***	−0.042***	−0.040***
	(3.82)	(−4.96)	(−3.64)	(−3.48)
Mshare	2.747***	−0.099	0.231***	0.224***
	(3.57)	(−0.86)	(4.01)	(3.85)
Mfee	0.017	0.012	0.015	0.013
	(0.59)	(0.41)	(0.62)	(0.55)
INST	−0.443	0.048	0.052	0.049
	(−0.76)	(0.54)	(0.57)	(0.54)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	21,147	21,147	21,147	21,147
adj. R^2	0.445	0.523	0.558	0.551

Notes: This table tests the talent agglomeration mechanism. Rdpera is the proportion of employees in tech-related industries (information transmission, R&D) to total urban employment; LnCit is the log of employment in tech-related industries. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

8.3. Regional innovation for AI development

AI is inherently linked to innovation, as the emergence and transformation of innovative developments contribute to the establishment of unique competitive advantages for entrepreneurs and new ventures. These dynamics foster the latent developmental potential of AI, and the protection of innovative achievements is vital for ensuring the sustainability of entrepreneurial growth. This study constructs two key variables to examine the mechanisms underpinning regional innovation.

First, consistent with the literature, we use the number of patents granted per 10,000 individuals in a given region as a proxy for measuring *regional innovation development*. This metric is widely recognized as an accurate indicator of the level of innovation produced within an area. Second, following the methodology suggested by Nie (2024), we quantify the level of *intellectual property protection* by calculating the ratio of

intellectual property cases in a region relative to the region's gross domestic product. This ratio serves as a measure of the effectiveness of intellectual property protection in promoting innovation.

Social credit helps protect intellectual property rights (such as by decreasing patent infringement cases) and promotes the circulation and sharing of regional AI patents—the AI patent citation rate in the pilot cities is 16.5% higher than that in the non-pilot cities (according to the IRPDB database). Enterprises can develop AI applications (such as intelligent monitoring systems) based on existing patents (such as image recognition technology), thus reducing R&D costs. At the same time, social credit promotes the formation of AI industrial clusters (the number of AI enterprises in pilot cities is 31% higher than that in non-pilot cities), and data sharing among enterprises (such as supply chain data) enhances the effectiveness of AI applications. As shown in Table 11, the coefficient is positive and significant, verifying this link.

Table 11
Mechanism test-regional innovation.

Variable	*Innovation development*	*Intellectual property protection*	*Lnpatents*	*Lnpatents*
	(1)	(2)	(3)	(4)
Trust	0.548***	0.418***	0.019***	0.016***
	(4.15)	(7.30)	(2.65)	(2.40)
Innovation Development			0.028***	
			(3.20)	
Intellectual Property Protection				0.032***
				(3.80)
Size	−0.446***	0.004	0.135***	0.128***
	(−7.00)	(0.12)	(10.50)	(9.80)
FirmAge	−0.790**	−0.306*	0.352***	0.338***
	(−2.37)	(−1.76)	(3.90)	(3.72)
Lev	−1.040***	0.117	−0.178***	−0.165***
	(−3.15)	(0.80)	(−3.50)	(−3.20)
Cashflow	−1.422**	1.813***	0.425***	0.408***
	(−2.09)	(6.86)	(4.10)	(3.90)
Dual	0.288**	0.172***	0.021	0.019
	(2.42)	(3.58)	(0.98)	(0.91)
Indep	1.381	−1.682***	0.168	0.155
	(1.36)	(−3.32)	(0.82)	(0.76)
Board	0.256	−0.273*	0.085	0.081
	(0.84)	(−1.85)	(1.25)	(1.19)
Top10	−1.007***	−0.162	−0.205***	−0.198***
	(−2.66)	(−0.86)	(−2.80)	(−2.60)
SA	−0.295	−0.223	0.487***	0.472***
	(−0.66)	(−1.00)	(3.85)	(3.68)
Growth	0.281***	−0.011	−0.037***	−0.035***
	(2.92)	(−0.39)	(−3.35)	(−3.18)
Mshare	1.250***	0.841***	0.215***	0.207***
	(3.45)	(5.68)	(3.70)	(3.55)
Mfee	0.019	0.014	0.012	0.010
	(0.77)	(0.53)	(0.50)	(0.42)
INST	−0.145	0.123	0.048	0.045
	(−0.51)	(1.08)	(0.53)	(0.50)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	21,147	20,425	21,147	20,425
Adj. R^2	0.250	0.450	0.283	0.485

Notes: This table tests the regional innovation mechanism. Innovation Development is patents granted per 10,000 people; Intellectual Property Protection is the ratio of IP cases to regional GDP. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

The regression analysis incorporating the regional innovation development level is presented in Column (1) of Table 11, and the analysis that includes the level of regional *intellectual property protection* is shown in Column (2) of Table 11. The results indicate that the strengthening of scientific research integrity and intellectual property protection, facilitated by the construction of a social credit system, significantly supports regional innovation and development. This enhanced framework is crucial for invigorating the vitality of regional AI development. By fostering a robust environment for innovation and safeguarding intellectual property, the social credit system plays a pivotal role in stimulating both the creative and entrepreneurial aspects of AI initiatives.

Our findings suggest that the social credit system not only enhances regional innovation but also strengthens intellectual property protection. These factors collectively contribute to a more dynamic environment for the development of AI at the firm level.

9. Heterogeneity analysis

9.1. Degree of market integration

In the context of advancing market integration, various factors such as fiscal decentralization and the phenomenon of "political tournaments" driven by economic growth contribute to intense inter-regional competition. This competition often results in significant local protectionism and market segmentation at the provincial level—conditions that impede the cross-regional flow of goods, capital and talent, create artificial trade barriers and elevate transaction costs for enterprises. These frictions not only distort resource allocation but also pose unique challenges to regional AI development, as AI innovation relies heavily on the free movement of critical inputs (e.g., technical talent, R&D funds) and the sharing of data resources.

As shown in Columns (1) and (2) of Table 12, the heterogeneous effects of the social credit system pilot policy on AI development are strongly tied to regional market integration levels: the policy exerts a far more pronounced promotional effect in regions with low market integration. Specifically, in the low market integration group, the coefficient of the core explanatory variable *Trust* (social credit pilot designation) is 0.063 ($p < 0.1$), indicating that enterprises in these regions have, on average, 6.3% more AI patents than their counterparts in non-pilot regions. In contrast, the coefficient of *Trust* in the high market integration group is 0.029 ($p > 0.1$), which is not statistically significant and less than half the magnitude of the low integration group's coefficient.

This heterogeneity can be explained by the compensatory role of social credit systems in underdeveloped market environments. In regions with low market integration, information asymmetry is more severe, and resource mobility is highly constrained. For example, local protectionism may limit cross-regional bank lending, or talent may be reluctant to relocate due to unclear institutional guarantees. These barriers significantly increase the costs of AI development, as enterprises struggle to obtain sufficient financing for algorithm R&D and to recruit core technical personnel (e.g., data scientists).

The social credit system mitigates these frictions by establishing a cross-regional credit information sharing mechanism, standardizing market transaction rules and reducing information asymmetry between market participants. For instance, by improving enterprise credit ratings, the system enables firms in low integration regions to secure inter-regional loans for AI equipment procurement; by enhancing transparency in talent mobility (e.g., verifying professional credentials via credit platforms), it also facilitates the agglomeration of AI-related human capital. These improvements directly address the core bottlenecks of AI development in low integration regions, thereby amplifying the policy's effectiveness.

In summary, the social credit system pilot policy serves as a critical market supplement in regions with low market integration: by reducing transaction costs and breaking down resource flow barriers, it compensates for the deficiencies of underdeveloped market mechanisms and improves AI development potential. In high market integration regions, where existing market institutions already support efficient resource allocation, the marginal contribution of the social credit system to AI development is relatively weaker. This finding highlights the importance of tailoring institutional policies to regional market conditions when promoting technological innovation.

Table 12
Heterogeneity test-marketization level.

	Lnpatents	*Lnpatents*
	(1) High marketization level	*(2) Low marketization level*
Trust	0.029	0.063*
	(1.09)	(1.71)
Size	0.154***	0.133***
	(8.81)	(8.29)
FirmAge	0.366***	0.403***
	(2.94)	(3.56)
Lev	−0.171***	−0.222***
	(−2.63)	(−3.24)
Indep	0.143	0.208
	(0.53)	(0.74)
Board	0.064	0.121
	(0.76)	(1.36)
Top10	−0.204**	−0.248***
	(−2.07)	(−2.67)
SA	0.483***	0.595***
	(2.89)	(3.74)
Dual	0.021	0.022
	(0.84)	(0.73)
Growth	−0.051***	−0.025*
	(−3.07)	(−1.81)
Cashflow	0.556***	0.283**
	(4.04)	(2.16)
Mshare	0.257***	0.145*
	(3.53)	(1.80)
Mfee	−0.057	0.029
	(−0.47)	(1.51)
INST	0.093	0.223***
	(1.54)	(3.74)
Year FE	*Yes*	*Yes*
Firm FE	*Yes*	*Yes*
N	12,332	8813
Adj. R^2	0.199	0.219

This table reports the results of the heterogeneity analysis. "*High marketization level*" *and "Low marketization level*" *are g*rouped by the "Marketization Index" (Liu et al., 2024). When the marketization index is higher than the median, it is defined as the "high marketization level", and when the marketization index is lower than the median, it is defined as the "low marketization level". Variable definitions are reported in Appendix A. All regressions include firm and year fixed effects. The t statistics reported in parentheses are based on standard errors clustered at the firm level. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

9.2. Heterogeneity based on industry type

Industries differ substantially in their dependence on external institutional environments and resource allocation patterns, which leads to heterogeneous responses to social credit system construction in terms of AI development (Abulibdeh et al., 2024). Drawing on the National Bureau of Statistics of China's industrial classification standards, this study divides the sample into two broad categories to avoid potential controversies from overly fine-grained segmentation: the service industry (tertiary industries) and non-service industries (primary and secondary industries). The service industry mainly includes sectors such as wholesale and retail, finance, information transmission, software and information technology services and resident services—fields

characterized by high trust intensity and heavy reliance on information sharing and customer credibility. Non-service industries cover primary industries (e.g., agriculture, forestry, animal husbandry, fishery) and secondary industries (e.g., manufacturing, construction, mining), which are more dependent on physical assets, production chains and factor inputs such as land and energy, with relatively lower direct demand for institutional trust in daily operations.

As shown in Columns (1) and (2) of Table 13, the core explanatory variable *Trust* (social credit system pilot) has a positive and significant coefficient in both industry groups, indicating that social credit construction promotes AI development across industries. However, the magnitude and significance of the effect differ markedly: in the service industry group, the coefficient of *Trust* is 0.062 ($p < 0.01$), meaning that enterprises in social credit pilot cities have, on average, 6.2% more AI patents than those in non-pilot cities. This difference is statistically significant at the 1% level. In contrast, the coefficient of *Trust* in the non-service industries group is 0.021 ($p > 0.1$), which is not statistically significant and only about one third of the size of the coefficient for the service industry.

Table 13
Heterogeneity test- industry type.

Variable	*Lnpatents*	*Lnpatents*
	Service Industry(Tertiary Industry)	Non-Service Industries (Primary & Secondary Industries)
Trust	0.062***	0.021
	(3.85)	(1.52)
Size	0.148***	0.125***
	(10.26)	(9.81)
FirmAge	0.376***	0.321***
	(4.02)	(3.65)
Lev	−0.189***	−0.162***
	(−3.72)	(−3.48)
Cashflow	0.453***	0.398***
	(4.38)	(4.10)
Dual	0.022	0.019
	(1.01)	(0.95)
Indep	0.179	0.156
	(0.88)	(0.82)
Board	0.091	0.078
	(1.30)	(1.22)
Top10	−0.221***	−0.195***
	(−2.95)	(−2.78)
SA	0.526***	0.475***
	(4.15)	(3.92)
Growth	−0.040***	−0.035***
	(−3.51)	(−3.30)
Mshare	0.228***	0.196***
	(3.89)	(3.67)
Mfee	0.018	0.014
	(0.72)	(0.58)
INST	0.051	0.046
	(0.56)	(0.53)
Year FE	Yes	Yes
Firm FE	Yes	Yes
N	12,186	18,352
Adj. R^2	0.235	0.182

This table reports the results of the heterogeneity analysis. ''Service industry'' and ''Non-Service Industries'' are grouped by *Industrial Classification of National Economy* (GB/T 4754-2017): ''Service Industry'' (tertiary industry) include wholesale/retail, finance, IT services, etc., ''Non-Service Industries'' include primary (agriculture) and secondary (manufacturing, construction) industries. Variable definitions are reported in Appendix A. All regressions include firm and year fixed effects. The t statistics reported in parentheses are based on standard errors clustered at the firm level. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

The underlying mechanism for this heterogeneity lies in the degree of matching between social credit functions and industry-specific needs for AI development. For the service industry, AI applications (e.g., intelligent customer service, credit risk control, precision marketing) are highly dependent on reliable information and trust endorsement—two core outputs of the social credit system. Social credit reduces information asymmetry between service providers and customers (e.g., through public enterprise credit ratings), which lowers the trial cost of AI-enabled services (e.g., consumers are more willing to use AI financial advisory tools backed by credit guarantees) and helps service enterprises obtain targeted resources (e.g., credit-based loans for AI system upgrades).

For non-service industries (e.g., manufacturing), AI development is more focused on production-side optimization (e.g., intelligent quality inspection, automated assembly lines), and it relies heavily on physical investments (e.g., industrial robots, sensor equipment) and supply chain stability. Although social credit may indirectly alleviate financing constraints for these enterprises, its marginal contribution to AI development is weaker compared with the service industry, as the core bottleneck for non-service industries' AI adoption is often technical compatibility with existing production lines rather than trust or information barriers. Although Geng et al. (2025) find that customer stability indeed generates supply chain spillover effects for non-service firms (e.g., through trade credit extension), the reliance on physical assets and production efficiency priorities makes AI development less sensitive to the trust and customer relationship mechanisms strengthened by social credit systems.

In conclusion, the social credit system pilot policy exerts a more prominent and statistically robust promotional effect on AI development in the service industry than in non-service industries. These findings highlight that the effectiveness of informal institutional policies (e.g., social credit construction) in driving technological innovation is closely tied to the trust sensitivity of the industry, providing targeted insights for formulating industry-specific AI promotion strategies.

9.3. SOEs vs. non-SOEs

State-owned enterprises (SOEs) and non-SOEs in China differ greatly in financing constraints, resource access and operational scales, which may lead to heterogeneous responses to social credit system construction in AI development. Columns (1) and (2) of Table 14 report the regression results for SOEs and non-SOEs, respectively.

The results show that the promotional effect of social credit system construction on AI development is more pronounced for non-SOEs: the coefficient of *Trust* for the non-SOE group is 0.059 ($p < 0.05$), indicating that non-SOEs in social credit pilot cities have 5.9% more AI patents on average than those in non-pilot cities. In contrast, the coefficient of *Trust* in the SOE group is 0.004 ($p > 0.1$), which is not statistically significant.

This heterogeneity arises from differences in resource dependence: SOEs typically enjoy preferential access to financing (e.g., policy-based loans) and stable talent pools because of their institutional advantages, so they are less reliant on the credit enhancement and information sharing functions of the social credit system. In contrast, non-SOEs often face severe financing constraints and information asymmetry (e.g., difficulty proving creditworthiness to banks), making the social credit system's role in improving credit ratings and reducing transaction costs far more critical for their AI development (e.g., securing funds for AI algorithm R&D, attracting technical talent).

9.4. Corporate governance

Drawing on Yin et al. (2022), we examine how corporate governance levels (measured by the Dibbo Internal Control Index, IC) affect the promotional effect of social credit system demonstration city construction on enterprise AI development. The grouped regression results are shown in Table 15.

The coefficient of *Trust* is positive for both groups, but its significance and magnitude differ considerably. In the low corporate governance group ($IC = 0$), the coefficient of *Trust* is 0.057 ($p < 0.05$), which is statistically significant and larger in magnitude. This indicates that the construction of social credit system demonstration cities has a more prominent promotional effect on AI development for enterprises with weaker internal con-

Table 14
Heterogeneity test- SOEs vs non-SOEs.

	Lnpatents	
	(1) SOEs	*(2) Non-SOEs*
Trust	0.004	0.059**
	(0.11)	(2.35)
Size	0.149***	0.136***
	(7.09)	(7.93)
FirmAge	0.461***	0.283**
	(2.97)	(2.35)
Lev	−0.404***	−0.102*
	(−4.01)	(−1.91)
Indep	0.192	0.122
	(0.55)	(0.55)
Board	0.133	0.024
	(1.17)	(0.35)
Top10	−0.515***	−0.076
	(−3.68)	(−0.90)
SA	0.621***	0.422**
	(3.04)	(2.45)
Dual	0.044	0.035*
	(0.92)	(1.66)
Growth	−0.015	−0.040***
	(−0.83)	(−2.87)
Cashflow	0.237	0.595***
	(1.28)	(5.06)
Mshare	0.346	0.227***
	(0.69)	(3.86)
Mfee	0.014	0.053
	(0.70)	(0.68)
INST	0.243**	0.083*
	(2.44)	(1.69)
Year FE	Yes	Yes
Firm FE	Yes	Yes
N	7458	13,686
adj. R^2	0.309	0.167

Notes: This table reports heterogeneity by ownership type. SOEs are firms controlled by central/local governments; non-SOEs are controlled by natural persons/private entities. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

trol. In the high corporate governance group (IC = 1), the coefficient of *Trust* is 0.032 ($p > 0.1$), which is not statistically significant.

This result can be explained by the complementary effect of informal and formal institutions. Enterprises with poor corporate governance often face higher information asymmetry (e.g., incomplete internal oversight, opaque financial reporting), making them more dependent on external institutional guarantees (such as the social credit system) to reduce risks for AI investors and talent. In contrast, enterprises with strong internal control already have effective internal oversight and information disclosure mechanisms, so the marginal value of the social credit system in promoting AI development is relatively limited.

9.5. Start-ups vs. mature firms

Firm age significantly impacts firms' technology development strategies. Start-ups, particularly those in the AI sector, often struggle with financing limitations and talent shortages, while mature firms benefit from internal resources and established foundations. This prompts an exploration of how social credit system construc-

Table 15
Heterogeneity test-corporate governance.

	Lnpatents	
	(1) IC = 1	*(2) IC = 0*
Trust	0.032	0.057**
	(1.31)	(2.15)
Size	0.148***	0.109***
	(10.82)	(7.06)
FirmAge	0.351***	0.307***
	(3.50)	(2.76)
Lev	−0.155***	−0.162***
	(−2.58)	(−2.93)
Indep	0.174	0.212
	(0.74)	(0.88)
Board	0.092	0.088
	(1.21)	(1.20)
Top10	−0.206***	−0.288***
	(−2.58)	(−3.07)
SA	0.488***	0.421***
	(3.51)	(2.84)
Dual	0.029	−0.000
	(1.28)	(−0.02)
Growth	−0.062***	−0.019
	(−4.52)	(−1.02)
Cashflow	0.482***	0.175
	(4.12)	(1.29)
Mshare	0.184***	0.331***
	(2.91)	(4.45)
Mfee	−0.024	0.017
	(−0.17)	(1.60)
INST	0.131***	0.161***
	(2.60)	(2.59)
Year FE	Yes	Yes
Firm FE	Yes	Yes
N	15,744	5400
adj. R^2	0.223	0.162

Notes: This table reports heterogeneity by corporate governance, measured by the Dibbo Internal Control Index (IC). High IC (IC = 1) means the index is above the sample median; low IC (IC = 0) means below the median. All observations are at the firm-year level. Standard errors are clustered at the firm level, with t-statistics in parentheses. ***, *, * indicate significance at the 1%, 5%, and 10% levels. Variable definitions are in Appendix A.

tion differently affects AI development among firms of varying ages to better understand policy effectiveness boundaries.

Following the classification criteria proposed by Crawford et al. (2018) and Wang et al. (2024a, 2024b), firm age is used as the basis for categorization. Start-up firms are firms aged ≤5 years, which typically include entities in the initial stages of AI R&D, as well as those expanding into the market. Mature firms, those aged >5 years, encompass well-established enterprises that are currently undergoing AI transformation. The dataset includes 4836 firm-year observations for start-ups (accounting for 22.8% of the sample) and 16,376 for mature firms (77.2%). This distribution mirrors real-world industry age profiles, ensuring the reliability of subsequent statistical analyses. The results are presented in Table 16.

As shown in Table 16, the promotional effect of social credit system construction on firms' AI development exhibits significant heterogeneity by firm age. For start-up firms, the coefficient of *Trust* is 0.068, significant at the 1% level ($t = 3.21$). This means that when we control for the other variables, the number of AI patents of start-up firms in cities piloting the social credit system is 6.8% higher on average than that of start-ups in non-

Table 16
Heterogeneity test- Start-up vs Mature.

Variable	*Lnpatents*	*Lnpatents*
	Start-up Firms (Age $\leq$ 5 Years)	Mature Firms (Age $>$ 5 Years)
Trust	0.068***	0.021
	(3.21)	(0.89)
Size	0.112***	0.156***
	(4.78)	(10.23)
FirmAge	0.205**	0.413***
	(2.15)	(4.56)
Lev	−0.198***	−0.172***
	(−3.01)	(−3.28)
Indep	0.245	0.158
	(0.92)	(0.71)
Board	0.087	0.093
	(1.05)	(1.32)
Top10	−0.226**	−0.201**
	(−2.31)	(−2.17)
SA	0.589***	0.476***
	(3.87)	(3.91)
Dual	0.035*	0.022
	(1.79)	(0.98)
Growth	−0.042**	−0.038**
	(−2.45)	(−2.93)
Cashflow	0.396***	0.468***
	(3.12)	(4.35)
Mshare	0.201**	0.198**
	(2.23)	(2.11)
Mfee	0.028	0.015
	(0.97)	(0.62)
INST	0.167**	0.132**
	(2.41)	(2.33)
Firm FE	Yes	Yes
Year FE	Yes	Yes
N	4836	16,376
Adj. R^2	0.235	0.201

Notes: This table reports the results of the heterogeneity analysis of "Start-up Firms" vs "Mature Firms", which is grouped by firm age. A sample firm is defined as "Start-up Firms" if its firm age is less than 5 years, and defined as "Mature Firms" if firm age is more than 5 years. Firm age is the number of years between the observation year and establishment year..''Urban Digitalization Level" is measured by "the proportion of urban digital economy scale in GDP" (data sourced from *China Digital Economy Development Report 2023*). Values in parentheses are t-statistics clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Definitions of control variables are consistent with the baseline regression (see Appendix A for details).

pilot cities, which is significantly higher than the baseline effect of the full sample (4.4%). The core reasons are as follows. The technology R&D (e.g., algorithm iteration, data annotation) and scenario implementation (e.g., smart product prototype development) of start-up AI firms are highly dependent on external resources. The social credit system improves the credit ratings of start-ups (e.g., including them on the whitelist of the Credit-Based Lending platform), alleviating investors' concerns about their high risk and information opacity

and helping them secure critical financing such as seed rounds and angel rounds. In addition, the improved urban environment (e.g., education and medical support) driven by social credit is more attractive to core AI talents such as algorithm engineers and data scientists, who are crucial for start-ups to break through technological bottlenecks. For mature firms, the coefficient of *Trust* is 0.021, which is not significant at the 10% level ($t = 0.89$). This indicates that the promotional effect of social credit on AI development of mature firms is weak. The main reasons are as follows. Mature firms usually have internal fund reserves (e.g., reinvestment of profits) and stable talent echelons, so they are less dependent on external financing and attracting urban talent. Meanwhile, the AI transformation of mature firms is mostly based on existing businesses (e.g., traditional manufacturing enterprises introducing AI quality inspection) with relatively mature technical paths, leading to limited demand for external support such as information transparency and risk reduction provided by the social credit system.

10. Conclusion

A well-functioning social credit system is crucial for the effective alignment of supply and demand, as it can significantly enhance AI development and enable finance to better serve the real economy. This study uses the pilot policy of social credit system construction as a quasi-natural experiment to analyze the impact of establishing a demonstration city for social credit system development on enterprise AI advancement and to explore the underlying mechanisms. The findings indicate that the construction of a social credit system significantly enhances AI development. This conclusion remains robust when we conduct validity tests using a multi-time DID model and a series of robustness checks. Mechanism testing reveals that the establishment of a social credit system stimulates enterprise AI development by improving the breadth and depth of financial support, leveraging the agglomeration effects of high-quality talent and enhancing the development and protection of regional innovation outcomes.

Heterogeneity analysis demonstrates that the impact of social credit system construction on enterprise AI development varies significantly according to urban market integration. Specifically, in regions with lower levels of market integration, the positive effects of the social credit system on enterprise AI development are more pronounced than in areas with higher levels of market integration.

Industry-specific analysis indicates that the construction of the social credit system primarily enhances entrepreneurial vitality within the service industry, non-SOEs, firms with low internal control and start-up firms. These findings highlight the critical role of social credit systems in promoting enterprise AI development across varying industries and firms.

This study enriches the discourse surrounding social credit, broadens the research perspective on the influence of the informal institutional environment on entrepreneurship and provides valuable policy insights for advancing the construction of social credit systems and, consequently, fostering the development of AI in regional enterprises.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A.

Variables	Definition
Lnpatents	Natural logarithm of (1 + number of AI-related patent applications of a firm in a year).
Lnwords	Natural logarithm of (1 + frequency of AI-related keywords in the full text of a firm's annual report).

Appendix A (*continued*)

Variables	Definition
Lnwords_MDA	Natural logarithm of (1 + frequency of AI-related keywords in the "Management Discussion and Analysis (MD&A)" section of a firm's annual report).
Trust	Dummy variable: =1 if the city where the firm is located is designated as a "National Social Credit System Demonstration City" in the current year or later; =0 otherwise.
Dialect	Interaction term of "number of Chinese sub-dialect slices in a city" and "Trust's policy time dummy variable (1 for policy implementation years, 0 otherwise)".
HHI	Herfindahl-Hirschman Index: Calculated as the sum of the squares of the proportion of each commercial bank's branches in a city to the total number of bank branches in that city (HHI = Σ(branch count of bank i/total branches)2).
Trust Inclusive Financial Index	Standardized sub-index measuring the coverage and utilization depth of trust-based inclusive financial services (including credit, investment, and risk-sharing services).
Rdpera	Proportion of employees in "information transmission, computer services, software industry" and "research, technical services, geological exploration industry" to the total urban employment.
LnCit	Natural logarithm of (1 + number of employees in the urban "information transmission, computer services, and software industry" in a year).
Innovation Development	Number of granted patents per 10,000 people in a region (city-level).
Intellectual Property Protection	Ratio of the number of intellectual property (IP) cases in a region to the region's annual GDP (IP cases/GDP).
Size	Natural logarithm of a firm's total assets at the end of the year (unit: RMB 100 million).
Lev	Ratio of a firm's total liabilities to total assets at the end of the year (Lev = Total Liabilities/Total Assets).
Cashflow	Ratio of a firm's net cash flow from operating activities to total assets at the end of the year (Cashflow = Operating Net Cash Flow/Total Assets).
FirmAge	Natural logarithm of (1 + number of years since the firm's establishment to the observation year).
Top1	Ratio of the number of shares held by the largest shareholder to the firm's total share capital at the end of the year.
Top10	Ratio of the total number of shares held by the top 10 shareholders to the firm's total share capital at the end of the year.
Dual	Dummy variable: =1 if the firm's chairman and general manager are the same person; =0 otherwise.
Indep	Ratio of the number of independent directors to the total number of directors on the firm's board.
Board	Natural logarithm of (1 + total number of directors on the firm's board).
SA	SA financing constraint index: Calculated as SA = −0.737 × Size + 0.043 × Size2 − 0.040 × Age, where "Size" is the natural logarithm of total assets (unit: RMB 100 million), and "Age" is the number of years since the firm's establishment.
Growth	Total asset growth rate: (End-of-year total assets − Beginning-of-year total assets)/ Beginning-of-year total assets.
Mshare	Ratio of the number of shares held by management (directors, supervisors, and senior executives) to the firm's total share capital at the end of the year.
Mfee	Ratio of the total annual compensation of a firm's management (directors, supervisors, and senior executives) to the firm's annual operating income.
INST	Ratio of the number of shares held by institutional investors (e.g., mutual funds, insurance companies) to the firm's total share capital at the end of the year.

Appendix B. The dictionary of AI keywords

AI; Computer vision; Image recognition; Map of knowledge; Smart Education; Augmented reality; Smart government; Feature extraction; Business intelligence; Smart pension; Support vector Machine (SVM); Representation of knowledge; Pattern recognition; Internet of Things; Man-machine dialogue; AI products; Human-computer interaction; Data mining; Smart Banking; Intelligent customer service; Virtual reality; Autonomous driving; Unmanned driving; Smart Finance; Big Data Marketing; Long and Short Term Memory (LSTM); Smart chip; Edge computing; Cloud computing; Deep neural network; AI chip; Deep learning; Identification of features; Smart insurance; Smart Retail; Smart Medical care; Intelligent transportation; Smart home; Recurrent neural network; Big data risk control; Robotic process automation; Wearable products; Big data platform; Enhanced Intelligence; Big data operation; Machine translation; Neural network; Synthesis of speech; Human-machine collaboration; Smart agriculture; Smart speaker box; Convolutional neural network; Question answering system; Reinforcement learning; Big data Analytics; Natural language processing; Big data management; Intelligent computing; Voice interaction; Machine learning; Biological identification; Speech recognition; Smart regulation; Intelligent investment advisory; Intelligent voice; Voice print recognition; Face recognition; An agent; Big data processing; Distributed computing; Smart sensor; Intelligent search; Intelligent environmental protection.

References

Abulibdeh, A., Zaidan, E., Abulibdeh, R., 2024. Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: challenges, opportunities, and ethical dimensions. J. Clean. Prod. 140527

Atanassov, J., 2013. Do hostile takeovers stifle innovation? Evidence from antitakeover legislation and corporate patenting. J. Financ. 68 (3), 1097–1131.

Bertrand, M., Mullainathan, S., 2003. Enjoying the quiet life? Corporate governance and managerial preferences. J. Polit. Econ. 111 (5), 1043–1075.

Brockman, P., El Ghoul, S., Guedhami, O., Zheng, Y., 2020. Does social trust affect international contracting? Evidence from foreign bond covenants. J. Int. Bus. Stud., 1–34

Chen, H., Zhang, M., Zeng, J., Wang, W., 2024. Artificial intelligence and corporate risk-taking: evidence from China. China J. Account. Res. 17 (3) 100372.

Cornée, S., Szafarz, A., 2014. Vive la différence: social banks and reciprocity in the credit market. J. Bus. Ethics 125, 361–380.

Corradini, C., 2022. Social trust and new firm formation: a regional perspective. Small Bus. Econ. 58 (1), 169–184.

Crawford, G.S., Pavanini, N., Schivardi, F., 2018. Asymmetric information and imperfect competition in lending markets. Am. Econ. Rev. 108 (7), 1659–1701.

Demirgüç-Kunt, A., Klapper, L., 2013. Measuring financial inclusion: explaining variation in use of financial services across and within countries. Brook. Pap. Econ. Act. 2013 (1), 279–340.

Diamond, D.W., 1989. Reputation acquisition in debt markets. J. Polit. Econ. 97 (4), 828–862.

Dutta, N., Meierrieks, D., 2021. Financial development and entrepreneurship. Int. Rev. Econ. Financ. 73, 114–126.

Erdélyi, O.J., Goldsmith, J., 2018. Regulating artificial intelligence: proposal for a global solution. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 95–101.

Fukuyama, F., 1995. Social capital and the global economy. Foreign Aff. 74, 89.

Garrett, J., Hoitash, R., Prawitt, D.F., 2014. Trust and financial reporting quality. J. Account. Res. 52 (5), 1087–1125.

Gehrig, T., Stenbacka, R., 2007. Information sharing and lending market competition with switching costs and poaching. Eur. Econ. Rev. 51 (1), 77–99.

Geng, Y., Wang, Y., Wu, S., 2025. Acquaintance means booster? Why stable customers matter for firm productivity. China J. Account. Res. 18 (3) 100435.

Gutiérrez-Nieto, B., Serrano-Cinca, C., Camón-Cala, J., 2016. A credit score system for socially responsible lending. J. Bus. Ethics 133, 691–701.

Horváth, R., 2013. Does trust promote growth? J. Comp. Econ. 41 (3), 777–788.

Kimmich, C., Wenzlaff, F., 2022. The structure–agency relation of growth imperative hypotheses in a credit economy. New Political Econ. 27 (2), 277–295.

Kong, G., Kong, T.D., Qin, N., Yu, L., 2023. Ethnic diversity, trust and corporate social responsibility: the moderating effects of marketization and language. J. Bus. Ethics 187 (3), 449–471.

Kostka, G., Antoine, L., 2020. Fostering model citizenship: behavioral responses to China's emerging social credit systems. Policy Internet 12 (3), 256–289.

Liang, F., Das, V., Kostyuk, N., Hussain, M.M., 2018. Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. Policy Internet 10 (4), 415–453.

Liu, B., Huang, W., Chan, K.C., Chen, T., 2022. Social trust and internal control extensiveness: evidence from China. J. Account. Public Policy 41 (3) 106940.
Liu, C., Kong, D., 2021. Business strategy and sustainable development: evidence from China. Bus. Strateg. Environ. 30 (1), 657–670.
Liu, D., Wang, Y., Li, M., 2024. Comparable but is it informative? Accounting information comparability and price synchronicity. J. Financ. Stab. 73 101297.
Mickiewicz, T., Rebmann, A., 2020. Entrepreneurship as trust. Found. Trends® Entrep. 16 (3), 244–309.
Ng, A., Ibrahim, M.H., Mirakhor, A., 2016. Does trust contribute to stock market development? Econ. Model. 52, 239–250.
Nicholls, A., 2010. Institutionalizing social entrepreneurship in regulatory space: reporting and disclosure by community interest companies. Acc. Organ. Soc. 35 (4), 394–415.
Nie, S., 2024. Does intellectual property rights protection matter for low-carbon transition? The role of institutional incentives. Econ. Model. 140 106842.
Padilla, A.J., Pagano, M., 1997. Endogenous communication among lenders and entrepreneurial incentives. Rev. Financ. Stud. 10 (1), 205–236.
Pagano, M., Jappelli, T., 1993. Information sharing in credit markets. J. Financ. 48 (5), 1693–1718.
Raisch, S., Krakowski, S., 2021. Artificial intelligence and management: the automation–augmentation paradox. Acad. Manag. Rev. 46 (1), 192–210.
Shapiro, C., 1986. Exchange of cost information in oligopoly. Rev. Econ. Stud. 53 (3), 433–446.
Wang, X., Han, L., Huang, X., 2020. Bank competition, concentration and EU SME cost of debt. Int. Rev. Financ. Anal. 71 101534.
Wang, Y., Cui, R., Gao, H., Lu, X., Hu, X., 2024a. Can the construction of the social credit system improve the efficiency of corporate investment? Int. Rev. Econ. Financ. 96 103510.
Wang, Z., Zhang, T., Ren, X., Shi, Y., 2024b. AI development rate and corporate green innovation efficiency: evidence from Chinese energy companies. Energy Econ. 132 107499.
Xiao, M., Chen, N., 2025. How does social trust promote enterprises' financialization? Int. Rev. Financ. Anal. 97 103819.
Yin, J.H., Song, H.Y., Zeng, K.X., 2022. Does a smart business environment promote corporate investment? A case study of Hangzhou. PLoS One 17 (7) e0269089.
Yu, L., Li, X., Tang, L., Zhang, Z., Kou, G., 2015. Social credit: a comprehensive literature review. Financ. Innov. 1, 1–18.
Zhang, C., 2020. Governing (through) trustworthiness: technologies of power and subjectification in China's social credit system. Crit. Asian Stud. 52 (4), 565–588.
Zhao, Q., Wu, S.Z., 2024. The deviation characteristics and influencing factors of urban scaling laws in China: industrial and sectoral perspective. Int. J. Urban Sci., 1–20
Zuo, J., Huang, C., Qiu, B., Mai, R., 2023. The construction of social credit system and corporate innovation: Evidence from China. Pacific-Basin Finance Journal 81, 102–116.

China Journal of Accounting Research 19 (2026) 100488

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Business groups and affiliated firms' digital transformation

Hui Xu [a], Haomin Wu [b], Wei Zheng [c], Kam C. Chan [d], Qiaoling Fang [c,*]

[a] Accounting/Audit School, Yunnan University of Finance and Economics, Kunming, China
[b] Journal Center, Yunnan University of Finance and Economics, Kunming, China
[c] School of Economics and Management, Fuzhou University, Fuzhou, China
[d] Research Center of Finance, Shanghai Business School, Shanghai, China

ARTICLE INFO

Article history:
Received 27 March 2025
Accepted 26 March 2026
Available online 16 April 2026

Keywords:
Business groups
Internal market
Digital transformation
Spillover effect

ABSTRACT

Considering the practical challenges of digital transformation (DT), this study examines whether business groups facilitate affiliate firms' DT through internal markets. Using a sample of Chinese firms, we find that group-affiliated firms exhibit higher levels of DT than comparable unaffiliated firms. The effect operates primarily through internal intangible markets—labor, technology and entrepreneurship—rather than internal capital markets. Central government-controlled business groups promote DT more effectively, whereas greater intra-group cultural diversity can impede DT. This impact is stronger under high product market competition, low economic policy uncertainty, low regional digitalization, low industrial structure upgrading and strong government support, and in provinces where DT is emphasized. Moreover, an affiliate's total factor productivity and Tobin's Q both increase after its DT is promoted by a business group.

© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rapid development of the digital economy worldwide (Goldfarb and Tucker, 2019), firms increasingly apply data-driven digital technologies to enhance their organization, investment and operations to maintain competitiveness and improve performance (Chen and Srinivasan, 2024). However, digitalization remains a complex task. Austin and Williams (2021) show that blockchain application, as a digitalization tool, is highly

* Corresponding author.
E-mail addresses: hui.xu@ynufe.edu.cn (H. Xu), whm_1977@126.com (H. Wu), wei.zheng@fzu.edu.cn (W. Zheng), Johnny.chan@wku.edu (K.C. Chan), qiaolingfang@fzu.edu.cn (Q. Fang).

https://doi.org/10.1016/j.cjar.2026.100488
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

complex and references to blockchain and technology-related critical audit matters prompt lower investment from non-professional investors. Similarly, Zhai et al. (2022) find that although corporate digital transformation (DT) is useful and enhances value, firms generally experience substantial difficulty and inefficiency in its implementation. In addition, a McKinsey and Company survey reports that large technology firms invested between USD 20 and 30 billion in artificial intelligence (AI) in 2016 alone.[1] While technology firms are familiar with digital technologies, non-technology firms face greater challenges in DT because it requires substantial investments in capital, human resources and technology. These obstacles are particularly pronounced among non-technology firms. In practice, many firms, particularly in emerging markets, belong to business groups. Thus, we argue that just as business groups operate internal capital markets (Claessens et al., 2006) and internal labor markets (Xu et al., 2023a), they also operate internal intangible markets related to technology and entrepreneurship.[2] Given that DT faces practical constraints associated with human capital and technology (Zhai et al., 2022), this study investigates whether a business group's internal intangible market facilitates DT adoption by affiliated firms. We hypothesize that business groups ease DT-related challenges by reallocating production factors, particularly labor, technology and entrepreneurship, from affiliated firms to a focal firm within the business group. Consequently, the focal firm exhibits a higher level of DT than an otherwise similar unaffiliated firm. In this study, we distinguish between tangible internal markets (capital) and intangible internal markets (labor, technology and entrepreneurship) because DT depends on both types of inputs.

We use a sample of Chinese firms for two reasons. First, unlike many other emerging markets, China's digital economy has developed rapidly. According to the Global Digital Economy Development Index Report (TIMG 2024), China's digital economy ranked 22nd worldwide in 2013 but rose to fifth place by 2023, surpassing several developed economies, including Australia, France, Canada, the Netherlands and Japan. Among emerging and developing economies, China is the only country ranked within the top 20 of the Digital Technology Index. Consistent with this rapid progress, China's Digital Technology Index score increased from 65.16 in 2013 to 84.46 in 2023. The rapid development of China's digital economy is strongly associated with government support. However, statistics from the Inspur Research Institute (2020) indicate that non-technology firms in China continue to lag behind in DT.

Second, previous studies indicate that business groups are widespread in emerging market economies as an institutional arrangement to address incomplete external markets (Khanna and Yafeh, 2007). Business groups are prevalent not only in Asian economies such as South Korea (Shin and Park, 1999), China (He et al., 2013) and India (Khanna and Palepu, 2000) but also in European civil-law countries such as Germany. Business groups improve internal capital allocation efficiency (He et al., 2013), promote research and development (R&D) and investment (Huang and Chen, 2011) and facilitate the reallocation of surplus labor among affiliated firms (Xu et al., 2023a). Hence, given that DT is particularly challenging for non-technology firms, it is crucial to examine how business groups in China use internal markets to enhance the DT of affiliated firms. Such analysis provides insights not only for China but also for other economies in which business groups are prevalent and for countries that continue to lag in DT.

Our findings reveal that business group-affiliated firms exhibit a higher level of DT than otherwise similar unaffiliated firms, supporting our hypothesis. This finding is robust across a battery of robustness tests and after accounting for endogeneity. Additionally, the mediation analysis indicates that business groups enhance affiliated firms' DT by promoting R&D investment, patent output and the hiring of executives with relevant educational and technological backgrounds. Hence, business groups primarily drive affiliated firms' DT by reallocating resources within their internal intangible markets rather than the tangible market.[3]

We additionally analyze four related aspects. First, we examine the strength of governmental support. From the perspective of state ownership, we find that the DT of firms affiliated with business groups controlled by the central government in China is more strongly promoted than that of firms affiliated with business groups controlled by local governments or non-state-owned business groups. Regarding governmental

[1] The relevant data are obtained from "Artificial intelligence the next digital frontier?," McKinsey Global Institute, 2017.

[2] In the economic literature, technology is typically embedded in the production function. "Entrepreneurship" refers to a value-creating factor of production in a firm.

[3] We consider labor, technology and entrepreneurship as intangible assets because of their heterogeneous nature. In contrast, capital is homogeneous and therefore tangible.

support, when an affiliated firm is located in a province that attaches greater importance to digitalization, receives more government subsidies or faces lower tax rates, the effect of the business group on the affiliated firm's DT is more pronounced. Second, we consider the characteristics of business groups. Greater cultural diversity in focal firms leads to cognitive bias, communication barriers and a lack of trust, which weakens the governance effect of a business group's internal intangible markets, and in turn, impedes the DT of business group-affiliated firms. In addition, a time difference exists in the spillover effects of the business group's DT. When the subsidiary in which the headquarters holds the highest equity stake undertakes DT first, this transformation can subsequently promote DT in other subsidiaries. Third, we examine the characteristics of firms' external environments. High product market competition, low economic policy uncertainty, low regional digitalization level or weak industrial structure upgrading cause business groups to exert a more pronounced effect on affiliated firms' DT. Fourth, we find that after a business group enhances an affiliated firm's DT, the affiliated firm's total factor productivity (TFP) and Tobin's Q both increase, suggesting that internal intangible markets are value enhancing.

This study makes several contributions to the literature. First, it extends the literature on corporate DT. Although previous studies extensively examine the effects of DT, including reductions in agency costs (Fang et al., 2023), enhancements in corporate competitive advantage (Benner and Waldfogel, 2023), improvements in corporate performance (Zhai et al., 2022; Chen and Srinivasan, 2024) and reductions in stock price crash risk (Wu et al., 2022), the contribution of business groups to their affiliates' DT remains underexplored. This study demonstrates that business groups facilitate firms' DT efforts.

Second, this study advances the literature on business groups. Previous studies mainly focus on internal capital markets within business groups in East Asia (Claessens et al., 2006), Korea (Shin and Park, 1999), China (He et al., 2013) and India (Khanna and Palepu, 2000). Some studies further show that business groups can reallocate resources through internal technology markets (Huang and Chen, 2011) or internal labor markets (Xu et al., 2023a). However, corporate decisions, especially risky ones, typically require a coordinated evaluation of multiple production factors. Our findings are consistent with those of Huang and Chen (2011) and Xu et al. (2023a) in demonstrating that internal intangible markets extend beyond internal capital markets within business groups. Specifically, we document the existence of internal intangible markets and show that the reallocation of internal resources, especially intangible resources, plays a positive role in the DT of business group-affiliated firms. Notably, we find that the internal capital market has a minimal effect on DT.

Finally, because business groups are common in both China and other emerging markets, there is strong demand for DT application. By leveraging internal intangible markets, aligning with government support, optimizing group governance and adapting to external contexts, business groups in other emerging markets can effectively overcome resource shortages and implementation challenges associated with DT, as demonstrated by Chinese business groups. This approach helps individual affiliates improve their DT capabilities and competitiveness and contributes to the overall development of the digital economy in emerging markets, thereby narrowing the gap with developed economies in the global digital landscape.

2. Literature review and hypothesis development

2.1. Business groups and internal markets

Business groups are common worldwide (Almeida and Wolfenzon, 2006) and play an important role in the economy, especially in emerging markets where external markets are less developed. When standalone firms form a group, the headquarters typically controls the strategy, finances and personnel of the group's subsidiaries. However, business groups have also been described as either "paradigms" or "parasites" (Khanna and Yafeh, 2007). Therefore, business groups have not only a bright side (Khanna and Tice, 2001) but also a dark side (Ozbas and Scharfstein, 2010).

The related literature mainly focuses on the internal capital markets of business groups. Evidence of the bright side shows that internal capital markets provide additional funding to group divisions (Claessens et al., 2006; He et al., 2013) and improve capital allocation efficiency in in business groups (Scharfstein and Stein, 2000). In contrast, the dark side is reflected in inefficient capital allocation through internal capital

markets (Ozbas and Scharfstein, 2010). Because of dual agency conflicts between group headquarters and affiliates (Wulf, 2009), parent firm dividend policies may increase tax avoidance among subsidiaries (Chen et al., 2025), and internal capital allocation may depend more on the power of divisional executives than on the expected returns of divisional investments (Glaser et al., 2013).

Research on business group internal markets also examines internal technology and labor markets. Huang and Chen (2011) and Cai et al. (2019) show that the technological development of group affiliates benefits from R&D investment by other affiliates through technological and knowledge spillover effects in business groups. An emerging body of literature suggests that business groups have an internal labor market. Horizontal employee mobility within a business group can smooth intra-group labor volatility (Faccio and O'Brien, 2021), thereby improving the group's overall labor productivity (Xu et al., 2023a).

2.2. Corporate DT

The literature on corporate DT can be broadly divided into two strands. The first concerns a firm's business environment. External factors, such as the development of digital technologies (Verhoef et al., 2021) or banking competition (Bai et al., 2024), can accelerate corporate DT. Internal factors, such as effective strategic management (Porfírio et al., 2021), employee digital quality (Kozanoglua et al., 2021) and CEO characteristics (Zhang et al., 2023a), also facilitate DT.

The second strand examines the economic implications of corporate DT. At the macro level, digital economy can increase national welfare (Torres and Augusto, 2020). At the micro level, corporate DT can reduce agency costs (Fang et al., 2023), exert a substitution effect on accounting employees (Chen et al., 2024), improve business operation efficiency (Zhang et al., 2023b) and enhance sustained corporate competitive advantage (Benner and Waldfogel, 2023). Through these channels, DT improves firm performance and valuation (Zhai et al., 2022; Chen and Srinivasan, 2024) and reduces stock price crash risk (Wu et al., 2022).

2.3. Business group-affiliated firms and DT

Because a business group's internal market has both bright and dark sides, we analyze its effect on a subsidiary's DT from two perspectives.

First, we examine the motivations underlying DT among group-affiliated firms. From an external institutional and economic perspective, the development of a country's digital economy is closely linked to its government's policies. The Chinese government considers digital economy as a core driver of future economic growth and has therefore actively promoted digital technologies. During our study period from 2008 to 2019, China's central and local governments introduced more than 420 policy initiatives related to the development of the digital economy and digital technologies. Policies such as the development plan for new-generation AI have strongly promoted digital economic development. Large business groups are often key targets of these policy initiatives. Business group-affiliated firms whose executives are appointed by group headquarters are more likely to follow government policies and are thus more strongly motivated to engage in DT.

From an internal motivational perspective, large business groups typically control two to four affiliated firms, with some groups holding up to 22 public firms in a given year. DT helps reduce agency costs and improve internal controls and corporate governance (Fang et al., 2023). The use of digital technologies, such as financial sharing centers, can enhance the group headquarters' control capacity and reduce operating costs and expenses (Na et al., 2024). Therefore, business groups have strong incentives to encourage their affiliated firms to engage in DT.

Second, we examine the ability of business groups to support DT. Large business groups often view DT as a strategic decision that faces substantial financial, human resource and technological barriers in external markets. However, a business group's internal market can serve as an effective mechanism for overcoming these barriers. In this setting, internal capital, labor, technology and entrepreneurship become "public resources" among business group-affiliated firms. In other words, these firms share their internal market resources.

When a group subsidiary requires production factors for DT, the group headquarters can reallocate these factors from other affiliates at a low cost, thereby alleviating external market barriers. Consequently, DT becomes more feasible for the focal firm. Based on this reasoning, we propose the following hypothesis:

H1. *The extent of DT in a business group-affiliated firm is greater than that in an otherwise comparable unaffiliated firm.*

2.4. Spillover effects of business groups' internal market and business group-affiliated firms' DT

Corporate production requires factors of production (FOPs). Traditional FOPs include capital, labor, land and entrepreneurship. In the internal market setting of a business group, land, including physical land and natural resources, is immobile. In addition, technology is typically embedded in the production function, whereas entrepreneurship is implicitly reflected in innovation activities. To streamline the discussion, we classify internal FOPs into tangible and intangible internal markets. The tangible internal market is relatively homogeneous, whereas the intangible internal market is heterogeneous. We argue that business groups can promote affiliated firms' DT through the productive spillover effects of internal tangible and intangible markets, as discussed below.

2.4.1. Capital spillovers

Corporate DT requires capital to support its transformation process. Financing costs in external markets are often high, particularly for high-risk investments. A business group's internal capital market can provide funding for affiliated firms' high-risk investments at a lower cost. Additionally, a business group headquarters can accurately identify and allocate funds to affiliated firms with stronger high-risk investment prospects. However, because DT is a distinct type of high-risk investment that typically attracts attention, it is likely to be supported strategically by affiliated firms themselves, which makes it easier for them to obtain capital from external sources. Capital, as a homogeneous production factor, may therefore be less critical to an affiliated firm's DT. Accordingly, an affiliated firm's DT does not necessarily require capital from its internal capital markets. Based on these considerations, we propose two competing hypotheses:

H2a. *A business group facilitates its affiliated firm's DT through capital spillovers in its internal capital market.*

H2b. *A business group facilitates its affiliated firm's DT but not through capital spillovers in its internal capital market.*

2.4.2. Technology and entrepreneurship spillovers

Firms, especially those in traditional industries, often face substantial technological barriers during DT because digitalization requires the effective application of new technologies such as big data and AI in their operations. A business group's internal technology market can reduce technology acquisition costs (Claessens and Fan, 2002), help overcome technological barriers (Huang and Chen, 2011) and alleviate information scarcity (Cai et al., 2019). The challenges associated with applying new technologies make it time-consuming and costly for affiliated firms to develop these technologies themselves or acquire them from external markets. In contrast, a business group's internal technology market can generate digital technology spillover effects at a lower cost, thereby accelerating the DT process of the focal firm.

Moreover, the application of technology requires an innovative mindset and involves uncertainty in its execution. Kim and Oh (2024) show that corporate entrepreneurship is critical for aligning the DT process with firm growth. Similarly, Yunis et al. (2018) reveal that corporate entrepreneurship contributes to improved technological innovation and organizational performance. Haier, a representative Chinese business group, has integrated the R&D resources and digital technologies of its home appliance, equipment manufacturing, smart home and related businesses into a shared platform. Haier's smart home appliance R&D unit shares its accumulated Internet of Things sensor technologies and user behavior data analysis models with other business units within the group, generating technology spillovers and enabling affiliated firms to deploy digital solutions more rapidly. We argue that entrepreneurial executives are better able to allocate budgets, arrange personnel and formulate strategies for the DT process. Accordingly, we propose the following hypothesis:

H3. *A business group facilitates its affiliated firm's DT through its internal entrepreneurship and technology markets.*

2.4.3. Human resource spillovers

DT depends on both digital technologies and adequate technical personnel. Claessens and Fan (2002) argue that when external markets are weak, the contract compliance rate is low. In such settings, business groups can provide public goods to affiliated firms through internal labor markets. For example, the Hyundai Group has established training centers and research institutes to develop skilled workers for the entire group (Khanna and Yafeh, 2007). Depending on business needs, labor mobility between affiliated firms within the group can be facilitated through the internal labor market (Xu et al., 2023a). Recruiting digitally skilled personnel from external markets is expensive. In contrast, labor sharing within business groups can improve the efficiency of digital labor reallocation and facilitate affiliated firms' DT. Sinopec, a leading centrally controlled state-owned business group in China, introduced human resource sharing as early as 2015 and achieved full coverage of human resource management and control services across its subsidiaries. By integrating human resource data, the group has implemented human resource analytics and centralized workforce planning. This approach improves the group's management efficiency, supports decision-making and human resource innovation and enhances the group's DT. Based on this reasoning, we propose the following hypothesis:

H4. *A business group facilitates its affiliated firm's DT through its internal labor market.*

3. Research methods

3.1. Data sources and sample selection

After the 2008 global financial crisis, China's economic growth model changed. In line with DT-related policies, some economically developed provinces and cities in China began implementing digital technologies in 2008. Therefore, in our study, we select Chinese listed firms from 2008 to 2022 as the initial sample. The final sample excludes firms in the financial industry, observations with missing data, firms designated as ST (special treatment) or PT (pause trading) and firms with abnormal financial conditions, defined as asset–liability ratios greater than one. The resulting sample comprises 33,500 firm-year observations, of which 8754 are affiliates with business groups. We manually collect data on listed firms affiliated with business groups and their controlling headquarters. All other financial and governance data are obtained from the China Securities Market & Accounting Research database.

3.2. Variable measurement

3.2.1. DT

Some studies perform textual analysis of financial reports to measure firms' DT levels. For instance, Chen and Srinivasan (2024) measure the extent of a firm's DT by extracting DT-related keywords from audited annual reports (10-K reports). Following this approach, we construct our main measures of firms' DT as follows:

First, we comprehensively identify digital technology categories based on research reports and academic studies from various research institutes, practitioners and scholars in recent years.[4] We classify digital technologies into 10 categories: intelligence, digitalization, automation, AI, machine learning, cloud computing, big data, blockchain, the Internet and the Internet of Things.

[4] The main research reports include "Analysis of the Digital Economy Situation in 2020–2021" (National Research Center for Industrial Information Security and Development); "White Paper on Future Economy 2021"; "Digital Economy Industry Development Index" (Big Data Analytics and Technology Innovation Center, Beijing Big Data Research Institute); "The 2021 Statistical Classification of the Digital Economy and Its Core Industries" (National Bureau of Statistics); "The 2019 Accenture China Enterprise Digital Transformation Index" (Accenture); "China Digital Economy Report" (McKinsey); and others. The referenced academic papers are listed in the reference section. Government policy documents include the 12th, 13th and 14th Five-Year Development Plans of the State Council and local governments, government work reports, and various central and local government policy documents, such as the Guiding Opinions on Actively Promoting the Action of Internet Plus, the Industrial Plan for Big Data Development and the New Generation Artificial Intelligence Development Plan. We use these research reports and documents to determine the categories of digital technology.

Second, we manually develop a DT keyword dictionary. Given the complexity of Chinese sentences, instead of using machine learning methods, we identify keywords through manual reading based on the established categories of digital technologies. Using the management discussion and analysis section of annual reports, we examine disclosures related to the 10 selected digital technology categories and identify DT-related keywords. Specifically, we organize two groups of research assistants. Each group manually reviews 100 text segments from listed firms and assigns relevant word roots appearing in the texts to different digital technology categories. The vocabularies generated by the two groups are then consolidated, after which the other group adds or removes word roots as appropriate by referring to existing research reports, academic references and Internet dictionaries. The final set of word roots for the 10 digital technology categories constitutes the digital technology information dictionary (Table 1).

Third, we conduct textual analysis using Python and the Jieba Chinese word segmentation algorithm. We incorporate the manually screened word roots into the Jieba thesaurus and apply the algorithm to segment relevant text in the annual reports of listed firms from 2008 to 2022. For each firm-year observation, we then calculate the frequency of words associated with each digital technology category.

Fourth, textual information on digital technology extracted from annual reports does not represent a firm's actual DT activities because firms may engage in strategic disclosure to attract external market attention. In addition, DT is a dynamic process. Measures based only on textual analysis therefore need to account for the degree of digitalization as well as its dynamic and sustained implementation over time. To address these concerns, we further refine the DT variables. The main purpose is to remove data that do not reflect a firm's actual implementation of DT. Hence, we exclude digital-related content in annual reports that pertains to forward-looking development plans, the economic environment and industry development as well as disclosures concerning the DT implementation of the firm's upstream suppliers and downstream customers. This procedure ensures that the retained text indicates the firm's own DT activities.

Based on this refined text, we define a dummy variable, *Digital*, which equals 1 if the firm has implemented DT and 0 otherwise. Furthermore, we distinguish the application of digital technologies across different segments of the firm by constructing three additional dummy variables: *Digital_P* for product-related DT, *Digital_R* for R&D-related DT and *Digital_O* for operational DT.

3.2.2. Business group affiliation

A business group is generally defined as a set of independent firms bound by formal or informal ties, most commonly through ownership (e.g., Khanna and Rivkin, 2001; Claessens et al., 2006). Following previous studies on Chinese business groups (e.g., He et al., 2013; Fang et al., 2017), we identify group-affiliated firms using ownership structure charts disclosed in firms' annual reports. We define a business group as one in which a common shareholder controls at least two listed firms. Consistent with Fang et al. (2017), we require a business group to control at least two listed firms with available financial data. Appendix A summarizes the variable definitions.

Table 1
Word Roots' List of Digital Transformation.

Dimension	Composition
Intelligence	8 roots of words related to Intelligent.
Digitalization	15 roots of words related to digital technology.
Automation	10 automation-related word roots.
Artificial Intelligence (AI)	15 roots of words related to artificial intelligence.
Machine Learning	16 roots of words related to machine learning.
Cloud Computing	8 roots of words related to cloud computing.
Big Data	24 roots of words related to big data.
Blockchain	5 roots of words related to blockchain.
Internet	1 root word related to Internet.
Internet of Things (IoT)	1 root word related to IoT.

This table presents word roots' list of digital transformation, which including ten categories of digital technologies.

3.3. Regression models

3.3.1. Business group affiliation and DT

Using Eq. (1), we test the impact of business group affiliation on firms' DT (i.e., H1).

$$Digital_{i,t} = \beta_0 + \beta_1 Group_{i,t} + \beta_i Control + Industry + Year + Region + \varepsilon_{i,t} \quad (1)$$

where *Digital* is an indicator variable that equals 1 if the firm undertakes substantial DT and 0 otherwise. To further examine the dimensions of firms' DT influenced by business group affiliation, we distinguish the application of digital technologies across product (*Digital_P*), R&D (*Digital_R*) and operational (*Digital_O*) segments. *Group*, the main explanatory variable, equals 1 if the listed firm belongs to a business group and 0 otherwise. Following Bai et al. (2024), we include a set of DT-related control variables capturing firm financial characteristics, including the nature of property rights (*SOE*), the natural logarithm of total assets (*Size*), the debt-to-total assets ratio (*Lev*), return on assets (*Roa*), firm age (*EstAge*), the intangible asset ratio (*Intaratio*), capital intensity (*CapiInte*), the current asset turnover ratio (*Current*), annual sales growth (*Growth*), the percentage of shares held by the largest shareholder (*Top1*), size of the board of directors (*Boardsize*), the proportion of independent directors (*Ind*), the management shareholding ratio (*Mshare*) and whether the firm is audited by a top four accounting firm (*Big4*). We also include industry, year and region fixed effects in Eq. (1) to control for factors specific to an industry, year or region. Following Guan et al. (2016), we classify industries according to the China Securities Regulatory Commission classification system, using one- and two-digit codes for non-manufacturing and manufacturing sectors, respectively. *Region* is a dummy variable indicating whether a firm is located in eastern, central or western China. Using logistic regression, we adjust standard errors for the clustering effect at the firm level in Eq. (1). To exclude the influence of outliers, we winsorize all continuous variables at the top and bottom 1% levels.

Because the effect of business group affiliation on DT may occur with a time lag, we lag *Group* by one to three periods. We expect the coefficients of *Digital* and *Group* to be significant and positive, indicating that business group-affiliated firms are more likely to undergo DT.

3.3.2. Business group affiliation's path analysis and DT

From H2 to H4, we argue that business group affiliation promotes DT through several spillover effects of internal tangible and intangible production factors. Hilary et al. (2016) suggest that a mediation model can be used as a form of path analysis. Path analysis uses a structural equation model to determine how a source variable affects an outcome variable through both direct and indirect paths via mediating variables. Accordingly, we use a mediation approach to analyze the path through which spillovers of internal tangible and intangible production factors link business group affiliation to DT.

We test whether specific production factors exert a spillover effect within business groups that promote DT through fund allocation (capital), technological R&D (technology and entrepreneurship) and human resource optimization (labor). We use the cash dividend level (*Dividend*) and interest coverage ratio (*ICR*) as proxies for internal capital factor mediating variables, R&D expenditure (*RD*) and patent output (*Patent*) as proxies for technology and entrepreneurship factor mediating variables, and executive education (*EDU*) and executives' technological background (*Manage_tech*) as proxies for human factor mediating variables.

Following Baron and Kenny (1986) and the mediation effect testing approach of Hayes (2018), we formulate Eq. (2) to examine whether business groups increase investment in three types of production factors, namely capital, technology and entrepreneurship and labor.

$$Intermediate_{i,t} = \gamma_0 + \gamma_1 Group_{i,t} + \gamma_i Control + Industry + Year + Region + \mu_{i,t} \quad (2)$$

where *Intermediate* includes *RD*, *Patent*, *EDU*, *Manage_tech*, *Dividend* and *ICR*. We then formulate Eq. (3) to test whether business groups promote the DT of affiliated firms using the aforementioned mediating variables.

$$Digital_{i,t} = \beta_0 + \varphi_1 Group_{i,t} + \varphi_2 Intermediate + \varphi_i Control + Industry + Year + Region + \varepsilon_{i,t} \quad (3)$$

According to the mediation effect-testing procedure described above, we sequentially examine the coefficients γ_1 and φ_2 in Eqs. (2) and (3). If both coefficients are statistically significant, it indicates that at least part of the effect of *Group* on *Digital* operates through the *Intermediate* variables. Hayes (2018) decomposes the total effect of a mediation model into direct and indirect components, where φ_1 represents the direct effect and$\gamma_1 \times \varphi_2$represents the indirect effect. If only one of the coefficients γ_1 or φ_2 is statistically significant, we further test for the presence of a mediating effect using the Sobel (1982) test.

3.3.3. Internal markets

In addition to using mediation analysis to test H2 to H4, we use a sample of only business group-affiliated firms to examine whether the spillover of factors in internal capital, technology and entrepreneurship and human resource markets within the group promotes the DT of affiliated firms. Accordingly, we formulate Eq. (4):

$$Digital_{i,t} = \beta_0 + \beta_1 ICM\ (ITM\ /\ IHM)_{i,t} + \beta_i Control + Industry\ +\ Year\ +\ Region\ + \varepsilon_{i,t} \quad (4)$$

Building on the findings of He et al. (2013), we examine three types of internal markets, namely capital, technology and entrepreneurship and labor. *ICM* represents the internal capital market and is measured by total free cash flow and total cash flow as a percentage of assets for all firms within a business group, excluding the focal firm, denoted as *ICM1_Group* and *ICM2_Group*. *ITM* represents the internal technology and entrepreneurship market and is measured using the natural logarithm of R&D expenditures plus 1 for all firms within a business group, excluding the focal firm (*RD_Group*), and the aggregate number of invention patents held by firms within the same business group, excluding the focal firm (*Patent_Group*). Similarly, *IHM* represents the internal labor market and is measured by the aggregated education level of senior executives of firms within a business group, excluding the focal firm (*EDU_Group*), and the aggregated value of the dummy variable *Manage_tech* of firms within a business group, except the focal firm (*Manage_tech_Group*).

3.4. Descriptive statistics and univariate analysis

Table 2 reports the descriptive statistics of the main variables. *Digital* has a mean value of 0.372, indicating that approximately 37% of firms have implemented DT, which suggests that DT adoption remains relatively limited. *Digital_P*, *Digital_R* and *Digital_O* have mean values of 0.215, 0.225 and 0.230, respectively, indicating that approximately 22% and 23% of firms have implemented DT in products and R&D and operations, respectively. The mean value of *Group* is 0.23, indicating that approximately 23% of firms in China are affiliated with business groups.

4. Empirical results

4.1. Baseline regression results

Table 3 reports the results for H1. The dependent variables are *Digital*, *Digital_P*, *Digital_R* and *Digital_O*. Column (1) shows that the coefficient of *Group* is significant and positive at the 5% level, indicating that business group-affiliated firms are more likely to undergo DT than unaffiliated firms. This result is consistent with H1. These results are also economically significant. The estimated coefficient of 0.158 implies that a business group-affiliated firm has an approximately 17% (i.e., $e^{0.158} - 1$) higher probability of engaging in DT than an unaffiliated firm.

We further report regression results for the relationship between business group affiliation and different DT levels in firms. In the product- and R&D-related DT segments, the coefficients of *Group* are significant and positive at the 5% level, with magnitudes of 0.182 and 0.183, respectively. These estimates imply that business group-affiliated firms have approximately 20% higher probabilities of engaging in product- and R&D-related DT, respectively, than unaffiliated firms. However, the coefficient of *Group* in Column (4), which captures operation-related DT, is nonsignificant. Overall, the results indicate that business groups promote DT among their affiliated firms, with the effect primarily concentrated in products and R&D activities.

Table 2
Descriptive statistics.

Variable	N	Mean	Std	Min	P25	Median	P75	Max
Panel A: Full sample								
Digital	33,500	0.372	0.483	0.000	0.000	0.000	1.000	1.000
Digital_P	33,500	0.215	0.411	0.000	0.000	0.000	0.000	1.000
Digital_R	33,500	0.225	0.417	0.000	0.000	0.000	0.000	1.000
Digital_O	33,500	0.230	0.421	0.000	0.000	0.000	0.000	1.000
Digital_D	33,500	0.784	1.409	0.000	0.000	0.000	1.099	5.136
Digital_RD	27,679	0.142	0.349	0.000	0.000	0.000	0.000	1.000
TFP	33,500	6.650	0.897	3.972	6.031	6.555	7.190	8.993
TobinQ	33,500	2.030	1.374	0.467	1.070	1.590	2.559	5.667
Group	33,500	0.261	0.439	0.000	0.000	0.000	1.000	1.000
RD	31,074	0.019	0.019	0.000	0.003	0.016	0.028	0.099
Patents	33,500	1.713	1.532	0.000	0.000	1.609	2.944	5.714
EDU	33,500	3.413	0.551	1.889	3.043	3.440	3.778	4.929
Manage_tech	33,500	1.272	0.677	0.000	0.693	1.386	1.792	2.639
Dividend	33,500	0.135	0.190	0.000	0.010	0.070	0.180	1.040
ICR	33,500	6.148	25.065	−54.626	1.224	3.195	10.726	118.387
SOE	33,500	0.374	0.484	0.000	0.000	0.000	1.000	1.000
Size	33,500	22.152	1.311	19.058	21.200	21.983	22.921	25.707
Lev	33,500	0.424	0.202	0.053	0.263	0.418	0.576	0.940
Roa	33,500	0.045	0.063	−0.341	0.017	0.042	0.074	0.224
EstAge	33,500	2.744	0.456	0.000	2.485	2.833	3.045	3.555
Intaratio	33,500	0.143	0.223	0.000	0.020	0.042	0.104	0.840
CapiInte	33,500	2.528	2.407	0.400	1.295	1.879	2.835	20.312
Current	33,500	1.251	0.953	0.098	0.637	0.994	1.543	5.453
Growth	33,500	0.193	0.464	−0.717	−0.021	0.112	0.283	3.043
Top1	33,500	0.349	0.150	0.091	0.232	0.329	0.450	0.755
Boardsize	33,500	7.539	2.945	1.609	7.000	9.000	9.000	15.000
Ind	33,500	0.373	0.050	0.000	0.333	0.333	0.429	0.500
Mshare	33,500	0.154	0.240	0.000	0.000	0.004	0.274	1.183
Big4	33,500	0.058	0.233	0.000	0.000	0.000	0.000	1.000
Panel B: Sample of business groups								
RD_Group	7921	0.058	0.089	0.000	0.004	0.023	0.068	0.762
Patents_Group	8754	7.675	11.830	0.000	1.099	3.526	8.823	92.367
EDU_Group	8754	3.309	0.689	0.000	3.087	3.450	3.722	4.576
Manage_tech_Group	8754	4.745	5.957	0.000	1.386	2.303	5.545	30.787
ICM1_Group	8754	0.090	0.351	−1.075	−0.036	0.092	0.237	1.325
ICM2_Group	8754	−0.256	0.634	−3.166	−0.452	−0.143	0.064	1.269
Digital_Group	8754	1.297	2.118	0.000	0.000	1.000	1.000	15.000
Digital_P_Group	8754	0.715	1.339	0.000	0.000	0.000	1.000	11.000
Digital_R_Group	8754	0.823	1.561	0.000	0.000	0.000	1.000	10.000
Digital_O_Group	8754	0.784	1.486	0.000	0.000	0.000	1.000	11.000

This table presents descriptive statistics of variables in the main analysis. Variable definitions are presented in Appendix A. Continuous variables are winsorized at the top and bottom 1% percentiles.

4.2. Robustness tests

4.2.1. Propensity score matching

One limitation of the above-mentioned regression results is that unaffiliated firms have more observations than group-affiliated firms. This imbalance may biased the findings. Furthermore, the sample is not randomly distributed between group-affiliated and unaffiliated firms. To overcome this limitation, we use propensity score matching (PSM) to identify unaffiliated firms with characteristics similar to those of group-affiliated firms. First, we construct a logit model using the control variables in Eq. (1) as explanatory variables and

Table 3
Business group affiliation and digital transformation.

	(1) *Digital*	(2) *Digital_P*	(3) *Digital_R*	(4) *Digital_O*
Group	**0.158*****	**0.182****	**0.183****	**−0.008**
	(2.723)	**(2.304)**	**(2.257)**	**(−0.124)**
SOE	−0.158**	−0.193**	−0.077	−0.153**
	(−2.575)	(−2.277)	(−0.885)	(−2.217)
Size	0.181***	0.170***	0.178***	0.219***
	(7.723)	(5.648)	(5.396)	(8.090)
Lev	−0.260*	−0.211	−0.037	−0.191
	(−1.820)	(−1.165)	(−0.189)	(−1.219)
Roa	0.707**	−0.328	0.658	1.144***
	(2.170)	(−0.849)	(1.607)	(3.298)
EstAge	−0.195***	−0.316***	−0.309***	−0.047
	(−3.187)	(−4.188)	(−3.733)	(−0.620)
Intaratio	−0.328**	0.084	0.199	−0.312**
	(−2.123)	(0.452)	(1.238)	(−2.012)
CapiInte	−0.079***	−0.062***	−0.100***	−0.106***
	(−5.476)	(−3.147)	(−4.387)	(−6.646)
Current	−0.165***	−0.255***	−0.284***	−0.059*
	(−5.672)	(−5.530)	(−5.797)	(−1.835)
Growth	0.017	0.074**	0.076**	−0.010
	(0.535)	(2.095)	(1.986)	(−0.281)
Top1	−0.037	−0.234	−0.059	0.399**
	(−0.235)	(−1.130)	(−0.264)	(2.261)
Boardsize	0.017	0.011	0.034*	0.023
	(1.060)	(0.524)	(1.747)	(1.266)
Ind	0.406	0.371	1.200**	0.123
	(0.930)	(0.655)	(1.995)	(0.253)
Mshare	0.337***	0.284**	0.260*	0.289***
	(3.382)	(2.511)	(1.895)	(2.729)
Big4	−0.040	−0.142	−0.069	0.164
	(−0.408)	(−1.123)	(−0.537)	(1.486)
Intercept	−4.082***	−5.004***	−4.732***	−5.553***
	(−6.905)	(−6.265)	(−5.491)	(−8.065)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	33,500	33,500	33,500	33,500
pseudo R^2	0.195	0.174	0.198	0.186

This table presents results on the relationship between business group affiliation and digital transformation within sample. The dependent variables are *Digital*, *Digital_P*, *Digital_R*, and *Digital_O*. Variable definitions are presented in Appendix A. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Group as the dependent variable to ensure that the characteristics do not significantly differ between the treatment group (business group-affiliated firms) and the control group (unaffiliated firms). Then, based on the calculated propensity scores, we conduct 1:1 matching of the treatment and control groups using the nearest-neighbor method without replacement and obtain 8754 samples for the treatment group and an equal number for the control group. Untabulated post-PSM results indicate that the means of the covariates do not significantly differ between the two groups. Next, we present the regression results for Eq. (1) using the PSM sample in Columns (1) (4) of Table 4, with *Digital*, *Digital_P*, *Digital_R* and *Digital_O* as dependent variables. The coefficients of *Group* remain significant and positive at the 5% level in Columns (1) (3), indicating the robustness of the results for H1.

Table 4
Robustness tests.

Panel A:

	PSM				IV		Heckman		Degree of digital	Digital R&D
	(1) *Digital*	(2) *Digital_P*	(3) *Digital_R*	(4) *Digital_O*	(5) *Group*	(6) *Digital*	(7) *Group*	(8) *Digital*	(9) *Digital_D*	(10) *Digital_RD*
Group	**0.146****	**0.197****	**0.193****	**−0.029**		**0.502*****		**0.146****	**0.041*****	**0.164***
	(2.349)	**(2.290)**	**(2.238)**	**(−0.437)**		**(3.759)**		**(2.507)**	**(2.611)**	**(1.859)**
ProvinceGroup					**0.264*****					
					(6.118)					
IMR								**0.392**		
								(1.644)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	−6.936***	−5.865***	−8.196***	−9.596***	−13.931***	0.125	−6.788***	−6.383***	−1.106***	19.319
	(−9.335)	(−5.743)	(−7.895)	(−11.370)	(−8.395)	(0.795)	(−5.963)	(−4.197)	(−6.189)	(0.000)
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	17,508	17,508	17,508	17,508	33,500	33,500	33,500	33,500	33,500	33,500
pseudo R^2/adj R^2	0.222	0.198	0.228	0.216	0.280	0.101	0.272	0.196	0.279	0.548

Panel B:

	Lagging *Group*			Adding control variables				
	(1) *Digital*	(2) *Digital*	(3) *Digital*	(4) *Digital*	(5) *Digital*	(6) *Digital_P*	(7) *Digital_R*	(8) *Digital_O*
Group				**0.136****	**0.165*****	**0.206****	**0.205****	**−0.015**
				(2.336)	**(2.814)**	**(2.568)**	**(2.499)**	**(−0.235)**
Group $_{t\ 1}$	**0.143****							
	(2.455)							
Group $_{t\ 2}$		**0.138****						
		(2.349)						
Group $_{t\ 3}$			**0.124****					
			(2.077)					
lntotalword				**1.241*****				
				(13.816)				
ControllerSeperation					**−0.004**	**−0.011****	**−0.011****	**0.002**
					(−1.111)	**(−2.548)**	**(−2.363)**	**(0.448)**
IsCocurP					**0.087**	**0.174****	**0.234*****	**0.126****
					(1.557)	**(2.498)**	**(3.157)**	**(2.045)**
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	−4.153***	−4.191***	−4.353***	−14.929***	−4.162***	−5.148***	−4.934***	−5.686***
	(−7.000)	(−7.039)	(−7.265)	(−15.087)	(−7.023)	(−6.447)	(−5.711)	(−8.225)
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	33,311	33,188	32,879	33,500	33,500	33,500	33,500	33,500
pseudo R^2/adj R^2	0.197	0.197	0.199	0.204	0.196	0.175	0.200	0.186

This table presents results of robustness tests on the relationship between business group affiliation and digital transformation within samples. Columns (1) to (4) use propensity score matching (PSM) samples. Columns (5) and (6) present the results using IV. Columns (7) and (8) present the results using Heckman two-stage analysis. Column (9) uses a Degree of DT as the dependent variable; Column (10) uses a firm's R&D expenditures related to DT as the dependent variable. Panel B Columns (1) to (3) use lagged one to three years of business group values as explanatory variables. Column (4) to (8) includes additional control variable. Variable definitions are presented in Appendix A. OLS regression is adopted for Column (9) of Panel A. Logit regression is adopted for other columns. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm-clustering effects.

4.2.2. Instrumental variables

To alleviate the endogeneity problem arising from mutual causality, we use the total asset size of listed firms within the province in which the focal firm is located (*ProvinceGroup*) as an instrumental variable. Column (5) of Panel A in Table 4 reports the first-stage regression results, indicating a significant and positive correlation between the instrumental variable (*ProvinceGroup*) and business group affiliation (*Group*). The F-statistic of the first-stage regression is 151, which is much greater than 10, indicating that the instrumental variable meets expectations. These tests further demonstrate the robustness of the conclusions drawn in this paper.

4.2.3. The Heckman two-step method

There may also be a sample self-selection problem, which could lead to estimation bias. We address this problem using the Heckman two-step method. Specifically, we calculate the inverse Mills ratio (*IMR*) based on the regression result reported in Column (7) of Panel B and include *IMR* as an additional control variable in Eq. (1) for the ordinary least squares regression reported in Column (8). The coefficient of *Group* remains significant and positive, further demonstrating the robustness of the core results of this study. Moreover, the coefficient of *IMR* is nonsignificant, suggesting that any sample selection bias, if present, is not severe.

4.2.4. Alternative DT metric

We use multiple variables to replace DT variables. First, following Wu et al. (2021), we construct a DT indicator (*Digit_D*) by extracting keyword frequencies related to DT from the annual reports of listed firms. Column (9) of Panel A presents the regression results for *Group* and *Digit_D*. The coefficient of *Group* is significant and positive. Second, we analyze the text of notes on R&D expenditures using Python and then extract R&D activities related to DT (*Digit_RD*) to identify whether firms are undergoing DT. *Digit_RD* is an indicator variable that equals 1 if a firm engages in DT-related R&D in a given year and 0 otherwise. Column (10) of Panel A presents the regression results for *Group* and *Digit_RD*. The coefficient of *Group* is significant and positive. Overall, these results confirm the robustness of the findings of this study.

4.2.5. Lagged explanatory variables

There may be endogeneity between business group affiliation and DT, which could lead to a simultaneous observation of the sustained effects of DT after firms become affiliated with a business group. To mitigate this concern, we lag the main explanatory variable *Group* by one to three periods. The coefficients of *Group* in Columns (1)–(3) of Panel B remain significant and positive at the 5% level, indicating that as firms join a business group over time, the DT promoted by the business group follows a pattern of continuous improvement.

4.2.6. Adding the number of annual report words and governance-related variables as control variable

The DT variables in our study are mainly collected through manual reading of annual reports, which raises the possibility of sample selection bias resulting from differences in the number of words in annual reports. To address this concern, we add a control variable for the number of words in annual reports (*lntotalword*) in Column (4) of Panel B. *lntotalword* is defined as the natural logarithm of the total number of words in an annual report plus 1. The coefficient of *Group* remains significant and positive after controlling for the number of words in annual reports, further demonstrating the robustness of the core assumptions of this study. The effect of business group affiliation on corporate DT may also be influenced by factors such as internal governance quality within the group and resource allocation across hierarchical levels. To further ensure robustness, we additionally control for the separation between control rights and cash flow rights (*ControllerSeperation*) and whether executives hold concurrent positions in shareholder firms (*IsCocurP*). The results reported in Columns (5)–(8) of Panel B remain consistent, confirming the robustness of our findings.

4.3. Mechanism tests based on internal tangible and intangible factors

Panel A of Table 5 presents the regression results of the mechanism tests for the internal capital factor. When *Dividend* and *ICR* are used as explanatory variables, the coefficient of *Group* is nonsignificant. Following the steps of the mechanism tests, we add *Dividend* or *ICR* to Eq. (1). As shown in Columns (3) and (4), the coefficients of *Dividend* and *ICR* are also nonsignificant. The Sobel (1982) test results further indicate that

Table 5
Mechanism tests based on internal tangible and intangible factors.

Panel A: capital factor				
	(1) *Dividend*	(2) *ICR*	(3) *Digital*	(4) *Digital*
Group	**−0.006**	**−0.638**	**0.159*****	**0.157*****
	(−1.391)	**(−1.333)**	**(2.745)**	**(2.718)**
Dividend			**0.186***	
			(1.653)	
ICR				**−0.001**
				(−1.352)
Controls	Yes	Yes	Yes	Yes
Intercept	−0.478***	4.573	−3.985***	−4.080***
	(−8.722)	(0.911)	(−6.722)	(−6.903)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Soble Z value			−1.113	0.800
N	33,500	33,500	33,500	33,500
adj. R^2	0.328	0.103		
pseudo R^2			0.196	0.196
Panel B: technology and entrepreneurship factor				
	(1) *RD*	(2) *Patents*	(3) *Digital*	(4) *Digital*
Group	**0.001****	**0.173*****	**0.131****	**0.131****
	(2.054)	**(4.880)**	**(2.273)**	**(2.276)**
RD			**11.676*****	
			(8.471)	
Patents				**0.156*****
				(9.059)
Controls	Yes	Yes	Yes	Yes
Intercept	0.038***	−6.052***	−4.747***	−3.049***
	(7.421)	(−17.632)	(−7.961)	(−5.118)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Sobel Z value			1.996	4.296
N	31,074	33,500	31,074	33,500
adj. R^2	0.416	0.524		
pseudo R^2			0.185	0.199
Panel C: labor factor				
	(1) *EDU*	(2) *Manage_tech*	(3) *Digital*	(4) *Digital*
Group	**0.118*****	**0.039***	**0.118****	**0.153*****
	(6.046)	**(1.892)**	**(2.045)**	**(2.645)**
EDU			**0.340*****	
			(8.497)	
Manage_tech				**0.143*****
				(3.925)
Controls	Yes	Yes	Yes	Yes
Intercept	2.052***	1.223***	−4.772***	−4.257***
	(9.818)	(5.789)	(−8.028)	(−7.175)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Soble Z value			4.926	1.704

Table 5 (*continued*)

Panel C: labor factor				
	(1) *EDU*	(2) *Manage_tech*	(3) *Digital*	(4) *Digital*
N	33,500	33,500	33,500	33,500
adj. R^2	0.123	0.307		
pseudo R^2			0.200	0.196

This table presents results on the mechanism tests. Panel A presents the results of the capital factor. The dependent variables are *Dividend, ICR,* and *Digital*. Panel B presents results of technology factor. The dependent variables are *RD, Patents,* and *Digital*. Panel C presents results of human resource factor. The dependent variables are *EDU, Manage_tech,* and *Digital*. Variable definitions. are presented in Appendix A. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Dividend and *ICR* as capital factors do not exhibit a mediating effect. These findings suggest that business groups do not promote their affiliated firms' DT by altering financing constraints, thereby supporting H2b.

Panel B of Table 5 presents the regression results of the mechanism test for internal technology and entrepreneurship factors. Columns (1) and (2) present the regressions based on Eq. (2), which examines the impact of business groups on the technological and entrepreneurial capacity of affiliated firms. The coefficients of *Group* are all significant and positive when the explanatory variables are *RD* and *Patents*, indicating that business groups enhance both the R&D input and innovation output of affiliated firms. Columns (3) and (4) report the results of Eq. (3), in which the mediating variables *RD* and *Patents* are added to the baseline regression in Eq. (1). The coefficients of *RD*, *Patents* and *Group* are all significant and positive in Columns (3) and (4). Following Hayes (2018), we decompose the mediation effects into direct and indirect components. When the mediating variable is *RD*, the direct effect, denoted by φ_1, is 0.131, and the indirect effect is $\gamma_1 \times \varphi_2 = 0.001 \times 11.676 = 0.012$. When the mediating variable is *Patents*, the direct effect, denoted by φ_1, is 0.131, and the indirect effect is $\gamma_1 \times \varphi_2 = 0.173 \times 0.156 = 0.027$. These results indicate that firms' R&D expenditure and patent output play mediating roles and indirectly affect the relationship between business group affiliation and DT, thereby supporting H3.

Finally, Panel C of Table 5 presents the regression results of the labor factor mechanism test. When *EDU* and *Manage_tech* are explanatory variables, the coefficients of *Group* are significant and positive in Columns (1) and (2), indicating that business groups improve the labor quality of affiliated firms. Columns (3) and (4) reports the results of Eq. (3), in which the mediating variables *EDU* and *Manage_tech* are added to the main regression in Eq. (1). The coefficients of *EDU*, *Manage_tech* and *Group* are all significant and positive, as shown in Columns (3) and (4). When *EDU* is the mediating variable, the direct effect, denoted by φ_1, is 0.118, and the indirect is $\gamma_1 \times \varphi_2 = 0.118 \times 0.340 = 0.040$. When *Manage_tech* is the mediating variable, the direct effect, denoted by φ_1, is 0.153 and the indirect effect is $\gamma_1 \times \varphi_2 = 0.039 \times 0.143 = 0.006$. These results indicate that executives' ducational attainment and technological background play mediating roles and indirectly affect the relationship between business group affiliation and DT, thereby supporting H4.

4.4. Direct evidence based on internal intangible factor spillover of intra-group markets

The aforementioned results indicate that business groups can promote affiliated firms' DT by optimally allocating technology, entrepreneurship and labor factors. Based on this finding, we examine direct evidence from an internal market perspective that business groups may use internal intangible markets to drive affiliated firms' DT.

First, Columns (1) and (2) in Table 6 report the regression results on whether the internal capital market affects business group-affiliated firms' DT. The coefficients of *ICM1_Group* and *ICM2_Group* are nonsignificant, indicating that the internal capital markets of business groups do not affect their affiliated firms' DT.

Second, Columns (3) and (4) present the regression results for the internal technology and entrepreneurship markets. The coefficients of *RD_Group* and *Patents_Group* are significant and positive, indicating that technology and entrepreneurship spillovers from business groups' internal intangible markets promote affiliated firms' DT.

Table 6
Direct evidence based on produce factor spillovers of internal market.

	Internal capital market		Internal technology and entrepreneurship markets		Internal labor market	
	(1) *Digital*	(2) *Digital*	(3) *Digital*	(4) *Digital*	(5) *Digital*	(6) *Digital*
ICM1_Group	**−0.083**					
	(−0.925)					
ICM2_Group		**0.012**				
		(0.236)				
RD_Group			**2.413*****			
			(4.347)			
Patents_Group				**0.008****		
				(2.087)		
EDU_Group					**0.114****	
					(2.100)	
Manage_Thec_Group						**0.016*****
						(3.305)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	−5.412***	−5.419***	−5.781***	−5.564***	−5.768***	−5.556***
	(−4.672)	(−4.676)	(−4.993)	(−4.820)	(−4.924)	(−7.203)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
N	8754	8754	7921	8754	8754	8754
pseudo R^2	0.210	0.210	0.203	0.211	0.211	0.211

This table presents results of the impact of internal markets' produce factor spillovers on digital transformation within samples of the business groups. The dependent variables are *Digital*. Variable definitions. are presented in Appendix A. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Similarly, Columns (5) and (6) report the regression results for the internal human resource market. The coefficients of *EDU_Group* and *Manage_Tech_Group* are all significant and positive, indicating that the internal labor market significantly contributes to business group-affiliated firms' DT. Overall, the regression results for the intra-group sample are consistent with those reported in Table 5 and support H2b, H3 and H4.

5. Further analysis and discussion

Based on multifaceted evidence, the core results are robust. We further test for heterogeneity in the strength of governmental support, business groups' own characteristics and the external environment, followed by a discussion on the economic consequences of business groups' facilitation of DT.

1.1. Strength of governmental support

5.1.1. Nature of property rights

In China, state ownership is a key factor in measuring the extent to which firms are motivated by government policy. Typically, private firms, controlled by individuals or families, prioritize the firms' economic interests as their primary objective, meaning that they aim to maximize firm value. Therefore, driven by economic interests, private firms may be more motivated to undergo DT. The significant and negative coefficient of *SOE* in the baseline regression in Table 3 supports this argument. However, privately owned business groups may be less motivated to promote DT. Compared with state-owned business groups, privately owned groups are smaller in scale, typically comprising two or three listed firms, and their headquarters have less control power, making them less likely to drive DT. Panel A of Table 7 presents the regression results. The coefficient of *Group* is nonsignificant in the non-state-owned enterprises (non-SOE) sample in Column (1), supporting our previous argument.

Table 7
Policy motivated tests.

Panel A: State-ownership

	(1) *Digital* non-SOE	(2) *Digital* SOE	(3) *Digital* Central SOE	(4) *Digital* Local SOE
Group	**0.118**	**0.201****	**0.477*****	**0.124**
	(1.355)	**(2.495)**	**(2.603)**	**(1.257)**
Controls	Yes	Yes	Yes	Yes
Intercept	−3.804***	−5.438***	−7.489***	−5.024***
	(−5.017)	(−5.111)	(−4.382)	(−3.609)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	20,960	12,540	4299	8241
pseudo R^2	0.175	0.230	0.227	0.241

Panel B: Government support

	(1) *Digital* Government report involves digitization	(2) *Digital* Government report not involves digitization	(3) *Digital* High subsidies	(4) *Digital* Low subsidies	(5) *Digital* High tax ratio	(6) *Digital* low tax ratio
Group	**0.203*****	**0.000**	**0.185*****	**0.111**	**0.108**	**0.203*****
	(3.170)	**(0.002)**	**(2.616)**	**(1.455)**	**(1.416)**	**(2.679)**
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	−3.133***	−4.314***	−3.227***	−3.917***	−3.533***	−5.122***
	(−4.656)	(−4.534)	(−4.161)	(−4.962)	(−4.572)	(−6.502)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
N	26,511	6989	16,783	16,717	17,602	15,898
pseudo R^2	0.178	0.183	0.201	0.193	0.171	0.230

This table presents results of various policy motivated tests. Panel A shows the policy motivated effect of state-ownership. We divided the samples into several sub-samples: non-SOE, SOE, SOE controlled by central state (Central SOE), and SOE controlled by local state (Local SOE). Panel B presents results of the impact of government support on the relationship of business groups and digital transformation. We divided the samples into several sub-samples: Government report involves digitization, Government report not involves digitization. High subsidies, Low subsidies, High tax ratio, and Low tax ratio. Variable definitions are presented in Appendix A. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Government-controlled state-owned enterprises (SOEs) have both economic and social goals, such as providing employment and stabilizing the national economy. Meanwhile, SOE decisions are strongly influenced by government policies. For independent SOEs, DT might be relatively weak because they must pursue these multiple objectives simultaneously. However, state-owned business groups in China are relatively large in scale, and their headquarters have strong control capabilities. When firms are affiliated with state-owned groups, the group headquarters may promote DT through centralized control and resource allocation. The coefficient of *Group* is significant and positive in the SOE sample in Column (2), indicating that SOEs affiliated with a business group have a higher probability of undergoing DT.

SOEs in China are not homogeneous and can be further divided into central SOEs, controlled by the central State-owned Assets Supervision and Administration Commission (SASAC), and local SOEs, controlled by local SASACs. Because of significant differences in their characteristics, there may be differences in the extent to which business groups controlled by central and local governments promote their affiliated firms' DT. First, compared with local SOEs, central SOEs are typically larger in scale, financially stronger and control important industries and key sectors of the national economy. In addition, they have stronger resource endowments to supporte DT. Second, business groups controlled by central SOEs are more policy-driven.

During SOE reforms, many policies have been initiated and implemented by central SOEs. Hence, central SOEs are strongly motivated to implement government policies because of the annual unified performance evaluations conducted by the central SASAC. The coefficient of *Group* is significant and positive in the central SOE sample and nonsignificant in the local SOE sample in Columns (3) and (4), respectively, indicating that business groups controlled by central SOEs are more likely to promote DT than those controlled by local governments.

5.1.2. Impact of government policies

During economic development, governments support firms' innovation to achieve economic transformation goals by supporting specific industries or firms, thereby promoting innovation and the commercialization of R&D outcomes (Tian and Xu, 2022). We further examine whether government support affects the relationship between business groups and DT. First, we divide the sample into two subsamples based on the median value of digitalization-related words in government reports, which measures the extent to which the province where the affiliated firm is located attaches importance to DT. Panel B of Table 7 reports the regression results. The coefficient of *Group* is significant and positive in the subsample with higher levels of digitalization-related content in government reports. Second, we divide the sample into two subsamples based on the median value of annual government subsidies per industry. The coefficient of *Group* is significant and positive in the high government subsidy subsample. A similar pattern is observed for the taxation subsample: the coefficient of *Group* is significant and positive in the low taxation subsample. These results indicate that when governments provide strong support through subsidies and tax incentives, business groups can better use their governance mechanisms to promote their affiliated firms' DT.

5.2. Impact of the characteristics of business groups

5.2.1. Cultural differences in business groups

In China, business group headquarters typically control many listed firms through a pyramid-type shareholding structure. Although headquarters hold controlling interests in listed firms and can directly influence their affiliated firms' decision-making, they also use various informal mechanisms to exert influence. For example, headquarters disseminate the core values of the business group to affiliated firms, and the cross-transfer of executives or employees between headquarters and affiliated firms creates an informal network of relationships between the headquarters and affiliated firms. Through these channels, headquarters can indirectly transmit organizational culture, such as risk tolerance, across affiliated firms.

In recent years, increasing research attention has been paid to spillover effects from the perspective of interactions between firms within a business group. Xu et al. (2023b) suggest that cultural differences between affiliated firms within a business group can reduce corporate innovation. Significant cultural differences between firms within a business group can lead to cognitive bias, communication barriers and a lack of trust. Consequently, these differences weaken the group's influence on individual firms and reduce the degree of DT.

To test the impact of cultural differences within a business group on DT, we construct variables to measure the degree of cultural differences within the business group. As the ultimate controlling entity, the group headquarters has the final authority over subsidiaries' decisions. Thus, to measure cultural differences, we need to quantify the culture of both the headquarters and its subsidiaries. There are two main approaches to measuring corporate culture. One approach is to analyze executives' cultural traits through textual analysis. Listed subsidiaries disclose substantial textual information, whereas headquarters disclose limited information and therefore lack sufficient textual data. In addition, our DT variable is derived from textual analysis. Inclusion of these two text-based indicators in the same regression model may create serious endogeneity concerns. Another approach is to use geographic proximity to nearby clan, religious or cultural sites as a proxy for cultural influence. We adopt the distance to Confucian temples as the cultural proxy because Confucianism is China's traditional cultural foundation, and headquarters' location data can be manually collected. This method allows us to quantify the culture of both headquarters and their affiliates simultaneously.

First, we measure firm-level culture using Confucian cultural influence by counting the number of Confucian temples within a radius of 200 (300) kilometers, denoted as *Culture200* (*Culture300*). Second, we manually collect and identify the headquarters of each business group and count the number of Confucian temples

within 200 (300) kilometers of the headquarters, denoted as *HQCulture200* (*HQCulture300*). We then subtract the firm-level cultural indicator (*Culture200* or *Culture300*) from its headquarters' cultural indicator (*HQCulture200* or *HQCulture300*) and take the absolute value to obtain *Cdiff200* or *Cdiff300*.

Panel A of Table 8 reports the regression results. The coefficients of *Cdiff200* and *Cdiff300* are all significant and negative in the business group sample, indicating that greater intra-group cultural differences reduce affiliated firms' DT.

5.2.2. Time difference in spillover effects

Because a group headquarters simultaneously controls multiple subsidiaries, DT among subsidiaries may not occur synchronously, and the corresponding spillover effects of DT may also differ in timing. It is common for group headquarters to designate one subsidiary as a pilot to implement DT, after which subsidiaries that adopt DT earlier may generate spillover effects for other subsidiaries. Accordingly, we propose that the subsidiary in which the headquarters holds the highest percentage of stock equity and that adopts DT first may

Table 8
Business groups characteristics.

Panel A: Cultural difference within a business group

	(1) *Digital*	(2) *Digital*
Cdiff200	**−0.031***** **(−2.621)**	
Cdiff300		**−0.019***** **(−2.908)**
Controls	Yes	Yes
Intercept	−5.163*** (−4.410)	−5.075*** (−4.360)
Industry	Yes	Yes
Year	Yes	Yes
Region	Yes	Yes
N	8754	8754
pseudo R^2	0.212	0.212

Panel B: Time difference in spillover effects

	(1) *Digital*	(2) *Digital*
Top1Digital$_{t-1}$	**0.060***** **(2.744)**	
Top1Digital$_{t-2}$		**0.098***** **(4.030)**
Controls	Yes	Yes
Intercept	−2.442*** (−4.386)	−2.393*** (−4.285)
Industry	Yes	Yes
year	Yes	Yes
Region	Yes	Yes
N	5760	5752
pseudo R^2	0.304	0.308

This table presents results of various moderating analyses. Panel A presents results of the impact of cultural differences in a business group. Panel B presents results of the impact of the external environment. We divided the samples into several sub-samples: High competition, Low competition, Large EPU, and Small EPU. Variable definitions are presented in Appendix A. Panel C presents results of the impact of time difference in digital spillover effect, *Top1Digital* refers to the degree of digital transformation of the firm with the highest control rights within a business group in the previous year, measured by the frequency of digital-related terms in its annual report.cultural differences in a business group. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

promote DT in other subsidiaries. Panel B of Table 8 shows that the regression coefficients of $TopDigital_{t-1}$ and $TopDigital_{t-2}$ are significant and positive, indicating that timing differences exist in the DT spillover effects within business groups.

5.3. Impact of the external environment

The external environment in which the market and government interact is an important factor influencing firms' investment decisions in China's capital market. Hence, we examine whether the external environment moderates the relationship between business groups and DT.

First, we examine product market competition by dividing the sample into two subsamples based on the median value of product market competition within an industry per year. Columns (1) and (2) of Panel A in Table 9 present the regression results. The coefficient of *Group* is significant and positive only in the low product market competition subsample in Column (2), suggesting that when the external corporate governance mechanism provided by product market competition is weak, business groups are more likely to replace it with external governance effects, thereby promoting affiliated firms' DT.

Table 9
Impact of external environment.

Panel A

	Product market competition		Economic policy uncertainty	
	High competition (1)	Low competition (2)	Large EPU (3)	Small EPU (4)
Group	**0.116**	**0.212****	**0.036**	**0.268*****
	(1.511)	**(2.554)**	**(0.514)**	**(3.410)**
Controls	Yes	Yes	Yes	Yes
Intercept	−5.336***	−3.174***	−5.246***	−3.603***
	(−6.299)	(−4.148)	(−7.478)	(−4.495)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	15,981	17,519	15,072	18,428
pseudo R^2	0.212	0.187	0.119	0.130

Panel B

	Regional digitalization level		Industrial structure upgrading	
	High regional digitalization (1)	Low regional digitalization (2)	High industrial structure upgrading (3)	Low industrial structure upgrading (4)
Group	**0.022**	**0.332*****	**0.054**	**0.273*****
	(0.324)	**(4.168)**	**(0.776)**	**(3.199)**
Controls	Yes	Yes	Yes	Yes
Intercept	−3.891***	−4.877***	−4.802***	−3.300***
	(−4.791)	(−7.224)	(−6.502)	(−3.863)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	16,688	16,812	16,890	16,610
pseudo R^2	0.121	0.137	0.180	0.194

This table presents results of the impact of the external environment. We divided the samples into several sub-samples: High competition, Low competition, Large EPU, and Small EPU in Panel A; High regional digitalization, Low regional digitalization, High industrial structure upgrading and Low industrial structure upgrading in Panel B. Variable definitions are presented in Appendix A. Logit regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Second, changes in government policy can worsen firms' financing environment and reduce operating income and cash flow performance (Demir et al., 2017). We use the economic policy uncertainty index developed by Baker et al. (2016) to measure policy uncertainty in China and divide the sample into two subsamples based on the median value of the economic policy uncertainty index (*EPU*). The coefficient of *Group* is significant and positive only in the low *EPU* subsample, indicating that high economic policy uncertainty weakens the positive effect of business groups on DT.

Third, regional digital infrastructure, such as the coverage of high-speed broadband networks and the deployment of 5G networks, provides essential conditions for corporate DT. Moreover, digital technologies within a region may generate spillover effects, such that a higher level of regional digitalization facilitates the DT of local enterprises. However, when the local digitalization level is relatively low, it remains unclear whether business groups can compensate for external environmental deficiencies and promote their own affiliated firms' DT. To examine this, we use the provincial digitalization level index from the Peking University Digital Inclusive Finance Index and divide the sample into high- and low-digitalization groups based on the median. The results reported in Panel B of Table 9 indicate that when enterprises are located in provinces with a lower digitalization level, business groups can leverage their internal markets to offset the lack of external resources and promote corporate DT.

Fourth, industrial structure upgrading generally involves the transition from traditional industries to high-tech and high-value-added sectors, thereby creating an industrial environment conducive to corporate digitalization. We measure the level of regional industrial structure upgrading by calculating the ratio of the added value of the tertiary industry to that of the secondary industry. The sample is divided into high- and low-indus trial-structure-upgrading groups based on the median. The results in Panel B of Table 9 show that when enterprises are located in provinces with a less advanced industrial structure, business groups can compensate for external environmental limitations and promote corporate DT.

5.4. Economic consequences of DT

A crucial question is whether business groups improve their productivity while implementing DT. To address this, we examine TFP. Table 10 presents the results. The coefficient of *Digital* × *Group* is significant and positive at the 10% level in Column (1), indicating that, conditional on being affiliated with a business group, DT implementation improves affiliated firms' TFP, that is, production efficiency. Similarly, the coefficients of *Digital_P* × *Group* and *Digital_R* × *Group* are significant and positive in Columns (2) and (3), respectively. However, the coefficient of *Digital_O* × *Group* is nonsignificant, indicating that when promoted by a business group, DT implementation at the product and R&D levels more effectively improves affiliated firms' TFP.

Furthermore, we examine the impact of group-affiliated DT on firm market value. As shown in Columns (4)–(8) of Table 10, with Tobin's Q as the dependent variable, the coefficients of *Digital* × *Group*, *Digital_P* × *Group* and *Digital_R* × *Group* are all significant and positive. This finding indicates that firm-wide DT driven by business group affiliation as well as product- or R&D-focused DT can significantly enhance firm market value. However, the coefficient for operational digitalization (*Digital_O* × *Group*) is positive but statistically nonsignificant, suggesting that operational DT does not have a significant positive effect on market value.

6. Conclusion

We examine the bright side of business groups' internal markets in terms of risky decisions. Specifically, we use a research design based on Chinese A-share listed firms with and without business group affiliations to examine how affiliation affects a focal firm's DT.

Consistent with our hypotheses, we find that group-affiliated firms apply digital technologies more effectively than otherwise similar unaffiliated firms, especially through internal intangible markets rather than tangible markets.

Table 10
The economic consequences of digital transformation.

	(1) *TFP*	(2) *TFP*	(3) *TFP*	(4) *TFP*	(5) *TobinQ*	(6) *TobinQ*	(7) *TobinQ*	(8) *TobinQ*
Digital* × *Group	0.054***				0.101***			
	(2.585)				(3.017)			
Digital_P* × *Group		0.082***				0.117**		
		(3.104)				(2.576)		
Digital_R* × *Group			0.095***				0.077*	
			(3.765)				(1.695)	
Digital_O* × *Group				0.017				0.049
				(0.781)				(1.344)
Digital	−0.026**				0.026			
	(−2.420)				(1.335)			
Digital_P		−0.004				0.027		
		(−0.305)				(1.152)		
Digital_R			−0.008				0.061**	
			(−0.580)				(2.212)	
Digital_O				−0.035***				−0.007
				(−2.975)				(−0.307)
Group	0.037*	0.039**	0.045**	0.061***	0.042	0.055**	0.060**	0.067**
	(1.918)	(2.218)	(2.297)	(3.124)	(1.473)	(1.991)	(2.149)	(2.417)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	−2.724***	−2.720***	−2.752***	−2.767***	11.750***	11.744***	11.740***	11.734***
	(−15.192)	(−15.178)	(−15.323)	(−15.381)	(35.019)	(35.007)	(35.053)	(34.985)
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	33,500	33,500	33,500	33,500	33,500	33,500	33,500	33,500
pseudo R^2	0.726	0.726	0.726	0.726	0.478	0.478	0.478	0.478

This table presents results of the economic consequences of digital transformation within samples. The dependent variable is *TFP*. Variable definitions. are presented in Appendix A. OLS regression is adopted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests and standard errors adjusted for firm clustering effects.

Moreover, we find that business groups controlled by the Chinese central government promote affiliated firms' DT more effectively than those controlled by local governments or non-state groups, particularly under stronger government support. However, business groups can impede affiliated firms' DT when there is greater cultural diversity among affiliated firms, and timing differences exist in the DT spillover effects within business groups. The external enviroment tests show that business groups promote affiliated firms' DT more strongly when product market competition is higher and economic policy uncertainty is lower as well as when affiliated firms receive more government subsidies or have lower tax rates. Finally, we find that when business groups promote affiliated firms' DT, affiliated firms experience increases in their TFP and Tobin's Q value.

Overall, our findings suggest that traditional firms often face substantial obstacles when attempting DT. When such firms affiliate with a business group, they can use the group's internal markets, especially intangible ones, to reallocate and optimize labor, technology and entrepreneurship within the group. This internal reallocation enables them to better adopt and implement DT. This study provides a new perspective that helps external market stakeholders understand the bright side of business groups. Specifically, although business groups can reallocate productive resources through their internal markets, the effectiveness of such reallocation depends on various moderating factors, which may result in differences in resource allocation efficiency.

Although this study focuses on business groups in China, business groups are also common in other emerging markets, such as India's Tata Group, Southeast Asia's Salim Group and Latin America's Grupo Bimbo, where firms face similar DT challenges and strong demand for DT implementation. Therefore, our study offers valuable and actionable lessons for firms in emerging markets.

First, business groups should prioritize the development of internal intangible markets rather than over-relying on internal capital markets. Our findings confirm that internal intangible resources—rather than tangible capital—are the core drivers of affiliate firms' DT, including technology transfer, talent mobility and entrepreneurial resource sharing. Therefore, business groups in emerging markets can establish cross-subsidiary technology-sharing platforms to facilitate the circulation of patents, digital tools and technical solutions among affiliates. They can also implement internal talent rotation mechanisms, such as seconding digital technology experts from technologically advanced subsidiaries to non-tech affiliates that face difficulties with DT. This approach directly addresses the technology and talent gaps that constrain many non-tech firms in emerging markets.

Second, business groups should proactively align with local government digital policies to amplify the effect of their internal intangible markets. As observed in China, government support significantly strengthens the role of business groups in promoting affiliate firms' DT. For enterprises in emerging markets, this implies actively engaging with local governments. For instance, Southeast Asian business groups can apply for government grants for digital upgrading and locate DT-focused affiliates in government-developed digital industrial parks to access shared infrastructure at lower costs. Similarly, African business groups can collaborate with local governments to participate in digital literacy training programs for employees, thereby leveraging public resources to reduce the cost of internal talent development.

Third, business groups should optimize internal governance to reduce obstacles to DT spillovers. Our study identifies two key group-level factors: cultural diversity and the pioneer subsidiary effect. For business groups in emerging markets, this suggests two practical actions. First, they should reduce internal cultural frictions by aligning DT goals across affiliates and standardizing digital collaboration processes. Second, they can adopt a "pioneer-driven" strategy—prioritizing DT in subsidiaries where the headquarters holds strong control and where business foundations are more mature, then replicating successful models and sharing technical teams with other affiliates.

Fourth, business groups should adjust DT strategies based on the external environment to enhance the adaptability of their internal intangible markets. Our findings show that the effect of business groups on affiliates' DT is more pronounced under conditions such as high product market competition, low economic policy uncertainty and low regional digitalization levels. For enterprises in emerging markets, this implies the need for contextualized adaptation. In regions with high policy volatility, business groups can initially implement low-risk, short-cycle DT projects to prevent losses resulting from policy changes and then expand to more complex initiatives once policy conditions stabilize. In regions with underdeveloped digital infrastructure, business groups can collaborate with peer firms to co-invest in basic digital infrastructure, thereby reducing the cost of DT for individual subsidiaries.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of the manuscript entitled.

Acknowledgements

We are thankful for comments and suggestions from the referee at the 2025 China Journal of Accounting Research Spring Workshop in Kunming. Xu thanks the National Natural Science Foundation of China (Approval Number: 72362033) and the Basic Research Project of Yunnan Province (202401AT070288) for financial support.

Appendix A:. Variable definitions

Variable	Definition
Dependent variables	
Digital	An indicator variable equal to 1 for observations undergoing a substantial digital transformation, and 0 otherwise.
Digital_P	An indicator variable equal to 1 for observations applying digital technologies in its product offerings or customer-facing applications, and 0 otherwise.
Digital_R	An indicator variable equal to 1 for observations applying digital technologies in its R&D processes, and 0 otherwise.
Digital_O	An indicator variable equal to 1 for observations applying digital technologies in its at the operational processes, and 0 otherwise.
Digital_D	It is constructed by extracting keyword frequencies related to DT from the annual reports of listed firm, following Wu et al. (2021).
Digital_RD	An indicator variable equal to 1 for a firm's R&D related to digital transformation, and 0 otherwise.
TFP	The firm's total factor productivity, defined as in Tian and Twite (2011) who use a version of Cobb-Douglas production function to derive TFP.
TobinQ	Stock market value of firm divided by total assets at the end of year t.
Variable of interest	
Group	An indicator variable equal to 1 for observations affiliated with a business group, and 0 otherwise.
Intermediary variables	
RD	R&D expenditure divided by total assets at the end of year t.
Patents	The natural logarithm of domestic and foreign invention patents plus 1 at the end of year t.
EDU	Mean value of education of senior executives. 1=Secondary's degree and below, 2=College's degree, 3 = Bachelor's degree, 4 = Master's degree, 5 = Doctoral degree.
Manage_tech	The number of directors, supervisors, or executives with a research and development background and added 1, and then taken the natural logarithm.
Dividend	Pre tax cash dividends per share at the end of year t
ICR	Interest coverage ratio at the end of year t.
Variables of internal market	
RD_Group	The total of variable *RD* of all firms in a business group, except the focal firm's R&D expenditures.
Patents_Group	The total of variable *Patents* of all firms within a business group, except the focal firm.
EDU_Group	Aggregate value of the variable *EDU* of firms within a business group, except the focal firm.
Manage_tech_Group	The total of variable *Manage_tech* of all firms within a business group, except the focal firm.
ICM1_Group	The total free cash flow as a percentage of assets for firms within a business group, except the focal firm.
ICM2_Group	The total cash flows as a percentage of assets for firms within a business group, except the focal firm.

Appendix A (*continued*)

Variable	Definition
Other Variables	
Culturediff200/300	Confucian cultural differences between a business group headquarter and affiliated firms.
SOE	An indicator variable equal to 1 for observations with state owned property rights, and 0 otherwise.
Size	The natural logarithm of total assets at the end of year t.
Lev	The ratio of liability to total assets at the end of year t.
Roa	The ratio of net income to total assets at the end of year t.
EstAge	The natural logarithm of the number of years the firm has been in existence.
Intaratio	The ratio of intangital assets to total assets at the end of year t
CapiInte	Total assets divided by operating income at the end of year t.
Current	Current asset turnover ratio at the end of year t.
Growth	Annual growth in sales at the end of year t.
Top1	The percentage of shares owned by the largest shareholder at the end of year t.
Boardsize	The number of the board of directors at the end of year t.
Ind	The proportion of independent directors to the number of board members at the end of year t.
Mshare	Senior executives' shareholding ratio at the end of year t.
Big4	An indicator variable equal to 1 for observations audited by top 4 accounting firms, and 0 otherwise.

References

Almeida, H.V., Wolfenzon, D., 2006. A theory of pyramidal ownership and family business groups. J. Financ. 61 (6), 2637–2680.

Austin, A.A., Williams, L.T., 2021. Evaluating company adoptions of blockchain technology: how do management and auditor communications affect nonprofessional investor judgments?. J. Account. Public Policy 40 (5) 106882.

Bai, Z., Ban, Y., Hu, H., 2024. Banking competition and digital transformation. Financ. Res. Lett. 61 105068.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131 (4), 1593–1636.

Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. 51 (6), 1173–1182.

Benner, M.J., Waldfogel, J., 2023. Changing the channel: digitization and the rise of "middle tail" strategies. Strateg. Manag. J. 44 (1), 264–287.

Cai, W., Ni, X., Zhao, P., Yang, T., 2019. The impact of business groups on innovation outputs: evidence from Chinese manufacturing firms. China Industr. Econ. 1, 137–155 (In Chinese).

Chen, W., Srinivasan, S., 2024. Going digital: implications for firm value and performance. Rev. Acc. Stud. 29 (2), 1619–1665.

Chen, X., Lin, Z., Xu, D., 2025. Parent firm dividends, financial pressure transmission and tax avoidance among subsidiaries: evidence from China. China J. Account. Res. 18 (4) 100447.

Chen, Y., Lu, C., Li, Z., 2024. How does digital transformation affect corporate accounting employees? China J. Account. Res. 17 (4) 100391.

Claessens, S., Fan, J.P., 2002. Corporate governance in Asia: a survey. Int. Rev. Financ. 3 (2), 71–103.

Claessens, S., Fan, J.P., Lang, L.H., 2006. The benefits and costs of group affiliation: evidence from East Asia. Emerg. Mark. Rev. 7 (1), 1–26.

Demir, E., Ersan, O., 2017. Economic policy uncertainty and cash holdings: evidence from BRIC countries. Emerg. Mark. Rev. 33, 189–200.

Faccio, M., O'Brien, W.J., 2021. Business groups and employment. Manag. Sci. 67 (6), 3468–3491.

Fang, J., Pittman, J., Zhang, Y., Zhao, Y., 2017. Auditor choice and its implications for group-affiliated firms. Contemp. Account. Res. 34 (1), 39–82.

Fang, Q., Yu, N., Xu, H., 2023. Governance effects of digital transformation: from the perspective of accounting quality. China J. Acc. Stud. 11 (1), 77–107.

Glaser, M., Lopez-De-Silanes, F., Sautner, Z., 2013. Opening the black box: internal capital markets and managerial power. J. Financ. 68 (4), 1577–1631.

Goldfarb, A., Tucker, C., 2019. Digital economics. J. Econ. Lit. 57 (1), 3–43.
Guan, Y., Su, L., Wu, D., Yang, Z., 2016. Do school ties between auditors and client executives influence audit outcomes? J. Account. Econ. 61 (2–3), 506–525.
Hayes, A.F., 2018. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, second ed. The Guilford Press, New York.
He, J., Mao, X., Rui, O.M., Zha, X., 2013. Business groups in China. Finance 22, 166–192.
Hilary, G., Hsu, C., Segal, B., Wang, R., 2016. The bright side of managerial over-optimism. J. Account. Econ. 62 (1), 46–64.
Huang, J., Chen, X., 2011. Business group and R&D investment: an analysis from the perspective of knowledge spillover and internal capital market. Econ. Res. J. 46 (6), 80–92, In Chinese.
Khanna, N., Tice, S., 2001. The bright side of internal capital markets. J. Financ. 56 (4), 1489–1528.
Khanna, T., Palepu, K., 2000. Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups. J. Financ. 55 (2), 867–891.
Khanna, T., Rivkin, J.W., 2001. Estimating the performance effects of business groups in emerging markets. Strateg. Manag. J. 22 (1), 45–74.
Khanna, T., Yafeh, Y., 2007. Business groups in emerging markets: paragons or parasites? J. Econ. Lit. 45 (2), 331–372.
Kim, J., Oh, I., 2024. Adoption of emerging technologies and growth of manufacturing firms: the importance of technology types and corporate entrepreneurship. Technol. Anal. Strat. Manag. 36 (10), 2476–2488.
Kozanoglu, D.C., Abedin, B., 2021. Understanding the role of employees in digital transformation: conceptualization of digital literacy of employees as a multi-dimensional organizational affordance. J. Enterp. Inf. Manag. 34 (6), 1649–1672.
Na, C., Chen, X., Xu, H., 2024. Can financial sharing reduce cost stickiness? Based on the perspective of group control. Nankai Bus. Rev. 27 (5), 198–208 (In Chinese).
Ozbas, O., Scharfstein, D.S., 2010. Evidence on the dark side of internal capital markets. Rev. Financ. Stud. 23 (2), 581–599.
Porfírio, J.A., Carrilho, T., Felício, J.A., Jardim, J., 2021. Leadership characteristics and digital transformation. J. Bus. Res. 124, 610–619.
Scharfstein, D.S., Stein, J.C., 2000. The dark side of internal capital markets: divisional rent-seeking and inefficient investment. J. Financ. 55 (6), 2537–2564.
Shin, H.H., Park, Y.S., 1999. Financing constraints and internal capital markets: evidence from Korean chaebols. Finance 5 (2), 169–191.
Sobel, M.E., 1982. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol. Methodol. 13, 290–312.
Tian, X., Xu, J., 2022. Do place-based policies promote local innovation and entrepreneurship? Eur. Finan. Rev. 26 (3), 595–635.
Torres, P., Augusto, M., 2020. Digitalisation, social entrepreneurship and national well-being. Technol. Forecast. Soc. Chang. 161 120279.
Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N., Haenlein, M., 2021. Digital transformation: a multidisciplinary reflection and research agenda. J. Bus. Res. 122, 889–901.
Wu, K., Fu, Y., Kong, D., 2022. Does the digital transformation of enterprises affect stock price crash risk? Financ. Res. Lett. 48 102888.
Wulf, J., 2009. Influence and inefficiency in the internal capital market. J. Econ. Behav. Organ. 72 (1), 305–321.
Xu, H., Chan, K.C., Na, C., Fang, Q., 2023a. The bright side of the internal labor market: evidence from the labor cost stickiness of firms affiliated with privately owned business groups in China. Finance 78 102356.
Xu, H., Cheng, B., Fang, Q., 2023b. Does culture diversity affect innovation? Evidence from Chinese business group affiliated firms. Econ. Res.-Ekonomska Istraživanja 36 (1) 2180415.
Yunis, M., Tarhini, A., Kassar, A., 2018. The role of ICT and innovation in enhancing organizational performance: the catalysing effect of corporate entrepreneurship. J. Bus. Res. 88, 344–356.
Zhai, H.Y., Yang, M., Chan, K.C., 2022. Does digital transformation enhance a firm's performance? Evidence from China. Technol. Soc. 68 101841.
Zhang, W., Lu, C., Liang, S., 2023a. More words but less investment: rookie CEOs and firms' digital transformations. China J. Account. Res. 16 (3) 100305.
Zhang, Y., Ma, X., Pang, J., Xing, H., Wang, J., 2023b. The impact of digital transformation of manufacturing on corporate performance—the mediating effect of business model innovation and the moderating effect of innovation capability. Res. Int. Bus. Financ. 64 101890.

HOSTED BY

Contents lists available at ScienceDirect

China Journal of Accounting Research

journal homepage: www.elsevier.com/locate/cjar

Corrigendum

Corrigendum to "Business groups and affiliated firms' digital transformation" [China J. Account. Res. 19 (2026) 100488]

Hui Xu [a], Haomin Wu [b], Wei Zheng [c], Kam C. Chan [d], Qiaoling Fang [c,*]

[a] *Accounting/Audit School, Yunnan University of Finance and Economics, Kunming, China*
[b] *Journal Center, Yunnan University of Finance and Economics, Kunming, China*
[c] *School of Economics and Management, Fuzhou University, Fuzhou, China*
[d] *Research Center of Finance, Shanghai Business School, Shanghai, China*

The authors regret that the direction of the moderating effect of product market competition was incorrectly stated in four places in the published article due to typos. Our empirical results in Table 9 show that the coefficient of Group is significant and positive only in the low product market competition subsample, indicating that business groups promote affiliated firms' digital transformation (DT) more strongly when product market competition is low. Accordingly, the following corrections should be made, with the relevant text underlined in both the original and corrected versions:

1. Abstract (page 1). The sentence "This impact is stronger under high product market competition, low economic policy uncertainty, low regional digitalization, low industrial structure upgrading and strong government support, and in provinces where DT is emphasized." should read "This impact is stronger under low product market competition, low economic policy uncertainty, low regional digitalization, low industrial structure upgrading and strong government support, and in provinces where DT is emphasized."
2. Introduction (page 3). The sentence "High product market competition, low economic policy uncertainty, low regional digitalization level or weak industrial structure upgrading cause business groups to exert a more pronounced effect on affiliated firms' DT." should read "Low product market competition, low economic policy uncertainty, low regional digitalization level or weak industrial structure upgrading cause business groups to exert a more pronounced effect on affiliated firms' DT."
3. Introduction (page 3), the study period was incorrectly stated. "During our study period from 2008 to 2019, China's central and local governments introduced more than 420 policy initiatives related to the development of the digital economy and digital technologies." should read "During our study period from 2008 to

☆ DOI of original article: 10.1016/j.cjar.2026.100488.
* Corresponding author.
E-mail address: qiaolingfang@fzu.edu.cn (Q. Fang).

https://doi.org/10.1016/j.cjar.2026.100514
1755-3091/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Sun Yat-sen University. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2022, China's central and local governments introduced more than 420 policy initiatives related to the development of the digital economy and digital technologies." The correct sample period is 2008-2022, consistent with the sample selection described in Section 3 (Research methods).

4. Conclusion (page 22). The sentence "The external environment tests show that business groups promote affiliated firms' DT more strongly when product market competition is higher and economic policy uncertainty is lower as well as when affiliated firms receive more government subsidies or have lower tax rates." should read "...more strongly when product market competition is lower and economic policy uncertainty is lower as well as when affiliated firms receive more government subsidies or have lower tax rates."
5. Practical implications in the Conclusion (page 23). The sentence "Our findings show that the effect of business groups on affiliates' DT is more pronounced under conditions such as high product market competition, low economic policy uncertainty and low regional digitalization levels." should read "...more pronounced under conditions such as low product market competition, low economic policy uncertainty and low regional digitalization levels."

The authors would like to apologise for any inconvenience caused.