Journal of Accounting, Business and Finance Research
ISSN: 2521-3830
Vol. 18, No. 1, pp. 12-25, 2024
DOI: 10.55217/102.v18i1.753

Does digital transformation promote innovation performance? Evidence from listed Chinese firms

Zhiwei Jin[1*]
Jie Qin[2]

[1,2]*Graduate School of Economics, Ritsumeikan University, Japan.*
[1]*Email: gr0602hi@ed.ritsumei.ac.jp*
[2]*Email: khata@ec.ritsumei.ac.jp*

Abstract

This study examines the impact of digitalization on firm innovation. Using a comprehensive sample of publicly listed Chinese firms, we construct a micro-level digitalization indicator using a textual analysis of annual financial reports. It was found that digitalization leads to a significant upswing in a firm's innovative output. Rigorous robustness checks and stability tests were conducted, including variable substitutions and different models, such as the Tobit and Poisson models. Additionally, the study utilizes a treatment effects model to account for unobserved heterogeneity and control for potential biases, ensuring the reliability and validity of the findings. Whether an enterprise is state-owned (SOEs) or non-state owned (non-SOEs), the digitalization of enterprises has dramatically enhanced their innovation capabilities. Digitalization has a significant effect on financially constrained firms. Compared to non-SOEs, the effect of financial constraints is more pronounced among SOEs. Highly digitalized firms also experienced higher growth rates and lower leverage ratios relative to firms with low digitalization, and they are more likely to receive governmental subsidies. Furthermore, the findings suggest that policymakers should prioritize initiatives aimed at promoting digitalization across all sectors of the economy. By fostering an environment conducive to digital innovation, governments can stimulate economic growth and enhance firms' competitiveness on a global scale. In essence, the study underscores the transformative potential of digitalization in driving innovation and economic progress.

Keywords:
Digitalization
Financial constraints
Innovation performance
Ownership structure.

JEL Classification:
O33; L25; G32.

Publisher:
Scientific Publishing Institute

Received: *28 December 2023*
Revised: *26 February 2023*
Accepted: *6 March 2024*
Published: *14 March 2024*

(* *Corresponding Author*)

Funding: This research is supported by Chinese Scholarship Council (Grant number: 202208360044).
Institutional Review Board Statement: Not applicable.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

1. Introduction

The revolution in technology is commonly seen as the primary driver behind the significant overhaul of the worldwide economy throughout the 1990s (Akcigit & Kerr, 2018). It is widely acknowledged that the surge in technology played a pivotal role in reshaping economic landscapes during that era. The process of digital transformation encompasses embracing digital technologies and reshaping the fundamental organizational framework to attain an enhanced and automated competitive edge (Brynjolfsson & Hitt, 2003; Chen & Srinivasan, 2023; Vial, 2019). As a pivotal stage of the technological revolution, this shift empowers companies to improve their production procedures and corporate structures with the incorporation of digital technology (Jiang, Du, & Chen, 2022). However, much debate has occurred on whether digital technology enhances or impedes innovation (Acemoglu, Autor, Dorn, Hanson, & Price, 2014; Kong, Lin, Wei, & Zhang, 2022). In this study, we construct a micro-level digitalization indicator and examine how digitalization affects firm innovation.

As micro-level actors in the market, enterprises are crucial agents in the complex web of economic relationships that underpin modern economies. They represent critical nodes for integrating digital technology into the real economy. Thus, this study focuses on how digitalization influences the productivity innovation of micro-enterprises. We develop a comprehensive digital lexicon by leveraging the semantic representations of governmental policies relevant to the digital economy and construct a micro-level digitalization indicator using a Python-based textual analysis of annual financial reports for all publicly listed Chinese firms.

In regressions that control for various factors and firm characteristics (including governance and transparency), highly digitalized firms were found to have significantly increased innovation compared to firms with low digitalization. These primary results remain robust even after conducting tests to ensure their validity, such as replacing the explanatory variables and using alternative model specifications.

In China, state-owned enterprises (SOEs) and non-SOEs exhibit distinctive characteristics in terms of ownership structure, governance mechanisms, and strategic decision-making processes (Allen, Qian, & Qian, 2005; Choi, Lee, & Williams, 2011; Song, Storesletten, & Zilibotti, 2011). We explore the influence of ownership type on innovation and observe a positive relationship exists between digitalization and firm innovation for both state-owned and privately owned firms. We further classify the firms by size and divide them into subgroups. The correlation remains positive and retains statistical significance even after controlling for firm size.

Further, we investigate the influence of financing constraints. Previous studies have found that unstable funding sources or restricted access to funds hinder companies' research and development (R&D) and innovation activities (Goldfarb & Tucker, 2019; Hsu, Tian, & Xu, 2014). Moreover, the lack of transparency and information asymmetry exacerbates the financing constraints that companies face in pursuing innovation endeavors (Ellis, Smith, & White, 2020; Kong et al., 2022). This study uses the Kaplan–Zingales (KZ) and Whited–Wu (WW) indices to measure financial constraints.[1] We found that digitalization chiefly propels firm innovation by curbing the finance constraints that are otherwise imposed on firms. In the case of SOEs, the influence of alleviating financing constraints and fostering innovation through digitalization is particularly prominent. However, for non-SOEs, this relationship lacks statistical significance. We also examine the influence of digitalization on corporate revenue performance and government subsidies. Our research demonstrates that firms with elevated digitalization levels exhibit superior growth, have reduced leverage, and have an increased probability of securing government grants compared with their counterparts.

Our study contributes to the existing literature in several ways. First, it contributes to the emerging literature on digitalization and innovation at the micro-firm level. Stiroh (2002) found that digitalization is crucial in improving productivity and promoting firm innovation, particularly in high-tech manufacturing and service industries. A related study by Ardito (2023) found that an increased level of digitalization within companies has a favorable impact on the probability of simultaneously introducing innovations in both environmental and social domains. Using a sample of micro and small businesses in South Africa, Gaglio, Kraemer-Mbula, and Lorenz (2022) demonstrated how digital transformation positively influences firm innovation. This study contributes to the existing literature by exploring the role of firm ownership and financing constraints in this process.

Moreover, we develop a comprehensive indicator to measure the digitalization level in enterprises. Vial (2019) noted that digitalization is a broad and complex concept; no uniform definition exists in extant literature. Some studies use IT investment, telecommunication expenditure, and the proportion of intangible assets associated to digitalization to gauge enterprise informatization density (e.g., (Stiroh, 2002; Tambe, Hitt, Rock, & Brynjolfsson, 2020; Wu, Lou, & Hitt, 2019)). Following the method of Yuan, Xiao, Geng, and Sheng (2021), we made some adjustments to reflect the influence of government policies. This indicator and its composition method may help future research on digitalization at the micro-firm level.

Finally, our study adds to the vast literature on the effects of digitalization. Akcigit and Kerr (2018) showed that the rapid development of a range of emerging digital technologies, including the internet, significantly impacts economic growth. Reimers and Waldfogel (2021) suggested that the digitalization trend has resulted in an abundance of fresh merchandise and significantly amplified the importance of pre-buy data for consumers. Bresnahan, Brynjolfsson, and Hitt (2002) showed that advances in firm digitalization, particularly IT and IT-enabled organizational change, have raised the requirements for an adept workforce. In this study, we explore the influence of digitalization on financial constraints, revenue performance, and government subsidies.

The remainder of the paper is structured as follows: Section 2 conducts a review of the literature, Section 3 explains the data, variables, and descriptive statistics, Section 4 provides the empirical results and robustness tests, Section 5 presents further analysis, and Section 6 concludes.

[1] This study measures financial constraints by employing a method initially proposed by Kaplan and Zingales (1997). Whited and Wu (2006) suggested an alternative gauge of financial constraints utilizing a Euler equation method derived from a structural investment model to formulate the WW index.

2. Literature Review

2.1. Digitalization and Firm Innovation

Innovation, which is the process of harnessing and integrating current knowledge to generate fresh insights, has garnered considerable attention in contemporary times. As we enter what Brynjolfsson and McAfee (2014) refer to as the "second machine age" marked by the rapid rise of digital technologies, machine learning and artificial intelligence, the landscape of labor markets, business competition and innovation are set to experience seismic transformations.

Evidence suggests that digitalization positively impacts business innovation, though the effects of this trend are complex. Several studies show that applying digital technologies can facilitate breakthrough innovation and enhance the efficiency of the innovation process (Kong et al., 2022; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Digitalization blurs the boundaries of innovation stages, making digital innovation products and services characterized by rapid iterations while also contributing to the transformation of firms. For example, Stiroh (2002) showed that IT is crucial in improving productivity and promoting firm innovation, particularly in high-tech manufacturing and service industries. Similarly, Goldfarb and Tucker (2019) noted that digital technology accelerates capital deepening through the accumulation of information and communications technology capital, thereby improving capital support for firms' innovative R&D investments and enhancing productivity.

On the other hand, some scholars have voiced concerns about the negative impact of digitalization on firm innovation. The increasing complexity of the market and technological environment, rising R&D costs for new technologies and products, shorter product life cycles, and the need for faster product iteration pose significant challenges for firms. The sustainability of their innovation activities is increasingly being put to the test. The productivity of R&D activities is declining, and it is becoming increasingly difficult for firms to develop new technologies independently. Acemoglu et al. (2014) argued that the productivity growth rate of the US manufacturing industry has declined following the widespread use of IT, which is consistent with the Solow paradox. The authors posit that a threshold exists for the impact of IT adoption on productivity and that when a certain level of IT adoption is reached, further applications no longer significantly increase productivity.

Thus, while digitalization presents unprecedented opportunities and challenges to business development, it is evident that it also poses significant risks. These arguments indicate the intricate nature of the digitalization–innovation nexus, and further investigation is imperative to comprehensively grasp its implications. The above discussion leads to our first hypothesis.

Hypothesis 1a: Digitalization has a positive impact on firm innovation.

Hypothesis 1b: Digitalization has a negative impact on firm innovation.

2.2. Digitalization, Ownership Type and Firm Innovation

The process of innovation is seen as a dynamic and interactive learning and cumulative procedure within the context of a nation's economic framework and institutional arrangement (Lundvall, 1992). Diverse nations harbor unique institutional foundations and regulations, leading to differences in how corporate governance and ownership arrangements impact the innovation endeavors of firms.

China exhibits notable characteristics in corporate governance, including state ownership, foreign capital, and centralized ownership. Institutional factors impact firms' innovative output by providing resources for creating new technologies. The government plays a pivotal role in enhancing enterprises' capacity for innovation via direct intervention as well as crafting industrial and technological policies (Aghion, Van Reenen, & Zingales, 2013; Choi et al., 2011). To stimulate innovation activities, the government has devised a series of incentive policies to guide enterprises in conducting core technology research and strengthen financial support, such as venture capital, for innovative endeavors. Furthermore, the government has developed a guiding catalog of key areas where the state encourages enterprises to conduct R&D, guiding technological innovation based on national needs. SOEs benefit from superior infrastructure and a talented workforce, facilitating innovative activities. Government ownership positively impacts firm innovation performance and fosters patent R&D activities.

However, certain research highlights the adverse effects of governmental ownership directly on the innovation output of firms (Boardman & Vining, 1989; Dewenter & Malatesta, 2001). Researchers posit that the genesis of this issue emanates from the government's pursuit of profit maximization and social stability as policy imperatives. During the initial phases of economic restructuring, China prioritized the privatization of SOEs, transferring property rights from the state sector to various non-state sectors (Fang, Lerner, & Wu, 2017). According to these scholarly works, government ownership negatively affects firm innovation due to its inefficient structure, lax management, and lack of incentives for innovation. Moreover, corruption and crony capitalism have fostered unhealthy relationships within businesses. Collectively, these studies and our related discussion lead to our second hypothesis.

Hypothesis 2a: Digitalization has a stronger impact on SOEs.

Hypothesis 2b: Digitalization has a stronger impact on non-SOEs.

2.3. Digitalization, Financial Constraints and Firm Innovation

Unlike other investment projects, innovation activities possess unique attributes such as high risk and information asymmetry. The first challenge stems from innovation activities that often encounter risks such as technological barriers, brain drain, and market competition. Despite investing substantial capital and human resources, firms may not necessarily yield actual returns beyond the value of their inputs. Second, competitive and imitative behaviors limit the incentive for firms to disclose details of their innovative R&D activities. Thus, investors are often unable to fully comprehend the true state of firms. This lack of transparency and information asymmetry only exacerbates the financing constraints on firms' innovation investments (Ellis et al., 2020; Kong et al., 2022).

Nevertheless, the advent of digitalization presents a promising avenue for mitigating the challenges posed by corporate financing constraints. On the one hand, digitalization can expand the financing sources available to enterprises. The profound amalgamation of digital technology and finance empowers enterprises to leverage the benefits of digital data sharing, facilitating access to timely and advantageous financing insights, thus attracting funds from social investors at a lower cost. This expands the number of financing channels available to firms and provides much-needed financial support for firms' innovation activities (Whited & Wu, 2006). On the other hand, digitalization can effectively lower the borrowing threshold and financing cost for enterprises. The new generation of digital IT empowers firms to improve the intensity of business disclosure, which allows banks and other financing institutions to use big data to make more accurate judgments on enterprise credit and financial information.

Hence, digitizing enterprises can efficiently mitigate the issue of information asymmetry amid financial institutions and businesses, promote capital flow with information flow, achieve a high degree of linkage between banks and enterprises, and further augment capital support for firm innovation and R&D investment (Goldfarb & Tucker, 2019; Hsu et al., 2014). These studies suggest that digitalization enhances firm innovation by alleviating financing constraints. This leads to our third hypothesis.

Hypothesis 3: Digitalization improves firm innovation by reducing financing constraints.

3. Data, Variables and Descriptive Statistics

3.1. Measures of Key Variables

3.1.1. Measuring Firm Innovation

This study uses the China Stock Market & Accounting Research (CSMAR) database, which includes records of patent applications and grants for over 3,000 firms from 1990 to 2020. Each record includes the patent owner's identity, year of application, year of grant, patent type, and other information. Following previous studies, we use the number of patent applications to measure a firm's innovation activity.[2] We also use the number of patent grants and the ratio of R&D expenditure to sales as additional innovation metrics in our robustness test.

3.1.2. Measuring Digitalization

Measuring digitalization is a complex and challenging process, particularly when accurately characterizing it at the micro-enterprise level. Existing research has primarily concentrated on the macro scale, employing regional or industry-based metrics to gauge digitalization trends (Vial, 2019). Few studies have been conducted at the micro-enterprise level. These studies use various digitalization measures, including information assets, IT personnel, and application of information systems. However, these measures have some deficiencies and limitations. For instance, some use IT investment, telecommunication expenditure, and the percentage of intangible assets associated with digitalization to gauge enterprise informatization density (e.g., (Bresnahan et al., 2002; Stiroh, 2002; Tambe et al., 2020)). These measures are intuitive but may not accurately reflect actual application level, because conspicuous investments easily influence investments and assets.

Thus, to avoid these problems, we proxy for firms' digital activities by counting the number of digital-related keywords in the firms' disclosures using text analysis (e.g., (Ardito, 2023; Chen & Srinivasan, 2023; Drechsler, Müller, & Wagner, 2023; Kindermann et al., 2021; Wu et al., 2019)). We created a dictionary of digital-related keywords on five topics: artificial intelligence, blockchain, cloud computing, big data, and digital technology applications. To count the instances of digital terms in the disclosures, we mainly followed the method of Yuan et al. (2021) and made some adjustments to reflect the influence of government policies. The specific words within these topics' groups are outlined in Appendix A. By counting the number of digital-related keywords for each firm each year using Python-based text analysis, we created a comprehensive indicator of digitalization for 3,800 listed firms.[3]

3.1.3. Control Variables

Following the previous literature, we control for several firm characteristics that may affect innovation productivity—firm size, age, leverage, growth, return on total assets (ROA), book-to-market ratio (BM), fixed

[2] Griliches, Hall, and Pakes (1988) showed that the year of filing a patent better reflects the actual timing of innovation (see also Acharya and Xu (2017); Ellis et al. (2020) and Mukherjee, Singh, and Žaldokas (2017)). This approach was adopted for this study.
[3] Appendix B describes the construction of the enterprise digitalization indicator.

assets ratio (PPE), percentage of independent directors (Indep), and growth opportunity (Tobin's Q). Table 1 presents a list of the variables and their definitions.

Table 1. Definition of variables.

Category	Variable	Definition
Dependent variables	Apply	Firm's total number of patent applications filed (And eventually granted)
	LnApply	Natural logarithm of one plus firm's total number of patent applications filed
Independent variable	Digital	Natural logarithm of one plus the number of firm's digital-related keywords
Control variables	Size	Natural logarithm of total assets
	Age	Natural logarithm of the number of years since the firm's first listing
	Lev	Book value of debt divided by total assets
	Growth	Yearly growth rate of total assets
	ROA	Return on assets, defined as operating income before depreciation divided by total assets
	BM	Book-to-market ratio
	PPE	Fixed assets scaled by total assets
	Indep	Percentage of independent directors
	Tobin's Q	Ratio of the market value of assets to the book value of assets

3.2. Descriptive Statistics

To construct the study sample, we obtained firms' annual reports from the CNINFO website[4] and their accounting data from the CSMAR database, a primary provider of Chinese data. The sample covers 2011–2020, excluding financial, ST (Special treatment) and ST* firms. To minimize the effects of potential outliers, we winsorized all variables at the 1st and 99th percentiles. After excluding firms with missing financial information, the final sample consists of over 3,341 unique firms and 26,538 firm-year observations.

Table 2 provides descriptive statistics for the variables. The mean and variance of total patent applications are 33.9358 and 85.4365, respectively, indicating that significant differences exist in the patenting activities of different firms.

Table 2. Descriptive statistic results.

Variable	Obs.	Mean	SD	Min.	Median	Max.
Apply	26,538	33.9358	85.4365	0.0000	8.0000	634.0000
LnApply	26,538	2.1760	1.6456	0.0000	2.1972	6.4536
Digital	26,538	3.0327	1.0514	0.0000	2.9957	5.4681
Size	26,538	22.1508	1.2887	19.8480	21.9693	26.0936
Age	26,538	2.8028	0.3718	0.0000	2.8332	4.1271
Lev	26,538	0.4149	0.2076	0.0506	0.4047	0.8889
Growth	26,538	0.1956	0.3512	−0.2952	0.0995	2.0333
ROA	26,538	0.0451	0.0615	−0.2186	0.0418	0.2206
BM	26,538	0.6256	0.2446	0.1174	0.6291	1.1576
PPE	26,538	0.2094	0.1604	0.0019	0.1757	0.6965
Indep	26,538	0.3757	0.0535	0.3333	0.3571	0.5714
Tobin's Q	26,538	2.2116	1.4596	0.8964	1.7366	9.6175

4. Empirical Results

4.1. Baseline Regression

To examine how digitalization affects firm innovation, we followed prior work (e.g., (Li, Yan, & Song, 2020; Nanda & Rhodes-Kropf, 2013)) and estimated the following:

$$LnApply_{it} = \alpha + \beta Digital_{it} + \gamma Controls_{it} + \theta_j + \mu_t + \varepsilon_{it} \quad (1)$$

Where $lnApply_{it}$ is the innovation output of firm i in year t, measured by the total number of patent applications filed; $Digital_{it}$ is the digitalization level of firm i in year t, measured by the number of digital-related keywords in the firm's disclosures; $Controls_{it}$ is a set of characteristic variables that affect a firm's innovation activities, including *Size*, *Age*, *Lev*, *Growth*, *ROA*, *BM*, *PPE*, *Indep*, and *Tobin's Q*; θ_j controls for industry fixed

[4] CNINFO is the official website (http://www.cninfo.com.cn/) designated by the China Securities Regulatory Commission (CSRC) for information disclosure of listed firms.

effects, while μ_t represents time fixed effects. The coefficient β estimates the effect of digitalization on firm innovation. Robust standard errors are clustered at the firm level.

Table 3 contains the baseline regression results. In column (1), the coefficient of Digital is 0.3070, which is significant at the 1% level. The preliminary results indicate that with the improvement of enterprise digitalization, there is a corresponding increase in innovative output, as evidenced by patent activity. Column (2) includes the baseline set of control variables. The result shows that a one percentage point increase in digitalization would increase firm innovation by 0.1986, which is an increase of approximately 9.12% (0.1986/2.1760 × 100%) relative to the mean value of the firm innovative output of 2.1760 over the sample period. To mirror the enduring aspect of investing in innovation over the long haul, we extended the observation period, drawing inspiration from Ding, Gu, and Peng (2022). In columns (3) and (4), the dependent variable is replaced with the number of patents filed for lagged one and two years, respectively. The results reveal that digitalization is strongly and consistently correlated to firm innovation; therefore, hypothesis 1a is confirmed.

Table 3. The impact of digitalization on firm innovation.

Variable	**(1) LnApply t**	**(2) LnApply t**	**(3) LnApply t+1**	**(4) LnApply t+2**
Digital	0.3070*** (14.7642)	0.1986*** (11.2877)	0.1954*** (10.5137)	0.1964*** (10.0873)
Size		0.6523*** (35.8465)	0.6527*** (33.5689)	0.6395*** (31.1468)
Age		−0.1307*** (-2.7199)	−0.1546*** (−3.0551)	−0.1552*** (−2.9826)
Lev		-0.1244 (-1.3637)	−0.0805 (−0.8311)	−0.0671 (−0.6678)
Growth		−0.1269*** (-5.0913)	−0.0634** (−2.2648)	−0.0680** (−2.2662)
ROA		0.5604*** (2.7396)	1.2838*** (5.5915)	1.5469*** (6.0732)
BM		−0.4790*** (−4.9680)	−0.6120*** (-5.7723)	−0.6907*** (−5.9141)
PPE		0.1110 (0.8815)	0.1722 (1.2977)	0.2359* (1.7026)
Indep		−0.2167 (−0.8566)	−0.2046 (−0.7590)	−0.1219 (-0.4266)
Tobin's Q		−0.0251** (−2.0219)	−0.0304** (−2.2106)	−0.0360** (−2.4923)
_cons	1.2450*** (19.1484)	−12.0452*** (-30.5884)	−11.8844*** (-28.5341)	−11.5442*** (−26.5259)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
N	26,538	26,538	22,046	18,709
Adj. R^2	0.3720	0.5421	0.5442	0.5435

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are clustered at the firm level and are reported in parentheses.

4.2. Robustness Tests

4.2.1. Alternative Dependent Variables

The CSMAR database provided us with firm-specific patent data, such as the number of patents issued and pending. Table 4 reports the results of the alternative proxy for firm innovation. In models (1) and (2), the number of patents granted is used as an additional proxy variable to measure firms' innovative output. Specifically, in model (2), the coefficient of Digital stands at 0.1860, indicative of the incremental impact of digitalization on the outcome variable, presumably patents granted.

Moreover, the effect brought by the digitalization of enterprises may increase innovation output by affecting their innovation input. Therefore, the ratio of R&D expenditure to sales is used as a proxy of firm innovation input for robustness testing. Many companies do not disclose their R&D expenditure data, resulting in a sharp reduction in the sample size. The regression results are shown in models (3) and (4). Our analysis indicates that digitalization continues to exert a statistically significant and positive influence on firm innovation, affirming the consistency with prior research findings in this domain.

Table 4. Robustness tests: Alternative dependent variables.

Variable	(1) LnGrant	(2) LnGrant	(3) RD/Sales	(4) RD/Sales
Digital	0.2849*** (13.2331)	0.1860*** (10.4924)	0.6633*** (8.8006)	0.5841*** (8.3953)
Size		0.7314*** (37.7918)		0.1900*** (3.2783)
Age		−0.0810* (−1.6630)		−1.0817*** (−6.3484)
Lev		−0.1029 (−1.1045)		−5.3185*** (−14.6279)
Growth		−0.3935*** (-15.7475)		0.3515*** (3.2380)
ROA		−0.1374 (−0.6647)		−10.9842*** (−11.2727)
BM		−0.7563*** (−7.5911)		−2.5604*** (-7.6275)
PPE		0.1417 (1.1119)		−1.4941*** (−3.7916)
Indep		−0.0537 (-0.2075)		1.0880 (1.2356)
Tobin's Q		-0.0125 (-0.9961)		0.1981*** (2.7540)
_Cons	1.3481*** (20.1452)	−13.7088*** (-32.5771)	2.5915*** (10.8948)	5.2066*** (4.2719)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
N	26,538	26,538	21,273	21,273
Adj. R^2	0.3461	0.5538	0.3570	0.4282

Note: In models (1) and (2), the dependent variable is the logarithm of one plus the number of granted patents in year t. In models (3) and (4), the dependent variable is RD/Sales. Robust standard errors clustered at the firm level are reported in parentheses. * and *** denote statistical significance at the 10% and 1% levels, respectively.

4.2.2. Alternative Models

Table 5 reports the results of the robustness checks of the baseline model with alternative models. In model (1), a dummy variable is constructed based on whether the number of firms' patent applications is zero to examine their willingness to innovate. Column (1) reports the results obtained using the Logit model. Particularly noteworthy is the revelation that for each incremental unit increase in the degree of digitalization, firms' inclination toward innovation increases by 0.38 units.

Since the number of patents of listed companies has many zero observations, as in Faleye, Kovacs, and Venkateswaran (2014), we also estimate the same variables in column (2) as a Tobit model. Column (3) repeats the specifications of the first two columns but uses a Poisson count data model. Notably, even with the adoption of different regression methodologies, the coefficient estimates for digitalization persistently exhibit a significant positive association with firm innovation. This consistency underscores the robustness of our baseline regression findings. Specifically, the coefficients of digitalization remain at 0.2662 and 0.1321 in columns (2) and (3) respectively, affirming the substantial impact of digitalization on fostering innovation within firms.

Table 5. Robustness tests: Alternative models.

Variable	(1) Logit	(2) Tobit	(3) Poisson
Digital	0.3872*** (9.6760)	0.2662*** (12.2220)	0.1321*** (5.0757)
Size	0.7298*** (16.9451)	0.7578*** (36.4931)	0.7432*** (37.5569)
Age	−0.2895*** (−2.6502)	−0.1806*** (−3.1627)	−0.1034 (−1.4761)
Lev	−0.3762* (−1.8267)	−0.1782 (−1.5347)	−0.1027 (−0.6968)
Growth	−0.1811*** (-2.8756)	−0.1434*** (-4.5458)	−0.1870*** (−3.7627)
ROA	0.4213 (0.9141)	0.7872*** (3.0304)	1.4745*** (3.3933)

Variable	(1) Logit	(2) Tobit	(3) Poisson
BM	−0.5686** (-2.5631)	−0.5725*** (-4.7815)	−0.3664** (−2.2894)
PPE	0.1218 (0.4231)	0.1449 (0.9111)	0.3297* (1.8842)
Indep	−1.1420* (−1.9535)	−0.3474 (−1.1486)	0.1140 (0.3570)
Tobin's Q	−0.0634** (−2.3585)	−0.0555*** (−3.4499)	−0.0232 (−1.1059)
Year fixed effects	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
N	26,538	26,538	26,538
Pseudo R^2	0.3161	0.2041	

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are clustered at the firm level and reported in parentheses. Pseudo R^2 indicates the pseudo-determination coefficient. Since the Logit, Tobit, and Poisson models are nonlinear regression models, the dependent variable measures of the three models differ significantly from the previous benchmark regression models, and the coefficients of digital regression cannot be directly compared with the baseline regression model.

4.3. Treatment Effects Model

There may be unobservable factors that drive both innovation and digitalization. We mitigate this problem using the treatment effects model proposed by Maddala (1986). Dummy variables are constructed based on whether a firm's digitalization level is zero or not to examine its motivation to go digital. The first step is a Probit model of the firm's decision to go digital, also known as the choice equation:

$$Digital_{it} = \begin{cases} 1 & if \;\; Digital_{it}^* > 0 \\ 0 & if \;\; Digital_{it}^* \leq 0 \end{cases} \quad Digital_{it}^* = \pi + \delta Z_{it} + \vartheta_{it} \qquad (2)$$

Where Z is a set of firm characteristic variables that affect firm digitalization decisions. The two-step approach requires the selection equation to include at least one explanatory variable that satisfies the exclusivity condition, i.e., the variable only affects firm digitalization decisions and does not directly affect firm innovation output. Given that Lev and Growth do not have a significant effect on patent applications but may influence corporate digitalization decisions, and since corporate digitalization has a continuum, corporate digitalization decisions in the previous year will affect the next year's digitalization decisions. Thus, we use Lev and Growth as control variables and add the previous year's corporate digitalization decisions (L.Digital) to the selection equation using a two-step approach to estimate the treatment effects model. In the first step, the probability of a firm going digital is estimated based on the Probit model in Equation 2.

In the subsequent stage, the analysis involves augmenting Equation 1 of the benchmark regression by adding the inverse Mills ratio as a control variable to mitigate potential self-selection bias. Throughout this study, a treatment effects model is consistently applied to all firms included in the analysis. The estimation results of the treatment effects model are reported in Table 6. The regression results in column (2) show that the coefficient of the inverse Mills ratio is significantly negative, indicating the presence of bias in the regression results caused by self-selection bias. The magnitude of the coefficient on firm digitalization changes with the inclusion of the inverse Mills ratio but is consistent with the previous results at the 1% significance level. After using the treatment effects model to correct for the endogeneity problem caused by the self-selection bias, the findings of this study still hold.

Table 6. Treatment effects model.

Variable	(1) First	(2) Second
Digital		1.5152*** (3.9355)
Growth	0.3929** (2.2132)	0.1073** (2.0835)
Lev	−0.7155*** (-4.3646)	−0.3359** (-2.4862)
L.Digital	2.1448*** (25.0083)	
Mills		−0.4150*** (−3.2954)
Size		0.5458*** (20.4973)
Age		−0.7503*** (−10.8118)
ROA		0.1110 (0.3706)

Variable	(1) First	(2) Second
BM		−1.2418*** (−9.3936)
PPE		0.0308 (0.1903)
Indep		−0.0345 (−0.0854)
Tobin's Q		−0.1425*** (−8.9723)
N	22,046	22,046
Adj. R^2		0.1439

Note: This table reports the treatment effects model's first and second step estimation results. ** and *** denote statistical significance at the 5% and 10% levels, respectively. Two-step consistent standard errors are reported in parentheses.

5. Ownership Type, Financial Constraints and Operating Performance

To provide a further understanding of the underlying mechanism(s) of innovation and digitalization, we examine the role firm ownership type, financing constraints, operating performance and government subsidies in this section.

5.1. State-Owned and Non-State-Owned Enterprises

In this section, we classify companies by ownership type and size to investigate the effect of firm ownership on digitalization and innovation. First, all listed companies are divided into SOEs and non-SOEs. Second, top-ranked firms in terms of size are considered large companies, while those with lower rankings are classified as small companies.

As shown in Table 7, the Digital coefficients for SOEs and non-SOEs are positive and significant at the 1% level, indicating that digitalization enhances firms' innovative output. Also, the p-value indicates that there is no statistically significant difference between SOEs and non-SOEs. Digitalization significantly impacts the innovation levels of companies with different ownership types. Furthermore, firm size does not affect the significance of the coefficient.

Table 7. Ownership and firm innovation.

Panel A: SOEs and non-SOEs		
Variable	(1) SOEs	(2) non-SOEs
Digital	0.1567*** (9.8443)	0.2061*** (19.0532)
_cons	−14.0911*** (−42.6491)	−11.1996*** (−40.9520)
Control variables	Yes	Yes
Year fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
N	9,052	17,486
Adj. R^2 P-value of test SOEs = non-SOEs	0.6486 0.3467	0.4903

Panel B: Ownership and firm size				
Variable	SOEs		Non-SOEs	
	(1) Large companies	(2) Small companies	(3) Large companies	(4) Small companies
Digital	0.1666*** (8.5342)	0.1626*** (6.1682)	0.1999*** (10.9579)	0.2145*** (16.3246)
_cons	−15.1762*** (−34.9095)	−12.4316*** (−11.8830)	−11.7709*** (−21.1618)	−11.3020*** (−23.4870)
Control variables	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
N	6,463	2,589	6,803	10,683
Adj. R^2	0.6624	0.5343	0.5683	0.4147

Note: In Panel A, all listed companies are further divided into two subsamples: state-owned enterprises (SOEs) and non-state-owned enterprises (non-SOEs). In Panel B, top-ranked firms in terms of size are considered large companies, while those with low rankings are classified as small companies. *** denotes statistical significance at the 10% level. Standard errors are clustered at the firm level and reported in parentheses.

5.2. Financial Constraints and Firm Innovation

For most enterprises, financing R&D activities can be challenging. Limited internal capital often falls short, thus external financing is a crucial source of R&D funds to ensure the continuity of these projects. As noted by Kaplan and Zingales (1997), cutting back on new product development investments due to financing sources or severe financing constraints is a common scenario that hinders innovation activities.

Brown and Petersen (2009) pointed out that digitalization can provide firms with greater access to finance, thereby reducing financing constraints and enabling more innovation. Expanding digital technology coverage improves the supply of financial services and offers enterprises convenient and diversified financing channels, thus promoting firm innovation output. In this study, to examine whether the digitalization of firms alleviates the inhibitory effect of financing constraints on firm innovation, the specific model is constructed as follows:

$$LnApply_{it} = \alpha + \beta_1 Digital_{it} * \mathrm{FC} + \beta_2 Digital_{it} + \beta_3 \mathrm{FC} + \gamma Controls_{it} + \theta_j + \mu_t + \varepsilon_{it} \quad (3)$$

Where FC is the financial constraints of firm i in year t, measured by the Kaplan–Zingales (KZ) and Whited–Wu (WW) indices; $Digital_{it} * FC$ is the inter term of digitalization and financial constraints; and the other variables are set in the same way as Equation 1.

Throughout each year within the observed period, we assessed companies using the KZ and WW indices, categorizing them into either financially constrained or unconstrained groups. In the top (bottom) tertile, firms are considered constrained (unconstrained). The regression results are presented in Table 8. In Panel A, the coefficient of the interaction term between digitalization and financing constraints is notably positive, suggesting that increased digitalization levels mitigate the inhibitory effect of financing constraints on firms' innovation output, which remains significant even after controlling for industry and year effects. Hypothesis 3 is therefore confirmed. All listed companies were further divided into two subsamples: SOEs and non-SOEs. For SOEs, the effect of reducing financing constraints through digitalization and promoting the innovation level of enterprises is noticeable. For non-SOEs, it is insignificant. The findings indicate that the digital transformation within enterprises expands financing channels, easing financial constraints faced by enterprises; furthermore, the greater availability of funds compels enterprises to boost R&D investments, thereby enhancing innovation outcomes.

Table 8. Financial constraints and firm innovation.

Panel A: Financial constraints measured by the KZ_INDEX			
Variable	**(1) Full sample**	**(2) SOEs**	**(3) non-SOEs**
Digital*KZ	0.0136*** (3.0281)	0.0245*** (3.1673)	0.0064 (1.2004)
Digital	0.1749*** (9.1183)	0.1061*** (3.1026)	0.1948*** (8.5257)
KZ	−0.0277* (−1.7568)	−0.0174 (−0.6501)	−0.0232 (−1.2281)
_cons	−11.9962*** (−28.5132)	−12.9973*** (−16.8723)	−10.2750*** (−18.8035)
Controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
N	24,165	8,719	15,445
Adj. R2	0.5482	0.6517	0.4937
Panel B: Financial constraints measured by WW_INDEX			
Variable	(1)	(2)	(3)
	Full Sample	SOEs	Non-SOEs
Digital*WW	0.0018*** (2.6479)	0.0158*** (6.0959)	−0.0086 (−0.4365)
Digital	0.1916*** (10.1895)	0.1625*** (5.1288)	0.1929*** (6.2020)
WW	−0.0001** (−2.1614)	−0.0001*** (−3.0020)	0.0163 (0.5138)
_cons	−12.0955*** (−27.9537)	−13.0593*** (−17.0966)	−10.1670*** (−17.5126)
Controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
N	21,983	8,310	13,671
Adj. R2	0.5491	0.6529	0.4910

Note: In Panel A, financial constraints are measured using the KZ index. In Panel B, financial constraints are measured using the WW index. The measurement is detailed in Equations 2 and 3. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are clustered at the country level and are reported in parentheses.

5.3. The Effect of Digitalization on Operating Performance and Government Subsidies

In this section, we examine firms' operating performance and government subsidies to explore in further detail the possible sources of motivation and the output of highly digitalized firms in gaining access to conduct innovation. The regression models for performance and government subsidies are estimated over the sample period.

The regression results are shown in Table 9. After controlling for the industry and year effects, the coefficients of growth and government subsidies are significantly positive, indicating that enterprises with high levels of digitalization have better profitability and can receive more government subsidies than their counterparts. In addition, the coefficient of the asset-liability ratio is significantly negative, indicating that digitalization can help reduce corporate leverage. This further shows that digitalization helps to improve corporate profitability and government subsidies, and companies' motivation to invest in R&D correspondingly increases, thereby increasing innovation output.

Table 9. The effect of digitalization on operating performance and government subsides.

Variable	(1) growth	(2) lev	(3) subsidy
Digital	0.0204*** (7.0130)	−0.0129*** (−5.0513)	0.2803*** (12.4420)
Size	−0.0472*** (−17.9106)	0.0812*** (41.3667)	0.9215*** (45.7598)
Age	−0.1035*** (−13.2668)	0.0498*** (7.9790)	−0.2135*** (−4.0797)
Lev	0.1594*** (9.7295)		0.1971 (1.5941)
ROA	2.1583*** (45.0766)	−1.1765*** (−38.4555)	1.4150*** (5.6526)
BM	0.3035*** (15.9031)	−0.0665*** (−4.4135)	−0.6455*** (−4.6845)
PPE	−0.3908*** (−22.0648)	0.0709*** (4.0427)	0.6235*** (4.0436)
Indep	0.0159 (0.3999)	−0.0249 (−0.7630)	−0.2620 (−0.8687)
Tobin's Q	0.0020 (0.7422)	−0.0016 (−0.6038)	−0.0720*** (−2.9622)
Growth		0.0361*** (9.5138)	−0.1755*** (−4.1966)
_Cons	1.1873*** (21.0086)	−1.3980*** (−30.9113)	−3.9212*** (−9.2561)
Year fixed effects	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
N	26,538	26,538	12,317
Adj. R^2	0.2238	0.4971	0.4984

Note: This table shows the effect of digitalization on operating performance and government subsidies. The dependent variables are sales growth rate, the asset-liability ratio, and government subsidies in year t. *** denotes statistical significance at the 10% level. Standard errors are clustered at the country level and are reported in parentheses.

6. Conclusion

This study explores the impact of digitalization on innovation from a micro-firm perspective. We construct a micro-level digitalization indicator using textual analysis of corporate annual financial reports for all publicly listed Chinese firms. After controlling for various firm characteristics, it was found that firms with high levels of digitalization have more innovative outputs than their counterparts. The results remain robust even after conducting tests to ensure their validity, such as replacing the explanatory variables and using alternative model specifications.

We also examined the potential mechanisms through which digitalization affects innovation. Regardless of whether an enterprise is state-owned or non-state owned, digitalization has dramatically enhanced their innovation capabilities. Digitalization has a stronger effect on financially constrained firms. Compared to non-SOEs, the effects of financial constraints are evidently more pronounced among SOEs. Highly digitalized firms also experience higher growth rates and lower leverage ratios relative to firms with low digitalization, and they are more likely to receive government subsidies.

Overall, the results of this research underscore the critical role of digitalization in enhancing innovation and driving economic development. The convergence of the digital economy with the real economy heralds a seismic transformation in the global economic landscape. In this new digital era, digitalization will keep acting as the driving force of firm innovation, economic development, and social reforms.

References

Acemoglu, D., Autor, D., Dorn, D., Hanson, G. H., & Price, B. (2014). Return of the Solow paradox? IT, productivity, and employment in US manufacturing. *American Economic Review, 104*(5), 394-399. https://doi.org/10.3386/w19837

Acharya, V., & Xu, Z. (2017). Financial dependence and innovation: The case of public versus private firms. *Journal of Financial Economics, 124*(2), 223-243. https://doi.org/10.3386/w19708

Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership. *American economic review, 103*(1), 277-304. https://doi.org/10.3386/w14769

Akcigit, U., & Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy, 126*(4), 1374-1443. https://doi.org/10.3386/w16443

Allen, F., Qian, J., & Qian, M. (2005). Law, finance, and economic growth in China. *Journal of financial economics, 77*(1), 57-116. https://doi.org/10.2139/ssrn.419481

Ardito, L. (2023). The influence of firm digitalization on sustainable innovation performance and the moderating role of corporate sustainability practices: An empirical investigation. *Business Strategy and the Environment*, 1-21. https://doi.org/10.1002/bse.3415

Boardman, A. E., & Vining, A. R. (1989). Ownership and performance in competitive environments: A comparison of the performance of private, mixed, and state-owned enterprises. *The Journal of Law and Economics, 32*(1), 1-33. https://doi.org/10.1086/467167

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics, 117*(1), 339-376. https://doi.org/10.3386/w7136

Brown, J. R., & Petersen, B. C. (2009). Why has the investment-cash flow sensitivity declined so sharply? Rising R&D and equity market developments. *Journal of Banking & Finance, 33*(5), 971-984. https://doi.org/10.1016/j.jbankfin.2008.10.009

Brynjolfsson, E., & Hitt, L. M. (2003). Computing productivity: Firm-level evidence. *Review of Economics and Statistics, 85*(4), 793-808. https://doi.org/10.2139/ssrn.290325

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* New York: WW Norton & Company. https://psycnet.apa.org/record/2014-07087-000.

Chen, W., & Srinivasan, S. (2023). Going digital: Implications for firm value and performance. *Review of Accounting Studies*, 1-47. https://doi.org/10.26226/morressier.5f0c7d3058e581e69b05d0ea

Choi, S. B., Lee, S. H., & Williams, C. (2011). Ownership and firm innovation in a transition economy: Evidence from China. *Research Policy, 40*(3), 441-452. https://doi.org/10.1016/j.respol.2011.01.004

Dewenter, K. L., & Malatesta, P. H. (2001). State-owned and privately owned firms: An empirical analysis of profitability, leverage, and labor intensity. *American Economic Review, 91*(1), 320-334. https://doi.org/10.1257/aer.91.1.320

Ding, N., Gu, L., & Peng, Y. (2022). Fintech, financial constraints and innovation: Evidence from China. *Journal of Corporate Finance, 73*, 102194. https://doi.org/10.1016/j.jcorpfin.2022.102194

Drechsler, K., Müller, S., & Wagner, H. T. (2023). The "digital premium": Why does digitalization drive stock returns? http://dx.doi.org/10.2139/ssrn.3972173

Ellis, J., Smith, J., & White, R. (2020). Corruption and corporate innovation. *Journal of Financial and Quantitative Analysis, 55*(7), 2124-2149. https://doi.org/10.2139/ssrn.2862128

Faleye, O., Kovacs, T., & Venkateswaran, A. (2014). Do better-connected CEOs innovate more? *Journal of Financial and Quantitative Analysis, 49*(5-6), 1201-1225. https://doi.org/10.2139/ssrn.1913906

Fang, L. H., Lerner, J., & Wu, C. (2017). Intellectual property rights protection, ownership, and innovation: Evidence from China. *The Review of Financial Studies, 30*(7), 2446-2477. https://doi.org/10.3386/w22685

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. *Technological Forecasting and Social Change, 182*, 121785. https://doi.org/10.1016/j.techfore.2022.121785

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature, 57*(1), 3-43.

Griliches, Z., Hall, B., & Pakes, A. (1988). *R&D, patents, and market value revisited: Is there evidence of a second technological opportunity related factor?* (No. 2624) National Bureau of Economic Research, Inc.

Hsu, P.-H., Tian, X., & Xu, Y. (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics, 112*(1), 116-135. https://doi.org/10.4324/9781351068284-9

Jiang, K., Du, X., & Chen, Z. (2022). Firms' digitalization and stock price crash risk. *International Review of Financial Analysis, 82*, 102196. https://doi.org/10.1016/j.irfa.2022.102196

Kaplan, S. N., & Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics, 112*(1), 169-215. https://doi.org/10.1162/003355397555163

Kindermann, B., Beutel, S., de Lomana, G. G., Strese, S., Bendig, D., & Brettel, M. (2021). Digital orientation: Conceptualization and operationalization of a new strategic orientation. *European Management Journal, 39*(5), 645-657. https://doi.org/10.1016/j.emj.2020.10.009

Kong, D., Lin, C., Wei, L., & Zhang, J. (2022). Information accessibility and corporate innovation. *Management Science, 68*(11), 7837-7860. https://doi.org/10.2139/ssrn.3291811

Li, C., Yan, X., & Song, M. (2020). Fintech and corporate innovation: Evidence from listed companies on the new third board. *China Industrial Economics, 1*, 81-98.

Lundvall, B. A. (1992). *Towards a theory of innovation and interactive learning*. New York: Pinter.

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.
McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60-68.
Mukherjee, A., Singh, M., & Žaldokas, A. (2017). Do corporate taxes hinder innovation? *Journal of Financial Economics, 124*(1), 195-221. https://doi.org/10.2139/ssrn.2585458
Nanda, R., & Rhodes-Kropf, M. (2013). Investment cycles and startup innovation. *Journal of Financial Economics, 110*(2), 403-418. https://doi.org/10.2139/ssrn.1950581
Reimers, I., & Waldfogel, J. (2021). Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *American Economic Review, 111*(6), 1944-1971. https://doi.org/10.3386/w26776
Song, Z., Storesletten, K., & Zilibotti, F. (2011). Growing like China. *American Economic Review, 101*(1), 196-233.
Stiroh, K. J. (2002). Information technology and the US productivity revival: What do the industry data say? *American Economic Review, 92*(5), 1559-1576. https://doi.org/10.2139/ssrn.923623
Tambe, P., Hitt, L., Rock, D., & Brynjolfsson, E. (2020). NBER working paper No. w28285. https://doi.org/10.3386/w28285
Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118-144.
Whited, T. M., & Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies, 19*(2), 531-559. https://doi.org/10.2139/ssrn.410816
Wu, L., Lou, B., & Hitt, L. (2019). Data analytics supports decentralized innovation. *Management Science, 65*(10), 4863-4877. https://doi.org/10.2139/ssrn.3351982
Yuan, C., Xiao, T., Geng, C., & Sheng, Y. (2021). Digital transformation and division of labor between enterprises: Vertical specialization or vertical integration. *China Industrial Economics, 9*, 137-155. https://doi.org/10.2139/ssrn.4628066

APPENDIX

Appendix A. Digital dictionary.

Artificial intelligence	Artificial intelligence, intelligence, algorithms, intelligent machines, intelligent networks, intelligent terminals, intelligent technology, intelligent homes, intelligent grids, intelligent transportation, intelligent supply chains, knowledge management, brain-like computing, intelligent data analysis, intelligent customer service, intelligent investment, intelligent tourism, intelligent robots, machine learning
Blockchain	Blockchain, data security, information systems, technology development, industry chain, bits, coding, virtualization, OneNet, credit, distributed computing, digital currency
Cloud computing	Cloud computing, cloud services, cloud platforms, cloud storage, key technologies, cloud ecology, internet thinking, multi-party secure computing, platform economy, digital supply chain, cloudification, in-memory computing, open banking, heterogeneous data
Big data	Big data, data mining, data analysis, databases, public data, data processing, data management, data centers, data sharing, data services, data platforms, digital communication, text mining, data visualization
Digital technology applications	Informatization, digitalization, communication, networking, broadband, information security, cyberspace, e-government, e-commerce, internet security, technology transformation, energy network, automation, high-tech, monitoring network, multimedia, internet protocol, satellite communications, intelligent algorithms, industry 4.0, intelligent manufacturing, robots

Appendix B. Constructing an enterprise digitalization indicator.

Part 1: Description of digital dictionary

The primary stage in formulating a digital lexicon entails the creation of a compendium of terms specific to enterprise digitalization. Given the absence of a specialized lexicon within the digital economy realm, this research undertakes the task of constructing such a lexicon by leveraging the semantic underpinnings of national policies.

Pertinent digital facets associated with enterprise digitalization were identified through a meticulous examination of 30 seminal, national-level policy documents concerning the digital economy, spanning the period from 2011 to 2020, sourced from the official websites of the Central People's Government and the Ministry of Industry and Information Technology. After employing Python-based word segmentation techniques and manual scrutiny, 230 digitalization-oriented terms, occurring with a frequency threshold of no fewer than five instances, were discerned, thus delineating the enterprise digitalization terminology compendium.

Part 2: Constructing an enterprise digitalization indicator

In the subsequent stage, an examination was conducted on phrases pertinent to annual reports. Leveraging the "jieba" Chinese lexical database integrated into the Python package, the digital terminology lexicon undergoes augmentation. Employing machine learning methodologies, the discourse within the "Management Discussion and Analysis" segment of listed companies' annual reports was meticulously analyzed. This systematic approach enabled the determination of the prevalence of terms associated with corporate digitalization within the reports. The study uses a logarithm to obtain the overall index of the degree of digitalization of enterprises to account for the "right bias" in this type of data. Overall, this study provides a more comprehensive and accurate measure of digitalization at the micro-enterprise level, which can help inform policymakers, researchers, and practitioners to understand the current state and trends of enterprise digitalization.