China Journal of Accounting Research 17 (2024) 100374

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

China Journal of Accounting Research

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

Does big data tax administration expand bank credit loans?

Xinwu Li, Zixi Ling *, Zhe Li, Liyi Zhu

School of Accountancy Central University of Finance and Economics Beijing 100081, China

ARTICLE INFO

Article history:
Received 31 May 2023
Accepted 17 June 2024
Available online 10 July 2024

JEL:
D21
G30
H26

Keywords:
Tax Administration with Data
Bank Credit
Information Quality

ABSTRACT

The application of big data technology to global tax management is becoming increasingly widespread. China has been implementing increasingly mature technologies for tax governance using big data systems in recent years. By collecting data through web scraping on the earliest implementation times of big data tax administration in various provinces of China, we explore the relationship between big data tax administration and corporate bank credit in emerging markets. Our results show that big data tax administration enhances firms' ability to obtain bank loans. Mechanism tests indicate that big data tax administration improves the quality of corporate information disclosure, facilitating access to bank credit loans. We find that big data tax administration improves the corporate financing environment, enhancing the efficiency of resource allocation in the credit market.

1. Introduction

The digital economy has expanded rapidly around the world. Driven by continuous upgrades in Internet functionality and the widespread application of big data, significant changes are occurring in governments, corporate business models and in people's daily lives (Chen and Srinivasan, 2024). Taxation departments provide a good example of such changes, as big data technology is expanding the traditional auditing model. "Big data tax administration" combines big data with tax auditing; it involves acquiring big data from Internet platforms and integrating and comparing multiple sources of data (Bassey et al., 2022). Its implementation, which has become a new trend in national tax governance, reflects the modernization of national governance

* Corresponding author at: Central University of Finance and Economics, 39 South Xueyuan Road, Haidian District, Beijing, China.
E-mail address: lzx13522124054@163.com (Z. Ling).

https://doi.org/10.1016/j.cjar.2024.100374
1755-3091/

capabilities and systems within the tax system and significantly enhances the efficiency of governments' tax collection (Canares, 2016). A key question is whether, looking beyond strict tax enforcement, modernized tax governance can guide firms to improve the quality of their information disclosure, thereby enhancing the efficiency of resource allocation in capital markets? Examining this question yields insights relevant to governments worldwide who are implementing governance based on big data technology.

Bank loans are an important financial resource and whether they are allocated in a timely and appropriate manner is an important issue from both theoretical and practical perspectives. An extensive body of research discusses the various factors that influence firm credit, including corporate characteristics and the policy environment. However, few studies explore whether innovations and modernizations within the tax governance system can improve the financing environment for firms. Studying the effects and mechanisms of big data tax administration on corporate bank credit enhances the understanding of the social effects of modernized governance and provides empirical evidence for the motivational impact of "tax administration with data" on firms.

This paper examines the impact of big data tax administration on corporate bank credit in China. China is selected as the research setting for two reasons. First, developing countries such as China face greater difficulties than developed countries in successfully implementing e-government practices, such as big data tax administration. Indeed, Heeks (2005) suggests that the failure rate for e-government initiatives in developing countries could be as high as 85 %. Therefore, developing countries need successful case studies to build their confidence in e-governance based on big data technology. China is the largest developing country in the world and is actively promoting big data technology. The 14th Five-Year Plan elevates big data to a national strategy[1] and, in 2022, the State Council issued documents specifically emphasizing 'using big data to strengthen economic monitoring and early warning' and 'enhancing the precision level of supervision with digital means' to strengthen the construction of a digital government.[2] Thus, choosing China as a research setting provides an analysis of the impact of e-government on the allocation of credit resources in a representative developing country.

Second, the allocation of financial resources in China's capital markets is heavily influenced by government macro-level controls. For instance, China's state-owned banks frequently help the government implement its planned investment policies (Carpenter et al., 2021). However, high levels of information asymmetry between the government and enterprises can result in the misallocation of resources when implementing planned investment policies. Therefore, it is worthwhile exploring whether enhancing the digital capabilities of tax departments with big data technology can open up multiple channels of information and make corporate information more public than at present. Increasing the transparency of corporate information will enhance the government's ability to optimize decisions on corporate planned investments. Therefore, choosing China as a setting for this research allows us to elucidate the achievement of optimal resource allocation in the capital market from the perspective of government macro-control.

Initially, to verify the practicality of our research, we conducted interviews with relevant enterprises on the topic of big data tax administration. Notably, during one interview with a financial technology firm, we asked about the impact of big data tax administration on micro-enterprises. The interviewee mentioned that it facilitated enterprises in obtaining bank credit, explaining the reasons as follows: "After big data tax administration, corporate information for tax and accounting has become much more standardized … Banks are actually very sensitive to the credibility of information. Once they detect an increase in the credibility of information, they feel much more at ease in granting loans … Big data governance methods have improved the social credit environment."

Building on this information about a potential correlation between big data tax administration and bank credit, we conduct theoretical analysis and collect practical evidence. We find that the impact of big data tax administration on firm credit may involve opposing effects. From the perspective of optimizing the information environment, in the context of big data tax administration, firms' motives to make opaque disclosures to conceal tax evasion behaviors decline. Moreover, firms develop corporate digitalization strategies when inter-

[1] Data source: https://www.gov.cn/zhengce/2021-12/01/content_5655197.htm.
[2] Data source: https://www.gov.cn/zhengce/content/2022-06/23/content_5697299.htm.

facing with tax authority systems, which improves the quality of their information disclosure. Simultaneously, the tax authorities achieve a system of "bilateral" integration with banks, which are third-party financial institutions. Banks can utilize part of the tax-related information about enterprises provided by the tax authorities, which reduces the information asymmetry between the banks and the enterprises, in turn relaxing the loan approval conditions for enterprises and promoting a more efficient allocation of credit resources. Therefore, enterprises will be able to obtain more bank credit after the implementation of local big data tax administration.

In addition, big data tax administration may encourage enterprises to obtain more bank credit by increasing their motivation to obtain funding. It curbs tax evasion by enterprises and increases their tax expenditure, which then reduces their operational cash flow. Under these circumstances, enterprises need more funds than before to meet cash flow expenditure, which could lead to them needing more credit financing from banks. As such, big data tax administration may positively impact firms' access to bank credit through improving the quality of information disclosure and strengthening the need for funding.

Conversely, however, the reduction in free cash flow could weaken enterprises' debt repayment capacity. Banks that can identify the increased risk associated with enterprise debt repayment may lower the credit limits for these enterprises, which could ultimately lead to a decrease in bank credit for the enterprises.

It should be clarified that big data tax administration is fundamentally different in nature from China's 'Golden Tax Phase III' tax collection and management project. The 'Golden Tax Phase III' is known as 'tax administration with invoices' and involves digitizing paper invoices to achieve an 'Internet-based' tax administration system, which allows all paper invoices and related tax activities to be monitored via the Internet. For instance, tax authorities can track input and output invoices under the same taxpayer identification number via the Internet to check if an enterprise is engaged in illegal activities, such as issuing false invoices. In contrast, big data tax administration employs big data technologies and applications to implement "tax administration with data," which breaks away from the traditional reliance on invoices and shifts from tracking tax-related activities to tracking economic activities. As an example, illegal "public-to-private" transfers do not generate invoices and cannot be detected and tracked by the Golden Tax project due to the lack of invoice documentation. However, under big data tax administration, when bank data are integrated with the tax system, the tax authorities can quickly capture such anomalous economic behaviors. Therefore, the "tax administration with data" that we discuss differs fundamentally from the "tax administration with invoices" of the Golden Tax Phase III project because the mechanisms of their effects on micro-enterprises are essentially distinct.

We explore the impact of big data tax administration on corporate credit acquisition. Our results show that big data tax administration can expand corporate bank loans, especially short-term bank loans. Mechanism tests reveal that big data tax administration affects bank credit by improving the corporate information environment. Further research suggests that the effects are more pronounced in firms subject to stronger (vs. weaker) financial constraints. Our conclusions provide empirical evidence for tax authorities to strengthen their cooperation with online third parties and actively promote big data tax administration. In addition, we uncover the unexpected effects of big data tax administration on micro-enterprises.

Our research makes three main contributions. First, in contrast with studies that focus on the direct expected effects of big data tax administration on the fairness of regional tax burdens and on corporate tax compliance, our study explores the spillover effects of big data tax administration. To the best of our knowledge, it is the first study to do so. Our research finds that big data tax administration can enhance enterprises' ability to obtain bank loans. We enrich the research on the economic consequences of big data tax administration and provide the first empirical evidence of the economic consequences of modernizing the governance system.

Second, whereas some studies examine the impact of existing micro-behaviors on enterprises' ability to obtain bank loans, our paper is one of the few to explore the impact of tax governance modernization, driven by the digital economy, on corporate credit capacity. We provide a new perspective on the factors affecting corporate financing capabilities.

Third, from a practical perspective, our paper explores the sustainability of strengthening modern tax administration, which has significant real-world relevance. Studies mainly focus on the benefits of big data tax administration to tax authorities, but we uncover the unexpected benefits for the corporate financing environment. We show that stricter tax supervision can have beneficial effects on enterprises, providing a theoretical basis for emerging market countries to implement "tax administration with data" practices.

2. Theoretical analyses and development of hypotheses

Big data tax administration can enhance enterprises' ability to obtain bank credit by reducing information asymmetry and improving the quality of information disclosure. We suggest that in the context of big data tax administration, firms will reduce opportunistic disclosure behavior and enhance digital infrastructure, thereby improving the quality of information disclosure. Banks will achieve "bilateral" integration with the information channels of the tax authorities, allowing them to conveniently access high-quality enterprise disclosures. Hence, the big data tax administration may ultimately enhance the efficiency of banks' credit resource allocation and promote the ability of enterprises to obtain more bank credit financing than in the absence of such a tax administration system.

First, under a big data tax administration system, enterprises' incentives to conceal tax evasion activities through opaque information disclosure will decline. When information asymmetry exists, executives can hide complex tax evasion activities within opaque information disclosures, allowing covert tax evasion to go unnoticed by external information users (Desai and Dharmapala, 2009). However, with the implementation of big data tax administration, tax authorities regulate corporate tax behavior more strictly, which improves corporate tax compliance (Pomeranz, 2015). In this scenario, the transparency of the corporate information environment is enhanced (Sun and Shi, 2022), making complex tax evasion behaviors more detectable, and illegal tax evasion and avoidance activities easier to discover than under a traditional tax administration system. Consequently, there is an increased likelihood of enterprises being penalized for tax violations. When concealing information related to tax evasion does not reduce corporate tax expenses, the motivation for enterprises to enhance transparency in information disclosure increases. Research reveals that higher transparency in information disclosure can improve enterprises' financing capabilities and reduce the interest rates on their bank loans (Chiu et al., 2018; Wang and Zeng, 2019). Therefore, big data tax administration can enhance enterprises' ability to obtain bank loans by strengthening their motivation to make high-quality information disclosures.

Second, big data tax administration enhances the digitalization level of enterprises, thereby raising the quality of information. Big data tax administration cannot be achieved solely through government efforts, but also requires the cooperation of enterprises. The government, by acquiring big data platforms and auditing technologies through procurement and other means, aims to better integrate its tax administration system with the data systems of enterprises. To align with the tax authorities' big data tax administration, enterprises must enhance their own digitalization level to integrate their data systems with the tax authorities' auditing systems (Du and Wang, 2023). As enterprises improve their digitalization, previously unstandardized data embedded in various processes are excavated and transformed into effective, comparable information outputs, enhancing the quality of information disclosure. The higher the quality of an enterprise's information disclosure, the greater is its creditworthiness in the eyes of banks and the lower its debt financing costs (Li and Wang, 2011). Thus, in the context of a big data tax administration that includes tax-related information, the more comprehensive and higher quality the corporate information disclosures made by enterprises, the more likely it is that the enterprises will be favored by banks when they seek to obtain credit.

Thus, we ask the following question: after enterprises have improved the quality of their information disclosure, will banks be able to access tax-related and other relevant information about enterprises more conveniently through the big data tax administration system? We obtain evidence to answer this question by searching online media sources and interviewing tax authority personnel.

From the perspective of the tax authorities, we find that tax authorities actively share tax-related information with third-party agencies and financial service institutions by leveraging big data technology. For instance, according to a report by the China Taxation News on big data tax administration in Shandong Province,[3] the Shandong Provincial Department of Finance started building an integrated tax information-

[3] Data source: https://news.sdufe.edu.cn/info/1022/15589.htm.

sharing platform in 2016, breaking down the information barriers between tax authorities and third parties. In 2017, the Shandong Provincial Local Taxation Bureau collaborated with the Provincial Insurance Regulatory Bureau to optimize the process of collecting and paying vehicle and vessel taxes, partnered with the Provincial Department of Housing and Urban–Rural Development to link online second-hand house contract prices directly with tax collection data and worked with the Provincial Price Bureau to establish a third-party public welfare tax dispute and relief mechanism. Banks, as important financial service institutions, are naturally part of this information-sharing network. The director of the financial bureau in a prefecture-level city in Shandong Province states that banks are authorized to use corporate tax credit rating information from the shared platform to identify quality clients. As of December 2017, the Shandong Local Taxation Bureau had signed "tax–bank interaction" agreements with 17 municipal bureaus, 178 county (city, district) bureaus and development zone branches and 895 banks, enabling 11,400 enterprises to secure loans worth 39.28 billion yuan.

In addition, we interviewed tax authority personnel in a prefecture-level city in Jiangsu Province. When asked if banks could access tax-related information about enterprises through the tax authorities, the official stated, "Banks can obtain some tax-related information about enterprises. For example, in our system integration with Bank A, the tax authorities provided the bank with information about the enterprise's export tax refund amount. The bank can use this information to understand the enterprise's operational status and assess its loan requirements, thereby enhancing the bank's credit resource allocation efficiency." Thus, it is evident that after the implementation of big data tax administration, tax authorities can bilaterally open up information channels with banks and other third parties, providing them with certain tax-related information to help enhance the efficiency of credit resource allocation by banks.

Big data tax administration drives banks to establish more convenient platforms for information communication and transmission, enabling them to obtain more comprehensive and accurate corporate information than before. For instance, according to a report by China UnionPay on Hubei Bank, in 2019, Hubei Bank launched a "Tax Easy Loan Platform" based on big data tax administration. This platform uses open application programming interface (API) technology and digital technology to improve data collection mechanisms and obtain more comprehensive external data about enterprises, including tax, business registration and credit information. By the end of 2021, the "Tax Easy Loan Platform" had issued more than 22,000 loans, totaling 2.798 billion yuan. This demonstrates that in the context of big data tax administration, banks actively participate in the construction of big data platforms to better integrate their systems with the tax authorities' systems.

In summary, based on the theoretical analysis and practical evidence of improved information disclosure quality, we contend that the quality of corporate information disclosure is enhanced in the context of big data tax administration. Furthermore, the tax authorities' systems in regions implementing big data tax administration become integrated with banking systems, achieving a bilateral information flow. Consequently, banks can access high-quality corporate information, which ultimately enhances the efficiency of banks' credit resource allocation and promotes greater access to bank credit financing for enterprises.

When a big data tax administration system is implemented, enterprises' need for and motivations to obtain funds may rise, leading them to seek more bank credit. Allingham and Sandmo (1972) develop an A–S deterrence model, which demonstrates that the optimal tax evasion choices of enterprises are related to the probability of being penalized and risk aversion preferences. Therefore, the more stringent the monitoring of corporate tax evasion and tax avoidance behaviors, the lower the motivations for tax evasion and the higher the tax compliance. Studies find that in an environment of big data tax administration, enterprises will increase their tax compliance (Sun and Shi, 2022). Furthermore, as tax avoidance behaviors decrease, tax expenses correspondingly increase, reducing the enterprises' operating cash flow and increasing their debt pressures. Thus, with reduced free cash flow and increased debt pressure, enterprises motivations to obtain funding rise, prompting them to seek more credit financing from banks, which is ultimately reflected in an increase in the scale of bank credit obtained by enterprises.

Potentially, however, there may be opposing effects arising from this mechanism that increase enterprises' need for funds. As noted, when enterprises reduce tax evasion due to the implementation of big data tax administration, their debt pressure increases. If banks are strongly regulated and identify the decrease in the enterprises' debt repayment capacity, they may reduce the credit limits of such enterprises, which ultimately weakens the enterprises' ability to obtain bank loans (Ivanov and Wang, 2023). Based on the above analysis, we propose our hypothesis in a competing form:

H1a: Following implementation of big data tax administration, the bank credit resources obtained by enterprises significantly increase.
H1b: Following implementation of big data tax administration, the bank credit resources obtained by enterprises significantly decrease.

3. Research design

3.1. Sample and data sources

We obtain data from China's A-share listed public firms during the period of 2014 to 2021. We adopt the approach of Sun and Shi (2022), conducting information retrieval through search engines such as Baidu and Bing and using web crawlers to capture and identify news content. This enables us to identify the earliest year in which the various provinces implement tax administration using big data. This process enables us to create a big data tax administration variable, *Bigdata*. The related financial indices and governance variables that we use in the study are sourced from the China Stock Market and Accounting Research database. After excluding observations with missing relevant indices, we obtain 23,007 firm-year observations. We winsorize all of the continuous variables at the 1 % and 99 % levels.

3.2. Variable definitions and regression model

The dependent variable in our study is the enterprises' bank credit, *Loan_all*, calculated as the new loans acquired by the enterprise in the current year divided by the total assets at the beginning of the year. The core explanatory variable is big data tax administration. We confirm the earliest year of implementation of big data tax administration for each region through a textual analysis of online news. Following the approach of Sun and Shi (2022), we conduct identification matching on network news with three sets of vocabulary. The first set includes terms relating to data-carrying platforms, such as "internet" and "database." The second set comprises specific technical method terms, such as "big data" and "crawler." The third set includes terms related to tax administration, such as "tax administration" and "tax collection." When news from a region contains terms from all three groups, this indicates that big data tax administration has been implemented in that region. Thereby, we define the dummy variable *Bigdata*, which equals 1 when big data tax administration

Table 1
Variable definitions.

Name	Definitions
Dependent Variable	
Loan_all	Incremental bank loans scaled by total assets
Independent Variable	
Bigdata	Dummy variable that equals 1 if big data tax administration is implemented in the province in which the enterprise is located in the current year or thereafter, and 0 otherwise
Control Variables	
Size	Natural logarithm of total assets
Roa	Earnings scaled by total assets
Lev	Total liabilities scaled by total assets
Rdfee	R&D expenditure scaled by operating income
Top1	The largest shareholder's shareholding percentage
Dir	Natural logarithm of the number of board directors + 1
Indir	Independent directors scaled by board directors
Dual	Dummy variable that equals 1 if the chair also serves as the general manager, and 0 otherwise
GT_Phase	Dummy variable that equals 1 if 'Golden Tax Phase III' is implemented in the province in which the enterprise is located

is implemented in an enterprise's location, and 0 otherwise. The definitions of the other variables are provided in Table 1.

Model (1) is used to test our Hypotheses 1a and 1b. The dependent variable, *Loan_all,* is a proxy for bank credit, as previously described. We use *Bigdata* as the core explanatory variable in our primary tests. It is assigned a value of 1 when big data tax administration is implemented in the province where the enterprise is registered, and 0 when it is not yet implemented. The model controls for firm size (*Size*), return on assets (*Roa*), leverage (*Lev*), the largest shareholder's holdings (*Top1*), board size (*Dir*), the proportion of independent directors (*Indir*) and combined chairperson and general manager roles (*Dual*). To avoid interference from other major policies that might affect the intensity of taxation, our model also controls for the implementation of the "Golden Tax Phase III" program (*GT_Phase*). Furthermore, Model (1) controls for year and firm fixed effects.

$$Loan\_all = \alpha + \beta_1 Bigdata + \beta_2 \sum Controls + \sum Firm + \sum year + \mu \tag{1}$$

4. Empirical results

4.1. Descriptive statistics

The descriptive statistics of the main variables are presented in Panel A of Table 2. The average ratios of total loans (*Loan_all*), short-term loans (*Loan_st*) and long-term loans (*Loan_lt*) are 16.9 %, 10.4 % and 4.9 %, respectively. The mean of *Bigdata* is 0.398, indicating that 39.8 % of the observations in the sample are influenced by big data tax administration. The average company size (*Size*) is 22.419. The means of the return on assets (*Roa*) for the leverage ratio (*Lev*) and the largest shareholder's holding (*Top1*) are 0.026, 0.468 and 33.4 %, respectively. The means of board size (*Dir*), the proportion of independent directors (*Indir*) and the

Table 2
Descriptive statistics for the main variables.

Panel A: Descriptive statistics for the full sample

Variables	*Mean*	*Std.Dev.*	*Min*	*P50*	*Max*
Loan_all	0.169	0.136	0.001	0.142	0.619
Bigdata	0.398	0.489	0.000	0.000	1.000
Size	22.419	1.350	19.863	22.221	26.501
Roa	0.026	0.078	–0.383	0.033	0.184
Lev	0.468	0.200	0.093	0.457	0.979
Top1	0.334	0.148	0.086	0.310	0.750
Dir	2.234	0.178	1.792	2.303	2.773
Indir	0.377	0.054	0.333	0.364	0.571
Dual	0.696	0.460	0.000	1.000	1.000
GT_Phase	0.692	0.462	0.000	1.000	1.000

Panel B: Univariate difference analysis

Variables	***Bigdata*** = 0 *Mean*	***Bigdata*** = 1 *Mean*	*DIFF*
Loan_all	0.162	0.179	–0.017***
Size	22.244	22.684	–0.440***
Roa	0.030	0.018	0.012***
Lev	0.450	0.496	–0.046***
Top1	0.339	0.327	0.013***
Dir	2.235	2.234	0.001
Indir	0.376	0.379	–0.003***
Dual	0.683	0.715	–0.032***
GT_Phase	0.568	0.879	–0.311***

Note: In Panel B, ***, ** and * in the group differences column indicate significance at the 1%, 5% and 10% levels, respectively.

incidence of dual chairperson–general manager roles (*Dual*) are 2.234, 37.7 % and 69.6 %, respectively. The proportion of observations that have implemented "Golden Tax Phase III" in the sample is 69.2 %.

Panel B of Table 2 reports the univariate analysis of differences for *Bigdata*. As theorized, big data tax administration is expected to promote corporate bank credit activities. The sample is divided into subsamples based on whether firms have been impacted by big data tax administration. The results show that observations impacted by big data tax administration have larger bank loans (*Loan_all*) compared with those not affected by it. Therefore, the inter-group differences in results generally align with the expectations set out in the theoretical analysis. Furthermore, firms in the subsample that have been affected by big data tax administration tend to have a larger company size (*Size*), higher leverage ratio (*Lev*) and a lower proportion of shares held by the largest shareholder (*Top1*) than firms not impacted.

4.2. Main results

The baseline regression results are shown in Table 3. Columns (1) and (2) report the regression results for the impact of big data tax administration on bank credit, with and without control variables, respectively. In both columns, the coefficients for *Bigdata* are positive and significant, indicating that big data tax administration significantly enhances the ability of enterprises to acquire bank loans. Our results support Hypothesis H1a, which states that big data tax administration can promote enterprise bank credit activities, enabling enterprises to obtain more bank loans than without big data tax administration.

Table 3
Baseline regression results.

Variables	(1)	(2)
	Loan_all	
Bigdata	0.020**	0.017**
	(2.528)	(2.219)
Size		–0.032***
		(–4.714)
Roa		–0.239***
		(–5.832)
Lev		0.494***
		(18.354)
Top1		–0.054
		(–1.161)
Dir		–0.022
		(–0.630)
Indir		–0.059
		(–0.645)
Dual		–0.004
		(–0.442)
GT_Phase		–0.021*
		(–1.729)
Constant	0.177***	0.749***
	(24.190)	(4.376)
Firm	YES	YES
Year	YES	YES
N	23,007	23,007
Within R^2	0.002	0.032

Note: In this table and all tables below, ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively. Values shown in parentheses are *t* values.

5. Robustness tests

5.1. Alternative measures of the independent variable

Based on the theoretical analysis and mechanism tests of our paper, the implementation of big data tax administration can break down data barriers between tax authorities and enterprises, enabling a more efficient transmission of more comprehensive tax-related data to the tax authorities than a traditional tax administration system. Simultaneously, the information barriers between tax authorities and banks break down, allowing for a bidirectional flow of information between the banks and tax authorities. Ultimately, this enhances the quality of corporate information disclosure, which improves information capture by banks and thus optimizes the efficiency of banks' credit resource allocation to enterprises.

Following this specific theoretical logic, in the robustness tests, we redefine the independent variable based on identifying and textually analyzing news articles on big data tax administration. First, we expand the three groups of vocabulary used to identify the original explanatory variable *Bigdata* by adding a fourth group that describes the specific tax administration methods used by government agencies or third parties to facilitate tax-related information channels; we identify vocabulary such as "data integration," "breaking down barriers" and "system integration." Local regions in which the local news articles match the criteria for the four groups of vocabulary criteria are defined as having implemented big data tax administration. Second, for each local region, the year in which such news articles are published is used to define the year of implementation of big data tax administration in that region. For local enterprises, this variable takes a value of 1 for that year and subsequent years, and 0 otherwise, resulting in a new alternative variable for big data tax administration, *Bigdata_alternative*.

The results after incorporating *Bigdata_alternative* into the main regression model and re-running the regression are shown in Table 4. It is evident that *Bigdata_alternative* is positive and significant at the 5 % level, and thus our main conclusions hold.

Table 4
Alternative measures for the independent variable.

Variables	(1) ***Loan all***
Bigdata_alternative	0.020**
	(2.383)
Size	–0.031***
	(–4.673)
Roa	–0.238***
	(–5.813)
Lev	0.494***
	(18.374)
Top1	–0.054
	(–1.161)
Dir	–0.023
	(–0.669)
Indir	–0.061
	(–0.658)
Dual	–0.003
	(–0.406)
GT_Phase	–0.028**
	(–2.091)
Constant	0.747***
	(4.366)
Firm	YES
Year	YES
N	23,007
Within R^2	0.032

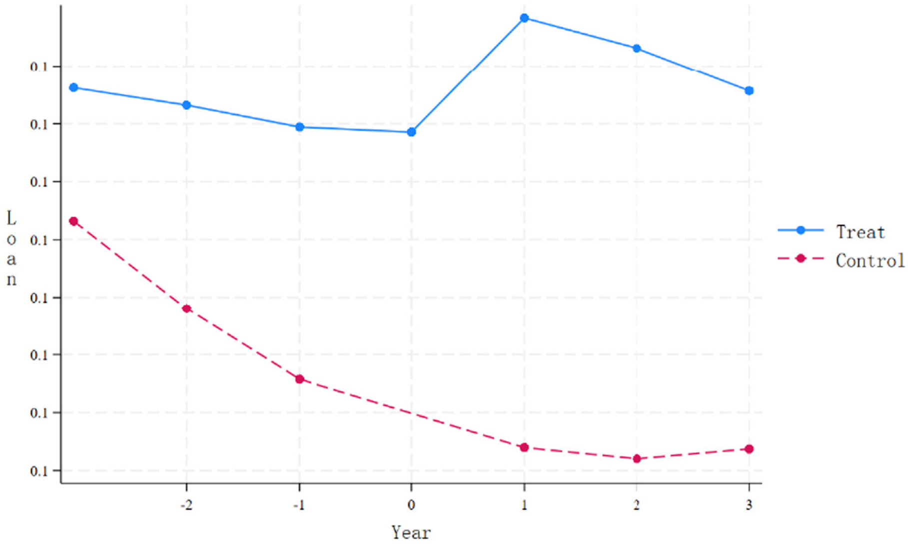

Fig. 1. Parallel trend test.

5.2. Parallel trends test

In this section, we test whether the treatment and control groups exhibit the same temporal trends before the implementation of big data tax administration. Following the approach of Jacobson et al. (1993) and Wang and Ge (2022), we aggregate and statistically analyze observations from 2 years before to 3 years after the shock, comparing the mean differences between the treatment and control groups relative to year –2 to year 0 around the shock year.[4] We focus on the differences between the treatment and control groups before the start of the shock to test whether there are significant time-related effects on corporate credit before the impact commenced. Fig. 1 reports the test results; the blue solid (red dashed) line represents the mean values of observations in the treatment (control) group before and after the shock. The results indicate that in the years before the implementation of big data tax administration, the corporate credit conditions of the treatment and control groups present broadly parallel trends, suggesting that the baseline results meet the assumption of parallel trends over time.

5.3. Adding city fixed effects

During the sample period, there are few instances of enterprises changing their operating locations. However, considering that regional factors are of strong importance to our analysis, changes in the location of enterprises could potentially affect the validity of our baseline results. Furthermore, it is possible that relevant tax administration policies other than other big data tax administration are introduced in an enterprise's location during the sample period (Hu et al., 2022). To eliminate this potential factor, we rerun the baseline regression, controlling for city fixed effects. Table 5 presents the results. After including city fixed effects, the coefficient on *Bigdata* remains positive and significant, indicating that the conclusions of the baseline tests are robust.

[4] There are few observations outside the period 2 years before or 3 years after the shock. Therefore, for simplicity and conciseness, we aggregate the data beyond the second year before the shock with the data for the second year before the shock. Similarly, data from beyond the third year after the shock are aggregated with the data for the third year after the shock.

Table 5
Adding city fixed effects.

Variables	(1)	(2)
	Loan_all	
Bigdata	0.046***	0.014**
	(7.727)	(2.320)
Size		–0.011***
		(–4.749)
Roa		–0.335***
		(–9.711)
Lev		0.484***
		(31.738)
Top1		0.019
		(1.119)
Dir		0.012
		(0.695)
Indir		–0.001
		(–0.015)
Dual		–0.000
		(–0.038)
GT_Phase		–0.010
		(–0.979)
Constant	0.216***	0.201***
	(9.907)	(3.098)
City	YES	YES
Year	YES	YES
N	23,007	23,007
Adj R^2	0.0242	0.0929

5.4. Propensity score matching (PSM)

The results of the baseline tests may be influenced by differences in control variables between groups. To eliminate interference from related factors, we employ the counterfactual inference method of propensity score matching (PSM). After performing a 1:1 matching without replacement, we obtain 6,535 observations in the treatment group (*Bigdata* = 1) and 6,535 observations in the control group (*Bigdata* = 0). Panel A of Table 6 shows the effects of the PSM matching. It can be observed that PSM effectively reduces inter-group differences in the sample compared with the univariate difference analysis results in Table 2. Panel B of Table 6 presents the regression results after using PSM. The coefficients on *Bigdata* remain positive and significant in columns (1) and (2), indicating that the conclusions of the baseline tests remain robust after PSM.

5.5. Placebo test

In the baseline regression results, it is possible that the impact of *Loan_all* on *Bigdata* is driven by random factors that we have overlooked. To dispel such concerns, we conduct a placebo test using the following basic method. The original *Bigdata* variable values are shuffled and randomly assigned to each firm-year observation to create a new dummy variable, *Bigdata_random*. Next, we rerun the regression to determine the coefficient of *Loan_all* on *Bigdata_random*, repeating this process 50, 100 and 200 times. The results are shown in Table 7. The coefficients of *Loan_all* on *Bigdata_random* are –0.003, 0.000 and –0.002 for the 50, 100 and 200 repetitions, respectively. The probabilities of this coefficient being significant and positive or negative are both small and roughly equal. The above analysis indicates that the random variable *Bigdata_random* does not have an effect on bank credit, thus confirming the robustness of the main results.

Table 6
Propensity score matching.

Panel A: Inter-group differences after PSM

Variables	***Bigdata*** = 0 (N = 6,535) *Mean*	***Bigdata*** = 1 (N = 6,535) *Mean*	*DIFF*
Size	22.403	22.424	–0.021
Roa	0.024	0.025	–0.001
Lev	0.468	0.468	0.000
Top1	0.328	0.331	–0.003
Dir	2.230	2.233	–0.003
Indir	0.378	0.377	0.000
Dual	0.704	0.694	0.010
GT_Phase	0.841	0.841	0.000

Panel B: Regression results after PSM

Variables	(1)	(2)
	Loan_all	
Bigdata	0.036**	0.039**
	(2.290)	(2.504)
Size		–0.060***
		(–4.134)
Roa		–0.236***
		(–3.029)
Lev		0.463***
		(8.420)
Top1		–0.092
		(–0.947)
Dir		0.025
		(0.346)
Indir		–0.069
		(–0.367)
Dual		–0.006
		(–0.381)
GT_Phase		–0.005
		(–0.204)
Constant	0.177***	1.283***
	(8.833)	(3.500)
Firm	YES	YES
Year	YES	YES
N	13,070	13,070
Within R^2	0.001	0.016

6. Further analysis

6.1. Mechanism analysis

The theoretical analysis of our paper suggests that big data tax administration improves the quality of corporate information, increasing banks' trust in corporate information and thereby promoting bank lending to corporations. In the logical framework and regression results of the previous sections, the improvement in corporate information quality plays a dominant role in increasing the ability of firms to obtain bank loans. To further verify the importance of the mechanism of enhanced information disclosure quality over the capital demand mechanism, in this section, we validate the mediating mechanism of corporate information disclosure quality.

The quality of corporate information disclosure is measured using the level of detail in the disclosures (*DQ*) and the information disclosure evaluation ratings (*DE*) provided by the Shenzhen and Shanghai Stock

Table 7
Placebo test.

Variables	*Bigdata_random*
50 repetitions	
The mean coefficient β on *Bigdata_random*	–0.000
[%β > 0 & α ≤ 5 %; %β < 0 & α ≤ 5 %]	[0.0 %; 0.0 %]
[%β > 0 & α ≤ 1 %; %β < 0 & α ≤ 1 %]	[0.0 %; 0.0 %]
100 repetitions	
The mean coefficient β on *Bigdata_random*	0.000
[%β > 0 & α ≤ 5 %; %β < 0 & α ≤ 5 %]	[2.0 %; 2.0 %]
[%β > 0 & α ≤ 1 %; %β < 0 & α ≤ 1 %]	[0.0 %; 0.0 %]
200 repetitions	
The mean coefficient β on *Bigdata_random*	–0.000
[%β > 0 & α ≤ 5 %; %β < 0 & α ≤ 5 %]	[0.0 %; 0.5 %]
[%β > 0 & α ≤ 1 %; %β < 0 & α ≤ 1 %]	[0.0 %; 0.0 %]

Exchanges for listed companies. Following Chen et al. (2015), we divide the accounts in China's financial statements (k) into five categories: current assets, non-current assets, current liabilities, non-current liabilities and equity. Variable *DQ* is calculated using Eq. (2), where Non-Missing Items represents the number of accounts that are not missing, Total Items represents the total number of accounts in category k, Assets_k refers to the total amount of non-missing account items in category k and Total Assets represents the total assets of the enterprise. Following Quan and Wu (2010), we utilize the quality rating for information disclosure determined by the Shenzhen and Shanghai Stock Exchanges. We assign a value of 1 to enterprises that achieve a rating of qualified or above in a given year, and 0 to enterprises rated as unqualified, resulting in the variable *DE*.

Table 8
Mechanism tests.

Variables	(1) *DQ*	(2) *DE*
Bigdata	0.004*	0.019***
	(1.869)	(2.732)
Size	0.005***	0.028***
	(2.599)	(4.705)
Roa	0.024**	0.497***
	(2.015)	(13.606)
Lev	–0.025***	–0.440***
	(–3.190)	(–18.355)
Top1	0.014	0.223***
	(1.020)	(5.331)
Dir	0.016	0.027
	(1.601)	(0.867)
Indir	0.050*	0.051
	(1.858)	(0.618)
Dual	–0.000	–0.024***
	(–0.010)	(–3.340)
GT_Phase	–0.512***	0.387***
	(–141.682)	(35.424)
Constant	0.696***	0.019***
	(13.790)	(2.732)
Firm	YES	YES
Year	YES	YES
N	23,007	23,007
Within R^2	0.850	0.315

$$\mathrm{DQ} = \sum\nolimits_{k=1}^{5}\left\{\left(\frac{\#\mathrm{Non}-\mathrm{Missing\ Items}}{\#\mathrm{Total\ Items}}\right)_{k} \times \frac{\mathrm{Assets}_{k}}{\mathrm{Total\ Assets}}\right\}\tilde{A}\cdot 2 \tag{2}$$

We use Model (2) for the mechanism tests, incorporating *DQ* and *DE* as dependent variables for the regression analysis. The results are shown in Table 8. Both *DQ* and *DE* have positive and significant regression coefficients on *Bigdata*, indicating that big data tax administration significantly improves the quality of corporate information disclosure, which in turn promotes enterprises' access to bank credit. This finding is consistent with our theoretical analysis.

6.2. Test to exclude alternative hypotheses

According to the theoretical analysis, after the implementation of big data tax administration, enterprises might improve the quality of information disclosure and banks could obtain more comprehensive tax information, reducing information asymmetry and thereby enabling enterprises to secure more credit financing. Conversely, however, the intensification of tax collection efforts reduces the possibilities for tax evasion and avoidance, thus increasing corporate tax expenditure and decreasing the disposable cash flow of enterprises. Facing cash flow pressures, enterprises might seek more credit financing based on funding needs. We posit that if enterprises seek more credit financing based on funding needs, those with weaker financial constraints will obtain more credit to meet needs for free cash flow. We use two methods to confirm the information disclosure mechanism and exclude this alternative hypothesis.

We use the Kaplan and Zingales (KZ) index, the nature of property rights and the political connections of enterprises to characterize the level of financial constraints faced by enterprises. First, a higher KZ index indicates stronger financial constraints (Kaplan and Zingales, 1997); we use the median industry-year KZ index to classify the sample into two subsamples above and below the industry-year median. The subsample above (be-

Table 9
Heterogeneity analysis of financial constraints.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Loan all					
	Higher KZ index	Lower KZ index	Non-SOEs	SOEs	Not politically connections	Politically connected
Bigdata	0.034**	–0.000	0.005***	0.035*	0.027**	0.001
	(2.146)	(–0.268)	(2.624)	(1.709)	(2.475)	(0.184)
Size	–0.051***	0.005***	–0.003	–0.080***	–0.032***	–0.028***
	(–4.254)	(2.769)	(–1.575)	(–4.327)	(–3.321)	(–9.824)
Roa	–0.271***	–0.100***	–0.098***	–0.671***	–0.108*	–0.120***
	(–3.954)	(–6.891)	(–9.753)	(–5.197)	(–1.915)	(–7.480)
Lev	0.485***	0.451***	0.448***	0.590***	0.472***	0.553***
	(9.527)	(57.943)	(63.666)	(7.605)	(12.489)	(49.650)
Top1	–0.081	0.001	–0.010	–0.115	–0.102	–0.039**
	(–0.902)	(0.113)	(–0.722)	(–0.966)	(–1.451)	(–2.245)
Dir	–0.033	–0.020***	–0.018*	0.001	–0.002	0.002
	(–0.477)	(–2.621)	(–1.871)	(0.015)	(–0.039)	(0.180)
Indir	–0.093	–0.005	–0.042	–0.080	–0.118	–0.046
	(–0.514)	(–0.254)	(–1.620)	(–0.360)	(–0.922)	(–1.261)
Dual	–0.009	0.003	–0.002	–0.003	–0.008	–0.001
	(–0.577)	(1.555)	(–1.095)	(–0.141)	(–0.722)	(–0.172)
GT_Phase	–0.033	–0.024***	–0.027***	–0.002	–0.028	–0.006
	(–1.378)	(–8.219)	(–7.491)	(–0.078)	(–1.586)	(–1.367)
Constant	1.244***	–0.117**	0.099**	1.803***	0.750***	0.560***
	(3.880)	(–2.508)	(2.122)	(3.968)	(3.093)	(7.775)
Diff	0.034**(p = 0.022)		0.029**(p = 0.042)		0.027**(p = 0.012)	
Firm	YES	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES	YES
N	12,870	10,137	14,345	8,307	14,968	7,914
Within R^2	0.023	0.383	0.349	0.025	0.021	0.381

low) the median faces stronger (weaker) financial constraints. Second, state-owned enterprises (SOEs) have more stable channels for obtaining loans and face less uncertainty than non-SOEs. Consequently, in general, they face weaker financial constraints than non-SOEs (Almeida et al., 2004; Fang, 2007). We classify non-SOEs as having strong financial constraints and SOEs as having weak constraints. Third, enterprises with political connections tend to have weaker financial constraints and are more likely to obtain bank credit than those lacking such connections (Claessens et al., 2008). Political connections can also play a governance role in enterprises, alleviating problems with information quality (Yu et al., 2012). Therefore, we classify enterprises without (with) political connections as having strong (weak) financial constraints.

The results of the group regression based on the strength of financial constraints are shown in Table 9. They indicate that big data tax administration significantly increases the ability of enterprises with stronger financial constraints, non-SOEs and enterprises without political connections to obtain bank credit. Conversely, enterprises with weak financial constraints, SOEs and those with political connections do not obtain more bank credit. Therefore, this confirms that the motivation for enterprises to obtain more bank credit after big data tax administration is not an increase in funding needs, thus mitigating the potential interference from the alternative hypothesis proposed in our research.

Second, to exclude the possibility that an increase in tax expenses due to big data tax administration leads to increased enterprise loan demand, we design alternative dependent variables to test this mechanism. We calculate *Loan_Adj* by subtracting the current year's income tax expense from the credit funds of the same year, then adding 1 and taking the logarithm. In addition, we standardize the credit funds by calculating (credit funds of the current year – income tax expense of the current year) / credit funds of the current year, resulting in the *Loan_Adj_Ratio*. *Loan_Adj* and *Loan_Adj_Ratio* as alternative dependent variables. After rerunning the main regression model with these variables, the results are reported in Table 10. They show that after excluding the tax expense factor, the coefficients on *Bigdata* remain positive and significant. These results reject the alternative hypothesis of funding needs driving the increase in the scale of bank credit.

Table 10
Excluding the impact of tax expenses.

Variables	(1) ***Loan_Adj***	(2) ***Loan_Adj_Ratio***
Bigdata	0.033**	0.231**
	(1.963)	(2.050)
Size	1.087***	–0.027
	(75.935)	(–0.277)
Roa	–0.325***	0.252
	(–3.738)	(0.422)
Lev	3.219***	0.278
	(55.665)	(0.708)
Top1	0.003	–0.650
	(0.035)	(–0.950)
Dir	–0.072	–0.039
	(–0.978)	(–0.078)
Indir	–0.081	0.107
	(–0.416)	(0.080)
Dual	0.020	0.000
	(1.161)	(0.002)
GT_Phase	–0.222***	–0.437**
	(–8.420)	(–2.439)
Constant	–5.414***	0.753
	(–14.814)	(0.301)
Firm	YES	YES
Year	YES	YES
N	23,007	23,007
Within R^2	0.457	0.001

6.3. Regional-level analysis

The baseline tests use data at the firm-year level to examine the impact of big data tax administration on bank credit. However, the effects of big data tax administration are regional in nature, affecting the information quality of all local enterprises and the credit approval processes of banks generally. Therefore, in this section, we use region-year observations to further analyze the impact of implementing big data tax administration on local bank credit issuance. We replace the dependent variable with the natural logarithm of the total amount of new loans issued in the region for that year, *Loan_local*, and rerun the regression. Columns (1) and (2) of Table 11 report the results with and without control variables, respectively; both include province and year fixed effects. In columns (1) and (2), the coefficients of *Loan_local* on *Bigdata* are positive and significant, confirming that the regional government's application of big data tax administration promotes local credit activities at the macro level.

7. Conclusions and implications

The widespread adoption and application of the Internet and big data have reduced the cost of information flow, and the multi-party storage and sharing of big data has gradually become a new form of social capital. The use of big data by tax authorities not only improves the capabilities and efficiency of tax collection but also promotes the rational allocation of financial capital at the national level and facilitates enterprise financing. We investigate the effects of big data tax administration on corporate credit and find that implementing big data tax administration (1) has a positive impact on corporate bank credit; (2) improves the quality of corporate information, thereby promoting corporate credit and (3) has a more pronounced effect on enterprises with stronger (vs. weaker) financial constraints. The conclusions of our study are beneficial because they enhance academic understanding of the economic consequences of big data tax administration, and also have practical significance for tax authorities actively seeking external forces to improve tax collection procedures. Based on our theoretical and empirical analyses, we put forward the following policy recommendations.

First, the government should enhance cooperation between tax authorities and third-party online platforms, synchronously strengthening tax administration under the "delegation, management and service" framework. We show that by enhancing information collecting capabilities, big data tax administration improves the quality of corporate information disclosure, thereby optimizing corporate financing efficiency. In the context of the digital transformation of enterprises and the increasing application of big data in society, tax inspection departments should make full use of the data at the societal level, engaging in cooperation with online platforms and online assessment agencies. By utilizing the correlation between multi-party data and the data provided by enterprises, tax inspection departments can strengthen their verification of corporate information. This will help eliminate the corporate practices of inflating projects and manipulating information disclosure, and encourage and guide firms to improve their information quality.

Second, it is crucial to enhance the promotion of tax information technology and increase enterprises' willingness to improve information disclosure. Enhanced promotional guidance for enterprises is required to take full advantage of the governance role of big data tax administration in improving the financing environment

Table 11
Regional-level analysis.

Variables	(1)	(2)
	Loan_local	
Bigdata	0.797***	0.136*
	(6.700)	(1.832)
Controls		YES
Province		YES
Year		YES
N	374	374
Within R^2	0.618	0.893

for enterprises and the efficiency of resource allocation in society. Tax authorities should enhance communications with enterprises, focusing on providing explanations of and guidance regarding tax policies and should emphasize that "tax administration with big data" is not merely equivalent to "strict tax administration," but has benefits for building the enterprises' own credit and reputation systems.

Faced with a rapidly changing business environment and a gradually improving information taxation system, the value orientations and proactive attitudes of corporate decision-makers will determine whether they can respond positively and seize the opportunities to optimize and adjust their operations. As big data tax administration identifies and blocks many non-standard information disclosure practices, promotion and guidance by government departments for enterprises, along with the implementation of relevant supporting policies, are crucial in assisting enterprises to progress and achieve a win–win situation for both government and 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.

References

Allingham, M.G., Sandmo, A., 1972. Income tax evasion: a theoretical analysis. Tax.: Crit. Perspect. World Econ. 3 (1), 323–338.

Almeida, H., Campello, M., Weisbach, M.S., 2004. The cash flow sensitivity of cash. J. Financ. 59 (4), 1777–1804.

Bassey, E., Mulligan, E., Ojo, A., 2022. A conceptual framework for digital tax administration—a systematic review. Gov. Inf. Q. 39 (4) 101754.

Canares, M.P., 2016. Creating the enabling environment for more transparent and better-resourced local governments: a case of e–taxation in the Philippines. Inf. Technol. Dev. 22 (sup1), 121–138.

Carpenter, J.N., Lu, F., Whitelaw, R.F., 2021. The real value of China's stock market. J. Financ. Econ. 139 (3), 679–696.

Chen, S., Miao, B., Shevlin, T., 2015. A new measure of disclosure quality: the level of disaggregation of accounting data in annual reports. J. Account. Res. 53 (5), 1017–1054.

Chen, W., Srinivasan, S., 2024. Going digital: implications for firm value and performance. Rev. Acc. Stud. 29, 1619–1665.

Chiu, T.T., Guan, Y., Kim, J.B., 2018. The effect of risk factor disclosures on the pricing of credit default swaps. Contemp. Account. Res. 35 (4), 2191–2224.

Claessens, S., Feijen, E., Laeven, L., 2008. Political connections and preferential access to finance: the role of campaign contributions. J. Financ. Econ. 88 (3), 554–580.

Desai, M.A., Dharmapala, D., 2009. Corporate tax avoidance and firm value. Rev. Econ. Stat. 91 (3), 537–546.

Du, J., Wang, H., 2023. Application of synthetic data in the digital transformation of tax collection and administration. Tax. Res. 07, 62–69 (in Chinese).

Fang, J., 2007. Ownership, institutional environment, and allocation of credit funds. Econ. Res. J. 12, 82–92 (in Chinese).

Heeks, R., 2005. E–government as a carrier of context. J. Publ. Policy 25 (1), 51–74.

Hu, T., Chen, R., Tu, Z., 2022. Environmental information disclosure evaluation and market value: the impact of third–party institutions. J. World Econ. 45 (11), 150–176 (in Chinese).

Ivanov, I.T., Wang, J.Z., 2023. Bank supervision and corporate credit supply. Manag. Sci. 70 (5), 3338–3361.

Jacobson, L., LaLonde, R.J., Sullivan, D.G., 1993. Earnings losses of displaced workers. Am. Econ. Rev. 83 (4), 685–709.

Kaplan, S.N., Zingales, L., 1997. Do investment–cash flow sensitivities provide useful measures of financing constraints? Q. J. Econ. 112 (1), 169–215.

Li, Z., Wang, S., 2011. Monetary policy, information disclosure quality, and corporate debt financing. Account. Res. 288 (10), 56–62 (in Chinese).

Pomeranz, D., 2015. No taxation without information: deterrence and self–enforcement in the value added tax. Am. Econ. Rev. 105 (8), 2539–2569.

Quan, X., Wu, S., 2010. CEO power, information disclosure quality, and the volatility of company performance—An empirical study based on companies listed on the Shenzhen Stock Exchange. Nankai Manage. Rev. 13 (04), 142–153 (in Chinese).

Sun, K., Shi, L., 2022. The effects of enterprise internet use and big data tax administration. Econ. Res. J. 57 (05), 176–191 (in Chinese).

Wang, F., Ge, X., 2022. Does low-carbon transition impact employment? Empirical evidence from low-carbon city pilots. China Ind. Econ. 05, 81–99 (in Chinese).

Wang, X., Zeng, J., 2019. Annual report risk information disclosure and bank loan interest rates. J. Financ. Res. 01, 54–71 (in Chinese).

Yu, W., Wang, M., Jin, X., 2012. Political connections and financing constraints: information effects and resource effects. Econ. Res. J. 47 (09), 125–139 (in Chinese).