sustainability

MDPI

Article

Regional Big Data Application Capability and Firm Green Technology Innovation

Guixiang Cao [1], Xintong Fang [2,*], Ying Chen [3] and Jinghuai She [1]

1 College of Business Administration, Capital University of Economics and Business, Beijing 100070, China
2 School of Accounting, Capital University of Economics and Business, Beijing 100070, China
3 Business School, University of Sydney, Sydney, NSW 2006, Australia
* Correspondence: fangxintong@cueb.edu.cn

Abstract: This study aims to investigate the impact of regional big data application capability (RBDAC) on the green technology innovation (GTI) of manufacturing firms. Based on the data from Shanghai and Shenzhen A-share listed manufacturing firms in China from 2010 to 2020, the difference-in-differences method is used for the analysis. The results show that RBDAC can significantly improve the GTI in manufacturing firms. Further research shows that government subsidy and analyst coverage have strengthened the positive effect of RBDAC on GTI. Extensive analysis validates the heterogeneity of RBDAC in influencing the GTI based on financial constraints, tax administration strengths, regions, property rights, and top management team. The economic outcome test shows that RBDAC also improves firms' environmental, social, and governance performance. Our findings contribute to the literature on big data application capability and GTI, as well as provide practical enlightenment for manufacturing firms to engage in digital and green practices.

Keywords: big data; government subsidy; analyst coverage; GTI

check for updates

Citation: Cao, G.; Fang, X.; Chen, Y.; She, J. Regional Big Data Application Capability and Firm Green Technology Innovation. *Sustainability* **2023**, *15*, 12830. https://doi.org/10.3390/su151712830

Academic Editors: Luigi Aldieri, Jianhua Zhu, Jiaoping Yang, Yanming Sun and Chaoan Lai

Received: 20 June 2023
Revised: 13 August 2023
Accepted: 23 August 2023
Published: 24 August 2023

1. Introduction

With the popularization of emerging technologies, the combination of data elements and traditional elements produces a multiplier effect. The soaring progress of the big data industry provides a technical basis for the intelligent evolution of the manufacturing industry. Digital China Development Report (2022) emphasizes that the scale of the domestic big data industry reached CNY 157 million expanded by 18% year-on-year in 2022, but the development level of regional big data is imbalanced in China. Applying big data and other underlying digital technologies to promote digital transformation is a key challenge for governments and firms. In 2016, the construction of the big data comprehensive pilot zone in China provided an open platform for data sharing; therefore, the pilot will provide a strategic resource in the process of digital transformation. As the applications of regional big data have taken root, firms can grasp the opportunity to develop big data application capability. Big data can integrate various business models and data to extract internal value [1]. Green innovation requires the sustainable development of firms in terms of economy and environment, posing new demands for firm innovation. Thus, flexible use of big data and other digital technologies can improve the efficiency of resource allocation of firms, achieve sustainable competitive advantages [2], and enhance firms' social and environmental performance [3].

The present literature has examined the drivers of green technology innovation (GTI) like environmental regulation [4]. Figure 1 presents the literature about GTI published from the WOS database; we selected 683 articles by searching for "green technology innovation" or "green technological innovation" in the title or abstract. As can be seen from Figure 1, the thematic research on GTI has gradually increased in recent years. However, few studies have investigated the effect of big data on GTI. Our study attempts to fill the research gap by investigating the relationship between regional big data application capability

(RBDAC) and GTI. Even though firms are exposed to the lasting economic shocks of COVID-19, RBDAC can give a much-needed boost to digital economy development and firm digital transformation. On the one hand, according to the information asymmetry theory, RBDAC enables firms to obtain information data more easily and promptly. Then, firms can consider and adapt to big data in innovation decision-making, improve their sustainable competitive advantage [5], and thus promote the GTI [6]. On the other hand, according to the resource-based view and organizational learning theory, RBDAC integrates firm information resources. As a type of innovative resource, big data can improve firm innovation [7]. However, "big data is not always better data", meaning that mining better data from imbalanced data types is impractical for firms [8]. Big data are vulnerable to network attacks that can cause the information to vanish or be altered [9]. The velocity and variety of big data can improve innovation performance, whereas data volume has no significant impact as massive data may impede sound decision-making [10]. Once the usage of big data exceeds a particular threshold, more big data may have a negative impact on sustainable innovation performance [2]. Based on the above analysis, the effect of RBDAC on GTI requires further investigation at a micro-level.

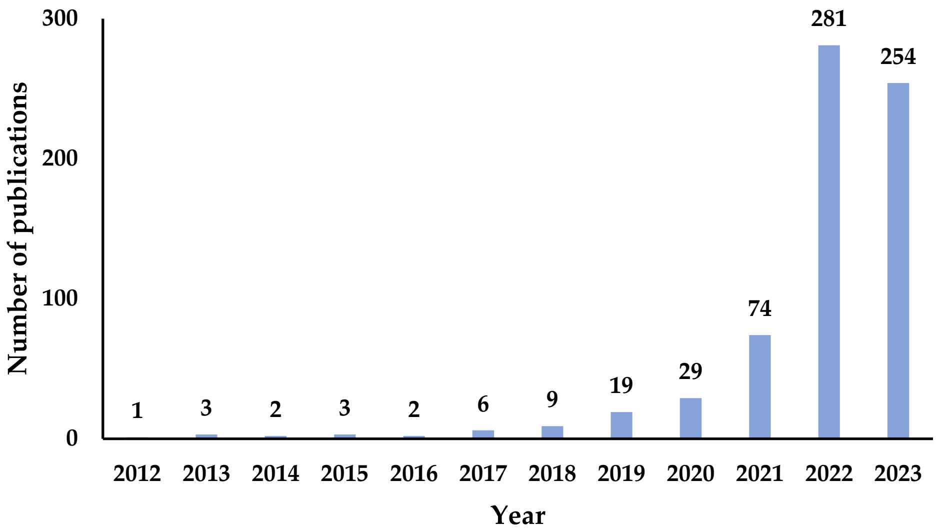

Figure 1. Publications statistics on GTI.

The establishment of the national big data comprehensive pilot zone in 2016 is adopted as a quasi-natural experiment in this study. With the data of Shanghai and Shenzhen A-share listed firms in the Chinese manufacturing industry from 2010 to 2020, this empirical study mainly answers the following questions: (1) Can RBDAC promote the levels of GTI in the manufacturing firms? (2) Can government subsidy and analyst coverage moderate the association between RBDAC and GTI? (3) Is there heterogeneity of different finance constraints, tax administration strengths, regions, property rights, and top management team in terms of the promotion effect of RBDAC on GTI?

The potential contributions to this article will be introduced as follows. Firstly, previous literature about the effect of RBDAC on GTI has drawn inconsistent conclusions. Thus, we collect the specific practices in the Chinese manufacturing firms as examples to fill the literature gap of the relationship between RBDAC and GTI. Secondly, this paper analyzes the mechanism of RBDAC in stimulating the GTI from several external factors' perspectives, and enriching the research situation of the drivers of GTI. Thus, we find that government subsidy and analyst coverage can strengthen the positive relationship between RBDAC and GTI. Furthermore, the GTI effect of RBDAC mainly exists in regions with low levels of financial constraints and tax administration strengths, as well as firms located in the eastern region. Finally, this study expands ideas about digital governance for firms and proposes management enlightenment to manufacturing firms. The vital stimulating elements in the effect of RBDAC on GTI not only depend on the external environment, but also have a stronger relationship in state-owned firms and firms with a high heterogeneity of the top management team in terms of age, gender, and professional background. In conclusion, the empirical analysis in this article confirms that RBDAC can promote the GTI. Our study provides new insights into the digital governance and green practices of firms.

The remaining essay is structured as follows: Section 2 is a literature review; Section 3 is a presentation of the theoretical foundations and research hypotheses; Section 4 introduces the econometric model, variables, and data; Section 5 reports the empirical results and the moderating effects of government subsidy and analyst coverage; Section 6 further explores the heterogeneous of the GTI effect of RBDAC from the perspectives of the external institutional environment (financing constraints, tax administration strengths, and regions) and internal conditions (property rights and top management team); and Section 7 offers major conclusions, implications, and limitations to this paper.

Table 1 lists the abbreviations used in our study.

Table 1. The list of abbreviations.

Abbreviations	Full Form
RBDAC	Regional big data application capability
GTI	Green technology innovation
SEM	Structural equation model
DID	Difference-in-differences
PLS	Partial least squares
OLS	Ordinary least squares
FE	Fixed effects
R&D	Research and development
GMM	Generalized method of moment
ESG	Environmental, Social, and Governance

2. Literature Review

2.1. Studies on Big Data

As a core engine of the data era, big data have the characteristics of large volume, wide variety, and high velocity, helping firms to reveal potential economic value [11]. Thus, scholars are more interested in whether big data can improve organizational performance and how firms can efficiently use the information value of big data. Existing studies related to big data mainly focus on the definition, dimensions, and measurement of big data analysis and big data capability [1]. Scholars also are concerned about the economic consequences of big data, such as the impact of big data on firm innovation [12]. The general methods for measuring big data mainly include surveys [13]. Some scholars also quantify the development level of big data from multiple dimensions [14]. In addition, scholars' empirical measurements of big data are mainly based on the structural equation model (SEM) and the difference-in-differences (DID) method for comparing the economic consequences before and after the application of big data [15]. With technological progress, a few scholars use updated research methods, such as using machine learning for data mining, which has expanded the application context of big data [16]. Aljumah et al. argue that organizations can adapt to environmental changes faster and better than competitors in the same industry based on the dynamic capability view [17]. The successful keys of firms do not simply depend on the integration of existing resources and capabilities, but rather on the ability to allocate resources and capabilities continuously and effectively in dynamic environments. Hence, the dynamic capability theory outlines different specific guidelines for firms to respond to different environmental changes [18]. Digital technology tools, such as big data and machine learning, have been widely applied [19]. The promotion of digital technology in the fields of economy, education, and healthcare has a profound impact on economic life [20]. Digital technology can improve the relationship between stakeholders. According to 214 questionnaires from Chinese firm executives, Lin and Lin claim that big data analysis is chosen as the most popular digital technology because digital technology can enable firms to develop in the direction of customization and enhance good customer relationship performance [21]. Based on a survey of 341 employees from manufacturing enterprises in Saudi Arabia, Jaouadi concludes that big data analysis technology and staff capability should be considered simultaneously in the decision-making process to improve the sustainability of supply chain performance [22]. Lozada et al. confirm that big data analytic capabilities assist stakeholders in working on collaborative innovation from the tangible, intangible, and human resource aspects [23]. Grounded in knowledge-based theory, Tunc-Abubakar et al. regard big data as a tool and investigate 97 employees working with big data in Turkey's firms, finding that the usage of big data drives product and process innovation, and high-quality big data features can boost the innovative output of firms [24].

To sum up (see Table 2), as a highly efficient information exchange channel, big data can alleviate the negative impacts brought by information asymmetry. Big data make it easier for firms to identify the target stakeholders, so firms might become more attractive in the investment market. Big data may also help firms solve the "last mile" issue in the supply chain by digitizing and upgrading logistical networks. The usage of big data has brought intended benefits to firms, including optimizing stakeholder relationships, expanding innovation capabilities, reducing innovation costs [25], and improving resource efficiency [7].

Table 2. Summary of selected literature on big data and performance outcome.

Performance Outcome	Literature	Data	Observations	Method
Sustainable competitive advantage	[5]	Firm-level	117	Survey; PLS-SEM
	[26]	Firm-level	229	Survey; Regression
Firm innovation	[1]	Firm-level	2706	Probit models
	[12]	Firm-level	179	Survey; PLS
Supply chain innovation	[22]	Firm-level	341	Survey; SEM
Co-innovation	[23]	Firm-level	112	Survey; PLS-SEM
Product and process innovation	[24]	Firm-level	97	Survey; PLS-SEM
Organizational performance	[7]	Firm-level	140	Survey; SEM
	[27]	Firm-level	297	Survey; PLS-SEM
	[17]	Firm-level	295	Survey; PLS-SEM
Innovation performance	[10]	Firm-level	239	Survey; PLS-SEM
Environmental performance	[28]	Firm-level	201	Survey; two-stage hybrid factorial analysis-SEM
Social and environmental performance	[3]	Firm-level	205	Survey; PLS-SEM
Sustainable innovation performance	[2]	Firm-level	1109	Quadratic regression model
Green total factor productivity	[14]	Province-level	240	Panel data model
	[15]	City-level	3640	Difference-in-differences (DID)

Source: Authors. PLS refers to partial least squares. SEM refers to structural equation model.

2.2. Studies on GTI

The rapid expansion of industrialization and modernization has resulted in increasingly widespread environmental challenges. The efforts of governments and businesses to establish a sustainable ecological community have attracted academic interest. GTI is considered as a vital factor in implementing a sustainable development strategy, which is good for enhancing environmental performance and improving the competitive position of firms [29]. Currently, research on the drivers of GTI is mainly concentrated on financial policy, environmental legislation, and tax policy at the macro level, as well as on leadership personality, internal control, and corporate governance at the micro level. The quantitative approach to evaluating the GTI efficiency is primarily the DEA method [30]. Some scholars also use the number and quality of patents for green innovations as proxy variables [31].

Castellacci and Lie divide the GTI into four main categories and empirically investigate that practical characteristics of green innovations may differ, as should associated policies, and thus financial institutions may invest in diversified portfolios of GTI at the same time [32]. In addition, there is an increasing amount of literature analyzing the drivers of GTI. The "Porter Hypothesis" suggests that appropriate environmental regulation can stimulate innovation activities, offset the expenditure by stricter regulations, and strengthen the competitive advantages of firms [33]. Thus, environmental regulation and environmental responsibility significantly improve the GTI, increase sustainable competitive advantage, and reduce the regulatory cost [4]. Fan et al. support the validity of digital strategy and find that the development of digitalization may directly or indirectly affect the GTI by alleviating financing constraints and enhancing corporate social responsibility [34]. Rui and Lu conduct a questionnaire of 255 executives from Chinese manufacturing firms to identify the firms' reactions to the triple pressure from stakeholders (including governments, consumers, and competitors) [35]. Stricter environmental regulations, increased green consumption, and more intense competition all drive firms to choose environmentally friendly behaviors voluntarily or involuntarily, thereby bolstering the GTI.

Currently, there is insufficient empirical research on the effects of big data and GTI (see Table 3). Firms that specialize in leveraging big data are more likely to have four competitive advantages: difficult-to-imitate data resources, technical infrastructure for optimizing management efficiency, the ability to conduct real-time market analysis, and the capability to adapt strategy based on the economic environment. Considering the above four advantages play a key part in generating significant value for firms in a complicated environment, Zhang et al. conclude that firms with better big data capabilities are more willing to implement innovation to achieve sustainable competitive advantages [26]. During the period of intense competition and business transformation, firms tend to employ digital technology more widely, such as artificial intelligence, big data, and cloud computing. Digital technology can upgrade information processing capabilities, reduce search costs of external information, optimize the supply chain, and re-establish the system of innovative collaboration. In addition, digital technology has achieved outstanding results in the GTI by advancing the development of innovation networks and processes [25].

Table 3. Summary of selected literature on the influence factors of firm's GTI.

Category	Influencing Factors	Literature	Observations	Method
		Macro-level		
Financial policy	Green credit	[36]	12,048	DID
	Green finance	[37]	18,775	DID
	Green bonds	[38]	7835	FE
Environmental legislation	Environmental regulation	[4]	3121	Mediating effects model
		[39]	308	Survey; Structural equation model (SEM)
Tax policy	Tax incentives	[40]	9744	FE
Digitalization	Digital transformation	[41]	15,029	Panel data models
	Digitalization	[34]	3547	FE
		[42]	13,140	FE
		Micro-level		
Leadership personality	Female CEOs	[43]	9997	FE
	Board gender diversity	[31]	15,615	FE
Internal control	Internal control	[44]	23,215	Regression
Corporate governance	Corporate governance	[45]	31,659	Ordinary least squares (OLS)
Stakeholder	Stakeholder pressure	[35]	255	Hierarchical regression
	Institutional investors	[46]	5473	FE
	Media attention	[47]	9637	FE

Source: Authors. DID refers to difference-in-differences. FE refers to fixed effects models.

The cited literature above mainly focuses on digital technology-driven GTI, providing theoretical guidance for the research direction. Even though GTI is widely recognized as a breakthrough advancement in China, the inevitable debate over whether firms will gain environmentally and economically from big data-driven innovation needs more academic attention. Therefore, this article constructs a paradigm of "RBDAC-GTI" to investigate the direct impact of regional big data application on GTI, the moderating roles of government subsidy and analyst coverage on the relationship of RBDAC and GTI, and the heterogeneity at the level of external institutional environment and internal firm level.

3. Theoretical Analysis and Research Hypotheses

3.1. The Impact of RBDAC and GTI

The arrival of big data conforms to the development trend of the digital era. The development of big data meets the technical requirements of the digital economy and serves as an effective means to improve firm performance. RBDAC and the practice of big data are critical for firms. As a pillar industry of the national economy in China, the green transformation of the manufacturing industry is urgent and may be undertaken with the assistance of GTI aims to achieve the targets of carbon peak and neutrality. Big data are a fundamental strategic resource for the country and bring remarkable and immediate influences on the economy. The popularization of the internet is a booster for the development of big data. Meanwhile, effective policies provided by the national and local governments provide strong support for the diffusion of big data technology. Nowadays, there is an increasing investment enthusiasm from diverse firms in the integration of big data with other new technologies, such as artificial intelligence, 5G, and Blockchain. The development of big data is progressing year by year. In 2014, a topic related to big data strategy was officially introduced by the Chinese government work report for the first time. In 2015, the release of the Action Plan for Promoting Big Data Development by the State Council of China aimed to expand big data technology to developing industries, emerging industries, and national mass entrepreneurship. In February 2016, as the first batch of launching big data pilot zones in China, Guizhou led the development of big data by establishing the first Data Exchange, which provides a platform for data trading and the development of the digital economy. In October 2016, the Chinese government launched its second batch of big data comprehensive pilot zones, including Beijing-Tianjin-Hebei Urban Agglomeration, Pearl River Delta, Shanghai, Henan, Chongqing, Shenyang, and Inner Mongolia. Henceforth, the eight national big data comprehensive pilot zones have led the development of the domestic big data industry, achieving the collaborative development of the East, West, South, and North regions. The regional development of big data plays a role in promoting the development of the national digital economy from a macroeconomic perspective, and in strengthening innovation and performance for the microeconomic entities through data governance and trading. However, studies on explaining whether big data can drive industrial innovations are scarce mainly because the development of big data is still in its early stages.

Despite several studies on how big data affect organizational performance, supply chain management, and technological innovation, research on the implications of big data development on green innovation is still sparse. Big data analysis capabilities may realize higher environmental performance alongside technological progress. Belhadi et al. confirm that both the Lean Six Sigma strategy and green manufacturing can mediate the effect of big data analysis capabilities on environmental performance [28]. According to dynamic capability theory and contingency theory, big data and predictive analytics have a multidimensional and resilient structure that consists of technical skills, managerial skills, organizational learning, and data-driven decision-making. Dubey et al. find that the capability of big data and predictive analytics can improve both the social and environmental performance of Indian manufacturing firms and enhance supply chain sustainability [3]. Big data have become a powerful driver of green economic development, as big data can promote firm innovation and productivity, drive technological progress for industrial development, and improve macro resource allocation. Thus, Wang et al. contend that firms can use big data to improve green technological progress and application efficiency in order to gain green competitive advantages [14]. Wang et al. believe that digitalization can realize the sustainable GTI by easing the information asymmetry, reducing financing constraints on innovation, raising the tolerance of risk-taking, and optimizing resource allocation [42]. El-Kassar and Singh find that the application of big data has a direct impact on the innovation of green processes and green products; therefore, big data bring more opportunities to green innovation practices [6]. Therefore, in order to achieve long-term survival in the industry, manufacturing firms should make full use of the big data policy and technique.

To begin with, data comprise the core competency of firms according to resource-based concepts and information asymmetry. RBDAC raises the degree of local big data development and removes barriers to regional information exchange. Therefore, firms find it easier to access and process information, improve information transparency, reduce the cost of time lag, and acquire information. Furthermore, by incorporating big data into the production, operation, and decision-making processes, firms can classify and filter information, improve information transparency, and meet the differentiated and customization needs of consumers. Firms update existing products or offer new items in response to consumer feedback in order to satisfy market demand on time. As a result, data information exchange can help firms develop a sustainable competitive advantage. Second, the sustainable advantages of firms are inseparable from continual learning and advancement considering organizational learning and dynamic capabilities. Organizations with learning skills will maintain competitiveness as new technologies emerge and they seek new solutions in a continuously changing consumer market. RBDAC has promoted the development of local big data industries, which has an unexpected effect that decision makers who benefit from gathering huge resources are more willing to help big data grow. The upgrade of digital technology will strengthen the ability of knowledge management and decision-making. The high velocity and liquidity characteristics of big data enable firms to quickly adjust their green research and development (R&D) strategies. Thus, firms can fully utilize technology to integrate resources, increase productivity, reduce resource waste and R&D risks, thereby enhancing the GTI.

In summary, the development of RBDAC promotes firms to integrate big data into innovation decision-making, accelerates the matching process between the GTI output and the market, reduces green R&D risks, and promotes the GTI. Hence, this article proposes the following assumption:

Hypothesis 1 (H1). *RBDAC has a positive impact on the GTI of manufacturing enterprises.*

3.2. Moderating Role of Government Subsidy

There is a high capital requirement for the GTI investment, and the process of producing innovation outputs is time-consuming, requiring firms to analyze the advantages and costs of innovation regularly. The ultimate purpose of firms is to maximize shareholder profit; thus, if there are no external supports and their equity capital or capital market financing is the only source of funding for the high-risk GTI, firms will become risk-averse. Therefore, government subsidy, as one of the key fiscal means, has drawn researchers' attention as a result of the need for government departments to support firms to convert towards green development [48]. During economic transformation, government subsidy can directly fund firm R&D to lower R&D costs and improve the GTI capacities of firms [49]. Adequate R&D funds can ease the immediate financial strain and encourage firms to initiate independent R&D. Therefore, government subsidy can promote R&D investment through resource effects. Furthermore, firms can use government subsidy as a positive signal for the investment market in order to mitigate the negative effects of information asymmetry [50], as well as stimulate the inflow of finances and human capital [51]. As a result, this not only enhances the financing ability of firms but also compensates for the lack of external funds of GTI [52].

To begin with, from the perspective of external government subsidy, the growth of RBDAC helps the government in supervising firms through data, while firms with better performance in GTI are more likely to apply for financial subsidies. Under the signal effect of government subsidy, the external financing level of firms increases, and the probability of achieving the expected subsidy effect rises. Thus, GTI will be improved, forming a strong subsidy incentive innovation impact. Ultimately, a virtuous cycle of innovation financing has been formed by government subsidy. Secondly, from the perspective of internal resources within firms, government subsidies act as funds invested in production and R&D and have a positive effect on R&D [53]. Human resources are one crucial link in R&D. Government subsidies are a reliable guarantee for highly skilled talents because the government must strictly examine the qualification of receiving subsidies. External financial capital infusions, along with internal human resource infusions, create a perfect environment for big data applications and accelerate the development of RBDAC [54]. Finally, these initiatives supply firms with funds and highly skilled talents for GTI, hence increasing the level of firm GTI. In conclusion, government subsidy can increase the effects of big data capability development on firm GTI. Thus, another assumption has emerged:

Hypothesis 2 (H2). *Government subsidy positively moderates the relationship between RBDAC and GTI, such that the effect will be stronger when government subsidy is high and vice versa.*

3.3. Moderating Role of Analyst Coverage

The 13th Five-Year Plan of China highlights that "innovation is the primary driver of development". In order to achieve the period targets, the Chinese government attaches great importance to industrial development and firm innovation, and has successively issued a series of policies to encourage more practitioners to carry out innovation. However, industrial policies may not be foolproof and may also have the possibility of opportunistic behavior. The appearance of analyst coverage can be seen as a rational management mechanism outside the organization, which can effectively reduce information asymmetry, managers' self-interest, and agent conflicts [55]. Furthermore, increased analyst coverage boosts R&D intensity and improves firm innovation [56]. As a mediator between corporations and investors, the media may be a double-edged sword for firms. While firms profit from a strong social image brought about by positive reporting, they must also bear the pressure of public opinion on environmental protection issues, which eventually leads to the expansion of GTI [47]. In a digital era where retail investors have easy access to firm information, firms that receive more attention from financial analysts and the media are more likely to be liked by investors. Following that, investor support highlights the impact of firm innovation [57]. Furthermore, big data provide a technical framework for consumers to collaborate with firms to achieve mutual goals, such as how consumer participation in big data analysis improves product innovation [58]. The regions become more attractive to investors when they increase the development level of RBDAC, but also lead to providing investors, customers, and analysts with a convenient platform for obtaining, processing, and integrating firm data. Firms can use the big data platform to promote information that is beneficial to themselves and guide public opinion. Investors can acquire firm information through internet search engines, community forums, python machine language, and mobile phone applications. The platform uses "data flow" to guide investor sentiment to change. The value of big data to firms should not be underestimated.

To begin with, according to the theory of information asymmetry, RBDAC accelerates information circulation, hence increasing the information collection and analysis capacities of stakeholders. In addition, big data development helps financial analysts in developing a better grasp of firm information, eliminating financial constraints imposed by information asymmetry, and resulting in higher innovation efficiency and green innovation performance [59,60]. According to the reputation theory, RBDAC improves the transparency of firm information, making it easier for investors to acquire both good and bad information about the firms. Under supervisory pressure from analyst coverage, managers value environmental challenges and reputation implications, reduce opportunistic behaviors, and focus on long-term investment in green innovation [47]. A higher analyst coverage leads to better market monitoring and a greater willingness of firms to innovate and change. Furthermore, analysts are more likely to use big data in regions with more mature big data systems to identify the effectiveness of firm innovation, enhance the efficiency of data flow among stakeholders, increase the efficiency of monitoring and reputation, and promote the GTI. In conclusion, more attention from financial analysts leads to a higher commitment to promoting the firms' GTI through the application of big data. Thus, we propose another following assumption:

Hypothesis 3 (H3). *Analyst coverage positively moderates the relationship between RBDAC and GTI, such that the effect will be stronger when analyst coverage is high and vice versa.*

The theoretical model (see Figure 2) is developed to clarify the effect of RBDAC on GTI, and the moderating effect of government subsidy and analyst coverage.

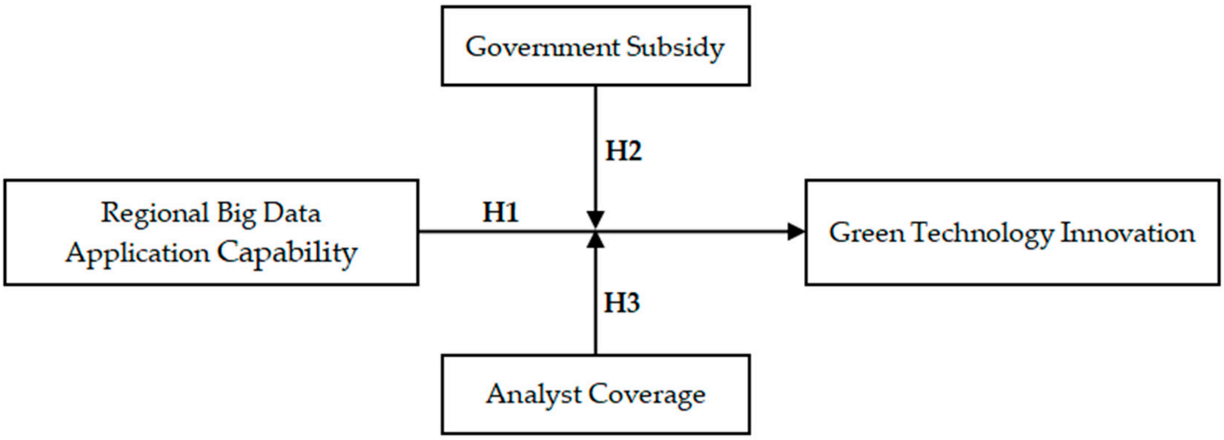

Figure 2. Theoretical model.

4. Methodology and Data

4.1. Model Specification

The purpose of this report is to test whether RBDAC can encourage more firms in the manufacturing industry to engage in GTI, so a difference-in-differences (DID) model is commonly used to evaluate the validity of the policy effect [61]. The emergence of the big data comprehensive pilot zone in 2016 is for research conducted by the government, so the pilot policy can be treated as a quasi-natural experiment. Therefore, the DID model adapted can avoid the causality of endogeneity and reverse. We take the cities or provinces with the big data pilot zone as experimental groups and other provinces without a zone as control groups. The year of the pilot zone policy implementation, 2016, is the starting point of the statistics collection, and the benchmark regression model is constructed as follows:

$$GIPA_{i,t+1} = \beta_0 + \beta_1 DID_{i,t} + \beta_2 TREAT_{i,t} + \beta_3 TIMEe_{i,t} + \beta_4 CV_{i,t} + \alpha_i + \varepsilon_{i,t} \quad (1)$$

where subscripts "i" and "t" denote city and year, respectively. *GIPA* represents the level of GTI. Because there is no consensus-based standard for evaluation of "level", the number of green invention patent applications is selected as a variable to lend concreteness to the developmental level. Additionally, there is a time lag of the innovation outputs, so the number of innovation patent applications lagging one period is used as the dependent variable. *DID* represents the interaction term between the policy variable (*TREAT*) and the time variable (*TIME*), i.e., *DID* equals to *TREAT* $\times$ *TIME*; TREAT is to indicate whether the firm is in the big data pilot zone, and it is a dummy variable that equals "1" if the firm is in the treatment zone, and "0" otherwise; *TIME* is another dummy variable that equals "1" if the year selected is after 2016 and onward, and "0" otherwise. Control variables (*CVs*) are a control variable matrix of firms, including firm size (*SIZE*), asset–liability ratio (*LEV*), total asset growth rate (*GROWTH*), the ratio of independent boards (*OUT*), share ratio of the managers (*MSR*), whether to pay cash dividends (*DIV*) and the duality of chairman and general manager (*DUAL*). In the benchmark analysis, the coefficient of the interaction term (β_1) needs to be emphasized in this article. Because the DID model can eliminate the deviation caused by differences between the experimental group and the control group or other deviations caused by time factors, the DID model obtains real feedback on the effectiveness of policies. If it is positive, it represents that the big data pilot zone promotes the GTI. Moreover, representing fixed individual effects is an error term that clusters the errors in the model at the provincial level.

4.2. Variable Selection and Interpretation

4.2.1. Dependent Variable: Green Technology Innovation

Guided by Wang et al. [47], this report uses the number of green invention patent applications as a variable to measure the level of green innovation development. But as there is a time lag of innovation output, the number chosen is the number of green invention patent applications lagging one year (*GIPA*).

4.2.2. Explanatory Variables: Variables Related to Big Data Pilot Zone

Following the previous literature, we construct a DID model in the province-year dimensions to identify the microeconomic effects of policies for the pilot zone. The concrete content is divided into three sections:

(1) Whether the area is in the big data comprehensive pilot zone (*TREAT*), Guizhou, Beijing, Tianjin, Hebei, Guangdong, Shanghai, Henan, Chongqing, Liaoning, and Inner Mongolia are selected for the big data pilot zone, according to the State Council's Action Plan for Promoting the Development of Big Data and other documents issued by pilot cities;
(2) Whether the time point is after the implementation of the big data pilot zone (*TIME*), 2016 is identified as the first year when the policy is operational;
(3) The interaction term between TREAT and TIME (*DID*) is used to assess the micro-effects of regional big data pilot policy.

4.2.3. Moderator Variables

In this article, there are two moderator variables:

(1) Government assistance. Following the practice of Xue et al. [41], the natural logarithm of the government subsidy in that year as recorded in the financial statement is used as a variable to assess the government subsidy (*SUBSIDY*);
(2) Coverage by analysts. According to the study of Liu and Xu [62], the number of financial analysts is used as a variable to assess the intensity of analyst coverage (*ANALYST*).

4.2.4. Control Variables

Following the previous literature, the control variables are selected at the firm level as follows:

(1) Firm size (*SIZE*), expressed as the natural logarithm of the firm total asset;
(2) Asset–liability ratio (*LEV*), expressed as the ratio of liabilities to assets;
(3) Total asset growth rate (*GROWTH*), expressed as the ratio of the difference between the total assets at the end of the year and the total assets at the beginning of the year to the total assets at the beginning of the year;

(4) The ratio of independent boards (*OUT*), expressed as the percentage of the number of independent directors on the board of directors;
(5) The share ratio of the managers (*MSR*), expressed as dividing the number of shares held by executives by the number of total shares;
(6) Whether to pay cash dividends (*DIV*), expressed as a dummy variable that equals "1" if the firm pays cash dividends in that year, and "0" otherwise;
(7) The duality of chairman and general manager (*DUAL*), expressed as a dummy variable that equals "1" if the chairman and general manager are the same person, and "0" otherwise.

In addition, this article simultaneously controls the annual (*Year*) and individual (*Firm*) fixed effects, respectively.

The specific definitions of the above variables are detailed in Table 4.

Table 4. Variable definitions.

Types	Variables	Symbols	Descriptions	Literature Source
Dependent Variables	Green technology innovation	GIPA	The number of green innovation patents lagging one year	[47]
Explanatory variables	Whether it is the area for a big data pilot zone	TREAT	Dummy variables of policy. "1" for if the firm is in Guizhou, Beijing, Tianjin, Hebei, Guangdong, Shanghai, Henan, Chongqing, Liaoning, or Inner Mongolia, otherwise"0"	[15]
	Whether time is after 2016	TIME	Dummy variable of time, "1" for 2016 and onward, otherwise "0"	[61]
	The interaction term of DID	DID	The interaction term of TREAT × TIME	[36]
Moderator variables	Government subsidy	SUBSIDY	The logarithm of government subsidy in that year noted in the financial statement	[41]
	The coverage of financial analysts	ANALYST	Number of financial analysts	[62]
Control variables	Firm size	SIZE	The logarithm of the total assets	[63]
	Asset–liability ratio	LEV	Total liability/Total assets	[39]
	Total asset growth rate	GROWTH	(Total assets at the end of the year—Total assets at the beginning of the year)/Total assets at the beginning of the year	[64]
	The ratio of independent boards	OUT	Number of independent boards/Number of the board of directors	[34]
	The share ratio of the managers	MSR	Number of shares held by executives/Number of total shares	[38]
	Whether to pay cash dividends	DIV	"1" for if the firm pays cash dividends in that year, otherwise "0"	[65]
	The duality of chairman and general manager	DUAL	"1" for if the chairman and general manager are the same person, otherwise "0"	[57]

4.3. Data Collection and Descriptive Statistics

Because the GTI development is primarily focused on the manufacturing sector, data from Shanghai and Shenzhen A-share listed manufacturing firms in China is a desirable and appropriate data source for the research in this article. Furthermore, as the national big data comprehensive pilot zone started in 2016, the sample period ranges from 2010 to 2020. The sources for the number of green patents originate from matching WIPO's list of international patent classifications with data from the State Intellectual Property Office. Then, other data generally comes from the CSMAR and WIND databases in China. Furthermore, the data are filtered by excluding listed firms with primary businesses in finance and insurance, anomalous deviations in stock returns throughout the sample period (ST, PT, or delisted from the stock market), and missing core statistics. All continuous variables are winsorized at 1% and 99% to reduce the influence of extreme values. Finally, we have 15,993 samples from 2399 manufacturing firms.

Table 5 presents the descriptive statistical results of the main variables. All variables lie in a reasonable range. The explanatory variable *GIPA* has a mean of 1.105, a maximum of 450, and a minimum of 0. These results show that the disparity in the number of green innovation patent applications among sample firms is relatively substantial and that the GTI development level of selected samples is below average.

Table 5. Descriptive statistics.

Variables	Sample	Mean	SD	Minimum	Median	Maximum
GIPA	15,993	1.105	8.237	0	0	450
DID	15,993	0.221	0.415	0	0	1
TREAT	15,993	0.389	0.487	0	0	1
TIME	15,993	0.563	0.496	0	1	1
SUBSIDY	15,993	16.284	2.466	0	16.485	20.758
ANALYST	15,993	1.451	1.190	0	1.386	3.932
SIZE	15,993	22.085	1.152	19.630	21.941	25.750
LEV	15,993	0.413	0.193	0.033	0.406	0.934
GROWTH	15,993	0.169	0.377	−0.371	0.090	5.116
OUT	15,993	0.375	0.054	0.286	0.333	0.571
MSR	15,993	0.094	0.156	0	0.001	0.690
DIV	15,993	0.739	0.439	0	1	1
DUAL	15,993	0.286	0.452	0	0	1

5. Empirical Results and Analysis

5.1. Baseline Results

According to the baseline regression assumed above, this section of the article aims to examine the impacts of RBDAC on the GTI of listed firms in a regression model with an ordinary least square (OLS) or fixed effect (FE). In columns (2) and (4) of Table 6, we control the fixed effects of Firm and Year, and empirical results are outlined in Table 6. According to Table 6, the coefficient of *DID* is positive and statistically significant at the 5% level no matter whether the control variables at the firm level are considered. After incorporating control variables at the firm level into the regression, the coefficient of *DID* without or with fixed effects of firm and year is 0.416 and 0.777, respectively, which are positive and statistically meaningful at the 5% significance level. Generally, the positive coefficient of *DID* directly indicates that the policy of the big data pilot zone effectively promotes that RBDAC is effective in promoting the patent application of GTI. In a deeper sense, the policies of RBDAC advance the development of GTI. Therefore, the first hypothesis (H1) is accepted. The results indicate that RBDAC has improved the digitalization level of firms, enabling firms to use data technology to reorganize consumer-tailored demands for green products and green process design, improve resource utilization efficiency and product competitiveness, and thus promote the development of firm GTI capabilities. The conclusion of the baseline regression illustrated above is consistent with reference to previous studies. Wang et al. find that implementing big data has significantly improved the green total factor productivity among different cities by about 0.4% [14]. Wei and Zhang advocate the big data pilot zone policy by providing the example that the average growth of the green total factor productivity among cities is about 7.7% [15]. Niebel et al. find that the application of big data analytics has increased the proportion of product innovation by approximately 6.5% [1].

Figure 3 shows the regression coefficients of the relationship between RBDAC and GTI.

Table 6. Impacts of RBDAC on GTI in manufacturing.

	(1)	(2)	(3)	(4)
Variables	**OLS-no CVs**	**FE-no CVs**	**OLS-CVs**	**FE-CVs**
DID	0.418 **	0.764 **	0.416 **	0.777 **
	(0.199)	(0.324)	(0.168)	(0.330)
TREAT	0.895 ***		0.653 **	
	(0.254)		(0.257)	
TIME	0.336 ***	0.753 **	−0.220	0.590 *
	(0.086)	(0.345)	(0.238)	(0.327)
SIZE			1.495 **	0.210 *
			(0.573)	(0.106)
LEV			−0.039	−0.600
			(0.411)	(1.134)
GROWTH			−0.556 **	−0.099
			(0.212)	(0.096)
OUT			−0.024	−3.715
			(1.479)	(2.624)
MSR			1.187 *	−0.462
			(0.618)	(0.654)
DIV			0.104	0.217
			(0.170)	(0.218)
DUAL			0.743	0.179
			(0.797)	(0.375)
Constant	0.476 ***	0.359	−32.413 **	−2.654
	(0.062)	(0.234)	(12.426)	(3.134)
Observations	15,993	15,993	15,993	15,993
Firm fixed effects	No	Yes	No	Yes
Year fixed effects	No	Yes	No	Yes
Adjust R^2	0.00538	0.0106	0.0445	0.0111

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

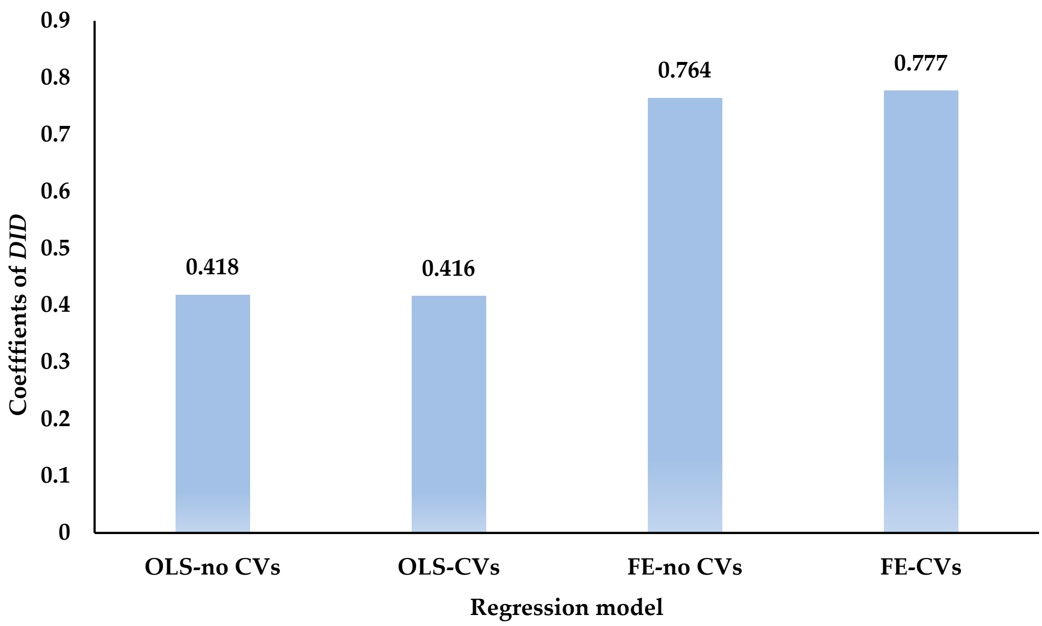

Figure 3. Regression coefficients of RBDAC on GTI.

5.2. Robustness Test

This article will analyze the conclusion by conducting robust testing in six aspects to ensure the reliability of the research conclusions. Recognized methods include parallel trend assumptions, PSM-DID, placebo test, removing interference from other policies, substituting the dependent variables, replacing the moderator factors, and dynamic panel estimates.

5.2.1. Parallel Trend Assumptions

There is a significant important requirement when using the DID model, which is that the experimental group and the control group should meet a generally consistent parallel trend before policy implementation. Otherwise, DID model is unable to evaluate the effectiveness of the policy. Following Bertrand and Mullainathan [66], the second regression will be constructed as follows:

$$\begin{aligned} GIPA_{i,t+1} = \beta_0 + \beta_1 Before2_{i,t} + \beta_2 Before1_{i,t} + \beta_3 Current_{i,t} + \beta_4 After1_{i,t} + \beta_5 After2_{i,t} + \beta_6 After3\_{}_{i,t} \\ + \beta_7 TIME_{i,t} + \beta_8 TREAT_{i,t} + \beta_9 CV_{i,t} + \alpha_i + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Concerning the equation outlined above, we define *Before*2, *Before*1, *Current*, *After*1, *After*2, and *After*3_ as dummy variables. If the values are data recorded in the second and first years before the policy of the big data pilot zone, the values of *Before*2 and *Before*1 are "1", otherwise "0". If it is data from the policy year, the value of *Current* is "1", otherwise "0". Moreover, if the data are recorded after the first, second, and third years or more of the policy implementation, *After*1, *After*2, and *After*3_ are "1", otherwise "0". Figure 4 shows the result of the parallel trend assumptions under fixed effects, and curve of coefficients demonstrates that there is consistent development trend of the experimental group and the control group before 2016. Thus, the conclusion of this article passed the parallel trend test.

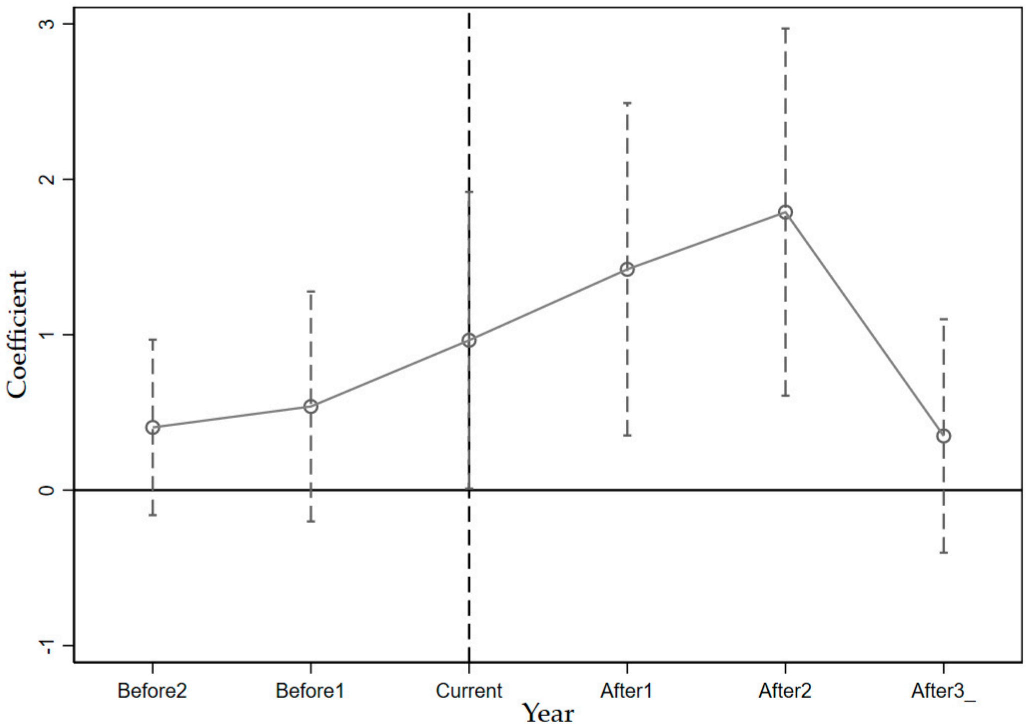

Figure 4. Parallel trend test.

Columns (1) and (2) in Table 7 show the results of OLS regression and FE regression by controlling *Firm* and *Year* fixed effects, respectively, and the data for both regressions show a consistent trend. The coefficients of *Before*1 and *Before*2 are not statistically significant at the 10% significance level, whereas the coefficients of *Current*, *After*1, and *After*2 are statistically significant at the 5% level. The coefficient of *Current*, *After*1 and *After*2 grows year by year, indicating that the number of patent applications for green innovation increases steadily following the implementation of the big data pilot zone policy. Nevertheless, the *After*3 coefficient is insignificant. The reason could be that the policy effect is gradually fading. Furthermore, the use of big data pilot zones in other Chinese cities has increased, resulting in present zones having less impact on GTI. As a result, the DID model in this paper already satisfies the constraints of parallel trend assumptions.

Table 7. Robust test results I.

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**PTT_OLS**	**PTT_FE**	**PSM-DID**	**Placebo Test 1**	**Placebo Test 2**	**Other Policies**
Before2	0.071	0.404				
	(0.267)	(0.276)				
Before1	0.415	0.538				
	(0.275)	(0.362)				
Current	0.896 *	0.965 **				
	(0.481)	(0.467)				
After1	1.227 **	1.421 **				
	(0.555)	(0.524)				
After2	1.573 **	1.789 ***				
	(0.644)	(0.578)				
After3_	−0.362	0.349				
	(0.429)	(0.368)				
DID			0.726 *	−0.318	0.958 *	−1.132 **
			(0.370)	(0.465)	(0.474)	(0.434)
TIME	−0.221	0.522 *	−0.580	0.345 ***	0.224	1.366 ***
	(0.156)	(0.266)	(0.597)	(0.114)	(0.192)	(0.371)
SIZE	1.495 ***	0.209 *	0.243	0.061	−0.065	0.180 *
	(0.357)	(0.107)	(0.161)	(0.224)	(0.245)	(0.106)
LEV	−0.054	−0.673	−1.446	−2.236	0.191	−0.817
	(0.377)	(1.147)	(1.754)	(1.694)	(0.410)	(1.152)
GROWTH	−0.599 ***	−0.096	−0.112	0.232	−0.004	−0.078
	(0.179)	(0.096)	(0.169)	(0.350)	(0.100)	(0.099)
OUT	0.016	−3.676	−6.214 *	−3.977	−0.659	−3.475
	(2.341)	(2.605)	(3.438)	(3.529)	(1.507)	(2.641)
MSR	1.203 *	−0.445	−0.956	−3.524	0.789	−0.354
	(0.639)	(0.655)	(0.781)	(3.633)	(0.610)	(0.651)
DIV	0.070	0.194	0.338	0.079	0.090	0.250
	(0.145)	(0.220)	(0.357)	(0.091)	(0.166)	(0.218)
DUAL	0.757 *	0.185	0.352	0.108	0.241	0.196
	(0.453)	(0.384)	(0.601)	(0.224)	(0.301)	(0.375)
Constant	−32.390 ***	−2.600	0.055	1.910	2.160	−2.356
	(7.723)	(3.147)	(4.081)	(6.968)	(5.782)	(3.107)
Observations	15,993	15,993	10,283	5547	7805	15,993
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjust R^2	0.0465	0.0122	0.0118	0.00533	0.0133	0.0119

Note: 10 heavy polluting industries in Chinese manufacturing are selected as follows: (C15) wine, beverage, and refined tea manufacturing; (C17) textiles; (C19) leather, fur, feather and its products, and shoes; (C22) paper and paper products; (C25) petroleum, coal, and other fuel processing; (C26) chemical raw materials and chemical manufacturing; (C27) pharmaceutical manufacturing; (C28) chemical fiber manufacturing; (C29) rubber and plastic products; (C31) ferrous metal smelting and calendaring; (C32) nonferrous metal processing. ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

5.2.2. PSM-DID

In the DID model, there is still the possibility of sample selection bias. To prevent the risk of endogeneity difficulties, this article uses a propensity score matching with the DID (PSM-DID) method. The model also incorporates firm-level control variables such as *SIZE*, *LEV*, *GROWTH*, *OUT*, *MSR*, *DIV*, and *DUAL*. In the regression, there are also Year and Firm fixed effects being controlled. Then, we will conduct a 1:1 matching between firms in cities with big data pilot zones and firms in cities without pilot zones, and outline a regression based on the matching results of paired samples. According to the results in column 3 of Table 7, the *DID* coefficient is statistically significant at the 10% level, which is consistent with the result of baseline regression.

5.2.3. Placebo Test

To make the conclusion more thoughtful and meticulous, this article uses two placebo tests to check whether there are any other factors influencing the impacts of RBDAC on GTI. (1) We advance the time of the big data pilot project and reconstruct the time dummy variables. If the coefficients after adjustments are still significant, the result suggests that there are other factors influencing the GTI development. Otherwise, the result demonstrates that the boosting effect on green innovation is related to the effects of the big data pilot project rather than another random cause. The results of the counterfactual test, as shown in column 4 of Table 7, indicate that the coefficient of *DID* is not significant if the pilot project is advanced to 2012 and the period of the sample is controlled from 2010 to 2014. The negative and statistically insignificant *DID* coefficient suggests that the positive changes in GTI are the result of big data pilot initiatives rather than other variables. (2) We carry the start date of the pilot policy to 2016, while shortening the sample period from 2014 to 2018. The results are summarized in column 5 of Table 7. The fact that the coefficient of *DID* is still positive and significant at the 10% level supports the argument that big data development is still playing a significant role in fostering the GTI in the three years before and following the implementation of the policy.

5.2.4. Excluding the Interference from Other Policies

Although the studies mentioned above confirm the accuracy of the DID model, some existing regulations may work against the benefits of big data pilot efforts. The 2016 changes to the resource tax and pollutant discharge fee will also directly influence the GTI for firms. Based on the "Guidelines for Environmental Information Disclosure of Listed Firms (2010)" and the 2012 China Securities Regulatory Commission industry classification, this research selects 10 Chinese manufacturing industries with high pollution levels; see the notes in Table 7 for details. Manufacturing industries with high levels of pollution are selected as the experimental group since they are more vulnerable to the effects of tax changes, while others are designated as the control group. It is also critical to keep the time factor constant; therefore, the time set for the renewal experiment is in 2016. Column 6 in Table 7 displays the regression findings. The *DID* coefficient is significantly negative, indicating that the 2016 tax reform is ineffective in GTI. As a result, the robustness of the preceding results is indirectly demonstrated.

5.2.5. Replacing the Dependent Variable

The dependent variables are replaced as the final check on the reliability of the research conclusions. *GIPA* will be replaced at this stage by the number of green utility model patent applications (*GUPA*), the total number of green patent applications (*GPA*), the number of green invention patents granted (*GIPG*), the number of green utility model patents granted (*GUPG*), and the overall number of green patents granted (*GPG*). Then, each alternative variable will be tested separately.

Table 8 shows that except for insignificant *DID* coefficients when replacing with *GUPA* and *GUPG*, the coefficients of *DID* with other replacing variables are all accepted at the 5% significance level. Importantly, the findings validate the reliability of the conclusion in this article. In comparison to green innovation patents, green utility patents have a lesser difficulty and cost of R&D, which is why RBDAC has a negligible impact on green utility model patents.

Table 8. Robust test results II.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	**GUPA**	**GPA**	**GIPG**	**GUPG**	**GPG**	**GMM**	**GMM-FE**
L.GIPA						0.507 ***	0.553 ***
						(0.06)	(0.07)
DID	0.249	1.026 **	0.550 ***	0.470	1.020 **	1.587 ***	0.395 **
	(0.182)	(0.472)	(0.194)	(0.296)	(0.452)	(0.28)	(0.18)
TREAT						−0.902 ***	−0.155 *
						(0.14)	(0.09)
TIME	0.467 ***	1.056 **	0.354 ***	0.627 ***	0.981 ***	−0.753 ***	−0.337 ***
	(0.151)	(0.393)	(0.073)	(0.210)	(0.261)	(0.13)	(0.12)
SIZE	0.201 ***	0.411 ***	0.018	0.262 ***	0.280 **	0.357 ***	0.198 ***
	(0.061)	(0.144)	(0.087)	(0.082)	(0.130)	(0.05)	(0.06)
LEV	0.031	−0.569	0.229	−0.344	−0.115	0.04	0.15
	(0.321)	(1.413)	(0.518)	(0.660)	(1.066)	(0.15)	(0.14)
GROWTH	−0.098 **	−0.197 *	−0.090 **	−0.170 **	−0.260 ***	−0.073 *	−0.061 *
	(0.039)	(0.104)	(0.035)	(0.065)	(0.078)	(0.04)	(0.03)
OUT	−0.595	−4.310	0.723	−0.240	0.484	0.43	0.66
	(0.404)	(2.696)	(0.747)	(0.722)	(1.164)	(0.46)	(0.81)
MSR	−0.474 **	−0.935	0.446	−0.065	0.381	0.247 *	0.12
	(0.206)	(0.799)	(0.268)	(0.268)	(0.375)	(0.14)	(0.23)
DIV	−0.035	0.182	0.173 *	−0.061	0.112	0.127 ***	0.395 **
	(0.037)	(0.249)	(0.093)	(0.070)	(0.149)	(0.05)	(0.16)
DUAL	−0.027	0.152	0.089	0.065	0.154	0.05	(0.01)
	(0.131)	(0.502)	(0.196)	(0.185)	(0.371)	(0.06)	(0.08)
Constant	−3.705 ***	−6.359 *	−0.939	−5.097 ***	−6.036 **	−7.344 ***	−4.522 ***
	(1.266)	(3.570)	(1.826)	(1.829)	(2.937)	(1.18)	(1.26)
Observations	15,993	15,993	15,993	15,993	15,993	15,194	15,194
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	No	Yes
Adjust R^2	0.0135	0.0139	0.0159	0.00875	0.0152		
AR (1) test						0.000	0.000
AR (2) test						0.168	0.262
Sargan test						0.000	0.000
Hansen test						0.159	0.129

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

5.2.6. Dynamic Panel Estimates

Since the possible endogeneity between RBDAC and GTI may lead to regression bias, this paper reports the robustness tests on the dynamic panel data. We set up a dynamic model by taking in a lagged measure of dependent variables, using the two-step system generalized method of moments (GMM) to address this concern. Column (6) in Table 8 shows the outcomes of the system GMM without controlling the fixed effects of Year, while Column (7) controls. The regression results show that coefficients of DID and lagged variable are significantly positive. The p value of AR (1) test is less than 0.1, the p value of AR (2) is greater than 0.1, and the p value of Hansen test is greater than 0.1, indicating that there is no second-order autocorrelation and over-identification of instrumental variables in the regressions. The regression and further tests show that the results of our model are reliable.

Figure 5 shows the regression coefficients of robustness tests.

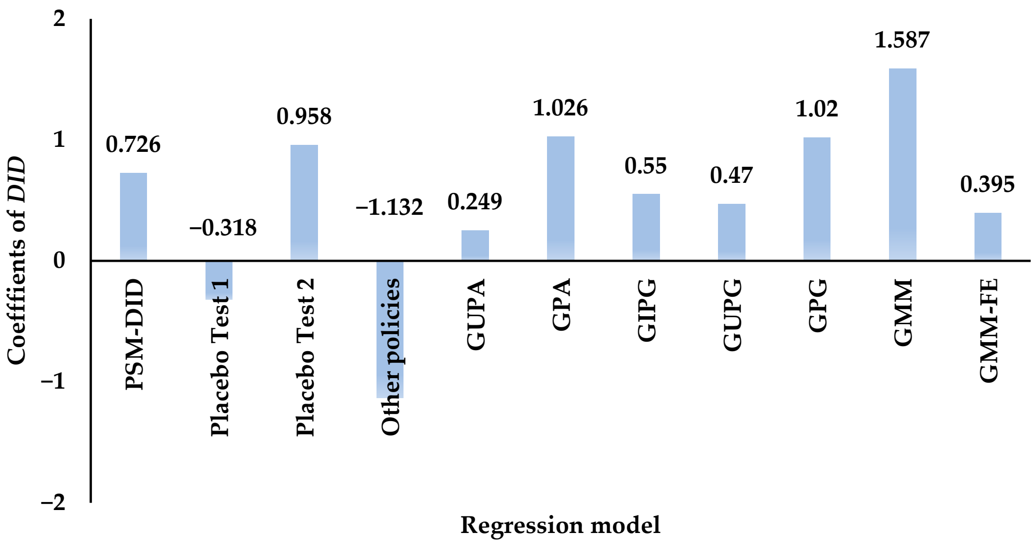

Figure 5. Regression coefficients of robustness tests.

6. Extensive Analysis

6.1. Test of the Impact Mechanism

The previous content explains that big data development has a direct positive impact on the advancement of GTI. The further study verifies the mechanism of government subsidy and analyst coverage, as well as investigates the internal mechanism of promoting the GTI in firms through RBDAC. The goal of this section is to resolve the question of whether RBDAC can demonstrate a higher level of GTI when the level of government subsidy or analyst coverage has been improved.

The new equation will be generated based on the benchmark regression by sequentially adding moderator variables and interaction terms between moderating factors and explanatory variables. The equations are presented as follows.

$$GIPA_{i,t+1} = \beta_0 + \beta_1 SUBSIDY_{i,t} \times DID_{i,t} + \beta_2 SUBSIDY_{i,t} \times TREAT_{i,t} + \beta_3 SUBSIDY_{i,t_i} \times TIME_{i,t} + \beta_4 DID_{i,t} + \beta_5 TREAT_{i,t} + \beta_6 TIME_{i,t} + \beta_7 SUBSIDY_{i,t} + \beta_8 CV_{i,t} + \alpha_i + \varepsilon_{i,t} \quad (3)$$

$$GIPA_{i,t+1} = \beta_0 + \beta_1 ANALYST_{i,t} \times DID_{i,t} + \beta_2 ANALYST_{i,t} \times TREAT_{i,t} + \beta_3 ANALYST_{i,t} \times TIME_{i,t} + \beta_4 DID_{i,t} + \beta_5 TREAT_{i,t} + \beta_6 TIME_{i,t} + \beta_7 ANALYST_{i,t} + \beta_8 CV_{i,t} + \alpha_i + \varepsilon_{i,t} \quad (4)$$

The moderator variables in this section are government subsidy (*SUBSIDY*) and analyst coverage (*ANALYST*). Columns 1 and 2 in Table 9 demonstrate that the coefficient estimates for all interaction terms are significant at the 5% level. In particular, the moderator variable in column 1 of Table 9 is government subsidy, and the coefficient for the interaction term between *SUBSIDY* and *DID* is 0.515 and statistically significant at the 5% level, implying that firms receiving more government subsidy are more efficient in developing big data to stimulate the GTI. Therefore, the second hypothesis (H2) is accepted. The results indicate that the government's financial subsidies to firms can help to develop the positive effect of RBDAC on firms' GTI output. The improvement of big data application capability has attracted enterprises that are expected to apply for subsidies, while the issuance of government subsidies has sent a positive signal to investors, improved firm financing, increased capital investment in innovation, and provided convenience for firms to innovate in green technology. The conclusion is consistent with reference to previous studies [62], arguing that government subsidy has a positive effect on green innovation.

Furthermore, the moderator variable in column 2 of Table 9 is *ANALYST*. The interaction term has a coefficient of 1.146 and is significant at the 5% level, implying that firms receiving more analyst attention can strengthen the positive relationship between RBDAC and GTI. Therefore, the third hypothesis (H3) is accepted. The results indicate that the application of big data technology has improved the transparency of firm information and analyst information recognition ability, while firms that obtain financial analyst attention receive more social publicity and supervision, reduce managers' opportunistic behavior, and improve managers' enthusiasm for R&D. The conclusion is consistent with reference to previous studies [60], arguing that analyst coverage has a positive effect on firms' green innovation performance.

Table 9. Test of the impact mechanism.

	(1)	(2)	(3)	(4)
Variables	**Government Subsidy**	**Analyst Coverage**	**Government Subsidy**	**Analyst Coverage**
SUBSIDY × DID	0.515 **		3.367 **	
	(0.213)		(1.564)	
SUBSIDY × TREAT	−0.434 **		−4.164 **	
	(0.205)		(1.788)	
SUBSIDY × TIME	0.173 ***		1.813 ***	
	(0.043)		(0.405)	
SUBSIDY	−0.088 ***		−0.700 **	
	(0.030)		(0.288)	
ANALYST × DID		1.146 **		0.895 **
		(0.427)		(0.354)
ANALYST × TREAT		−0.449 *		−0.381 *
		(0.258)		(0.209)
ANALYST × TIME		0.350 ***		0.276 ***
		(0.115)		(0.090)
ANALYST		−0.199 ***		−0.150 **
		(0.070)		(0.056)
DID	−7.646 **	−1.024 **	−1.765	−0.984 **
	(3.157)	(0.435)	(1.058)	(0.457)
SIZE	0.191	−0.103	0.200 *	−0.094
	(0.115)	(0.183)	(0.105)	(0.187)
LEV	−0.521	−0.458	−0.656	−0.462
	(1.093)	(1.081)	(1.143)	(1.072)
GROWTH	−0.034	0.015	−0.081	0.007
	(0.113)	(0.129)	(0.099)	(0.127)
OUT	−3.889	−3.950	−3.731	−3.983
	(2.674)	(2.600)	(2.647)	(2.584)
MSR	−0.653	−0.730	−0.362	−0.710
	(0.673)	(0.690)	(0.641)	(0.693)
DIV	0.192	0.171	0.213	0.172
	(0.221)	(0.211)	(0.217)	(0.215)
DUAL	0.172	0.179	0.180	0.173
	(0.391)	(0.361)	(0.386)	(0.359)
Constant	1.853	4.804	−0.724	4.601
	(3.982)	(5.240)	(3.340)	(5.337)
Observations	15,993	15,993	15,993	15,993
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Adjust R^2	0.0177	0.0183	0.0118	0.0180

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

This paper also replaces *SUBSIDY* with "ratio of government subsidy to operating income" and *ANALYST* with "the logarithm of the number of attention level by securities report adding 1". According to Table 9, in columns 3 and 4, the substitution variables of government subsidy and analyst coverage can also significantly positively regulate the relationship between RBDAC and GTI. As a result, the robustness of moderator variables has been demonstrated.

Figure 6 shows the regression coefficients of the interaction term between RBDAC and moderator variables (government subsidies, analyst coverage) on GTI.

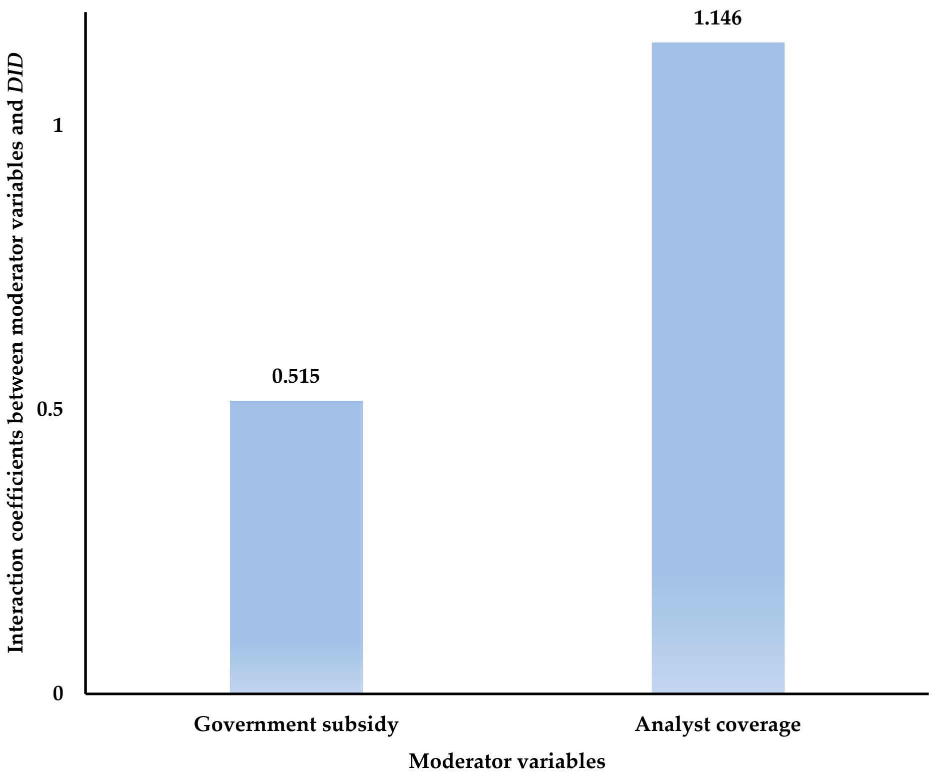

Figure 6. Regression coefficients of moderator variables on GTI.

6.2. Test of Heterogeneity

Previous evidence confirms that RBDAC promotes GTI in the manufacturing firms. Furthermore, government subsidy and analyst coverage positively regulate the impact of RBDAC on GTI. However, two points need special attention. The first is that the RBDAC innovation incentives are influenced by factors other than government subsidy and analyst coverage. Furthermore, RBDAC has various effects on firms. As a result, the heterogeneity of the institutional environment and firms will be tested. The institutional heterogeneity research examines three factors: finance constraints, tax administration strengths, and region. The examination of firms' heterogeneity that follows is based on property rights and top management team. This study examines the heterogeneity factors above in groups and regresses by controlling the Year and Firm fixed effects.

6.2.1. Heterogeneity of Institutional Environment

The institutional and economic environments influence the heterogeneity of the RBDAC innovation effect. This article chooses finance constraints and tax administration strengths to measure institutional pressure on firms, while regional heterogeneity assesses the level of economic development in the region where firms are located. High finance constraints and tax administration strengths may be detrimental to firm digitalization and reduce the RBDAC innovation effects. In developed regions, finances and legislative support for digital transformation may be more conducive to GTI.

(1) Finance constraints. To begin with, Hadlock and Pierce inspire to construct the SA index model related to the degree of finance constraints [65]. The equation is formatted as follows:

$$SA = -0.737SIZE + 0.043SIZE^2 - 0.04AGE \tag{5}$$

where SIZE is the logarithm of the asset size of a firm, and AGE is the listing age of a firm. Furthermore, we use the logarithm of the absolute value of the SA index as a proxy indicator for financing constraints (*FC*). The bigger the value of *FC*, the greater the pressure on firms from financing constraints. Columns 1 and 2 of Table 10 show the coefficients of explanatory variables when firms face lesser or higher financing constraints, respectively. According to the empirical findings, RBDAC may promote greater innovation willingness in GTI in a group with lower financial constraints. The genesis of this phenomenon could be attributed to the fact that when firms have lower financial constraints, they are under less financial pressure to upgrade big data to drive the GTI, and as a result, they are more inclined to contribute more innovation funds.

(2) Tax administration strengths. The institutional context influences the heterogeneity of the RBDAC innovation impact, so this article selects tax administration strengths to assess firm institutional pressure. However, the high tax administration may be detrimental to firm digitalization. Meanwhile, the potential to innovate driven by RBDAC is also inhibited. Based on the results of Cao et al. [67], the following tax revenue prediction model is formed as follows:

$$TAX/GDP = \alpha_0 + \alpha_1 * IND1/GDP + \alpha_2 * IND2/GDP + \alpha_3 * OPEN/GDP + \varepsilon \quad (6)$$

where *TAX* is the tax revenue in each region, *IND*1 and *IND*2 are the annual output values of the primary and secondary industries in each region, *OPEN* is the domestic import and export value in each region, and *GDP* is the gross domestic product in each region. We use the model's goodness-of-fit test to estimate tax income for each region. Furthermore, the proxy variable for tax administration strengths (*TE*) is calculated by dividing the actual annual tax in each region by the expected tax from the revenue-fitted model. The larger the *TE* value, the higher the tax administration strength in this region. Columns 3 and 4 of Table 10 show the grouping results of low and high tax administration strengths. Empirical evidence demonstrates that when compared to firms with higher tax administration pressure, firms with lower tax administration pressure are more ready to participate in big data development to facilitate the GTI. The reason the correlation coefficient of groups with high tax administration strengths is not significant may be the rising tax pressure, which reduces the cash flow available to firms for developing big data and suppresses innovation investment.

(3) Regional heterogeneity. As the geographical environment of firms influences the innovation heterogeneity induced by RBDAC, this article investigates the effects of regional heterogeneity on GTI. The sample area is divided into two groups in this article: the eastern region (including Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan) and the non-eastern region. The paper then investigates if regional heterogeneity influences the beneficial effects of regional big data progress on GTI. As shown in columns 5 and 6 of Table 10, the grouping results of the eastern area and non-eastern region reveal that big data development boosts the GTI more significantly for firms located in the eastern region. However, the *DID* coefficient for the eastern region group is close to that of the non-eastern region. The fact that the coefficients of the eastern region and non-eastern region are not significantly different may be due to governments in the central and western regions recognizing the ability of big data to drive innovation and beginning to deeply integrate big data with the industrial economy, thereby stimulating firm innovation in the region.

Table 10. Result of heterogeneity test I.

	Financing Constraints		**Tax Administration Strengths**		**Regional Heterogeneity**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Variables	**Low**	**High**	**Low**	**High**	**Non-Eastern Region**	**Eastern Region**
DID	1.781 **	0.337	1.101 ***	0.458	0.707	0.779 *
	(0.688)	(0.569)	(0.281)	(0.470)	(0.590)	(0.419)
TIME	−0.489	0.803 **	0.255	0.110	−0.321	−0.471
	(0.959)	(0.328)	(0.413)	(0.231)	(0.298)	(0.588)
SIZE	−0.021	0.274	0.370 **	0.237 *	0.409 *	0.082
	(0.256)	(0.193)	(0.148)	(0.121)	(0.236)	(0.110)
LEV	−2.209	−0.001	−2.867	0.865 **	−0.172	−0.790
	(3.244)	(0.417)	(2.165)	(0.414)	(0.641)	(1.743)
GROWTH	0.069	−0.156	−0.111	−0.099 *	−0.212	−0.017
	(0.213)	(0.106)	(0.216)	(0.054)	(0.146)	(0.144)
OUT	−8.302	0.906	−7.406	−0.101	−0.677	−5.653
	(5.632)	(1.340)	(5.617)	(0.808)	(1.355)	(3.971)
MSR	−0.137	−0.067	−1.163	0.154	0.641	−0.751
	(1.074)	(0.622)	(0.835)	(0.602)	(0.461)	(0.819)
DIV	0.428	0.053	−0.087	0.337	0.160	0.247
	(0.499)	(0.097)	(0.202)	(0.310)	(0.176)	(0.345)

Table 10. *Cont.*

	Financing Constraints		Tax Administration Strengths		Regional Heterogeneity	
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Low	High	Low	High	Non-Eastern Region	Eastern Region
DUAL	0.634	−0.050	0.590	−0.233 ***	−0.311	0.410
	(0.847)	(0.208)	(0.732)	(0.079)	(0.216)	(0.523)
Constant	3.386	−5.840	−3.257	−5.470 *	−8.492	1.002
	(6.549)	(4.485)	(5.649)	(2.738)	(5.042)	(3.989)
Observations	7997	7996	8048	7945	5478	10,515
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjust R^2	0.0152	0.0155	0.0121	0.0193	0.0207	0.00985

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

6.2.2. Heterogeneity of Firms

Property rights and top management team heterogeneity are two aspects of firms' heterogeneity.

(1) The heterogeneity of property rights. From the nature of property rights, different property rights have an impact on innovative heterogeneity for developing big data applications. State-owned firms are more responsive to policies than non-state-owned firms; therefore, they are more likely to obtain government support and have stronger financing capabilities. However, more research is needed to find out if the incentive effect of RBDAC can be reinforced for firms that are both state-owned and located in a big data pilot zone. This article divides the sample firms into state-owned and non-state-owned groups based on the real control of firms, and then investigates whether the nature of property rights and regional heterogeneity affect the innovation incentive effect from RBDAC. The grouping results of non-state-owned and state-owned firms are presented in columns 1 and 2 of Table 11, confirming that state-owned firms benefit more from the innovative effects of big data development than non-state-owned firms. This could be because state-owned firms have a natural advantage of obtaining government support, and firms located in big data pilot zones are more likely to gain technical support. Thus, the state-owned firms located in big data pilot zones perform better in boosting the GTI of the national manufacturing industry.

(2) The heterogeneity of the top management team. According to the upper echelons theory, big data is an entirely new technology, hence its development is vulnerable to decisions made by the top management team. As a result, the level of heterogeneity in senior executives is critical for decision-making. This article chooses three factors for top management team heterogeneity testing: age, gender, and professional experience. The age heterogeneity is measured by the age coefficient of variation which is the ratio of the standard deviation of age to the age mean. In addition, the gender (female and male) and professional background heterogeneity (including production, R&D, design, human resources, management, marketing, finance, law, etc.) are measured through the Herfindal–Hirschman Index. The formula is constructed as follows:

$$H = 1 - \sum_{i=1}^{n} P_i^2 \tag{7}$$

where the proportion of *i*-class members is in the team. The interpretation of the index is that a higher Herfindal–Hirschman Index reflects greater gender and professional background variability. The regression results of age, gender, and professional background heterogeneity are shown from column 3 to column 8 in Table 11. The results imply that senior executives in organizations with higher heterogeneity of age, gender, and work background variability have varying support in creating synergistic effects between RBDAC and GTI. Presumably, a top management team with a larger age heterogeneity has more youthful senior executives who are more likely to be risk-takers and prepared to invest in risky new technologies. Furthermore, a top management team with a higher level of gender heterogeneity means more female senior executives are involved who may prefer to invest in GTI to reduce pollution. Finally, a top management team with higher professional background heterogeneity might provide new perspectives on how to maximize the benefits of RBDAC, thereby enhancing the decision-making quality.

Table 11. Result of heterogeneity test II.

	Heterogeneity of Top Management Team							
	Property Rights		**Age**		**Gender**		**Professional Background**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Variables	**Non-State-Owned**	**State-Owned**	**Low**	**High**	**Low**	**High**	**Low**	**High**
DID	0.029	1.863 *	0.723	1.360 *	0.359	1.702 **	0.253	0.748 *
	(0.127)	(0.926)	(0.623)	(0.761)	(0.475)	(0.632)	(0.208)	(0.369)
TIME	−1.258	1.158 **	−0.879	0.256	−0.676 ***	0.966 **	−0.055	−1.281
	(0.845)	(0.476)	(0.711)	(0.206)	(0.243)	(0.380)	(0.119)	(1.165)
SIZE	0.247 ***	0.589 *	0.190	0.130	0.409 **	−0.114	0.083	0.320 *
	(0.087)	(0.301)	(0.154)	(0.140)	(0.162)	(0.210)	(0.085)	(0.176)
LEV	−0.271	−0.647	0.161	−1.515	0.427	−0.609	0.169	−1.560
	(0.472)	(2.871)	(1.233)	(1.568)	(0.468)	(1.784)	(0.180)	(2.083)
GROWTH	−0.025	−0.446 **	0.020	−0.056	−0.173 *	−0.010	−0.015	−0.116
	(0.104)	(0.213)	(0.193)	(0.080)	(0.099)	(0.083)	(0.063)	(0.228)
OUT	−2.121	−6.783	−3.412	−2.134	−1.139	−5.523 **	1.647	−8.388
	(2.140)	(4.898)	(2.612)	(3.188)	(2.950)	(2.578)	(1.183)	(4.997)
MSR	−0.898	−1.345	−2.089	0.351	−0.585	1.098 *	0.116	−1.163
	(0.943)	(2.017)	(2.624)	(0.613)	(0.804)	(0.571)	(0.519)	(1.375)
DIV	0.023	0.421	0.549	−0.177	0.001	0.428	0.024	0.268 **
	(0.046)	(0.656)	(0.373)	(0.162)	(0.168)	(0.467)	(0.109)	(0.116)
DUAL	−0.020	0.834	−0.163	0.478	0.169	0.257	−0.085	0.336
	(0.136)	(1.128)	(0.275)	(0.545)	(0.300)	(0.433)	(0.170)	(0.387)
Constant	−3.074	−10.091 *	−2.326	−1.518	−8.152 *	3.771	−2.271	−1.997
	(2.764)	(5.814)	(4.310)	(3.995)	(4.072)	(4.489)	(1.973)	(5.784)
Observations	10,763	5230	8130	7863	8780	7213	8089	7904
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjust R^2	0.00779	0.0204	0.0148	0.00837	0.0187	0.0106	0.0132	0.0124

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

Figure 7 shows the regression coefficients of heterogeneity tests.

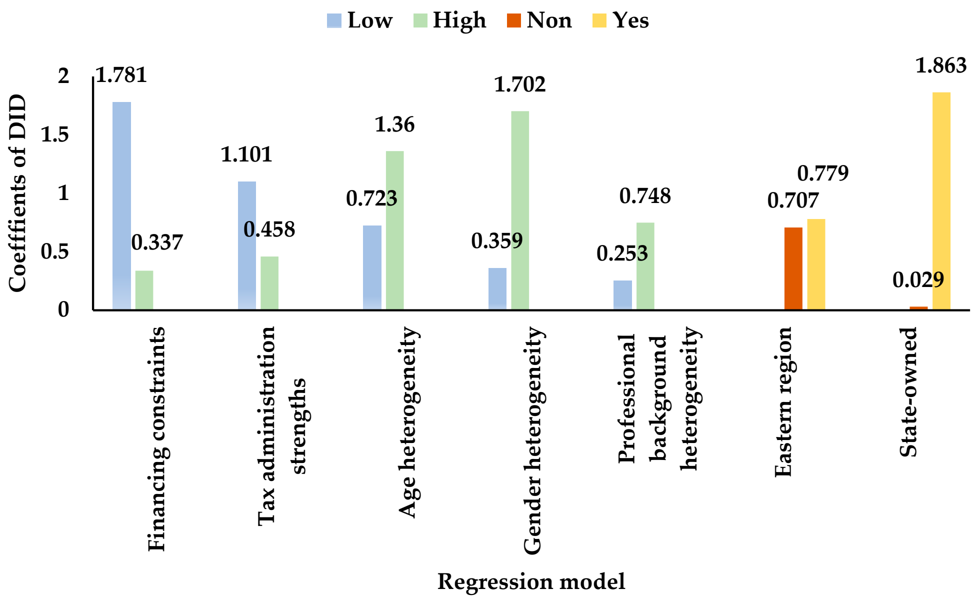

Figure 7. Regression coefficients of heterogeneity tests.

6.3. Test of Economic Outcomes

The preceding section has shown that the development of RBDAC may promote the GTI in manufacturing firms. The goal of using green innovation is to encourage environmental protection and green governance for firms. As a result, it is worth considering whether the improved level of GTI as a result of regional big data growth will affect the economic consequences for firms and improve environmental quality. Table 12 shows the estimated results of the economic effects of RBDAC in this article. The explanatory variable shown in column 1 of Table 12 is firm performance on Environmental, Social, and Governance (*ESG*). Furthermore, the performance will be evaluated using the ESG rating using the Sino-Securities Index of the WIND databases in China. The explanatory variables in columns 2 and 3 are ESG performance with one and two lagging periods (*L1.ESG* and *L2.ESG*), respectively. According to the findings in Table 12, the improvement in GTI brought about by RBDAC will continue to improve firm ESG performance. The conclusion is consistent with reference to previous studies [28], arguing that big data have a positive effect on environmental performance. Specifically, the development of regional big data has promoted the innovative development of firms' application of new information technology and has become an important technological support to improve firms' ESG practice. First, from the perspective of environmental performance, firms use big data and other information technologies to achieve the arrangement and supervision of production processes, improve production efficiency, reduce resource waste and pollution emissions, provide technical support for green innovation, and thus improve the performance of firms in environmental sustainable development [68]. For example, China's steel sector uses digital technology to intelligently manage production and decision-making, track product carbon footprint data, and achieve low-carbon transformation. Second, from the perspective of social performance, the application of big data meets the interests of shareholders, and firms are more willing to fulfill social responsibilities, which will encourage firms to actively participate in environmental projects, adopt energy-saving and environmental protection technologies, carry out green innovation activities, improve firm reputation [69], and thus improve firm performance in fulfilling social responsibilities [70]. Firms with good ESG performance are more likely to receive government subsidy and investor attention, forming an effective cycle between good social reputation and firm ESG practices. Third, from the perspective of corporate governance, the application of big data has improved firm operation capability, accelerated the exchange of firm information [71], and improved resource allocation efficiency, all of which are conducive to obtaining more green technology support and improving the level of firm ESG practice [72]. In addition, this good impact will result in a long-term beneficial cycle for firms and the environment. Finally, firms will achieve better ESG performance through the development of capabilities in RBDAC, resulting in positive economic effects.

Table 12. Impact of RBDAC on ESG performance.

	(1)	**(2)**	**(3)**
Variables	**ESG**	**L1.ESG**	**L2.ESG**
DID	0.066 *	0.073 *	0.089 **
	(0.039)	(0.039)	(0.041)
TIME	0.073	−0.323 ***	−0.110 ***
	(0.056)	(0.038)	(0.038)
SIZE	0.184 ***	0.155 ***	0.098 ***
	(0.029)	(0.034)	(0.036)
LEV	−0.606 ***	−0.250 **	−0.013
	(0.092)	(0.099)	(0.105)
GROWTH	0.004	−0.054 ***	−0.059 ***
	(0.020)	(0.019)	(0.021)
OUT	−0.472 **	−0.258	−0.063
	(0.240)	(0.250)	(0.261)
MSR	0.486 ***	0.924 ***	0.631 ***
	(0.109)	(0.200)	(0.213)
DIV	0.183 ***	0.108 ***	0.059 ***
	(0.022)	(0.022)	(0.021)
DUAL	−0.043	−0.010	0.026
	(0.029)	(0.030)	(0.030)
Constant	2.367 ***	3.098 ***	4.258 ***
	(0.622)	(0.755)	(0.813)
Observations	15,993	13,456	11,080
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Adjust R^2	0.0541	0.0460	0.0339

Note: ***, **, and * represent significant levels at 1%, 5%, and 10%, respectively.

Figure 8 shows the regression coefficients of RBDAC on ESG (current, lagging one period and lagging two periods).

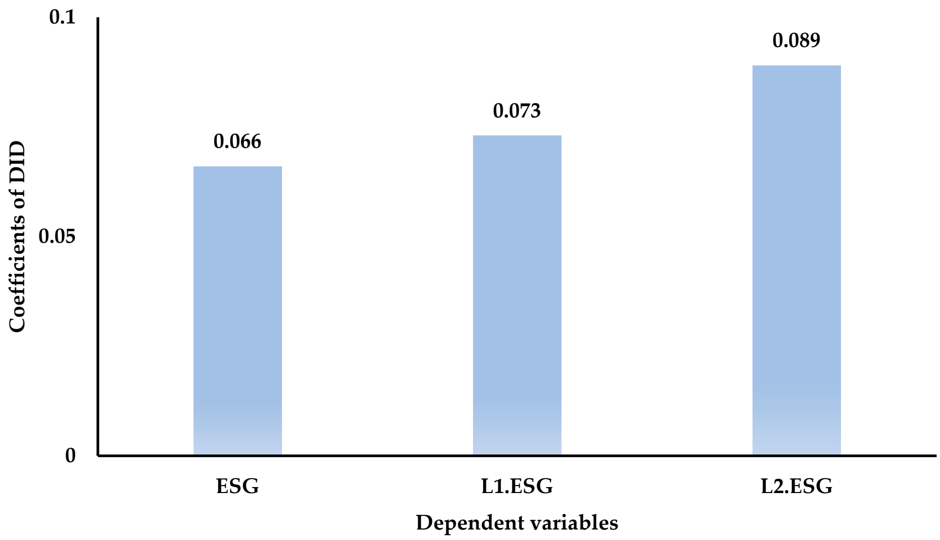

Figure 8. Regression coefficients of RBDAC on economic consequence variables.

7. Conclusions

7.1. Research Conclusions

Recently, the Chinese government has stated its strong support for the development of the digital economy by issuing several policies to support the development of the digital industry. RBDAC can become one of the fundamental infrastructures to support the development of the digital economy, which will have a far-reaching impact on GTI. A DID model is constructed around the new economic paradigm of "RBDAC-GTI" in this article, and China's national comprehensive pilot zone of big data was built in 2016 to serve as a quasi-natural experiment. The main conclusions are shown as follows:

(1) Based on data from manufacturing firms in Shanghai and Shenzhen A-shares from 2010 to 2020, the research finds that RBDAC has promoted the GTI level of manufacturing firms to a certain extent. This is in line with the literature [10], supporting that big data play an important role in promoting firm innovation. We have answered the goal of hypothesis 1, that is regional big data development can drive firms' digital transformation, so as to develop the customization requirements for products in a green way, and thus advance firm GTI.

(2) The GTI incentive effect of RBDAC is moderated by the positive effect of government subsidy and analyst coverage, that is, the innovation incentive level of RBDAC is more significant in firms with high government subsidy and high analyst coverage. This is in line with the literature [62], supporting that government subsidy and analyst coverage play an important role in promoting firms' green innovation. We have answered the goal of hypothesis 2 and 3, which is that regional big data development can attract firms that are expected to apply for subsidies and improve analyst information recognition ability, then the government subsidy and analyst attention can attract more investors' attention. Thus, government subsidy and analyst coverage can improve firm financing and provide financial support for green innovation practices.

(3) The sub-sample regression test examines the heterogeneity effects of the external institutional environment and internal conditions of firms on the impact of RBDAC on GTI. Regarding the aspect of the external institutional environment, the result shows that firms in regions with low levels of financial constraints and tax administration strengths and firms located in the eastern region can strengthen the effect of RBDAC on GTI. On the other aspect of internal conditions of firms, the result shows that state-owned firms and firms with a high heterogeneity of the top management team in terms of age, gender, and professional background have a stronger influence on the positive effects of RBDAC on GTI.

(4) The economic consequence test demonstrates that RBDAC improves firm ESG performance and that this influence sustains over time. This is in line with the literature [3], supporting that big data play an important role in promoting firms' environmental and social performance. The result clarifies that the improvement of regional big data development can help firms to reduce resource waste, raise resource efficiency, increase firm return, and provide support for green governance. As the widespread use of new technologies promotes environmental performance

and fosters a positive image of actively engaging in social duties, firm ESG practice will be elevated to a new and higher level.

7.2. Implication

This paper presents the topic of RBDAC and the development of GTI in manufacturing with the following insights:

(1) Promote the development of regional big data strategy, and enhance the incentive role of big data in GTI. There is an increasing consciousness of the value of developing the RBDAC to accelerate the advancement of GTI. Since RBDAC promotes the GTI in manufacturing firms, the established national big data comprehensive pilot zone should continue to work on the implementation and supplementation of big data policies, as well as the acceleration of big data infrastructure construction. The provinces that did not participate in the pilot should also follow the national policy orientation, with a focus on promoting big data initiatives and driving the innovative role of regional big data in empowering the local economy.

(2) The government should improve government subsidy policies related to digital technology and green innovation, while the capital market should regulate a set of criteria for analyst coverage to monitor firms. Existing government subsidy policies are not so efficient at determining whether funds are earmarked for GTI. As a result, big data should be used to standardize the review and approval process of subsidy applications, as well as conduct follow-up investigations on the actual usage situation of the subsidies to ensure the high efficiency of using government subsidies. Subsidy requirements should be moderately lowered if firms perform excellently in green development, so that they might be encouraged to scale up green innovation and achieve a win-win situation for the economy and the environment. Analyst coverage is one of the most efficient channels for connecting firms and the market. The government supervision of financial analysts is to enable analysts to better leverage their important function in reducing information asymmetry and promoting innovation. The government should use big data to help investors understand the lags in the economic and environmental performance of firms, encourage analysts to identify high-quality firms with strong GTI capabilities, and motivate firms to carry out effective GTI activities.

7.3. The Research Limitations and Agenda

This study cites only a small portion of the big data literature that is relevant to major research interests. Even though those references have a deep understanding of big data, this article's comprehensive understanding of big data has several limitations.

(1) The performance of GTI is simply quantified as the number of patent applications. Quantity is a crucial indicator of the GTI development, but the quality of innovation is also an assignable component. As a result, future research should consider innovation quality.

(2) This study only discusses the relationship between RBDAC and GTI and confirms a positive effect of RBDAC on GTI. However, a study of the interaction mechanism is lacking, thus future research should focus on investigating this mechanism.

(3) The manufacturing industry is used as a case study to investigate the relationship between RBDAC and GTI, but GTI has far-reaching implications for heavily polluted industries. As a result, greater research into the effects of GTI on polluting firms is required.

Author Contributions: Conceptualization, J.S. and G.C.; methodology, G.C. and X.F.; software, G.C.; data curation, G.C.; writing, G.C. and Y.C.; investigation, X.F.; supervision, X.F. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Key Research Base Project of Philosophy and Social Sciences in Jiangxi Province, grant number 21JDJC01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available from the corresponding author on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Niebel, T.; Rasel, F.; Viete, S. BIG data–BIG gains? Understanding the link between big data analytics and innovation. *Econ. Innov. New Technol.* **2018**, *28*, 296–316. [CrossRef]
2. Hao, S.; Zhang, H.; Song, M. Big Data, Big Data Analytics Capability, and Sustainable Innovation Performance. *Sustainability* **2019**, *11*, 7145. [CrossRef]
3. Dubey, R.; Gunasekaran, A.; Childe, S.J.; Papadopoulos, T.; Luo, Z.; Wamba, S.F.; Roubaud, D. Can big data and predictive analytics improve social and environmental sustainability? *Technol. Forecast. Soc. Chang.* **2019**, *144*, 534–545. [CrossRef]
4. Wang, Y.; Yang, Y.; Fu, C.; Fan, Z.; Zhou, X. Environmental regulation, environmental responsibility, and green technology innovation: Empirical research from China. *PLoS ONE* **2021**, *16*, e0257670. [CrossRef] [PubMed]
5. Ramadan, M.; Shuqqo, H.; Qtaishat, L.; Asmar, H.; Salah, B. Sustainable Competitive Advantage Driven by Big Data Analytics and Innovation. *Appl. Sci.* **2020**, *10*, 6784. [CrossRef]
6. El-Kassar, A.N.; Singh, S.K. Green innovation and organizational performance: The influence of big data and the moderating role of management commitment and HR practices. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 483–498. [CrossRef]
7. Ghasemaghaei, M. Improving Organizational Performance Through the Use of Big Data. *J. Comput. Inform. Syst.* **2018**, *60*, 395–408. [CrossRef]
8. Fan, W.; Bifet, A. Mining big data. *ACM SIGKDD Explor. Newsl.* **2013**, *14*, 1–5. [CrossRef]
9. Mazhar, T.; Irfan, H.M.; Khan, S.; Haq, I.; Ullah, I.; Iqbal, M.; Hamam, H. Analysis of Cyber Security Attacks and Its Solutions for the Smart grid Using Machine Learning and Blockchain Methods. *Future Internet* **2023**, *15*, 83. [CrossRef]
10. Ghasemaghaei, M.; Calic, G. Assessing the impact of big data on firm innovation performance: Big data is not always better data. *J. Bus. Res.* **2020**, *108*, 147–162. [CrossRef]
11. Akoka, J.; Comyn-Wattiau, I.; Laoufi, N. Research on Big Data–A systematic mapping study. *Comput. Stand. Inter.* **2017**, *54*, 105–115. [CrossRef]
12. Xie, W.; Zhang, Q.; Lin, Y.; Wang, Z.; Li, Z. The Effect of Big Data Capability on Organizational Innovation: A Resource Orchestration Perspective. *J. Knowl. Econ.* **2023**, *ahead-of-print*. [CrossRef]
13. Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big data analytics and firm performance: Findings from a mixed-method approach. *J. Bus. Res.* **2019**, *98*, 261–276. [CrossRef]
14. Wang, L.; Wu, Y.; Huang, Z.; Wang, Y. How big data drives green economic development: Evidence from China. *Front. Environ. Sci.* **2022**, *10*, 1055162. [CrossRef]
15. Wei, J.; Zhang, X. The Role of Big Data in Promoting Green Development: Based on the Quasi-Natural Experiment of the Big Data Experimental Zone. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4097. [CrossRef]
16. Soofi, A.; Awan, A. Classification Techniques in Machine Learning: Applications and Issues. *J. Basic Appl. Sci.* **2017**, *13*, 459–465. [CrossRef]
17. Aljumah, A.I.; Nuseir, M.T.; Alam, M.M. Organizational performance and capabilities to analyze big data: Do the ambidexterity and business value of big data analytics matter? *Bus. Process. Manag. J.* **2021**, *27*, 1088–1107. [CrossRef]
18. Prange, C.; Verdier, S. Dynamic capabilities, internationalization processes and performance. *J. World Bus.* **2011**, *46*, 126–133. [CrossRef]
19. Haq, I.; Mazhar, T.; Malik, M.A.; Kamal, M.M.; Ullah, I.; Kim, T.; Hamdi, M.; Hamam, H. Lung Nodules Localization and Report Analysis from Computerized Tomography (CT) Scan Using a Novel Machine Learning Approach. *Appl. Sci.* **2022**, *12*, 12614. [CrossRef]
20. Ali, T.M.; Nawaz, A.; Ur Rehman, A.; Ahmad, R.Z.; Javed, A.R.; Gadekallu, T.R.; Chen, C.; Wu, C. A Sequential Machine Learning-cum-Attention Mechanism for Effective Segmentation of Brain Tumor. *Front. Oncol.* **2022**, *12*, 873268. [CrossRef]
21. Lin, S.Z.; Lin, J.B. How organizations leverage digital technology to develop customization and enhance customer relationship performance: An empirical investigation. *Technol. Forecast. Soc. Chang.* **2023**, *188*, 122254. [CrossRef]
22. Jaouadi, M.H.O. Investigating the influence of big data analytics capabilities and human resource factors in achieving supply chain innovativeness. *Comput. Ind. Eng.* **2022**, *168*, 108055. [CrossRef]
23. Lozada, N.; Arias-Perez, J.; Perdomo-Charry, G. Big data analytics capability and co-innovation: An empirical study. *Heliyon* **2019**, *5*, e02541. [CrossRef] [PubMed]
24. Tunc-Abubakar, T.; Kalkan, A.; Abubakar, A.M. Impact of big data usage on product and process innovation: The role of data diagnosticity. *Kybernetes* **2022**, *ahead-of-print*. [CrossRef]
25. Li, H.Y.; Liu, Q.; Ye, H.Z. Digital Development Influencing Mechanism on Green Innovation Performance: A Perspective of Green Innovation Network. *IEEE Access* **2023**, *11*, 22490–22504. [CrossRef]
26. Zhang, Z.; Shang, Y.; Cheng, L.; Hu, A. Big Data Capability and Sustainable Competitive Advantage: The Mediating Role of Ambidextrous Innovation Strategy. *Sustainability* **2022**, *14*, 8249. [CrossRef]
27. Wamba, S.F.; Gunasekaran, A.; Akter, S.; Ren, S.J.-F.; Dubey, R.; Childe, S.J. Big data analytics and firm performance: Effects of dynamic capabilities. *J. Bus. Res.* **2017**, *70*, 356–365. [CrossRef]
28. Belhadi, A.; Kamble, S.S.; Zkik, K.; Cherrafi, A.; Touriki, F.E. The integrated effect of Big Data Analytics, Lean Six Sigma and Green Manufacturing on the environmental performance of manufacturing companies: The case of North Africa. *J. Clean. Prod.* **2020**, *252*, 119903. [CrossRef]

29. Sahoo, S.; Kumar, A.; Upadhyay, A. How do green knowledge management and green technology innovation impact corporate environmental performance? Understanding the role of green knowledge acquisition. *Bus. Strateg. Environ.* **2022**, *32*, 551–569. [CrossRef]
30. Deng, Q.; Zhou, S.; Peng, F. Measuring Green Innovation Efficiency for China's High-Tech Manufacturing Industry: A Network DEA Approach. *Math. Probl. Eng.* **2020**, *2020*, 8902416. [CrossRef]
31. Lin, X.; Yu, L.; Zhang, J.; Lin, S.; Zhong, Q. Board Gender Diversity and Corporate Green Innovation: Evidence from China. *Sustainability* **2022**, *14*, 15020. [CrossRef]
32. Castellacci, F.; Lie, C.M. A taxonomy of green innovators: Empirical evidence from South Korea. *J. Clean. Prod.* **2017**, *143*, 1036–1047. [CrossRef]
33. Porter, M.E.; Linde, C.V.D. Toward a New Conception of the Environment-Competitiveness Relationship. *J. Econ. Perspect.* **1995**, *9*, 97–118. [CrossRef]
34. Fan, Y.; Su, Q.; Wang, X.; Fan, M. Digitalization and green innovation of enterprises: Empirical evidence from China. *Front. Environ. Sci.* **2023**, *11*, 1120806. [CrossRef]
35. Rui, Z.; Lu, Y. Stakeholder pressure, corporate environmental ethics and green innovation. *Asian J. Technol. Innov.* **2020**, *29*, 70–86. [CrossRef]
36. Zhang, Z.; Duan, H.; Shan, S.; Liu, Q.; Geng, W. The Impact of Green Credit on the Green Innovation Level of Heavy-Polluting Enterprises-Evidence from China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 650. [CrossRef]
37. Han, S.; Zhang, Z.; Yang, S. Green Finance and Corporate Green Innovation: Based on China's Green Finance Reform and Innovation Pilot Policy. *J. Environ. Public Health* **2022**, *2022*, 1833377. [CrossRef]
38. Rao, H.; Chen, D.; Shen, F.; Shen, Y. Can Green Bonds Stimulate Green Innovation in Enterprises? Evidence from China. *Sustainability* **2022**, *14*, 15631. [CrossRef]
39. Wang, G. Research on the Influence of Environmental Regulation on Enterprise Green Innovation Performance. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *647*, 012179. [CrossRef]
40. Wang, C.; Chen, P.; Hao, Y.; Dagestani, A.A. Tax incentives and green innovation—The mediating role of financing constraints and the moderating role of subsidies. *Front. Environ. Sci.* **2022**, *10*, 1067534. [CrossRef]
41. Xue, L.; Zhang, Q.; Zhang, X.; Li, C. Can Digital Transformation Promote Green Technology Innovation? *Sustainability* **2022**, *14*, 7497. [CrossRef]
42. Wang, C.; Yan, G.; Ou, J. Does Digitization Promote Green Innovation? Evidence from China. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3893. [CrossRef] [PubMed]
43. Javed, M.; Wang, F.; Usman, M.; Ali Gull, A.; Uz Zaman, Q. Female CEOs and green innovation. *J. Bus. Res.* **2023**, *157*, 113515. [CrossRef]
44. Wang, P.; Bu, H.; Liu, F. Internal Control and Enterprise Green Innovation. *Energies* **2022**, *15*, 2193. [CrossRef]
45. Amore, M.D.; Bennedsen, M. Corporate governance and green innovation. *J. Environ. Econ. Manage.* **2016**, *75*, 54–72. [CrossRef]
46. Zhao, J.; Qu, J.; Wei, J.; Yin, H.; Xi, X. The effects of institutional investors on firms' green innovation. *J. Prod. Innovat. Manag.* **2022**, *40*, 195–230. [CrossRef]
47. Wang, F.; Sun, Z.; Feng, H. Can Media Attention Promote Green Innovation of Chinese Enterprises? Regulatory Effect of Environmental Regulation and Green Finance. *Sustainability* **2022**, *14*, 11091. [CrossRef]
48. Huang, Z.; Liao, G.; Li, Z. Loaning scale and government subsidy for promoting green innovation. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 148–156. [CrossRef]
49. Luo, L.; Yang, Y.; Luo, Y.; Liu, C. Export, subsidy and innovation: China's state-owned enterprises versus privately-owned enterprises. *Econ. Political Stud.* **2016**, *4*, 137–155. [CrossRef]
50. Li, J.; Lee, R.P.; Wan, J. Indirect effects of direct subsidies: An examination of signaling effects. *Ind. Innov.* **2019**, *27*, 1040–1061. [CrossRef]
51. Wang, H.; Xue, C.; Li, Y.; Zhao, P. The impact of government subsidies on the human resources and innovation output of makerspaces according to signal theory. *Int. J. Technol. Manag.* **2023**, *92*, 139–158. [CrossRef]
52. Lee, Y.N.; Walsh, J.P. Inventing while you work: Knowledge, non-R&D learning and innovation. *Res. Pol.* **2016**, *45*, 345–359. [CrossRef]
53. Xu, J.; Wang, X.; Liu, F. Government subsidies, R&D investment and innovation performance: Analysis from pharmaceutical sector in China. *Technol. Anal. Strateg. Manag.* **2020**, *33*, 535–553. [CrossRef]
54. Song, M.; Tao, W.; Shen, Z. The impact of digitalization on labor productivity evolution: Evidence from China. *J. Hosp. Tour. Technol.* **2022**, *ahead-of-print*. [CrossRef]
55. Guo, B.; Pérez-Castrillo, D.; Toldrà-Simats, A. Firms' innovation strategy under the shadow of analyst coverage. *J. Financ. Econ.* **2019**, *131*, 456–483. [CrossRef]
56. Gentry, R.J.; Shen, W. The impacts of performance relative to analyst forecasts and analyst coverage on firm R&D intensity. *Strategic. Manag. J.* **2013**, *34*, 121–130. [CrossRef]
57. Hao, J. Retail investor attention and corporate innovation in the big data era. *Int. Rev. Financ. Anal.* **2023**, *86*, 102486. [CrossRef]
58. Zhang, H.; Xiao, Y. Customer involvement in big data analytics and its impact on B2B innovation. *Ind. Market. Manag.* **2020**, *86*, 99–108. [CrossRef]
59. Zhong, R. Transparency and firm innovation. *J. Acc. Econ.* **2018**, *66*, 67–93. [CrossRef]

60. Han, M.; Lin, H.; Sun, D.; Wang, J.; Yuan, J. The Eco-Friendly Side of Analyst Coverage: The Case of Green Innovation. *IEEE Trans. Eng. Manag.* **2022**, *ahead-of-print*. [CrossRef]
61. He, L.; Gan, S.; Zhong, T. The impact of green credit policy on firms' green strategy choices: Green innovation or green-washing? *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 73307–73325. [CrossRef] [PubMed]
62. Liu, Y.; Xu, H.; Wang, X. Government subsidy, asymmetric information and green innovation. *Kybernetes* **2021**, *51*, 3681–3703. [CrossRef]
63. Fan, J.; Zhou, Y. Empirical Analysis of Financing Efficiency and Constraints Effects on the Green Innovation of Green Supply Chain Enterprises: A Case Study of China. *Sustainability* **2023**, *15*, 5300. [CrossRef]
64. Wu, Y.; Sun, H.; Sun, H.; Xie, C. Impact of Public Environmental Concerns on the Digital Transformation of Heavily Polluting Enterprises. *Int. J. Environ. Res. Public Health* **2022**, *20*, 203. [CrossRef] [PubMed]
65. Hadlock, C.J.; Pierce, J.R. New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index. *Rev. Financ. Stud.* **2010**, *23*, 1909–1940. [CrossRef]
66. Bertrand, M.; Mullainathan, S. Enjoying the Quiet Life? Corporate Governance and Managerial Preferences. *J. Polit. Econ.* **2003**, *111*, 1043–1075. [CrossRef]
67. Cao, Y.; Hu, X.; Lu, Y.; Su, J. Customer Concentration, Tax Collection Intensity, and Corporate Tax Avoidance. *Emerg. Mark. Financ. Trade* **2019**, *56*, 2563–2593. [CrossRef]
68. Tang, H. The Effect of ESG Performance on Corporate Innovation in China: The Mediating Role of Financial Constraints and Agency Cost. *Sustainability* **2022**, *14*, 3769. [CrossRef]
69. Peattie, K.; Ratnayaka, M. Responding to the green movement. *Ind. Market. Manag.* **1992**, *21*, 103–110. [CrossRef]
70. Xu, J.; Liu, F.; Shang, Y. R&D investment, ESG performance and green innovation performance: Evidence from China. *Kybernetes* **2021**, *50*, 737–756. [CrossRef]
71. Subramaniam, M.; Youndt, M. The Influence of Intellectual Capital on the Types of Innovative Capabilities. *Acad. Manag. J.* **2005**, *48*, 450–463. [CrossRef]
72. Wang, F.; Sun, Z. Does the Environmental Regulation Intensity and ESG Performance Have a Substitution Effect on the Impact of Enterprise Green Innovation: Evidence from China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8558. [CrossRef] [PubMed]