ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
Received: 24 March 2026; Accepted: 8 May 2026
https://doi.org/10.17559/TV-20260324003477
Original scientific paper

The Green Governance Effects of Data Assets: A Technology Innovation Perspective

Yanmin SHI*, Yan CHEN

Abstract: This study conducts an empirical investigation into the green governance effects of data assets, utilizing a sample of Shanghai and Shenzhen A-share listed companies spanning the period from 2007 to 2023. The findings indicate that data assets facilitate enterprises' green technology innovation. Data assets enhance green technology innovation by "financial support" and "technology-driven" effects. Further examination demonstrates that ESG performance and digital transformation amplify the positive impact of data assets on green technology innovation. The promotion effect of data assets is more pronounced in state-owned enterprises, heavily polluting firms, and high-tech industries. Extended analysis reveals that the influence of data assets predominantly enhances the quality; moreover, the green technology innovation contributes to advancing the green transformation efforts.

Keyword: data assets; green governance; productive factors; technology driven; technology innovation

1 INTRODUCTION

Currently, human society has transitioned into the era of the digital economy, and data volume is experiencing exponential growth. The economic benefits and societal value derived from data are becoming increasingly significant, rendering data capitalization an inevitable trend. The Interim Provisions on Accounting Treatment Related to Enterprise Data Assets operationalizes the recognition of data assets, marking a significant shift whereby data has transitioned from being regarded as a natural resource to being acknowledged as an economic asset. The data asset is regarded as data legally owned or controlled by an organization, which is recorded electronically or by other means. Such data can be quantified or traded and can generate direct or indirect economic and social benefits.

As China advances into a new phase of development, enhancing the quality of economic growth and achieving sustainable development have emerged as primary objectives for both economic and social progress. Green technology innovation plays the pivotal role in facilitating the economy's transition toward environmentally sustainable development. Distinct from general technological innovation, green technology innovation involves enterprises adopting novel technologies and methods to enhance resource utilization efficiency and reduce pollutant emissions, thereby improving environmental performance and achieving the dual objectives of sustainable economic development.

Regarding data assets, these are defined as resources that can guide enterprise production activities, assist managerial decision-making, and generate economic benefits for firms. Compared to other asset types, data assets are characterized by extensive sharing, non-consumptive use, exchangeability, dynamism, and diverse modes of value creation [1]. Theoretically, data assets can enhance an enterprise's cash flow and mitigate external financing constraints, thereby exerting a "financial support effect". Concurrently, data assets facilitate technological renewal and broaden innovation collaboration networks, constituting a "technology-driven effect". Consequently, data assets may exert a positive influence on enterprises' capacity for green technology innovation. Based on this, the present study empirically investigates how data assets influence corporate green technology innovation.

This study distinguishes itself from prior research through several key contributions. First, it empirically examines the green governance of data assets, thereby enriching the body of literature concerning the determinants of green innovation from the perspective of production factors. While existing studies have extensively investigated the drivers of green technology innovation focusing on external pressures and internal governance mechanisms, there remains a notable gap. Addressing this gap, the present paper systematically explores how data assets influence corporate green technology innovation, thus contributing to the literature on factors influencing green technology innovation.

Second, this research provides novel evidence demonstrating the green governance of data assets. Prior studies have predominantly approached data assets qualitatively, analyzing them at the macro level such as their characteristics, statistical treatment, and accounting considerations and at the micro level, including recognition, measurement, and reporting within firms. Empirical investigations have largely focused on the economic consequences of data assets in relation to small and medium-sized enterprises' financing constraints, human capital development, innovation investment, and equity capital costs. However, scant attention has been paid to green innovation. By empirically assessing the green governance of data assets, this paper offers incremental insights into the economic outcomes associated with data assets.

Third, the study broadens the theoretical framework employed to analyze the economic consequences of data assets. Existing literature has predominantly utilized theories such as signaling theory, resource-based theory, incomplete contract theory, financing constraint theory, credit rationing theory, and preferential financing theory to interpret these consequences. In contrast, this paper adopts a novel theoretical lens by integrating information asymmetry theory, technology innovation theory, and innovation diffusion theory.

2 LITERATURE REVIEW

Scholars have engaged extensively in theoretical investigations concerning the conceptualization, characteristics, rights delineation, valuation, and pricing of data assets. Building upon these foundations, a subset of researchers has undertaken empirical analyses to examine the economic implications of data assets.

From a theoretical standpoint, the conceptual understanding, defining attributes, valuation, and pricing mechanisms of data assets remain in preliminary exploratory phases. Data assets can be reutilized in production processes and thus possess economic value [2]. Data inherently functions as an information carrier and medium of expression, exhibiting properties of sharing and non-rivalry. The non-rivalrous nature of data engenders ambiguities in property rights, which subsequently diminishes the value of data [3]. To quantify data value, Bergemann and Bonatti (2019) [4] introduced an integrated pricing model for data assets within the context of advertisers acquiring consumer characteristic data.

Empirically, as research on the conceptualization, characteristics, valuation, and pricing of data assets has deepened, scholars have begun to empirically investigate the economic outcomes associated with data assets. Zheng and Zhang (2023) [5] argue that data assets can mitigate financing constraints, foster inter-firm R&D collaboration, and enhance firms' capacity to assimilate valuable information, thereby promoting innovation investment. Y. Li et al. (2024) [6] suggest that data assets positively influence corporate monitoring mechanisms, which in turn facilitate improvements in corporate ESG performance. Additionally, some researchers have examined the economic effects of data asset disclosure, asserting that data assets can provide forward-looking insights, enhance information transparency, and alleviate financing constraints [7].

The current body of literature has produced valuable theoretical advancements concerning the concepts, attributes, valuation, and pricing of data assets. Additionally, certain researchers have empirically examined the economic implications of data assets, particularly regarding innovation inputs and the efficiency of capital market pricing. Nevertheless, there remains a gap in the scholarly investigation of the economic effects of data assets.

3 THEORETICAL AND HYPOTHESIS

3.1 Data Assets and Enterprise Green Technology Innovation

Enterprises can enhance resource utilization efficiency and reduce pollutant emissions via green technology innovation. Such innovation supports the objective of sustainable economic development, representing a critical aspect of green governance practices and ecological responsiveness.

Nevertheless, green technology innovation presents challenges. Firstly, it accentuates the capital-intensive nature of enterprises, thereby intensifying their financing constraints. Given limited financial resources, green technology innovation may exert a "crowding out effect" on productive investments. Secondly, green technology innovation is characterized by extended development cycles, elevated risks, and externalities. Moreover, enterprises predominantly engage in green technology innovation by focusing on end-of-pipe pollutant emission reductions, resulting in relatively low levels of innovation content.

Data assets, serving as repositories of information regarding an enterprise's production, operational activities, and other relevant data, facilitate improved management and more informed decision-making, thereby enhancing economic returns and increasing internal cash flow [8]. Additionally, data assets enable upstream suppliers and downstream customers to gain a clearer understanding of the enterprise's production and business conditions, mitigating information asymmetry, which in turn fosters greater commercial credit availability. Furthermore, data assets assist external financial institutions in more accurately assessing firms' credit worthiness, easing firms' financing constraints [9]. The resultant increase in internal cash flow, enhanced commercial credit financing, and alleviation of external financing constraints collectively create favorable financial conditions that support corporate green technology innovation.

As an emerging factor of production, data can be deeply integrated with traditional production factors [10], thereby enhancing the synergy and interconnection among all production inputs. This integration facilitates the full realization of the potential inherent in various production factors within green technology innovation processes, ultimately promoting enterprises' engagement in green technology innovation. Moreover, data factors exhibit more pronounced externalities compared to conventional production factors. Consequently, data assets can give rise to digital industry clusters through the facilitation of information flow, technology transfer, capital movement, and material circulation, enabling digital synergistic linkages and positive feedback mechanisms [11]. Such dynamics create favorable conditions for enterprises to pursue green technology innovation. Additionally, data assets possess attributes of sharing, exchangeability, and non-rivalry [1], which enable the sharing and exchange of research and development outputs, technological knowledge, and other relevant information among firms. This not only enhances enterprises' awareness of advancements in cutting-edge green innovation but also provides critical data support for acquiring, analyzing, and learning from the successful experiences of other firms in this domain.

This study advances the following hypothesis:

Hypothesis 1: Data assets can promote enterprise green technology innovation.

3.2 Data Assets, Financial Support and Green Technology Innovation

According to the lens of the pecking order theory, given the substantial capital demands of green technology innovation, firms often require external financing, particularly in contexts where internal cash flow is generally limited. However, green technology innovation entails higher risks and complexities relative to conventional technological innovation, exacerbating information asymmetry between firms and financial institutions. Consequently, financial institutions impose higher risk premiums, thereby increasing the cost of external financing. External financing costs act as a significant barrier to firms' engagement in green technology innovation [12].

According to information asymmetry theory, banks possess less information than firms, which results in financing constraints for the latter. The advent of the digital economy modifies the applicability of the asset specialization theory within traditional transaction cost economics. Unlike intangible assets, data assets intrinsic to the digital economy are more generalized and can serve as credit collateral, thereby enhancing firms' credit financing capacity and mitigating financing constraints. Furthermore, data assets function as repositories of information regarding firms' production and operational activities, enabling banks to acquire enterprise information and assess financial risks more efficiently and at lower costs. This improved access to information allows financial institutions to better understand firms' operational status and more accurately evaluate their creditworthiness, thereby reducing information asymmetry. Consequently, financing constraints faced by firms are alleviated. Financing constraints constitute a critical impediment to firms' green technology innovation [13].

In the context of the macroeconomy, data has emerged as a novel production factor. Data assets are increasingly recognized as irreplaceable competitive resources. According to the resource-based view, the integration of data assets with conventional resources not only enhances operational efficiency, improves product and service quality, and reduces overall costs, but also provides more comprehensive, accurate, and predictive information to support strategic decision-making [14]. This integration facilitates access to high-quality investment opportunities, thereby improving profitability, economic returns, and internal cash flow. Internal cash flow is a significant determinant of firms' green technology innovation [15].

This study advances Hypothesis 2:

Hypothesis 2: Data assets promote enterprise green technology innovation through the financial support effect.

3.3 Data Assets, Technology Drivers and Enterprise Green Technology Innovation

Presently, many firms predominantly employ low-technology, end-of-pipe management approaches rather than advanced, source-oriented technological strategies for environmental governance, resulting in a substantial disparity between environmental management practices and intended objectives.

Drawing on the theory of technological innovation, such innovation entails the integration of resources through novel technologies to generate new economic value. Firstly, processes involving data collection, processing, analysis, and application impose elevated demands on the human capital of enterprises. As data assets are transformed into information and further internalized as knowledge, enterprises' technical capabilities in human capital are significantly enhanced, laying a talent foundation for green technological innovation. Secondly, data assets, being scarce, irreplaceable, and difficult to replicate, enable comprehensive, multi-dimensional, and full-chain transformations within enterprises. These transformations bolster competitive advantage and market positioning by improving management efficiency, reducing costs, increasing operational effectiveness, and upgrading business models to attract highly skilled and innovative personnel, thereby fostering green technology innovation.

According to innovation diffusion theory, knowledge spillover serves as a critical source of technological innovation. Data assets facilitate the formation of digital industry clusters through the synergistic flow of information, technology, capital, and materials, thereby enabling digital synergistic linkages and positive feedback mechanisms [11]. This process expands the network of innovation collaboration among enterprises. Innovation collaboration networks provide platforms for sharing knowledge and experiential resources related to green technology innovation. Cooperative relationships between enterprises with strong green technology innovation performance enable resource sharing and complementation, reduce research costs, improve R&D efficiency, and collectively advance green technology innovation.

This study proposes Hypothesis 3:

Hypothesis 3: Data assets promote enterprise green technology innovation through the technology-driven effect.

4 RESEARCH DESIGN

4.1 Samples and Data

This study utilizes A-share listed companies from the Shanghai and Shenzhen stock exchanges spanning the period from 2007 to 2023. The sample was refined through:

(1) exclusion of companies classified as ST, *ST, and PT;

(2) removal of firms operating within the financial sector; and

(3) elimination of observations with missing data for key variables.

The resulting dataset comprises 18,691 firm-year observations. To mitigate the impact of outliers on regression analyses, continuous variables were subjected to Winsorization at the 1st and 99th percentiles.

Data on asset information were sourced from annual financial reports. Information regarding green patent applications was obtained from the Chinese Research Data Service Platform (CNRDS), while all other variable data were extracted from the China Stock Market & Accounting Research Database (CSMAR).

4.2 Model

This paper establishes the model for regression analysis:

$$\begin{aligned} Green_{i,t} &= \alpha_0 + \alpha_1 Data_{i,t} + \alpha_2 Controls_{i,t} + \\ &+ \sum Industry + \sum Year + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where i denotes firm and t denotes year. *Green* denotes enterprise green technology innovation, and *Data* denotes enterprise data assets. *Controls* denote a set of control variables. This paper also controls for enterprise-fixed effects (*Industry*) and time-fixed effects (*Year*) to control the impact of unobservables that do not vary with enterprise and do not vary with time on the results. ε denotes the random error term.

4.3 Variable Definitions

4.3.1 Green Technology Innovation

Given that green patent applications directly reflect the output of enterprises' green technology innovation, they offer a more immediate and effective measure. Consequently, following B. Wang et al. (2024) [16], we use green patent applications as a proxy of corporate green technology innovation. Due to the right-skewed distribution of green patent application counts, the measure is transformed by adding one to the count and then applying the natural logarithm.

4.3.2 Data Asset

This study employs text analysis methodologies to quantify the frequency of data asset-related keywords in corporate financial disclosures. Following the approach of Fang and Liu (2024)[17], data assets are quantified by calculating the proportion of data asset keywords relative to the total word count in the annual reports, multiplied by 100.

4.3.3 Control Variables

Referring to X. Li et al. (2023) [18] and Xu and Lin (2025) [19], this paper controls variables that may affect enterprises' green technology innovation, such as enterprise size (*Size*), gearing ratio (*Level*), cash flow (*Flow*), growth capacity (*Grow*), profitability (*ROA*), enterprise age (*Age*), proportion of shares held by the first largest shareholder (*Top*1), size of the board of directors (*Board*), proportion of independent directors (*Indep*), and two-tier concurrent positions (*Dual*).

Specific variables are defined as shown in Tab. 1.

Table 1 Definitions of variables

Variables	Definition
Green	Natural logarithm of the number of green patent applications + 1
Data	The proportion of keyword word of data assets in enterprise annual report to the total word of the annual report, multiplied by 100
Size	Natural logarithm of total assets
Level	Total liabilities to total assets
Flow	Net cash flows from operating activities to total assets
Grow	Growth rate of operating revenue
ROA	Net profit to total assets
Age	Natural logarithm of the time of establishment of the enterprise + 1
Top1	Number of shares held by the largest shareholder as a percentage of total number of shares
Board	Natural logarithm of total number of board members
Indep	Number of independent directors as a percentage of the total number of board members
Dual	Dummy variable. Assigns a value of 1 if the chairman of the board of directors and the general manager are the same person, otherwise assigns a value of 0

5 RESULTS AND ANALYSIS
5.1 Descriptive Statistics

Tab. 2 presents the descriptive statistics for the variables analyzed in this study. The average number of green patent applications submitted by enterprises is 1.4989, suggesting that, on average, sample firms file approximately one to two green patent applications. The standard deviation is 5.6250, indicating considerable variability and suggesting that some firms have yet to submit any green patent applications. The median value is 0, which is substantially lower than the mean, implying that more than half of the enterprises have not engaged in green technology patenting. This finding highlights substantial potential for firms to enhance their green technology innovation efforts.

Regarding data assets, the mean value is 0.0386, exceeding the median value of 0. This suggests that, on average, keywords related to data assets constitute 0.0386% of the total vocabulary in firms' annual reports. Nonetheless, at least half of the enterprises have not disclosed information pertaining to data assets in their reports. The minimum value for data assets is 0, the maximum is 0.9017, and the standard deviation is 0.1239, reflecting variability in the extent of data asset disclosure across firms.

The descriptive statistics for the remaining variables align with those reported in prior studies.

Table 2 Descriptive statistics of variables

Variables	Obs	Mean	Std	Min	P50	Max
Green	18691	0.3280	0.7708	0.0000	0.0000	3.7612
Total	18691	1.4989	5.6250	0.0000	0.0000	42.0000
Data	18691	0.0386	0.1239	0.0000	0.0000	0.9017
Size	18691	22.6131	1.4197	19.9416	22.4484	26.7214
Level	18691	0.4929	0.1995	0.0722	0.4989	0.9204
Flow	18691	0.0466	0.0704	–0.1696	0.0463	0.2489
Grow	18691	0.1513	0.4130	–0.5884	0.0881	2.6359
ROA	18691	0.0308	0.0536	–0.1869	0.0295	0.1844
Age	18691	2.9444	0.3474	1.9459	2.9957	3.5835
Top1	18691	36.8645	15.4772	9.4434	35.2001	76.1285
Board	18691	2.1892	0.1915	1.6094	2.1972	2.7081
Indep	18691	0.3697	0.0532	0.3000	0.3333	0.5714
Dual	18691	0.1379	0.3448	0.0000	0.0000	1.0000

5.2 Benchmark Regression Analysis

Tab. 3 presents the findings from the benchmark regression analyses.

Table 3 Benchmark regression results

	Green	Green	Green	Green
	(1)	(2)	(3)	(4)
Data	0.7510*** (11.626)	0.7234*** (11.505)	0.3837*** (5.351)	0.3438*** (5.015)
Size		0.1426*** (23.853)		0.1389*** (22.298)
Level		–0.0557* (–1.849)		0.1020*** (3.142)
Flow		0.0162 (0.222)		0.0583 (0.812)
Grow		–0.0518*** (–5.170)		–0.0552*** (–5.419)
ROA		0.1689 (1.552)		0.4479*** (4.090)
Age		–0.1106*** (–6.737)		–0.0940*** (–4.282)
Top1		–0.0020*** (–4.989)		–0.0003 (–0.845)
Board		0.1106*** (2.950)		0.1536*** (4.235)
Indep		0.0525 (0.426)		0.1125 (0.963)
Dual		0.0391** (2.439)		0.0254* (1.657)
Cons	0.2991*** (53.156)	–2.7606*** (–17.950)	0.3132*** (54.916)	–2.9764*** (–17.322)
Industry	No	No	Yes	Yes
Year	No	No	Yes	Yes
Obs	18691	18691	18691	18691
Adj_R^2	0.0145	0.0781	0.1159	0.1802

Note: Obs denotes observed values. *, **, and *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. The numbers in parentheses are *t*-values. The regression results are based on robust standard errors. The same applies to the tables below.

Column (1) reports the results of a regression examining the relationship between data assets and firms' green technology innovation. Column (2) extends this analysis by incorporating control variables that may influence firms' green technology innovation. Column (3) displays the results of regressions that include only firm and year fixed effects. Column (4) combines both the control variables and firm and year fixed effects, with the coefficient for data assets continuing to be significantly positive at the 1% level.

In terms of economic significance, a one-unit increase in data assets leads to an 83.48% increase in a company's green technological innovation.

5.3 Endogeneity Test

5.3.1 Instrumental Variable Method (IV)

Green technological innovation in the corporate sector may, in turn, have an impact on data assets. This study employs the 2SLS estimation approach. Referring to M. Wang et al. (2024) [20], the average data assets of other firms within the same industry and year are utilized as an instrumental variable. This choice of instrument satisfies the relevance and exclusion criteria. Empirical tests support the validity of this instrument: the Kleibergen-Paaprk LM statistic yields a p-value of zero, and the Kleibergen-Paaprk Wald F statistic substantially exceeds the Stock-Yogo critical value of 16.38 at the 10% significance level.

Tab. 4 presents the 2SLS regression outcomes. Column (1) reports the first-stage regression, where the coefficient of the average data assets of peer firms is positive and statistically significant at the 1% level. Column (2) displays the second-stage results, revealing a positive and significant effect of data assets on green technology innovation at the 1% significance level.

5.3.2 Difference-in-Difference Method (DID)

This study applies the DID methodology to further address potential endogeneity concerns. In 2016, the Plan for the Development of the Big Data Industry (2016-2020) (hereafter referred to as the "Plan") was promulgated. This policy catalyzed rapid advancements in big data innovation across multiple sectors, providing a favorable context for firms to accumulate data assets. Consequently, the introduction of the Plan serves as an exogenous shock. This paper employs the difference-in-differences model (Eq. (2)) to conduct the regression analysis:

$$Green_{i,t} = \alpha_0 + \alpha_1 Treat_i * Post_t + \alpha_2 Controls_{i,t} + \sum Industry + \sum Year + \varepsilon_{i,t} \quad (2)$$

where $Treat$ denotes the policy dummy variable, grouped according to the median data assets of each firm in 2016, if the enterprise's industry's data assets in 2016 are less than the median data assets of all industries in that year, then it is the experimental group and $Treat$ takes the value of 1, and vice versa as the control group, $Treat$ takes the value of 0. $Post$ denotes a time dummy variable taking 1 for 2016 and subsequent years, and vice versa 0.

Column (3) of Tab. 4 presents the regression outcomes derived from DID. The coefficients associated with the interaction terms between the policy indicator and the temporal dummy on firms' green technology innovation is positive and significant. These regression results further demonstrate the robustness of the study's conclusions.

5.3.3 Explanatory Variables Lagged One Period

To further address the endogeneity issue arising from potential reverse causality, this study re-estimates the model by introducing a one-period lag for the explanatory variables. Column (4) of Tab. 4 presents the outcomes. The coefficient for data assets lagged by one period is positive and statistically significant.

Table 4 Results of endogeneity test 1

	Data	Green	Green	Green
	IV		DID	Explanatory variables lagged one period
	(1)	(2)	(3)	(4)
Data		0.7138*** (4.157)		
Data_iv	0.7520*** (21.742)			
Treat*Post			0.1247*** (5.145)	
L. Data				0.3235*** (4.221)
Size	0.0114*** (7.825)	0.0235*** (2.989)	0.1402*** (22.473)	0.1439*** (20.917)
Level	0.0080 (1.201)	0.0293 (0.881)	0.0926*** (2.849)	0.1175*** (3.265)
Flow	–0.0083 (–0.920)	–0.0629 (–1.146)	0.0363 (0.507)	0.0570 (0.712)
Grow	0.0015 (1.054)	–0.0095 (–1.367)	–0.0546*** (–5.352)	–0.0496*** (–4.330)
ROA	–0.0133 (–0.722)	0.0400 (0.476)	0.4466*** (4.075)	0.4297*** (3.530)
Age	–0.0188* (–1.872)	0.3689*** (5.091)	–0.0986*** (–4.497)	–0.0959*** (–3.838)
Top1	–0.0006*** (–6.207)	–0.0014*** (–2.696)	–0.0003 (–0.756)	–0.0002 (–0.378)
Board	0.0051 (0.919)	0.0461 (1.263)	0.1447*** (3.968)	0.1660*** (4.102)
Indep	–0.0073 (–0.436)	0.3590*** (3.123)	0.1099 (0.940)	0.0834 (0.660)
Dual	–0.0056*** (–2.659)	0.0004 (0.031)	0.0283* (1.836)	0.0278 (1.643)
Cons	0.0114*** (7.825)	0.0235*** (2.989)	–2.9763*** (–17.295)	–3.1098*** (–16.383)
Industry	Yes	Yes	Yes	Yes
Firm	Yes	Yes	No	No
Year	Yes	Yes	Yes	Yes
Kleibergen-Paaprk LM	240.771 (0.000)			
Kleibergen-Paaprk Wald F	472.732			
Obs	18649	18649	18691	16007
Adj_R^2	0.6898		0.1799	0.1810

Note: *p*-values in parentheses for Kleibergen-Paaprk LM statistic.

5.3.4 Heckman

The baseline regression analysis may be subject to endogeneity concerns arising from selection bias. To address this issue, this study employs the Heckman model. In the first stage, a Probit regression is conducted, where the dependent variable indicates whether a firm discloses data asset information in its annual report (coded as 1 if disclosed, 0 otherwise). The instrumental variable utilized is the average data asset value of other firms within the same industry and year. In the second stage, the inverse Mills ratio derived from the first stage is incorporated as a control variable in the primary model.

The results of the Heckman regression are presented in the first two columns of Tab. 5. Column (1) reports the Probit regression, revealing that the coefficient of the average data asset value of peer firms is positive and statistically significant at the 1% level, indicating its relevance in explaining disclosure decisions. Column (2) presents the second-stage regression results, where the coefficient of data assets on firms' green technology innovation is positive and significant at the 5% level.

5.3.5 Propensity Score Matching (PSM)

This study employs the PSM technique to reselect the research sample. The full sample is categorized into two groups based on whether firms disclose data assets in their annual reports: a disclosure group (treatment group) and a non-disclosure group (control group). Using the previously identified control variables as covariates, a 1:1 nearest-neighbor matching procedure with a caliper of 0.1 is applied.

Tab. 6 presents the results of the balance test. Prior to PSM, significant differences exist between the treatment and control groups across several covariates. After matching, these covariates show no significant differences between the two groups, except for the gearing ratio.

Column (3) of Tab. 5 presents the regression outcomes based on the sample obtained after applying propensity score matching. The coefficient of data assets in relation to firms' green technology innovation is positive and significant. These regression results further demonstrate the robustness of the study's conclusions.

Table 6 Results of the balance test

	Unmatched	Mean			% reduct	*t*-test	
	Matched	Treated	Control	% bias	|bias|	*t*	*p*>| *t* |
Size	U	23.021	22.431	41.8		26.74	0.000
	M	23.021	22.991	2.1	95.0	1.09	0.275
Level	U	0.484	0.497	–6.4		–4.01	0.000
	M	0.484	0.476	4.3	33.1	2.28	0.022
Flow	U	0.047	0.046	1.8		1.10	0.273
	M	0.047	0.048	–0.9	48.3	–0.5	0.612
Grow	U	0.138	0.157	–4.8		–3.01	0.003
	M	0.138	0.134	0.8	83.4	0.44	0.656
ROA	U	0.032	0.030	2.2		1.41	0.158
	M	0.032	0.032	–0.5	77.5	–0.27	0.785
Age	U	3.070	2.888	55.3		34.17	0.000
	M	3.070	3.068	0.7	98.7	0.40	0.686
Top1	U	36.575	36.994	–2.7		–1.71	0.087
	M	36.575	36.755	–1.2	56.9	–0.62	0.538
Board	U	2.183	2.192	–5.0		–3.19	0.001
	M	2.183	2.178	2.6	48.7	1.39	0.165
Indep	U	0.373	0.368	8.1		5.14	0.000
	M	0.373	0.373	–0.2	97.3	–0.12	0.908
Dual	U	0.146	0.134	3.4		2.14	0.032
	M	0.146	0.144	0.4	86.7	0.24	0.812

5.3.6 Firm and its Joint Fixed Effects with Year

This study incorporates fixed effects at the industry, year, and firm levels, and additionally controls for combined "industry-year" fixed effects.

Table 5 Results of endogeneity test 2

	Data_dum	Green	Green	Green	Green
	Heckman		PSM	Firm fixed effects	Joint fixed effects
	(1)	(2)	(3)	(4)	(5)
Data		0.2064** (2.433)	0.2366*** (2.861)	0.2714*** (4.546)	0.1344** (2.243)
Data_iv	1.1564*** (3.367)				
Size	0.1463*** (13.093)	0.1738*** (6.554)	0.0313** (2.156)	0.1490*** (14.862)	0.0292*** (3.856)
Level	–0.0439 (–0.561)	0.3833*** (4.872)	0.0001 (0.001)	0.0995* (1.831)	0.0328 (0.995)
Flow	–0.5175*** (–2.756)	0.3368 (1.627)	–0.0314 (–0.325)	–0.0595 (–0.461)	–0.0731 (–1.339)
Grow	0.0132 (0.453)	–0.1160*** (–4.898)	–0.0087 (–0.761)	–0.0735*** (–4.432)	–0.0086 (–1.257)
ROA	1.3870*** (5.000)	1.6994*** (5.396)	0.1097 (0.826)	0.7291*** (4.029)	0.0269 (0.324)
Age	–0.2001*** (–4.131)	–0.0387 (–0.735)	0.3831*** (2.661)	–0.1136*** (–3.174)	0.3674*** (5.066)
Top1	–0.0008 (–0.917)	–0.0023** (–2.538)	–0.0014 (–1.629)	–0.0003 (–0.438)	–0.0017*** (–3.195)
Board	0.0248 (0.336)	0.0491 (0.570)	0.0335 (0.512)	0.2246*** (3.731)	0.0480 (1.321)
Indep	–0.0861 (–0.351)	–0.2613 (–0.911)	0.3521 (1.643)	0.3119 (1.589)	0.3534*** (3.080)
Dual	0.0770** (2.242)	0.0361 (0.970)	0.0370* (1.784)	0.0305 (1.249)	–0.0025 (–0.206)
Imr		–0.0138 (–0.059)			
Cons	–4.8496*** (–15.210)	–3.5781*** (–4.886)		–3.3510*** (–11.876)	–1.6114*** (–5.565)
Firm	Yes	Yes	No	Yes	No
Year	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes
Industry* Year	No	No	No	No	Yes
Obs	17279	5019	7892	18653	18645
Adj_R^2		0.2097	0.1870	0.6671	0.6703
Pseudo R^2	0.2640				

The final two columns of Tab. 5 present the regression outcomes. Specifically, column (4) reports results from regressions controlling for both industry and year fixed effects, revealing that the coefficient on data assets with respect to firms' green technology innovation is positive and significant. Column (5) displays results that further control for firm fixed effects, year fixed effects, and joint "industry-year" fixed effects, where the coefficient on data assets remains positively significant.

5.4 Robustness Test

5.4.1 Substitution of Explained Variable Measures

This study employs an alternative measurement approach for the explanatory variables. Referring to J. Chen et al. (2024) [21], the ratio of green innovation patent applications to the total patent applications within the same year is utilized as a proxy variable ("Green_bl").

The results corresponding to this alternative measurement are presented in Column (1) of Tab. 7. The coefficients of data assets remain significantly positive, aligning with the previous findings.

5.4.2 Substitution of Explanatory Variable Measures

This study employs an alternative approach to address potential measurement errors in the explanatory variable and their effects on the regression outcomes. Referring to Cheng et al. (2024) [22], the natural logarithm of one plus the frequency count of data asset-related keywords disclosed in corporate annual reports is utilized as a proxy variable for data assets ("Data_ln"). Furthermore, this paper measures data assets ("Data_or") using a binary indicator reflecting whether data asset information is disclosed in the firm's annual report.

Columns (2) and (3) of Tab. 7 present the results. The explanatory variables also exhibit statistically significant relationships with firms' green technology innovation.

Table 7 Results of robustness tests

	Green_bl	Green	Green	Green	Green	Green
	Substitution of explained variable measures	Substitution of explanatory variable measures		Tobit	Excluding other interfering factors	
	(1)	(2)	(3)	(4)	(5)	(6)
Data	0.0264** (2.286)			0.8294*** (4.384)	0.3269*** (4.716)	0.4998*** (5.082)
Data_ln		0.0293*** (6.734)				
Data_or			0.0674*** (4.785)			
Size	0.0006 (0.315)	0.1360*** (21.886)	0.1379*** (22.080)	0.4781*** (22.445)	0.1383*** (21.029)	0.1313*** (19.590)
Level	0.0022 (0.267)	0.1011*** (3.116)	0.0989*** (3.048)	0.6691*** (4.329)	0.1142*** (3.297)	0.1662*** (4.652)
Flow	0.0013 (0.108)	0.0565 (0.787)	0.0471 (0.656)	–0.3793 (–0.952)	0.0740 (0.945)	–0.0523 (–0.642)
Grow	–0.0000 (–0.024)	–0.0550*** (–5.394)	–0.0544*** (–5.339)	–0.2471*** (–3.794)	–0.0530*** (–4.748)	–0.0515*** (–4.308)
ROA	–0.0201 (–0.967)	0.4215*** (3.864)	0.4262*** (3.907)	2.8891*** (5.246)	0.4807*** (3.986)	0.4574*** (3.835)
Age	0.0157 (1.033)	–0.0928*** (–4.231)	–0.0940*** (–4.287)	–0.6251*** (–6.966)	–0.0985*** (–4.190)	–0.1599*** (–6.417)
Top1	–0.0001 (–0.886)	–0.0004 (–0.983)	–0.0004 (–1.134)	–0.0018 (–1.090)	–0.0004 (–0.888)	–0.0012*** (–2.818)
Board	0.0103 (1.278)	0.1491*** (4.095)	0.1459*** (3.999)	0.6885*** (5.091)	0.1701*** (4.358)	0.0863** (2.216)
Indep	0.0802*** (3.316)	0.1145 (0.980)	0.1107 (0.946)	–0.0181 (–0.038)	0.1277 (1.017)	0.2087 (1.599)
Dual	0.0000 (0.002)	0.0236 (1.538)	0.0245 (1.592)	0.0409 (0.616)	0.0285* (1.725)	0.0090 (0.533)
Cons	–0.0658 (–1.029)	–2.9201*** (–16.982)	–2.9375*** (–17.060)	–14.6960*** (–23.927)	–2.9940*** (–16.361)	–2.4892*** (–13.500)
Year	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Obs	18653	18691	18691	18691	16682	14317
Adj_R^2	0.3694	0.1809	0.1793		0.1802	0.1718
Pseudo R^2				0.1353		

5.4.3 Tobit Model

The findings from the preceding descriptive statistics indicate that over half of the enterprises have not yet adopted green technology innovation. Given that the majority of firms have not filed for green technology patents, this study uses a Tobit model to address selection bias. Column (4) of Tab. 7 presents the results, revealing that the coefficient of data assets on firms' green technology innovation is positive and statistically significant.

5.4.4 Excluding Other Interfering Factors

Concerning the temporal dimension, the sample period utilized in the preceding benchmark regression spans from 2007 to 2023. Notably, the global financial crisis of 2008 and the stock market volatility experienced in 2015 exerted substantial effects on firms' production and operational activities, which could, in turn, influence their engagement in green technology innovation. Additionally, the adoption of new Accounting Standards in China in 2008 may have further impacted corporate green technology innovation endeavors.

Regarding the spatial dimension, the initial benchmark regression sample encompasses listed firms across all regions of China. However, the unique political and economic characteristics of China's municipalities, specifically Beijing, Tianjin, Shanghai, and Chongqing, may influence the regression results due to the distinct nature of firms located therein.

Tab. 7 presents the results. Column (5) reports the findings after removing the 2008 and 2015 samples. Column (6) displays the results following the exclusion of firms located in the four municipalities, with the coefficient of data assets on green technology innovation maintaining significance.

6 MECHANISM ANALYSIS

To empirically examine the mechanisms, this study develops Model (2):

$$M_{i,t} = \alpha_0 + \alpha_1 Data_{i,t} + \alpha_2 Controls_{i,t} + \\ + \sum Industry + \sum Year + \varepsilon_{i,t} \tag{2}$$

where, M represents mechanism variables, including external financing constraints (SA), internal cash flow (Fcf, Ica), high-skilled technical personnel ($R\&D\_N$, $R\&D\_R$), and external innovation cooperation ($Union\_N$, $Union\_R$). Referring to Y. Li et al. (2024) [6], this paper adopts SA index to measure enterprises' external financing constraints. For internal cash flow, this paper selects two indicators to measure free cash flow (Fcf) and internal cash flow adequacy (Ica). Specifically, referring toBeladi et al.(2021)[23], the ratio of the difference between operating cash flow and capital expenditures to total assets is used to measure free cash flow, and the ratio of operating cash flow to liquid liabilities is used to measure internal cash flow adequacy. Referring to Jiang et al. (2023) [24], this paper uses the number of research and development personnel ($R\&D\_N$) and the ratio of the ratio ($R\&D\_R$) to measure the firm's high-skilled technical personnel. Referring to Liu and Wang (2023) [25], this paper adopts the number ($Union\_N$) and percentage ($Union\_R$) of enterprises' joint applications for green patents to measure enterprises' external innovation cooperation.

Tab. 8 presents the regression analysis examining the mechanisms. Column (1) employs external financing constraints as the mediating variable, revealing a statistically significant negative coefficient for data assets. Columns (2) and (3) utilize internal cash flow as the mechanism variable, with regression coefficients for data assets on free cash flow being significantly positive at the 10% level, and on internal cash flow sufficiency at the 5% level.

Columns (4) and (5) focus on internal high-skilled technical personnel, where data assets exhibit significantly positive coefficients for both the number and proportion of R&D personnel. Columns (6) and (7) examine external innovation collaboration, with data assets showing significantly positive effects on both the number and proportion of joint green patent applications.

These findings demonstrate that data assets exert both financial support and technology-driven effects.

Table 8 Results of mechanism tests

	SA	Fcf	Ica	R&D_N	R&D_R	Union	Union_R
	financial support			technology–driven			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Data	–0.0389*** (–3.583)	0.0188* (1.740)	0.0614** (2.039)	7.4688*** (7.111)	0.1541*** (15.320)	0.0719** (2.002)	0.0294** (2.435)
Size	0.0545*** (33.722)	0.0042*** (7.190)	–0.0069*** (–4.167)	7.7956*** (26.399)	–0.0099*** (–9.484)	0.1019*** (23.695)	0.0097*** (10.658)
Level	–0.0638*** (–9.135)	–0.0005 (–0.109)	0.4052*** (24.065)	–6.5243*** (–7.141)	–0.0456*** (–5.733)	–0.0918*** (–4.864)	0.0187*** (3.251)
Flow	0.0519*** (3.460)	–0.9049*** (–100.412)	–2.8588*** (–90.323)	–3.9313 (–1.616)	–0.1841*** (–9.077)	0.0874** (2.080)	–0.0124 (–1.019)
Grow	–0.0063** (–2.539)	0.0124*** (7.289)	0.0116*** (2.883)	–1.2239*** (–3.222)	0.0045 (1.310)	–0.0311*** (–5.235)	–0.0054*** (–3.059)
Roa	–0.2850*** (–12.251)	0.0417*** (2.637)	–0.3680*** (–7.159)	–5.0925* (–1.656)	0.0895*** (3.189)	0.0768 (1.203)	0.0773*** (4.113)
Age	–0.7069*** (–165.004)	–0.0214*** (–9.371)	0.0166** (2.190)	–3.7738*** (–4.669)	–0.0217*** (–4.609)	–0.0320** (–2.193)	–0.0217*** (–5.113)
Top1	–0.0000 (–0.677)	–0.0002*** (–4.579)	–0.0001 (–0.907)	–0.0092 (–0.693)	–0.0003*** (–3.665)	0.0001 (0.264)	–0.0001 (–1.566)
Board	–0.0082 (–1.349)	0.0031 (0.912)	–0.0154 (–1.517)	–0.7076 (–0.568)	–0.0199*** (–2.920)	0.0173 (0.800)	0.0020 (0.349)
Indep	0.0960*** (4.539)	–0.0024 (–0.202)	0.0432 (1.357)	16.1936*** (3.778)	0.0227 (1.077)	0.1246* (1.694)	–0.0125 (–0.612)
Dual	0.0172*** (5.633)	0.0049*** (2.834)	–0.0033 (–0.647)	1.4219*** (2.951)	–0.0004 (–0.122)	0.0020 (0.224)	–0.0019 (–0.726)
Cons	–2.9541*** (–79.834)	0.0087 (0.580)	–0.1119** (–2.563)	–1.6e+02*** (–24.410)	0.4950*** (17.706)	–2.1216*** (–18.825)	–0.1342*** (–5.107)
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	17309	9265	9265	8537	8537	18691	18691
Adj_R^2	0.7998	0.5849	0.7031	0.3669	0.3937	0.1325	0.0422

7 FURTHER ANALYSIS
7.1 Moderating Effect
7.1.1 ESG Performance

Drawing upon stakeholder theory, ESG factors are recognized as influential in shaping stakeholder behavior. Firstly, ESG disclosures enhance information transparency in capital markets, thereby improving resource allocation efficiency and directing capital flows toward firms experiencing financing constraints. Secondly, ESG practices mitigate information asymmetry between firms and banks, facilitating access to bank credit, particularly green financing. Thirdly, ESG strengthens collaborative relationships with suppliers and customers, which can lead to enhanced commercial credit financing and alleviate firms' financing limitations. Lastly, superior ESG performance signals a firm's dedication to employee welfare and development, attracting highly skilled and technologically proficient talent. Collectively, these factors suggest that robust ESG performance can provide both financial and technical support, thereby mitigating firms' financing and technological constraints and amplifying the positive impact of data assets.

This study adopts the Huazheng ESG ratings as a measure of ESG performance. Specifically, the Huazheng ESG ratings, ranging from AAA to C, are numerically coded from 9 to 1 in descending order of performance.

The results presented in Tab. 9, Column (1), reveal that the coefficient of the interaction term between ESG performance and data assets is positive and significant. This finding indicates that ESG performance significantly strengthens the positive influence of data assets on enterprises' green technology innovation.

7.1.2 Digital Transformation

Within organizational contexts, not all data qualify as data assets; the development of data assets necessitates proactive management and effective governance of data. Digital transformation has the potential to augment the quantity of enterprise data assets while simultaneously enhancing capabilities for data management and control. Specifically, digital transformation improves the clarity and accessibility of information, thereby significantly reducing information asymmetry between enterprises and financial institutions, as well as between suppliers and customers. This reduction in information asymmetry helps to mitigate the financing constraints faced by firms.

Furthermore, digital transformation can improve the efficiency and utilization of resource allocation, enabling firms to achieve greater innovation outcomes within their existing innovation resource constraints, thereby strengthening their willingness and motivation to innovate. Accordingly, digital transformation can amplify the contribution of data assets.

Referring to Cheng et al. (2024) [22], this study measures digital transformation through the natural logarithm of the frequency of digital transformation-related keywords (plus one).

The results presented in Column (2) of Tab. 9 demonstrate a significant positive coefficient for the interaction term between digital transformation and data assets in relation to enterprises' green technology innovation.

Table 9 Results of moderating effects

	Green	Green
	ESG performance	Digital transformation
	(1)	(2)
ESG*Data	0.2612*** (3.547)	
Digi*Data		0.1587** (2.186)
Data	–0.8348*** (–2.612)	–0.6077* (–1.654)
ESG	0.0696*** (9.685)	
Digit		0.0382*** (6.880)
Size	0.1229*** (17.814)	0.1370*** (20.789)
Level	0.1926*** (5.272)	0.1166*** (3.394)
Flow	0.0424 (0.511)	0.0928 (1.198)
Grow	–0.0424*** (–3.369)	–0.0657*** (–5.852)
ROA	0.3765*** (3.016)	0.4755*** (4.099)
Age	–0.0961*** (–3.963)	–0.0855*** (–3.727)
Top1	–0.0005 (–1.110)	–0.0003 (–0.712)
Board	0.1561*** (3.788)	0.1532*** (4.007)
Indep	0.0070 (0.054)	0.0749 (0.612)
Dual	0.0309* (1.852)	0.0245 (1.513)
Cons	–2.8862*** (–15.341)	–3.0342*** (–16.884)
Industry	Yes	Yes
Year	Yes	Yes
Obs	16551	17646
Adj_R^2	0.1886	0.1823

7.2 Heterogeneity Test
7.2.1 Property Nature

Within the context of China's distinctive political and economic framework, substantial disparities exist in the business objectives and resource endowments among enterprises with differing ownership structures. Specifically, political backing facilitates SOEs' access to financing from banks, suppliers, and customers. Conversely, private firms encounter more stringent financing constraints due to ownership-based discrimination, which increases the difficulty and cost of securing bank financing. Furthermore, SOEs bear greater policy burdens and exhibit lower levels of innovation. In contrast, private enterprises operate under more competitive market incentives, with economic profit as their primary objective, and innovation serving as a critical mechanism for achieving profitability. The financial support derived from data assets can mitigate the financing constraints faced by private enterprises, thereby fostering their green technology innovation. Simultaneously, the technology-driven benefits of data assets may enhance the innovation capacity of SOEs, promoting their green technological advancements.

This study categorizes the sample into two groups based on the nature of the beneficial controller's equity: state-owned enterprises and private enterprises.

The first two columns of Tab. 10 present the regression outcomes. Column (1) reports the results for SOEs, revealing a significant positive coefficient for data assets. Column (2) displays the results for private enterprises, where the coefficient for data assets is positive but statistically insignificant. These findings imply that data assets play a more pivotal role in advancing green technology innovation in SOEs compared to private enterprises.

7.2.2 Heavily Polluting Enterprises

Enterprises with varying pollution intensities exhibit distinct demands for capital and technology. Specifically, in the context of achieving the "dual carbon" objectives, heavily polluting enterprises encounter more stringent financing constraints. Data assets can alleviate their financing limitations. Furthermore, heavily polluting firms are compelled not only to enhance productivity but also to assume greater environmental responsibilities, which incentivizes them to pursue technological advancements.

The study segregates the overall sample into two groups: heavily polluting enterprises and non-heavily polluting enterprises. The results, presented in the middle two columns of Tab. 10, reveal the impact of data assets on green technology innovation across these groups. Column (3) reports the regression outcomes for heavily polluting enterprises, where the coefficient for data assets is positive and statistically significant. Similarly, column (4) displays results for non-heavily polluting enterprises, with the data assets coefficient also being positive and significant. A formal test comparing the coefficients between the two groups confirms a significant difference. Therefore, relative to non-heavily polluting enterprises, data assets play a more substantial role in advancing green technology innovation among heavily polluting firms.

7.2.3 High-Tech Industries

The preceding analysis was confined to the micro-level of individual firms, neglecting the influence of industry-specific characteristics. Such an omission may constrain the breadth and applicability of the findings. High-technology industries typically entail substantial capital requirements and extended investment return horizons, rendering high-tech firms more susceptible to financing constraints. Although national policies and formal regulatory frameworks have alleviated some of these financial barriers, data assets - considered a form of informal regulation - can synergistically complement formal mechanisms to further mitigate financing challenges faced by high-tech enterprises.

Table 10 Results of heterogeneity analysis

	Green	Green	Green	Green	Green	Green
	Property Nature		Heavily Polluting		High-tech industries	
	(1)	(2)	(3)	(4)	(5)	(6)
Data	0.3449*** (3.966)	0.0294 (0.321)	0.7515*** (3.068)	0.3251*** (4.580)	0.4198*** (4.830)	–0.2261*** (–3.023)
Size	0.1410*** (19.723)	0.0945*** (6.552)	0.1335*** (9.553)	0.1367*** (19.251)	0.2137*** (22.319)	0.0548*** (7.524)
Level	0.0445 (1.129)	0.2292*** (3.875)	0.4500*** (5.650)	–0.0178 (–0.509)	0.2131*** (4.352)	–0.0931** (–2.268)
Flow	0.1736* (1.946)	–0.1763 (–1.469)	0.2056 (0.926)	0.0055 (0.074)	0.0059 (0.049)	–0.0770 (–1.088)
Grow	–0.0582*** (–4.530)	–0.0202 (–1.100)	–0.1226*** (–4.022)	–0.0385*** (–3.721)	–0.0968*** (–5.186)	–0.0232** (–2.367)
ROA	0.3427** (2.474)	0.7287*** (3.922)	0.5011* (1.819)	0.3997*** (3.341)	0.5761*** (3.743)	0.0381 (0.258)
Age	–0.0264 (–1.060)	–0.2439*** (–5.409)	–0.1628*** (–2.716)	–0.0963*** (–4.100)	–0.1687*** (–5.138)	–0.0664** (–2.365)
Top1	0.0004 (0.971)	–0.0019** (–2.178)	0.0013 (1.412)	–0.0007* (–1.668)	–0.0024*** (–3.892)	0.0019*** (4.383)
Board	0.1409*** (3.413)	0.2481*** (2.863)	0.3048*** (3.608)	0.1002** (2.529)	0.2014*** (3.547)	0.0693* (1.734)
Indep	0.1532 (1.169)	–0.1149 (–0.466)	0.9913*** (3.512)	–0.2210* (–1.858)	0.1423 (0.809)	–0.0427 (–0.332)
Dual	0.0717*** (3.484)	–0.0795*** (–3.538)	0.1053*** (2.591)	0.0049 (0.310)	0.0275 (1.266)	0.0216 (1.237)
Cons	–3.2228*** (–16.841)	–1.6423*** (–3.506)	–3.4538*** (–8.891)	–2.6350*** (–13.698)	–4.3761*** (–17.950)	–1.0442*** (–4.664)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Difference in coefficients between groups ($b_0 - b_1$)	–0.3155** （0.012）		–0.4264*** （0.000）		–0.6459*** （0.000）	
Obs	14582	3209	4510	14181	10322	7367
Adj_R^2	0.1862	0.1554	0.1940	0.1706	0.1842	0.1656

Note: b_0 presents the regression coefficients of data assets on green technology innovation of private enterprises, non-heavily polluted enterprises, and enterprises in non-high-tech industries, respectively. b_1 represents the regression coefficients of data assets on green technology innovation of state-owned enterprises, heavily polluted enterprises, and enterprises in high-tech industries, respectively. *p*-values are in parentheses for differences in coefficients between groups.

Drawing upon the classification framework proposed by Cui and Mak (2002) [26], this research designates the manufacturing sector and the information transmission, software, and information technology services sector as high-tech industries.

The final two columns of Tab. 10 present the results. Specifically, column (5) reports the results for enterprises within high-technology industries, where the coefficient for data assets is positive and significant. Column (6) displays the regression outcomes for enterprises in non-high-technology industries, revealing a negative and statistically significant coefficient for data assets. Consequently, data assets exert a more pronounced and favorable influence on promoting green technology innovation in high-technology industry enterprises relative to their non-high-technology counterparts.

7.3 Extensive Research

7.3.1 Quantity and Quality of Green Technology Innovation

It is important to distinguish between the "quality" and "quantity" of green technology innovation. Although the volume of patent applications in China has surged substantially, there has been no commensurate improvement in patent quality, with the phenomenon of a "patent bubble" being particularly pronounced [27]. The financial support function of data assets alleviates external financing constraints faced by firms, thereby enhancing their internal cash flow. However, this effect is likely more evident in the increased number of green technology innovations. Conversely, the technology-driven role of data assets attracts highly skilled personnel and strengthens external innovation collaborations, which may more significantly enhance the quality of corporate green technology innovation.

The number of green utility model patent applications filed by firms is employed as a proxy for the quantity of green technology innovation, while the number of green invention patent applications serves as a measure of innovation quality.

The results are presented in the first two columns of Tab. 11. In column (1), where the dependent variable represents the quality of green technology innovation, the coefficient for data assets is positive and significant. In contrast, column (2) shows that when the dependent variable is the quantity of green technology innovation, the coefficient for data assets is negative and significant. Taken together with the findings from the baseline regression, these results suggest that data assets more substantially enhance innovation quality, thereby overall promoting green technology innovation.

Table 11 Results of extensive research

	Green_invent	Green_new	Trans_green	Trans_green	Trans_green
	Quantity and quality of green technology innovation		Green transformation		
	(1)	(2)	(3)	(4)	(5)
Data	0.4180*** (6.503)	−0.0832*** (−2.750)			
Green			0.0661*** (8.010)		
Green_invent				0.0612*** (6.240)	
Green_new					0.1018*** (8.212)
Size	0.1163*** (21.366)	0.0776*** (18.764)	0.1411*** (26.706)	0.1431*** (27.081)	0.1426*** (27.258)
Level	0.0714*** (2.609)	0.0571*** (2.644)	0.1823*** (5.222)	0.1844*** (5.282)	0.1825*** (5.229)
Flow	0.0683 (1.151)	0.0023 (0.047)	0.1010 (1.169)	0.1010 (1.168)	0.1031 (1.193)
Grow	−0.0452*** (−5.470)	−0.0293*** (−4.342)	0.0238* (1.777)	0.0229* (1.709)	0.0232* (1.734)
ROA	0.3838*** (4.199)	0.1317* (1.840)	0.0681 (0.557)	0.0735 (0.601)	0.0835 (0.683)
Age	−0.0567*** (−3.149)	−0.0564*** (−3.811)	−0.2553*** (−11.741)	−0.2581*** (−11.865)	−0.2561*** (−11.770)
Top1	−0.0005 (−1.435)	0.0004 (1.539)	0.0007* (1.807)	0.0007* (1.813)	0.0006 (1.616)
Board	0.1304*** (4.190)	0.1097*** (4.530)	−0.0990*** (−3.051)	−0.0970*** (−2.991)	−0.1003*** (−3.093)
Indep	0.1598 (1.606)	0.0725 (0.918)	−0.3811*** (−3.356)	−0.3837*** (−3.374)	−0.3802*** (−3.352)
Dual	0.0275** (2.104)	0.0205** (2.003)	−0.0350** (−2.215)	−0.0349** (−2.212)	−0.0352** (−2.227)
Cons	−2.6192*** (−17.489)	−1.7188*** (−14.755)	−0.4159*** (−2.952)	−0.4512*** (−3.193)	−0.4391*** (−3.139)
Industry	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Obs	18691	18691	19541	19541	19541
Adj_R^2	0.1639	0.1398	0.3582	0.3572	0.3584

7.3.2 Enterprise Green Transformation

Enterprises advance existing pollution control technologies through green technology innovation, thereby increasing resource use efficiency and decreasing pollutant discharge. Such efforts support the dual objectives of sustainable economic development. The concept of green transformation aligns closely with these goals.

Adopting a measurement approach inspired by Y. Chen et al. (2023) [28], this study utilizes the natural logarithm of the frequency of green transformation-related keywords plus one as an indicator of enterprise green transformation

("Trans green").

The results are presented in the final three columns of Tab. 11. Column (3) reports that green technology innovation exhibits a significant positive coefficient. Column (4) shows that the quality of green technology innovation similarly has a significantly positive effect. Likewise, column (5) demonstrates that the quantity of green technology innovation positively and significantly influences enterprise green transformation at the 1% level. These findings collectively underscore the importance of both the quality and quantity of green technology innovation in advancing the green transformation of enterprises.

8 CONCLUSIONS AND IMPLICATIONS

8.1 Conclusions

This study empirically investigates how data assets influence corporate green technology innovation. The key findings are as follows:

Data assets significantly enhance enterprises' green technology innovation. Mechanism analysis reveals that data assets alleviate external financing constraints, increase internal cash flow, attract highly skilled technical personnel, and strengthen external innovation collaborations, thereby fostering green technology innovation.

ESG performance and digital transformation serve as effective moderators. The promotive effect of data assets is more pronounced in state-owned enterprises, heavily polluting firms, and high-tech industry enterprises.

The enhancement of green technology innovation is primarily reflected in the quality dimension. Improvements in green technology innovation are conducive to enterprises' green transformation efforts.

8.2 Implications

This study derives the following implications:

(1) Governmental bodies and relevant authorities should advance the implementation of China's strategic framework for data elements by enhancing top-level design and fostering the integration of institutional support with technological empowerment. Fiscal and tax policies should be strengthened, including the provision of substantial tax deductions or incentives for enterprises pursuing digitalization strategies and engaging in data asset financing. Policymakers should also provide targeted incentives to private enterprises. For industries that are not major sources of pollution, the lack of accumulated data assets could be addressed through greater openness of government data resources. For non-high-tech enterprises, "digital and green transformation" assessments and training programs should be conducted to lower barriers to technology adoption.

(2) Enterprises and other market participants should expedite their digital transformation processes, promote digital development, and establish comprehensive enterprise data asset management platforms. Concurrently, organizations should actively recruit and cultivate digital talent, emphasizing the enhancement of employees' data literacy to build a robust human resource foundation that supports the pursuit of green technology innovation.

8.3 Limitations and Prospects

Two primary limitations remain. First, the measurement of data assets relies on a text analysis methodology, which, although widely adopted and possessing a certain level of validity and accuracy, is not without imperfections and leaves room for refinement.

Second, the investigation is confined to the context of China. Subsequent studies could examine the economic implications of data assets across various economies, countries, and regions.

Lastly, the measurement of green technological innovation which relies on the number of green patent applications appears somewhat limited due to data availability. In the future, the methodology for measuring green technological innovation could be refined by considering the various technological fields in which green patents are filed or by analyzing the citation rates of such patents.

Data Availability Statement

Restrictions apply to the availability of the data, which were used under license for this study. Data are available at https://www.sse.com.cn/disclosure/listedinfo/regular/, https://www.szse.cn/disclosure/ listed/fixed/index.html, https://www.cnrds.com and https://data.csmar.com with the permission of the Shanghai Stock Exchange and the Shenzhen Stock Exchange, CNRDS and CSMAR.

9 REFERENCES

[1] Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, *54*(2), 442-492. https://doi.org/10.1257/jel.54.2.442

[2] Rassier, D. G., Kornfeld, R. J., & Strassner, E. H. (2019). Treatment of Data in National Accounts.

[3] Rubinfeld, D. L. & Gal, M. S. (2017). Access Barriers to Big Data. *Arizona Law Review*, *59*(2), 339-382.

[4] Bergemann, D., & Bonatti, A. (2019). Markets for Information: An Introduction. *Annual Review of Economics*, *11*, 85-107. https://doi.org/10.1146/annurev-economics-080315-015439

[5] Zheng, Y. & Zhang, Q. (2023). Digital transformation, corporate social responsibility and green technology innovation- based on empirical evidence of listed companies in China. *Journal of Cleaner Production*, *424*, 138805. https://doi.org/10.1016/j.jclepro.2023.138805

[6] Li, Y., Zhang, Y., & Geng, L. (2024). Digital finance, financing constraints and supply chain resilience. *International Review of Economics & Finance*, *96*, 103545. https://doi.org/10.1016/j.iref.2024.103545

[7] Sun, X. & Du, Z. (2024). Enhancing capital market efficiency: The role of data assets disclosure in reducing stock price synchronicity. *International Review of Economics & Finance*, *94*, 103351. https://doi.org/10.1016/j.iref.2024.05.030

[8] Veldkamp, L. & Chung, C. (2024). Data and the Aggregate Economy. *Journal of Economic Literature*, *62*(2), 458-484. https://doi.org/10.1257/jel.20221580

[9] Carrière-Swallow, Y. & Haksar, V. (2019). The Economics and Implications of Data: An Integrated Perspective. https://doi.org/10.5089/9781513511436.087

[10] Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The quarterly journal of economics*, *135*(2), 645-709. https://doi.org/10.1093/qje/qjaa004

[11] Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big Data and Firm Dynamics. *AEA Papers and Proceedings*, *109*, 38-42. https://doi.org/10.1257/pandp.20191001

[12] Xiang, X., Liu, C., & Yang, M. (2022). Who is financing corporate green innovation? *International Review of Economics & Finance*, *78*, 321-337. https://doi.org/10.1016/j.iref.2021.12.011
[13] Lv, C., Shao, C., & Lee, C.-C. (2021). Green technology innovation and financial development: Do environmental regulation and innovation output matter? *Energy Economics*, *98*, 105237. https://doi.org/10.1016/j.eneco.2021.105237
[14] Hannila, H., Silvola, R., Harkonen, J., & Haapasalo, H. (2022). Data-driven Begins with DATA; Potential of Data Assets. *Journal of Computer Information Systems*, *62*(1), 29-38. https://doi.org/10.1080/08874417.2019.1683782
[15] Ghosh, D. & Dutta, M. (2022). Environmental behaviour under credit constraints - Evidence from panel of Indian manufacturing firms. *Structural Change and Economic Dynamics*, *63*, 490-500. https://doi.org/10.1016/j.strueco.2022.07.004
[16] Wang, B., Khan, I., Ge, C., & Naz, H. (2024). Digital transformation of enterprises promotes green technology innovation – The regulated mediation model. *Technological Forecasting and Social Change*, *209*, 123812. https://doi.org/10.1016/j.techfore.2024.123812
[17] Fang, X. & Liu, M. (2024). How does the digital transformation drive digital technology innovation of enterprises? Evidence from enterprise's digital patents. *Technological Forecasting and Social Change*, *204*, 123428. https://doi.org/10.1016/j.techfore.2024.123428
[18] Li, X., Guo, F., Xu, Q., Wang, S., & Huang, H. (2023). Strategic or substantive innovation? The effect of government environmental punishment on enterprise green technology innovation. *Sustainable Development*, *31*(5), 3365-3386. https://doi.org/10.1002/sd.2590
[19] Xu, B. & Lin, B. (2025). Green credit and green technology innovation: Impact mechanism and nonlinear effect test. *Environmental Impact Assessment Review*, *110*, 107652. https://doi.org/10.1016/j.eiar.2024.107652
[20] Wang, M., Li, Y., Wang, Z., Shi, Y., & Zhou, J. (2024). The synergy impact of external environmental pressures and corporate environmental commitment on innovations in green technology. *Corporate Social Responsibility and Environmental Management*, *31*(2), 854-878. https://doi.org/10.1002/csr.2604
[21] Chen, J., Wu, L., Hao, L., Yu, X., & Streimikiene, D. (2024). Does the import of green products encourage green technology innovation? Empirical evidence from China. *Technological Forecasting and Social Change*, *200*, 123137. https://doi.org/10.1016/j.techfore.2023.123137
[22] Cheng, W., Li, C., & Zhao, T. (2024). The stages of enterprise digital transformation and its impact on internal control: Evidence from China. *International Review of Financial Analysis*, *92*, 103079. https://doi.org/10.1016/j.irfa.2024.103079
[23] Beladi, H., Deng, J., & Hu, M. (2021). Cash flow uncertainty, financial constraints and R&D investment. *International Review of Financial Analysis*, *76*, 101785. https://doi.org/10.1016/j.irfa.2021.101785
[24] Jiang, C., Gao, Y., Li, S. H., Luo, L. H., & Zhu, L. N. (2023). Characteristics and factors for the innovation performance of New R&D Institutes at start-up stages: an exploratory study from China. *R & D MANAGEMENT*, *53*(3), 408-433. https://doi.org/10.1111/radm.12579
[25] Liu, S. & Wang, Y. (2023). Green innovation effect of pilot zones for green finance reform: Evidence of quasi natural experiment. *Technological Forecasting and Social Change*, *186*, 122079. https://doi.org/10.1016/j.techfore.2022.122079
[26] Cui, H. & Mak, Y. T. (2002). The relationship between managerial ownership and firm performance in high R&D firms. *Journal of Corporate Finance*, *8*(4), 313-336. https://doi.org/10.1016/S0929-1199(01)00047-5
[27] Wang, H., Qi, S., Zhou, C., Zhou, J., & Huang, X. (2022). Green credit policy, government behavior and green innovation quality of enterprises. *Journal of Cleaner Production*, *331*, 129834. https://doi.org/10.1016/j.jclepro.2021.129834
[28] Chen, Y., Ma, X., Ma, X., Shen, M., & Chen, J. (2023). Does green transformation trigger green premiums? Evidence from Chinese listed manufacturing firms. *Journal of Cleaner Production*, *407*, 136858. https://doi.org/10.1016/j.jclepro.2023.136858

Contact information:

Yanmin SHI
(Corresponding author)
School of Accounting,
Shandong University of Finance and Economics,
Jinan, 250014, China
E-mail: shiyanmin1226@163.com

Yan CHEN
School of Accounting,
Shandong University of Finance and Economics,
Jinan, 250014, China