Data Elements and New Quality Productivity-Quasi-natural Experiments Based on a National-level Comprehensive Big Data Pilot Area

Yue Xue and Yue Wang*

School of Economics and Management, Qinghai Minzu University, Xining, China
Emai: 48320970@qq.com (Y.X.); 790032856@qq.com (Y.W.)
*Corresponding author

Manuscript received July 19, 2024; revised August 2, 2024; accepted August 21, 2024; published December 30, 2024.

***Abstract*—This paper uses the pilot of a national-level big data comprehensive experimental zone as a quasi-natural experiment to explore the impact of data elements on the new-quality productivity of micro-enterprises based on the data of A-share listed companies in Shanghai and Shenzhen during the period of 2011–2022. It is found that data elements significantly promote the new quality productivity of enterprises, and the conclusion still holds after a series of robustness tests. Heterogeneity analysis shows that data elements enhance the new productivity of high-tech firms more significantly than that of non-high-tech firms; data elements promote the new productivity of firms in the eastern region, but have no significant effect on the new productivity of firms in the central and western regions. Mechanism analysis shows that enterprise resource allocation efficiency and digital transformation level play the role of path, and enterprise internal control level and scientific and technological innovation level play a moderating role in the correlation between data elements and enterprise new quality productivity. This paper provides empirical evidence for the further improvement of enterprise new quality productivity.**

***Keywords*—data elements, new quality productivity, double difference models, triple difference models**

I. Introduction

In 2024, General Secretary Xi Jinping, while presiding over the eleventh collective study of the Political Bureau of the twentieth Central Committee, pointed out that "the development of new-quality productive forces is an intrinsic requirement and an important point of focus in promoting high-quality development, and it is necessary to continue to do a good job of innovation, and to promote the accelerated development of new-quality productive forces." With the reshaping of the production process by scientific and technological progress, the data factor has formally become a basic production factor alongside land, labor, capital and technology, and its new characteristics such as non-competitiveness, virtual replicability, and zero marginal cost enable the data factor to break through the constraints of resource scarcity and expand the connotations and boundaries of the traditional economic theories, and the data, as a new type of production object, has become a key element of the new-quality productivity (Wang *et al.*, 2024). In January 2024, seventeen departments issued a circular on the "Three-Year Action Plan of "Data Factor x" (2024–2026)," aiming to give full play to the multiplier effect of data factors and empower economic and social development. So, can data elements drive the development of new quality productivity of microenterprises? Is there any difference in the impact of data elements on the new productivity of different types of enterprises and in different regions? What is the path for data elements to drive the new quality productivity of enterprises? This paper attempts to analyze and explore these questions.

II. Literature Review

A. Studies Related to Data Elements and New Quality Productivity of Enterprises

From a theoretical perspective, at this stage, data is already a production factor, with qualitative characteristics such as amplification and superposition, dependence and multiplication, composite generalization, dynamic precision, cumulative spillover effect, scale payoff increment, Marshallian externalities, etc., which has been integrated into all elements of the labor process and the whole elements of the new-quality productive forces, and is able to nourish new types of laborers, generate new types of labor materials, breed new types of labor objects, and empower the optimal combination of the whole elements. The optimization of the combination of data elements to drive the innovation and development of the new quality productivity also constitutes the 4.0 stage of productivity development (Li, 2024). The data factor drives the innovative development of new quality productivity and constitutes the 4.0 stage of productivity development. At the same time, the marketization process of data elements has given rise to many new forms and modes of business, which has also laid a solid industrial foundation for the formation and development of new quality productivity (Feng *et al.*, 2024). The combination of data, arithmetic power and algorithms firstly causes the decision-making revolution of productivity, and secondly causes the tool revolution, labor revolution, production factor revolution and technology-economic paradigm revolution of productivity, which further promotes the digital development of new quality productivity (Ren *et al.*, 2024). From a practical perspective, based on panel data of 230 prefecture-level cities across China from 2011 to 2021, constructed a new quality productivity indicator system at the city level with the three dimensions of scientific and technological productivity, green productivity and digital productivity, regarded the pilot policy of the national-level big data comprehensive experimental zone as a quasi-natural experiment, and constructed a multi-period double-difference model to empirically test the effects of the pilot policy, and found that the national-level big data comprehensive. It is found that the national-level big data comprehensive experimental zone significantly empowers the new quality productivity enhancement (Zhao *et al.*, 2024).

doi:10.18178/joebm.2024.12.4.838

B. Relevant Studies on the Mechanisms of the Impact of Data Elements on the New Quality Productivity of Enterprises

1) Data elements, resource allocation efficiency and new quality productivity of enterprises

Neoclassical economic theory believes that government intervention can correct the market failure problem in the process of economic operation. Government intervention can play the role of macro-control of resources, promote the rational allocation of resources, reduce resource mismatch, and improve total factor productivity (Zeng *et al.*, 2024). Among the existing studies, some scholars take enterprise resource allocation as an entry point, utilize the "Internet + government service" pilot city policy announced in 2016 as a quasi-natural experiment, and adopt the successive double-difference method to conclude that digital government can enhance enterprise resource allocation efficiency through cost-saving effects, alleviating financing constraints and innovation incentives (Liao *et al.*, 2024).

2) Data elements, level of digital transformation and new quality productivity of the enterprise

The new quality productivity has high-tech, high-efficiency and high-quality characteristics, which is in line with the new development concept of advanced productivity qualities, and the process of enterprise digital transformation can further accelerate the development of new quality productivity (Jin *et al.*, 2024). Some scholars point out that in order to realize the blueprint of "Digital China" to empower new quality productivity, it is urgent to precisely release the dividend effect of digital transformation in the future (Chen *et al.*, 2024). Some scholars based on panel data from 30 provinces (autonomous regions and municipalities) in China from 2011 to 2021, found that the digital transformation of commercial banks can effectively channel new productivity through the market transformation of technological achievements and capital agglomeration (Lin *et al.*, 2024). These studies suggest that digital transformation plays a crucial role in promoting new quality productivity of enterprises, and by improving the level of digital transformation, it promotes the innovation and development of enterprises.

3)Data elements, level of internal control and new quality productivity of enterprises

Internal control is a series of dynamic means implemented by a company to achieve a certain purpose, the essence of which is part of the production and operation, but also continuously monitor the process of production and operation to ensure that the production and operation is carried out in the established direction (Gu *et al.*, 2020). On the one hand, Some scholars based on the quasi-natural experiment of the "fusion of two standards" pilot, used the data of China's A-share listed manufacturing companies, and found that the quality of internal control was significantly improved after the digital transformation of enterprises using the double-difference model test (Zhang *et al.*, 2022). On the other hand, both high-quality internal control and social auditing can effectively increase the total factor productivity of enterprises and exert substitution effects (Guo *et al.*, 2020). These studies suggest that the level of internal control plays an important moderating role in the process of new quality productivity enhancement of enterprises by coordinating the activities of various departments and improving the efficiency of decision implementation.

4)Data elements, STI levels and firms' new quality productivity

Enterprises use data elements to tap into the deep needs of consumers and gain high-level insights from the data to positively influence their innovation behavior. According to the theory of factor allocation, data factor empowerment can be regarded as an extension of digital technology's participation in factor allocation and the creation of value, which largely extends to enterprise innovation activities (Ma *et al.*, 2024). This value will be largely extended to enterprise innovation activities. Science and technology innovation forms new quality productivity by acting on the components of productivity, and promotes the formation of new quality productivity at different levels of micro, meso and macro; breakthrough technological innovation plays an important role in the formation of new quality productivity (Du *et al.*, 2024).

The possible marginal contributions of this paper are, first, to study the theoretical logic of data elements for new quality productivity with micro enterprises as the starting point; second, the paper validates the theory from an empirical point of view by constructing a double-difference model, which introduces resource allocation efficiency and the level of digital transformation of the enterprise as mechanism variables to analyze the paths, and also introduces the level of internal control of the enterprise and the level of scientific and technological innovation as triple difference variable to further analyze the influence mechanism. This helps to better understand the factor-driven path of enterprises' new quality productivity, and provides policy reference and useful experience for the development of new quality productivity.

III. Theoretical Analysis and Research Hypotheses

A. Direct Effects of Data Elements to Enable New Quality Productivity in Enterprises

The concept of factors of production has evolved from the earlier dichotomy of land and labor to a broader theoretical framework covering capital, organizational effectiveness, and technological innovation. Some scholars expanded the theory of endogenous growth by taking big data as a factor of production, divorced big data as a new type of factor of production from material capital, introduced the endogenization of big data into the production function under the framework of the theory of creative destruction, constructed a multisectoral Schumpeterian quality ladder model, and theoretically deducted the "multiplier effect" and the "multiplier effect" of big data to promote the height of intermediate quality ladder, and the "multiplier effect" of big data to promote the height of intermediate goods quality ladder (Yang *et al.*, 2022). The theoretical interpretation of big data to promote the height of the quality ladder of intermediate products "multiplier effect" and cause "R & D mode transformation". Specifically, the original C-D production function is $Y_t = A_t K_t^{\alpha} L_t^{\beta}$ When the pilot policy was released, in order to accurately reflect the role of data elements in the production process, the data elements (D) were included in the classic C–D production function, and it was expanded into a new production function with the following new structure integrating technology, labor, capital and data elements: $Y_{t+1} = A_{t+1}K_{t+1}^{\alpha}L_{t+1}^{\beta}D_{t+1}^{\gamma}$, where Y represents output, A represents the rate of technological progress, K represents capital, L represents labor, and Y_{t+1}/Y_t i.e., the growth effect of the new quality productivity (Nie, 2024). Based on this, this paper proposes the following hypothesis:

Hypothesis 1: Data elements can enhance the new quality productivity of a firm.

B. Indirect Effects of Data Elements to Enable New Qualitative Productivity in Enterprises

1) Resource allocation efficiency

The mechanism of resource allocation efficiency is mainly embodied in the following: first, taking advantage of big data technology, enterprises can effectively reduce information asymmetry and improve the allocation efficiency of capital and labor resources. Secondly, data-driven resource management improves the reaction speed and flexibility of enterprises, which can quickly respond to market changes and customer needs, thus maintaining an advantage in competition. Third, data elements support the refined management of resources, and through detailed data analysis of various types of resources, enterprises are able to better grasp the use of resources and optimize production and operational processes.

2) Level of digital transformation

The digital transformation level mechanism is mainly reflected in the following: first, the complete data factor marketization can break the information silos and industry information barriers, promote the aggregation of high-value elements, and provide sufficient guarantee for enterprises to carry out data transactions, data technology application and business model transformation, thus realizing the digital transformation of enterprises (Xu *et al.*, 2024). Secondly, data elements empower the decision-making process of enterprises, through the analysis of market data, customer behavior data and internal operation data, enterprises can make more scientific and accurate decisions, and improve the flexibility and market responsiveness of enterprises.

3) Level of internal control

The mechanism of the level of internal control is mainly reflected in the following: first, internal control can prevent data leakage and errors through the establishment of a sound data management system and process, so as to enhance the effectiveness of the utilization of data elements, and internal control can also have a positive impact on the quality of enterprise development by promoting the intensity of research and development, and improving the efficiency of labor output (Zhang *et al.*, 2020). Second, enterprises with a high level of internal control are able to better monitor and adjust resource allocation in data application, identify and solve potential problems in a timely manner, and improve resource utilization efficiency and productivity. Third, a good internal control mechanism promotes transparency and coordination of information within the enterprise, enhances data sharing and collaboration among departments, and promotes enterprise innovation and flexible response to market changes.

4) Level of scientific and technological innovation

The mechanism of the level of science, technology and innovation is mainly embodied in the following: on the one hand, the new quality of productivity is the advanced quality of productivity in which innovation plays a leading role and science, technology and innovation is the core element; a high level of science, technology and innovation capacity not only enhances the ability to collect, analyse and apply data, but also builds a digital environment conducive to innovation, which promotes the in-depth application of enterprises in the area of data-driven decision-making and production optimization. On the other hand, the wide application of data elements not only promotes the sharing of information among enterprises, but also promotes the sharing of R&D resources, technological knowledge and market channels, while STI further supports enterprises in developing and applying advanced data analysis tools and intelligent algorithms, which are able to more accurately predict market trends and optimize production processes, thereby improving production efficiency and product quality.

Based on the above analysis, the following hypotheses are proposed in this paper:

Hypothesis 2: Data elements can enhance the new quality productivity of firms through the path of improving the efficiency of resource allocation in firms, and hence, the new quality productivity of firms;

Hypothesis 3: The data elements can improve the digital transformation level of the enterprise through the path of improving the enterprise's digital transformation level, which in turn improves the enterprise's new quality productivity;

Hypothesis 4: The relationship between data elements and firms' new quality productivity is moderated by the level of firms' internal controls;

Hypothesis 5: The relationship between data elements and firms' new quality productivity is moderated by the level of firms' STI.

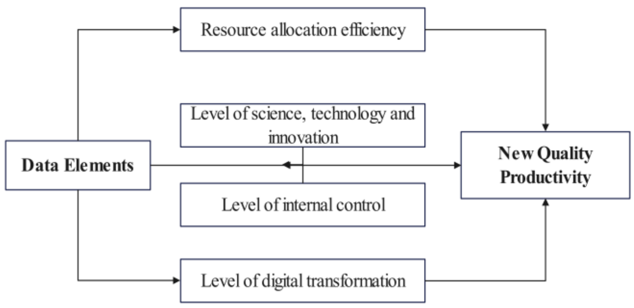

Fig. 1. Conceptual model.

IV. Research Design

A. Modeling

$$\mathrm{Npro}_{i,t} = \beta_0 + \beta_1 \mathrm{Treat}_i \times \mathrm{Post}_t + \beta_2 \mathrm{Control}_{i,t} + \mu_i + \gamma_t + \varepsilon_{i,t} \tag{1}$$

where i,t denotes the ith enterprise and year t, respectively; Npro represents the explanatory variable firm new quality productivity; Treat × Post represents the core explanatory variable data elements; Control denotes a set of control variables; μ_i denotes industry fixed effects; γ_t denotes time fixed effects; $\varepsilon_{i,t}$ denotes residual term.

B. Selection of Variables

1) Explained variables

The level of new quality productivity of enterprises. The core of new quality productivity is innovation (Song *et al.*, 2024), based on the theory of two factors of productivity and considering the role and value of labor objects in the production process, the entropy method is used to measure the new quality productivity.

Table 1. New quality productivity indicator system

Level 1 indicators	Secondary indicators	Tertiary indicators	Description of indicator values	Weight (%)
Labor force	Labor	Percentage of R&D salaries	Research and development expenses - salaries and wages/operating income	28
		Percentage of R&D staff	Number of R&D staff / Number of employees	4
		Percentage of highly educated personnel	Number of undergraduates and above / Number of employees	3
	Physical labor (objects of labor)	Fixed assets as a percentage	Fixed assets/total assets	2
		Manufacturing costs as a percentage	(Subtotal of cash outflows from operating activities + depreciation of fixed assets + amortization of intangible assets + provision for impairment - cash paid for purchases of goods and services - wages paid to and for employees)/(Subtotal of cash outflows from operating activities + depreciation of fixed assets + amortization of intangible assets + provision for impairment)	1
Production tool	Hard technology	R&D depreciation and amortization as a percentage of	R&D expenses-depreciation and amortization/operating income	27
		R&D lease payments as a percentage of	Research and development expenses - lease payments/operating income	2
		R&D direct investment as a percentage	R&D expenses - direct inputs/operating income	28
		Intangible assets as a percentage	Intangible assets/total assets	3
	Soft technology	Total asset turnover	Operating income/average total assets	1
		Inverse equity multiplier	Owners' equity/total assets	1

2) Core explanatory variables

Data elements. The enterprise registration place becomes a national-level big data comprehensive pilot zone as a proxy variable for the data element (Stan *et al.*, 2024), and the sample enterprises are divided into experimental and control groups, Treat=1 indicates that the sample enterprises are the experimental group; on the contrary, Treat=0 indicates that the enterprises are the control group. Post indicates whether the implementation of the national-level big data comprehensive pilot zone has started or not, and if the enterprise registration place belongs to the region in the current year and the following years belongs to the comprehensive big data pilot area, then Post=1; conversely, Post=0.

3) Control variables

In this paper, we choose the nature of ownership, firm size, gearing ratio, return on net assets, fixed asset ratio, proportion of independent directors, two positions, proportion of shares held by the first largest shareholder, number of years on the stock market, and whether there are four largest as control variables.

Table 2. Description of variables

Variable type	Variable name	Variable symbol	Variable Definition
Explanatory variable	New mass productivity	Npro	Entropy measurement
Core explanatory variables	process variable	Treat	1 if the region where the enterprise is registered is a policy pilot region, otherwise 0
	time dummy variable	Post	1 for 2016 and beyond, 0 otherwise
control variable	Nature of property rights	SOE	1 for state-controlled enterprises, 0 for others
	Company size	Size	Natural logarithm of total assets for the year
	gearing	Lev	Total liabilities at year-end/total assets at year-end
	return on net assets	ROE	Net profit/average balance of owners' equity
	Fixed assets as a percentage	FIXED	Net fixed assets/total assets
	Proportion of independent directors	Indep	Independent directors divided by number of directors
	two jobs in one	Dual	The chairman and general manager are the same person as 1, otherwise 0
	Shareholding ratio of the largest shareholder	Top1	Number of shares held by the largest shareholder/total number of shares
	Number of years listed	ListAge	ln(current year - year of listing + 1)
	Are the four major	Big4	1 if the company is audited by the Big 4 (PricewaterhouseCoopers, Deloitte, KPMG, Ernst & Young) and 0 otherwise.

C. Data Sources

The sample selected for this paper is the annual data of listed companies in Shanghai and Shenzhen A-shares from 2011 to 2022. The data of this paper comes from the annual reports of listed companies and related website disclosure, CSMAR database and so on. The sample excludes listed companies in the financial industry, listed companies in ST and listed companies with missing financial data, and gets 30,990 valid observations. In order to avoid the effect of extreme values, continuous variables are subjected to Winsor2 shrinkage of top and bottom 1%.

V. Empirical Analysis

A. Descriptive Statistics

In this paper, the descriptive statistics of the main variables were carried out, and the results are shown in Table 3. From the point of view of enterprise new quality productivity, the mean value of Npro of the sample enterprises is 4.924, the maximum value is 14.611, and the minimum value is 0.656, indicating that there is a large difference in the new quality productivity of listed companies in China. From the point of view of enterprise data elements, the mean value of Treat is 0.395, indicating that the proportion of enterprises subject to policy shocks in the sample enterprises is roughly 40%, which indicates that the data selection in this paper is more homogeneous. From the point of view of each control variable, its mean value is roughly consistent with existing studies.

Table 3. Descriptive statistics

Variant	**Sample size**	**Average value**	**(statistics) standard deviation**	**Minimum value**	**Maximum values**
Npro	30990	4.901	2.214	0.656	14.611
Treat	30990	0.395	0.489	0	1
Post	30990	0.683	0.465	0	1
SOE	30990	0.354	0.478	0	1
Size	30990	22.241	1.280	19.585	26.452
Lev	30990	0.421	0.203	0.032	0.908
ROE	30990	0.064	0.128	–0.926	0.437
FIXED	30990	0.209	0.154	0.002	0.719
Indep	30990	0.376	0.054	0.286	0.600
Dual	30990	0.285	0.452	0	1
Top1	30990	0.345	0.147	0.080	0.758
ListAge	30990	2.135	0.830	0.000	3.401
Big4	30990	0.059	0.236	0	1

B. Benchmark Regression

Table 4 reports the results of the baseline regressions, column (1) is the regression results without the inclusion of control variables and fixed effects, column (2) is the regression results with the inclusion of control variables but without the inclusion of fixed effects, column (1) is the regression results without the inclusion of control variables but with the inclusion of fixed effects, column (3) is the regression results with the inclusion of control variables and without the inclusion of fixed effects, and column (4) is the regression results with the inclusion of data elements (Treat×Post). Column (3) is the regression result with the inclusion of control variables and without fixed effects, column (4) is the regression result with the inclusion of control variables and fixed effects, the results show that the regression coefficients of the data elements (Treat×Post) are all significantly positive at 1% level, which indicates that the data elements can significantly improve the new quality productivity of the enterprise, and Hypothesis 1 has been verified.

Table 4. Benchmark regression

VARIABLES	(1)	(2)	(3)	(4)
	Npro	Npro	Npro	Npro
Treat×Post	0.581***	0.840***	0.139**	0.241***
	(9.652)	(14.352)	(2.193)	(3.921)
SOE		0.393***		0.523***
		(5.532)		(7.951)
Size		0.173***		0.090***
		(6.186)		(3.474)
Lev		–1.433***		–0.418***
		(–9.356)		(–3.075)
ROE		–0.512***		0.154
		(–3.501)		(1.180)
FIXED		5.687***		6.611***
		(33.520)		(37.940)
Indep		0.696		0.009
		(1.622)		(0.024)
Dual		0.091*		0.014
		(1.786)		(0.307)
Top1		–1.678***		–0.620***
		(–8.684)		(–3.606)
ListAge		–0.258***		–0.163***
		(–7.129)		(–5.005)
Big4		0.046		0.267***
		(0.397)		(2.805)
Constant	4.747***	0.987*	4.865***	1.982***
	(140.485)	(1.702)	(154.835)	(3.630)
observed value	30,990	30,990	30,990	30,990
Industry FE	NO	NO	YES	YES
Year FE	NO	NO	YES	YES
R-squared	0.013	0.191	0.255	0.385

Note: (1) ***, **, and * denote significance levels at 1%, 5%, and 10%, respectively; (2) t-statistics in parentheses; (3) Based on the fact that the reghdfe command automatically excludes samples with only a single period in the sample as well as reasons such as the replacement or addition of variables in this paper, the sample sizes in some of the tables in this paper will appear to be changed from the descriptive statistics. The same as below.

C. Robustness Tests

1) Parallel trend test

In assessing the policy effects of the data elements, it is necessary to ensure that there is consistency in the growth trends of the treatment and control groups before the implementation of the policy, which is verified in this paper by the parallel trend test. In order to avoid the problem of multicollinearity in the regression analysis, the data of the period before the implementation of the policy are excluded from the study. According to the results of the parallel trend test in Fig. 2, the estimated coefficients of the impact on the new quality productivity of enterprises before the implementation of the policy of the national-level comprehensive pilot zone for big data are relatively small and statistically insignificant, whereas the current period of the policy implementation and after it have a positive and statistically significant impact on the level of the new quality productivity of enterprises, which indicates that there is no significant systematic difference between the treatment group and the control group before the implementation of the policy, and it meets the requirements for the DID analysis. the parallel trend assumption required to conduct DID analysis. Therefore, the method of double difference analysis adopted in this study is reasonable in assessing the effectiveness of the impact of data factor policies on the level of new quality productivity of enterprises.

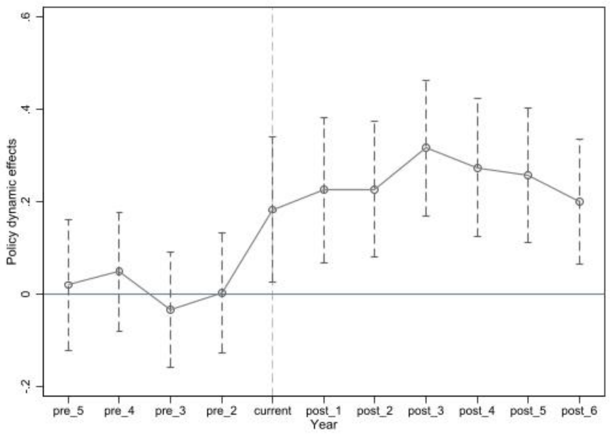

Fig. 2. Parallel trend test.

2) Placebo test

In order to ensure that the results of the benchmark model estimation are not affected by the omission of unobserved variables (Wang *et al.*, 2022). The study employs a placebo test to model the impact of data factor policies on firms' new quality productivity by randomly selecting 1,743 firms as the hypothetical treatment group and treating the remaining firms as the hypothetical control group. The above process is repeated 500 times to obtain 500 regression coefficients and their corresponding p-values. By plotting the kernel density distribution and p-values of these 500 coefficient estimates as shown in Fig. 3, it can be seen that the regression coefficients fall around the value of 0 and follow a normal distribution, with the vast majority of the regressions being insignificant. The coefficient estimates in the benchmark regression are located on the right side of the distribution of spurious regression coefficients, which are small probability events in the placebo test. Accordingly, it can be ruled out that the benchmark estimates in this paper are due to unobservables.

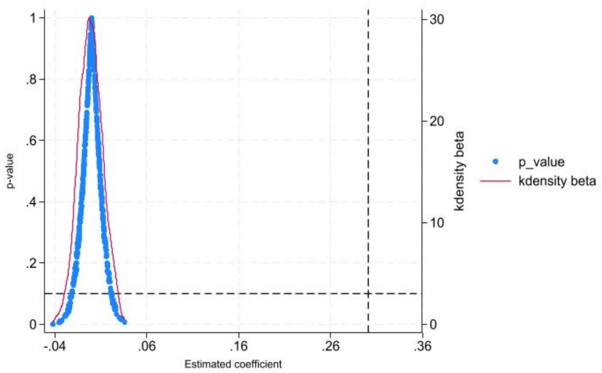

Fig. 3. Placebo test.

3) Substitution of explanatory variables

Considering that the new quality productivity mainly contains the optimization and perfection of the allocation mode of production factors, the revolutionary advancement of traditional technology, the deepening and upgrading of regional industries and other important aspects, and takes the improvement of total factor productivity as the core symbol, therefore, this paper adopts the LP method to measure the total factor productivity of enterprises as a measure of the new quality productivity of enterprises and then re-regress it, and the results, as shown in Column (1) of Table 5, are shown in column (1) of Table 5, and the data elements of the regression coefficients are significantly positive at the 1% level, consistent with the benchmark regression results.

4) PSM-DID

The initial conditions of the pilot and non-pilot regions of the comprehensive big data pilot area may differ significantly, and this systematic difference may make the pilot areas lack randomness in selection, which leads to poor comparability of the sample firms. For this reason, this paper attempts to use Propensity Score Matching (PSM) to address the endogeneity problem caused by the sample selection bias problem. Specifically, using the control variables as covariates, there is a putback for 1:1 near-neighbor matching, selecting from the sample firms similar to the firms subject to policy shocks Vickers subject to policy shocks as the control group, and rerunning the regressions in accordance with the model in equation (1), the estimation results are shown in column (2) of Table 5, and the interaction term (Treat×Post) is significantly positive at the 1% level, which further verifies the findings' Robustness.

5) Removing policy interference

Inevitably, there are also major policies that could potentially affect firms' new quality productivity during the time interval of the study, thus making the estimated effects of the Big Data Comprehensive Pilot Zone policies overestimated or underestimated. In order to identify and exclude the interference of the policy stacking effect caused by such related policies, this paper combed through other policy events enacted during the implementation of the Big Data Pilot Zone, focusing on the possible interference of broadband China, by adding the policy dummy variable of "broadband China" in the baseline regression. The regression results in column (3) of Table 5 show that after considering the above policy impacts, the impact of the construction of the comprehensive big data pilot zone on the new quality productivity of enterprises is still significantly positive at the 1 percent level.

Table 5. Robustness test

VARIABLES	(1) Substitution of explanatory variables	(2) PSM-DID	(3) Removing policy interference
	TFP_LP	Npro	Npro
Treat×Post	0.047***	0.145**	0.196***
	(2.904)	(2.063)	(3.194)
"Broadband China" pilot program			0.298***
			(5.928)
Constant	–4.673***	1.948***	1.872***
	(–27.111)	(2.769)	(3.431)
control variable	YES	YES	YES
observed value	30,003	15,326	30,990
Industry FE	YES	YES	YES
Year FE	YES	YES	YES
R-squared	0.767	0.390	0.388

6) Endogenous processing

a) Instrumental variables approach

Data elements can drive firms' new quality productivity, but at the same time, the improvement of firms' new quality productivity also affects the development of data elements, so there may be a bidirectional causality, and considering that there may be omitted variables, this paper uses the Two-Stage Least Squares (2SLS) method for testing.

First of all, the construction of data factor market is closely related to the level of local information infrastructure construction, and the cost of information infrastructure construction is greatly affected by the topography of the city, the greater the degree of topographic relief, the greater the cost of network infrastructure construction, and at the same time, topographic relief also affects the flow of information and data elements, affecting the construction of the data factor market, therefore, the degree of topographic relief meets the correlation hypothesis of instrumental variable setting. correlation assumption, and digital industry innovation relies more on factors such as R&D investment, human capital, science and technology, and is less directly affected by terrain relief, which also meets the condition of exogeneity, so the selection of terrain relief as an instrumental variable has a certain degree of rationality (Zhao *et al.*, 2024). Secondly, in the case of Internet and data elements, it is not easy for the data elements to be used as instruments. Secondly, in areas with higher levels of development of Internet and data elements, their communication business tends to develop better, which satisfies the condition of instrumental variable, therefore, this paper also selects the number of fixed telephones per 100 people at the end of 1984 as the instrumental variable.

Considering that the degree of topographic relief and the 1984 data are cross-sectional data, The instrumental variables for the data elements are represented by using the cross-multiplier terms of the degree of terrain undulation and the number of fixed telephone sets per 100 people at the end of 1984 with the national IT service revenue in the previous year (Zhu *et al.*, 2024), respectively, and the results are shown in Column (1) of Table 6. The regression results show that the regression coefficients for the data elements in the second stage remain significantly positive, eliminating the interference of the endogeneity problem in the conclusions of this paper and verifying the accuracy of its conclusions.

b) Lagged one-period explanatory variables

Considering that after the implementation of the pilot policy of "Comprehensive Experimental Zone of Big Data", there is a certain construction period, resulting in the policy effect is not realized in the current period, so this paper adopts the lagged one-period method of the explanatory variables, to exclude the interference of endogeneity on the empirical results, and the results are shown in Column (2) of Table 6. The results show that the regression coefficients of the data elements (Treat×Post) are still significantly positive at the 1% level, indicating that the conclusions of this paper are reliable.

Table 6. Endogenous treatments

VARIABLES	(1) Instrumental variables approach		(2) Lagged one-period explanatory variables
	Phase I	Phase II	
	Treat×Post	Npro	Npro
IV1	0.013***		
	(88.492)		
IV2	–0.001***		
	(–15.111)		
Treat×Post		0.921***	0.191***
		(17.047)	(3.053)
Kleibergen-Paap rk LM statistic	3856.729		
	[0.000]		
Kleibergen-Paap rk Wald F-statistic	9452.632		
	{19.93}		
Constant	–0.450***	1.757***	2.277***
	(–8.068)	(6.066)	(4.057)
control variable	YES	YES	YES
observed value	30,990	30,990	25,999
Industry FE	YES	YES	YES
Year FE	YES	YES	YES
R-squared	0.408	0.371	0.373

Note: ***, **, and * denote significance levels at 1%, 5%, and 10%, respectively, t-statistics in ( ), p-values in [ ], and values in { } are critical values at the 10% level of the Stock-Yogo weak identification test.

D. Heterogeneity Analysis

1) Heterogeneity analysis based on firm nature

Referring to some scholars study (Peng *et al.*, 2017), whether the enterprise belongs to high-tech enterprises or not is divided into two parts for the difference study, and the results are shown in column (1) and column (2) of Table 7. The results show that when the firms belong to high-tech firms, the role of data elements on the new quality productivity of the firms is more obvious than that of non-high-tech firms, and the Chow test also shows that the coefficients are significantly different between the two groups. The possible reason for this is that high-tech firms usually have a more advanced technological base and stronger data processing capabilities, and are able to transform data into productivity-enhancing drivers more efficiently. In addition, these firms rely more on data analysis in their decision-making process, have stronger innovation capabilities and market agility, can quickly adapt to market changes and identify new opportunities and trends, and invest more in R&D to enhance the efficiency of technological advancement and product innovation through data application. In contrast, non-high-tech firms may be relatively deficient in terms of technological base, data-processing capability and R&D investment, resulting in a relatively weak impact of data elements on new quality productivity.

2) Heterogeneity analysis based on the nature of the region

In this paper, the provinces where the research sample enterprises are registered are categorized into two parts: the east and the central and western parts for the differentiation study, and the results are shown in columns (3) and (4) of Table 7. The results show that the regression coefficients of data factors are significantly positive at 1% level in the eastern region, and the regression coefficients of data factors are not significant in the central and western regions, which also pass the Chow test at 5% level. It indicates that data factors can significantly increase the new quality productivity of enterprises in the eastern region, but it is not yet significant in the central and western regions. The possible reasons are that the eastern region has more perfect digital infrastructure and high-quality information technology talents, the economic structure is dominated by high-tech industry and service industry, which is more dependent on data, while the government has invested more resources and policy support in promoting the development of the digital economy, and the fierce competition in the market has prompted the enterprises to actively adopt the data-driven production and management methods. On the other hand, in the central and western regions, on the one hand, the construction of digital infrastructure is relatively insufficient, resulting in the role of data elements on enterprise productivity improvement is not obvious, on the other hand, the central and western regions account for a large proportion of traditional manufacturing industries and resource-based industries, the demand for data in these industries is relatively low, which also restricts the play of data elements.

Table 7. Heterogeneity analysis

VARIABLES	(1) High-tech enterprises	(2) Non-high-tech enterprises	(3) Eastern Region	(4) Midwest Region
	Npro	Npro	Npro	Npro
Treat×Post	0.236***	0.270***	0.335***	–0.026
	(2.839)	(3.175)	(4.927)	(–0.190)
Constant	1.466*	2.220***	1.863***	2.020*
	(1.913)	(2.951)	(2.882)	(1.898)
control variable	YES	YES	YES	YES
observed value	18,366	12,624	22,170	8,815
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
R-squared	0.289	0.518	0.388	0.398
Difference in coefficients between groups	0.280***		0.334**	

Note: Differences in coefficients between groups are Chow test estimates.

VI. Scalability Studies

A. *Mechanism of Action Analysis*

Based on the theoretical analysis in the previous section, this paper argues that data elements may enable channels for enterprises to improve the level of resource allocation efficiency and digital transformation, which in turn will increase the new quality productivity of enterprises. Based on this, this part will try to test the above potential channels of influence. Referring to some scholars the discussion about channel testing (Jiang *et al.*, 2022), the first test is whether the core explanatory variables act on the mechanism variables (M). On this basis, in order to avoid the problem that the theoretical argument for the causal effect of M on Y may not be sufficient, we further test the effect of M on Y, thus supplementing the correlation evidence support. The specific model is as follows:

$$\mathrm{M}_{i,t} = \beta_0 + \beta_1 \mathrm{Treat}_i \times \mathrm{Post}_t + \beta_2 \mathrm{Control}_{i,t} + \mu_i + \gamma_t + \varepsilon_{i,t} \quad (2)$$

$$\mathrm{Npro}_{i,t} = \beta_0 + \beta_1 \mathrm{M}_{i,t} + \beta_2 \mathrm{Control}_{i,t} + \mu_i + \gamma_t + \varepsilon_{i,t} \quad (3)$$

1) Path to more efficient resource allocation

This paper refers to some scholars study (Ni *et al.*, 2022), which first estimates the reasonable level of firms' investment in the current year and then calculates the degree of inefficient investment to measure firms' investment efficiency. Columns (1) and (2) of Table 8 report the regression results of the mechanism analysis of resource allocation paths. Column (1) shows that the regression coefficient of data elements on inefficient investment is -0.010, which is significantly negative at the 1% level, and the regression coefficient of inefficient investment on firms' NQP in Column (2) is also significantly negative at the 1% level. It indicates that data elements will increase the level of new quality productivity through the channel of improving the efficiency of resource allocation in enterprises, which in turn will increase the level of new quality productivity. Hypothesis 2 is verified.

2) Improve the level of enterprise digital transformation path

This paper refers to some scholars (Wu *et al.*, 2021), we search, match, and count the word frequencies of the feature words of digital transformation extracted from the text of annual reports of listed enterprises, and then categorize the word frequencies of the key technology directions and form the final summed word frequencies. Since this kind of data has the typical characteristic of "right skewedness", logarithmization is carried out to get the overall indicators for the portrayal of enterprises' digital transformation. Overall indicators of enterprise digital transformation. Columns (3) and (4) of Table 8 report the regression results of the mechanism analysis of the path of enterprise's digital transformation level. Column (3) shows that the regression coefficient of data elements on digital transformation level is 0.130, which is significantly positive at 1% level, and the regression coefficient of the digital transformation level on the enterprise's new quality productivity in Column (4) is also significantly positive at 1% level. It indicates that data elements will increase the level of new quality productivity of enterprises through the channel of increasing the level of digital transformation of enterprises, and then increase the level of new quality productivity of enterprises. Hypothesis 3 is verified.

Table 8. Mechanism analysis

VARIABLES	(1)	(2)	(3)	(4)
	Ineff	Npro	DCG	Npro
Treat×Post	–0.010***		0.128***	
	(–3.790)		(3.965)	
Ineff		–0.434***		
		(–3.779)		
DCG				0.173***
				(7.993)
Constant	0.231***	2.063***	–2.530***	2.299***
	(5.776)	(3.622)	(–8.748)	(4.248)
control variable	YES	YES	YES	YES
observed value	27,683	27,683	30,790	30,790
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
R-squared	0.100	0.393	0.486	0.390

B. Triple Difference Models

In this paper, we adopts the following triple difference model to further test the mechanism of the influence of data elements on firms' new quality productivity (Xiao *et al.*, 2021):

$$\mathrm{Npro}_{i,t} = \beta_0 + \beta_1 \mathrm{Treat}_i \times \mathrm{Post}_t \times \mathrm{IC}_i + \beta_2 \mathrm{Treat}_i \times \mathrm{Post}_t + \beta_3 \mathrm{Treat}_t \times \mathrm{IC}_i + \beta_4 \mathrm{Post}_t \times \mathrm{IC}_i + \beta_5 \mathrm{Control}_{i,t} + \mu_i + \gamma_t + \varepsilon_{i,t} \quad (4)$$

$$\mathrm{Npro}_{i,t} = \beta_0 + \beta_1 \mathrm{Treat}_i \times \mathrm{Post}_t \times \mathrm{Patent}_i + \beta_2 \mathrm{Treat}_i \times \mathrm{Post}_t + \beta_3 \mathrm{Treat}_t \times \mathrm{Patent}_i + \beta_4 \mathrm{Post}_t \times \mathrm{Patent} + \beta_5 \mathrm{Control}_{i,t} + \mu_i + \gamma_t + \varepsilon_{i,t} \quad (5)$$

1) Level of internal control

In this paper, the Boddy internal control index is used to measure the level of internal control of the firms, and the dummy variable for the level of internal control of the firms, IC, is set. IC takes the value of 1 when the level of internal control of the firms is higher than the median, otherwise IC takes the value of 0. The regression results of the model (4) are shown in the Table 9 column (1). As can be seen from Table 7, the estimated coefficient of the key explanatory variable Treat × Post × IC is 0.217 and is significant at 1% level. This indicates that the level of internal control of firms plays a significant positive moderating role in the effect of data elements on firms' new quality productivity. Thus hypothesis 4 is proved.

2) Level of scientific and technological innovation

In this paper, whether the enterprise applies for a patent or not is used to set the dummy variable Patent for the enterprise's level of science and technology innovation. When the enterprise applies for a patent in the current year, Patent takes the value of 1, otherwise Patent takes the value of 0. The regression results of model (4) are shown in column (1) of Table 9. As can be seen from Table 7, the estimated coefficient of the key explanatory variable Treat×Post×Patent is 0.354 and is significant at 1% level. This indicates that the level of science, technology and innovation of the firm plays a significant positive moderating role in the effect of data elements on the new quality productivity of the firm. Thus hypothesis 5 is proved.

Table 9. Triple difference test

VARIABLES	(1)	(2)
	Npro	Npro
Treat×Post×IC	0.217***	
	(2.902)	
Treat×Post×Patent		0.204**
		(2.133)
Treat×Post	0.165**	0.158**
	(2.499)	(2.269)
Post×IC	–0.015	
	(–0.607)	
Treat×IC	–0.071	
	(–1.417)	
Post×Patent		0.353***
		(6.315)
Treat×Patent		–0.018
		(–0.271)
Constant	0.270***	0.284***
	(2.830)	(2.991)
control variable	YES	YES
observed value	30,990	1.827***
Industry FE	YES	(3.366)
Year FE	YES	YES
R-squared	0.385	0.390

VII. Research Conclusions and Policy Recommendations

A. Conclusions of the Study

Based on the data of A-share listed companies in Shanghai and Shenzhen from 2011 to 2022, this paper constructs a double-difference model to empirically analyze the correlation between data elements and the new productivity of enterprises, and draws the following conclusions: first, data elements significantly promote the enhancement of the new productivity of enterprises. Second, the heterogeneity analysis shows that data factors enhance the new productivity of high-tech enterprises more significantly than that of non-high-tech enterprises; data factors promote the new productivity of enterprises in the eastern region, but have no significant effect on the new productivity of enterprises in the central and western regions. Third, the mechanism analysis shows that data elements promote enterprise new productivity through promoting enterprise resource allocation efficiency and digital transformation level, and then promote enterprise new productivity, meanwhile, the triple difference model shows that in the process of promoting enterprise new productivity, the level of enterprise internal control and scientific and technological innovation play a positive role in regulating the data elements.

B. Policy Recommendations

Based on this, the paper makes the following policy recommendations:

Firstly, promote the widespread application of data elements in various industries. The government and enterprises should strengthen the collection, management and utilization of data resources, and thoroughly implement the Three-Year Action Plan of "Data Elements x" (2024–2026), especially in high-tech industries, where they should increase investment in data elements, enhance data processing and analysis capabilities, and promote enterprises' technological innovation and productivity enhancement.

Secondly, optimize the distribution of regional data elements. More support should be given to the central and western regions in terms of policy to promote the construction of data infrastructure and the equitable distribution of data elements, narrow the regional gap, and promote the improvement of new quality productivity of enterprises nationwide.

Thirdly, it will enhance the efficiency of resource allocation and the level of digital transformation of enterprises. Through policy guidance and financial support, we will help enterprises optimize their resource allocation and improve their digitalization level. Strengthen the internal digital management and operation of enterprises, enhance internal control and management efficiency through information technology, and promote the overall productivity and innovation capacity of enterprises.

Finally, we should strengthen the internal control of enterprises and their capacity for science, technology and innovation. Enterprises should strengthen internal control, clarify the division of responsibilities and improve the efficiency of decision-making and implementation. At the same time, they should increase investment in Research and Development (R&D), enhance their capacity for scientific and technological innovation, and drive productivity enhancement through innovation. The government can promote the enthusiasm of enterprises in technological innovation by providing incentives for scientific and technological innovation and financial support for research and development.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Y.X. contributed to data collection, data analysis, interpretation, manuscript preparation, and editing; Y.W. contributed to study design, as well as the introduction and conclusion sections of the manuscript manuscript; both authors had approved the final version.

References

Chen, X., Liu, S., & Shen, H. 2024. Enhancing new productivity with digital transformation. ***Xinjiang Social Science***, (02): 41–45.

Du, C., Li, Y. 2024. Mechanism of strengthening the ability of scientific and technological innovation to accelerate the formation of new quality productivity. ***Journal of Hunan University of Science and Technology (Social Science Edition)***, 27(01): 100–109.

Feng, Y., Lin, H. 2024. Data elements enabling new quality productivity:theoretical logic and practical approach. ***The Economist***, (05): 15–24.

Gu, H., Bian, Y. 2020. Internal control, board interlock and enterprise innovation: based on the evidence from listed companies on GEM in China. ***Journal of Management***, 33(06): 48–60.

Guo, L., Li, X. 2020. Internal control, social audit and enterprise total factor productivity: Collaborative supervision or mutual substitution. ***Journal of Statistics and Information***, 35(11): 77–84.

Jiang, T. 2022. Mediating effects and moderating effects in causal inference. ***China Industrial Economy***, (05): 100–120.

Jin, X., Sun, Q., & Jin, R. 2025. Digital transformation, new quality productive forces and corporate innovation performance. *Journal of Hainan University (Humanities & Social Sciences)*, 43(01): 86–96.

Li, Z. 2024. A study on the theoretical logic and practical progression of data elements empowering new quality productive forces—An analysis based on Marx's labor process theory. ***Shanghai Journal of Economics***, (05): 25–36.

Liao, X., Bai, J. 2024. Can digital government improve the efficiency of enterprise resource allocation? ***Modern Economic Research***, (04): 60–75+92.

Lin, C., Ge, X., & Sun, Y. Digital transformation and new productivity enhancement in commercial banks. ***The Theory and Practice of Finance and Economics***, 1–8 Available: http://kns.cnki.net/kcms/detail/43.1057.f.20240612.1419.008.html

Ma, S., Xu, M., & Zhang, C. 2024. Application of data factor and enterprise innovation effect: Empirical evidence from Chinese A-share listed companies. ***Journal of Business Economics***, (02): 50–67.

Ni, T., Wang, Y. 2022. Regional administrative integration, factors' marketization and firms' resource allocation efficiency. ***Journal of Quantitative & Technological Economics***, 39(11): 136–156.

Nie, Y., Ma, X., & Zheng, J. 2024. The level of data factor development and economic growth: analysis of impact effects and mechanisms. ***China Circulation Economy***, 38(08): 56–68.

Peng, H., Mao, X. 2017. Government subsidies for innovation, company executives background and R&D investment-evidence from the high-tech industry. ***Finance and Trade Economics***, 38(03): 147–161.

Ren, B., Dou, Y. 2024. Combination of data, arithmetic and algorithms reflecting the level of digital development of the new quality productivity. ***Journal of Zhejiang Gongshang University***, (03): 91–100.

Shi, D., Sun, G. 2024. Data elements and new quality productive forces: A perspective from total factor productivity of enterprises. ***Economic Theory and Business Management***, 44(04): 12–30.

Song, J., Zhang, J., & Pan, Y. 2024. Research on the impact of ESG development on new quality productive forces of enterprises—Empirical evidence from Chinese A-share listed companies. ***Contemporary Economic Management***, 46 (06): 1–11.

Wang, F., Ge, X. 2022. Can low-carbon transition impact employment—Empirical evidence from low-carbon city pilot policy. ***China Industrial Economics***, (05): 81–99.

Wang, L., Liao, Z., Jia, N. 2024. New horizons for urbanization: Innovation-driven and information-pulled mechanisms of data factors. ***China Rural Economy***, (06): 25–49.

Wu, F., Hu, H., et al. 2021. Enterprise digital transformation and capital market performance: Empirical evidence from stock liquidity. ***Journal of Management World***, 37(07): 130–144+10.

Xiao, J., Xiao, Y., & Shi, W. 2021. The growth path of Chinese enterprises' OFDI under the belt and road initiative. ***International Business Research***, (07): 107–119+137.

Xu, Y., Wang, Z. 2024. Data elements marketization and enterprise digital transformation: A quasi-natural experiment based on data trading platform. ***Soft Science***, 38(09): 24–29+39.

Yang, J., Li, X., & Huang, S. 2022. Big data, technical progress and economic growth: An endogenous growth theory introducing data as production factors. ***Economic Research Journal***, 57(04): 103–119.

Zeng, J., Yang, M. 2024. Tax effort, resource misallocation and total factor productivity of enterprises. ***Statistics & Decision***, 40(02): 144–149.

Zhang, G., Meng, M. 2020. Internal control, media attention and high-quality development of manufacturing companies. ***Modern Economic Research***, (05): 81–87.

Zhang, Q., Yang, M. 2022. Corporate digital transformation and internal control quality—A quasi-natural experiment from integration of informatization and industrialization. ***Auditing Research***, (06): 117–128.

Zhao, F., Jiang, G., & Ma, W. 2024. Data factor market empowers digital industry innovation: Evidence from Quasi-natural experiments. ***Economic Review***, (03): 109–125.

Zhao, P., Zhu, Y., & Zhao, L. 2024. National big data comprehensive experimental zone and new quality productivity: Based on empirical evidence from 230 cities. ***Journal of Chongqing University (Social Science Edition)***, 30(04): 62–78.

Zhu, Q., Zou, X., Shi, X. 2024. Can the development of digital economy reduce the dependence of local governments on land finance? ***Contemporary Finance and Economics***, (07): 45–58.