Pacific International Journal, Vol. 7(6); 2024
ISSN (Print) 2663-8991, ISSN (Online) 2616-4825[1]
DOI: 10.55014/pij.v7i6.719
https://rclss.com/index.php/pij

Data Assets and Enterprise Total Factor Productivity

Feng Liu [1, 2] , Xinran Wang [2] , Hongjuan Wang [3]

[1]Philippine Christian University，1004，Manila City，Republic of the Philippines;
[2]School of Economics and Management, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353 , China;
[3] Lyceum of the Philippines University , 1002 Intramuros, NCR, City of Manila, Republic of the Philippines
Email: 153918494@qq.com, wxr2514@163.com, hongjuanwang2015@163.com

Abstract: The progression of digital transformation has rendered the assetization of data an inevitable trend. This paper investigates the relationship between data assets and total factor productivity using a sample of A-share listed companies in Shanghai and Shenzhen. We find that data assets enhance firms' total factor productivity via two primary mechanisms: augmenting human capital and fostering innovation within firms. Moreover, this effect is particularly pronounced in large-scale firms, state-owned enterprises (SOEs), as well as firms located in regions characterized by high levels of digitization and economic policy uncertainty. This paper contributes to the literature by enriching and expanding the understanding of the economic consequences of data assets on firms' total factor productivity, while providing a policy rationale for further promoting data assetization as a catalyst for innovation and development.

Keywords: data assets; total factor productivity; human capital; level of innovation

I. Introduction

In recent years, the advancement of digital transformation has spurred the emergence of data-driven industries and new business models, significantly improving the ability to generate, collect, store, and process data, thereby leading to massive data growth. The latest information from the National Data Work Conference indicates that total data production in China is projected to exceed 32 zettabytes (ZB) by 2023. As the scale of data expands, its value as a resource continues to be unlocked and generated. In April 2020, the CPC Central Committee and the State Council issued the "Opinions on Building a More Market-oriented Mechanism for Factor Allocation," officially recognizing data as a key factor of production alongside labor, capital, land, and technology, and calling for the accelerated development of the data factor market. In August 2023, the Ministry of Finance (MoF) issued the "Interim Provisions on Accounting for Data Resources of Enterprises" (the "Interim Provisions"), clarifying the scope, standards, and requirements for the presentation and disclosure of data resources. These provisions took effect on January 1, 2024, marking a strategic milestone in guiding Chinese enterprises in the accounting and reporting of data assets. Compared to traditional production factors, data as a factor of production is intangible, non-rivalrous, exhibits increasing returns to scale, and possesses strong positive externalities, unclear property rights, and derivative attributes. It not only functions as a production factor but also enhances the interaction among other factors, generating a multiplier effect (Xu et al., 2021; Li et al., 2021). Data assetization has become an inevitable trend, and examining the impact of data assets on enterprises has become a critical issue.

Existing research on data assets can be divided into two main areas: the definition of the data asset concept and the economic consequences of data assets. Regarding the concept of data assets, the term was first introduced by Richard Peters in 1974, who regarded data assets as including items such as government bonds, corporate bonds, and physical bonds. However, this concept gained limited attention until the advent of information technology, which significantly broadened the definition and scope of data assets. Bughin et al. (2010) define data assets as intelligent resources capable of enhancing process efficiency, adding new functions to products, and fostering innovative business models. Chen et al. (2020) define data assets from the perspective of acquisition and storage, viewing them as resources generated internally or acquired externally by an enterprise for specific purposes, which can be stored, processed, and either traded in the market or used to support production and operational activities. Zhang et al. (2020) emphasize that data assets must meet the basic criteria of assets, be owned or controlled by the enterprise, and are expected to generate economic benefits. In essence, data assets refer to resources that are either internally produced or externally acquired, owned or controlled by the enterprise, stored in digital or physical form, and capable of delivering future economic benefits.

Regarding the economic consequences of data assets, at the macro level, Veldkamp et al. (2019) propose a producer model where the output of goods (Y) depends on the amount of data (Ω) and the labor (L) employed in production. The combination of data and labor yields increasing returns to scale, thereby promoting economic growth. Xu et al. (2020) develop an economic growth model of data capital, illustrating both the direct impact and the spillover effects of data capital on economic growth. Zhao (2024) highlights the strategic significance of accounting for data assets, arguing that their inclusion will fully unleash the potential of the digital economy. At the micro level, research primarily explores the

[Received 01 October 2024; Accepted 18 December 2024; Published (online) 20, December, 2024]

impact of data assets on enterprise innovation investment, firm value, and total asset growth (Gunasekara et al., 2017; Yuan et al., 2022).

Existing studies increasingly examine the role of data assets in empowering enterprise development, but most focus on measuring corporate data asset disclosures through textual analysis and evaluating their impact on enterprise value and related metrics. However, there is a lack of sufficient evidence regarding the relationship between data assets and total factor productivity, as well as limited explanations of the mechanisms through which data assets exert influence. To address this gap, this paper constructs a conceptual model and uses Chinese A-share listed companies as a sample to examine whether, and how, data assets affect total factor productivity.

The potential contributions of this paper can be summarized in three key areas: (1) By constructing a theoretical model of the impact of data assets on enterprise total factor productivity, it expands the understanding of the economic consequences of data assets on total factor productivity, verifies their positive influence, and offers new insights for enterprise development. (2) This paper analyzes the impact of data assets on total factor productivity through two intermediary channels, namely human capital and innovation level, and reveals the underlying mechanisms through which data assets enhance total factor productivity, thereby providing a theoretical framework for how enterprises can leverage data assets to drive innovation and development. (3) By incorporating firm-specific characteristics and external macroeconomic factors into the empirical analysis, the paper examines the heterogeneous effects of data assets on total factor productivity, providing a theoretical reference for more targeted and effective government policy implementation.

II. Literature review and research hypothesis

Total factor productivity (TFP) refers to the level of output per unit of input and is typically understood as the "surplus" in total output attributed to factors like technological progress, managerial expertise, and production innovations beyond traditional factor inputs (Baier et al., 2010). TFP is influenced by a variety of factors, including external factors such as macroeconomic conditions, national policies, and infrastructure, as well as internal factors like human capital, enterprise innovation, financing constraints, and corporate governance. Several scholars have explored the impact of human capital on TFP. As enterprises enhance their human capital, high-quality intellectual and human capital become integrated into production and operations, improving business processes, reducing production and transaction costs, facilitating technology diffusion, enhancing innovation capacity, and ultimately boosting productivity (Teixeira & Fortuna, 2010; Liu et al., 2018). Additionally, current domestic and international studies on enterprise innovation and TFP are extensive, with most confirming that innovation positively contributes to the improvement of TFP (Mohnen and Hall, 2013; Amable et al., 2016; Liu, 2022; Ye et al., 2016). Data assets possess key characteristics such as derivativeness, virtualization, non-competition, sharing, exclusivity, and inherent asset value (Zhu et al., 2018), which can elevate human capital levels, foster enterprise innovation, and consequently enhance TFP.

Data assets can elevate human capital levels and thereby enhance TFP. Human capital encompasses the collective knowledge, skills, and experience held by an enterprise's employees (Yuan et al., 2022). Data assets are intelligent resources that aggregate information across product design, manufacturing, operations, maintenance, and supply chains. By collecting and analyzing this data, enterprises can achieve fine-tuned management of the product life cycle, thereby improving operational efficiency and reducing costs (Xiao, 2020). The acquisition, management, and application of data assets not only demand appropriate technological infrastructure but also require high-level human capital proficient in data mining, analysis, and processing technologies. To expedite the acquisition and application of data assets, and maximize their value, enterprises must recruit data-related technical experts or train existing staff in relevant technologies, thereby enhancing the enterprise's human capital. Moreover, in the digital economy era, enterprises must fully leverage data assets to achieve sustainable development. The presence of data assets in an enterprise highlights its emphasis on soft power resources such as data, information, and knowledge, which enriches decision-making, optimizes service matching and supply (Sun et al., 2019), and enhances the enterprise's core market competitiveness. Thus, data assets enhance the enterprise's image, and firms with robust data assets are often viewed as having greater development potential, attracting high-level human capital and further improving the enterprise's human capital.

Data assets can foster enterprise innovation and thereby enhance TFP. Data assets can effectively alleviate enterprises' financing constraints, thereby providing the necessary financial support for innovation activities. Innovation activities are typically characterized by high investment, significant risk, and long cycles, making them highly dependent on funding. First, data assets are processed and valuable resources that help financial institutions better understand an enterprise's production, operational status, and creditworthiness. This reduces information asymmetry between banks and enterprises, alleviating issues of adverse selection and credit risk, thus lowering financing costs and easing financing constraints (Begenau et al., 2018). Additionally, data assets provide a transparent view of an enterprise's real condition, allowing investors to more accurately assess the rationality and profit potential of their investments (Tang et al., 2021). This reduces information collection and decision-making costs, as well as risks caused by information asymmetry, enabling investors to offer enterprises more substantial financing (Lin et al., 2007), thereby alleviating financing constraints. Second, the non-competitive nature of data production factors ensures high utilization efficiency and significant potential economic value (Jones et al., 2020). Data assets, as valuable resources owned or controlled by enterprises, possess inherent value. Enterprises can organize, sell, or transfer all or part of their data assets to generate income through exclusivity (Varian, 2018), further mitigating their financing constraints. Furthermore, data assets can transform enterprise innovation modes and improve innovation efficiency. First, their virtual and shareable nature allows innovation activities to overcome time and geographical restrictions, enabling inter-enterprise resource sharing and cross-enterprise, cross-regional, and cross-industry R&D collaboration. Innovation is inherently a trial-and-error process, and the vast data

generated from failed attempts during R&D is of great value to enterprises (Akcigit et al., 2016). Exchanging and sharing data during the R&D process among enterprises reduces trial-and-error rates, significantly improves innovation efficiency, and enhances production efficiency. Second, data assets are reusable; once analyzed and valued, they can be recycled for further analysis and data mining in other segments, continuously generating economic value (Nolin, 2020). Accelerated data flow and sharing facilitate open, networked, collaborative innovation across enterprise systems, supply chains, and consumers (Tu et al., 2023). Concurrently, knowledge spillovers further enhance overall innovation capabilities.
Based on the above analysis, the following hypothesis is proposed:
H1: Data assets positively influence firms' total factor productivity (*TFP*).

III. Research design

3.1Sample selection and data sources

We use A-share listed companies in Shanghai and Shenzhen from 2008 to 2020 as the base sample and excludes certain firms based on established research criteria: (1) companies categorized as ST, *ST, or PT; (2) companies with missing data for relevant variables; and (3) companies in the financial industry. The data for the empirical analysis are primarily sourced from the CSMAR, Wind, and China Research Data Service Platform (CNRDS) databases.

3.2Definition and interpretation of variables

3.2.1 Explained Variables. This paper employs the LP method (Levinsohn, 2003) to calculate total factor productivity (*lnTFP*) and applies a logarithmic transformation to total factor productivity.
3.2.2 Core Explanatory Variables. Following the approach of Lu (2023), we calculate firms' data assets (*dataAsset*) using the formula ln (market value − fixed assets − financial assets − intangible assets).
3.2.3 Mediating Variables. (1) Firm Innovation Level (*lnApply*): Firm innovation level (*lnApply*) is measured as the logarithmic value of the number of patent applications filed by the firm in the current year, following Lu et al. (2023). (2) Human Capital (*talent*): Human capital (*talent*) is quantified as the percentage of R&D personnel within the firm.
3.2.4 Control Variables. In accordance with established studies (Song et al., 2021), we select both firm-level and regional-level variables as control variables. The firm-level control variables include firm size (*Size*), firm age (*FirmAge*), cash holdings (*CashHolding*), gearing ratio (*Lev*), firm growth (*Growth*), return on total assets (*ROA*), and nature of property rights (*SOE*). The regional-level control variables encompass the level of economic development (*lnPerGDP*). Specific definitions of the variables can be found in Table 1.

Table 1 List of variable definitions

Variable abbreviation	Variable	Definition
lnTFP	Total factor productivity	LP method takes the natural logarithm
TFP_OP	Total factor productivity	OP method
dataAsset	data asset	ln (market value -- fixed assets -- financial assets -- intangible assets)
dataAsset1	data asset	Data asset keywords as a percentage of total word frequency in the annual report
SA	Financing constraints	SA=0.043*size*sa_size-0.04*Age-0.737*size
agencyCost	agency cost	(Administrative expenses + Selling expenses) divided by sales revenue
lnApply	Innovation level	Logarithmic value of the number of patent applications filed by the enterprise in the year
Size	Enterprise size	Natural logarithm of total assets
FirmAge	Age of business	Difference between the current year and the year of registration
CashHolding	cash holdings	Cash and cash equivalents divided by total assets at end of period
Leverage	gearing	Liabilities divided by assets
Growth	Corporate Growth	Growth rate of operating income
ROA	return on total assets	Net profit/total assets
SOE	Nature of property rights	State-owned enterprises are assigned a value of 1, while others are assigned a value of 0.
lnperGDP	Level of economic development	Natural logarithm of GDP per capita
MI_FG	Marketization Index	Marketization index

3.3Modeling：

In order to examine the effect of data assets on total factor productivity, we perform the following baseline regression estimates:

$$lnTFP_{i,t} = \alpha_0 + \alpha_1 \times dataAsset_{i,t} + \alpha_2 \times FirmCtrl + \alpha_3 \times LocalCtrl + FE(Year/Industry) + \varepsilon_{i,t} \quad (1)$$

where $lnTFP_{i,t}$ represents the total factor productivity of firm *i* in year *t*. The variable $dataAsset_{i,t}$ denotes the data assets owned by firm *i* in year *t*. The model also includes *FirmCtrl*, control variables related to firm-specific characteristics, and

LocalCtrl, control variables related to regional factors. Additionally, fixed effects at the year and industry levels are controlled for.
To examine the specific mechanism through which data assets affect total factor productivity, this paper employs econometric model (2) and tests the mechanism of action based on models (1) and (2).

$$Medium_{i,t} = \beta_0 + \beta_1 \times dataAsset_{j,t} + \beta_2 \times FirmCtrl + \beta_3 \times LocalCtrl + FE(Year/Industry) + \varepsilon_{i,t} \quad (2)$$

where *i* and *t* represent the firm and year, respectively. Medium refers to the mediating variable, which includes $talent_{i,t}$ (the level of human capital) and $lnApply_{i,t}$ (the level of firm innovation). The remaining variables are consistent with those in equation (1). This model is used to test whether the core explanatory variable influences the mediating variable (*Medium*). If the regression coefficient β_1 is significant, it indicates that the core explanatory variable, data assets (*dataAsset*), can impact the mediating variable (*Medium*). On this basis, the theoretical relationship between Medium and the dependent variable (*lnTFP*) is further analyzed.

IV. Analysis of empirical results

4.1Descriptive statistics

Table 2 presents the descriptive statistics for the key variables under study. The findings indicate that the mean and standard deviation of total factor productivity are 8.253 and 1.071, respectively, suggesting significant variations in total factor productivity across firms. Additionally, the mean and standard deviation of enterprise data assets are 22.451 and 1.157, respectively, highlighting substantial differences in the extent of data assets owned by various enterprises. This variability establishes a solid foundation for the regression analysis conducted in this study.

Table2: Description of main variables

	N	mean	sd	min	p25	p50	p75	max
lnTFP	37963	8.253	1.071	5.859	7.524	8.161	8.888	11.104
TFP_OP	37963	6.653	0.903	4.661	6.028	6.559	7.191	9.091
dataAsset	41722	22.451	1.157	20.409	21.618	22.285	23.109	26.025
dataAsset1	41722	1.733	1.297	0.308	0.956	1.338	2.029	7.953
SA	43032	-3.784	0.266	-4.424	-3.959	-3.788	-3.605	-3.066
agencyCost	43343	0.072	0.089	0.000	0.019	0.041	0.086	0.483
lnApply	43889	0.559	1.378	0.000	0.000	0.000	0.000	5.737
Size	43813	22.031	1.357	19.074	21.090	21.861	22.806	26.135
FirmAge	43882	25.212	5.511	12.000	22.000	25.000	30.000	39.000
CashHolding	43813	0.213	0.157	0.013	0.100	0.168	0.284	0.741
Leverage	43875	0.424	0.214	0.051	0.253	0.413	0.579	0.969
Growth	42677	0.383	1.084	-0.753	-0.041	0.123	0.407	7.754
ROA	43875	0.038	0.072	-0.326	0.014	0.040	0.071	0.219
SOE	43031	0.356	0.479	0.000	0.000	0.000	1.000	1.000
lnperGDP	35437	4.187	0.530	2.825	3.827	4.228	4.588	5.215
MI_FG	43871	8.665	2.083	3.370	7.240	8.890	10.290	12.100

4.2 Baseline regression

Table 3 presents the results of estimating the impact of data assets on firms' total factor productivity in the benchmark regression analysis. Column (1) includes only the core explanatory variables, whereas column (2) accounts for year and industry fixed effects based on the specifications in column (1). The results indicate that the regression coefficients for data assets are significantly positive at the 1% level. To examine the heterogeneity across firms and the influence of macro-environmental factors in different regions, firm-level control variables are included in column (3), and regional-level control variables are further added in column (4). The findings reveal that the regression coefficient for data assets remains significantly positive. Ultimately, the results suggest that data assets can significantly enhance total factor productivity, aligning with the initial hypothesis.

Table3: Regression of data asset on TFP

	(1)	(2)	(3)	(4)
	lnTFP	lnTFP	lnTFP	lnTFP
dataAsset	0.680***	0.704***	0.131***	0.131***
	(0.003)	(0.003)	(0.006)	(0.007)
Size			0.498***	0.494***
			(0.006)	(0.007)
FirmAge			-0.000	-0.001
			(0.001)	(0.001)
CashHolding			0.509***	0.503***
			(0.027)	(0.028)
Leverage			0.854***	0.912***
			(0.024)	(0.025)

Growth			-0.029***	-0.025***
			(0.004)	(0.004)
ROA			2.239***	2.082***
			(0.065)	(0.073)
SOE			0.032***	0.042***
			(0.007)	(0.008)
lnperGDP				0.060***
				(0.013)
MI_FG				0.033***
				(0.003)
_cons	-7.061***	-7.562***	-6.396***	-6.669***
	(0.069)	(0.076)	(0.073)	(0.081)
Year/Industry	No	Yes	Yes	Yes
N	37324	37324	37258	30689
adj_R^2	0.532	0.603	0.707	0.721

This table reports ordinary least square (OLS) regression of data assets measured by *dataAsset* on total factor productivity measured by *lnTFP* All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors are reported in parentheses. *, **, *** indicate 0.1, 0.05, and 0.01 level of significance, respectively.

To reveal the impact of data assets at different levels of total factor productivity, this paper conducts a quantile regression in addition to the benchmark regression, with results presented in Table 4. Columns (1) to (4) display the regression results for quartiles 0.2, 0.4, 0.6, and 0.8 sequentially. The results indicate that the regression coefficients for data assets are significantly positive at the 1% level across all quartiles, demonstrating a substantial positive effect of data assets on total factor productivity. Analysis of the regression coefficients for data assets across different quartiles reveals a gradual increase in coefficients with ascending quartiles, stabilizing at higher levels. Specifically, the coefficients for the four quartiles are 0.074, 0.087, 0.109, and 0.108, respectively. These results suggest a structural difference in the impact of data assets on total factor productivity, particularly at the mid- and high quartiles (0.6 and 0.8). This indicates that data assets are particularly effective in enhancing the efficiency of high-productivity firms. This phenomenon may be attributed to the higher technological capabilities, greater innovation, and more efficient resource allocation exhibited by firms with elevated total factor productivity, enabling them to utilize data assets more effectively.

Table4: Quantile Regression of data asset on TFP

	(1) lnTFP	(2) lnTFP	(3) lnTFP	(4) lnTFP
dataAsset	0.074***	0.087***	0.109***	0.108***
	(0.009)	(0.007)	(0.007)	(0.009)
Size	0.536***	0.527***	0.511***	0.507***
	(0.008)	(0.006)	(0.006)	(0.008)
FirmAge	-0.002**	-0.000	0.001	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
CashHolding	0.387***	0.448***	0.499***	0.526***
	(0.036)	(0.029)	(0.030)	(0.038)
Leverage	0.886***	0.954***	0.984***	1.023***
	(0.028)	(0.023)	(0.023)	(0.030)
Growth	-0.017***	-0.020***	-0.022***	-0.026***
	(0.004)	(0.004)	(0.004)	(0.005)
ROA	3.302***	2.625***	2.079***	1.651***
	(0.073)	(0.059)	(0.060)	(0.078)
SOE	0.053***	0.053***	0.039***	0.045***
	(0.010)	(0.008)	(0.008)	(0.011)
lnperGDP	0.046***	0.048***	0.051***	0.025
	(0.017)	(0.014)	(0.014)	(0.018)
MI_FG	0.038***	0.032***	0.028***	0.029***
	(0.004)	(0.003)	(0.003)	(0.004)
_cons	-6.759***	-6.561***	-6.439***	-6.008***
	(0.108)	(0.088)	(0.090)	(0.116)
Year/Industry	Yes	Yes	Yes	Yes
N	30689	30689	30689	30689

4.3 Endogeneity test

4.3.1 Instrumental variables approach

Firms with high total factor productivity (*TFP*) may possess greater capacity and willingness to invest in data assets, resulting in reverse causality between data assets and firms' TFP, which triggers endogeneity issues. To minimize the interference of such endogeneity issues on the estimation results, this paper employs an instrumental variables approach to mitigate potential effects.

We re-estimate the baseline regression using the two-stage least squares method, with the mean value of data assets within the same industry (excluding the enterprise itself) serving as the instrumental variable (Tang et al., 2020). The rationale for this choice is that the mean value of data assets among enterprises in the same industry reflects the level of data asset ownership in that industry; thus, a higher mean indicates a greater presence of data assets, satisfying the correlation requirement of instrumental variables. Additionally, individual enterprise data assets do not have a direct impact on the total factor productivity of the entire industry, thereby satisfying the exogeneity requirement of the instrumental variable.

From the test results, Table 4 column (1) shows the results of the first stage test, the coefficient of instrumental variables is positive and significant at 1% level, which indicates that the instrumental variables selected in this paper do not have the problem of weak instrumental variables, and satisfy the relevance conditions of selected instrumental variables; Table 4 column (2) shows the results of the second stage test, the coefficient of data assets is significantly positive at 1% level, which indicates that after mitigating this kind of endogeneity in the form of reverse causality, the core conclusion of this paper still holds. After mitigating the endogeneity problem of reverse causality, the core conclusion of this paper still holds, i.e., data assets can significantly enhance the total factor productivity of enterprises.

4.3.2 Fixed effects model

Industries and regions exhibit varying trends over time, which may influence the total factor productivity of firms. For instance, certain industries may experience substantial increases in productivity during specific periods due to industry-specific policies and technological advancements, while other industries may not. Likewise, regions may exhibit distinct patterns of total factor productivity over time, influenced by local policies and infrastructure development. Employing a fixed effects model for the main regression can effectively control for these heterogeneities, thereby reducing omitted variable bias.

As presented in Table 4, columns (3) and (4) further incorporate ***Industry*** × ***Year*** fixed effects and ***Region*** × ***Year*** fixed effects, respectively, in addition to controlling for industry and year fixed effects. The regression results in these two columns indicate that the coefficients for data assets are significantly positive at the 1% level, reinforcing the core conclusions of this paper.

Table5: Endogeneity test

	Instrumental variable		Increase fixed effects	
	(1) dataAsset	(2) lnTFP	(3) lnTFP	(4) lnTFP
instrument	0.230^{***}			
	(0.024)			
dataAsset		0.469^{***}	0.074^{***}	0.087^{***}
		(0.158)	(0.009)	(0.007)
Size	0.769^{***}	0.238^{**}	0.536^{***}	0.527^{***}
	(0.003)	(0.120)	(0.008)	(0.006)
FirmAge	0.004^{***}	-0.001	-0.002^{**}	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
CashHolding	0.239^{***}	0.334^{***}	0.387^{***}	0.448^{***}
	(0.022)	(0.083)	(0.036)	(0.029)
Leverage	0.195^{***}	0.848^{***}	0.886^{***}	0.954^{***}
	(0.020)	(0.040)	(0.028)	(0.023)
Growth	0.012^{***}	-0.028^{***}	-0.017^{***}	-0.020^{***}
	(0.003)	(0.004)	(0.004)	(0.004)
ROA	1.533^{***}	1.543^{***}	3.302^{***}	2.625^{***}
	(0.064)	(0.268)	(0.073)	(0.059)
SOE	-0.020^{***}	0.052^{***}	0.053^{***}	0.053^{***}
	(0.007)	(0.009)	(0.010)	(0.008)
lnperGDP	0.109^{***}	0.023	0.046^{***}	0.048^{***}
	(0.011)	(0.022)	(0.017)	(0.014)
MI_FG	-0.008^{***}	0.035^{***}	0.038^{***}	0.032^{***}
	(0.003)	(0.003)	(0.004)	(0.003)
_cons	-0.468	-8.268^{***}	-6.759^{***}	-6.561^{***}
	(0.518)	(0.754)	(0.108)	(0.088)
Year/Industry	Yes	Yes	Yes	Yes
Industry*Year	No	No	Yes	No
Province*Year	No	No	No	Yes
N	34122	30680	30689	30689
adj_R^2	0.816	0.697	0.727	0.729
Kleibergen-Paap rk LM stat		76.79		
Cragg-Donald		79.158		

Wald F stat				

4.4 Robustness tests

4.4.1 Variable and method substitution

(1) Replacement of explanatory variables.

In this paper, total factor productivity is measured using the OP method, and the regression results are presented in column (1) of Table 6. The results indicate that, following the replacement of the total factor productivity measure, the coefficient for data assets remains significantly positive at the 1% level, aligning with the findings of the benchmark regression.

(2) Replacement of explanatory variables.

This paper measures total factor productivity using the OP method, with the regression results displayed in column (1) of Table 5. The findings demonstrate that, after replacing the measure of total factor productivity, the coefficient for data assets continues to be significantly positive at the 1% level, consistent with the benchmark regression results.

(3) Replacement estimation method.

This paper references the study by Tang Song et al. (2020) and employs a Tobit model for re-testing, with the results presented in column (4) of Table 6. The results indicate that the coefficient for data assets remains significantly positive at the 1% level, demonstrating that data assets continue to have a significant impact on total factor productivity.

Table6: robust test

	(1) TFP_OP	(2) lnTFP	(3) lnTFP	(4) lnTFP
dataAsset	0.091^{***}			0.131^{***}
	(0.007)			(0.007)
dataAsset1		0.027^{***}		
		(0.003)		
dataAsset2			1.290^{***}	
			(0.029)	
Size	0.352^{***}	0.604^{***}	0.633^{***}	0.494^{***}
	(0.006)	(0.003)	(0.003)	(0.007)
FirmAge	-0.000	-0.001	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
CashHolding	0.355^{***}	0.545^{***}	0.304^{***}	0.503^{***}
	(0.028)	(0.028)	(0.028)	(0.028)
Leverage	0.746^{***}	0.930^{***}	0.859^{***}	0.912^{***}
	(0.024)	(0.025)	(0.025)	(0.025)
Growth	-0.007^{*}	-0.024^{***}	-0.037^{***}	-0.025^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
ROA	1.837^{***}	2.206^{***}	1.999^{***}	2.082^{***}
	(0.068)	(0.074)	(0.069)	(0.073)
SOE	0.039^{***}	0.037^{***}	0.066^{***}	0.042^{***}
	(0.008)	(0.008)	(0.008)	(0.008)
lnperGDP	0.168^{***}	0.070^{***}	0.037^{***}	0.060^{***}
	(0.013)	(0.013)	(0.013)	(0.013)
MI_FG	0.011^{***}	0.033^{***}	0.029^{***}	0.033^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
_cons	-4.548^{***}	-6.294^{***}	-7.639^{***}	-6.669^{***}
	(0.080)	(0.080)	(0.083)	(0.081)
Year/Industry	Yes	Yes	Yes	Yes
N	30689	30689	30689	30689
adj_R^2/ Pseudo_R^2	0.619	0.719	0.738	0.436

(4) Replacement samples

To mitigate the influence of specific events, industries, and regions on the estimation results, this paper excludes certain samples and re-estimates the model. First, the analysis excludes the effects of the 2010 financial crisis and the 2015 stock market crash, with the regression results presented in Table 7, Column (1). Second, this paper excludes samples from key digital economy industries, which typically involve significant data assets; the regression results are shown in Column (2). Lastly, considering the unique economic characteristics of China's municipalities, this paper excludes samples from these regions, with the regression results displayed in Column (3). The results indicate that the coefficients for data assets remain significantly positive at the 1% level following the sample exclusions, consistent with the conclusions of the baseline regression.

Table7: robust test

	(1) lnTFP	(2) lnTFP	(3) lnTFP

dataAsset	0.136***	0.128***	0.137***
	(0.008)	(0.007)	(0.008)
Size	0.484***	0.499***	0.506***
	(0.007)	(0.007)	(0.007)
FirmAge	-0.000	-0.002**	-0.003***
	(0.001)	(0.001)	(0.001)
CashHolding	0.436***	0.560***	0.496***
	(0.030)	(0.032)	(0.031)
Leverage	0.926***	0.848***	0.824***
	(0.028)	(0.027)	(0.027)
Growth	-0.024***	-0.023***	-0.030***
	(0.004)	(0.004)	(0.005)
ROA	2.071***	2.145***	1.945***
	(0.081)	(0.081)	(0.081)
SOE	0.027***	0.045***	0.053***
	(0.009)	(0.009)	(0.009)
lnperGDP	0.062***	0.069***	0.058***
	(0.015)	(0.014)	(0.019)
MI_FG	0.029***	0.031***	0.035***
	(0.003)	(0.003)	(0.004)
_cons	-6.622***	-6.673***	-6.973***
	(0.093)	(0.087)	(0.101)
Year/Industry	Yes	Yes	Yes
N	25038	25929	24601
r2_a	0.724	0.730	0.714

4.5Mechanism testing

The preceding theory posits that data assets can enhance an enterprise's total factor productivity by improving its human capital level and promoting innovation. To verify this mechanism, this paper employs a mediation effect model for empirical testing, with the regression results presented in Table 8. Column (1) presents the regression results demonstrating that data assets significantly enhance the human capital level of enterprises, with the coefficient being positive at the 1% level. This enhancement subsequently increases the total factor productivity of the enterprises. Column (2) displays the regression results concerning the promotion of enterprise innovation by data assets. The regression results indicate that the coefficient for data assets is significantly positive at the 1% level, demonstrating that data assets can significantly promote enterprise innovation. This innovation-driven effect further enhances the total factor productivity of the enterprises.

Table8: mediation analysis

	(1) talent	(2) lnApply
dataAsset	0.044***	0.079***
	(0.002)	(0.015)
Size	-0.040***	0.074***
	(0.002)	(0.013)
FirmAge	-0.002***	-0.016***
	(0.000)	(0.002)
CashHolding	0.059***	0.505***
	(0.008)	(0.057)
Leverage	-0.076***	-0.096**
	(0.006)	(0.044)
Growth	0.013***	0.028***
	(0.001)	(0.006)
ROA	-0.160***	0.092
	(0.016)	(0.115)
SOE	-0.003	0.118***
	(0.002)	(0.018)
lnperGDP	0.026***	0.106***
	(0.004)	(0.028)
MI_FG	0.001*	0.016**
	(0.001)	(0.007)
_cons	-0.034	-3.356***
	(0.034)	(0.200)
Year/Industry	Yes	Yes

N	19214	34132
r2_a	0.341	0.063

4.6 Heterogeneity test

To further investigate the heterogeneous impact of data assets on total factor productivity based on various firm micro characteristics and external macro environments, this paper aims to establish differentiated policy orientations. This paper conducts a heterogeneity analysis from four perspectives: firm size, property rights, regional digitization level, and economic policy uncertainty.

4.6.1 Heterogeneity test based on micro characteristics of enterprises

We conduct group regression tests based on the micro characteristics of enterprises, specifically focusing on enterprise size and the nature of property rights. First, the sample is divided into large-scale and small-scale enterprises by constructing dummy variables based on whether their size exceeds the industry median. Second, the sample is categorized into state-owned and non-state-owned enterprises based on the nature of their property rights. The regression results are presented in Table 8. Columns (1) and (2) present the test results for small-scale and large-scale enterprises, respectively. Data assets significantly affect the total factor productivity of enterprises of varying sizes, with a more pronounced effect observed in large-scale enterprises. This result suggests that data assets exhibit a scale effect, indicating that larger enterprises are more capable of utilizing data assets to enhance their total factor productivity. Columns (3) and (4) present the test results for non-state-owned and state-owned enterprises, respectively. Data assets significantly impact the total factor productivity of enterprises with varying property rights, with a greater effect observed in state-owned enterprises. This result may be attributed to state-owned enterprises' advantages in capital, scale, scientific research, and policy (Li Zheng and Lu Yinhong, 2014), enabling them to more effectively acquire, manage, and utilize diverse data assets to enhance their total factor productivity.

4.6.2 Heterogeneity test based on the external macro environment

We conduct group regression tests based on the external macro-environment in which enterprises operate, specifically focusing on regional digitization levels and economic policy uncertainty. First, dummy variables are constructed based on whether the digitization level of different regions exceeds the median; a level above the median indicates high digitization, while a level below indicates low digitization. Second, dummy variables are constructed based on whether the economic policy uncertainty index in different regions exceeds the median; levels above the median indicate high economic policy uncertainty, while those below indicate low uncertainty. The regression results are presented in Table 7. Columns (5) and (6) present the test results for firms in regions with low and high digitization levels, respectively. Data assets significantly impact firms' total factor productivity across regions, with a slightly greater effect observed in regions characterized by high digitization levels. This is attributed to the more developed digital economy infrastructure in high-digitization regions, which facilitates the accumulation and application of data assets by firms, thereby leading to a more substantial increase in total factor productivity. Columns (7) and (8) present the test results for firms in regions with low and high economic policy uncertainty, respectively. Data assets exert a greater influence on firms' total factor productivity in regions characterized by high economic policy uncertainty. This result may be explained by the fact that in environments of high economic policy uncertainty, firms tend to make more cautious decisions. Data assets provide a richer information base for decision-making, thereby reducing information asymmetry (Lukasz et al., 2016) and enabling firms to better navigate uncertainty.

Table9: Heterogeneity test

	(1) lnTFP	(2) lnTFP	(3) lnTFP	(4) lnTFP	(5) lnTFP	(6) lnTFP	(7) lnTFP	(8) lnTFP
dataAsset	0.115***	0.154***	0.085***	0.167***	0.126***	0.129***	0.127***	0.137***
	(0.012)	(0.009)	(0.009)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)
Size	0.534***	0.454***	0.530***	0.472***	0.503***	0.480***	0.504***	0.477***
	(0.013)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
FirmAge	-0.002	0.000	-0.003***	0.005***	0.000	-0.001	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CashHolding	0.399***	0.639***	0.353***	0.801***	0.482***	0.504***	0.477***	0.559***
	(0.039)	(0.042)	(0.033)	(0.053)	(0.039)	(0.041)	(0.038)	(0.041)
Leverage	0.852***	0.964***	0.929***	0.822***	0.778***	1.080***	0.895***	0.985***
	(0.036)	(0.035)	(0.033)	(0.041)	(0.034)	(0.038)	(0.034)	(0.039)
Growth	-0.026***	-0.024***	-0.036***	-0.017***	-0.031***	-0.020***	-0.024***	-0.027***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
ROA	1.911***	2.172***	2.101***	2.237***	2.093***	2.092***	2.637***	1.723***
	(0.092)	(0.118)	(0.088)	(0.129)	(0.103)	(0.103)	(0.103)	(0.098)
SOE	0.033**	0.049***			0.070***	0.028**	0.085***	-0.008
	(0.013)	(0.010)			(0.010)	(0.012)	(0.010)	(0.013)
lnperGDP	0.109***	0.022	0.053***	0.061***	0.055***	0.057***	0.029	0.062***
	(0.020)	(0.017)	(0.017)	(0.020)	(0.018)	(0.019)	(0.018)	(0.020)
MI_FG	0.031***	0.034***	0.039***	0.024***	0.036***	0.027***	0.050***	0.020***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
_cons	-7.194***	-6.261***	-6.428***	-6.986***	-6.678***	-6.475***	-6.772***	-6.502***

	(0.242)	(0.120)	(0.119)	(0.119)	(0.112)	(0.143)	(0.110)	(0.134)
Year/Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	13899	16790	18251	12438	16915	13774	17822	12867
r2_a	0.453	0.649	0.689	0.736	0.702	0.737	0.723	0.718

V. Conclusions and discussion

This paper provides a theoretical analysis and empirical testing of the impact of data assets on enterprise total factor productivity, exploring the underlying mechanisms and examining heterogeneity. The main conclusions are as follows: (1) Data assets significantly enhance enterprise total factor productivity, and this finding remains robust even after addressing endogeneity and conducting robustness tests; (2) Data assets enhance enterprise total factor productivity through two mechanisms: improving the level of human capital and fostering innovation; (3) The effect of data assets on enterprise total factor productivity exhibits heterogeneity across different enterprise and regional characteristics. Notably, this enhancement is particularly pronounced in large-scale enterprises, state-owned enterprises, firms in regions with higher digitization levels, and those in regions with greater economic policy uncertainty.

Based on the conclusions drawn from the preceding research, this paper presents the following recommendations: (1) To achieve the widespread adoption and utilization of data assets, it is essential to expand the trading market and broaden application prospects. The government should establish a comprehensive system for data asset rights, value assessment, accounting, auditing, and taxation. Additionally, a standardized data trading platform should be developed, along with improvements in the trading model and the establishment of a robust legal framework to facilitate the smooth circulation and trading of data assets. (2) Data assets are a crucial product of digital transformation. In this context, many enterprises, particularly small-scale and non-state-owned enterprises, encounter challenges related to concepts, systems, management, technology, and talent acquisition. The government should support and encourage small-scale and non-state-owned enterprises to expedite their digital transformation. This can be achieved by enhancing the acquisition, management, and application of data assets through various initiatives, including policy support, financial assistance, technical guidance, and training programs. (3) Digital infrastructure serves as the foundation for the accumulation and utilization of data assets. The government should support and guide regions with low levels of digitization to enhance the development of digital infrastructure. Concurrently, it should encourage enterprises to increase investments in data assets and participate in the construction of digital management platforms, thereby providing essential support for enterprises to fully explore and utilize these assets. (4) With the rapid development of the digital economy, data assets have emerged as a vital strategic resource for enterprises. Enterprises should acknowledge the significance of data assets and concentrate on the development and utilization of these resources. They must strengthen their digital infrastructure, optimize processes and systems, promote the efficient flow of data, foster cooperation and sharing, and further expand the application areas and profit models of data resources. Additionally, they should explore data resource realization methods tailored to their unique characteristics and enhance their ability to manage uncertainty. Moreover, the management and application of data assets must align with the prevailing technological standards. Enterprises should strengthen the cultivation and recruitment of data-related technical talent, enhancing their capabilities in mining, managing, and applying data assets. This will maximize the value of these assets and drive innovation and development within the organization.

REFERENCES

[1].Amable, B., Ledezma, I., & Robin, S. (2016). Product market regulation, innovation, and productivity. Research Policy, 45(10), 2087-2104.
[2].Baier, S. L., Dwyer, G. P., & Tamura, R. (2010). How important are capital and total factor productivity for economic growth? Economic Inquiry, 44(1), 23-49.
[3].Akcigit, U., & Liu, Q. (2016). The role of information in innovation and competition. Journal of the European Economic Association, 14(4), 828-870.
[4].Acquisti, A., et al. (2016). The economics of privacy. Journal of Economic Literature, 54(2), 442-492.
[5].Ackerberg, D., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. Econometrica, 83(6), 2411-2451.
[6].Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. Journal of Monetary Economics, 97, 71-87.
[7].Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. McKinsey Quarterly, 56(1), 75-86.
[8].Chen, G., Zeng, D., Wei, Q., & Others. (2020). Paradigm shifts and enabling innovation in decision-making under big data. Management World, 36(2), 95-105.
[9].Cheng, X., & Song, C. (2023). Futures investment, financing constraints, and total factor productivity of firms. Jianghuai Forum, 2023(5), 39-48.
[10].Gunasekara, A., Papadopoulos, T., & Dubey, R. (2017). Big data and predictive analytics for supply chain and organizational performance. Journal of Business Research, 70, 308-317.
[11].Hua, J., Liu, C., & Zhu, D. (2022). Digital transformation, financing constraints, and total factor productivity of firms. Southern Finance, 2022(7), 54-65.

[12].He, L. (2022). Data factor marketization reform, entrepreneurship, and digital transformation of manufacturing. Journal of Hunan University of Science and Technology (Social Science Edition), 25(6), 65-76.
[13].Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. American Economic Review, 110(9), 2819-2858.
[14].Jones, C. I., & Tonetti, C. T. (2020). Nonrivalry and the economics of data. NBER Working Paper, No. 26260.
[15].Kshetri, N. (2014). Big data impact on privacy, security, and consumer welfare. Telecommunications Policy, 38(11), 1134-1145.
[16].Kubina, M., et al. (2015). Use of big data for competitive advantage of companies. Procedia Economics and Finance, 26(1), 561-565.
[17].Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. The Review of Economic Studies, 70(2), 317-341.
[18].Li, Z., & Lu, Y. (2014). Do state-owned enterprises really lack innovation capability? An empirical analysis and comparison of the ownership nature and innovation performance of listed companies. Economic Theory and Business Management, 2014(2).
[19].Li, H., & Zhao, L. (2021). Data as a production factor: Characteristics, mechanisms, and evolution of value forms. Shanghai Economic Research, 2021(08), 48-59.
[20].Liu, Y. (2014). A brief discussion on the recognition and measurement of big data assets. Commercial Accounting, 2014(18), 3-4.
[21].Liu, W., & Ni, H. (2018). Servitization of manufacturing inputs and firms' technological progress: Effects and mechanisms. Finance and Trade Economics, 2018(8).
[22].Liu, G. (2024). The value creation logic and path of data assets enabling high-quality enterprise development. Finance & Accounting Communications, 2024(4), 3-12.
[23].Lin, Y., & Zhang, Y. (2007). The relationship between information transparency and firm performance. Accounting Research, 2007(9), 26-34+95.
[24].Lukasz, O., Stanislaw, S., & Zbigniew, T. (2016). Leverage big data analytics for dynamic informed decisions with advanced case management. Journal of Decision Systems, 25(1).
[25].Lu, Z., Zhou, T., Wang, L., & Liao, Z. (2023). Data assets and enterprise development: Empirical evidence from Chinese listed companies. Industrial Economic Research, 2023(4), 128-142.
[26].Mohnen, P., & Hall, B. H. (2013). Innovation and productivity: An update. Eurasian Business Review, 3(1), 47-65.
[27].Nolin, J. M. (2020). Data as oil, infrastructure, or asset? Three metaphors of data as economic value. Journal of Information, Communication and Ethics in Society, 18(1), 28-43.
[28].Peters, R. E. (1974). A cross-section study of the demand for money: The United States, 1960-62. The Journal of Finance, 1974(1), 73-88.
[29].Shen, G., & Yuan, Z. (2020). The impact of internet adoption on innovation and exports of Chinese firms. Economic Research Journal, 55(1), 33-48.
[30].Sun, Y., & Chen, S. (2021). Data assets and the high-quality development of science and technology service enterprises: A study based on the "Broadband China" quasi-natural experiment. Wuhan University Journal (Philosophy and Social Sciences Edition), 74(5), 132-147.
[31].Sun, X., & Wang, Y. (2014). R&D investment and firm productivity: A PSM analysis based on Chinese industrial firm microdata. Science Research Management, 2014(11), 92-99.
[32].Sun, X., Qian, Y., Zhang, M., & Li, J. (2019). Mechanisms of big data-driven agility in supply chains. Management World, 2019(9), 133-151.
[33].Tang, Y., Ma, W., & Xia, L. (2021). Environmental information disclosure quality, internal control "level", and firm value: Empirical evidence from listed companies in heavily polluting industries. Accounting Research, 2021(7), 69-84.
[34].Tang, S., Wu, X., & Zhu, J. (2020). Digital finance and firms' technological innovation: Structural characteristics, mechanism identification, and effects under financial regulation. Management World, 36(5), 52-66.
[35].Teixeira, A. A. C., & Fortuna, N. (2010). Human capital, R&D, trade, and long-run productivity: Testing the technological absorption hypothesis for the Portuguese economy, 1960–2001. Research Policy, 39(3), 335-350.
[36].Tu, X., & Yan, X. (2022). Digital transformation, knowledge spillover, and total factor productivity of firms: Empirical evidence from listed manufacturing firms. Industrial Economic Research, 2022(2), 43-56.
[37].Varian, H. (2018). Artificial intelligence, economics, and industrial organization. NBER Working Paper, No. 24839.
[38].Veldkamp, L., & Chung, C. (2019). Data and the aggregate economy. Journal of Economic Literature, forthcoming.
[39].Von Hippel, E. (1986). Lead users: A source of novel product concepts. Management Science, 32(7), 791-805.
[40].Xiao, J. (2020). Cross-system digital transformation and management adaptation reform of firms. Reform, 2020(4), 37-49.
[41].Xu, X., Li, K., & Tian, X. (2021). Research progress on data as a production factor. Journal of Economic Dynamics, 2021(4), 142-158.
[42].Xu, X., & Zhao, M. (2020). Data capital and economic growth paths. Economic Research Journal, 55(10), 38-54.
[43].Ye, J., & Lin, J. (2016). Innovation and total factor productivity of firms: Evidence from Chinese manufacturing firms. Learning and Exploration, 2016(5), 105-111+160.
[44].Yuan, Z., Yu, X., & Li, M. (2022). Data asset disclosure, institutional investor heterogeneity, and firm value. Modern Finance, 2022(11), 32-47.

[45].Zhao, C., Wang, W., & Li, X. (2021). How does digital transformation affect firms' total factor productivity? Finance and Trade Economics, 2021(7), 114-129.
[46].Zhang, Y., & Guo, X. (2023). Big data development and total factor productivity of firms: An empirical analysis based on national big data comprehensive pilot zones. Industrial Economic Research, 2023(2), 69-82.
[47].Zhang, J., Wei, Y., & Song, X. (2020). Research on the accounting treatment and information disclosure of enterprise data assets. Accounting and Economic Research, 2020(3), 3-15.
[48].Zhu, Y., & Ye, Y. (2018). Data assets from the perspective of data properties. Forum, 2018(2), 65-76.