# Data-driven, Digital Transformation and Enterprise Total Factor Productivity — Empirical Evidence from A-share Listed Companies in China

**Ming Chen, Zhifan Wang**

School of Economics, Management and Law, University of South China, Hengyang 421001, Hunan, China

**Abstracts:** With the continuous development of the digital economy, economies around the world have taken the promotion of big data as an important driver for high-quality economic development. This paper empirically examines the role and influence mechanism of big data-driven enterprise total factor productivity using a multi-period double-difference model with exogenous policy shocks of the National Comprehensive Pilot Zone on Big Data set up in China, with the help of enterprise data of China's A-share listed companies from 2010 to 2020. The results of the study show that big data development can significantly increase firms' total factor productivity, which still holds after a series of robustness tests such as Machine Learning, and is more pronounced in larger firms. Promoting the digital transformation of enterprises is an important mechanism of action for the development of big data to promote the total factor productivity of enterprises. This paper enriches the mechanism of the role of big data in driving the total factor productivity of enterprises, and provides empirical references for the regional emphasis on the development of big data and the transformation of enterprise databasing.

*Keywords*: big data, digital transformation, TFP, DID, machine learning

## 1. Introduction

As the digital economy continues to grow, data-driven is receiving attention from major economies around the world, which have introduced a series of policies to support regional big data development. Take China for example, which has issued documents such as The Outline of Action for Promoting the Development of Big Data and The Plan for The Development of Big Data Industry (2016-2020) since 2015 to support the development of big data in the country. According to the White Paper on the Global Digital Economy published by the Chinese Academy of Communications. The digital economy of the world's major economies is up to $41.2 trillion in 2022, accounting for 46.1% of GDP, Digital elements, as the digital economy and the core development engine, have become an important driver of economic growth. Promoting enterprise total factor productivity improvement is the basis for improving the total factor productivity of the national economy of the country, and it is the inevitable way for the world economy to move towards high-quality development, so exploring how to improve the total factor productivity improvement of enterprises through the development of big data has an important theoretical and practical value.

From the perspective of current research, the current research on total factor productivity of enterprises is mainly centered on macroeconomic policies[1] (Kaihao Liu, 2024), property rights protection[2] (Xicang Zhao, 2023), R&D investment, and the Digital Economy[3] (Delong Hu, 2023), It can be seen that the development of the digital economy is a powerful driving force for the improvement of business effectiveness. As for the development of Big Data as the core of the digital economy, the current relevant research mainly focuses on the innovation effect[4] (Lin Xu, 2022), the industrial ecological effect[5] (Wei Lili et al., 2022) and the carbon emission reduction effect [6] (Chang Haoliang et al., 2023) it exerts, and so on. As for the enhancement of enterprises, only with Yihao Zhang[7] (2023) and others explored the role of big data development in enhancing the total factor productivity of enterprises from the perspectives of enhancing enterprise innovation and improving operational efficiency. However, the development of big data has another important purpose: to promote the full digital transformation of more enterprises, so as to cope with the complex changes in the market environment and to ensure the stable functioning of the regional economic system. In this regard, our concern about whether the development of big data promotes the digital transformation of enterprises and thus contributes to the improvement of total factor productivity of enterprises is not reflected here. In addition, the current research methodology here mainly utilizes fixed panel models, which face problems such as artificially set model bias, for which their scientific validity needs to be verified. Based on this, this paper empirically examines the role of big data in driving the total factor productivity

of enterprises and the influence mechanism of big data by using a multi-period double-difference model with the data of A-share listed companies in the Chinese region from 2010 to 2020, taking the big data comprehensive experimental zone set up in China as an exogenous policy shock.

The possible marginal contribution of this paper is: Firstly, as far as the research perspective is concerned, this paper enriches the mechanism of the role of big data in driving the total factor productivity improvement of enterprises from the perspective of promoting the digital transformation of enterprises. Secondly, from the empirical method, the traditional econometric model faces problems such as the bias in the results caused by the artificial setting model, and this paper excludes these disturbances through the dual machine learning model in the robustness proposal to ensure the reliability of the empirical results in this paper.

## 2. Theoretical mechanisms and research hypotheses

The National Big Data Comprehensive Pilot Zone is an important initiative of China to promote regional big data development and accelerate the construction of a digital power, aiming to promote the development of digital industry through big data and give full play to the driving role of digital elements in promoting the high-quality development of regional economy. On the one hand, from the point of view of enterprise management, big data makes the information more transparent and more liquid, providing a rich basis for enterprise decision-making. The level of corporate decision-making depends not only on aspects such as the competence of decision makers, but also on the stock of information available. The development of big data in the region will continue to enhance the stock of valid information in the market, improve the quality of business decisions, promote the efficiency of business operations, and thus increase total factor productivity. On the other hand, big data reduces the degree of resource mismatch in the market, and the increase in the stock of effective information in the market can make the resources flow to the demand side, reduce the waste of resources, improve the utilization rate of resources, and reduce the cost of production for enterprises. In addition, data elements can replace the traditional inefficient labor production methods, and the aggregation of digital elements brought by big data renews the production methods of enterprises and improves the total factor productivity of enterprises.

Improved big data capabilities help enterprises to digitally transform themselves and thus increase their total factor productivity. On the one hand, regions with a high level of big data will attract more diversified enterprises to move in, forcing local enterprises to carry out digital transformation to improve their core competitiveness, in addition, big data can incubate more digital industries, attracting more digital technology talents to enhance the human capital of enterprises, and accelerating the process of enterprise digitization[8] (Yutang Shi, 2023). On the other hand, the degree of enterprise digital transformation can reduce the enterprise financing constraints, Internet information technology to break the previous information barriers, broaden the enterprise financing channels, and big data as the core of the application of digital technology to enhance the market information search capabilities, so that the supply and demand for funds to accurately match the two sides to reduce the cost of financing, and to promote enterprise production. In addition, an increase in the degree of digital transformation of enterprises can promote the development of innovation within the enterprise, and the application of new technologies in production to enhance the total factor productivity of enterprises.

In summary this paper leads to the following hypothesis:

Hypothesis 1: Big Data-Driven can promote total factor productivity of enterprises.

Hypothesis 2: Big Data-Driven can increase enterprise total factor productivity by increasing the degree of digital transformation of the enterprise, which in turn increases enterprise total factor productivity.

## 3. Empirical testing

### 3.1 Empirical modeling

The purpose of this paper is to explore the big data drive can improve the total factor productivity of enterprises, so this paper constructs the following multi-period DID model:

$$TFP_{it} = \alpha_0 + \alpha_1 Data_{it} + \sum \alpha_j X_{it} + \mu_i + \nu_t + \varepsilon_{it} \tag{1}$$

In which: $TFP_{it}$ is the explanatory variable, representing the total factor productivity of the first enterprise in the first year, using the number of enterprise employees as the labor input, net fixed assets as the capital input, the enterprise electricity consumption as the energy input, the enterprise business revenue as the desired output, the industrial waste generated by the enterprise as the non-desired output, using the LP method to synthesize; $Data_{it}$ is the policy dummy variable of the National

Comprehensive Pilot Zone of Big Data, the enterprise is in the pilot region after assigning a value of 1, and the rest are 0; $X_{it}$ is the aggregate generated by control variables, specifically: Enterprise Age, Return on Assets, Tobin's Q value, Equity Concentration, Asset Growth Rate, and Asset Turnover; $\mu$ is the individual fixed effect; $\nu$ is the time fixed effect; and $\varepsilon$ is the random error.

In addition, the premise of applying the double difference method needs to pass the parallel trend test, by setting the dynamic effect model to test the results are as follows, that is, there is no significant difference between the experimental group and the control group of the city in the pre-policy period, and the difference after the implementation of the policy has increased significantly, and the parallel trend test is passed. (Figure 1)
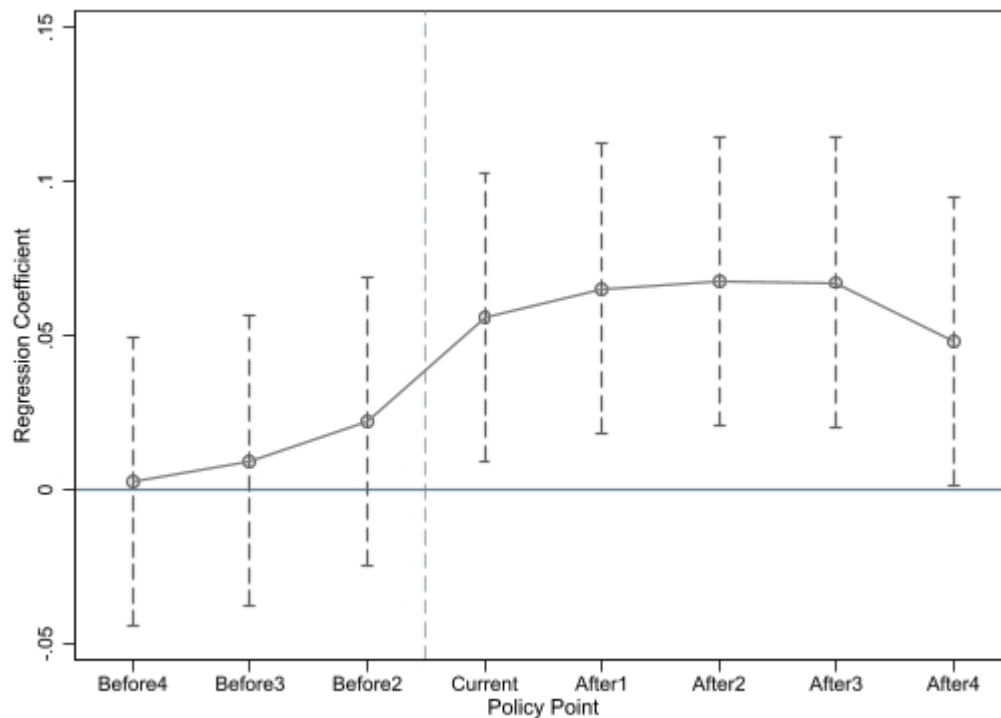


Figure 1. Parallel Trend Test Plot

## 3.2 Empirical results

This paper analyzes the role of big data development in affecting firms' total factor productivity using a multi-period DID model, whose empirical results are shown in columns (1)-(3) of Table 1. The results are significant at the 1% level after gradually adding control variables, individual fixed effects and time fixed effects. In order to exclude the bias caused by the artificial setting of the model, this paper draws on the research ideas of Chernozhukov[9], the use of dual machine learning model to re-test the results, its sample split ratio of 1:4, using the random forest algorithm to solve the model, the regression results as shown in Table 1, column (6), the results are still significantly positive at the 5% level, which can be concluded that the big data-driven significantly enhance the enterprise's total factor productivity, Hypothesis 1 has been verified.

In order to verify whether the big data drive enhances enterprise total factor productivity by promoting enterprise digital transformation, this paper refers to the mechanism test idea of Ting Jiang[10] (2022), and conducts a regression after replacing the explanatory variables of the formula (1) species with the degree of enterprise digital transformation, in which the degree of enterprise digital transformation utilizes python's text analysis technology to analyze the annual reports of the enterprises about digital transformation. content of the enterprise annual report about digital transformation using python text analysis technology for word frequency processing, so as to synthesize the enterprise digital transformation index ( $DIGI$ ). The regression results are shown in Table 1, Column 4. From the results, it can be seen that the development of big data significantly improves the degree of enterprise digital transformation, which promotes the improvement of enterprise total factor productivity, and hypothesis 2 is verified.

Considering the moderating effect of enterprise size in big data-driven enterprise total factor productivity improvement,

this paper constructs a dummy variable for enterprise size ( *Size* ), Large-scale enterprises are assigned a value of 1, while small and medium-scale enterprises are assigned a value of 0. Then the cross-multiplier term ( $Size \times DID$ ) of the dummy variable of enterprise size and big data test area is regressed as the core explanatory variable of Model 1, and the regression results are shown in Column 5 of Table 1, which shows that larger enterprises are more capable of seizing opportunities in the areas where big data are developed to enhance their total factor productivity through the advantage of scale.

**Table 1. Table of empirical results**

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
|  | *TFP* | *TFP* | *TFP* | *DIGI* | *TFP* | *TFP* |
| *Data* | 0.405*** (0.022) | 0.129*** (0.012) | 0.044*** (0.012) | 0.324*** (0.067) |  | 0.307** (0.152) |
| $Size \times DID$ |  |  |  |  | 0.124*** (0.014) |  |
| *X* | No | Yes | Yes | Yes | Yes | Yes |
| *Firm* | No | No | Yes | Yes | Yes | Yes |
| *Year* | No | No | Yes | Yes | Yes | Yes |
| *N* | 9295 | 9295 | 9295 | 9287 | 9295 | 9295 |

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

## 4. Conclusions of the study

This paper empirically examines the role and influence mechanism of big data-driven enterprise total factor productivity using a multi-period double-difference model with the help of enterprise data of Chinese A-share listed companies from 2010 to 2020, taking the National Big Data Comprehensive Pilot Zone set up in China as an exogenous policy shock. The results show that big data development can significantly improve the total factor productivity of enterprises, which still holds after a series of robustness tests such as machine learning, and is more obvious in larger companies. Promoting enterprise digital transformation is an important mechanism for big data development to promote enterprise total factor productivity. In summary this paper makes the following policy recommendations.

The region should pay attention to the big data-driven capability, the region should continue to improve the digital infrastructure to promote the development of big data, encourage local enterprises to carry out digital innovation, the government should be guided by the relevant policies, and enterprises to jointly promote the development of the local digital industry, to attract more digital technology-based talents to provide human support for the enterprise, and to empower the enterprise digital transformation so as to enhance the total factor productivity of the enterprise.

## Acknowledgments

## References

[1] Kaihao Liu, Yiran Wang, Yunfei Hao. How does the tax reduction policy affect the total factor productivity of enterprises? — Empirical evidence from the "two-tax merger"[J/OL]. Modern Finance and Economics,2024(02):99-113. (in Chinese)

[2] Xicang Zhao, Liang Xu, Yusen Luo. Does Intellectual Property Protection Enhance Total Factor Productivity of Enterprises? — Evidence from Chinese A-share Listed Companies[J]. *Journal of Jiangsu University (Social Science Edition)*,2023,25(06):68-83. (in Chinese)

[3] Delong HU, Manzhen SHI. Research on the impact of digital economy on enterprise total factor productivity[J]. *Contemporary Finance and Economics*,2023(12):17-29. (in Chinese)

[4] Lin XU, Linqi HOU, Guangbin CHENG. Research on the Innovation Effect of State-level Big Data Comprehensive Pilot Zone[J]. *Science and Technology Progress and Countermeasures*,2022,39(20):101-111.

[5] Lili WEI, Hongyan XIU, Yuqi HOU. Research on the impact of digital economy on the ecologization of urban industry — a quasi-natural experiment based on the establishment of national-level big data comprehensive pilot zone[J]. *Urban*

*Issues*,2022(11):34-42. (in Chinese)

[6] Haoliang CHANG, Bei JIN, Fei XUE. Impact of big data strategy on carbon emissions from electricity consumption — a quasi-natural experiment based on a national-level big data comprehensive pilot zone[J]. *Economic and Management Research*,2023,44(05):93-109. (in Chinese)

[7] Yihao Zhang, Xiaohui Guo. Big data development and enterprise total factor productivity--an empirical analysis based on the national-level big data comprehensive pilot zone[J]. *Industrial Economics Research*,2023(02):69-82. (in Chinese)

[8] Yutang Shi, Xiaodan Wang. Can the Establishment of Comprehensive Big Data Pilot Zone Drive Digital Transformation of Enterprises? — An empirical study based on a quasi-natural experiment[J/OL]. Research in Science:1-18. (in Chinese)

[9] Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters [J]. *The Econometrics Journal*, 2018, 21(1):C1-C68.

[10] Ting Jiang. Mediating and moderating effects in empirical studies of causal inference[J]. *China Industrial Economy*,2022(05):100-120. (in Chinese)