

The Impact of Data Assetization on the Innovation Efficiency of Manufacturing Enterprises

Miao Wang

Department of Economics, Northwest Normal University, Lanzhou, Gansu, China

Abstract: In recent years, as data has been recognized as a new type of production factor, how to empower manufacturing industry innovation has become a key issue. Based on the data of A-share manufacturing companies from 2015 to 2023, this paper uses text mining technology to construct indicators for data assetization of manufacturing enterprises, employs the stochastic frontier analysis method to measure innovation efficiency, and builds a theoretical framework of "data capital accumulation - knowledge reorganization - dynamic efficiency optimization". It systematically examines the impact path and heterogeneity characteristics of data assetization on the innovation efficiency of manufacturing enterprises. The research findings are as follows: First, data assetization significantly improves the innovation efficiency of manufacturing enterprises, and this conclusion still holds after controlling for endogeneity and multiple robustness tests; Second, the industry synergy effect is prominent - the data transformation of enterprises is highly dependent on the industry environment, and the overall data level of the industry has a positive spillover effect on the individual innovation efficiency; Third, the effect shows structural differentiation: large-scale enterprises benefit more significantly due to their resource endowment advantages, the response intensity of non-state-owned enterprises is higher than that of state-owned enterprises, and the improvement effect of non-key cities enterprises is better than that of key cities; At the regional level, it presents a gradient feature of "central region > eastern region > western region > northeastern region"; Fourth, the mechanism test shows that data assetization affects the innovation efficiency of enterprises by enhancing the overall value creation ability of enterprises, improving the accuracy of R&D investment, and stimulating the synergy effect between data and human

resources and capital. Based on this, it is suggested to establish a classification certification system for enterprise data capabilities to guide differentiated transformation paths, implement regional adaptation strategies to break through institutional transformation barriers, and improve the valuation and circulation infrastructure of data assets, providing systematic support for the "digital-real integration" of manufacturing.

Keywords: Data Assetization; Innovation Efficiency; Text Analysis; Stochastic Frontier Analysis; Simultaneous Equation Model

1. Introduction

With the in-depth development of the new round of technological revolution and industrial transformation, the global digitalization wave is accelerating, and data has become a key production factor driving economic and social development. However, China's data resources generally face the problem of "high quantity and low quality", and most enterprises have not yet established a complete data governance system. There are also significant obstacles in data ownership confirmation, assessment, and compliance. In this context, how to effectively promote the transition from "data resource accumulation" to "data value creation" - the process of data assetization - becomes particularly important for enterprises. Although data assetization has attracted widespread attention, its research system is still in its early stages and is not yet mature, indicating that this field has significant research potential and development prospects [1]. Most current research on data assetization mainly focuses on theoretical analysis, mainly including accounting confirmation and value assessment of data assets [2], theoretical models [3], and implementation paths. Existing research mainly uses text analysis methods to quantify the degree of data assetization, extracting key words' frequency

from enterprise annual reports through constructed text dictionaries. The construction of text dictionaries mainly includes two types: one is to select core seed words directly related to data assetization, such as "information", "network", "digital", "data", etc., and combine the Word2Vec model to generate similar word sets [4, 5]; the other is to construct text dictionaries according to the formation stage and management framework of enterprise data assets [6].

The increasingly valued data assets have had a significant impact on enterprise innovation. As the core driving force for the continuous development of enterprises, innovation efficiency has always been a hot topic of scholars' research. External conditions are the main factors affecting enterprise innovation. Existing research mainly focuses on financing constraints [7], corporate ESG performance [8], digital inclusive finance [9], artificial intelligence [10], digital transformation [11], and financial strict supervision [12]. However, there are relatively few relevant literature directly exploring the impact of data assetization on the innovation efficiency of manufacturing enterprises, and it has not deeply explored its heterogeneity performance and internal mechanism among different types of enterprises. Therefore, whether data assetization can enhance the innovation efficiency of manufacturing enterprises, the extent of its impact, and the specific path of its effect still need further empirical verification and theoretical deepening. Traditional theory holds that problems such as difficult data ownership confirmation, ambiguous valuation, and high transaction costs may hinder the effective allocation of enterprise innovation resources [13]. However, recent studies have shown that enterprises' data assets can enhance the ability to raise external funds, ensure the sufficiency and liquidity of funds needed for enterprise innovation activities, help enhance the subjective and objective innovation intentions of enterprises, and thereby improve financial stability and innovation investment [14]. Especially in the field of intelligent manufacturing, data assetization-driven predictive maintenance and personalized customization have significantly shortened the product development cycle [15]. Data assetization can help enterprises optimize production and management decisions, improve the efficiency of information and knowledge

conversion, improve process and product quality, thereby enhancing labor productivity and use value, and ultimately enabling enterprises to gain more economic profits [16]. Secondly, compared with other types of factors, data assets have stronger positive externalities. They not only can enhance the production efficiency of enterprises, but also can catalyze the formation of digital industrial clusters through the interweaving and integration of information flow, technological flow, capital flow and material flow, and thereby achieve digital collaborative linkage and positive feedback cycle effects [17], which can help improve the financial stability of enterprises. In the era of digital economy, enterprises with higher data processing and application capabilities often can raise funds at a lower cost. Through the deep integration of data assets with traditional production factors such as capital, labor, etc., these enterprises can effectively stimulate new driving forces for economic development, thereby promoting the enhancement of enterprise innovation willingness and the optimization of innovation resource allocation [18].

Most existing studies focus on the macroeconomic growth effect of data elements, and there is still a lack of systematic dissection of the mechanism of how data assetization affects innovation efficiency at the micro-enterprise level. Therefore, this paper uses the data of A-share manufacturing listed companies from 2015 to 2023 to test the effect of data assetization on the innovation efficiency of manufacturing enterprises. Firstly, a four-sector model integrating endogenous growth, knowledge production, and data resource accumulation is constructed. Then, the text mining technology is used to construct the data assetization indicator by taking the logarithm of the frequency of characteristic words plus one. The random frontier analysis method is used to measure innovation efficiency, and the generalized method of moments estimation (GMM), two-stage least squares method (2SLS), and simultaneous equation model are mainly used for estimation.

The marginal contribution of this paper lies in: First, it incorporates data assetization into the research framework of manufacturing enterprise innovation efficiency, and empirically verifies the promoting effect of data elements on total factor productivity at the micro level, filling the gap in existing literature that focuses more on

the macro level and slightly on the micro level. Second, it enriches the academic achievements related to data assetization and innovation, constructs a theoretical framework of "data capital accumulation - knowledge recombination - dynamic efficiency optimization" and a relatively complete influence path chain, explains the mechanism and contextual differences between the two, and expands the boundaries of existing enterprise development theories.

2. Research Design

2.1 Theoretical Model and Hypothesis

This paper constructs a four-sector model that integrates endogenous growth, knowledge production, and data resource accumulation, to explore the impact of enterprise data assetization on innovation, and emphasizes the value creation path of data as a new type of production factor. The basic logic of model construction is that data assetization promotes the accumulation of data capital, realizes knowledge production and reorganization, enhances the innovation efficiency of enterprises, and ultimately achieves the growth of innovation output.

2.1.1. Departmental setup

(1) Production department

Introduce data capital into the production function of the enterprise. The representative enterprise's production function is given in the form of C-D (Cobb-Douglas):

$$Y_t = A_t K_t^\alpha L_t^\beta D_t^\gamma, \alpha + \beta + \gamma = 1 \quad (1)$$

Among them, the value added by enterprise Y_t , A_t represents the technological level (which can also be understood as total factor productivity TFP), is determined by the R&D department, K_t represents the stock of physical capital, L_t represents the labor input, and D_t represents the stock of data capital. α , β and γ respectively represent the output elasticity of physical capital, labor input, and data capital.

(2) Research and development department

Integrating data capital into Romer's knowledge production function, an extended Romer knowledge production function is obtained:

$$\dot{A}_t = \theta R_t^\lambda D_t^\eta A_t^\phi \quad (2)$$

Among them, \dot{A}_t represents the knowledge increment (or TFP growth rate); R_t represents the R&D investment, including R&D personnel L_t^R and R&D funds K_t^R . For simplicity, assume that $R_t = s_R Y_t$, where s_R is the R&D investment

ratio; D_t represents the data capital stock, which is assumed to be shared by the production sector; θ is the R&D efficiency parameter; λ is the elasticity of R&D investment (usually $0 < \lambda \leq 1$); η is the elasticity of data capital on innovation output, and it is expected that $\eta > 0$; ϕ is the elasticity of the existing knowledge stock; if $\phi > 0$, it indicates positive feedback in knowledge accumulation, $\phi = 1$ indicates that the knowledge stock has a scale effect and sustainable growth, if $\phi < 1$, it indicates that the knowledge accumulation efficiency decreases and a steady state exists.

(3) Data management department

It is assumed that the accumulation of enterprise data capital is affected by three factors. The first is that enterprises convert raw data into productive data capital through data governance, which is a function of $s_D Y_t$, where s_D is the data investment rate; the second is that they obtain the required data M_t through the data factor market and convert it into data capital σM_t , where σ is the data integration efficiency, representing the enterprise's ability to transform and absorb the purchased data; the third is that data capital depreciation, $\delta_D D_t$, where δ_D is the data capital depreciation rate. The data capital accumulation equation:

$$\dot{D}_t = s_D Y_t - \delta_D D_t + \sigma M_t \quad (3)$$

(4) Employee behavior

Assuming that the government and international trade are not taken into account, under the budget constraint $\dot{K}_t = Y_t - C_t - I_t + \delta_t K_t$, the representative employee provides labor and consumes final products, with the goal of maximizing the final utility function as follows:

$$U = \int_0^\infty e^{-\rho t} \frac{C_t^{1-\sigma} - 1}{1-\sigma} dt \quad (4)$$

2.1.2 Market clearing

Suppose that when the market clears, the total labor force in the industry is fixed at \bar{L} . Then, in the labor market, $L_t = L_t^Y + L_t^R = \bar{L}$; in the capital market, $K_t = K_t^Y + K_t^R + K_t^D$, where K_t^D represents the physical management-related physical capital and can be simplified to assume $K_t^D = \kappa D_t$, with κ being the proportion coefficient; in the product market, $Y_t = C_t + \dot{K}_t + \delta_k K_t + R_t + I_{Dt}$, where $R_t = s_R Y_t$ and $I_{Dt} = s_D Y_t$.

Assuming that on the equilibrium growth path, the growth rates of output g_Y , consumption g_C , and capital g_K are equal, i.e., $g_Y = g_C = g_K$. Taking the logarithmic derivative of the production function, the growth relationship can be obtained:

$g_Y = g_A + \alpha g_K + \beta g_L + \gamma g_D$ (5) Since the labor force L_t is fixed ($g_L = 0$), we have:

$$g_Y = g_A + \alpha g_K + \gamma g_D \quad (6)$$

On the balanced growth path, we have $g_K = g_Y$. From the capital accumulation equation $\dot{K}_t = s_K Y_t - \delta_t K_t$ (where s_K is the rate of investment in physical capital), we can obtain:

$$g_Y = \frac{s_K Y_t}{K_t} - \delta_k \quad (7)$$

On the balanced growth path, we have $\dot{D}_t = g_D D_t$ where g_D is the growth rate of data capital D_t . Assuming that the purchased data flow M_t is proportional to output Y_t (i.e., $M_t = m Y_t$, where m is a constant), and using the capital accumulation equation $\dot{D}_t = s_D Y_t - \delta_D D_t + \sigma M_t$ we can obtain:

$$g_D D_t = s_D Y_t - \delta_D D_t + \sigma m Y_t \quad (8)$$

Final translation:
Organized as follows:

$$g_D = (s_D + \sigma m) \frac{Y_t}{D_t} - \delta_D \quad (9)$$

On the balanced growth path, we have $\dot{A}_t = g_A A_t$, where g_A is the growth rate of knowledge stock. Let the research and development investment be $R_t = s_R Y_t$. Substituting this into the knowledge production function gives:

$$g_A A_t = \theta (s_R Y_t)^\lambda D_t^\eta A_t^\phi \quad (10)$$

Divide both sides by A_t and take the logarithmic derivative, we obtain:

$$g_A = \frac{\lambda g_Y + \eta g_D}{1 - \phi} \quad (11)$$

2.1.3. Solve for the equilibrium growth rate

From equation (6), we have: $g_Y - \alpha g_K - \gamma g_D = g_A$

On the equilibrium growth path, $g_K = g_Y$, therefore:

$$g_Y(1 - \alpha) - \gamma g_D = g_A \quad (12)$$

Combining (11), we have:

$$g_Y = \frac{\eta + \gamma(1 - \phi)}{(1 - \alpha)(1 - \phi) - \lambda} g_D \quad (13)$$

$$g_A = g_Y \frac{\lambda + \eta(1 - \alpha)}{\eta + \gamma(1 - \phi)} = g_D \frac{\lambda + \eta(1 - \alpha)}{(1 - \alpha)(1 - \phi) - \lambda} \quad (14)$$

Here, g_D is determined by data capital accumulation and is proportional to g_Y and g_A in the steady state.

From Equation (14), it can be seen that the elasticity η of data capital on innovation output is larger. Under the equilibrium growth path, the impact of data capital D_t on the knowledge output efficiency of the R&D department becomes greater. Thus, Proposition 1 of this paper is derived.

Proposition 1: There is a positive relationship between enterprise data assetization and innovation efficiency. The greater the innovation output elasticity of data capital, the greater the impact of data assetization on innovation efficiency.

The output elasticity γ of data assetization is greater than 0. Data capital increases final output, as shown in Equation (13), increases resources available for research and development, increases the growth rate of knowledge stock, and promotes the improvement of innovation efficiency. Therefore, data assetization indirectly plays a role in guiding enterprise innovation by increasing output, leading to Proposition 2 of this paper.

Proposition 2: Enterprise data assetization indirectly promotes the improvement of enterprise research and development efficiency through the enterprise value creation effect.

The knowledge stock elasticity ϕ is greater than 0, indicating that the existing knowledge stock enhances future research and development efficiency. On the one hand, it reduces the cost of technology route selection through predictive analysis, and on the other hand, it reduces resource misallocation losses through data matching, ultimately improving the efficiency of research and development investment and strengthening the self-incremental effect of knowledge stock. This leads to Proposition 3 of this paper.

Proposition 3: Enterprise data assetization indirectly promotes the improvement of enterprise innovation efficiency by enhancing the efficiency of enterprise research and development investment.

Through the interaction of the research and development investment elasticity λ and the elasticity of data capital on innovation output η , data capital forms a complementary effect with the investment of research and development personnel and research and development expenses, thereby enhancing the innovation efficiency of the enterprise. This collaborative essence is the permeability reorganization of data elements on traditional production elements [19]. This leads to Proposition 4 of this paper.

Proposition 4: Enterprise data assetization generates a synergistic effect with research and development personnel, research and development expenses, and other research factors, enhancing the innovation efficiency of the enterprise.

2.2 Measurement Model

2.2.1. Baseline regression model

Based on the aforementioned theoretical model and Proposition 1, the following baseline regression model is constructed:

$Eff_y_{it} = \beta_0 + \beta_1 DA_{i,t} + \gamma \sum Controls_{i,t} + \varepsilon_{i,t}$ (15) Among them, the explained variable Eff_y_{it} it represents the enterprise innovation efficiency of enterprise i in the t -th year, the core explanatory variable $DA_{i,t}$ represents the data assetization of enterprise i in the t -th year, $Controls_{i,t}$ is the control variable, and ε is the random error term.

2.2.2. Mediating effect model

Based on the aforementioned theoretical model and Propositions 2 and 3, the following mediating effect model is constructed:

$$Lnvadd_{it} = \delta_0 + \delta_1 DA_{i,t} + \gamma \sum Controls_{i,t} + \zeta_{i,t}$$

$(16) Absinv_{it} = \delta_0 + \delta_1 DA_{i,t} + \gamma \sum Controls_{i,t} + \zeta_{i,t}$ (17) Among them, the dependent variable $Lnvadd_{it}$ in formula (16) represents the enterprise's added value in year t for enterprise i the dependent variable $Absinv_{it}$ in formula (17) represents the enterprise's investment efficiency, which is measured by the deviation index of enterprise R&D investment. Referencing the relevant studies on investment efficiency [20], the investment efficiency indicator is calculated according to the following formula:

$$Inv_{it} = \alpha_0 + \alpha_1 Rev_{it-1} + \alpha_2 Cash_{it-1} + \alpha_3 Age_{it-1} + \alpha_4 Size_{it-1} + \alpha_5 Roe_{it-1} + \alpha_6 Lev_{it-1} + \alpha_7 Inv_{it-1} + \sum YEAR + \sum IND + \rho_{it}$$
 (18)

Among them, Inv_{it} represents the R&D investment level of enterprise i in the t -th year, $Inv_{it} = Trde_{it} / Indx_{it}$ where $Trde$ represents the R&D expenditure invested by the enterprise, and $Indx$ represents the consumer price index of each province in the t -th year based on the base year of 2000. Rev is the growth rate of operating income, $Cash$ is the net cash flow generated from operating activities, Age is the natural logarithm of the age of the enterprise plus 1, $Size$ is the natural logarithm of the enterprise's total assets plus 1, Lev is the enterprise's debt-to-equity ratio, Roe is the enterprise's return on equity, and the model controls the R&D investment expenditure of the previous period. The residual of equation (18) represents the inefficient investment of the enterprise. According to existing research, taking the absolute value of the residual gives the indicator of the deviation degree of the enterprise's R&D investment from the expected level. The larger this indicator is, the greater the deviation between the actual investment and the expected investment of the enterprise is, and the lower the investment efficiency of the enterprise is.

The core explanatory variable $DA_{i,t}$ is the data assetization of enterprise i in the t -th year,

$Controls_{i,t}$ is the control variable, and ζ is the random error term.

2.2.3. Moderation effect model

Based on the aforementioned theoretical model and Proposition 4, the following moderation effect model is constructed:

$$Eff_y_{it} = \eta_0 + \eta_1 DA_{i,t} + \eta_2 RDA_{it} + \gamma \sum Controls_{i,t} + \tau_{i,t}$$

$$(19) Eff_y_{it} = \eta_0 + \eta_1 DA_{i,t} + \eta_2 NDA_{it} + \gamma \sum Controls_{i,t} + \tau_{i,t}$$

(20) In formula (19), the variable RDA_{it} represents the product of the logarithm of data assetization and R&D investment of enterprise i in the t -th year. In formula (20), the variable NDA_{it} represents the product of the logarithm of data assetization and the number of R&D personnel of enterprise i in the t -th year. $Controls_{i,t}$ is the control variable, and τ is the random error term.

2.2.4. Endogeneity discussion

This paper adopts the method of Greene (2018) [21] and constructs a system of equations including the equation of enterprise innovation efficiency and the equation of data assetization. It uses full information maximum likelihood estimation to alleviate endogeneity. The equation of data assetization (DA) plays a role in model identification, and it needs to be relatively exogenous. The role of parameter ρ is to determine the bias direction of the least squares parameter estimation. If there is a bias problem caused by omitted variables in the equation of enterprise innovation efficiency (Eff_y), the omitted variables will enter the random error term ε , and ρ is the correlation coefficient between the omitted factor and data assetization. Assuming the omitted variable is x , under the condition of not considering other factors, the relationship between the least squares estimation bias and the true value can be obtained as follows:

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \rho$$
 (21)

Among them, $\hat{\beta}_1$ represents the parameter of the least squares estimation. β_1 and β_2 are the true parameters, indicating the influence of the omitted variable on the innovation efficiency of the enterprise. ρ is the correlation coefficient between the omitted variable x and the data assetization. If $\rho > 0$ and $\beta_2 > 0$, or $\rho < 0$ and $\beta_2 < 0$, the least squares estimator is overestimated; if $\rho > 0$ and $\beta_2 < 0$, or $\rho < 0$ and $\beta_2 > 0$, the least squares estimator is underestimated. Compared with other methods for alleviating endogeneity, the simultaneous equation model can simultaneously identify β_1 and ρ , and can also explain the causes of

endogeneity.

The simultaneous equation model of enterprise innovation efficiency (Eff_y) and data assetization (DA) is as follows:

$$\begin{cases} Eff_{y_{it}} = \beta_0 + \beta_1 DA_{i,t} + \gamma \sum Controls_{i,t} + \varepsilon_{i,t} \\ DA_{i,t} = \alpha_0 + \alpha_1 MDA_{i,t} + \gamma' \sum Controls'_{i,t} + \mu_{i,t} \end{cases}$$

(22) The explained variable $Eff_{y_{it}}$ represents the enterprise innovation efficiency of enterprise i in the t -th year. The core explanatory variable $DA_{i,t}$ represents the data assetization of enterprise i in the t -th year. $MDA_{i,t}$ represents the average data assetization of the industry where the enterprise is located. $Controls$ and $Controls'$ are control variables, and ε and μ are random error terms. It is assumed that ε and μ follow a bivariate normal distribution with mean vector u and variance matrix Σ , that is, $\begin{pmatrix} \varepsilon \\ \mu \end{pmatrix} \sim N_2(u, \Sigma)$ where:

$$u = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad (23) N_2(\cdot, \cdot)$$

represents the joint distribution function of the bivariate normal distribution, where σ_1 and σ_2 are the standard deviations of ε and μ respectively, and ρ is the correlation coefficient. Then, the log-likelihood function lnL_i of enterprise i is as follows:

$lnL_i = -\frac{M}{2} ln 2\pi - \frac{1}{2} ln |\Sigma| - \frac{1}{2} \gamma_i' \Sigma^{-1} \gamma_i$ (24) Perform a complete information maximum likelihood estimation on Equation (22), and obtain the parameter estimate value $\hat{\beta}_1$ of the coefficient β_1 for the enterprise data assetization. The $MDA_{i,t}$ in the enterprise data assetization equation refers to the industry average data assetization of the enterprise in the current year.

3. Data Sources and Descriptive Statistics

3.1 Sample Selection and Data Sources

This paper selects the financial reports and annual report texts of Chinese A-share manufacturing companies listed on the stock market from 2015 to 2023 as the research data source. To avoid data bias from interfering with the research, the following screening and processing are carried out on the selected data: eliminating samples of companies that experienced ST, *ST, and financial-related issues in the same year; eliminating samples where the main empirical data are missing. By using Python text mining and keyword frequency analysis, key variable data, asset-based data are obtained; the original data

of other related variables come from the WIND database. Through strict screening, this study finally obtained 20,539 annual observations of enterprises.

3.2 Variable Definition

3.2.1. Dependent variable

Enterprise innovation efficiency (Eff_y). Currently, the academic community mainly uses two types of methods, namely parametric and non-parametric, to measure enterprise innovation efficiency. The parametric method is represented by stochastic frontier analysis (SFA), which estimates the efficiency frontier by setting the form of the production function [22]; the non-parametric method, such as data envelopment analysis (DEA), assesses the relative efficiency of multi-input-output systems through linear programming [23]. Considering the significant industry technological heterogeneity and random interference factors in the manufacturing innovation process, as well as the single-output characteristic of patent output, this paper adopts the stochastic frontier analysis method for measurement.

In terms of specific measurement, drawing on the model setting of Wu Yanbing (2017) [24], the innovation output is represented by the logarithm of the enterprise's annual patent application volume plus 1 ($lnApat$), and the innovation input factors are the logarithm of the research and development capital stock ($lnCapital_rd$) and the logarithm of the number of research and development personnel ($lnNrd$). The research and development capital stock is calculated using the perpetual inventory method, and the research and development personnel data are processed through logarithmic transformation to eliminate the influence of units of measurement. Based on the Cobb-Douglas production function form, a stochastic frontier model is constructed:

$$ln\hat{y}_{it} = \beta_0 + \beta_1 ln\hat{x}_{it} + \beta_2 ln\hat{z}_{it} + v_{it} - u_{it}$$

(25) To control for industry technological differences, based on the two-digit codes of the "Guidelines for Industry Classification of Listed Companies" (2012 Edition) issued by the China Securities Regulatory Commission, an estimation was conducted to cover 13-42 sub-sectors of the manufacturing industry. The time-varying random effects model (tre) was employed to capture the dynamic evolution characteristics of efficiency, and it was assumed that the inefficiency term followed a truncated

normal distribution. Finally, the technical efficiency values were calculated using the JLMS estimator proposed in [22]:

$Eff_{y_{it}} = \exp\{f_{it}\}(-u_{it}) \in (0,1)$ (26) The economic meaning of this indicator is the ratio of actual patent output to the optimal industry frontier output. When the IE value is closer to 1, it indicates that the enterprise is getting closer to the optimal production level of the industry under the given R&D investment conditions.

3.2.2. Core explanatory variable

Enterprise data assetization (DA). This study selecting legal and regulatory texts related to data assets to build a professional corpus, determining five core terms "digital", "data", "internet", "information", and "intelligence" as initial seed words, and using the Word2Vec word vector model and deep learning technology to train the corpus, generating a set of similar words for each seed word and constructing a data assetization text dictionary. Then, using Python software, the standardized word frequency of corresponding characteristic words in the annual reports of listed companies is extracted, and finally, the standardized word frequency of the characteristic words is incremented and then logarithmically transformed to construct the main explanatory variable - data assetization.

3.2.3. Control variables

Following the research approach of He Ying et al. (2024) [5], this paper incorporates a series of control variables: the enterprise basic dimension includes the natural logarithm of the enterprise age+1 (Age) and the natural logarithm of the enterprise total assets + 1 (Size), the financial structure dimension introduces the debt-to-asset ratio (Leve) and the current ratio (Liq), the governance mechanism dimension controls the proportion of the top five shareholders' holdings (Top5_Sh) and the proportion of independent

directors (ID_Ratio), and the innovation policy dimension embeds the intensity of R&D investment (R&D_Int) and government subsidies (PS). Finally, through the three-dimensional fixed effect layer of industry-region-year, the systematic biases caused by differences in industrial structure, regional factor market segmentation, and macro cycle fluctuations are absorbed, ensuring that the estimated values of the core explanatory variables meet the condition independence assumption.

3.3 Descriptive Statistics of Variables

Table 1 shows the descriptive statistics of the main variables: The median innovation efficiency of the sample enterprises is 0.219, which is lower than the mean of 0.242. This indicates that more than half of the manufacturing enterprises have an innovation efficiency lower than the industry average, reflecting the current situation of insufficient overall innovation momentum in China's manufacturing sector. The minimum value of data assetization level is 0 and the maximum value is 6.713, showing a significant structural differentiation. It indicates that the leading enterprises have established a complete data management system, while some enterprises have not yet initiated data transformation. The practical dilemma lies in that the median of the enterprise's current ratio is 1.902, which is lower than the safety threshold of 2. Combined with the extreme value of the asset-liability ratio of 284.9%, it confirms that the debt risk has systematically occupied the space for innovation resource investment. The difference between the median of government subsidies (0) and the maximum value of 39.85 highlights the significant structural bias in policy resource allocation.

Table 1. Descriptive Statistics of Key Variables

Variable	Observation Value	Average	Standard Deviation	Minimum Value	Median	Maximum Value
Corporate Innovation Efficiency	20 539	0.242	0.175	0.000	0.219	0.977
Data Monetization	20 539	2.850	1.066	0.000	2.833	6.713
Natural logarithm of the Company's Age + 1	20 539	3.007	0.296	1.386	3.045	4.234
Natural Logarithm of Total Enterprise Assets + 1	20 539	22.120	1.181	17.810	21.940	27.640
Asset-Liability Ratio	20 539	38.310	19.400	0.836	37.130	284.900
Current Ratio	20 539	2.876	3.279	0.094	1.902	80.660
Percentage of Independent Directors	20 539	0.378	0.066	0.111	0.333	1.000
Top Five Shareholders' Shareholding Ratios	20 539	53.280	15.000	6.910	53.140	99.230
Proportion Of R&D Expenses to Revenue	20 539	0.051	0.043	0.001	0.040	0.254
Government Subsidies	20 539	0.105	0.728	-0.048	0.000	39.850

4. Empirical Research

4.1 Baseline Regression

This paper first estimates the impact of data assetization on the innovation efficiency of manufacturing enterprises from 2015 to 2023 based on Equation (15). Table 2 presents the benchmark regression results. In the (1) and (2) columns, after controlling for relevant factors, the coefficient of the core explanatory variable is positive and highly significant at the 1% level. For every 1-unit increase in the level of data assetization of enterprises, the innovation efficiency increases by an average of 1.6 percentage points. In the (3) column, after including the enterprise attribute variables such as specialized and innovative enterprises and strategic emerging enterprises, the constant term significantly decreases, indicating that the

inherent innovation advantages of policy-supported high-tech enterprises have been effectively separated. In the (4) and (5) columns, after further controlling for enterprise ownership, urban and regional variables, the core coefficients remain stable and are highly significant at the 1% level, indicating that the innovation promoting effect of data assetization has the universality across regions and property rights. With the gradual improvement of the control variable system, the goodness-of-fit R^2 value monotonically increases from 0.182 to 0.193, and the maximum likelihood estimate value increases by 141.684, confirming that the model setting conforms to the principle of asymptotic validity in econometrics. Supporting the proposition 1 of this paper, that is, data assetization has a significant promoting effect on the innovation efficiency of manufacturing enterprises.

Table 2. Regression Results of the Impact of Data Assetization on Enterprise Innovation Efficiency

Variable	(1)	(2)	(3)	(4)	(5)
	Corporate Innovation Efficiency				
Enterprise Data Monetization	0.016*** (12.410)	0.015*** (12.300)	0.016*** (12.630)	0.016*** (12.630)	0.016*** (12.870)
Constant Term	0.416*** (14.480)	0.433*** (14.170)	0.370*** (11.910)	0.374*** (10.840)	0.311*** (8.860)
Control Variables	Yes	Yes	Yes	Yes	Yes
Enterprise Scale	No	Yes	Yes	Yes	Yes
Specialized, Refined, Unique, And Innovative	No	No	Yes	Yes	Yes
Strategic Emerging Industries	No	No	Yes	Yes	Yes
Enterprise Ownership Structure	No	No	No	Yes	Yes
City of Residence	No	No	No	No	Yes
Location	No	No	No	No	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
N	20 539	20 539	20 539	20 539	20 539
R^2 Value	0.182	0.183	0.188	0.189	0.193
Maximum Likelihood Estimate	8 761.643	8 771.740	8 831.571	8 847.525	8903.327

4.2 Endogenous Treatment

4.2.1. Estimation using GMM and 2SLS methods

To address potential endogenous bias, this paper selects the mean of industry data assetization as the instrumental variable based on the technological homogeneity of the manufacturing industry and the characteristics of the industrial cluster. It then conducts regression estimation using the Generalized Method of Moments (GMM) and the Two-Stage Least Squares (2SLS) method. The regression results are shown in columns (1) and (2) of Table 3. The regression coefficients are all positive and statistically significant at the 1% level, indicating that data

assetization has a significant positive impact on the innovation efficiency of enterprises. The coefficient value is nearly 6.06 times higher than that of the baseline regression, confirming that the Ordinary Least Squares method has a significant underestimation bias.

4.2.2. Systematic estimation using a system of equations model

To overcome the limitations of the instrumental variable method when dealing with bidirectional causal relationships, this paper employs a system of equations model for comprehensive estimation. The regression results are shown in the third column of Table 3. The results of the first-stage regression indicate that after controlling for the bidirectional feedback

mechanism, the regression coefficients are significantly enhanced compared to the baseline model. The data assetization level has a significant positive effect on the innovation efficiency of enterprises, verifying the robust enabling effect of data elements on innovation efficiency. The second-stage regression confirmed that for every 1-unit increase in the average level of industry data assetization, it can significantly promote the degree of data transformation of individual enterprises by 0.811

units, highlighting the dominant role of the technological diffusion effect in the manufacturing industry. Further, the residual correlation coefficient is significantly negative, indicating that there are systematic factors not captured by the observed variables that inhibit the coordinated evolution of the data assetization process and innovation performance. The promotion effect of data assetization on the innovation efficiency of enterprises is significantly underestimated.

Table 3. Endogeneity Test

Variable	(1) Gmm	(2) 2sls	(3) Estimation of Systems of Simultaneous Equations
Equation	Corporate Innovation Efficiency		
Enterprise Data Monetization	0.112*** (10.760)	0.113*** (10.630)	0.119*** (10.430)
Constant Term	0.486*** (10.750)	0.488*** (11.050)	0.466*** (10.640)
Control Variables	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Equation 2	Enterprise Data Monetization		
Industry Average Data Monetization	-	-	0.811*** (11.430)
Constant Term	-	-	-0.0410 (-0.140)
Control Variables	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Residual Correlation Coefficient	-	-	-0.503*** (-12.170)
Sample Size	20 539	20 539	20 539
Maximum Likelihood Estimate	-	-	-17 329.862

4.3 Robustness Test

4.3.1. Replacing the explanatory variables

To verify the reliability of the basic regression conclusion, this paper first changes the measurement method of the core explanatory variables and re-constructs the data assetization indicators by using the TF-IDF technology based on the text mining of enterprise annual reports. The test results are shown in column (1) of Table 4. The regression coefficient of the data assetization variable measured by TF-IDF is 0.213, and it is statistically significant at the 1% level. The sign of this coefficient is consistent with the baseline estimation results, and the direction is significantly positive, confirming the robustness of the baseline regression results.

4.3.2. Replace the dependent variable

To enhance the generalizability of the research conclusions, this paper further changes the measurement method of the dependent variable. Following the approach of Jin Huan et al. (2024) [25], this method combines the ACF method with the modified semi-parametric OP method to

re-calculate the innovation efficiency of manufacturing enterprises. Among them, the innovation output variable ($\ln Apat$) is measured by taking the logarithm of the total number of patent applications by enterprises plus 1. The innovation input variable, the investment in R&D personnel ($\ln Nrd$) is measured by taking the logarithm of the total number of scientific and technological activity personnel of enterprises. The R&D capital input ($\ln Capital_rd$) is estimated using the perpetual inventory method: $Capital_{it} = (1 - \delta)Capital_{it-1} + Inv_{it}$, where Inv_{it} represents the R&D investment level of enterprise i in the t -th year, $Inv_{it} = Trde_{it} / Indx_{it}$, $Trde_{it}$ represents the R&D expenditure of enterprises, measured as the sum of the internal funds expenditure for scientific and technological activities and the funds expenditure for conducting scientific and technological activities by entrusting external units, and $Indx$ represents the consumer price index of each province in the t -th year based on the base year of 2000, with δ set at 15%, indicating the depreciation rate of R&D capital.

The test results show that the regression coefficient of the data assetization variable is 1.177, and it remains highly significant at the 1% significance level. Compared with the benchmark model, this coefficient has stronger statistical significance, indicating that after using the OP-ACF joint estimation method, the marginal improvement effect of data assetization on innovation efficiency is more prominent, and the regression conclusion is more robust.

4.3.3. Trimming of outliers

To control the potential bias of outlier samples on the estimation results, this study applies a two-tailed 1% trimming to the data assetization variables. This method effectively mitigates the distortion of extreme observations on parameter estimation. The results of the simultaneous equation estimation show that the data assetization coefficient after trimming is 0.118, which is statistically significant at the 1% level, and its direction is exactly the same as that of the benchmark model.

4.3.4. Lagged one period of core explanatory variables

To alleviate the potential interference of reverse

causality, this paper incorporates the explanatory variable of enterprise data assetization lagged by one period into the simultaneous equation system. The results show that the lagged coefficient is 0.165 and is highly significant at the 1% level. This confirms that the accumulation of data elements has a continuous promoting effect on innovation efficiency.

4.3.5. Subsample regression

To assess the potential impact of sample selection bias on the estimation results, this study employed a 90% random sampling to reconstruct the observational data set for simultaneous equation estimation. The results of the subsample regression showed that the coefficient of data assetization was 0.123 and remained highly significant at the 1% significance level, with its direction being completely consistent with the benchmark estimation of the full sample. Reducing the sample size to 15,404 observations did not lead to the failure of statistical inference, and the standard errors of the core variables remained within a reasonable range.

Table 4. Robustness Test Results

Variable	(1)	(2)	(3)	(4)	(5)
Equation 1	Corporate Innovation Efficiency				
Tf-idf	0.213*** (6.470)	-	-	-	-
Enterprise Data Monetization	-	1.177*** (9.990)	-	-	0.123*** (9.890)
Data Assetization After Tail Trimming	-	-	0.118*** (10.970)	-	-
Data Lag One Period Behind	-	-	-	0.165*** (10.520)	-
Constant Term	0.324*** (8.900)	3.118*** (6.890)	0.461*** (10.690)	0.505*** (9.270)	0.450*** (9.000)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Equation 2	Enterprise Data Assetization				
Industry Average Data Monetization	0.946*** (25.820)	0.661*** (8.890)	0.829*** (11.840)	0.504*** (6.990)	0.798*** (9.970)
Constant Term	0.624*** (10.030)	1.961*** (5.930)	1.392*** (4.190)	1.627*** (4.400)	1.588*** (4.150)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Residual Correlation Coefficient	-0.203*** (-6.210)	-0.520*** (-12.800)	-0.495*** (-12.630)	-0.655*** (-16.920)	-0.519*** (-11.820)
Sample Size	20 539	20 539	20 539	17 304	15 404
Maximum Likelihood Estimation	17 143.043	-64 966.976	-17 129.859	-14 561.926	-12 911.679

4.4 Heterogeneity Analysis

This study sequentially introduces the dummy variables of enterprise characteristics and location into the benchmark regression model to

reveal the potential heterogeneity impact of data assetization on the innovation efficiency of manufacturing enterprises.

4.4.1. Heterogeneity of enterprise scale

Based on the enterprise scale classification

standard provided by Wind, this paper divides enterprises into three categories: large, medium, and small. Due to the small sample size of small-scale enterprises, they are combined with medium-sized enterprises for heterogeneity analysis. The results are shown in Table 5: In large-scale enterprises, the promotion effect of enterprise data assetization on innovation efficiency is significant, with a coefficient of 0.130 and statistically significant at the 1% level. However, in medium-sized and small-scale enterprises, the coefficient is only 0.055, which is relatively small. This indicates that data assetization improvement has a more obvious effect on the innovation efficiency of large-scale enterprises.

Table 5. Heterogeneity in Firm Size

Variable	(1) Large-scale Enterprises	(2) Small And Medium-sized Enterprises
Equation 1	Corporate Innovation Efficiency	
Enterprise Data Monetization	0.130*** (10.550)	0.055*** (2.790)
Constant Term	0.424*** (8.730)	0.741*** (6.670)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Equation 2	Enterprise Data Monetization	
Industry Average Data Monetization	0.771*** (10.200)	1.096*** (6.520)
Constant Term	2.170*** (5.720)	-1.874** (-2.220)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Residual Correlation Coefficient	-0.515*** (-12.550)	-0.265** (-2.270)
Sample Size	15 466	5 073
Maximum Likelihood Estimate	-13 013.567	-3 917.298

4.4.2. Heterogeneity of enterprise ownership structure

The property attributes of enterprises have a significant impact on the relationship between data assetization and the innovation efficiency of enterprises. According to the property attributes of the actual controllers, this study divides the sample enterprises into state-owned enterprises and non-state-owned enterprises. The empirical results are shown in Table 6. When the dependent variable is the innovation efficiency of enterprises, the coefficient of enterprise data

assetization for state-owned enterprises is 0.065, and that for non-state-owned enterprises is 0.137, both of which are significant at the 1% level. This indicates that data assetization has a more significant promoting effect on the innovation efficiency of non-state-owned enterprises.

Table 6. Heterogeneity in Enterprise Ownership Structure

Variable	(1) State-owned Enterprise	(2) Non-state-owned Enterprise
Equation 1	Corporate Innovation Efficiency	
Enterprise Data Monetization	0.065*** (4.240)	0.137*** (10.270)
Constant Term	0.0800 (1.040)	0.586*** (10.710)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Equation 2	Enterprise Data Monetization	
Industry Average Data Monetization	1.090*** (7.650)	0.701*** (8.950)
Constant Term	3.168*** (4.390)	1.243*** (3.280)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Residual Correlation Coefficient	-0.281*** (-3.950)	-0.571*** (-13.390)
Sample Size	5 123	15 416
Maximum Likelihood Estimate	-4 469.581	-12 446.409

4.4.3. Heterogeneity of the city type where the enterprise is located

The administrative level of the city where the enterprise is located varies, and the impact of data assetization on innovation efficiency will also differ. Based on whether the registered location of the enterprise belongs to a provincial capital or a sub-provincial city, this study divides the entire sample into two groups: "Provincial capital or sub-provincial city" and "Non-provincial capital or sub-provincial city", and re-estimates the model separately. The empirical results are shown in Table 7: Firstly, data assetization can significantly enhance the innovation efficiency of enterprises in both types of cities, but the promoting effect is stronger in enterprises in non-provincial capital or sub-provincial cities, with an estimated coefficient of 0.121, which is significant at the 1% level, higher than 0.112 in the provincial capital/sub-provincial city group.

Table 7. Heterogeneity of Enterprise Location by City Type

Variable	(1)	(2)
	Provincial Capital or Sub-provincial City	Non-provincial Capital or Sub-provincial City
Equation 1	Corporate Innovation Efficiency	
Enterprise Data Monetization	0.112*** (6.540)	0.121*** (9.160)
Constant Term	0.560*** (6.040)	0.417*** (8.430)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Equation 2	Enterprise Data Monetization	
Industry Average Data Monetization	0.762*** (5.080)	0.811*** (10.290)
Constant Term	4.185*** (5.440)	1.060*** (2.840)
Control Variables	Yes	Yes
Year Fixed Effect	Yes	Yes
Industry Fixed Effects	Yes	Yes
Residual Correlation Coefficient	-0.489*** (-7.820)	-0.510*** (-10.700)
Sample Size	4 713	15 826
Maximum Likelihood Estimate	-3 929.476	-13 083.973

4.4.4 Heterogeneity of the region where the enterprise is located

Based on the regional classification standard of the National Bureau of Statistics, this paper divides the samples into four groups: the eastern region, the central region, the western region and the three northeastern provinces. The empirical results are shown in Table 8: In the innovation efficiency equation, the coefficient of the enterprise data assetization from high to low is: the central region, the eastern region, the western region, and the coefficient of the northeastern region is negative and not significant.

Table 8. Heterogeneity of Enterprises' Locations

Variable	(1)	(2)	(3)	(4)
	Eastern Region	Central Region	Western Region	Manchuria
Equation 1	Corporate Innovation Efficiency			
Enterprise Data Monetization	0.097*** (8.600)	0.338*** (4.710)	0.069*** (3.140)	-0.0180 (-0.310)
Constant Term	0.465*** (9.550)	0.375* (1.860)	0.464*** (4.500)	0.355 (0.730)

Control Variables	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Equation 2	Enterprise Data Monetization			
Industry Average Data Monetization	0.840*** (10.020)	0.319** (2.550)	1.067*** (5.350)	0.962** (2.550)
Constant Term	1.856*** (4.400)	2.978*** (3.590)	2.809*** (3.110)	-8.241*** (-4.390)
Control Variables	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Residual Correlation Coefficient	-0.422** (-8.770)	-0.864** (-17.960)	-0.267** (-2.610)	0.147 (0.400)
Sample Size	14 535	2 958	2 334	712
Maximum Likelihood Estimate	-12 439.247	-2 236.787	-1 548.075	-313.923

5. Further Analysis

According to the theoretical model presented earlier, this promoting effect is mainly achieved by enhancing the overall value creation ability of the enterprises, improving the accuracy of R&D investment, and stimulating the mechanisms of data-human capital synergy and data-research and development capital synergy. Based on this, the following text will conduct a systematic examination of these mechanisms to verify the validity of the theoretical hypotheses.

5.1 The Effect of Enterprise Value Creation

The data assetization process, by driving the optimization of production and operation, forms the foundation for the accumulation of innovative resources. To verify this mechanism, the enterprise added value (Ln_{vadd}) was introduced as an intermediary variable. This added value was calculated using the income method and was obtained by summing up and taking the logarithm of labor compensation, net production tax, fixed asset depreciation, and operating surplus. The regression results are shown in the first column of Table 9. The coefficient of data assetization is 0.043 and is

significant at the 1% level. For every 1 unit of data assetization, the enterprise added value increases by 4.3 percentage points. This indicates that data assetization enhances the overall operational efficiency, provides more abundant financial reserves for enterprise

activities, and directly drives the investment in research and development funds, providing a continuous capital supply material basis for innovative activities. This verifies Proposition 2 of this paper.

Table 9. Mechanism Verification

Variable	(1)	(2)	(3)	(4)
	Enterprise Value Added	R&D Investment Efficiency	Innovation Efficiency	Innovation Efficiency
Enterprise Data Monetization	0.043*** (8.660)	-0.011*** (-3.160)	-0.017** (-2.220)	-0.006 (-1.300)
Data Monetization × Logarithm of R&D Investment	-	-	0.004*** (4.300)	-
Data Monetization × Logarithm of R&D Personnel Count	-	-	-	0.004*** (4.690)
Constant Term	-2.551*** (-22.450)	-1.535*** (-18.810)	0.618*** (12.920)	0.440*** (11.440)
Control Variables	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
R ²	0.782	0.210	0.178	0.183
Sample Size	19 527	19 862	17 297	20 542
Maximum Likelihood Estimate	-18 141.069	-12 069.697	7 579.241	8 776.701

5.2 Optimization Effect of R&D Investment

The accuracy of R&D resource allocation is the key to solving the problem of mismatch of innovation resources. This paper introduces enterprise investment efficiency ($Absinv_{it}$) as a mediating variable. This variable is measured by the deviation degree of R&D investment. The regression results are shown in column (2) of Table 9. The coefficient of data assetization is significant at the 1% level, and the negative coefficient indicates that data assetization reduces the deviation degree of R&D investment, that is, the improvement of data assetization can promote the improvement of enterprise investment efficiency. To verify the proposition 3 of this paper.

5.3 Factor Synergy Effect

The release of data value is highly dependent on the deep integration with traditional production factors. Therefore, in the equation of innovation efficiency, the cross-term of data assetization and R&D factors is introduced.

5.3.1. Funding synergy

The interaction term of data assetization and the logarithm of R&D investment (RDA) was introduced. The regression results are shown in the third column of Table 9. The coefficient of this interaction term is positive and significant at the 1% level. In light of the previous conclusion regarding the effect of enterprise value creation, data assetization can enhance enterprise value creation, which means that enterprises will have

more abundant funds flowing into their innovation activities. Adequate innovation funds further ensure the smooth conduct of enterprise R&D activities and improve the efficiency of R&D.

5.3.2. Human resource synergy

The interaction term of data assetization and the logarithm of the number of R&D personnel (NDA) was introduced. As shown in the fourth column of Table 9, the coefficient of this interaction term is positive and significant at the 1% level. This indicates that human capital is the key carrier for decoding data value. On one hand, data assetization, through activating the data application ability of human capital, forms a multiplier effect, thereby systematically improving the efficiency of innovation resource allocation; on the other hand, data transparency effectively eliminates information barriers between departments, making it possible to visualize the entire innovation chain, thereby reducing agency costs. To verify Proposition 4 of this paper.

The results of the above mechanism tests show that data assetization, on the one hand, accumulates innovation resources through value creation, and on the other hand, optimizes the accuracy of R&D allocation to reduce decision-making costs. Ultimately, it achieves a leap in innovation efficiency through deep synergy with enterprise R&D factors.

6. Conclusion and Implications

This study, based on the data of A-share

manufacturing companies from 2015 to 2023, measures innovation efficiency using stochastic frontier analysis, combines text mining technology to construct indicators for the assetization of enterprise data, empirically examines the impact of data assetization on innovation efficiency and its heterogeneity characteristics, and further analyzes the mechanism of this impact. The main conclusions are as follows: The data assetization significantly enhances the innovation efficiency of manufacturing enterprises. However, its impact varies depending on the characteristics of the enterprises, with large-scale, non-state-owned enterprises and enterprises in non-provincial capital cities showing a stronger promoting effect. Regionally, the effect is most prominent in the central region. Mechanism analysis indicates that data assetization enhances the enterprise's value creation ability, improves the accuracy of R&D investment, and leverages its synergy with human resources and capital, thereby promoting the improvement of enterprise innovation efficiency. Based on these conclusions, the following suggestions are proposed:

First, comprehensively deepen the strategy of data assetization for manufacturing enterprises, and build a market support system for data assetization. Accelerate the formulation of national standards for the valuation of manufacturing data, providing valuation benchmarks for financial institutions to carry out data pledge financing; build regional platforms for the registration and confirmation of data elements to reduce the compliance costs of data transactions; cultivate third-party data governance service providers to provide auditing, custody, and other professional services for enterprises.

Second, implement differentiated and precise support policies to bridge the "data gap" among enterprises. Focus on providing data empowerment support to small, medium, and micro manufacturing enterprises, offering low-cost or subsidy-based cloud data service tools, shared data resource libraries, professional consulting and training services. In terms of regional policies, support the central region to leverage its significant advantages and establish national manufacturing data assetization application demonstration zones; guide the eastern region to upgrade to data-intensive application and value extraction; increase

investment in basic data facilities and talent introduction in the western region; for the northeastern region, conduct in-depth research on the specific obstacles in the data assetization transformation process and provide targeted solutions, such as establishing a special relief fund or pilot projects for data assetization transformation.

Third, deepen the reform of the data innovation mechanism of state-owned enterprises. Establish a data application error-tolerant and liability exemption system that grants risk tolerance to innovative projects that adopt industry-standard algorithms and have compliant processes; set up data asset special zones in suitable regions to pilot measures for facilitating cross-border data flow; promote state-owned listed companies to regularly disclose data resource management reports and incorporate data element input-output into the corporate governance evaluation framework. Through institutional innovation to activate the potential of state-owned enterprises' data, form an innovation pattern that complements the advantages of private enterprises.

Fourth, deepen the collaborative investment of data elements and innovation elements, and create a "data × talent × capital" multiplier effect. When formulating innovation strategies, enterprises must integrate data resource investment with R&D funds and the introduction of high-end talents in a bundled and upgraded manner. On one hand, continuous investment in advanced data analysis tools and the cultivation of a R&D team with "technology + business + data" comprehensive capabilities are necessary to ensure that the full value of data can be exploited. On the other hand, when evaluating R&D projects, data support capabilities should be regarded as a key consideration factor, and projects with a solid data foundation and a clear data application path should be given priority.

References

- [1] Han Xiulan, Cui Mengtao, Li Baoyu. Analysis of the Current Situation and Research Frontiers of Domestic Data Assets. *Statistics and Decision*, 2024, 40(12): 18-23.
- [2] Luo Mei, Li Jinpu, Tang Ke. Enterprise Data Assetization: Accounting Confirmation and Value Evaluation. *Journal of Tsinghua University (Philosophical and Social Sciences Edition)*, 2023, 38(05): 195-209.
- [3] Xu Tao, You Jianxin, Zeng Caixia, Shi

- Yongjiang. Exploration of Enterprise Data Assetization Practice and Theoretical Model Construction. *Foreign Economics and Management*, 2022, 44(06): 3-17.
- [4] Chen Fuyong, Zeng Xiao. "Data" as a New Engine: A New Path for Driving High-Quality Development of Enterprises through Data Assetization. *Economic Issues*, 2025(06): 30-37.
- [5] He Ying, Chen Lili, Du Yaguang. Can Data Assetization Alleviate Financing Constraints for "Specialized, Innovative, Technologically Advanced, and Small and Medium-sized" Enterprises? *China Industrial Economy*, 2024(08): 154-173.
- [6] Peng Zhengyin, Meng Siyuan, Luo Guanqing. The Impact of Data Assetization on Enterprise Growth: "Adding Butter to Bread" or "Helping in Difficult Times". *Contemporary Finance*, 2025: 1-15.
- [7] Ju Xiaosheng, Lu Di, Yu Yihua. Financing Constraints, Working Capital Management, and the Sustainability of Enterprise Innovation. *Economic Research*, 2013, 48(01): 4-16.
- [8] Fang Xianming, Hu Ding. Enterprise ESG Performance and Innovation: Evidence from A-share Listed Companies. *Economic Research*, 2023, 58(02): 91-106.
- [9] Qi Hongqian, Zhang Jixin, Chen Miao. Research on the Promotion of Small and Medium-sized Enterprise Innovation Efficiency by Digital Inclusive Finance: From the Perspective of the Innovation Value Chain. *Macroeconomic Research*, 2023(10): 23-40.
- [10] Li Yuhua, Lin Yuxin, Li Dandan. How Does the Application of Artificial Intelligence Technology Affect Enterprise Innovation? *China Industrial Economy*, 2024(10): 155-173.
- [11] Tang Tian, Deng Beibei. The Impact Effect and Mechanism of Digital Transformation on Innovation: Empirical Study Based on Data of Chinese Listed Companies. *Statistics and Decision*, 2024, 40(21): 167-172.
- [12] Chen Wenrui, Hu Peiwen, Shi Xin, Liu Xinghe. Research on the Impact of Financial Strong Supervision on Enterprise Innovation. *Research Management*: 1-17.
- [13] Jones, Tonetti. Nonrivalry and the Economics of Data. *American Economic Review*, 2020, 9(110): 2819-2858.
- [14] Li Jian, Dong Xiaofan, Zhang Jinlin, Tao Yunqing. Research on the Impact of Data Assets on Enterprises' Innovation Investment. *Foreign Economics and Management*, 2023, 45(12): 18-33.
- [15] Li Haijun, Zhao Li. Data Becomes a Production Factor: Characteristics, Mechanisms and Evolution of Value Forms. *Shanghai Economic Research*, 2021(08): 48-59.
- [16] Farboodi. Long-Run Growth of Financial Data Technology. *American Economic Review*, 2020, 110(8): 2485-2523.
- [17] Farboodi. Big Data and Firm Dynamics. *AEA Papers and Proceedings*, 2019, 109: 38-42.
- [18] Juliane Begenau. Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 2018, 97: 71-87.
- [19] Bresnahan, Brynjolfsson, Hitt. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, 2002, 117(1): 339-376.
- [20] Richardson. Over-investment of free cash flow. *Review of Accounting Studies*, 2006, 11(2-3): 159-189.
- [21] Greene. *Econometric Analysis*. 8th. London: Pearson Education Limited, 2018
- [22] James Jondrow. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 1982, 19(2-3):233-238.
- [23] Xiao Wen, Lin Gaobang. Government Support, R&D Management and Technological Innovation Efficiency: An Empirical Analysis Based on Chinese Industrial Sectors. *Management World*, 2014(04): 71-80.
- [24] Wu Yanbing. Biased Investment under the Chinese-style Decentralization. *Economic Research*, 2017, 52(06): 137-152.
- [25] Jin Huan, Yu Lihong, Xu Yuanbin. Digital Economy, Factor Marketization and Enterprise Innovation Efficiency. *Economic Review*, 2024(05): 20-36.