

FACTORS INFLUENCING DATA ELEMENTS EMPOWERING NEW QUALITY PRODUCTIVE FORCES AND THEIR SPATIOTEMPORAL EVOLUTION: A STUDY BASED ON PANEL DATA FROM 31 PROVINCES IN CHINA

YuWen Zhu^{1*}, Ze Song², XiaoLing Xu¹

¹*School of Management, Tianjin University of Technology, Tianjin 300382, China.*

²*School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300382, China.*

Corresponding Author: YuWen Zhu, Email: 19515672598@163.com

Abstract: In the context of the deep integration between the digital and real economies, data elements have become a core force driving economic growth and social transformation. China has positioned market-oriented allocation and value realization of data elements as a national development strategy; however, challenges such as regional development imbalances and widening digital divides persist in the empowerment of new-quality productive forces. This study employs panel data from 31 Chinese provinces and municipalities between 2012 and 2022, integrating entropy-based TOPSIS models, ARIMA models, and spatial econometric models to systematically explore the impact mechanisms and spatiotemporal evolution characteristics of data elements on new-quality productive forces. The ARIMA model predicts that China's national coupling coordination degree will show a yearly growth trend over the next three years, with eastern regions outperforming central and western regions, though regional disparities are gradually narrowing. The spatial econometric model further reveals heterogeneous response characteristics of data element policies: "rapid effectiveness in eastern regions with delayed effects in central and western areas." This study aims to bridge the digital divide, optimize data governance, and provide robust support for achieving high-quality development goals in China's digital economy during the 14th Five-Year Plan period.

Keywords: Data element; New quality productivity; Regional development imbalance; Coupling coordination degree; ARIMA models

1 INTRODUCTION

Amid the rapid advancement of digital technologies and the ongoing Fourth Industrial Revolution, data elements are emerging as a new type of production factor, gradually becoming the core engine driving economic growth and social transformation. The World Economic Forum's 2023 Global Competitiveness Report indicates that data contributes over 15% to global economic growth, with its permeability, non-exclusivity, and technology-integration characteristics playing pivotal roles in restructuring production functions [1]. Theoretically, Liu Huachu and Tang Tang [2], building on Marx's theory of productive forces, propose that data drives qualitative transformations in new productivity through reshaping labor tools, objects, and workforce structures—a perspective echoing the assertion in The Second Machine Age [3] about exponential efficiency gains from data-driven intelligent technologies. Currently, China is vigorously promoting deep integration between the digital economy and real economy [4], elevating market-oriented allocation and value realization of data elements to a national strategic level. However, multiple challenges hinder empowering new productivity in practice: On one hand, the China Digital Economy Report (2023) [5] shows that China's digital economy reached 50.2 trillion yuan (41.5% of GDP) in 2022, yet significant regional development imbalances have widened spatial disparities in new productivity [6]; On the other hand, obstacles such as redundant platform construction, low utilization rates of public data, and insufficient professional service capabilities constrain value realization, further exacerbating resource allocation conflicts between emerging industries and traditional sectors [7].

Prior to this, some scholars have proposed perspectives on data elements and new quality productive forces. Wang Haijie and Wang Kaiyang [8] pointed out that data, as a new type of production factor, possesses characteristics such as non-excludability, high reusability, and technological integration. It can permeate the entire production chain, giving rise to intelligent production tools, high-end labor objects, and a digital workforce. However, there is a lack of concrete empirical data to support this. Although some countermeasures are mentioned, most remain at the macro level without specific implementation details or operational steps, making them difficult to implement. Xu Zhongyuan and Zheng Huangjie [9] proposed the theoretical foundation for empowering new-quality productive forces through marketization of data elements, namely data property rights theory, information sharing theory, and transaction cost theory, establishing a systematic, holistic, and collaborative data element market system. The shortcomings are as follows: While mentioning the practical challenges and constraints faced by current data element marketization, the analysis and case studies of these issues are relatively superficial, failing to thoroughly examine the actual challenges encountered during the empowerment process of new-quality productive forces. Hui Ning and Shi Xiaorong [10] used panel data from A-share manufacturing listed companies between 2015 and 2022 as research samples. By constructing fixed-effect models and conducting descriptive statistics, benchmark regression analysis, and various tests, they explored the impact

of data elements on enterprises' new-quality productive forces. However, they did not further analyze the time-space effects triggered by data elements or provide predictions for future trends in the development of new-quality productive forces.

Therefore, this study conducts a comprehensive analysis of existing research gaps, employing a dynamic-spatial dual-dimensional framework to quantify the direct impact of data elements on new-quality productivity and their spatiotemporal evolution patterns. Through a three-dimensional evaluation system encompassing "new-quality laborers, production materials, and work objects," we empirically demonstrate how data elements drive regional disparities via technology spillovers and resource optimization. By reconstructing regional differentiation patterns through an economic-geographic composite weighting matrix, this work challenges the conventional view that regional gaps stem solely from infrastructure limitations. Furthermore, integrating ARIMA forecasting with spatial effect decomposition, we propose a "regional coordination-stage adaptation" policy framework that translates spatiotemporal characteristics into actionable measures. This innovation provides both theoretical depth and practical value for optimizing data element market allocation and bridging the digital divide.

2 MODEL

2.1 Entropy TOPSIS Model

Data elements and new-quality productivity are the core research subjects of this study. The Entropy Value TOPSIS method not only avoids subjective factors affecting weight determination by automatically adjusting weights based on data variability, but also enables clear prioritization of outcomes through "positive ideal solutions" and "negative ideal solutions" to enhance visual clarity. Therefore, this study employed this method to evaluate new-quality productivity levels. The model calculation steps are as follows.

1) Min-max standardization of indicators

Forward pointer:

$$X_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad (1)$$

Negative indicators:

$$X_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \quad (2)$$

2) Calculate the information entropy of indicators

$$e_j = -\frac{1}{\ln(m)} \sum_{i=1}^n P_{ij} \ln(P_{ij}) \quad (3)$$

$$P_{ij} = \frac{X_{ij}}{\sum_{i=1}^m X_{ij}} \quad (4)$$

3) Calculate the proportion of each indicator

$$w_j = \frac{1 - e_j}{\sum_{j=1}^m e_j} \quad (5)$$

4) Calculate the weighted matrix

$$X = (x_{ij})_{m \times n} \quad (6)$$

$$R = (r_{ij})_{m \times n}, r_{ij} = w_j \times x_{ij} \quad (7)$$

($i = 1, 2, \dots, m; j = 1, 2, \dots, n$)

5) Determine the positive and negative ideal solutions

$$R_j^+ = \max(r_{1j}, r_{2j}, \dots, r_{nj}), R_j^- = \min(r_{1j}, r_{2j}, \dots, r_{nj}) \quad (8)$$

6) Calculate the Euclidean space distance between each object and the positive and negative ideal solutions

$$d_i^+ = \sqrt{\sum_{j=1}^m (R_{ij} - R_j^+)^2} \quad (9)$$

$$d_i^- = \sqrt{\sum_{j=1}^m (R_{ij} - R_j^-)^2} \quad (10)$$

7) Calculate the comprehensive evaluation index of each object

$$C_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (11)$$

2.2 ARIMA Model

Based on the calculated coupling coordination degree between data elements and new-quality productive forces, this study employs an ARIMA model to conduct a short-term trend prediction analysis. The ARIMA model primarily consists of four types: Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Moving Average with differencing (ARIMA). Given the non-stationary nature of the time series in this study, the ARIMA model was adopted for processing.

The basic formula for the ARIMA model is as follows:

$$\varphi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t \quad (12)$$

where Y_t represents the value of the time series at time t , L is the lag operator defined as $LY_t = Y_{t-1}$, $\varphi(L)$ is the autoregressive polynomial with the form $\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_m L^m$, where denotes the coefficients of the autoregressive component, $(1-L)^d$ is the difference operator representing the order difference operation applied to the original sequence for stationarity, $\theta(L)$ is the moving average polynomial with the form $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_n L^n$, where denotes the coefficients of the moving average component, ε_t represents the white noise error term, and c is the constant term.

1) Data stabilization

The ADF test is used to determine whether the series is stationary or not. If there is a unit root (p value > 0.05), the series should be differenced by order, which is shown in the formula $(1-L)^d Y_t$.

2) Determine model parameters

The difference order is typically determined by the results of the Autoregressive Integrated Function (ADF) test, where a value of 1 or 2 generally indicates non-stationarity can be eliminated. The autoregressive order can be identified by observing the partial autocorrelation function (PACF) plot of the differenced series. If the plot shows significant truncation beyond the confidence interval, an AR(p) model should be selected. For the moving average order, the Autocorrelation Function (ACF) plot serves as the diagnostic tool. When truncation occurs after the specified order, a MA(q) model becomes the appropriate choice.

3) Model fitting and parameter estimation

The parameters are solved by maximum likelihood estimation or least squares method according to the formula. The parameter combination is optimized using AIC/BIC criterion and the model with the minimum AIC/BIC value is selected.

4) Model test and diagnosis

White noise test (residual test): The residual sequence should satisfy the independence (ACF has no significant correlation), which is verified by Ljung-Box test.

5) Model prediction

The fitting model is used to predict the future value, and the prediction effect is verified by actual data.

2.3 Spatial Measurement Model

To further explore the spatial correlation between provincial data elements and new quality productivity and their radiation effect on neighboring provinces, this study adopts a spatial econometric model for more in-depth and detailed research. The core purpose of this model is to separate the direct effect (the effect of local variables) and the indirect effect (spatial spillover effect) generated by the variables.

2.3.1 Build the spatial weight matrix

The spatial weight matrix is mainly divided into the adjacency weight matrix, geographical distance weight matrix, and economic distance weight matrix. A single type of matrix cannot completely reflect the economic and geographical correlation of provinces; therefore, this study adopts a combination of both to construct the economic-geographical distance weight matrix.

$$W_{ij} = \begin{cases} \frac{|\overline{Q_i} - \overline{Q_j}|}{(Y_i - Y_j)^2}, & i \neq j \\ 0, & i = j \end{cases} \quad (13)$$

where $\overline{Q_i}$ and $\overline{Q_j}$ are the average GDP of the two provinces, and $(Y_i - Y_j)^2$ is the square of the distance between the two provinces.

2.3.2 Spatial autocorrelation test

Global Moran's I test:

$$I = \frac{n}{\sum_i \sum_j W_{ij}} \cdot \frac{\sum_i \sum_j W_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \quad (14)$$

If the value is significant and greater than zero, there is a positive spatial correlation.

Local Moran's I test:

$$I' = \frac{Y_i - \bar{Y}}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \cdot \sum_{j=1}^n W_{ij} (Y_j - \bar{Y}) \quad (15)$$

Similarly, if the value is significant and greater than zero, there is a spatial positive correlation. Based on this, Moran's I scatter plot was drawn to identify local hot spots (High-High) and cold spots (Low-Low) regions.

2.3.3 Classification of spatial measurement models

Spatial econometric models are mainly divided into three types: spatial lag model (SLM), spatial error model (SEM), and spatial Durbin model (SDM).

Space lag model (SLM):

$$y = \lambda W y + X \beta + \varepsilon \quad (16)$$

where y is the dependent variable vector, λ is the lag coefficient of the dependent variable, which represents the intensity of spatial spillover, W is the spatial weight matrix, $W y$ is the spatial lag term, X is the explanatory variable matrix, β is the corresponding coefficient vector, and ε is the error term vector.

Spatial error model (SEM):

$$y = X \beta + \varepsilon, \varepsilon = \rho W \varepsilon + \mu \quad (17)$$

where β contains spatial correlation, ρ is the spatial autoregressive coefficient, and μ is an independent, identically distributed random error vector.

Spatial Dubin model (SDM):

$$y = \lambda W y + X \beta + W X \theta + \varepsilon \quad (18)$$

where λ and θ are the lag coefficients of the dependent and independent variables, respectively.

2.3.4 Model test and selection

When selecting a spatial econometric model, the process begins with the least maximum (LM) test to determine the appropriate model type. The Hausman test then determines whether to use a fixed-effects or random effects model. If a fixed-effects model is chosen, further verification is required to decide between individual fixed effects and time-fixed effects. If both tests are passed, the two-way fixed-effects model becomes viable. Additionally, if the LM test selects a spatial Durbin model, Wald and likelihood ratio (LR) tests must be conducted separately to assess potential degradation into spatial lag or spatial error models.

2.3.5 Space effect decomposition

After selecting the research model, the sample data of 31 provinces were decomposed by spatial effects to analyze the significance degree of explanatory variables and control variables in direct, indirect, and total effects, so as to explore the influence degree on the explained variable in the province and surrounding areas.

3 RESULTS AND ANALYSIS

3.1 Evaluation Index System and Benchmark Model Construction

The numerical indicators related to new productive forces discussed in this study primarily draw from the research of Han Wenlong et al., while those concerning data elements mainly reference studies by Ye Lu and colleagues. The specific data acquisition channels included the CSMAR database, EPSDATA official website, and annual statistical yearbooks. The control variable data involved in this study were mainly obtained through authoritative sources, such as the National Bureau of Statistics official website, provincial statistical yearbooks, and the China Statistical Yearbook, ensuring data accuracy and consistency.

3.1.1 Evaluation index system and benchmark model construction

In this study, the new quality productivity data collected were divided into three index layers, and the comprehensive evaluation index system of new quality productivity was established, as shown in Table 1. The entropy TOPSIS method is used to assign certain weights to each third-level index to calculate the annual level of new quality productivity in each province.

Table 1 Comprehensive Evaluation Index System of New Quality Productivity

The first indicator	The second indicator	Third indicator	Direction
New quality workers	Workers in emerging industries	Total number of employees in emerging industries	+
	Technologically innovative workers	The number of R&D personnel in high-tech enterprises	+
New qualitative means of production	Digitization of labor data	Robot installation density	+
		Number of mobile users	+
		Integrated circuit production	+
		Total telecom business volume	+
		Software revenue	+
		E-commerce sales	+
		Number of Internet broadband access ports	+
		Optical cable line length/area of the region	+

		Mobile Internet access traffic	+
	Flexible and customized labor data	Research and development investment of high-tech enterprises	+
		The number of R&D institutions in high-tech enterprises	+
		Regulate industrial innovation funds for industrial enterprises above designated size	+
		Full-time equivalent of R&D personnel in industrial enterprises above designated size	+
		Number of patents granted by region	+
		Revenue from high-tech industries	+
	Environmental protection and energy conservation of labor materials	Energy consumption/GDP	-
		Industrial water use/gross domestic product	-
		Industrial wastewater discharge/gross domestic product	-
		Industrial SO ₂ emissions/gross domestic product	-
		Comprehensive utilization/production of industrial solid waste	+
New quality objects of labor	The object of digitalization and informatization	Number of data exchanges	+
	Green environmental protection and sustainable development of labor objects	Investment in industrial pollution control	+
	The object of labor for high-end equipment and intelligent manufacturing	Number of enterprises with e-commerce transactions	+
		Number of AI enterprises	+
	The object of labor in the future industry	Proportion of new energy generation	+
		Number of ultra-high voltage transmission lines	+
		New energy utilization efficiency	+
		Output value of new materials	+
		Number of new material enterprises	+

3.1.2 Data element evaluation index system

In this study, starting from the entire process of data value conversion, data elements are categorized into three levels: data generation, data transformation, and data application. Each level is further explained using the corresponding indicators, as shown in Table 2. The entropy value TOPSIS method was applied to assign weights to each indicator, thereby determining the annual data element levels for each province.

Table 2 Comprehensive Evaluation Index System of Data Elements

The first indicator	The second indicator	Third indicator	Direction
Data elements	Data generation	Number of pages	+
		Number of domain names	+
		There are 100 websites owned by every 100 enterprises	+
	Data transformation	Revenue from information technology services	+
		Information security revenue	+
		Digital financial insurance	+
	Data applications	Digital financial payments	+
		Degree of digitization of digital financial data	+

3.2 Short-term Trend Analysis of Coupling Coordination Degree

To explore the short-term trend of the coupling coordination degree nationwide and regionally, the ARIMA model was used for prediction and analysis. Based on the model selection and white noise test results, considering the limited sample size, we fitted the national and regional coupling coordination levels to predict their development trends from 2023 to 2025, with specific results shown in Figure 1.

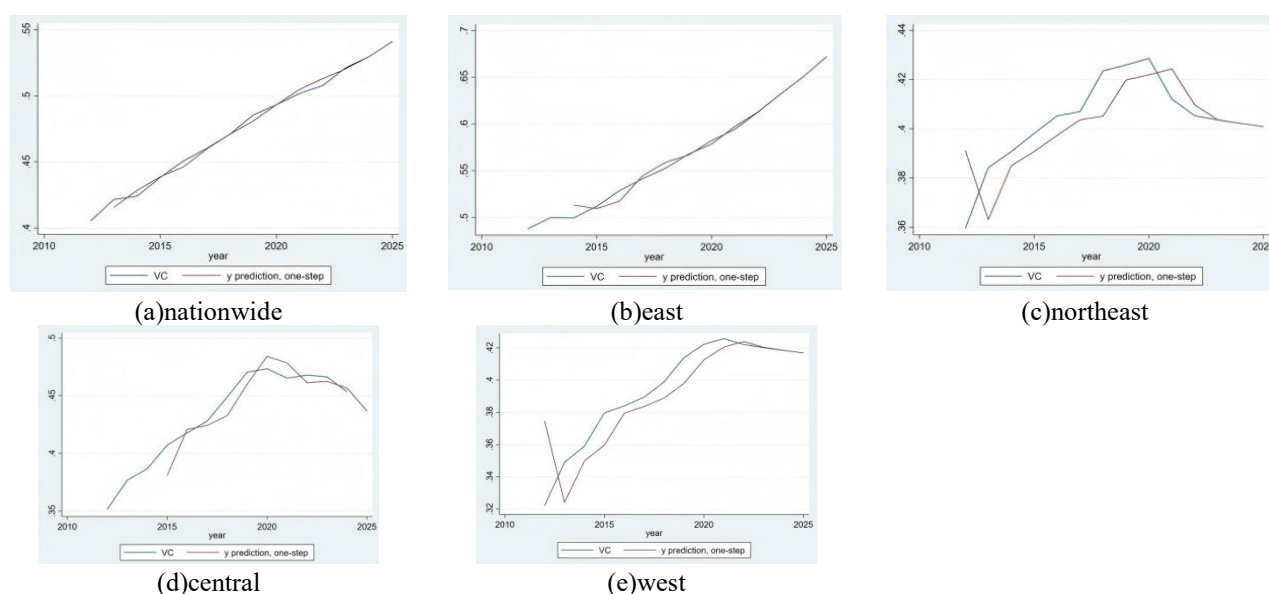


Figure 1 Trend Forecast of Coupling Coordination Degree Nationwide and Regionally

The fitting results indicate that the national coupling coordination level exhibits a steady upward trend. The predicted values lagged by one period compared to the actual values closely matched, demonstrating strong predictive validity and suggesting continued growth over the next three years. The eastern regions exhibited similar upward trends with consistent predictions, maintaining rising levels through the forecast period. The coupling coordination in Northeast China peaked around 2020 before declining, likely due to pandemic's impact, although the overall upward trend persisted with close alignment between the actual and predicted values. This region is expected to maintain this level over the next three years. Similarly, the central regions experienced a decline around 2020 due to the pandemic effects, with discrepancies between the predictions and actual values caused by the second-order lag data. However, their overall downward trend aligns with the predicted outcomes, which is likely attributed to the uneven development of data elements. The western regions experienced a notable slowdown in growth around 2015, followed by rapid recovery and stabilization post-2020. The predictions closely matched the actual values, indicating a probable maintenance of stable trends over the next three years.

3.3 Spatial Evolution Analysis

3.3.1 Correlation analysis between data elements and collaborative development space of new quality productivity

To further study the spatial correlation between the coordinated development of data elements and new quality productivity levels in different provinces, this study first uses the Moran index to analyze the spatial autocorrelation of the coupling coordination degree.

Table 3 Global Moran Index of Spatial Autocorrelation of Coupling Coordination Degree in 31 Provinces from 2012 to 2022

year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Moran'I	0.082	0.074	0.064	0.052	0.069	0.073	0.072	0.072	0.065	0.069	0.080
Price Z		2.069	1.866	1.642	1.971	2.056	2.043	2.048	1.939	2.009	2.203
	2.191										
Price P	2.029	0.039	0.062	0.101	0.049	0.040	0.041	0.041	0.053	0.045	0.028

As shown in Table 3, the global Moran indices for both subsystems remained above zero from 2012 to 2022, indicating a significant spatial positive correlation in their coordinated development. Only three years showed P-values exceeding the 0.05 significance threshold, with the overall test passing statistical validity for the remaining years. Notably, around 2015, when eastern coastal provinces such as Guangdong and Zhejiang initiated big data industry policy pilots, this initiative may have temporarily disrupted spatial correlation stability. Meanwhile, delayed policy responses in the central and western regions resulted in a slight decline in the global spatial correlation during that period.

3.3.2 Spatial econometric regression results and spatial effect analysis

After choosing the Dubin effect two-way fixed model, regression analysis was performed on the model to explore the significance of the spatial influence of data elements and a series of control variables on new quality productivity. The regression results are presented in Table 4.

Table 4 Regression Results of Spatial Dubin Model

	(1) Back β to results β	(2) WX Back to results WX
<i>Del</i>	0.179*** (0.040)	0.320*** (0.231)
<i>Led</i>	0.003*** (0.002)	0.056** (0.009)
<i>Rdi</i>	3.794 (0.067)	-2.268 (3.249)
<i>Dow</i>	-0.046* (0.019)	0.305* (0.107)
<i>Hcl</i>	-2.262 (0.195)	-0.212 (0.924)
<i>Gil</i>	0.090** (0.037)	0.064 (0.216)

***p<0.01, **p<0.05, *p<0.10

Regression analysis reveals that both data factor levels and economic development levels reach the 1% significance level in terms of β values, with coefficients of 0.179 and 0.003, respectively. This indicates a positive spatial interaction between these factors and new-quality productivity. The degree of openness to the outside world and government intervention shows significant effects at the 10% and 5% levels, with coefficients of -0.046 and 0.09, respectively. These findings suggest that increased openness may lead to a slight loss of data factors, thereby affecting new quality productivity levels, while government policies play a positive regulatory role.

The WX value further explains the spatial spillover effect based on β . The data element level, economic development level, and degree of opening to the outside world show significant effects at the 1%, 5%, and 10% levels, respectively, with coefficients of 0.320, 0.056, and 0.305. This indicates that all three factors exhibit positive spatial spillover effects, demonstrating that neighboring provinces exert a positive transmission effect on their own province's new quality productivity level.

4 CONCLUSIONS AND OUTLOOKS

Through an in-depth analysis of panel data from 31 provinces and municipalities spanning 2012-2022, this study systematically explores the impact mechanisms and spatiotemporal evolution characteristics of data elements on new-quality productive forces. The results reveal that data elements play a significant positive role in driving the development of new-quality productive forces, while the influence of economic development level and R&D intensity remains noteworthy. In terms of temporal evolution, ARIMA model predictions indicate that the national coordination level will continue to rise over the next three years, signaling a new phase in the coordinated development of data elements and new quality productive forces. Regarding spatial evolution, spatial autocorrelation tests revealed spatial correlations between the data elements and the new-quality productive forces.

Although this study developed a comprehensive model to analyze the impact of data elements on new-quality productivity and their spatiotemporal evolution characteristics, several limitations remain. First, the model primarily relies on panel data for processing, which may fail to fully capture dynamic changes and nonlinear relationships in the time series. Second, although the model incorporates multiple dimensions of indicators when constructing the coupling coordination degree, it still exhibits subjectivity and limitations in indicator selection, failing to comprehensively reflect all relevant factors. Additionally, the prediction component mainly depends on ARIMA models, which could affect accuracy when dealing with complex economic environments and policy changes. Future research can enhance the model's explanatory power and predictive precision by introducing additional data sources and refining the modeling methodologies.

COMPETING INTERESTS

The authors declare no relevant financial or non-financial interests.

REFERENCES

- [1] Klaus Schwab. The Fourth Industrial Revolution: The Power of Transformation. Investment & Cooperation, 2016(12): 105.
- [2] Liu Huachu, Tang Tang. Data Elements Empowering the Emergence of New Quality Productive Forces — Theoretical Cornerstones, Internal Logic, and Operational Mechanisms. Journal of Beijing University of Aeronautics and Astronautics Social Sciences Edition, 2025, 38(3): 1-9. DOI: 10.13766/j.bhsk.1008-2204.2024.1422.
- [3] Eric Brynjolfsson, Andrew McAfee. The Second Machine Revolution. Theory and Contemporary, 2014(9): 53.
- [4] Han Wenlong, Zhang Ruisheng, Zhao Feng. Measurement of New Quality Productivity Level and New Drivers of China's Economic Growth. Journal of Quantitative & Technical Economics, 2024, 41(6): 5-25.

- [5] Chong T T L, Wang S, Zhang C. Understanding the digital economy in China: Characteristics, challenges, and prospects. *Economic and Political Studies*, 2023, 11(4): 419-440.
- [6] Liu W, Lin X. Data transaction mode and its legal regulation in the context of market-oriented allocation of data elements. In: Hong, W., Kanaparan, G. (eds) *Computer Science and Education. Computer Science and Technology. ICCSE 2023. Communications in Computer and Information Science*, Singapore. 2023, 42-51. DOI: https://doi.org/10.1007/978-981-97-0730-0_5.
- [7] Wagner H. Facility Updates and Machine Upgrades. *Synchrotron radiation news*, 2023, 36(1): 2-2.
- [8] Wang Haijie, Wang Kaiyang. Mechanisms, Challenges and Countermeasures for Data-Driven Development of New Quality Productivity. *China Circulation Economy*, 2025, 39(1): 3-13.
- [9] Xu Zhongyuan, Zheng Huangjie. Empowering New Productive Forces: Legal Allocation of Data Element Marketization. *Theory and Reform*, 2024(6): 63-80.
- [10] Hui Ning, Shi Xiaorong. Research on the Impact of Data Elements on the New Quality Productivity of Enterprises. *Journal of Northwest University (Philosophy and Social Sciences Edition)*, 2025, 55(3): 81-98. DOI: 10.16152/j.cnki.xdxbsk.2025-03-008.