China's Economic Growth: The "Two-Dimensional Driving Effect" of Data Factors

Yan Yang, Li Wang, Yujia Li, Zujun Liao*

Data factors have become one of the five essential production factors, but their role in economic growth has always been ambiguous. Starting from AI technologies, this paper establishes an endogenous growth model of data factors affecting economic growth, constructs the generation path and value path of data factors, and estimates the value of new data factors at the provincial level in China from 1999 to 2018 accordingly. Based on theoretical analyses and empirical tests, it clarifies that data factors have a "two-dimensional driving effect" on China's economic growth, that is, data factors can drive growth both directly through its own economic growth effect and indirectly by promoting technological progress. Furthermore, this paper makes three extended discussions, aiming to make a trial study on the impacts of local government big data transaction platforms on data factors and their growth effects, discuss whether it is possible to reduce the uncertainties of local economic policy based on the nature of data factors, and make a preliminary survey of the output elasticity of data factors between 1999 and 2018.
Keywords: data factors, economic growth, AI; economic, growth effect

1. Introduction

The core role of production factors is to promote economic growth. Data, as a key factor and one of the five essential production factors (namely data, land, labor, capital and technology) listed in the *Guidelines on Establishing Better Systems and Mechanism for Market-Based Allocation of Factors of Production* issued in April 2020. The positive effects of land, labor, physical capital and technological progress on economic growth have been confirmed in literature. Nevertheless, as for the impacts of data factors on economic growth, and such issues as whether data factors have economic growth effects, there has not been a definite conclusion yet.

* Yan Yang, Professor at the School of Economics, Sichuan University; Li Wang (email: 447161772@qq.com) , Doctoral student of the School of Economics, Sichuan University; Yujia Li, Assistant Research Fellow of the School of Economics, Sichuan University; Zujun Liao, Research Fellow and Director of the Institute of Regional Economy of Sichuan Academy of Social Sciences. Fund project: "Research on System Regulation on High-quality Supply of Data Factors under the Framework of 'Market + Government + Community' Collaborative Governance", a National Social Science Fund Project for 2022. (22BJL033).

Open Access. © 2023 The Author(s), published by De Gruyter.
This work is licensed under the Creative Commons Attribution 4.0 License.

As seen from existing researches, data factors may influence economic growth by making "two-dimensional driving effect", namely, directly promote economic growth through data factors and indirectly by facilitating the generation and accumulation of new knowledge and ideas, namely making technological progress. To empirically test the existence of the two-dimensional driving effect of data factors, we need to acquire "data" of data factors and find "technologies" that are closely related to the indirect driving effect of data factors. At the present stage, the promoting effect of AI technologies on economic growth can best reflect the indirect driving effect of data factors due to the following reasons. First, AI needs to achieve self-accumulation, update and iteration by relying on data (Lu, 2021), and data factor is an absolute impetus to the development of AI (Jones and Tonetti, 2020). Second, the advancement of AI technologies can help the society acquire and use data factors more effectively (Veldkamp and Chung, 2022; Farboodi and Veldkamp, 2021). Third, traditional production factors such as material capital and common labor may not give a significant boost to economic growth any more, while AI is one of key factors determining future economic growth (Aghion *et al.*, 2019; Lu, 2021).

In view of this, starting from the direct driving effect of data factors and their indirect driving effect on AI technological progress, this paper takes overall consideration of the two-dimensional driving effect of data factors on China's economic growth. First, we establish an endogenous growth model including data factors and AI accumulation, theoretically describing the direct and indirect driving effects of data factors on economic growth. Second, by using Statistics Canada (2019a, 2019b) as well as relevant researches of Xu and Zhao (2020), we construct the generation path and value path of data factors, and estimate the value of new data factors at the provincial level in China from 1999 to 2018 in accordance with China's reality to represent the scale of new data factors, and make it "data" of the data factors in the follow-up empirical analyses of this paper. Finally, we conduct empirical tests to prove that data factors can promote economic growth through the "two-dimensional driving effect", thereby proving that data factors have also effects on economic growth.

The three main contributions of this paper are shown in the following three aspects. First, theoretically, starting from the basic assumption that data come from household consumption, we include data factors into the equation of accumulation of AI technologies and the function of enterprise production, deduce the two-dimensional driving effect of data factors on economic growth and its mechanism, and proves theoretically that data factors have both "level effect" and "growth effect". Second, in the aspect of value calculation of data factors, by referring to existing documents and based on the social and economic environments of China, this paper constructs the generation path and value path of data factor and estimates the annual value of new data factors at the provincial level in China, so as to provide practical references for follow-up researches of this paper and relevant researches of other scholars. Third, in reality, it has confirmed that data factors have obvious promoting effects on economic growth, providing strong supports for the rationality that China lists data factors together with land, labor, capital and technology as fundamental production factors, and helping governments at all levels attach greater importance to and strengthen application of data factors.

The following structure of this paper is arranged as follows. Starting from AI technologies, Section 2 establishes a "two-dimensional driving effect" model about the effect of data factors on economic growth; Section 3 estimates the value of new data factors at the provincial level in China from 1999 to 2018; Section 4 introduces empirical strategies, variables and data; Section 5 analyzes the process and results of empirical tests; Section 6 summarizes research results and policy implications.

2. Theoretical Analyses

2.1. Environmental Settings of Generation and Use of Data Factors

On the principles of "data are by-products of economic activities" (Jones and Tonetti, 2020; Cong *et al*., 2021; Farboodi and Veldkamp, 2021) and the implication that China deems "data" the new-type production factors, this paper transforms such issues as the generation of data factors and data factors facilitates innovation and AI technological progress into a macro-economic model to represent economic growth. To make the analyses simple, we assume that an economic entity consists of large quantities of representative households and representative firms, firms are owned by households, the economic environment is definite and the duration of household is continuous and infinite.

2.1.1. Representative Household

This paper sets data factors as by-products of consumption:[1]

$$D = Nc \tag{1}$$

Existing foreign literature do not distinguish data factors from data. However, among the massive data generated in social and economic activities, only some valuable "effective data" can help social and economic operations and conform to the

[1] In some literature, data are set as output of capital. For example, Farboodi and Veldkamp (2021) proposed the concept of data point, considering that the number of data points a firm observes at the end of each period depends on their capital output. Just like Jones and Tonetti (2020), this paper deems data factors by-products of consumption, aiming mainly to simplify the analyses and not consider data factors in the process of capital accumulation.

connotations of data factors in China. Therefore, the differences between this paper and Jones and Tonetti (2020) and Cong *et al*. (2021) regarding data generation settings are shown below: 1 unit consumption can generate $N \in [0, \infty)$ unit data factors D rather than 1 unit data. Households own data factors, which fully depreciate after being used in each period.[1]

Assuming that homogenous representative households own 1 unit time, of which u is used for study, v is used for research and development of AI technologies (AI), and the remaining $1-u-v$ is used for producing goods. Households input material capital s in AI research and development and $1-s$ for producing goods. It is worth noting that, due to the non-competitiveness of data, a household can use all data factors for AI research and development and production of goods. Similar to Lucas (1988), household human capital is accumulated through learning. Based on the view of Lu (2021), AI can achieve self-enhancement and accumulation by relying on deep learning. In addition, AI research and development need not only computers (material capital) and programmers (human capital), but also mass data for algorithm training. Therefore, the accumulative process of human capital and AI is shown below:

$$\dot{h} = Buh \tag{2}$$

$$\dot{A} = M(vh)^{\phi}(sk)^{\theta} D^{\varphi} A^{1-\phi-\theta-\varphi} \tag{3}$$

Wherein, B and $M > 0$ measure the cumulative efficiency of human capital and AI, respectively, and the value range of ϕ, θ, φ and $1-\phi-\theta-\varphi$ is (0,1). Indicating k, h, A as material capital, human capital and AI, and its initial stock is $k(0)$, $h(0)$, $A(0)$, respectively. The budget constraint of a representative household is shown below:

$$\dot{k} = w(1-u-v) + r(1-s)k + r_A + r_D + \pi - c \tag{4}$$

Wherein, π denotes corporate profit, w and r denote wage rate and rental rate, respectively, r_A and r_D denote return of AI and return of data factors, respectively.

The utility of a representative household is shown below:

[1] Although data can be accumulated, they will not attenuate or be consumed like machine, equipment, building and natural resources (Farboodi *et al.*, 2019). However, for most economic activities, the effectiveness of data reduces over time, and their economic value may depreciate, so do data factors. Therefore, in order to avoid excessive state variables in analyses, this paper assumes that data factors fully depreciates after being used in each period.

$$U = \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\sigma} - 1}{1-\sigma} dt \tag{5}$$

2.1.2. Representative Firm

Different from traditional machines replacing human for simple physical labor, AI, which keeps learning and enhancing by relying on data factors, can gradually deal with increasingly extensive complex realistic scenes (for examples, automatic driving algorithm, medical image processing, etc.) in place of human. Meanwhile, the non-competitiveness of AI can make it used for both R&D and production in each period. In a competitive environment, representative firms produce products by employing labors and renting material capital, AI and data factors, and its production function is set below:

$$y = F[(1-s)k]^{1-\alpha-\delta} \{(1-\alpha)[(1-u-v)h]^{\beta} + aA^{\beta}\}^{\frac{\alpha}{\beta}} D^{\delta} \tag{6}$$

Wherein, $F > 0$ is a firm's production efficiency, α, δ, $1-\alpha-\delta \in (0,1)$ are share of labor, data factors and material capital, respectively. $\{(1-\alpha)[(1-u-v)h]^{\beta} + aA^{\beta}\}^{\alpha/\beta}$ shows that AI and human labor can replace each other, $\beta < 1$ is the substitution parameter of human capital and AI, and $\alpha \in (0,1)$ measures the importance of AI for human capital.

Within a single time period *t*, representative firm's profit π is:

$$\pi = y - w(1-u-v)h - r(1-s)k - r_A A - r_D D \tag{7}$$

2.2. "Two-Dimensional Driving Effect" on the Balanced Growth Path

The balanced growth path consists of the quantity path $\{c(t), v(t), u(t), y(t), k(t), h(t), A(t), D(t)\}_{t=0}^{\infty}$ and the price path $\{r(t), r_k(t), r_A(t), r_D(t)\}_{t=0}^{\infty}$, maximizing household lifetime utility and corporate profit. Meanwhile, commodity market, capital market, labor market and data market meet the conditions of market clearing. From this we can obtain economic growth rate g_y,[1]

$$g_y = Bu = \psi(\frac{k}{h}) \times (\frac{D}{h})^{\delta} + \chi(\frac{k}{h}, \frac{D}{h}) \times (\frac{A}{h})^{\beta} \tag{8}$$

[1] Due to space constraints, the detailed process of derivation has been omitted. If you are interested in it, please contact the authors. The same below.

Wherein,

$$g_y = \frac{\dot{y}}{y} = \frac{\dot{c}}{c} = \frac{\dot{k}}{k} = \frac{\dot{h}}{h} = \frac{\dot{A}}{A} = \frac{\dot{D}}{D} = Bu \tag{9}$$

$$\psi(\frac{k}{h}) = \frac{1-\alpha}{a} Fu\sigma(1-s)^{-\alpha-\delta}(1-a)(1-u-v)\beta(\frac{k}{h})^{-\alpha-\delta} \tag{10}$$

$$\chi(\frac{k}{h},\frac{D}{h}) = \frac{1-\alpha}{a} Fu\sigma(1-s)^{-\alpha-\delta} a(\frac{k}{h})^{-\alpha-\delta}(\frac{D}{h})^{\delta} \tag{11}$$

$$\frac{k}{h} = \{\frac{B^{-\varphi}}{M} uv^{-\phi} s^{-\theta} N^{-\varphi} (\frac{1-s}{1-\alpha-\delta})^{-\varphi} [\frac{(1-a)v}{a\phi}(\frac{1}{u}-1+\phi+\theta+\varphi)(1-u-v)^{\beta-1}]^{\frac{\phi+\theta+\varphi}{\beta}}\}^{\frac{1}{\theta+\varphi}} \tag{12}$$

The "two-dimensional driving effect" of data factors on economic growth is reflected in Formula (8): on the one hand, the exponent of $(D/h)^{\delta}$ is positive, indicating that the more data factors in an economic entity, the faster the economy grows, reflecting the direct drive effect of data factors and indicating that data factors have an effect on economic growth; on the other hand, the exponent of $(A/h)^{\beta}$ is positive and $\dot{A}$ is under positive impact of D in Formula (3), indicating that when the data factors in an economic entity keeps increasing, the faster AI technologies progress, the faster the economy grows, showing the indirect drive effect of data factors.

3. Value Estimation of Data Factors

3.1. The Generation Path and Value Path of Data Factors

Accumulated data is a type of valuable assets (Farboodi and Veldkamp, 2021). Theoretically, the three methods used usually for assessing any asset value (income method, market method and cost method) should also apply to data factors. As far as China's reality at present, besides a small part of data factors that have been traded on various data transaction platforms, there are a mass of data factors which have not been sold on the market,[1] making income method and market method not

[1] In society there are massive public data in governments or non-profit organizations. These data are mostly secret-related or involve personal privacy, such as data of medical insurance, social insurance and public medical services. Such data are only limited to internal use and accumulation in governments or non-profit organizations and cannot be sold on the market for the moment.

suitable for value calculation of data factors. Data is a type of intangible capital, and its value is generally calculated by using the cost method (Veldkamp and Chung, 2022). Therefore, it is feasible to evaluate the value of data factors at present by indirectly estimating the total costs for generating, collecting, sorting and using data factors.

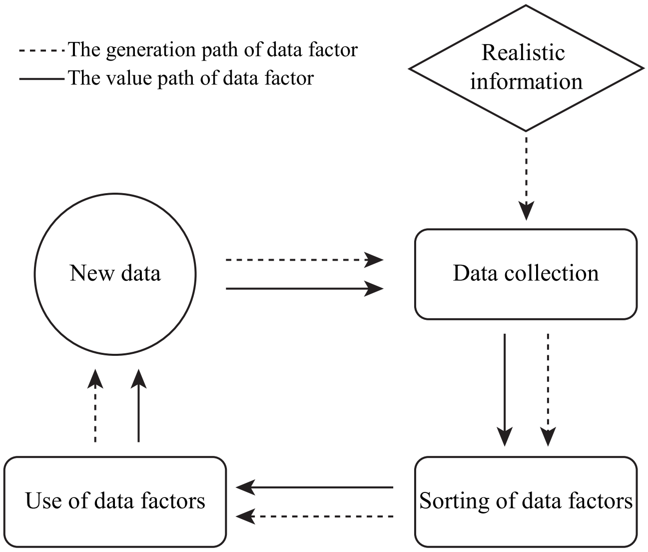

Figure 1. The Generation Path and Value Path of Data Factors

In this paper, the generation path and value path of data factors are summarized in Figure 1. Next, this paper tries to make a preliminary estimation of the value of provincial annual new data factors in China[1] based on the value path of data factor, by using the thoughts in Statistics Canada (2019b) on measuring data values for reference, and referring to the researches of Xu and Zhao (2020) on using microscopic database to estimate the value of data asset stock.

3.2. Value Estimation of New Data Factors and Rationality Test

In the value path of data factors, if the cost method is used to estimate the value of data factors, it shall involve the wage cost as well as relevant non-direct labor costs and other costs for such activities as data generation, collection, arrangement and use of

[1] Based on the theoretical assumption in Section 2: “The data factors in each period will fully depreciate after being used and will not be accumulated as data assets”. Therefore, this paper only estimates the value of data factors that are newly added each year without considering the stock value of data assets, and makes the estimated value of newly added annual data factors proxy indicators in follow-up empirical analyses.

data factor,[1] however, laborers engaging in such activities do not use all their working time for generating data value. Therefore, this paper, first of all, on the basis of the occupational and industrial codes in the China Family Panel Studies Database, sorts and classifies occupations related to such activities as data collection, organization and use of data factors (see Table 1), and then conducts online and offline interviews and surveys on 20 domestic organizations with relatively centralized laborers who are engaged in the above occupations,[2] so as to roughly obtain the activities of their participation in collection, organization and use of data factors in reality, average daily effective working hours (based on an eight-hour day working system, and overtime is not considered for the moment) which can directly generate the value of data factors, as well as relevant average indirect wages and other costs. Although there are certain errors in the practice, and the interview samples cannot represent the general situation of all years, however, limited by objective conditions, the practice in this paper using research interviews to acquire cost data has improved subjective settings in Statistics Canada (2019b) such as "indirect wages and other costs account for 50% of the wage cost" and "the share of production activity data is 10%~100%", and is thus more suitable for China' realistic environment. Table 1 presents the average effective working hours of those who are engaged in data collection, organization and use of data factors, and average share of indirect wages and other costs in wages. The formula for value estimation of annual new data factors is set below:

$$Value\_Data_t = \sum_j \sum_i wage_{ijt} \left(\frac{work\_time_j}{8} + other_j \right) \tag{13}$$

Wherein, subscript i denotes occupation, j denotes type of data activity, t denotes year. $Value\_Data_t$ denotes the value of new data factors in year t. $wage_{ijt}$ denotes the wage of occupation i in activity j in year t. $work\_time_j$ denotes the average effective working hours per day of all occupations in activity j, $other_j$ denotes the mean value of the proportion of indirect wages and other costs related to all occupations in activity j to wages of corresponding occupations.[3]

[1] Non-labor direct costs and other costs roughly include relevant costs of water, electricity, building maintenance and rental, telecommunications services, social insurance, medical insurance, office equipment and HR management.

[2] The 20 institutions include universities, scientific research institutions, government departments, provincial big data centers, big data firms, telecom operators and some financial institutions.

[3] As is seen from the interview, though wages of various occupations increase over time, "average effective working hours per day (hour)" and "average share of indirect wages and other costs in wages (%)" were always within certain limits and did not fluctuate significantly through the years. Therefore, this paper assumes that they do not change over time and have no subscript t.

Table 1. Average Proportions of Occupations, Average Effective Working Hours, Indirect Wages and Other Costs

Type of activities	Relevant occupations	Average effective working hours per day (hour)	Average share of indirect wages and other costs in wages (%)
Data collection	Engineering technician, agricultural technician, aircraft and ship technician, professional medical workers, financial service personnel, reporter, administrative staff, postal and telecommunications personnel, geological surveyor, surveying and mapping personnel, environmental monitoring and waste disposal personnel, inspection and metrology personnel	5.5	45
Organization of data factor	Economic and business personnel, literature and archives personnel, editor, proofreader and translator.	5	20
Use of data factor	Scientific researcher	7	30

In accordance with Formula (13) for value estimation of annual new data factors, we make a preliminary estimation of provincial new value of data factors in China between 1999 and 2018 on the basis of indicators in Table 1 by using relevant data in statistical yearbooks of local governments, *China Labor Statistical Yearbook*, China Family Panel Studies (CFPS), China Labor-force Dynamics Survey (CLDS), and Chinese General Social Survey (CGSS). Figure 2 randomly presents the estimated results of the annual new value of data factors of some provinces and cities[1], from which we can see that after the first year of the era of big data (2013), the values of new data factors have significantly increased in all regions, which is in correspondence with the big data development of China.

Next, this paper uses two major indicators (Table 2[2]) to conduct rationality tests on the estimated value of new data factors based on type and time period. The first type of indicators are individual indicators, of which the first four indicators represent scale and volume of data factors and the last two indicators themselves are data. The second types of indicators are comprehensive indicators, including “Digital Financial Inclusion Index” and “Digital Economy Comprehensive Development Index”, the former is from Peking University Digital Financial Inclusion Index (2011–2020), while the latter is calculated by referring to Zhao *et al.* (2020). All variable indicators have been standardized in tests.

[1] Due to limited space, the estimated results of the values of annual new data factors of all provinces are omitted. If you are interested in these estimated results, please contact the authors.

[2] All data of individual indicators are from the macroeconomic database of the EPS (Easy Professional Superior) data platform. The years with statistical data are different but temporally continuous, and no data is missing. The years of the last two comprehensive indicators are 2011–2018, and no data is missing. Therefore, it is more convincing to use indicators of different type and time period to test the rationality of value estimation results of annual new data factor.

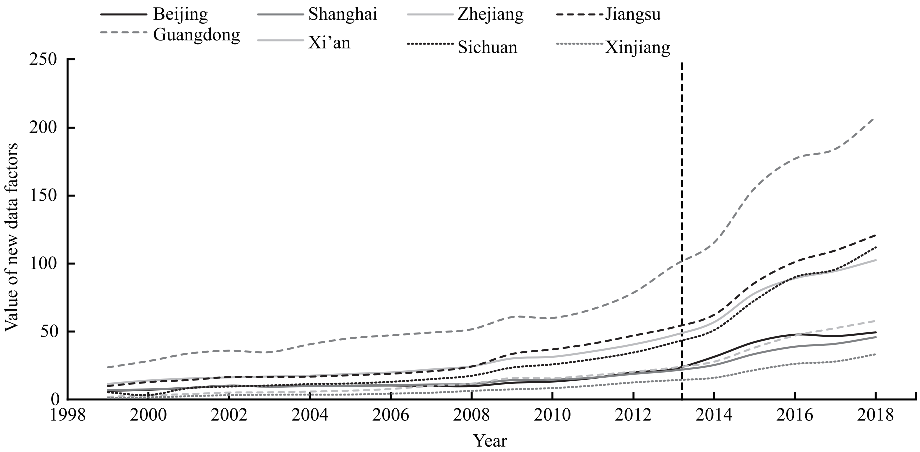

Figure 2. Values of New Data Factors in Some Provinces and Cities (1999–2018)

As is seen from the result of rationality test, the estimated values of new data factors have strong positive correlation with indicators of the first type and the second type,[1] indicating that the value of new data factors in the provincial level in China between 1999 and 2018 estimated by constructing the value path of data factors has certain rationality both theoretically and practically.

Table 2. Indicators for Rationality Test of Value of New Data Factors

Type of indicator	Relevant indicators	Time span (year)	Number of observations	Minimum value	Maximum value
Individual indicators	Mobile base station (10000)	2013−2018	180	1.18	65.06
	IT service revenue (RMB 100 million)	2014−2018	150	0.36	6280.05
	Number of enterprises with e-commerce trading activities	2013−2018	180	41.00	12240.00
	R&D expenditures of industrial enterprises above designated size (RMB 10 million)	2012−2018	210	65.03	21072.03
	Mobile Internet access traffic (1000000 GB)	2014−2018	150	106.97	84584.12
	Geologic information data	2001−2011	330	49.00	264247.00
Comprehensive indicator	Digital Financial Inclusion Index	2011−2018	240	18.33	377.74
	digital economy Comprehensive Development Index	2011−2018	240	−1.03	5.80

[1] Due to limited space, the rationality test results of indicators of the first type and the second type are omitted. If you are interested in the results, please contact the authors.

4. Empirical Strategies, Variables and Data

4.1. Research Thoughts and Model Setting

This paper, first of all, uses the “three-step method” to intuitively display the “two-dimensional driving effect” of data factor and its mechanism.

Step 1: Directly study the impacts of data factors on economic growth, the fact that coefficient a_1 is significantly positive is exactly the precondition of proving data factor has direct driving effects. The measurement model is shown below:

$$Gpgdp_{it} = \alpha_0 + \alpha_1 \ln Data_{it} + \alpha_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{14}$$

Wherein, the subscript i denotes province, t denotes year; μ_i and λ_t denote provincial fixed effect and yearly fixed effect, respectively; ε_{it} denotes random disturbance term. $Gpgdp_{it}$ is the growth rate of per capital GDP ($pgdp_{it}$) measured with 1998 as the constant price, used for measuring economic growth of this paper. The computational formula is $Gpgdp_{it} = \ln pgdp_{i,t} - \ln pgdp_{i,t-1}$. $Data_{it}$, represents new data factors and is one of the core explanatory variables of this paper. X_{it} is control variable set, covering human capital stock, population size, level of research and development, degree of marketization, urban and rural structure, capital factor, government scale, and foreign trade.

Step 2: Analyze the effects of data factors on AI technological progress, so as to verify the rationality set in Formula (3) in theoretical analyses in this paper. The measurement model is shown below:

$$\ln AI_{it} = \beta_0 + \beta_1 \ln Data_{it} + \beta_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{15}$$

Wherein, AI_{it} denotes the AI technological progress of province i in year t, and the fact that coefficient β_1 is significantly positive is the precondition of proving that data factors have indirect driving effects.

Step 3: Introduce AI technological progress on the basis of Formula (14), namely:

$$Gpgdp_{it} = \gamma_0 + \gamma_1 \ln Data_{it} + \gamma_2 \ln AI_{it} + \gamma_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{16}$$

At this time, γ_1 is the direct drive effect of data factors on economic growth, while $\beta_1 \times \gamma_2$ reflects to a certain extent the indirect drive effect of data factors on economic growth.

4.2. Measurement of AI Technological Progress

The development of AI technology is impossible without its patent protection, and the development of AI patent is impossible without corporate investment and application and promotion of AI patent. In view of this, this paper uses the interaction term of annual AI patent application quantity and number of registered AI enterprises to measure the progress of AI technologies. First of all, it uses AI, machine learning, machine (computer) vision, computer recognition, virtual reality, neural network, natural language processing, robot and biological recognition as keywords, searches the quantity of relevant patents between 1999 and 2018 in China Patent Full-text Database (CNKI Version),[1] totals and assigns them to sample provinces to obtain the quantity of new patents of AI technologies $patent\_ai_{it}$ newly applied by province i in year t; next, in order to get AI firms in line with the type of patents, this paper, based on the above keywords, uses "Qichacha"[2] to search the number of AI firms $firm\_ai_{it}$ newly registered in province i in t between 1999 and 2018; ultimately, the AI technological progress $AI_{it}$of province i in t is indicated by $AI_{it} = patent\_ai_{it} \times firm\_ai_{it}$.

4.3. Explanation on Control Variables and Data

The core explanatory variables of the model are $Data_{it}$ and AI_{it}, other control variables include: human capital stock *(h)*, which is a provincial human capital stock calculated by reference to Peng (2005); population size *(peo)*, indicated by year-end population of each province; research & development level *(rd)*, indicated by the annual average number of people engaging in high technology industries in the year; degree of marketization *(market*), indicated by marketization index;[3] urban and rural structure (*urb*), indicated by urbanization rate of permanent resident population of the year; capital factor (*k*), indicated by the gross capital formation of the year as a share of gross regional domestic product; government scale (*gov*),

---

[1] China Patent Full-text Database (CNKI Version) consists of three sub-databases of patents for invention, patents for utility models and design patents, respectively, accurately reflecting China's latest patented inventions. Users can retrieve information through retrieval items such as application number, application data, publication number, publication date, patent name, patent abstract, classification number, applicant, inventor and priority, http://cn.oversea.cnki.net/kns55/brief/result.aspx?dbprefix=scpd.

[2] "Qichacha" is an officially certified national credit reference system and its website is https://www.qcc.com/.

[3] Detailed sources of market-based indexes of sample provinces in this paper between 1999 and 2018: 1999–2007 data are from the NERI Index of Marketization of China's Provinces 2009 Report; 2008–2016 data are from the Marketization Index of China's Provinces: NERI Report 2018; 2017–2018 data are calculated by using the method in Chapter 4 of the Marketization Index of China's Provinces: NERI Report 2018.

measured by fiscal expenditure of local governments as a share of gross regional domestic product of the year; foreign trade (*trade*), indicated by regional total import & export volume of the year as a share of gross regional domestic product. All the data of control variables except for the marketization indicators between 1999 and 2016 are from the *China Statistical Yearbook* and statistical yearbooks of local governments.

This paper uses the panel data of China's 30 provinces (except for Hong Kong, Macao, Taiwan and Xizang of China) between 1999 and 2018 as research samples. Table 3 lists the descriptive statistical results of variables.

Table 3. Descriptive Statistical Results of Major Variables

Name of variable	Sample size	Mean value	Standard deviation	Minimum value	Maximum value
Gpgdp	600	0.0948	0.0280	−0.0233	0.2127
ln *Data*	600	2.5202	1.1447	−1.4720	5.3356
ln *AI*	600	6.2496	3.9364	−2.0392	17.8776
ln *h*	600	49.3156	15.3771	14.2526	89.6000
ln *peo*	600	8.1526	0.7589	6.1786	9.3366
ln *rd*	600	0.0059	0.0074	0.0002	0.0363
market	600	6.9390	0.8568	4.4515	9.2900
urb	600	6.0935	1.9606	1.7200	11.7100
k	600	0.5543	0.1675	0.2881	1.4847
gov	600	0.1972	0.0927	0.0630	0.6269
trade	600	0.3046	0.3789	0.0168	1.7223

5. Empirical Analyses

5.1. Preliminary Analysis

As seen from the benchmark regression results in Table 4, columns (1) and (2) are regression results based on Formula (14), showing that data factors, whether the control variables are included or not, have an obvious promoting effect on

China's economic growth, but it is unknown whether the effect comes from the indirect spillover of AI technological progress or directly driven by data factor as a fundamental production factor. Columns (3) and (4) are estimated results based on Formula (15). As seen from the significance of estimated coefficients, data factors can indeed promote AI technological progress, and the setting of this paper about Formula (3) is reasonable. Columns (5) and (6) are corresponding to the regression of Formula (16). Whenever the scale of data factors increases by 1%, it can directly increase China's economic growth rate by 0.0515%; whenever the scale of data factors increases by 1%, it can promote AI technological progress by 0.511% and indirectly increase the economic growth rate of China by 0.02%.

Table 4. The "Two-Dimensional Driving Effect" of Data Factors: Preliminary Estimation

	Gpgdp	*Gpgdp*	ln *AI*	ln *AI*	*Gpgdp*	*Gpgdp*
	(1)	(2)	(3)	(4)	(5)	(6)
ln *Data*	0.1630^{***}	0.0715^{***}	0.5910^{***}	0.5110^{***}	0.1180^{***}	0.0515^{*}
	(0.0259)	(0.0193)	(0.1216)	(0.1245)	(0.0231)	(0.0200)
ln *AI*					0.0759^{***}	0.0392^{***}
					(0.0076)	(0.0076)
ln *h*		0.0624^{**}		0.3690^{**}		0.0479^{*}
		(0.0201)		(0.1249)		(0.0198)
ln *peo*		-0.8240^{***}		-1.3850^{***}		-0.7690^{***}
		(0.0572)		(0.3336)		(0.0564)
ln *rd*		1.7990		45.8300^{***}		0.0020
		(1.0919)		(5.1484)		(1.1213)
market		0.0242^{***}		0.2220^{***}		0.0155^{**}
		(0.0062)		(0.0323)		(0.0060)
urb		0.0043^{***}		0.0320^{***}		0.0031^{***}
		(0.0009)		(0.0056)		(0.0008)
k		0.1720^{***}		0.0243		0.1710^{***}
		(0.0344)		(0.1843)		(0.0342)
gov		-0.6520^{***}		2.5830^{***}		-0.7530^{***}
		(0.1541)		(0.6745)		(0.1608)

	Gpgdp	*Gpgdp*	ln *AI*	ln *AI*	*Gpgdp*	*Gpgdp*
	(1)	(2)	(3)	(4)	(5)	(6)
trade		0.0222 (0.0241)		0.2580 (0.1330)		0.0121 (0.0234)
Direct drive effect			0.0515			
indirect drive effect			0.0200			
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Yearly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observed value	600	600	600	600	600	600
Within R^2	0.1110	0.5503	0.0597	0.3929	0.2438	0.5732

Note: Values in parentheses are the robust standard error. *, ** and *** indicate that they are statistically significant at 10%, 5% and 1%, respectively. The same below.

5.2. Endogenous Processing and Reidentification of "Two-Dimensional driving effect"

The "three-step method" may help intuitively understand the mechanism of the two-dimensional driving effect, but data factors and the proxy indicators of AI technological progress are basically produced internally from social and economic operations, plus such issues of endogeneity, missing variable and product error of Formulas (14)~(16), the conclusion in preliminary regression is not rigorous. Therefore, this paper uses the mechanism identification thought of Beck *et al.* (2010) to construct the following empirical model and re-estimate the two-dimensional driving effect of data factors on economic growth based on the two stage least square (2SLS) framework:

$$Gpgdp_{it} = \xi_0 + \xi_1 \ln Data_{it} + \xi_2 \ln Data_{it} \times \ln AI_{it} + \xi_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{17}$$

Wherein, ξ_1 denotes the direct drive effect of data factors on economic growth, $\xi_2 \times \overline{\ln AI_{it}}$ denotes the indirect drive effect of data factors on economic growth by promoting AI technological progress, and $\overline{\ln AI_{it}}$ is the sample mean value of $\ln AI_{it}$. Although the fixed effect panel model can mitigate the biased error of missing variable, reverse causality still exists. We still need instrumental variables that are related to endogenous variables ($\ln Data_{it}$ and $\ln Data_{it} \times \ln AI_{it}$) and do not directly influence explained variable ($Gpgdp_{it}$). Therefore, this paper first identifies the instrumental variable of data factors ($\ln Data_{it}$) and AI technological progress ($\ln AI_{it}$), namely per capita wage of state-owned enterprises (*wage*) and total human capital of other provinces (*exh*), and then uses the method of

Goldsmith-Pinkham *et al.* (2020), and uses interaction term $wage \times exh$ as the instrumental variable of $\ln Data_{it} \times \ln AI_{it}$.

5.2.1. Instrumental Variable of Data Factors: Per Capita Wage of State-Owned Enterprises ($wage$)

This paper uses per capita wage of state-owned enterprises between 1959 and 1978 as an instrumental variable of data factors[1] due to the reasons below. First, before the reform and opening up, China adopted the strategy of giving priority to developing the heavy industry and state-owned enterprises are an important organizational form for implementing the strategy. Benefiting from the highly centralized planned economy and strong support from the state, state-owned enterprises are not excessively impacted on the whole in the two decades, creating relatively stable internal and external environments that ensure state-owned enterprises can record their own wage data completely and on a long-term basis. When estimating the value of new data factors in the preceding text, the data of labor wages of state-owned enterprises are used, so the wages of state-owned enterprises are closely related to the value of data factor in this paper. Second, wage income can have an effect on economic growth through consumption. The per capita wage of state-owned enterprises between 1959 and 1978 might make an impact on the economic operations of the year and later, but it had little effect on China's economic growth between 1999 and 2018, meeting the exclusive requirements of instrumental variable.

5.2.2. Instrumental Variable of AI Technological Progress: Total Human Capital of Other Provinces (*exh*)

By reference to the thoughts of Yu and Liang (2019), this paper uses total human capital of other provinces of the year as the instrumental variable of AI technological progress of the province. On the one hand, human capital is an important factor facilitating the generation and accumulation of AI technologies. Just like knowledge, technology is also non-competitive. The innovation in AI technologies of a province is supported by local human capital, and is subject to the spillover effects of AI technologies developed by human capital in other provinces. On the other hand, those who can successfully develop AI technologies are generally high-level talents, and it is difficult for them to freely flow between different provinces in the short term, so it is difficult for human capital of other provinces of the year to have significant and direct

[1] The data of per capita wages of state-owned enterprises between 1959 and 1978 are from *China Compendium of Statistics 1949–2009*. Specifically, the overall data of Liaoning Province is missing, and is replaced by the mean value of the data of Jilin Province and Heilongjiang Province, while other small amount of data are supplemented by using the interpolation method.

impacts on the economic growth of the province. Therefore, total human capital of other provinces meets the relevance and exclusive conditions of instrumental variable.

Table 5 lists 2SLS regression results based on Formula (17). As the F test value of the first stage is much greater than 10, there is a reason to believe that there is no weak instrumental variable. The regression results of the second phase show that the two-dimensional driving effect of data factors on China's economic growth is consistent with the preliminary regression results both in direction and significance. Compared with the "three-step method", the direct drive effect and indirect drive effect have increased to 0.6805 and 0.0294, respectively, further proving data factors can promote China's economic growth by facilitating technological progress and have effects on economic growth.

Table 5. The "Two-Dimensional Driving Effect" of Data Factors: Estimation of Instrumental Variable

	2SLS estimation			
	(1)		(2)	
Regression in the first stage	$\ln Data$	$\ln Data \times \ln AI$	$\ln Data$	$\ln Data \times \ln AI$
wage	0.0015^{***}		0.0005^{**}	
	(0.0002)		(0.0002)	
$wage \times exh$		-0.0767^{***}		-0.0629^{***}
		(0.0038)		(0.0042)
F value in the first stage	549.7300		530.4800	
Regression in the second stage	*Gpgdp*		*Gpgdp*	
$\ln Data$	0.5372^{***}		0.6805^{**}	
	(0.0885)		(0.3269)	
$\ln Data \times \ln AI$	0.0044^{***}		0.0047^{*}	
	(0.0015)		(0.0026)	
Direct drive effect	0.5372		0.6805	
Indirect drive effect	0.0275		0.0294	
Control variable	No.		Yes	
Provincial fixed effects	Yes		Yes	
Yearly fixed effects	Yes		Yes	
Within R^2	0.9686		0.9756	

5.3. Heterogeneity and Robustness Tests

5.3.1. Spatial and Temporal Disparities of the Two-Dimensional Driving Effect

From the aspects of sub-region and full sample, this paper studies the spatial-temporal heterogeneity under the two-dimensional driving effects of data factor. The columns (1)~(4) of Table 6 report the test results of spatial-temporal heterogeneity in sub-regions. As can been seen from columns (1) and (2), whenever the scale of data factor in eastern China increases by 1%, it will directly and indirectly increase the economic growth of the region by 0.1815% and 0.0469%, respectively. As seen from columns (3) and (4), the two-dimensional driving effect on data factors in the central and western regions (including northeast China) was not significant before 2003; after 2003, whenever the scale of data factors increased by 1%, it directly and indirectly increased the economic growth rate of the region by 0.0722% and 0.0542%, respectively. In the aspect of indirect drive effect, the economic growth rate of central and western regions (including northeast China) was higher than that of eastern China by 0.0073%, but from the aspect of overall "two-dimensional driving effect", the economic growth rate of eastern China was higher than that of the central and western regions (including northeast China) by 0.102%. Columns (5)~(8) report the temporal heterogeneity test of full samples, the results show that the two-dimensional driving effect of data factors on China's economic growth was significant before and after the first year of the era of big data, since which, both the direct drive effect and indirect drive effect of data factors have significantly increased. Specifically, compared with that before the first year of the era of big data, the direct drive effect has increased by 44.6%.

Table 6. The "Two-Dimensional Driving Effect" of Data Factors: Spatial and Temporal Disparities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Spatial-temporal heterogeneity of sub-regional				Temporal heterogeneity of full sample			
	Eastern China		Central and western regions (including northeast China)		1999−2012		2013−2018	
	1999−2018		2003−2018					
	ln *AI*	*Gpgdp*	ln *AI*	*Gpgdp*	ln *AI*	*Gpgdp*	ln *AI*	*Gpgdp*
ln *Data*	0.6582^{***}	0.1815^{***}	1.6117^{***}	0.0722^{**}	0.2803^{**}	0.0985^{***}	1.1781^{*}	0.1424^{***}
	(0.1795)	(0.0269)	(0.2142)	(0.0357)	(0.1170)	(0.0225)	(0.6535)	(0.0473)
ln *AI*		0.0713^{***}		0.0336^{***}		0.0554^{***}		0.0140^{*}
		(0.0111)		(0.0091)		(0.0089)		(0.0077)
Direct drive effect	0.1815		0.0722		0.0985		0.1424	
Indirect drive effect	0.0469		0.0542		0.0155		0.0165	
Control variable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yearly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed value	200	200	340	340	420	420	180	180
Within R^2	0.4920	0.3645	0.0597	0.3929	0.1638	0.3843	0.0699	0.4621

SBJLD-《中国财政与经济研究》2023年第4期.indd 103 2023/12/14 13:15:14

5.3.2. A New Indicator of Economic Growth: Data of Night Satellite Light

This paper uses night satellite light intensity as a proxy indicator of economic growth (Henderson *et al.*, 2012), uses alternately corrected DMSP/OLS and NPP/VIIRS[1] to conduct a robustness test on the preliminary estimates in Table 4 in different time periods, and adopts the practices of Liao and Wang (2019) for light data correction. Columns (1)~(3) and columns (4)~(6) in Table 7 list the robustness test results of two types of night light data, indicating that the "two-dimensional driving effect" of data factors is still significant.

Table 7. The "Two-Dimensional Driving Effect" of Data Factors: Per Capita Night Light Intensity

	(1)	(2)	(3)	(4)	(5)	(6)
	DMSP/OLS 1999−2013			NPP/VIIRS 2013−2018		
	Glight	ln *AI*	*Glight*	*Glight*	ln *AI*	*Glight*
ln *Data*	0.1611***	0.5322***	0.1382***	0.2942***	1.1794*	0.2522***
	(0.0290)	(0.1223)	(0.0316)	(0.0893)	(0.6585)	(0.0915)
ln *AI*			0.0432***			0.0355**
			(0.0128)			(0.0163)
Control variable	Yes	Yes	Yes	Yes	Yes	Yes
Direct drive effect	0.1382			0.2522		
Indirect drive effect	0.0230			0.0419		
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Yearly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observed value	450	450	450	150	150	150
Within R^2	0.3158	0.2376	0.3366	0.0884	0.0699	0.1159

5.3.3. A New Indicator of AI: Industrial Robot Installation Density

By referring to the practices of Yan *et al.* (2020), this paper constructs the

[1]At present, the most-used night light data mainly include the two types below. First, data captured by former US military weather satellites (DMSP/OLS, time span: 1992–2013). Second, data captured by the US polar orbiting satellite (NPP/VIIRS, time span: 2013–till now). Due to the gradient failure of the USA's military weather satellite sensor, DMSP/OLS has been gradually replaced by NPP/VIIRS since 2013. Source of the image data of the above two types of satellite lights: https://www.ngdc.noaa.gov/ngdc.html.

provincial-level installation rate of industrial robots (*Robort*) as a replacement indicator of AI technologies. The method for estimate the provincial installation rate of industrial robots is shown below:

$$Robort_{it} = \sum_{j=1}^{N} \frac{Rob\_ind_{jt}}{L_{jt}} \times \frac{L_{ijt}}{L_{it}} \tag{17}$$

Wherein, t denotes year; Rob_ind_{jt} denotes the industrial robots stock of industry j, L_{jt} denotes the total number of persons employed in industry j; L_{ijt} denotes the number of persons employed in industry j of province i, L_{it} denotes the total persons employed in province i. The original data source of industrial robot stock is the International Federation of Robotics (IFR), the data of employers by region and industry data come from the *China Labor Statistical Yearbook*. The estimated results in Table 8 show that the conclusion that data factors have two-dimensional driving effect on economic growth is still robust.

Table 8. The "Two-Dimensional Driving Effect" of Data Factors: IFR Industrial Robots

	ln *Robort*	*Gpgdp*
	(1)	(2)
ln *Data*	1.0255***	0.1847***
	(0.2002)	(0.0364)
ln *Robort*		0.0916***
		(0.0116)
Direct drive effect	0.1847	
Indirect drive effect	0.0939	
Control variable	Yes	Yes
Provincial fixed effects	Yes	Yes
Yearly fixed effects	Yes	Yes
Observed value	390	390
Within R^2	0.9664	0.9845

6. Conclusions and Policy Implications

This paper constructs the generation path and value path of data factors, based on which estimates the value of new data factors at the provincial level in China from 1999 to 2018, and by introducing AI technologies, theoretically and empirically discusses about the "two-dimensional driving effect" of data factors on China's economic growth. The main conclusions of this paper are as follow. It is theoretically and practically rational to some extent to estimate the value of new data factors at the provincial level by following the value path of data factors; data factors can directly drive economic growth, namely data factors have economic growth effects and are the essential production factors; data factors can indirectly drive economic growth by promoting AI technological progress and so on.

The above research conclusions have important policy implications. First, only by keeping strengthening the development of China's data factors market, giving full play to the leading role of the data factors market and promoting autonomous, full and orderly flow of data factors in society, can we keep enhancing and give better play to the "two-dimensional driving effect" of data factors. Second, data factors are fundamental resources of the digital economy, and, more importantly, core strategic resources helping China take part in global competitions. At present, in the developed areas with the conditions for vigorously developing the digital economy, we should work hard to remove the institutional barriers to access to public data and fully increase the scale and mobility of data factors in social and economic activities while properly protecting data privacy; in areas with poorly-developed digital economy, it is recommended to first develop digital infrastructures related to data factors and give priority to collection, storage and maintenance of data factors. Finally, we should actively promote the development of local government big data transaction platform. It should be noted that, the trading of data factor through the platform are not limited by space, and local governments should make overall planning on the development of big data transaction platform so as to avoid waste of resources due to repeated construction.

References

Aghion, P., Jones, B, F., & Jones, C. I. (2019). *Artificial Intelligence and Economic Growth*. Chicago, IL: University of Chicago Press.

Beck, T., Levine, R., & Levkov, A. (2010). Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States. *The Journal of Finance*, 65(5), 1637–1667.

Cong, L., W, Xie, D., & Zhang, L. (2021). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science*, 67(10), 6480–6492.

Farboodi, M., & Veldkamp, L. (2021). A Growth Model of the Data Economy. National

Bureau of Economic Research.

Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big Data and Firm Dynamics. *American Economic Association Papers and Proceedings*, 109, 38–42.

Goldsmith-Pinkham, P., Sorkin., I, & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8), 2586–2624.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2), 994–1028.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9), 2819–2858.

Liao, Z., & W, L. (2019). Urban Sprawl and High-Quality Economic Development, The Research of Nighttime Satellite Light Data in China. *Finance & Economics* (*Caijing Kexue*), 6, 106–119.

Lu, C. (2021). The Impact of Artificial Intelligence on Economic Growth and Welfare. *Journal of Macroeconomics*, 69, 103342.

Lucas, Jr. R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1), 3–42.

Peng, G. (2005). China's Regional Income Gap, Total Factor Productivity and Its Convergence Analyses. *Economic Research Journal* (*Jingji Yanjiu*), 9, 19–29.

Rao, P., Yue, H., & Jiang, G. (2017). Research on Economic Policy Uncertainties and Enterprise Investment Behaviors. *The Journal of World Economy* (*Shijie Jingji*), 40(2), 27–51.

Statistics Canada. (2019a). Canada Data Framework Investment in Data, Databases and Data Science: Conceptual Framework . https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm.

Statistics Canada. (2019b). The Value of Data in Canada: Experimental Estimates. https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm.

Varian, H. (2019). *Artificial Intelligence, Economics, and Industrial Organization*. Chicago, IL: University of Chicago Press.

Veldkamp, L., & Chung, C. (2022). Data and the Aggregate Economy. *Journal of Economic Literature*, forthcoming.

Xu, X., & Zhao, M. (2020). Data Capital and Economic Growth Path. *Economic Research Journal* (*Jingji Yanjiu)*, 10, 38–54.

Yan, X., Zhu, B., & Ma, C. (2020). The Use of Industrial Robots and Employment in the Manufacturing Industry. Evidences from China. *Statistical Research* (*Tongji Yanjiu*), 37(1), 74–87.

Yu., W., & Liang., P. (2019). Uncertainty, Business Environment and Vitality of Private Enterprises. *China Industrial Economics (Zhongguo Gongye Jingji)*, 11, 136–154.

Zhao., T., Zhang, Z., & Liang, S. (2020). Digital Economy, Entrepreneurial Activity and High-Quality Development: Empirical Evidence from Chinese Cities. *Management World* (*Guanli Shijie*), 36(10), 65–76.