Transactions on Social Science, Education and Humanities Research
ISSN: 2960-1770, eISSN: 2960-2262 | Volume 17
CEPC 2026

The Impact of National Big Data Comprehensive Pilot Zones on Provincial Educational Disparities

Xinting Yuan *, Shiya Yin
Faculty of Education, Shandong Normal University, Jinan 250014, China
* Corresponding Author Email: yuanxinting2026@163.com

Abstract. Against the backdrop of the digital economy profoundly reshaping development patterns, the social welfare effects of big data strategies have attracted increasing attention. However, the dividends and divides generated by digital technology empowerment coexist, and its net effect on social equity, especially regional educational equity, remains unclear. Taking the staggered establishment of the National Big Data Comprehensive Pilot Zones as a quasi-natural experiment, this paper constructs a multi-period difference-in-differences model to systematically evaluate the causal effect of this policy on provincial educational disparities. The study finds that the implementation of the pilot-zone policy significantly narrows provincial educational disparities, and this conclusion remains valid after a series of robustness tests, including the parallel-trend test, placebo test, replacement of the dependent variable, and PSM-DID. Further analysis reveals regional heterogeneity in the policy effect: it is significant in eastern, central, and northeastern China, but insignificant in western China. Mechanism analysis indicates that the policy mainly optimizes the allocation efficiency and accessibility of educational resources through three transmission channels: strengthening fiscal support, promoting the upgrading of industrial structure, and improving the level of digital technology, thereby restraining interregional educational inequality. This study confirms that big-data development policies have the dual functions of promoting growth and improving equity, and provides empirical evidence for simultaneously advancing economic development and social equity in the process of digital transformation and for formulating more inclusive regional development policies.

Keywords: National Big Data Comprehensive Pilot Zones; provincial educational disparity; fiscal allocation effect; human-capital demand effect; technology empowerment effect.

1. Introduction

At present, the world is experiencing a new round of data-driven technological revolution and industrial transformation. New-generation information technologies represented by big data have become a key force in promoting economic and social development and reshaping national competitive advantages. High-quality development depends on the effective release of the potential of data factors and the construction of a healthy digital-economy ecosystem. As Xi Jinping (2022) [1] emphasized, it is necessary to "promote the implementation of the national big data strategy, accelerate the improvement of digital infrastructure, advance the integration and open sharing of data resources, and speed up the building of Digital China." Among these efforts, the National Big Data Comprehensive Pilot Zones (hereinafter referred to as the "pilot zones"), as a strategic forerunner at the national level for systematically advancing big-data development, undertake key pilot tasks such as exploring market-oriented allocation of data factors, promoting the deep integration of big data with the real economy, and innovating social governance models. However, the empowerment effects of digital technology are not released evenly, and the digital divide may aggravate imbalances in regional development. In China, the long-standing imbalance in educational resource allocation and development levels among provinces is a major constraint on educational modernization and the building of a strong education system. Therefore, to scientifically guide the implementation of the big-data development strategy and help address the imbalance in regional educational development, it is of practical urgency to systematically evaluate the net effect of the pilot-zone policy on provincial educational disparities.

A large body of literature has examined digital development and its economic and social impacts from multiple perspectives. Existing studies have affirmed the positive effects of policies such as the National Big Data Comprehensive Pilot Zones. For instance, Wu Yongchao and Li Jianjiang (2026) [2], using a difference-in-differences model, demonstrate that the establishment of big-data comprehensive pilot zones significantly promotes educational equity in the areas where they are established, raising the level of educational equity by 6.07%. With respect to the determinants of intra-provincial educational disparities, Luo Weiqing (2010a) [3] argues that China's excessively decentralized system of "centralized fiscal power and decentralized administrative responsibilities" gives local governments incentives to reduce education spending with positive externalities, thereby aggravating inequality in the supply of public education within provinces. In addition, Farley & Burbules (2022) [4] note that although technology has the potential to expand educational opportunities, it may also create new inequalities, and digitalization in education does not automatically lead to equity. Ventrella & Cotnam-Kappel (2024) [5], in their study of Canadian elementary schools, also find that students' digital capital plays a crucial role in shaping their digital experiences and offline academic outcomes, and that digital inequalities are intensifying. Nevertheless, literature focusing on the direct causal relationship between big-data development policies and intra-provincial educational disparities remains limited. Existing studies either emphasize theoretical discussion and feasibility analysis or are constrained by endogeneity problems, making it difficult to provide robust empirical evidence. On this basis, this paper treats the establishment of the National Big Data Comprehensive Pilot Zones as a quasi-natural experiment and systematically examines their causal impact, mechanisms, and heterogeneous effects on provincial educational disparities.

Theoretically, the establishment of National Big Data Comprehensive Pilot Zones may help narrow provincial educational disparities. First, one important direction of pilot-zone construction is to promote the development and application of educational big data, with the aim of building unified and open educational-resource platforms based on big-data resources and technologies. This can substantially reduce the exclusivity of high-quality educational resources, enabling less developed areas within provinces to share high-quality resources at lower cost and thereby narrow interregional opportunity gaps (Chen Shuangye et al., 2014) [6]. Second, big-data-driven educational management systems can provide precise decision support, assist educational administrators in conducting more scientific resource allocation and evaluation, and induce deep changes in the education-governance system, thereby optimizing the spatial distribution of educational resources within provinces and alleviating uneven allocation (Wang et al., 2026) [7]. Therefore, from the perspectives of resource inclusiveness and governance optimization, the pilot-zone policy may become an effective tool for bridging intra-provincial educational disparities.

However, the establishment of National Big Data Comprehensive Pilot Zones may also objectively widen intra-provincial educational disparities and generate a digital-divide effect. On the one hand, the application and development of big data require a high level of infrastructure, technical talent, and fiscal input. In the early stage of policy implementation, resources and policy dividends in pilot provinces may be absorbed first by provincial capitals and economically developed regions, leading to a Matthew effect in which the strong become stronger. This is similar to the logic observed by Luo Weiqing (2010b) [3], whereby excessive decentralization causes resources to tilt toward high-growth sectors, thereby possibly intensifying the gap in technology empowerment between core and peripheral regions within provinces. On the other hand, as Ventrella & Cotnam-Kappel (2024) [5] reveal, students' household digital capital has a significant effect on academic outcomes, and differences in such digital capital are closely related to family socioeconomic status, potentially resulting in the reproduction of educational inequality. Farley & Burbules (2022) [4] further point out that in online learning environments, students from disadvantaged backgrounds face structural barriers (such as insufficient devices and environmental interference) and instructional barriers (such as insufficient cultural inclusiveness), so the diffusion of technology may intensify rather than alleviate existing educational injustice. Therefore, from the perspectives of resource siphoning and regional competition, the effect of the pilot-zone policy is also uncertain.

In sum, the theoretical direction of the impact of National Big Data Comprehensive Pilot Zones on intra-provincial educational disparities remains unclear and requires rigorous empirical testing. Accordingly, this paper regards the staggered establishment of pilot zones since 2016 as a quasi-natural experiment and constructs a multi-period difference-in-differences model based on Chinese prefecture-level city panel data from 2003 to 2020. The study finds that the establishment of National Big Data Comprehensive Pilot Zones significantly restrains the expansion of intra-provincial educational disparities, and this result remains valid after robustness tests including the parallel-trend test, placebo test, replacement of the dependent variable, and instrumental-variable method. Heterogeneity analysis shows that the policy effect is more significant in central and western regions, in provinces with lower fiscal self-sufficiency, and in provinces with weaker initial human-capital levels. Mechanism tests indicate that optimizing the allocation of fiscal education resources (mitigating the drawbacks of fiscal decentralization) and improving the sharing level of digital education resources (technology-empowered services) are the two main channels. These results are consistent with the theoretical expectation that the pilot-zone policy bridges intra-provincial educational disparities by enhancing provincial coordination capacity and technology spillover effects.

The possible contributions of this paper are as follows. First, it extends research on the economic consequences of National Big Data Comprehensive Pilot Zones. Existing studies mainly focus on the economic performance of pilot zones (Wu Yongchao and Li Jianjiang, 2026) [2], while paying less attention to their spillover effects on the equalization of public services. Based on this, this paper explores the role of big-data development policies in promoting social welfare and improving social equity. The results show that the digital economy not only effectively promotes economic growth but also displays positive effects in optimizing income distribution and narrowing development gaps. Second, this paper expands the research field on the determinants of educational disparities. Previous studies have explained educational disparities mostly from single perspectives such as fiscal decentralization (Luo Weiqing, 2010) [3] or the micro-level digital divide (Ventrella & Cotnam-Kappel, 2024) [5]. In contrast, this study innovatively incorporates the fiscal resource allocation effect and technology-empowered service effect into a unified analytical framework and systematically examines the comprehensive impact of the National Big Data Comprehensive Pilot Zone policy on provincial educational disparities. This integrated perspective not only helps identify policy channels more comprehensively but also deepens understanding of the fiscal-technical synergy underlying the formation and evolution of educational disparities. Third, the paper has clear practical significance. Its findings indicate that implementing the big-data strategy is not only a lever for promoting high-quality economic development, but also an important instrument for addressing imbalances in regional educational development. This provides direct empirical evidence and policy reference for governments seeking to optimize educational resource allocation and promote common prosperity through cross-regional data sharing and collaborative governance in the context of digital transformation.

The remainder of this paper is organized as follows. Section II reviews the literature; Section III presents the research hypotheses and empirical design; Section IV reports and analyzes the empirical results; Section V provides further mechanism tests; and Section VI concludes.

2. Literature Review

2.1. Research on National Big Data Comprehensive Pilot Zones.

The National Big Data Comprehensive Pilot Zone is a regional digitalization pilot policy launched by the State Council in 2015. This policy is an important institutional vehicle for implementing China's big-data strategy. In August 2015, the State Council issued the Outline of Action for

Promoting Big Data Development (Guo Fa [2015] No. 50) [8], which explicitly proposed "carrying out regional pilots and promoting the construction of big-data comprehensive pilot zones in Guizhou and other places," marking the official launch of the National Big Data Comprehensive Pilot Zone policy. In September of the same year, Guizhou Province was the first to be approved to build the country's first big-data comprehensive pilot zone. In October 2016, the National Development and Reform Commission, the Ministry of Industry and Information Technology, and the Cyberspace Administration of China jointly approved the second batch of pilot zones, covering seven regions: Beijing-Tianjin-Hebei, the Pearl River Delta, Shanghai, Henan Province, Chongqing, Shenyang, and Inner Mongolia Autonomous Region. Since then, eight national-level big-data comprehensive pilot zones have been established nationwide, covering more than ten provinces in eastern, central, western, and northeastern China and forming a policy layout characterized by staged approval, regional differentiation, and classified implementation. The policy carries out systematic exploration in big-data institutional innovation, open sharing of public data, innovative big-data applications, and industrial agglomeration, with the goal of driving high-quality economic and social development through data factors. In the public-service dimension, Lan Xiangjie and Yang Lei (2026) [9] find that the policy significantly improves the level of regional public-service supply. At the micro-household level, Wang Qiankun (2024) [10] confirms the policy's income-increasing effect, which is more pronounced for low-income households. Wu Yongchao and Li Jianjiang (2026) [2], using a difference-in-differences model, confirm that the establishment of big-data comprehensive pilot zones generally promotes educational equity in the locations where they are established. Liu (2025) [11], using Chinese household survey data, finds that the digital economy significantly alleviates educational income inequality; Wang et al. (2026) [12], in an empirical study of high-school graduates in western China, find that digital literacy has a significantly positive impact on access to higher education through the mediating role of online social capital. In addition, Xu et al. (2025) [13] show that the digital economy empowers the high-quality development of higher education through three channels: increasing educational funding, driving educational technology innovation, and accelerating educational digital transformation. These studies provide multi-dimensional and multi-level empirical evidence for understanding how big-data policies affect educational disparities.

2.2. Manifestations, Causes, and Measurement of Educational Inequality.

Educational inequality refers to systematic differences among social groups in access to educational opportunities, participation in educational processes, and attainment of educational outcomes (Wen Xiaoguo, 2016) [14]. Its formation is rooted in the interaction of multiple structural factors. In terms of fiscal factors, Luo Weiqing (2010) [3] argues that China's excessively decentralized system of "centralized fiscal power and decentralized administrative responsibilities" gives local governments incentives to reduce education spending with positive externalities, thereby aggravating inequality in the supply of public education within provinces. In terms of technology diffusion, Vassilakopoulou & Hustad (2021) [15] indicate that the digital divide partly mirrors existing inequalities in socioeconomic resources, and education level is a key factor affecting individuals' Internet access and usage capacity. Farley & Burbules (2022) [4] further point out that although technology has the potential to expand educational opportunities, digitalization does not automatically lead to equity. Ventrella & Cotnam-Kappel (2024) [5], in an empirical study of Canadian elementary schools, find that students' digital capital plays a key role in shaping their digital experiences and offline academic outcomes. Chen Shuangye et al. (2014) [6] note that big-data-driven educational management systems can provide precise decision support, offering technological possibilities for bridging interregional gaps in educational quality. In terms of measurement tools, the education Gini coefficient has become a core indicator for comparing the equality of educational distribution across regions and over time (Wen Xiaoguo, 2016) [14].

2.3. Reasoning on the Impact of Big Data Comprehensive Pilot Zones on Educational Inequality.

Although empirical studies directly testing the causal relationship between pilot zones and educational inequality remain scarce, existing literature provides indirect inferential clues through verified channels of influence. In the resource-allocation channel, Lan Xiangjie and Yang Lei (2026) [9] confirm that pilot zones can expand the coverage of public services, while Wang Qiankun (2024) [10] confirms the income-increasing effect of digital inclusive finance on low-income groups. Together, these findings point to a possibility: by lowering the cost of information acquisition and breaking geographical barriers, digital policies may help bridge the educational-resource gap long maintained by the urban-rural dual structure. In the dimension of regional disparities, Zhang et al. (2025) [16] find that the digital economy significantly narrows the urban-rural income gap, and that this effect is most pronounced in western China; Nie & Wan (2023) [17], based on the quasi-natural experiment of "Broadband China," find significant regional heterogeneity in the impact of Internet infrastructure construction on the urban-rural income gap. However, from the perspective of social stratification, studies by Vassilakopoulou & Hustad (2021) [15], Farley & Burbules (2022) [4], and Ventrella & Cotnam-Kappel (2024) [5] reveal that the diffusion of digital technology may also reinforce inequality through its coupling with existing social stratification: groups with more economic, cultural, and social capital are able to use digital resources more effectively and thus occupy an advantageous position in the digitalization process. This digital-divide dilemma constitutes a theoretical proposition that urgently requires empirical testing.

2.4. Literature Review Summary.

In summary, existing studies have generated a substantial body of research along two main lines: pilot zones and educational inequality. However, these two strands of literature remain relatively independent. Evaluations of pilot-zone policy effects have not yet extended to the dimension of educational equity measured by the education Gini coefficient, and studies of educational inequality have not incorporated national-level digital strategy policies into their framework of exogenous shocks. This disconnection constitutes the core academic gap that this study seeks to fill.

3. Research Hypotheses and Empirical Design

3.1. Research Hypotheses.

The theoretical logic of this study is that the National Big Data Comprehensive Pilot Zone policy may influence provincial educational disparities through multiple channels. Whether its ultimate impact is positive, negative, or uncertain is the core question for empirical testing. Based on the theoretical analysis, this paper proposes the following hypotheses.

An important direction of National Big Data Comprehensive Pilot Zone construction is to promote the development and application of educational big data, with the aim of building unified and open educational-resource platforms based on big-data resources and technologies. Wu Yongchao and Li Jianjiang (2026) [2] find that this process can greatly reduce the exclusivity and rivalry of high-quality educational resources, allowing learners in less developed regions to enjoy high-quality educational resources equally at lower cost, thereby expanding the coverage of educational resources and creating conditions for educational opportunity equity. Meanwhile, Chen Shuangye et al. (2014) [6] argue that big-data-driven educational management systems are highly intelligent and precise, can assist educational administrators in conducting more scientific resource allocation, educational diagnosis, and evaluation, and can induce deep changes in the education-governance system, thereby making it possible to bridge interregional gaps in educational quality and improve the overall level of regional education. Wang et al. (2026) [7], in an empirical study of high-school graduates in western China, further find that digital literacy has a significantly positive impact on access to higher education through online social capital, and that this effect varies by family socioeconomic background, providing micro-level evidence for digital technology's empowerment of educational equity.

However, the development and application of big data also involve risks of a digital divide and may instead intensify inequality. On the one hand, big data has high requirements for infrastructure, technical talent, and fiscal investment. In the early stage of policy implementation, resources and policy dividends in pilot provinces may be absorbed first by provincial capitals and economically developed regions, producing a Matthew effect in which the strong become stronger and thereby widening the technology-empowerment gap between core and peripheral regions within provinces. On the other hand, students' household digital capital has a significant effect on academic outcomes, and differences in family background may transmit educational inequality across generations through the use of digital technology (Ventrella & Cotnam-Kappel, 2024) [5]. If the policy fails to effectively cover households with low socioeconomic status, the high-skilled labor demand generated by big-data development may attract talent to central cities, solidifying or even widening educational-output gaps based on human-capital differences.

In sum, this paper tends to regard the net effect as positive. Therefore, the core research hypothesis is proposed as follows:

H1: The establishment of National Big Data Comprehensive Pilot Zones can significantly narrow provincial educational disparities.

Pilot-zone construction can promote local economic growth and increase fiscal revenue, thereby strengthening provincial fiscal capacity. According to fiscal decentralization theory, increased fiscal capacity is the basis for governments to expand public-service supply. Provincial governments may therefore increase overall investment in education and strengthen special education transfer payments to weak regions within provinces, thus restraining interregional gaps in educational investment. Liu & Li (2024) [18] provide indirect support for this logic. Their analysis of China's decentralization reform shows that moderate decentralization can improve resource allocation efficiency and reduce regional disparities in access to higher education caused by the uneven distribution of key universities. Accordingly, this paper proposes:

H1a: The National Big Data Comprehensive Pilot Zone policy can narrow provincial educational disparities by optimizing the intra-provincial allocation of fiscal education resources.

Big-data development creates demand for high-skilled labor and generates "skill-biased technological progress." This increases the expected returns to high-skill education and significantly changes household decisions on educational investment, making households more inclined to invest in high-quality and STEM education. Structural changes on the demand side of households are transmitted to the supply side through labor-market signals, forcing local governments and schools, in response to industrial upgrading and social demand, to adjust the educational supply structure and increase resource allocation in STEM and related fields. Zhang et al. (2025) [16] show that the digital economy significantly narrows the urban-rural income gap by optimizing human-capital allocation, with the effect being most pronounced in western China, thereby providing important evidence for the human-capital pathway. Accordingly, this paper proposes:

H1b: The National Big Data Comprehensive Pilot Zone policy can affect provincial educational disparities by inducing industrial upgrading and demand for skilled talent, thereby forcing the intra-provincial educational supply structure to adjust toward STEM fields.

The upgrading of digital infrastructure driven by pilot-zone construction provides key support for building provincial smart-education platforms and promoting resource-sharing models such as the "Three Classrooms." This process helps overcome the spatial limitations in the distribution of high-quality educational resources and effectively improves access to high-quality educational content in areas with weak educational resources. Meanwhile, the application of big-data technology can promote refined and intelligent educational governance, realizing precise matching and dynamic adjustment between the supply and demand of educational resources through data analysis. This governance model of targeted policy implementation helps allocate educational resources more effectively to the regions and groups most in need, thereby more substantively promoting equity in educational outcomes while improving resource-use efficiency. Xu et al. (2025) [13] find that the digital economy promotes high-quality development of higher education precisely by accelerating educational digital transformation, such as the construction and application of network multimedia classrooms, and by promoting educational technology innovation. Accordingly, this paper proposes:

H1c: The National Big Data Comprehensive Pilot Zone policy can narrow provincial educational disparities by improving the sharing level of digital educational resources and the government's precise-governance capacity.

3.2. Empirical Design.

To test research hypothesis H1 and following the econometric method used by Wu Yongchao and Li Jianjiang (2026) [2] in evaluating the policy effect of National Big Data Comprehensive Pilot Zones, this paper regards the establishment of pilot zones as a quasi-natural experiment and constructs a multi-period difference-in-differences model for empirical analysis. The regression model is specified as follows:

$$Gap_{it} = \alpha + \beta(Treat_i \times Post_{it}) + \gamma X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{1}$$

In this model, i and t denote province and year, respectively. The dependent variable Gap_{it} measures the level of educational disparity in each province in a given year; the core explanatory variable $Treat_i \times Post_{it}$ reflects the policy effect of pilot-zone establishment; γX_{it} denotes a set of control variables that vary across both time and province; μ_{i} and λ_t control for province fixed effects and year fixed effects, respectively; and ε_{it} is the random error term. In the regressions, standard errors are clustered at the provincial level to address possible serial correlation.

3.2.1. Dependent variable:

Gap_{it} represents the provincial educational disparity. Existing literature measures educational equity mainly using comprehensive indicators such as the Gini coefficient (EGINI), Theil index, and coefficient of variation. For example, Wu Yongchao and Li Jianjiang (2026) [2] use the reciprocal of the education Gini coefficient as a proxy for educational equity. However, the Gini coefficient measures overall inequality and cannot precisely decompose the sources of inequality. Because this study focuses on between-group inequality among different prefecture-level cities within provinces, it draws on methods in inequality measurement and mainly adopts the between-group component of the decomposable education Theil index as the core dependent variable. This indicator directly and precisely measures interregional educational disparities within provinces.

3.2.2. Core explanatory variable:

$Treat_i \times Post_{it}$ is the policy treatment variable. $Treat_i$ is the treatment-group dummy variable; it equals 1 if province i was approved as a National Big Data Comprehensive Pilot Zone during the sample period, and 0 otherwise. $Post_{it}$ is the post-policy time dummy variable; for treatment-group provinces, it equals 1 from the year of approval (inclusive) onward and 0 before that, while for control-group provinces it is always 0. The coefficient β is the focus of this paper. If its estimate is significantly negative, it indicates that the National Big Data Comprehensive Pilot Zone policy significantly narrows provincial educational disparities, thus supporting H1.

3.2.3. Control variables:

γX_{it} is a set of time-varying provincial control variables used to control for other factors that may influence educational disparities. This paper controls for fiscal support (fsupp), economic development level (econ), urbanization rate (urban), industrial structure upgrading (isadv), financial development level measured by deposit and loan balances (fin), urban-rural residents' income gap (gap), human-capital level (hum), R&D intensity (rd), and digital technology level (digtech). In all regressions, standard errors are clustered at the provincial level.

Table 1. Definition and Construction of Variables.

Variable Type	Variable Symbol	Variable Construction
Dependent variable	Gap_{it}	Provincial educational disparity, measured by the between-group component of the education Theil index.
Core explanatory variable	$Treat_i \times Post_{it}$	Pilot-zone policy effect, the DID interaction term. Treat_i equals 1 if province i established a big-data comprehensive pilot zone and 0 otherwise; Post_it equals 1 if year t is after the establishment of the pilot zone (including the year of establishment), and 0 otherwise.
Control variable	fsupp	Fiscal support, measured by the share of local fiscal education expenditure in general public budget expenditure; it reflects the government's willingness and capacity to invest resources in education.
	econ	Economic development level, measured by per capita gross regional product (logged).
	urban	Urbanization rate, measured by the share of the urban resident population in the total year-end population.
	isadv	Industrial structure upgrading, measured by the ratio of value added in the tertiary industry to value added in the secondary industry; it reflects the shift toward a more service-oriented and high-end industrial structure.
	fin	Financial development level, measured by the sum of financial institutions' deposit and loan balances as a share of gross regional product.
	gap	Urban-rural residents' income gap, measured by the ratio of per capita disposable income of urban residents to that of rural residents.
	hum	Human-capital level, measured by the number of students enrolled in regular institutions of higher education per 100,000 people.
	rd	R&D intensity, measured by research and experimental development expenditure as a share of gross regional product.
	digtech	Digital technology level, measured by the number of Internet broadband access users per 100 persons (or a proxy indicator related to digital industries), reflecting the level of regional digital infrastructure and application.

3.3. Sample Sources.

This study selects panel data for 30 provincial-level regions from 2010 to 2024 as the research sample. Information on the batches and timing of the establishment of National Big Data Comprehensive Pilot Zones is compiled from public documents issued by the National Development and Reform Commission and other departments. The prefecture-level city data on average years of education per capita and population needed to calculate the dependent variable come from the China City Statistical Yearbook and provincial and municipal statistical yearbooks. Control variables mainly come from the China Statistical Yearbook, China Fiscal Yearbook, and the National Bureau of Statistics database. To eliminate the influence of extreme values, all continuous variables are winsorized at the upper and lower 1% levels.

3.4. Descriptive Statistics

Table 2. Descriptive Statistics.

VARIABLES	N	mean	sd	min	max
Gap_{it}	450	0.224	0.053	0.154	0.530
$Treat_i \times Post_{it}$	450	0.200	0.400	0.000	1.000
fsupp	450	0.289	0.198	0.105	1.326
econ	450	10.910	0.511	9.486	12.340
urban	450	0.593	0.132	0.222	0.898
isadv	450	2.552	1.140	0.173	17.48
fin	450	3.483	1.089	1.695	7.545
gap	450	2.528	0.407	1.718	3.735
hum	450	0.021	0.006	0.008	0.044
rd	450	0.018	0.012	0.002	0.080
digtech	450	0.001	0.001	0.000	0.006

Table 2 reports the descriptive statistics of the main variables. The meaning of the core dependent variable Gap_{it} is 0.224, with a standard deviation of 0.053 and minimum and maximum values of 0.154 and 0.530, respectively. This indicates significant differences in interregional educational disparities within Chinese provinces, providing necessary variation for this study. The mean of the core explanatory variable $Treat_i \times Post_{it}$ is 0.200, indicating that approximately 20.0% of the observations in the sample period were affected by the National Big Data Comprehensive Pilot Zone policy. The statistical characteristics of the control variables are basically consistent with existing studies, and the data distribution is reasonable.

4. Empirical Results and Analysis

4.1. Analysis of Baseline Regression Results.

Table 3. Baseline Regression Results.

VARIABLES	(1) Gap_{it}	(2) Gap_{it}	(3) Gap_{it}	(4) Gap_{it}
$Treat_i \times Post_{it}$	-0.011***	-0.011***	-0.011***	-0.012***
	(-3.293)	(-4.045)	(-4.264)	(-4.065)
econ		-0.030***	-0.029**	-0.027**
		(-3.001)	(-2.549)	(-2.664)
urban		0.040	0.018	-0.013
		(1.372)	(0.715)	(-0.373)
fin			0.004**	0.004**
			(2.645)	(2.370)
gap			-0.016**	-0.016**
			(-2.548)	(-2.740)
hum				0.953*
				(1.756)
rd				-0.012
				(-0.126)
Constant	0.165***	0.474***	0.492***	0.469***
	(80.746)	(4.409)	(3.849)	(4.071)
Observations	450	450	450	450
R-squared	0.978	0.979	0.979	0.980
Province FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Note: Values in parentheses are t-values adjusted for clustering; *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The same applies below.

This section gradually introduces control variables to test the causal effect of the National Big Data Comprehensive Pilot Zone policy $Treat_i \times Post_{it}$ on provincial educational disparities Gap_{it}. Table 3 reports the estimation results based on a two-way fixed-effects model (province fixed effects and year fixed effects). Column (1) is the baseline model, which controls only for province and year fixed effects in order to capture the net policy effect. The result shows that the coefficient of $Treat_i \times Post_{it}$ is -0.011 and significantly negative at the 1% level, preliminarily indicating that policy implementation helps narrow educational disparities. Columns (2) through (4) are extended models that gradually add control variables to exclude interference from other factors affecting educational disparities and thereby identify the policy effect more accurately. As control variables are added, the coefficient of $Treat_i \times Post_{it}$ remains stable between -0.011 and -0.012 and is consistently significantly negative at the 1% statistical level, indicating that the policy effect remains robust after controlling other influencing factors.

The regression results show that the coefficient of $Treat_i \times Post_{it}$ is significantly negative in all four model specifications (-0.011 to -0.012). Because the dependent variable is the education Gini coefficient, a negative coefficient means that the implementation of the National Big Data Comprehensive Pilot Zone policy significantly reduces educational inequality across provinces, that is, it narrows intra-provincial educational disparities. This finding strongly supports the research hypothesis of this paper. In addition, among the control variables,econ and gap are significantly negative, indicating that regions with higher levels of economic development and smaller internal economic differences have lower education Gini coefficients, which is consistent with theoretical expectations.

4.2. Robustness Tests.

To ensure the robustness of the baseline regression results, this paper adopts the following three testing strategies:

4.2.1. Parallel-Trend Test.

Using the event-study method, this study takes the establishment of the National Big Data Comprehensive Pilot Zones in 2016 as the event starting year and examines changes in the estimated coefficients before policy implementation. To avoid multicollinearity, the year immediately before policy implementation is excluded from the regression estimation. The test results show that in the years before the implementation of the National Big Data Comprehensive Pilot Zone policy, the DID regression coefficients are not significant (all include 0). This indicates that the treatment group and the control group have consistent trends without systematic differences; the assumption is satisfied and the estimation model is valid.

To verify the core premise of the difference-in-differences model—the parallel-trend assumption—that there were no systematic differences in the trends of educational inequality Gap_{it} between the treatment group and the control group before the policy shock, this study conducts an event-study test. Taking the initial year of the National Big Data Comprehensive Pilot Zone policy (the benchmark year) as the reference, we construct a series of dummy variables from five years before to four years after the policy occurred (before5 to after4) and include them in the regression model to observe the dynamic effects.

The results in Table 4 show that in the years before policy implementation (before5 to before2), except for the coefficient of before5, which is significantly positive at the 5% level (0.009**), the estimated coefficients of before4, before3, and before2 are not significantly different from 0. This indicates that, for most periods before the policy occurred, the trends in the education Gini coefficient EGINI between the treatment and control groups did not exhibit significant differences, basically satisfying the parallel-trend assumption. Although the significant coefficient of before5 suggests possible early trend fluctuations in a more distant period, the insignificant coefficient in the key year close to the policy period (before2) provides major support for the applicability of the difference-in-differences model.

In the year of policy implementation current and thereafter, the coefficients of after1, after2, and after3 are all significantly negative (-0.009**, -0.010***, and -0.008**), indicating that the policy produced a significant effect of narrowing educational disparities for several periods after implementation. Subsequently, the coefficient of after4 is no longer significant, which may imply that the policy effect weakens over time.

Table 4. Parallel-Trend Test.

VARIABLES	EGINI
before5	0.009**
	(2.624)
before4	0.005
	(1.282)
before3	0.005
	(1.237)
before2	0.005
	(1.273)
current	-0.005
	(-1.574)
after1	-0.009**
	(-2.564)
after2	-0.010***
	(-3.433)
after3	-0.008**
	(-2.462)
after4	-0.005
	(-1.480)
Controls	YES
Observations	450
R-squared	0.981

4.2.2. Placebo Test.

To exclude interference from other potential unobserved variables in the estimation results, this study conducts a placebo test by randomly constructing pseudo-treatment groups. Specifically, provinces in the sample are randomly assigned to a virtual treatment group, and this process is repeated 1,000 times. Each time, the same regression estimation as in the baseline model is performed, yielding 1,000 "pseudo policy effect" coefficients.

As shown in Figure 1, the vast majority of pseudo coefficients are densely distributed around 0, and their distribution differs markedly from the true estimate obtained in the baseline regression. This indicates that the policy effect captured by the baseline model is unlikely to be caused by unobserved factors or randomness, thereby confirming the robustness of the research conclusion.

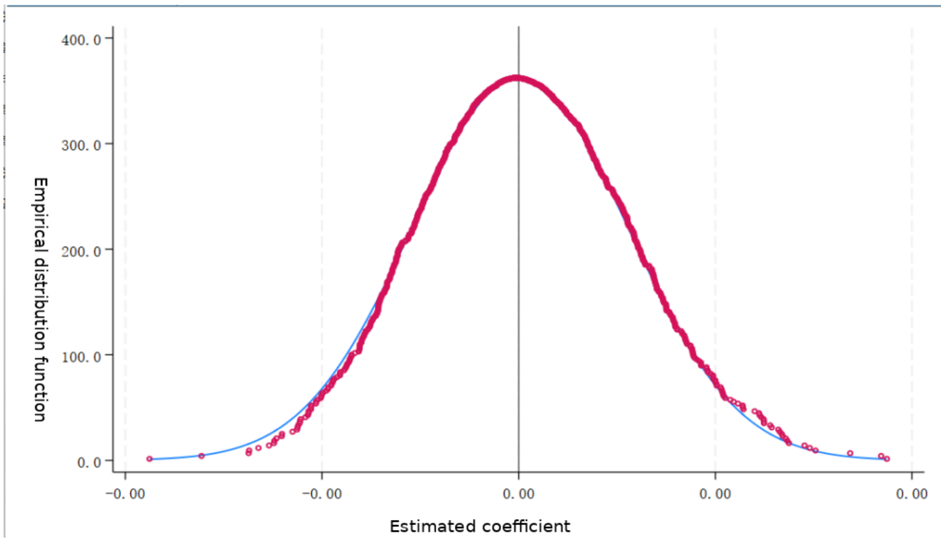

Fig 1. Placebo Test.

4.2.3. PSM-DID.

To alleviate systematic differences that may exist between the treatment and control groups before policy implementation (i.e., sample selection bias), this study uses the propensity-score-matching difference-in-differences method to conduct a robustness test of the baseline regression conclusion. Table 5 reports the PSM-DID estimation results.

Table 5. PSM-DID Estimation Results.

VARIABLES	Gap_{it}
$Treat_i \times Post_{it}$	-0.009*
Observations	105
R-squared	0.978
Controls	YES
Province FE	YES
Year FE	YES

The results show that the coefficient of the core explanatory variable is -0.009 and significant at the 10% level. This estimate is highly consistent with the baseline regression in terms of direction, magnitude (approximately -0.01), and significance.

After using the PSM method to construct a control group with characteristics more similar to those of the treatment-group provinces, thereby effectively alleviating sample selection bias, the negative effect of the National Big Data Comprehensive Pilot Zone policy ($Treat_i \times Post_{it}$) on narrowing provincial educational disparities (Gap_{it}) remains statistically significant. This strongly confirms that the policy's effect in promoting educational equity captured by the baseline regression is not attributable to inherent pre-policy differences between the treatment and control groups, and that the research conclusion has good robustness. Meanwhile, the signs of the control-variable coefficients are basically consistent with expectations; for example, econ and gap significantly aggravates educational inequality, which indirectly verifies the rationality of the model specification.

5. Further Analysis

5.1. Regional Heterogeneity Analysis.

The National Big Data Comprehensive Pilot Zone policy has a positive promotional effect on regional educational equity. However, because regions in different locations differ in economic foundations, educational resources, and digital infrastructure, the promotional effect of the policy may also vary. Wang Jin et al. (2026) [19] similarly find that the driving effect of pilot-zone establishment on regional technological innovation differs significantly among the eight pilot areas; among them, the effects in the Pearl River Delta, Shanghai, and Chongqing are insignificant, suggesting that the performance of policy effects is constrained by each region's economic foundation, innovation structure, and development goals. Zeng et al. (2025) [20] likewise find significant regional differences in the impact of pilot-zone policy on the scale and speed of capital allocation in different cities, further confirming that the effect of digital policies depends on cities' economic foundations and institutional environments. Existing research confirms significant regional heterogeneity in the effects of such digital policies. For example, Yi (2025) [21], based on the "Broadband China" strategy, finds that digital infrastructure has a more significant inequality-reducing effect on employment in eastern regions, high-human-capital cities, and municipalities directly under the central government. Ma & LIn (2023) [23], based on the quasi-natural experiment of "Broadband China," show that digital infrastructure construction has a stronger promotional effect on green economic transformation in eastern and economically developed cities. Bergantino et al. (2026) [24], in their study of broadband coverage and income inequality across Italian provinces, also find that broadband expansion aggravates inequality in southern provinces with structural vulnerability and limited human capital, indicating that the distributional effect of digital infrastructure has evident "threshold conditions": in regions lacking supporting human capital and institutional capacity, technology access may expand rather than bridge existing disparities. This cross-national comparative perspective further supports this study's finding that the insignificant policy effect in western China may not mean that the policy itself is ineffective; rather, it may be constrained by insufficient local human-capital accumulation and digital-technology absorption capacity, which have not yet reached the basic threshold needed for policy empowerment effects to be released. To test this heterogeneity, this study divides the full sample into four subsamples—eastern, central, western, and northeastern China—based on the geographical regional classification standards of the National Bureau of Statistics and then includes each subsample in the baseline model for regression.

Table 6. Regional Heterogeneity Analysis.

VARIABLES	(1) West	(2) East	(3) Central	(4) Northeast
$Treat_i \times Post_{it}$	-0.006	-0.015***	-0.009**	-0.015*
	(-1.471)	(-8.461)	(-2.260)	(-4.234)
Observations	150	90	165	45
R-squared	0.939	0.953	0.982	0.926
Controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

The regional heterogeneity regression results in Table 6 show significant regional differences in the policy effect. In the eastern region, the coefficient of the policy variable is significantly negative at the 1% level, indicating that the National Big Data Comprehensive Pilot Zone policy has the strongest and most significant effect in narrowing educational disparities there. In central and northeastern China, the policy coefficients are -0.009 and -0.015, respectively, both significantly negative at the 5% or 10% levels, suggesting that the policy also produces positive equity-promoting effects in these two regions, though with somewhat lower statistical significance or effect size than in the east. In western China, the policy coefficient is -0.006 and does not pass the 10% significance test (t = -1.471). This indicates that, in western China, the policy's contribution to educational equity has not yet fully manifested statistically. The regional heterogeneity results show that the effect of the National Big Data Comprehensive Pilot Zone policy on educational equity follows a gradient pattern: strongest in the east, followed by the central and northeastern regions, and insignificant in the west. The educational-equity effect of the policy is characterized by clear regional heterogeneity, and its effectiveness depends heavily on regional economic foundations and digitalization conditions.

5.2. Mechanism Analysis.

The theoretical hypothesis of this paper is that the National Big Data Comprehensive Pilot Zone policy does not directly affect educational outcomes but indirectly influences educational equity by activating three intermediary channels: fiscal allocation, industrial structure, and digital technology. Specifically, the policy provides a foundation for balanced allocation of educational resources by strengthening local fiscal capacity (the fiscal allocation effect); induces high-skilled talent demand through industrial upgrading and thereby forces supply-side adjustment in education (the human-capital demand effect); and empowers remote sharing of educational resources and precise governance by improving technological levels (the technology empowerment effect).

Table 7. Mechanism Analysis.

VARIABLES	(1) fsupp	(2) isadv	(3) digtech
$Treat_i \times Post_{it}$	0.028***	0.452**	0.000***
	(3.200)	(2.240)	(2.858)
Observations	450	450	450
R-squared	0.989	0.325	0.576
Controls	YES	YES	YES
Province FE	YES	YES	YES
Year FE	YES	YES	YES

Following the approach used by Wu Yongchao and Li Jianjiang (2026) [2] to test mediating effects, this study constructs a mediation-effect model and empirically tests the impact of the policy on the three mechanism variables described above:fsupp, isadv, and digtech.

Fiscal support fsupp is usually measured by the share of local fiscal education expenditure in public fiscal expenditure, reflecting the intensity of local government funding for education. The regression results show that the estimated coefficient of the policy variable $Treat_i \times Post_{it}$ on fsupp is 0.028 and significantly positive at the 1% level. This result indicates that implementation of the National Big Data Comprehensive Pilot Zone policy significantly strengthens the fiscal-support capacity of local governments and provides a fiscal foundation for provincial governments to increase educational investment and optimize the intra-provincial allocation of educational resources.

Isadv is usually measured by the ratio of value added in the tertiary industry to value added in the secondary industry (or by the share of tertiary-industry output), reflecting the trend of a region's industrial structure toward services and high-end development. The regression results show that the estimated coefficient of the policy variable $Treat_i \times Post_{it}$ on isadv is 0.452 and significantly positive at the 5% level. This result indicates that the pilot-zone policy effectively promotes industrial structure upgrading in pilot regions.

Digtech is usually measured by the number of Internet broadband access users per 100 persons (or by the share of value added in digital industries or digital infrastructure coverage), reflecting the penetration of regional digital infrastructure and the level of technology application. The regression results show that the estimated coefficient of the policy variable $Treat_i \times Post_{it}$ on digtech is 0.000 and significantly positive at the 1% level. This result indicates that the pilot-zone policy significantly improves the level of digital technology in pilot regions.

Combined with existing studies, these three pathways all point toward improvements in educational equity. In terms of fiscal support, Luo Weiqing (2010) [3] shows that optimizing the public expenditure structure under a fiscal-decentralization system can effectively alleviate shortages in educational supply caused by insufficient local fiscal capacity, thereby reducing educational inequality. In terms of industrial structure, Zhang et al. (2025) [25] argue that digital-economy-driven industrial upgrading significantly narrows the urban-rural income gap by optimizing human-capital allocation, indicating that skill-biased demand can strengthen the role of education in income adjustment and push the education system toward a more adaptive direction. In terms of digital technology, Xu et al. (2025) [13] show in practice that the digital economy breaks the geographical limitations of high-quality educational resources by accelerating educational digital transformation and promoting educational technology innovation, significantly improving the balance of educational resource allocation and narrowing urban-rural educational gaps. Wang et al. (2026) [22] similarly find that the effect of big-data development on improving urban environmental welfare performance is stronger in central and western non-old-industrial-base cities, but insignificant in city clusters with the lowest levels of economic development. This further supports the existence of "threshold conditions" for policy effects: only when economic foundations and digital infrastructure have accumulated to a certain level can the empowerment effects of the policy be effectively released. Yi et al. (2025) [21], based on firm-level empirical analysis, further find that big-data pilot-zone policies significantly promote firms' human-capital upgrading through data factors, providing direct micro-level evidence that data factors empower human-capital accumulation.

6. Conclusion

This paper takes the establishment of National Big Data Comprehensive Pilot Zones as a quasi-natural experiment and uses provincial panel data to examine the impact of big-data development policies on educational disparities within provinces. The empirical results show that the pilot-zone policy significantly narrows provincial educational disparities, and this conclusion passes a series of robustness tests, including the parallel-trend test, placebo test, and PSM-DID. Mechanism analysis indicates that the policy operates through three channels: strengthening local fiscal support for education, driving industrial structure upgrading to induce high-skilled education demand, and enhancing digital technology to empower educational resource sharing and precise governance. At the same time, heterogeneity analysis suggests that this effect is not uniform: it is significant in eastern, central, and northeastern China, but has not yet emerged in western China, indicating that the realization of policy effects depends on regional economic foundations and digitalization conditions.

Based on these conclusions, this paper proposes the following policy recommendations. First, in promoting the big-data strategy, attention should be paid to coordinated linkage with fiscal and industrial policies. Specifically, the policy dividends of pilot zones can be used to increase special education transfer payments to fiscally constrained localities, ensuring that they have sufficient financial capacity to improve school conditions. At the same time, instruments such as tax incentives and industrial guidance funds can be used to encourage firms in the digital-economy sector to participate in vocational-skills training and industry-education integration projects, so that industrial upgrading's demand for high-skilled talent can be effectively transmitted to the education supply side. Second, the education-empowering effect of digital technology should be strengthened, while risks of widening the digital divide must be guarded against. It is recommended that a unified provincial smart-education resource-sharing platform be built to extend high-quality digital educational resources to schools in remote areas, while establishing regular mechanisms for updating and evaluating the quality of digital educational resources to ensure that technological dividends truly benefit educationally weak regions. Third, policy implementation should be adapted to local conditions. For western and other weak-foundation regions, priority should be given to addressing shortcomings in digital infrastructure, increasing investment in education funds and teaching staff, and providing appropriate support in higher-education enrollment and vocational-education layout, so that data factors can release truly inclusive value in the field of educational equity. Only by coordinating efficiency and equity can the digital economy become a lasting driving force for common prosperity.

References

[1] State Council. The 14th Five-Year Plan for Digital Economy Development [Z]. 2022. (Guo Fa [2022] No. X)

[2] Wu Yongchao & Li Jianjiang. (2026). Big Data Development and Regional Educational Equity: Evidence from China's National Big Data Pilot Zones. Heilongjiang Higher Education Research, 44(03), 133-139. https://doi.org/10.19903/j.cnki.CN23-1074/G.2026.03.019.

[3] Luo Weiqing. (2010). Relationship between Fiscal Decentralization and Public Education Provision: Theoretical Model and Research on Prefectural Panel Data. Journal of Finance and Economics, 36(11), 39-50. https://doi.org/10.16538/j.cnki.jfe.2010.11.007.

[4] Farley, I. A., & Burbules, N. C. (2022). Online education is viewed through an equity lens: Promoting engagement and success for all learners. Review of Education, 10, e3367.

[5] Ventrella, F. M., & Cotnam-Kappel, M. (2024). Examining digital capital and digital inequalities in canadian elementary schools: insights from teachers. Telematics and Informatics, 86(Feb.), 102070.1-102070.13.

[6] Chen Shuangye, Meng Liujin & Zhang Haiyan. (2014). Education Policy Evidence in the Big Data Era:The Uses of Evidence-based Policy for Modern Education Governance and Scientific Educational Policy-making in China. Global Education, 43(02), 121-128.

[7] Wang, S., Alam, G. M., Skrbinjek, V., Bashir, K., & Miao, L. (2026). Managing digital bridges: How online social capital translates digital literacy into higher education access. Human Systems Management, 45(2), 139-154.

[8] State Council. Outline of Action for Promoting Big Data Development [Z]. 2015. (Guo Fa [2015] No. 50)

[9] Lan Xiangjie, Yang Lei.The Impact of Digital Infrastructure on the Level of Public Service Provision:A Quasi-Natural Experiment Based on National Big Data Comprehensive Pilot Zones [J]. East China Economic Management, 2026.

[10] Wang Qiankun. Household Income Increasing Effect of the National Big Data Comprehensive Pilot Zone [J]. West Forum, 2024, 34(5): 1-17.

[11] Liu, X. (2025). The impact of the digital economy on educational income inequity: evidence from household survey in china. Sustainability, 17.

[12] Wang, S., Alam, G. M., Skrbinjek, V., Bashir, K., & Miao, L. (2026). Managing digital bridges: How online social capital translates digital literacy into higher education access. Human Systems Management, 45(2), 139-154.

[13] Xu, Y., Fu, J., Zhou, W., & Bao, H. (2025). How can digital economy enable high-quality development of higher education? Evidence from an emerging country China. Industrial Management & Data Systems, 125(4), 1181-1202.

[14] Wen Xiaoguo. Exploring of Educational Development and Educational Equality in the Urban and Rural Areas of China: An Empirical Analysis Based on the Three Recent Nationwide Population Censuses [J]. Education & Economy, 2016(3).

[15] Vassilakopoulou, P., & Hustad, E. (2021). Bridging digital divides: a literature review and research agenda for information systems research. Information Systems Frontiers (2).

[16] Zhang, Y., Hainan, L., Feng, F., & Wu, X. (2025). Digital economy, education, human capital and urban–rural income disparity. Finance Research Letters, 71(000).

[17] Nie, C., & Wan, J. (2023). How does internet infrastructure construction affect urban-rural income gap? Evidence from a quasi-natural experiment in China. Progress in Development Studies, 23(3), 317-337.

[18] Liu, Q., & Li, T. (2024). When decentralization reduces regional disparity: Analyzing access to college in china. Sage Open, 14(1) doi: https://doi.org/10.1177/21582440231218679

[19] Wang Jin, Li Junyan & Wu Jianlong. (2026). The impact of the national big data comprehensive pilot zone on regional science and technology innovation capacity——Policy effect evaluation based on multi-period DID. Studies in Science of Science, 44(01), 166-177. https://doi.org/10.16192/j.cnki.1003-2053.20250120.001.

[20] Zeng, W., Wu, X., Zhu, K., & Zhu L. (2025). Can digitization policy improve capital misallocation in china? empirical evidence based on national big data comprehensive pilot zone policy. SAGE Open, 15(3).

[21] Yi, S., Duchin, F., Hagemann, H., Landesmann, M, Scazzieri, R, & Steenge, A., et al. (2026). Digital infrastructure and employment inequality: Evidence from the Broadband China Strategy.

[22] Wang, X., Song, J., Li, H, Beladi, H., & Chen, C. (2026). How does big data development affect urban environmental welfare performance in China? Evidence from national big data comprehensive pilot zones.

[23] Ma, R., & Lin, B. (2023). Digital infrastructure construction drives green economic transformation: Evidence from chinese cities. Humanities & Social Sciences Communications, 10(1), 460. doi: https://doi.org/10.1057/s41599-023-01839-z

[24] Bergantino, A. S., Fusco, G., Intini, M., Monturano, G., Andersson, M., & Kim, E., et al. (2026). Digital divide and income inequality: causal evidence from italian provinces. The Annals of Regional Science, 75(1).

[25] Zhang, Z., renxiaoting0103@163.com, Ren, X., 18153616900@163.com, & Yu, Y. (2025). How data factors drive to the new quality productive forces in China: the role of big data pilot zones.