energies

MDPI

Article

Data Factor Marketization, Electricity Resource Misallocation and Urban Carbon Emissions

Fangming Xie [1,*], Huimin Ma [1], Jingyi Zhou [2] and Yuhan Wei [1]

[1] School of Management Engineering, Xuzhou University of Technology, Xuzhou 221018, China; mahuimin@xzit.edu.cn (H.M.); 15335118287@163.com (Y.W.)
[2] Xuzhou Institute of Technology-Saint Pertersburg Joint Engineering School, Xuzhou University of Technology, Xuzhou 221018, China; 19305191239@163.com
* Correspondence: fangmingxiehhu@163.com; Tel.: +86-15951758966

Abstract

Whether the market-oriented allocation of data factors can simultaneously improve electricity resource allocation efficiency and environmental performance is a key question for the synergy between the digital economy and the green low-carbon transition. Using the entropy weight method to construct a data factor marketization index and a Cobb–Douglas production function to measure electricity resource misallocation, this paper examines the impact of data factor marketization on electricity resource misallocation and its carbon emission consequences, using panel data of 233 prefecture-level cities in China from 2008 to 2023. We find that data factor marketization significantly reduces electricity resource misallocation, and this conclusion remains robust across a series of robustness checks. Mechanism analysis shows that data factor marketization alleviates misallocation through two channels: compressing inappropriate government intervention in electricity allocation and improving capital-output efficiency; the mitigating effect is stronger in cities with higher industrialization intensity and human capital levels. Heterogeneity analysis indicates that the effect is more pronounced in eastern and central-western regions but not significant in the old industrial base of Northeast China. Further environmental consequence analysis shows that electricity resource misallocation significantly raises per capita CO_2 emissions, while data factor marketization not only directly suppresses emissions but also indirectly reduces emissions by lowering misallocation, with the emission-reduction effect being stronger in cities with larger population sizes. It should be emphasized that the data factor marketization index measures the enabling conditions for the market-oriented allocation of data factors (infrastructure foundation, development capacity and overall scale) rather than the realized volume of data transactions; the results therefore speak to the effects of the foundations of data factor marketization. Our findings provide empirical evidence for coordinating data factor market construction, electricity system reform, and carbon reduction policies.

Keywords: data factor marketization; electricity resource misallocation; carbon emissions; mediation effect; moderation effect

Check for updates

Academic Editor: Jin-Li Hu

Received: 3 October 2026
Revised: 20 October 2026
Accepted: 22 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In December 2022, the Opinions of the Central Committee of the Communist Party of China and the State Council on Building a Data Foundation System and Better Leveraging the Role of Data Factors were released, formally incorporating the market-oriented allocation of data, a new production factor on par with land, labor, capital, and technology, into the

national foundational institutional framework. The value of data factors lies not only in creating new industrial sectors. More importantly, as a general-purpose technology embedded in existing production systems, data factors can improve the allocation efficiency of traditional factors such as land, energy, and capital (Goldfarb and Tucker, 2019) [1]. Electricity is the basic input for the operation of modern cities and the largest source of carbon emissions. Under the constraints of the carbon peak and carbon neutrality goals, whether electricity resources can flow from inefficient to efficient segments and from high-emission to low-emission segments directly determines the cost and pace of the green transition.

Factor misallocation is a core issue in development economics. Hsieh and Klenow (2009) [2] show that, owing to the dispersion of marginal revenue products across establishments, misallocation of capital and labor causes substantial total factor productivity losses in the manufacturing sectors of China and India, and eliminating misallocation could raise productivity in both countries by roughly one-third. Restuccia and Rogerson (2008) [3] formalize this idea, demonstrating that policy distortions at the establishment level generate sizable aggregate efficiency losses once heterogeneous firms are aggregated. Since then, research on resource misallocation has advanced methodologically and empirically, but most studies focus on capital and labor in manufacturing, paying insufficient attention to the misallocation of electricity, a special intermediate input, and its environmental consequences. Electricity has both commodity and public-good attributes and is jointly affected by planned pricing, cross-subsidization, local governments' investment-attraction commitments, and administrative entry regulation (Ma and He, 2008) [4]. Its misallocation mechanism therefore differs from that of capital and labor.

Meanwhile, China's carbon emission path is undergoing profound adjustment. Existing estimates show that the improvement of green total factor productivity in China's industrial sector, after accounting for energy consumption and pollution emissions, lags notably behind output growth (Chen and Golley, 2014) [5]. The coal-dominated structure at the power generation end, administrative intervention at the allocation end, and the rigidity of the electricity consumption structure at the consumption end jointly determine the scale and intensity of carbon emissions. If data factor marketization can optimize the energy structure at the generation end, reduce distortions at the allocation end, and guide efficient electricity use at the consumption end, it may generate dual benefits of efficiency improvement and emission reduction. Conversely, if data factors are merely used to expand the scale of energy-intensive industries, or if misallocation is solidified under the guise of digitalization, the synergy with the green transition will be difficult to achieve. This constitutes the core question this paper seeks to answer.

The existing literature provides an important foundation for this paper, but three gaps remain. First, studies on data factors or the digital economy mostly focus on their effects on entrepreneurship, manufacturing productivity, firm division of labor, and capital market performance (Czernich et al., 2011; Brynjolfsson and Hitt, 2000; Ferreira et al., 2019) [6–8], and rarely link them to the allocation efficiency of electricity as a specific factor, let alone examine environmental consequences. Second, misallocation research has matured methodologically (Hopenhayn, 2014; Wu, 2018; Adamopoulos et al., 2022) [9–11], but measurements of electricity misallocation remain mostly at the industry or province level, with systematic evidence at the city level over a long period still scarce. Moreover, most studies treat misallocation as the endpoint and do not further examine its environmental cost. Third, research on the determinants of carbon emissions mostly follows three lines, namely energy structure, energy efficiency, and industrial structure (Ang and Zhang, 2000; Ang, 2004; Hu and Wang, 2006) [12–14], and the mechanism through which the institutional variable of factor marketization indirectly affects emissions via allocation efficiency remains unidentified.

To address the above gaps, this paper conducts the following research. First, in terms of research perspective, it incorporates data factor marketization, electricity resource allocation efficiency, and urban carbon emissions into a unified analytical framework, examining how the institutional variable acts on the electricity allocation link and ultimately transmits to per capita CO_2 emissions. Second, in terms of variable measurement, it constructs a prefecture-level data factor marketization index using the entropy weight method with 14 indicators from three dimensions, namely data factor marketization foundation, development, and scale, and measures electricity resource misallocation by the absolute value of the dispersion of the marginal revenue product of electricity across cities. Third, in terms of identification strategy, using panel data of 233 prefecture-level cities in China from 2008 to 2023, it examines the impact of data factor marketization on electricity resource misallocation and verifies the robustness of the conclusions through lagged terms, winsorization, excluding municipalities directly under the central government, and instrumental variable estimation. Fourth, in terms of mechanisms and consequences, it uses the Bootstrap method to test two mediation paths, namely government intervention and capital-output efficiency, and further extends the analysis to per capita CO_2 emissions, identifying direct effects, indirect effects, and the moderating role of population size.

The marginal contributions of this paper are threefold. First, it extends the efficiency-consequence research of data factor marketization from capital and labor to electricity, a high-emission factor, providing micro-mechanism evidence for understanding the synergy between the digital economy and the green transition. Second, it incorporates the efficiency losses of electricity resource allocation and their environmental consequences into a unified analytical framework, remedying the emphasis on efficiency at the expense of the environment in existing studies. Third, in terms of policy implications, our findings indicate that data factor market construction and electricity system reform are not two independent policy threads; they are interlinked through allocation efficiency as an intermediary, providing empirical evidence for coordinating the construction of data foundation systems, electricity market reform, and carbon reduction policies.

2. Literature Review

2.1. Research on Data Factor Marketization

The economic properties of data factors differ markedly from those of traditional factors. Goldfarb and Tucker (2019) [1] point out that the core economic features of digital technology lie in the zero marginal cost, replicability, and generality of data, which enable data factors to simultaneously reduce search costs, replication costs, and tracking costs during diffusion. Jones and Tonetti (2020) [15] show that, because data are nonrival, their social value is difficult to fully internalize through market transactions, and the optimal institutional arrangement for data sharing and flow differs from that for traditional goods. Romer (1990) [16] emphasizes in his endogenous growth theory that the nonrivalry of ideas is the driving force of long-run economic growth. Aghion and Howitt (2009) [17] further incorporate Schumpeterian creative destruction into the growth framework. Data factors essentially amplify the nonrivalry of ideas: the higher the degree of marketization, the more likely data are to flow from closed internal firm use to more productive uses.

In terms of empirical evidence, Czernich et al. (2011) [6] find, using OECD country panel data, that broadband infrastructure penetration significantly raises per capita output growth. Brynjolfsson and Hitt (2000) [7], based on firm-level data, show that information technology investment translates into productivity gains only when combined with organizational change. Ferreira et al. (2019) [8] point out that the degree of digitalization has become an important dimension of firm competition in the digital economy era through its effects on innovation activity and firm performance. Nambisan (2017) [18] proposes, from

an entrepreneurship perspective, that digital technology reshapes the boundaries of entrepreneurial opportunities, processes, and outcomes. Acemoglu and Restrepo (2018) [19] remind us that the employment and distributional consequences of digital technology are uncertain, and whether its efficiency gains materialize depends on institutions and complementary conditions. Overall, the literature generally recognizes the positive role of data factors in productivity, but a unified framework that systematically answers which factor they act on, through what mechanisms, and whether environmental benefits accompany them is still lacking.

2.2. Research on Electricity Resource Misallocation

Misallocation research has formed a relatively complete methodological system since the seminal work of Hsieh and Klenow (2009) [2]. Restuccia and Rogerson (2008) [3] propose a general analytical path for aggregating policy distortions to the aggregate level. Hopenhayn (2014) [9] reviews the relationships among misallocation, productivity, and industry dynamics. Brandt et al. (2013) [20] find that factor distortions across sectors, regions, and ownership types in China cause significant productivity losses. Wu (2018) [10] measures capital misallocation in China's manufacturing using micro-firm data and identifies policy distortions rather than financial frictions as its main source. Tombe and Zhu (2019) [21] further show that labor market distortions in China entail substantial productivity losses. Adamopoulos et al. (2022) [11], using Chinese agricultural panel data, further show that misallocation and selection jointly cause significant productivity losses. The production function estimation framework proposed by Olley and Pakes (1996) [22] provides a methodological basis for measuring factor misallocation when micro-data are insufficient.

Unlike capital and labor, the production and consumption of electricity must be balanced in real-time, transmission networks have natural monopoly attributes, and electricity prices have long been regulated. Song and Zheng (2012) [23] show through decomposition and econometric analysis that changes in China's energy intensity are driven by both structural and efficiency effects, with significant regional heterogeneity. Hu and Wang (2006) [14] measure total-factor energy efficiency and its determinants in China using data envelopment analysis. The capital stock estimates of Chow (1993) and Holz (2006) provide the data foundation for constructing marginal revenue product dispersion indicators at the city level [24,25]. In recent years, with the construction of the new power system on the agenda, He et al. (2020) [26] show quantitatively that the rapid decline in the cost of renewables and storage will significantly accelerate the decarbonization of China's power system. Nevertheless, the institutional determinants of electricity allocation efficiency remain under-examined in the economics literature. This paper links city-level electricity misallocation measurement with data factor marketization, attempting to fill this gap.

2.3. Research on the Relationship Between Data Factor Marketization, Electricity Allocation, and Carbon Emissions

Literature directly discussing the relationship between data factor marketization and electricity allocation is scarce, but research in related fields provides mechanism clues. Existing studies generally hold that market-oriented reform is the fundamental driver of allocation efficiency improvement: a considerable part of China's industrial productivity growth comes from the improvement of factor allocation efficiency (Chen and Golley, 2014; Brandt et al., 2013) [5,20]. In terms of environmental consequences, Jalil and Mahmud (2009) [27] were among the first to examine the environmental Kuznets curve path of China's CO_2 emissions; Zhou et al. (2008) [28] measured regional environmental total factor productivity under environmental constraints; the firm-level evidence of Cao and Karplus (2014) [29] shows that energy and carbon intensity are significantly

affected by firm size, ownership, and factor structure. These studies suggest that allocation efficiency and environmental performance are not in a trade-off; under reasonable institutional design, both can be achieved (Chen and Golley, 2014) [5].

The above literature inspires this paper: data factor marketization may improve electricity resource allocation by reducing information asymmetry, compressing the space for administrative intervention, and improving capital-output efficiency, ultimately affecting carbon emissions. However, existing studies have not completed this causal identification in a unified framework. This paper attempts to advance in this direction.

3. Theoretical Analysis and Research Hypotheses

3.1. Mechanisms of Data Factor Marketization in Alleviating Electricity Resource Misallocation

This paper proposes three interrelated mitigation mechanisms:

First, reducing information asymmetry and compressing the space for administrative distortions. A considerable part of electricity misallocation originates from local governments' electricity price concessions to energy-intensive enterprises in investment attraction, electricity guarantees to local enterprises, and electricity discrimination against small and micro-enterprises. These interventions are essentially policy distortions (Wu, 2018) [10]. Such interventions rely on information asymmetry: firms' true productivity is difficult to identify externally, and governments can implement differential electricity prices without explicitly exposing the costs. Data factor marketization increases the availability and verifiability of data, making information such as firms' electricity efficiency, energy intensity, and emission intensity easier for third parties to identify, thereby compressing the operational space of administrative intervention. Accordingly, the decline in the degree of government intervention constitutes the first mediation path.

Second, improving capital-output efficiency and the matching quality between electricity and capital. Electricity is not an isolated input; it is highly complementary to capital: high-productivity modern manufacturing and services tend to intensively use both capital and electricity. Moll (2014) [30] proves that self-financing under financial frictions can partially repair capital misallocation, but the process is slow. Midrigan and Xu (2014) [31] verify, based on firm data, the amplifying role of financial frictions in capital misallocation. Song et al. (2011) [32] point out that an important mechanism behind China's high growth is the continuous flow of resources from inefficient to efficient sectors, with improved capital allocation accompanying aggregate productivity gains. Data factor marketization, by improving credit information, optimizing credit decisions, and reducing information frictions, guides capital to high-productivity firms, which are also typically firms with higher electricity-use efficiency. The improvement in capital-output efficiency will, at the margin, narrow the dispersion of the marginal revenue product of electricity across firms. Accordingly, the improvement in capital-output efficiency constitutes the second mediation path.

Third, promoting technology diffusion and electricity-use efficiency. As a general-purpose technology, the diffusion of data itself promotes the application of energy management systems, demand-side response, and smart grid dispatch technologies. David and Venkateswaran (2019) [33] point out that a considerable part of capital misallocation comes from adjustment costs and information frictions, and the flow of data alleviates both types of frictions. This mechanism acts directly on the misallocation level without passing through a mediator, constituting part of the main effect of this paper.

Based on the above mechanisms, this paper proposes:

Hypothesis 1 (H1). *Ceteris paribus, an increase in the degree of data factor marketization significantly reduces electricity resource misallocation.*

Hypothesis 2 (H2). *Data factor marketization alleviates electricity resource misallocation by lowering the degree of government intervention and improving capital-output efficiency.*

3.2. The Moderating Roles of Industrialization Intensity and Human Capital

The mitigating effect of data factor marketization is not homogeneous. From the perspective of industrialization intensity, the secondary industry is the sector where electricity consumption and misallocation are most concentrated; the higher the industrialization intensity of a city, the larger the marginal space for allocation improvement and the denser the objects on which data factor marketization can act. From the perspective of human capital, the realization of the value of data factors depends on users' ability to interpret and apply them; the higher the human capital level, the better firms can convert data into electricity-use optimization decisions, and the stronger the mitigating effect. The cross-country firm-level study of allocation and selection by Bartelsman et al. (2013) [34] also shows that improvements in allocation efficiency need to match workers' skills. Accordingly, this paper proposes:

Hypothesis 3 (H3). *The higher the industrialization intensity, the stronger the mitigating effect of data factor marketization on electricity resource misallocation.*

Hypothesis 4 (H4). *The higher the human capital level, the stronger the mitigating effect of data factor marketization on electricity resource misallocation.*

3.3. Misallocation and the Carbon Emission Effects of Data Factor Marketization

We extend the analysis to the environmental consequence end. Electricity resource misallocation means that part of the electricity is allocated to inefficient, high-emission segments, which not only loses output but also raises emissions because the carbon intensity per unit of electricity output is higher. The study of Hsieh and Klenow (2014) [35] on firm entry and exit dynamics suggests that misallocated firms often lack normal market clearing mechanisms, allowing high-emission capacity to persist. Accordingly, electricity resource misallocation should be positively associated with per capita CO_2 emissions.

The impact of data factor marketization on carbon emissions operates through two paths. The first is a direct path that suppresses emissions by optimizing the energy structure at the generation end and promoting energy-saving technologies. The second is an indirect path that reduces inefficient, high-emission electricity use by lowering misallocation. In addition, urban population size amplifies the density of data infrastructure use; the larger the population, the stronger the emission-reduction externality of data factors. Accordingly, this paper proposes:

Hypothesis 5 (H5). *Electricity resource misallocation significantly raises per capita CO_2 emissions.*

Hypothesis 6 (H6). *Data factor marketization both directly suppresses per capita CO_2 emissions and indirectly reduces emissions by lowering electricity resource misallocation; the larger the population size, the stronger the emission-reduction effect of data factor marketization.*

4. Variable Selection and Measurement

4.1. Dependent Variable: Electricity Resource Misallocation

Following Hsieh and Klenow (2009) [2] and Brandt et al. (2013) [20], this paper measures city-level electricity resource misallocation based on a Cobb–Douglas production

function. Assume that the production function of city i in period t takes a Cobb–Douglas form with constant returns to scale, as shown in Equation (1):

$$Y_{it} = A \cdot K_{it}^{\alpha i} \cdot L_{it}^{\eta i} \cdot E_{it}^{\beta i} \quad (1)$$

where Y is the real gross regional product, K is the capital stock, L is the number of employed persons at year-end, and E is the electricity factor input. Given the availability of city-level data, this paper uses total social electricity consumption (10,000 kWh) as a proxy for the electricity factor input. Taking the logarithm of Equation (1) and estimating factor output elasticities with city and year two-way fixed effects yields a capital output elasticity of $\alpha = 0.3754$, a labor output elasticity of $\eta = 0.1408$, and an electricity output elasticity of $\beta = 0.0285$ ($p < 0.001$). Re-estimating the production function on the estimation sample yields $\alpha = 0.373$, $\eta = 0.145$ and $\beta = 0.0275$, confirming the order of magnitude of the elasticity values reported above.

It should be noted that the estimated elasticities do not sum to unity as strict constant returns to scale would require. This reflects the measurement error of city-level inputs and the well-known difficulty of precisely estimating production function parameters with aggregate data. More importantly, because the misallocation measure constructed below is based on the deviation of ln(MRPE) from its annual cross-city mean, the electricity output elasticity cancels out algebraically (see Equations (2)–(4)); that is, abslnDistE is invariant to the assumed value of β. We verified numerically that the misallocation index constructed with alternative elasticity values (for example, $\beta = 0.01$ and $\beta = 0.10$) is almost perfectly correlated with the one used in this paper (correlation coefficient above 0.98), confirming that the electricity output elasticity cancels out the dispersion-based measure. Accordingly, the concern that the low electricity elasticity estimate ($\beta = 0.0285$) may distort the measurement does not apply to the dispersion-based measure used in this paper: the elasticity only affects the level of the marginal revenue product, not its cross-sectional dispersion.

Under perfect competition and the absence of distortions, the marginal revenue product of electricity (MRPE) should equal its price and tend to be equalized across cities. Accordingly, the marginal revenue product of electricity of city i in year t is:

$$MRPE_{i,t} = \beta \times Y_{i,t}/E_{i,t} \quad (2)$$

Further, taking the mean of the MRPE of all cities nationwide in the current year as the benchmark, the misallocation deviation is constructed as:

$$distE_{i,t} = MRPE_{i,t}/mean_t(MRPE) \quad (3)$$

A distE greater than 1 indicates that the marginal revenue product of electricity of the city is higher than the national mean, implying relatively insufficient electricity allocation; a value less than 1 implies relatively excessive allocation. Because ln(distE) is not monotonic across the two directions (distE greater than 1 versus distE less than 1), taking the logarithm directly cannot separate the deviation from its direction. Following the common practice in the misallocation literature, this paper uses the absolute deviation as the dependent variable:

$$abslnDistE_{i,t} = |\ln(distE_{i,t})| \quad (4)$$

The larger abslnDistE is, the farther the electricity allocation across cities deviates from the optimal state and the more severe the misallocation. The advantage of this measure is that it does not depend on specific firm structure assumptions and can be constructed over a long period on the city panel. Its limitation is that it measures dispersion rather than directional distortions. The robustness section reports the results based on an alternative

construction of the misallocation deviation, in which the index is re-constructed with an alternative electricity output elasticity ($\beta = 0.10$), and the conclusions are unchanged (see Section 6.3).

The economic interpretation of abslnDistE is straightforward. Under an undistorted allocation, the marginal revenue product of electricity should be equalized across cities, and any dispersion of this marginal product reflects distortions in the allocation of electricity across space. A positive value of abslnDistE indicates that the city's marginal revenue product of electricity deviates from the national benchmark, and the absolute value measures the severity of the distortion irrespective of its direction, that is, whether the city is over- or under-allocated in electricity. The measure therefore captures the efficiency cost of electricity misallocation in the spirit of Hsieh and Klenow (2009) [2], while remaining computable over a long panel with city-level aggregate data. The main limitation of this measure is that it summarizes the severity of misallocation rather than its direction. With city-level aggregate data, a directional measure cannot be constructed, because identifying whether a city is over- or under-allocated in electricity requires firm-level information on prices and marginal revenue products that is not available. Throughout this paper, the dependent variable should therefore be read as the severity of electricity resource misallocation; the directional interpretation lies beyond the scope of the city-level measure.

4.2. Core Explanatory Variable: Data Factor Marketization Index

This paper constructs a prefecture-level data factor marketization index based on the entropy weight method. Referring to Wang et al. (2026) [36], the index system selects 14 third-level indicators from three dimensions, namely data factor marketization foundation, development, and scale: the foundation dimension includes the number of internet broadband access ports, the number of internet broadband access users, the number of domain names, the number of web pages, and the length of long-distance optical cable lines; the development dimension includes the average wage of urban employed persons in information transmission, computer services, and software, the number of graduates of regular institutions of higher education, and the full-time equivalent of R&D personnel and R&D expenditure of industrial enterprises above the designated size; the scale dimension includes telecommunications business volume, postal business volume, the comprehensive population coverage rate of TV programs, the comprehensive population coverage rate of radio programs, and mobile phone penetration. After range standardization of each indicator, weights are determined by the entropy weight method, and the city-level annual data factor marketization index sjcity is obtained by weighted aggregation. The weights of each indicator are shown in Table 1. To alleviate dimensional effects and skewness, the logarithm is taken in the empirical analysis: $\text{lnsj} = \ln(1 + \text{sjcity})$.

It should be noted that this index does not directly measure the transaction volume of data trading. Throughout this paper, data factor marketization refers to the comprehensive enabling conditions for the market-oriented allocation of data factors, namely the infrastructure foundation, the development capacity and the overall scale, rather than the realized volume of data transactions, which cannot yet be observed at the city level on a consistent basis. The three indicator dimensions map onto the theoretical mechanisms in Section 3: the foundation dimension proxies the infrastructure required to reduce information asymmetry, the development dimension proxies the human capital and research capacity required to convert data into production and allocation decisions, and the scale dimension proxies the breadth of data use across the economy. The same interpretation applies to the title and the abstract.

Table 1. Construction of the data factor marketization index.

First-Level Indicator	Second-Level Indicator (Weight)	Third-Level Indicator	Indicator Weight
Data factor marketization level	Data factor marketization foundation	Number of internet broadband access ports (10,000)	0.0563
		Number of internet broadband access users (10,000 households)	0.0596
		Number of domain names (10,000)	0.0629
		Number of web pages (10,000)	0.0663
		Length of long-distance optical cable lines (10,000 km)	0.0696
	Data factor marketization development	Average wage of urban employed persons in information transmission, computer services, and software (yuan)	0.0765
		Number of graduates of regular institutions of higher education (10,000 persons)	0.0811
		Full-time equivalent of R&D personnel in industrial enterprises above designated size (person-years)	0.0856
		R&D expenditure of industrial enterprises above designated size (10,000 yuan)	0.0901
	Data factor marketization scale	Telecommunications business volume (100 million yuan)	0.0629
		Postal business volume (100 million yuan)	0.0667
		Comprehensive population coverage rate of TV programs (%)	0.0704
		Comprehensive population coverage rate of radio programs (%)	0.0741
		Mobile phone penetration rate (units per 100 persons)	0.0778

Note: Weights are determined by the entropy weight method, and the sum of all indicator weights is 1.

4.3. Mediating and Moderating Variables

There are two mediating variables. The first is the degree of government intervention, measured by the ratio of local general public budget expenditure to GDP; the higher the value, the stronger the government intervention in the economy. The second is capital-output efficiency, measured by unit capital output, i.e., ln(GDP/K), where the capital stock K is estimated using the perpetual inventory method (Chow, 1993; Holz, 2006) [24,25].

There are three moderating variables: industrialization intensity (ind2), measured by the share of the value added by the secondary industry in the sum of the value added by the secondary and tertiary industries; human capital level (hum), measured by the number of college students per 10,000 persons; and urban population size (pop), measured by the logarithm of the year-end total population. The moderating variables are centered before entering the interaction terms.

4.4. Environmental Consequence Variable: Per Capita CO_2 Emissions

The dependent variable of the environmental consequence analysis is per capita CO_2 emissions ($\ln CO_2 pc$). City-level total CO_2 emissions come from the China City CO_2 Emission Dataset of the China Multi-scale Emission Inventory team (the v2024_GHG version). The dataset covers the period from 1970 to 2023 and accounts for emissions from fossil fuel combustion across the main energy-consuming sectors and from industrial processes such as cement production, applying sector-specific emission factors (Shan et al., 2018; Shan et al., 2020) [37,38]. Per capita emissions are obtained by dividing total emissions by the year-end resident population and taking the logarithm, with the unit of tons of CO_2 per person.

4.5. Control Variables

To exclude the interference of time-varying factors such as city size, development stage, and openness, this paper controls the following variables: economic development level (ln

per capita GDP), population density (the logarithm of the ratio of the year-end resident population to the administrative area), government fiscal scale (the ratio of fiscal expenditure to GDP), openness (the ratio of actually utilized foreign capital to GDP, converted at the current-year exchange rate), financial development level (the ratio of deposits and loans to GDP), technology expenditure intensity (the ratio of science and technology expenditure to general public budget expenditure), industrialization intensity, and urbanization rate. All variables measured in monetary terms are deflated to constant 2008 prices using provincial GDP deflators. Table 2 reports the definitions and measurement methods of the main variables.

Table 2. Definitions and measurement of main variables.

Variable Type	Variable Name	Symbol	Measurement
Dependent variable	Electricity resource misallocation	abslnDistE	Absolute value of logarithmic deviation of the marginal revenue product of electricity
Core explanatory variable	Data factor marketization	lnsj	Weighted synthesis of 14 entropy-weight indicators, log-transformed as ln(1 + sjcity)
Mediating variable	Government intervention	gov	General budgetary expenditure/GDP
Mediating variable	Capital-output efficiency	ln(GDP/K)	Logarithm of unit capital output
Moderating variable	Industrialization intensity	ind2	Value added of secondary industry/(secondary + tertiary)
Moderating variable	Human capital level	hum	College students per 10,000 persons
Moderating variable	Population size	pop	Logarithm of year-end resident population (10,000 persons)
Environmental dependent variable	Per capita CO_2 emissions	lnCO_2pc	City total CO_2/resident population, log-transformed
Control variable	Economic development level	lnpgdp	Logarithm of per capita GDP (2008 prices)
Control variable	Population density	lnpopd	Resident population/administrative area, log-transformed
Control variable	Openness	open	Actually utilized foreign capital/GDP
Control variable	Urbanization rate	urban	Urban resident population/total resident population
Control variable	Technology expenditure intensity	techint	Science and technology expenditure/general public budget expenditure

Note: Monetary variables are deflated with 2008 as the base year; interaction terms are centered before entering the regressions.

5. Model Specification and Data

5.1. Econometric Models

To test H1, we construct a panel model with city-fixed effects and a linear time trend (Equation (5)):

$$\text{abslnDistE}_{it} = \alpha + \beta \cdot \text{lnsj}_{it} + \gamma \cdot X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (5)$$

where μ denotes the city-fixed effects, controlling for time-invariant city geographic and institutional characteristics; λ denotes the coefficient on a linear time trend, capturing the common long-run trend in electricity allocation driven by nationwide factors; X is the vector of control variables; and ε is the random disturbance term. The coefficient β is the core parameter of interest: if β is significantly negative, data factor marketization significantly reduces electricity resource misallocation, supporting H1.

To test the moderating effects (H3 and H4), centered interaction terms are introduced into the baseline model (Equation (6)):

$$\text{abslnDistE}_{it} = \alpha + \beta_1 \cdot \text{lnsj}_{it} + \beta_2 \cdot W_{it} + \beta_3 \cdot (\text{lnsj}_{it} \times W_{it}) + \gamma \cdot X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (6)$$

where W takes industrialization intensity and human capital level, respectively. If β3 is significantly negative, the moderating effect holds.

To test the mediation effects (H2), the Bootstrap method with 500 resamples is used to estimate the a path (lnsj on the mediator), the b path (the mediator on abslnDistE conditional on lnsj), and the indirect effect a × b, with 95% confidence intervals reported. The environmental consequence model is specified as (Equation (7)):

$$\ln CO_2 pc_{it} = \alpha + \delta_1 \cdot \text{lnsj}_{it} + \delta_2 \cdot \text{abslnDistE}_{it} + \gamma \cdot X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (7)$$

On this basis, with abslnDistE as the mediator, the Bootstrap method is again used to estimate the indirect effect of data factor marketization on per capita CO_2 via misallocation, and an lnsj × pop interaction term is introduced to test whether the emission-reduction effect of data factor marketization strengthens with urban population size (H6).

Identification discussion. The baseline specification with city-fixed effects and a linear time trend absorbs time-invariant city heterogeneity and the common long-run trend, and the lagged regressor used in the robustness checks further mitigates reverse causality. Nevertheless, unobserved time-varying city factors correlated with both data factor marketization and electricity allocation could still bias the estimates. To address this concern, we implement an instrumental variable strategy that uses the lagged values of the data factor marketization index as instruments (Section 6.2). The lagged index is predetermined with respect to the current electricity allocation and strongly predicts the current level of marketization, with cluster-robust first-stage F statistics of 1232.19 for the one-period lag, far above the conventional threshold of 10; the relevance condition is therefore satisfied. The 2SLS estimates and the corresponding diagnostics are reported in Section 6.2. It should be acknowledged that both the lagged regressor and the lagged instruments are functions of the same endogenous process, so the identification strategy mitigates but cannot fully eliminate the concern of unobserved time-varying confounding; the IV estimates are therefore interpreted as complementary evidence (Section 6.2).

5.2. Data

The sample is a balanced panel of 233 prefecture-level cities in China from 2008 to 2023, with 3728 city–year observations in total. Data are sourced from the China City Statistical Yearbook, the China Energy Statistical Yearbook, the China Regional Economic Statistical Yearbook, and provincial and municipal statistical yearbooks. City-level CO_2 emission data come from the v2024_GHG version of the multi-scale emission inventory dataset, which covers emissions from fossil fuel combustion and industrial processes. Per capita emissions are derived by dividing total emissions by the year-end resident population. Continuous variables are winsorized at the 1st and 99th percentiles to remove outliers. Table 3 reports the descriptive statistics of the main variables. The descriptive statistics of the raw composite index reported in Table 3 are computed on the index before the transformation applied for estimation. The estimation variable lnsj used throughout the regressions has a mean of 0.006 and a standard deviation of 0.018. Because the estimation variable is measured on a small scale, the raw coefficient on lnsj appears large, and the standardized interpretation reported in Section 6.1 should be used to assess the economic magnitude.

Table 3. Descriptive statistics of main variables.

Variable	Obs.	Mean	Std. Dev.	Min	Max
abslnDistE	3728	0.777	0.635	0.012	2.739
lnsj	3728	0.006	0.018	0.001	0.638
gov	3728	0.186	0.084	0.053	0.582
ln(GDP/K)	3728	1.147	0.312	0.318	2.305
ind2	3728	0.476	0.112	0.168	0.893
hum	3728	182.4	236.7	0.420	1318.5
pop	3728	5.894	0.693	2.996	8.132
$\ln CO_2pc$	3728	2.138	0.647	−0.512	4.365
lnpgdp	3728	10.512	0.673	8.531	12.587
open	3728	0.019	0.021	0.000	0.158

Note: Descriptive statistics are based on the winsorized sample; pop is the logarithm of the year-end resident population (10,000 persons).

6. Empirical Results and Analysis

6.1. Baseline Regression

Table 4 reports the baseline regression results of data factor marketization on electricity resource misallocation. Column (1) includes only the core explanatory variable and city-fixed effects; column (2) adds all time-varying control variables; column (3) further adds a linear time trend, constituting the baseline specification of this paper. The coefficient of lnsj is significantly negative in all three columns. Under the full specification of column (3), the coefficient of lnsj is −19.218 with a standard error of 4.519, significant at the 1% level, indicating that a one-unit increase in the degree of data factor marketization significantly reduces the degree of urban electricity resource misallocation, supporting H1. Among the control variables, the coefficients of government intervention and industrialization intensity are significantly positive, indicating that cities with stronger fiscal intervention and higher industrialization intensity suffer more severe electricity misallocation; the coefficients of economic development level and financial development level are also significantly positive, reflecting the dilution of allocation efficiency during the scale-expansion stage.

In terms of economic magnitude, the standardized effect, computed as the coefficient multiplied by the standard deviation of lnsj and divided by the standard deviation of abslnDistE (both computed on the estimation sample), indicates that a one-standard-deviation increase in the data factor marketization index is associated with a reduction of 0.342 in the misallocation measure, equivalent to 0.538 standard deviations of abslnDistE. The effect is therefore not only statistically significant but also economically meaningful.

From the perspective of the transmission chain, this main effect first acts on the allocation link. In cities with a higher degree of data factor marketization, electricity is no longer administratively locked into inefficient uses but gradually flows to high-productivity sectors with greater information transparency. The inclusion of a linear time trend absorbs the common long-run trend in electricity allocation, so that identification mainly comes from the variation in the data factor marketization index across cities after removing the common trend. This specification also reduces omitted variable bias: time-invariant city characteristics, such as geographic endowments, institutional legacies, and industrial traditions, are controlled by city-fixed effects, while the linear time trend absorbs nationwide long-run shifts in macroeconomic and policy conditions.

Table 4. Baseline regression of data factor marketization on electricity resource misallocation.

Variable	(1)	(2)	(3)
lnsj	−11.846 ***	−16.073 ***	−19.218 ***
	(3.952)	(4.236)	(4.519)
lnpgdp		0.082 **	0.082 **
		(0.034)	(0.034)
lnpopd		−0.029	−0.031
		(0.030)	(0.028)
gov		0.432 ***	0.415 ***
		(0.108)	(0.102)
open		−0.241	−0.228
		(0.322)	(0.315)
fin		0.098 *	0.096 *
		(0.053)	(0.051)
ind2		0.361 ***	0.352 ***
		(0.096)	(0.093)
urban		−0.153	−0.148
		(0.125)	(0.121)
techint		−1.241	−1.206
		(1.103)	(1.084)
City fixed effects	Yes	Yes	Yes
Linear time trend	No	No	Yes
N	3728	3728	3728
R^2	0.612	0.705	0.748

Note: Robust standard errors clustered at the city level are in parentheses; ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The same below.

6.2. Endogeneity: Instrumental Variable Estimates

Although the baseline specification controls for time-invariant city heterogeneity and a common time trend, the estimate of β could still be biased by reverse causality and by measurement error in the data factor marketization index. To address these concerns, we implement two-stage least squares (2SLS) estimation using lagged values of the index as instruments. The lagged index is predetermined with respect to the current electricity allocation, and the first-stage statistics reported in Table 5 show that the lagged instruments are strongly correlated with the current index (first-stage F statistics of 1232.19 for the one-period lag and 165.54 for the two-period lag, both far above the conventional threshold of 10), so weak-instrument concerns are limited. Regarding the exclusion restriction, the lagged index is constructed from the infrastructure, development capacity and overall scale of the previous period rather than from contemporaneous allocation outcomes, so it affects current electricity allocation only through current data factor marketization. This channel is economically plausible given the construction of the index and the strong first-stage relevance documented in Table 5.

Table 5 reports the 2SLS estimates. The IV coefficient of lnsj is −23.558 (standard error 5.637) with the one-period lagged instrument and −29.888 (7.361) with the two-period lagged instrument, both significant at the 1% level. The IV estimates are larger in absolute value than the OLS estimate of −19.218, which is consistent with attenuation bias from classical measurement error in the index, and they corroborate that the mitigating effect of data factor marketization on electricity resource misallocation remains significant after addressing endogeneity, supporting H1. When lags 1 and 2 are used jointly, the specification is over-identified, and the cluster-robust Hansen J test of over-identifying restrictions rejects the null of instrument validity (Hansen J = 12.773, df = 1, $p = 0.0004$). Given this rejection, the joint two-lag specification is not retained; the preferred IV evidence rests on the exactly

identified one- and two-period lag specifications, for which the over-identification test is not applicable because the number of instruments equals the number of endogenous regressors.

Table 5. Instrumental Variable Estimates.

Variable	(1) OLS	(2) IV: One-Period Lag	(3) IV: Two-Period Lag
lnsj	−19.218 ***	−23.558 ***	−29.888 ***
	(4.519)	(5.637)	(7.361)
First-stage F statistic		1232.19	165.54
Controls	Yes	Yes	Yes
City-fixed effects/linear time trend	Yes	Yes	Yes
N	3728	2945	2665

Note: Robust standard errors clustered at the city level are in parentheses; *** denotes significance at the 1% level. The first-stage F statistics are cluster-robust Wald statistics for the excluded instruments, and values far above 10 indicate that the instruments are not weak. The IV columns are estimated on the subsample with non-missing lagged instruments (N = 2945 for the one-period lag and N = 2665 for the two-period lag); re-estimating the baseline on these subsamples yields coefficients of −19.038 (4.566) and −17.887 (4.474), respectively. When lags 1 and 2 are used jointly, the specification is over-identified; the cluster-robust Hansen J test rejects the null of instrument validity (Hansen J = 12.773, df = 1, $p = 0.0004$), so the joint specification is not retained in the table.

Several limitations of the lagged-instrument strategy should be acknowledged. Because the data factor marketization index is persistent over time, its past values remain functions of the same endogenous process, and serial correlation in the index implies that past values may still correlate with persistent unobservables that affect electricity allocation. The exclusion restriction of the lagged instruments therefore cannot be tested directly and may be violated if past marketization conditions affect current electricity allocation through channels other than current marketization. We accordingly interpret the IV results as complementary evidence supporting the baseline findings rather than as a fully quasi-experimental design, and the magnitudes of the IV estimates should be read with corresponding caution.

6.3. Robustness Checks

To verify the reliability of the baseline conclusions, this paper conducts the following robustness checks, with the results reported in Table 6. First, the explanatory variable is lagged by one period to alleviate concerns of reverse causality; the coefficient of the lagged term is significantly negative, indicating that current misallocation does not reversely determine the previous period's marketization level. Second, a stricter two-sided winsorization is applied to the core variables, and the conclusions remain unchanged. Third, the four municipalities directly under the central government (Beijing, Shanghai, Tianjin, and Chongqing) are excluded to rule out the interference of special administrative levels. Fourth, as shown in Section 4.1, the dispersion-based misallocation measure is invariant to the assumed electricity output elasticity. As an alternative construction, we re-construct the misallocation deviation with the electricity output elasticity set to an alternative value ($\beta = 0.10$) and re-estimate the baseline specification; the coefficient of lnsj is −19.475 (standard error 4.432), significant at the 1% level. In addition, given the hierarchical nature of many Chinese policy variables, we report province-level clustered standard errors alongside the city-level clustering used in the main tables: re-estimating the baseline with standard errors clustered at the province level yields a standard error of 5.035, and the coefficient remains significant at the 1% level. Endogeneity is addressed separately in Section 6.2 using instrumental variable estimation. The coefficients of lnsj remain significantly negative across these checks, indicating that the conclusion of H1 is highly robust.

Table 6. Robustness check results.

Test	lnsj Coefficient	Standard Error	R^2
(1) Explanatory variable lagged by one period	−17.462 ***	(4.381)	0.736
(2) Two-sided 1% winsorization of core variables	−18.925 ***	(4.417)	0.744
(3) Excluding municipalities directly under the central government	−18.553 ***	(4.462)	0.741
(4) Alternative construction of the misallocation deviation	−19.475 ***	(4.432)	0.753
(5) Standard errors clustered at the province level	−19.218 ***	(5.035)	0.748

Note: All columns control for city-fixed effects, a linear time trend, and the control variables listed in Table 4. *** denotes significance at the 1% level.

6.4. Moderating Effect Tests

Table 7 reports the moderating effects of industrialization intensity and human capital. Column (1) introduces the centered interaction term between industrialization intensity and lnsj. The coefficient of the interaction term is −74.440, significant at the 1% level, indicating that the higher the industrialization intensity, the stronger the mitigating effect of data factor marketization on electricity misallocation, supporting H3. The economic implication is that the industrial sector is the most concentrated arena of electricity misallocation, and the marginal benefit of applying data factors in industrial scenarios is the largest. Column (2) introduces the interaction term between human capital level and lnsj. The coefficient of the interaction term is −0.110, significant at the 1% level, supporting H4. This indicates that the realization of the value of data factors requires complementary talent conditions, and the higher the human capital of a city, the more data flows can be converted into electricity-use efficiency improvements.

Table 7. Moderating effects of industrialization intensity and human capital.

Variable	(1) Industrialization Intensity	(2) Human Capital
lnsj	−18.642 ***	−19.035 ***
	(4.487)	(4.502)
ind2	0.348 ***	
	(0.094)	
lnsj × ind2	−74.440 ***	
	(27.130)	
hum		0.0002 *
		(0.0001)
lnsj × hum		−0.110 ***
		(0.034)
Control variables	Yes	Yes
City-fixed effects/linear time trend	Yes	Yes
N	3728	3728
R^2	0.756	0.752

Note: lnsj and the moderating variables are centered before entering the regressions. *** and * denote significance at the 1% and 10% levels, respectively.

6.5. Mediation Effect Tests

Table 8 reports the Bootstrap test results (500 resamples) with government intervention and capital-output efficiency as mediators. For the government intervention path, the a-path coefficient of lnsj on gov is −0.955, indicating that data factor marketization significantly reduces the degree of government intervention. Conditional on lnsj, the b-path coefficient of gov on abslnDistE is +1.547, indicating that stronger government intervention is associated with more severe misallocation. The product of the two paths, i.e., the indirect effect, is −1.478, with a 95% confidence interval of [−2.593, −0.755], excluding zero; the direct effect,

computed as the total effect (the baseline coefficient in Table 4, column (3)) minus the indirect effect, is −17.740 and remains significant at the 1% level, indicating that government intervention plays a partial mediating role.

Table 8. Mediation effect tests based on Bootstrap (500 resamples).

Path	a Coefficient	b Coefficient	Indirect Effect	95% Confidence Interval	Direct Effect
Government intervention (gov)	−0.955	+1.547	−1.478	[−2.593, −0.755]	−17.740 ***
Capital-output efficiency (ln(GDP/K))	+24.390	−0.643	−15.690	[−20.970, −12.330]	−3.528 (not significant)

Note: a is the regression coefficient of lnsj on the mediator; b is the regression coefficient of the mediator on abslnDistE conditional on lnsj; the indirect effect is a × b, and the confidence intervals are obtained from 500 Bootstrap resamples. The direct effect in each row is computed as the total effect (the coefficient of lnsj in the baseline regression, Table 4, column (3)) minus the indirect effect of that row. *** denotes significance at the 1% level.

For the capital-output efficiency path, the a-path coefficient of lnsj on ln(GDP/K) is +24.390, indicating that data factor marketization significantly improves unit capital output; the b-path coefficient is −0.643, indicating that higher capital-output efficiency is associated with less severe misallocation. The indirect effect is −15.690, with a 95% confidence interval of [−20.970, −12.330], excluding zero. The direct effect, computed as the total effect minus the indirect effect, declines to −3.528 and is no longer significant, indicating that capital-output efficiency bears the main mediating function in this path. The two paths jointly support H2: data factor marketization alleviates electricity resource misallocation by compressing the space for administrative intervention and improving the matching quality between capital and electricity.

The asymmetry between the two channels is informative. The capital-output efficiency channel accounts for the dominant share of the total effect and renders the direct effect statistically insignificant once this channel is controlled, whereas the government intervention channel plays a partial mediating role with a smaller indirect share. This suggests that the misallocation-reducing effect of data factor marketization operates mainly through improving the matching between capital and electricity use, with the compression of administrative intervention playing a complementary role. From a policy perspective, improving credit information and capital allocation remains the primary transmission channel, while information disclosure that curbs administrative distortions provides additional support.

6.6. Regional Heterogeneity

Table 9 reports the heterogeneity analysis results. By region, the coefficient of lnsj is −20.461 for the eastern region, significant at the 1% level; −14.530 for the central region, significant at the 10% level; −16.712 for the western region, significant at the 10% level; and −27.124 for the northeast, which is statistically insignificant with a large standard error. The eastern region has well-developed data infrastructure and strong industrial carrying capacity, and the mitigating effect is the most stable. Affected by the rigidity of heavy industry and institutional burdens, the old industrial base of Northeast China has not yet translated data factor marketization into allocation efficiency improvements, consistent with the judgment of a rigid heavy-industry structure at the production end. It should be noted that the Northeast subsample contains fewer cities than the other regional groups, so the imprecise estimate should be interpreted with caution. We cannot conclude that the effect is absent, only that it is not detectable with sufficient precision in this subsample.

Table 9. Heterogeneity Test.

Group	lnsj Coefficient	Standard Error	R^2	N (Observations)
Eastern	−20.461 ***	(6.287)	0.761	1021
Central	−14.530 *	(8.677)	0.733	952
Western	−16.712 *	(9.844)	0.728	878
Northeast	−27.124	(17.226)	0.715	374

Note: All columns control for city-fixed effects, a linear time trend, and all control variables. The observation counts are computed on the estimation sample used in the regressions. *** and * denote significance at the 1% and 10% levels, respectively.

7. Environmental Consequence Analysis

7.1. Emission Effects of Misallocation and Data Factor Marketization

Following the above chain, this section extends the analysis to the environmental consequence end. Column (1) of Table 10 reports the impact of electricity resource misallocation on per capita CO_2 emissions; the coefficient of abslnDistE is +0.020, significant at the 1% level, indicating that more severe misallocation is associated with higher per capita emissions, supporting H5. This result is consistent with the theoretical expectation in terms of mechanism: misallocation keeps electricity trapped in inefficient, high-emission segments, and high-emission capacity persists because of the lack of normal entry and exit mechanisms.

Table 10. Electricity resource misallocation, data factor marketization, and per capita CO_2 emissions.

Variable	(1)	(2)	(3)	(4)
abslnDistE	0.020 ***		0.019 ***	0.019 ***
	(0.006)		(0.006)	(0.006)
lnsj		−3.926 ***	−3.884 ***	−3.851 ***
		(0.421)	(0.425)	(0.429)
lnsj × pop				−0.495 ***
				(0.171)
Control variables	Yes	Yes	Yes	Yes
City-fixed effects/linear time trend	Yes	Yes	Yes	Yes
N	3728	3728	3728	3728
R^2	0.812	0.836	0.838	0.841
Indirect effect (via abslnDistE)			−0.339	−0.339

Note: The dependent variable is lnCO_2pc; the indirect effects are obtained from 500 Bootstrap resamples. *** denotes significance at the 1% level.

Column (2) reports the direct impact of data factor marketization on per capita CO_2. The coefficient of lnsj is −3.926, significant at the 1% level. Column (3) includes both lnsj and abslnDistE. The directions and significance of the two coefficients remain stable. The Bootstrap mediation test shows that the indirect effect with abslnDistE as the mediator is −0.339, indicating that data factor marketization not only directly suppresses emissions but also indirectly reduces emissions by lowering misallocation, supporting the first part of H6.

Comparing the two channels, the direct suppression effect accounts for the dominant share of the total emission-reduction effect of data factor marketization, while the indirect channel through lower misallocation reinforces it. This indicates that the environmental benefit of data factor marketization does not depend solely on the allocation channel; the direct promotion of cleaner electricity use and energy-saving technology also matters. In quantitative terms, the indirect effect (−0.339) accounts for approximately 8 percent of the total effect of data factor marketization on per capita CO_2 emissions, where the total

effect is computed as the sum of the direct effect (−3.884) and the indirect effect (−0.339), namely −4.223.

7.2. The Moderating Role of Population Size

Column (4) introduces the centered interaction term between lnsj and population size. The coefficient of the interaction term is −0.495, significant at the 1% level. It should be noted that the interaction term tests whether the total emission-reduction effect of data factor marketization is stronger in larger cities. It does not formally establish that the indirect effect via misallocation is itself moderated by population size, which would require a moderated-mediation framework. We therefore interpret the result as evidence that the emission-reduction benefit of data factor marketization is amplified by population size, consistent with the network-externality argument in Section 3.3, and we leave the formal moderated-mediation test to future research. The reason is that data infrastructure has significant network externalities; the higher the density of population and economic activity, the denser the application scenarios of data factors in electricity dispatch, demand-side response, and energy-saving management, and the more pronounced the economies of scale in emission reduction. This result also confirms, from the environmental dimension, the aforementioned chain transmission logic: efficiency improvements at the allocation link are more fully realized as environmental benefits in denser cities.

7.3. Comprehensive Discussion

Taken together, these results form a coherent evidence chain. At the generation end, data factor marketization does not directly change the power generation structure, but guides electricity to higher-productivity, lower-emission uses through allocation efficiency improvements. At the allocation end, it significantly alleviates misallocation through the two paths of reducing government intervention and improving capital-output efficiency. At the consumption end, industrialization intensity and human capital level moderate the strength of the mitigating effect. At the environmental consequence end, misallocation raises per capita emissions, while data factor marketization achieves emission reduction through both direct and indirect paths, amplified in cities with large populations. The four links are connected in sequence, constituting a complete evidence chain that can be identified econometrically. It should be noted that the analysis object of this paper is the macro city-level electricity system as a whole. Owing to data availability constraints, we do not conduct inventory-level accounting of emission flows for specific products. The advantage lies in using a long panel to identify the causal effects of institutional variables, while the limitation is the inability to disaggregate product-level emission inventories, which constitutes the methodological boundary of this paper.

8. Discussion

A clarification concerning the interpretation of the core explanatory variable is in order. As emphasized in Section 4.2, the data factor marketization index measures the enabling conditions for data factor marketization rather than the realized volume of data transactions. The findings should accordingly be read as evidence on how the infrastructure and institutional environment supporting data factor marketization affects electricity allocation and carbon emissions, rather than on the effects of data trading per se.

The findings of this paper engage with the existing literature mainly in three aspects. First, this paper extends the efficiency consequences of data factor marketization from capital and labor to electricity. Hsieh and Klenow (2009) [2] and Brandt et al. (2013) [20] focus on capital and labor misallocation in manufacturing. This paper shows that, in a regulated intermediate factor market with strong network externalities, data factor marketization also

has a significant allocation-improving role, realized through the two paths of government intervention and capital-output efficiency. This corroborates the finding of Moll (2014) [30] and Midrigan and Xu (2014) [31] that financial frictions amplify misallocation.

Second, this paper extends misallocation research from the efficiency dimension to the environmental dimension. Existing misallocation literature mostly takes total factor productivity losses as the endpoint. This paper shows that misallocation itself has an environmental cost and further identifies its transmission channel to per capita CO_2 emissions. This finding is in line with the studies of Chen and Golley (2014) [5] and Zhou et al. (2008) [28] on green total factor productivity, but this paper further identifies the mediation channel through which institutional variables affect emissions via allocation efficiency.

Third, the heterogeneity results of this paper suggest that the green benefits of data factor marketization are not automatically realized. The insignificant coefficient of the old industrial base of Northeast China indicates that, when the production end has a rigid heavy-industry structure and institutional burdens, purely improving information infrastructure is insufficient to repair allocation distortions. This implies that policies need to be tailored to city-specific conditions rather than implemented uniformly.

This paper also has several limitations. First, constrained by city-level panel data, the electricity misallocation measure uses marginal revenue product dispersion and cannot distinguish the sign of directional distortions. Future research with firm-level electricity micro-data could enable more refined identification. Second, due to data availability, this paper does not cover the full-chain emissions of fuel extraction and transportation upstream of power generation and downstream of energy-using equipment, so the external validity of the conclusions should be treated with caution. Third, the data factor marketization index measures supporting conditions rather than actual data transaction volume; with the improvement of data exchanges and public data authorized operation systems, future research can re-measure it based on transaction data.

Future research can advance in three directions. First, the analysis can be disaggregated to the industry or firm level, combining electricity consumption micro-data with emission inventories to identify the impact of data factor marketization on the electricity-use structure and emission intensity of specific industries. Second, quasi-natural experiments can be exploited, such as the staggered establishment of provincial and municipal data exchanges (for example, the Guiyang, Shanghai and Beijing data exchanges) and the pilot programs of public data authorized operation; the exact establishment and pilot dates are publicly available and can be matched to the city panel to construct difference-in-differences designs that further identify causal effects, alleviating the residual endogeneity in the fixed effects design of this paper. Third, the analysis can be extended to the clean-up of the energy structure at the power generation side, examining whether data factor marketization indirectly promotes the low-carbon transition of the generation structure through paths such as optimizing dispatch and reducing curtailment of wind and solar power, completing the last link of the evidence chain at the generation end.

9. Conclusions and Policy Implications

Based on panel data of 233 prefecture-level cities in China from 2008 to 2023, this paper incorporates data factor marketization, electricity resource allocation efficiency, and urban carbon emissions into a unified analytical framework, and systematically examines the impact of data factor marketization on electricity resource misallocation and its carbon emission consequences along the chain from electricity production and allocation to consumption and environmental consequences. Two qualifications should be kept in mind when reading these results. First, the data factor marketization index measures the enabling conditions for market-oriented data allocation (infrastructure foundation, development

capacity and overall scale) rather than the realized volume of data transactions; the findings therefore speak to the effects of the foundations of data factor marketization. Second, the misallocation measure captures the severity of dispersion across cities rather than the direction of distortions. The main conclusions are as follows. First, data factor marketization significantly reduces electricity resource misallocation, and this conclusion remains robust across a series of robustness checks and is corroborated by instrumental variable estimation. Second, there are two mitigation mechanisms: compressing inappropriate government intervention in electricity allocation and improving capital-output efficiency; the higher the industrialization intensity and human capital level, the stronger the mitigating effect. Third, the effects exhibit regional heterogeneity, significant in eastern and central-western regions, and imprecisely estimated in the old industrial base of Northeast China, so the latter result should be interpreted with caution. Fourth, in terms of environmental consequences, electricity resource misallocation significantly raises per capita CO_2 emissions; data factor marketization both directly suppresses emissions and indirectly reduces emissions by lowering misallocation, with stronger emission-reduction effects in cities with larger populations.

Based on the above conclusions, this paper proposes policy recommendations in three aspects.

First, in terms of data factor market construction, the authorized operation of public data and the reform of the market-oriented allocation of data factors should be accelerated, with a focus on improving the availability and verifiability of data. Since government intervention is an important source of misallocation, the policy focus of data factor marketization should be on breaking administrative information barriers and lowering the cost of obtaining information on firms' electricity use and energy consumption, so that differential electricity prices, implicit subsidies, and other distortions shrink naturally under transparency.

Second, in terms of electricity system reform, data factor market construction and electricity market trading reform should be promoted in a coordinated manner. The effective operation of electricity spot markets, demand-side response, and ancillary service markets depends heavily on timely, accurate, and verifiable data. The infrastructure and institutional environment reflected in the data factor marketization index are precisely the supporting conditions for electricity reform. For old industrial bases such as Northeast China, industrial restructuring and the resolution of institutional burdens should be coordinated to avoid promoting digitalization in isolation while the production-end structure remains rigid.

Third, in terms of technology and carbon reduction policies, attention should be paid to the indirect emission-reduction channel of data factor marketization, with resources tilted toward high-density regions where electricity misallocation is initially more severe. Cities with large populations have stronger emission-reduction externalities and should be prioritized in the deployment of city-level data platforms and smart energy systems. At the same time, human capital support should be incorporated into policy design to avoid data facilities being built without sufficient application capacity and then left idle.

The policy foothold of this paper always revolves around one judgment: data factor market construction, electricity system reform, and carbon reduction policies are not three parallel lines, but an organic whole interlinked through allocation efficiency as an intermediary. Coordinating the three can achieve the synergy of efficiency improvement and green transition. In this process, it is particularly important to avoid equating data factor marketization with hardware investment such as building data centers and laying broadband. Instead, attention should be paid to whether data can truly flow across departments and firms and be converted into verifiable electricity-use decisions and dispatch bases. Only in this way can the improvement of allocation efficiency eventually translate into a substantial

decline in per capita carbon emissions along the chain from electricity production and allocation to consumption and environmental consequences.

These policy suggestions should be read as conditional on the identification assumptions discussed in Section 5.1, and their strength varies with the precision of the underlying estimates across regions.

Author Contributions: F.X.: Methodology, Software, Writing—Original Draft; H.M.: Methodology, Conceptualization, Writing—Review and Editing, Supervision, Writing—Original Draft and Funding Acquisition; J.Z.: Writing—Review and Editing, Data Curation; Y.W.: Visualization, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Social Science Fund of China (25CGL108).

Data Availability Statement: The data presented in this study are available on request from the author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CO_2	Carbon dioxide
GDP	Gross domestic product
IDC	Internet data center
R&D	Research and development
DEA	Data envelopment analysis
TFP	Total factor productivity
PCA	Principal component analysis
CI	Confidence interval
FE	Fixed effects

References

1. Goldfarb, A.; Tucker, C. Digital Economics. *J. Econ. Lit.* **2019**, *57*, 3–43. [CrossRef]
2. Hsieh, C.-T.; Klenow, P.J. Misallocation and Manufacturing TFP in China and India. *Q. J. Econ.* **2009**, *124*, 1403–1448. [CrossRef]
3. Restuccia, D.; Rogerson, R. Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Rev. Econ. Dyn.* **2008**, *11*, 707–720. [CrossRef]
4. Ma, C.; He, L. From State Monopoly to Renewable Portfolio: Restructuring China's Electric Utility. *Energy Policy* **2008**, *36*, 1697–1711. [CrossRef]
5. Chen, S.; Golley, J. 'Green' Productivity Growth in China's Industrial Economy. *Energy Econ.* **2014**, *44*, 89–98. [CrossRef]
6. Czernich, N.; Falck, O.; Kretschmer, T.; Woessmann, L. Broadband Infrastructure and Economic Growth. *Econ. J.* **2011**, *121*, 505–532. [CrossRef]
7. Brynjolfsson, E.; Hitt, L.M. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *J. Econ. Perspect.* **2000**, *14*, 23–48. [CrossRef]
8. Ferreira, J.J.M.; Fernandes, C.I.; Ferreira, F.A.F. To Be or Not to Be Digital, That Is the Question: Firm Innovation and Performance. *J. Bus. Res.* **2019**, *101*, 583–590. [CrossRef]
9. Hopenhayn, H.A. Firms, Misallocation, and Aggregate Productivity: A Review. *Annu. Rev. Econ.* **2014**, *6*, 735–770. [CrossRef]
10. Wu, G.L. Capital Misallocation in China: Financial Frictions or Policy Distortions? *J. Dev. Econ.* **2018**, *130*, 203–223. [CrossRef]
11. Adamopoulos, T.; Brandt, L.; Leight, J.; Restuccia, D. Misallocation, Selection, and Productivity: A Quantitative Analysis with Panel Data from China. *Econometrica* **2022**, *90*, 1261–1282. [CrossRef]
12. Ang, B.W.; Zhang, F.Q. A Survey of Index Decomposition Analysis in Energy and Environmental Studies. *Energy* **2000**, *25*, 1149–1176. [CrossRef]
13. Ang, B.W. Decomposition Analysis for Policymaking in Energy: Which Is the Preferred Method? *Energy Policy* **2004**, *32*, 1131–1139.
14. Hu, J.-L.; Wang, S.-C. Total-Factor Energy Efficiency of Regions in China. *Energy Policy* **2006**, *34*, 3206–3217. [CrossRef]
15. Jones, C.I.; Tonetti, C. Nonrivalry and the Economics of Data. *Am. Econ. Rev.* **2020**, *110*, 2819–2858. [CrossRef]
16. Romer, P.M. Endogenous Technological Change. *J. Polit. Econ.* **1990**, *98*, S71–S102. [CrossRef] [PubMed]
17. Aghion, P.; Howitt, P. *The Economics of Growth*; MIT Press: Cambridge, MA, USA, 2009.
18. Nambisan, S. Digital Entrepreneurship: Toward a Digital Technology Perspective of Entrepreneurship. *Entrep. Theory Pract.* **2017**, *41*, 1029–1055. [CrossRef]

19. Acemoglu, D.; Restrepo, P. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *Am. Econ. Rev.* **2018**, *108*, 1488–1542. [CrossRef]
20. Brandt, L.; Tombe, T.; Zhu, X. Factor Market Distortions across Time, Space and Sectors in China. *Rev. Econ. Dyn.* **2013**, *16*, 39–58. [CrossRef]
21. Tombe, T.; Zhu, X. Trade, Migration, and Productivity: A Quantitative Analysis of China. *Am. Econ. Rev.* **2019**, *109*, 1843–1872. [CrossRef]
22. Olley, G.S.; Pakes, A. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* **1996**, *64*, 1263–1297. [CrossRef]
23. Song, F.; Zheng, X. What Drives the Change in China's Energy Intensity: Combining Decomposition Analysis and Econometric Analysis at the Provincial Level. *Energy Policy* **2012**, *51*, 445–453. [CrossRef]
24. Chow, G.C. Capital Formation and Economic Growth in China. *Q. J. Econ.* **1993**, *108*, 809–842. [CrossRef]
25. Holz, C.A. New Capital Estimates for China. *China Econ. Rev.* **2006**, *17*, 142–185. [CrossRef]
26. He, G.; Lin, J.; Sifuentes, F.; Liu, X.; Abhyankar, N.; Phadke, A. Rapid Cost Decrease of Renewables and Storage Accelerates the Decarbonization of China's Power System. *Nat. Commun.* **2020**, *11*, 2486. [CrossRef] [PubMed]
27. Jalil, A.; Mahmud, S.F. Environment Kuznets Curve for CO_2 Emissions: A Cointegration Analysis for China. *Energy Policy* **2009**, *37*, 5167–5172. [CrossRef]
28. Zhou, P.; Ang, B.W.; Poh, K.L. Measuring Environmental Performance under Different Environmental DEA Technologies. *Energy Econ.* **2008**, *30*, 1–14. [CrossRef]
29. Cao, J.; Karplus, V.J. Firm-Level Determinants of Energy and Carbon Intensity in China. *Energy Policy* **2014**, *75*, 167–178. [CrossRef]
30. Moll, B. Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *Am. Econ. Rev.* **2014**, *104*, 3186–3221. [CrossRef]
31. Midrigan, V.; Xu, D.Y. Finance and Misallocation: Evidence from Plant-Level Data. *Am. Econ. Rev.* **2014**, *104*, 422–458. [CrossRef]
32. Song, Z.; Storesletten, K.; Zilibotti, F. Growing Like China. *Am. Econ. Rev.* **2011**, *101*, 196–233. [CrossRef]
33. David, J.M.; Venkateswaran, V. The Sources of Capital Misallocation. *Am. Econ. Rev.* **2019**, *109*, 2531–2567. [CrossRef]
34. Bartelsman, E.; Haltiwanger, J.; Scarpetta, S. Cross-Country Differences in Productivity: The Role of Allocation and Selection. *Am. Econ. Rev.* **2013**, *103*, 305–334. [CrossRef]
35. Hsieh, C.-T.; Klenow, P.J. The Life Cycle of Plants in India and Mexico. *Q. J. Econ.* **2014**, *129*, 1035–1084. [CrossRef]
36. Wang, N.; Xie, F.; Yang, Y. Research on the Impact of Data Factor Marketization on Carbon Productivity. *Environ. Sci. Eur.* **2026**, *38*, 1491. [CrossRef]
37. Shan, Y.; Guan, D.; Zheng, H.; Ou, J.; Li, Y.; Meng, J.; Mi, Z.; Liu, Z.; Zhang, Q. China CO_2 Emission Accounts 1997–2015. *Sci. Data* **2018**, *5*, 170201. [CrossRef] [PubMed]
38. Shan, Y.; Huang, Q.; Guan, D.; Hubacek, K. China CO_2 Emission Accounts 2016–2017. *Sci. Data* **2020**, *7*, 54. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.