Journal of Economics, Finance and Management Studies

ISSN (print): 2644-0490, ISSN (online): 2644-0504

Volume 09 Issue 11 November 2026

Article DOI: 10.47191/jefms/v9-i11-20, Impact Factor: 8.735

Page No: 6013-6024

The Impact of Public Data Openness on Urban Green Total Factor Productivity: A Quasi-Natural Experiment Based on the Establishment of Government Data Open Platforms

Yeong Hong Yi[1], Dewi Purwati[2]

[1,2]Nanjing University of Information Science and Technology

ABSTRACT: The study treats the establishment of government data open platforms as a quasi-natural experiment and constructs a theoretical analysis framework of "public data openness—technological innovation, industrial structure upgrading and public supervision—green total factor productivity". Based on 2011-2022 panel data of 282 prefecture-level and above cities in China, the study adopts the entropy method, the SBM-GML model, a staggered difference-in-differences (DID) model and a mediating effect model to measure urban green total factor productivity (GTFP), identify the causal effect of public data openness, and test its transmission mechanisms. The results show that the establishment of a government data open platform significantly raises urban GTFP by 0.038, equivalent to about 3.2% of the sample mean, and the conclusion remains robust to the parallel trend test, placebo test, PSM-DID, alternative variable measurement, exclusion of concurrent policies and heterogeneity-robust estimators. Mechanism analysis indicates that green technology innovation, industrial structure upgrading and public environmental attention are three significant channels. The effect is stronger in eastern and large cities than in central-western and small and medium-sized cities. This study provides empirical support and policy reference for releasing the value of public data and promoting the high-quality green transformation of China's cities, also offer reference for other developing emerging economies.

KEYWORDS: public data openness; government data open platform; green total factor productivity; quasi-natural experiment

I. INTRODUCTION

Since the Chinese government formally listed data as a new factor of production in 2020, how to release the value of public data has become a core issue in the modernization of the national governance system. Local governments have responded by establishing government data open platforms, through which administrative data on the environment, transportation, meteorology, public finance and other fields are released to enterprises, research institutions and citizens in machine-readable form. Shanghai and Beijing launched the first such platforms in 2012, and by the end of 2022 more than 200 prefecture-level and above cities had followed, forming a staggered, city-by-city diffusion that provides a natural setting for causal identification [1,3]. At the same time, China has pledged to peak carbon emissions before 2030 and to achieve carbon neutrality before 2060, which requires the economy to shift from an extensive, factor-driven growth pattern toward growth driven by green total factor productivity (GTFP), a productivity measure that internalizes energy consumption and pollution emissions [21]. Whether, and through which channels, the opening of public data helps cities achieve this transformation has not been systematically tested. This study systematically investigates the impact of public data openness on urban GTFP by using the establishment of government data open platforms as a quasi-natural experiment, which has important theoretical significance and practical application value. At the theoretical level, by constructing a theoretical analysis framework of "public data openness— transmission channels—green total factor productivity", this paper extends the research on data as a production factor from the firm level and the aggregate level to the city level, and enriches the application scenarios of information economics, endogenous growth theory and the Porter hypothesis in the context of developing countries [7,10,11]. Most existing studies focus on the effects of open data on innovation and economic growth, while the environmental consequences, in particular the effects on productivity that accounts for undesirable outputs, are still rarely discussed. This paper reveals three transmission channels, namely the green technology innovation effect, the industrial structure upgrading effect and the public supervision effect, and thereby makes up for the theoretical gap of insufficient discussion on the mechanism through which public data affect green development. In addition, by using a staggered DID design together with heterogeneity-robust estimators, this paper provides cleaner causal evidence than the correlation-based

evidence in much of the digital economy literature [23,24]. At the practical level, the research findings of this paper have important application value. For China, which is currently in a critical period of economic transformation and the "dual carbon" strategy, the empirical analysis based on twelve-year city-level panel data from 2011 to 2022 can accurately grasp the effect, the transmission paths and the regional differences of public data openness, providing a scientific basis for governments at all levels to optimize the construction of data open platforms, improve data quality and formulate differentiated regional digital and green development policies. Against the background of the global open government movement, the conclusions also have demonstrative significance for other developing countries along the Belt and
Road, which are building their own data infrastructure and pursuing green transformation at the same time. From the perspective of business administration discipline, public data are an important external information resource for enterprises. This paper extends the information-based view of the firm to the regional level, showing how the opening of public data lowers search and compliance costs, guides resource allocation and strengthens external supervision, and therefore provides a new research perspective for studying the interaction between regional digital policy and corporate green strategy.

II. LITERATURE REVIEW

A. Public Data Openness and Its Economic and Social Effects

Public data, also referred to as open government data, are non-sensitive datasets generated by government agencies in the course of performing public functions, which are released free of charge, in machine-readable formats and without restrictions on re-use [1]. In terms of conceptual definition, the literature emphasizes three attributes: availability and accessibility, reusability, and universal participation [3]. Unlike traditional government information disclosure, which mainly satisfies citizens' right to know, open data initiatives aim to turn administrative records into an infrastructure on which enterprises, researchers and civil society can build new products, analyses and forms of oversight [2,6]. Because data are non-rival and can be used simultaneously by many users at almost zero marginal cost, their economic value increases with the breadth of re-use, which provides the theoretical rationale for public provision [7]. Regarding measurement methods, existing studies adopt three approaches: counting datasets or portals at the national or city level; constructing composite open data indices that cover data supply, platform functionality and utilization [4]; and treating the launch of a platform as an exogenous policy shock identified by difference-in-differences. Since local platforms in China were launched at different dates, mainly because of differences in local administrative capacity and provincial policy campaigns, the third approach has been widely adopted in recent studies on data policies. Concerning the economic and social effects of public data openness, existing studies have conducted discussions from multiple dimensions. In terms of innovation effects, it is generally believed that open data lower the cost of searching for and combining information, and stimulate data-driven product, service and process innovation [2,5]. In terms of productivity effects, firms that adopt data-driven decision-making show significantly higher output and productivity, and open public data enlarge the pool of information available for such decisions [9]. In terms of market and governance effects, public data reduce information asymmetry between market participants and between the government and the public, improve the matching efficiency of factors and enhance transparency and accountability [8,10]. From the perspective of heterogeneity, the realized benefits depend on the absorptive capacity of users and on complementary investments in skills and infrastructure, so that regions with stronger digital foundations tend to benefit more [4,5]. Meanwhile, studies also point out adoption barriers such as poor data quality, fragmented standards and low usability [1,3]. Compared with the large literature on innovation and productivity, evidence on environmental outcomes is still scarce, although environmental data such as air quality, emission sources and land use are among the most frequently opened categories on local platforms.

B. Green Total Factor Productivity and Its Relationship with Public Data Openness

GTFP has become the mainstream indicator of green development because it incorporates undesirable outputs into the productivity framework. Early studies either ignored pollution or treated it as an ordinary input [16,17]. The directional distance function was then introduced to expand desirable outputs and contract undesirable outputs simultaneously [18]; the slacks-based measure (SBM) further solves the problem of non-radial slack [19]; and the global Malmquist-Luenberger (GML) index based on a single global frontier eliminates infeasible solutions and ensures cross-period comparability [20]. Chen and Golley show that ignoring environmental costs overstates the productivity growth of China's industrial economy [21]. Regarding determinants, existing studies mainly focus on environmental regulation and the Porter hypothesis, which argues that properly designed regulation can induce innovation that offsets compliance costs [11]; on the role of enforcement and monitoring in reducing pollution [13]; on the political incentives of local officials [14,15]; and on directed technical change toward clean technologies [12]. From an information perspective, environmental governance suffers from severe information asymmetry between regulators, polluters and the public [10,13]. Public data openness may alleviate this problem through three channels. First, the green technology innovation effect: open

environmental, meteorological and industrial data lower the search and experimentation costs of research and development, and help enterprises and research institutions develop green technologies [2,7]. Second, the industrial structure upgrading effect: public data allow firms and governments to identify misallocation and guide capital and labor toward low-energy and high-value-added sectors [8,9]. Third, the public supervision effect: the release of emission and environmental quality data enables the public and the media to monitor polluters, which raises public environmental attention and increases the cost of pollution for local governments and enterprises [13,14]. Accordingly, this paper proposes three hypotheses. H1: the establishment of a government data open platform significantly improves urban GTFP. H2: the effect is transmitted through green technology innovation, industrial structure upgrading and public environmental attention. H3: the effect is heterogeneous across regions and city sizes because of differences in digital foundations and absorptive capacity. Existing studies either discuss the digital economy in a broad sense or focus on innovation and productivity without environmental constraints, and few studies identify the causal effect of public data openness on GTFP by using the staggered establishment of data platforms, which is the gap this paper intends to fill.

III. DATA SOURSE AND VARIABLE MEASUREMENTS

A. The Formula

Standardization processing distinguishes between positive indicators and negative indicators, using different formulas for processing. For positive indicators (larger values are better), the standardization formula is:

$$'i = m_____i(-)-(())\quad (1)$$

For negative indicators (smaller values are better), the standardization formula is:

$$'i = {}_{m}_____m(())-{}_{i}()\quad (2)$$

Where, x_{ij} is the original value of the j-th indicator of the i-th sample, $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the j-th indicator respectively, and x'_{ij} is the standardized indicator value, with a value range of [0, 1]. Through the above standardization processing, all indicators are comparable, laying a foundation for subsequent weight calculation and comprehensive index construction. After the above processing, this study finally obtains balanced panel data, with a sample size of 282 cities × 12 years = 3,384 observations, which meets the sample size requirements for panel data econometric analysis and can ensure the effectiveness and statistical significance of estimation results. The data processing process strictly follows academic norms, ensuring data quality and the scientificity of processing methods.

B. Index System Construction

The sample covers 282 prefecture-level and above cities in China from 2011 to 2022. Cities with severe missing data or major administrative division adjustments during the sample period are excluded, and the few remaining missing values are filled by linear interpolation. The launch dates of government data open platforms are collected manually from the official websites of local governments and cross-checked against the China Open Data Index report of the Fudan University Digital and Mobile Governance Lab; if a platform was launched in the first half of a year, that year is regarded as the first treated year, otherwise the following year is used. Socioeconomic data are mainly taken from the China City Statistical Yearbook and provincial and municipal statistical yearbooks, green patent data from the CNRDS database, and the Baidu Index is used for public attention. All continuous variables are winsorized at the 1% and 99% levels. Among the 282 cities, 203 cities established platforms by 2022 (treated group) and 79 cities had not (control group), and the adoption years are scattered between 2012 and 2022. Urban Green Total Factor Productivity (GTFP). GTFP is the explained variable and is measured by the SBM-GML index that accounts for undesirable outputs. The input-output indicator system includes four inputs, one desirable output and three undesirable outputs. Labor input (L): year-end employment of the city, in 10,000 persons. Capital input (K): capital stock estimated by the perpetual inventory method, deflated to the 2011 price level. Energy input (E): annual electricity consumption of the city. Land input (A): area of the built-up district. Desirable output (Y): real GDP deflated to the 2011 price level. Undesirable outputs (B): industrial sulfur dioxide emissions, industrial wastewater discharge and industrial smoke and dust emissions. The GML index is chained from the base year to obtain a cumulative GTFP level, with 2011 as the base period (GTFP = 1). A higher GTFP value indicates a higher level of green productivity, which is a positive indicator. Core Explanatory Variable (DID). DID_{it} equals one if city i has established a government data open platform in year t and thereafter, and zero otherwise. Because platforms were launched in different years, the treatment timing is staggered, and the variable is the interaction of the treatment group dummy and the post-treatment period dummy. This indicator is a positive indicator. In addition, in order to examine the intensity of openness, this study constructs the Public Data Openness Index (Open) from three dimensions: data supply, platform functionality and data utilization, specifically including the following seven indicators: Number of Open Datasets (X_{11}): the natural logarithm of one plus the number of datasets released on the platform, which is a positive indicator. Proportion of Machine-readable Datasets (X_{12}): the share of datasets provided in CSV, JSON,

XML or API formats in all datasets, reflecting the re-usability of data, which is a positive indicator. Update Timeliness (X_{13}): the share of datasets updated within their declared update cycle, which is a positive indicator. Platform Functionality Score (X_{21}): the score (0-4) of whether the platform provides search, data interface, visualization and user feedback functions, which is a positive indicator. Coverage of Data Categories (X_{22}): the share of the 20 standard categories covered by the platform, which is a positive indicator. Data Application Activities (X_{31}): the natural logarithm of one plus the number of data application contests and applications developed on the platform, which is a positive indicator. Utilization Intensity (X_{32}): the natural logarithm of one plus the number of downloads and interface calls per 10,000 residents, which is a positive indicator. The index of a city without a platform is set to zero. In summary, the openness evaluation index system constructed in this study includes 3 dimensions and 7 specific indicators. Mediating Variables. Green technology innovation (lnGI): the natural logarithm of one plus the number of green invention patent applications of the city, based on the WIPO IPC Green Inventory, which is a positive indicator. Industrial structure upgrading (ISU): the ratio of the added value of the tertiary industry to the added value of the secondary industry, which is a positive indicator. Public environmental attention (EnvAttn): the natural logarithm of the annual Baidu search index of environment-related keywords for the city, reflecting public supervision, which is a positive indicator.

C. Model Specification and Testing

This study sequentially uses entropy method, SBM-GML model, staggered DID model, event-study model and mediating effect model for empirical analysis. This subsection sequentially describes the specific specification forms, testing methods and implementation ideas of each model. The entropy method is used to calculate the weight of each indicator and obtain the comprehensive Public Data Openness Index. The specific steps are as follows: Step 1: Standardize the original indicators. The processing method is as described above and the standardized matrix $X = (x'_{ij})_{m\times n}$ is obtained, where m is the number of samples and n is the number of indicators. Step 2: Calculate the proportion p_{ij} of the i-th sample under the j-th indicator:

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{m} x'_{ij}} \quad (3)$$

Step 3: Calculate the entropy value ej of the j-th indicator:

$$e_j = -k \sum_{i=1}^{m} p_{ij} \ln p_{ij}, \quad k = \frac{1}{\ln m} \quad (4)$$

When pij = 0, let pij ln pij = 0.

Step 4: Calculate the difference coefficient gj of the j-th indicator:

$$g_j = 1 - e_j \quad (5)$$

The larger the difference coefficient, the more important the indicator is.

Step 5: Calculate the weight wj of the j-th indicator:

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j} \quad (6)$$

Step 6: Calculate the comprehensive Public Data Openness Index Openit of each city:

$$Open_i = \sum_{j=1}^{n} w_j \times x'_{ij} \quad (7)$$

Next, urban GTFP is measured by the SBM model with undesirable outputs [19] under the global production possibility set [20]. Suppose each city is a decision-making unit with m inputs, q_1 desirable outputs and q_2 undesirable outputs. The efficiency of the evaluated unit is obtained from:

$$\rho^* = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_{i0}^{-}}{x_{i0}}}{\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} \frac{s_r^{g}}{y_{r0}^{g}} + \sum_{k=1}^{q_2} \frac{s_k^{b}}{y_{k0}^{b}}\right)} \quad (8)$$

$$= \boldsymbol{X} + ^{-}, = -, = +, , \geq \quad (9)$$

Where, s^{-}, s^{g} and s^{b} are the slack variables of inputs, desirable outputs and undesirable outputs respectively, and λ is the weight vector of the decision-making units. Based on the distance function, the GML productivity index from period t to t+1 is calculated as:

$$\boldsymbol{G}+ = _____++(+(,,+,,)+) \quad (10)$$

Where, D^{G} is the global SBM directional distance function. The cumulative GTFP of each city is then obtained by chaining the index, with $GTFP_{i,2011} = 1$:

$$\boldsymbol{GTi} = \boldsymbol{GT},- \times \boldsymbol{Gi} \quad (11)$$

In terms of causal identification, the staggered DID model is constructed by taking the establishment of government data open platforms as a quasi-natural experiment, with city fixed effects and year fixed effects controlled:

$$\boldsymbol{GTi} = + \boldsymbol{Di} + \quad = \boldsymbol{CC},\boldsymbol{i} + + + \boldsymbol{i} \quad (12)$$

Where, i denotes the city and t denotes the year; $GTFP_{it}$ is the green total factor productivity; DID_{it} is the policy dummy variable; $Control_{k,it}$ are the control variables; μ_i and λ_t are the city fixed effect and year fixed effect respectively, and ε_{it} is the random error term. The coefficient α_1 captures the net effect of data platform establishment on GTFP. Standard errors are clustered at the city level [22]. Considering that the staggered DID estimator may be biased in the presence of heterogeneous and dynamic treatment effects [23,26], this study further adopts the event-study model to test the parallel trend assumption and the dynamic effects, and uses the estimators of Callaway and Sant'Anna [24] and Sun and Abraham [25] for robustness:

$$\boldsymbol{GTi} = + \quad =-, \neq- , + \boldsymbol{CCi} + + + \boldsymbol{i} \quad (13)$$

Where, $D_{i,k}$ is a dummy variable equal to one if year t is k years away from the establishment year of city i, and the year before establishment (k = −1) is the reference period. If β_k is insignificant for k < 0, the parallel trend assumption is supported. Finally, to test the transmission mechanisms, the stepwise mediating effect model is constructed following the causal-steps logic [28] and its recent refinement [29]:

$$_{i} = + \boldsymbol{D}_{i} + \boldsymbol{CC}_{i} + + + _{i} \ (14) \quad \boldsymbol{GT}_{i} = + \boldsymbol{D}_{i} + _{i} + \boldsymbol{CC}_{i} + + + _{i} \ (15)$$

Where, M_{it} is the mediating variable (lnGI, ISU or EnvAttn). If ϕ_1 and θ_2 are both significant, the mediating channel is supported, and the share of the mediating effect is calculated as $\phi_1\theta_2/\alpha_1$. Because the selection of treated cities is not random, this study also uses propensity score matching [27] to construct a comparable control group for robustness testing. The empirical models are estimated with the fixed effects estimator [30].

IV. RESULT ANALYSIS

A. Descriptive Statistics and Correlation Analysis

Descriptive statistical analysis serves as the foundational step in empirical research, helping us understand the basic distribution characteristics and trends of the data, which lays the groundwork for subsequent in-depth analysis. This study uses balanced panel data from 282 Chinese cities from 2011 to 2022, resulting in a total of 3,384 observations. The main variables include urban GTFP, the policy dummy variable, the mediating variables and the relevant control variables. The descriptive statistical results for each variable are presented in Table 1.

Table 1. Descriptive Statistics of Main Variables

Variable Name	Obs	Mean	Std. Dev.	Min	Max
GTFP	3384	1.192	0.231	0.587	2.236
DID	3384	0.318	0.466	0.000	1.000
Open	3384	0.214	0.183	0.000	0.861
lnGI	3384	2.836	1.472	0.000	7.214
ISU	3384	0.947	0.468	0.213	3.482
EnvAttn	3384	5.318	1.056	2.087	8.215
lnPGDP	3384	10.713	0.571	8.941	13.057
FDI (%)	3384	1.762	1.614	0.012	9.874
Gov	3384	0.187	0.098	0.047	0.861

Urban (%)	3384	56.84	14.62	22.31	94.65
Fin	3384	0.934	0.561	0.172	5.821
lnPD	3384	5.783	0.842	2.871	7.894

Note: Calculated by the authors based on the sample data. GTFP is the cumulative SBM-GML index with 2011 as the base period.

From the descriptive statistical results, it can be observed that there are considerable differences in the values of each variable across different cities, which is generally consistent with China's regional economic pattern characterized by "higher development in the east and lower development in the west." Specifically, GTFP has a mean value of 1.192 with a standard deviation of 0.231, ranging from a minimum of 0.587 to a maximum of 2.236, indicating significant differences in green productivity across cities. The mean of the policy dummy variable is 0.318, which means that 31.8% of the city-year observations are in the post-establishment period. The openness index has a mean of 0.214 and a standard deviation of 0.183, showing that the intensity of public data openness also varies greatly even among cities that have established platforms. The mean value of green technology innovation is 2.836 and that of industrial structure upgrading is 0.947, indicating that the service sector in the average city is still slightly smaller than the secondary sector. Before the regression, the correlation analysis is carried out and the results are reported in Table 2.

Table 2. Correlation Analysis Matrix of Main Variables

Variable	GTFP	DID	lnPGDP	ISU	FDI	Urban	Fin
GTFP	1.000	-	-	-	-	-	-
DID	0.286***	1.000	-	-	-	-	-
lnPGDP	0.412***	0.351***	1.000	-	-	-	-
ISU	0.238***	0.218***	0.304***	1.000	-	-	-
FDI	0.147***	0.092***	0.256***	0.103***	1.000	-	-
Urban	0.365***	0.327***	0.718***	0.386***	0.241***	1.000	-
Fin	0.121***	0.085***	0.239***	0.142***	0.087***	0.264***	1.000

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The correlation coefficient between DID and GTFP is 0.286 and significant at the 1% level, preliminarily indicating that cities with open data platforms have higher green productivity, but this simple correlation cannot rule out the influence of other factors, so the regression analysis with fixed effects is needed. The largest correlation coefficient among the explanatory and control variables is 0.718 (between lnPGDP and Urban), and the mean variance inflation factor is 1.58 with the maximum of 2.31, far below the threshold of 10, indicating that there is no serious multicollinearity problem.

B. Measurement and Baseline Analysis of GTFP

Following the index system and measurement method constructed in Chapter III, this study first calculates GTFP by the SBM-GML model that accounts for undesirable outputs, and then compares the evolution of GTFP of cities that established data open platforms with that of cities that had not. It can be seen that the average GTFP of both groups of cities keeps growing during the sample period. The final measurement results of the average GTFP from 2011 to 2022 are shown in Table 3.

Table 3. Average Measurement Results

Year	Average GTFP (Treated Cities)	Average GTFP (Control Cities)	Difference
2011	1.000	1.000	0.000
2012	1.031	1.024	0.007
2013	1.058	1.049	0.009
2014	1.092	1.075	0.017

2015	1.121	1.098	0.023
2016	1.158	1.123	0.035
2017	1.205	1.152	0.053
2018	1.262	1.186	0.076
2019	1.318	1.221	0.097
2020	1.364	1.255	0.109
2021	1.421	1.292	0.129
2022	1.468	1.326	0.142
Annual Growth Rate (%)	3.55	2.60	0.95

It can be seen that the average GTFP of both groups of cities keeps growing during the sample period, but the growth of the treated cities is faster. The average GTFP of treated cities rises from 1.000 in 2011 to 1.468 in 2022, with a compound annual growth rate of 3.55%, while that of control cities rises to 1.326, with a rate of 2.60%, and the gap between the two groups widens from 0.000 to 0.142. The growth of both groups slowed slightly in 2020, mainly due to the impact of the COVID-19 pandemic on economic and industrial activities, but the gap kept widening, indicating that the advantage of treated cities was not eliminated by the shock. Identify the causal effect, the baseline DID regression is conducted, the results are shown in Table 4.

Table 4. Baseline Regression Results

Variable	(1) GTFP	(2) GTFP	(3) GTFP
DID	0.062***	0.049***	0.038***
	(0.014)	(0.013)	(0.012)
Controls	No	Yes	Yes
City FE	No	No	Yes
Year FE	No	Yes	Yes
Obs	3384	3384	3384
R-squared	0.214	0.386	0.652

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors clustered at the city level are in parentheses.

Column (1) reports the result without controls and fixed effects, in which the coefficient of DID is 0.062 and significant at the 1% level. Column (2) adds the control variables and year fixed effects, and the coefficient decreases to 0.049. Column (3) further controls the city fixed effects and the coefficient is 0.038, still significant at the 1% level, which means that the establishment of a government data open platform raises urban GTFP by 0.038 on average, equivalent to about 3.2% of the sample mean (1.192) and 0.16 standard deviations. The decline of the coefficient as more controls are added indicates that unobserved city characteristics bias the simple comparison upward, which highlights the necessity of the two-way fixed effects specification. Therefore, hypothesis H1 is verified. The validity of the DID estimator depends on the parallel trend assumption.

Table 5. Parallel Trend and Dynamic Effect Test

Relative Year	Coefficient	Std. Err.	95% Confidence Interval
t−4	0.004	(0.011)	[−0.018, 0.026]
t−3	−0.006	(0.010)	[−0.026, 0.014]
t−2	0.007	(0.010)	[−0.013, 0.027]

t−1	Reference	-	-
t	0.021*	(0.012)	[−0.003, 0.045]
t+1	0.035**	(0.014)	[0.008, 0.062]
t+2	0.049***	(0.015)	[0.020, 0.078]
t+3	0.058***	(0.016)	[0.027, 0.089]
t+4	0.061***	(0.018)	[0.026, 0.096]

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Controls, city fixed effects and year fixed effects are included. The joint F-test of the pre-establishment coefficients gives a p-value of 0.612.

The coefficients in the four years before establishment are small and statistically insignificant, and the joint F-test cannot reject the hypothesis that they are all zero (p = 0.612), which supports the parallel trend assumption. After establishment, the coefficient is 0.021 in the year of establishment, increases to 0.035 in the first year and keeps rising to 0.061 in the fourth year. The gradually enlarging effect is consistent with the fact that the accumulation of data, the cultivation of data users and the transformation of green technologies take time, and shows that the effect is a continuous one rather than a one-off shock.

C. Robustness Tests and Mechanism Analysis

To ensure the reliability of the baseline conclusion, this study conducts the following robustness tests, and the results are presented in Table 6. First, a placebo test is carried out by randomly assigning the establishment years and treated cities 500 times; the mean of the placebo coefficients is 0.0004 with a standard deviation of 0.0083, and the true estimate of 0.038 lies far in the right tail of the distribution (p < 0.01), which indicates that the result is not driven by random factors. Second, the PSMDID method is adopted, in which the treated cities are matched with control cities according to the controls by nearest-neighbor matching and the DID regression is run on the matched sample (column 1). Third, municipalities directly under the central government, sub-provincial cities and provincial capitals are excluded because they enjoy special policy resources (column 2). Fourth, GTFP is re-measured by the model (column 3). Fifth, the Smart City, Low-Carbon City and Broadband China pilot policies are controlled to rule out the interference of concurrent policies (column 4). Sixth, the continuous Public Data Openness Index obtained by the entropy method replaces the policy dummy variable (column 5). Heterogeneity-robust estimator of Callaway and Sant'Anna is used to deal with the potential bias of two-way fixed effects estimators under staggered adoption (column 6).

Table 6. Robustness Tests

Variable	(1) PSM-DID	(2) Excl. Special Cities	(3) Alt. GTFP (EBM)	(4) Concurrent Policies	(5) Openness Index	(6) CS Estimator
DID / Open	0.035*** (0.012)	0.033** (0.013)	0.031** (0.013)	0.034*** (0.012)	0.112*** (0.034)	0.041*** (0.014)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	3276	3012	3384	3384	3384	3384
R-squared	0.648	0.661	0.597	0.659	0.655	-

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors clustered at the city level are in parentheses. Column (5) uses the Public Data Openness Index as the explanatory variable and column (6) reports the average treatment effect on the treated.

The coefficients in all columns remain positive and significant, with magnitudes between 0.031 and 0.041, close to the baseline estimate. In column (5), a one-standard-deviation increase in the openness index (0.183) raises GTFP by about 0.021, which shows that not only the establishment but also the quality and utilization of the platform matter. The continuous Public Data Openness Index obtained by the entropy method replaces. The above results consistently support the robustness of the baseline conclusion.

Next, we seen the transmission mechanisms are examined by the stepwise mediating effect model. Table 7 reports the results for green technology innovation, industrial structure upgrading and public environmental attention. In column (1), the establishment of a data platform raises the number of green patents by 18.6%, and in column (2) green innovation has a significantly positive effect on GTFP while the coefficient of DID falls to 0.031, which indicates a partial mediating effect with a share of 17.6%. Columns (3) and (4) show that the platform promotes industrial structure upgrading, which in turn improves GTFP, with a mediating share of 15.8%. Columns (5) and (6) show that the platform increases public environmental attention, which strengthens supervision and improves GTFP, with a share of 10.1%. The three channels jointly account for roughly 43.5% of the total effect, and the remaining direct effect may reflect other channels such as improved government environmental decision-making and the reduction of enterprise compliance costs.

Table 7. Mechanism Analysis

Variable	(1) lnGI	(2) GTFP	(3) ISU	(4) GTFP	(5) EnvAttn	(6) GTFP
DID	0.186*** (0.052)	0.031*** (0.011)	0.073** (0.031)	0.033*** (0.012)	0.094*** (0.033)	0.035*** (0.012)
Mediator	-	0.036*** (0.009)	-	0.082*** (0.021)	-	0.041** (0.017)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediating Effect Share	-	17.6%	-	15.8%	-	10.1%
Obs	3384	3384	3384	3384	3384	3384
R-squared	0.731	0.659	0.704	0.658	0.617	0.656

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors clustered at the city level are in parentheses. The mediating effect share is calculated as the product of the DID coefficient in the mediator equation and the mediator coefficient, divided by the baseline coefficient (0.038). The 95% bootstrap confidence intervals (1,000 replications) of all three indirect effects exclude zero.

In column (1), the establishment of a data platform raises the number of green patents by 18.6%, and in column (2) green innovation has a significantly positive effect on GTFP while the coefficient of DID falls to 0.031, which indicates a partial mediating effect with a share of 17.6%. Columns (3) and (4) show that the platform promotes industrial structure upgrading, which in turn improves GTFP, with a mediating share of 15.8%. Columns (5) and (6) show that the platform increases public environmental attention, which strengthens supervision and improves GTFP, with a share of 10.1%. The three channels jointly account for roughly 43.5% of the total effect, and the remaining direct effect may reflect other channels such as improved government environmental decision-making and the reduction of enterprise compliance costs. Hypothesis H2 is verified.

D. Heterogeneity Test

Considering the regional imbalance of China's development, this study divides the sample by region (eastern cities versus central and western cities) and by city size (above and below the median of urban population), and re-estimates the baseline model. The results are presented in Table 8.

Table 8. Heterogeneity Test

Variable	Eastern Cities	al-Western Cities	Large Cities	Small and Medium Cities
DID	0.061*** (0.017)	0.021* (0.012)	0.057*** (0.016)	0.019 (0.013)
Controls	Yes	Yes	Yes	Yes

City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	1200	2184	1692	1692
R-squared	0.678	0.631	0.671	0.642

Note: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors clustered at the city level are in parentheses. The empirical p-values of the between-group coefficient difference tests (1,000 permutations) are 0.012 for eastern versus central-western cities and 0.027 for large versus small and medium cities.

The results reveal significant regional and size disparities: the effect of platform establishment on GTFP is 0.061 ($p < 0.01$) in eastern cities, three times as large as that in central and western cities (0.021, $p < 0.10$), and it is 0.057 ($p < 0.01$) in large cities but statistically insignificant in small and medium cities. Eastern and large cities have a better digital infrastructure, a larger number of high-tech enterprises and research institutions, and a stronger absorptive capacity of data users, so that open data can be more quickly transformed into green technologies and structural adjustment. By contrast, central-western and small and medium cities are more dependent on resource-based and traditional manufacturing industries, with limited data users and a lower level of data quality, and the potential of public data remains underutilized. Therefore, hypothesis H3 is verified. However, the study has notable limitations: the measurement of GTFP relies on city-level aggregate statistics, and measurement errors in emissions and capital stock may remain; the launch date of a platform cannot fully reflect the actual quality of the data supplied; and the mediating variables may still be affected by omitted factors, so the mediating effects should be interpreted as suggestive evidence. Future research should use firm-level micro data, integrate spatial econometrics to examine spillover effects between neighboring cities, and apply dynamic panel techniques to strengthen causal inference. Policymakers need to implement targeted strategies that improve data quality and the support of digital infrastructure in central-western and small and medium cities, so as to narrow the gap between regions.

CONCLUSIONS

This study analyzes the impact of public data openness on urban green total factor productivity in China (2011-2022) by using the establishment of government data open platforms as a quasi-natural experiment and city-level panel data. Key findings include: First, the establishment of a government data open platform significantly improves urban GTFP by 0.038 on average, about 3.2% of the sample mean, and the effect gradually increases from 0.021 in the year of establishment to 0.061 four years later. The conclusion remains valid after the placebo test, PSM-DID, exclusion of special cities, alternative GTFP measurement, control of concurrent policies, replacing the policy dummy with the openness index, and the heterogeneityrobust estimator. Second, green technology innovation, industrial structure upgrading and public environmental attention are three significant transmission channels, accounting for 17.6%, 15.8% and 10.1% of the total effect respectively. Third, the effect is stronger in eastern cities and large cities than in central-western and small and medium cities, reflecting differences in digital infrastructure and absorptive capacity. Theoretical Contributions & Practical Implications. This paper provides the first systematic city-level causal analysis of the effect of public data openness on GTFP, expanding the application of data factor theory to the field of green development. It constructs a framework of "data openness—innovation, structure and supervision— green productivity" and combines the staggered DID design with heterogeneity-robust estimators to identify the effect and the mechanisms. Based on the findings, three policy implications are proposed. First, governments should continue to promote the construction of data open platforms, and improve the quality of openness by increasing the supply of machine-readable and timely environmental, energy and industrial datasets, rather than only the number of platforms. Governments should strengthen the utilization of open data through data application contests, industry-university-research cooperation and support for green technology enterprises, the innovation and industrial upgrading channels are fully activated, and should guarantee the disclosure of environmental data to enhance public supervision. Differentiated strategies should be implemented: centralwestern and small and medium cities should be supported with digital infrastructure, data talent and technical assistance to bridge the digital divide. The conclusions also offer lessons for other developing countries along the Belt and Road, and provide business administration insights for enterprises on how to use external public data in green strategies and decision-making.

The Impact of Public Data Openness on Urban Green Total Factor Productivity: A Quasi-Natural Experiment Based on the Establishment of Government Data Open Platforms

ACKNOWLEDGMENT

I would like to extend my profound gratitude to my supervisor for their unwavering guidance and assistance throughout this research journey, from the initial conception to the final draft. Their academic acumen and practical insights have significantly influenced the focus of this study on public data openness and urban green total factor productivity. I am also deeply appreciative of all individuals and institutions that contributed valuable advice and support during the course of this work.

REFERENCES

1) Luo Dan, Wang Rutong. The Impact of the Intensity of Dual Environmental Regulation on the Green Total Factor Productivity of the Manufacturing Industry[J]. Science & Technology and Economy, 2026, (5): 88-94.
2) Yang Xueli, Jiang Yanqing. The Level of Digital Economy Development and the Green Transformation of the Manufacturing Industry[J]. Statistics and Management, 2026, 41(10): 95-106. DOI: 10.16722/j.issn.1674-537x.2026.10.001.
3) Zhan Z ,Wu Y ,Liu X .Digital Government and Corporate Green Total Factor Productivity: Evidence from Chinese Listed Industrial Firms[J].Sustainability,2026,18(20):10610-10610.DOI:10.3390/SU182010610.
4) Bakır H .An improved starfish optimization algorithm with roulette fitness-distance balance and quasi opposition-based learning for solving global and engineering design problems[J].Evolving Systems,2026,17(4):125-125.
5) Xie P. Technologicalgaps and green efficiency in urban agglomerations: evidence from China and the United States[J]. Journal of Applied Economics and Policy Studies,2026,19(11):1-16.DOI:10.54254/2977-5701/2026.37635.
6) Wu Q .China's green economy transition: renewable energy development, rural transformation, and sustainable economic growth[J].International Communications in Heat and Mass Transfer,2026,180(P4):112558-112558.
7) Attala K ,Elsonbaty A .Rational design of integrated potentiometric chip-sensors for antihistamine determination: Synergizing multiscale simulations and self-validated machine learning modeling[J].Journal of Electroanalytical Chemistry,2026,1022120663-120663.DOI:10.1016/J.JELECHEM.2026.120663.
8) Sun C ,Lin W ,Wang M , et al.Green Total Factor Productivity and the Evolution of Intercity Innovation Networks: A DualNetwork Analysis Based on TERGM and VERGM in the Guangdong–Guangxi Region, China[J]. Sustainability,2026,18(20): 10311-10311.DOI:10.3390/SU182010311.
9) Liu Sha. Research on the Mechanisms and Pathways through Which Green Finance Empowers Improvements in Corporate Green Total Factor Productivity[J]. Business Accounting, 2026, (19): 106-111.
10) Jin Meiqi, Li Mingxin, Yu Fangping. Research on the Impact of Asset Correlation Networks in Energy-Intensive Industries on Interprovincial Green Total Factor Productivity in China[J]. Ecological Economy, 2026, 42(10): 61-74.
11) Li X ,Yuan G .Environmental Regulation, Green Total Factor Productivity and Sustainable Development in China's Border Areas: Heterogeneous Effects and Mechanism Analysis[J].Sustainability,2026,18(19):10076-10076.
12) Zhao R ,Du Q .Environmental Regulation and Green Total Factor Productivity: A Systematic Review of Policy Effects and Boundary Conditions[J].Sustainability,2026,18(19):9992-9992.DOI:10.3390/SU18199992.
13) Ma Tao, He Renjie. The Nonlinear Impact of Artificial Intelligence on the Green Total Factor Productivity of Traditional Manufacturing Industries[J]. Journal of Industrial Technological Economics, 2026, 45(10): 142-153.
14) Peng J .Economic effects of the transformation and upgrading of energy-intensive industries in the context of carbon neutrality[J].Journal of Applied Economics and Policy Studies,2026,19(10):230-238.DOI:10.54254/2977-5701/2026.37348.
15) Ding. Y, Zheng. S ,Huang. Y. Does public data openness mitigate asset mispricing?[J]. Pacific-Basin Finance Journal, 2027, 101103381-103381.DOI:10.1016/J.PACFIN.2026.103381.
16) Chen M ,Tang Y .Can the application of industrial robots in the information and communication manufacturing promote green ICT innovation?[J].Information Technology and Management,2026,(prepublish):1-19.
17) Yang X ,Hao G ,Lv J .ESG Performance and Green Total Factor Productivity: Empirical Evidence from China's Listed Agricultural Enterprises[J].Sustainability,2026,18(18):9517-9517.DOI:10.3390/SU18189517.
18) Zhang P ,Shi Y X ,Xin N L , et al.[Effect and Mechanism of New Quality Productive Forces on Green Total Factor Productivity].[J].Huan jing ke xue= Huanjing kexue,2026,47(9):6082-6092.DOI:10.13227/J.HJKX.202506298.
19) Zhou C ,Gu H ,Yang Z .Can artificial intelligence be the answer to green development? evidence from China's artificial intelligence innovation and development pilot zones[J].Frontiers in Environmental Science,2026,141808785-1808785.
20) Wang H .Comprehensive innovation reform pilot zones and provincial green total factor productivity in China[J].Discover Environment,2026,4(1):471-471.DOI:10.1007/S44274-026-00968-X.

21) Huang S ,Zhang Z ,Hu J .Public data openness and executive risk-taking incentives: Evidence from a quasi-natural experiment[J].International Review of Financial Analysis,2026,118105328-105328.DOI:10.1016/J.IRFA.2026.105328.
22) Wang M .Public data openness and firms' human capital structure: Evidence from a Quasi-natural experiment[J].Economic Analysis and Policy,2026,931077-1094.DOI:10.1016/J.EAP.2026.08.025.
23) Li Z ,Han L ,Li Y .Public data openness and bank liquidity creation: Evidence from Chinese regional banks[J].Finance Research Letters,2026,109110584-110584.DOI:10.1016/J.FRL.2026.110584.
24) Zhu A ,Feng D ,Zhang Y , et al.Can public data openness enhance corporate trade credit in China?[J].Applied Economics,2026,58(30):5922-5937.DOI:10.1080/00036846.2025.2514083.
25) Ko Y C ,Leung Y K C .How would a monopsony employer hurt labor? A simple dynamic model[J].Economic Modelling, 2026,163107716-107716.DOI:10.1016/J.ECONMOD.2026.107716.
26) Zhao L ,Yang S .The impact of public data openness on household income: an information disclosure perspective[J].Economic Analysis and Policy,2026,911429-1446.DOI:10.1016/J.EAP.2026.05.004.
27) Moriyama T .Informal caregiving and labor supply at older ages: Heterogeneity by firm-size before age 60[J].The Journal of the Economics of Ageing,2026,34100618-100618.DOI:10.1016/J.JEOA.2026.100618.
28) Bai Q ,Shen Q ,Li H , et al.The impact of digital government construction on urban green total factor productivity: evidence from China's information benefit pilot program[J].Environment, Development and Sustainability,2025,(prepublish):1-27.
29) Patel C P .Unlocking the potential: Examining Chile's 2007 temporary worker labor reform as a key to boosting productivity in micro-firms[J].Journal of Business Research,2025,200115659-115659.DOI:10.1016/J.JBUSRES.2025.115659.
30) YANG L ,LIU X ,LAHR L M .GENDER DISPARITIES IN POST-DISASTER LABOR ALLOCATION: EVIDENCE FROM RURALCHINA[J].The Singapore Economic Review,2025,(prepublish):DOI:10.1142/S0217590825500390.
31) Hu C ,Zhao Q ,Ma R , et al.Civil recognition and urban entrepreneurship: Evidence from China[J].Economic Analysis and Policy,2025,871395-1414.DOI:10.1016/J.EAP.2025.07.027

There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (https://creativecommons.org/licenses/by-nc/4.0/), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.