sustainability

MDPI

Article

The Impact of Natural Resource Accountability Audit for Leading Cadres on Green Productivity: A Sustainability Perspective

Yichen Qin [1], Yalian Zhang [1,*], Wangwei Zhang [1] and Xin Guo [2]

[1] College of Economics and Management, Central South University of Forestry and Technology, Changsha 410004, China; m19976690708@163.com (Y.Q.); 2026213256@stu.sufe.edu.cn (W.Z.)
[2] School of Business and Creative Industries, University of the West of Scotland, Paisley PA1 2BE, UK; xin.guo@uws.ac.uk
* Correspondence: t20061325@csuft.edu.cn

Abstract

This study examines the relationship between the Natural Resource Accountability Audit for Leading Cadres (NRAALC) and green productivity using 104 city-year observations from 13 cities in Hunan Province during 2010–2017. A two-way fixed-effects difference-in-differences model relates staggered audit exposure to a composite index of technological, low-carbon, and digital productive capacity. The baseline coefficient yields a positive point estimate (0.0388; conventional OLS $p = 0.0077$), which remains positive after adding regional openness. Nevertheless, city-clustered standard-error inference ($p = 0.1585$) and wild-cluster bootstrap inference ($p = 0.1102$) fail to reject the null hypothesis of zero policy effect at the 5% significance level. While covariate-adjusted Goodman-Bacon decomposition reproduces the two-way fixed-effects baseline result, the Callaway-Sant'Anna cohort-specific estimator offers no statistically supportive evidence for the policy effect. Although audit exposure correlates positively with the promotion-pressure proxy variable, the full mediation regression and bootstrap confidence interval cannot confirm the indirect transmission mechanism. The positive moderating interaction between audit and officials' tenure is only statistically meaningful under conventional OLS rather than cluster-robust inference and should be interpreted as suggestive exploratory evidence. This study delivers context-specific correlational evidence regarding ecological accountability and regional green productive capacity; the magnitude and causal attribution of the policy effect remain inconclusive. Empirical conclusions are heavily bounded by the small number of treated cities and scarce post-treatment observations in our provincial sample.

Keywords: Natural Resource Accountability Audit for Leading Cadres; green productivity; sustainability

Check for updates

Academic Editor: Bernardino Benito López

Received: 2 September 2026
Revised: 22 October 2026
Accepted: 23 October 2026
Published: 26 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

Since the reform and opening-up, China's cadre evaluation system has gradually shifted from a political orientation to one focused on economic performance. The GDP-centric "promotion tournament" model has effectively driven the rapid growth of the national economy, but it has also given rise to extensive development problems such as excessive resource consumption and ecological and environmental degradation. To correct the one-sided performance incentives for local cadres, improve the ecological governance system, and advance the modernization of governance capabilities, the Chinese

government has implemented the Natural Resource Accountability Audit for Leading Cadres (NRAALC). This system has been steadily advanced over the years. In 2013, the Third Plenary Session of the 18th Central Committee of the Communist Party of China first proposed implementing this audit system and establishing a system of lifelong accountability for ecological and environmental damage. In 2015, the central government made arrangements to expand the pilot work, designating five locations, including Loudi, Hunan, as pilot cities. In 2018, the system was fully implemented nationwide. The 2025 Central Document No. 1 focuses on the development of new-quality productive forces in rural areas. Changes to audit evaluation in provinces such as Shandong illustrate efforts to connect the NRAALC with national development priorities. The policy is therefore relevant not only to environmental oversight but also to the incentives that shape local investment and production decisions. Whether those incentives are associated with higher green productivity is the empirical question addressed here.

The Fourth Plenary Session of the 20th CPC Central Committee places green economic and social transformation and the Beautiful China initiative among the priorities of the 15th Five-Year Plan. Meeting these objectives requires action by public agencies, firms, and other stakeholders. The NRAALC addresses the responsibilities of local officials by examining their management of natural resources and environmental protection during their terms. In principle, this can alter the balance between growth targets and ecological responsibilities in local decision-making. Officials facing closer scrutiny of resource use and pollution may give greater attention to environmental enforcement, the upgrading of existing industries, and investment in cleaner production. Accountability that extends beyond an official's term may also create incentives to consider the longer-term consequences of development decisions. These institutional arguments motivate the analysis; the pilot-period estimates below assess a specific provincial setting.

This paper studies the NRAALC pilot in Hunan Province to assess its relationship with regional green productivity. It examines promotion pressure as a possible transmission pathway and officials' tenure as a separate moderating condition. This distinction matters because the audit may alter performance incentives, while tenure may affect an official's ability to sustain a policy over time. The analysis builds on research into the environmental and economic consequences of the audit and uses a composite index covering technological, low-carbon, and digital development. The results can inform discussion of ecological assessment and local governance, including the green-development priorities of the 15th Five-Year Plan. Their application to current policy requires care, however, because the data describe the 2010–2017 pilot period rather than nationwide implementation.

The literature reviewed below covers the design and effects of the NRAALC, together with research on the measurement and development of green productivity. As the NRAALC is rooted in China's specific institutional setting, directly comparable international research is limited. Domestic research has mainly examined two areas: the design and implementation of the audit system and its policy and economic consequences, while related international studies provide broader insights into government auditing, public accountability, and external oversight mechanisms. The first direction concerns research on the construction and implementation of the audit system: Cui and Huang [1] propose that natural resource asset accounting serves as the key data foundation for evaluating officials' performance in fulfilling their resource and environmental responsibilities, and Liu Hanbin and Li Ziang [2] point out that current audits suffer from ambiguous scope and imperfect evaluation indicators, and that a scientific audit system should be constructed based on the value of natural resources. Zhang et al. [3] construct an evaluation index system for the NRAALC based on the Pressure–State–Response (PSR) model and the RS method and verify its applicability with a case study of Huaihua City, Hunan Province; and Zhang

et al. [4] confirm, from the perspective of main functional zones, that the NRAALC can significantly improve regional environmental performance. The second direction concerns research on the policy and economic effects of the NRAALC, in which scholars use the audit pilot as a quasi-natural experiment for empirical tests: Zeng [5], based on a difference-in-differences model, confirms that the NRAALC can improve the green innovation level of enterprises, with a more prominent incentive effect on green invention patents; Li and Guo [6] find, based on panel data of prefecture-level cities, that this policy can promote regional industrial upgrading and the iteration of green technologies. Huang Rongbing and Xie [7] further find that the NRAALC can significantly improve the environmental governance execution capacity of local governments; Ma Zhijuan and Yin [8] find that the NRAALC can improve environmental governance efficiency by enhancing the intensity and persistence of the government's environmental attention; and Ma et al. [9] further find that the NRAALC can improve water resource utilization efficiency and water environment quality. In terms of green total factor productivity, Zhu et al. [10] confirm that the NRAALC can significantly improve urban green total factor productivity, Wang et al. [11] find that it significantly improves the green total factor productivity of enterprises in pilot cities, Liu [12] confirms the same positive effect based on data from 110 prefecture-level cities in the Yangtze River Economic Belt, and Zhang et al. [13] further verify the "dual carbon" governance effect of the NRAALC. At the micro level, Han et al. [14], Sun and Guo [15], and Tang et al. [16] examine the effects and mechanisms of the NRAALC on corporate green innovation and green innovation efficiency, while Xu and Lian [17] document its role in reducing land-resource misallocation.

Foreign literature primarily focuses on government auditing, accountability systems, and external constraint mechanisms. Kim [18] argues that the design of the auditing system directly determines the effectiveness of accountability in the public sector; Rayamajhee and March [19] propose that economic liberalization enhances government accountability through media independence, public participation, and multi-party checks and balances; and Whiteley and Kölln [20] further distinguish the differentiated roles of party identification, noting that performance-based identification strengthens oversight and constraints, while affective identification weakens the effectiveness of accountability. Overall, international research generally reaches a consensus that government accountability requires the concerted efforts of multiple mechanisms, including audit oversight, institutional rules, and social supervision.

Chinese research on green productivity describes several ways in which production can become less resource-intensive. Liu [21] emphasizes the allocation of factors and changes in industrial composition, whereas Sun [22] examines technological upgrading, structural adjustment, and resource efficiency as sources of urban development. Other studies focus on the conditions needed for these changes. Dong [23] places green technological innovation at the centre of the process. Zhou and Wu [24] identify institutional weaknesses, technological limitations, and established consumption practices as obstacles. Together, these studies suggest that green productivity depends on both productive capabilities and the institutions governing their use.

Foreign scholars often analyse the logic behind the evolution of green productivity from the perspectives of environmental regulation, technological innovation, and market constraints. From a policy perspective, Tan and Zhang [25] and Salman [26] demonstrate that a combination of environmental policies and the optimization of the energy structure can positively drive the growth of green productivity; from a technological perspective, the digital economy, e-commerce, and artificial intelligence can all drive the growth of green productivity by reducing costs and improving efficiency through technological R&D; at the market and external-shock level, media coverage of ESG issues can compel enter-

prises to undergo green transformation through reputation and capital market constraints, while climate disasters can negatively affect agricultural green productivity by suppressing technological R&D and reducing production efficiency. Furthermore, regarding the measurement of green total factor productivity, Chen et al. [27] estimate the growth of green total factor productivity and its influencing factors based on data from China's industrial economy, while Mushtaq et al. [28] systematically examine the synergistic relationship between industrial green total factor productivity and the circular economy, providing methodological support for the quantitative research on green productivity.

Taken together, the existing literature has established a substantial foundation for understanding the environmental and economic consequences of the NRAALC. In particular, Zhu et al. [10] and Liu [12] provide city-level evidence that the NRAALC improves green total factor productivity, while related studies document its effects on corporate green productivity, green innovation, industrial upgrading, environmental governance, and policy transmission mechanisms [5–17]. Therefore, the relationship between the NRAALC and regional green performance should not be regarded as an unexplored research question. Nevertheless, further context-specific evidence is needed regarding the measurement, policy transmission mechanisms, and regional heterogeneity of the effects of the NRAALC on green development.

This paper adds context-bound provincial-level evidence to existing literature on ecological accountability and regional green development, addressing gaps left by prior city-level studies focusing mainly on green total-factor-productivity metrics [10–12]. First, we construct a multidimensional composite index capturing technological, low-carbon, and digital productive capacity, which conceptually differs from conventional green total-factor-productivity indicators and enriches measurement perspectives for green-development evaluation. Second, we explicitly separate the theoretical proposition of a promotion-pressure-driven mediating channel and tenure-based moderating condition, and comprehensively report both supportive and disconfirming statistical test outcomes instead of only reporting significant results [16]. Third, we systematically demonstrate the sensitivity of staggered DID estimation results to alternative inference strategies and different DID estimators. The core incremental contribution lies in transparently documenting empirical uncertainties and identification risks within a small provincial pilot panel, instead of attempting to identify a universal causal effect of the NRAALC policy.

2. Theoretical Analysis and Research Hypotheses

2.1. Natural Resource Accountability Audit for Leading Cadres and Green Productivity

The NRAALC, in accordance with laws and regulations, is an institutional audit process that oversees, evaluates, and holds leading cadres accountable for their performance in managing natural resource assets and protecting the ecological environment during their terms. It has the core functions of enforcing ecological responsibilities, overseeing governance, and empowering performance evaluations. It is a crucial institutional tool for advancing regional ecological governance and green development. Green productivity is an advanced form of productivity centred on green scientific and technological innovation, which balances resource conservation, pollution reduction, and production efficiency improvement to achieve synergistic benefits across three dimensions: economic benefits, social benefits, and ecological benefits. It is the core driving force behind the green transformation and high-quality development of regional economies.

Based on the above analysis, the NRAALC strengthens oversight and constraints on local governments' ecological responsibilities through institutionalized means, thereby reshaping regional economic development. From a macro perspective, it promotes green productivity by optimizing resource allocation and improving the institutional environ-

ment. As ecological responsibility constraints are strengthened, factors of production such as capital and labour will gradually shift from high-pollution, low-efficiency industries toward green, high-value-added sectors, thereby improving resource allocation efficiency. Furthermore, standardized and institutionalized audit oversight helps enhance government governance transparency and policy stability, thereby reducing the uncertainty faced by market entities when making green investments and further stimulating their enthusiasm for participating in green development.

From a regional perspective, the audit has incorporated responsibilities related to resources and the environment into the performance evaluation and accountability system for leading cadres, making ecological performance an important factor in assessing their achievements. This has shifted local governments' previous behaviour, which tended to prioritize short-term economic growth, and prompted them to focus more attention on ecological conservation and green development. Under this constraint mechanism, local governments, seeking to avoid the risk of accountability for environmental issues, will proactively strengthen the enforcement of environmental regulations to guide the adjustment of the regional economic structure toward a green and low-carbon direction. Therefore, this paper proposes the following research hypotheses:

Hypothesis 1. *The NRAALC is positively associated with the enhancement of regional green productivity. (Note: Given empirical inference fragility, this hypothesis is framed as a positive associational proposition rather than a deterministic causal prediction.)*

2.2. Promotion Pressure as a Possible Transmission Pathway

Performance evaluations influence the priorities of officials seeking promotion. Under a GDP-centred evaluation system, economic growth receives particular attention. Adding environmental criteria changes the set of outcomes for which officials are assessed: resource conservation and pollution control become relevant alongside growth. Promotion pressure can therefore provide a possible link between ecological accountability and local policy choices.

The NRAALC adds scrutiny of ecological responsibilities to these performance incentives. If officials expect poor environmental management to affect their evaluations or expose them to subsequent accountability, they may devote more effort to pollution enforcement and environmental supervision. They may also support cleaner industries or place tighter conditions on pollution-intensive investment. Firms responding to such measures could change their environmental spending, technology choices, or use of resources. This sequence provides a theoretical reason to examine promotion pressure as a potential transmission pathway. It does not imply that each link is established by the city-level regressions. We therefore propose the following hypothesis:

Hypothesis 2. *The NRAALC may exert an influence on regional green productivity via changing local officials' promotion pressure. This is a theoretical potential transmission pathway instead of a pre-assured empirically valid causal mechanism.*

2.3. Officials' Tenure as a Moderating Condition

The audit evaluates officials' handling of natural resources and environmental responsibilities during their terms. Many relevant activities, such as ecological restoration and the adoption of cleaner production methods, take time to implement. Officials who remain in office longer may be better placed to maintain these activities and observe their results. A shorter tenure may instead encourage measures with more immediate and visible outcomes. Tenure may therefore condition the relationship between audit exposure and

green productivity. This argument concerns the continuity that longer service may permit, rather than treating tenure length as a direct measure of policy quality.

From the perspective of corporate response, the stability of leading cadres' terms helps enhance the continuity and predictability of local policies, which in turn, reduces institutional uncertainty. In a relatively stable policy environment, enterprises will have greater confidence in researching and developing green technologies and making long-term environmental investments, thereby reinforcing the impact of the audit on green productivity. In contrast, when terms are unstable, policy implementation may fluctuate, and enterprises will face higher levels of uncertainty. This dampens their enthusiasm for green investment and their drive for innovation, thereby diminishing the effectiveness of the policies. In addition, term stability may also strengthen the role of promotion incentives and accountability mechanisms. When cadres have relatively long terms, they have more leeway to build a track record for promotion by continuously improving environmental performance. Thus, the incentive and constraint mechanisms created by the audit become more effective. Conversely, if terms are short, this may lead to a lack of incentives or deviations in behaviour, which in turn could affect the effectiveness of policy transmission.

The continuity that a longer tenure may permit could strengthen the relationship between audit exposure and green productivity. This gives rise to the following moderation hypothesis:

Hypothesis 3. *Officials' tenure may positively moderate the associational relationship between NRAALC implementation and green productivity. We treat this as an exploratory moderating proposition sensitive to inference approaches.*

3. Research Design

3.1. Sample Selection and Data Sources

China entered a preliminary policy and methodological exploration stage for the NRAALC in 2014. The city-level treatment coding used in the empirical tables identifies six treated prefecture-level cities: Huaihua and Loudi from 2015, Zhuzhou and Chenzhou from 2016, and Changsha and Yueyang from 2017. The system moved from pilot implementation to comprehensive rollout in 2018. Accordingly, this paper selects prefecture-level cities in Hunan Province from 2010 to 2017 as the research sample. Data on the NRAALC pilot program were collected from the official website of the Hunan Provincial Audit Office; green productivity data were obtained from the China Urban Statistical Yearbook and the Hunan Provincial Statistical Yearbook; and the remaining variables were sourced from the statistical communiqués of each city. To enhance the reliability of the results, Xiangxi Autonomous Prefecture was excluded due to missing values in key variables. The analysis uses the recorded Pro and Pre scores and the original control-variable values, without additional winsorization or logarithmic transformation. The final balanced panel comprises 13 prefecture-level cities observed annually from 2010 to 2017, yielding 104 city-year observations. Under this treatment coding scheme, merely six cities receive audit intervention, generating only 12 valid post-treatment city-year observations, corresponding to 11.54% of the full sample. Such sparse treated observations create well-known small-sample biases for DID estimation. Moreover, several time-varying control variables included in our specification may themselves be potential post-treatment outcomes affected by NRAALC policy, bringing post-treatment bias risk. We acknowledge this identification limitation explicitly when interpreting all regression outputs, rather than treating control variables as fully exogenous confounders.

3.2. Variable Definitions and Measurement

3.2.1. Dependent Variable

Pro represents our self-constructed composite index for regional green productive capacity. Drawing from the indicator framework proposed by Zhang and Han [29], this index integrates 14 sub-indicators covering technological, low-carbon, and digital development dimensions (see Table 1). The indicators are standardized and then aggregated using weights determined by the entropy-weighting method to construct the composite green productivity index. It is critical to conceptually separate Pro from widely used green total-factor productivity (GTFP). GTFP typically measures input-output production efficiency incorporating bad environmental outputs; by contrast, Pro synthesizes multi-faceted regional development conditions reflecting potential green-development capacity, rather than efficiency-oriented total-factor performance.

Table 1. Evaluation indicator system for green productivity development.

Primary Indicator	Secondary Indicator	Specific Indicator	Unit	Attribute
Technological productivity	Labour force	Number of employees in scientific research, technical services, and geological prospecting	10,000 persons	+
Technological productivity	Means of labour	Number of urban invention patent applications	applications	+
Technological productivity	Means of labour	Robot installation density	%	+
Technological productivity	Objects of labour	Number of artificial intelligence enterprises	enterprises	+
Low-carbon productivity	Labour force	Number of urban employees in water conservancy, environment, and public facility management	10,000 persons	+
Low-carbon productivity	Means of labour	Number of low-carbon technology patent applications	applications	+
Low-carbon productivity	Means of labour	Centralized treatment rate of wastewater treatment plants	%	+
Low-carbon productivity	Objects of labour	Industrial wastewater discharge	10,000 tons	−
Low-carbon productivity	Objects of labour	Industrial sulphur dioxide (SO_2) emissions	tons	−
Digital productivity	Labour force	Number of employees in information transmission, computer services, and software	10,000 persons	+
Digital productivity	Means of labour	Number of digital economy patent applications	applications	+
Digital productivity	Means of labour	Number of Internet broadband access subscribers	subscribers	+
Digital productivity	Means of labour	Number of mobile subscribers	10,000 subscribers	+
Digital productivity	Objects of labour	Total telecommunication business volume	100 million yuan	+

Note: A plus sign denotes an indicator for which a higher value contributes positively to the index; a minus sign denotes an indicator with the reverse direction.

3.2.2. Explanatory Variables

Audit exposure is represented by DID, the product of Treat and Post. The indicator follows the city-level treatment chronology represented in Table 2. Following Zeng et al. [5], Treat equals 1 for the six treated cities and 0 for the seven cities coded as untreated. Post equals 1 from each treated city's recorded treatment-onset year onward and 0 beforehand; it remains 0 for untreated cities throughout the sample period. The treatment year is included in the post-treatment period. Thus, DID_it = Treat_i × Post_it.

The treatment-onset years are 2015 for Huaihua and Loudi, 2016 for Zhuzhou and Chenzhou, and 2017 for Changsha and Yueyang. This coding yields 12 observations with DID = 1. Official pilot information was obtained from the Hunan Provincial Government and Audit Office portals.

Table 2. City-level treatment coding and observation counts, 2010–2017.

City	Treatment Status	Coded Treatment Year	Cohort	Pre-Treatment Obs.	Post-Treatment Obs.	Untreated Obs.
Changsha	Treated	2017	2017	7	1	0
Zhuzhou	Treated	2016	2016	6	2	0
Xiangtan	Untreated	—	—	0	0	8
Hengyang	Untreated	—	—	0	0	8
Shaoyang	Untreated	—	—	0	0	8
Yueyang	Treated	2017	2017	7	1	0
Changde	Untreated	—	—	0	0	8
Zhangjiajie	Untreated	—	—	0	0	8
Yiyang	Untreated	—	—	0	0	8
Chenzhou	Treated	2016	2016	6	2	0
Yongzhou	Untreated	—	—	0	0	8
Huaihua	Treated	2015	2015	5	3	0
Loudi	Treated	2015	2015	5	3	0

Note: This table reports the city-level treatment coding used in the empirical tables. The treatment-onset year is included in the post-treatment period. Pre-treatment observations refer to observations from treated cities before their treatment onset. Untreated observations refer to observations from cities coded as untreated throughout 2010–2017. Under this coding, the sample comprises 104 city-year observations, of which 12 are coded as DID = 1. The symbol "—" indicates not applicable.

Table 2 reports the treatment status, first treatment year, cohort, and numbers of pre- and post-treatment observations for all cities in the estimation sample. The table makes the treatment variation used in the regression models transparent.

3.2.3. Promotion Pressure and Officials' Tenure

Pre represents promotion-related performance pressure through a composite index. Following Tang et al. [16], its components are GDP growth, the environmental-protection share of fiscal expenditure, and unemployment. These indicators capture economic performance, environmental spending, and employment conditions that may matter in officials' evaluations. The index combines the components using entropy weights; Table 3 gives their definitions and directions. The analysis retains the recorded Pre scores. Pre is a proxy for promotion pressure and does not directly observe an official's expectations of promotion.

Table 3. Evaluation indicator system for promotion pressure of leading cadres.

Primary Indicator	Secondary Indicator	Specific Indicator	Unit	Attribute
Promotion pressure of leading cadres	Annual GDP growth	Annual growth rate of city GDP	%	+
Promotion pressure of leading cadres	Proportion of environmental protection expenditure	Environmental protection expenditure in year t/total fiscal expenditure in year t	%	+
Officials' promotion pressure	Urban unemployment rate	Registered urban unemployed persons/(employed persons + registered urban unemployed persons)	%	−

Ten records the number of years an official has served and is used as a proxy for term stability. The moderation regression contains Ten and the DID × Ten interaction in addition to the baseline controls and city and year fixed effects. Ten remains in years and is not mean-centred. The interaction tests whether the estimated audit association varies with length of service; it does not directly measure the continuity of a particular environmental policy.

3.2.4. Control Variables

Per Capita GDP (RGDP)

RGDP is regional GDP divided by the permanent resident population. It controls for differences in economic development that may be associated with investment and technology adoption. The regressions use its recorded level, with no logarithmic transformation.

Urban Innovation (Inn)

Inn is science and technology expenditure divided by regional GDP. It measures the scale of this expenditure relative to the local economy and enters the regressions as a ratio. The variable accounts for differences in public support for innovation that may be related to green productive capacity.

Financial Support (Fin)

Fin is the year-end balance of RMB loans from financial institutions divided by regional GDP. The ratio measures financial support relative to the size of the local economy. Access to finance may affect firms' ability to invest in cleaner technologies, although the variable does not distinguish green lending from other loans.

Level of Service Sector Development (Ser)

Service-sector development (Ser) is measured as the share of value added by the tertiary sector in regional GDP, expressed in percentage points. It enters the baseline regression in levels, without a logarithmic transformation. The service-sector structure may affect resource use and environmental performance, so Ser is included as a control variable.

Leadership Changes (Cha)

This paper constructs a dummy variable to indicate whether there was a change in key local leadership in a given year, with the specific settings shown in Table 4. If there was a change in the municipal party secretary or mayor in a given year, this variable is set to 1; if there was no change, it is set to 0. Leadership changes may lead to policy uncertainty or adjustments in governance style, which in turn may affect green productivity and the effectiveness of the related policy (NRAALC).

Table 4. Variables and definitions.

Variable	Definition	Measurement
Pro	Green productivity development index	Composite index based on technological, low-carbon and digital productivity
DID	NRAALC treatment indicator	Treat × Post; equals 1 from the city's first implementation year onwards
RGDP	Economic development	GDP per capita; entered in levels
Inn	Urban innovation	Science and technology expenditure/GDP; ratio form
Fin	Financial support	Year-end financial-institution loans/GDP
Ser	Service-sector development	Tertiary-sector value added/GDP × 100; entered in levels
Cha	Leadership changes	Equals 1 if the party secretary or mayor changed
Pre	Promotion pressure	Entropy-weighted index of GDP growth, environmental expenditure share and unemployment rate
Ten	Officials' tenure (proxy for term stability)	Recorded length of service, in years; uncentred

3.3. Construction of the Research Model

The reference model is a multi-period difference-in-differences regression with city and year fixed effects. The control vector X contains RGDP, Inn, Fin, Cha, and Ser. RGDP and Ser enter in levels; Inn and Fin enter as ratios. The estimates use the recorded indices and control values without additional winsorization or logarithmic transformation. Conventional homoskedastic OLS standard errors are reported for comparison with the reference specification. For the main policy coefficient, we also report city-clustered standard errors and null-imposed wild-cluster bootstrap inference [30]. With 13 city clusters, the bootstrap uses Rademacher weights and enumerates all 8.192 sign patterns; the requested replication count is 9999, and the random seed is 20261005. The small-cluster results govern the strength of the main inference. The baseline and promotion-pressure equations exclude Ten; the moderation equation includes both Ten and DID × Ten. Appendix A reports inference and specification sensitivities.

$$Pro_{i,t} = \alpha_0 + \alpha_1 DID_{i,t} + \alpha_2 X_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t} \quad (1)$$

In Equation (1), $Pro_{i,t}$ represents green productivity; $DID_{i,t}$ is a dummy variable that takes the value 1 if region i implements the NRAALC in year t, and 0 otherwise; $X_{i,t}$ denotes the control variables, and δ_i and μ_t represent the region and time fixed effects, respectively; $\varepsilon_{i,t}$ denotes the random disturbance term; α_0 and α_2 denote the constant term and the coefficients of the control variables, respectively; α_1 is the estimated coefficient of the dummy variable.

Following Jiang [31], we distinguish an audit-to-pressure association from a complete mediation pathway. Equation (2) estimates the coefficient a of DID in the Pre regression. Equation (3) includes both DID and Pre and estimates the remaining DID coefficient and the coefficient b of Pre. All equations contain the same five controls and city and year fixed effects. We estimate the indirect association as a × b and obtain its percentile confidence interval by resampling cities, assigning a new city identifier to each sampled copy, and re-estimating both equations. Of 9999 requested replications with seed 20261005, 9997 return valid indirect-effect estimates. These regressions do not establish the absence of unobserved mediator–outcome confounding; the mediation interpretation therefore requires more than statistical significance alone.

$$Pre_{i,t} = \beta_0 + \beta_1 DID_{i,t} + \beta_2 X_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t} \quad (2)$$

$$Pro_{i,t} = \gamma_0 + \gamma_1 DID_{i,t} + \gamma_2 Pre_{i,t} + \gamma_3 X_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t} \quad (3)$$

The tenure specification in Equation (4) adds Ten and DID × Ten to the baseline controls and the city and year effects. Because Ten is uncentred, the audit effect at T years of service is $\theta_1 + \theta_3 T$. The coefficient covariance matrix provides the conventional OLS standard errors and 95% confidence intervals for these effects. We evaluate them at two, three, four, and five years, covering the tenure values observed in treated city-years.

$$Pro_{i,t} = \theta_0 + \theta_1 DID_{i,t} + \theta_2 Ten_{i,t} + \theta_3 (DID_{i,t} \times Ten_{i,t}) + \theta_4 X_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t} \quad (4)$$

3.4. Descriptive Statistical Analysis

The sample summarized in Table 5 contains 104 city-year observations. Twelve are treated observations, giving DID a mean of 0.1154. The mean Pro score is 0.1061, and its standard deviation is 0.1357. Table 5 also reports the minimum, maximum, and median to describe the distribution beyond these averages.

Table 5. Descriptive statistics of the main variables.

Variable	N	Mean	Std. Dev.	Min	Max	Median
DID	104	0.1154	0.3210	0.0000	1.0000	0.0000
Pro	104	0.1061	0.1357	0.0250	0.9495	0.0667
RGDP	104	385.9718	234.8177	104.6800	1312.0700	325.8450
Inn	104	0.0017	0.0020	0.0006	0.0200	0.0013
Fin	104	0.5888	0.2775	0.2638	1.5652	0.5178
Cha	104	0.2788	0.4506	0.0000	1.0000	0.0000
Ser	104	57.6328	9.8922	36.5000	78.4900	56.4700
Pre	104	0.5353	0.1416	0.2688	0.9689	0.5089
Ten	104	3.7885	1.6171	2.0000	8.0000	3.0000

Note: RGDP and Ser are measured in levels. Inn and Fin are measured as ratios. Pre is the promotion-pressure index, and Ten measures officials' recorded length of service in years. The descriptive statistics and regressions use the same original variable values.

RGDP and Ser are reported in levels, with mean values of 385.9718 and 57.6328, respectively. Ser is expressed in percentage points. Inn and Fin are measured as ratios, with mean values of 0.0017 and 0.5888, respectively. Pre has a mean of 0.5353, while Ten has a mean of 3.7885 years.

Table 6 reports pairwise Pearson correlation coefficients among the main variables. Variance inflation factors are reported separately in Table 7 because pairwise correlations do not constitute a formal test of multicollinearity.

Table 6. Correlation coefficients among key variables.

	DID	Pro	RGDP	Inn	Fin	Cha	Ser	Pre	Ten
DID	1.0000	0.1752	0.1663	0.0563	0.1089	−0.0232	0.0312	0.4443	−0.1582
Pro	0.1752	1.0000	0.8839	0.1754	0.7684	0.0450	−0.1769	0.3347	0.1166
RGDP	0.1663	0.8839	1.0000	0.2605	0.6789	0.0377	−0.3866	0.5277	0.1473
Inn	0.0563	0.1754	0.2605	1.0000	0.1276	−0.0566	−0.2784	0.1520	0.1399
Fin	0.1089	0.7684	0.6789	0.1276	1.0000	−0.0297	0.0951	0.2606	0.1501
Cha	−0.0232	0.0450	0.0377	−0.0566	−0.0297	1.0000	0.0037	0.0147	0.3349
Ser	0.0312	−0.1769	−0.3866	−0.2784	0.0951	0.0037	1.0000	−0.2716	−0.1282
Pre	0.4443	0.3347	0.5277	0.1520	0.2606	0.0147	−0.2716	1.0000	0.0135
Ten	−0.1582	0.1166	0.1473	0.1399	0.1501	0.3349	−0.1282	0.0135	1.0000

Table 7. Variance inflation factor (VIF) test.

Variable	VIF	1/VIF
RGDP	3.150	0.317
Fin	2.617	0.382
Ser	1.729	0.578
Inn	1.125	0.889
DID	1.048	0.954
Cha	1.018	0.983
Mean VIF	1.781	—

The correlation between Pro and RGDP is 0.8839, and that between Pro and Fin is 0.7684. DID is positively correlated with Pro (0.1752) and Pre (0.4443). The correlation between DID and Ten is −0.1582. These are unconditional associations and should not be interpreted as causal relationships.

Some correlations are relatively high, particularly those between Pro and Fin and between Pro and RGDP. However, pairwise correlations alone do not provide a formal test

of multicollinearity. We therefore conduct a variance inflation factor (VIF) test in the next step. The VIF results provide a more appropriate assessment of multicollinearity among the explanatory variables included in the regression models.

Table 7 checks collinearity among DID, RGDP, Inn, Fin, Cha, and Ser. City and year indicators are excluded from this diagnostic. The largest VIF is 3.150 for RGDP, and the mean is 1.781; these values do not suggest severe collinearity among the substantive baseline regressors. Pre and Ten are assessed in the subsequent pathway and moderation analyses. A low VIF does not establish the assumptions needed for causal interpretation, address possible post-treatment bias, or resolve the precision problems of a small panel.

4. Empirical Analysis

4.1. Regression Results and Analysis

In Table 8, the estimated coefficient on DID is 0.038787. Its ordinary OLS standard error is 0.014188, yielding t = 2.7339 and $p = 0.0077$. Thus, the baseline estimate is positive and significant at the 1% level under conventional OLS inference, consistent with the direction proposed in Hypothesis 1. RGDP and Fin also have positive coefficients significant at 1%. The Ser coefficient is negative with $p = 0.0591$, while the Inn and Cha coefficients are not statistically significant.

Table 8. Baseline regression results.

Variable	Coefficient	Ordinary SE	t-Statistic
DID	0.038787 ***	0.014188	2.7339
RGDP	0.000995 ***	0.000075	13.1800
Inn	0.504807	2.051847	0.2460
Fin	0.531125 ***	0.122793	4.3254
Cha	0.009351	0.008246	1.1340
Ser	−0.002239 *	0.001169	−1.9156
Constant	−1.017405 ***	0.191947	−5.3005
City fixed effects	Yes		
Year fixed effects	Yes		
Controls	Yes		
Observations	104		
R-squared	0.9556		

Note: Pro is the outcome variable. Controls are RGDP, Inn, Fin, Cha, and Ser, with city and year fixed effects. RGDP and Ser are in levels. The table reports conventional homoskedastic OLS standard errors and the corresponding t-statistics for 104 observations across 13 cities. R-squared refers to the overall OLS fit. The significance symbols * and *** indicate the 10%, and 1% levels, respectively.

Appendix A, Table A1 reports inference for the same DID coefficient and the same estimation sample. The city-clustered standard error is 0.025790 ($p = 0.1585$). The wild-cluster bootstrap gives $p = 0.1102$ and a 95% confidence interval of [−0.00514, 0.12030]. Thus, the conventional significance in Table 8 does not persist under these procedures. This difference concerns uncertainty around the same point estimate, not a change in the estimated coefficient.

The baseline point estimate is consistent with the positive direction proposed in Hypothesis 1, but the inference comparisons do not establish a reliably positive policy effect. The results are reported as a positive conditional association with substantial uncertainty. Sample size alone cannot explain away the non-rejection of a zero effect.

The control-variable results describe conditional associations in the baseline model. The positive RGDP and Fin coefficients and negative Ser coefficient should be interpreted within the definitions and units used in Tables 4 and 5.

4.2. Identification Diagnostics and Sensitivity Analyses

4.2.1. Parallel Trends Test

Figure 1 reports the two-way fixed-effects event-study coefficients with the year before treatment (event time −1) omitted. The regression contains the five baseline controls and city and year fixed effects. The displayed 95% intervals use conventional OLS standard errors and match the underlying regression estimates. The joint test of the six pre-treatment coefficients gives F(6, 70) = 1.09 and $p = 0.3766$. This fails to reject joint equality to zero; it does not establish that the parallel-trends assumption is true.

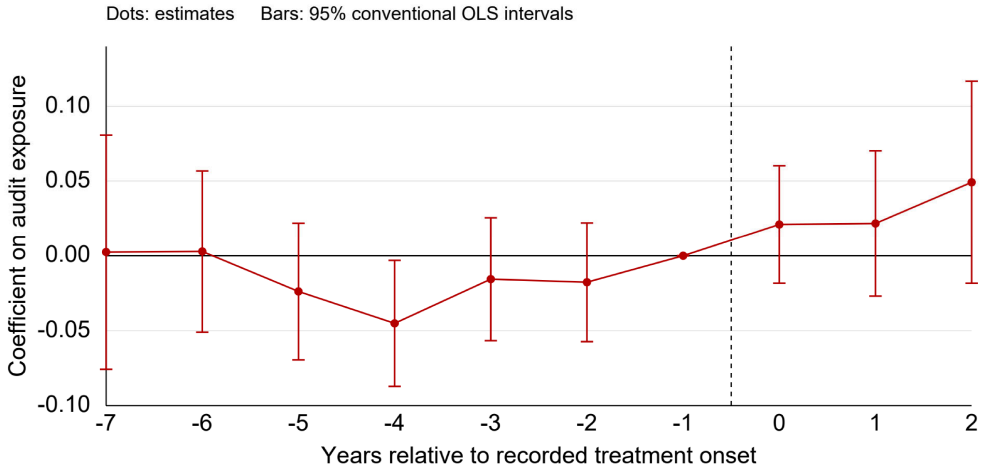


Figure 1. Event-study estimates with conventional OLS 95% confidence intervals. Red points represent the estimated coefficients, and the connecting red line illustrates their pattern over event time. Red vertical error bars indicate 95% confidence intervals. The vertical dashed line separates the pre-treatment and post-treatment periods; event time 0 denotes the recorded treatment-onset year. Event time −1 is the reference period.

Notably, the pre-treatment coefficient at event time −4 is individually statistically significant (−0.0451; $p = 0.036$), which raises concerns over the plausibility of the parallel-trends assumption. None of the three post-treatment event-study coefficients reaches 5% statistical significance (p = 0.292, 0.378, 0.151). Hence, Figure 1 provides no solid evidence for a delayed policy-driven rise in green productivity. Only the 2015 treated cohort supplies observations for event time +2, amplifying estimation uncertainty given our limited sample size. Further cohort-specific diagnostic tests in Appendix A reinforce our concern regarding pre-treatment imbalance across treated and control groups, which weakens causal identification credibility of our staggered DIDdesign.

4.2.2. Sensitivity to Earlier Treatment Dates

Table 9 shifts each treated city's recorded onset one, two, or three years earlier while retaining the full sample. The estimated coefficients are 0.0362, 0.0281, and 0.0292, with conventional OLS p-values of 0.006, 0.031, and 0.036. These estimates are significant under this inference method. Because the shifted indicators also equal one during the actual post-treatment periods, this exercise can capture the actual treatment association as well as earlier differences. It is a timing-sensitivity exercise, not a clean pre-treatment placebo, and cannot be used to claim that a placebo test has been passed.

Table 9. Full-sample regressions with treatment onset shifted earlier.

Variable	1 Year Earlier	2 Years Earlier	3 Years Earlier
Shifted DID	0.036161 ***	0.028131 **	0.029170 **
	(2.8266)	(2.1919)	(2.1329)
RGDP	0.000968 ***	0.000955 ***	0.000945 ***
	(12.8170)	(12.2108)	(11.8984)
Inn	0.452813	0.537491	0.715502
	(0.2215)	(0.2574)	(0.3405)
Fin	0.514485 ***	0.491506 ***	0.507819 ***
	(4.2337)	(3.9876)	(4.0903)
Cha	0.007081	0.009072	0.010431
	(0.8633)	(1.0828)	(1.2323)
Ser	−0.002236 *	−0.002077 *	−0.002063 *
	(−1.9188)	(−1.7501)	(−1.7353)
Constant	−0.973718 ***	−0.940675 ***	−0.957376 ***
	(−5.1324)	(−4.8740)	(−4.9495)
City fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	104	104	104
R-squared	0.9558	0.9541	0.9540

Note: Parentheses report conventional OLS t-statistics. The actual post-treatment observations remain in each regression. *, **, and *** denote $p < 0.10$, $p < 0.05$, and $p < 0.01$.

4.2.3. Treatment Timing and the Goodman–Bacon Decomposition

Table 10 reports the covariate-adjusted Goodman–Bacon decomposition [32] for the same 104 observations, outcome, treatment indicator, five controls, and city and year fixed effects as Table 8. The weights sum to one. Multiplying each component estimate by its weight and summing gives 0.0387872, reproducing the baseline DID coefficient up to rounding. The within component is part of the covariate-adjusted decomposition and should not be interpreted as a separate policy effect. Most weight is attached to treated-versus-never-treated comparisons, but this allocation alone does not establish robustness to treatment-effect heterogeneity. We further implement the regression-adjusted Callaway-Sant'Anna staggered DID estimator [33] with never-treated cities as the control group and identical five baseline covariates. The resulting aggregate ATT is 0.001893 ($p = 0.893$; Appendix A Table A2), close to zero and statistically indistinguishable from zero. This stands in sharp contrast with the positive point estimate from the conventional two-way fixed-effects TWFE model. Such divergence across different DID estimators demonstrates high estimation instability within our sample, implying that the baseline TWFE result cannot be regarded as robust policy evidence.

Table 10. Covariate-adjusted decomposition of the baseline TWFE coefficient.

Component	Estimate	Weight
Timing-group comparisons	−0.0619608355	0.1889504389
Treated versus never-treated	0.0500718310	0.7859255682
Within component	0.4434804320	0.0251239930
Weighted sum/total weight	0.0387872	1.0000000

Note: The weighted sum is (−0.0619608355 × 0.1889504389) + (0.0500718310 × 0.7859255682) + (0.4434804320 × 0.0251239930) = 0.0387872. The final row reports the weighted coefficient and the sum of weights. Component estimates and weights are rounded.

4.2.4. Adding Additional Control Variables

Table 11 adds regional openness (Open) to the baseline model, retaining the same observations, controls, and city and year fixed effects. The DID coefficient is 0.038586

(ordinary standard error = 0.014276; $p = 0.0085$), significant at the 1% level. This result is directionally consistent with Table 8. This comparison concerns the point estimate and conventional inference; it does not resolve the small-cluster inference limitations reported for the main model.

Table 11. Regression results with additional control variables.

Variable	(1) Pro	(2) Pro
DID	0.038787 ***	0.038586 ***
	(2.7339)	(2.7029)
RGDP	0.000995 ***	0.000999 ***
	(13.1800)	(13.0434)
Inn	0.504807	0.481781
	(0.2460)	(0.2334)
Fin	0.531125 ***	0.538841 ***
	(4.3254)	(4.3064)
Cha	0.009351	0.009470
	(1.1340)	(1.1413)
Ser	−0.002239 *	−0.002220 *
	(−1.9156)	(−1.8875)
Open		0.004150
		(0.3805)
Constant	−1.017405 ***	−1.046772 ***
	(−5.3005)	(−5.0357)
City fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	104	104
R-squared	0.9556	0.9557
Adjusted R-squared	0.9413	0.9407

Note: Parentheses report t-statistics based on conventional OLS standard errors. * and *** denote $p < 0.10$ and $p < 0.01$, respectively.

4.3. Promotion Pressure and Tenure

The audit may change the performance incentives facing local officials. We examine this proposed pathway using a promotion-pressure proxy, and then assess tenure as a separate moderating condition. The analyses are exploratory and use the definitions in Section 3.2.3.

Table 12 reports the baseline equation, the Pre equation, and the outcome equation containing both DID and Pre. The audit-to-pressure coefficient is positive (a = 0.107000; conventional OLS $p = 0.0052$). Once DID is included in the outcome equation, the coefficient of Pre is −0.010144 ($p = 0.8160$), while the remaining DID coefficient is 0.039873 ($p = 0.0096$ under conventional OLS inference). The estimated indirect association is −0.001085, with a city-resampled bootstrap percentile 95% confidence interval of [−0.010493, 0.005476]. The interval includes zero. These results do not support Hypothesis 2 as a complete mediation mechanism. Promotion pressure is retained as a theoretically possible transmission pathway, not a confirmed mediator.

Tenure is examined separately from promotion pressure. Longer service may give an official more time to implement environmental measures and observe their outcomes. The moderation specification tests whether the estimated audit association varies with tenure, rather than assuming that the audit changes tenure itself.

Table 13 reports the complete tenure moderation model. The DID × Ten coefficient is 0.027681 (ordinary standard error = 0.009509; $p = 0.0047$). The positive interaction is consistent with Hypothesis 3 under conventional OLS inference. With uncentred tenure, the DID main coefficient refers to tenure equal to zero, outside the observed range. The

marginal effects below therefore provide the relevant interpretation of the interaction. The interaction is not significant with city-clustered standard errors (SE = 0.023942; $p = 0.2701$). Accordingly, the evidence for Hypothesis 3 is sensitive to the inference method and does not establish robust moderation.

Table 12. Complete promotion-pressure regressions.

Variable	(1) Pro	(2) Pre	(3) Pro
DID	0.038787 ***	0.107000 ***	0.039873 ***
	(2.7339)	(2.8759)	(2.6560)
Pre	—	—	−0.010144
			(−0.2335)
RGDP	0.000995 ***	−0.000328	0.000992 ***
	(13.1800)	(−1.6568)	(12.8323)
Inn	0.504807	−9.162881 *	0.411859
	(0.2460)	(−1.7029)	(0.1959)
Fin	0.531125 ***	−0.135466	0.529751 ***
	(4.3254)	(−0.4207)	(4.2831)
Cha	0.009351	−0.006814	0.009282
	(1.1340)	(−0.3151)	(1.1181)
Ser	−0.002239 *	0.008765 ***	−0.002150 *
	(−1.9156)	(2.8600)	(−1.7394)
Constant	−1.017405 ***	0.550987	−1.011816 ***
	(−5.3005)	(1.0946)	(−5.1995)
City fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	104	104	104
R-squared	0.9556	0.7192	0.9556

Note: Parentheses contain conventional OLS t-statistics. All columns contain RGDP, Inn, Fin, Cha, and Ser. * and *** denote $p < 0.10$ and $p < 0.01$. The indirect association is −0.001085; its city-resampled bootstrap percentile 95% interval is [−0.010493, 0.005476], based on 9997 valid replications out of 9999 requested. The indirect-effect interval, rather than the significance of the audit-to-pressure coefficient alone, informs the mediation conclusion.

Table 13. Moderating effect test.

Variable	Coefficient	Ordinary SE
DID	−0.049154	0.033147
Ten	−0.001651	0.002684
DID × Ten	0.027681 ***	0.009509
RGDP	0.000944 ***	0.000075
Inn	0.359807	1.975980
Fin	0.488191 ***	0.120156
Cha	0.002288	0.009002
Ser	−0.001999 *	0.001141
Constant	−0.924903 ***	0.187668
City fixed effects	Yes	
Year fixed effects	Yes	
Observations	104	
R-squared	0.9600	
Adjusted R-squared	0.9459	

Note: Pro is the dependent variable. Ten is uncentred and measured in years. * and *** denote $p < 0.10$ and $p < 0.01$. The marginal audit effects are 2 years: 0.0062 (95% CI [−0.0289, 0.0413]; $p = 0.7258$); 3 years: 0.0339 (95% CI [0.0065, 0.0612]; $p = 0.0159$); 4 years: 0.0616 (95% CI [0.0303, 0.0929]; $p < 0.001$); 5 years: 0.0893 (95% CI [0.0453, 0.1332]; $p < 0.001$). These estimates represent discrete changes from DID = 0 to DID = 1 at the specified tenure values. They do not identify a tenure threshold. All reported marginal-effect intervals use conventional OLS inference; Appendix A Table A3 presents them for comparison. They should not be interpreted as city-cluster-robust effects.

4.4. Heterogeneity Analysis

Table 14 compares the Changsha–Zhuzhou–Xiangtan (CZT) region with other cities using separate regressions with the baseline controls and city and year fixed effects. The DID coefficients are 0.083037 (p = 0.2120) and −0.002533 (p = 0.5260), respectively. Neither estimate is statistically significant under conventional OLS inference. The subsamples contain 24 and 80 observations. These results do not establish a difference in the audit effect between the two regions. Parentheses report t-statistics; *, **, and *** denote $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

Table 14. Heterogeneity analysis.

Variable	CZT	Other Cities
DID	0.083037	−0.002533
	(1.3565)	(−0.6381)
RGDP	0.001463 ***	0.000223 ***
	(5.7611)	(3.5854)
Inn	−2.219364	0.821514
	(−0.5254)	(0.2627)
Fin	1.157244 **	0.068369 *
	(2.8095)	(1.9174)
Cha	0.006522	−0.000926
	(0.1543)	(−0.4314)
Ser	−0.001484	0.000477
	(−0.4706)	(0.9236)
Constant	−2.223742 ***	−0.028351
	(−3.7818)	(−1.0280)
City fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	24	80
R-squared	0.9810	0.9457
Adjusted R-squared	0.9455	0.9247

Note: Parentheses report t-statistics; *, **, and *** denote $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

5. Conclusions and Recommendations

The reference TWFE regression yields a positive audit coefficient of 0.0388, significant under conventional OLS inference. Adding regional openness leaves that point estimate largely unchanged. However, city-clustered and wild-cluster bootstrap inference do not reject a zero main effect at the 5% level, and the cohort-specific estimate is not statistically significant. The promotion-pressure proxy is positively associated with audit exposure, but the complete mediation regression and bootstrap interval do not support an indirect effect. The positive tenure interaction and its marginal effects are significant only under the reported conventional inference; clustered inference does not confirm the interaction. Neither regional subgroup estimate is statistically significant. Collectively, our empirical results only reveal tentative correlational patterns derived from a small provincial pilot-period panel dataset. Due to multiple identification threats, including fragile statistical inference, pre-trend concerns, estimator instability, sparse treated observations, potential post-treatment bias for time-varying controls, and unsupported mediation/moderation effects, we cannot draw robust causal inferences that NRAALC raises regional green

productivity, nor can we validate promotion-pressure-related mediating or tenure-related moderating mechanisms. All findings should be interpreted conditionally within Hunan's 2010–2017 pilot context.

This study connects ecological accountability with a measure of green productive capacity that covers technological, low-carbon, and digital development. It also separates two questions: whether promotion pressure may transmit the audit association, and whether that association varies with officials' tenure. The findings are specific to Hunan's pilot period. The 2010–2017 window retains variation between treated and untreated cities; after the nationwide rollout in 2018, the same untreated comparison is no longer available for a direct extension of this design. The estimates therefore describe the earlier institutional setting. Assessing current effects would require later data and an identification strategy appropriate to nationwide implementation.

The institutional discussion identifies clear ecological assessment criteria, follow-up on audit findings, and continuity of environmental responsibilities as matters for policy consideration. These are implications of the governance framework, not interventions whose effectiveness has been demonstrated by the present estimates. The subgroup regressions do not establish that the CZT region offers a superior model for other cities. Stronger policy recommendations require additional evidence on audit exposure, implementation, and outcomes, including a design suited to nationwide implementation after 2018.

6. Limitations and Directions for Future Research

The analysis uses a provincial city-level panel to examine audit exposure, green productivity, promotion pressure, and tenure. The following limitations define the scope of the findings and the evidence needed to extend them.

6.1. Sample Size, Treatment Support, and Generalizability

The balanced panel contains 13 cities and 104 observations, but only six cities and 12 city-year observations are treated. This limited support constrains precision and generalizability. Conventional significance is not robust to city clustering or the wild-cluster bootstrap; the latter is a useful diagnostic, not a guarantee of reliable inference with sparse treatment. The estimates are also sensitive to data transformation and to the inclusion of Changsha (Appendix A, Table A4). These limitations prevent a strong causal claim. Missing key variables for Xiangxi also restrict geographical coverage. Non-significance in this sample should not be interpreted as evidence that the true policy effect is zero.

6.2. Dynamic Effects over a Longer Time Horizon

The data cover 2010–2017, with coded treatment beginning between 2015 and 2017. This leaves only a short period in which to observe outcomes after the audit starts. Changes in technology, industrial composition, and environmental quality may take longer to appear. Later studies could examine whether audit effects persist, strengthen, or weaken over time, using a design that accounts for the subsequent nationwide rollout.

6.3. In-Depth Exploration of Micro-Level Transmission Mechanisms

The promotion-pressure index and recorded tenure measure are proxies; they do not trace officials' expectations or firms' responses to audit findings. The indirect-effect interval includes zero, and the city-level regressions cannot rule out mediator–outcome confounding. Contemporaneous controls may themselves respond to policy, which also complicates causal interpretation. Future work should link official appointment records, audit implementation documents, environmental investment, innovation, and emissions, and distinguish city-level exposure from audits implemented in constituent counties. The com-

posite Pro index measures green productive capacity rather than input–output productivity efficiency, and its construction should be documented alongside the underlying indicators.

Author Contributions: Methodology, Y.Q.; software, Y.Q.; formal analysis, Y.Q.; resources, Y.Z. and W.Z.; data curation, Y.Q.; writing—original draft, Y.Q.; writing—review and editing, Y.Q. and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: Support for this research was provided by the National Social Science Foundation of China (General Program), grant 25BJY131, and the Department of Natural Resources of Hunan Province, China, grant HBZ20260109.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

Appendix A. Additional Inference and Sensitivity Results

Table A1. Inference for the baseline DID coefficient.

Inference Method	DID	*p*-Value	95% Interval
Conventional OLS	0.038787	0.0077	[0.010542, 0.067033]
City-clustered, 13 cities	0.038787	0.1585	[−0.017405, 0.094979]
Wild-cluster bootstrap	0.038787	0.1102	[−0.00514, 0.12030]

Note: All rows use 104 observations, the five baseline controls, and city and year fixed effects. City-clustered inference uses 12 degrees of freedom. The null-imposed wild-cluster bootstrap uses Rademacher weights at the city level and enumerates all 8192 sign patterns. The reported interval is obtained by test inversion. A bootstrap with few treated clusters does not eliminate all small-sample uncertainty.

Table A2. Cohort-specific estimates with never-treated controls.

Specification	Aggregate ATT	SE	*p*-Value
No covariates	0.014953	0.016949	0.378
Five baseline covariates	0.001893	0.014037	0.893

Note: These estimates use the regression-adjustment implementation of Callaway and Sant'Anna [33], with never-treated cities as controls and treatment cohorts defined by first recorded audit exposure. The adjusted ATT has a reported 95% interval of [−0.025619, 0.029405]. The adjusted pre-treatment joint diagnostic returns $\chi^2(15) = 450.2573$ (reported $p < 0.001$); at event time −3, the estimate is 0.016580 ($p = 0.009$). These diagnostics raise concerns about pre-treatment comparability. They should also be interpreted cautiously: each treated cohort contains only two cities, and a 15-restriction joint statistic in such a small panel requires attention to the covariance specification and rank. The output is not treated as a definitive small-cluster-robust test. The cohort-specific ATT and TWFE coefficient use different aggregation weights and need not coincide.

Table A3. Conditional audit associations at selected tenure values.

Tenure (Years)	DID Association	*p*-Value	95% Interval
2	0.006209	0.7258	[−0.028926, 0.041343]
3	0.033890	0.0159	[0.006533, 0.061247]
4	0.061571	<0.001	[0.0303, 0.0929]
5	0.089253	<0.001	[0.045340, 0.133165]

Note: Estimates are the DID = 0 to DID = 1 contrast at each tenure value in Equation (4); intervals and *p*-values use conventional OLS inference. Ten is not centred. The interaction itself has a city-clustered *p*-value of 0.2701, so the conditional associations do not establish a robust tenure effect or a three-year threshold.

Table A4. Sensitivity to variable treatment and one-city exclusion.

Specification	DID	OLS *p*-Value	N
Recorded values; no added winsorization or logs	0.038787	0.0077	104
Pooled 1st/99th percentile winsorization; levels	0.020923	0.085	104
Same winsorization; log RGDP and Inn	−0.012624	0.544	104
Recorded values; exclude Changsha	0.004070	0.3286	96

Note: All specifications retain the five controls and city and year fixed effects. Winsorization applies to Pro, RGDP, Inn, Fin, and Ser across the pooled sample; DID and the binary Cha indicator are not winsorized. In the third row, logarithms of RGDP and Inn are taken after winsorization; Fin and Ser remain in levels. The Changsha exclusion is an influence diagnostic, not the preferred sample. The baseline retains all 13 cities. The baseline and alternative specifications are shown together to make sensitivity to transformation and sample composition explicit.

References

1. Cui, H.; Huang, L. The Logic of Natural Resource Asset Accounting and the Design of Its Audit Procedures for the Natural Resource Accountability Audit for Leading Cadres. *Financ. Account. Mon.* **2023**, *44*, 97–102.
2. Liu, H.; Li, Z. Research on the Innovation of the Evaluation System for the Natural Resource Accountability Audit for Leading Cadres. *Environ. Sci. Manag.* **2022**, *10*, 5–9, 98.
3. Zhang, Y.; Wang, C.; Zhang, W. Evaluation of the Natural Resource Accountability Audit for Leading Cadres Based on the PSR Model and RS Method: A Case Study of Huaihua City, Hunan Province. *J. Xinjiang Univ. Financ. Econ.* **2021**, *4*, 51–59.
4. Zhang, Y.; Zhang, W.; Cheng, T. Natural Resource Accountability Audit for Leading Cadres and Environmental Performance Improvement: From the Perspective of Main Functional Zones. *J. Cent. South Univ. For. Technol. (Soc. Sci.)* **2024**, *5*, 58–68.
5. Zeng, C.; Liu, L.; Li, J.; Li, L. Environmental Assessment and Corporate Green Innovation: A Quasi-Natural Experiment Based on the Pilot of the Natural Resource Accountability Audit for Leading Cadres. *Account. Res.* **2022**, *3*, 107–122.
6. Li, Z.; Guo, L. Study on the Impact of National Audit on Low-Carbon Transformation Development: Evidence from the Pilot of Audit for Outgoing Officials on Natural Resource Assets. *J. Nanjing Audit. Univ.* **2024**, *21*, 1–13.
7. Huang, R.; Xie, X. Can the Natural Resource Accountability Audit for Leading Cadres Improve the Environmental Governance Execution Capacity of Local Governments? *Audit. Econ. Res.* **2022**, *37*, 9–20.
8. Ma, Z.; Yin, F. Natural Resources Assets Departure Audit, Environmental Attention and Environmental Governance Efficiency. *J. Audit Econ.* **2023**, *38*, 1–11.
9. Ma, Z.; Ren, L.; Xu, J.; Li, X. Analysis of the Impact of the Natural Resource Accountability Audit for Leading Cadres on the Water Resource Environment. *Financ. Account. Mon.* **2022**, *6*, 88–95.
10. Zhu, Z.; Cai, Z.; Fang, Y. Natural Resource Accountability Audit for Leading Cadres and Urban Green Total Factor Productivity. *J. Nanjing Audit. Univ.* **2025**, *22*, 32–42.
11. Wang, B.; Tan, B.; Tang, K. Natural Resource Accountability Audit for Leading Cadres and Corporate Green Total Factor Productivity. *Financ. Res.* **2024**, *4*, 49–62.
12. Liu, Y. The Impact of the Natural Resource Accountability Audit for Leading Cadres on Green Total Factor Productivity: Evidence from 110 Prefecture-Level Cities in the Yangtze River Economic Belt. *Sci. Technol. Ind.* **2023**, *5*, 108–112.
13. Zhang, Y.; Liu, Y.; Peng, J. The "Dual Carbon" Governance Effect of the Natural Resource Accountability Audit for Leading Cadres. *J. Guangdong Univ. Financ. Econ.* **2026**, *41*, 129–145.
14. Han, F.; Zhao, X.; Sun, P. Natural Resource Accountability Audit and Corporate Green Innovation. *J. Environ. Econ.* **2024**, *4*, 165–186.
15. Sun, Y.; Guo, R. Natural Resource Accountability Audit for Leading Cadres and Corporate Green Innovation Efficiency. *Friends Account.* **2026**, 128–135.
16. Tang, Y.; Cui, Y.; Ma, W.; Wang, X. Research on the Impact and Mechanism of the Natural Resource Accountability Audit for Leading Cadres on Corporate Green Innovation. *Audit. Res.* **2025**, *5*, 15–26.
17. Xu, H.; Lian, R. How Does Environmental Assessment Affect Land Resource Misallocation? Empirical Evidence from the Pilot of the Natural Resource Accountability Audit for Leading Cadres. *China Land Sci.* **2024**, *38*, 47–58.
18. Kim, S.B. A Study of the Bid Protest System of the US Government Accountability Office. *Public Law* **2020**, *48*, 27–50. [CrossRef]
19. Rayamajhee, V.; March, R.J. Does economic liberalization increase government accountability? *J. Econ. Behav. Organ.* **2025**, *237*, 107143. [CrossRef]
20. Whiteley, P.; Kölln, A.K. How do different sources of partisanship influence government accountability in Europe? *Int. Political Sci. Rev.* **2019**, *40*, 502–517. [CrossRef]
21. Liu, C. Developing Green Productivity to Build Green Competitive Advantages: Theoretical Logic, Practical Requirements, and Paths. *Product. Res.* **2026**, *2*, 161.

22. Sun, G. The Impact of Green Productivity on Urban Low-Carbon High-Quality Development: Mechanism Tests Based on Industrial Structure Upgrading and Financial Innovation. *Commer. Res.* **2026**, *1*, 85–94.
23. Dong, H. The Theoretical Connotation, Value Implications, and Development Paths of Green Productivity. *Jiangnan Forum* **2026**, *3*, 38–43.
24. Zhou, Y.; Wu, X. The Value Implications, Practical Difficulties, and Implementation Paths of Cultivating and Developing Green Productivity. *Res. Gov. Mod.* **2025**, *41*, 87–93.
25. Tan, B.; Zhang, N. When carbon policy meets sulfur policy: How environmental policy portfolios enhance green productivity in Chinese cities. *Energy Econ.* **2026**, *154*, 109139. [CrossRef]
26. Salman, M. Green Productivity and Energy Transition under China's Belt and Road Policy: A Geopolitical Perspective. *Environ. Dev. Sustain.* **2026**, *28*, 10559–10590. [CrossRef]
27. Chen, C.; Lan, Q.; Gao, M.; Sun, Y. Green Total Factor Productivity Growth and Its Determinants in China's Industrial Economy. *Sustainability* **2018**, *10*, 1052. [CrossRef]
28. Mushtaq, B.; Halog, A.; Hussain, S. Exploring the Synergies between Industrial Green TFP and Circular Economy: A Systematic Review. *Circ. Econ. Sustain.* **2025**, *5*, 6991–7022. [CrossRef]
29. Zhang, J.; Han, D. Research on the Effect of New Energy Industry Agglomeration on Green Productivity. *Ecol. Econ.* **2025**, *41*, 77–84, 104.
30. Roodman, D.; Nielsen, M.Ø.; MacKinnon, J.G.; Webb, M.D. Fast and wild: Bootstrap inference in Stata using boottest. *Stata J.* **2019**, *19*, 4–60. [CrossRef]
31. Jiang, T. Mediation Effects and Moderation Effects in Causal Inference Empirical Research. *China Ind. Econ.* **2022**, *5*, 100–120.
32. Goodman-Bacon, A. Difference-in-differences with variation in treatment timing. *J. Econom.* **2021**, *225*, 254–277. [CrossRef]
33. Callaway, B.; Sant'Anna, P.H.C. Difference-in-Differences with multiple time periods. *J. Econom.* **2021**, *225*, 200–230. [CrossRef]

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.