Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence Firm Concentration and Urban Green Innovation

huan chen

15779268409@163.com

Guangdong University of Technology https://orcid.org/0009-0006-7507-1403

Research Article

Keywords: artificial intelligence firms, green innovation, patent composition, spatial association, temporal validation

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11805004/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Artificial Intelligence Firm Concentration and Urban Green Innovation

Abstract

The concentration of artificial intelligence (AI) firms may accompany local green patent activity without indicating a greener innovation portfolio or providing useful information about subsequent outcomes. This study evaluates these distinctions using a panel of 295 Chinese cities over 2013–2024. Spatial Durbin estimates combine a positive baseline direct effect of 0.139 with an indirect effect of −5.901; the latter is sensitive to patent definition and spatial weights. Within-city regressions do not establish a positive green-share association or a difference between green and nongreen patent responses. An expanding-window exercise then predicts outcomes one year ahead using only lagged indicators and benchmarks AI against prior patent performance, total patenting and economic conditions. Across 2020–2024, adding AI reduces the pooled RMSE for green invention applications by 1.43%, with a city-block bootstrap interval of 0.33–2.64%, but increases RMSE for all green applications by 2.95%. The green-share improvement is small and its interval includes zero. These results distinguish spatial association from environmental orientation and forward predictive information. They support an outcome-dependent assessment of AI enterprise counts as an innovation indicator, while providing no evidence of a causal policy effect or a general redirection toward green technologies.

Keywords: artificial intelligence firms; green innovation; patent composition; spatial association; temporal validation

1. Introduction

Local governments seek technological industries that can strengthen urban innovation and support environmental objectives. Enterprise counts offer a visible measure of industrial development, but their interpretation matters for regional policy: does a larger technological firm base indicate more green invention, a greener innovation portfolio, or information useful for monitoring future outcomes? These are different requirements. Evaluating them together helps determine what cities can infer from the firms they attract and whether a local innovation advantage extends across the surrounding spatial system.

Artificial intelligence can support research search, production optimization and environmental monitoring. However, the same capabilities can support inventions without an environmental purpose.

An AI enterprise count therefore identifies a local firm base, rather than its adoption by other firms or its use in environmental R&D. General inventive capacity, environmental orientation and the timing of observable patent outcomes are conceptually distinct (Bresnahan & Trajtenberg, 1995; Cockburn et al., 2018; Biggi et al., 2025).

China provides a setting in which technological enterprise concentration coexists with substantial differences in urban inventive capacity. For local and regional development agencies, an AI firm count may capture an established innovation center rather than an emerging environmental advantage. Its relevance therefore depends on the outcome being monitored, the geographic relationships considered and the information already available from a city's patent record.

Existing studies link AI adoption, AI-related patent knowledge and industrial agglomeration to environmental innovation (Wang et al., 2024; Biggi et al., 2025; Liu, 2026). He et al. (2024) examine favorable local and adverse spatial relationships. Cicerone et al. (2023) relate AI knowledge to regional green technological specialization, with a role for prior specialization. Building on this work, the present study assesses a city-level AI enterprise count as an indicator for regional innovation monitoring. It connects geographic association with tests of environmental composition and additional information beyond the existing patent base.

The assessment distinguishes local and intercity patent associations, environmental orientation and forward information. The first concerns the geographic distribution of green patent activity; the second concerns the green share of inventive activity and the comparison with nongreen patents; the third concerns incremental next-year information. This framework links the spatial analysis and supplementary prediction exercise to a common regional policy question: which aspects of environmental innovation can an observable enterprise count help cities monitor?

The analysis addresses three research questions:

(1) How does the AI firm base relate to local and spatially connected green patent output, and how sensitive is that relationship to its specification?

(2) Is the firm-count association environmentally specific when green and nongreen patent activity and patent composition are compared?

(3) Does the AI firm base add information for next-year green patent volume, invention applications and patent share beyond prior outcomes, total patenting and economic conditions?

The empirical design centers on spatial effect decomposition and regional heterogeneity. Within-city patent-composition regressions assess whether the association is environmentally specific. A supplementary expanding-window exercise evaluates next-year outcomes against an unchanged-last-year forecast and a model containing patent history. The contemporaneous model comparisons and temporal validation use different targets and information sets and are reported separately.

The study contributes to regional innovation research in three connected ways. It assesses the geographic reach of AI enterprise concentration by distinguishing local associations from indirect relationships and examining their sensitivity to outcome definition and spatial weights. It then evaluates environmental specificity through comparable green and nongreen patent regressions and the green patent share. Finally, it benchmarks the enterprise count against a city's established patent capabilities. Together, these comparisons connect a readily observed industrial indicator to the outcomes used to appraise urban green innovation.

The intended application is urban and regional innovation monitoring. Development agencies need to distinguish enterprise presence from environmental orientation and to test whether a new indicator adds information to existing records. This evaluation can inform the choice of monitoring measures; the effectiveness of policies that attract AI firms requires separate causal evidence.

The remainder of the paper is organized as follows. Section 2 reviews the relevant evidence and situates the indicator question. Section 3 develops the distinction between capability, environmental orientation and forward information. Section 4 describes the data and empirical designs. Section 5 presents spatial, composition and temporal-validation results. Section 6 discusses their implications and limitations, and Section 7 concludes.

2. Literature Review

2.1 Green innovation, policy incentives, and measurement

Green innovation differs from conventional innovation because both knowledge creation and environmental improvement generate benefits that may extend beyond the innovating firm. Rennings (2000) treats eco-innovation as a combination of technological, organizational, and institutional change. Porter and van der Linde (1995) emphasize the potential relationship between environmental regulation and innovation, while Jaffe et al. (2002) distinguish innovation from the adoption and diffusion of

environmental technologies. These distinctions imply that patent output captures one part of green technological change rather than the full environmental transition.

The direction of innovation also depends on incentives and accumulated capabilities. Popp (2002) examines the relationship between energy prices and energy-related invention, and Johnstone et al. (2010) use patent counts to investigate renewable-energy policy. Acemoglu et al. (2012) model the redirection of technical change toward cleaner inputs; Aghion et al. (2016) examine clean and dirty innovation in the automotive industry. Together, these studies motivate attention to the conditions under which AI capabilities are applied to environmental problems, without assuming that a larger AI firm base necessarily produces greener technologies.

2.2 AI-related capabilities and environmental innovation

The AI–green innovation literature measures several different objects. Biggi et al. (2025) examine AI content in green inventions; Wang et al. (2024) study enterprise adoption and innovation efficiency; Mijit et al. (2025) evaluate AI pilot-zone policy; and Liu (2026) relates AI industrial agglomeration to corporate breakthroughs in key green technologies. Cicerone et al. (2023) investigate regional AI knowledge stocks and specialization in environmental technologies. These objects differ from the number of AI-related enterprises in a city. Evidence about technological content, adoption or policy treatment cannot automatically establish the validity of an enterprise-count indicator.

2.3 Spatial organization and the research question

Knowledge spillovers can be geographically localized, as shown through patent citations by Jaffe et al. (1993), and the geographic organization of innovative activity need not match that of production (Audretsch & Feldman, 1996). However, location alone does not establish effective knowledge transfer. Boschma (2005) distinguishes geographic proximity from other forms of proximity, while Cohen and Levinthal (1990) emphasize the capabilities required to recognize and use external knowledge. These arguments motivate a distinction between a city's local AI firm base and the innovation opportunities available across connected cities.

The spatial evidence of He et al. (2024) and the specialization analysis of Cicerone et al. (2023) provide the starting points for the present study. Its focus is the environmental relevance of an observable AI enterprise base, assessed through spatial reach, portfolio composition and information about subsequent patent outcomes. The forward comparison conditions on the city's patent history and

total inventive capacity, allowing the enterprise count to be evaluated against information already available from established innovation activity. This comparison complements research on AI knowledge, adoption and agglomeration by examining the usefulness of the count-based measure for a defined monitoring task. Explanatory fit and predictive performance address different empirical questions (Shmueli, 2010), and temporal validation preserves the ordering of information and outcomes (Roberts et al., 2017).

2.4 Bibliometric overview

Figures 1–3 provide a descriptive overview of the thematic structure of the selected AI and green innovation literature over 2014–2025. They complement the substantive review and do not establish that the corpus is exhaustive or identify a research gap by themselves (van Eck & Waltman, 2010; Zupic & Čater, 2015).

Figure 1 presents the keyword co-occurrence network. Node size represents keyword frequency, while links indicate co-occurrence. AI, green innovation, and urban green innovation connect with terms relating to the digital economy, machine learning, green technologies, carbon emissions, and spatial analysis. These connections situate the present study at the intersection of digital technological capability, environmental innovation, and urban spatial relationships.

The clustering map in Figure 2 groups related keywords into themes concerning green innovation, AI and digital development, spatial spillovers, industrial agglomeration, and sustainable development. The clusters indicate overlapping research agendas rather than independent explanations of the AI–green innovation relationship. In particular, digital development and agglomeration should be distinguished: one concerns technological capability, whereas the other concerns the geographic organization of firms and resources.

Figure 3 places AI, machine learning and digital economy terms alongside environmental regulation, industrial agglomeration and carbon emissions over time. The map describes thematic relationships within the selected literature.

The substantive issue concerns the distribution of environmental innovation opportunities across cities. Knowledge diffusion can extend capabilities beyond a technological cluster, while resource competition can concentrate them. The distinction between local and intercity relationships motivates the spatial analysis.

The bibliometric overview serves as contextual evidence. The empirical contribution is evaluated through the city panel and the comparisons described below.

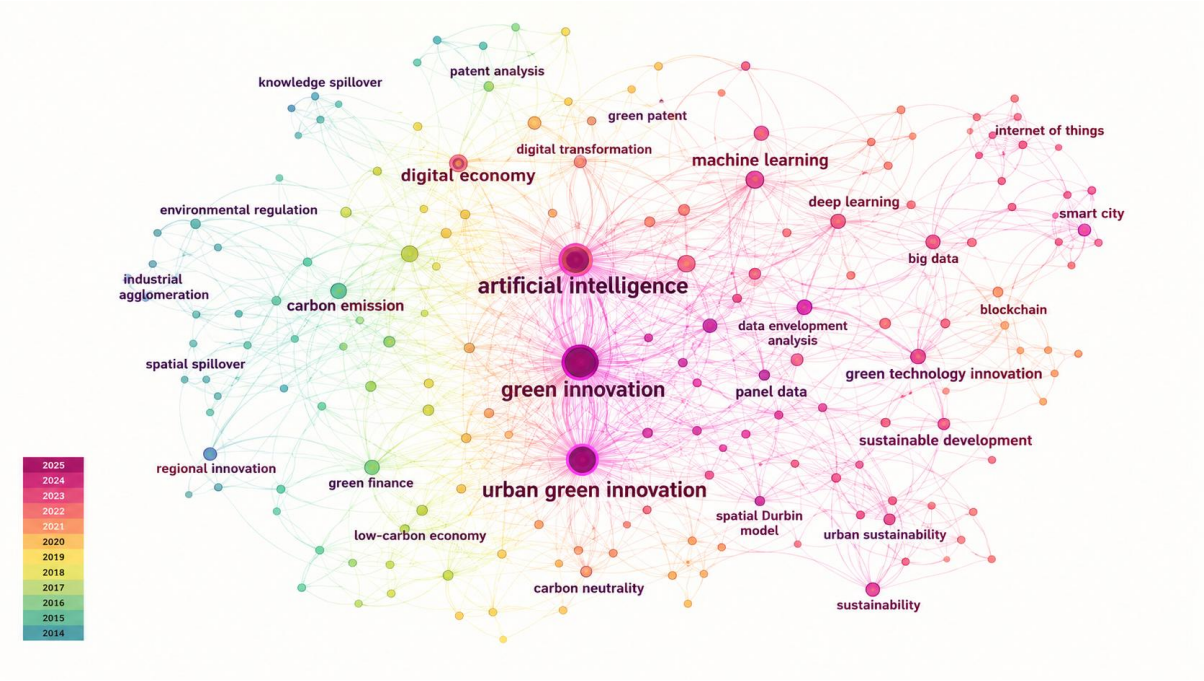


Fig. 1. VOSviewer Keyword Co-occurrence Network of AI and Green Innovation Research, 2014–2025

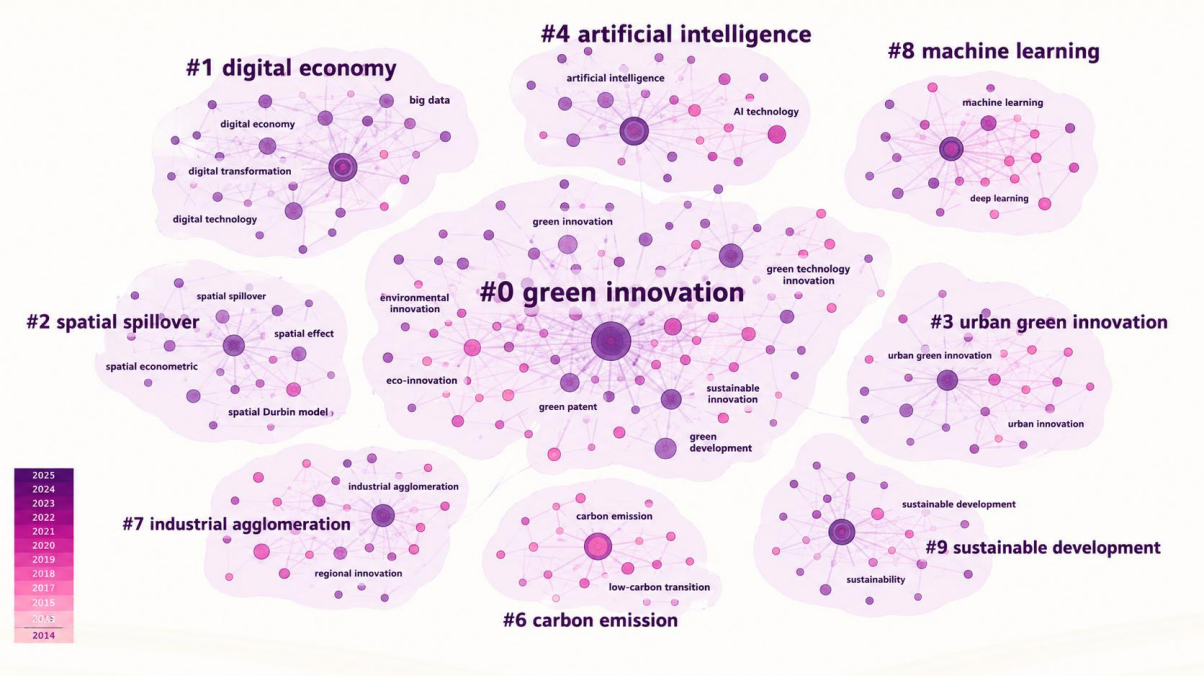


Fig. 2. VOSviewer Keyword Clustering Map of AI and Green Innovation Research, 2014–2025



Fig. 3. VOSviewer Timeline Evolution of Research Topics on AI and Green Innovation, 2014–2025

3. Theoretical Analysis

3.1 Local Capability and Patent Output

An urban AI enterprise base can improve access to specialized expertise, technical services and research partners. These capabilities can reduce search and coordination costs and help process experimental or production data, supporting research activities that include environmental applications. The local-capability argument predicts a favorable conditional association with green patent output where relevant demand and complementary knowledge are present. Because the same services also support other research, a positive local association alone does not determine the environmental direction of the resulting patent portfolio.

3.2 Environmental Direction of Inventive Activity

Environmental direction depends on the incentives and capabilities that guide the application of general technological inputs. Greater research capacity can increase green applications while leaving their share unchanged if nongreen activity expands similarly. The stronger environmental-direction argument therefore has two observable implications: a more favorable association with green than nongreen applications under a common specification, and a positive association with the green patent

share. The coefficient-difference and share tests assess these implications directly. Prior specialization can shape the use of digital knowledge (Cicerone et al., 2023), but the aggregate share in this dataset measures portfolio composition rather than specialization in individual technology classes.

3.3 Spatial Diffusion and Resource Competition

The local-capability argument does not determine the spatial counterpart. Collaboration, supplier relationships and technical services can transmit capabilities across cities, while competition for specialized workers and finance can concentrate them. The net relationship depends on both the connections through which knowledge moves and the capacities of receiving cities. Geographic proximity captures only part of this setting (Boschma, 2005). The spatial analysis therefore evaluates the reach and sensitivity of the local relationship; its indirect estimate is not assumed to identify a particular flow of talent or capital.

3.4 Incremental Information and Supporting Conditions

A useful monitoring indicator must also supply information that is not already contained in observed innovation capacity. Prior patent outcomes summarize accumulated capabilities, environmental orientation and persistence, while total patenting characterizes the general scale of inventive activity. The AI enterprise count could add information if it captures capabilities not yet visible in these records. This implication is assessed by predicting subsequent outcomes with and without the count while retaining prior patents and economic conditions in both models. The comparison distinguishes the indicator's additional information from the level of innovation in cities that already possess large research bases.

Figure 4 organizes three institutional and digital conditions that may shape the application of local capabilities. Fiscal science and technology expenditure can support research infrastructure and access to resources. Its association with AI-related enterprise activity is examined as contextual evidence; the proposed fiscal pathway to patent output is not tested in the reported outcome regressions.

Market conditions influence competition, financing and the allocation of innovation resources. They can support green investment when environmental applications offer attractive returns, but can favor other activities when those returns are uncertain. The marketization indicator does not directly measure short-term managerial pressure, so no fixed constraining direction is imposed.

Mobile communication user scale captures the reach of one form of digital connectivity. Connectivity can facilitate access and coordination, but subscriptions alone do not identify how digital services are used. Resource diversion toward consumer applications is not treated as an established mechanism.

The framework yields three connected assessment criteria. Local capability is evaluated through the SDM direct effect and its spatial robustness checks. Environmental direction is evaluated through the green–nongreen coefficient contrast and patent-share regression. Incremental information is evaluated through one-year-ahead comparisons against patent-history and persistence benchmarks. The related institutional regressions provide context for applying capabilities, while Figures 4 and 5 organize the proposed links and spatial workflow. Together, these components connect the theoretical distinction between broad capability and environmental application to observable empirical comparisons.

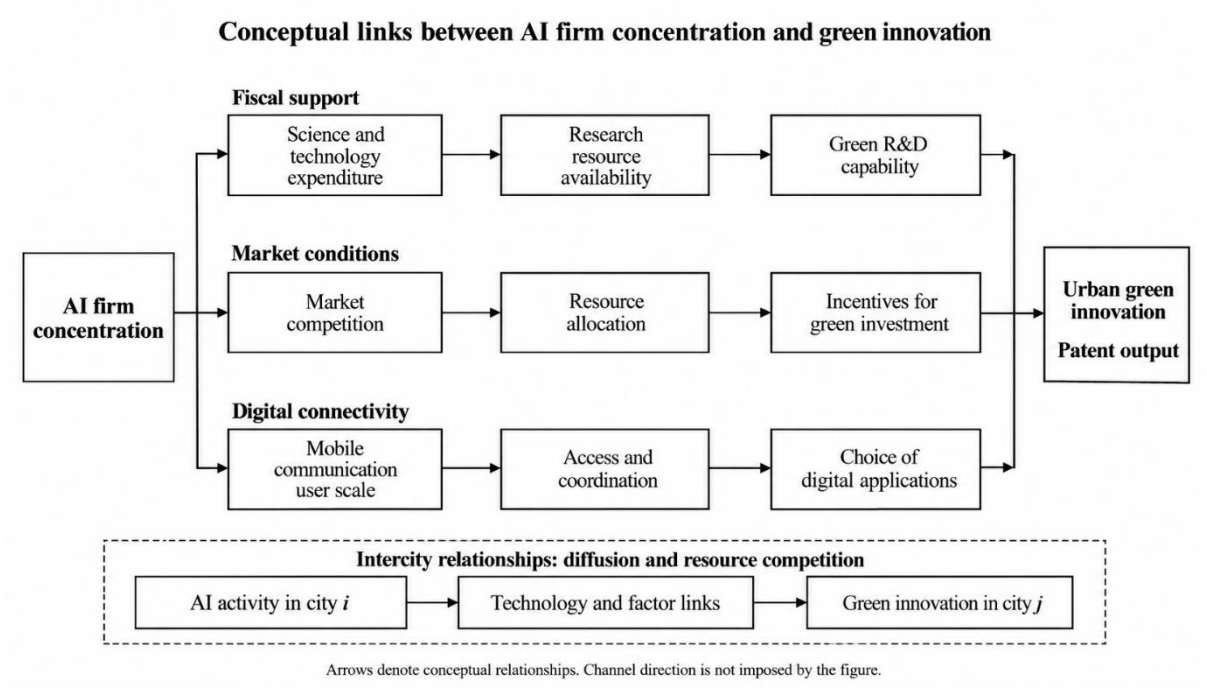


Fig. 4. Conceptual relationships between AI firm concentration and urban green innovation. Arrows denote proposed links; the sign and causal status of the institutional and digital relationships are not assumed.

4. Methodology

Figure 5 summarizes the spatial workflow and related-condition regressions. The analysis is extended by two separate indicator assessments: common-specification patent-composition diagnostics

and one-year-ahead temporal validation against patent-history benchmarks. The latter uses dated predictors and precedes any claim that the firm count could inform monitoring of subsequent innovation outcomes.

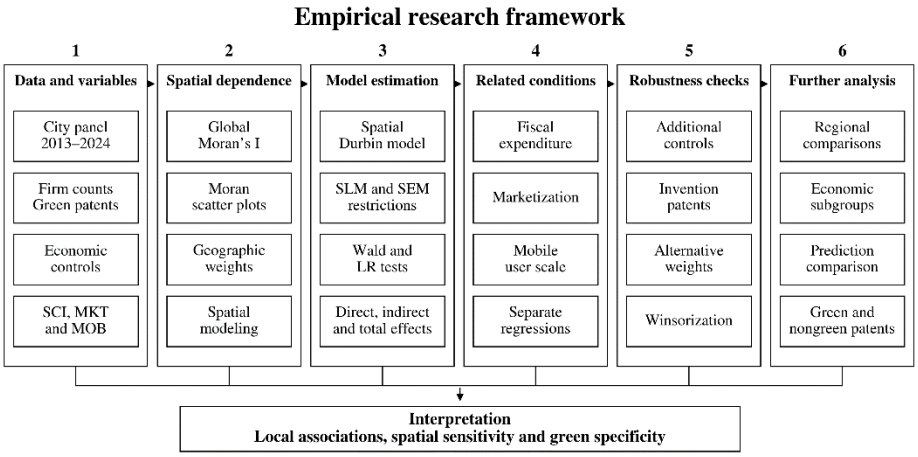


Fig. 5. Empirical workflow and proposed analytical components.

4.1 Variable Selection

4.1.1 Dependent Variable: Urban Green Innovation (GI)

Urban green innovation (GI) is measured using city-level green patent applications. Patent applications capture observable inventive activity related to environmental technologies, but not all forms of green innovation or their realized environmental benefits. The measure is therefore interpreted as patent output rather than a comprehensive indicator of green transition performance (Griliches, 1990; Haščič & Migotto, 2015).

The transformation retains observations with no patent applications and reduces the influence of large counts:

$$GI=ln(Green\ patent\ applications+1)$$

Adding one before taking the natural logarithm retains zero counts. Because the transformed variable is ln(1 + patents), coefficients should not be treated as exact elasticities of raw patent counts, particularly for observations with few applications.

Green invention patent applications (GII) provide an alternative outcome. They distinguish invention applications from the broader patent total, but do not directly establish patent quality, commercial value, or environmental impact. The alternative measure assesses whether the relationship extends to this more specific category of inventive activity.

GII is defined analogously:

$$GII=ln(Green\ invention\ patent\ applications+1)$$

4.1.2 Independent Variable: Artificial Intelligence Firm Concentration (AI)

AI is constructed from the city-year enterprise-count field ai_firm, using the natural logarithm of one plus the recorded count. The count follows the AI-related enterprise classification in the underlying city-level dataset; no additional enterprise-level search or reclassification is applied in this analysis. It measures the scale of the recorded local firm base, rather than a location quotient, density measure or firm-level adoption rate.

The count is transformed as follows:

$$AI=ln(Number\ of\ AI\text{-}related\ firms+1)$$

The ai_firm field contains nonnegative integer counts with no missing city-year values in the initial panel. The recorded AI series equals ln(1 + ai_firm) to numerical precision. The transformation retains zero counts. Because the analytical file contains aggregated counts rather than enterprise records, it does not permit independent verification of business-scope keyword selection, duplicate-enterprise handling, legal status or whether each year measures active firms, new registrations or a cumulative stock. The series is therefore described as recorded AI-related enterprise counts, without assigning one of these unverified administrative definitions.

The analysis preserves the recorded city-year counts, including observed declines, and does not replace them with a monotonic cumulative series. The measure identifies enterprise presence as classified in the dataset; it does not establish technical specialization, environmental applications or AI use by other firms. Its geographical interpretation follows the dataset's city identifiers, rather than an independently reconstructed allocation of individual enterprises.

4.1.3 Control Variables

The baseline model includes economic development (PGDP), government intervention (GOV), financial development (FIN), and openness (OPEN). PGDP is represented by logged per capita GDP;

the other indicators characterize fiscal expenditure intensity, local financial development, and foreign trade intensity. These controls account for selected economic conditions associated with both AI-related activity and patent output. Their inclusion reduces some sources of confounding but does not eliminate omitted variable bias or reverse causality.

Population density (DENS) and industrial structure (IND) are added in a robustness specification to account for urban concentration and industrial composition. The industrial structure variable is interpreted as the reported index, rather than assumed to measure technological upgrading or industrial dependence without reference to its construction.

4.1.4 Potential Channel Variables

Three variables are examined as potential channels or supporting conditions: fiscal science and technology expenditure (SCI), marketization (MKT), and mobile communication user scale (MOB). Their inclusion permits an assessment of related conditional associations. Establishing mediation requires additional assumptions and evidence beyond the significance of individual coefficients.

SCI is the natural logarithm of fiscal expenditure on science and technology in the city-level panel. It represents public financial support for scientific and technological activity, rather than expenditure directed exclusively toward AI or green R&D. City-level indicators are sourced from the China City Database.

MKT represents the marketization indicator supplied in the city-level panel. Its interpretation concerns market development and resource allocation, subject to the scale and construction of the underlying index. It is not treated as a direct measure of firms' short-term profit pressure.

MOB is the natural logarithm of mobile communication users in the city-level panel sourced from the China City Database. It captures the scale of subscriptions or users recorded in the source, rather than the quality of digital infrastructure or the intensity of green applications.

4.1.5 Other Variables

Spatial relationships are represented by a weight matrix and spatially lagged variables; unobserved time-invariant city differences and common annual shocks are represented by fixed effects.

The baseline spatial weight matrix (W) uses inverse geographic distances between city coordinates, with a zero diagonal and row normalization. It assigns positive weights to geographically connected

cities, including cities beyond immediate administrative neighbors. Indirect effects therefore refer to the spatial system represented by this matrix, not solely to adjacent cities.

The spatial lag WGI captures the relationship between a city's green innovation and the weighted green innovation of other cities. A significant spatial autoregressive coefficient indicates conditional spatial dependence; it does not independently identify technology diffusion or a causal transmission channel (Manski, 1993).

City fixed effects (μ_i) control for time-invariant city characteristics, and year fixed effects (λ_t) account for shocks common to all cities. Neither absorbs time-varying city-specific factors that may jointly influence AI-related activity and green innovation.

Robustness specifications use green invention patents, additional controls, alternative spatial weights and winsorization. The alternative weights use min–max standardization of the geographic distance matrix.

Table 1 summarizes the variables, symbols, and definitions used in the analysis.

[Insert Table 1 here]

Table 1. Description of variables

Classification	Variable	Symbol	Definition
Dependent variable	Green innovation	GI	ln(1 + green patent applications)
Alternative dependent variable	Green invention innovation	GII	ln(1 + green invention patent applications)
Independent variable	AI firm concentration	AI	ln(1 + AI-related firms)
Control variable	Economic development level	PGDP	Natural logarithm of per capita GDP
Control variable	Government intervention	GOV	Government fiscal expenditure intensity
Control variable	Financial development	FIN	Local financial development level
Control variable	Openness	OPEN	Foreign trade intensity
Additional control variable	Population density	DENS	Population density of the city
Additional control variable	Industrial structure	IND	Industrial structure indicator

Classification	Variable	Symbol	Definition
Potential channel variable	Fiscal support for science and technology	SCI	Logarithm of science and technology fiscal expenditure
Potential channel variable	Marketization	MKT	Marketization level
Potential channel variable	Mobile communication user scale	MOB	Logarithm of mobile communication users

4.2 Data Sources

The study combines city-level indicators from the China City Database with green patent data from Chinese Research Data Services (CNRDS). Enterprise-related information is sourced from the China Stock Market & Accounting Research Database (CSMAR, Guotai An) and Shenzhen Stock Exchange disclosures. The analytical panel is organized at the city-year level. The source attribution for enterprise-related information does not establish that exchange disclosures alone cover the population of AI-related enterprises; the ai_firm indicator retains the classification and coverage represented in the aggregated city panel.

The initial panel contains 297 city identifiers over 2013–2024, with 3,564 city-year records and no duplicate id–year keys. Green patent applications and green invention applications are missing in two years for Sansha and all twelve years for Danzhou, giving 14 missing records for each outcome. To retain complete annual outcome coverage for the baseline spatial panel, all 24 records belonging to these two cities are excluded, including Sansha's ten nonmissing records. The resulting baseline contains 295 cities and 3,540 observations. This is a city-level complete-coverage restriction, not a restriction that deletes only the 14 missing rows.

Observations are ordered and matched by the dataset's city identifier and year; city names alone are not used as merge keys. GI and GII equal ln(1 + green applications) and ln(1 + green invention applications), while AI equals ln(1 + ai_firm). Recorded zero patent counts are retained. Within the 295-city baseline, GI, GII, AI, the four baseline controls and city coordinates are complete. No zero replacement or interpolation is applied to these variables in the composition or temporal analyses. The data also include fiscal expenditure, marketization and communication indicators; the sample rules used below are summarized in Appendix Table A4.

SCI has one missing baseline value, for Huludao in 2023. The regression relating SCI to AI reports 3,539 observations, consistent with omission of that missing value. SCI is examined only as a supporting condition associated with AI; no SCI outcome equation is used to assess its relationship with green patent output. The MKT and MOB regressions report 3,540 observations.

For the supplementary patent-composition regressions, nongreen applications equal total minus green applications; this difference is nonnegative in every baseline record. Counts retain all 3,540 observations. Green patent share is defined only where total applications are positive. Rikaze and Nagqu each have zero total applications in 2014; those two records are omitted only from the share regression, leaving 3,538 observations. Missing shares are not set to zero.

4.3 Model Selection and Construction

4.3.1 Spatial Autocorrelation Test

Spatial autocorrelation tests assess whether observations exhibit geographic dependence under the chosen weights. Such dependence can arise from shared conditions, common shocks, or interaction between cities. It motivates consideration of a spatial specification but does not establish the direction of the AI–green innovation relationship (Anselin, 1995).

Global Moran's I is calculated for GI and AI:

$$I=\left[\frac{N}{SO}\right]\times\frac{\sum i\sum j w_{ij}(x_i-\overline{x})(x_j-\overline{x})}{\sum i(x_i-\overline{x})^2} \quad (1)$$

Here, n is the number of cities, w_{ij} is the spatial weight between cities i and j, x_i and x_j are the observed values, $\bar{x}$ is their mean, and S_0 is the sum of the weights. A significantly positive statistic indicates positive spatial autocorrelation under the specified matrix.

4.3.2 Spatial Durbin Model (SDM)

The SDM includes a spatial lag of the dependent variable and spatial lags of the explanatory variables. This permits the estimated relationship with AI to extend beyond the local regression coefficient. Compared with spatial lag and spatial error restrictions, the full specification allows a broader representation of spatial dependence (Elhorst, 2010; LeSage & Pace, 2009).

First, the basic spatial panel model is specified as follows:

$$GI_{it} = \alpha + \rho\sum_j w_{ij}\, GI_{jt} + \beta_0\, AI_{it} + \beta_1\, X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (2)$$

Furthermore, the Spatial Durbin Model is constructed as follows:

$$GI_{it} = \alpha + \rho\sum_j w_{ij}\, GI_{jt} + \beta_0\, AI_{it} + \beta_1\, X_{it} + \theta_0\sum_j w_{ij}\, AI_{jt} + \theta_1\sum_j w_{ij}\, X_{jt} + \mu_i + \lambda_t + \varepsilon_{it} \quad (3)$$

In equations (2) and (3), i indexes cities and t indexes years. GI_{it} is logged green patent output, AI_{it} is the logged AI-related firm count, and X_{it} contains PGDP, GOV, FIN, and OPEN. W is the spatial weight matrix, ρ is the spatial autoregressive coefficient, and the θ parameters multiply spatially lagged explanatory variables. μ_i and λ_t denote city and year fixed effects, respectively, and ε_{it} is the disturbance.

The baseline spatial weight matrix is constructed based on inverse geographic distance. The element Wij is defined as follows:

$$w_{ij} = 1/d_{ij}, \quad i \neq j \quad (4)$$

$$w_{ij} = 0, \quad i = j \quad (5)$$

Here, d_{ij} is the geographic distance between cities. The baseline matrix is row-normalized. The alternative specification changes the matrix standardization, allowing an assessment of sensitivity to the representation of spatial connections (Halleck Vega & Elhorst, 2015).

Wald and likelihood-ratio tests assess the restrictions under which the SDM reduces to a spatial lag or spatial error model. The reported tests reject both restrictions. The baseline specification includes city and year fixed effects and is estimated by maximum likelihood. Direct, indirect, and total effects are interpreted using the spatial multiplier rather than the coefficients of AI and W × AI alone. For AI, the derivative matrix is $(I - \rho W)^{-1}(\beta_{AI} I + \theta_{AI} W)$. The average diagonal summarizes the direct effect, the average off-diagonal row sum summarizes the indirect effect, and their sum is the total effect. This decomposition incorporates feedback through the spatial system.

4.3.3 Channel-Related Regressions

The additional regressions examine SCI, MKT, and MOB as potential pathways or supporting conditions. A first equation relates each variable to AI and the controls. For MKT and MOB, a second equation includes the variable and its spatial lag in the GI equation. SCI is included only in the first equation. These regressions characterize conditional relationships; they do not by themselves establish causal mediation (Imai et al., 2010).

The first-stage equation is specified as follows:

$$M_{it} = \alpha_0 + \alpha_1\, AI_{it} + \alpha_2\, X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (6)$$

The corresponding spatial outcome equation is:

$$GI_{it} = \delta_0 + \rho\sum_j w_{ij}\, GI_{jt} + \delta_1\, AI_{it} + \delta_2\, M_{it} + \delta_3\, X_{it} + \theta_1\sum_j w_{ij}\, AI_{jt} + \theta_2\sum_j w_{ij}\, M_{jt} + \theta_3\sum_j w_{ij}\, X_{jt} + \mu_i + \lambda_t + \varepsilon_{it} \quad (7)$$

M_{it} denotes SCI, MKT, or MOB in the supporting-condition equation, and MKT or MOB in the GI outcome equation. The remaining variables, spatial weights, and fixed effects follow the baseline specification. WGI, WAI, and WM represent spatial lags. The outcome model permits both local coefficients and spatially lagged covariate terms, but its full effects also depend on the spatial multiplier.

The two equations describe conditional associations with each institutional or digital variable. They do not identify a mediation effect without additional assumptions and a corresponding spatial decomposition (Imai et al., 2010). In particular, the coefficient on a spatially lagged covariate does not itself equal its indirect effect.

4.4 Robustness Checks and Predictive Comparison

The robustness checks assess sensitivity to outcome measurement and specification. First, GII replaces GI:

$$GII_{it} = \alpha + \rho\sum_j w_{ij}\, GII_{jt} + \beta_0\, AI_{it} + \beta_1\, X_{it} + \theta_0\sum_j w_{ij}\, AI_{jt} + \theta_1\sum_j w_{ij}\, X_{jt} + \mu_i + \lambda_t + \varepsilon_{it} \quad (8)$$

Second, DENS and IND are added to the control vector:

$$GI_{it} = \alpha + \rho\sum_j w_{ij}\, GI_{jt} + \beta_0\, AI_{it} + \beta_1\, X_{it} + \beta_2\, Z_{it} + \theta_0\sum_j w_{ij}\, AI_{jt} + \theta_1\sum_j w_{ij}\, X_{jt} + \mu_i + \lambda_t + \varepsilon_{it} \quad (9)$$

Z_{it} contains these additional controls.

Further specifications change the spatial weights and winsorize continuous variables. These comparisons assess sensitivity to modeling choices.

Separately, eight predictive models compare specifications with and without AI. The models are Linear Regression, Ridge, Lasso, SVR, Random Forest, Extra Trees, Gradient Boosting, and XGBoost. The model set combines linear specifications with ensemble and boosting approaches (Breiman, 2001; Chen & Guestrin, 2016; Friedman, 2001), providing complementary ways to assess predictive information (Varian, 2014). Performance is summarized by root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2):

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2\right]} \quad (10)$$

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i| \quad (11)$$

$$R^2 = 1 - [\sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2] \quad (12)$$

Lower RMSE and MAE and higher R^2 indicate better performance within the evaluation sample. Prediction comparisons concern incremental information rather than a causal effect. Temporal and geographic generalization require an evaluation design that separates training from the intended prediction setting (Roberts et al., 2017).

A supplementary green-specificity diagnostic uses the baseline city panel in non-spatial regressions with city and year fixed effects. The outcomes are ln(1 + green applications), ln(1 + nongreen applications), the green share of all applications and the difference between the two logged counts. Nongreen applications equal total applications minus green applications. The share equation excludes two city-years with zero total applications, for which a patent share is undefined; it uses 3,538 observations. The count and contrast equations retain all 3,540 observations. The difference equation directly tests the green–nongreen slope contrast with the same controls. Standard errors are clustered by city, with a finite-sample correction accounting for absorbed fixed effects; two-sided p-values use 294 degrees of freedom. These regressions are specification diagnostics and are not substitutes for the SDM effect decomposition.

4.5 One-Year-Ahead Validation Beyond Patent History

The temporal exercise predicts GI, GII and green patent share in target years 2020–2024. Each fold trains on target years from 2014 through the year preceding the test year. Predictors are the previous year's target outcome, ln(1 + total patent applications), PGDP, GOV, FIN and OPEN, plus a linear calendar trend. The augmented model adds the previous year's AI measure. A separate persistence forecast carries the previous outcome forward unchanged. All three forecasts are assessed on the same observations within each outcome and year.

Both fitted specifications use ridge regression. The penalty is selected from 0.1, 1, 10 and 100 using the latest preceding target year as validation data and earlier years as the inner training set. After selection, the model is refitted on all preceding target years. Standardization is fitted within each training

split. No target-year outcome enters training, tuning or preprocessing. Share predictions are clipped to [0, 1]. The outcomes and temporal windows are fixed across specifications, rather than selected according to statistical significance. Each outcome has 295 evaluated cities per year and 1,475 pooled predictions.

Pooled RMSE reductions compare the augmented model with the patent-history model. Positive values indicate improved accuracy; negative values indicate deterioration. Percentile uncertainty intervals use 2,000 city-block bootstrap draws, resampling each city's complete five-year error trajectory with seed 20261013. These intervals preserve within-city dependence but condition on the fitted predictions and observed test years. Appendix Table A3 reports each year, so a pooled gain cannot conceal an adverse final-year result. Since the dataset supplies historical values rather than archived release vintages, this is a retrospective lagged-information exercise, not a verified real-time forecast deployment.

5. Results and Analysis

5.1 Descriptive Statistics and Correlation Analysis

Table 2 summarizes the baseline sample. GI and GII have means of 5.2188 and 4.3121 and standard deviations of 1.6918 and 1.7647. Their ranges indicate substantial differences in logged inventive activity across city-years.

The AI firm-count measure has a mean of 6.4810 and a standard deviation of 1.6995. Population density ranges from 1.2277 to 9008.5128. All listed variables have 3,540 observations except SCI, which has 3,539.

Table 2. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
GI	3540	5.2188	1.6918	0.0000	10.4288
GII	3540	4.3121	1.7647	0.0000	10.1765
AI	3540	6.4810	1.6995	0.0000	12.2368
PGDP	3540	10.8991	0.5557	9.0369	12.5614
GOV	3540	0.2227	0.1482	0.0383	1.5541

Variable	Obs	Mean	Std. Dev.	Min	Max
FIN	3540	1.2074	0.7019	0.1842	12.8171
OPEN	3540	0.1790	0.2930	0.0000	2.7233
DENS	3540	488.6033	718.2047	1.2277	9008.5128
IND	3540	0.4270	0.1088	0.0957	0.7936
SCI	3539	10.5992	1.5210	6.6254	15.6647
MKT	3540	15.7232	1.1044	11.8117	19.0367
MOB	3540	5.8734	0.8260	2.5649	8.4604

Figure 6 shows that AI is positively correlated with both patent outcomes, with a pooled correlation of 0.786 in each case. Pooled correlations combine within-city changes with persistent differences between cities. Appendix Table A1 reports VIFs between 1.197 and 4.301 for the listed specifications.

5.2 Spatiotemporal Patterns of AI and GI

Figures 7 and 8 map green patent activity and the AI firm base. Figure 9 aggregates these measures by province, while Figure 10 depicts their annual distributions. These figures describe the geography of the panel.

High values of both measures cluster in major eastern urban regions, including the Yangtze River Delta, Pearl River Delta and Beijing–Tianjin–Hebei area. The maps also show substantial variation across inland cities.

Figure 10 displays the distributions on a rescaled horizontal axis. These values are distinguished from the logged units in Table 2 and describe variation in the displayed distributions.

Figure 9 shows an uneven provincial distribution of patent applications and AI-related firms. Provincial totals reflect the size and activity of the constituent cities and are not measures of innovation efficiency.

The geographic overlap motivates the conditional spatial analysis, which accounts for observed economic conditions and fixed effects.

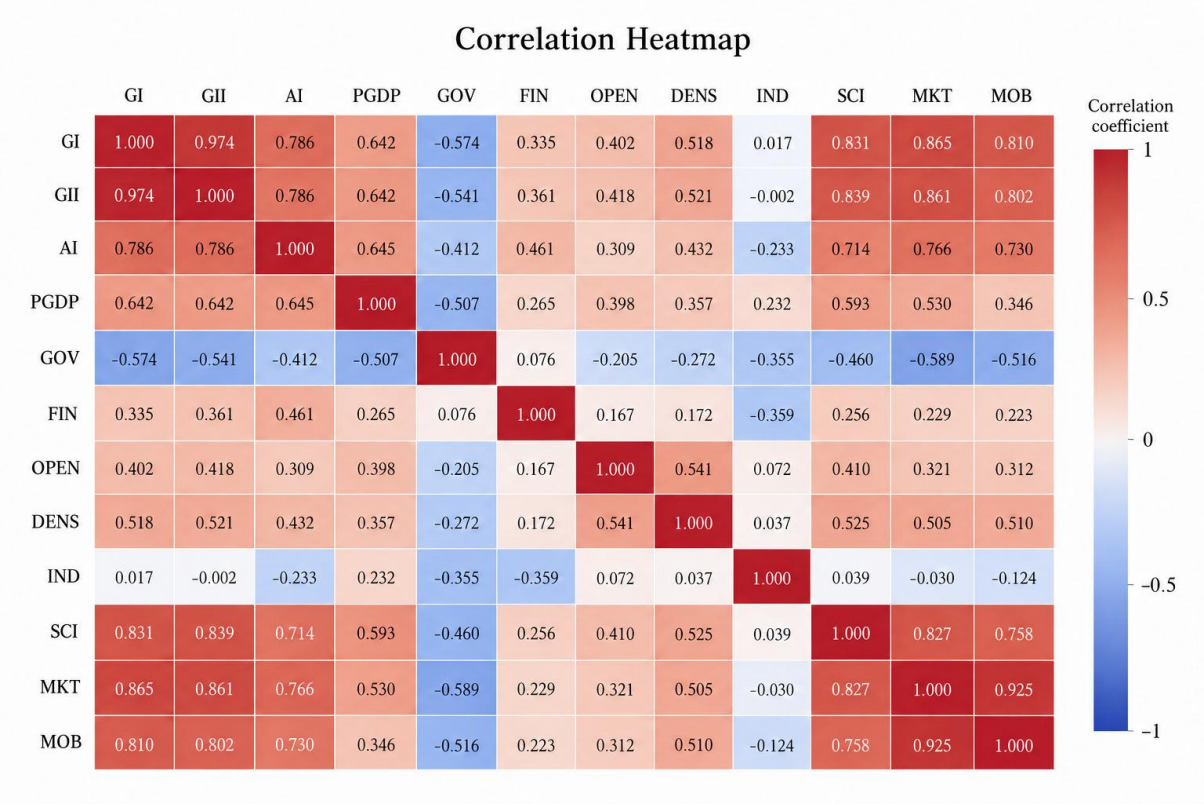


Fig. 6. Correlation Heatmap of Main Variables

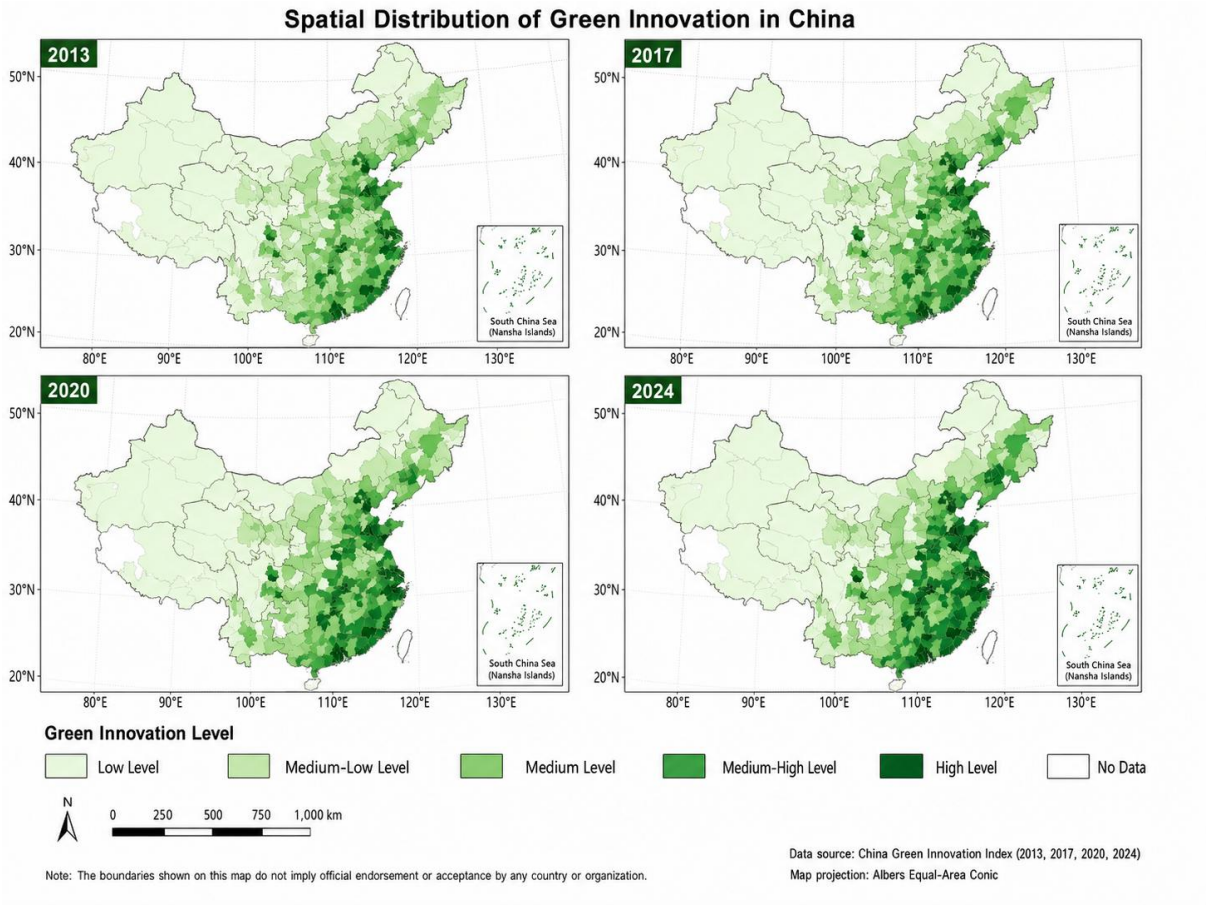


Fig. 7. Spatial Distribution of Urban Green Innovation in China, 2013–2024

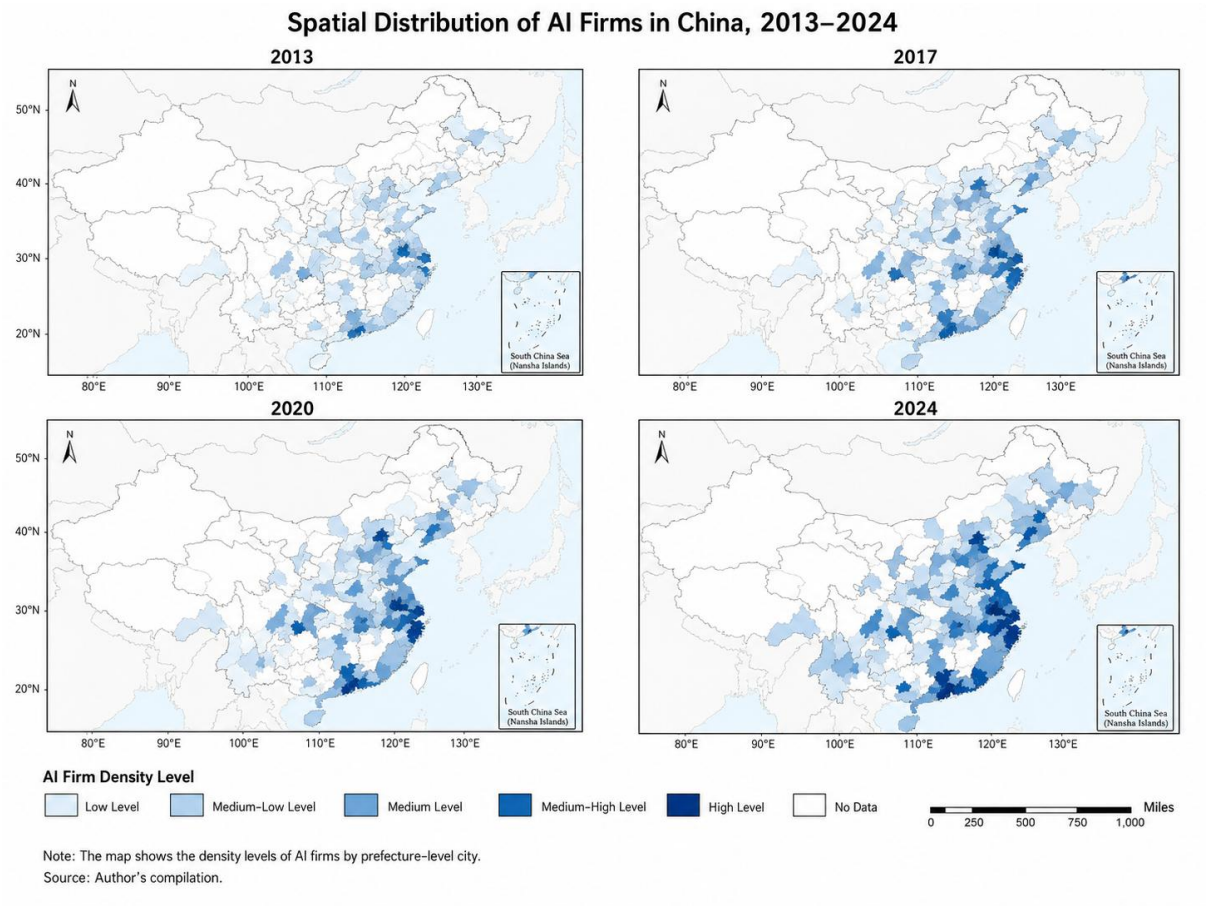


Fig. 8. Spatial Distribution of AI Firm Concentration in China, 2013–2024

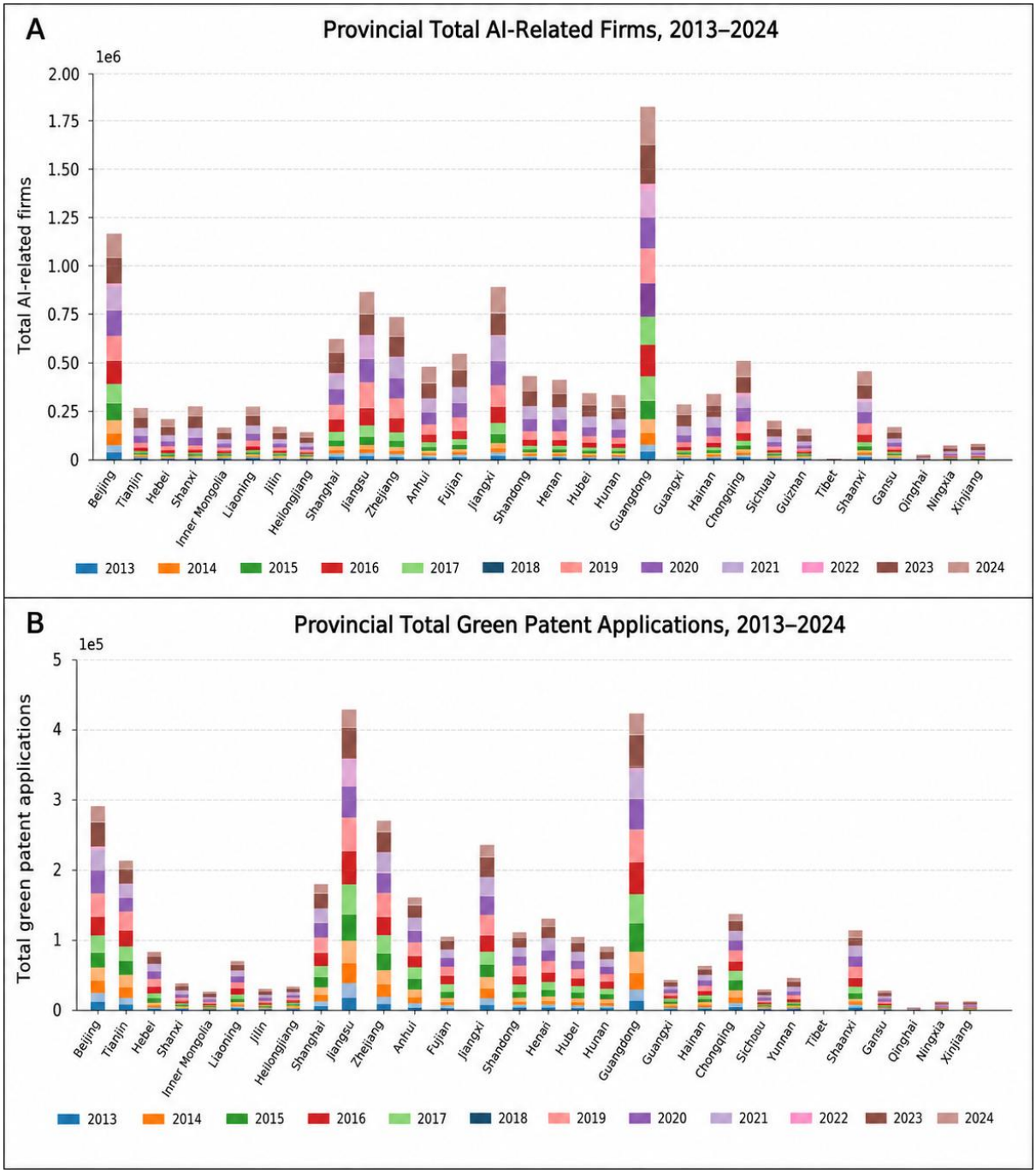


Fig. 9. Provincial Total AI-Related Firms and Green Patent Applications, 2013–2024

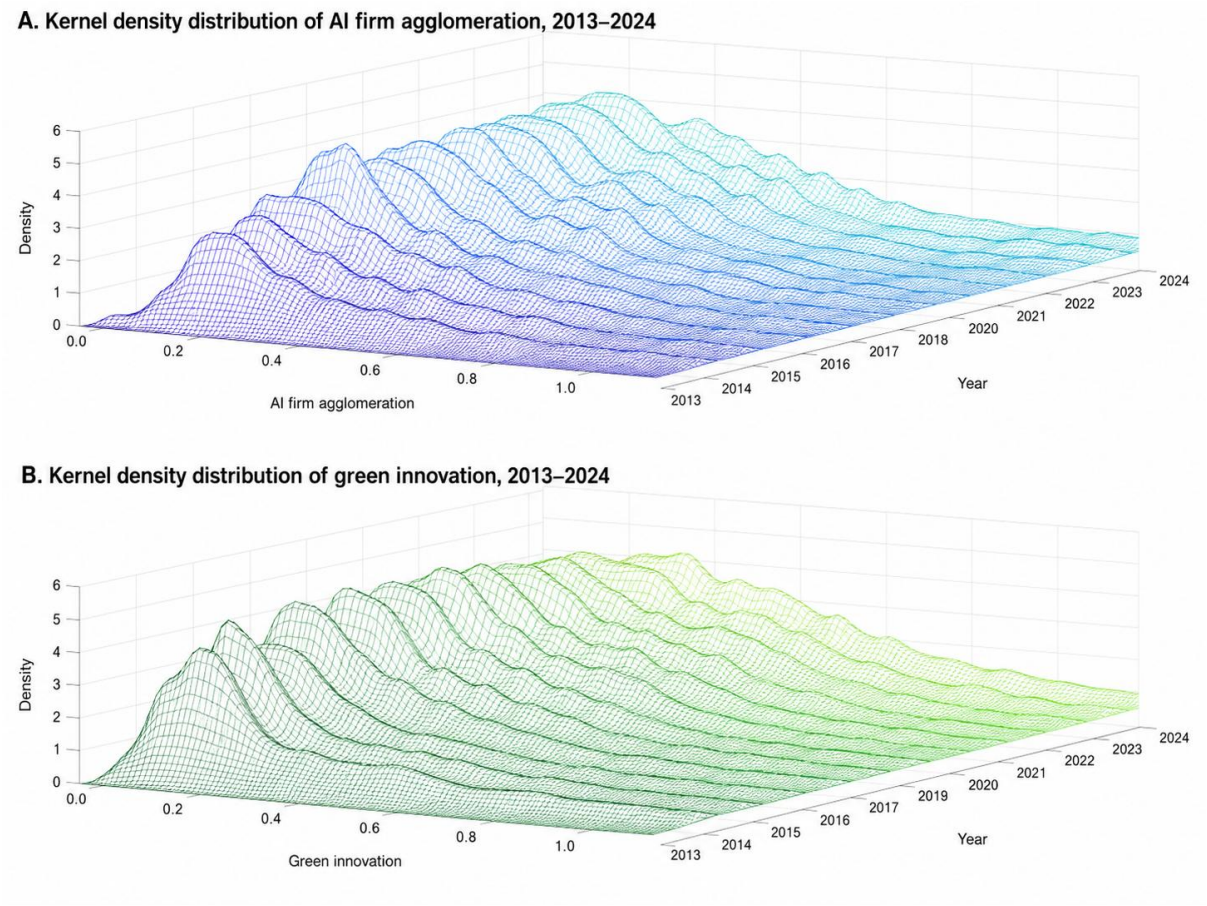


Fig. 10. Kernel Density Distribution of AI Firm Concentration and Green Innovation, 2013–2024

5.3 Spatial Autocorrelation Tests

Figures 11 and 12 report Moran scatter plots for GI and AI in 2013, 2017, 2020, and 2024.

For GI, Moran's I equals 0.0994, 0.1097, 0.1114, and 0.0973 in the four selected years, respectively; all displayed p-values are below 0.001. The positive statistics indicate positive spatial autocorrelation under the specified weights. The pattern is not a monotonic increase in clustering over time.

For AI, the corresponding values are 0.0535, 0.0529, 0.0617, and 0.0458, again with displayed p-values below 0.001. AI exhibits positive spatial autocorrelation, although the statistics are smaller than those for GI in each selected year. Statistical significance should be distinguished from the magnitude of this dependence.

These results establish spatial dependence under the specified weights and motivate the SDM. They do not identify the direction of an AI spillover.

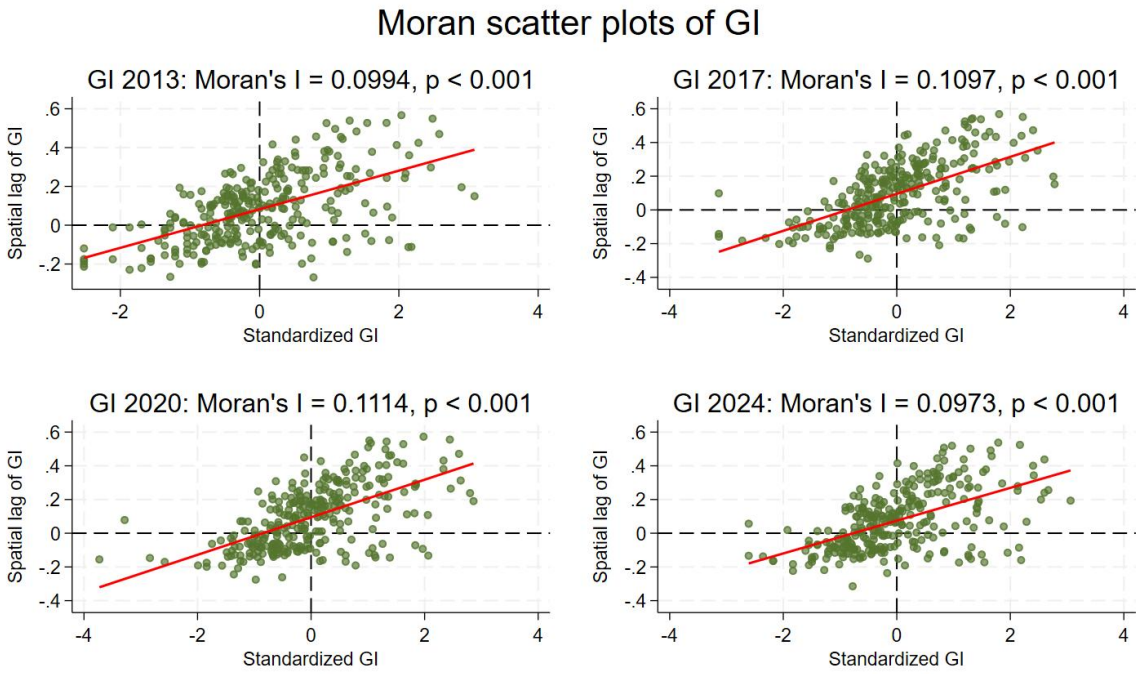


Fig. 11. Moran Scatter Plots of Urban Green Innovation (GI)

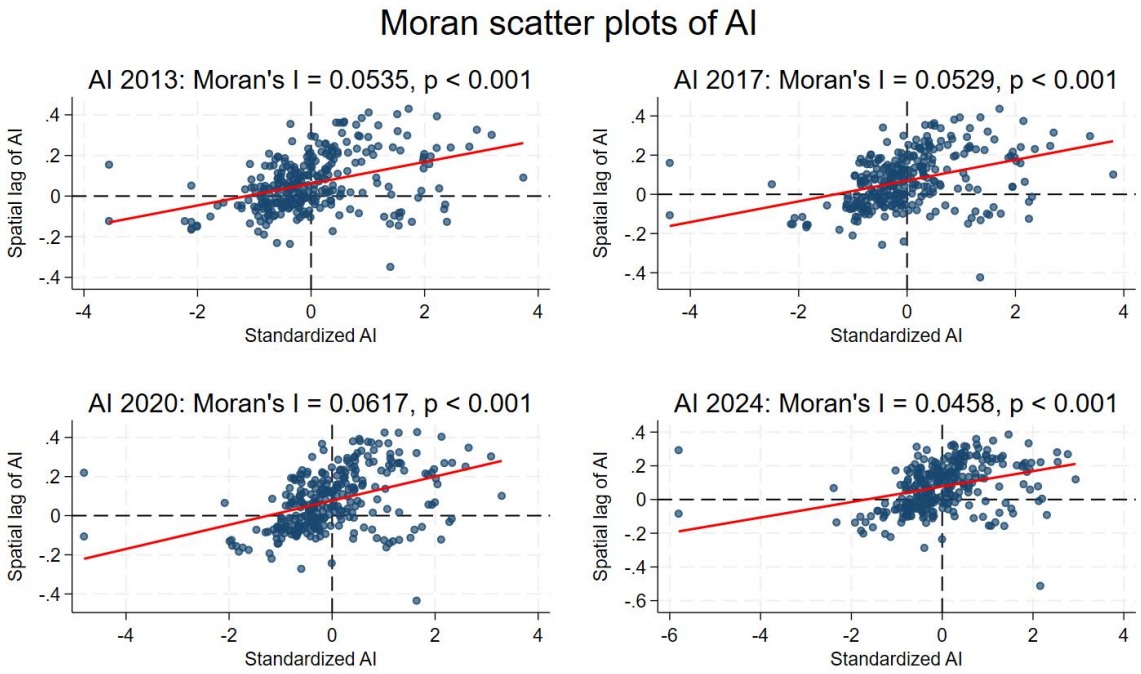


Fig. 12. Moran Scatter Plots of AI Firm Concentration (AI)

5.4 Spatial Impact of AI on GI

Table 3 reports the baseline SDM coefficients and effect decomposition. The Wald statistic for the spatial terms is 500.36. The reported Wald and likelihood-ratio restriction tests reject simplification to

either a spatial lag or spatial error model, with $p < 0.001$. The model includes city and year fixed effects and uses 3,540 observations from 295 cities.

AI has a positive coefficient of 0.159 and an average direct effect of 0.139, both significant at the 1% level. The direct effect is the relevant summary of the local relationship because it includes spatial feedback. It indicates that a larger AI firm base is associated with higher local logged green patent output, conditional on the model. The spatial autoregressive coefficient is 0.849 and significant at the 1% level.

The baseline AI indirect effect is −5.901, and the total effect is −5.763; both are significant at the 5% level. The estimates combine a positive direct association with a negative net relationship elsewhere in the weighted system. The magnitude of the indirect effect reflects spatial feedback as well as the coefficient on spatially lagged AI, making comparison across weights important.

PGDP and OPEN have positive direct effects of 0.182 and 0.111. FIN has negative direct and indirect effects of −0.069 and −10.654, while the GOV effects are not statistically significant.

The baseline estimates distinguish the within-city relationship from the net spatial association. Their stability is assessed in the following specifications.

Table 3. SDM regression results

Variables	Main (1)	Wx (2)	Direct effect (3)	Indirect effect (4)	Total effect (5)
AI	0.159***	-1.030***	0.139***	-5.901**	-5.763**
	(0.035)	(0.314)	(0.034)	(2.577)	(2.578)
PGDP	0.204***	-1.150***	0.182***	-6.438**	-6.257**
	(0.055)	(0.401)	(0.054)	(3.138)	(3.135)
GOV	-0.156	-0.709	-0.176	-5.552	-5.728
	(0.176)	(1.691)	(0.171)	(11.076)	(11.075)
FIN	-0.032	-1.589***	-0.069***	-10.654***	-10.723***
	(0.020)	(0.243)	(0.022)	(3.294)	(3.304)
OPEN	0.131**	-0.962*	0.111*	-5.611	-5.500
	(0.066)	(0.561)	(0.065)	(3.945)	(3.946)

Variables	Main (1)	Wx (2)	Direct effect (3)	Indirect effect (4)	Total effect (5)
Spatial rho	0.849***				
	(0.043)				
Wald test of spatial terms	500.36***				
Wald_lag	76.87 (Prob > chi2 = 0.0000)				
Wald_err	74.91 (Prob > chi2 = 0.0000)				
LR_lag	76.18 (Prob > chi2 = 0.0000)				
LR_err	74.46 (Prob > chi2 = 0.0000)				
City FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	3,540	3,540	3,540	3,540	3,540
R-squared	0.723	0.723	0.723	0.723	0.723
Number	295	295	295	295	295

Notes: Parentheses contain standard errors. * p < 0.10, ** p < 0.05, *** p < 0.01. Direct, indirect and total effects are distinguished from the Main and Wx coefficients where reported.

5.5 Robustness Checks

Table 4 reports four robustness specifications. The AI direct effect is positive and significant in every column, ranging from 0.114 to 0.166. The indirect effect is negative and significant with additional controls and winsorization, but is not significant with green invention patents or alternative weights. The evidence is therefore more stable for the local relationship within the spatial specifications than for the negative indirect effect.

5.5.1 Adding Additional Control Variables

Adding DENS and IND yields an AI coefficient of 0.164 and a direct effect of 0.142, both significant at the 1% level. The indirect effect remains negative at −6.244 and significant at the 5%

level. The association persists after this additional adjustment, although the specification cannot rule out other omitted factors.

5.5.2 Replacing the Dependent Variable

With GII as the outcome, the AI coefficient is 0.121 and the direct effect is 0.114, both significant at the 1% level. The indirect effect is −2.011 and not statistically significant. The positive local relationship thus extends to green invention applications, while evidence of a negative spillover is not retained for this outcome.

5.5.3 Changing the Spatial Weight Matrix

Under the alternative spatial matrix, the direct effect is 0.126 and significant at the 1% level. The indirect effect is −0.111 and the total effect is 0.015; neither is statistically significant. The spatial autoregressive coefficient remains positive at 0.574. This specification supports the local relationship but shows substantial sensitivity of the indirect and total effects to the spatial weights.

5.5.4 Accounting for Extreme Values

Winsorization produces a direct effect of 0.166 and an indirect effect of −3.841. The former is significant at the 1% level and the latter at the 10% level. The signs resemble the baseline pattern, although the indirect effect has weaker statistical support.

Across these spatial specifications, the direct effect remains positive. The indirect relationship depends on the outcome measure and spatial weights. This distinction is retained in the conclusions.

Table 4. Robustness test results

Variables	(1)Adding Control Variables	(2)Replacing Dependent Variable	(3)Alternative Spatial Matrix	(4)Winsorized Variables
AI	0.164***	0.121***	0.126***	0.179***
	(0.035)	(0.044)	(0.033)	(0.032)
W × AI	-1.156***	-0.429	-0.168	-0.779***
	(0.316)	(0.396)	(0.174)	(0.298)
Direct effect	0.142***	0.114***	0.126***	0.166***
	(0.034)	(0.042)	(0.033)	(0.031)

Variables	(1)Adding Control Variables	(2)Replacing Dependent Variable	(3)Alternative Spatial Matrix	(4)Winsorized Variables
Indirect effect	-6.244**	-2.011	-0.111	-3.841*
	(2.527)	(2.418)	(0.199)	(2.052)
Total effect	-6.101**	-1.898	0.015	-3.675*
	(2.527)	(2.414)	(0.200)	(2.051)
Spatial rho	0.837***	0.838***	0.574***	0.837***
	(0.046)	(0.046)	(0.094)	(0.047)
Wald spatial terms	471.90***	378.32***	174.07***	441.15***
Control variables	Yes	Yes	Yes	Yes
Spatial FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	3,540	3,540	3,540	3,540
R-squared	0.706	0.717	0.729	0.727
Number	295	295	295	295

Notes: Parentheses contain standard errors. * p < 0.10, ** p < 0.05, *** p < 0.01. Direct, indirect and total effects are distinguished from the Main and Wx coefficients where reported.

5.6 Heterogeneity Test

Tables 5 and 6 compare geographic and economic subgroups. The estimates describe variation across samples; formal cross-group tests are needed to establish statistically significant differences between effects.

5.6.1 Regional Heterogeneity

In eastern China, the direct effect is 0.347 and the indirect effect is 2.381, both significant at the 1% level. The total effect is 2.728. This pattern is consistent with positive local and spatial associations within the eastern subgroup. Stronger intercity collaboration is a possible explanation, but is not directly measured by these estimates.

In central China, the direct effect is −0.067 and the indirect effect is −0.340; neither is statistically significant. The estimates do not provide clear evidence of a local or indirect association in this

subgroup. Insignificance should not be interpreted as proof that AI-related activity has no relevance to central cities.

In western China, the direct effect is 0.275 and the indirect effect is −6.445, both significant at the 1% level. The total effect is −6.170. The signs resemble the national baseline, but the proposed roles of resource competition and limited diffusion require evidence on actual intercity linkages.

5.6.2 Heterogeneity by the Hu Line

East of the Hu Line, the direct effect is 0.168 and significant at the 1% level; the indirect and total effects are not statistically significant. This subgroup therefore provides clearer evidence of a local association than of a net spatial association.

West of the Hu Line, the direct effect is 0.918 and significant at the 1% level, and the total effect is 2.213 and significant at the 5% level. The indirect effect and spatial autoregressive coefficient are not significant. This subgroup contains only 30 cities, compared with 265 east of the line, which limits direct comparison. The Hu Line partition also differs from the eastern, central, and western regional classification.

5.6.3 Heterogeneity by Economic Development Level

High-PGDP cities have a direct effect of 0.222, significant at the 1% level; low-PGDP cities have a direct effect of 0.086, significant at the 10% level. The point estimates suggest a stronger local association in the higher-income subgroup, but a formal cross-group test is needed to establish this difference statistically.

Neither PGDP subgroup has a statistically significant indirect effect. The evidence therefore concerns variation in local point estimates rather than a clear difference in spatial spillovers.

5.6.4 Heterogeneity by Industrial Structure

In the high-IND subgroup, the direct effect is −0.040 and not statistically significant. This result is described in terms of the industrial structure index; it does not directly demonstrate industrial lock-in or high adjustment costs.

In the low-IND subgroup, the direct effect is 0.222 and significant at the 1% level, while the indirect effect is −2.922 and significant at the 1% level. The total effect is −2.700 and significant at the

5% level. These results indicate a positive local and negative indirect association in this subgroup, without identifying the resource flows responsible.

The national estimate does not describe every regional sample: the eastern indirect effect is positive, whereas the western and low-IND indirect effects are negative. Regional and Hu Line partitions represent different groupings and are interpreted separately.

Table 5. Geographic subgroup estimates

Variables	(1) Eastern region	(2) Central region	(3) Western region	(4) East of Hu Line	(5) West of Hu Line
AI	0.328***	-0.064	0.331***	0.165***	0.934***
	(0.059)	(0.051)	(0.068)	(0.034)	(0.176)
W × AI	1.205***	-0.069	-3.038***	0.044	1.775
	(0.354)	(0.308)	(0.434)	(0.310)	(1.230)
Direct effect	0.347***	-0.067	0.275***	0.168***	0.918***
	(0.057)	(0.050)	(0.068)	(0.033)	(0.176)
Indirect effect	2.381***	-0.340	-6.445***	0.837	1.296
	(0.723)	(0.905)	(1.413)	(1.464)	(1.027)
Total effect	2.728***	-0.407	-6.170***	1.005	2.213**
	(0.710)	(0.905)	(1.420)	(1.461)	(1.037)
Spatial rho	0.438***	0.673***	0.561***	0.793***	-0.224
	(0.109)	(0.080)	(0.103)	(0.057)	(0.208)
Wald spatial terms	65.25***	106.48***	316.55***	243.60***	24.08***
Control variables	Yes	Yes	Yes	Yes	Yes
Spatial FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	1,200	1,200	1,140	3,180	360
Pseudo R-squared	0.719	0.720	0.672	0.671	0.648
Number	100	100	95	265	30

Notes: Parentheses contain standard errors. * p < 0.10, ** p < 0.05, *** p < 0.01. Direct, indirect and total effects are distinguished from the Main and Wx coefficients where reported.

Table 6. Subgroup estimates by economic development and industrial structure

Variables	(1) High PGDP	(2) Low PGDP	(3) High industry	(4) Low industry
AI	0.236***	0.090*	-0.027	0.240***
	(0.052)	(0.048)	(0.054)	(0.045)
W × AI	-0.811*	-0.267	-0.526	-1.334***
	(0.420)	(0.384)	(0.383)	(0.381)
Direct effect	0.222***	0.086*	-0.040	0.222***
	(0.049)	(0.047)	(0.052)	(0.045)
Indirect effect	-2.063	-0.535	-1.873	-2.922***
	(1.358)	(0.962)	(1.358)	(1.080)
Total effect	-1.841	-0.449	-1.913	-2.700**
	(1.350)	(0.961)	(1.351)	(1.080)
Spatial rho	0.688***	0.606***	0.711***	0.595***
	(0.082)	(0.095)	(0.077)	(0.099)
Wald spatial terms	168.64***	84.81***	218.55***	108.37***
Control variables	Yes	Yes	Yes	Yes
Spatial FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	1,776	1,764	1,776	1,764
Pseudo R-squared	0.705	0.722	0.702	0.703
Number	148	147	148	147

Notes: Parentheses contain standard errors. * p < 0.10, ** p < 0.05, *** p < 0.01. Direct, indirect and total effects are distinguished from the Main and Wx coefficients where reported.

5.7 Evidence on Potential Channels

Table 7 presents regressions involving SCI, MKT and MOB. Columns (1), (2) and (4) relate each condition to AI, while columns (3) and (5) include MKT and MOB, respectively, and their spatial lags in the GI equation.

For SCI, the AI coefficient is 0.142 and is not statistically significant. This regression does not establish an association between AI-related enterprise concentration and fiscal science and technology expenditure. No inference about the SCI–patent relationship or fiscal mediation is drawn from the reported specifications.

For MKT, the AI coefficient is 0.122 and significant at the 5% level. The local MKT coefficient in the GI equation is 0.019 and not significant, whereas W × MKT is −0.545 and significant at the 5% level. This pattern does not establish a local constraining pathway, and the spatial-lag coefficient is distinguished from a decomposed indirect effect.

For MOB, AI has a coefficient of 0.046, significant at the 1% level. The local MOB coefficient in the GI equation is 0.052 and not significant; W × MOB is −4.982 and significant at the 1% level. The pattern concerns communication scale, not an observed diversion of innovation resources.

The regressions identify related institutional and digital conditions but do not establish mediation through any of the three variables. The spatial AI indirect effect therefore remains distinct from the proposed explanations in Figure 4.

Table 7. Regressions involving potential channel variables

Variables	(1) SCI	(2) MKT	(3) GI	(4) MOB	(5) GI
AI	0.142	0.122**	0.158***	0.046***	0.174***
	(0.108)	(0.056)	(0.035)	(0.016)	(0.035)
MKT			0.019		
			(0.042)		
MOB					0.052
					(0.067)
W × AI			-0.885***		-0.579*
			(0.321)		(0.318)

Variables	(1) SCI	(2) MKT	(3) GI	(4) MOB	(5) GI
W × MKT			-0.545**		
			(0.254)		
W × MOB					-4.982***
					(0.678)
Spatial rho			0.841***		0.823***
			(0.046)		(0.050)
Wald spatial terms			456.07***		454.59***
Control variables	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	3,539	3,540	3,540	3,540	3,540
R-squared	0.478	0.642	0.782	0.573	0.611
Number	295	295	295	295	295

Notes: Parentheses contain standard errors. * p < 0.10, ** p < 0.05, *** p < 0.01. Direct, indirect and total effects are distinguished from the Main and Wx coefficients where reported.

5.8 Supplementary Predictive Analysis

Table 8 reports contemporaneous model comparisons, with GI in Panel A and GII in Panel B; Figures 13 and 14 display the corresponding values. The augmented models add AI to economic controls. The evaluation does not establish performance on a documented temporal or geographic holdout, so these comparisons are treated as supplementary evidence rather than estimates of future-year accuracy.

For GI, Table 8 shows lower RMSE and MAE and higher R^2 after AI is added in all eight models. Linear Regression RMSE decreases from 0.9804 to 0.7950, and its R^2 rises from 0.6327 to 0.7586. Ridge has the largest reported proportional RMSE reduction, 19.15%. XGBoost has the lowest augmented-model RMSE, 0.5577, and an R^2 of 0.8812. The proportional improvement and the lowest final error are distinct comparisons; they need not favor the same model.

For GII, Table 8 likewise reports improvements across the eight models. Lasso RMSE decreases from 1.0574 to 0.8797, a reported reduction of 16.81%. Gradient Boosting has the lowest augmented-model RMSE, 0.6217, followed by SVR at 0.6252. Its R^2 increases from 0.8361 to 0.8644. These results indicate additional predictive information for green invention applications, not a direct measure of invention quality or environmental benefit.

Table 8 evaluates the addition of AI to economic controls. It does not assess whether AI supplies information beyond patent persistence or total inventive capacity. Section 5.10 addresses that question using expanding-window temporal validation. The two exercises differ in their targets, information sets and evaluation designs.

5.9 Supplementary Green-Specificity Diagnostics

Appendix Table A2 reports within-city regressions for green and nongreen patent activity. The AI coefficient is 0.0717 for green patents ($p = 0.398$) and 0.0255 for nongreen patents ($p = 0.791$). The coefficient for the green patent share is −0.0017 ($p = 0.500$), and the green–nongreen log-count contrast is 0.0462 ($p = 0.642$). None is statistically significant at conventional levels. These diagnostics do not establish a green-specific association or equivalence between outcomes. They also show that the positive direct association in the SDM is not reproduced as a significant coefficient in a non-spatial within-city specification.

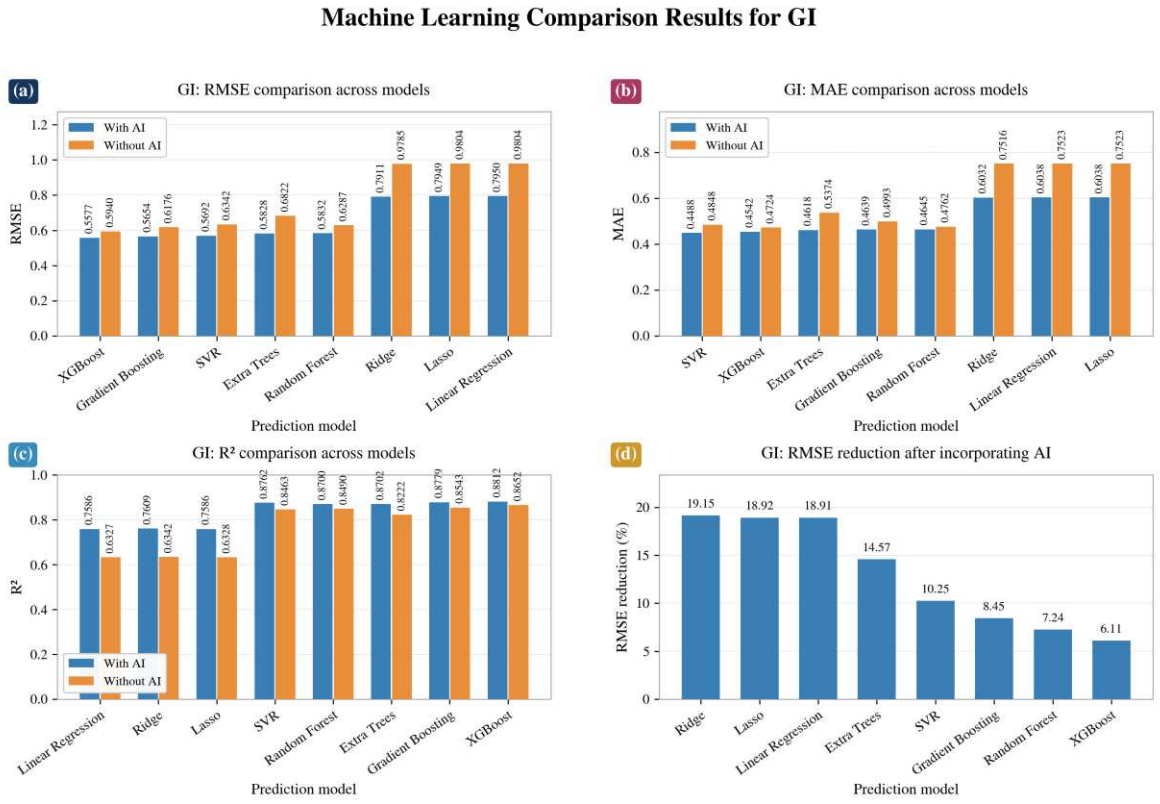


Fig. 13. Predictive comparisons for green innovation (GI).

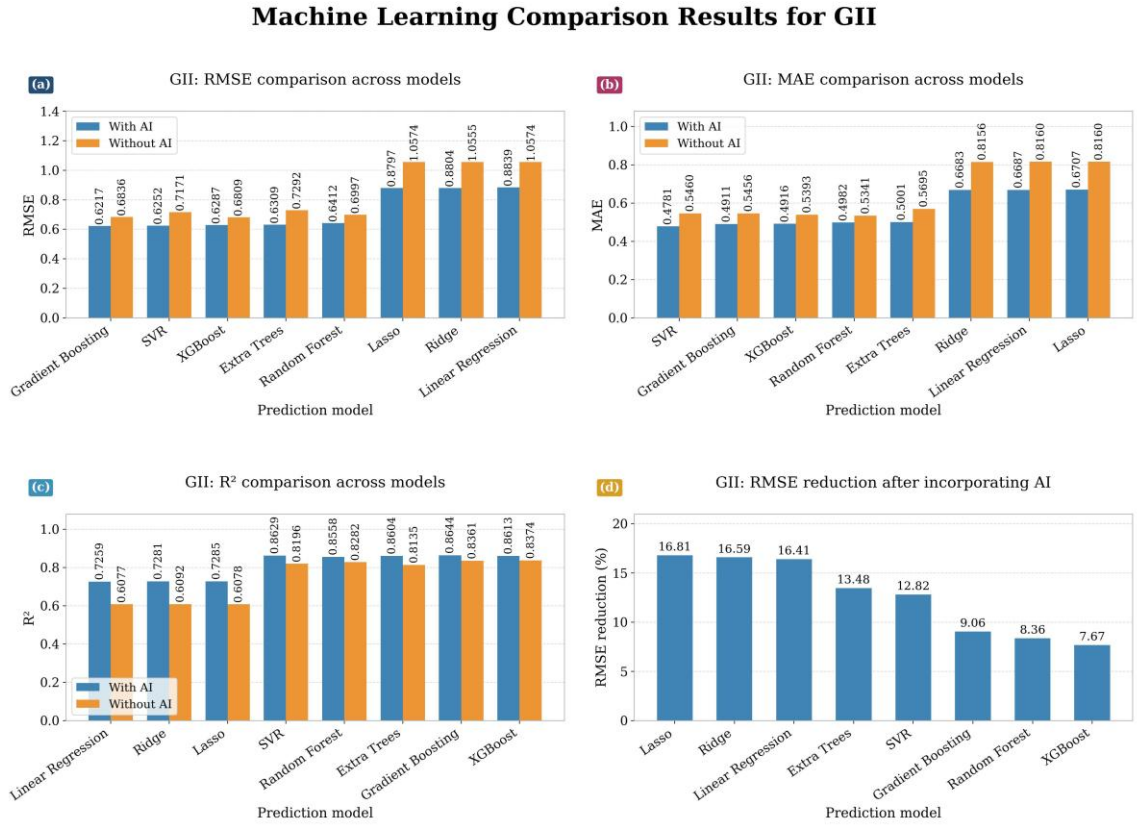


Fig. 14. Predictive comparisons for green invention innovation (GII).

Table 8. Predictive performance with and without AI

Panel A. Prediction performance for green innovation (GI)

Prediction model	Without AI RMSE	Without AI MAE	Without AI R-squared	With AI RMSE	With AI MAE	With AI R-squared	RMSE reduction (%)
Linear Regression	0.9804	0.7523	0.6327	0.7950	0.6038	0.7586	18.91
Ridge	0.9785	0.7516	0.6342	0.7911	0.6032	0.7609	19.15
Lasso	0.9804	0.7523	0.6328	0.7949	0.6038	0.7586	18.92
SVR	0.6342	0.4848	0.8463	0.5692	0.4488	0.8762	10.25
Random Forest	0.6287	0.4762	0.8490	0.5832	0.4645	0.8700	7.24
Extra Trees	0.6822	0.5374	0.8222	0.5828	0.4618	0.8702	14.57
Gradient Boosting	0.6176	0.4993	0.8543	0.5654	0.4639	0.8779	8.45
XGBoost	0.5940	0.4724	0.8652	0.5577	0.4542	0.8812	6.11

Panel B. Prediction performance for green invention innovation (GII)

Prediction model	Without AI RMSE	Without AI MAE	Without AI R-squared	With AI RMSE	With AI MAE	With AI R-squared	RMSE reduction (%)
Linear Regression	1.0574	0.8160	0.6077	0.8839	0.6687	0.7259	16.41
Ridge	1.0555	0.8156	0.6092	0.8804	0.6683	0.7281	16.59
Lasso	1.0574	0.8160	0.6078	0.8797	0.6707	0.7285	16.81
SVR	0.7171	0.5460	0.8196	0.6252	0.4781	0.8629	12.82
Random Forest	0.6997	0.5341	0.8282	0.6412	0.4982	0.8558	8.36
Extra Trees	0.7292	0.5695	0.8135	0.6309	0.5001	0.8604	13.48
Gradient Boosting	0.6836	0.5456	0.8361	0.6217	0.4911	0.8644	9.06
XGBoost	0.6809	0.5393	0.8374	0.6287	0.4916	0.8613	7.67

Notes: RMSE is root mean squared error, MAE is mean absolute error, and R-squared is the coefficient of determination. The baseline predictive specification includes the controls; the augmented specification adds AI. Reported RMSE reductions compare the two specifications within each model.

5.10 Forward Information Beyond Existing Patent Capabilities

Table 9 compares a persistence forecast, the fitted patent-history benchmark and the same benchmark augmented by AI. For green invention applications, pooled RMSE decreases from 0.3690 to 0.3637, a reduction of 1.43%. The city-block bootstrap interval for the reduction is 0.33–2.64%. MAE decreases from 0.2833 to 0.2782. Four annual folds improve, while 2024 deteriorates by 4.16%. The gain is therefore modest and outcome-dependent, rather than evidence of uniformly reliable prediction.

For all green applications, AI increases pooled RMSE from 0.4745 to 0.4886. Expressed as an RMSE reduction, the change is −2.95%, with an interval of −3.93 to −1.99%. Only two years improve. For green share, the pooled reduction is 0.47%, with an interval of −0.89 to 1.95%; MAE rises slightly from 0.01380 to 0.01390. Moreover, carrying last year's share forward gives RMSE 0.01580, below both fitted models. The share comparison therefore supplies no clear evidence that AI improves the information needed to anticipate environmental orientation.

The temporal comparisons and Table 8 answer different questions. Adding AI to economic controls and adding it to prior patent information use different benchmarks; contemporaneous and next-year outcomes also differ. Beyond patent history, the temporal exercise finds incremental information for invention applications but no general improvement across green outcomes. This is evidence about prediction in the historical panel, not the effect of increasing enterprise counts.

Table 9. One-year-ahead prediction beyond patent history

Outcome	N	Persistence RMSE	History RMSE	History + AI RMSE	RMSE reduction %	95% interval %
Green applications	1475	0.46381	0.47455	0.48857	-2.95	[-3.93, -1.99]
Green invention applications	1475	0.38557	0.36902	0.36374	1.43	[0.33, 2.64]
Green patent share	1475	0.01580	0.01845	0.01836	0.47	[-0.89, 1.95]

Notes: Target years are 2020–2024; all inputs other than the known calendar trend are lagged by one year. The history model includes the lagged outcome, total patent count and four economic controls. The augmented version adds lagged AI. All outcomes use 1,475 observations

from 295 cities. Positive reductions indicate improvement relative to the history model. Intervals use 2,000 city-block draws conditional on fitted predictions. Share is on a 0–1 scale. The temporal exercise has a different target and information set from Table 8.

6. Discussion

The results indicate that AI enterprise concentration and local green patent activity are positively associated in the spatial specifications, while the relationship across cities is less stable. For regional innovation monitoring, this makes the geographic scale of assessment consequential. A local enterprise base can accompany a local patent advantage without supporting the same interpretation for connected cities or for the environmental composition of their inventions.

The baseline negative indirect estimate changes with patent definition and spatial weights, and the eastern subgroup produces a positive estimate. These patterns are consistent with geographic variation in the balance between diffusion and resource competition, but the regressions do not identify either process. Regional appraisal should therefore report the sensitivity of intercity associations rather than translate a single estimated sign into a uniform claim about cluster benefits or losses.

The subgroup estimates also depend on how the geography is partitioned. The western region and the area northwest of the Hu Line contain different cities and produce different results. Economic subgroup point estimates suggest variation in the direct association, but no formal cross-group test establishes that the effects differ.

The supporting-condition regressions provide limited context for these patterns. AI is not significantly associated with fiscal science and technology expenditure. Marketization and communication scale have negative spatial-lag coefficients in their GI equations, while their local coefficients are not significant. These covariate relationships remain distinct from decomposed indirect effects and do not establish mediation.

The temporal results reveal a distinction within green inventive activity. Adding AI to a patent-history model improves pooled prediction of green invention applications by 1.43%, with improvement in four of five years, but worsens the pooled broad-green prediction by 2.95%. The green-share gain is uncertain, and persistence provides a stronger pooled share forecast. Thus, a firm count can carry modest information for one category without becoming a reliable indicator of portfolio greening. This observation is compatible with the broad applicability of AI, but it does not identify a technological or institutional mechanism.

For regional policy research, the contribution is the joint evaluation of an industrial indicator across geographic reach, environmental composition and monitoring horizons. The composition tests provide no clear evidence that the firm-count association is greener than general inventive activity. The temporal exercise identifies a modest invention-category gain beyond patent history, alongside deterioration for broader green counts. These comparisons specify where the indicator supplies additional information and where established patent measures remain necessary.

The interpretation remains bounded by the underlying panel. The AI count follows the dataset's enterprise classification, but its firm-level search rules and administrative status definition cannot be independently audited from the aggregate file. It does not measure adoption or environmental application. Patent applications do not establish realized environmental gains. Spatial estimates do not resolve reverse causality or time-varying confounding. The temporal exercise uses lagged values from the current historical dataset, without archived release vintages. Its bootstrap conditions on fitted predictions and five observed test years, leaving re-estimation and future common-shock uncertainty outside the interval. The invention gain is modest and is not retained in 2024. These limitations bound the indicator application described here.

7. Conclusions and Policy Recommendations

7.1 Research Conclusions

Using 295 Chinese cities over 2013–2024, the study evaluates the AI enterprise count through spatial association, environmental composition and one-year-ahead prediction. The baseline SDM direct effect is 0.139 and the indirect effect is −5.901. The direct association remains positive across the four spatial checks, whereas indirect estimates depend on patent definition, weights and region.

The negative indirect relationship is conditional. It becomes statistically insignificant with green invention patents and alternative spatial weights. Eastern estimates indicate positive indirect effects, while western estimates indicate negative effects. A local patent advantage is therefore not sufficient evidence of a uniform regional benefit or loss.

Within-city estimates do not establish a positive green-share association or a difference between green and nongreen patent responses. In the temporal exercise, AI reduces pooled next-year green-invention RMSE from 0.3690 to 0.3637, a gain of 1.43% beyond patent history and economic conditions. It instead raises broad-green RMSE from 0.4745 to 0.4886. The green-share gain of 0.47%

has a bootstrap interval containing zero. These comparisons are separate from the contemporaneous metrics in Table 8.

The substantive implication is that the environmental relevance of the urban AI enterprise base depends on the outcome and geographic scale being assessed. Local green patent association, portfolio greening and next-year monitoring usefulness are empirically distinct. Regional innovation assessments can make this distinction explicit by combining enterprise presence with patent composition, intercity relationships and historical benchmarks.

7.2 Policy Recommendations

Urban development agencies can report AI enterprise counts alongside green patent volume and green patent share, with each measure assigned a clear purpose. Enterprise presence describes the local industrial base, patent volume describes inventive activity, and the share addresses environmental orientation. The present evidence does not justify collapsing these measures into a single score of green development.

Regional cluster assessments should report both local and intercity estimates and their sensitivity to geographic definitions and spatial weights. The contrast between eastern and western estimates motivates attention to context, while the lack of a formal cross-group test limits claims about differences in effect size. Data on collaboration, inventor mobility and supply-chain links would help evaluate the processes behind these spatial patterns.

For next-year monitoring, the relevant benchmark is information already available from the city's patent history. AI counts show a modest incremental gain for green invention applications in this panel, but that gain is uneven across years and does not extend to all green applications. Agencies considering the indicator should evaluate it against persistence and historical-patent forecasts and retain years with deterioration in their assessment.

Environmental monitoring should also distinguish invention applications from technological quality, adoption and realized environmental performance. Further evidence on technological content and use is needed to assess whether AI-related capabilities are directed toward environmental problems. Neither an enterprise-count increase nor a better patent forecast establishes that outcome.

These recommendations concern the selection and evaluation of indicators in the studied urban panel. They guide monitoring design; decisions about incentives, subsidies or firm attraction require evidence on intervention effects and their distribution across cities.

References

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The environment and directed technical change. *American Economic Review, 102*(1), 131–166. https://doi.org/10.1257/aer.102.1.131

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy, 124*(1), 1–51. https://doi.org/10.1086/684581

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review, 86*(3), 630–640. https://asu.elsevierpure.com/en/publications/rampd-spillovers-and-the-geography-of-innovation-and-production/

Biggi, G., Iori, M., Mazzei, J., & Mina, A. (2025). Green intelligence: The AI content of green technologies. *Eurasian Business Review, 15*, 803–840. https://doi.org/10.1007/s40821-024-00288-1

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics, 65*(1), 83–108. https://doi.org/10.1016/0304-4076(94)01598-T

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Cicerone, G., Faggian, A., Montresor, S., & Rentocchini, F. (2023). Regional artificial intelligence and the geography of environmental technologies: Does local AI knowledge help regional green-tech specialization? Regional Studies, 57(2), 330–343. https://doi.org/10.1080/00343404.2022.2092610

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The impact of artificial intelligence on innovation* (NBER Working Paper No. 24449). National Bureau of Economic Research. https://doi.org/10.3386/w24449

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128–152. https://doi.org/10.2307/2393553

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis, 5*(1), 9–28. https://doi.org/10.1080/17421770903541772

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 28*(4), 1661–1707. https://www.nber.org/system/files/chapters/c8351/c8351.pdf

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science, 55*(3), 339–363. https://doi.org/10.1111/jors.12188

Haščič, I., & Migotto, M. (2015). *Measuring environmental innovation using patent data* (OECD Environment Working Papers No. 89). OECD Publishing. https://doi.org/10.1787/5js009kf48xw-en

He, X., Ruan, J., & Bian, C. (2024). Does artificial intelligence application promote green innovation in three major city clusters of the Yangtze River Economic Belt? Based on the perspectives of digital dividend and digital divide [In Chinese]. *Economic Geography, 44*(8), 137–147. https://doi.org/10.15957/j.cnki.jjdl.2024.08.016

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309–334. https://doi.org/10.1037/a0020761

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics, 108*(3), 577–598. https://doi.org/10.2307/2118401

Jaffe, A. B., Newell, R. G., & Stavins, R. N. (2002). Environmental policy and technological change. *Environmental and Resource Economics, 22*(1–2), 41–70. https://doi.org/10.1023/A:1015519401088

Johnstone, N., Haščič, I., & Popp, D. (2010). Renewable energy policies and technological innovation: Evidence based on patent counts. *Environmental and Resource Economics, 45*(1), 133–155. https://doi.org/10.1007/s10640-009-9329-1

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press. https://doi.org/10.1201/9781420064254

Liu, S. (2026). Artificial intelligence industrial agglomeration and innovation in key green technologies: Evidence from China. *Journal of Environmental Management, 412*, Article 130161. https://doi.org/10.1016/j.jenvman.2026.130161

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531–542. https://doi.org/10.2307/2298123

Mijit, R., Hu, Q., Xu, J., & Ma, G. (2025). Greening through AI? The impact of Artificial Intelligence Innovation and Development Pilot Zones on green innovation in China. *Energy Economics, 146*, Article 108507. https://doi.org/10.1016/j.eneco.2025.108507

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review, 92*(1), 160–180. https://doi.org/10.1257/000282802760015658

Porter, M. E., & van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *Journal of Economic Perspectives, 9*(4), 97–118. https://doi.org/10.1257/jep.9.4.97

Rennings, K. (2000). Redefining innovation—Eco-innovation research and the contribution from ecological economics. *Ecological Economics, 32*(2), 319–332. https://doi.org/10.1016/S0921-8009(99)00112-3

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289–310. https://doi.org/10.1214/10-STS330

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Wang, Z., Zhang, T., Ren, X., & Shi, Y. (2024). AI adoption rate and corporate green innovation efficiency: Evidence from Chinese energy companies. *Energy Economics, 132*, 107499. https://doi.org/10.1016/j.eneco.2024.107499

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629

Appendix

Table A1. Variance inflation factors by specification

Panel A. Baseline model

Variable	VIF	Tolerance
AI	2.174	0.460
PGDP	2.095	0.477
GOV	1.540	0.649
FIN	1.435	0.697
OPEN	1.197	0.835

Panel B. Model with additional controls

Variable	VIF	Tolerance
AI	3.070	0.326
PGDP	2.559	0.391
GOV	1.748	0.572
FIN	1.498	0.668
OPEN	1.531	0.653
DENS	1.616	0.619
IND	1.822	0.549

Panel C. Specification including MKT

Variable	VIF	Tolerance
AI	3.732	0.268
MKT	3.219	0.311
PGDP	2.136	0.468
GOV	1.906	0.525
FIN	1.443	0.693
OPEN	1.223	0.817

Panel D. Specification including MOB

Variable	VIF	Tolerance
AI	4.301	0.233
MOB	3.111	0.321
PGDP	2.629	0.380
GOV	1.921	0.521
FIN	1.442	0.693
OPEN	1.274	0.785

Table A2. Non-spatial within-city diagnostics of green specificity

Variable	Green patents	Nongreen patents	Green patent share	Log-count contrast
AI	0.0717	0.0255	-0.0017	0.0462
	(0.0847)	(0.0963)	(0.0026)	(0.0992)
PGDP	0.1671	0.2305**	-0.0034	-0.0634
	(0.1027)	(0.1137)	(0.0097)	(0.1293)
GOV	-0.1813	0.7780**	-0.0416	-0.9593*
	(0.3416)	(0.3906)	(0.0370)	(0.5218)
FIN	-0.0749	-0.1362	-0.0009	0.0613
	(0.0653)	(0.0869)	(0.0034)	(0.0376)
OPEN	0.0637	-0.0853	0.0053	0.1491
	(0.1549)	(0.1506)	(0.0045)	(0.0911)
AI p-value	0.3981	0.7914	0.5000	0.6420
Observations	3540	3540	3538	3540
Cities	295	295	295	295
Within R^2	0.0169	0.0400	0.0045	0.0216
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
City-clustered SE	Yes	Yes	Yes	Yes

Notes: Green and nongreen patents use ln(1 + applications). Green patent share is green applications divided by all applications, on a 0–1 scale; the two zero-denominator city-years are excluded from that equation. The log-count contrast equals ln(1 + green applications) minus ln(1 + nongreen applications). All regressions include the same economic controls and city and year fixed effects. Parentheses contain city-clustered standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference uses 294 degrees of freedom. The contrast is a non-spatial slope comparison, not a spatial effect.

Table A3. Annual folds of the one-year-ahead validation

Outcome	Year	N	Persistence RMSE	History RMSE	History + AI RMSE	Reduction %
Green applications	2020	295	0.37233	0.27549	0.26286	4.58
Green applications	2021	295	0.26191	0.40300	0.41813	-3.75
Green applications	2022	295	0.24625	0.28482	0.30337	-6.51
Green applications	2023	295	0.21702	0.23282	0.21929	5.81
Green applications	2024	295	0.87214	0.86737	0.89968	-3.73
Green invention applications	2020	295	0.45044	0.36878	0.35872	2.73
Green invention applications	2021	295	0.38018	0.40954	0.40055	2.20
Green invention applications	2022	295	0.34162	0.33246	0.32485	2.29
Green invention applications	2023	295	0.32510	0.33427	0.31441	5.94
Green invention applications	2024	295	0.41653	0.39356	0.40992	-4.16
Green patent share	2020	295	0.01671	0.01694	0.01681	0.78
Green patent share	2021	295	0.01572	0.01692	0.01683	0.56
Green patent share	2022	295	0.01297	0.01764	0.01782	-1.01
Green patent share	2023	295	0.01161	0.01572	0.01615	-2.77
Green patent share	2024	295	0.02047	0.02387	0.02327	2.53

Notes: Each annual fold trains only on earlier target years and selects its ridge penalty on the most recent preceding target year. Predictions concern the same 295 cities with predictors dated one year earlier. The pooled reductions in Table 9 are calculated from all prediction errors, not the arithmetic mean of annual percentages. Negative reductions are retained and indicate deterioration. Share forecasts are clipped to 0–1.

Table A4. Enterprise-count definition and sample accounting

Component	Count or definition	Processing and interpretation
Enterprise measure	ai_firm; AI = ln(1 + ai_firm)	Recorded AI-related classification; no new enterprise-level selection. Original keywords, legal status and deduplication rules are not recoverable from the aggregate file.
Initial panel	297 cities; 3,564 records	2013–2024; unique city-identifier/year keys.
Complete-coverage restriction	2 cities; 24 records removed	Sansha: 2 missing outcome years and 10 nonmissing years; Danzhou: 12 missing outcome years. Both complete city histories are excluded.
Baseline and count diagnostics	295 cities; 3,540 records	No missing baseline outcome, AI, economic-control or coordinate values. Recorded zero counts retained; no imputation in the supplementary regressions.
Patent-share diagnostic	295 cities; 3,538 records	Rikaze 2014 and Nagqu 2014 omitted because total applications equal zero. No zero replacement of undefined shares.
SCI supporting-condition regression	295 cities; 3,539 records	Huludao 2023 omitted because SCI is missing. SCI is not included in the retained GI outcome equations.
MKT and MOB regressions	295 cities; 3,540 records	Sample sizes as reported in Table 7.
One-year-ahead validation	295 cities; 1,475 tests per outcome	Each outcome evaluated on 2020–2024, 295 records/year. Predictor values lagged by one year. Training uses earlier target years; model-specific complete inputs required.
Source attribution	City indicators: China City Database; green patents: CNRDS	Enterprise-related inputs: CSMAR (Guotai An) and Shenzhen Stock Exchange disclosures. Specific modules, download dates and firm-level selection rules are not available in the retained analytical file.