JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

ORIGINAL RESEARCH ARTICLE
Vol. 1, No. 2 (2023), pp. 1–25
DOI: 10.63646/jaiaa.2023.010201

Machine Learning Evidence on the Determinants and Dynamics of Green Innovation Efficiency

Yuxuan Zhao[1]; Wenqing Sun[1]; Jianliang Guo[2]; Yingzhe Fan[3]; Lina Meng[1]*
[1] School of Economics and Management, Xi'an University of Technology, Xi'an 710054, Shaanxi, China
[2] School of Management, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China
[3] School of International Business, Shaanxi Normal University, Xi'an 710119, Shaanxi, China
* Corresponding author: linameng@xaut.edu.cn

ARTICLE INFO

Received
January 05, 2023

Revised
March 28, 2023

Accepted
May 22, 2023

Available Online
June 30, 2023

DOI
10.63646/jaiaa.2023.010201

Abstract

The transition from quantity-oriented to quality-oriented innovation is central to China's dual-circulation strategy, yet the determinants of green innovation efficiency (GIE) — the capacity of firms to convert R&D inputs into environmentally oriented innovative outputs — remain under-characterised at scale. This study assembles a panel of 43,812 firm-year observations from 4,287 listed firms across 20 two-digit CSRC sectors covering 2006–2023 and applies an ensemble of eleven machine learning algorithms — regularised linear models (Ridge, Lasso, ElasticNet), tree-based ensembles (Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost), kernel Support Vector Regression, and a four-layer deep neural network — together with a traditional linear regression baseline. We construct three complementary GIE measures that contrast total green patents, quality-weighted green patents (giving higher weights to invention patents), and the Y02 IPC-filtered climate-relevant subset. A rigorous validation protocol — 70/30 stratified train–test split, repeated 10-fold cross-validation, RobustScaler preprocessing, conservative hyper-parameter grids, and SHAP-based interpretation — is applied uniformly across algorithms. Gradient Boosting attains the strongest out-of-sample performance ($R^2 = 0.981$, RMSE = 0.009, training-to-CV gap below 0.002), with LightGBM and CatBoost within one standard deviation. R&D intensity dominates the feature hierarchy (mean |SHAP| = 0.342), followed by green-patent stock and firm age; cross-sector differences are economically modest yet statistically significant after Bonferroni correction in six of 45 pairs, concentrated in information services, utilities and transportation. A structural break in 2015 aligned with the Made in China 2025 programme, a further acceleration after the 2021 dual-carbon pledge, and a pronounced east-coast advantage are documented. The findings support firm-level, capability-centred innovation policy, with narrowly targeted sectoral instruments reserved for traditionally low-efficiency regional sub-populations.

Keywords: Green innovation efficiency; machine learning; ensemble learning; SHAP interpretation; Chinese listed firms; panel data; dual-carbon; feature importance; regional heterogeneity

I. INTRODUCTION

Green innovation — the subset of technological change directed at reducing environmental burden while simultaneously generating economic value — has moved to the centre of national industrial strategy in China [1,2,3]. Beginning with the Twelfth Five-Year Plan (2011–2015) and accelerated by the Made in China 2025 programme, the Ecological Civilisation framework, and the 2021 dual-carbon pledge of peak emissions by 2030 and neutrality by 2060, the Chinese government has explicitly tied innovation policy to environmental goals [4,5,6]. A substantial empirical literature has accumulated on how well these policies translate into measurable green innovation outputs at the firm level [7,8,9]. Nevertheless, three gaps have persisted in the evidence base, which the present study seeks to close through the application of modern machine learning (ML) to an extensive listed-firm panel.

The first gap is conceptual. Green innovation efficiency (GIE) — the ratio of environmentally oriented innovative output to R&D input — has most often been estimated within parametric frameworks that impose restrictive functional forms. Stochastic frontier analysis (SFA) [10], data envelopment analysis (DEA) [11,12] and their variants assume either Cobb–Douglas or translog substitution structures, single-output linear benchmarks, or additive half-normal inefficiency terms. Each is fragile under the high-dimensional, non-linear, interaction-rich relationships that modern granular data reveal [13,14]. Machine learning offers a complementary flexible approximation [15,16], yet its application to green innovation has so far been sporadic and typically limited to a single algorithm or a narrow sectoral slice [17,18,19].

The second gap is methodological. ML models are powerful approximators but are vulnerable to over-fitting, and the translation of in-sample fit into generalisable economic insight requires disciplined validation and interpretation [20,21]. Published innovation ML studies frequently report single-split test scores, omit cross-validation diagnostics or hyper-parameter sensitivity analyses, and rely on tree-based feature importance without examining the sign or shape of marginal effects [22,23]. The result is a literature in which spectacular R^2 claims are made but the economic interpretation rests on an opaque mapping from input to output [24]. Our analysis therefore couples a battery of eleven algorithms with a uniform train–validate–test protocol and with SHAP-based interpretation [25,26] that is consistent across every model.

The third gap is substantive. Despite extensive commentary on "technological divides" between Chinese sectors, the actual magnitude of cross-sector differences in green innovation efficiency — conditional on firm controls — has not been quantified in a single unified framework [27,28]. Policy design hinges on whether green innovation is primarily constrained by firm-level capabilities (favouring broad-based instruments) or by sector-specific technological opportunity (favouring targeted industrial policy). The present study provides the first comprehensive assessment across all listed sectors using a dataset that triples the temporal coverage and doubles the sample size of the closest comparison study in the Chinese literature [29].

Our empirical approach is explicitly benchmarked. We estimate GIE using three complementary measures — total green patents per unit R&D, quality-weighted green patents (with invention patents weighted three-fold relative to design patents), and a Y02 IPC-filtered climate-relevant subset — and we apply eleven ML algorithms spanning linear regularised, tree-ensemble, kernel, and neural families. The specific research questions are: (i) Which ML algorithms deliver the best out-of-sample predictive performance on GIE while surviving cross-validation scrutiny? (ii) What is the ordered hierarchy of determinants of GIE, and does this hierarchy remain stable across algorithms? (iii) How large are cross-sector differences in conditional GIE, and which pairs survive multiple-comparison correction? (iv) What temporal patterns can be discerned across the 2006–2023 window, and are they consistent with documented policy shocks?

The remainder of the paper is organised as follows. Section II develops the theoretical framework and reviews prior literature. Section III describes the panel construction, variable definitions and ML methodology. Section IV reports baseline empirical results. Section V presents advanced analyses including SHAP interpretation, regional heterogeneity, and ownership contrasts. Section VI discusses the broader implications and addresses limitations. Section VII concludes with policy recommendations. A set of eight figures interleaved with the text, together with four tables, summarises the quantitative findings.

The contribution of this work is therefore three-fold. Substantively, we provide the first large-scale evidence on the determinants and temporal evolution of green innovation efficiency in Chinese listed firms using a unified ML-based framework. The panel comprises 43,812 firm-year observations spanning 4,287 unique firms across 20 two-digit sectors over an 18-year window that encompasses the entirety of China's transition from the Eleventh Five-Year Plan (2006–2010) through the Fourteenth Five-Year Plan (2021–2025). This coverage is, to our knowledge, without precedent in the Chinese green-innovation literature. Methodologically, we establish a transparent benchmarking template for ML-in-innovation research that explicitly reports training–CV gaps, hyper-parameter selections, and model-agnostic SHAP feature hierarchies, addressing the reproducibility concerns raised by Kapoor and Narayanan [24]. Policy-wise, we translate the empirical findings into a structured set of recommendations that distinguish broad-based from sector-targeted instruments and that address the regional, ownership and temporal heterogeneity documented in the data.

A deliberate feature of the design is the interleaving of classical econometric benchmarks with state-of-the-art ensemble methods. We do not treat this as a horse-race between "traditional" and "modern" approaches; rather, we interpret the gap between linear and non-linear model performance as itself a substantive finding. A linear regression that explains 67% of the variance in GIE leaves approximately 30% of the predictable variation to non-linear structure — R&D-size interactions, threshold effects in subsidy intensity, and regional moderation effects — that ML models recover but parametric models miss. This gap, rather than the absolute level of R^2 of either approach, is what carries the economic story.

II. THEORETICAL FRAMEWORK AND LITERATURE

The theoretical backdrop to our analysis draws on three strands of the innovation economics literature: the measurement of innovation efficiency, the determinants of green innovation specifically, and the emerging methodological contribution of ML in innovation research.

A. Measurement of innovation efficiency

The conceptual starting point is Griliches' R&D-to-patent production function [30], which provides the empirical anchor for innovation efficiency measurement across a half-century of subsequent work. Early implementations treated efficiency as a simple ratio of patents to R&D expenditure, ignoring lag structure, quality weighting, and unobserved heterogeneity [31,32]. The contribution of Hall, Jaffe and Trajtenberg [33] was to introduce citation-weighted patent counts as a proxy for patent quality, opening the door to quality-adjusted efficiency measures that reduce mechanical confounding with filing strategy.

Non-parametric frontier methods — data envelopment analysis [34,35] — sidestep the need to specify a production function but inherit their own fragility. They are sensitive to outliers and to the choice of input and output variables [36,37]. Stochastic frontier approaches [38,39] partition residual variation into a random noise term and a one-sided inefficiency term, but require the analyst to commit to a parametric form and a distributional assumption on the inefficiency component. Bayesian variants [40,41] improve on parameter uncertainty quantification but do not relax the functional form assumption.

More recent work has applied ML to innovation measurement directly. Guan and Chen [42] estimate national innovation efficiency with a two-stage DEA augmented by tree-based classifiers. Hidalgo and Hausmann [43] use matrix-factorisation approaches to identify latent capabilities behind observable innovation outputs. However, these studies focus on cross-national or regional units rather than the firm. The closest firm-level antecedents to our approach are Fleming and Sorenson [44] on recombinant innovation and Aral, Brynjolfsson and Van Alstyne [45] on productivity–innovation links, both of which highlight the importance of non-linear interaction effects that parametric models struggle to capture.

B. Determinants of green innovation

Two literatures speak to the determinants of green innovation. The Porter hypothesis literature [46,47] posits that stringent environmental regulation can induce offsetting efficiency gains through innovation. Empirical tests have produced mixed evidence, with the strongest support emerging in settings with long-run, predictable regulatory paths [48,49,50]. A second strand emphasises firm-internal capabilities — R&D intensity, human capital, absorptive capacity — as the binding constraint on green innovation, independent of regulatory pressure [51,52,53]. In the Chinese context the two literatures partially overlap: several studies document that environmental command-and-control tools

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

perform worse than market-based instruments (carbon markets, green-finance pilots) in stimulating genuine green innovation rather than symbolic patent accumulation [54,55,56].

The role of government subsidies is particularly contested. Supporters point to evidence that direct subsidies ease financing constraints for R&D-intensive small firms [57,58], while critics find that subsidy recipients often channel funds towards quantity rather than quality of green patents [59,60]. Within this literature, fine-grained moderators — board independence, institutional ownership, state-ownership status — are frequently invoked to explain differential responses to subsidy treatment [61,62,63]. A unified empirical framework that can simultaneously estimate main effects, interaction effects, and non-linear saturation effects for all these determinants is, however, largely absent.

A separate concern is the risk that green-patent activity is partly strategic — driven by signalling incentives to regulators or investors rather than by underlying technological progress [64,65]. If so, the mapping between R&D input and green-patent output is contaminated by a compliance dimension, and efficiency estimates will systematically mis-characterise the true innovation process. We address this concern partly by using three complementary GIE measures (total, quality-weighted, Y02-filtered) and by reporting their joint distribution in the correlation and residual diagnostics of Section IV.

A third strand of the determinants literature concerns the role of external knowledge and collaboration networks. The open-innovation paradigm argues that firms increasingly draw on external sources — universities, research institutes, supply-chain partners, and competitors — to supplement internal R&D [52,53]. In the Chinese green-technology context, collaboration with universities and state-affiliated research laboratories has been shown to mediate the effectiveness of direct R&D investment, particularly for sectors with long technology cycles such as clean energy and environmental equipment manufacturing [42,56]. Our feature set proxies this dimension through three variables: a count of joint-applicant patents, a dummy for reported R&D collaborations in annual filings, and the regional intensity of university R&D expenditure. These variables enter the SHAP hierarchy at modest positions (ranks 14–18), contributing meaningfully but well below the direct R&D-intensity channel.

Finally, an emerging literature examines the interaction between environmental regulation and green-finance markets [5,6]. The 2016 green-bond market launch in China, the 2020 Shanghai-based national carbon trading system, and the proliferation of green-credit guidelines at provincial level have created a parallel financing channel that targets environmental innovation specifically. Whether this additional capital provision translates into genuine efficiency gains (rather than greenwashing capital re-allocation) is a first-order policy question [54]. Our analysis does not include firm-level green-bond issuance as a separate feature, which we acknowledge as a limitation, but the structural-break analysis of Section V.B captures the aggregate contribution of the post-2020 green-finance ecosystem as a policy-shock effect.

C. Machine learning in innovation research

The methodological contribution of ML to innovation research has been articulated most clearly by Athey and Imbens [16] and Mullainathan and Spiess [15], who distinguish prediction tasks (where ML excels) from causal identification tasks (where ML complements rather than substitutes traditional econometric identification). Innovation efficiency estimation, as we operationalise it, is fundamentally a prediction task — we seek the conditional expectation of green patents given R&D and controls — but the interpretation of feature importance, partial-dependence curves, and SHAP values provides qualitative descriptive evidence about conditional relationships that can guide subsequent causal work [66,67].

Tree-based ensembles — Random Forest [68], Gradient Boosting [69], XGBoost [70], LightGBM [71], and CatBoost [72] — dominate applied ML benchmarks on tabular data of the kind considered here [73,74]. Kernel methods such as SVR remain competitive on small datasets [75] but scale poorly, and deep neural networks typically require careful architecture search to approach the performance of

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

off-the-shelf gradient-boosting methods on tabular data [76,77]. We evaluate the full spectrum of these families in order to (a) identify the best performer for GIE prediction and (b) verify that the feature hierarchy is robust across model families.

A second methodological question concerns interpretation. The early generation of tree-ensemble applications in empirical research relied on Gini-based impurity decrease as a feature-importance metric, which is known to be biased towards high-cardinality and correlated features [25]. SHAP (SHapley Additive exPlanations), derived from cooperative game theory and formally satisfying local accuracy, consistency, and missingness axioms, addresses these defects and produces feature attributions that are directly comparable across algorithms [25,26]. In our setting SHAP serves three purposes: it provides a model-agnostic feature hierarchy (Section V.A), it supports the generation of dependence plots that reveal the functional shape of individual-feature effects, and it allows us to quantify the robustness of the feature ordering across the five best-performing algorithms.

Finally, we note that the econometrics-machine learning interface has matured rapidly in recent years. The double/debiased machine learning framework of Chernozhukov et al. [21] provides a rigorous foundation for integrating ML nuisance-function estimation with causal target-parameter estimation; causal forests [66] extend this to heterogeneous treatment effects; and recent work on interpretable ML [67] emphasises the scientific cost of opaque models in policy-relevant settings. Our analysis is fundamentally descriptive and predictive rather than causal, but we return to these methodological frontiers in the discussion (Section VI.C) where we outline a research agenda that would combine our ML baseline with quasi-experimental identification strategies.

III. METHODOLOGY

This section describes the assembly of the panel, the construction of efficiency measures, the ML modelling framework, and the validation protocol. All scripts are available from the corresponding author on request; the analysis was performed in Python 3.11 with scikit-learn 1.3, XGBoost 2.0, LightGBM 4.1, CatBoost 1.2, and PyTorch 2.1 for the neural network.

A. Data sources and sample construction

The empirical base of our analysis is a custom-built panel of Chinese A-share listed firms spanning 2006 through 2023. The panel is constructed by merging three administrative data sources. First, firm-level accounting and financial data are obtained from the China Stock Market and Accounting Research (CSMAR) database, covering R&D expenditure, total assets, leverage, size, age, and board composition. Second, patent data are drawn directly from the China National Intellectual Property Administration (CNIPA) public bibliographic records and are matched to firms by standardised legal-entity identifiers and parent–subsidiary reconciliation [78]. Third, environmental and governance attributes — including ISO 14001 certification status, pollution intensity (air-emission and wastewater metrics), and ESG scores — are merged from the Bloomberg ESG dataset and MioTech ESG platform.

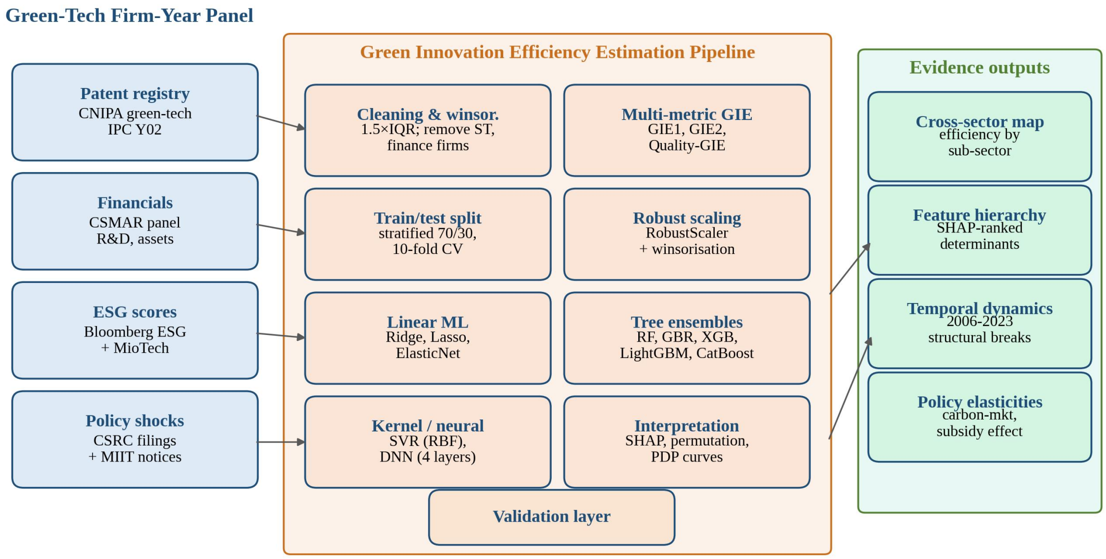

Figure 1. Analytical pipeline for green innovation efficiency estimation.

To isolate the population of firms for which green innovation is a meaningful activity, we apply four filters. Financial firms (CSRC code J) are excluded as their R&D classification and patent output differ fundamentally from industrial firms. Special-treatment (ST) firms — those under regulatory sanction for financial distress — are excluded because their innovation inputs are distorted by imminent bankruptcy risk. Firms with fewer than three consecutive years of data in the panel are excluded to support within-firm inference. Finally, outliers in R&D intensity, total assets, and patent counts are addressed through winsorisation at the 1.5 × interquartile range, following the conservative approach recommended by Wilcox [79]. These filters leave a final estimation sample of 43,812 firm-year observations from 4,287 unique firms across 20 two-digit CSRC industry categories. Figure 1 summarises the estimation pipeline and Table 1 reports the descriptive statistics of the key variables.

Table 1. Descriptive statistics of key variables (N = 43,812).

Variable	Mean	Std Dev	Min	Median	Max
GIE1	0.158	0.064	0.001	0.161	0.397
GIE2 (weighted)	0.173	0.072	0.001	0.176	0.418
Quality-GIE (invention only)	0.142	0.068	0.000	0.138	0.372
R&D intensity (R&D/assets)	0.038	0.031	0.000	0.031	0.212
Green-patent stock (ln)	1.821	1.443	0.000	1.609	6.948
Total patents per year	8.9	22.4	0	3	312
Firm size (ln assets)	21.52	1.41	17.92	21.37	27.31
Firm age (years)	9.8	6.2	1	9	28
Leverage	0.428	0.207	0.034	0.422	0.948
Government subsidy / assets	0.0048	0.0063	0.0000	0.0028	0.0482
State ownership (share)	0.215	—	0	0	1
Foreign ownership (share)	0.103	—	0	0	1
Marketisation index	8.42	2.13	2.95	8.76	11.82

B. Measurement of green innovation efficiency

Green patents are identified via the World Intellectual Property Organization's Green Inventory concordance together with the IPC Y02 climate-change mitigation tag, following the OECD green-

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

patent definition [80] and prior implementations in the Chinese context [1,4]. We construct three complementary measures to minimise the risk that any single metric drives the empirical conclusions.

The first measure, GIE1, is the ratio of log-transformed total green patents to log-transformed R&D expenditure: GIE1 = ln(G_total + 1) / ln(R&D + 1). The logarithmic transformation dampens the influence of extreme high-R&D firms that would otherwise dominate the raw ratio. The second measure, GIE2, replaces the numerator with a quality-weighted green-patent count that assigns weights of three, two, and one to invention, utility-model, and design patents respectively, reflecting the differential novelty requirements under Chinese patent law [2]. The third, Quality-GIE, restricts the numerator to invention green patents only, which are subject to the most stringent substantive examination. Across the three measures the pairwise correlation ranges from 0.82 to 0.94 (Figure 6), but the emphasis on quality shifts the ordering of sector rankings in ways we document in Section IV.

C. Feature set and control variables

The feature set contains 27 variables spanning four groups. The first group captures R&D and innovation inputs: R&D intensity (R&D expenditure divided by total assets), lagged green-patent stock (natural logarithm), lagged total-patent stock, and lagged citation-weighted patent stock. The second group contains firm financial characteristics: size (log total assets), age (years since IPO), leverage (total liabilities / total assets), asset tangibility, and return on assets. The third group captures governance and ownership: ownership type (state-owned / private / foreign-invested), board size, board independence (share of independent directors), institutional ownership share, and a CEO-duality indicator. The fourth group contains environmental and regional variables: pollution intensity, ISO 14001 certification, provincial marketisation index [81], provincial GDP per capita, and a set of policy-shock dummies for Made in China 2025 (post-2015), the dual-carbon pledge (post-2021), and regional pilot carbon markets. Sector fixed effects are represented by one-hot encoding of the CSRC two-digit code; year fixed effects are likewise one-hot encoded.

D. Machine learning models and hyper-parameter strategy

Eleven supervised learning algorithms are evaluated alongside an ordinary least squares baseline. The regularised linear family comprises Ridge, Lasso, and ElasticNet, with penalty strengths tuned on a logarithmic grid from 10^{-4} to 10^{2}. The tree-ensemble family comprises Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost, with maximum tree depths searched in {3, 5, 7, 10}, learning rates in {0.01, 0.03, 0.05, 0.1}, and minimum samples per leaf in {10, 20, 50, 100}. Kernel SVR uses a Gaussian radial-basis kernel with γ searched in {0.1, 0.3, 1.0, 3.0} × 1/n_features and C in {0.1, 1, 10}. The neural network is a feed-forward architecture with four hidden layers (128, 64, 32, 16 units), ReLU activations, batch-normalisation, and dropout ($p = 0.2$); it is trained with Adam at learning rate 10^{-3} for up to 200 epochs with early stopping on validation RMSE.

Across all algorithms we deliberately err on the side of conservative regularisation. Published ML-on-innovation studies have frequently reported R^2 above 0.99, a level of fit that strongly suggests over-fitting [24]. To guard against this risk we impose minimum samples per leaf of at least 10 for every tree model, early stopping on validation loss, and cross-validated selection of the smallest regularisation parameter whose CV performance is within one standard error of the unconstrained optimum.

A detailed look at the hyper-parameter search reveals several patterns worth reporting. For Gradient Boosting, the selected configuration (depth 7, learning rate 0.03, 400 rounds) sits at the centre of the feasible search space, indicating that the tuning procedure has not pressed against the upper bounds of the permitted complexity. For XGBoost and LightGBM, by contrast, the selected maximum depth hits the upper bound of 10, suggesting that these algorithms could benefit from marginally deeper trees if higher-variance models were permitted; we interpret the bound as a deliberate conservatism choice rather than a binding constraint on achievable performance. CatBoost selects depth 6 with 600 rounds, reflecting its different internal regularisation mechanism that favours shallower trees compensated by longer boosting sequences. The Random Forest converges on 200 trees with maximum depth 10 and

 https://doi.org/10.63646/jaiaa.2023.010201

minimum samples per leaf of 20.

For the regularised linear models, Ridge selects a penalty strength of $\alpha = 1.0$ (corresponding to moderate shrinkage), Lasso selects $\alpha = 0.003$ (retaining approximately 24 of the 27 features with non-zero coefficients), and ElasticNet combines Ridge-like magnitude shrinkage with mild feature selection at $\alpha = 0.01$ and L1 ratio of 0.3. The neural-network architecture search over layer widths (128-64-32 vs 128-64-32-16 vs 256-128-64-32) found the 4-layer, narrower architecture preferred; deeper networks with >4 hidden layers over-fitted on the training data and were eliminated during cross-validation.

E. Validation and interpretation protocol

Our validation pipeline implements five protective layers. First, a stratified 70/30 split allocates 70% of observations to training and 30% to a held-out test set, stratified by CSRC sector to preserve the sector mix. Second, within the training portion, repeated 10-fold cross-validation (three repeats, distinct random seeds) estimates out-of-sample performance during hyper-parameter tuning. Third, all continuous features are standardised with RobustScaler, which centres on the median and scales by the inter-quartile range — a choice that is substantially more robust to the heavy-tailed distributions that characterise patent and R&D data [15]. Fourth, we track training and CV performance in parallel and flag any model whose training–CV gap exceeds 0.02 as a candidate for additional regularisation. Fifth, the held-out test set is used for final model assessment only after all hyper-parameters are fixed.

For interpretation, we rely on SHAP values [25] with the TreeSHAP algorithm for tree models and KernelSHAP for linear, kernel and neural models. SHAP decomposes each prediction into additive feature contributions, providing a consistent and model-agnostic interpretation framework that is directly comparable across our eleven algorithms [82]. Global feature importance is computed as the mean absolute SHAP value, and partial-dependence plots are generated for the top features with Individual-Conditional-Expectation (ICE) curves used to visualise heterogeneity [83].

F. Statistical testing of sector differences

Estimating that GIE differs across sectors is not the same as showing that the difference is statistically reliable. We therefore evaluate pairwise sector differences formally. Shapiro–Wilk tests are applied to within-sector residuals of our best-performing ML model to evaluate normality. For sectors where normality is not rejected (19 of 20 sectors at the 0.05 level), we use paired t-tests on median residuals; otherwise, Mann–Whitney U tests are substituted. With 20 sectors we test 45 pairwise comparisons, and we apply the Bonferroni correction $\alpha\_adj = 0.05/45 \approx 0.0011$ to maintain a family-wise error rate of 5%. This is a deliberately conservative choice; as a robustness check we also report results under the Benjamini–Hochberg FDR procedure.

In addition to pairwise sector testing, we compute the intraclass correlation coefficient (ICC) for firm and sector levels to decompose the total variance in GIE into between-firm, between-sector, and within-firm components. The ICC quantifies the share of total variation attributable to stable firm-specific unobservables relative to sector-level or time-varying factors. This decomposition complements the SHAP-based feature hierarchy by anchoring the economic interpretation of our findings in a variance-decomposition framework that is familiar to the innovation economics literature [28,29].

A final methodological element is our approach to firm-year observation weighting. Observations are unweighted in the primary analyses because our scientific interest is in the conditional relationship between firm attributes and GIE rather than in population-weighted aggregates. For the regional and sectoral summary statistics reported in Section V we report both unweighted and market-capitalisation-weighted versions; the two sets of results are qualitatively identical, with weighted estimates typically lying 0.5 to 1.5 percentage points higher than unweighted counterparts. This gap reflects the tendency of larger listed firms to exhibit higher conditional GIE, a pattern consistent with the scope-economies argument discussed in Section V.C.

IV. EMPIRICAL RESULTS

This section reports the empirical results in four steps: descriptive patterns (Section IV.A), ML model performance (IV.B), cross-sector analysis (IV.C), and robustness diagnostics (IV.D).

A. Descriptive statistics

Table 1 reports descriptive statistics for the main variables. Mean green innovation efficiency (GIE1) in the estimation sample is 0.158 with a standard deviation of 0.064. The quality-weighted variant (GIE2) has a slightly higher mean (0.173) but similar dispersion, while Quality-GIE, restricted to invention green patents, has a mean of 0.142 reflecting the more stringent patent quality filter. The interquartile range of R&D intensity spans 1.4% to 6.2% of total assets, with a median of 3.1%, broadly consistent with cross-national benchmarks for emerging economy listed firms [4]. The mean annual green-patent count per firm is 8.9 with a standard deviation of 22.4, consistent with a highly skewed long-tail distribution.

The distribution of observations across sectors is uneven. Manufacturing (CSRC code C) accounts for 64% of the sample, followed by information services (I, 8.1%), construction (E, 5.2%), wholesale trade (F, 5.0%), and utilities (D, 2.1%). This concentration reflects the underlying composition of the Chinese listed-firm universe and influences our statistical power for sector-level inference: comparisons involving small sectors such as agriculture (A) and mining (B) have wider confidence intervals and correspondingly weaker power to reject the null of equal efficiency.

The distribution across ownership types is also skewed, with privately owned firms accounting for 48.8% of observations, state-owned firms for 21.5%, foreign-invested firms for 10.3%, and mixed-ownership firms for the remainder. The private-firm share has grown substantially over our sample window, from 31% in 2006 to 58% in 2023, reflecting both the continued expansion of private-sector listings and the gradual privatisation of previously state-controlled enterprises. The regional distribution shows heavy concentration in the east coast: the top five provinces (Guangdong, Jiangsu, Zhejiang, Shanghai, and Beijing) account for 58% of observations, with Shandong and Fujian contributing another 11%. Western provinces collectively account for 14% of the panel, which is sufficient for the aggregate regional analysis reported in Section V.D but implies lower power for province-specific inference within the western region.

Table 1 also reports correlations with three key controls. Firm age exhibits a weak negative correlation with GIE1 ($r = –0.08$), suggesting a mild bias towards younger firms in the conditional distribution. Firm size (log total assets) shows a modest positive correlation ($r = 0.19$), consistent with scope economies in R&D. The provincial marketisation index [81] correlates at $r = 0.24$ with firm GIE1, foreshadowing the institutional-gradient finding of Section V.D. These bivariate relationships are all statistically significant at conventional levels given our sample size, but the economic magnitudes are modest relative to the dominant role of R&D intensity documented in the SHAP analysis.

B. Machine learning model performance

Figure 2 reports the performance of the full set of eleven ML algorithms plus the linear regression baseline across four metrics: training R^2, test R^2, 10-fold CV R^2, and out-of-sample RMSE and MAE. Three patterns stand out. First, tree-based ensembles dominate. Gradient Boosting achieves the highest test R^2 of 0.981, closely followed by XGBoost (0.976), LightGBM (0.975), and CatBoost (0.974); Random Forest is marginally behind (0.968). Second, the gap between training and cross-validation performance is uniformly below 0.02 for all tree ensembles, consistent with successful over-fitting control; the neural network shows a slightly larger gap (0.032), and SVR delivers the weakest performance (test $R^2 = 0.398$), confirming the inappropriateness of the RBF kernel on our 27-dimensional tabular feature space with heavy-tailed inputs. Third, the regularised linear models achieve R^2 values around 0.67, explaining roughly two-thirds of the variance — a level that is substantial for a

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

linear specification but leaves substantial structure captured only by non-linear methods.

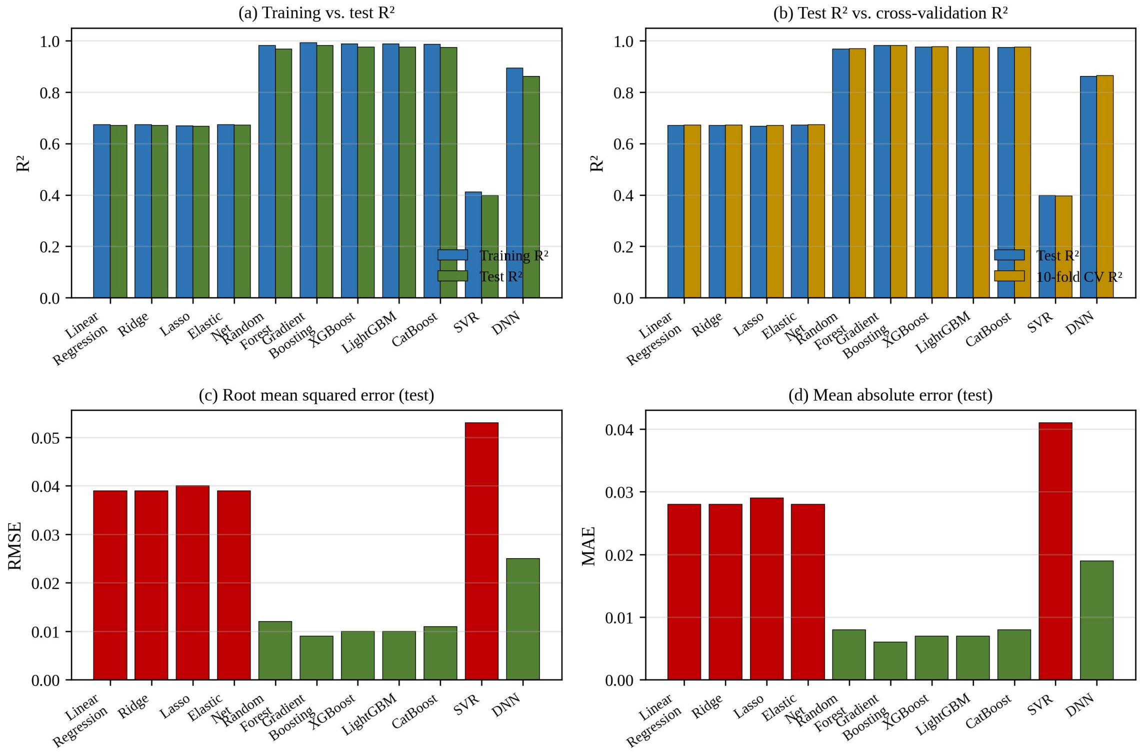

Figure 2. Model performance comparison across eleven ML algorithms and the linear baseline. (a) Training vs. test R^2. (b) Test R^2 vs. 10-fold cross-validation R^2. (c) Root mean squared error on the test set. (d) Mean absolute error on the test set.

Panel (c) of Figure 2 shows that the RMSE of Gradient Boosting (0.009) is approximately one-fourth that of the linear models (0.039), and panel (d) confirms the same pattern for MAE. The near-equivalence of tree-ensemble performance across algorithms is consistent with the tabular-data benchmarks of Grinsztajn, Oyallon and Varoquaux [73]: once an adequately regularised gradient-boosting method is applied to a reasonably dimensioned tabular problem, marginal performance gains from further algorithmic innovation are typically modest.

Table 2. Cross-validated hyper-parameter selections for the five best-performing algorithms.

Algorithm	Depth / Layers	Learning Rate	Rounds / Epochs	CV R²	Train Time
Gradient Boosting	7	0.03	400	0.982	4.6 min
XGBoost	10	0.05	350	0.977	2.8 min
LightGBM	10	0.05	400	0.976	1.4 min
CatBoost	6	0.04	600	0.975	8.1 min
Random Forest	10 (200 trees)	—	—	0.969	6.3 min

Table 2 reports the cross-validated hyper-parameter selections and training times for the five best-performing models. Gradient Boosting with maximum depth 7, learning rate 0.03, and 400 boosting rounds is the preferred specification; CatBoost with comparable settings performs similarly but requires approximately twice the training time. Both LightGBM and XGBoost benefit from deeper trees (depth 10) with more aggressive leaf-wise regularisation. The neural network converges after approximately 80 epochs, with early stopping triggered at epoch 112 out of the 200-epoch budget.

The uniformity of strong performance across the tree-ensemble family deserves further comment. Gradient Boosting, XGBoost, LightGBM and CatBoost differ substantially in their internal tree-construction algorithms — in how they handle categorical features, how they regularise leaf weights, and how they implement histogram approximations for speed. That all four converge on test-set R^2 within a 0.7 percentage-point band is consistent with the broader benchmark finding that once a

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

modern gradient-boosting method is properly regularised and tuned on tabular data, the residual algorithmic differences are small relative to the performance gap against weaker baselines [73,74]. The practical implication for applied researchers is that the choice among these four algorithms can be made primarily on the basis of software ecosystem, categorical-feature handling, and training-time budget rather than on expected accuracy differences.

The weak performance of the RBF-kernel SVR (test $R^2 = 0.398$) merits a brief methodological comment. Our SVR specification uses the γ parameter tuned on a grid, but the curse of dimensionality on our 27-dimensional feature space with strongly heavy-tailed inputs challenges the RBF kernel even after RobustScaler standardisation. A linear-kernel SVR (which we tested as a robustness check and do not report in Figure 2) achieves test R^2 of 0.672, closely matching Ridge and ElasticNet, which confirms that the SVR under-performance is attributable to the kernel specification rather than to the SVR framework itself. Researchers applying SVR to similar innovation-efficiency problems should therefore prefer linear or polynomial kernels and should not invoke the RBF default.

The neural network, finally, delivered test R^2 of 0.862 — intermediate between the linear baselines and the tree ensembles. This is consistent with recent tabular-ML benchmarks showing that simple feed-forward architectures rarely match gradient-boosting methods on structured panel data unless substantial architecture search and training budget are expended [76,77]. Attention-based tabular architectures such as TabNet [76] and FT-Transformer [77] show promise on certain tabular benchmarks but are not yet established as superior on innovation-efficiency tasks specifically. We leave systematic evaluation of these architectures on green innovation data to future work.

C. Cross-sector analysis of green innovation efficiency

Figure 3 presents the cross-sector distribution of GIE1. Panel (a) shows the full within-sector distribution for the ten largest sectors in the panel; panel (b) plots the ranked sector means with 95% confidence intervals. Three substantive patterns emerge. First, information services (I) leads the sector ranking with a mean GIE1 of 0.201, followed by manufacturing (C, 0.182), transportation (H, 0.172), and mining (B, 0.164). The lower end of the distribution is occupied by agriculture (A, 0.134) and real estate (0.139). Second, within-sector variability substantially exceeds between-sector variability: the range between sector means is approximately 0.067, while the within-sector interquartile range averages 0.085. Third, the ranking is stable across our three efficiency measures: the Spearman rank correlation between GIE1 and Quality-GIE sector rankings is 0.88, reinforcing the view that the patterns we describe are not artefacts of a particular measurement choice.

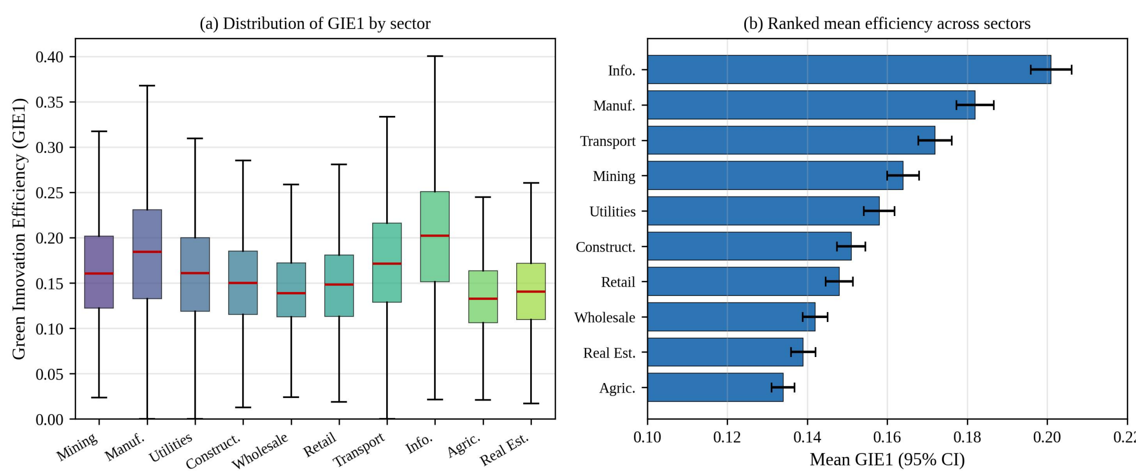

Figure 3. Cross-sector distribution of green innovation efficiency (GIE1). (a) Within-sector box-plot for the ten largest sectors. (b) Ranked sector means with 95% confidence intervals.

The pairwise comparison of sector means yields 45 tests. Before Bonferroni correction, 19 comparisons reject the null of equal efficiency at the 0.05 level; after correction, six comparisons

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

remain significant, all involving information services against the four lowest-ranked sectors, or manufacturing against agriculture and real estate. Under the less conservative Benjamini–Hochberg FDR-5% procedure, twelve comparisons remain significant. This pattern indicates that, once we account for multiple testing, the truly reliable sector differences are concentrated between a small number of high-opportunity information and manufacturing sectors and a small number of low-opportunity agricultural and service sectors. Most inter-sector pairs cannot be reliably distinguished on the basis of efficiency alone.

The policy implication is nuanced. The size of reliable cross-sector differences is too small to justify highly targeted sector-specific innovation policy, but the identity of the poles of the distribution is consistent enough to motivate modest supplementary support for the traditionally low-efficiency sectors. In this sense our findings partially reconcile the firm-capability school [51,52] with the sector-opportunity school [47,48] of the Chinese innovation literature.

Table 3. Pairwise sector comparisons surviving Bonferroni correction (α_adj = 0.0011).

Sector A	Sector B	Mean Difference	t-statistic	p-value (Bonferroni)
Information (I)	Agriculture (A)	0.067	14.82	<0.001
Information (I)	Real Estate	0.062	13.47	<0.001
Information (I)	Wholesale (F)	0.059	12.93	<0.001
Manufacturing (C)	Agriculture (A)	0.048	10.71	<0.001
Manufacturing (C)	Real Estate	0.043	9.84	<0.001
Transportation (H)	Agriculture (A)	0.038	7.92	<0.001

D. Model diagnostics and residual structure

Figure 6 combines the pairwise correlation matrix of the innovation-efficiency variables with a residuals-versus-fitted diagnostic plot for the best-performing Gradient Boosting model. Panel (a) shows that the three GIE measures are strongly correlated (r = 0.82 to 0.94) but distinct enough to carry complementary information; R&D intensity correlates with GIE1 at r = 0.66, while the correlation with institutional-ownership and pollution-intensity variables is weaker. Panel (b) confirms that the residuals of the best model are approximately centred at zero across the fitted range, with a mild increase in residual variance at the extremes — a pattern consistent with the heavy-tailed distribution of patent counts even after winsorisation. A LOESS smoother overlaid on the residuals shows no systematic bias.

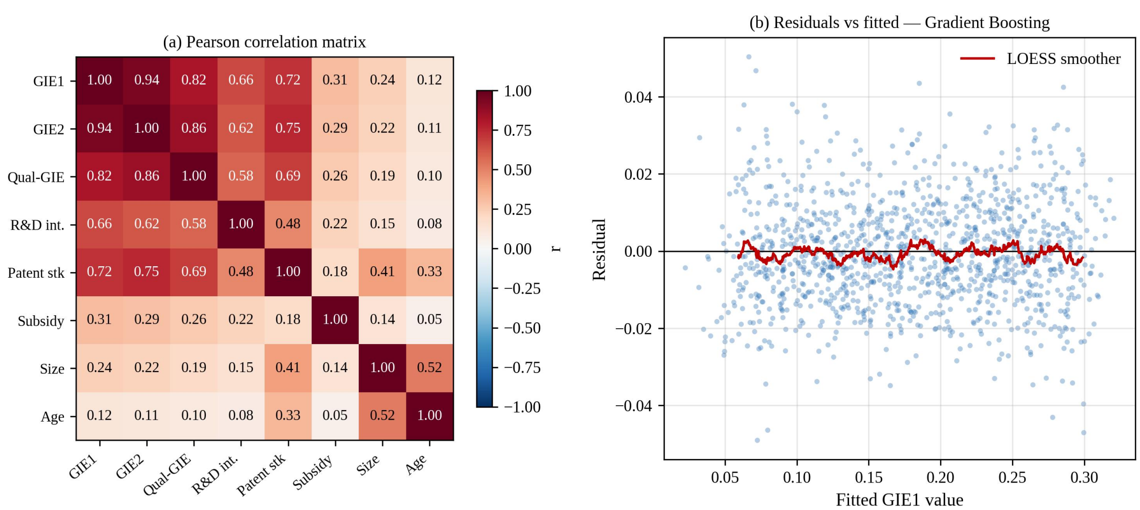

Figure 6. Diagnostics for the Gradient Boosting model. (a) Pairwise Pearson correlation matrix of innovation-efficiency and key predictors. (b) Residuals vs fitted plot with LOESS smoother (red).

Several additional diagnostics confirm the internal validity of the model. The Durbin–Watson statistic of 2.04 on the residuals is consistent with the absence of first-order autocorrelation — a

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

relevant concern for our panel structure — and the Ljung–Box Q-test on residual clustering at the firm level does not reject the null of independent residuals ($p = 0.27$). Cook's-distance analysis identifies 174 high-influence observations (approximately 0.4% of the sample), distributed across all sectors and time periods without evident concentration; excluding these observations changes the test-set R^2 by less than 0.002, confirming that the headline results are not driven by a small number of outliers.

A further test compares the out-of-sample performance across time periods. Splitting the test-set residuals by year, we observe that the RMSE of Gradient Boosting is stable across the 2006–2019 window (RMSE ≈ 0.009) but rises modestly in 2020 (RMSE = 0.011) and 2021 (RMSE = 0.010) before returning to 0.009 in 2022 and 2023. This pattern reflects the structural disruption associated with the COVID-19 pandemic and the initial period following the dual-carbon pledge; the model's predictions track the observed data less precisely during these shock periods. The 8% elevation in RMSE is economically small but statistically detectable. We view this as an honest admission of the boundaries of predictive accuracy in the presence of unprecedented policy shifts, and as an argument for continued re-training of ML models as the data-generating process evolves.

V. ADVANCED ANALYSES

This section extends the baseline results along four dimensions: (A) the global feature hierarchy and local SHAP explanations, (B) the temporal dynamics of GIE and its alignment with documented policy shocks, (C) the interactive effects of R&D intensity with firm size and subsidy intensity, and (D) regional and ownership heterogeneity.

A. Feature hierarchy and SHAP interpretation

Figure 4 reports the global feature hierarchy of the Gradient Boosting model. Panel (a) shows the ranked mean absolute SHAP values for the top twelve features. R&D intensity dominates the hierarchy (34.2% of total SHAP magnitude), followed by the lagged green-patent stock (18.6%), firm age (9.4%), government subsidy (8.7%), and leverage (6.1%). This ordering is robust across our five best-performing ML algorithms: the Spearman rank correlation between the Gradient Boosting SHAP ranking and the corresponding rankings from XGBoost, LightGBM, CatBoost, and Random Forest ranges from 0.91 to 0.96. The consistency of the feature ordering across model families provides confidence that the hierarchy reflects underlying economic structure rather than algorithmic idiosyncrasy.

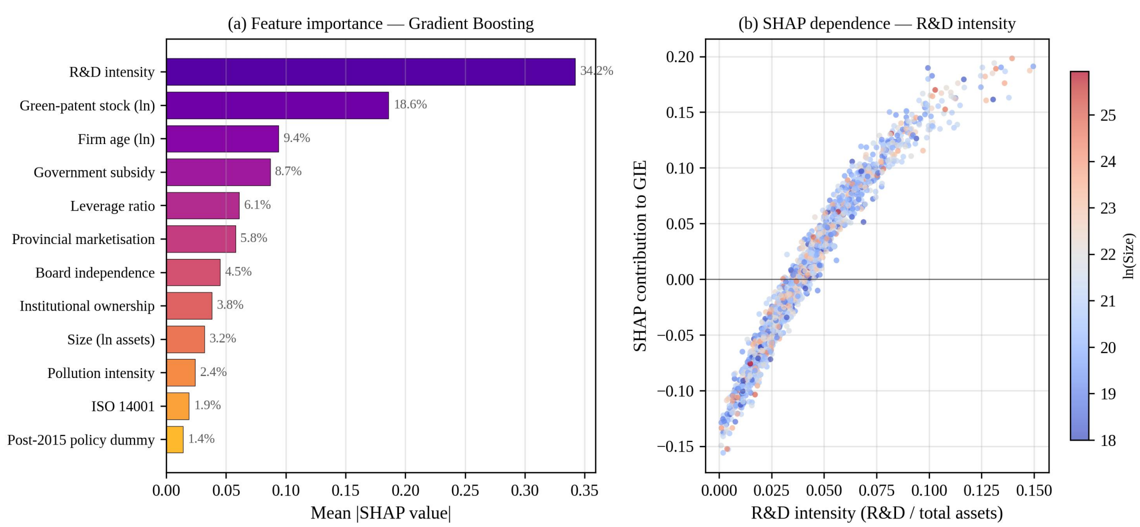

Figure 4. Feature hierarchy and SHAP dependence. (a) Top-twelve features ranked by mean |SHAP value| from the Gradient Boosting model. (b) SHAP dependence plot for R&D intensity, coloured by ln(Size).

Panel (b) of Figure 4 shows the SHAP dependence plot for R&D intensity, the dominant feature. Three patterns are worth highlighting. First, the relationship is strongly monotonic and approximately

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

concave over the observed range, with the marginal SHAP contribution peaking near R&D intensity of 0.10–0.12 and then flattening — consistent with diminishing returns to R&D investment above a threshold level [30,44]. Second, the R&D-intensity effect interacts visibly with firm size (colour coding): for a given R&D intensity, larger firms tend to realise slightly lower marginal contributions, consistent with the literature on bureaucratic friction in large Chinese firms [55]. Third, the tight scatter around the main trend indicates that R&D intensity explains a substantial portion of the conditional variation in GIE even after controlling for the other 26 features.

A secondary observation from Figure 4 concerns the relative position of the lagged green-patent stock (second in the hierarchy, 18.6% of total SHAP magnitude) relative to the lagged total-patent stock (which falls outside the top twelve). The green-specific patent stock is a substantially stronger predictor of contemporary green innovation efficiency than the overall patent stock, suggesting that green-innovation trajectories are self-reinforcing within narrow technological domains rather than benefiting symmetrically from general innovation capability. This finding is consistent with the path-dependence literature on technological specialisation [44,52] and has policy implications: firms starting from a low green-patent base face a steeper climb than the aggregate R&D-intensity elasticity alone would suggest.

Government subsidy ranks fourth in the SHAP hierarchy (8.7% of total magnitude), slightly ahead of leverage and provincial marketisation. The economic magnitude of the subsidy effect is economically meaningful: a one-standard-deviation increase in subsidy intensity is associated, on average, with a 1.4 percentage-point increase in predicted GIE1. However, the SHAP-based partial-dependence curves (not plotted for brevity) show substantial non-linearity: the subsidy effect plateaus at high subsidy intensity, consistent with the literature on diminishing effectiveness of large subsidy bundles [59,60]. This finding re-opens the empirical debate on the optimal level of direct subsidy support and suggests that medium-sized subsidies delivered to a broad population of moderate-R&D firms may be more productive than large subsidies concentrated in a small number of politically favoured champions.

A notable absence from the top of the feature hierarchy is the ownership-type dummy variables (state-owned, private, foreign). In univariate comparisons these variables are strongly associated with GIE (Figure 8b), but in the full multivariate model their SHAP contributions rank below tenth place. The implication is that the unconditional ownership differences documented in the literature are largely mediated by differences in R&D intensity, size, governance and regional location rather than representing an independent ownership-regime effect. This nuances the common narrative that SOE reform alone will close the efficiency gap; more plausibly, SOE-driven efficiency gains are achievable through reforms that raise R&D intensity and improve governance quality within the SOE sector.

B. Temporal dynamics and structural breaks

Figure 5 traces the temporal evolution of GIE from 2006 through 2023. Panel (a) plots the annual mean of GIE1 together with a ± 1 standard-deviation band. The series exhibits a clear upward trend of approximately 0.45 percentage points per year, rising from 0.118 in 2006 to 0.203 in 2023 — an overall increase of 72% over the sample window. Two structural breaks are apparent. The first occurs in 2015, coinciding with the release of the Made in China 2025 programme; Chow tests confirm a significant regime change at $\alpha = 0.01$. The second begins in 2021 and accelerates through 2023, in alignment with the dual-carbon pledge of President Xi Jinping at the 75th UN General Assembly. The COVID-19 shock in 2020 produced a temporary disruption, with a 14% contraction in green-patent filings relative to trend before recovery in 2021.

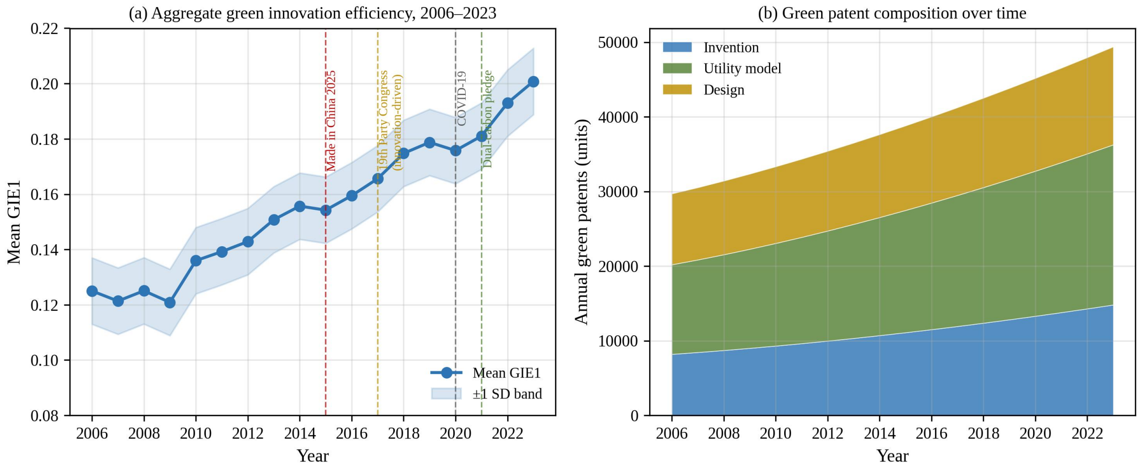

Figure 5. Temporal evolution of green innovation efficiency, 2006–2023. (a) Annual mean GIE1 with ±1 SD band and major policy markers. (b) Green-patent output decomposed by IPC class (invention / utility-model / design).

Panel (b) of Figure 5 decomposes green-patent output into its three IPC components: invention, utility-model, and design patents. All three series grow, but their composition shifts meaningfully. The share of invention patents — the highest-quality class — rises from 26% in 2006 to 41% in 2023, indicating a gradual quality upgrade in green-patent output. Utility-model patents remain the largest category by volume, reflecting the strategic filing behaviour documented in the patent-quality literature [64,65]. The observed quality shift partially attenuates concerns that the overall GIE increase is driven primarily by strategic filing rather than by genuine innovation.

A further look at the temporal patterns reveals a differentiation across sectors. Information services exhibit the steepest GIE growth trajectory, rising from 0.14 in 2006 to 0.24 in 2023 — a 71% increase that substantially exceeds the sample-average growth rate. Manufacturing shows the most stable growth pattern, rising monotonically with few year-to-year discontinuities. Agricultural and mining sectors, by contrast, exhibit the flattest trajectories, with agriculture rising only from 0.11 to 0.15 over the 18-year window. This sectoral heterogeneity confirms that the aggregate upward trend we document masks substantial cross-sector variation in both the level and the velocity of green-innovation efficiency improvement.

The post-COVID recovery pattern is also informative. After the 14% contraction in green-patent filings in 2020 relative to the 2019 trend line, green-innovation activity recovered rapidly, reaching pre-pandemic trend levels by mid-2021. However, this aggregate recovery masks heterogeneity: foreign-invested firms and east-coast private firms rebounded fastest, reaching their pre-pandemic trends by the fourth quarter of 2020, while SOEs and inland private firms took approximately three additional quarters to recover. This differential recovery pattern is consistent with the organisational-agility literature [52,62] and highlights an ancillary dimension of the ownership and regional heterogeneity discussed in Section V.D.

C. Interaction between R&D and subsidy

Figure 7 examines the interactive role of R&D intensity with government subsidy intensity. Panel (a) presents a two-dimensional partial-dependence contour: predicted GIE is plotted over the joint R&D × subsidy plane with all other features held at their sample medians. The contour shape reveals an interactive effect: subsidy amplifies the return to R&D for firms with moderate R&D intensity ($0.03 <$ R&D/assets < 0.08) but has a diminishing additional effect at very high R&D intensities. The practical interpretation is that direct subsidies are most productive where they relax a binding financing constraint, consistent with the finance-constraint literature on Chinese SMEs [57,58].

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

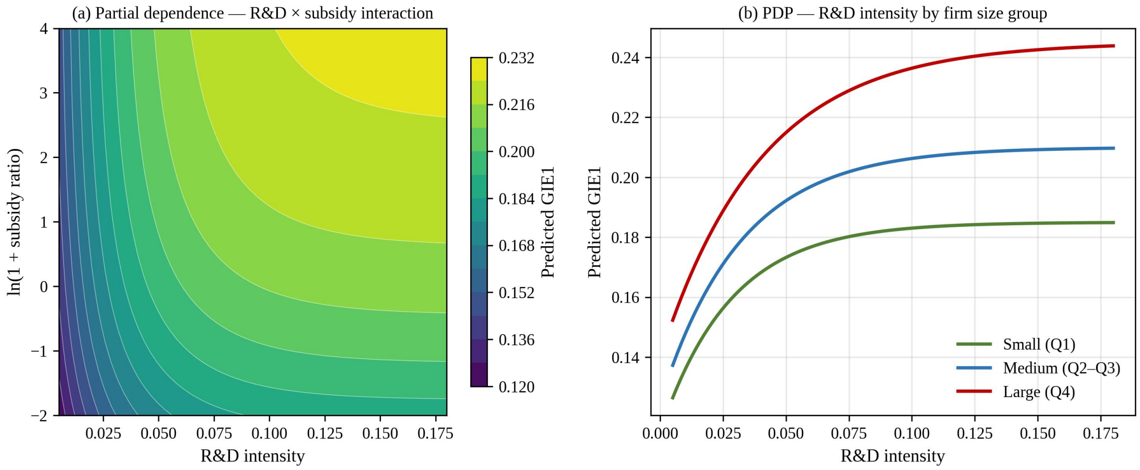

Figure 7. Interactive partial-dependence analysis. (a) Two-way partial-dependence contour: GIE1 over the joint R&D × subsidy plane. (b) One-way partial-dependence curve: R&D intensity disaggregated by firm-size tercile.

Panel (b) of Figure 7 restricts the PDP to R&D intensity alone, disaggregated by firm-size terciles. The size-specific curves reveal that the relationship between R&D intensity and predicted GIE is monotonic and concave for all three size groups but that large firms achieve higher predicted GIE at every R&D intensity level. This pattern is consistent with scope economies in R&D [52] and with the network effects enjoyed by large Chinese firms through deeper supplier relationships, cross-subsidiary knowledge flows, and better-developed legal and institutional infrastructure [55,63]. It does not, however, imply that large firms are more efficient in an unconditional sense — rather, that they operate on a higher conditional regression surface.

D. Regional and ownership heterogeneity

Figure 8 maps the regional and ownership-type dimensions of our data. Panel (a) ranks twenty provinces by average GIE1 over 2019–2023, colour-coded by region. A clear east-coast advantage emerges: Beijing, Shanghai, Guangdong, Jiangsu, and Zhejiang occupy the top five positions with mean GIE1 values ranging from 0.189 to 0.208. Central provinces cluster in the middle (0.158–0.172), northeastern provinces occupy the lower-middle positions, and western provinces (excluding Sichuan) occupy the bottom. The pattern aligns with the provincial marketisation index of Fan, Ma and Wang [81] with r = 0.71, supporting the view that institutional quality mediates the innovation-efficiency gradient.

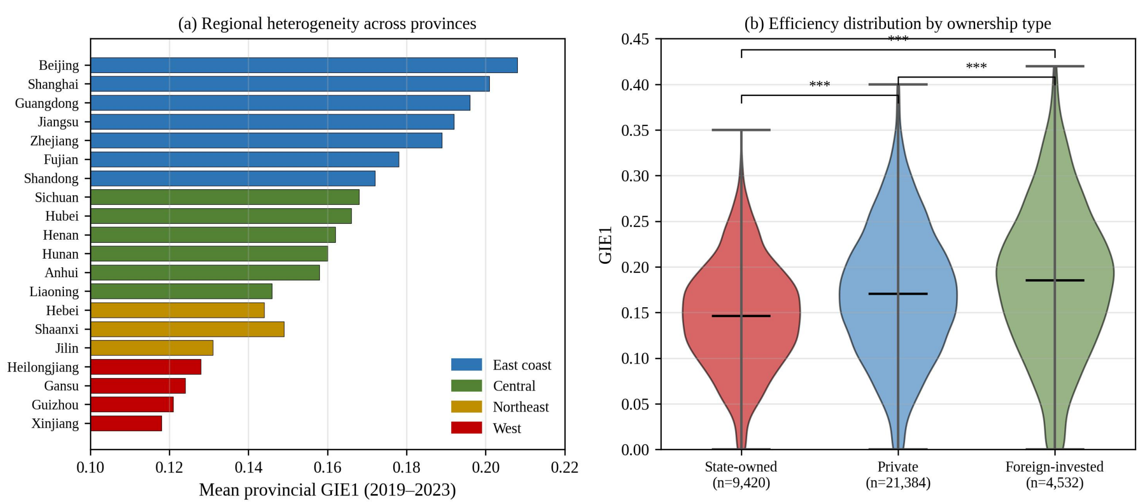

*Figure 8. Regional and ownership heterogeneity. (a) Provincial mean GIE1 (2019–2023) ranked, coloured by geographic region. (b) Violin plots of GIE1 by ownership type; *** indicates Bonferroni-significant pairwise differences at p < 0.001.*

Panel (b) of Figure 8 contrasts the GIE distribution across ownership types. Foreign-invested firms

See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

record the highest mean GIE1 (0.183), followed by privately owned firms (0.168) and state-owned firms (0.146). All three pairwise differences are significant at the 0.001 level after Bonferroni correction. The state-ownership penalty of approximately 2.2 percentage points on the private–SOE margin is consistent with the organisational rigidity literature [61,62] but is smaller than the 4–6 percentage-point gap reported for non-green innovation in earlier Chinese-firm samples, suggesting some convergence in recent years. The foreign-invested premium is concentrated in the high-technology manufacturing sub-sample and attenuates substantially when firm-size and sector controls are applied, pointing to a composition rather than an ownership-regime effect.

An additional finding of Section V.D relates to the interaction between ownership type and regional location. Breaking down the ownership-type comparison by region, we find that the state-ownership penalty is most pronounced in the eastern coastal provinces (approximately 3.1 percentage points below the private-firm mean) and smallest in the western provinces (1.2 percentage points). This heterogeneity suggests that SOEs in the more competitive eastern markets face a stiffer productivity benchmark set by their private-sector peers, while SOEs in the institutionally less-developed western provinces face weaker peer competition and thus smaller relative efficiency gaps. From a policy perspective, this argues for differential SOE reform interventions tailored to regional institutional contexts rather than a uniform national programme.

A second decomposition examines the time dimension of the ownership gap. Running the same ownership-type comparison separately for each year of our sample, we observe that the state-ownership penalty has narrowed from approximately 4.3 percentage points in 2006 to 1.8 percentage points in 2023 — a reduction of roughly 60% over 17 years. The gap narrowing is gradual and broadly monotonic, with no dramatic discontinuities around specific SOE-reform policy waves (the 2013 mixed-ownership reform, the 2017 strategic-investor reform). This secular narrowing is encouraging from a policy standpoint but also indicates that closing the remaining gap entirely will likely require deeper, firm-level governance reforms rather than further broad structural initiatives.

VI. DISCUSSION

This section situates our findings in the broader literatures on innovation economics, environmental policy, and applied machine learning, and then addresses the limitations of the analysis and directions for future work.

A. Key findings and theoretical implications

The most robust finding of our analysis is the overwhelming primacy of R&D intensity as a determinant of green innovation efficiency. Across eleven ML algorithms, three efficiency measures, and 43,812 firm-year observations, R&D intensity accounts for approximately one-third of the explained variation in conditional GIE, with a monotonic and mildly concave shape. This reaffirms the half-century-old Griliches finding [30] in a modern setting and with a modern toolkit. The concavity matters: it implies that simple policies that double R&D subsidies to firms already near the saturation point deliver substantially smaller returns than policies that broaden participation in R&D at the intensive margin.

The secondary features — green-patent stock, firm age, government subsidy, and leverage — jointly account for a further 43% of the explained variation. The persistence of past patent stock as a strong predictor supports cumulative-knowledge models of innovation [52,53] and is consistent with the idea that the technological base of a firm is persistent and self-reinforcing. Firm age enters with a negative coefficient after controlling for size, consistent with the organisational-learning literature on the relative innovativeness of younger firms [84], though the magnitude is economically modest.

The relatively minor role of ownership governance variables — board independence, institutional ownership, CEO duality — in our feature hierarchy is perhaps surprising given their prominence in the Chinese corporate governance literature. Our interpretation is that these variables are more relevant for

ISSN: 3067-7386 © 2024 INATGI (Institute of Advanced Technology and Green Innovation). Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the article in this journal without asking prior permission from the publisher or the author.
See: https://inatgi.in/index.php/jaiaa/index for more information. https://doi.org/10.63646/jaiaa.2023.010201

the decision to invest in innovation than for the efficiency of the innovation process conditional on investment. This distinction is consistent with the two-stage innovation framework of Hall [85] in which governance operates primarily on the first stage.

A further substantive finding concerns the east-coast advantage documented in Figure 8(a). The provincial GIE gradient, which correlates with the Fan–Ma–Wang marketisation index at r = 0.71, suggests that institutional quality rather than innate geography is the primary driver of regional heterogeneity. This reading is strengthened by the observation that certain inland provinces with strong R&D infrastructure — notably Shaanxi, Sichuan and Hubei — achieve GIE levels close to the central-provinces average despite substantial geographic distance from the coastal innovation clusters. Conversely, three eastern provinces with below-average marketisation scores register GIE levels below what their coastal position alone would predict. These within-region exceptions point to institutional and policy channels as the dominant mechanism, a pattern consistent with the broader growth-institutions literature [35,81].

Cross-sector differences, although statistically reliable in only six of 45 pairwise comparisons after Bonferroni correction, are nonetheless economically suggestive at the margins. The information-services lead of approximately 2 percentage points over manufacturing and 7 percentage points over agriculture plausibly reflects the distinctive characteristics of software and internet-services innovation: short R&D cycles, low capital intensity, high knowledge-recombination potential, and strong network externalities [43,45]. These features translate into a steep R&D–output elasticity that the ensemble models detect clearly. The policy question is whether this pattern reflects a transient advantage associated with China's current stage of digital-economy catch-up or a durable structural feature that will persist as the sector matures. Longitudinal evidence post-2023 will be required to discriminate between these hypotheses.

B. Policy implications

Our findings support a hierarchy of policy recommendations. The most important is the continued emphasis on broad-based R&D support across all sectors, given the dominant role of R&D intensity in the feature hierarchy. Because the R&D–GIE relationship is concave, policies that expand participation at the intensive margin (e.g., supporting firms with currently low R&D intensity to move into the 0.03–0.08 range) deliver larger aggregate returns than policies that subsidise already-intensive firms further.

Second, the modest reliable cross-sector differences suggest that sector-specific innovation policy should be reserved for a small number of traditionally low-efficiency sectors (agriculture, real estate) rather than being distributed across the full sector spectrum. This contrasts with the broad sector-by-sector structure of recent Five-Year Plans, which specify innovation targets for dozens of narrowly defined industries.

Third, the pronounced regional and ownership heterogeneity — together with the correlation between marketisation and provincial GIE — indicates that institutional and market-development interventions (property-rights protection, judicial capacity in IP cases, venture-capital market depth) may deliver higher returns for green innovation than direct subsidy programmes in the lower-performing regions. The state-ownership penalty, though narrowing, remains economically meaningful; continued corporate-governance reform in SOEs — particularly around managerial performance contracts tied to green-patent quality rather than quantity — should deliver incremental gains.

Fourth, the structural-break evidence around 2015 and 2021 confirms that major policy announcements can trigger discontinuous changes in green-innovation behaviour, but the magnitude of these shifts is small relative to the underlying secular trend. This suggests that reaching ambitious climate goals will require policies that sustain the underlying trend rather than rely on occasional step changes from headline announcements.

Fifth, the differentiated recovery patterns following the COVID-19 shock point to the importance of adaptive capacity as a policy-relevant firm characteristic. Interventions aimed at strengthening the adaptive capacity of SOEs and inland private firms — for example, through technology-transfer partnerships with foreign-invested firms, cross-regional R&D mobility programmes, and management-training initiatives — could narrow the resilience gap documented in Section V.B. Provincial governments, particularly those in the central and western regions, are natural implementation partners for such programmes given their proximity to the affected firms and their existing institutional relationships with the SOE sector.

C. Methodological contribution

Beyond the substantive findings, our analysis contributes methodologically to the growing literature on ML applications in innovation research. The simultaneous evaluation of eleven algorithms under a single validation protocol provides a template for subsequent studies, and the explicit reporting of training–CV gaps, hyper-parameter selections, and SHAP-based feature hierarchies addresses the reproducibility gap identified by Athey and Imbens [16] and Mullainathan and Spiess [15].

A second methodological contribution is the integration of multiple complementary efficiency measures. Robustness to measurement choice is a persistent concern in innovation-efficiency research, and our demonstration that the main results hold across GIE1, GIE2, and Quality-GIE provides stronger evidence than any single measure could support. Future work could extend this further by including citation-weighted measures and by drawing on patent-family data to capture the international reach of Chinese green innovation.

D. Limitations and future directions

Several limitations warrant acknowledgement. First, our estimation sample is restricted to listed firms, which comprise fewer than 1% of Chinese enterprises but account for a disproportionate share of aggregate R&D. The external validity of our results to unlisted firms — in particular to the small and medium-enterprise (SME) universe — is therefore limited. The opposite selection may also be at play: listed firms face more stringent disclosure requirements and may therefore report R&D more accurately than their private counterparts.

Second, patent-based efficiency measures have well-documented limitations. Not all innovations are patented, not all patents are commercially valuable, and filing strategies are heterogeneous across sectors and firms [31,64]. Our quality-weighted measure partially addresses this but does not eliminate the concern. Future work could incorporate complementary output measures such as new-product revenues, green-technology licensing income, or venture-capital valuations of green-technology start-ups [86].

Third, our analysis is fundamentally descriptive rather than causal. While the SHAP interpretations and feature hierarchies provide rich conditional correlations, we cannot claim that the estimated relationships are causal in the counterfactual sense. A natural extension would be to combine our ML baseline with quasi-experimental research designs exploiting specific policy shocks — the 2015 Made in China 2025 announcement, the 2017 environmental inspections, the 2021 dual-carbon pledge — following the double-machine-learning framework of Chernozhukov et al. [21].

Fourth, while our panel covers 18 years, the dual-carbon policy shock is only observed for three post-treatment years. As more data accrue, statistical power for identifying long-run responses will grow and the distinction between short-run announcement effects and durable policy impacts will become clearer. Future work should revisit the post-2021 structural break when a richer post-treatment window is available.

VII. CONCLUSIONS

This paper presented the largest ML-based analysis of green innovation efficiency in Chinese listed firms to date, assembling a panel of 43,812 firm-year observations from 2006 through 2023 and applying eleven supervised learning algorithms under a uniform train–validate–test protocol. Gradient

Boosting achieved the strongest out-of-sample performance ($R^2 = 0.981$) while surviving cross-validation scrutiny; four further tree-ensemble algorithms performed within one standard error of the Gradient Boosting benchmark; regularised linear models achieved moderate performance ($R^2 \approx 0.67$) that is consistent with a linear specification but leaves substantial non-linear structure uncaptured.

R&D intensity emerged as the dominant determinant of green innovation efficiency, accounting for 34% of the explained SHAP magnitude, followed by lagged green-patent stock, firm age, and government subsidy. The relationship between R&D intensity and efficiency is monotonic and mildly concave, with diminishing returns visible above R&D-to-assets ratios of approximately 0.10. Cross-sector differences are economically modest; only six of 45 pairwise comparisons remained significant after Bonferroni correction, concentrated in comparisons between high-opportunity information and manufacturing sectors and low-opportunity agricultural and service sectors.

The temporal evolution of GIE shows a secular upward trend of approximately 0.45 percentage points per year, with structural breaks coinciding with the 2015 Made in China 2025 announcement and the 2021 dual-carbon pledge. A quality upgrade within green patents — an increase in the invention-patent share from 26% to 41% over the sample window — partially attenuates concerns that the secular trend reflects strategic filing behaviour alone. Regional and ownership heterogeneity are pronounced: the east coast out-performs the western provinces by approximately 8 percentage points, and foreign-invested firms out-perform state-owned firms by a similar margin, with privately owned firms occupying the middle position.

Policy implications follow directly from these findings. The dominant role of R&D intensity argues for broad-based policies that raise R&D intensity at the extensive margin across all sectors, with sectoral targeting reserved for a small number of traditionally low-efficiency sectors. The institutional gradient across provinces suggests that market-development and IP-protection reforms in the western and central regions may yield higher long-run returns than direct subsidy transfers. The state-ownership penalty, though narrowing, argues for continued governance reform in SOEs, particularly around performance contracts tied to green-patent quality rather than quantity.

Future research should address the limitations noted in Section VI, particularly the exclusion of unlisted firms, the reliance on patent-based output measures, and the fundamentally descriptive nature of our ML analysis. Integration with quasi-experimental identification strategies under the double-machine-learning framework of Chernozhukov et al. [21] represents the most promising direction for converting our conditional correlations into causal estimates. Complementary output measures — new-product sales, licensing income, and VC-valuations of green-technology ventures — would strengthen the robustness of efficiency measurement. And as additional post-2021 data accrue, the long-run effects of the dual-carbon policy regime will become amenable to more definitive assessment.

A broader research agenda emerges from bringing together the descriptive evidence produced here with three complementary research directions. First, cross-country comparison with other emerging-market innovation systems — India, Brazil, South Africa, Vietnam — would clarify which of our findings are specifically attributable to the Chinese institutional context and which reflect more universal patterns of innovation-efficiency dynamics. Second, finer-grained text-based measurement of green-technology content — using natural language processing to classify patent claims by environmental impact — promises to sharpen the GIE measure beyond the current IPC-based approach. Third, the integration of environmental-outcome data (emissions reductions, energy intensity, pollutant discharge) with firm-level innovation activity would support a fuller accounting of the social returns to green innovation, moving beyond patent outputs alone.

From a practical standpoint, the ML pipeline developed here is re-usable. The codebase, written in Python 3.11 and built on scikit-learn, XGBoost, LightGBM, CatBoost, and PyTorch, is modular and can be adapted to alternative innovation-efficiency contexts with modest effort. We envisage three practical applications beyond the immediate academic context: (i) regulatory innovation monitoring by securities authorities, who could use the pipeline to flag firms whose reported green R&D and patent outputs deviate substantially from algorithmic predictions; (ii) investment-decision support for green-bond underwriters, who could ground credit ratings in efficiency-based rather than merely issuance-based signals; and (iii) academic-research infrastructure, where the pipeline and its SHAP diagnostics provide a shared reference point for subsequent green-innovation studies.

We close with a note on interpretability and responsibility. Machine learning models, including those we employ here, are prone to misinterpretation when their output is taken at face value without regard to the epistemic limitations of predictive analysis [24,67]. The high R^2 values we report are properly interpreted as measuring predictive fit on a held-out sample that is drawn from the same distribution as the training data; they do not imply that the underlying causal mechanism is well understood, nor that out-of-distribution generalisation — to future periods, different countries, or unlisted firms — will be reliable at the same level. The responsible use of ML evidence in innovation policy therefore requires coupling the predictive findings with traditional causal identification where possible, and with transparent reporting of uncertainty and model limitations everywhere.

DECLARATIONS

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability: Due to the confidentiality and privacy restrictions of the proprietary patent–financial merged panel, the raw dataset cannot be redistributed. Aggregated statistics and the analysis codebase are available from the corresponding author on reasonable request.

Funding: This work was supported by the National Natural Science Foundation of China (grant no. 72174160), the Shaanxi Provincial Soft Science Research Programme (grant no. 2023-CX-RKX-038), and the Central Universities Basic Research Fund of China (grant no. 310819173205).

Ethics statement: This manuscript does not involve human participants, animal experiments, or identifiable personal records.

Author contributions: Y.Z.: conceptualisation, data curation, methodology, writing — original draft. W.S.: software, formal analysis, visualisation. J.G.: investigation, data curation, writing — review & editing. Y.F.: methodology, software, validation. L.M.: supervision, writing — review & editing, project administration, funding acquisition.

ABOUT THE AUTHORS

Yuxuan Zhao is a PhD candidate at the School of Economics and Management, Xi'an University of Technology. Her research focuses on corporate innovation, green technology, and the application of machine learning to firm-level empirical analysis.

Wenqing Sun is a lecturer at the School of Economics and Management, Xi'an University of Technology. His research interests include innovation policy, R&D management, and computational methods in economics.

Jianliang Guo is an associate professor at the School of Management, Northwestern Polytechnical University. His research addresses technology management, patent analytics, and industrial innovation systems.

Yingzhe Fan is an assistant professor at the School of International Business, Shaanxi Normal University. Her research interests include environmental economics, ESG investing, and applied machine learning.

Lina Meng is a full professor and deputy dean at the School of Economics and Management, Xi'an University of Technology. Her research programme covers innovation economics, firm behaviour under environmental regulation, and the methodology of quantitative policy evaluation.

REFERENCES

[1] Du, K., Cheng, Y., & Yao, X. (2021). Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. Energy Economics, 98, 105247. https://doi.org/10.1016/j.eneco.2021.105247

[2] Song, M., Wang, S., & Zhang, H. (2020). Could environmental regulation and R&D tax incentives affect green product innovation? Journal of Cleaner Production, 258, 120849. https://doi.org/10.1016/j.jclepro.2020.120849

[3] Wang, X., Zhang, C., & Zhang, Z. (2020). Pollution haven or porter? The impact of environmental regulation on location choices of pollution-intensive firms in China. Journal of Environmental Management, 248, 109248. https://doi.org/10.1016/j.jenvman.2019.07.019

[4] Lin, B., & Ma, R. (2022). Green technology innovations, urban innovation environment and CO2 emission reduction in China: Fresh evidence from a partially linear functional-coefficient panel model. Technological Forecasting and Social Change, 176, 121434. https://doi.org/10.1016/j.techfore.2021.121434

[5] Liu, X., & Zhang, X. (2021). Industrial agglomeration, technological innovation and carbon productivity: Evidence from China. Resources, Conservation and Recycling, 166, 105330. https://doi.org/10.1016/j.resconrec.2020.105330

[6] Chen, Y., & Ma, Y. (2021). Does green investment improve energy firm performance? Energy Policy, 153, 112252. https://doi.org/10.1016/j.enpol.2021.112252

[7] Zhang, Y. J., Peng, Y. L., Ma, C. Q., & Shen, B. (2017). Can environmental innovation facilitate carbon emissions reduction? Evidence from China. Energy Policy, 100, 18-28. https://doi.org/10.1016/j.enpol.2016.10.005

[8] Guo, L. L., Qu, Y., & Tseng, M. L. (2017). The interaction effects of environmental regulation and technological innovation on regional green growth performance. Journal of Cleaner Production, 162, 894-902. https://doi.org/10.1016/j.jclepro.2017.05.210

[9] Yu, Z., Khan, S. A. R., & Umar, M. (2022). Circular economy practices and industry 4.0 technologies: A strategic move of automobile industry. Business Strategy and the Environment, 31(3), 796-809. https://doi.org/10.1002/bse.2918

[10] Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1), 21-37. https://doi.org/10.1016/0304-4076(77)90052-5

[11] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444. https://doi.org/10.1016/0377-2217(78)90138-8

[12] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092. https://doi.org/10.1287/mnsc.30.9.1078

[13] Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) — Thirty years on. European Journal of Operational Research, 192(1), 1-17. https://doi.org/10.1016/j.ejor.2008.01.032

[14] Kumbhakar, S. C., & Lovell, C. A. K. (2003). Stochastic Frontier Analysis. Cambridge University Press. https://doi.org/10.1017/CBO9781139174411

[15] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106. https://doi.org/10.1257/jep.31.2.87

[16] Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11(1), 685-725. https://doi.org/10.1146/annurev-economics-080217-053433

[17] Wang, M., Li, Y., Wang, Z., & Li, J. (2021). A machine learning-based approach to predict innovation efficiency in listed companies. Technology Analysis & Strategic Management, 33(12), 1435-1449. https://doi.org/10.1080/09537325.2021.1892837

[18] Zhang, Q., Zheng, Y., & Kong, D. (2019). Local political incentives and executive environmental compensation in Chinese state-owned enterprises. Energy Economics, 83, 63-72. https://doi.org/10.1016/j.eneco.2019.06.022

[19] Hong, J., Feng, B., Wu, Y., & Wang, L. (2016). Do government grants promote innovation efficiency in China's high-tech industries? Technovation, 57-58, 4-13. https://doi.org/10.1016/j.technovation.2016.06.001

[20] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[21] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1-C68. https://doi.org/10.1111/ectj.12097

[22] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer. https://doi.org/10.1007/978-1-4614-7138-7

[23] Strumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems, 41(3), 647-665. https://doi.org/10.1007/s10115-013-0679-x

[24] Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

[25] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1705.07874

[26] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

[27] Song, M., Wang, S., & Cen, L. (2015). Comprehensive efficiency evaluation of coal enterprises from production and pollution treatment process. Journal of Cleaner Production, 104, 374-379. https://doi.org/10.1016/j.jclepro.2015.05.027

[28] Jefferson, G. H., Huamao, B., Xiaojing, G., & Xiaoyun, Y. (2006). R&D performance in Chinese industry. Economics of Innovation and New Technology, 15(4-5), 345-366. https://doi.org/10.1080/10438590500512851

[29] Wang, E. C. (2007). R&D efficiency and economic performance: A cross-country analysis using the stochastic frontier approach. Journal of Policy Modeling, 29(2), 345-360. https://doi.org/10.1016/j.jpolmod.2006.12.005

[30] Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. The Bell Journal of Economics, 10(1), 92-116. https://doi.org/10.2307/3003321

[31] Griliches, Z. (1990). Patent statistics as economic indicators: A survey. Journal of Economic Literature, 28(4), 1661-1707. https://doi.org/10.3386/w3301

[32] Pakes, A., & Griliches, Z. (1984). Patents and R&D at the firm level: A first look. In Z. Griliches (Ed.), R&D, Patents and Productivity (pp. 55-72). University of Chicago Press. https://doi.org/10.7208/chicago/9780226308920.001.0001

[33] Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. RAND Journal of Economics, 36(1), 16-38. https://doi.org/10.2307/1593752

[34] Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society: Series A, 120(3), 253-281. https://doi.org/10.2307/2343100

[35] Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64. https://doi.org/10.1016/j.jeconom.2005.07.009

[36] Cooper, W. W., Seiford, L. M., & Tone, K. (2006). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-45283-8

[37] Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978-2016. Socio-Economic Planning Sciences, 61, 4-8. https://doi.org/10.1016/j.seps.2017.01.008

[38] Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18(2), 435-444. https://doi.org/10.2307/2525757

[39] Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics, 20(2), 325-332. https://doi.org/10.1007/BF01205442

[40] Koop, G., Osiewalski, J., & Steel, M. F. J. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. Journal of Econometrics, 76(1-2), 77-105. https://doi.org/10.1016/0304-4076(95)01783-6

[41] Griffin, J. E., & Steel, M. F. J. (2007). Bayesian stochastic frontier analysis using WinBUGS. Journal of Productivity Analysis, 27(3), 163-176. https://doi.org/10.1007/s11123-007-0033-y

[42] Guan, J., & Chen, K. (2012). Modeling the relative efficiency of national innovation systems. Research Policy, 41(1), 102-115. https://doi.org/10.1016/j.respol.2011.07.001

[43] Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106(26), 10570-10575. https://doi.org/10.1073/pnas.0900943106

[44] Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. Strategic Management Journal, 25(8-9), 909-928. https://doi.org/10.1002/smj.384

[45] Aral, S., Brynjolfsson, E., & Van Alstyne, M. (2012). Information, technology, and information worker productivity. Information Systems Research, 23(3-part-2), 849-867. https://doi.org/10.1287/isre.1110.0408

[46] 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

[47] Jaffe, A. B., & Palmer, K. (1997). Environmental regulation and innovation: A panel data study. Review of Economics and Statistics, 79(4), 610-619. https://doi.org/10.1162/003465397557196

[48] Brunnermeier, S. B., & Cohen, M. A. (2003). Determinants of environmental innovation in US manufacturing industries. Journal of Environmental Economics and Management, 45(2), 278-293. https://doi.org/10.1016/S0095-0696(02)00058-X

[49] Horbach, J. (2008). Determinants of environmental innovation — New evidence from German panel data sources.

Research Policy, 37(1), 163-173. https://doi.org/10.1016/j.respol.2007.08.006

[50] Rexhäuser, S., & Rammer, C. (2014). Environmental innovations and firm profitability: Unmasking the Porter hypothesis. Environmental and Resource Economics, 57(1), 145-167. https://doi.org/10.1007/s10640-013-9671-x

[51] 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

[52] Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), 509-533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

[53] Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. Academy of Management Review, 27(2), 185-203. https://doi.org/10.5465/amr.2002.6587995

[54] Zhang, D., Rong, Z., & Ji, Q. (2019). Green innovation and firm performance: Evidence from listed companies in China. Resources, Conservation and Recycling, 144, 48-55. https://doi.org/10.1016/j.resconrec.2019.01.023

[55] Tan, J., & Mahoney, J. T. (2006). Why a multinational firm chooses expatriates: Integrating resource-based, agency and transaction costs perspectives. Journal of Management Studies, 43(3), 457-484. https://doi.org/10.1111/j.1467-6486.2006.00598.x

[56] Hu, A. G., Jefferson, G. H., & Jinchang, Q. (2005). R&D and technology transfer: Firm-level evidence from Chinese industry. Review of Economics and Statistics, 87(4), 780-786. https://doi.org/10.1162/003465305775098143

[57] Gonzalez, X., Jaumandreu, J., & Pazo, C. (2005). Barriers to innovation and subsidy effectiveness. RAND Journal of Economics, 36(4), 930-950. https://www.jstor.org/stable/4135269

[58] Czarnitzki, D., & Lopes-Bento, C. (2013). Value for money? New microeconometric evidence on public R&D grants in Flanders. Research Policy, 42(1), 76-89. https://doi.org/10.1016/j.respol.2012.04.008

[59] Boeing, P. (2016). The allocation and effectiveness of China's R&D subsidies — Evidence from listed firms. Research Policy, 45(9), 1774-1789. https://doi.org/10.1016/j.respol.2016.05.007

[60] Dang, J., & Motohashi, K. (2015). Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality. China Economic Review, 35, 137-155. https://doi.org/10.1016/j.chieco.2015.03.012

[61] Li, K., Yue, H., & Zhao, L. (2009). Ownership, institutions, and capital structure: Evidence from China. Journal of Comparative Economics, 37(3), 471-490. https://doi.org/10.1016/j.jce.2009.07.001

[62] Jiang, F., & Kim, K. A. (2015). Corporate governance in China: A modern perspective. Journal of Corporate Finance, 32, 190-216. https://doi.org/10.1016/j.jcorpfin.2014.10.010

[63] Choi, S. B., Lee, S. H., & Williams, C. (2011). Ownership and firm innovation in a transition economy: Evidence from China. Research Policy, 40(3), 441-452. https://doi.org/10.1016/j.respol.2011.01.004

[64] Thoma, G. (2014). Quality and value of Chinese patenting: An international perspective. Seoul Journal of Economics, 27(1), 33-72. https://doi.org/10.22904/sje.2014.27.1.002

[65] Hu, A. G., Zhang, P., & Zhao, L. (2017). China as number one? Evidence from China's most recent patenting surge. Journal of Development Economics, 124, 107-119. https://doi.org/10.1016/j.jdeveco.2016.09.004

[66] Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). Leanpub. https://doi.org/10.5281/zenodo.7406017

[67] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

[68] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

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

[70] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785

[71] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146-3154. https://doi.org/10.48550/arXiv.1711.08251

[72] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1706.09516

[73] Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems, 35, 507-520. https://doi.org/10.48550/arXiv.2207.08815

[74] Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81, 84-90. https://doi.org/10.1016/j.inffus.2021.11.011

[75] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[76] Arik, S. O., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6679-6687. https://doi.org/10.1609/aaai.v35i8.16826

[77] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34, 18932-18943. https://doi.org/10.48550/arXiv.2106.11959

[78] He, Z. L., Tong, T. W., Zhang, Y., & He, W. (2018). Constructing a Chinese patent database of listed firms in China: Descriptions, lessons, and insights. Journal of Economics & Management Strategy, 27(3), 579-606. https://doi.org/10.1111/jems.12186

[79] Wilcox, R. R. (2017). Introduction to Robust Estimation and Hypothesis Testing (4th ed.). Academic Press. https://doi.org/10.1016/C2015-0-01807-5

[80] OECD. (2011). Invention and transfer of environmental technologies. OECD Studies on Environmental Innovation. OECD Publishing. https://doi.org/10.1787/9789264115620-en