Article

A Hybrid Data Envelopment Analysis–Random Forest Methodology for Evaluating Green Innovation Efficiency in an Asymmetric Environment

Limei Chen [1,2,*], Xiaohan Xie [1], Yao Yao [1], Weidong Huang [1,2] and Gongzhi Luo [1]

1   School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210023, China;
    xie_xiao_han@163.com (X.X.); yaoyaozh0826@163.com (Y.Y.); huangwd@njupt.edu.cn (W.H.);
    lgzlyg@163.com (G.L.)
2   Key Research Base of Philosophy and Social Sciences in Jiangsu, Information Industry Integration Innovation
    and Emergency Management Research Center, Nanjing 210023, China
*   Correspondence: chenlm@njupt.edu.cn; Tel.: +86-025-13776693396

Abstract: The accurate evaluation of green innovation efficiency is a critical prerequisite for enterprises to achieve sustainable development goals and improve environmental performance and economic efficiency. This paper evaluates the green innovation efficiency of 72 new-energy enterprises by using a hybrid method of Data Envelopment Analysis (DEA) and a random forest model. The non-parametric DEA model is combined with the parametric SFA model to analyze the real green innovation efficiency on the basis of removing environmental factors and random factors. Then, the random forest model based on a nonlinear relationship is used to evaluate factors impacting green innovation efficiency. This paper proposes a comprehensive evaluation method designed to assess the green innovation efficiency of new-energy enterprises. By applying this method, companies can gain a comprehensive understanding of the current performance in green innovation, facilitating informed decision-making and accelerating sustainable development.

Keywords: three-stage DEA model; random forest method; green innovation efficiency; new-energy companies

1. Introduction

With the increasing global emphasis on sustainable development, corporate green innovation has emerged as a crucial driver for economic growth and environmental protection. However, accurately assessing the efficiency of corporate green innovation faces numerous challenges, among which the issue of data asymmetry is particularly prominent. The diversity and difficulty in quantifying the input–output indicators of green innovation, the nonlinearity of the innovation process, as well as the complexity of firm heterogeneity and external environments have all led to difficulties in data collection, processing, and analysis for measuring green innovation efficiency.

The asymmetry of corporate data stems from various factors such as different business areas, market competition, economic cycle fluctuations, differences in management levels, and external environments, resulting in unbalanced performance across various indicators. This complexity reflects the diversity and challenges of the business environment, necessitating a comprehensive consideration to better understand and address the asymmetry of corporate data, thereby supporting effective decision making and management. In dealing with asymmetric data, the DEA (Data Envelopment Analysis) model exhibits significant advantages, making it a powerful tool for evaluating efficiency. DEA model is applicable to situations with multiple inputs and outputs. As a non-parametric method, it does not require assumptions about probability distribution of data or the form of production functions, thus being more suitable for handling various types of data, including asymmetric

data. DEA assesses the performance level of units by comparing their relative efficiency rather than relying on a specific mathematical model. This allows DEA to effectively handle asymmetric data and provide judgments of relative efficiency, making it more flexible and widely applicable in practical settings.

DEA is a non-parametric efficiency evaluation method used to assess the efficiency of decision-making units. DEA was first proposed by Charnes, Cooper, and Rhodes in 1978 [1], leading to the CCR model. Nowadays, this model has been used in numerous applications [2–4]. Subsequently, Banker, Charnes, and Cooper (1984) further expanded on the DEA method, introducing the BCC model [5], which addressed the issue of constant returns to scale present in CCR model. However, Fried et al. pointed out that [6] enterprises' inefficiency is not only impacted by internal mismanagement but also by external environments and random errors, thus proposing a three-stage DEA model. Nevertheless, traditional three-stage DEA models also have shortcomings. The efficiency measurement in the first stage of traditional three-stage DEA model operates under the premise of equal contraction proportions for every input [7]. In reality, however, different inputs exhibit different elasticities and do not decrease proportionally, ignoring the slackness in resource utilization. This can lead to biases in evaluation results and the failure to fully reflect on decision-making units' efficiency level. A model based on slack variables, SBM model introduced by Tone [8], can effectively address this deficiency.

According to the above analysis, this article embeds SBM model into three-stage DEA model, adopting a non-parametric and non-oriented SBM model in the first and third stages of the three-stage DEA model. By considering slackness in resource utilization, it evaluates the efficiency level of DMU more comprehensively and provides more accurate evaluation results. In practical applications, many regression problems exhibit nonlinear relationships, and traditional linear regression models often have difficulties effectively capturing the complex patterns in data. As a powerful ensemble learning method, random forests demonstrate significant advantages in handling nonlinear regression problems due to their non-parametric, highly flexible, and robust nature. This article will delve into the application of random forests in studying the factors impacting green innovation efficiency, aiming to provide decision makers with more detailed and comprehensive information that can aid in taking appropriate actions.

This paper's remaining sections are arranged as follows: A summary of previous research on green innovation efficiency is given in Section 2. The formulas for three-stage DEA model and random forest model utilized in this article are presented in Section 3. In Section 4, green innovation efficiency measured by three-stage DEA model is examined and elements that influence green innovation are discussed. The study findings are outlined in Section 5, along with the paper's limitations and future directions.

2. Literature Review

The rapid evolution of the modern economy has brought environmental concerns to the forefront, prompting a heightened societal focus on ecological issues. Green innovation has emerged as a critical research area attracting growing scholarly attention. Green innovation, rooted in the idea of "sustainable development," was initially used in the 1980 "World Conservation Strategy Report." [9]. Subsequently, literature related to sustainable innovation [10,11], eco-innovation [12,13], green innovation [14–17], and environmental innovation [18] has gradually increased. Scholars hold different views on the understanding of green innovation. Chen et al. [19] defined "green innovation" as advancements in hardware or software that contribute to eco-friendly products or processes. These innovations encompass areas such as sustainable product design, and environmentally responsible corporate management practices. Wu et al. [20] believed that green innovation is a product of the combination of innovation theory and ecological views, which aims to maximize economic benefits while obtaining new knowledge and technologies to reduce environmental pollution. Rennings [21] argued that green innovation has a "double externality," with spillover effects both in the production stage and in the diffusion stage,

resulting in a certain degree of reduction in internal costs and external environmental costs. Bernauer et al. [22] discussed the concept of green innovation as being within the same category as environmental innovation and eco-innovation. Zhang [23] and Schiederig [14], among other scholars, conducted detailed literature reviews and comparative analyses of definitions, revealing that green innovation, eco-innovation, environmental innovation, and sustainable innovation share a high degree of consistency in their core concerns and goals. Disregarding the subtle differences in their definitions, they are often interchangeably used or even equated in many literature sources. Currently, there are three major interpretations of green innovation definition in academia: equating green innovation with innovations that contribute positively to the ecological environment, equating green innovation with innovations that introduce environmental performance, and equating green innovation with environmental innovation or the optimization and innovation of environmental performance [23].

When evaluating green innovation efficiency and its influencing factors, numerous scholars have adopted diverse strategies. Most input–output indicators are constructed using Stochastic Frontier Analysis (SFA) [24,25], DEA [26–29], and related methods. Some scholars have also comprehensively assessed green innovation efficiency through spatial econometrics [30] and evaluated it using the entropy method [31]. Xiao et al. [32] used an improved SFA model to conduct a thorough assessment of green innovation efficiency in Yangtze River Economic Belt. However, SFA model application requires the presetting of a production function, which, to a certain extent, increases the subjectivity of the evaluation. In contrast, DEA model operates without requiring assumptions regarding the production function form and can make evaluation results more objective and accurate. Thus, DEA has become the mainstream method for scholars to study green innovation efficiency. The following Table 1 presents relevant studies that use the DEA model to measure green innovation efficiency.

Table 1. Relevant DEA studies on green innovation efficiency.

Author	Research Method	Research Object
Wang and Ren (2022) [33]	Dynamic Network SBM Model	30 provinces in China
Li et al. (2023) [34]	Slack-Based Measure (SBM) Model	Construction enterprises
Xu et al. (2023) [35]	Bounded Concave Cone (BCC) Model	42 listed textile enterprises
Xiao et al. (2021) [36]	Two-Stage Network DEA Model	84 resource-based cities
Xu et al. (2022) [37]	Super-efficiency SBM Model	79 cities
Wang et al. (2022) [38]	Epsilon-Based Measure–Geometric Mean (EBM-GML) Model	2177 listed companies

Regarding the evaluation index system, the existing literature primarily constructs such a system from the following two aspects: input and output. This encompasses the following three dimensions: green innovation efficiency input variable, desirable output, and undesirable output, as illustrated in Table 2. Tian et al. [39] divided the input–output indicators into the following two stages: scientific and technological research and development (R&D) and achievement transformation. For R&D stage, the input–output indicators include the number of R&D personnel, whereas for achievement transformation stage, indicators encompass technology introduction and transformation expenditure, sales revenue of new products, etc. Zhang et al. [40] categorized inputs into human, material, and financial resources, while the selected innovation output indicators are broadly divided into the following two types: scientific and technological outcomes and economic benefits. Ma and Zhu [41] distinguished innovation inputs into R&D investment and production investment. R&D investment is represented by R&D funding and personnel, while pro-

duction investment is expressed by employee compensation. For output indicators, they selected the number of patent applications and intangible assets.

Table 2. Overview of research on green innovation efficiency indicator systems.

Grade I	GradeII	Source of Literature
Input variable	Number of R&D personnel	Tian et al. (2023) [39]
	R&D expenditure	Wang et al. (2022) [42]
	Expenditure on technology import	Kang (2017) [43]
	Expenditure on new product development	Wu (2020) [44]
	Employee compensation	Ma et al. (2023) [41]
Desirable output	Turnover of technology market	Tao and Dai (2018) [45]
	Sales revenue of new products	Tian et al. (2023) [46]
	Number of patent applications	Lv and Qiao (2019) [47]
Undesirablele output	Industrial wastewater and waste gas emissions	Li et al. (2020) [48]
	Solid waste discharge	Wang and Wang (2023) [49]

When examining factors impacting green innovation efficiency, scholars have primarily focused on two levels: macro-environment and micro-level factors. The macro-environment encompasses the institutional landscape [50–52], market industry [34,53], and related international trade relations [54]. At the micro-level, internal factors related to enterprises mainly include the level of awareness of enterprise personnel [55], enterprise costs [56,57], and social responsibility [58,59]. It is observable that empirical research on green innovation efficiency differs depending on research questions. Hong et al. [60] analyzed the influencing factors of innovation efficiency in China's pharmaceutical manufacturing industry and found that two external macro-factors, namely, market competition intensity and government policy support, as well as the internal micro-factor of the enterprise size, are essential for achieving higher levels of innovation efficiency. Wenbo [61] studied the impact of production factors, economic benefits, internal management, and the social environment on green innovation. Kang et al. [62] examined whether and how environmental regulations drive green innovation, aiming to explore the influencing mechanism of green innovation efficiency. Yalabik [63] found that factors such as market competition, consumption, and environmental protection pressure can significantly affect firms' green technology innovation efficiency. Gong et al. [64] provided a detailed analysis of how factors such as the agglomeration effect of outward foreign direct investment influence industrial green innovation efficiency. Kuang et al. [65] tested the influencing mechanism of green innovation efficiency from the perspective of the shadow economy, exploring potential pathways to enhance green innovation efficiency.

Regarding research methods for influencing factors, the random forest model, as an integrated learning method, exhibits good robustness and generalization capabilities, and is suitable for various types of datasets and problems. In 1995, Ho [66] first proposed the concept of random decision forests. He suggested creating a classifier based on decision trees that contained an infinite number of decision trees, which were combined in a complementary or weighted manner to construct a new classifier, namely, the random decision forest. Random decision forests address the issue of overfitting that can occur with single decision trees. In 2001, Breiman [67] integrated bagging algorithms, random subspace algorithms, and classification and regression trees to propose the traditional random forest. Subsequently, the traditional random forest has been widely applied in numerous fields such as ecology [68–70], medicine [71–73], management [74,75], and economics [76,77], and has achieved good results in solving routine classification or regression problems. Xu et al. [78] applied the random forest to observe data from gastric cancer patients to predict their postoperative survival status and assist doctors in assessing treatment decisions. Xie

et al. [75] integrated sampling techniques and cost penalties into the random forest and used bank customer data as an example to predict customer churn. Susana et al. [79] applied the random forest method to unbalanced samples to enable public institutions to direct public investment subsidies to identified groups of enterprises based on this identification.

The traditional DEA model has limitations in efficiency evaluation, which does not consider the impact of environmental variables and random factors on the green innovation efficiency, resulting in bias in efficiency evaluation results. The measurement of green innovation efficiency mainly stays at the macro-level, such as the province and industry, and there are few studies on the enterprise level. As an important force to promote green low-carbon transformation and achieve sustainable development, research on measuring green innovation efficiency in new-energy companies using DEA model remains limited. The research on the influencing factors of green innovation efficiency is mainly based on linear regression models, which cannot effectively analyze nonlinear relationships. There is a gap in the research on the nonlinear influence relationship, and it is difficult to accurately evaluate the factors affecting the green innovation efficiency.

Against this backdrop, this paper establishes a research framework that combines a three-stage DEA model with an SBM model, excluding environmental and random factors, to provide a accurate measure of green innovation efficiency. Considering the advantages of random forest model in exploring influencing factors, this paper selects the random forest model to analyze the influencing factors of green innovation efficiency. The main contributions of this paper are as follows:

Firstly, by embedding SBM model into the three-stage DEA, this paper comprehensively evaluates the efficiency level of DMUs by considering the slackness of resource utilization, providing more accurate evaluation results. By combining the parametric SFA model with the non-parametric DEA model, this paper fully utilizes their respective advantages to better handle asymmetric data, thereby more comprehensively assessing the efficiency level of units and proposing improvement suggestions.

Secondly, unlike other linear regression methods, the random forest model adopted in this paper can not only provide rankings of influencing factors but can also visually demonstrate the nonlinear characteristics of influencing factors on green innovation efficiency by plotting partial dependence plots. This facilitates a deeper understanding of how various factors influence green innovation efficiency.

3. Research Methodology

3.1. The Three-Stage DEA Model

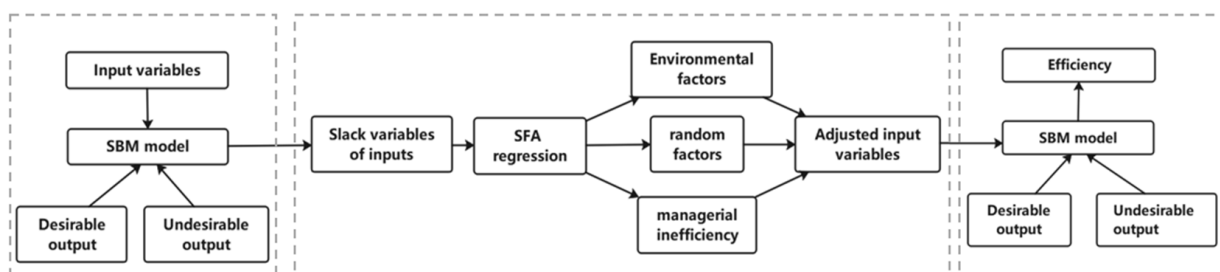The three-stage DEA model framework for analyzing green innovation efficiency is illustrated in Figure 1.

Figure 1. Research framework.

3.1.1. The First Stage: SBM Model

To assess efficiency from both input and output perspectives, this paper utilizes the non-oriented SBM model. The SBM model formula is as follows:

$$min \ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{ik}}}{1 + \frac{1}{s_1+s_2}\left(\sum_{r=1}^{s_1}\frac{s_r^g}{y_{rk}^g} + \sum_{r=1}^{s_2}\frac{s_r^b}{y_{rk}^b}\right)}$$

$$\text{s.t.} \begin{cases} x_k = X\lambda + s^- \\ y_k^g = Y^g\lambda - s^g \\ y_k^b = Y^b\lambda + s^b \\ s^- \geq 0, s^g \geq 0, s^b \geq 0, \lambda \geq 0 \end{cases}$$

The SBM model incorporates slack variables to account for differences in input and output levels. $s^- \in R^m$ represents slack in input resources, $s^b \in R^{s_2}$ reflects slack in undesirable outputs, and s^g represents slack in desirable outputs. The model considers m input variables, s_1 desirable output variables, and s_2 non-desirable output variables. $\left(x_k, y_k^g, y_k^b\right)$ represents the input, desirable output, and non-desirable output values for the k-th decision-making unit. The collective data for all decision-making units is represented by $\left(X, Y^g, Y^b\right)$. The weights assigned to each of the n decision-making units are represented by $\lambda \in R^n$.

The efficiency value of the evaluated DMU is denoted by ρ. Technical efficiency (TE) is determined under the assumption of constant returns to scale (CRS), while pure technical efficiency (PTE) is calculated assuming variable returns to scale (VRS). Scale efficiency (SE) is calculated as the ratio of TE to PTE (SE = TE/PTE).

3.1.2. The Second Stage: SFA Model

In the second stage, this paper decomposes input slack variables into components representing environmental factors, random factors, and managerial inefficiency. By excluding the environmental and random factors, this paper obtains the input redundancy attributable solely to managerial inefficiency. This can be expressed as:

$$S_{nk} = f^n(Z_k; \beta^n) + V_{nk} + U_{nk}$$
$$n = 1, 2, \ldots, N; k = n = 1, 2, \ldots, K$$

In this expression, S_{nk} denotes the slack variable associated with the n-th input of the k-th decision-making unit. The influence of environmental factors is denoted by $f^n(Z_k; \beta^n)$, typically calculated as $f^n(Z_k; \beta^n) = Z_k\beta^n$, where Z_k represents observed environmental variables and β^n is the corresponding parameter vector. The mixed error term $V_{nk} + U_{nk}$ incorporates both random factors $V_{nk} \in N(0, \sigma_{vn}^2)$ and managerial inefficiency $U_{nk} \in N(\mu_u, \sigma_{un}^2)$.

Frontier 4.1 software is used to perform SFA regression analysis, yielding estimates for β^n, σ^2, and the parameter γ. These estimates are then used to calculate σ_{vn} and σ_{un} using the formulas below:

$$\sigma^2 = \sigma_{vn}^2 + \sigma_{un}^2, \gamma = \frac{\sigma_{vn}^2}{\sigma_{vn}^2 + \sigma_{un}^2}$$

The parameter γ quantifies the proportion of variance attributed to managerial inefficiency within the total variance. When γ is close to 1, managerial inefficiency has a more significant impact. Conversely, when γ is close to 0, random factors have a greater influence.

The managerial inefficiency term can be isolated using the following formula:

$$E[U_{nk}|V_{nk} + U_{nk}] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\varphi\left(\frac{\varepsilon\lambda}{\sigma}\right)}{\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)} + \frac{\varepsilon\lambda}{\sigma}\right]$$

The mixed error term is represented by $\varepsilon = V_{nk} + U_{nk}$, with $\lambda = \sigma_{un}/\sigma_{vn}$. φ denotes the probability density function and Φ denotes the distribution function of the standard normal distribution.

After isolating the managerial inefficiency term U, the random factor term V can be calculated using the following formula:

$$E[V_{nk}|V_{nk} + U_{nk}] = S_{nk} - f^n(Z_k; \beta^n) - E[U_{nk}|V_{nk} + U_{nk}]$$

Next, the input variables are adjusted using the SFA model to derive new input values, which are calculated as follows:

$$X_{nk}^* = X_{nk} + [max(Z_k\beta^n) - Z_k\beta^n] + [max(V_{nk}) - V_{nk}]$$
$$n = 1, 2, \ldots, N; k = n = 1, 2, \ldots, K$$

In this formula, X_{nk}^* represents the adjusted input, while X_{nk} denotes the original input. The term $[max(Z_k\beta^n) - Z_k\beta^n]$ accounts for the adjustment to the influence of environmental factors. The term $[max(V_{nk}) - V_{nk}]$ accounts for the adjustment of random factors influence, ensuring that all decision-making units are evaluated under equivalent conditions.

3.1.3. The Third Stage: The SBM Model after Adjusting the Input Variables

By reintroducing adjusted inputs X_{nk}^* and original outputs into the SBM model, this paper can re-evaluate efficiency. This approach removes the influence of environmental and random factors, resulting in a more accurate representation of green innovation efficiency.

3.2. Random Forest Model

In this paper, a random forest model is used to analyze the factors influencing new-energy companies' green innovation efficiency. The random forest model is generally implemented through the following steps:

The bootstrap method is used to extract subsamples with sample size n from the original data, and m feature variables are determined to form the dataset $D = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}, y_i\}(i \in [1, m])$.

A regression tree is constructed for each subsample, denoting the regression tree as $t_j(x)$.

The results of all regression trees are summarized to obtain the optimal estimate, $t(x) = \sum_j t_j(x)$.

Compared with the traditional multiple regression analysis, the advantages of the random forest are very obvious. Not only does it not need to set the function form, it can also rank the importance of the independent variables and further give the partial correlation graph.

4. Research Results

4.1. Variable Selection and Data Sources

Input variables are chosen based on three aspects: labor, capital, and energy. Labor input: Selecting the number of R&D personnel as an indicator can directly reflect the human resource investment of enterprises in green innovation. Capital investment: R&D expenditure, as a measure of capital investment, reflects the financial support of enterprises in green technology research and development. Energy input: The comprehensive energy consumption can reflect the energy consumption level of the enterprise in the production and operation process, and is an important indicator to measure the energy utilization efficiency and green development level. The selection of these three indicators takes into account the characteristics of green innovation and can better reflect the enterprises' investment in green innovation.

Output variables are categorized as either desirable or undesirable. Desirable outputs are selected based on technological and economic factors. Technological output is measured by the number of green patent applications. These patents represent innovations in

environmentally friendly technologies, products, or solutions, reflecting a company's commitment to sustainable development. Main business income serves as the economic output variable, representing the sales revenue generated through core operations. Greenhouse gas emissions are chosen as the undesirable output variable, reflecting the new-energy companies' contribution to advancing the dual carbon target.

A detailed description of input and output variables, environmental factors, and data sources employed in the study is presented in Table 3.

Table 3. Green innovation efficiency index system.

Primary Index	Secondary Index	Three-Level Index	Data Source
Input variable	Labor input	The number of R&D personnel	Annual report
	Capital input	R&D expenditure	Annual report
	Energy input	Comprehensive energy consumption	ESG Report/Social responsibility reports
Desirable output	Technological output	The number of green patent applications	CNRDS database
	Economic output	The main business income	Annual report
Undesirable output	Environmental pollution	Greenhouse gas emissions	ESG Report/Social responsibility reports
Environmental factor	Environmental regulation intensity	Investment in industrial pollution control	
	Technological market environment	Technology market turnover	Statistical yearbook
	Educational environment	Local education expenditure	
	Economic development level	Per capita GDP	
	Regional openness	Foreign investment	

Data Source Description: This paper analyzes A-share listed companies in the new-energy sector. Companies with ST or *ST designations, those without disclosed ESG or social responsibility reports, and those with missing indicators are excluded. This resulted in a sample of 72 new-energy listed companies. Data for 2022 is collected from company annual reports, ESG reports, CNRDS database, and statistical yearbooks.

4.2. Three-Stage DEA Model for Green Innovation Efficiency Analysis

4.2.1. Green Innovation Efficiency Analysis in the First Stage

As seen in Figure 2 and Table 4, the first stage green innovation efficiency of 72 new-energy enterprises in 2022 was determined using the MAXDEA software, which was based on SBM model. Environmental and random factors are not excluded in this calculation.

Table 4. Green innovation efficiency in the first stage.

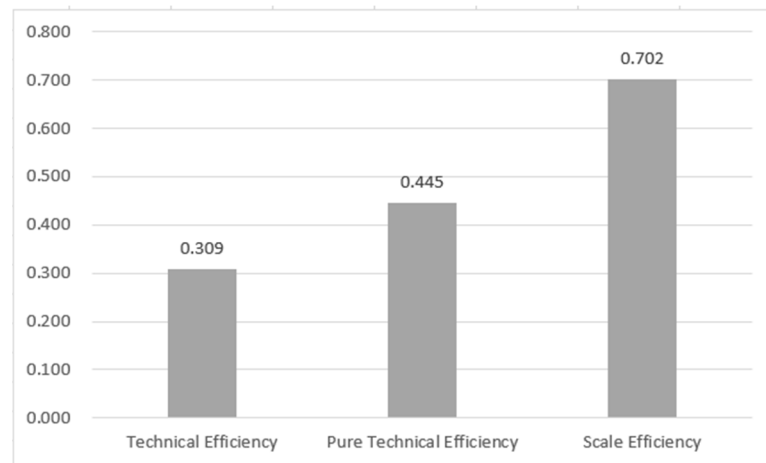| Efficiency Interval	Technical Efficiency		Pure Technical Efficiency		Scale Efficiency	
	Quantity	Proportion	Quantity	Proportion	Quantity	Proportion
<0.2	40	55.56%	31	43.06%	4	5.56%
[0.2,0.5)	17	23.61%	17	23.61%	11	15.28%
[0.5,0.8)	1	1.39%	0	0.00%	27	37.50%
[0.8,1)	0	0.00%	0	0.00%	16	22.22%
1	14	19.44%	24	33.33%	14	19.44%

Figure 2. The efficiency mean of the first stage.

Figure 2 indicates a relatively low mean technical efficiency of 0.309 for the 72 new-energy companies. Pure technical efficiency, also averages 0.445, suggesting a low level for technology and management within the sample. Scale efficiency, representing the rationality of company size and its influence on efficiency, averages 0.702. This suggests that scale efficiency is higher than pure technical efficiency within the sample.

According to Table 4, both technical efficiency and pure technical efficiency have a large number of enterprises in the range of less than 0.5, followed by a large number of enterprises with an efficiency value of 1, and a small number of enterprises in the range of 0.5–1. The scale efficiency is the largest number of enterprises in the range of 0.5–0.8, accounting for the largest proportion. It shows that different new-energy enterprises have a large gap in green innovation efficiency.

DEA is a non-parametric efficiency evaluation method which evaluates the relative efficiency of each DMU by constructing the efficiency front of the DMU. DEA does not need to set the weight of the input–output index in advance, thereby avoiding the influence of subjective factors on the weight setting. The relative importance of each input and output index can be indirectly reflected through the analysis of slack variables.

Table 5 reveals that input improvement values are negative across labor, capital, and energy inputs, indicating excessive resource utilization. Companies appear to use more resources than necessary to achieve outputs. While economic output is relatively close to the target value, suggesting a focus on economic benefits, there's a significant gap between target and actual values for technical output. DMUs have a large improvement in the output index of green patent applications, which indicates that the output has a great impact on the efficiency of DMUs. Through the analysis of the slack variables, the improvement direction for efficiency is provided.

Table 5. Slack variable analysis in the first stage.

Variable	Original Value	Slack Value	Projection Value	Improvement Ratio
The number of R&D personnel	3684.833	−997.660	2687.173	−27.07%
R&D expenditure	217,692.420	−63,610.829	154,081.591	−29.22%
Comprehensive energy consumption	1,285,105.964	−261,737.918	1,023,368.046	−20.37%
The number of green patent applications	47.708	39.570	87.279	82.94%
The main business income	5,614,479.023	22,079.575	5,636,558.599	0.39%
Greenhouse gas emissions	4,432,341.918	−902,530.917	3,529,811.001	−20.36%

4.2.2. SFA Regression Analysis in the Second Stage

The input slack variables from the first stage are used as dependent variables in a regression analysis and environmental factors are as independent variables. The SFA regression analysis, conducted using Frontier 4.1, is summarized in Table 6.

Table 6. Regression results of the SFA model.

Variable.	Slack Variable of the Number of R&D Personnel	Slack Variable of R&D Expenditure	Slack Variable of Comprehensive Energy Consumption
Constant term	−1770.583 ***	−7814.858 ***	128073.570 ***
Environmental regulation intensity	−0.000062	0.015838	0.387586 ***
Technological market environment	−0.000009	0.000077	0.002989 ***
Educational environment	0.219	2.936 ***	−117.948 ***
Economic development level	0.011 *	0.027	−1.163 ***
Regional openness	−0.009	−1.173 ***	−1.023
Sigma-squared	14687403	38875426000	1064423000000
Gamma	1	1	1
LR test of the one-sided error	63.28 ***	65.71 ***	66.27 ***

Note: *** and * represent the significance levels of 1% and 10%, respectively.

Table 6 shows that the one-sided error LR test is significant at the 1% level, rejecting the hypothesis of no managerial inefficiency. This implies that the slack variables of the three inputs are impacted by management inefficiency. The gamma value of 1 indicates that managerial inefficiency dominates, while random factors have a limited impact on green innovation efficiency. These findings support the use of the SFA model. Although the regression coefficients of the environmental variables on the slacks of the individual input variables are not significant, the LR one-sided error test passes at the 1% significance level. Therefore, the adjustment of the input variables still needs to take into account all five of the environmental variables mentioned above.

Environmental regulation intensity is positively correlated with comprehensive energy consumption slack variable at the 1% significance level. The increased intensity of environmental regulations may require companies to adjust or improve production processes, which may lead to some energy consumption increases.

A positive correlation is observed between technological market environment and the slack variable for comprehensive energy consumption at the 1% significance level. The improvement of the technological market environment may encourage new-energy companies to undergo technological updates and transformations, accompanied by a certain increase in energy consumption. However, as technology gradually matures, companies are expected to ultimately achieve a reduction in energy consumption through new technologies and more efficient production methods.

At the 1% level of significance, the educational environment is positively correlated with the slack variable of R&D expenditure, but negatively correlated with the slack variable of comprehensive energy consumption. Increased competition in technological innovation, often driven by a higher local education level, may prompt companies to boost R&D expenditure to remain competitive. The increase in local educational expenditure may offer new energy companies better access to talent and technological support, facilitating the transition from high-energy-consumption stages to more efficient and sustainable production modes.

Economic development level exhibits a positive correlation ($p < 0.10$) with R&D personnel slack variable and a negative correlation ($p < 0.01$) with comprehensive energy

consumption slack variable. This suggests that as regional economies grow, more investment opportunities and innovative projects arise, leading to increased demand for R&D personnel. With the gradual advancement of technological progress, production optimization, and economic structural adjustments, a trend towards a reduction in energy consumption may be observed.

Regional openness exhibits a negative correlation ($p < 0.01$) with R&D expenditure slack variable. This suggests that open regions, with the favorable innovation ecosystems, facilitate more efficient utilization of R&D funds by fostering external cooperation, bringing in advanced technology, innovative management practices, and R&D resources.

4.2.3. Green Innovation Efficiency Analysis in the Third Stage

The SBM model was used to re-evaluate the green innovation efficiency of new-energy companies, using adjusted input variables in place of the originals while keeping output variables constant. This re-evaluation, illustrated in Figure 3 and Table 7, provides a more accurate assessment of efficiency by eliminating the influence of environmental and random factors.
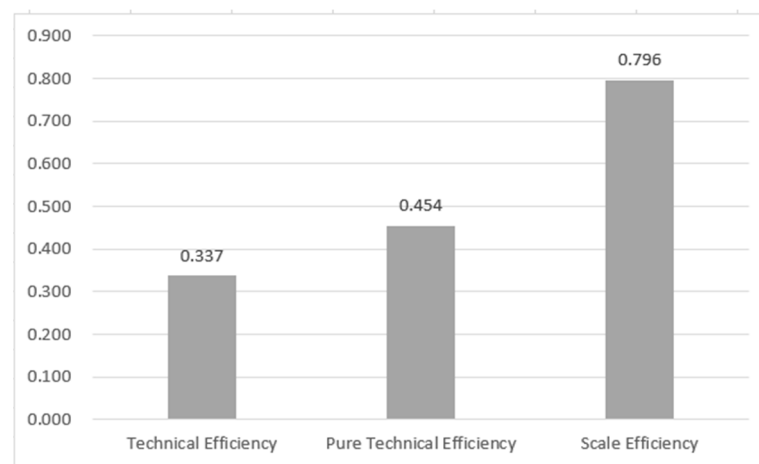

Figure 3. The efficiency mean of the third stage.

Table 7. Green innovation efficiency in the third stage.

Efficiency Interval	Technical Efficiency		Pure Technical Efficiency		Scale Efficiency	
	Quantity	Proportion	Quantity	Proportion	Quantity	Proportion
<0.2	37	51.39%	31	43.06%	4	5.56%
[0.2,0.5)	20	27.78%	17	23.61%	7	9.72%
[0.5,0.8)	1	1.39%	0	0.00%	16	22.22%
[0.8,1)	0	0.00%	0	0.00%	31	43.06%
1	14	19.44%	24	33.33%	14	19.44%

A comparison of the green innovation efficiencies in the first and third stages reveals that all efficiency types have improved after removing environmental and random factors. The average technical efficiency increased from 0.309 to 0.337, the average pure technical efficiency increased from 0.445 to 0.454, and the average scale efficiency increased from 0.702 to 0.796. This suggests that the initial assessment of green innovation efficiency was underestimated due to environmental impacts, highlighting the constraints imposed on new-energy companies by external conditions. While improvements were observed after adjustment, significant room for further improvement remains.

In the third stage, the number of companies achieving DEA effectiveness remains at 14. Technical efficiency and pure technical efficiency are still the largest number of enterprises in the range of less than 0.5, accounting for more than half, while scale efficiency is the largest number of enterprises in the range of 0.8–1. This shows that the level of technical

efficiency and pure technical efficiency of most enterprises is low, but the level of scale efficiency is high, so the emphasis should be placed on the improvement of the enterprise technology and management level.

As can be seen from Table 8, the improvements in slack variables in the third stage are similar to those in the first stage, where both input variables have redundant phenomena, the number of green patent applications in the output variable has a large room for improvement, and economic output closely approaches the target value.

Table 8. Slack variable analysis in the third stage.

Variable	Original Value	Slack Value	Projection Value	Improvement Ratio
The number of R&D personnel	4160.857	−1036.586	3124.272	−24.91%
R&D expenditure	225,664.976	−63951.125	161,713.851	−28.34%
Comprehensive energy consumption	1,425,256.872	−297,192.427	1,128,064.446	−20.85%
The number of green patent applications	47.708	44.970	92.679	94.26%
The main business income	5,614,479.023	131,797.358	5,746,276.382	2.35%
Greenhouse gas emissions	4,432,341.918	−709,067.834	3,723,274.084	−16.00%

Analyzing input redundancy and output insufficiency allows for an evaluation of resource utilization efficiency and provides insights into improving both input and output inefficiencies. The insights can empower managers to make informed decisions that promote rational resource allocation, improve green innovation efficiency, and drive sustainable development.

After the removal of environmental factors and random factors, 72 new-energy companies are classified into four groups according to their pure technical efficiency and scale efficiency levels. The scatter points on the graph represent sample companies. Taking scale efficiency as the *Y*-axis and pure technical efficiency as the *X*-axis, and bounded by the mean value (0.454, 0.796), it is divided into the following four types: high-tech high-scale, high-tech low-scale, low-tech high-scale, and low-tech low-scale, as shown in Figure 4.
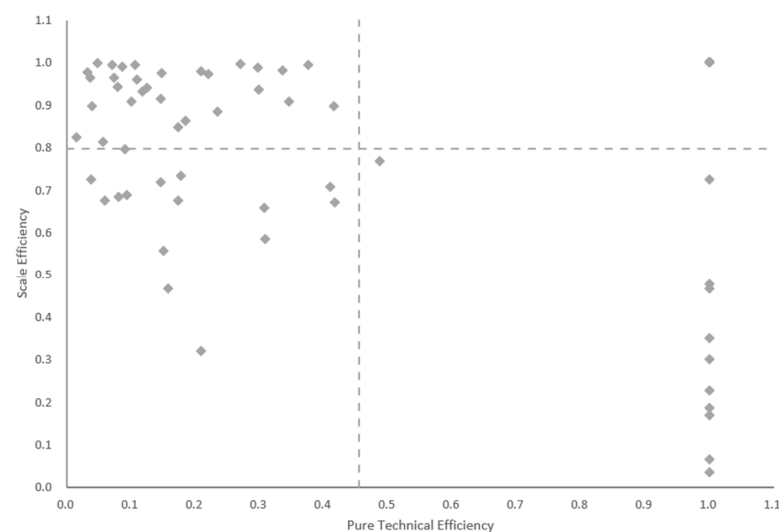

Figure 4. Efficiency classification chart.

High-tech High-scale: There are 14 enterprises, accounting for 19.44% of the total, and PTE and SE of these 14 enterprises are one, reaching the forefront of efficiency. Although a business has reached the DEA efficiency frontier, it can still achieve further development by looking for new growth opportunities, maintaining sensitivity to competitive dynamics, and adapting strategies to capitalize on new opportunities.

High-tech Low-scale: Including 11 enterprises, accounting for 15.28% of the total, the PTE is at a high level, but the SE is low. Therefore, these enterprises should focus on scale efficiency, consider multiple factors such as strategy, market demand, capital and resources, and risk assessment, and reasonably control the scale of enterprises and provide efficient products and services at an appropriate scale.

Low-tech High-scale: Including 32 enterprises, accounting for 44.44% of the total, accounting for the largest proportion, its PTE is low, while the SE is at a higher level. Therefore, these enterprises should focus on pure technical efficiency, and improve the level of enterprise technology, management, and resource utilization by rationally allocating R&D resources, optimizing management processes, and improving the professional competence and innovation consciousness of employees.

Low-tech Low-scale: Including 15 enterprises, accounting for 20.83% of the total, their PTE and SE are at a low level. These enterprises should not only focus on enhancing technological capabilities, management practices, and resource utilization but also consider the optimal size for operations.

4.3. Analysis of Influencing Factors of Green Innovation Efficiency Based on Random Forest Model

This paper analyzes the importance of factors influencing new-energy companies' green innovation efficiency based on the random forest model. To further enhance the interpretability of the random forest model, the influencing factors are analyzed based on partial dependence plots. Traditional regression analysis methods only represent the influence of independent variables on dependent variables in terms of average trends through regression coefficients, but the random forest model can intricately demonstrate the effects of independent variables on the dependent variable at different levels through partial dependence plots.

The mean square error is 0.039, the root mean square error is 0.198, and the average absolute error is 0.157. All three of these values are small, indicating that the error between the actual value and the predicted value is small, and the prediction effect of the model is better.

Figure 5 presents the ranking of the importance of the factors influencing green innovation efficiency based on the random forest model. It is evident that, among these influencing factors, ownership concentration, R&D personnel structure, and operational capacity hold the top three positions in terms of importance, exerting significant influence on new-energy companies' green innovation efficiency.

The relationship between the ownership concentration and green innovation efficiency is illustrated in Figure 6. When the concentration of ownership is high, the resources of an enterprise are more likely to be concentrated in the hands of a few major shareholders. These large shareholders usually have a stronger decision-making ability and resource allocation ability, and can promote the implementation of green innovation projects more efficiently. High ownership concentration means that the interests of major shareholders are more consistent with the interests of the enterprise as a whole. In this case, major shareholders have more incentive to promote green innovation because it not only helps to enhance the social image and brand value of the company, but also can bring long-term economic benefits.
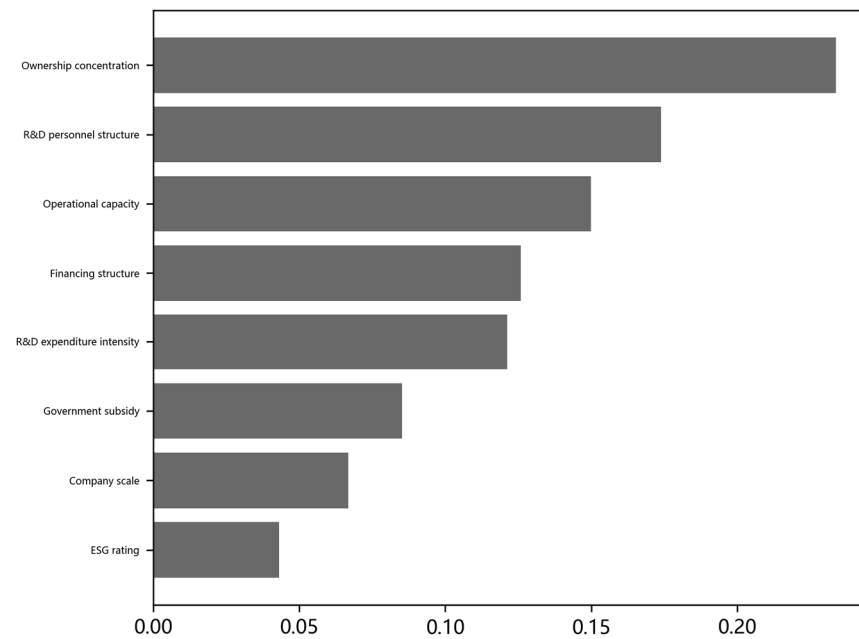
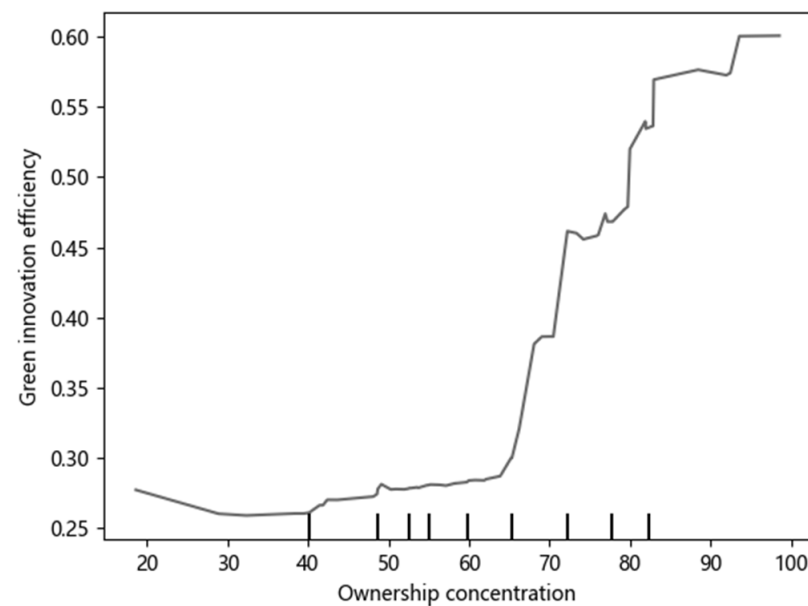**Figure 5.** Ranking of influencing factors.

Figure 6. Ownership concentration partial dependence graph.

The relationship between the R&D personnel structure and green innovation efficiency is illustrated in Figure 7. Initially, the newly added R&D personnel require time to adapt to the company's working environment, products, and technology, leading to a temporary decrease in efficiency. Over time, the company's R&D team gradually establishes a more mature collaborative mechanism and accumulates experience in green innovation, resulting in an increase in green innovation efficiency. Overall, this change may stem from the developmental process of the R&D team, starting from the initial adaptation period to subsequent synergistic effects, ultimately leading to improved efficiency in green innovation.
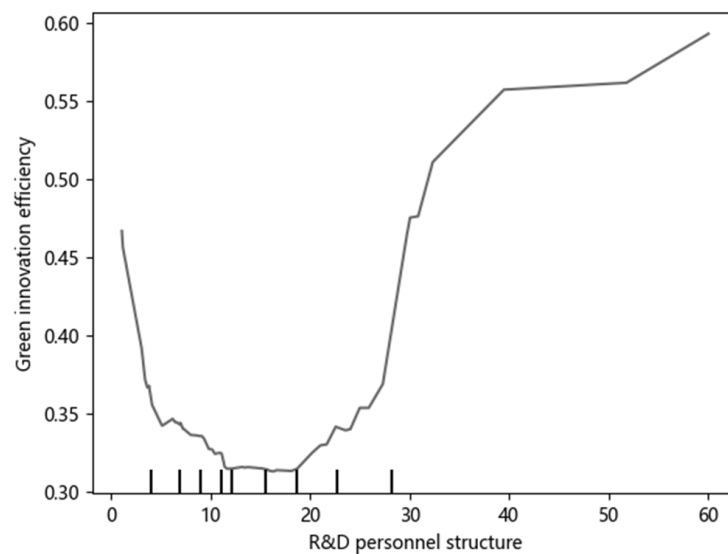
Figure 7. R&D personnel structure partial dependence graph.

The relationship between operational capability and green innovation efficiency is illustrated in Figure 8. The enhancement of a company's operational capability necessitates optimizing resource allocation to improve production efficiency and accelerate asset turnover. Such changes in resource allocation may have a short-term impact on the input and efficiency of green innovation. However, with the continuous optimization of resource allocation and the adoption of new technologies, green innovation efficiency is expected to gradually increase and achieve long-term improvements.
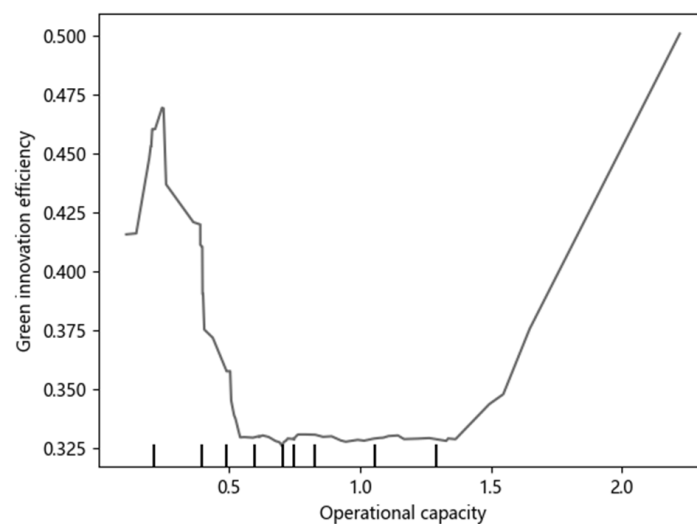
Figure 8. Operational capacity partial dependence graph.

5. Conclusions and Discussion

5.1. Conclusions

The three-stage DEA model reveals that, after the second stage of SFA adjustments, TE, PTE, and SE all demonstrate some improvement. However, significant potential for further enhancement remains, highlighting the impact of external environmental constraints on new-energy companies' green innovation efficiency. Despite adjustments, SE consistently surpasses PTE. The number of enterprises in the state of low-tech high-scale is the largest, accounting for the largest proportion. Therefore, improving green innovation efficiency requires a focus on increasing pure technical efficiency through advancements in technology and management practices.

The factors affecting green innovation efficiency of new-energy companies are studied based on random forest model. Meanwhile, so as to further improve the interpretability of random forest model, important influencing factors are analyzed based on partial dependence plots. The study found that, among these influencing factors, the ownership concentration, R&D personnel structure, and operational capacity hold the top three positions in terms of importance, exerting important influence on new-energy companies' green innovation efficiency.

In order to improve the green innovation efficiency, it is necessary to work together on the following:

At the enterprise level, strengthen technological innovation capacity building, increase investment in research and development, strengthen key core technologies, and develop more efficient, clean, and low-carbon new-energy technologies and products. Optimize the energy management system, actively promote clean production, and reduce pollutant emissions. Strengthen the construction of the talent team, introduce and cultivate green innovation talents, and enhance the talent support ability of green innovation in enterprises.

At the government level, improve the policy support system and increase the policy support for the green innovation of new-energy enterprises. Strengthen industry supervision, establish a sound green innovation standard system, strengthen the supervision and management of green innovation activities, and guide the green and healthy development of enterprises. Foster a favorable environment for innovation, strengthen intellectual property protection, and create a market environment for fair competition.

5.2. Discussion

In the field of new energy, the green innovation efficiency serves as a pivotal indicator for measuring sustainable development ability and competitiveness of enterprises. With the enhancement of global environmental awareness and the transformation of energy structures, the new-energy industry is facing unprecedented opportunities and challenges, and improving green innovation efficiency is crucial to promoting the high-quality development of the new-energy industry. Green innovation efficiency is the key for new-energy enterprises to achieve win–win economic and environmental benefits. China's new-energy industry is developing rapidly, but it also faces challenges such as tight resource and environmental constraints and the need for the breakthrough of core technologies. Improving the efficiency of green innovation can promote the development of the new-energy industry into the high-end, intelligent, and green direction, thereby getting rid of the dependence on traditional resources and achieving sustainable development.

The managerial implications of this study for new-energy companies lies in the following. (1) Pointing out the improvement direction and improving the performance of green innovation: The research results can help management to find the shortcomings of enterprises in green innovation, such as a low resource allocation efficiency and poor control of undesirable output, and take targeted improvement measures to improve green innovation performance of enterprises. (2) The research results can help management to deeply understand the factors affecting the green innovation efficiency, identify the advantages and disadvantages of enterprises, and provide a scientific basis for formulating green transformation and upgrading strategies, thereby optimizing resource allocation and enhancing enterprises competitiveness. (3) Promote the change in management concepts and strengthening the awareness of green development: This study emphasizes the importance of green innovation and sustainable development, and encourages enterprises to fully integrate green development principles into management practices.

This research also needs to be further explored from the following aspects: (1) Due to the limited years in which green data, such as the comprehensive energy consumption and greenhouse gas emissions of enterprises, can be obtained, this study only selects 2022 as the research period. It is suggested that the research time scope should be further expanded in future studies to explore the dynamic evolution trend of the green innovation efficiency of enterprises. (2) The selected index system is not complete enough, which affects the depth

and breadth of the conclusion. In future studies, qualitative indicators can be added on the basis of quantitative indicators, and the two can be combined for the research so as to further improve the index system.

References

1. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
2. Guan, Z.C.; Zhao, E.S. Application of preferred DEA in the evaluation of relative efficiency in scientific research institutions. *Sci. Sci. Technol. Manag.* **2008**, *8*, 40–45.
3. Wong, W.P.; Wong, K.Y. Supply chain performance measurement system using DEA modeling. *Ind. Manag. Data Syst.* **2007**, *107*, 361–381. [CrossRef]
4. Mariappan, P.; Sreeaarthi, G. Application of data envelopment analysis (DEA) to study the performance efficiency of the scheduled commercial banks of India. *Int. J. Mark. Technol.* **2013**, *3*, 52–70.
5. Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical and scale efficiency in data envelopment analysis. *Manag. Sci.* **1984**, *30*, 1078–1099. [CrossRef]
6. Fried, H.O.; Lovell, C.A.K.; Schmidt, S.S.; Yaisawarng, S. Accounting for environmental effects and statistical noise in data envelopment analysis. *J. Prod. Anal.* **2020**, *17*, 157–174. [CrossRef]
7. Tang, Y.; Ding, H.; Shan, X.; Wang, X. Application of the novel three-stage DEA model to evaluate total-factor energy efficiency: A case study based on 30 provinces and 8 comprehensive economic zones of China. *Results Eng.* **2023**, *20*, 101417. [CrossRef]
8. Tone, K. A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **2001**, *130*, 498–509. [CrossRef]
9. Song, Z.M.; Zhang, N. Research on Green Innovation Efficiency of Listed Chinese Energy Companies Based on Triple Bottom Line. *Complexity* **2020**, *1*, 3450471. [CrossRef]
10. Schot, J.; Geels, F.W. Strategic niche management and sustainable innovation journeys: Theory, findings, research agenda, and policy. *Technol. Anal. Strateg. Manag.* **2008**, *20*, 537–554. [CrossRef]
11. Boons, F.; Lüdeke-Freund, F. Business models for sustainable innovation: State-of-the-art and steps towards a research agenda. *J. Clean. Prod.* **2013**, *45*, 9–19. [CrossRef]
12. Hellström, T. Dimensions of environmentally sustainable innovation: The structure of ecological innovation concepts. *Sustain. Dev.* **2007**, *15*, 148–159. [CrossRef]
13. Olsson, P.; Galaz, V. Social-ecological innovation and transformation. In *Social Innovation*; Palgrave Macmillan: London, UK, 2012; pp. 223–247.
14. Schiederig, T.; Tietze, F.; Herstatt, C. Green innovation in technology and innovation management—An exploratory literature review. *R&D Manag.* **2012**, *42*, 180–192.
15. Cheng, C.C.J. Sustainability orientation, green supplier involvement, and green innovation performance: Evidence from diversifying green entrants. *J. Bus. Ethics* **2020**, *161*, 393–414. [CrossRef]
16. El-Kassar, A.-N.; Singh, S.K. Green innovation and organizational performance: The influence of big data and the moderating role of management commitment and HR practices. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 483–498. [CrossRef]
17. Zhang, D.; Rong, Z.; Ji, Q. Green innovation and firm performance: Evidence from listed companies in China. *Resour. Conserv. Recycl.* **2019**, *144*, 48–55. [CrossRef]
18. Fernando, Y.; Chiappetta Jabbour, C.J.; Wah, W.-X. Pursuing green growth in technology firms through the connections between environmental innovation and sustainable business performance: Does service capability matter? *Resour. Conserv. Recycl.* **2019**, *141*, 8–20. [CrossRef]
19. Chen, Y.; Lai, S.; Wen, C. The Influence of green innovation performance on corporate advantage in Taiwan. *J. Bus. Ethics* **2006**, *67*, 331–339. [CrossRef]
20. Wu, C.; Yang, S.W.; Tang, P.C. Construction of green innovation efficiency improvement model in China's heavily polluting industries. *Chin. J. Popul. Resour. Environ.* **2018**, *28*, 40–48.
21. Rennings, K. Redefining innovation—Eco-innovation research and the contribution from ecological economics. *Ecol. Econ.* **2000**, *32*, 319–322. [CrossRef]

22. Bernauer, T.; Engel, S.; Kammerer, D.; Sejas Nogareda, J. Explaining green innovation: Ten years after Porter's Win-Win Proposition: How to study the effects of regulation on corporate environmental innovation? *Soc. Sci. Electron. Publ.* **2007**, *39*, 323–341.

23. Zhang, G.; Zhang, X.J. Several basic issues in green innovation research. *China Sci. Technol. Forum* **2013**, *4*, 12–15.

24. Wang, Q.; Sun, Y.L.; Zhao, Z. Two-stage innovation efficiency of new energy enterprises in China: A non-radial DEA approach. *Technol. Forecast. Soc. Chang.* **2016**, *112*, 254–261. [CrossRef]

25. Gao, Y.; Tsai, S.B.; Xue, X.; Ren, T.; Du, X.; Chen, Q.; Wang, J. An empirical study on green innovation efficiency in the green institutional environment. *Sustainability* **2018**, *10*, 724. [CrossRef]

26. Lin, S.; Sun, J.; Marinova, D.; Zhao, D. Evaluation of the green technology innovation efficiency of China's manufacturing industries: DEA window analysis with ideal window width. *Technol. Anal. Strateg. Manag.* **2018**, *30*, 1166–1181. [CrossRef]

27. Namazi, M.; Mohammadi, E. Natural resource dependence and economic growth: A topsis/dea analysis of innovation efficiency. *Resour. Policy* **2018**, *59*, 544–552. [CrossRef]

28. Wang, W.; Yu, B.; Yan, X.; Yao, X.; Liu, Y. Estimation of innovation's green performance: A range-adjusted measure approach to assess the unified efficiency of China's manufacturing industry. *J. Clean. Prod.* **2017**, *149*, 919–924. [CrossRef]

29. Broekel, T.; Rogge, N.; Brenner, T. The innovation efficiency of German regions—A shared-input DEA approach. *Rev. Reg. Res.* **2018**, *38*, 77–109. [CrossRef]

30. Guan, J.C.; Liu, S.Z. The study on impact of institutions on innovation efficiency in regional innovation systems. *Stud. Sci. Sci.* **2003**, *2*, 98–102.

31. Tseng, M.L.; Chiu, A.S.F. Grey-entropy analytical network process for green innovation practices. *Procedia Soc. Behav. Sci.* **2012**, *57*, 10–21. [CrossRef]

32. Xiao, Q.L.; Xiao, L.M. Spatial-Temporal Differences and Responses of Coupling and Coordination between Green Innovation Efficiency and Ecological Governance Performance: A Case Study of 108 Cities in the Yangtze River Economic Belt. *World Reg. Stud.* **2022**, *31*, 96–110.

33. Wang, Q.; Ren, S. Evaluation of green technology innovation efficiency in a regional context: A dynamic network slacks-based measuring approach. *Technol. Forecast. Soc. Chang.* **2022**, *182*, 121836. [CrossRef]

34. Li, X.; Liu, X.; Huang, Y.; Li, J.; He, J. Theoretical framework for assessing construction enterprise green innovation efficiency and influencing factors: Evidence from China. *Environ. Technol. Innov.* **2023**, *32*, 103293. [CrossRef]

35. Xu, X.; Cui, X.; Zhang, Y.; Chen, X.; Li, W. Carbon neutrality and green technology innovation efficiency in Chinese textile industry. *J. Clean. Prod.* **2023**, *395*, 136453. [CrossRef]

36. Xiao, H.; Wang, D.; Qi, Y.; Shao, S.; Zhou, Y.; Shan, Y. The governance-production nexus of eco-efficiency in Chinese resource-based cities: A two-stage network DEA approach. *Energy Econ.* **2021**, *101*, 105408. [CrossRef]

37. Xu, Y.; Liu, S.; Wang, J. Impact of environmental regulation intensity on green innovation efficiency in the Yellow River Basin, China. *J. Clean. Prod.* **2022**, *373*, 133789. [CrossRef]

38. Wang, G.; Hou, Y.; Du, S.; Shen, C. Do pilot free trade zones promote green innovation efficiency in enterprises?—Evidence from listed companies in China. *Heliyon* **2023**, *9*, E21079. [CrossRef]

39. Tian, Z.; Wang, R.; Xiao, Q.; Ren, F. Research on Green Technology Innovation Efficiency of Advanced Manufacturing Industry in the Yangtze River Delta Region. *J. Anhui Normal Univ.* **2021**, *49*, 137–147.

40. Zhang, H.; Mao, J.; Wang, T.Y. Key Influencing Factors of Technological Innovation Efficiency Improvement in China's Smart Manufacturing Enterprises: An Analysis Based on the Three-Stage DEA-Tobit Model. *Sci. Technol. Manag. Res.* **2023**, *43*, 95–101.

41. Ma, W.B.; Zhu, H. Measurement of Innovation Efficiency and Influencing Factors of Green and Low-carbon Enterprises: Based on the Three-Stage DEA and Tobit Model. *Soft Sci.* **2024**, 1–12.

42. Wang, H.; He, X.Y. Industrial Agglomeration, Environmental Regulation, and Green Innovation Efficiency. *Stat. Decis.* **2022**, *38*, 184–188.

43. Kang, S.J. Measurement of Innovation Value Chain Efficiency in High-tech Industries from the Perspective of Industry Heterogeneity: An Empirical Analysis Based on the Three-stage DEA Model with SFA Correction. *Sci. Technol. Manag. Res.* **2017**, *37*, 7–12.

44. Wu, J.R. Evaluation of Technological Innovation Capability of China's Manufacturing Industry from the Perspective of Innovation-Driven Development. *J. Ind. Technol. Econ.* **2020**, *39*, 74–80.

45. Tao, X.; Dai, Q. Research on the Evaluation of Green Technology Innovation Capability in Anhui Province. *J. Bengbu Coll.* **2018**, *7*, 64–68.

46. Tian, Z.; Fang, Q.; Ju, Y.; Ren, Y. Evaluation of Green Transformation Efficiency and Influencing Factors of Manufacturing Industries in China's Three Major River Basins. *Resour. Environ. Yangtze Basin* **2023**, *32*, 2072–2084.

47. Lv, H.Y.; Qiao, P.H. Regional Differences and Driving Factors of Green Innovation Efficiency in China's Industrial Sector. *East China Econ. Manag.* **2020**, *34*, 28–36.

48. Li, D.Q.; Zhong, C.L.; Hu, J.W. Environmental Regulation, Government Support, and Green Technology Innovation Efficiency: An Empirical Study Based on Large-scale Industrial Enterprises from 2009 to 2017. *J. Jianghan Univ. (Soc. Sci. Ed.)* **2020**, *37*, 38–49+125.

49. Wang, Q.L.; Wang, Z.Y. Calculation of Green Technology Innovation Efficiency in China's Manufacturing Industry Based on DEA-Malmquist Method. *Sci. Technol. Sq.* **2023**, *2*, 65–78.

50. Watkins, A.; Papaioannou, T.; Mugwagwa, J.; Kale, J. National innovation systems and the intermediary role of industry associations in building institutional capacities for innovation in developing countries: A critical review of the literature. *Res. Policy.* **2015**, *44*, 1407–1418. [CrossRef]

51. Hong, J.; Feng, B.; Wu, Y.R.; Wang, L.B. Do government grants promote innovation efficiency in China's high-tech industries? *Technovation* **2016**, *57–58*, 4–13. [CrossRef]

52. Liu, P.Z.; Huang, T.; Shao, Y.T.; Jia, B. Environmental regulation, technology density, and green technology innovation efficiency. *Heliyon* **2024**, *10*, e23809. [CrossRef] [PubMed]

53. Ji, Y.; Dou, J. Study on stage impacts of factor price distortion on chinese technology innovation based on data mining. *Comput. Theor. Nanosci.* **2016**, *13*, 10504–10513. [CrossRef]

54. Ozkan, O.; Sharif, A.; Mey, L.S.; Tiwari, S. The dynamic role of green technological innovation, financial development and trade openness on urban environmental degradation in China: Fresh insights from carbon efficiency. *Urban Clim.* **2023**, *52*, 101679. [CrossRef]

55. Chen, Z.W.; Liang, M. How do external and internal factors drive green innovation practices under the influence of big data analytics capability: Evidence from China. *J. Clean. Prod.* **2023**, *404*, 136862. [CrossRef]

56. Qinqin, W.; Qalati, S.A.; Hussain, R.Y.; Irshad, H.; Tajeddini, K.; Siddique, F.; Gamage, T.C. The effects of enterprises' attention to digital economy on innovation and cost control: Evidence from A-stock market of China. *J. Innov. Knowl.* **2023**, *8*, 100414. [CrossRef]

57. Wang, N.N.; Cui, D.F.; Dong, Y. Study on the impact of business environment on private enterprises' technological innovation from the perspective of transaction cost. *Innov. Green Dev.* **2023**, *2*, 100034. [CrossRef]

58. Yuan, B.L.; Cao, X.Y. Do corporate social responsibility practices contribute to green innovation? The mediating role of green dynamic capability. *Technol. Soc.* **2022**, *68*, 101868. [CrossRef]

59. Khan, S.A.; Sheikh, A.A.; Ahmad, Z. Developing the interconnection between green employee behavior, tax avoidance, green capability, and sustainable performance of SMEs through corporate social responsibility. *J. Clean. Prod.* **2023**, *419*, 138236. [CrossRef]

60. Hong, J.; Li, J.F.; Li, X.F. Analysis of Technological Innovation Efficiency and Influencing Factors in China's Pharmaceutical Manufacturing Industry from the Perspective of Two Stage Innovation Value Chain. *J. Northwestern Polytech. Univ. (Soc. Sci. Ed.)* **2013**, *33*, 51–56.

61. Wenbo, L. Comprehensive evaluation research on circular economic performance of eco-industrial parks. *Energy Procedia* **2011**, *5*, 1682–1688. [CrossRef]

62. Kang, P.H.; Ru, S.F. Bilateral effects of environmental regulation on green innovation. *China Popul. Resour. Environ.* **2020**, *30*, 93–104.

63. Yalabik, B.; Fairchild, R.J. Customer, Regulatory, and Competitive Pressure as Drivers of Environmental Innovation. *Int. J. Prod. Econ.* **2011**, *131*, 519–527. [CrossRef]

64. Gong, X.S.; Li, M.J.; Zhang, H.Z. Does OFDI improve the efficiency of industrial green innovation in China: An empirical study based on the effect of agglomeration economy. *J. Int. Trade* **2017**, *11*, 127–137.

65. Kuang, C.E.; Wen, Z.Z.; Peng, W.B. The threshold effect of shadow economy on green innovation efficiency. *Econ. Geogr.* **2019**, *39*, 184–193.

66. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; IEEE: Piscataway, NJ, USA, 1995; pp. 278–282.

67. Breiman, L.; Cutler, R.A. Random Forests Machine Learning. *J. Clin. Microbiol.* **2001**, *2*, 199–228.

68. Mellor, A.; Boukir, S.; Haywood, A.; Jones, S. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *J. Photogramm. Remote Sens.* **2015**, *15*, 155–168. [CrossRef]

69. Martinez-Sanchez, L.; See, L.; Yordanov, M.; Verhegghen, A.; Elvekjaer, N.; Muraro, D.; d'Andrimont, R.; Van der Velde, M. Automatic classification of land cover from LUCAS in-situ landscape photos using semantic segmentation and a Random Forest model. *Environ. Model. Softw.* **2024**, *172*, 105931. [CrossRef]

70. Zhang, H.; Ren, X.Z.; Chen, S.K.; Xie, G.Q.; Hu, Y.S.; Gao, D.D.; Tian, X.G.; Xiao, J.; Wang, H.Y. Deep optimization of water quality index and positive matrix factorization models for water quality evaluation and pollution source apportionment using a random forest model. *Environ. Pollut.* **2024**, *347*, 123771. [CrossRef] [PubMed]

71. Cao, L.C.; Wei, S.Q.; Yin, Z.Y.; Chen, F.; Ba, Y.; Weng, Q.; Zhang, J.H.; Zhang, H.Z. Identifying important microbial biomarkers for the diagnosis of colon cancer using a random forest approach. *Heliyon* **2024**, *10*, e24713. [CrossRef]

72. Palmal, S.; Arya, N.; Saha, S.; Tripathy, S. Integrative prognostic modeling for breast cancer: Unveiling optimal multimodal combinations using graph convolutional networks and calibrated random forest. *Appl. Soft Comput.* **2024**, *154*, 11379. [CrossRef]

73. Alladio, E.; Trapani, F.; Castellino, L.; Massano, M.; Di Corcia, D.; Salomone, A.; Berrino, E.; Ponzone, R.; Marchiò, C.; Sapino, A.; et al. Enhancing breast cancer screening with urinary biomarkers and Random Forest supervised classification: A comprehensive investigation. *J. Pharm. Biomed. Anal.* **2024**, *244*, 116113. [CrossRef] [PubMed]

74. Larivière, B.; Van den Poel, D. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Syst. Appl.* **2005**, *29*, 472–484. [CrossRef]

75. Xie, Y.Y.; Li, X.; Ngai, E.W.T.; Ying, W.Y. Customer churn prediction using improved balanced random forests. *Expert Syst. Appl.* **2009**, *36*, 5445–5449. [CrossRef]
76. Supsermpol, P.; Thajchayapong, S.; Chiadamrong, N. Predicting financial performance for listed companies in Thailand during the transition period: A class-based approach using logistic regression and random forest algorithm. *J. Open Innov. Technol. Mark. Complex.* **2023**, *9*, 100130. [CrossRef]
77. Meher, B.K.; Singh, M.; Birau, R.; Anand, A. Forecasting stock prices of fintech companies of India using random forest with high-frequency data. *J. Open Innov. Technol. Mark. Complex.* **2024**, *10*, 100180. [CrossRef]
78. Xu, C.; Wang, J.; Zheng, T.L.; Cao, Y.; Ye, F. Prediction of prognosis and survival of patients with gastric cancer by a weighted improved random forest model: An application of machine learning in medicine. *Arch. Med. Sci.* **2022**, *18*, 1208–1220.
79. Álvarez-Diez, S.; Baixauli-Soler, J.S.; Lozano-Reina, G.; Rey, D.R. Subsidies for investing in energy efficiency measures: Applying a random forest model for unbalanced samples. *Appl. Energy* **2024**, *359*, 122725. [CrossRef]